RECOMMENDED PRACTICES FOR THE IMPLEMENTATION OF RENEWABLE ENERGY FORECASTING SOLUTIONS

- Part 1: FORECAST SOLUTION SELECTION PROCESS -

2. EDITION 2021

Draft for Review by Stakeholders prior to submission to the Executive Committee of the International Energy Agency Implementing Agreement

Prepared in 2021 as part of the IEA Wind Task 36, WP 2.1.

October 12, 2021

Contents

Preface

This 2^nd edition of this recommended practice document is the result of a collaborative work that has been edited and authored by the undersigning authors in alignment with many discussions at project meetings, workshops and personal communication with colleagues, stakeholders and other interested persons throughout the phase 1 (2016-2018) and phase 2 of the IEA Wind Task 36 (2018-2021) as part of workpackage 2.1.

The editors want to thank participants, contributors and supporter of the meetings, workshops and sessions that contributed in the discussions, provided feedback or other input throughout the past 6 years.

IEA Wind Task 36, October 12, 2021

Editors and Authors:
Dr. Corinna Möhrlen (WEPROG, Denmark) <com@weprog.com>
Dr. John Zack (UL AWS Truepower, USA) <john.zack@ul.com>

Contributing Authors:
Dr. Jeffrey Lerner (ENFOR, Denmark)
Mikkel Westenholz (ENFOR, Denmark)
Supported by:
Operating Agent Dr. Gregor Giebel (Danish Technical Univerity, DTU Wind, Denmark)

Chapter 1

Background and Objectives

1.1 BEFORE YOU START READING

This is the first part of a series of three recommended practices that deal with the development and operation of renewable energy forecasting solutions in the power market. The first part Forecast Solution Selection Process, which is the current document, deals with the selection and background information necessary to collect and evaluate when developing or renewing a forecasting solution for the power market. The second part Benchmarks and Trials, of the series offers recommendation on how to best conduct benchmarks and trials in order to test or evaluate different forecasting solutions against each other and the fit-for-purpose. The third part Forecast Evaluation, provides information and guidelines regarding effective evaluation of forecasts, forecast solutions and benchmarks and trials. If you already have experience in setting up a forecast solution and you have an up-to-date IT infrastructure, then it is recommended to go straight to part 2 or 3. However, if you are considering renewal of your IT infrastructure, require new forecasting products, need to extend or reduce the amount of vendors engaged or you are starting from scratch to build a forecasting solution, then the information in this recommended practices guideline will provide input to important considerations in this process. An overview of the decision support tool to help develop structured processes in the design and planning for a new or renewal of a forecasting solution can be found in chapter 3, while chapters 1 and 2 provide background information and initial considerations. It is recommended to use the table of contents actively to find the topics that are most relevant for you.

1.2 BACKGROUND

The effectiveness of forecasts in reducing the variability management costs of power generation from wind and solar plant is dependent upon both the accuracy of the forecasts and the ability to effectively use the forecast information in the grid management decision-making process. Therefore, there is considerable motivation for stakeholders to try to obtain high quality forecasts and effectively use this information as input to other operational processes or trading. This document is intended to provide guidance to stakeholders who are seeking a forecasting solution that fits their purpose and enables them to work efficient and economically responsible. In recent years, carrying out trials or benchmarks seemed to be an industry practice with an easy and straight forward decision process for many. In reality, trials are often expensive for both the end-user and the vendor, are not straightforward, nor entirely conclusive. Benchmarks have little value for commercial vendors, except in their start-up phase, and end-users can often not count on results that reflect state of the art. Further, if trials and benchmark studies lead to a dissatisfying result, forecasting solutions become increasingly criticized for their value. And, providers that may have had the most technically qualified solution at hand, but did not score best at a specific (maybe simplified) test, may be deselected. This recommended practices document will therefore focus on the key elements to consider when seeking to establish or renew a forecasting solution that fits ones purpose. In summary, this document provides recommendations and a decision support tool to establish procedures for an effective selection process.

1.3 OBJECTIVES

This document is intended to serve as guidance and standard for private industry, academics and government for the process of obtaining an optimal wind or solar power forecast solution for their applications and, in particular, it provides guidance to the design and requirements for effective renewable energy forecasting solutions. These guidelines and best practices are based on years of industry experience and intended to achieve maximum benefit and efficiency for all parties involved.

1.4 DEFINITIONS

In the discussion of the process of obtaining the best possible forecasting solution, there are a number of terms and concepts that are used. Several of the key terms and concepts are defined in the following. Note, these definitions are kept as general as possible with a focus on forecasting processes in the power industry and may not have such a completely general character to be applied to other areas of business.

- **Request for Information (RFI)**: a RFI serves the client to get information about the state-of-the-art business practices and available commercial products in the preparation or design of a forecast application or solution for a specific target process. By providing information about the target application, a client can ask vendors for their recommendations and experience to solve specific tasks. Such information is useful in the preparation and design of a new system, but also for systems that need to be rebuilt due to changing requirements.

- **Request for Proposal (RFP)**: a RFP is a tender process, where the client prepares a document laying out the system design of a forecasting solution and asking vendors to propose a solution and price quote. Usually, a set of minimum requirements are provided that become part of a contractual agreement for the awarded vendor. Renewable Energy Forecast Benchmark: an exercise conducted to determine the features and quality of a renewable energy forecast such as wind or solar power. The exercise is normally conducted by an institution or their agent and usually includes multiple participants from private industry forecast providers or applied research academics.

- **Renewable Energy Forecast Trial**: an exercise conducted to test the features and quality of a renewable energy forecast such as wind or solar power. This may include one or more participants and is normally conducted by a private company for commercial purposes. A trial is a subset of a Renewable Energy Forecast Benchmark.

Chapter 2

INITIAL CONSIDERATIONS

> ***Key Points***
> *This part provides guidelines for those whose task is to provide a plan and justification for a forecasting solution selection process. It intends to assist in finding the necessary information when navigating through the vast jungle of information, opinions and possibilities end ensures that crucial details are being considered.*

2.1 TACKLING THE TASK OF ENGAGING A FORECASTER FOR THE FIRST TIME

The most important considerations and first question to answer, when starting out to plan a forecasting solution is to be clear about the desired outcome. A lot of time and resources can get wasted for all involved parties on trials and benchmarks that are not aligned with requirements, also when planned and conducted by personnel with little or no experience in the subject. To avoid this, the recommended practice is to carry out a market analysis in the form of a request for information (RFI) and to establish a requirement list (see also APPENDIX B). In some cases, it can be beneficial to test vendors or solutions prior to implementation. The difficulty with this method lies in the evaluation of tests, especially, when they are short in time. In many cases they do not answer the questions an end-user needs answered, because such tests mostly are simplified in comparison to the real-time application and, but still require significant resources. For such cases, this guideline provides other methods for an evaluation of different forecast solutions/vendors. The pitfalls and challenges with trials and/or benchmarks are the topic of part 2 of this series of recommended practices. The following table shall summarise some of the aspects and help the decision process as to where and when such pilot projects, trials or benchmarks may not the the best choice when designing and choosing a forecast solution. The column recommendation in Table provides other methodologies that may be used to evaluate a forecast solution. Additionally, a typical set of questions to be asked to service providers will be provided in APPENDIX A**??**.

Table 2.1: Recommendations for initial considerations prior to forecast solution selection for typical end-user scenarios

Scenario	Limitation	Recommendation
Finding best service provider for a large portfolio (> 1000MW) distributed over a large area	Test of entire portfolio is expensive for client and service provider in terms of time and resources. Simplifying test limits reliability of result for entire portfolio.	RFI and RFP, where service providers methods are evaluated and incentive scheme on the contract terms provides more security on performance.
Medium sized Portfolio (500MW< X < 1000MW) over limited area	Test of entire portfolio is expensive for client and service provider in terms of time and resources. Simplifying tests limits reliability of result for entire portfolio.	RFP, where service providers methods are evaluated. Building of a system that enables change of service provider and incentive scheme may be more efficient than a test in the long run. (More detail on incentive schemes are found in section 3.9.3.3 and Part 3 of this guideline [**ieawindtask36RP2019**]).
Finding best service provider for small sized portfolio (< 500MW)	Test of portfolio requires significant staff resources, a budget and a minimum of 6 months.	Difficult to achieve significance on target variable in comparison to required costs and expenses trial costs makes solution more expensive. Test is possible, but expensive. Cheaper to setup an incentive scheme and a system, where the suppliers may be exchanged relatively easily.
Micro portfolio (< 100MW) or single plants	Cost of a trial with many parties can easily be higher than the cost of 1 year of forecasting. Time for a trial can delay real-time experience by up to 1 year.	Evaluation of methodologies and setting up the internal system with an incentive scheme and ease of service provider exchange is more beneficial. (More detail on incentive schemes are found in section 3.9.3.3 and Part 3 of this guideline [**ieawindtask36RP2019**])

Scenario	**Limitation**	**Recommendation**
Sale of generation at power market	Best score difficult to define, as sale is dependent on market conditions and a statistical score like RMSE or MAE cannot reflect the best marketing strategy, considering the uncertainty of a forecast and the associated costs	Strategic choice of forecast provider and incentive scheme better than real-time test. Strategic choice may be: choice of vendor in comparison to others that use different, uncorrelated weather forecasts, uncorrelated weather-to-power model, unique forecast methodology, flexibility, expandable, etc. Incentive scheme ensures resources and incentive for continuous performance improvements (see section 3.9.3.3 , Part 3[**ieawindtask36RP2019**]).
Market share of service provider is high	Monopolies in the power market mean that forecast errors are correlated among generators.	This could lead to increased balancing costs. The forecast error might be low, but the costs for errors may be disproportionately high. Ask about the market share of a provider and do not choose one with a share > 30% as the only provider!
System operation in extreme events	Today, extreme (or rare) events are better forecasted,when considering weather uncertainty.	Statistical approaches relying solely on historic information may not be sufficient. A PoE50 (probability of exceedance of 50%) needs to have equally high probability in every time step above and below. The IEA Task 36 WP 3 has been dealing with uncertainty forecasting and provides recommendations for such situations. See section . Forecasting solution needs to be weather and time dependent, i.e. only physical methodologies (ensemble forecast systems) fulfill such tasks
Critical Ramp forecasts	Critical ramp forecasts are part of an extreme event analysis and require probabilistic methods with time dependency	Consider difference between a ramp forecast and a critical ramp as extreme event analysis that requires time + space dependent prob. methods such as ensemble forecasts. See references for uncertainty forecasts.

Scenario	Limitation	Recommendation
Blind forecasting, i.e. no measurement data available for the park or portfolio	Only useful for portfolios, where small errors are canceled out and indicative regarding performance. Without measurements, forecast accuracy will be non-representative of what accuracy can be achieved by training forecasts with historical data.	Evaluation can only be carried out for day-ahead or long-term forecasts, if measurements are collected throughout the trial. If you have a portfolio > 500MW, a blind test against a running contract can provide an inexpensive way to test the potential of a new provider. For single sites, the benefits of training are so large (>50% of error reduction at times) that blind forecasting is not recommended. It wastes resources for everybody without providing useful results.
Dynamic reserve	Deterministic forecasts cannot solve reserve requirements.	It is necessary to apply probabilistic methods for reserve calculation for intermittent resources such as wind and solar. See section **??**.

2.2 Purpose and Requirements of a Forecasting Solution

Once the limitations are defined, the next step is to define what objectives the project has. As outlined in Table, it poses very different forecasting strategies to the project, if the objective is e.g. system balance of renewables or selling generated electricity at the power market. In the first task, extremes must be considered and risks estimated; mean error scores are not that important. Large errors are most significant, as they could potentially lead to lack of available balancing power. In the second case, it is important to know the uncertainty of the forecast and use a forecast that is uncorrelated to others. The mean error of a forecast is important, but not a priority target, if the target e.g. is to use a forecast that generates low balancing costs. This is not always the same, because errors that lie within the forecast uncertainty are random. Such errors can only be reduced by strategic evaluations and decisions, not by methodology. If the objective is to calculate dynamic reserve requirements, probabilistic forecasts are required and should be part of the requirement list. When choosing a forecast solution, understanding the underlying requirements is key. It is not enough to ask for a specific forecast type without specifying the target objective. For this reason, defining the objective is most important. And, if there is no knowledge in the organisation regarding the techniques required to reach the objective, it is recommended to start with a RFI (see section 1.4) from different forecast providers and thereby gain an understanding and overview of the various existing solution and their capabilities.

2.3 Adding Uncertainty Forecasts to Forecasting Solutions

In any power system that is in the transition to carbon neutrality, wind and solar generating resources are part of the solution. In order to integrate larger amounts of these intermittent and variable energy resources, forecast uncertainty needs to be reflected and taken into account in grid related operational decision-making processes. The future of renewable energy systems cannot be economically operated without taking uncertainty into account.

In the world meteorological organization's (WMO) guidelines on ensemble prediction [**wmo2012**], the WMO warns about ignoring uncertainty in forecasts, if an end-user receives a deterministic forecast. The WMO argues that *if a forecaster issues a deterministic forecast the underlying uncertainty is still there, and the forecaster has to make a best guess at the likely outcome. Unless the forecaster fully understands the decision that the user is going to make based on the forecast, and the impact of different outcomes, the forecaster's best guess may not be well tuned to the real needs of the user*.

Weather related decision-making hence requires a deeper understanding of weather uncertainty, the way any weather service provider produces uncertainty of weather forecasts, and how such forecasts are to be translated into end-user applications. In [**bessa2017**], a thorough review of uncertainty forecasting techniques, methods and applications has been made. This review will be basis to the following definitions and recommendations for the selection of forecast solutions, where uncertainty forecasts are to be incorporated. There will be named gaps and pitfalls and how to best apply uncertainty forecasts in power system applications.

2.4 INFORMATION TABLE FOR SPECIFIC TASKS AND TARGETS

Table 2 lists a number of targets and points to the chapter or part of this guideline series, where the topic is described in detail. The table provides some typical targets and where to find information on how to achieve the best solution for that target.

Table 2.2: Information table for specific targets

Target	Information
How to find the best forecast solution	Section 3
Creating a requirements list	Section 3.3, 2.2, 3.2.1, and 3.2.2
Deterministic versus Probabilistic	Section 3.2.2 and **??**
Decision support tool and practical guide to forecasting	Figure 3.1

Target	**Information**
Evaluation of vendors: interviewing or conducting trial?	Section 3.9 and References in section 5
Do I need to test reliability and consistency?	Section 3.2.1 and 3.9.2.1
How do I know which forecast solution fits my purpose best ?	Section 2.2 and 3.1 , APPENDIX A
How do I build up sufficient IT infrastructure for a trial?	Part 2: Trial Execution
Which metrics for what purpose?	Part 3: Evaluation of forecasts
Step-by-step guide for trials and benchmarks	Part 2: Trial Execution

Chapter 3

Decision Support Tool

Practical usage of the Decision Support Tool: *The decision support tool in Figure 3.1 provides a high-level overview of the process for finding the most suitable forecast solution and vendor, respectively. The sections provide guidance in how to use the decision support tool with detailed descriptions and explanation to provide the low-level information for the detailed planning and design of the decision process.*
Notice for the practical usage of the Decision Support Tool: *To find the de-tailed recommendations, the numbers in the boxes of Figure 3.1 correspond to the head-lines in the following sections.*

From an end-user perspective, it is a non-trivial task to decide which path to follow, when implementing a forecasting solution for a specific application. Whether this is at a system operator, energy management company, a power producer or power trader, there are always multiple stakeholders involved in the decision-making process. A relatively straight forward way to decide for one path or another is to use a decision support tool.

Visualisation of the Decision Support Tool
Figure 3.1 shows a decision support tool aimed to high- level decisions of managers and non-technical staff when establishing a business case for a forecasting solution. Independent on the experience with forecasting solutions, the high-level thought construct shown in Figure 3.1 is targeted to assist in considering the required resources and involvement of departments and staff for the decision process. The decision tool is constructed to begin with initial considerations to establish a "Forecast System Plan". The tool aims to assist in taking a decision on the major dependencies to the planned item. There are cross references in the decision tool and referrals to a different decision streams, dependent on the answer at each step of the decision flow.

Starting at the very top, the first major dependency when planning a new or renewal of

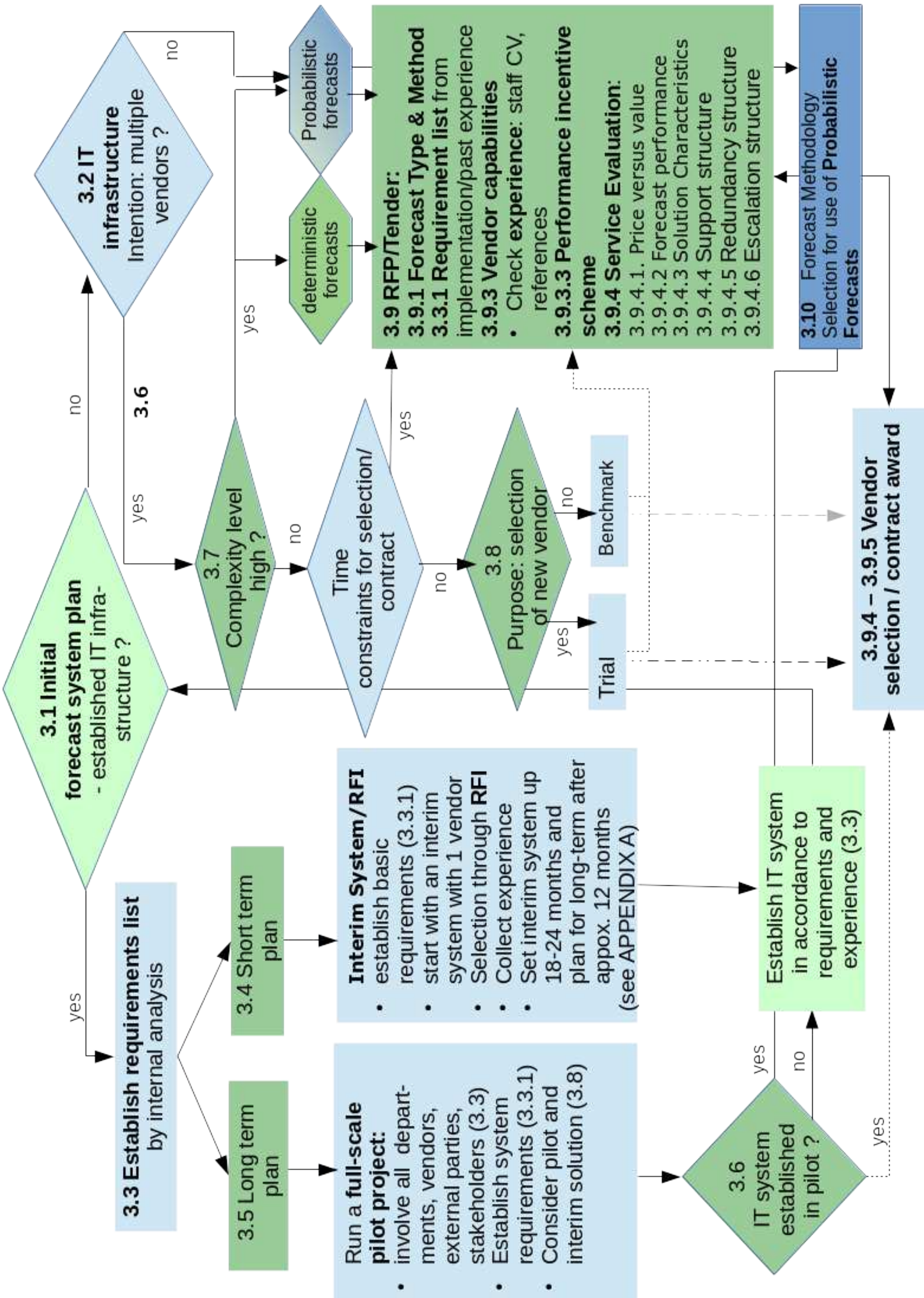


Figure 3.1: Decision Support Tool

a forecasting system is the IT infrastructure. Dependent on the status of IT infrastructure, the recommended procedure splits up here and follows in different paths. This is not to be understood that the IT infrastructure has higher priority over the forecasting solution itself. It is rather to sharpen the awareness that if the IT infrastructure is not in place yet or needs renewal for a new technology to be implemented, the IT needs to be part of the decision process from the very beginning.

The decision support tool in Figure 3.1 provides a high-level overview of the process for finding the most suitable forecast solution and vendor, respectively. The following sections provide guidance in how to use the decision support tool with detailed descriptions and explanation to provide the low-level information for the detailed planning and design of the decision process.

3.1 INITIAL FORECAST SYSTEM PLANNING

The planning of a forecasting system for renewables is a complex task and highly individual. This guideline therefore focuses solely on aspects that are of general planning and management tasks specific to the implementation of wind power or solar power forecasts into an operational environment. Note that the limited information and considerations about forecast technologies or methodologies has the objective to provide guidelines on the impacts of commonly implemented technologies in the implementation and decision process. On the other hand, there is strong focus on the IT infrastructure as one of the most crucial tasks in the implementation and integration of forecast solutions that are prone to become limiting factors for changes at later stages. For that reason, it is recommended that the IT infrastructure is established or, if already available, evaluated together with the planning of the forecast solution and methodology, in accordance to its ability to develop along with changes in forecast practices, possible statutory changes, etc. Databases are prone to have limitations that prevent changes to incorporate more information or store information in a different way. Such consideration need to take place prior to and should be part of the decision process and the requirement list (see section 3.3).

3.2 IT INFRASTRUCTURE CONSIDERATIONS

The starting point of the tool is the IT infrastructure. If a company has already built an appropriate infrastructure, finding a solution or vendor is more straight forward. The reason for this is that in this case, the forecast provider will need to conform to file formats, communication protocols or security constraints, for example. If an IT infrastructure for the forecasting solution is to be established or renewed it needs to be closely following the technical requirements of the solution. In the other case, i.e. no IT infrastructure has been built yet, an internal analysis of the needs are required. In this analysis, it is important to know, whether there is a short-term goal with an objective to be reached with time constraints, or whether it is a long-term plan that needs to be satisfied. Usually such differentiation is dependent on the political landscape and adopted policies for the development of renewables in the country. The important aspects in the IT infrastructure to be considered are:

- database structure
- communication layer
- monitoring and error handling
- data storage and historic data accessibility

In general a forecast system interface, whether in-house or outsourced requires multiple data streams, starting from measured power and weather variables. Usually, there is a connection to the power units SCADA (Supervisory control and data acquisition) system. However, the measurement data needs storage and a data flow to the forecaster needs to be added as one more of the various internal data flow processes.

It needs to be decided, whether there is a need to access other external data sources, such as NWP data, or the forecast data itself.

Dependent on the setup of the forecasting solution, it is necessary to evaluate how fast accessible historic data has to be, for example to carry out internal analysis, external data delivery to vendors, etc.

3.2.1 IT requirements for single versus multiple forecast vendors

Impacts on multiple vendor solution:

- infrastructure more complex
- database requirements are higher due to higher data volumes
- Strategy required for forecast: mixing versus primary/secondary forecast

IT infrastructure Impacts for single vendor solution:

- reliability requirement of solution high
- monitoring requirement higher for up-time
- higher requirements for quality control of forecasts
- less data volume than for multiple-vendor solutions
- database structure less complex than for multiple-vendor solutions

3.2.2 IT requirements for deterministic versus probabilistic forecasts

From an IT infrastructure and architectural perspective, deterministic and probabilistic forecasting solutions are quite different. The database requirements are by a factor of 10 to 100 higher for the latter. Dependent on the way the probabilistic forecasts are used, they add significant amounts to the storage requirements. Nevertheless, storage and computational resources are changing with changing requirements in industry and hence should not per se be considered a barrier or limitation for the integration or implementation of new technologies. But, they need consideration and careful planning. The advantages and disadvantages of the deterministic versus the probabilistic solution from a IT perspective are similar to single versus multiple providers in section 3.2.1 .

3.3 ESTABLISHMENT OF REQUIREMENT LIST

Establishing a requirement list for a forecasting solution is highly individual and depends on many factors. Every end-user will have very specific needs to fulfill. There are however common areas that require consideration. This is how the following recommendation list has to be interpreted.

Two of the fundamental aspects when establishing a requirements list are:

1. Description of the current situation
   In this process, it is imperative to describe exactly all processes, where forecasting is required and how these processes are interlinked. Here it is essential to get the different departments involved, also the IT department. The more accurate you can describe the situation at hand, (e.g. integration plans, use of forecasts, market situation, statutory aspects, IT restrictions, limitations and methods for data exchange exist, current or future challenges, etc.), the more straight forward it will be to (1) ask questions to the vendors regarding forecasting methodology, but also (2) get clarity of the involved processes enabling forecasting.

2. Engage vendors, stakeholders and independent consultants
   Questions to vendors should be of technical character regarding forecast methodology, but also on available data exchange methodologies, required input data for the models and system support. If you already have a forecast vendor, it is recommended to engage with the forecaster to discuss the current situation and where the forecaster sees limitations and potential for improvements. Often, forecast providers need to adopt their forecasts to a specific need and even though a new technology may be available, it is not used due to current limitations. Other vendors, stakeholders and independent consultants may at any stage be engaged, not only when it comes to establishing a new or renewal of a forecasting system. For new systems, it is recommended to engage different forecast vendors and stakeholders to provide insight from a variety of experiences. In all cases, it is essential to describe the planned objective and name

limitations, if they are already known. The more information that can be shared the better a vendor, stakeholder or consultant can evaluate what is considered the most appropriate solution.

Appendix A contains an additional listing of recommended considerations that are applicable also for RFIs.

3. Description of the envisaged Situation
The description of the envisaged situation is most important for the implementation of a solution. Analysis of the current situation, the forecast vendor(s) input and other organizational and statutory requirements should lay the basis for an envisaged new system. It is recommended to put as much detail into this part as possible. The following requirement list assists in defining all aspects for the planning phase of a forecasting system. Recommendation in short: Describe (1) the current situation, (2) engage vendors and stakeholders and (3) describe the envisaged situation in great detail. Ask specific questions that are required to get the highest possible level of detail for the decision process.

> ***Recommendation in short****: Describe (1) the current situation, (2) engage vendors and stakeholders and (3) describe the envisaged situation in great detail. Ask specific questions that are required to get the highest possible level of detail for the decision process.*

3.3.1 Requirement List

The following areas are recommended to be considered in the list:

IT infrastructure:

- communication/data exchange with the forecast vendor(s)
- communication/data exchange with the assets (wind/solar parks)
- database and storage implications
- accessibility of data information of internal users
- application interfaces to internal tools (e.g. graphics, models, verification, metering)
- information security policies

Forecast Methodology and Attributes:

- Weather input
- Methodology of weather to power model
- Application/model background for each forecast product
- Forecast time horizons
- Forecast frequency
- Forecast uncertainty

Support and Service:

- service level for each product (e.g. 24/7, business hours etc.)
- system recovery
- failure notifications and reporting
- escalation procedures
- service documentation
- contact list for different services
- staff training

Contracting:

- contract length
- amendment possibilities
- additional work outside contract
- licenses
- confidentiality (NDA)
- insurances
- sub-contracting
- Price table for each product category

Performance and Incentivization:

- verification methods
- verification parameter
- definition of payment structure (boolean or sliding areas)
- expected accuracy for each forecast horizon

3.4 SHORT-TERM SOLUTION

In this case, current requirements should be listed and analyzed in accordance with possible time limitations. It is recommended that a short-term solution is sought, if the political situation does not seem to be stable to make long-term investments, or a here-and-now issue needs to be solved and experience gained. In such cases, a relatively simple methodology that can be implemented fast and easy is the best way forward. Today, this can be found by carrying out a RFI, where vendors can suggest how to best and easiest fulfill very specific needs. Due to IT constraints in many organizations, such solutions sometimes are set up with delivery by Email. This is not a recommended practice for security and reliability reasons, but can help to fill a gap between a long-term solution and an urgent need.

Despite the shortcomings, interim solutions are recommended as they are valuable in respect to experience with forecasting data and its handling inside the organization. If such solutions are employed while a long-term plan is being developed, it can be of great benefit for the long-term solution. Such solutions should last approx. 18-24 months. Planning for a long-term solution should ideally start after 12 months.

The danger lies in staying with an interim solution, if it has real limitations on security (e.g. email delivery) and reliability, as such limitations may not be problematic for a long time, but reliance on non-redundant systems can cause sudden uncontrollable situations. For this reason, we posted the question about the IT system at the end of the short-term solution, as this is a crucial part in the next step. We recommend that this is taken as a priority topic, once practical experience with forecasting has been gained.

3.5 LONG-TERM SOLUTION

Developing a long-term solution can be cumbersome and difficult, as many aspects have to be considered, from policies to governmental plans or corporate strategies. A practical way forward is to conduct a full-scale pilot project, where different solutions are tested and verified over a period of at least 1 year. The advantage of such a pilot project is that there is the possibility to verify and evaluate different solutions and their fit for purpose over a longer time span.
Moreover, a pilot project is characterised by:

1. Involvement of all relevant departments (internal and external parties/stakeholders)
2. Establish system requirements
3. Pilot maybe used as interim solution

The disadvantage is that it takes a long time and hence is costly and it is not given that there is a very clear winning solution to a specific area or task. On the other hand, to find the most appropriate long-term solution needs many considerations, not only technically, but also economically and whether a solution is future compatible. So, the experience of the vendor in adjusting, maintaining and developing a solution with changing needs may be a challenge for some and the business philosophy for others. Such vendor policies can be identified and clarified when carrying out long-term tests. The box therefore feeds into the question about an appropriate IT system. If this has not been established, it is recommended to prioritise the IT before going further.

The end of a pilot project has therefore 3 further paths:

1. vendor selection
2. redefining requirements to start a solution bottom up
3. carrying out a RFP with the identified requirements.

3.6 GOING FORWARD WITH AN ESTABLISHED IT SYSTEM

In the case an IT system has been established and new vendors or a renewal of the system is the objective for the project, there are various possibilities to move forward. Crucial in this phase is again to set target and objectives. If the target is to find out, whether there exist forecast vendors on the market that may provide forecasts with other methods or for a lower price, it may be a good way forward to carry out a trial or benchmark. Dependent on the structure of the system, or complexity of the system and time constraints, a benchmark/trial or a RFP as alternative are recommended. One crucial criterion when deciding on the two alternatives RFP or trial/benchmark in existing IT environments is whether the IT structure can handle multiple suppliers.
If this is not the case, any evaluation against an existing supplier can be cumbersome and at times impossible. The recommended practices guideline part 2 is going into detail with this topic, which is mostly related to:

- **representative** (including consistency)
- **significant** (including repeatable)
- **relevant** (including fair and transparent)

These are the key points when carrying out a comparison.

3.7 COMPLEXITY LEVEL OF THE EXISTING IT SOLUTION

Apart from accuracy or statistical skills of forecasts, there are also other aspects to be considered when choosing a forecast supplier. It has been observed that such evaluations based on non-technical skills or skills leading to forecast performance for a specific purpose have been underestimated in their importance. One aspect is the ability to improve, which is fully excluded with a trial/benchmark as sole decision-making criterion (besides price) as capability of vendors. It is often forgotten that long-term experience in a specific area can provide significant advantages. On the other hand, verifying only a small part of a complex system for practical reasons may result in a misleading result (see 3.6 representative, significant and relevant).

The complexity of a system that a forecast solution must adapt to, but also the data flow that complex systems inherit, is seldom easy to simulate in trials and will always disqualify some participants, when it comes to the real system. To conclude, the complexity of a system and the purpose of a forecast within a complex corporate structure are significant aspects to consider in a forecast solution selection.

> ***Recommendation:*** *The path to follow in case of complex structures and requirements are hence best performed by a RFP process, where core capabilities should be evaluated, when choosing a forecasting solution.*

3.8 SELECTION OF A NEW VENDOR VERSUS BENCHMARKING EXISTING VENDOR

If there are no time constraints and the complexity level of the running system is not too high, or a new system is in the process of being built, a trial or a benchmark exercise can be very useful in order to gain some experience in the building process.

> ***Recommendation:*** *Conduct a trial in case a new vendor has to be selected and a trial can be carried out in such a way that the results are fair, transparent, representative and significant. Carry out a benchmark, if the purpose is not from the outset to engage a new vendor, but also to compare the capabilities of a vendor with other vendors or against newer technology. In both cases the invited vendors need to be notified of the purpose of the exercise.*

3.9 RFP EVALUATION CRITERIA FOR A FORECAST SOLUTION

If complexity levels are high and if time constraints do not allow for a lengthy trial or benchmark, the RFP should be compiled with care in order to fulfill all requirements and yet not ask for more than needed. The most important evaluation criteria for a forecast solution to be defined in a RFP is:

- the type of forecast that is required (e.g., hours-, day-, or week-ahead)
- the methodology that is applied to generate these forecasts

- compliance to requirements

It is recommended that this first step should be vendor independent. And, if this cannot be defined, it is recommended to first conduct an RFI to scan the industry on their capabilities and their recommendation which type and methodology should be applied for the specific needs. Appendix B contains typical questions for an RFI. Only when the forecast type and methodology is defined, the vendor comes into play. The important factors to consider here are:

- capabilities (experience)
- support and maintenance services

The sections below describe these considerations in detail.

3.9.1 Forecast Solution Type

Most users will agree that they want to obtain forecasts with the best possible forecast accuracy for their application. A benchmark or a trial has in the past often been viewed as a way to determine which provider is most likely to deliver the best possible forecast performance. In theory, this is a reasonable objective. In practice, it is not recommended to rely solely on a test. The following subsections will address a number of key issues associated with the dilemma of finding the best forecasting solution with a simple and non-costly exercise for both the end-user and the forecast provider.

3.9.1.1 Forecast solution Type

Single versus multiple forecast providers It has been widely documented (e.g. Nielsen et al., 2007, Sanchez, 2008) that a composite of two or more state-of-the-art forecasts will often achieve better performance (accuracy) than any of the individual members of the composite over a statistically meaningful period of time. Indeed, many of the FSPs internally develop their approach and services on that basis. And, there are well founded reasons for an end-user to consider the use of multiple FSPs to achieve better forecast accuracy. However, in a practical sense, there are several advantages and disadvantages that should be considered. When building up a solution, it is recommended to consider the following aspects: Benefits of using multiple vendors:

1. There are a number of FSPs in todays forecast market that exhibit performance that is close to the state-of-the-art. It may be advantageous for reliability to assemble a set of state-of-the-art forecasts, unless they are highly correlated.

2. Higher forecast accuracy can often be achieved by blending forecasts from multiple uncorrelated1 FSPs.

3.9.1.2 Drawbacks of using multiple vendors

The benefits of having multiple vendors also contain inherent challenges for the end-user:

1. Increased internal costs, even if two cheap vendors may be less costly than one high-end forecast vendor, employing multiple vendors increases internal costs significantly due to increased amounts of data and IT processes.

2. Blending algorithms need to be intelligent. Multiple forecasts can be beneficial, but only, if the algorithm is intelligent to only blend/mix, if all forecasts are available and easy to retrain, if forecast statistics change. With two forecast vendors this is relatively easy. If there are more than two, it becomes more complex.

3. Forecast improvements are difficult to achieve with a multi-forecast provider solution. When improvements are achieved on the vendor side, the blending algorithm is becoming inconsistent and can result in worse scores than before, unless long-term historic data can be delivered. In other words, the handling and the improvement of forecasts are complex and difficult with multiple forecasts.

4. Multi-vendor Solutions cannot be incentivized as easily to achieve continuous performance increase over time. Although incentive schemes can be a good way to provide resources to the FSP for continuous improvements, in a multi-vendor environment, this can be counter productive, as changing statistical characteristics of forecasts can have a bad influence on the resulting blended forecast. Any end-user needs to be aware of this pitfall, when choosing a solution and take mitigating measures.

5. Multiple points of failure - with multiple forecast providers, the IT infrastructure needs to contain more logic to deal with one or more data streams when there are, for example, delivery disruptions, timeliness, or quality issues.

3.9.1.3 Deterministic versus Probabilistic

Many forecasting tasks need a discrete answer. For that reason, forecasting solutions have been mostly fed with deterministic forecasts in the past. Although weather forecasts and hence also power forecasts of intermittent resources such as wind and solar power, contain inherent uncertainties, probabilistic forecast products have been associated with forecasts not being discrete. The probability of a generic power generation at time x cannot be used in a trading application with the purpose to bid into the market. As penetration of variable generation resources increase and digitialization increases, the uncertainty information for decision taking can and is being processed by algorithms, also those whose output needs a discrete answer. Deterministic forecasts by default suppress the underlying uncertainty in the forecasts. By using probabilistic forecasts, this uncertainty can be taken into consideration in the decision processes.

The most common products of uncertainty or probabilistic forecasts are the probability of exceedance (PoE) values, typically given as PoE05, PoE50 and Poe95, quantiles, or percentiles or confidence bands (see Glossary for definitions).

The advantage of probabilistic/uncertainty forecasts in comparison to the deterministic best guesses is the possibility to act upon the probability of an event to occur, rather than being surprised, when the deterministic forecast is wrong. In power markets, for example, a probability of exceedance of 50% (PoE50) is an important parameter for a system operator, as such forecasts prevent the market to be able to speculate against system imbalance. Extreme ramping, high-speed shut-down risk, unit commitment and dynamic reserve allocation are other examples, where probabilistic forecasts are beneficial or required. In other words, wherever there are some kinds of uncertainty and extreme to be considered that may have impact on a decision or the costs of a process, probabilistic forecasts provide the necessary information to an end-user to take a decision upon some objective uncertainty criteria.

Recommendation: *When establishing or renewing a forecasting system, the question should not be posed on advantages and disadvantages for deterministic or probabilistic forecast solution, but rather whether a deterministic solution can fulfill the objective of the application. Section 3.10 describes uncertainty forecasts and how to select the appropriate probabilistic methodology for specific applications.*
A thorough academic review about probabilistic methodologies can be found in the References Material under Uncertainty Forecast Information in section5

3.9.1.4 Forecast horizons

The forecast horizons play a major role in the ability to plan using forecasts. Today, there are 5 types of forecast horizons applied in the power industry:

1. Minute-ahead forecasts or nowcasts (0-120min)
2. Hours-ahead forecasts (0-12 hours)
3. Day-ahead forecasts (0-48 hours)
4. Week-ahead forecasts (48-180 hours)
5. Seasonal forecasts (monthly or yearly)

The Minute-ahead forecasts are in literature also sometimes referred to as ultra-short term forecasts or nowcasts and are mainly used in areas with high penetration and high complexity in system operation or significant risk for high-speed shut down and extreme events. These forecasts are either based on a statistical extrapolation of measurements or weather input together with measurements generated on minute basis. The recommended practice depends on the severity and costs of the target value. For situational awareness, a simple extrapolation of measurements may be sufficient. For extreme events (e.g. ramps, high-speed shut down) the involvement of weather related forecasts in high time resolution is recommended.

Hours-ahead forecasts or sometimes referred to as short-term forecasts correct a day-ahead forecast by using real-time measurements and extrapolate from local real-time observations an improved view of the current state and the next few hours.

There are different methods available from simple extrapolation of measurements to advanced weather and distance- dependent algorithms. Its recommended to get details of a short-term forecast methodology described by the vendors, as quality and usability can differ strongly with availability of data, quality of measurement data etc.

If the target is e.g. ramp forecasting, system control, a very large fleet or quality issues with measurement data not dealt with by the end-user, simple algorithms are often not capable of providing a good enough picture of the next few hours.

The Day-ahead forecasts are widely-used forecasts for general system operation, trading and short-term planning. Traditionally, they are based on a combination of weather models and statistical models.

The Week-ahead forecasts, sometimes referred to as long-term forecasts, are usually applied in cases where the focus is not on forecast accuracy, but on forecast skill, e.g. in situations, where trends prevail over granularity. These forecasts are most valuable as a blending of a number of different forecasts or from an ensemble predication system, where the small-scale variability is reduced. If this is done, such forecasts can serve to reduce reserve costs and generate more dynamic reserve allocation as well as auctions. The Seasonal forecasts sometimes referred to as ultra-long-term forecasts, predict variations due to seasonal and or climate variability. They may be derived based on climatology, correlation to various climate indices and oscillatory phenomena, climate models, or a combination of these methods. Ensemble methodologies are the most preferable method due to the inherent uncertainty on such time frames. The most simple method is to analyze past measurements.

> ***Recommendation:*** *Key when choosing a methodology is to carefully analyze the accuracy requirements of the task to solve. For trading of futures in a trading environment a simple methodology may be sufficient. Tasks such as grid balancing, grid infrastructure planning or long-term capacity planning however require more advanced methodologies. It is recommended to choose the method according to the need to capture quantities only (simple method) or capture also climatic extremes (advanced method).*

3.9.2 Vendor Capabilities

3.9.2.1 Experience and Reliability

Experience is a key element of a successful vendor and implementation of the forecasting solution. It can usually be evaluated by the selected references that are provided and measured by conducting interviews with customers of similar type or by asking for information about the vendors background and experience with similar customers. If a vendor is new to the market that may not be possible. In this case, staff resources and experience of the key staff is usually indicating, whether the experience level for the minimum requirements is given. Reliability is also connected to experience, as it implies the reliable implementation and real-time operation of a forecasting service. It is an important aspect and may be derived by requiring examples of similar projects and interviewing references. It can also save a lot of work and resources in comparison to carrying out a trial, if reliability and experience with respect to e.g. complex IT infrastructure, security aspects, reliable delivery and provision of support etc. are a more crucial aspect than specific statistical performance scores.

> ***Recommendation:*** *Ask vendors to describe their experience and provide references and CV of key staff members.*

3.9.2.2 Ability to maintain state-of-the-art performance

The previous section provided an overview of all of the considerations for the technical aspects of forecast type and methodology. In order to assure that the forecast vendor can maintain state-of-the-art performance it is recommended to verify, whether the provider engages in ongoing method refinement/development and forecast improvement activities. Recommendation: Evaluate by asking the vendor to provide information about

- research areas and engagement
- references to staff publications of e.g. their methodology, project reports
- references of participation in conferences/workshops
- percent of revenue reinvested into research and development

3.9.2.3 Performance incentive Schemes

A performance incentive scheme is the most effective way to ensure that a forecaster has an incentive to improve forecasts over time and also allocates resources to it. By setting up a performance incentive scheme, the client acknowledges that development requires resources and vendors have not only an economic incentive to allocate resources for further developments, but can also influence their reputation. Incentive schemes do not have to be enormously high, but usually range between 10-30

Establishing a performance scheme What must be key to a performance incentive scheme is that it reflects the importance of the forecast parameters that are incentivized for the client! The evaluation of such forecast parameters should be selected according to:

1. the objective of the forecasting solution

2. the use/application of the forecasts
3. the available input at forecast generation time

The objective (1) in this context is defined as the purpose of the forecast. For example, if a forecast is used for system balance, an evaluation should contain a number of statistical metrics and ensure that there is an understanding of the error sources that the forecaster can improve on. A typical pitfall is to measure performance only with one standard metric, rather than a framework of metrics reflecting the cost or loss of a forecast solution. For example, if a mean absolute error (MAE) is chosen to evaluate the performance in system balance, an asymmetry in price for forecast errors will not be taken into account. Also, if e.g. large errors pose exponentially increasing costs, an average metric is unsuitable.

The use or application of forecasts (2) is defined in the context of where forecasts are used in the organization and where these have impact and influence on internal performance metrics or economic measures. For example, a wind power forecast that a trader uses for trading the generation of a wind farm on a power market has two components: revenue and imbalance costs. The revenue is defined by the market price for each time interval, whereas the cost is defined by the error of the forecast, the individual decision that may have been added to the forecast and the system balance price. When evaluating a forecast in its application context, it is important to choose an evaluation that incentivizes the vendor to tune the forecast to the application. A forecast that is optimized to avoid large errors may create lower revenue. However, if income is evaluated rather than revenue, such a forecast may be superior due to lower imbalance costs. On the other hand, if the end-user makes changes to the forecast along the process chain, the forecast evaluation must stop, where it is outside the forecast vendors influence.

The available input at forecast generation time (3) is most important when evaluating short-term forecasts that use real-time measurements. For example, if the forecast is evaluated against a persistence forecast with corrected measurements rather than with the measurements that were available at the time of forecast generation, the evaluation is to the disadvantage of the forecaster. The same applies, if aspects that affect the forecast such as curtailments, dispatch instructions, turbine availability, are not taken out of the evaluation or are corrected.

Recommendation: *When incentivizing a forecast solution with a performance incentive, the evaluation need to consider the non-technical constraints in the forecast and the parts that a forecaster does not have influence upon. A fair performance incentive scheme needs to measures the performance of a forecast by blacklisting any measurement data that is incorrect or corrupt, that contains curtailments, dispatch instructions, reduced availability or other reductions outside of the forecasters influence. Evaluation against persistence forecasts also need to be done with the available data at the time of forecast generation to not give advantage to persistence.*

Additionally, single standard statistical metric (e.g. MAE or RMSE) alone cannot be recommended. More details on the purpose and interconnection of statistical metrics for evaluation of incentive schemes can be found in part 3 of this recommended practice and in the references under Evaluation and Metrics. Structure of a performance incentive payment The structure of performance incentive scheme is an individual process and contractual matter between parties. When establishing the structure of a performance incentive it is recommended to consider that by choosing a maximum and minimum, the maximum value provides budget security to the end-user, also when e.g. changing from a very simple solution to an advanced one with much higher performance. The latter provides security to the forecaster to ensure that the basic costs for generation of forecasts are covered. Adding a sliding structure in between ensures the forecaster always has an incentive to improve, also when it is foreseeable that the maximum may not be achievable.

> ***Recommendation:*** *it is recommended to apply a maximum incentive payment and a maximum penalty or minimum incentive. A sliding change is preferable over for a boolean (yes|no) decision for incentive payments, as it always encourages forecast improvement efforts.*

3.9.3 Evaluation of services

The recommended practice in any evaluation is to consider a number of factors that contribute to the value that a user will obtain from a forecast service. It is not possible to provide a complete list of factors to consider. However, the most important factors that should be addressed are the following elements:

- Price versus value and quality
- Forecast Performance
- Solution Characteristics
- Speed of delivery
- Support structure
- Redundancy structure

The issues associated with each of these aspects will be addressed in the following subsections in more detail.

3.9.3.1 Price versus Value and Quality

The value of a forecast may or may not be directly measurable. In most cases however, the value can be defined for example in terms of cost savings or obligations and in that way provide an indication of the expected value from a certain solution. Prices are difficult to evaluate. A low price often indicates that not all requirements may be fulfilled in operation or not all contractual items are accepted and left to the negotiations. For these reasons, care has to be taken in the evaluation process. Some services and methods are more expensive than others on e.g. computational efforts, required licenses, database requirements, reliability, etc. Unless prices are driven by competition in a overheated market, a service price is normally coupled to the requirements and acceptance of contractual items. Some items such as reliability, customer support or system recovery can have high prices, but can always be negotiated to a different level. In an RFP end-users need to be aware of the relation between cost, value and associated service level to prevent vendors from speculation on negotiable item in the requirement list.

Recommendation: *Following a decade of experience in the forecasting industry, the recommended practice on price evaluation is to connect technical and contractual aspects to the price and consider to let vendors detail contractual aspects that may be associate with high service costs separately, especially, if a fixed cost price is requested. An example could be the requirement of full system recovery within 2 hours in a 24/7/365 environment. If there is no penalty associated, a vendor may ignore this requirement, which may result in a much lower price. Requesting transparent pricing eases evaluation and makes sure that speculations regarding negotiable aspects of a service can be clearly compared.*

3.9.3.2 Forecast Performance

Forecast performance evaluation should contain a number of metrics that are representative for the need to the forecast user. It is recommended to establish an evaluation framework for the performance evaluation. How to establish such a framework is dealt with in Part 3 of this recommended practice guideline.

3.9.3.3 Solution Characteristics

The solution characteristics of a forecast service also contains much value for an end-user and should get attention in the evaluation. It can be defined in terms of the available graphical tools, ease of IT services for retrieving data or exchanging data in real-time as well as historical data, customer support setup and staff resources connected to the forecasting solution. This can be key for the operational staff to accept and be comfortable with a forecast service as well as having confidence in the service. Additional work that may be connected, but outside the scope of the operational service can also be key elements for a well functioning service. Recommendation: Ask the vendor to describe how the system will be built up, how communication and support is envisaged and let them provide examples of graphics (if applicable).

3.9.3.4 Support Structure

Customer service is often under-estimated and in most cases second to an accuracy metric when selecting a vendor. Support can be a costly oversight if, for example, costs are related to a continuously running system or extreme events, where the user needs an effective warning system and related customer service. Support can have a relatively large cost in a service contract and may provide a false impression on service prices, if, for example support is only offered at business hours.

Key elements for the customer support is:

- the responsiveness of the provider, when issues arise
- live support in critical situations

A support structure and its management for operational processes additionally need to bind the following strategic areas together:

1. Customer Support
2. Operations Software and Service
3. IT Infrastructure

The customer support (a) should be handled by a support platform, ideally with different forms for contact, e.g. telephone hotline and email ticket system.

Any end-user needs to ensure that third-party software used in the operational environment (b) is licensed and renewed and maintained according to the licensing partys recommendations.

The IT infrastructure (c) should ideally be ISO 9001 and ISO 27001 certified in cases, where real-time operation and security is of paramount importance.

> ***Recommendation:*** *Definition of the required support structure should be part of the requirement list for any forecasting solution. For real-time forecasting solutions end-user need to ensure that there is an appropriate support structure in place. Considerations of the real-time environment, own resources and which of the forecasting business practices are of significance to the user should be carried out. Especially, where processes are supposed to run every day in the year.*

3.9.3.5 Redundancy Structure

Redundancy depends very much on the end-users needs to maintain a frictionless and continuous operation. Forecasting is mostly carried out in real-time, which has an inherit requirement of being functional all the time. While there are many processes and targets for forecasting that may not require large redundancy and permanent up-time, the following recommendation is targeted to those end-users where forecasting is to some extend mission critical. There are a number of different redundancy levels that need consideration and that can be achieved in various ways:

1. Physical delivery of the service IT infrastructure
2. Content of the delivery – Forecasting methods

The delivery of the service (1) is connected to the IT infrastructure. Redundancy measures may be a combination of any of these:

- Delivery from multiple locations to mitigate connectivity failures
- Delivery from multiple hardware/servers to mitigate individual server failure
- Delivery with redundant firewalls to mitigate hardware failure
- Delivery through a ISP using Email, etc.

The redundancy of the forecast content is equally important as the physical delivery of the data, but often neglected. It is recommended to consider any combination of the following redundancy measures for correct forecast content:

- redundant providers of weather input
- redundant/multiple providers of forecast service
- redundant input and mitigation strategy for weather models
- redundant input and mitigation strategy to power conversion models

> ***Recommendation:*** *Define the required redundancy level according to the importance of a permanent functioning service and the impact of delivery failure to other internal critical processes.*

3.9.3.6 Escalation Structure

It is recommended for high-level contracts, where forecasting is critical to the end-users processes to get information about escalation structures in case of failure. This is especially important when employing only one forecast provider. Recommendation: An end-user needs to have a description about structure and corresponding responsibilities for their operations staff in order to incorporate such information into own escalation structures in case of emergencies.

Each level of escalation ideally contains the following structured process:

Table 3.1: Recommendation of a three tier escalation structure.

Escalation Level	Forecast service provider coordination	End-user coordination
Level 1:failure to deliver service	Technical Staff	Operations Staff Project manager
Level 2: failure to recover or implement service	Project manager	Project/Department manager
Level 3: failure to solve failure/recovery	General management	General management

- Formulation of the problem/failure
- Root cause analysis
- Coordination of action plan for troubleshooting inclusive responsibilities
- Coordinated action plan progression
- Escalation to the next level or closure of escalation procedure

3.10 Forecast Methodology Selection for use of Probabilistic Forecasts

Currently, used methodologies of generating probabilistic uncertainty forecasts for the power industry have proven concepts and are integrated in today's business practices. Looking into these applications, it becomes apparent that uncertainty forecasts have found their place in the power industry, but are on the other hand far from being exploited to a level that could be expected and may be necessary in the future, considering the value that uncertainty forecasts already today can provide to many processes and applications.

The following definitions and recommendations aim to assist in the implementation of uncertainty or more general probabilistic forecast methods into operational processes. While this guide aims to be comprehensive, it is not possible to provide all details that may be necessary for a first time implementation or planning of a fully integrated probabilistic forecast solution. Nevertheless, the information provided is taken from existing documentation, partially coordinated by the IEA Wind Task 36, but also general publications. This information can be found in in the References Material under Uncertainty Forecast Information 5, especially the reviews on probabilistic methods for the power industry [**Bessa_2017**] and on uncovering wind power forecasting uncertainty origins and development through the whole modelling chain [**yan2021**].

3.10.1 Definitions of Uncertainty

In order to establish a common language of uncertainty forecasts, the most common definitions of uncertainty are explained. These are:

1. forecast error spread:
   the historically observed deviation of a forecast to its corresponding observation at a specific time. It can also refer to an average error provided by an error metric, e.g. variance or standard deviation.

2. confidence interval:
   A confidence interval displays the probability that an observed value will fall between a pair of forecast values around the mean. Confidence intervals measure the degree of uncertainty or certainty in a *sampling method*, not the forecast [1]. They are often constructed using confidence levels of 5%, 95% etc.

3. forecast uncertainty:
   is defined as a possible range of forecast values in the future. In meteorology this range is defined by the uncertainty of the atmospheric development in the future and represented in ensemble forecasts by applying perturbations to initial and boundary conditions and expressing model physics differences.

4. forecast interval:
   determined uncertainty band representing **forecast uncertainty** and containing the respective probability of the real value being contained in the range of forecasted values, which will only be observed in the future.

Forecast intervals are the most common used visualisation for forecast uncertainty. They can be derived from

(a) parametric (e.g. Gaussian distribution)

(b) non-parametric (e.g. empirical distribution functions, kernel density estimation) representations of uncertainty

(c) a larger number of NWP forecasts in an ensemble forecasting system that represent the forecast uncertainty of the target variable

From these probability density functions (PDFs), quantiles or percentiles[2] can be extracted and higher-order statistics such as skewness and kurtosis can be calculated. This

[1]One of the common misunderstandings is that a **confidence interval** is showing the uncertainty of a forecast. This is not the case. By adding and subtracting for example one standard deviation to the deterministic forecast of wind speed and converting it to wind power, such intervals represent a measure of the deviation to climatology and do not represent current or geographically distributed uncertainty.

[2]In statistics and the theory of probability, quantiles are cut points dividing the range of a probability distribution into contiguous intervals with equal probabilities. The 100-quantiles are called percentiles.

is where the distinction is most pronounced: from a statistical error measure like standard deviation, it is not possible to derive quantiles or percentiles.

For applications like reserve predictions, ramp constraints or optimization tasks for storage applications, this distinction is imperative. Such applications also require that the geographical distribution of the variables are captured by scenarios of ensembles of possible outcomes of a pre-defined value.

3.10.2 Uncertainty Forecasting Methods

Forecast uncertainty for application in the power industry are today based on three main processes and procedures (fig. 3.2):

1. **Statistical methods of probabilistic forecasts:**
   This method is based on statistical processing of past (historic) data in order to derive a probability density function of the possible forecasting spread. The advantage of such methods are that they are computationally extremely cheap and simple to apply. The disadvantage is that none of these methods produce a realistic representation of the forecast uncertainty in a spatial and temporal manner. There is also no physical dependency on the forward results, as the spread is based on past climatology. Typically, statistical learning algorithms (e.g., neural networks, machine learning) are used to fit historical time series of weather parameters from a NWP model to their corresponding power generation data. From the fitting process, a PDF can be derived and used forward in time. A newer, more intelligent method is the analogue ensemble method (AnEn) that searches through historical forecasts for those past events that are most similar or "analogous" to the current forecast. The observations with the best fit form the probability distribution of the forecast uncertainty. So far the method is one-dimensional and hence does not take geographical or temporal aspects of uncertainty into account. To be able to benefit from integration of information from geographically distributed time series or from a grid of NWP the methods needs to add a second dimension. This is in the focus of some recent research [**spirati2017**], where each grid point in an area, where wind farms are located, is treated independently, using meteorological analysis instead of observations.

2. **Statistically-based ensemble scenarios:**
   With this method statistically-based scenarios are produced that are a result of statistical generation of scenarios from the probability distributions produced by statistical models based on the copula theory. We define them as scenarios, as the further processing of the approach contains x independent results in contrast to the statistical method, producing a PDF function. Such scenarios are quite similar to the third methods, the physically-based ensembles. However, the uncertainty representation of the statistical scenarios today only capture the spatial variability of the forecast, like ramps. We therefore distinguish them here as scenarios rather then ensembles. Outliers that indicate extreme events, for example above cut-out wind speeds of wind turbines can only

be detected with probability characterisation and require an extreme event analysis. This is due to the conversion to power taking place in the first step of the statistical training in the same way as for deterministic forecasts. Extremes in wind power are in that way difficult to detect, because the flat part of the power curve prevents extremes that would be visible in the wind speeds to show up in the power scenarios. The clear advantage of the statistically based scenarios is that they are computationally much cheaper than physical ensembles as they are built from a deterministic weather forecast. They also generate a much more realistic uncertainty representation than the pure statistical approach, while only being slightly more computationally costly.

3. **Physically based ensemble forecasts:**
The third type of methodologies, the "physically based ensembles" can be considered a post-processing of a set of NWP ensemble members, which are a set of NWP forecasts produced by perturbing the initial or boundary conditions and/or model physics perturbation, the result from different parameterisation schemes of one NWP model "multi-scheme" approach) or complete different NWP models "multi model" approach), converted in a subsequent phase into power with a curve fitting method (see e.g. [**bessa2017**]). The NWP ensemble is configured to represent the physical uncertainty of the weather ahead of time rather than uncertainty as a function of past experience. In practice, this means that the NWP ensembles, especially the multi-scheme approach, are event driven, produce outliers and also catch extremes, even those with a return periods of 50 years. This is a clear distinction from statistical methods, because even long time-series of historic data contain too few extreme events to have impact in the learning algorithms. Often ensemble prediction systems (EPS) are found "under-dispersive", i.e. the uncertainty spread does not cover or represent the uncertainty of the target variables. This can have many reasons, some often found reasons being that:

    (a) the ensemble is not targeted to the variable of interest of the end-user
    (b) the time or spatial resolution is too coarse to capture the small scale phenomena of the target variable
    (c) insufficient information is extracted or used in the conversion to wind power to represent a realistic uncertainty. Mostly such deficiencies can be mitigated by calibration methods ((see e.g. [**bessa2017**]).

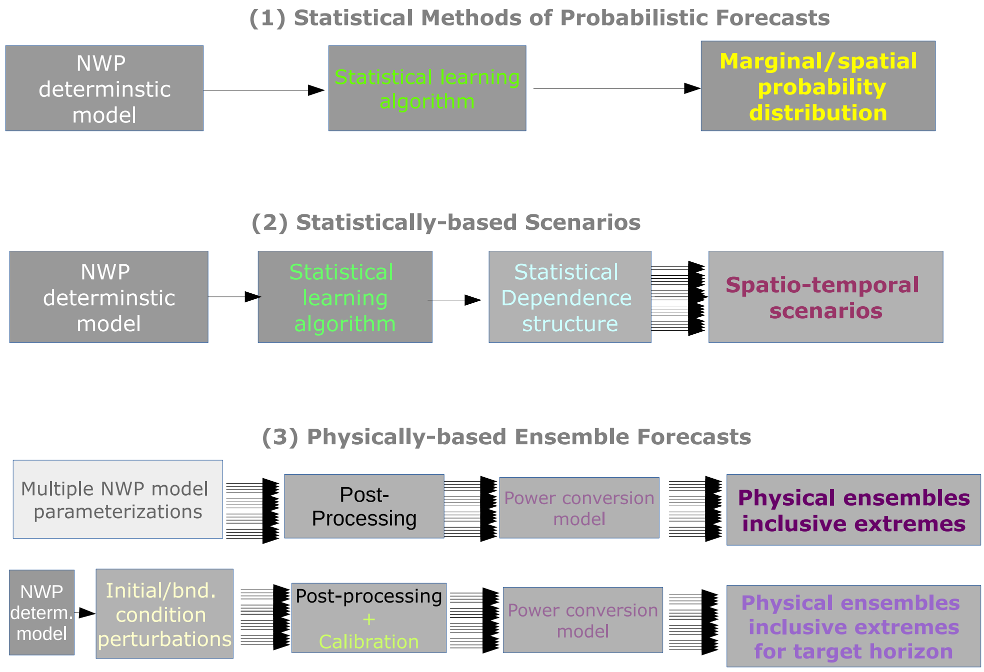


Figure 3.2: Standard methods of uncertainty forecast generation to be used in wind power and PV forecasting. The black arrows indicate whether the so-called ensemble members stem from a statistical procedure or are individual scenarios.

Recommendation: *when selecting a forecast method for a specific application, it is important to know, whether or not a specific method is suitable for the application or not. There are 3 major branches of uncertainty generating forecasting methods:*

1. *Statistical methods of probabilistic forecasts*
2. *Statistically-based ensemble scenarios*
3. *Physically based ensemble forecasts*

We have provided some basic guidance and a graph (fig. 3.2) summarising these methods in order to differentiate between the methods, showed the spatial and temporal dependencies of some methods and emphasised that the statistical methods are not suitable for applications that have such dependencies to the uncertainty measure of interest. These are e.g. applications that deal with extremes that may not happen frequently or where the uncertainty estimate is required in each hour of the forecasts rather than over a forecast period of a day or a week.
*More information about probabilistic methodologies can be found in the References Material under Uncertainty Forecast Information, especially in a review on probabilistic methods for the power industry [***bessa***].*

3.10.3 Applications of Uncertainty Forecasts in the Energy Industry

Typical applications in the energy industry, where it is recommended to use uncertainty forecasts, are:

1. **Balancing/trading of wind/solar power:**
   For balancing and trading applications the optimal bid/schedule, from the expected value decision paradigm, consists in a quantile calculated from the forecasted imbalance costs [**bremnes2004probabilistic**] or a percentile calculated from an ensemble forecasting system [**moehrlen2012a**, **moehrlen2012b**]. The calibration of uncertainty is a critical requirement for the end-user and has a non-marginal economic impact. Moreover, in electricity markets with high integration levels of wind/solar power, the combination of extreme forecast errors and high imbalance prices is critical and demands for risk modelling techniques and uncertainty forecasts with high accuracy in detecting extreme events (e.g., cut-out wind speed, ramps) ([**bessa2017**]). If the portfolio includes also energy storage units, the temporal dependency of forecast uncertainty is a primary requirement [**Haessig2015**]. For this use case, the end-user should request ensemble forecasts from physically based methods (see 3.10.2).

2. **Dynamic reserve and ramping reserve requirements:**
   The use of uncertainty forecasts for setting the power system reserve requirements is probably the most well-accepted business case for the energy industry. A critical requirement is minimum deviation from perfect calibration to avoid under- and overestimation of the risk (i.e., loss of load probability, probability of curtailing renewable energy) [**Bessa2012a**]. Another criteria in the design of dynamic reserve allocation are the boundaries that need to be defined. The following aspects are crucial boundaries in this respect:

   - **Use the correct type of ensemble data input**
     Physical NWP ensemble: e.g. multi-scheme approach
     Deterministic reserves do not provide uncertainty
     It is the weather uncertainty that generates the errors

   - **Clear definition of the forecast objective**
     Which types of errors are critical
     How to handle outliers
     What type of reserve fits to the end-users objective:
     typical scenarios are: static, security or dynamic/economic

   - **Definition of the time scales that needs to be forecasted**
     Required ramping capabilities

- **Forecast uncertainty required for all weather dependent sources & sinks**
  The uncertainty term should ideally be built upon load+wind+solar
- **Definition of a noise term to handle the non-local imbalances**
  imbalances from interconnections (small system <-> large system)

Allocating reserve dynamically requires probabilistic forecasts and the value for the system operator is well defined. Yet, the following challenges also need to be addressed, when implementing probabilistic forecasts for dynamic reserve requirements and allocation:

(i) communication and visualisation of forecast uncertainty and extreme events in TSO dispatch centers

(ii) training of human operators to understand and exploit the probabilistic information, i.e. move from a deterministic/ real-time paradigm to probabilistic/predictive operation paradigm.

An example of a dynamic reserve visualisation tool is illustrated and described in Appendix C.

3. **Extreme event warning such as High-Speed shut down warnings:**
   For risk indices, it is imperative that there exists a well-justified and transparent underlying computation of the conditions that may lead to a shut-down event impacting system security, that should be provided rather early than late. It is generally accepted that a planned scheduled shut-down at a slightly lower wind speed extend the lifetime of the gearbox system in wind turbines. Therefore, one could argue that there is modest economic loss by executing controlled shutdowns to reduce the ramp-rate in a power system.

   The value of such alert systems is gained with early detection of extreme events. This can for example be accomplished by introducing a gradual artificial transition from full load to no generation already at 22.5m/s. The starting point of such an index will be discussed below. A simple argument for 22.5m/s is that 2m/s is the typical forecast accuracy at such high wind speeds.

   A "high-speed wind event" can be defined as active, if the hub height wind speed is above 24.5m/s, while there is no event, if the wind speed is below 22.5m/s. Table 3.2 shows how such an index may be defined.

   The required low level forecast information to raise alerts can be generated in a typical 6-hourly cycle, although it may be coupled with a short-term forecast on a shorter frequency dependent on the importance of critical ramps for system security. One of the major challenges for such an alarm system is in fact the strategy of dissemination of the warning information to the user in the control room. If a critical event is discovered about 5 days in advance, the question is how often a warning should be issued, also

Table 3.2: Definitions of a high-speed shut down index for a control area with a high penetration level of wind power and a wind resource with a high variability and wind speeds often exceeding 25m/s.

wind speed in 100m	index value
0 - 22.5 m/s	0.00%
22.5 - 24.5 m/s	0 -> 100%
24.5 m/s	100.00%

in order to avoid too many false alarms or forecast misses. Threshold values for alert generation therefore has to be a function of lead time, time of the day and week day.

The more alerts there are generated, the less serious they are taken and the higher the likelihood that a critical event is overlooked. Nevertheless, there are periods where events should create alertness, even though they may not result in a sufficient strong concurrent shutdown. Typical examples could be:

- An alert at a 6 day horizon issued on a Thursday valid for Tuesday morning following a long holiday weekend may be desirable even if the likelihood is low.
- An alert to cause attention on a change of expected ramp rate 6 hours ahead, even though there has already been raised an alert for the event from previous forecasts

The objective for such an alert must always be to avoid costly actions to be initiated, if there is a critical ramp rate in the forecast far enough away that an economic solution can be prepared.

As briefly discussed below under *situational awareness*, if a major fraction of the power generation is wind dependent, it would be considered best practice, if the operator is aware of the risk of high-speed shutdown, even if the likelihood is low, but still justifiable. The same applies to the ramp rate caused by a fast moving low pressure system, where the center wind speeds may be below the cut-in level. Both event types can simultaneously amplify the ramp down rate and call therefore for a ramp rate based consideration instead of an isolated high-speed shutdown consideration.

4. **Situational awareness:**
For system operators, but also wind farm operators or trader, information from uncertainty forecasts can be integrated at two levels:

(a) *Visualization and cognition*: provision of alarms and early warnings to human operators about predefined events with impact in the frequency control tasks, e.g. large ramps, wind turbines tripping, large forecast errors. With this information, the human operator can use his/her experience or operating practice to derive a set of control actions (e.g., change current dispatch, activate reserve) that mitigate the effects of renewable energy uncertainty and variability in the system's frequency.

(b) *Technical evaluation of network constraints*: uncertainty forecasts can be integrated in a power flow module, available in commercial energy management systems (EMS), to detect voltage and congestion problems with a certain probability threshold [**Usaola2009**]. With this information the human operator can plan preventive actions in advance, e.g. change the market dispatch, define a cape for market offers in a specific network area/node.

The following requirements should be requested by the end-user for the forecasting provider:

(a) high accuracy in detecting extreme events related to RES uncertainty and variability

(b) capacity to capture the temporal and spatial dependency of forecast errors

5. **Flexibility management in smart power grids:**
The deployment of smart grid technology enables the control of distributed energy resources (DER), e.g. storage and demand response, which flexibility can contribute to increase the RES hosting capacity while maintaining the standard quality of supply levels. The combination of forecasting systems and optimal power flow tools can be used by transmissions and distribution system operators to pre-book flexibility for the next hours in order to handle the technical constraints of their electrical network [**Soares2017**].

Presently, distribution system operators are starting to explore RES forecasts in the following use cases: a) forecast grid operating conditions for the next hours; b) improved scheduling and technical assessment of transformer maintenance plans; c) contract and activate flexibility from DER to solve technical problems.

In all these cases, a primary requirement is the need to have a spatial-temporal representation of forecast uncertainty, where the temporal component is only relevant, if inter-temporal constraints are required (e.g., operation of storage devices, control of capacitor banks and on load tap changers).

Finally, a current topic of interest is the coordination between the transmission and distribution systems. Different frameworks for information management and exchange are under discussion [**Jong2016**]. It is clear that uncertainty forecasts can be used to provide future information about nodal consumption/injection in the interface between the two networks. For example, the FP7 European Project evolvDSO developed the concept of flexibility maps, where RES forecasts are used to quantify the operating point and flexibility range in the TSO-DSO interface [**Silva2017**]. This paves the way to combine information about forecast uncertainty and flexibility, as proposed in [**Bucher2015**].

Recommendation:
The transition of the energy systems towards a CO_2-free power generation with large-scale integration of renewables on a global basis also requires a restructuring of the power system operation processes. The intermittened generating units driven by wind and solar resources call for a more dynamic and weather driven structure of the operating practice. Probabilistic forecasts can suuport that dynamic structure and provide the possibility to deal with the uncertainties associated with the non-linearities of weather deveoplment as well as extremes that can affect the power system and cause large-scale blackout.
No forecasting solution today should be designed without the uncertainty of weather driven energy resources in mind. The minimum integration of uncertainty forecsats today are for the following application types:

1. Balancing/trading of wind/solar power
2. Dynamic reserve and ramping reserve requirements
3. Extreme event warning such as High-Speed shut down warnings
4. Situational awareness
5. Flexibility management in smart power grids

The basics of these methodologies have been described in this section. Detailed implementation information about the described probabilistic methodologies can be found in the References Material under Uncertainty Forecast Information, especially in a review on probabilistic methods for the power industry [**bessa2017**, **Wuerth2019**, **Haupt2020**, **yan2021**].

3.10.4 Visualisation of forecast uncertainty

The visualisation of uncertainty forecasts has shown to be a difficult topic, especially for the not so experienced user. While an expert immediately can see the difference between a chart generated with a statistical approach and an ensemble approach as defined in section 3.10.2, it can be difficult for a beginner. The descriptions of *Fan charts* and usefulness of the so-called spaghetti plots has the purpose of providing a general overview and assist in distinguishing of various methods for the graphical visualisation of uncertainty forecasts.

1. “**FAN CHART**”

   The “fan chart” is a common way of visualising a set of *forecast intervals* that are aggregated in one plot. Visualizations as shown in Figure 3.3 may however provide misleading information to a decision-maker. For example, if the decision-maker interprets each one of the quantiles as a possible evolution of wind power production in time, he needs to be sure that the visualization tool uses the data that he expects to interpret the information correct.

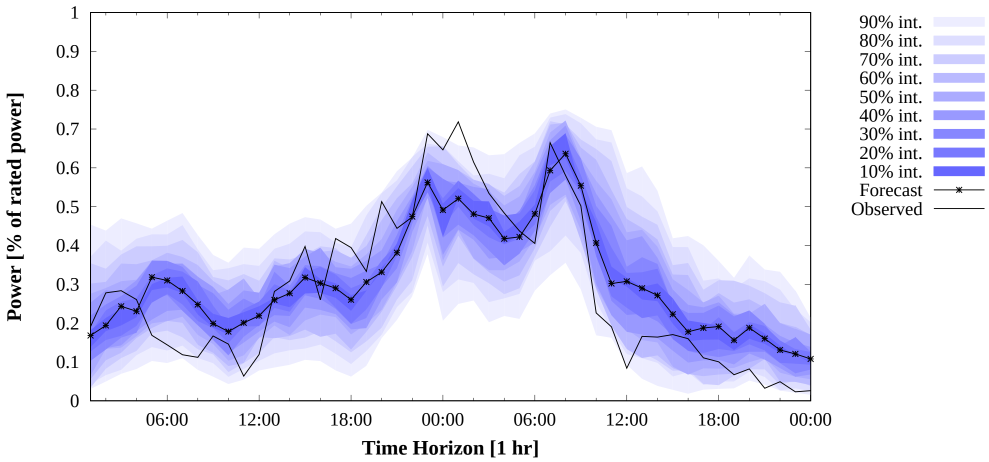


Figure 3.3: Example of a “Fan chart” of wind power production at a single wind farm built marginal forecast intervals of a statistical method.

Differentiation of forecast methods used in the fan chart:

- **statitical method**:
  A fan chart generated with a statistical method visualizes the “marginal forecast interval”, meaning each interval is only confined to separated forecast lead-times and does not have information about the joint probability distribution across the full set of lead times, or in other words, these intervals are not modeling the inter-temporal dependency structure of forecast uncertainty. These intervals are different for each lead-time. Figure 3.3 shows an example of a fan chart where the intervals were generated with a statistical model. The lead-time dependence is visible through the relatively equal intervals in size over the entire forecast. The observations (black solid line) are covered, except for a short period around midnight of the first day.

  In that hour there is a probability of around $\alpha = 90\%$ (limited by quantiles 95% and 5%) that the observed value is within approximately $P_{t+k}^{\tau^L} = 0.18$ and $P_{t+k}^{\tau^H} = 0.65$. This is the typical interpretation. Looking at the observations, another way to interpret is that there is a 5% likelihood that the observations are within $P_{t+k}^{\tau^L} = 0.63$ and $P_{t+k}^{\tau^H} = 0.65$.

2. **Ensemble Method**:
   In Figure 3.4 we also see forecast intervals for the same wind farm and day. This time, the intervals were formed of 300 wind speeds in 4 different heights by a 75 member

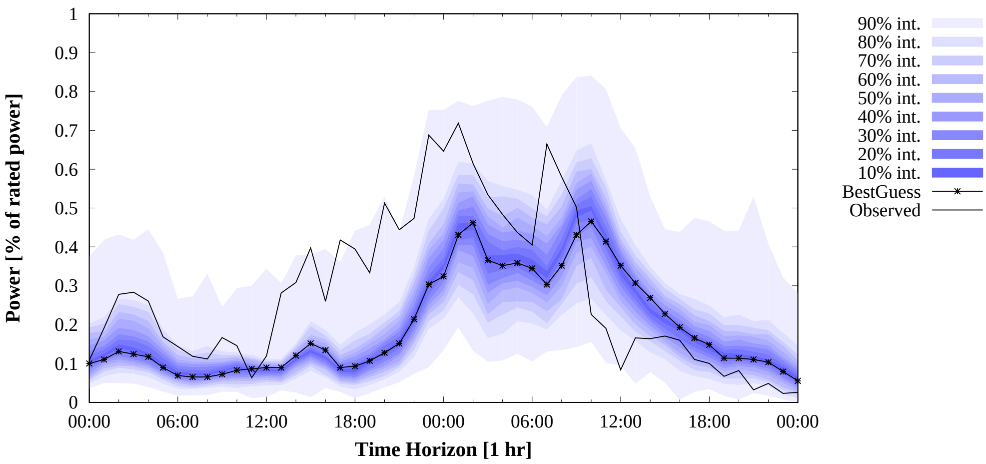


Figure 3.4: Example of a “Fan chart” of wind power forecasts at the same time and wind farm as in 3.3, built by a 75 member multi-scheme NWP ensemble system (MSEPS).

multi-scheme NWP based ensemble prediction system (MSEPS). These intervals look very different from the statistically generated intervals. Even though the 90% probability is within approximately $P^{\tau^L}_{t+k} = 0.21$ and $P^{\tau^H}_{t+k} = 0.75$, the 5% probability that the observations is found within the upper quantile has an interval size of 0.25 (range $P^{\tau^L}_{t+k} = 0.50$ and $P^{\tau^H}_{t+k} = 0.75$). That means the interval size is larger by a factor of 10. Compared to the statistical method, this result indicates that the current weather development contains a low probability for a high uncertainty range towards increased production.

Especially the physical based multi-scheme approach that provides the uncertainty in each time step quantifies the uncertainty with the knowledge of now and different physical approaches to compute the future development, rather than comparing the situation with past data.

Any application that may be subject to extreme events that may not have happened within the last months or years, should use uncertainty forecasts from this mthod and make sure that this is reflected in the visualisation of the forecasts.

3. **“Spaghetti Plot”**

Figure 3.5 shows the same wind farm, forecast days and method as in Figure 3.4, but as individual forecasts in a so-called spaghetti-plot where each of the 300 wind power forecasts are one line. In this way, it becomes apparent, how individual ensemble forecast “members” generated outliers.

In comparison to the lead-time dependent approach, the physical approach forms a large outer quantile band and a more condensed inner part, indicating that many of the 75 forecasts are aligned in their atmospheric development, while there are a small number of forecasts that result in higher power generation. The difference here is that

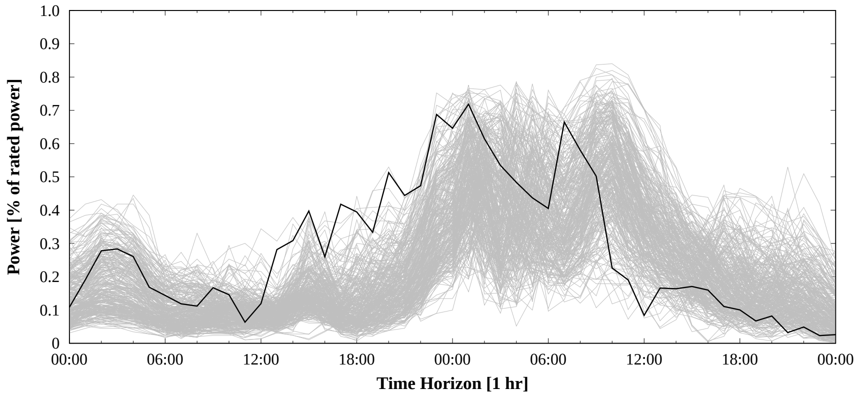


Figure 3.5: Example of a spaghetti plot of 300 wind power forecasts at the same time and wind farm and method as in 3.4.

the intervals are a result of the NWP ensembles reproducing the physical uncertainty of the current atmospheric processes that generate the power and are fully independent of the lead-time. Here, a large spread can be generated based on a very low likelihood or probability, also if such events have not been observed before.

An operator or trader has a number of ways to interpret such a forecast. Two likely scenarios could be:

(a) ignoring the outer interval and acting upon the highest probability ranges

(b) verifying the system upon issues or the market price that could arise, if the low probability of high generation would become reality

Whether the operator or trader acts upon such a forecast depends on their business practices. Nevertheless, it shows that the information contained in the forecast intervals have a direct practical application.

Take-away from the Visualisation of Uncertainty:
The different results also illustrate that the successful interpretation of such information depends on the algorithm used to generate these intervals and an understanding of the methods with which such intervals have been created. The major difference here is that one method is based on current atmospheric conditions (NWP ensemble) and the other relies on historical documentation of the atmospheric conditions. Here, the intervals of extremes are usually smaller and less pronounced unless there are long time series available that contain a significant number of such extremes to impact the spread in given weather conditions.

Chapter 4

DATA COMMUNICATION

Key Points
This section provides recommendations for forecast data-related terminology, data description and data formats and exchange protocols

- The **terminology section** (4.1) provides definitions for terms commonly used in the description and exchange of data between forecast providers and users
- The **description section** (4.2) provides a specification of the mandatory and optional types of data required to train and operate a forecasting system
- The **exchange section** (4.3) provides recommended standards for data formats and exchange protocols. The standards are presented for two levels of users:

  Level 1: basic data format and data exchange for groups with limited IT knowledge and/or experience

  Level 2: more detailed description of more sophisticated formats and exchange protocols that enable groups with more extensive IT knowledge, experience and resources to achieve higher level of data exchange robustness and efficiency.

Currently, there exists no best practices or standard for (1) definitions of the various inputs that are used to configure and operate a renewable energy forecasting system, (2) data formats and (3) data exchange protocols. Forecast suppliers and consumers use different terminology and can end up spending many hours on unnecessary communication. This is a problem which ultimately comes at a cost to the supplier, energy forecast purchaser and overall cost of operating renewable energy projects. Similarly, there is no standard or recommended best practice for the format of the data used by the forecasting systems and the methods for delivering this data. This also creates additional delays and inefficiencies in the forecast setup process.

Two user groups are targeted in this recommendation that are referred to as Level 1 and Level 2 users in this document:

Level 1 users: this group typically has limited IT resources or experience. Input documents are geared towards manual input of forecast specifications.

Level 2 users: this user group will have a deeper IT knowledge or experience than Level 1 users. Input file format examples and exchange methods presented here may be programmatically adopted for real-time use.

Once adherence to the data exchange standard is attained, the benefits are numerous and tangible. For the supplier of renewable energy forecasts this is:

1. More efficient onboarding Level 1 consumers (i.e., with less experience using forecasts or less IT expertise) standard templates are followed; for more advanced, Level 2 users, the onboarding process can be mostly automated.

2. Back-and-forth communication time is minimised - Online references to standard documentation reduces communication blockages between forecast supplier and consumer.

3. Greater automation of adding/removing renewable power plants to forecast engine. Adhering to a standard reduces (doesnt eliminate) the need for customised software development

For the renewable energy forecast consumer, some benefits of adhering to the data exchange interface include:

- A much more efficient process to benchmark different forecast providers
- Easier to add, remove, or switch forecast providers
- Quicker turnaround time in adding or removing forecast projects to asset portfolio
- No need to develop new processes for different weather-impacted renewable technologies

This recommended practice can be applied to other types of renewables. The focus of this specification is forecasting of individual renewable projects. An aggregate or area-level forecast is not specifically addressed in this document as these configurations are less common. However, aggregations and area forecasts may be treated like a single project or be specifically coded by a unique ID (e.g., Wind Region 01). Data field definitions are also provided in the hopes of promoting a standard definition for these fields. However, we recognise that there are many factors that make data definition standards more challenging and beyond the scope of this recommendation. For example:

- Industry-specific standards several industries are involved in the Renewable Energy industry and each has its own data definitions for similar fields. This includes OEM (turbine manufacturers), independent engineers (consultants), SCADA software companies, renewable plant developers, utilities, and TSO or ISOs.
- Geographic location terminology and translation from one language to another.

Different standardisation bodies governing different industries also can present challenges, but for renewable energy forecasting, the International Electrotechnical Commission (IEC) is the generally accepted organisation that establishes, promotes, and updates data definition standards.

4.1 Terminology

For clarification, several terms are defined that appear repeatedly in this section and therefore are defined below.

Renewable Energy Forecast Customer/Consumer: an institution or corporation that requires a forecast of power from a renewable energy generation facility with a look-ahead time of minutes to days.

Renewable Energy Forecast Provider: an institution or corporation that delivers renewable energy forecasts with a look-ahead time of minutes to days. In most cases, the provider is a company whos business includes selling forecasts to customer/consumer.
Renewable Energy Forecast Trial: an exercise conducted to test the features and quality of a renewable energy forecast such as wind or solar power. This may include one or more participants and is normally conducted by a private company for commercial purposes. A trial is a subset of a Renewable Energy Forecast Benchmark.
Renewable Energy Forecast Benchmark: an exercise conducted to determine the features and quality of a renewable energy forecast such as wind or solar power. The exercise is normally conducted by an institution or their agent and multiple participants including private industry forecast providers or applied research academics.
Online measurements: These are observations used for tuning a renewable energy forecast system and adjusting intra-day renewable energy forecasts. Measurements are usually power or energy since that is the target variable of interest to the consumer. However, other weather variables might be included with online measurements. Online measurements are also referred to as real-time measurements and are transferred between forecast customer and provider on a regular basis.
Offline measurement: These are observations used for tuning a renewable energy forecast system. As opposed to online, offline observations are historical and dont directly impact short term (< 12 hours) forecast horizons. Measurements are usually power or energy since that is the target variable of interest to the consumer. Other weather variables might be used for energy forecast training, especially for a newly operational renewable energy plant.

4.2 Data Description

This standard interface is defined on two levels. Ideally renewable energy forecast customers and renewable energy forecast providers should comply with both level 1 and level 2 of the standard, but only complying with level 1 can also provide significant efficiency gains during the setup of a forecasting system. The two levels are:
Level 1: A high-level description of the information and data required to carry out a successful trial and operation of a specific forecast solution. This level of standardisation provides a common terminology that will enable the renewable forecast customer to prepare and organise data facilitating an efficient system configuration process.

Level 2: A detailed specification of both the format and method, which should be used to exchange data between the renewable forecasting provider and the renewable energy forecasting customer. This level of standardisation enables an efficient, repeatable and scalable configuration process applicable to trials as well as operational forecast systems. Compliance with level 2 facilitates renewable energy forecast customers to efficiently carry out trials/benchmarks as well as enabling renewable energy forecast providers to participate in trials/benchmarks efficiently and at low costs.
Online/real-time measurements are suited for both intraday and day-ahead forecasting whereas offline measurements are best for day-ahead forecasting only.

Mandatory and optional data
The metadata tables and definitions below specify some data to be mandatory in order to setup a meaningful forecasting system. Other data is considered optional as those data may improve forecast accuracy, take into account a future operating state of the renewable energy plant, or might be less common forecast consumer requirements. All renewable energy forecast consumers should be able to provide the mandatory data and all forecast providers should be able to process mandatory data, whereas optional data depends on specifics to each forecast installation.

4.2.1 LEVEL 1 Data description

Table 4.1 provides an overview of the different data types required by a forecasting system. This data needs to be available to the forecast provider for both training and operating a forecasting system. Table 4.2 provides an overview of the different types of meta data, which describes the attributes of the data types listed in Table 4.1. Tables 4.3 to Table 4.7 contain the data field and definitions that a forecast solution will require configuring an operational forecast.

Table 4.1: List of the different types of input and output data needed in a forecast setup

Data	Type of Data	Description of the Type of Data
Master Data	Site information	A specification/description of the site(s). A description can contain one or more sites. A site can be an aggregate of multiple sites. All sites in the same description must have similar data structure as specified in the associated meta data descriptions. If the data structures are not similar, then the sites need to be split up into multiple Sites and multiple meta data descriptions.
Online Data	Measurements	Observational data from a site which will be used as input for training models produced by the forecast system.
	Future Availability	The data about expected future availability of the site(s) due to maintenance, curtailment or other planned schedules. Used as input to the forecast.
	Forecasts	The output data (results) produced by the forecast system.

It is important to note that Measurements data should be made available both as historical data (also referred to as Offline Data) for training of models and as operational or real-time data (also referred to as Online Data) for operational forecasting. If available, the forecast customer should provide a minimum of 3 months of historical Measurements, but ideally 1-2 years of historical data to capture both seasonal and inter-annual variability.

Metadata, also referred to as Master Data, has a broad definition but generally refers to information that describes the forecast configuration or the data itself. In most cases, it should not change often. The different types of metadata needed for renewable energy forecast include information about the sites, the measurements, and the forecast configuration. This is detailed in Table 4.2 .

The metadata about the renewable energy site (or power plant) itself is often the first information a forecast supplier will need to initially set up a forecast installation and is, therefore, essential to be accurate and from the forecast consumer. Table 4.3 details the renewable energy site metadata.

Table 4.2: List of different types of meta data needed in a forecast setup

Type of meta data	Description of the meta data
Site specification	This is a description of the renewable energy power plant characteristics. Most of the characteristics dont change often with time.
Measurement specification	This is a description of the attributes of the observational data sent to the forecasting system. Measurement meta data may be separated into realtime (online) and historical (offline) measurements because these can often be described differently or represent different parameters.
Forecast Time Series specification	This is a description of the forecast system product which is the time series output.
Scheduled Availability specification	This is a description of the forecasted or scheduled availability at the power plant.
Forecast system specification	A description of the necessary inputs and outputs that tells the forecasting system how to model the target forecast variable(s). This includes forecast system and output attributes such as units, variables, timing, and temporal resolution.

Table 4.3: Specification of the data required for describing a renewable energy site

Sites:
The sites description can contain data about one or more site(s). A site can be comprised of an aggregate of sites or, in the case of wind, an aggregate of individual turbines. For each site the following data needs to be filled out.

Description	Mandatory /optional	Wind/ PV/Both	Type	Data field
Plant Name ID	Unique ID (name) used for identifying the site(s)	Mandatory	Both	String
Generation Type	Either Wind or Solar	Mandatory	Both	String
Latitude	Latitude coordinate of Plant Name in decimal degrees.	Mandatory	Both	Float
Longitude	Latitude coordinate of Plant Name in decimal degrees.	Mandatory	Both	Float
Capacity	Capacity of the site (often also referred to as the rated power of the site) for which the forecast should not exceed (kW)	Mandatory	Both	String
Hub height	The average height of the wind turbine hubs (meters)	Mandatory	Wind	Float

Number of turbines	Number of turbines that comprise wind farm	Mandatory	Wind	Integer
Wind turbine make and model	Turbine manufacturer and model name	Mandatory	Wind	String
Turbine power curve*	Default power curve table of wind speeds and corresponding power capacity factor. May be turbine manufacturers specification. Normalized by rated capacity in the range [0, 1]	Optional	Wind	Float
Solar Technology	Description of the PV technology. Fixed-tilt, single axis or dual axis are most common.	Mandatory	PV	String
Minimum panel orientation angle	The minimum orientation angle of the PV panels in degrees from north (0 to 359, where east is 90ř and west is 270ř). If the system is single axis tracking or fixed tilt, then minimum and maximum orientation should be the same.	Mandatory	PV	Integer
Maximum panel orientation angle	The maximum orientation angle of the PV panels in degrees from north (0 to 359, where east is 90ř and west is 270ř). If the system is single axis tracking or fixed tilt, then minimum and maximum orientation should be the same.	Mandatory	PV	Integer
Minimum panel inclination angle	The minimum angle of the PV panels in degrees from horizontal (-90 - +90). If the system is fixed tilt, then minimum and maximum inclination should be the same.	Mandatory	PV	Integer
Maximum panel inclination angle	The maximum angle of the PV panels in degrees from horizontal (-90 - +90). If the system is fixed tilt, then minimum and maximum inclination should be the same.	Mandatory	PV	Integer

Panel tested capacity	The total solar panel capacity for the site (kW) based on standard test conditions (1000W/m^2) at panel temperature of 25řC	Optional	PV	Float
Panel temperature sensitivity	The temperature sensitivity of the PV panels (%/řC)	Optional	PV	Float
Inverter capacity	The capacity (kW) of the inverters	Optional	PV	Float
Inverter efficiency	The inverter efficiency at 95% load	Optional	PV	Float

* Turbine power curve table is often appended or delivered separately.

Measurement data for the forecast site is often provided prior to the configuration of the forecast system to calibrate the forecast model and thus reduce forecast error. If the renewable energy site has been in operations, then this would likely include power observations since that is typically the target forecast variable. However, for new renewable energy sites, a history of wind speed or irradiance is often provided to the forecast supplier which may help reduce forecast error until a suitable history of power observations is obtained.

Table 4.4 details the specific fields that describe the measurement data required for renewable energy forecast model training. This table should be filled out more than once if the historical measurement data differs in any way from the online (realtime) data being sent to the Forecast Provider. Differences may arise due to the type of observation (e.g., SCADA power versus settlement power observations, wind speed from a met tower versus nacelle anemometer average wind speed).

Table 4.4: Specification of the data required for describing the forecast system input measurements

Measurements:
The Measurements should be delivered for each site and contain the information described below. All values must be available with the same granularity, e.g. every 5 minutes and with a fixed update frequency.

Data field	**Description**	**Mandatory /optional**	**Wind/ PV/Both**	**Type**
TIME SERIES FIELDS				
Plant Name ID	Unique ID (name) used for identifying the site(s)	Mandatory	Both	String
StartTime	Date and time stamp indicating the start of the period for which the measurements are observed.	Mandatory	Both	String
EndTime	Date and time stamp indicating the end of the period for which the measurements are observed.	Mandatory	Both	String
Power	The power production (kW) when the measurement was observed	Mandatory	Both	Float
Available capacity	The observed available capacity (kW) of the site due to a reduction in available generators. If wind turbines, solar panels or inverters are not available (due to maintenance, break downs or similar) the capacity of the site is temporarily reduced.	Optional	Both	Float
Limitation level	The limitation (kW) of the site due to curtailment, set point level, do not exceed limitations or down regulation. This includes non-scheduled changes by grid operator.	Optional	Both	Float
Wind speed	Wind speed (m/s) from a representative instrument. For example, a mean of the turbine nacelles, a meteorological tower anemometer or LIDAR instrument).	Optional	Both	Float
Wind direction	Wind direction measured in degrees from north (0-359ř). East = 90ř, South= 180ř, West = 270ř, North = 0ř.	Optional	Wind	Float
Temperature	Air temperature (degrees Celsius)	Optional	Both	Float

Pressure	Atmospheric air pressure (hectopascals, hPa)	Optional	Both	Float
Relative Humidity	Relative humidity of the air (%)	Optional	Both	Float
Precipitation	The amount of rain or snow that has fallen on the ground (millimeters, mm)	Optional	Both	Float
Global Horizontal Irradiance	The total short-wave radiation from the sky falling onto a horizontal surface (W/m^2)	Optional	PV	Float
Global tilted irradiance	The total short-wave radiation from the sky falling onto a tilted surface (W/m^2)	Optional	PV	Float
Direct solar irradiance	The short-wave radiation that has not experienced scattering in the atmosphere (W/m^2). The radition comes from the disc of the sun.	Optional	PV	Float
Diffuse irradiance	The short-wave radiation from light scattered by the atmosphere excluding from the disc of the sun (W/m^2)	Optional	PV	Float
META DATA				
Time zone	The time zone of the timestamp in IANA Time Zone (TZ) database format (e.g, Europe/Barcelona)	Mandatory	Both	String
Time interval label	Describes what time the measurement point represents. Can be instantaneous, period beginning average (leading), or period ending average (trailing)	Mandatory	Both	String
Power measurement type	This field is a text description of the power measurement field. It can be, for example: substation meter, SCADA power, active power, potential power, settlement power.	Optional	Both	String
Wind speed measurement type	Specify if the wind measurement for the site is from turbine nacelle, average of nacelle, met mast or other.	Optional	Both	String
Wind speed measurement height	The height of the wind speed measurement in meters.	Optional	Both	Float

Most forecast suppliers have the ability to incorporate scheduled changes to the renewable plants availability on the forecasted output. This comes in the form of reduced capacity owing to reduction in available units (e.g., turbines or inverters) or due to a generating limit of the power plant (e.g., curtailment or transmission limit).). The important distinction between an outage and limitation is that an outage is a proportional reduction in the plants capacity for all wind or solar irradiation conditions. The limitation is a maximum capped output (e.g., set point) of the plant based on the available capacity. This information needs to be described as it will routinely be sent from the forecast consumer to the forecast supplier. Table 4.5 details this information.

Table 4.5: Specification of the data required for describing the future availability and curtailments

Scheduled Availability: Future availability and curtailments for each site, should contain the information described below.				
Data field	**Description**	**Mandatory /optional**	**Wind/ PV/Both**	**Type**
Plant Name ID	Unique ID (name) used for identifying the site(s)	Mandatory	Both	String
StartTime	Date and time stamp indicating the start of the period for which the measurements are observed.	Mandatory	Both	String
EndTime	Date and time stamp indicating the end of the period for which the measurements are observed.	Mandatory	Both	String
Outage level	The expected available capacity (kW) of the site. If wind turbines, solar panels or inverters are not available (due to maintenance, break downs or similar) the Capacity of the site is temporarily reduced to available power capacity.	Optional	Both	Float
Limitation level	The expected available capacity (kW) of the site due to a limiting factor such as curtailment, setpoint instruction from grid operator or temporary limit on the maximum allowable production.	Optional	Both	Float
Reason of unavailability	Editable text that gives a reason for the reduction in available capacity (e.g., Maintenance, plant limitation). Often entered by plant manager or remote operations center.	Optional	Both	Float

The forecast deliverable is a point time series generated by the forecast software system. This product has characteristics that will vary depending on the consumers needs. Table **??** contains a description of the forecast file metadata.

Table 4.6: Forecast time series specification metadata

Forecast Time Series: The attributes of the output forecast time series are described below.				
Data field	**Description**	**Mandatory /optional**	**Wind/ PV/Both**	**Type**
Plant name ID	Unique ID (name) used for identifying the site(s)	Mandatory	Both	String
StartTime	Date and time indicating the beginning of the forecast interval	Mandatory	Both	String
EndTime	Date and time indicating the ending of the forecast interval	Mandatory	Both	String
Power	The power production forecast (kW) for the period	Mandatory	Both	Float
Power quantiles	Probabilistic power production forecast corresponding to a specific quantile level	Optional	Both	Float
Wind speed	Wind speed forecast (m/s)	Optional	Both	Float
Wind direction	Wind direction forecast in degrees from north (0-359ř). East = 90ř, South= 180ř, West = 270ř, North = 0ř.	Optional	Wind	Float
Temperature	Air temperature forecast (degrees Celsius)	Optional	Both	Float
Pressure	Atmospheric air pressure forecast (hectopascals, hPa)	Optional	Both	Float
Relative Humidity	Relative humidity of the air forecast (%)	Optional	Both	Float
Precipitation	Forecast of the amount of rain or snow that has fallen on the ground (millimeters, mm)	Optional	Both	Float
Global Horizontal Irradiance	The forecast total short-wave radiation from the sky falling onto a horizontal surface (W/m^2)	Optional	PV	Float
Global tilted irradiance	The forecast total short-wave radiation from the sky falling onto a tilted surface (W/m^2)	Optional	PV	Float

Direct solar irradiance	The forecast short-wave radiation that has not experienced scattering in the atmosphere (W/m^2). The radition comes from the disc of the sun.	Optional	PV	Float
Diffuse irradiance	The forecast short-wave radiation from light scattered by the atmosphere excluding from the disc of the sun (W/m^2)	Optional	PV	Float
Event Forecast	Forecast value of the custom-defined power forecast event (e.g., ramp rate probability, %)	Optional	Both	Float

The forecast deliverable is the product of many configuration parameters within a forecast software system. Not all forecast software systems have similar built-in features, but Table 4.7 highlight some of the salient details that are important to a forecast system specification regardless of software system implementation details.

Table 4.7: Specification of the data required for describing the Forecast System configuration

System specification: This specification describes general aspects (meta data) of the forecast system				
Data field	**Description**	**Mandatory /optional**	**Wind/ PV/Both**	**Type**
Plant name ID	Unique ID (name) used for identifying the site(s)	Mandatory	Both	String
Power unit	Unit of power quantities (MW, KWh)	Mandatory	Both	String
Time zone	The time zone of the timestamp in IANA Time Zone database format (e.g, Europe/Barcelona)	Mandatory	Both	String
Daylight Savings Time flag	Flag to indicate whether daylight savings time applies to forecast file (True or False)	Mandatory	Both	String
Time stamp format	Format of the date and time stamp (e.g., yyyy-MM-ddTHH:mm:ss).	Mandatory	Both	String
Forecast interval label	Describes what the time of forecast point represents. Can be instantaneous, period beginning average (leading), or period ending average (trailing)	Mandatory	Both	String
Issue time of day	The time of day that a forecast is issued specified in HH:MM format, e.g. 00:30. For forecast runs issued multiple times within one day (e.g. hourly), this specifies the first issue time of day.	Mandatory	Both	String
Forecast update frequency	Define how often the forecast time series is updated (in minutes)	Mandatory	Both	Integer
Forecast interval length	The length of time (in minutes) each forecast point represents	Mandatory	Both	Integer
Measurement delay	The expected time delay from when a value is measured until it is available to the forecasting system in minutes	Mandatory	Both	Integer
Forecast maximum horizon	Horizon (or maximum look-ahead) of the forecast in hours	Mandatory	Both	Integer

Forecast quantiles	Quantile of the forecast distribution given to the nearest integer. Specify a single or list of quantiles (e.g., P10, P25, P50, P75, P90, P99)	Optional	Both	String
Forecast weather variable units	Units corresponding to weather forecast variables	Optional	Both	String
Event fore-cast	Name and description of custom fore-cast variable	Optional	Both	String
Event fore-cast units	Units of custom forecast event	Both	String	

4.3 Data Format and Exchange

4.3.1 LEVEL 1 Data Format and Exchange

Two main considerations in recommending a suitable data format for Level 1 users of forecasts is ease of use for the forecast client and, for the forecast provider, the ability to programmatically read in the necessary metadata and time series input files. Comma Separate Value (CSV) satisfies these conditions and is also an acceptable format for Level 2 users. While Level 2 users can utilize high-level programming languages to generate and process CSV files, Level 1 users can still interact with CSV format through Microsoft Excel or other ascii-text editors. Another advantage of CSV for Level 1 users is that they can shift to become a Level 2 users should experienced IT resources be procured. It can be done incrementally by keeping the same CSV format or shifting to JSON or XML as is described below.

Level 1 users that have less IT experience or very limited IT resources on hand, it is recommended that data transfer between forecast consumer and forecast provider be done via *SSH File Transfer Protocol (SFTP)* which is a secure and reliable standard of sharing data. Forecast providers typically will host an SFTP server and issue a unique login credential to the forecast consumer. The forecast consumer can choose among several SFTP applications common in both the Windows and Linux operating system environments. The SFTP application should have the functionality to create batch job configurations such that online measurements can be automatically uploaded between forecast consumer and provider.

Command line SFTP works out-of-the-box with Linux operating system installations. Software applications such as Filezilla that support SFTP can also be installed on a server running Linux. Many file transfer applications built for Windows operating system have SFTP functionality. This includes Filezilla and WinSCP.

Most cloud storage commercial systems (e.g., OneDrive, GoogleDrive, Dropbox, AWS) now include SFTP. As API's become more user-friendly and accessible for people without computer programming experience, this recommendation might have to be revised as REST API (along with SFTP) are the Level 2 recommended data exchange method between forecast provider and consumer.

4.3.2 LEVEL 2 - Data Format and Exchange

Level 2 data exchange is for both providers and consumers of renewable energy forecasts that wish to programmatically exchange data. Thus, the most widely used and extensible tools and formats were selected to allow for more seamless integration. The following three data formats are the most commonly used in the exchange of renewable energy forecasting data and the measurements that feed into forecasting systems:

1. CSV (Comma Separated Value)
2. JSON (JavaScript Object Notation)
3. XML (eXtensible Markup Language)

Each format comes with its own advantages and weaknesses, but all are in ascii, human-readable form. Examples using representative data are useful for illustrating how metadata, measurement, and forecast data are organized and may be copied as a template. CSV-format files can be used for both Level 1 and Level 2 users as it has the distinct advantage of being opened and converted in the widely-used Microsoft Excel application. However, for Level 2 users that want to automate processes or configure many forecasts, JSON and XML are much more efficient formats.

Both JSON and XML formats are accompanied by a schema whose main purpose is to enforce consistency and data validity. They also serve as human readable documentation for the forecast system metadata, measurements, and forecast files.

Two widely used modes of data exchange are recommended in this best practice. They are:

- Secure File Transfer Protocol (SFTP), and
- Application Programming Interface (API) that is RESTful

One of the main advantages of programmatically generating metadata and time series measurement and forecast files is the wide number of applications that can be used to exchange this data from forecast provider to forecast consumer and vice-versa. Many commonly used programming languages such as Python, Javascript, R, and Ruby have packages and libraries that make parsing and interacting with JSON and XML formats easier. Additionally, and often overlooked, is that internal applications can rely on the same data exchange methods further standardizing code that has to be developed and maintained.

In recommending a data exchange standard, there are several important issues that have been considered including:

- International support and usage
- Upfront, transitional and integration resource costs (financial, human)
- Extensibility
- IT Security
- Ease of use

International support and usage can not be overstated. If a method of data exchange becomes obsolete or doesnt have a very large user base, it will not adapt to evolving communication and security standards. Additionally, the number of people familiar with the exchange method is smaller thus putting operational support of automated processes more at risk in case of a disruption in data flows.

The **upfront, transitional and integration costs** all have to do with how much an organization must invest to build, transition, or support the methods of data transfer. If the exchange method is open source and works across operating systems, this is a huge benefit to the organization. If the data exchange methods can be integrated into existing software (e.g., using a simple URL to make requests), this lowers the integration and development costs. Transitional costs associated with adopting the standard or best practice will be a key consideration on whether an organization chooses to incorporate a new data exchange method. This touches upon all the important considerations listed above.

Extensibility is a key issue in choosing an exchange method as society goes through an energy transition away from fossil fuels and towards an energy system with greater electrification, distributed generation and load, and faster internet. Will what works today for sending data back and forth be around in 10 years time? This is why we recommend not one data exchange method, but two, since its difficult to predict how computers, electronic devices and communications will change over time. RESTful APIs are quickly becoming the international standard in data communications as it is not constrained to one format (like SOAP is with XML) and doesnt have to be http-based although that is the primary protocol in use today with RESTful APIs.

IT security is crucial in the energy industry as malicious malware usage and cyber attack incidents have grown in number and scale every year. Both SFTP and RESTful API supports secure standard protocols for data sharing. Secure shell in the case of SFTP and HTTPS using TLS encryption as one example in RESTful APIs. Although these protocols dont eliminate security risks entirely, they reduce the risks of cyber attacks to your organization.

Both SFTP and RESTful APIs are supported by most software applications and are **easy to use** since they can be invoked from the most commonly used scripting languages. Ease of use may be the primary factor in deciding whether to adopt a new method to transfer data for the purpose of renewable energy forecasting. A forecast provider can develop a web application to accept any of the recommended formats. There are many tools and applications which then allow the forecast consumer to verify or view the metadata, measurements and forecasts in a web browser. A good example of this can be seen in the US Department of Energy Solar Forecast Arbiter (SFA) project (Hansen et al., 2019). This project developed an open-source RESTful API that uses JSON formatted messages. Forecast site metadata, measurements, and forecast time series are exchanged through POST and GET commands. Once the data is uploaded via the API, a dashboard has been built that allows the forecast provider and consumer to visualize, download, and create verification reports.

Sample formatted template files and accompanying schema files will be made available in CSV, JSON, and XML format for download accessible via the IEA Task 36 Work Package 3 website. This is currently under construction.

Chapter 5

FINAL AND CONCLUDING REMARKS

While every forecasting solution contains very individual processes and practices, there are a number of areas that all forecasting solutions have in common. For any industry it is important to establish standards and standardise practices in order to streamline processes, but also ensure security of supply with a healthy competition structure.

This document is providing state of the art practices that have been carefully collected by experts in the area and reviewed by professionals and experts in an appropriate number of countries with significant experience in wind energy forecasting. The recommendations are to encourage both end-users and forecast service providers to bring focus to areas of practice that are common to all solutions. The document will be updated as the industry moves towards new technologies and processes.

The key element of this recommended practice is to provide basic elements of decision support and thereby encourage end-users to analyse their own situation and use this analysis to design and request a forecasting solution that fits their own purpose rather than applying a doing what everybody else is doing-strategy.

This document is also intended to serve forecast service providers new to the market or those wanting to evolve to a new level of service and support as a guideline to state of the art practices that should be incorporated into business practices.

REFERENCE MATERIAL

NOTE: Access to references at IEA Wind Task 36 webpage:
http://www.ieawindforecasting.dk

Forecast solutions, Trials and Benchmarks IEA Wind Task 36: Recommended Practices Guideline for the Implementation of Wind Power Forecasting Solutions Part 2: Designing and executing forecasting bench-marks and trials. Online access: http://www.ieawindforecasting.dk

Corinna Möhrlen, John Zack, Jeff Lerner, Aidan Tuohy,ăJethro Browell, Jakob W. Messner, Craig Collier, Gregor Giebel, Recommended Practices for the Implementation of Wind Power Forecasting Solutions Part 1: Forecast Solution Selection Process, Proc. 17th Int. Workshop on Large-Scale Integration of Wind Power into Power Systems,ăStockholm, Sweden, October 2018. Online Access: ăhttp://download.weprog.com/wiw18-133_recommended-practice_selection-process.pdf

C. Möhrlen, C. Collier , J. Zack , J. Lerner , Can Benchmarks and Trials Help Develop new Operational Tools for Balancing Wind Power?, Proc. of 16th International Workshop on the Large-Scale Integration of Wind Power into Power Systems, Paper WIW-292, Berlin, Germany, 2017. Online access: http://download.weprog.com/WIW2017-292_moehrlen_et-al_v1.pdf Evaluation and Metrics

IEA Wind Task 36: Recommended Practices Guideline for the Implementation of Wind Power Forecasting Solutions Part 3: Evaluation of forecast solutions. Online access: http://www.ieawindforecasting.dk

C. Möhrlen, C. Collier , J. Zack , J. Lerner , Recommended Practices for the Implementation of Wind Power Forecasting Solutions Part 2&3: Designing and executing forecasting benchmarks and trials and evaluation of forecast solutions, Proc. of 16th International Workshop on the Large-Scale Integration of Wind Power into Power Systems, Paper WIW-160, Berlin, Germany, 2017.
Online access: http://download.weprog.com/wiw18-160_recommended-practice_

benchmark-evaluation.pdf

Anemos.Plus Project DELIVERABLE D-1.3 (), Towards the definition of a standardised evaluation protocol for probabilistic wind power forecasts. Online available: http://www.anemos-plus.eu/images/pubs/deliverables/aplus.deliverable_d1.3-protocol_v1.5.pdf http://www.anemos-plus.eu/images/pubs/deliverables/aplus.deliverable_d1.3-protocol_v1.5.pdf

Gensler, Andr and Sick, Bernhard and Vogt, Stephan. (2016). A Review of Deterministic Error Scores and Normalization Techniques for Power Forecasting Algorithms. 10.1109/SSCI.2016.7849848. Online access: https://ieeexplore.ieee.org/document/7849848

Jensen, T., Fowler, T., Brown, B. Lazo, J., Haupt S.E. (2016), Metrics for evaluation of solar energy forecasts, NCAR Technical Note NCAR/TN-527+STR. Online available: http://opensky.ucar.edu/islandora/object/technotes:538 Nielsen, H.A., Nielsen, T.S., Madsen, H., San Isidro Pindado, M.J., Marti, I.: Optimal combination of wind power forecasts, Wind Energy 10(5), pp. 471-482, 2007. Online:
https://onlinelibrary.wiley.com/doi/abs/10.1002/we.237

Frías Paredes, L., Stoffels, N., Statistical analysis of wind power and prediction errors for selected test areas, EU 7th Framework project Safewind, Deliverable Dp-7.1. Statistical analysis of wind power and prediction errors for selected test areas. Online available:
http://www.safewind.eu/images/Articles/Deliverables/swind.deliverable_dp-7.1_statistical_analysis_v1.6.pdf
http://www.safewind.eu/images/Articles/Deliverables/swind.deliverable_dp-7.1_statistical_analysis_v1.6.pdf

Sánchez, I.: Adaptive combination of forecasts with application to wind energy. International Journal of Forecasting 24(4), pp. 679693, 2008. Online:
https://doi.org/10.1016/j.ijforecast.2008.08.008

Uncertainty Forecast Information

Yan, J; Möhrlen, C.; Göçmen, T.; Kelly, M.; Wessel, A.; Giebel, G. Uncovering wind power forecasting uncertainty origins and development through the whole modelling chain. Submitted to Sustainable and Renewable Energy Reviews, Sept. 2021.

Möhrlen, C.; Bessa, R.J.; Fleischhut, N.; Decision-Making Experiment under Wind Power Forecast Uncertainty. Submitted to Meteorological Applications, Aug. 2021.

S. E. Haupt, M. Garcia Casado, M. Davidson, J. Dobschinski, P. Du, M. Lange, T. Miller, C. Möhrlen, A. Motley, R. Pestana, and J. Zack. The use of probabilistic forecasts: Applying them in theory and practice.IEEE Power and Energy Magazine, 17(6):4657, 2019.

Bessa, R.J.; Möhrlen, C.; Fundel, V.; Siefert, M.; Browell, J.; Haglund El Gaidi, S.;

Hodge, B.-M.; Cali, U.; Kariniotakis, G. Towards Improved Understanding of the Applicability of Uncertainty Forecasts in the Electric Power Industry. Energies 2017, 10, 1402. Online access: `https://www.mdpi.com/1996-1073/10/9/1400`

Dobschinski, J., Bessa, R., Du, P., Geisler, K., Haupt, S.-E., Lange, M., Möhrlen, C., Nakafuji, D., Rodriguez, M. d.l.T., Uncertainty Forecasting in a Nutshell: Prediction Models Designed to Prevent Significant Errors, IEEE Power and Energy Magazine, vol. 15, no. 6, pp. 40-49, Nov.-Dec. 2017. doi: 10.1109/ MPE.2017.2729100.

C. Möhrlen and J.U. J/orgensen, Chapter 3: The Role of Ensemble Forecasting in Integrating Renewables into Power Systems: From Theory to Real-Time Applications, Integration of Large-Scale Renewable Energy into Bulk Power Systems - From Planning to Operation, Editors: Du, Pengwei, Baldick, Ross, Tuohy, Aidan (Eds.), pp 79-134.

Möhrlen, C., Bessa, R., Giebel, G., Jørgensen, J.U., Uncertainty Forecasting Practices for the Next Generation Power System, Proc. 16th International Workshop on Large-Scale Integration of Wind Power into Power Systems as well as on Transmission Networks for Offshore Wind Power Plants,Germany, of 16th International Workshop on the Integration of Solar Power into Power Systems, 2017. Online available: `www.ieawindforecasting.dk/publications`

Related Data Standards and Projects

European Network of Transmission System Operators for Electricity (ENTSO-E), Weather Process and Energy Prognosis Implamentation Guide:
https://www.entsoe.eu/Documents/EDI/Library/cim$_{b}ased/Weather_{I}G_{V}1r2.pdf$

European Network of Transmission System Operators for Electricity (ENTSO-E), Common Information Model, https://www.entsoe.eu/digital/common-information-model/cim-for-energy-markets/

C. W. Hansen et al., "The Solar Forecast Arbiter: An Open Source Evaluation Framework for Solar Forecasting," 2019 IEEE 46th Photovoltaic Specialists Conference (PVSC), 2019, pp. 2452-2457, doi: 10.1109/PVSC40753.2019.8980713

US Department of Energy Solar Forecast Arbiter Project - https://solarforecastarbiter.org

IEA Task 43 Work Package 4 - Digital Wind Resource Assessment Data Standard - https://www.ieawindtask43.org/work-package-4-digital-wra

Presentations:

Möhrlen, C., Collier, C. ,Zack , J., Lerner, J.A., Can Benchmarks and Trials Help Develop new Operational Tools for Balancing Wind Power?, Proc. of 16th International Workshop on the Integration of Solar Power into Power Systems, Paper WIW-126, Berlin, Germany, 2017. Online access: www.ieawindforecasting.dk/publications

RECOMMENDED PRACTICES FOR THE IMPLEMENTATION OF WIND POWER FORECASTING SOLUTIONS
Corinna Möhrlen, John Zack, Jeff Lerner, Aidan Tuohy,ăJethro Browell, Jakob W. Messner, Craig Collier, Gregor Giebel
Part 1: FORECAST SOLUTION SELECTION PROCESS
Part 2 and 3: DESIGNING AND EXECUTING FORECASTING BENCHMARKS AND TRIALS AND EVALUATION OF FORECAST SOLUTIONS

Möhrlen, C., Zack, J., Lerner, J.A., Tuohy, A., Browell, J., Messner, J.W., Collier, C., Giebel, G. RECOMMENDED PRACTICES FOR THE IMPLEMENTATION OF WIND POWER FORECASTING SOLUTIONS - Part 1: FORECAST SOLUTION SELECTION PROCESS and Part 2 and 3: DESIGNING AND EXECUTING FORECASTING BENCHMARKS AND TRIALS AND EVALUATION OF FORECAST SOLUTIONS, Proc. of 17th International Workshop on the Integration of Solar Power into Power Systems, Paper SIW-126, Berlin, Germany, 2018.
Online access: www.ieawindforecasting.dk/publications

GLOSSARY AND ABBREVIATIONS

Ensemble Forecasting:
Ensemble forecasts are sets of different forecast scenarios, which provide an objective way of evaluating the range of possibilities and probabilities in a (weather or weather related) forecast.

Probabilistic Forecast:
General description of defining the uncertainty of a forecast with objective methods. These can be ensemble forecasts, probability of exceedance forecasts, or other forms of measures of uncertainty derived by statistical models.

Quantile:
A quantile is the value below which the observations/forecasts fall with a certain probability when divided into equal-sized, adjacent, subgroups.

Quartile:
quantiles that divide the distribution into four equal parts.

Percentile:
Percentiles are quantiles where this probability is given as a percentage (0-100) rather than a number between 0 and 1.

Decile:
Quantiles that divide a distribution into 10 equal parts.

Median:
the 2nd quantile, 50th percentile or 5th decile, i.e. the value, where the distribution has equally many values above and below that value.

Abbreviations

The following abbreviations are used in this document:

FSP	Forecast service provider
NWP	Numerical Weather Prediction
EPS	Ensemble Prediction System
NDA	Non-disclosure Agreement
RFI	Request for Information
RFP	Request for Proposals
TSO	Transmission system operators
ISO	Independent system operator

Appendix A

CLARIFICATION QUESTIONS FOR FORECAST SOLUTIONS

In order to define the objectives and possible solutions for a forecasting system, it is recommended to follow an overall structure: 1. Describe your situation In this process, it is imperative to describe exactly those processes, where you need forecasting in the future. Here it is essential to get the different departments involved, especially the IT department. The more accurate you can describe the situation you need to solve with forecasting (e.g. which IT restrictions, limitations and methods for data exchange exist, current or future challenges, etc.), the more straight forward it will be to (1) ask questions to the vendors regarding forecasting methodology, but also (2) get clarity of the involved processes enabling forecasting.

Ask Questions to the vendors

The questions to the vendors should be of technical character regarding forecast methodology, but also on available data exchange methodologies, required input data for the models and system support.

TYPICAL QUESTIONS FOR PART 1

- Processes: Which processes require forecasting
- Data
  - How will the data flow internally be solved: data storage, data exchange, data availability ?
  - Which data do we collect that may assist the forecaster to improve accuracy

- Data Formats:
  - Which formats are required for applications, data exchange and storage ?
- Applications
  - Who/which department will use the forecasts, are new applications required to make use of the forecasts ?
- Education
  - Is it required to train staff in how to use forecasts ?
- Policies
  - Are there policies, political or legal restrictions to be aware of when exchanging data with a forecaster ?

TYPICAL QUESTIONS FOR PART 2

The following are typical questions to get some overview of what is state-of-the-art in forecasting for renewables and what products are available on the market for a specific purpose.

- Describe the methodology you will use when generating forecast for (wind|solar|)
- How many years of experience do you have in this specific area or related areas
- Required data fields for the forecasting model for the trial
- Time scales and IT requirements for the data for the forecasting model
- Required data for vendor's model, if adopted and used live
- Applicable charges for a trial with vendor
- Vendors forecast model forecast horizons

Appendix B

TYPICAL RFI QUESTIONS PRIOR TO OR IN AN RFP

- **Methodology**
  - What unique services can you provide that may address our needs ?
  - What input weather data is used
  - What methodology is used for power generation for the long-term (>1 days ahead) and short-term forecasting (0...24h).
  - Can uncertainty forecasts or probability bands be provided ?1 If yes, which methodology is being used.
  - What are the minimum requirements for wind farm site data?
  - Can a Graphical User Interface be provided to visualise forecasts ? If yes, please describe it in detail (e.g. platform dependence, user management, in-house installation or web-based).
- **Service Level**
  - What kind of service level does the provider offer (ticket system, personal support, call center, online support, etc.)
  - What kind of service level is recommended for the specific service.
  - Does the provider have outage recovery guarantee
- **Contract and Pricing**
  - What are restrictions and preferences on the pricing structure of your service (e.g. price per park, per MW, per parameter, per time increment)?
  - What restrictions/preferences does the provider have in responding to RFPs ?
- **Experience**

- Can the vendor provide minimum of 3 examples of your work that is applicable to our needs (e.g. forecast accuracy, references, methodology)?
- Does the company have significant market shares in the market/area of business
- Additionally, can your company supply products or information that you consider relevant for us when setting out an RFP ?

Appendix C

Application Examples for Use of Probabilistic Uncertainty Forecasts

C.0.1 Example of the Graphical Visualization of an Operational Dynamic Reserve Prediction System at a System Operator

Figure C.1 shows an example of the graphical visualization of an operational dynamic reserve prediction system at a system operator, where operators requested to have various intervals in order to evaluate which of the intervals was economically or from a system security aspect the better choice in a given situation.

The reserve requirement is built with a NWP ensemble approach where the ensemble spread is related and calibrated to the expected forecast error of wind power, demand and an estimated cross-border exchange requirement. The mean of the computed reserve requirement is scaled to zero and the possible positive and negative requirement intervals are plotted in form of 4 percentiles up and down, respectively.

The red circles indicate areas, where the requirements would have been higher than what e.g. a P20-P80 interval would have covered, if this was the uncertainty range the operators would have requested. It also illustrates why the operators wanted to be "aware" of such outliers, even if they may not have pre-allocated according to the outer ranges or boundaries.

Figure C.1: Example of the graphical visualization an operational dynamic reserve prediction at a system operator. The reserve requirement with the

C.0.2 High-Speed shut down warning system

In a typical area where high-speed shut down is a challenge for the grid security, the development of low pressure systems are frequent and the variability of the wind resources are relatively high. Thus, an alert system concerning high-speed shutdown of wind power must be established based on probabilities computed from a probabilistic prediction system that can take the spatial and temporal scales into consideration in order to capture the temporal evolution and spatial scale of such low pressure systems that contain wind speeds leading to large scale shut-down of wind farms.

This can for example be provided by a physical approach based on a NWP ensemble that ideally contains all extreme values inherent in the approach without the requirement of statistical training. Alternative solutions may exist from statistical approaches (see **??** by employing an extreme event analysis to a statistical ensemble of type 2. This is due to the requirement that such forecasts must be able to provide probabilities of extreme events, where each "forecast member" provides a valid and consistent scenario of the event. The probabilities need to be suitable solutions for a decision process. They can be computed for very critical and less critical events, dependent on the end-users requirements.

Figure C.2: Example of a high-speed shut-down example, where within 5 days 2 extreme events showed up in the risk index of the system (upper graph), showing the probability of occurrence in terms of probability ranges as percentiles P10...P90 of a high speed-speed shutdown. The second graph shows the 5-day wind power forecast inclusive uncertainty intervals as percentile bands P10...P90 and the observations (black dotted line). The red circles indicate the time frame in which the alarms were relevant.

Figure C.2 shows an example of a real-time setup of such a high speed shut down warning system. The example exhibits 2 events. The first graph shows the risk index in probability space of a high-speed shutdown event to occur. The second graph shows the wind power forecast with uncertainties inclusive the observations (black dotted line) of what happened. From the upper graph, the operator can directly read out the following:

- Case 1 at 26. January:
  - 10% probability of 50% shutdown
  - 8% probability of 90% shutdown
  - 90% probability of 5% shutdown

- Case 2 on 31. January:
  - 10% probability of 50% shutdown
  - 15% probability of 90% shutdown
  - 90% probability of 10% shutdown

The reality is shown by the observations in the lower graph of figure C.2, where it can be seen that the first case's peak value was 35% high-speed shut-down and the second case exhibited a peak value of 45% of high-speed shut-down.

Practical experience from evaluating high-speed shutdown events and discussing the alert system with the operators, showed that it is absolutely crucial that the operators understand the alerts and are capable of checking and verifying themselves in a graphical way, what they may receive as written alert. Therefore, the impact of a false alarm needs to be evaluated, decided upon and documented in the design phase, so that the operators have a clear reference system to relate an alert to. Technically, the frequency of the alert generation should be adjusted to:

a lead time of the alert

b change of severity level since previous alert

c initial and valid week day and time of the day

d severity of the event computed from a ramp-rate perspective and actions required

e the need and possibility to call back and/or revert actions

The strategy of issuing an alert should include (1) issuing of every alert according to a simple scheme and (2) reduction of the amount of alerts to a level that prevents that critical alerts are not accidentally overlooked.

It was also found that the Use of sliding interval from 23-25m/s was an important introduction into the design to ensure that tje warning is issued **before** the event.

RECOMMENDED PRACTICES FOR THE IMPLEMENTATION OF RENEWABLE ENERGY FORECASTING SOLUTIONS

- Part 2: DESIGNING AND EXECUTING FORECASTING BENCHMARKS AND TRIALS -

2. EDITION 2021

Draft for Review by Stakeholders prior to submission to the Executive Committee of the International Energy Agency Implementing Agreement

Prepared in 2021 as part of the IEA Wind Task 36, WP 2.1.

Document Version: 2.0
October 12, 2021

Contents

Preface

This 2^nd edition of this recommended practice document is the result of a collaborative work that has been edited by the undersigning authors in alignment with many discussions at project meetings, workshops and personal communication with colleagues, stakeholders and other interested persons throughout the phase 2 of the IEA Wind Task 36 (2018-2021) as part of workpackage 2.1.
The editors want to thank everybody that has been part of the meetings, workshops and sessions and contributed in the discussions, provided feedback or other input throughout the past two and a half years.

IEA Wind Task 36, October 12, 2021

Editors and Authors:
Dr. Corinna Möhrlen (WEPROG, Denmark) <com@weprog.com>
Dr. John Zack (UL AWS Truepower, USA) <john.zack@ul.com>

Contributing Authors:
Dr. Craig Collier (Energy Forecasting Solutions, USA)
Dr. Aidan Tuohy, EPRI, USA
Dr. Jakob W. Messner (MeteoServe Wetterdienst, Austria)
Dr. Jeffrey Lerner (ENFOR, Denmark)
Dr. Jethro Browell (University of Glasgow, United Kingdom)
Dr. Justin Sharp (Sharply Focused, USA)
Mikkel Westenholz (ENFOR, Denmark)

Supported by:
Operating Agent Dr. Gregor Giebel (Danish Technical Univerity, DTU Wind, Denmark)

Chapter 1

Background and Objectives

1.1 BEFORE YOU START READING

This is the second part of a series of three "recommended practices" documents that deal with the development and operation of forecasting solutions. This document "Execution of Benchmarks and Trials" deals with the configuration and steps for carrying out a benchmark or trial of different forecasting solutions prior to selection.
The first part "Forecast Solution Selection Process" deals with the selection and background information necessary to collect and evaluate when developing or renewing a forecasting solution. The third part "Forecast Evaluation" provides information and guidelines regarding effective evaluation of forecasts, forecast solutions and benchmarks and trials. If your main interest is in selecting a forecasting solution or verifying the quality of your forecast solution, please move on to part 1 or part 3 of this recommended practices guideline, respectively.

1.2 BACKGROUND

The effectiveness of forecasts in reducing the variability management costs of power generation from wind and solar plants is dependent upon both the accuracy of the forecasts and the ability to effectively use the forecast information in the user's decision-making process. Therefore, there is considerable motivation for stakeholders to try to obtain the most effective forecast information as input to their respective decision tools.
This document is intended to provide guidance to stakeholders on a primary mechanism that has been used extensively in the past years to assess the accuracy of potential forecasting solutions: benchmarks and trials.

This guideline focuses on the key elements to carry out a successful trial or benchmark and on typical pitfalls. It will also provide recommendations as to when it is beneficial or too risky or expensive in terms of resources to carry out a trial or benchmark.

1.3 DEFINITIONS

The two main terms and concepts "trial and benchmark" that are used in this recommended practice shall be defined in the following. Note, the focus has been on forecasting processes in the power industry and the definition may not have a completely general character to be applied to other areas of business. Additionally, it should be noted that "forecasting trials and benchmarks" will be abbreviated with "t/b" throughout this document for simplicity.

1.3.1 Renewable Energy Forecast Trial

A "renewable energy forecast trial" is in this document defined as an exercise conducted to test the features and quality of a renewable energy forecast such as wind or solar power. This may include one or more participants and is normally conducted by a private company for commercial purposes. A trial is a subset of a Renewable Energy Forecast Benchmark.

1.3.2 Renewable Energy Forecast Benchmark

A "renewable energy forecast benchmark" is in this document defined as an exercise conducted to determine the features and quality of a renewable energy forecast such as wind or solar power. The exercise is normally conducted by an institution or their agent and multiple participants, including private industry forecast providers or applied research academics.

1.4 Objectives

The guidelines and best practices recommendations are based on years of industry experience and intended to achieve maximum benefit and efficiency for all parties involved in such benchmark or trial exercises. The entity conducting a trial or benchmark taking the recommendations provided in this guideline into consideration will have the following benefits:

1. Being able to evaluate, which of a set of forecast solutions and forecast service providers (FSP) fits best the need, specific situation and operational setup

2. Short term internal cost savings, by running an efficient t/b
3. Long term cost savings of forecast services, by following the trial standards and thereby help reduce the costs for all involved parties

In the discussion of the process of obtaining the best possible forecasting solution, there are a number of terms and concepts that are used. Several of the key terms and concepts are defined in the following.
Note, these definitions are kept as general as possible with a focus on forecasting processes in the power industry and may not have such a completely general character to be applied to other areas of business.

Chapter 2

INITIAL CONSIDERATIONS

> ***Key Points***
> *This section is targeted to the task of engaging a forecast service provider (FSP) and how to navigate through the vast amount of information.*

2.1 Deciding whether to CONDUCT a Trial or Benchmark

The most important initial consideration when planning a forecasting trial or benchmark (t/b) is to be clear about the desired outcome.
The following tables provide information about the benefits and drawbacks of conducting a t/b as a key part of the selection process. Before a decision is made to conduct a t/b it is recommended to go through these tables and determine, if the effort is warranted.
A possibly attractive alternative approach for a forecast user that wishes to evaluate a set of forecast solutions for their ability to meet the user's needs is to engage an independent trial administrator. An experienced and knowledgeable administrator can act as a neutral third party and advocate for both the vendors and the end-users in the design and execution of a t/b and the evaluation and interpretation of the results. Such an arrangement builds trust in the process among all parties.
An effective administrator can take the requirements from the user and ensure they are realistically incorporated into the trial design. There obviously is a cost to engage such an administrator, but it may be more cost effective for the user and generate more reliable information for the user's decision-making process.

2.2 Benefits of Trials and Benchmarks

Table 2.1: Decision support table for situations in which trials/benchmarks are determined to be beneficial

Situation	**Benefit**
Real-time trial for an entire portfolio	High cost but information gain is greater and more representative; provides the best estimate of the error level and which solution/FSP is best for the target applications
Real-time trial for a selected number of sites	Lower cost but still a substantial information gain if sites are well selected; provides a reasonable idea about the error level and a good indication of which solution/FSP fits is best for the target applications
Retrospective benchmark with historic data for a specific time period separate from a supplied training data set	Low cost: In multi-FSP systems, the error level of an additional FSP is secondary, while the correlation with other FSPs determines whether the additional FSP improves the overall error of a multi-FSP composite forecast
Blind forecast without historic measurements	Test to get an indication of the accuracy of forecasts from an FSP in the upstart phase of a project, where no historical data are available. Excludes statistical methods, which need historical data. An inexpensive way to get an indication of forecast accuracy for larger portfolios (> 500MW), where measurement data handling is complex. NOTE: There is an inherent risk that the result may be random and FSP use different methods for blind forecasting and forecasting with measurement data. See also Table 2.2 for limitations of this approach.

2.3 Limitations with Trials and Benchmarks

Table 2.2: Decision support table for situations in which trials/benchmarks are determined to contain limitations and a t/b is not recommended.

Situation	**Limitation**	**Recommendation**
Finding best service provider for large portfolio (> 1000MW) distributed over a large area	Trial for entire portfolio is expensive for client and FSP in terms of time and resources.	Limiting scope of trial limits representativeness of results for entire portfolio. RFI and RFP in which FSP's methods are evaluated and the use of an incentive scheme in the contract terms provides more security of performance than a limited trial.
Finding best service provider for a medium sized portfolio (500MW< X < 1000MW) over a limited area	Trial for entire portfolio is expensive for client and service provider in terms of time and resources.	Limiting scope of trial limits representativeness of results for entire portfolio. RFP in which FSP's methods are evaluated. Design of a system that enables an easy change of FSP and use if an incentive scheme is more a more cost effective approach than a trial.
Finding best service provider for small sized portfolio (< 500MW)	Trial for entire portfolio usually requires significant staff resources for about 6 months	Trial is feasible, but expensive. Difficult to achieve significance on target variable in comparison to required costs and expenses – trial costs makes solution more expensive. Less expensive to setup an incentive scheme and a system where the FSPs can be changed relatively easily.
Finding best service provider for micro portfolio (< 100MW) or single plants	Cost of a trial with many parties can easily be higher than the cost of a 1-year forecasting contract.	Time for a trial can delay operational forecast utilization by up to 1 year! Select FSP based on an evaluation of methods and experience.

Situation	**Limitation**	**Recommendation**
Design a system that enables an easy change of FSP and use an incentive scheme for FSP performance	Power marketing	Best score difficult to define, as sale of energy is also dependent on market conditions and a statistical forecast performance score such as RMSE or MAE does not reflect the best marketing strategy More efficient and timely to perform back test of historical forecasts combined with historical prices, or make a strategic choice with an performance incentive.
Market share of FSP in a specific power market is high	FSP monopolies in a specific power market mean that forecast errors are correlated and hence increase balancing costs.	Ask about the market share of a provider and do not choose one with a share > 30% as the only provider!
Blind forecasting, i.e. no historic measurements data available	Without measurements the value of a trial is very limited due to the significant improvement from statistically training forecasts and the importance of recent data for intra-day forecasts	Evaluation can only be meaningfully done for day- ahead or longer forecasts.
Some FSP may us different methods for forecasting with and without historic data (statistical methods need historical data to function!)	Results are limited to testing quality on upstart phase of new projects, where no historical data exist (see also Table 1).	For single sites, the benefits of training are so large (>50% of error reduction at times) that blind forecasting is not recommended. For larger portfolios it can provide an indication of quality - for physical conversion methods only!

2.4 Time lines and Forecast periods in a Trial or Benchmark

Time lines and forecast periods need to be set strictly in a trial or benchmark in order to achieve a fair, transparent and representative exercise.
The following time lines should be considered:

1. Start and stop dates of the t/b must be fixed
2. Start and stop dates must be the same for all FSPs
3. Pre-trial setup and test dates for IT infrastructure (including any required security protocols) for trial must be specified and enforced
4. Delivery times of forecasts must be set and enforced
5. Forecasts for periods with missing forecasts from one FSP must be excluded for all FSPs

2.5 1-PAGE "Cheat sheet" Checklist

The following checklist is provided to help trial organizers save time, apply best practices, and avoid common pitfalls when designing and executing forecast trials. It has been compiled by leading forecast vendors and researchers with many years experience.

Forecast Trial Checklist

--Preparation--

☐ Determine outcomes / objectives
☐ Consult expert with experience
☐ Establish timeline and winning criteria
☐ Decide on live or retrospective trial
☐ If live trial with datafeed, begin datafeed setup
☐ Gather metadata (use IEA checklist spreadsheet)
☐ Determine if adequately resourced to carry out
☐ Obtain historical data
☐ Invite forecast service providers
☐ Distribute historical and meta-data
☐ Finalize datafeed configuration (if applicable)
☐ Allow two weeks Q&A prior to start
☐ Begin

--During Trial--

☐ Develop validation report
☐ Check interim results
☐ Provide interim results (if no live data being provided)
☐ End

--Post Trial--

☐ Provide final results
☐ Notify winner(s)
☐ Contract with winner(s)
☐ Start Service

Figure 2.1: "Cheat sheet" Checklist

Chapter 3

Deterministic Trials and Benchmarks

***Key Points**:*
Deterministic trials have become an established way to test different forecast venders or test the compatibility and benefits of combining various forecast methods in the forecast solution selection process. Such trials are complicated and the required ressounces to conduct fair, transparent and representative results are often unceres-timated. In order to generate valuable results, such trials need to follow a specific structure, which is characterised by three phases:

- *Phase 1: Preparation*
- *Phase 2: During Trial*
- *Phase 3: Post Trial*

These three main phases of a trial exercise, preparation ahead of the trial, actions during the trial, and post-trial follow up are described in detail in the following.

3.1 Phase 1: PREPARATION

The time required for the pre-trial preparation is significant and should not be underestimated to insure a successful outcome. If the operatorr of the trial has no experience in renewable energy forecasting or running a t/b, it would be prudent to contact an experienced individual, organization or forecast provider to obtain feedback on what can reasonably be accomplished given the target time line and objectives. Part 1 of this recommended practice contains a decision support path that may be useful for determining the proper course of action.

3.1.1 Key Considerations in the Preparation Phase

Once the objectives of the t/b are known (see Section 1.1 Background and 1.2 Objectives), there are some key decisions to be made that will play a major role in determining the complexity of the trial. They are:
(1) Choice of forecast horizon Are forecast horizons less than 6 hours operationally important? If the answer is "no", establishing a live data feed may not be necessary. Although there are advantages of running a trial with a live data feed, it is one of the most time consuming aspects of trial preparation. Are forecast lead times greater than "day-ahead" operationally important? If the answer is no, this will reduce the volumes of data that need to be processed saving time and resources. If many lead times are of operational importance, consider that the performance of different providers will likely vary across lead times, therefore, different lead times, e.g. hour-ahead, day-ahead and week-ahead, should be evaluated separately.
(2) Weather conditions for the exercise: Will the benchmark take place during periods of more difficult to predict weather conditions that reflect the organization's difficulties in handling renewable generation, e.g. windy or cloudy periods? The answer here should be "Yes" to insure the sample size of harder-to-forecast events is sufficient. If the answer is "No", the trial operator should strongly consider doing a retrospective forecast (also known as "backcast") that includes the types of conditions that are critical for the user's application.
(3) Historical data/observations for the exercise: For locations in which there are significant seasonal differences in weather conditions and the associated renewable generation levels and variability, it is best to provide 12 months or more of historical data from the target generation facilities to the FSPs for the purpose of training their forecast models. However, if it is not feasible to make this amount of data available or if the target location does not exhibit much seasonal variation, most FSPs can typically train their forecast models reasonably well with 3-6 months of on-site historical observations.
It should be noted that advanced machine learning methods often exhibit significantly greater performance improvement over less sophisticated methods as the training sample size increases. Thus, FSPs that employ the latest and most advanced machine learning prediction tools may not be able to demonstrate the ultimate value of their approaches, if only short historical data sets are provided. If 6-12 months of data are not available, the trial operator might consider another location or conduct a longer trial on the order of 4-6 months to monitor forecast improvements over time as more data becomes available to the FSPs to improve the quality of the training of their prediction models.
In general it is recommended that the t/b operator should provide a dataset of the typical length that is available data for the application that is the target of the t/b. If more historical data is available for a t/b than in the typical application, care should be taken in the evaluation of methods, as e.g. machine learning methods

might outperform e.g. physical methods in the trial, but perform worse in the real application due to the benefits associated with the longer data sets.
(4) Representativeness: Is the benchmark location representative from a wind-climatology perspective of the scope of locations for which the operator will ultimately require operational forecast services? That is, the trial operator should select a location that is needed for subsequent forecasting or a location with a similar climatology. Operators should also be aware of the randomness of forecast performance on single locations, if a large area with many sites is the target. It should be noted that forecast performance exhibits a significant "aggregation effect". That is the magnitude and patterns of forecast errors vary substantially depending on the size and composition of the forecast target entity. Thus, the characteristics of forecast errors for an individual turbine, a single wind park and a portfolio of wind parks will typically be quite different and the forecast evaluator should be very careful when inferring forecast performance characteristics from one scale of aggregation (e.g. a single wind park) to a different scale (e.g. a geographically diverse portfolio of wind parks) (see also part 3 of this recommended practice for more details on evaluation methods).
(5) Metrics: Are the metrics that will be used to evaluate the forecasts meaningful to the success of my project? There are a wide variety of well-documented error metrics that penalize forecast errors differently. For example, root mean squared error penalizes large errors more than small errors. It is important to choose a metric, or set of metrics, that reflects the value of an improved forecast to the user's application and can discriminate between different forecast solutions. Please refer to part 3 of this recommended practice for details on metric selection.

3.1.2 Metadata Gathering in the Preparation Phase

Details of the forecast trial, such as location and capacity of the target generator, are required by all FSPs and comprise the trial Metadata. Appendix A "Metadata Checklist" provides the information that is typically needed by FSPs for participation in a trial and is designed to be used as a spreadsheet form that is completed during the preparation phase of a t/b. This should also include the desired format (filename and content) of the forecasts you'll be comparing. The best way to communicate the forecast file format to multiple FSPs is to provide an example file.

3.1.3 Historical Data Gathering in the Preparation Phase

On-site observations of power production or the renewable resource (e.g., irradiance or wind speed at hub height) are critical for helping the FSPs statistically "train" their forecast models and thus reduce error and bias in the forecasts. Good quality data is critical. "Good quality" means that the data does not, for example,

contain many gaps or unrepresentative values. Curtailed power data should be accompanied by plant availability or a curtailment flag.
Data time intervals should be regular and there should be a clear documentation of the units, how the observations were averaged, the time zone of the data, and whether there's a shift in time due to daylight savings time. Appendix A of this document has a concise list of the necessary historical data attributes required to efficiently start a t/b.

3.1.4 IT/Data Considerations in the Preparation Phase

Most organisations have constraints on the amount of IT resources available for a t/b. Therefore, it is best to plan ahead or keep the sending and receiving of data very simple. The primary IT issue is typically the selection and setup of data formats and communication protocols that will be used for the t/b operator to send data to the FSPs and for the FSPs to send forecasts to a platform designated by the t/b operator.

Data formats:
There are many possibilities for data formats, which range from a simple text file with comma separated variables (CSV) to more sophisticated XML or openAPI formats. Similarly, there are a wide range of communication protocols that can be used. These range from the relatively simple Secure Shell File Transfer Protocol (SFTP) to more sophisticated web service or API structures. The more sophisticated structures have advantages and there are many IT companies and resources that support these structures but they almost unavoidably increase the complexity of the setup.
Unless adequate IT resources or knowledge are available for all participants (especially the operator) it is recommended that simple data formats and communication resources be employed for a t/b. This typically means the use of the CSV data format and an SFTP data communications protocol.

Live trial considerations:
If a live trial is planned (most common), but real-time data will not be made available to the FSPs, then a place for each FSP to send forecast files will need to be setup. One of the metrics that is often used to evaluate an FSP is the timeliness of forecast delivery. In this case, it is important that a mechanism to verify the time of delivery be established. If real-time data is provided by the t/b conductor, it is typically easiest to create a common password-protected file server directory from which FSPs can download the data via a protocol such as SFTP. Another approach is to use SFTP to push data files to each FSP. This typically requires more effort, especially for the t/b operator.

Historical data can be provided to FSPs in the same data format via the same communication protocol. However, it often requires a SCADA engineer or expert on third party software to extract the historical data for the SCADA (or other) data archive.

Legal Agreements:
Another often-overlooked data-related issue is the legal agreements required to disseminate data from possibly multiple data provider entities (e.g. the wind facility owners/operators) to multiple data user entities (e.g. the FSPs in the t/b). This may be relatively simple in cases in which the user (such as a generator fleet operator) owns all the data and is willing to make it available for the t/b with few restrictions. However, it be a very complex and time consuming process in cases in which the user (e.g. a system operator) does not own the data and merely serves as a conduit from the multiple data owners with different data dissemination restrictions to the data users.
In such cases, the process of formulating and executing the required legal documents (such as non-disclosure agreements (NDAs)) can cause substantial delays in the initiation of a t/b and perhaps even change its scope.
See Appendix B for example formats in csv and xml.

3.1.5 Communication in the Preparation Phase

Transparency:
Anonymising the FSPs for all communication is considered a best practice as it ensures transparency of the available information, promotes competition and entry from smaller FSPs trying to become more established in the industry. Communication via email therefore should always be consistent with blind copies to all FSPs.
Consistency:
Consistent in this context means always sending and sharing emails with the same group of FSP users. Common information sharing engenders trust and the perception of fairness in the benchmark or trial process. In the preparation phase, it is not uncommon that the FSPs will have questions that could affect how the trial is conducted.
For this reason, it is recommended to have a 2-week question and answer period before the official start date to allow FSP participants to ask questions that then can be answered in a living document that contains all questions and answers up to the present time. All participants should be notified whenever this document is updated.
Frequency:
The importance of frequent and clear communication cannot be overstated when

conducting a t/b. Not only will the t/b operator receive the most accurate forecasts, it will make it much easier the next time a t/b is executed to gage the state-of-the-art in forecasting technologies and features.

3.1.6 Test run in the Preparation Phase

It is recommended to that a minimum of one-week is allocated for a test period before the official start date of the t/b to identify and remove any technical issues that could invalidate forecast results. This helps to improve the likelihood that all results can be included in the final validation calculations without the need for omitting the first part of the t/b.

3.2 Phase 2: DURING BENCHMARK/TRIAL

Verification & Validation Report preparation Often the most successful forecast provider is one that can show steady improvement over time. Providing an interim validation report will not only prepare the trial operator for the final validation report but will give important feedback to the FSPs – not only throughout the trial or benchmark, but also in the daily operations.
validation Strategy:
Part 3 of this recommended practice provides information about validation and verification that incentivices the FSP, where it is beneficial for the end-user.
Verification strategy:
In **??**, a verification and validation strategy is described that emphasizes that verification of validation code is an essential part of a validation. In the case of a trial or benchmark, it is recommended that the verification strategy and the input data for the validation is shared with the FSP. In that way, the verification code is tested as recommended by **??** and there is transparency on the results. If the FSPs result differs from the end-user's result, the errors can be detected and solved.

3.2.1 Communication during the T/B

In a well-designed t/b, most of the communication between the trial operator and FSPs should be during the pre-trial period. However, issues often arise especially during a live trial with a real-time data feed. It may be helpful to all t/b participants to establish an open forum during the first part of the live t/b period (e.g. the first 2 weeks) to provide a way to effectively and uniformly resolve all issues early in the t/b period However, it is strongly recommended that if any attributes of the t/b are changed at any point during the live part of the t/b, the changes should be communicated to all participants immediately as they might require action on the FSP's part.

Examples might include: changing the forecast validation metric, if there are unreported outages that should be omitted for future model trainings, or if the location of the data feed or forecast file destination has changed. It should be emphasized that all communications related to the t/b should be distributed to all FSPs without exception. Additional communication with individual FSPs (including forecast incumbents) can be interpreted as bias on the part of the operator of the t/b and in some cases may actually bias the t/b result due to information that impacts forecast design, production or delivery not being equally available to all FSPs.

3.2.2 Forecast Validation and Reporting during the T/B

Forecast validation reports are often compiled during the t/b. With forecast data coming in at regular intervals, the t/b operator has real data to feed into the validation report. If the t/b has a duration of several months (i.e., >3 months), it is recommended to provide at least one interim report to FSPs that include anonymized results from all FSPs. This benefits the trial operator as errors in the evaluation process or the report generation can be flagged earlier and ways to make the report generation more efficient can be realized. The interim report benefits the FSPs as course-corrections can be made during the t/b to improve the forecasts.
If there are several FSPs participating, efficiencies can be realized by automating part or most of the validation metrics especially as the forecast file format should be the same from all FSPs.

3.3 PHASE 3: POST TRIAL OR BENCHMARK

The post trial phase is an important aspect of the t/b because FSP selection will likely occur during this phase based on the criteria set out at the start of the t/b. (see recommended practices part 1 on "evaluation of services and decision support").

3.3.1 Communication at the end of the T/B

If the trial operator hasn't already done so, an email should be sent within a week before the end date of the t/b to alert FSPs that the end of the trial is near and to communicate the timeline for sharing results and re-iterate the specifications of the FSP selection process.

3.3.2 Forecast Validation and Reporting at the end of the T/B

If an interim report was provided during the trial, then the final report can either be an updated version of the validation report expressing the bulk metrics or appended month-by-month forecast validation results. For transparency and to promote further forecast improvements, it is recommended that the t/b operator share the anonymized forecast results from each FSP at the time-interval frequency that forecasts were being made at (e.g., hourly). This will help FSPs discover where forecasts are similar or different from the competition which may spawn improved methodologies.

Chapter 4

Probabilistic Trials and Benchmarks

Key Points:
Testing, verification and validation of probabilistic forecast methods and forecast solutions need to be handled fundamentally different than for deterministic methods. The latter can be aggregated, combined and compared and has in the past been mostly used to foster improvements on basic statistic metrics.
Probabilistic forecast methods on the other hand deal with uncertainties in the forecast process chain and can be compared, but not - in a straight forward or easy way – aggregated or combined.
While it is possible to make trials and benchmarks with probabilistic solutions, the verification has to be done by method and in most cases with event based verification metrics such as:

(i) "Event evaluation"
Examples are categorial event analysis with contingency tables, critical success index (CSI), measuring the ratio of correct event forecasts to the total number of forecasted and observed events.

(ii) "Cost or Loss Functions"
Such functions measure the sensitivity of a user's application to the forecast error which for example can be wether the observations is within the forecasted uncertainty spread, uncertainty measures such as quantiles or precentiles.

Probabilistic trials and benchmarks are in this section defined as a verification and validation of probabilistic forecast solutions for specific applications in the energy industry.

Because the probabilistic forecast solution have very distinct and different attributes in comparison to deterministic forecast solutions, their testing also needs specific requirements and attention. For example, if an end-user request quantiles or percentiles of a specific variable to be delivered by different forecast vendors, it is **not possible** to:

- aggregate variable's quantiles or percentiles
- aggregate location's quantiles or percentiles
- average quantiles or percentiles over time

The phases described in section 3 are also valid for probabilistic forecast solutions. However, both the testing and the evaluation phase need slightly different considerations.

4.1 Phase 1: PREPARATION

4.2 Phase 2: DURING BENCHMARK/TRIAL

4.3 PHASE 3: POST TRIAL OR BENCHMARK

Chapter 5

BEST PRACTICES

Although there are many different ways that a t/b may be conducted, there are some common elements of a successful t/b that provide the t/b operator with the best forecast solution and the participants with useful knowledge of where their forecast ranks among the competition.
The following are some selected best practice recommendations:

1. A clear purpose for the t/b exercise
2. Pre-defined and explicit accuracy metrics and solution selection criteria
3. A clear time line (start/end dates, selection announcement, contract award)
4. Anonymized forecast results. Ask FSP's approval to share results. This helps FSPs find ways to improve their forecast accuracy and see their shortcomings.
5. Question & answer period before benchmark period begins (1-2 weeks)
6. Sufficient time allocated for testing the transfer of data between participant(s) and operator
7. Prompt communication to participants regarding any changes or answers to questions that arise
8. Consistent forecast file format requested of all - example file sent to all
9. Consistent data formats (both observations and forecast files) ideally as close to (if not identical to) what the trial operator needs, once contract is executed.
10. Providing the same historical and project metadata to all participants
11. Allocation of sufficient resources by the t/b conductor to furnish data and perform validation

12. PITFALLS TO AVOID The following list describes a few common mistakes and how to avoid them in the design, setup and execution of a forecast t/b. The consequences of errors and omissions in trials are often underestimated. However, if results are not representative, the efforts that have gone into a t/b can effectively be wasted. Some of these common pitfalls can be expensive to the operator because they result in placing the operator in a position of making a decision without having truly objective and representative information to base it on.

    (a) Poor Communication
    All FSPs should receive the same information. Answers to questions should be shared with all FSPs. Fairness, and perception of fairness, are important when running and evaluating the results of trials.

    (b) Unreliable Validation Results
    Don't compare forecasts from two different power plants or from different time periods. Forecast performance will vary depending on location and specific time periods. Only forecasts for the same period and location/power plant/portfolio should be compared.

    (c) Examples of Bad Design
        i. A trial with 1 month length during a low-wind month
        ii. No on-site observations shared with forecast providers
        iii. Hour-ahead forecasts initiated from once a day data update
        iv. Data only processed in batches or at the end of a real-time trial – this is an invitation for cheating to the FSPs. In most cases, there will be some that use the opportunity to do so

    (d) Examples of Missing or Non-communicated Data
        i. daylight savings time changes are not specified
        ii. data time stamp represents interval beginning or ending not specified
        iii. plant capacity of historical data differs from present capacity
        iv. data about curtailment and maintenance outages not provided

    (e) Possibility of Cheating
    In any type of competition, cheating is a reality. If there are not taken precautions, results may be biased and decisions are taken upon incorrect results. It is recommended that the possibility of cheating is considered with seriousness and avoided, where possible.

Typical situations, where cheating is being observed are:

- Forecast t/b being carried out for a period of time for which FSPs are given data. Recommendation: separate historical data from t/b period.

- Forecast t/b being carried out for a period of time for which FSPs are given data. Recommendation: separate historical data from t/b period.

- if there is one or more incumbent FSP with a longer history of data, this should be taken into consideration in the evaluation, as such an FSP may not be able or willing to modify forecast models for the purpose of being "comparable" in a t/b. Recommendation: see limitations in Table 2 and part 3 of this recommended practice.

Other observed situations, where cheating is happening is:

- Missing forecasts: FSP leave out "difficult situations" as missing forecasts are often not penalized. However, missing data may bias "average" forecast metrics, potentially resulting in the formulation of incorrect conclusions. Recommendation: remove dates where forecasts are missing for one FSP for all FSPs

- If delivered forecasts from a FSP as part of a live trial are not downloaded, moved or copied in accordance with the operational process being simulated, and certainly before the time period being forecast, FSPs can potentially renew forecasts with high accuracy due to fresher information being available. Recommendation: Such an omission should not be underestimated and care taken for the evaluation.

REFERENCE MATERIAL

J. Kehler and D. McCrank, Integration of wind power into Albertas electric system and market operation, Proc. of IEEE Power and Energy Society General Meeting - Conversion and Delivery of Electrical Energy in the 21st Century, Pittsburgh, PA, pp. 1-6. doi: 10.1109/PES.2008.4596824, 2008.

E. Lannoye, A. Tuohy, J. Sharp, V. Von Schamm, W. Callender, L.Aguirre, Solar Power Forecasting Trials and Trial Design: Experience from Texas, Proc. of 5th International Workshop on the Integration of Solar Power into Power Systems,, Brussels, Belgium, ISBN: 978-3-9816549-2-9, 2016.

E. Lannoye, A Tuohy, J Sharp, and W Hobbs, Anonymous Solar Forecasting Trial Outcomes, Lessons learned and trial recommendations, Proc. of 7th International Workshop on the Integration of Solar Power into Power Systems, Paper SIW-126, Berlin, Germany, 2017.

C. Möhrlen, C. Collier , J. Zack , J. Lerner , Can Benchmarks and Trials Help Develop new Operational Tools for Balancing Wind Power?, Proc. of 7th International Workshop on the Integration of Solar Power into Power Systems, Paper SIW-126, Berlin, Germany, 2017. Online access: `http://download.weprog.com/WIW2017-292_moehrlen_et-al_v1.pdf`.

Bessa, R.J.; Möhrlen, C.; Fundel, V.; Siefert, M.; Browell, J.; Haglund El Gaidi, S.; Hodge, B.-M.; Cali, U.; Kariniotakis, G. Towards Improved Understanding of the Applicability of Uncertainty Forecasts in the Electric Power Industry. Energies 2017, 10, 1402. Online access: `http://www.mdpi.com/1996-1073/10/9/1402`

Conference Papers

Corinna Möhrlen, Recommended Practices for the Implementation of Wind Power Forecasting Solutions Part 1: Forecast Solution Selection Process, Proc. 17th International Workshop on Large-Scale Integration of Wind Power into Power Systems as well as on Transmission Networks for Offshore Wind Power Plant, Stockholm, Sweden, October 17.-19, 2018. Online Access: http://www.ieawindpowerforecasting.dk/publications

Corinna Möhrlen, John Zack, Jeff Lerner, Aidan Tuohy, Jethro Browell, Jakob W. Messner, Craig Collier, Gregor Giebel

Part 23: DESIGNING AND EXECUTING FORECASTING BENCHMARKS AND TRIALS AND EVALUATION OF FORECAST SOLUTIONS, Proc. 17th International Workshop on Large-Scale Integration of Wind Power into Power Systems as well as on Transmission Networks for Offshore Wind Power Plant, Stockholm, Sweden, October 17.-19, 2018
Online Access: http://www.ieawindpowerforecasting.dk/publications

C. Möhrlen, R. Bessa, Understanding Uncertainty: the difficult move from a deterministic to a probabilistic world, Proc. 17th International Workshop on Large-Scale Integration of Wind Power into Power Systems as well as on Transmission Networks for Offshore Wind Power Plant, Stockholm, Sweden, October 17.-19, 2018
Online Access: http://www.ieawindpowerforecasting.dk/publications

C. Möhrlen, R. Bessa, G. Giebel, J. Jørgensen,G. Giebel, Uncertainty Forecasting Practices for the Next Generation Power System, Proc. 16th Int. Workshop on Large-Scale Integration of Wind Power into Power Systems as well as on Transmission Networks for Offshore Wind Power Plant, Berlin (DE), 26-29 June 2017.
Online Access: http://www.ieawindpowerforecasting.dk/publications
C. Möhrlen (WEPROG, Denmark), C. Collier (DNV GL, USA), J. Zack (AWS Truepower, USA), J. Lerner (Vaisala, USA) Can Benchmarks and Trials Help Develop new Operational Tools for Balancing Wind Power?, Proc. 16th Workshop on Large-Scale Integration of Wind Power into Power Systems as well as on Transmission Networks for Offshore Wind Power Plant , Berlin (DE), 26-29 June 2017.
Online Access: http://www.ieawindpowerforecasting.dk/publications

Presentations

C. Möhrlen, C. Collier , J. Zack , J. Lerner , Can Benchmarks and Trials Help Develop new Operational Tools for Balancing Wind Power?, Proc. of 7th International Workshop on the Integration of Solar Power into Power Systems, Paper SIW-126, Berlin, Germany, 2017. Online access:
http://download.weprog.com/WIW17-292_MOEHRLEN-ET-AL_PRESENTATION_20171028.pdf

J. W. Zack (2017), Wind and solar forecasting trials experience: do's and don'ts, Part 2 UVIG 2017 Forecasting Workshop, Atlanta (US), 21-22 June 2017

Online access: http://www.ieawindforecasting.dk/-/media/Sites/IEA_task_36/Publications/forecast_trials_session_4_uvig2017_jzack.ashx?la=da

C. Collier (2017), Why Do Forecast Trials Often Fail to Answer the Questions for which End-Users Need Answers: A Forecaster's Point of View UVIG Forecasting Workshop, Atlanta (US), 21-22 June 2017. Online access: http://www.ieawindforecasting.dk/-/media/Sites/IEA_task_36/Publications/forecast_trials_session_4_uvig2017_ccollier.ashx?la=da

T. Maupin (2017), Wind and Solar Forecasting Trials: Do's and Don'ts, Part 1 Best practices. UVIG 2017 Forecasting Workshop, Atlanta (US), 21-22 June 2017. Online access: http://www.ieawindforecasting.dk/-/media/Sites/IEA_task_36/Publications/forecast_trials_session_4_uvig2017_ccollier.ashx?la=da

GLOSSARY AND ABBREVIATIONS

Ensemble Forecasting:
Ensemble forecasts are sets of different forecast scenarios, which provide an objective way of evaluating the range of possibilities and probabilities in a (weather or weather related) forecast.

Probabilistic Forecast:
General description of defining the uncertainty of a forecast with objective methods. These can be ensemble forecasts, probability of exceedance forecasts, or other forms of measures of uncertainty derived by statistical models.

Quantile:
A quantile is the value below which the observations/forecasts fall with a certain probability when divided into equal-sized, adjacent, subgroups.

Quartile:
quantiles that divide the distribution into four equal parts.

Percentile:
Percentiles are quantiles where this probability is given as a percentage (0-100) rather than a number between 0 and 1.

Decile:
Quantiles that divide a distribution into 10 equal parts.

Median:
the 2nd quantile, 50th percentile or 5th decile, i.e. the value, where the distribution has equally many values above and below that value.

Abbreviations

The following abbreviations are used in this document:

FSP	Forecast service provider
NWP	Numerical Weather Prediction
EPS	Ensemble Prediction System
NDA	Non-disclosure Agreement
RFI	Request for Information
RFP	Request for Proposals
TSO	Transmission system operators
ISO	Independent system operator

Appendix A

Metadata Checklist

The following checklist (Table A.1), when filled out, will greatly aid FSPs in configuring forecasts efficiently. Many of the essential questions relevant to benchmark and trial forecast model configuration are provided here.

Note that the following table is an example and may not contain all necessary information required for the FSP to setup a solution for your purpose. The table is meant to serve as a guideline and can be copied, but should be carefully adopted to the specific exercises before sending out to FSP with questions filled in. If this is done with care, it will expedite forecast configuration and save back and forth communication time.

Table A.1: Wind Power Forecast Trial Checklist

Metadata	**Input**
Name of site(s) as it should appear in datafile	
Name of site(s) as it should appear in datafile	
Latitude and longitude coordinates of sites	
Nameplate capacity of each site	
Will a graphical web tool be needed?	
Turbine make/model/rating	
Number of turbines	
Hub height of turbines	
Please attach suitable plant power curve	
Forecast output information	

Metadata	**Input**
Forecast output time intervals (e.g., 15-min, 1-hourly)	
Length of forecast required	
Timezone of forecast datafile	
Will local daylight savings time be needed?	
Forecast update frequency (e.g., once a day, every hour)	
Value of Forecast	
Which variables will be forecasted and validated?	
Which forecast horizons are being validated?	
Which metrics are being used to gage forecast performance?	
List criteria for determining winning forecast provider	
Will results be shared as a report? Will results be anonymized?	
On what frequency will results be shared with forecast provider?	
Historical Data Checklist	
Is the data in UTC or local time?	
Is the data interval beginning or ending or instantaneous?	
What are the units of the data?	
If met tower histories being provided, indicate height of measurements.	
Realtime Data Checklist (if applicable)	
Is the data in UTC or local time?	
Is the data interval beginning or ending or instantaneous?	
What are the units of the data?	
Email and Telephone number of technical point of contact (POC)	
Email and Telephone of datafeed POC	
Name and email of users that need website access	
Person name and email that filled out this checklist	

Appendix B

Sample forecast file structures

Back and forth communication can sometimes delay the start of a trial or benchmark. One of these delays is getting the forecast file output format just right for the beginning of the trial.
Standardisation of the format will make the trial operators life much easier when time comes to validating forecasts. A best practice here is for the trial operator to use a format that is already in use or a format that has already proven to work in operations.

Plant Output	Acme Wind Farm	1.11.2017 4:00	1.11.2017 5:00	1.11.2017 6:00	1.11.2017 7:00
Power	MW	41.43	41.43	41.43	40.89
Windspeed	m/s	11	10	10	10
Time zone: Central European Summer Time (CEST)					
Intervals: hour ending					
Date time format: dd.mm.yyyy hh:mm (e.g., 06.08.1969 08:30)					

Figure B.1: Example forecast file with the first few fields.

B.0.1 XSD template example for forecasts and SCADA

The following are typical XSDs for forecasts and SCADA data in a b/t, usable also with WebServices

```xml
<?xml version="1.0" encoding="utf-8"?>
<xs:schema attributeFormDefault="unqualified" elementFormDefault="qualified"
xmlns:xs="http://www.w3.org/2001/XMLSchema">
  <xs:element name="WindForecast">
    <xs:complexType>
     <xs:attribute name="VendorCode" type="xs:string" use="required"/>
     <xs:attribute name="ImportTime" type="xs:dateTime" use="required"/>
     <xs:sequence>
       <xs:element name="CUSTOMER">
         <xs:complexType>
          <xs:attribute name="name" type="xs:string" use="required"/>
           <xs:sequence>
            <xs:element name="Forecast">
              <xs:complexType>
               <xs:attribute name="MWaggregated" type="xs:double" use="required"/>
               <xs:attribute name="time" type="xs:dateTime" use="required" />
               <xs:sequence>
                 <xs:element name="Probability">
                   <xs:complexType>
                     <xs:attribute name="P95" type="xs:double" use="required"/>
                     <xs:attribute name="P50" type="xs:double" use="required"/>
                     <xs:attribute name="P05" type="xs:double" use="required"/>
                     <xs:attribute name="max" type="xs:double" use="required"/>
                     <xs:attribute name="min" type="xs:double" use="required"/>
                    </xs:complexType>
                  </xs:element>
                 <xs:element name="WindFarms">
                 <xs:complexType>
               <xs:sequence>
                 <xs:element name="WindPark1">
                  <xs:complexType>
                      <xs:attribute name="id" type="xs:string" use="required"/>
                       <xs:attribute name="mw" type="xs:double" use="required"/>
                   </xs:complexType>
                   </xs:element>
               </xs:sequence>
                   </xs:complexType>
                   </xs:element>
               </xs:sequence>
                </xs:complexType>
              </xs:element>
            </xs:sequence>
          </xs:complexType>
        </xs:element>
      </xs:sequence>
    </xs:complexType>
  </xs:element>
</xs:schema>
```

B.0.2 XSD SCADA template for exchange of real-time measurements


```
<?xml version="1.0" encoding="utf-8"?>
<xs:schema attributeFormDefault="unqualified" elementFormDefault="qualified"
xmlns:xs="http://www.w3.org/2001/XMLSchema">
 <xs:element name="WindSCADA">
  <xs:complexType>
    <xs:sequence>
      <xs:element maxOccurs="unbounded" name="WindPark">
        <xs:complexType>
          <xs:attribute name="ID" type="xs:string" use="required"/>
          <xs:attribute name="Time" type="xs:dateTime" use="required"/>
          <xs:attribute name="Mw" type="xs:decimal" use="required"/>
          <xs:attribute name="Availabilty" type="xs:decimal" use=" optional"/>
          <xs:attribute name="CurrentActivePower" type="xs:decimal" use=" optional"/>
          <xs:attribute name="Curtailment" type="xs:string" use="optional"/>
          <xs:attribute name="WindSpeed" type="xs:decimal" use="optional"/>
          <xs:attribute name="WindDirection" type="xs:decimal" use="optional"/>
          <xs:attribute name="AirTemperature" type="xs:decimal" use="optional"/>
          <xs:attribute name="AirPressure" type="xs:decimal" use="optional"/>
        <xs:attribute name="Outage" type="xs:decimal" use="optional"/>
      </xs:complexType>
    </xs:element>
   </xs:sequence>
  </xs:complexType>
 </xs:element>
</xs:schema>
```

RECOMMENDED PRACTICES FOR THE IMPLEMENTATION OF RENEWABLE ENERGY FORECASTING SOLUTIONS

- Part 3: Forecast Solution Evaluation -

2. EDITION 2021

Draft for Review by Stakeholders prior to submission to the Executive Committee of the International Energy Agency Implementing Agreement

Prepared in 2021 as part of the IEA Wind Task 36, WP 2.1.

Document Version: 2.0
October 12, 2021

Contents

Preface

This recommended practice document is the result of a collaborative work that has been edited by the undersigning authors in alignment with many discussions at project meetings, workshops and personal communication with colleagues, stakeholders and other interested persons throughout the phase 1 (2016-2018) and the phase 2 (2019-2021) of the IEA Wind Task 36 as part of workpackage 2.1.

The editors want to thank participants, contributorscontributors and supporter of the meetings, workshops and sessions that contributed in the discussions, provided feedback or other input throughout the past 6 years.

IEA Wind Task 36, October 12, 2021

Editors and Authos:
Dr. Corinna Möhrlen (WEPROG) <com@weprog.com>
Dr. John Zack (UL AWS Truepower) <john.zack@ul.com>

Contributing Authors:
Dr. Jakob W. Messner (MeteoServe Wetterdienst, Austria)
Dr. Jethro Browell (University of Glasgow, UK)

Contributions:
Dr. Craig Collier (Energy Forecasting Solutions, USA)
Dr. Aidan Tuohy, (EPRI, USA)
Dr. Stephan Vogt (Fraunhofer Institute for Energy Economy and Energy System technology (IEE)), Germany

Supported by:
Operating Agent Dr. Gregor Giebel (Danish Technical Univerity, DTU Wind, Denmark)

Chapter 1

Background and Objectives

1.1 BEFORE YOU START READING

This is the third part of a series of three recommended practice documents that deal with the development and operation of forecasting solutions in the power market. The first part Forecast Solution Selection Process deals with the selection and background information necessary to collect and evaluate when developing or renewing a forecasting solution for the power market. The second part Design and Execution of Benchmarks and Trials, of the series deal with benchmarks and trials in order to test or evaluate different forecasting solutions against each other and the fit-for-purpose. The third part Forecast Solution Evaluation, which is the current document, provides information and guidelines regarding effective evaluation of forecasts, forecast solutions and benchmarks and trials.

1.2 Introduction

The evaluation of forecasts and forecast solutions is an obligation for any forecast provider as well as end-user of forecasts. It is important, because economically significant and business relevant decisions are often based on evaluation results. Therefore, it is crucial to design and outline forecast evaluations with this importance in mind, give this part the required attention and thereby ensure that results are significant, representative and relevant. Additionally, forecast skill and quality has to be understood and designed in the framework of forecast value in order to evaluate the quality of a forecast on the value it creates in the decision processes. This first edition of the recommended practices guideline focuses on a number of conceptual processes to introduce a framework for evaluation of wind and solar energy forecasting applications in the power industry. A comprehensive outline of forecast metrics is not part of this guideline. There are a number of very useful and comprehensive publications available (e.g. [1], [2], [3], [4]) which will also specifically be referenced. A state-of-the-art of forecast evaluation is also not part of this guidelines, as the process of standardization has only just started in the community. This topic will be covered in one of the next versions of this guideline.

This first version of the recommended practices guideline focuses on:

1. *Impact of forecast accuracy on application*
First, it's often difficult to define the forecast accuracy impact to the bottom line as forecasts are just one of many inputs. Second, trials or benchmarks often last longer than anticipated or too short to generate trustworthy results. Thus, the Forecast User is often under pressure to either wrap up the evaluation quickly or to produce meaningful results with too little data. As a consequence, average absolute or squared errors are employed due to their simplicity, even though they seldom reflect the quality and value of a forecast solution for the Forecast User's specific applications.

2. *Cost-Loss Relationship of forecasts*
A forecast that performs best in one metric is not necessarily the best in terms of other metrics. In other words, there exists no universal best evaluation metric. Using metrics that do not well reflect the relationship between forecast errors and the resulting cost in the Forecast User's application, can lead to misleading conclusions and non-optimal (possibly poor) decisions. Knowing the cost-loss relationship of their applications and to be able to select an appropriate evaluation metric accordingly is important. This becomes especially important as forecasting products are becoming more complex and the interconnection between errors and their associated costs more proportional. Apart from more meaningful evaluation results, knowledge of the cost-loss relationship also helps the forecast service provider to optimize forecasts and develop custom tailored forecast solutions for the intended application.

Evaluation of forecast solutions is a complex task and it is usually neither easy nor recommended to simplify the evaluation process. As a general recommendation, such a process needs to follow an evaluation paradigm with three principles for an evaluation to be:

1. **representative**
2. **significant**
3. **relevant**

How to setup an evaluation process and achieve these principles is the core of this recommended practices guideline.

In chapter 2 these three main principles are outlined and the general concept of evaluation uncertainty is explained as this should be the basis for any evaluation task. In chapter 3, the uncertainty of measurement data collection and reporting is explained as the second base principle of evaluation and verification tasks. If forecasts are evaluated against data that inherit errors, results may still show some significance, but may no longer be considered trustworthy, nor relevant and representative. In chapter 4 metrics for evaluation and verification will be conceptualized and categorized in order to provide an issue oriented guideline for the selection of metrics in a evaluation framework. The last chapter 5 introduces the concept of developing such an evaluation framework and provides practical information on how to maximize value of operational forecasts, how to evaluate benchmarks and trials and new forecasting techniques or developments. Lastly, recommendations are made for a number of practical use cases for power industry specific applications.

Chapter 2

Overview of Evaluation Uncertainty

Key Points
All performance evaluations of potential or ongoing forecast solutions have a degree of uncertainty, which is associated with the three attributes of the performance evaluation process: (1) representativeness, (2) significance and (3) relevance.
A carefully designed and implemented evaluation process that considers the key issues in each of these three attributes can minimize the uncertainty and yield the most meaningful results.
A disregard of these issues is likely to lead to uncertainty that is so high that the conclusions of the evaluation process are meaningless and therefore decisions based on the results are basically random.

Uncertainty is an inherent characteristic of the forecast evaluation process. The objective of the design and execution of a forecast evaluation procedure is to minimize the uncertainty and thereby reduce its impact on the decisions association with forecast selection or optimization. In order to minimize forecast evaluation uncertainty it is useful to understand the sources of uncertainty on the evaluation process.

The sources of forecast evaluation uncertainty can be linked to three key attributes of the evaluation process: (1) representativeness (2) significance and (3) relevance. If any one of these are not satisfactorily addressed, than an evaluation will not provide meaningful information to the forecast solution decision process and the resources employed in the trial or benchmark will essentially have been wasted. Unfortunately, it may not be obvious to the conductor of a forecast evaluation or the user of the information produced by an evaluation whether or not these three attributes have been satisfactorily addressed. This section will present an overview of the key issues associated with each attribute. Subsequent sections of this document will provide guidance on how to maximize the likelihood that each will be satisfactorily addressed.

2.1 Representativeness

Representativeness refers to the relationship between the results of a forecast performance evaluation and the performance that is ultimately obtained in the operational use of a forecast solution. It essentially addresses the question of whether or not the results of the evaluation are likely to be a good predictor of the actual forecast performance that will be achieved for an operational application. These are many factors that influence the ability of the evaluation results to be a good predictor of future operational performance.

Four of the most crucial factors are:

1. size and composition of the evaluation sample,
2. quality of the data from the forecast target sites,
3. the formulation and enforcement of rules governing the submission of forecasts (sometimes referred to as fairness),
4. availability of a complete and consistent set of evaluation procedure information to all evaluation participants (sometimes referred to as transparency)

2.1.1 Size and composition of the evaluation sample

The size of the evaluation sample is one of the most important representativeness factors. The size of the sample is a key factor in determining the extent to which the results are influenced by random variation, or noise, compared to true differences in forecast skill. The use of a small sample increases the probability that the conclusions from the evaluation will be due to noise (random and unrepresentative events) in the sample. For example, the occurrence of very unusual weather events for a few days in a short sample may dominate the evaluation results. The predictability of these events is often lower (i.e. higher forecast errors) than that of typical weather conditions. Therefore, a small sample that contains such very unusual events may lead to an overestimation of the typical magnitude of forecast errors. Conversely, a small sample that has no difficult-to-forecast events may lead to an underestimation of the typical forecast error. However, the performance of the forecasts under unusual weather conditions may be very important to the user's application and therefore an assessment of how different forecast systems perform under these conditions may be very valuable information to the solution selection process. Thus,there are two key points that the user should keep in mind when using a small evaluation sample. First, conclusions from a small sample will always be less reliable (i.e. more uncertain) than those from a larger sample. Second, the user should make an effort to understand the composition of the small sample by examining the relationship between the weather conditions in the sample relative to an estimate of the climatological (i.e. long-term) distribution (e.g. was the sample dominated by typical conditions or were there one or more atypical events?) for the site or region and also by examining the forecast error distributions (e.g. were almost all of the forecast error magnitudes clustered around the average magnitude or were there a significant number of outliers?) (see also 5.1.1, 4.1.4,).

That leads to the question of how large of a sample is adequate? A commonly used target sample size guideline when gathering data for statistical analysis is 30. If all the sample points are independent then a sample of 30 provides a reasonable adequate minimization that sampling noise will impact the conclusions. But the key phrase is that the sample data points must be independent (uncorrelated) for this guideline to be valid. However, weather processes are typically highly correlated over time periods of 3 to 4 days. This means that an adequate sample from a continuous evaluation period should be 3 to 4 times larger than 30 or in other words, 90 to 120.

The composition of an evaluation sample is another key issue. The composition should be constructed so that all significant modes of variation of the forecast variable (e.g. wind power production) are included in the evaluation sample. For example if there is a high wind season and a low wind season then both should have a representative number of cases in the evaluation sample. However, if this is not practical than at least there should at least be a representative sample of the most important modes for the application (e.g. high wind season when the speeds are near cutout or periods when the wind speed is frequently in the highly sensitive steeply sloped part of the turbine power curve).

2.1.2 Data Quality

The quality of the data used in the forecast evaluation process can be a major source of uncertainty. The data from the forecast target location is typically used for two purposes: (1) as training data for the statistical components of each forecast system and (2) evaluation of the forecast performance. If the data has many quality issues then the representativeness of both applications is compromised. The quality issues may include: (1) out of range or locked values, (2) biased values due to issues with measurement devices or location of measurement, (3) badly or not at all calibrated instruments and (4) values that are unrepresentative of meteorological conditions because of undocumented outages or curtailments. If a substantial of data with these issues is used is used in the evaluation process for either of the two purposes, the results will likely not be representative of the true skill of the forecasting solutions that are being evaluated.

2.1.3 Forecast Submission Control

A third important factor is the formulation and enforcement of rules for the submission of forecasts in the evaluation process. This is sometimes noted as a fairness issue and it is indeed an issue of fairness to the forecast providers who are typically competing to demonstrate the skill of their system and thereby obtain an award of a contract for their services. However, from the users perspective it is a representativeness issue. If it is possible to for some forecasting solution providers to provide forecasts with unrepresentative skill then the conclusions of the entire evaluation process are questionable. A couple of examples can illustrate this point. One is example is a situation in which there is no enforcement of the forecast delivery time. In this case it would be possible for a forecast provider to deliver forecasts at a later time (perhaps overwriting a forecast that was delivered at the required time) and use later data to add skill to their forecast or even wait until the outcome for the forecast period is known. Although one might think that such explicit cheating is not likely to occur in this type of technical evaluation, experience has indicated that it is not that uncommon if the situation enables its occurrence.

A second example, illustrate how the results might be manipulated without explicit cheating by taking advantage of loopholes in the rules. In this example the issue is that the evaluation protocol does specify any penalty for missing a forecast delivery and the evaluation metrics are simply computed on whatever forecasts are submitted by each provider. As a forecast provider it is not difficult to estimate the difficulty of each forecast period and to simply not deliver any forecasts during periods that are likely to be difficult and therefore prone to large errors. This is an excellent way to improve forecast performance scores. Of course, it makes the results unrepresentative of what is actually needed by the user. Often it is good performance during the difficult forecast periods that are most valuable to a user.

2.1.4 Process Information Dissemination

A fourth key factor is the availability of a complete and consistent set of information about the forecast evaluation process to all participants. Incomplete or inconsistent information distribution can occur in many ways. For example, one participant may ask a question and the reply is only provided to the participant who submitted the inquiry. This can contribute to apparent differences in forecast skill that are associated with true differences in the skills of the solution. This of course results in unrepresentative evaluation of the true differences in forecast skill among the solutions.

2.2 Significance

Significance refers to the ability to differentiate between performance differences that are due to noise (quasi-random processes) in the evaluation process and those that are due to meaningful differences in skill among forecast solutions. Performance differences that stem from noise have basically no meaning and will not represent the performance that a user will experience in a long-term operational application of a solution. *Real* performance differences on the other hand should be stable and should not change if an evaluation process is repeated, e.g., one year later. A certain degree of noise is inevitable in every evaluation task but both, minimization of noise and awareness of the uncertainty it causes are crucial to base reliable decisions on the evaluation results.

As mentioned above, repeatability is a good practical indication of significance in evaluation results. The highest potential for achieving repeatability is the use of a representative evaluation sample. This means the sample should cover as many potential weather events, seasons, and perhaps forecast locations as possible. Otherwise, there is a high probability that the results will be different for features that are not well represented in the evaluation sample. Thus, significance is highly related to representativeness and very much depends on the evaluation sample size and composition.

2.2.1 Quantification of Uncertainty

In addition to noise minimization through the use of representative evaluation data sets, it is also very useful to quantify the significance (i.e. the uncertainty) of the evaluation results. Quantification of the uncertainty is important for decision making. For example, if a number of forecast solutions are evaluated with a specified metric, but their differences are much smaller than the uncertainty in the result due to e.g. measurement uncertainty, the meaning of their ranking is actually very limited and should not be used for important decisions.

2.2.1.1 Method 1: Repeating the evaluation task

The simplest approach to estimate evaluation uncertainty would be to repeat the evaluation task several times on different data sets. This approach is often effective, because the variation or uncertainty of the evaluation results is typically attributable largely to their dependence on the evaluation data set and therefore results often vary among different evaluation data sets. However, since evaluation data sets are usually very limited, this is often not a feasible approach.

2.2.1.2 Method 2: Bootstrap Resampling

A simple alternative method is to simulate different data sets, through the use of bootstrap resampling process. In this approach an evaluation data set of the same length as the original data set is drawn from the original data set with replacement and the evaluation results are derived on this set. By repeating this "N" times, "N" different evaluation results become available and their range can be seen as the evaluation uncertainty. Alternatively, parametric testing can also provide information on the significance of evaluation results. Typically two sample paired t-tests applied on the sets of error measures for each event provide a good estimate of the significance of the results. Diebold et al. [5] proposed a variation of this t-test to account for temporal correlations in the data and can therefore provide a more accurate significance quantification. Messner et al. [6] also describes different parametric testing or bootstrap resampling approaches that can be employed to quantify the evaluation uncertainty.

If it is found, that the forecast that is identified as the "best" an evaluation process does not exhibit significantly better performance than some of the other benchmark participants, the final selection of forecast solutions should only consider differences among forecast solutions that are significant.

2.3 Relevance

Relevance refers to the degree of alignment between the evaluation metrics used for an evaluation and the true sensitivity of a users application(s) to forecast error. If these two items are not well aligned then even though an evaluation process is representative and the results show significant differences among solutions, the evaluation results may not be a relevant basis for selecting the best solution for the application. There are a number of issues related to the relevance factor.

1. Best Performance Metric
   First, the selection of the best metric may be complex and difficult. The ideal approach is to formulate a cost function that transforms forecast error to the application-related consequences of those errors. This could a monetary implication or it might be another type of consequence (for example a reliability metric for grid operations). However, if it is not feasible to do this, another approach is to use a matrix of performance metrics that measure a range of forecast performance attributes.

2. Multiple Performance Metrics
   If there is a range of forecast performance attributes that are relevant to a users application, it most likely will not be possible to optimize a single forecast to achieve optimal performance for all of the relevant metrics. In that case, the best solution is to obtain multiple forecasts with each being optimized for a specific application and its associated metric.

3. Multiple Forecast Solutions
   Another type of issue arises when the user intends to employ multiple (N) forecast solutions and create a composite forecast from the information provided by each individual forecast. In this case it may be tempting to select the best N performing forecasts in the evaluation according to the metric or metrics identified as most relevant by the user. However, that is not the best way to get the most relevant answer for the multiple provider scenario. In that case the desired answer is to select the N forecasts that provide the best composite forecast. This may not be the set of N forecasts that individually perform the best. It is the set of forecasts that best complement each other. For example, the two best forecasts according to a metric such as the RMSE may be highly correlated and provide essentially the same information. In that case, a forecast solution with a higher (worse) RMSE may be less correlated with the lowest RMSE forecast and therefore be a better complement to that forecast.

Chapter 3

Measurement Data processing and Control

Key Points

- *Measurements from the forecast target facilities are crucial for the forecast production and evaluation process and therefore much attention should be given to how data is collected, communicated and quality controlled*
- *Collection and reporting of measurement data requires strict rules and formats, as well as IT communication standards in order to maximize its value in the forecasting process; standards and methods for collecting and reporting data are available from multiple sources referenced in this section*
- *An effective quality control process is essential since bad data can seriously degrade forecast performance; standard quality maintenance and control procedures have been documented and some are noted in this section*

In any evaluation the measurements or observations are alpha and omega for trustworthy results. For this reason, this section is dedicated to the importance of data collection, verification and the identification of the measurement uncertainty. In the evaluation of wind power forecasts, power data is most important but also meteorological measurements are often provided to the forecast providers as input to improve their forecast models. Furthermore, failure, service periods, curtailment and other disturbances in the power measurements can have significant impact on the results of an evaluation. The following section deal with these aspects and provide recommendations for a correct handling of such data for the evaluation phase.

3.1 Uncertainty of instrumentation signals and measurements

All data are derived from different measurement devices and depending on the quality of these devices the measurements can deviate from the reality to a certain degree. In fact, measurement errors can never be avoided completely and can potentially affect the significance of evaluation results. Therefore, it is crucial to assure and maintain specific quality requirements for the measurement devices to obtain data of good quality and thus keep the measurement uncertainty to a low level. This will not only improve the significance of evaluation results but also assure an optimum quality of forecasts that use the measurements as input.

For power data, the measurement quality is usually ensured by existing grid code standards that are verified in the commissioning phase and are serviced as part of the turbines SCADA system maintenance.

Recommendations on minimum technical requirements is going beyond the scope of this recommended practice guideline. For anyone intending to collect and process bankable wind measurements, the following standards and guidelines provide a basis for the adaptation into real-time operational applications :

1. the International Electrotechnical Committee (IEC)
2. the International Energy Agency (IEA)
3. the International Network for Harmonised and Recognised Wind Energy Measurement (MEASNET)
4. United States Environmental Protection Agency (EPA)

If these requirements are fulfilled, the measurement error is usually negligible compared to other sources of uncertainty in the evaluation procedure.

- For *relevant* evaluation results, minimum standards for measurement data precision and quality have to be ensured and maintained.

3.2 Measurement data reporting and collection

Once wind farms are operational and the production data are measured it is important to collect, store and report them properly, which requires strict rules and formats, as well as IT communication standards. Standard protocols for collecting and reporting power data are usually enforced by jurisdictional grid codes. There are however a number of aspects that are not covered in the grid codes that are essential for verification or evaluation of forecasting tools. This section will discuss the main aspects to be considered for any measurement data collection and archiving. In the following we limit the description for the purpose of verification or evaluation of forecasts in a real-time operational framework or a forecast test framework.

3.2.1 Non-weather related production reductions

Raw power production data contains a number of non-weather related reductions that need consideration in the collection or archiving of measurement data, such as

- failure of turbines in a wind park (availability)
- scheduled and non-scheduled maintenance
- curtailment
- reductions due to environmental constraints (noise, birds, ...)

The so-called Net to Grid signal is often disturbed by such technical constraints that are usually not part of the wind power forecasting task. Therefore, to evaluate the actual forecast quality such events have to be filtered in the evaluation. Especially in the case of curtailment, the forecast user needs to decide whether the target parameter is the real power production or available power. If it is the latter, data with curtailment should be removed from the evaluation data set, because errors are not meaningful for the forecast performance, unless the curtailments are predicted as well.

- To receive *relevant* results, remove events from the evaluation data set that are effected by non-weather related production constrains unless these are to be predicted as well.

3.2.2 Aggregation of measurement data in time and space

Often, temporally or spatially aggregated data (averages, sums) are more useful in power applications than instantaneous signals. The aggregation level, or if no aggregation over time is carried out, for example, if hourly values are provided that are not hourly averages of higher resolution data, but instantaneous values taken at the start of the hour, this should be communicated to the forecast provider to assure optimum forecast performance for the intended application. Furthermore, it is strongly recommended to aggregate the measurement data according to the intended applications before comparing, analysing and verifying forecasts. Otherwise, the evaluation results might not be relevant for the forecast user.

When aggregating measurement data over parks, regions, control zones or other aggregation levels, it is important to consider non-weather related events as discussed in Section 3.2.1. In particular

- Non-reporting generation units
- IT communication failures or corrupt signals

have to be identified and reported and the aggregated data should be normalized accordingly. Such failures are impossible to predict by the forecast vendor and should therefore not be part of the evaluation process.

- For *relevant* results, average the measurement data over a time frame that is also useful for the intended application.
- For *representative* results, non-weather related events should be identified and the aggregated signals normalized accordingly.

3.3 Measurement data processing and archiving

In any real-time environment, measurements should be delivered as is, but flagged, if they are considered wrong (1) at the logger level and (2) after a quality control before employing measurements in a forecast process.

Archiving data is dependent on the way the further processing of the data is planned. In most cases, it is useful to archive data in a database. There are many different structures of data bases available today. Such structural decisions are out of the scope of this guideline. Nevertheless, there are general considerations when planning and designing a database for operational data. While measurements are available only at one specific time, forecast data have overlapping time periods and need to be separated from measurement data. At the design level it is necessary to consider the following aspects.

1. single or multiple time points per measurement signal in database
2. flagging at each data point and
   (a) possibility to overwrite corrupt data in database
   (b) possibility to add correct data point in database
   (c) knowledge of time averaging level of data signal
3. single or multiple measurement points per wind farm
4. ability to expand and upscale the database: expansion with increasing number of measurement points/production units
5. importance of access to historical data

The database dimensions and setup of tables has to take such decisions and requirements into consideration.

3.4 Quality assurance and quality control

Quality of data is a crucial parameter for any real-time forecasting system. If the data that real-time forecasts are based on are corrupt or misleading, the result can be worse than not having measurements or observations at all. Therefore, any real-time system using measurements needs a quality control mechanism to discard bad data. However, bad, corrupt or misleading data signals can have an almost unlimited amount of reasons, which means that specific limits, operating ranges and validity checks need to be established when dealing with observational data. While this is critical in real-time environments, the quality of measurement data in the verification phase is equally important. For example, if a wind power forecast is verified against observations from a wind farm and a maintenance schedule or a curtailment from the system operator is not filtered out or marked in the data time series, then the result may be bad for the wrong reason. Trustworthiness in data can only be a result of control and maintenance of both the hardware and the corresponding software and data archiving. The following sections outline the most important parts of a quality control that should be carried out regularly in real-time environments and prior to verification or evaluation exercises.

- For *relevant* evaluation results, the data has to be of high quality, and faulty or corrupt data has to be detected, flagged and disregarded for the evaluation process.

3.5 Filtering processes and Data Preparation

The filtering process and data preparation are crucial whenever dealing with measurements or observational data in the evaluation process. A number of parameter have been identified as being important to consider in the preparation phase of any verification/evaluation. Messner et al. [2018]) recommended the following requirements:

- **Data set representation and composition:**
  The selected data set should be representative for the application and forecasts should be compared with exactly the same data sets. Results of different locations, seasons, lead times etc. are in general not comparable. The composition should be constructed so that all significant modes of variation of the forecast variable (e.g. wind power production) are included in the evaluation sample. For example if there is a high wind season and a low wind season then both should have a representative number of cases in the evaluation sample. However, if this is not practical than at least there should at least be a representative sample of the most important modes for the application (e.g. high wind season when the speeds are near cutout or periods when the wind speed is frequently in the highly sensitive steeply sloped part of the turbine power curve).

- **Data set length:**
  The size of the evaluation sample is one of the most important representativeness and significance factors. The size of the sample is a key factor in determining to what extent results are influenced by random variation, or noise, compared to true predictive performance. The use of a small sample increases the probability that any conclusions reached from the evaluation will be due to noise (random and unrepresentative events) in the sample. For example, the occurrence of very unusual weather events for a few days in a short sample may dominate the evaluation results.

That leads to the question of how large of a sample is adequate? A commonly used target sample size guideline when gathering data for statistical analysis is 30. If all the sample points are independent then a sample of 30 provides a reasonable adequate minimization that sampling noise will impact the conclusions. But the key phrase is that the sample data points must be independent (uncorrelated) for this guideline to be valid. However, weather processes are typically highly correlated over time periods of 3 to 4 days. This means that an adequate sample from a continuous evaluation period should be 3 to 4 times larger than 30 or in other words, 90 to 120 days.

- **Data set consistency:**
  For a fair evaluation of a forecast, whether against other forecasts, measurements or persistence, it is very important to use the same data set to derive the evaluation results. If a certain forecast is not available for a specific time, this time has to be disregarded for all the other forecasts or persistence as well. Else, if forecasts are for example missing for days that are particularly difficult to predict, they would in total perform much better than forecasts that are expected to have high errors at these days. This also applies for curtailment data. It is important to evaluate a forecast against the weather related performance and remove all non-weather related impacts that are out of the forecasters control. Especially, if forecasts are evaluated against a persistence forecast, especially in minute- or hour scale forecasts, where models are adopted to measurements that may contain curtailment or failures due to turbine unavailability or communication issues, the corresponding persistence need to be computed accordingly. If this is not done, the forecast performance of the persistence will be overestimated and the performance of the forecast underestimated.

Chapter 4

Assessment of Forecast Performance

Key Points

- *All performance evaluations of potential or ongoing forecast solutions have a degree of uncertainty*
- *The uncertainty is associated with three attributes of the performance evaluation process evaluation process: (1) representativeness, (2) significance and (3) relevance*
- *A carefully designed and implemented evaluation process that considers the key issues in each of these three attributes can minimize the uncertainty and yield the most meaningful results*
- *A disregard of these issues is likely to lead to uncertainty and/or decisions based on unrepresentative information*

The relevance of different aspects of forecast performance depends on the user's application. For instance, one user may be concerned with the size of *typical* forecast errors, while another my only be concerned with the size and frequency of particularly large errors. There are a wide range of error metrics and verification methods available to forecast users, but their relationship to different attributes is not always clear. This chapter deals with the issues around evaluating specific attributes of forecast performance including metric selection, verification and the use of some specific metrics in forecast optimization.

4.1 Forecast Attributes at Metric Selection

Forecast users may be interested in either a single attribute, or a range of attributes. When evaluating forecasts to either track performance changes or discriminate between different forecasts, it is important to consider those attributes relevant to the forecasts intended use. Where a forecast is used in multiple applications there is not guarantee that these attributes will be aligned and it may be necessary to compromise or procure multiple forecast products. Selecting an appropriate metric, or set of metrics, is a key requirement in order to to produce a representative evaluation forecast performance which is relevant to the forecast's end use.

Quantitative evaluation methods are usually the core of the evaluation framework since they allow to objectively rank different forecast models. Typical choices of quantitative metrics are the (root) mean squared error, the mean absolute error or the quantile score (see [6] for details) for continuous forecasts and various quantities derived from contingency tables for forecasts of binary forecasts.

As emphasized in Section 5.1.4, the selection of metrics should be informed by the forecast user's intended use, and if a forecast is intended to be used for multiple applications, different basic metrics may be applied and merged into a weighted sum. Below, a range of forecast attributes and their relation to different evaluation metrics are discussed.

4.1.1 Typical Error Metrics

The most common error metrics used in the wind industry summarize 'typical' error by averaging the absolute value of errors, or squared errors, often normalized by installed capacity. Such metrics are simple to produce and give a high-level view of forecast performance. They give equal weighting to all errors included, which may be appropriate if the forecast is used to inform decisions at any time, as opposed to only when a particular event is predicted.

In energy trading, for example, the forecast is used to inform decisions for every trading period and the cost implication of a forecast error is usually proportional to the error. In this case, absolute value of the error is directly related to the forecast's end-use so mean squared error would not be as informative as mean absolute error.

However, average error metrics hide some information which may be of interest. For example, a forecast with mostly small errors and occasional large errors could return a similar mean score to one with all moderate errors. In some cases this may not be an issue, but some users may prefer to experience fewer large errors even if that means fewer small errors too.

Examples of typical error metrics are discussed in section 5.1 and especially in section 5.1.1.

4.1.2 Outlier/Extreme Error

Another important attribute is the prevalence of large errors. Some applications aim to prepare for large errors, such as managing reserve energy or other risk management. Calculating metrics based on historic errors is more challenging than for 'typical' errors as large errors are more effected by specific situations. It is recommended that different root causes of large errors are considered separately, and positive and that negative errors are treated separately.

For example, large errors at a single wind farm during a period of high wind speed may be caused by high speed shut down, but are unlikely if the wind speed is only just above rated. If considering aggregated production from multiple wind farms, large errors may be caused by wind speed forecast errors in the vicinity of large areas of concentrated capacity.

4.1.3 Empirical Error Distribution

The empirical distribution of past forecast errors gives a detailed picture of how frequent errors of different sizes have been. It can be useful to examine the distribution of errors for specific situations, such as when power was forecast to be 70±2%, as the shape of the distribution will depend on power level, particularly for individual wind farms.

4.1.4 Binary or Multi-criteria events

Some attributes of forecast performance relate to the prediction of events such as ramps (or particular rate and duration) which may span multiple lead-times and spatial scales. Furthermore, events typically have multiple attributes, such as timing and magnitude. Different attributes may be of more or less interest depending on the use case for the forecast. In these cases, average error metrics may not be representative of the desired forecast attribute.

For example, ramp rate may be of most importance to one user, whereas the timing or ramp magnitude may be of more importance to another. This effect is illustrated in Figure 4.1. Timing or *phase* errors are penalized heavily by mean absolute error so the forecast which best predicted both the ramp rate and magnitude appears worse by this measure. A similar principal applies to events such as the duration of high or low power periods. In general, average error metrics favour 'smooth' forecasts rather than those which capture the precise shape of specific events.

Contingency tables provide a framework for quantifying the prediction of categorical events, which can be defined to match the user's decision making process. For example, the user may define a particular ramp event with some tolerance for phase and level error and then evaluate the performance of a particular forecast solution at predicting such events. There are four possibilities for each predicted and/or actual event: a true positive (hit), true negative (correct negative), false positive (false alarm) or false negative (miss). From these, a range of metrics can be calculated and used for comparison with other forecast systems. Furthermore, if the cost implications of decisions based on the forecast are known (or can be estimated) then the relative value of forecasting systems may be calculated.

Examples on how to verify outliers can be found in section 5.1, and 5.5.2.1.

4.1.5 Prediction Intervals and Predictive Distributions

Prediction intervals may be supplied to provide situational awareness or to information or quantitative risk management. These intervals predict an upper and lower bound which the observation will fall between with some probability. It is therefore an important attribute that observations do in fact fall between the interval with the prescribed frequency. This property is call 'reliability' and can by evaluated by simply counting the frequency of observations

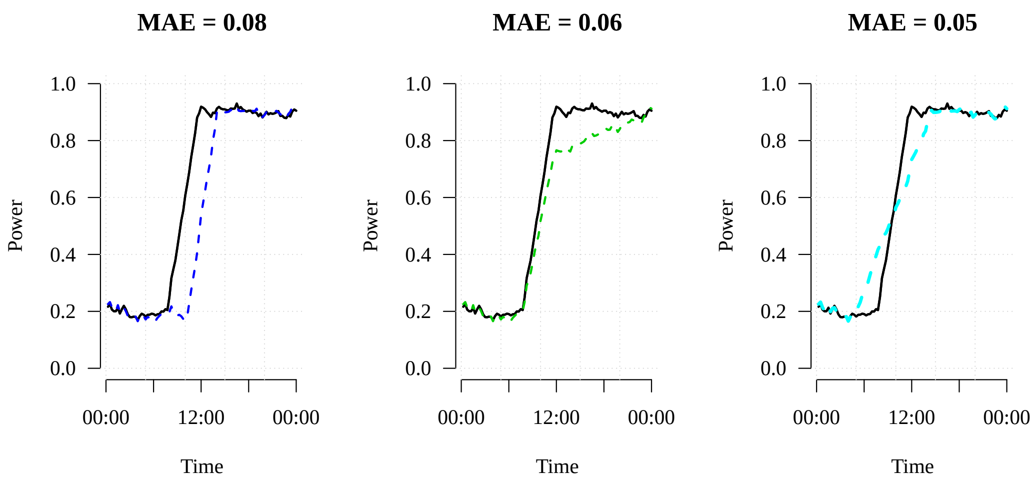


Figure 4.1: Examples of different types of ramp forecast error. Actual power is shown as solid black lines, forecasts are colored dashed lines. From left to right: phase or timing error, level error and ramp rate error. The mean absolute error (MAE) for each forecast is shown above the plots. Despite being the only forecast the correctly predict the ramp rate and duration, the forecast with a phase error has the largest MAE.

within and outside the interval. A more accurate forecasts with a narrower interval is said to be 'sharp' and provides greater confidence than a wide interval, but must be reliable in order to inform risk-based decision making. Therefore, prediction intervals should be evaluated following the principal of *sharpness subject to reliability*.

A predictive distribution is a smooth probability density function for the future value. It provides full information about probability of all possible value ranges rather than a single interval. In this case the principal of *sharpness subject to reliability* still applies but sharpness and reliability needs to be evaluated for a range of probability levels.

In quantitative decision making under uncertainty the optimal decision is often a *quantile*, i.e. the value that is forecast to be exceeded with some probability. For example, if the cost of taking precautionary action is C to protect against an uncertain adverse effect with potential loss L, then the precautionary action should be take in the probability of the adverse effect happening is greater than the cost-loss ratio C/L.

In applications of wind power forecasting, the adverse event could be exposure to imbalance costs, or holding insufficient energy reserves. In most cases, the values of C and L will be changing continuously and the decision maker will be aiming to select a future value of energy production which will be achieved with some probability $p = C/L$. Therefore, it is necessary to have access to the full predictive distribution in order to make an appropriate decision. Where the cost-loss ratio is known, the relative economic value of different forecasting systems can be calculated.

4.2 Probabilistic Forecast Assessment Methods

Probabilistic forecast evaluation is a complex topic. There are a number of classical metrics, just like for deterministic forecasts. However, the evaluation of probabilistic forecasts places greater importance on an end-user's knowledge of a cost function that provides a good indication of how well the forecast performance has met the requirements of the user's application. (see 5.1.4).

The considerations from chapter 2 on the performance evaluation and it's inherent uncertainty are even more important here. The three attributes *(1) representativeness, (2) significance* and *(3) relevance* are equally important to consider when setting up evaluation of probabilistic forecasts.

In same cases, it might be best to only use a graphical inspection of how well observations lie within forecast intervals. This can then be extended to an interval evaluation to provide objective values to the visual impression from the graph. This is similar to the "dichotomous event evaluation" described in 5.1.2.1 for predefined events. These scores can also be used for probabilistic/uncertainty forecasts, if the application is about how well the probabilistic forecasts or forecast intervals have captured the observations.

It is therefore important to follow the recommendations in the "best practice recommendations" 5 on how to built up a evaluation platform that reflects the purpose of the forecasts an provides and incentive to the forecast provider to match these criteria with the appropriate methods.

Therefore, the following description of metrics only provide a set of possible tools that can be used for the evaluation of probabilistic forecasts and the user must select the most appropriate set depending on the characteristics of the user's application and objectives of the forecast evaluation.

4.2.1 Brier Scores

The Brier score [8] is probably the most prominent and widely-used probabilistic forecast performance metric. It is a useful measure for a general assessment of the performance of probabilistic forecasts. However, the formulation of the basic Brier makes it suitable only for the evaluation of probabilistic forecasts of binary events (i.e. occurrence or non-occurrence of a defined event)

The Brier Score (BS) is the equivalent to the mean-squared error (MSE) for probabilistic forecasts with the same limitations as for deterministic forecasts. That means, the Brier Score is sensitive to the climatological frequency of events in the sense that the rarer an event, the easier it is to get a good BS without having any real skill. The BS is defined as,

$$BS = \frac{1}{N}\sum_{i=1}^{N}(f_i - o_i)^2 \tag{4.1}$$

where f_i is the *forecast probability* at time i, o is the observation at time i, and N is the

number of forecasts. The forecast probabilities range in value between 0 and 1, the observed values are either 0, if the event does not occur, or 1, if the event occurs.
Equation 4.1 is bound mathematically to values between 0 and 1. A lower Brier score, similar to the MSE, indicates greater accuracy. The maximum squared error is 1, because all squared errors will lie between 0 and 1. A *perfect accuracy* is reflected in the Brier score with 0, i.e. there is no difference between scores of an event and someones probabilistic predictions for those events. The opposite, i.e. a Brier score of 1, reflects perfect inaccuracy, which means that there are probabilities of 0 given to events that occur and probabilities of 1 to events that do not occur.

In order to gain further insight into the behavior of the Brier score, it can be decomposed algebraically into three components:

$$BS = CAL - RES + UNC \tag{4.2}$$

where, the CAL is the Calibration and is also sometimes referred to as the "reliability" , RES is the resolution, UNC is the uncertainty. The first components use the predicted probability to determine the performance of the forecast's ability to predict an event occurring with the provided probabilities as well as the observed frequency of that event, binned by forecast probability.

In [7], these three components are explained in a way that is easy to understand and relate to applications:

- *CAL* is a squared function of forecasted probability (f_p) and the mean probability ($_p$) and measures whether the forecasted values consistently represent the frequencies with which events occur (i.e., is the forecasted probability too large or too small on average?). For example, does the event occur 30% of the time when a forecast of 0.30 is issued? Specifically, CAL measures the difference between the actual frequency of occurrence and the forecast prediction. This is also referred to as the "reliability" of a probabilistic forecast.

- *RES* is a squared function of ($_p$) and () and measures how much the frequency of event occurrence varies among the forecasts. It measures the ability of the forecast to distinguish between event and non-event. For example, if the average frequency of event occurrence across all forecasts is 0.50, the relative frequency of occurrence ($_p$) should be much smaller for events, when the forecast is 0.10 (low likelihood of event) and much larger when the forecast probability is 0.90 (high likelihood of event). Higher RES scores indicate more skill and therefore appears in equation (4.1) with a negative sign. In the worst case, when the same probability (for example, the climatological probability) is always forecast, the resolution is zero.

- *UNC* is a function of () only and does not specifically measure how well the forecasts predict the event. Instead, UNC is an important measure of the difficulty of the forecasting situation. Large values of UNC (e.g., when the event is very rare) indicate that the forecasting situation is more difficult. It is inappropriate to compare forecasts for systems with significantly different UNC values.

In [7] a very useful list of specific questions that the Brier score answers have been listed:

1. ***Brier Score (BS)*** answers how accurate the probability forecasts are
2. ***Calibration (CAL)*** answers how well does the conditional relative frequency of occurrence of the event match a situation?
3. ***Resolution (RES)*** answers how well does the forecast separate events according to whether they occur or dont occur
4. ***Uncertainty UNC)*** answers how difficult/uncertain is the forecast situation

4.2.2 Ranked Probability (Skill) Score (RP(S)S)

The Ranked Probability Score (RPS) and Ranked probability Skill Score [9] is widely used for multi-category probability forecasts that have a magnitude order to them (such as generation forecasts). The RPS is the multi-category extension of the Brier score, and the "Skill" part refers to the skill relative to a reference forecast. The RPSS is the same calculation as the RPS, except that the comparison is to a one reference forecast (category).

In other words, the RPS measures cumulative squared error between categorical forecast probabilities and the observed categorical probabilities, and the RPSS measures the error relative to a reference (or standard baseline) forecast (climatology, persistence, a reference forecast). The observed categorical probabilities are 100% in the observed category, and 0% in all other categories [9].

$$RPS = \sum_{cat=1}^{Ncat} (P_{cumF(cat)} - P_{cumO(cat)} \tag{4.3}$$

Where $Ncat = 3$ for tercile forecasts. The "cum" implies that the summation is done first for cat 1, then cat 1 and 2, then cat 1 and 2 and 3 [9].

The higher the RPS, the poorer the forecast. RPS=0 means that the probability given to the category that was observed was 100%. The RPSS is based on the RPS for the forecast compared to the RPS For a reference forecast such as one that simply gives climatological probabilities.
RPSS > 0 when RPS for actual forecast is smaller than RPS for the reference forecast.

$$RPSS = 1 - -\frac{RPS_{forecast}}{RPS_{observation}} \tag{4.4}$$

The RPSS is made worse by three main factors [10]:

(1) Mean probability biases

(2) Conditional probability biases (including amplitude biases)

(3) The lack of correlation between forecast probabilities and obs

(1) and (2) are calibration factors and (3) involves discrimination. The tercile category system can be seen as a two category system if the two tercile boundaries are considered one at a time: below normal vs. not below normal above normal vs. not below normal.

4.2.3 Reliability Measures

There are a number of reliability measures that measure or depict the same attribute: the agreement between forecasted probabilities and observed frequencies.
The differences and similarities of the various measures, rank historgram, reliability diagrams and calibration diagrams are explained and discussed in the following sections, so that the use and benefits of combining some of these measures become clear.

It is also worth noting that the *CAL* term in the Brier Score (BS) is basically a quantification of what is seen in these diagrams.

4.2.3.1 Rank Histogram

Rank histograms measure the consistency and reliability and assumes that the observation is statistically indistinguishable from the ensemble members.

The rank histograms are developed by ranking the N ensemble members from lowest to highest and identify the rank of observation with respect to the forecasts. Figure 4.3 show typical distributions and their characteristics with respect to their skill.

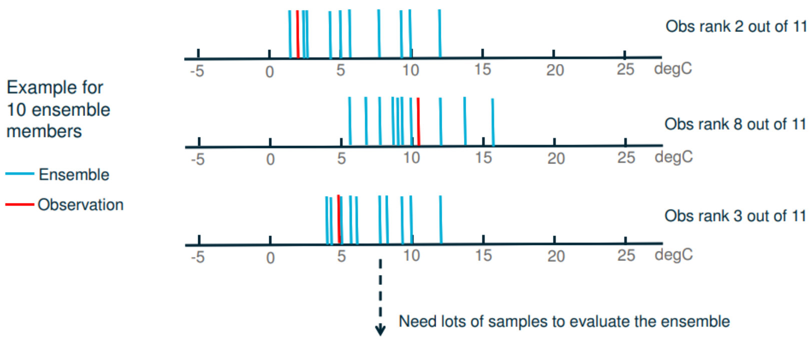


Figure 4.2: One rank histograms ©[10]

It is important to note that the flat rank histogram does not necessarily indicate a skillful forecast. Rank histograms show conditional/unconditional biases, but does not necessarily provide a full picture of the skill, because it[10]:

- only measures whether the observed probability distribution is well represented by the ensemble

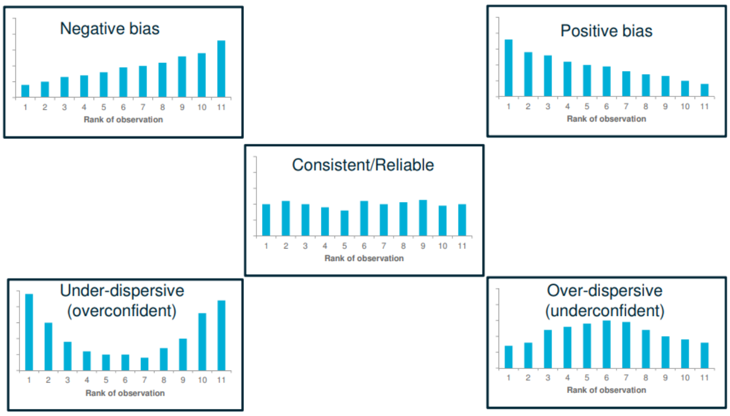


Figure 4.3: Examples of a rank histograms ©[10]

- does NOT show sharpness climatological forecasts are perfectly consistent (flat rank histogram) but not useful

4.2.3.2 Reliability Diagram

The reliability diagrams tell how well predicted probabilities of an event correspond to their observed frequencies.
The reliability diagram plots the observed frequency against the forecast probability, where the range of forecast probabilities is divided into K bins (for example, 0-5%, 5-15%, 15-25%, etc.). The sample size in each bin is often included as a histogram or values beside the data points [12].

The characteristics of the reliability is indicated by the proximity of the plotted curve to the diagonal. The deviation from the diagonal gives the conditional bias. If the curve lies below the line, this indicates over-forecasting (probabilities too high); points above the line indicate under-forecasting (probabilities too low). The flatter the curve in the reliability diagram, the less resolution it has. A forecast of climatology does not discriminate at all between events and non-events, and thus has no resolution. Points between the "no skill" line and the diagonal contribute positively to the Brier skill score. The frequency of forecasts in each probability bin (shown in the histogram) shows the sharpness of the forecast [12]. Figure 4.4 show this principle and Figure 4.5 show typical examples of reliability diagrams for various forecast flaws.

The reliability diagram is conditioned on the forecasts (i.e., given that an event was predicted, what was the outcome?), and can be expected to give information on the real meaning of the forecast. It is a good partner to the ROC, which is conditioned on the observations. Some users may find a reliability table (table of observed relative frequency associated with each forecast probability) easier to understand than a reliability diagram.

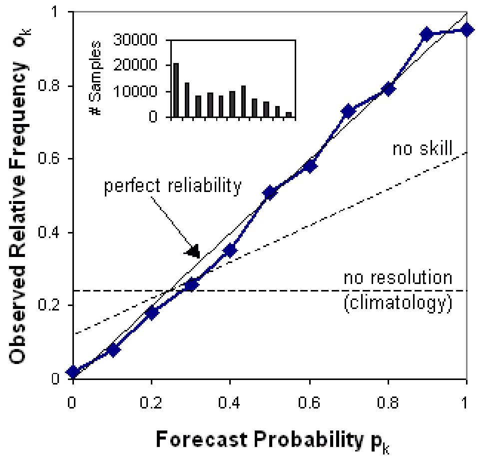


Figure 4.4: Connection between rank histograms and reliability diagrams ©[12]

4.2.4 Calibration Diagram

The calibration diagram provides insight into how well calibrated a probabilistic forecast is and is a complementary metric to the Brier scores (4.2.1 and the Relative Operating Characteristics (ROC) curve (4.2.5). It represents a basic requirement for a probability forecast to be useful [7]. In mathematical terms, it depicts the degree of agreement between the relative frequency of occurrence of an event and the forecasted probability of the event.

A calibration diagram presents pairs of forecast probabilities and relative frequencies of occurrence of an event of interest (e.g., a ramp), with the forecast probabilities represented on the x-axis and the observed relative frequencies on the y-axis.

A calibration diagram is created by first sorting the probability forecasts into probability categories. For example, the categories might include all forecasts with values between 0 and 0.10, 0.10 and 0.20, 0.20 and 0.30, and so on. Then the observed relative frequency of the event occurrence can be computed for each category by counting the number of times when the event occurred in each category and dividing by the number of forecasts in that category.

4.2.5 Event Discrimination Ability: Relative Operating Characteristic (ROC)

This metric shows a probabilistic forecast's ability to predict the occurrence of events and non-occurence non-events.

In the *ROC* diagram the performance of forecasts at different probability thresholds is visualised. One important aspect of the ROC is that it ignores calibration of the forecasts. That is, a poorly calibrated forecast will not be penalized by the ROC. Thus, it is important to pair the ROC evaluation with an evaluation of forecast calibration, such as the calibration diagram, which is discussed in the next section.

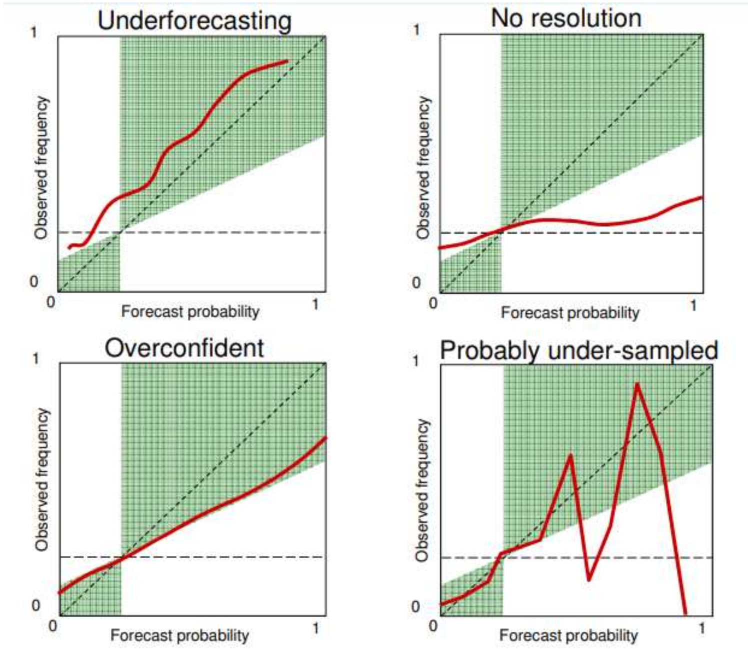


Figure 4.5: Examples of reliability diagrams. The left upper and lower figure correcpond to the histograms for over- and underdispersive ditributions in Figure 4.3. ©D. Hudsen, "Ensemble Verification Metrics" Presentation at ECMWF Annual Seminar[10]

The ROC is based on computing two categorical statistics (see 5.1.2.1):

1. the Probability of Detection (POD), Hit Rate (HR) or true positive rate (TPR)
2. the False Alarm Rate (FAR) or False Positive Rate (FPR)

The ROC curve is created by plotting the true positive rate (TPR) or the probability of detection (POD) against the false positive rate (FAR) or false alarm rate (FAR) at various thresholds. The true-positive rate is also known as sensitivity, recall or probability of detection in machine learning. The false-positive rate is also known as probability of false alarm and can be calculated as (1 specificity) [13].
Figure 4.6 shows how to generate a ROC curve. The orange line represents a forecasting system with little skill, the green with moderate (better) skill and the blue line a forecasting system with reasonable skill.
As shown in Figure 4.6, when the ROC curve falls below the diagonal line the forecasts are random classifiers, or in other words have no skill according to this metric. The blue line shows a good, or better forecast skill, where the curve is pushed up towards in the upper left corner (TPR = 1.0). The area under the ROC curve provides a useful measure of forecast skill[13]).

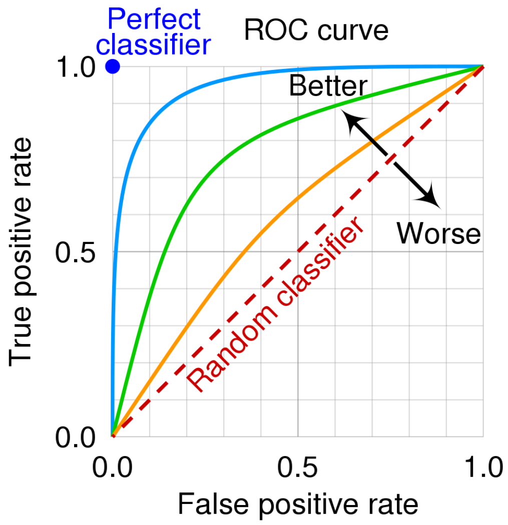


Figure 4.6: Example of a "Relative Operating Curve" (ROC) curve ©Wikipedia [13]

It can also be thought of as a plot of the power as a function of the Type I Error of the decision rule (when the performance is calculated from just a sample of the population, it can be thought of as estimators of these quantities).
The ROC curve is thus the sensitivity or recall as a function of fall-out.
In general, if the probability distributions for both detection and false alarm are known, the ROC curve can be generated by plotting the cumulative distribution function (area under the probability distribution from $-\infty$ to the discrimination threshold) of the detection probability in the y-axis versus the cumulative distribution function of the false-alarm probability on the x-axis[13].

4.2.6 Uncertainty in Forecasts: Rény Entropy

General forecast metrics such as MAE and RMSE do not measure the uncertainty of the forecast and are only considered unbiased, if the error distribution is Gaussian, which is seldome the case. In order to define this, and compare it with uncertainty forecasts, it is recommended to use the Rény entropy, defined as the variation of wind or solar forecast errors in a specified time period [7] (chapter 6).
The Rényi entropy is defined as:

$$H_\alpha(X) = \frac{1}{1-\alpha} log_2 \sum_{i=1}^{n} p_i^\alpha$$

where α (where $\alpha > 0$ and $\alpha \neq 1$) is the order of the Réényi entropy, which allows to create a spectrum of Réényi entropies with p_i being the probability density of the i discrete section of the distribution. Large values of α favor higher probability events, while smaller values of α weigh all instances more evenly. The value of α is specified by the metric user.

4.3 Metric-based Forecast Optimization

Once the most important attributes of a forecasting system and an evaluation metric or matrix has been decided, it may be possible to optimize the forecasting system to have desirable properties. Many forecasting solutions are tuned/optimized for specific performance criteria either at the post-processing stage (conversion of weather forecasts to power forecasts) or even in the numerical weather models themselves. For example, many statistical post-processing techniques allow the user to specify whether to minimize (root) mean squared error or mean absolute error. The former is implicit in *ordinary lest squares*, a widely used method for estimating the parameters of linear models or methods that are based on maximum likelihood estimation assuming Gaussian (or 'Normally') distributed errors. The latter has no closed form solution for estimating linear models so requires the application of numerical methods to solve.

It is recommended that the desired properties of a forecasting solution are considered from the outset and are known to those responsible for the solution's development and implementation.

Chapter 5

Best Practice Recommendations

Key Points
The recommendations in this section are based on the following set of principles:

- *Verification is subjective*
  it is important to understand the limitations of a chosen metric
- *Verification has an inherent uncertainty*
  due to its dependence on the evaluation data set
- *Evaluation should contain a set of metrics*
  in order to measure a range of forecast performance attributes
- *Evaluation should reflect a cost function*
  i.e. the metric combinations should provide an estimate of the value of the solution

In this last chapter, the principles developed in the previous chapters are brought to the application level. In other words, the somewhat theoretical considerations from the previous chapters are now applied to real-world problems. In the second chapter 2, the concept of forecast evaluation uncertainty was introduced with the three attributes “representative”, “significant” and “relevant” to help minimize this type of uncertainty in the evaluation. The following chapter 3, introduced the concept of measurement uncertainty with the associated uncertainty in the evaluation process and how to minimize the errors in the evaluation due to this type of uncertainty. In the previous chapter 4 the performance assessment was described in general terms and with examples that are relevant for all types of evaluation in the power sector.

5.1 Developing an Evaluation framework

> ***Key Points***
> *The construction of a comprehensive evaluation framework is an alternative to a one-metric forecast evaluation approach and can be an effective way to mitigate the "relevance" issues associated with the tuning (optimization) of forecasts to target metrics that are not optimal indicators of value for an end user's application.*

The "typical forecasting task" is defined in this context as forecasts generated to fulfill operational obligations in electric system operation, trading and balancing of renewable energy and in particular wind power in power markets. There are certainly many other tasks and applications of weather and power forecasts in the power industry that can also benefit from the following best practice recommendations. However, the primary target for the following recommendations is the evaluation of forecasts for these particular applications. Section 5.2 deals with the evaluation to maximize value from operational forecasts, section 5.3 with the evaluation of trials and benchmarks and in the use cases section 5.5 there are example evaluations for energy trading and balancing, power ramps and reserve.

5.1.1 Analyses of Forecasts and Forecast errors

In this discussion, forecast errors are defined as forecast minus observation ($fc - obs$). Errors in forecasting are inevitable. The primary objective is, of course, to minimize the magnitude of the error. However, a secondary objective may be to shape the error distribution in ways that are beneficial to a specific application. A direct and deep analysis of the prediction errors can provide considerable insight into the characteristics of forecast performance as well as information that can allow users to differentiate situations in which forecasts are likely to be trustworthy from those that are likely to produce large errors.

The construction of a frequency distribution of errors (also referred to as density functions or probability density functions) is an effective way to obtain insight about forecast error patterns. These are created by sorting errors and visualizing their distribution as e.g.,

- (probability) density curve
- histogram (frequency bars)
- box plot

All of these chart types show the same basic information but with different degrees of detail. Density curves provide the most detail since they depict the full probability density function of the forecast errors. Histograms provide an intermediate level of detail by showing the frequency of a specified number of error categories. Box plots condense this information into several quantiles (see 5.1.2.2). Errors of a well calibrated forecast model should always be scattered around zero. A frequency distribution that has a center shifted from zero indicates a systematic error (also known as a bias).

For power forecasts one will often see positively skewed error distributions, which are due to the shape of the power curve which has flat parts below the cut-in wind speed and at wind speeds that produce the rated power production. The skewed distribution is often the result of the fact that forecasts close to zero cannot have large negative errors. The inverse is true for forecasts of near rated power (i.e. large positive errors cannot occur) but forecasts of rated power are often less frequent than near zero forecasts and hence have less impact on the error distribution.

5.1.2 Choice of Deterministic Verification methods

When evaluating forecasts one or several evaluation methods or metrics to measure and compare the forecast performance have to be selected. There is not a single best metric that can be effectively used for all applications. The definition of "best metric" highly depends on the user's intended application and should be based on a quantification of the sensitivity of a user's application to forecast error. For example, if a user has to pay a penalty for forecast errors that are proportional to the squared error, a mean squared error metric is well suited for evaluation.

However, if the penalty is proportional to the absolute error, a mean absolute error metric would be a better choice. If the user is interested in predictions of specific events such as high wind shutdown or large wind ramps, the mean squared or absolute error metrics are not good choices, because they do not provide any information about the ability of a forecast to predict these events due to their averaging characteristics. In this case, an event-based metric should be employed. An example of this type of metric is the critical success index (CSI), which measures the ratio of correct event forecasts to the total number of forecasted and observed events.

5.1.2.0.1 “Loss function:” In order to get forecast performance information that is relevant for a user's application, it is crucial to carefully select the evaluation metrics and ideally they should be based on the so-called “loss function” for the user's application. The “loss function” is also often referred to as a “cost function”, especially when related to costs that can be associated with specific forecast errors. Conceptually, a well-formulated "loss" or "cost" function measures the sensitivity of a user's application to forecast error. If one forecast is used for different applications with different loss functions, a set of metrics should be derived. If a single metric is desired, then a composite metric can be constructed by weighting the individual application-based metrics by the relative importance. More details on how to develop such loss functions and evaluation matrices can be found in 5.1.4 .

5.1.2.1 Dichotomous Event Evaluation

One may quantify desirable qualities by considering a range of of dichotomous (yes/no) events such as high-speed shut-down or ramps. A forecast might imply that "yes, a large ramp will happen" and trigger the user to take action, but the ability of a forecasting system to make such predictions is not clear from the average error metrics. Therefore, one should employ a quantitative verification approach to assess this ability by analyzing the number of correct positive, false positive, correct negative and false negative predictions of particular events [14], [1]. Table 5.1 provides an example table to carry out such categorical evaluations.

Table 5.1: Example of a dichotomous evaluation table

		Observations	
		YES	NO
Fore-cast	YES	a correct event forecast	b false alarm
	NO	c surprise events	d no events

Recommendation for applications with (Extreme) Event Analyses:
Categorical statistics that can be computed from such a yes/no contingency table. The list below is an except of a comprehensive list of categorical statistics tests published by the Joint World Weather Research Program (WWRP) and Working Group Numerical Experimentation on Forecast Verification (WGNE) and provides the most common used metrics and their characteristics, relevant for forecast applications in the power industry. Details, equations and more comprehensive explanation on the use of these as well as references can be found (online) in [1]. It is recommended to apply these categorical statistics in particular for applications, where standard average metrics do not provide a measure of the true skill of a forecast to predict a specific event. In wind energy forecasting applications this is in particular important for extreme event analysis, ramping and high-speed shutdown forecasting etc. In such applications, it is important to distinguish between *quality of a forecast* (the degree of agreement between the forecasted and observed conditions according to some objective or subjective criteria) and *value of a forecast*(the degree to which the forecast information helps a user to achieve an application objective such as improved decision-making). Wilks [22] and Richardson [21] present concepts for the value versus skill for deterministic and probabilistic forecast evaluation of that type, respectively.

- **Accuracy**
  Answers the question: Overall, what fraction of the forecasts were correct?
  Range: 0 to 1. Perfect score: 1

- **Bias score**
  Answers the question: How did the forecast frequency of "yes" events compare to the

observed frequency of "yes" events?
Range: 0 to 1. Perfect score: 1

- **Probability of detection (POD)** Answers the question: What fraction of the observed "yes" events were correctly forecast?
  Range: 0 to 1. Perfect score: 1
- **False alarm ratio (FAR)**
  Answers the question: What fraction of the predicted "yes" events actually did not occur (i.e., were false alarms)?
  Range: 0 to 1. Perfect score: 0
- **Probability of false detection (POFD)**
  Answers the question: What fraction of the observed "no" events were incorrectly forecast as "yes"?
  Range: 0 to 1. Perfect score: 0
- **Success ratio**
  Answers the question: What fraction of the forecast "yes" events were correctly observed?
  Range: 0 to 1. Perfect score: 1
- **Relative value curve (versus skill)** for deterministic forecast
  Answers the question: For a cost/loss ratio C/L for taking action based on a forecast, what is the relative improvement in economic value between climatological and perfect information? Range: -1 to 1. Perfect score: 1.

5.1.2.2 Analyzing Forecast Error Spread with Box and Wiskers Plots

The box-and-whiskers plot is a visualization tool to analyze forecast performance in terms of the error spread when comparing forecasts with different attributes such as forecast time horizons, vendors, methodologies. Figure 5.4 shows the principle of a box and whiskers plot. This type of charts can be used to illustrate the spread of forecast performance in each hour of the day-ahead horizon can be visualized. It can also show that some forecasts in some hours have very low errors compared to the average error in that hour, as well as occasionally very high errors. In section 5.4.2, a use case for the application of box plots is demonstrated to verify significance of results.

5.1.2.3 Visualising the error frequency distribution with histograms

Histograms allow one to (1) quantify the frequency of occurrence of errors below or above a specified level or (2) visualise the forecast error distribution for specified error ranges. In case (1) the graphical or table presentation can be directly used to derive a metric that indicates that errors are less than x% of the installed capacity in y% of the time. In this

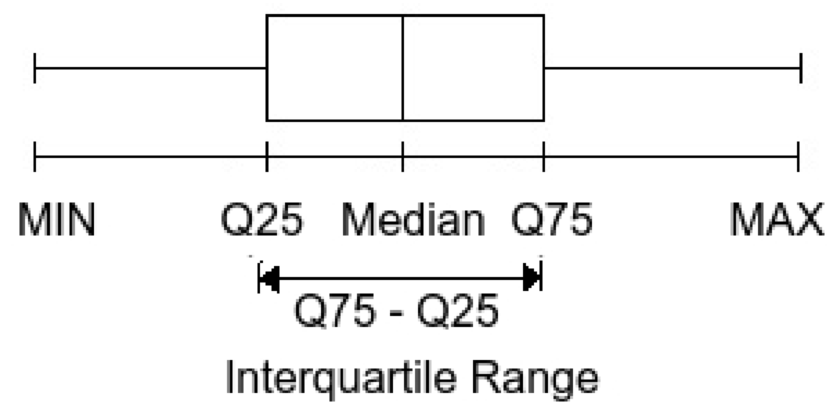


Figure 5.1: Principle of a box-and whiskers plot. The plot displays a five-number summary of a set of data, which is the minimum, first quartile, median, third quartile, and maximum. In a box plot, a box from the first quartile to the third quartile is drawn to indicate the interquartile range. A vertical line goes through the box at the median.

way, histograms function as a metric providing the percentage of time that errors are within a given margin [[4]]. In case (2) the error distribution of a forecast can be derived the graphical or tabular presentation of the histogram information. This enables an easy identification of the frequencies of large errors and provides the possibility to analyze and possibly modify the forecast system to minimize these errors. In summary, histograms visualize two main attributes:

- Robustness of a forecast
- Large Errors in an error distribution

In Madsen et al. [4] an example can be found for the way histograms help to interpret statistical results and error distributions. In their example, they directly determined that a 1 hour-ahead prediction contained errors less than 7.5% of the available capacity in 68% of the time, while a 24 hour-ahead prediction showed errors of that size only in 24% of the time. For large errors, they determined from the histogram that the same 1 hour-ahead prediction's largest errors were 17.5% of available capacity in only 3% of the time.

Recommendation: If the application requires that specified error sizes should occur less than specified percentages of the time, a histogram analysis should be used to directly identify, whether or not a forecast's performance fulfills such criteria.

Figure 5.2 provides two example histograms with typical frequency distribution of errors for a 2-hour forecast horizon (left) and a day-ahead horizon (right) as described in [4].

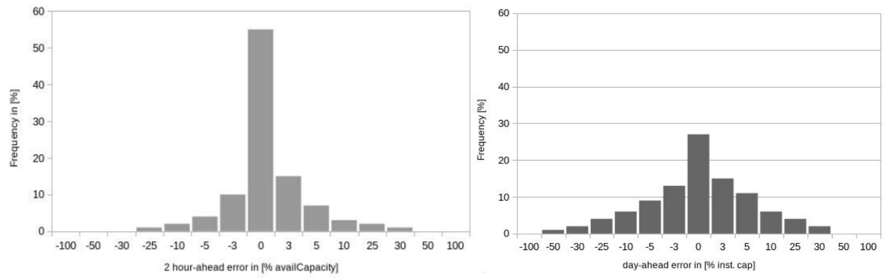


Figure 5.2: Examples of two histograms showing typical frequency distribution of errors for a 2-hour forecast horizon (left) and a day-ahead horizon (right).

5.1.3 Specific Probabilistic Forecast Verification

As in the case of the verification of deterministic forecasts, it is recommended that multiple verification scores also be employed for the evaluation of probabilistic forecasts. A well chosen set of probabilistic forecast evaluation metrics will provide an indication of several key aspects of the forecast skill and thus provide a more comprehensive representation of forecast performance than a single metric. It can also assist users and providers in the determination of which aspects of forecast should be the focus of forecast improvement efforts.

An evaluation of probabilistic forecasts should ideally three components:

1. a metric that measures overall quality (discrimination and calibration together), such as the Brier Score (BS) or Ranked Probability Score (RPS)
2. a metric that measures discrimination alone such as the ROC
3. a metric or chart that provides an indication of the reliability (calibration) such as the ranked histogram. reliability diagram or CAL component of the Brier Score.

This combination of metrics will provide a broad perspective on forecast performance and also can assist in the identification of forecast performance issues. For example, when discrimination is good but calibration (biases) issues are degrading the overall quality, a reliability diagram can reveal the nature of the calibration problems. Details about how to compute or construct each of these metrics and diagrams can be found in section 4.2

5.1.4 Establishing a Cost Function or Evaluation Matrix

Due to the complexity of the task and the fact that the objectives of forecast users are not the same, the following section is an introduction to the concept of a evaluation framework in which structured procedures for the evaluation and verification of forecasts are established. The structure may be shortened and adapted depending on the size of the forecasting system and the importance in the overall business processes.

Best practice in this context is to following a procedure, where the evaluation/verification reflects the importance of a forecasts in it's role of the business processes and provides incentives for the forecast service provider to generate forecasts that fit the outlined (and verified) purpose.

As a minimum requirement when establishing such an evaluation framework the following set of procedures should be considered:

1. **Definition of the forecast framework**
   It is important to exactly define the forecast application, the key time frames and a ranking of relative importance.

2. **Base performance evaluation on a clearly defined set of forecasts**
   The base performance should contain "typical error" metrics in order to monitor an overall performance level.
   - time frame: minimum 3 months, ideally 1 year
   - "typical error" metrics: nMAE, nRMSE, BIAS
3. **Quality assessment of the evaluation sample data**
   The detection of missing or erroneous data and a clear strategy how to deal with such missing data needs to be made at the outset of any evaluation period to ensure that verification and forecasting is fair and transparent.
4. **Specific Performance evaluation on a set of error metrics**
   - Visual Inspection
   - Use of more specific metrics:
     (a) deterministic: SDE, SDBIAS, StDev, VAR, CORR
     (b) probabilistic: Brier Score, ROC curve, Probability Interval Forecast Evaluation (4.2)
   - Use of histogram or boxplot for evaluation of outliers
   - Use of contingency tables for specific event analysis
   - Use of improvement scores relative to a relevant reference forecast for comparisons
   -

Note, details on the framework and evaluation metrics can be found in [4] and [6], specific metrics and explanation of metrics can be found in [2], [16] for deterministic forecasts inclusive solar forecasting and for probabilistic forecast metrics in [3]. Significant tests can be found e.g. in [17].

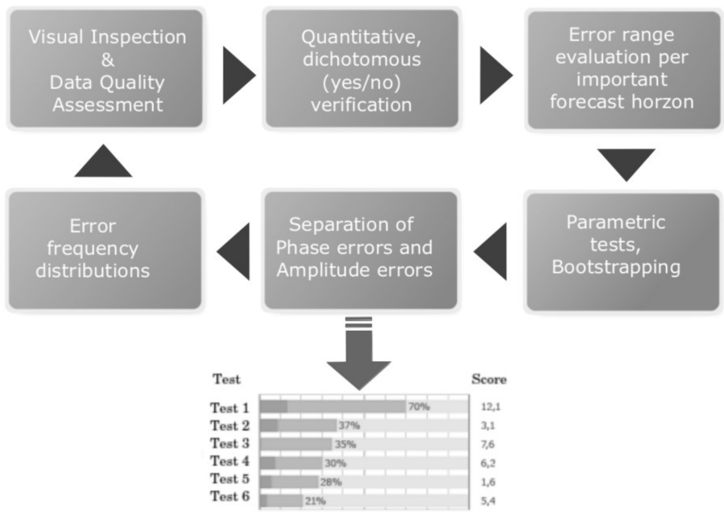


Figure 5.3: Example of an evaluation matrix that verifies forecasts against 6 test metrics and displays the scores for a holistic overview of the forecast performance.

5.1.4.1 Evaluation Matrix

Establishing an evaluation matrix is complex, but can be straight forward if the principles of forecast uncertainty and choice of appropriate metrics are incorporated into the evaluation strategy.

Best practice for the establishment is to go through the various steps outlined in section 5.1.4 to choose the components for the evaluation framework. The core concept is to use this framework to define a formal structure and then add multiplication factors to weight each of the selected individual metrics according to their relative importance.

The matrix can be setup in a spreadsheet environment with macros or within a database environment, where all data is available and metrics may even be directly computed though the database software. The key point of the matrix is that the forecast performance results can be collected, multiplied with an "importance factor", normalised and transferred into the summary table to visualize the scores. For example the scores can be visualized with a bar chart that indicates the performance in a scale from e.g. 0 to 1 or 0 to 100 as shown in 5.3.

Such a evaluation matrix provides important information in a comprehensive way and can be applied for comparisons as well as for the analysis of the potential for forecast improvement.

5.2 Operational Forecast Value Maximization

Key Points

- *Once operational forecasts have been established it is important to monitor the quality of generation facility data supplied to the forecast system(s) and used for forecast evaluation; often attention to this diminishes after a benchmark is completed*
- *Ongoing deep analysis of forecast performance and effective provider user communication is critical for maintaining and refining forecast performance*
- *Focus should be on maximizing forecast value for the application and not on maximizing performance of standard metrics; this may include identifying or refining the cost function for a users application and/or working with the provider to optimize forecasts for the application(s)*
- *A plan should be developed to motivate and reward providers to continually refine forecast methods and adapt new approaches from the latest research; this may include financial incentive schemes*

Operational forecasts should be evaluated in the context of their end-use. Different use cases will have different cost functions, some of which may be complex or impossible to define. Organizations evaluate operational forecasts for a variety of reasons and on a wide range of scales, from individual wind farms to entire fleets, and from short lead times to horizons spanning several days.

Simple evaluation metrics such as MAE or RMSE can be used to get an overview of general forecast performance and to provide an indication of forecast performance for decisions with (symmetric) linear or quadratic loss functions, respectively. However, in most cases the true cost of wind power forecast errors will be more complex and depend on externalities.

Systematic evaluation of operational forecasts is however an important business function for forecast users. Whether this is monitoring the quality of the forecasts produced in-house or procured from vendors, regular evaluation supports continuous improvement in forecast performance and end-use. This section provides a guide to the best practices in evaluation of operational forecasts. It begins by reviewing common motivations for continuous and periodic evaluation of operational forecasts, and then discusses different evaluation paradigms for specific use-cases.

5.2.1 Performance Monitoring

Continuous monitoring of forecast performance is best practice in order to develop an understanding of forecast capability and to identify and respond to issues with raw forecast data or its processing. While failure of forecasting systems is extremely rare, weather models, IT systems, and the forecast target (e.g. individual wind farm, portfolio of wind farms, national wind output) are constantly evolving. This has the potential to introduce new and unforeseen sources of error.

5.2.1.1 Importance of Performance Monitoring for Different Time Periods

Short Periods (monthly): While error metrics or contingency tables calculated over short periods do not provide reliable measures of overall performance they can provide an indication of problems with a forecasting system and large errors should be logged and investigated. Abrupt changes in forecast performance can result from errors in data processing, such as incorrect availability information during maintenance.

Long Periods (> 6 months): Changes in performance over longer time scales may be a result of changes to a supplier's numerical weather model(s) or changes in the behaviour of wind power plant as they age. Slow changes may be more difficult to detect, but over time can accumulate significant biases which should also be investigated.

For both cases, it is necessary to dis-aggregate forecast metrics to identify some sources of error. Important factors to consider when dis-aggregating errors are to include lead-time, time of day, power level and weather type.

Regular reporting and tracking of forecast performance over relevant periods can help foster understanding of forecast capability across business functions and support staff and process development.

Recommendation:

- Forecasts performance should be monitored continuously to quickly identify technical problems
- Large errors should be investigated and recorded for future analysis
- Error metrics should be dis-aggregated by appropriate factors, e.g. lead-time, power level
- Regular reporting for error metrics supports forecast users' interpretation of forecast information

5.2.2 Continuous improvement

Forecast evaluation is the first stage in identifying areas for potential improvement in forecasting systems. Periodically evaluating operational forecast performance and its impact on wider business functions can be a valuable exercise. For example, changes in the way forecasts are used, or the importance of different lead-times or variables may be a cause to change the way forecasts are produced or communicated internally.

In situations where multiple operational forecasts are produced or supplied, regular benchmarking can add value as different services are upgraded over time or exhibit different performance characteristics.

Recommendation:

- Evaluation underpins forecast improvement and insights should be shared with both forecasters and end-users
- Evaluation and improvement should be driven by end-use and business value

5.2.3 Maximization of Forecast Value

Forecast value can be maximized by continuously monitoring and evaluating operational processes of both forecasts and measurement quality. Additionally, the use of forecasts and the interaction with other business processes need to be taken into consideration as well, if they can impact the quality of the forecasts or the correctness and trustworthiness of the evaluation.

The use of a single metric such as a mean absolute or root mean squared error for forecast evaluation may be a way to start a process and can be helpful in identifying errors in the system that can cause unwanted costs. This is a valid and useful approach. It is however recommended to use such simplified methods only for monitoring purposes and not as the primary verification tool (see also chapter 2, especially sections 2.2, 2.3 and 5.1).

Recommendation: The following aspects should be taken into consideration when identifying a "loss function" or "cost function" in the selection process of performance metrics for operational forecasts. Details on some metrics can be found in the Appendix A, a comprehensive database for metrics can be accessed online [1] together with the concepts of the metrics and valuable combinations of metrics, which have also been described in more detail in section 5.1.

- Evaluation should contain a selection of metrics:
  - One metric alone is not indicative of overall forecast performance
  - Use de-compositions of errors to identify the origin of errors. e.g. look at bias and variance alongside MAPE or RMSE.

   - Selected metrics should reflect the costs of errors or security constraints to the greatest extent possible based on the user's knowledge of the application's characteristics
   - Box plots, histograms and scatter plots reveal additional important information compared to a "typical error" metric

- Evaluation metric combinations can provide a representative approximation of a "cost function":

   1. subjective evaluation through visual inspection
   2. quantitative, dichotomous (yes/no) verification of critical events such as high-speed shut-down or ramps with e.g. contingency tables
   3. error ranges per important forecast horizon
   4. error ranges per hour of day or forecast hour
   5. error frequency distributions in ranges that have different costs levels
   6. separation of phase errors and amplitude errors according to their impact
   7. parametric tests, bootstrapping can be used to look on individual error measures before averaging

5.2.4 Maintaining State-of-the-Art Performance

If expensive long-term solutions have been established it can be challenging for an end-user to ensure that state-of-the-art performance is maintained. This can be due to the stiffness of the established IT solution (see also Part 1 of this recommended practice), but also due to the fact that there is no monitoring of the performance.

Recommendation: It is recommended that a performance monitoring takes place, where those forecasts that are relevant for the business processes are compared against a suitable and objective measure. The most common measures are climatology values, persistence values or comparison to previous periods, such as the previous calendar year. Such techniques can provide motivation and can be set up with a reward scheme for the forecast provider to improve forecasts with time and improved knowledge of the specific challenges and needs of the end-user's forecast problem. (see Table 5.2)

5.2.5 Incentivization

Operational forecasts may be tied to an incentive scheme by which monies are exchanged based on forecast performance. Examples of such arrangements exist in both commercial forecast services and regulation of monopoly businesses. As the terms of the incentive scheme typically include details of how forecasts are evaluated, performing this evaluation poses few risks. However, the evaluation methodology should be carefully considered when

Table 5.2: List of possible performance monitoring types useful for evaluation of operational forecasts, incentive scheme benchmarks, tests and trials. The types are not meant to be stand-alone and may also be combined.

Performance Measure	**Comment/Recommendation**
Improvement over persistence	comparison against persistence is the same as comparing not having a forecast to having one. Useful measure for short-term forecasts as a mean of evaluating the improvement of applying forecast information to measurements. Note: be aware of data quality issues when evaluating, especially in the case of constant values that benefit persistence, while the forecast provides a realistic view.
Improvement over past evaluation period / forecast	If improvement is important, the comparison to a past evaluation can be useful, especially in long-term contracts. In this way, the forecaster is forced to continue to improve and the target is moved with the improvements. The payment structure however needs to incorporate the fact that improvements reduce over time and have an upper limit.
Comparison against set targets	If the required performance of a forecasting system can be defined, clear targets should be set and the payment directed according to a percentage from 0-100% of the achieved target.
Categorised error evaluation	An effective evaluation format is to not set one error target, but categorise errors instead e.g. large, medium and small errors. If large errors pose a critical issue, then improvement on these may be incentivized higher and vice versa. The end-user can in that way steer the development and focus of improvements.

negotiating or subscribing to such incentive schemes.

Incentives may take the form of a linear relationship between reward/penalty and a forecast metric such as Mean Absolute Error, which may be normalized to installed capacity, and capped at some minimum/maximum reward/penalty. Similarly, incentives may be based on an event-based metric, accuracy or hit-rate for example, for specific events such as ramps or within-day minimum/maximum generation. The time period over which such an incentive is calculated and settled will have a large impact on it's volatility as evaluation metrics may vary greatly on short time scales. Longer timescales are conducive to a stable incentive reflective of actual forecast performance rather than variations in weather conditions. The basic evaluation rules developed in section 2 and 4 are equalyy valid here and are recommended to be applied.

In summary, the recommendation is that the formulation of an incentive schemes should consider four factors:

1. selection of relevant target parameters (see section 2.3)
2. selection of relevant metrics (see sections 5.2,5.1, 5.1.4, 5.4.1)
3. selection of relevant verification horizons (see section 2.2)
4. exclusion principles (see chapter 3 and section 3.2 and 3.5)

The selection process of relevant target parameters is highly dependent on the forecasting solution. The objective and proper setup of verification as well as evaluation metrics and frameworks can be found in 2, 4 and sections 5.1, 5.1.1, 5.3.1.

Recommendation: A set of relevant target parameters needs to be defined to provide a focus area for the forecaster. Comparison to a previous period, to a persistence forecast or a set target that is realistic can circumvent a number of constraints that are difficult to exclude in an evaluation. The most important consideration for any performance incentive scheme is that the scheme should put emphasis on the development and advancement of forecast methods for exactly those targets that are important for the end-user's applications.

Table 5.2 provides a list of possible benchmark types for an incentive scheme.

5.3 Evaluation of Benchmarks and Trials

Key Points
In order to maximize the probability of selecting an optimal forecast solution for an application the performance evaluation uncertainty process should be minimized and non-performance attributes of a forecast solution should be effectively considered. Evaluation uncertainty can be minimized by a well-designed and implemented performance benchmark or trial protocol. A benchmark should have three well designed phases: (1) preparation, (2) execution and (3) performance analysis that each address the key issues associated of three primary attributes of an evaluation process.

As a general guideline, the evaluation needs to follow the three principles of being:

1. **representative**
2. **significant and repeatable**
3. **relevant, fair and transparent**

The principles have been explained in detail in Chapter 2. In this section specific considerations and the application of these principles in benchmarks and trials are provided.

5.3.1 Applying the 3 principles: representative, significant, relevant

The three key attributes of a forecast solution evaluation associated with a trial or benchmark (T/B) are (1) representativeness (2) significance and (3) relevance. If any one of these are not satisfactorily achieved the evaluation will not provide meaningful information to the forecast solution decision process and the resources employed in the trial or benchmark will effectively have been wasted. Unfortunately, it many not be obvious to the conductor of a T/B or the user of the information produced by the T/B whether or not these three attributes have not been achieved in the evaluation. This section will present the issues associated with each attribute and provide guidance on how to maximize the likelihood that each will be achieved.

The conductors of a T/B should consider all of the factors noted in the three key areas for a T/B. Part of these are described in detail in section 2 in sections 2.1, 2.2 and 2.3. The following is a reminder with specifics for the T/B case:

1. **Representativeness**
   Representativeness in this context refers to the relationship between the results of a trial or benchmark evaluation and the performance that is ultimately obtained in the operational use of a forecast solution. It essentially addresses the question of whether or not the results of the evaluation are likely to be a good predictor of the actual forecast performance that will be achieved for an operational application. There are many factors that influence the ability of the T/B evaluation results to be a good predictor of future operational performance. Four of the most crucial factors here are:
   (a) size and composition of the evaluation sample,
   (b) quality of the data from the forecast target sites,
   (c) the formulation and enforcement of rules governing the submission of T/B forecasts (sometimes referred to as fairness),
   (d) availability of a complete and consistent set of T/B information to all T/B participants (sometimes referred to as transparency)
2. **Significance** (see section 2.2) For benchmarks and trials it is specifically important that a result obtained now, should also be obtainable when doing a second test. Or, if a test runs over 1 month, the same result should be obtainable over another randomly selected month.

   Often, especially in short intervals, this is not possible due to the different climatic and specific weather conditions that characterize specific periods of a year. In this case, it is necessary to establish mitigating measures in order to generate results that provide a correct basis for the respective decision making.

Such a mitigating measure could be to consume potentially new forecasts in real-time and

(a) compare or blend them with a running system in order to test the value of such a new forecast

(b) evaluate the error structure of a potential new forecast to the error structure of your running system

The both tests can be relatively easy incorporated and tested against the main forecast product, such as a day-ahead total portfolio forecast. It will not reflect the potential or performance and quality of a new forecast in its entirety, but comparing error structures in form of for example error frequency distributions, ensures that a bias due to a lack of training or knowledge about operational specifics does not provide a misleading impression on quality. Chapter 4 details principles and section 5.1 provides details on suitable metrics.

3. **Relevance** (see section 2.3) Results obtained must reflect relevance in respect to the associated operational task and forecasts for energy applications should follow physical principles and be evaluated accordingly. That means in fact that the b/t task must in some way reflect the future function of the forecasts. If this is not so, the results from a b/t should not be used to select a solution of vendor. Instead it may be used to evaluate other performance measures, such as service, support, delivery etc. Fairness in the evaluation, specific for benchmarks and trials then means that the forecast providers are informed about this different objective. Forecasts also need to be evaluated on the same input and output. If assumptions are made, these assumptions must also be provided in a transparent way to alll participants.

A useful approach is to create a evaluation plan matrix that lists all of the factors noted in the discussion in this section and how the users evaluation plan addresses them.

5.3.2 Evaluation Preparation in the Execution Phase

The evaluation of a T/B should start in the execution phase in order to prevent errors along the way from making results unusable. Since there is usually a time constraint associated with T/B's there are a number of aspects that should be considered to ensure meaningful results.

Recommendations for the the execution phase:
Data monitoring:
Measurement data and forecast delivery should be monitored and logged in order to prevent data losses and to ensure that all relevant data is available for the evaluation. It is recommended that the data monitoring should contain the following tasks:

- test accuracy and delivery performance for fairness and transparency

- monitor forecast receipt to test reliability
- exclude times, where forecasts are missing to prevent manipulation on performance

Consistent Information

The fourth key factor is the availability of a complete and consistent set of T/B information to all participants in the T/B. Incomplete or inconsistent information distribution can occur in many ways. For example, one participant may ask a question and the reply is only provided to the participant who submitted the inquiry.

Develop and refine your own evaluation scripts:

Independent whether is is a first time b/t or a repeated exercise, the execution phase is the time, where the following evaluation has to be planned and prepared. It is recommended to verify metrics scripts or software tool and input/output structures as well as exclusion principles.

5.3.3 Performance Analysis in the Evaluation Phase

The performance analysis has a number of key points that need consideration. These are:

1. Application-relevant accuracy measures of the forecasts
   The key point here is that the metrics that are used in the verification must have relevance for the application. For example, if a ramp forecast is tested, a mean average error only provides a overall performance measure, but is not relevant for the target application. If a vendor knows that performance is measured with an average, the incentive would be to dampen forecasts to reduce the overall average error, which is the opposite of what is required for the application to work. Such an application would have to use a scoring system for hits, misses and false alarms of pre-defined ramping events.

2. Performance in the timely delivery of forecasts
   The key pitfalls in an T/B are often associated with the failure to closely monitor the following aspects:

   (a) Lack of check or enforcement of forecast delivery time
   If forecast delivery is not logged or checked, it is possible for a forecast provider to deliver forecasts at a later time (perhaps overwriting a forecast that was delivered at the required time) and use fresher information to add skill to their forecast or even wait until the outcome for the forecast period is known. Although one might think that such explicit cheating is not likely to occur in this type of technical evaluation, experience has indicated that it is not that uncommon if the situation enables its occurrence.

(b) Selective delivery of forecasts
This example illustrates how the results might be manipulated with explicit cheating by taking advantage of loopholes in the rules. In this example the issue is that the B/T protocol does specify any penalty for missing a forecast delivery and the evaluation metrics are simply computed on whatever forecasts are submitted by each provider. As a forecast provider it is easy to estimate the difficulty of each forecast period and to simply not deliver any forecasts during periods that are likely to be difficult and therefore prone to large errors.
This is an excellent way to improve forecast performance scores. Of course, it makes the results unrepresentative of what is actually needed by the user. Often it is good performance during the difficult forecast periods that are most valuable to a user.

3. Ease of working with the forecast provider
In a T/B support in understanding forecast results and error structures may be a good time to test and evaluate for the future. It should however be considered to communicate to the vendors, if it is a decision criteria, especially in non-refunded situations, where resources are used differently than in contractual relationships.

5.3.4 Evaluation examples from a benchmark

Figure 5.4 shows an example of a forecast evaluation using a box-and-whiskers-plot to visualize the spread in MAPE (mean absolute error as percentage of nominal power) of 5 forecasts of different day-ahead time periods (each column) at two different sites. The distribution within each time period is shown for the 5 forecasts errors. In that way, the spread of forecast performance in each hour of the day-ahead horizon can be visualized. It also shows how some forecasts in some hours show very low errors compared to the average error in that hour, as well as occasionally very high errors.

Figure 5.5 shows an example of an evaluation of errors by time of day for a fixed lead time of 3 hours. It illustrates a very large spread in errors during certain times of the day, as would be expected.
Nevertheless, if such evaluations are compared between different forecast providers an evaluation of the "most costly errors" may reveal a very different result than, if only an average metric per forecaster would be used.

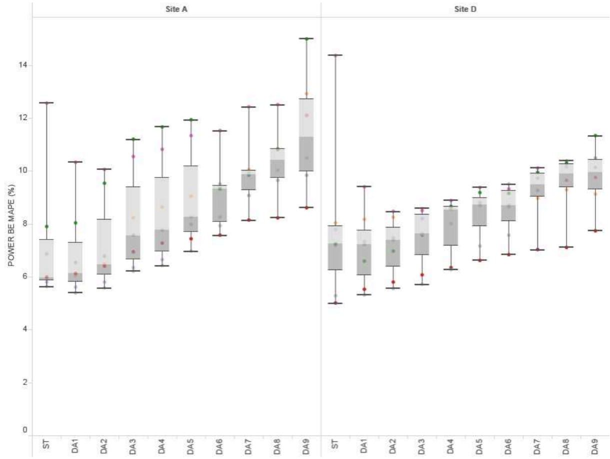


Figure 5.4: Example of a box-and-whisker-plot verification at two different sites (left and right panel) for different look ahead times (x-axis; DAx is $x^t h$ hour of day-ahead forecast) and mean absolute percentage error (MAPE; y-axis).

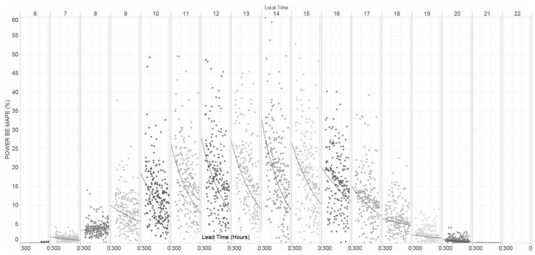


Figure 5.5: Example of a forecast error scatter plot by time of the day (top x-axis) for 3-hours lead times and forecast error (y-axis)

5.4 Evaluation of Development Techniques

Key Points

Keeping State of the Art in forecasting is an important aspect for any end-user, but especially for those with complex IT infrastructure systems or multiple suppliers of forecasts that are bound to statistically consistent forecasts over a period of time for highest performance.

This Section outlines how analysis, diagnostics and evaluation of improvements need to be structured in order to ensure sustained improvement over time without radical changes in existing infrastructures and the typical pitfalls associated with such evaluations.

5.4.1 Forecast Diagnostics and Improvement

The improvement of a forecast over time is especially important in an operational environment, where the IT infrastructure is complex and the amount of resources required to exchange a forecast service provider is in no relation to the gain in forecast performance. Other cases of this type may be a statistical dependence of a or multiple forecasts going into a tool for further processing. The following recommendations may therefore be applied for any of such cases, where an end-user is bound to a forecast solution.

Improvements over time and the importance of a forecast solution being able to develop over time in a real-time environment is difficult to measure. Also, the improvement of forecasts may have a steep curve in the first years, or when constant changes in the system become less frequent.
However, over time any forecast has a limit and the rate of improvement reduces. This needs to be taken into account equally much as the ability of a forecast solution for develop over time to keep a state of the art character.

Table 5.2 is a guideline for the evaluation of forecasts and diagnostics for such improvement monitoring (see also 5.2.5).

5.4.2 Significance Test for new developments

Forecast vendors and researchers are always seeking for improvements and new developments, testing and investigating new technology or techniques to add value to specific tasks in the forecasting arenas. Whenever a new development is ready for testing, the researchers or technical staff are confronted with the question, whether the new technique outperforms the older or current state of the art. Due to time constraints, data limitations or lack of historical available forecasts or measurements, this is often a difficult question to answer.

The following example demonstrates such a typical situation and presents and outlines the overall considerations that need to be taken, followed by the choice of metrics and test on significance on the results.

Initial Considerations
A forecasting model that can take various inputs, such as online measurements in an auto-regressive manner, weather forecasts or other predictive features, generates power forecasts, which estimate the future electricity production. In order to decide which model is most suitable, it is necessary to evaluate its quality by comparing the forecast against power measurements. Typically, the errors of a separate test data are compared against each other in order to then decide in favor of one of the models. Which error measure is chosen should be individually adjusted to the corresponding application.

The evaluation should be performed strictly on test data that were not used to calibrate the respective model. Otherwise it can easily happen that models are favored, which have adapted too much to the training data without being able to generalize for future unknown situations. If several models are compared, they should also have been jointly trained on data that does not originate from the test set.

In the case of wind power forecasts, it is furthermore essential to select the test data from a continuous period. The data cannot be considered temporally independent. If one were to randomly assign individual samples to a training and a test set, one would assign both sets to random samples that share a large part of the information. As a result, preference would also be given to models that are over-adapted to the training data.

In addition to the error measure, other aspects can also play a role. For example, one is faced with the question of whether an established model should be replaced. For several reasons it may seem attractive not to replace it even though another one shows a smaller error. For instance, because confidence in the model functionality has been built up, or because a change in the model requires additional effort. Such or similar cases make it necessary to examine the significance of the estimated error values. The critical question behind this is whether the extent of the test data considered is sufficient to form the basis for a decision.

Evaluation of Significance

One way to evaluate the significance of the error values is to evaluate the distribution of the error measures of a model across different locations. In the following, the relevant aspects of the results of the study in [17] are summarized. It compared different machine learning models for weather forecasting and real-time measurement based forecasting. The box plot shown in Figure 5.6 shows the distribution of the error measures of 29 wind farms in northern Germany. The error measure used here is the root mean square error (RMSE) which is applied to nominal power normalized time series. The individual boxes represent the error distribution of one of the six models used. The triangular markers indicate the confidence range of the median. If these ranges do not overlap for two models, the medians are different under normal distribution assumption to a 5% significance level. This corresponds to a visual representation of a t-test.

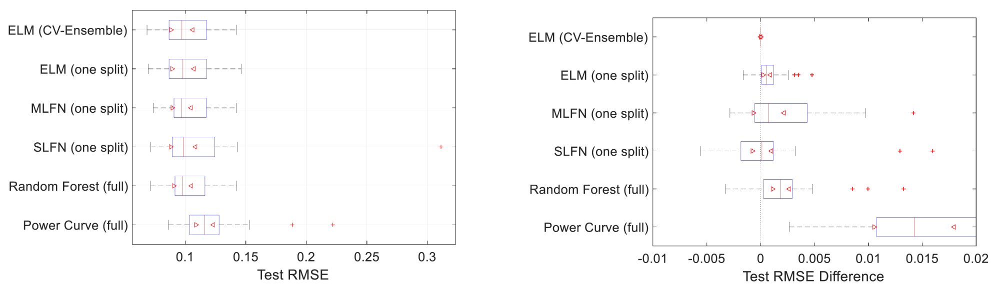


Figure 5.6: RMSE distribution for six different forecasting models forecasting for 29 wind farms in the North of Germany (left figure). Pairwise differences RMSE for each single model in comparison to the wind farm RMSE of the reference model ELM (CV-Ensemble) [17] (right figure).

Figure 5.6 (left) shows, that only the power curve model has a significantly higher RMSE.

All others cannot be clearly distinguished. The reason for this can be found in the broad distribution. This can be explained to a greater extent by the different local properties, such as the number of turbines per wind farm or the local orography. When considering the paired differences, local influences can be partially eliminated.

Figure 5.6 (right) shows the distribution of the difference between a model and a reference model (ELM (CV-Ensemble)) across all 29 wind farms. If the distribution of a model is significantly in the positive range, it can be assumed that the reference model is significantly better. Thanks to these pairwise differences, it can now be established that two other models have a significantly worse result.

5.5 Use cases

Key Points
The section presents a number of use cases that illustrate how an evaluation in a specific part of the power and energy sector should ideally be carried out. In the ***Energy Trading and Balancing, ramping forecast in general and for reserve allocation****, forecasts are today a crucial part of the processes at balance responsible parties, but also system operators. And yet, many mistakes are made in the evaluation and incentivization of forecasts that effectively often lead to results that are unsatisfactory and create mistrust in the ability of forecast service providers to have skills to provide useful forecasts.*

5.5.1 Energy Trading and Balancing

In energy trading forecasts of multiple variables are used in order to provide situational awareness and support quantitative decision making. Costs accrue on the basis of forecasts and energy prices at multiple look-ahead times. An example is forecasts used at the day-ahead stage and then again at an intra-day look-ahead time frame for the same trading period, and the relative price of buying and selling energy at different times.

Furthermore, prices, particularly imbalance prices, may be influenced by the cumulative forecasts and forecast errors of all market participants creating dependency between wind power forecast errors and the price at which resulting imbalances are settled. Similarly, unrelated events may cause large price movements that result in an otherwise unremarkable forecast error having a large financial impact. Therefore, care must be taken when designing an evaluation scheme that it is reflective of forecast performance and not externalities.

5.5.1.1 Forecast error cost functions

If trading decisions are based on a deterministic power production forecast, it is tempting to try and evaluate the 'cost' of forecast errors based on energy prices.

For example by taking the cost of under forecasting to be equal to the difference between the day-ahead price and the system sell price (the opportunity cost of having to sell at the system sell price rather than day-ahead price), and taking the cost of under forecasting to be equal to the difference between the system buy price and the day-ahead price (the cost of having to buy back the energy not produced at a higher price than it was sold for).

This approach has several problems:

1. price asymmetry:
   Traders are aware of the asymmetry in imbalance prices and have a view of whether the market is likely to be long or short, as such they do not naively trade the forecast production and will hedge against penalizing prices. It is therefore not representative to assume the day-ahead forecast is contracted.

2. adjustment opportunities:
   The intra-day market and flexibility within the traders portfolio provide opportunities for adjustment between the day-ahead market and imbalance settlement which may influence both the value and volume of traded energy, and potentially the imbalance price.

3. Forecast error correlation:
   Wind power forecast errors are highly correlated across the entire market and therefore to the market length and total imbalance. As a result, evaluating forecast errors based on imbalance cost will not discriminate between forecast performance and correlation with imbalance prices and one may incorrectly interpret reduced 'cost' as improved forecast skill.

For these reasons it is recommended that (normalized) mean absolute error be used as part of an evaluation matrix of other relevant metrics when evaluating deterministic wind power forecast performance for trading applications (see 4, 5.1). Additionally, a real-example of a market analysis and evaluation of how different trading strategies influence tne costs in comparison to the revenue can be studied at [**?**], and [**?**].

If trading decisions are based on probabilistic power production forecasts those forecasts should be evaluated as described in section 4.1.5. If probabilistic forecasts of both power production and prices are used it is important that the dependency structure between power and price forecast errors is correct. Various metrics exists to measure this, such as the multivariate energy score [**?**] and p-variogram score [**?**]. Details are beyond the scope of this document.

5.5.2 General Ramping Forecasts

Power ramps can have significant impact on power system and electricity market operation and are of interest to decision-makers in both domains. However, as ramps comprise a sequence of two or more forecasts, metrics that only compare predictions and observations at single time points are not suitable for evaluating ramp forecasts. Event-based evaluation in the form of contingency tables and associated metrics provide a tool-set for evaluating these forecasts.

Once an event is defined, such as ramp defined as a particular change in wind energy production over a particular time period, occurrences in forecasts and observations can be labeled and a table of true-positive, false-positive, true-negative and false-negative forecasts can be produced. From this, the skill of the forecast at predicting such events can be evaluated.

The definition of a ramp will influence the forecast tuning and evaluation results. It is recommended that the definition reflects the decision(s) being influenced by the forecast. For example, this could be related to a commercial ramp product definition, or the ramp rates of thermal power plant used in balancing. Furthermore, if an economic cost can be assigned to each outcome, then the forecasting system can be tuned to minimize costs, and the relative value of different forecasting systems can be compared.

In general terms, the following methods and metrics are recommended as basis for the evaluation of ramp forecasts:

- Contingency tables and statistics derived from the tables provide an evaluation framework
- Ramp definitions should reflect operational decision-making
- The cost implications of different types of errors should be considered when comparing different forecasting systems

In the next sections, a number of examples are described to demonstrate how evaluation should be planned and that illustrates the pitfalls in the metric selection process.

5.5.2.1 Amplitude versus Phase

Ramping events cause shortage or overproduction and risk for congestion in the power system for relatively short time frames. For this reason, many system operators have different levels of reserve time frames and also forecasting time frames that provide the possibility to allocate different types of reserve to counteract ramps that have been forecasted insufficiently strong (amplitude) and/or are wrong in phase. On system operator level it is often described that the amplitude is more important than the exact timing (phase).

In this case, it is necessary that the evaluation method does not punish the forecaster stronger for a phase error than an amplitude error. This means for example that using a root mean square error to evaluate ramps is incentivizing a forecaster to dampen amplitudes and optimize on phase. Sometimes it is referred to the **"forecaster's dilemma"** when the end-user defines a metric for evaluation such that the target is opposite of what the end-user asks for and needs. The forecast provider then either tunes forecasts to the metric or to what the end-user likes to see and risks to be punished (e.g. loose a contract), when evaluated. See also [**?**].

Recommendation: When a forecaster should be incentivized for amplitude in a ramp forecast, the evaluation metric cannot be an average error measure such as mean absolute error or root mean square error. If these average error metrics are used, the data to be evaluated has to be prepared to:

- reflect only cases that contain ramps of a certain strength
- widen ramp events with a forward/backward window of 1 – 2 hours to allow for phase errors

Additionally, either a contingency test with hit rate, misses and false alarms have to be used in the evaluation of the forecasts to reflect the focus on amplitude.

5.5.2.2 Costs of false alarms

Ramps can have different costs in a power system. In some systems, too fast up-ramping causes congestion or in some way over-production that needs to be dealt with (case 1). The opposite case, the down-ramping can cause that there is power missing on the grid that is not available and the fast primary reserve causes high costs (case 2). In case 1, the system operator has to be able to reduce ramping capacity of the wind farms or have other highly flexible resources on the grid to level out the overproduction. In case 2, lacking energy can cause high costs for fast ramping resources on primary reserve level or outages, which are unwanted.

The consequence is that the cost profile for up-ramping and down-ramping is usually different. Also, the cost of not forecasting a ramp that occurs (false-negative) can be significantly higher than the cost of preparing for a ramp, which does not occur (false-positive). The only way to verify, whether a forecast is sufficiently good in predicting a specific type of ramping event is to use contingency tables, where the forecast skill can be computed and visualised.

5.5.3 Evaluation of probabilistic Ramp forecasts for Reserve Allocation

The primary scope of reserve predictions is to reduce balancing costs via dynamic allocation of reserve and if possible with the help of non-fossil fuel capacity.
If a system operator (SO) or balance responsible party (BRP) can schedule reserve more dynamic, the costs for imbalances become lower and the energy system more efficient.

This was the scope of a study that will be presented as an example of the evaluation of a real-time environment application that needed a practical solutions in order to reduce costs for reserve allocation for the end-user [23]. The evaluation strategy and results of the study can be considered a kind of guideline on how to best manage renewable energy imbalances in a market system.

In this sample control area there are approximately 40 wind farms. The permanent allocation of reserves for the control area amounted at the outset to +/-10% and up to +/- 30% of installed capacity of wind, dependent on the time of the year, i.e. there are large seasonal reserve allocation differences. In our example area the wind generation is correlated and strong ramps occur. However, it is seldom to observe that the wind generation ramps down in a dramatic speed. Ramp-ups are faster than down-ramps and it is very unlikely that an instant total wind ramp down to zero can occur in the control area.

5.5.3.1 Definition of Error Conditions for the Forecast

Fundamental for forecasting is that a criteria for success and error can be defined. Given the fact that certain swings in the data are unrealistic or possibly so extreme that the operational cost of self-balancing would be too high, there was need to work with probabilities. One way of doing this is to define that, if a forecast value lies within a band, the result is a success and if it lies outside the band, it is a false alarm. A constant very wide reserve band would imply 100% success, but would not be affordable.

The gain lies in finding a balanced criteria considering the following questions:

- How many failures can be tolerated ?
- What is the allowed maximum error ?
- Which frequency of reserve under-prediction is allowed ?
- What is the cost of spilled reserve ?

These questions are related or determined by the SO's operational experience and standards to which the SO must be conform. Figure 5.7 illustrates the challenges of deciding how many outliers can be accepted to reduce costly spill, a dilemma every balance responsible party has to deal with. The static allocation of reserves is very expensive, especially if all extremes should be covered. Even, if extremes are not covered always, there is a lot of spill (black areas in Figure 5.7) in comparison to a dynamic allocation of reserves.

The difficulty for such a situation is to find objective criteria suitable for evaluation of a model result, which relates to operation and presents incentives for the forecaster to reduce the spill by maximizing coverage of extremes. Standard statistical metrics do not provide answers to this optimization task, because (1) it is not the error of 1 forecast any more and (2) the target is whether the allocation was sufficient and cheaper than allocating with a constant "security band".

With contingency statistics it is possible to ask the right questions:

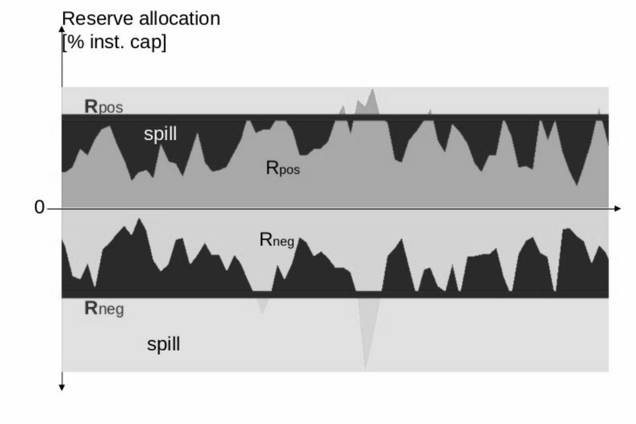


Figure 5.7: Illustration of the "reserve allocation dilemma" of costly spill versus covering all possible ramping events. Here, R_{pos} is the dynamic positive reserve, R_{neg} is the dynamic negative Reserve, the upper linear borders R_{pos} and R_{neg} are the static reserve allocation, the black area and the outer light gray areas are the spill for the dynamic and static allocation of reserves, respectively.

Hits and Misses Analysis show the percentage of time the band was too small Positive and negative reserve allocation can be split up to reflect use of tertiary reserve allocation (cheaper) instead of primary reserve (high expenses)

Table 5.3: Applied metrics in the evaluation matrix for the reserve allocation example in [23]. The input forecasts are split up in 9 percentile bands from P10..P90 and a minimum and maximum.

Metrics		Purpose	Input forecasts
BIAS		average to gain overview	MIN
MAE		average to gain overview	P10
RMSE		average to gain overview	P20
Inside Band		consistency forecast-deployment	P30
$R_{coverage}$		forecasted reserve deployment	P40
Hit rate	Total	achievable percent of activated reserve	P50
	R_{pos}	as above for pos reserve	P60
	R_{neg}	as above for neg. reserve	P70
Misses	Total	avg under-predicted reserve	P80
	R_{pos}	as above for pos reserve	P90
	R_{neg}	as above for neg. reserve	MAX
Spill	Total	avg over-predicted reserve	
	R_{pos}	as above for pos reserve	
	R_{neg}	as above for neg reserve	

The following analysis was carried out to reflect these objectives:

1. A BIAS, MAE and RMSE provide an overview of the plain statistical capabilities of the various forecasts

2. Contingency tables for hit rate, misses, spill and reserve coverage have been computed to provide metrics for further optimization of the task

Table 5.3 shows the evaluation matrix of metrics and their purpose in the verification and further optimisation. The study [23] concluded that the real reserve deployment will not be able to cover the shortage or overcapacity for about two hours per day in average. Their 5760 hours of evaluation was not considered very robust to draw final conclusions and to set long-term strategies, it was found that the results provided the information necessary to enhance the optimisation task and follow it's progress closely over some time.

Bibliography

[1] WWRP/WGNE Joint Working Group on Forecast Verification Research, Proc. of 7th International Verification Methods Workshop on Forecast Verification methods Across Time and Space Scales, Berlin, Germany, May 3-10, doi: http://www.cawcr.gov.au/projects/verification, 2017.

[2] Jensen, T., Fowler, T., Brown, B. Lazo, J., Haupt S.E. (2016), Metrics for evaluation of solar energy forecasts, NCAR Technical Note NCAR/TN-527+STR. Online available: `http://opensky.ucar.edu/islandora/object/technotes:538`

[3] Anemos.Plus Project DELIVERABLE D-1.3, Towards the definition of a standardised evaluation protocol for probabilistic wind power forecasts, 2012.
Online available: `http://www.anemos-plus.eu/images/pubs/deliverables/aplus.deliverable_d1.3-protocol_v1.5.pdf`

[4] Madsen H., Pinson P., Kariniotakis G., Nielsen HA, Nielsen TS., Standardizing the Performance Evaluation of Short-Term Wind Power Prediction Models, Wind Engineering, 29(6), 475489. doi:10.1260/030952405776234599, 2005.

[5] Diebold FX, Mariano RS, Comparing Predictive Accuracy, Journal of Business & Economic Statistics, 13(3), 253263. doi:10.1080/07350015.1995.10524599, 1995.

[6] Messner, JW, Pinson, P, Browell, J, Bjerregård, MB, Schicker, I. Evaluation of wind power forecastsAn up-to-date view. Wind Energy. 2020; 23: 1461 1481. https://doi.org/10.1002/we.2497

[7] Jensen, T. L., Fowler, T. L., Brown, B. G., Lazo, J. K., Haupt, S. E. (2016). Metrics for Evaluation of Solar Energy Forecasts (No. NCAR/TN-527+STR). doi:10.5065/D6RX99GG

[8] BRIER, G. W. (1950). VERIFICATION OF FORECASTS EXPRESSED IN TERMS OF PROBABILITY. Monthly Weather Review 78, 1, 1-3, https://doi.org/10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2.

[9] Epstein, E. S. (1969). A Scoring System for Probability Forecasts of Ranked Categories, Journal of Applied Meteorology and Climatology, 8(6), 985-987. Retrieved

Oct 12, 2021, from https://journals.ametsoc.org/view/journals/apme/8/6/1520-0450_1969_008_0985_assfpf_2_0_co_2.xml.

[10] Hudson, D., Ensemble Verification Metrics, European Center for Medium Range Forecasting (ECMWF) Annual Seminar 2017, https://www.ecmwf.int/sites/default/files/elibrary/2017/17626-ensemble-verification-metrics.pdf

[11] Murphy, A. H. (1973). A New Vector Partition of the Probability Score, Journal of Applied Meteorology and Climatology, 12(4), 595-600, https://journals.ametsoc.org/view/journals/apme/12/4/1520-0450_1973_012_0595_anvpot_2_0_co_2.xml

[12] WWRP/WGNE Joint Working Group on Forecast Verification Research, Methods for probabilistic forecasts, https://www.cawcr.gov.au/projects/verification/#Methods_for_probabilistic_forecasts, Accessed: 2021-10-11.

[13] Wikipedia, Receiver operating characteristic, howpublished = https://en.wikipedia.org/wiki/Receiver_operating_characteristic, note = Accessed: 2021-10-11

[14] Hamill, T.M. and Juras, J. (2006), Measuring forecast skill: is it real skill or is it the varying climatology?. Q.J.R. Meteorol. Soc., 132: 2905-2923. https://doi.org/10.1256/qj.06.25.

[15] Wilks 2011: Statistical Methods in the Atmospheric Sciences, Third Edition.

[16] Zhang 2015: A suite of metrics for assessing the performance of solar power forecasting. Solar Energy 111, 157.

[17] Vogt, S., Braun, A., Koch, J., Jost, D., Dobschinski, R.J.. Benchmark of Spatio-temporal Shortest-Term Wind Power Forecast Models, Proc. of the 17th International workshop on large scale integration of wind power into power systems as well as on transmission networks for offshore wind power plants, Stockholm, 2018.

[18] Gensler, A., Sick, B., Vogt, S., A Review of Deterministic Error Scores and Normalization Techniques for Power Forecasting Algorithms, 2016 IEEE Symposium Series on Computational Intelligence (SSCI), Athens, 2016, pp. 1-9. DOI: 10.1109/SSCI.2016.7849848, 2016. Online: https://www.sciencedirect.com/science/article/pii/S0169207008001003

[19] Friás Peredes, L., Stoffels, N., Statistical analysis of wind power and prediction errors for selected test areas, EU 7th Framework project Safewind, Deliverable Dp-7.1. Statistical analysis of wind power and prediction errors for selected test areas. Online available: http://www.safewind.eu/images/Articles/Deliverables/swind.deliverable_dp-7.1_statistical_analysis_v1.6.pdf

[20] Sańchez, I.: Adaptive combination of forecasts with application to wind energy. International Journal of Forecasting 24(4), pp. 679693, 2008. Online: https://doi.org/10.1016/j.ijforecast.2008.08.008

[21] Richardson, D.S., 2000: Skill and relative economic value of the ECMWF ensemble prediction system. Quart. J. Royal Met. Soc., 126, 649-667.

[22] Wilks, D.S., 2001: A skill score based on economic value for probability forecasts. Meteorol. Appl., 8, 209-219.

[23] Möhrlen, C., Jørgensen J.U., Reserve forecasting for enhanced Renewable Energy management, Proc. 12th Int. Workshop on Large-Scale Integration of Wind Power into Power Systems as well as on Transmission Networks for Offshore Wind Farms, Berlin, 2014.

Appendix A

Standard Statistical Metrics

Mean Absolute Error (MAE): *The average of all absolute errors for each forecast interval. Measures the average accuracy of forecasts without considering error direction.*

$$\frac{1}{n}\sum_{i=1}^{n}(f_i - m_i)$$

Mean Absolute Percent Error (MAPE): *This is the same as MAE except it is normalized by the capacity of the facility.*

Root Mean Square Error (RMSE): *Measures the average accuracy of forecasts without considering error direction and gives a relatively high weight to large errors*

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(f_i - m_i)^2}$$

Root Mean Square Percent Error (RMSPE): *As above normalize by plant capacity.*

BIAS: *Indicates whether the model is systematically under- or over-forecasting*

$$\frac{1}{n}\sum_{i=1}^{n}(f_i - m_i)$$

Correlation: Correlation is a statistical technique that is used to measure and describe the STRENGTH and DIRECTION of the relationship between two variables.

$$r(x,y) = \frac{COV(x,y)}{STD_x \cdot STD_y} = \frac{\sum (x-\bar{x}) \cdot (y-\bar{y})}{N \cdot STD_x \cdot STDy}$$

where f are the forecasted values, m are the measurements, COV is the covariance, STD is the standard deviation.

Standard Deviation: A measure of the spread or dispersion of a set of data. The more widely the values are spread out, the larger the standard deviation. It is calculated by taking the square root of the variance.

$$\mathrm{STD} = \sqrt{\left(\frac{\sum \left((f_i - \bar{f}_i)^2\right)}{n}\right)}$$

Variance: A measure of the average distance between each data point and the data mean value; equal to the sum of the squares of the difference between each point value and the data mean.

$$\sigma^2 = \frac{\sum \left((f_i - \bar{f}_i)^2\right)}{n}$$

RECOMMENDED PRACTICES FOR THE IMPLEMENTATION OF RENEWABLE ENERGY FORECASTING SOLUTIONS

- Part 4: Meteorological and Power Data Requirements for real-time forecasting Applications-

1. DRAFT EDITION 2021

Draft for Review by Stakeholders prior to submission to the Executive Committee of the International Energy Agency Implementing Agreement in September 2021

Prepared in 2021 as part of the IEA Wind Task 36, WP 3.3.

Document Version: 1.0
October 12, 2021

Contents

Preface

This recommended practice document is the result of a collaborative work that has been edited by the undersigning editors and authors in alignment with many discussions at project meetings, workshops and personal communication with colleagues, stakeholders and other interested persons throughout the phase 1 of the IEA Wind Task 36 (2019-2021) as part of workpackage 3.3.

The editors want to thank everybody that has been part of the meetings, workshops and sessions and contributed in the discussions, provided feedback or other input throughout the past 3 years.

IEA Wind Task 36, October 12, 2021

Editors and Authors:
Dr. Corinna Möhrlen (WEPROG) <com@weprog.com>
Dr. John Zack (UL AWS Truepower) <john.zack@ul.com>

Contributing Authors:
Dr. Jethro Browell (University of Glasgow, UK)
Dr. Stefan Wilbert (German Aerospace Center, DLR, Spain)
Ines Würth (University of Stuttgart, Germany)
Dr. Martin Felder (ZSW Baden-Württemberg, Germany)

Contributions:
Dr. Jie Yan (North Chine Electric Power University, P.R. China),
Rui Pestana (REN, Portugal)
Anton Kaifel (ZSW Baden-Württemberg, Germany)
Jan Remund (Meteo , Switzerland)
Dr. Carmen Köhler (P3R Solutions, Germany)

Supported by:
Operating Agent Dr. Gregor Giebel (Danish Technical Univerity, DTU Wind, Denmark)

Chapter 1

Background and Objectives

1.1 BEFORE YOU START READING

This is the forth part of a series of four recommended practice documents that deal with the development and operation of forecasting solutions in the power market. The first part Forecast Solution Selection Process deals with the selection and background information necessary to collect and evaluate when developing or renewing a forecasting solution for the power market. The second part Design and Execution of Benchmarks and Trials, of the series deal with benchmarks and trials in order to test or evaluate different forecasting solutions against each other and the fit-for-purpose. The third part Forecast Solution Evaluation, which is the current document, provides information and guidelines regarding effective evaluation of forecasts, forecast solutions and benchmarks and trials.

1.2 Introduction

Meteorological measurements provide an independent measure of the wind and solar resource and weather situation at any given time. This information can and is, as technology enhances, not only an obligation that stems from technical requirements of the system operator, but is also used to optimise the operation of renewable power plants and electricity grids. For both the system operator and the power plant operator, these measurements are an independent signal at the power plant that can warn about critical weather and provide an indication on whether the power plants work at their expected performance level. For the transmission system operator, such measurements can additionally be used for situational awareness of the weather in the control area that may affect the transmission lines. They also provide a second means of verification, whether the power signal at a given power plant is malfunctioning in situations that may be critical in terms of grid operation.

Data assimilation with independent measurements from power plants have also been tested by meteorological centres (e.g. [39, 21]). One of the most important findings so far is that the quality of data provided is the most essential issue to be solved in order to gain higher quality forecasts with such measurements. In other words, if there is no specific effort put into standardisation of requirements in the power industry, the benefits can not be achieved.

This recommended practice (RP) aims to provide background information on meteorological instrumentation, their recommended setup, maintenance, quality control and use in the real-time environment of wind and solar power generation plants. Information on use of real-time power and other operational data in forecasting systems is also included. The document provides practical guidelines for quality control and recommended standards to be applied in the setup and calibration of instrumentation when entering the operational phase of wind and solar projects and for applications relevant to plant operators, system operators, power traders or balance responsible parties.

1.3 Using this document

This document is structured such that the chapters are ordered following a decision making process for the use of measurements for real-time forecasting applications.

- **Meteorological instrumentation** To forecast wind or solar power, meteorological conditions such as wind speed or solar radiations need to be known. There are several measurement devices in use, which all come with different properties and requirements. Chapter 2 gives an overview of those instruments.

- **Power measurements** as a key input to very short-term forecasts (minute to hours ahead) and to check the operational capacity and actual produced power produced to validate forecasts. Chapter 3 gives an overview of relevant data types and useage.

- **Setup and calibration** When working with measurement data, it is crucial to follow best practices for the measurement setup and to calibrate the instruments properly to ensure the highest standard of data quality. Chapter 4 gives an overview of different measurement setups and calibration procedures.

- **Instrumentation performance** There is a saying “he who measures measures garbage”, meaning that measurements are prone to errors and therefore checking the performance of measurement instrumentation and signals is crucial to ensure high quality of forecasts. Chapter 5 gives an overview of measures for quality control of measurements.

- **Best practices Summary** Chapter 6 is the short-cut to the recommendations in a nutshell. Here, we collect and summarise all recommendations for all steps in the use of measurements in real-time forecasting applications wihout explanations or justification and link them to the respective backgound information.

1.4 Use and Application of real-time Meteorological Measurements

In this section, we will define and discuss the use of meteorological measurements for real-time application and especially make the distinction to non-real-time use and applications such as resource assessment.

1.4.1 Application-specific Requirements for Meteorological Data

Meteorological measurements made at the sites of current or future wind or solar generation facilities can be used for several different types of applications. Although there are many common requirements, there are significant differences in the required attributes of the meteorological data gathering among these applications. The primary applications for on-facility meteorological data are:

1. **Short-term Forecasting**
   Short-term (minutes to days ahead)forecasting utilizes on-facility meteorological data for the training of forecast systems ans well as input into the production of forecasts. It is the "input to forecast production" application that creates the most significant requirements since in order to have maximum value the data must be quickly quality controlled and available to the forecast production process with minimal time lags. This application is the primary focus of this document .

2. **Resource Assessment**
   The resource assessment application employs meteorological data to generally assess the generation resource over a region (resource mapping) as guidance for the planning of future development or at a specific location (generation yield estimation)for the purpose of estimating the fture production of a proposed project in order to assess its the economic viability. This does not require near-real-time access to the meteorological data .

3. **Generation Asset Performance Assessment**
   The performance assessment application uses meteorological data to estimate the expected generation from a wind or solar facility in order to assess the performance of the generation hardware at a specific facility relative to the manufacturer's performance specifications. This does not generally require near-real-time access to the meteorological data

4. **Turbine Control**
   The turbine control application is specific to wind generation facilities and uses meteorological data to optimally orient the turbine for maximum generation efficiency. Typically only wind speed and direction data is required in real-time so that timely adjustments to the turbine position can be made.

5. **Plant Monitoring and Situational Awareness**
   The plant monitoring and situational awareness application employ meteorological data to facilitate the awareness of plant operators and possibly also grid system operators to the operating status of the generation facility and the likelihood of very-short-term changes in production. Meteorological data is employed as a independent check on the generation and availability status of the facility and also to identify very short term trends in the resource that may not be evident from the generation data (such as in the case of wind generation facilities that are operating at rated power but may be near thresholds sich as high speed shutdown). The near real-time availability of data is important for this application.

A summary of the differences in measurement attributes among these applications is provided in Table 1.1. The most significant point is that the short-term forecasting application has specific requirements that are different from those of other applications and therefore a meteorological measurement system designed primarily for use in a different type of application may not be ideal or even adequate for short-term forecasting. :

1.4.2 Applications in System Operation, Balancing and Trading

The key applications in system operation, balancing and trading of wind and solar power, where real-time meteorological measurements are required are:

- **Situational awareness in critical weather events**
  Critical weather events can cause severe security risks in grid operation. Forecasting and measurements assist to a large extend to make such events predictable and provide the necessary information to be able to act in advance. Such situational awareness is an important planning tool in grid operation, where penetration levels are above ca. 30% of energy consumption. Especially for wind generating power extreme winds can only be predicted by meteorological signals, due to the flattening of wind power curves in the wind ranges > 12m/s. In this range, the power signal provides no local information to the system operator or the wind farm operator/balance responsible.

- **High-Speed Shutdown events**
  During storm events, the critical parameter for the grid operation is the proportion of wind farms that are expected to enter into high-speed shutdown (HSSD) in any high penetration area. The risk and increased uncertainty for HSSD during storm events can result in the System Operator having to limit the wind generation in advance so that sufficient reserve is available.

- **Grid related down-regulation or curtailments**
  Grid related down-regulations or curtailments can be due to extreme weather events or technical problems at the electrical lines or controllers. Wind or solar radiation measurements can provide an independent signal to the system operator on the available

Table 1.1: *Differences in required attributes of meteorological measurements among the major renewable energy applications*

Attribute	**Forecasting**	**Resource Assessment**	**Performance Assessment**	**Turbine Control**	**Plant Monitoring**
		General			
Availability in near real-time	essential	not necessary	not necessary	essential	essential
Availability in extreme weather conditions	essential	not critical	not critical	essential	essential
Measurement bias	not critical if trends are accurate	near zero bias critical	low bias important	not critical	not critical
Additional met variables (other than resource variables)	useful especially in specific weather scenarios	minimal value	not needed	not needed	useful in some weather scenarios
Plant outages/curtailment	essential	not relevant	not relevant	not relevant	useful
		Wind Power related			
Deployment location considers wakes	important	not relevant	not an issue	not an issue	not critical
		Solar Power related			
Obstacle Shading	must be representative of facility	must be representative of site	must be representative of facility	not applicable	not critical

active power potential, where this is not broadcasted and also provide the possibility to compute the lost power production due to the down/regulation.

- **Short-term Forecasting with updates from measurements**
  Short-term forecasting with time horizons of minutes up to a few hours need to rely on measurements as additional information to weather data from NWP models, due to the 6-hourly schedule of most operational NWP model deliveries. Improvements can however only be expected, if the real-time measurements to update the forecasts have a sufficient quality. If too much of the data needs to be blacklisted, or the filters have difficulties to capture erroneous measurement values, such data can actually make

the forecast worse than without the additional data. Quality of these data is therefore imperative.

- **Intra-day Power plant balancing**
  Intra-day power plant balancing is equally dependent on a high quality measurement data quality as the short-term forecasting, and usually relies on short-term forecasts.

1.4.3 Applications in Wind turbine and wind farm operation

The key applications for wind plant operation, where real-time meteorological measurements are required, are:

- **Wind turbine control**
  Due to wake effects on nacelle anemometer, independent site data from a met mast or LIDAR can assist the turbine controller to work more safe and efficient. Preview information of the turbine inflow from a nacelle-based lidar system can be used to induce blade pitch action and thus reduce loads and improve turbine power performance.
- **Wind farm control**
  Wake measurements of scanning lidars or nacelle-based lidars that measure the wake of turbines, can assist to redirect those wakes and thus reduce loads on downstream turbines and increase their power production.
- **Condition Monitoring**
  Knowledge about the wind conditions that affect a wind turbine or wind farm help to estimate the load budget that the turbines have experience during their lifetime. This information can be used for lifetime extension measures thus increasing the energy yield of a wind farm and its profit.

1.4.4 Solar/PV plant operation

The key applications for solar plant operation, where real-time meteorological measurements are required are:

- Yield forecasting for the next minutes, hours and days
  - ramp rate control in PV power plants
  - dispatch optimisation of power plans with storage
  - optimisation of power plant operation for thermal plants (e.g. start up decisions)
- Power Plant Control
  - optimisation of tracking of tracked PV or concentrating collectors
    * stow position for high wind speeds/gusts

    * optimal tracking angle for non concentrating tracked PV according to sky radiance distribution and shading
    * de-focusing of concentrating collectors at high DNIs
  – mass flow control in thermal collectors
  – aim point control for solar tower plants

Most of the real-time meteorological measurements are only required for large scale power plants. For tracking collectors of any size wind speed and direction measurements are required to allow for securing the collectors in the stow positions at high wind speeds. Depending on the solar thermal technology also further meteorological measurements may be required even for small systems.

In the handbook [55] and [56], it is noted that measurements of irradiance are complex and more expensive as general meteorological instruments, especially for measuring direct normal irradiance (DNI). The applications named in [55] and [56] developers utilise for irradiance data are:

- Site resource analysis
- System design
- Plant operation
- Developing, improving and testing models that use remote satellite sensing techniques or available surface meteorological observations
- Developing, improving and testing solar resource forecasting techniques.

1.5 Available and applicable Standards for real-time Meteorological and Power Measurements

There are to date no standards available that deal with real-time meteorological and power measurements for real-time wind and solar forecasting applications. Nevertheless, there are a number of standards available from the three areas of meteorological monitoring, wind and solar power resource assessment and monitoring that provide useful guidelines for measurement collection, monitoring, setup and calibration of instrumentation, verification and validation of measured data etc. In this section we introduce existing guidelines and standards from these three areas and analyse them for their applicability in real-time wind and solar forecasting environments.

1.5.1 Standards and Guidelines for Wind Measurements

For resource or site assessment in the planning phase of a wind farm an IEC standard exists [IEC, 2005] with an updated version 2 (IEC 61400-12-2:2013), that specifies which tests and what kind of criteria the instrumentation has to fulfil when used for the required tests to be carried out. The IEC 61400-12-2:2013 rules contain the following items:

- Extreme winds
- Shear of vertical wind profile
- Flow inclination
- Background turbulence
- Wake turbulence
- Wind-speed distribution

The results of these tests have to be within a pre-defined range to be acceptable. In Appendix F of the 61400-12-1:2005 "Cup anemometer calibration procedure" the calibration of the instruments for measuring wind are specified.

The use of remote sensing for wind measurement was introduced in a new version 61400-12-1:2017. In Annex L guidelines for the classifications of remote sensing devices, for the verification of the performance and for the evaluation of uncertainties of the measurements are given.

MEASNET (MEASuring NETwork), the "international network for harmonised and recognized measurements in wind energy" has defined so called "Round Robin rules" for calibration of cup anemometers for wind energy [MEASNET, 2009], which are widely used. MEASNET has also under the EU project ACCUWIND published a number of guidelines regarding instrument calibration and measurement campaigns for the wind industry (Dahlberg et al., 2006, Pedersen et al. 2006, Eecen, 2006]. Lee [2008] found a way of calibrating wind direction sensors with an optical camera.

In 2016 MEASNET published Version 2.0 of a Procedure for the evaluation of site-specific wind conditions [41]. This document gives guidance on measuring wind characteristics, and comprises an annex on wind measurement using remote sensing devices. It also includes guidance on how to set up measurement campaigns depending on the data required.

IEA Wind Task 32 and Task 11 published recommended practices RP-15 for ground-based remote sensing for wind resource assessment in 2013 [65]. It covers different aspects of using lidars and sodars. An updated review version of 2020 [12] identifies recommendations from the relevant normative documents (RP-15, MEASNET 2016, IEC 61400-12-1:2017 Annex L) concerning characterisation, installation, operation, data analysis and verification of wind lidar .

The Annex D in IEC 61400-12-1:2005 standard states that the "implicit assumption of the method of this standard is that the 10 min mean power yield from a wind turbine is fully explained by the simultaneous 10 min mean wind speed measured at hub height, and the air density" [IEC, 2005, Annex D, Table D.1] and describes the associated measurement uncertainty evaluation principles. In this respect, the standard refers to the "ISO Guide to the expression of Uncertainty in Measurements" [20, 6], and its 2 supplements [59, 60] from the Joint Committee for Guides in Meteorology (JCGM), where there are two types of measurement uncertainty that are to be accounted for in any standardised measurement taking:

1. systematic errors, often associated with offsets/bias of the measured quantity

2. random errors, which are associated with the fact that 2 measurements of the same quantity are seldom the same

In section 3.1.2 of the guide, [[59, 60] it is stated that "the result of a measurement .. is only an approximation or estimate .. of the value of the measurand and thus is complete only when accompanied by a statement of the uncertainty ... of that estimate". Considering this definition, all measurements should ideally have an uncertainty term associated with it. This is impractical in real-time operations, where the value of the measurements lies in the availability of the data at a given time. Therefore, it is unrealistic to request uncertainty measures. However, it could be a standing data value that is determined at the setup of the instrument and provided as part of the standing data. In that way, the instrument specific uncertainty could be accounted for in the handling of measurements (for other mitigation methods see 5.3.2.

In the introduction to the Guide [6], it is stated that ..the principles of this Guide are intended to be applicable to a broad spectrum of measurements, including those required for:

- maintaining quality control and quality assurance in production
- complying with and enforcing laws and regulations
- calibrating standards and instruments and performing tests throughout a national
- measurement system in order to achieve tractability to national standards developing, maintaining, and comparing international and national physical reference standards, including reference materials

To summarise, the handling and integration of wind power into the electric grid is an equally important step to harness the full potential of the energy resource in an efficient and environmentally friendly way.

This requires that measurements are trustworthy and maintained to a quality that allows for their use in forecasting tools in order to produce high quality forecasts and thereby reduce the need of reserves. These guides in combination with the IEC 61400-1 standard would provide a good foundation for any grid code technical requirement specifications.

1.5.2 Standards and Guidelines for Solar Measurements

A general guideline for meteorological measurements for solar energy is available in the IEA PVPS handbook [56]. The handbook summarises important information for all steps of a solar energy project - reaching from required measurements and the design of measurement stations to forecasting the potential solar radiation. Measurement instruments and their application as well as other sources for solar measurement data are presented. The handbook links to relevant further standards and guidelines. Here we briefly summarise those standards and guidelines that are relevant for real-time applications. For resource assessment purposes, we refer to the IEA PVPS handbooks original Version [54], version 1[55] and 2 [56].

One of the frequently cited guidelines for measuring radiation is chapter 7 of the WMO CIMO guide [66]. There, guidelines are presented for meteorological measurements in general which mostly apply also for solar energy related measurements. However, some guidelines are different for solar energy as the conditions relevant for the power systems must be measured. For example, these conditions are often affected by nearby objects such as the solar collectors themselves, trees or buildings which should be avoided for other meteorological measurements.

In the ISO 9060 standard radiometers are classified according to their measurement errors caused by different effects, such as sensor temperature, or calibration stability. Classes are defined for both pyranometers (used to measure global radiation of the hemisphere above the sensor) and pyrheliometers (used to measure the direct normal irradiance). There are several ISO and ASTM standards for the radiometer calibration (ISO, 9059, ISO 9847, ISO 9846, ASTM G207, ASTM824, ASTM167). Guidelines for the application of pyranometers can be found in ISO 9901 and guidelines for the application of pyranometers and pyrheliometers are found in ASTM 183.

IEC 61724 on PV system performance monitoring describes how radiometers and other meteorological instruments should be integrated and used in PV plants. Accuracy classes of the resulting monitoring systems are defined. Also the number of sensors depending on the peak power of the PV system is given. Additionally, the standard defines cleaning and calibration intervals for pyranometers.

1.6 Standards and Guidelines for general meteorological measurements

As mentioned in section 1.5.1 and 1.5.2, the *Guide to Instruments and Methods of Observation* from the WMO (world meteorolgy organisation) [66] is standardising instrumentation for surface winds in chapter 5 and for radiation in chapter 7. Other meteorological parameters covered in the WMO Guide are measurement of:

- Surface and upper air temperature ([66] in chapter 2 and 10)
- Atmospheric and upper air pressure ([66] in chapter 3 and 10)
- Surface and upper-air Humidity ([66] in chapter 4 and 10)
- Surface and upper air wind ([66] in chapter 5 and 13)
- Sunshine Duration (([66] in chapter 6)
- Visibility ([66] in chapter 8)
- Evaporation ([66] in chapter 9)
- Clouds (([66] in chapter 15)

- Atmospheric Compisitions ([66] in chapter 16)

The United States Environmental Protection Agency (EPA) provides a Meteorological Monitoring Guidance for Regulatory modelling Applications [42], which is a guideline on the collection of meteorological data for use in regulatory modelling applications such as air quality. It provides recommendations for instrument, measurement and reporting for all main meteorological variables used in meteorological modelling. In Section 4 of the guideline, the EPA provides recommended system accuracies and resolutions for especially wind speed, wind direction, ambient and dew point temperatures, humidity, pressure and precipitation, which are useful for wind and solar applications as well and will be discussed in the measurement setup and calibration section 4.

These guidelines and recommendations have been assessed for the purpose of wind and/or solar projects and have been basis for a number of our best practice recommendations in section 6.

1.7 Data Communication

An important component of the gathering of on-facility meteorological data for short-term forecasting is the timely retrieval of the data from the sensors and dissemination to the forecast production process. This requires efficient and standardized data communication protocols and well-documented descriptions and formats of the data elements so that the data can be correctly and efficiently interfaced with the forecast production process. Part of this process will be dictated by the type of sensor hardware and software used to make the measurements and retrieve the data from the sensor site.
However, there will typically also be a component of that process that provides an interface between the process that gathers data from the sensors and the process that disseminates data to the forecast providers. Proposed standards and options for the data descriptions, format and exchange protocols for the latter process are presented in section 4 of this Recommended Practice Part 1 “Forecast Solution Selection Process”.

Chapter 2

Meteorological Instrumentation for real-time operation

> ***Key Points***
> *In this section currently applied as well as instrumentation that is under development is being described...*

The purpose of meteorological measurements as supplement to the power measurements at wind and solar plants is to provide a measure of the resource and weather situation at the specific location and any given time. This information should not only come from an obligation to wind and solar plant operators defined in the technical requirements of system operators, but has equally much become a tool to optimise the operation of wind turbines and solar plants by the operators.

For both the system operator and the plant operator, these measurements are an independent signal at the plant location that can warn about critical weather and provide an indication on whether the wind turbines and solar panels work at their expected performance level. For the transmission system operator, such measurements can additionally be used for situational awareness of the weather in the control area that may affect the transmission lines. They also provide a second means of verification, whether the power signal at a given wind or solar plant is malfunctioning in situations that may be critical in terms of grid operation. Meteorological centres have also shown an interest in these data for the data assimilation of the numerical weather prediction models (e.g.[39, 21]) that typically lack observational information at hub heights.

One of the most important findings so far has been that the quality of data provided is the most essential issue to be solved in order to gain higher quality forecasts with such measurements. In fact, it has been identified that if there is no specific effort put into standardisation of requirements in the power industry, the benefits cannot be achieved.

The following is a list of typical instrumentation used for wind and solar projects that will be subject for this recommended practice guideline and recommendations made in how to setup these instruments and which implications the use of the various instruments have on data quality and usability in the operational real-time context in the energy industry.

2.1 Instrumentation for Wind projects

Typical instrumentation for meteorological measurements in wind power context are divided into two categories:

- Met mast
  1. Lattice masts
  2. Telescope masts
- Steel cabinet
  1. Data logger
  2. Communication system
  3. Components for the power supply
  4. Additional system components
     - Anemometers
     - Wind vanes
     - Temperature humidity sensors
     - Air pressure sensors
     - hygrometer sensors
     - precipitation sensors rain gauges
- Remote Sensing Systems
  1. LiDAR
     - Wind Profiling LiDAR
     - Scanning LiDAR (Long-Range and short-range)
     - Nacelle-based LiDAR
  2. SoDAR

Not so common instrumentation or additional instrumentation for wind farms are:

- Microwave Radiometers (measures energy emitted at sub-millimetre-to-centimetre wavelengths at frequencies of 11000GHz)
- Ceilometer (light source to determine the height of a cloud base. Ceilometers can also be used to measure the aerosol)
- Microbarographs (measures atmospheric pressure)

These instrumentation are more commonly used in research measurement campaigns and in meteorological projects. Literature on these types such as microwave radiometers are described by e.g. [62, 40], for ceilometers by [49], or microbarographs by [48].

Meteorological mast are still the most commonly used measurement instrumentation for the planning phase and operation of wind farms. An example is shown in Figure 2.1.

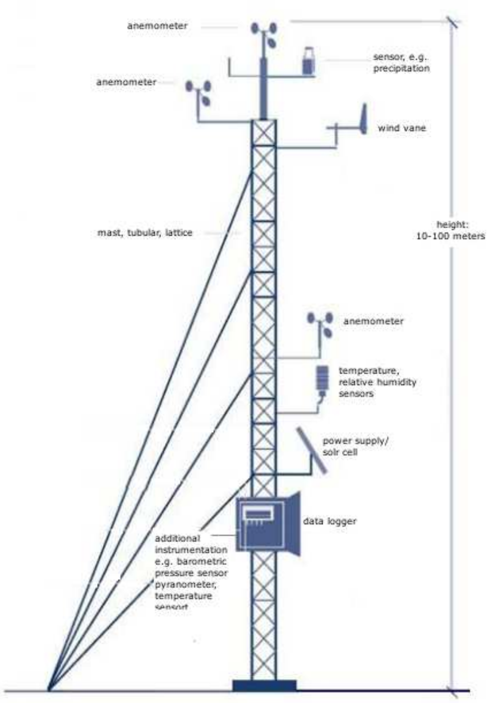


Figure 2.1: Example of a typical met mast with various instrumentation, data logger, power supply and wiring. Typical instrumentation are cup anemometers or sonic anemometers, wind vanes, temperature, pressure and air density sensors, pyranometers and precipitation and humidity sensors.

2.1.1 Remote Sensing Instrumentation for wind farms

Remote sensing has a long tradition in geology, atmospheric science, hydrology and oceanography and other earth sciences. The earliest remote sensing "devices" were aerial photography that was analysed with the heights and the geographical space in which the pictures were taken.

Today, remote sensing devices are used in wind energy applications to measure wind speed and direction [13]. They emit radiation or sound that interacts with the atmosphere, e.g. with aerosols that move along with the wind speed. The sensor measures the signal that is reflected or back-scattered from the particles and has experienced a frequency shift due to the Doppler effect. The frequency shift is proportional to the line-of-sight (LOS) wind speed along the direction of the emitted beam. Horizontal wind speeds and directions are determined then determined using wind field reconstruction algorithms.

Examples of such devices are RADARS (Radio detection and ranging Sensor), SODARS (SOund Detection and Ranging Sensor) and LiDAR (Light Detection and Ranging Sensor). These sensors can be deployed on the ground, on aircraft or satellites. For wind energy applications such as site assessment, mostly ground-based devices are used that measure the LOS wind speed vertically and then a horizontal wind speed and wind direction is derived.

The quality of remote sensing data is very much dependent on its spatial, spectral, radiometric and temporal resolutions.

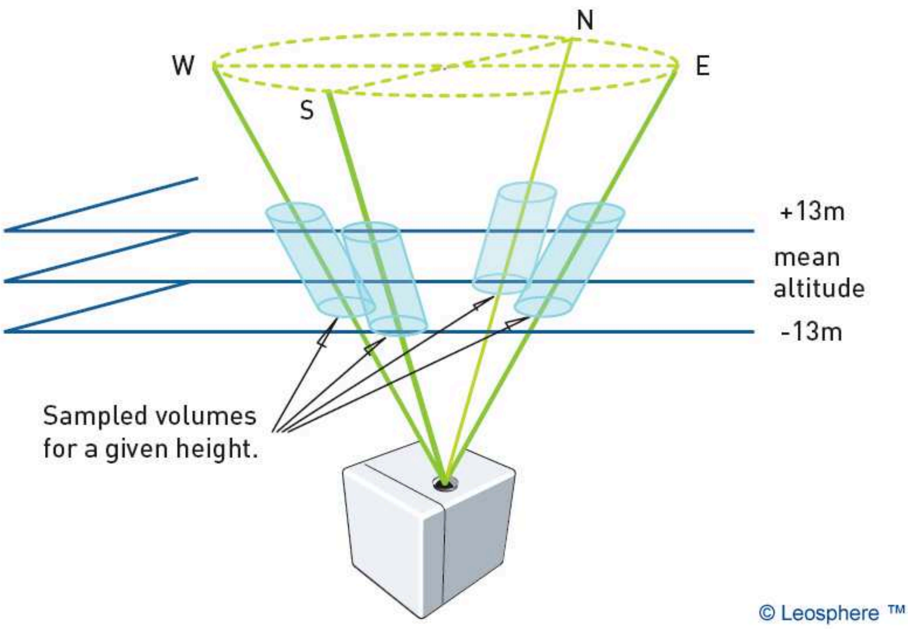


Figure 2.2: Principle of the remote sensing devices scanning. Picture shows a Windcube from Leosphere.

Ground based remote sensing in wind energy has until recently mostly be driven by a desire to find alternative measurements for expensive and at times difficult installation and erection of met masts. Especially with increasing hub heights, met mast heights have grown to a size, where the erection requires planning permission and cranes of significant size. Hence, it has become so expensive that previously never considered alternatives from the remote sensing area have become price competitive.

The main driver of recent developments has been the competitiveness in price, the ease of installation and the increasing heights of wind turbines and size of the projects, where it is often no longer sufficient to measure at only one site. Nevertheless, the disadvantage of not directly measuring the target value is still present [68]. With increasing experience and technical advances in computational science and technology, the remote sensing devices have however become a real alternative.

This has also been reflected in the IEC 61400-12:2005 standard, where remote sensing devices have been incorporated as possible devices to carry out wind measurements for wind energy applications in the 2017 update (IEC 61400-12:2017).

Looking at the benefits outlined by manufacturers and users of remote sensing devices the following list of key advantages of using remote sensing devices in wind energy applications can be summarised to:

- Minimal environmental impact
- Short installation time
- Highly portable
- Short lead times
- Wind profiling over the whole rotor area up to 300m
- Flexible measurement heights

The main technical advantage to be considered is the ability to measure over an area or volume rather than at pre-defined fixed heights above the ground. This is also how forecasting models work. NWP models compute variables across grid cells as area averages and area verification of variables are widely applied for model verification in meteorology.

The drawbacks of remote sensing devices so far have been for both SODAR and LiDAR inaccuracies of signals in complex terrain. According to the white paper of the Deutsche Windguard and DTU [5] and also Bradley [11], especially "in complex terrain sites, influence of the relatively large scanning volume of LiDARs and SODARs must be carefully considered in terms of its influence on the measurement accuracy.". This has been a general observation and a large research topic [see e.g. [19, 11, 10, 38, 28]]. Issues and preliminary recommendations of using lidar in complex terrain are summarised by a group from IEA Wind Task 32 in [3]. The task also formed a working group which is carrying out a group study on several different transfer methods for the use of lidar in complex terrain. Results are expected in 2021.

When used in real-time applications, remote sensing devices require special treatment. Due to the physics of the measurement principle, the measurement availability is depending on environmental conditions [67]. If there are too many (fog) or too few aerosols in the atmosphere (after a rain shower), a LiDAR does not provide a useful signal. Filter algorithms need to detect these data gaps and the real-time forecast algorithms needs to be able to handle them. The common findings of all the experimental measuring campaigns as well as real-time testing is that the instruments need to be well serviced and are maintained similar to any other real-time instrument operating under changing conditions throughout the yearly cycles. If this is not done, echoes, interfering noise sources, laser beam disturbances deteriorate the instruments and make the further processing of the data impossible. It is also commonly understood that it requires skilled personnel to install and maintain such instrumentation, if it should run continuously and reliably. For a real-time application it is additionally crucial that the measurement signals can be used as is and need no further processing.

2.1.2 Nacelle instrumentation and measurements

Among the nacelle measurement devices there are three types that are commonly used:

- cup anemometers
- horizontally mounted LiDAR
- (ultra-) sonic anemometers

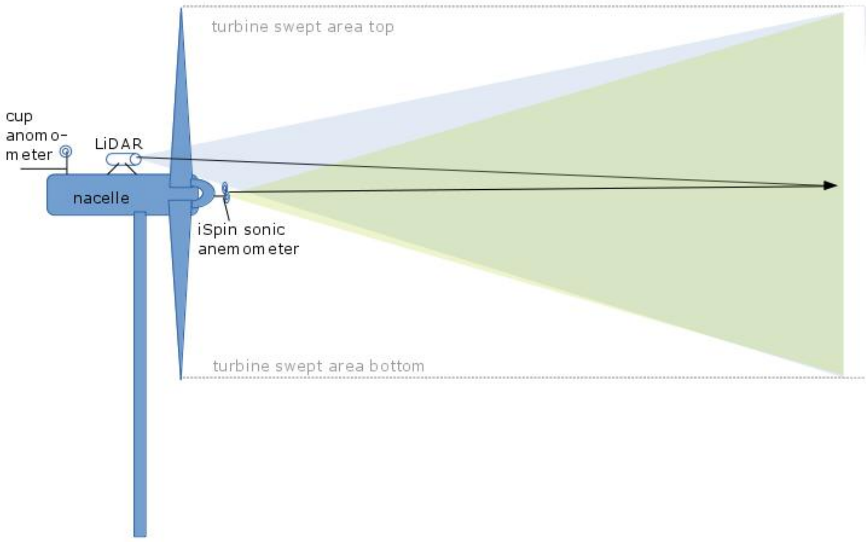


Figure 2.3: Schematic of the nacelle mounted instruments cup anemometer, LiDAR and iSpin ultra-sonic anemometer. The latter two instruments look forward into the wind eye and cover the rotor swept area.

.

Figure 2.3 shows a schematic of the instruments and how and where they are mounted at the turbines nacelle. The cup or (less often) sonic anemometers are typically mounted at the back of the nacelle. The horizontally mounted LiDAR is mounted approximately in the middle of the nacelle with a slight displaced angle in order to cover the total swept area of the rotor upwind of the turbine. The iSpin ultra-sonic anemometers are mounted at the front of the spinner, measuring wind characteristics directly in front of the rotor.

2.1.3 Cup anemometers

Most commonly cup anemometers with wind vanes for direction measurements are installed at the nacelle. There are the IEC 61400-12-1, the 61400-12-2 and the ISO/IEC 17025 standards that describe how these instruments must be calibrated and mounted as well as describing the process and the integrity of the measurement processes and design of the mast, instruments and measuring procedures. This will also be discussed in the standards

analysis in Section **??**. In this section, we only discuss, whether and how the data from cup anemometer instrumentation at the nacelle can add value to forecasting.
The cup anemometers at the nacelle have one distinctive advantage over any other instruments: they are installed at the turbine and connected to the SCADA system that is delivering data to the system operator. However, this advantage comes with a downside: the measurements taken at the nacelle are affected by two major disturbances: (1) the rotating turbine blades, which extract energy from the wind speed and increase turbulence and (2) wake effects from other upstream wind turbines. To counteract (1) the cup anemometers are usually calibrated with a nacelle transfer function which transfers the signal measured behind the rotor into a free stream measurement. However, in the worst collection hours both phenomena disturb the signal of the nacelle instrument and the signals can cause deterioration of the forecast, if they are used in the data assimilation phase.

Figure 2.4: MEASNET certified cup anemometer from Cambell Scientific and a cup anemometer 40C from RNRG.

2.1.4 Sonic and ultra-sonic anemometers

The sonic and ultra-sonic 3D anemometers have a long tradition in atmospheric science and meteorology in relation to boundary layer studies of turbulence intensity and phenomena like low level jets. These instruments are well tested and can be used for real-time operations, but are mostly considered too expensive for traditional wind measurements [e.g. Berg et al. 2012, Popinet et al., 2006, Basu et al., 2004, Lundquist, 2014].

Figure 2.5 shows a mast with an ultra-sonic anemometer at the NREL test site in Golden, Colorado and a well-tested 3D sonic anemometer from Campbell Scientific.

A newer type of sonic anemometer are the so-called ultra-sonic 3D spinner anemometer instruments, short iSpin, which have found their way into instrumentation for wind energy. Figure 2.6 shows the principle of the iSpin anemometer from ROMOWind and an installed spinner anemometer example from METEK. With the update of the IEC standard 61400-12-2, the iSpin technology has become part of the measurement types to define the absolute

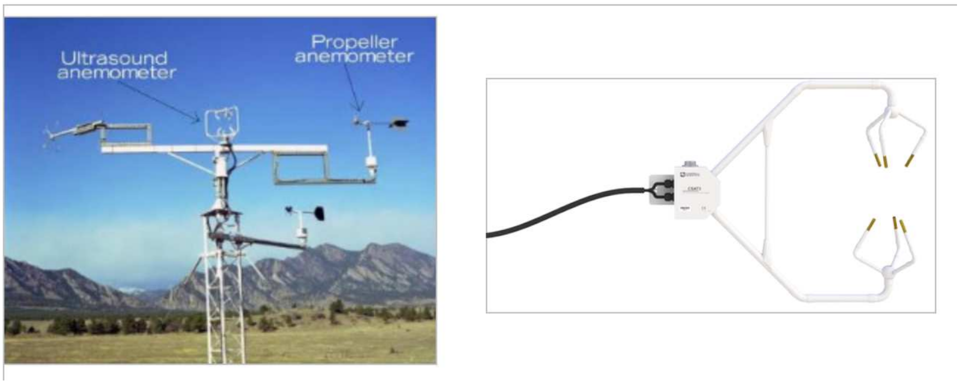


Figure 2.5: Example of a 3D ultra sonic anemometer and a propel anemometer at the NREL test site in Colorado (left) and a 3D sonic anemometer from Campbell scientific (right).

power curve. The iSpin technology strictly speaking are sonic anemometers that are mounted at the tip of the spinner of the rotor, in front of the turbine blades, looking forward in wind direction and rotating with the blades. This means that the velocity of the rotating blades is taken into the computation of the signals and wake effects and yaw misalignment are measured instead of the signal being disturbed by the rotating blades.

There are a number of studies that have been carried out since 2011, when the instruments were first launched by ROMO Wind. Reviews made by DNV GL [22] and ECN [63] provide a comprehensive overview of the technology and it's development from 2011 to today.

DTU Wind (Risø) has published documents describing the technology characteristics and basic principles [23, 24, 58]. The so-called spinner ultra-sonic 3D anemometer technology iSpin has been installed for the first time in 2016 on a fleet of wind turbines in Denmark. The iSpin devices are delivered, implemented and maintained by ROMOWIND [30, 29]. ROMOWIND run an open-access project that was supported by the Danish EUDP program where data was collected from 1st November 2014 to 17th December 2015. A follow up research and demonstration project "Open access to data from the Performance Transparency Project" (PTP) provides access to data from the iSpin technology to Wind turbine performance evaluations in all terrains . The project is a systematic demonstration project for wind measurement systems with around 90 iSpin systems on wind turbines in 9 different wind farms, 3 different turbine types, each of them installed at wind farms in 3 different terrain classes like flat, semi-complex, complex or offshore [4].

2.1.5 Horizontally mounted nacelle LiDAR

Another nacelle mounted wind measurement device is a horizontally mounted LiDAR at the turbine nacelle. Theoretically, every LiDAR can be mounted in that way. However, the space and requirements on top of the wind turbine are very different to the ground. There are several commercially available instruments on the market at present (Figure 2.7). All belong to the remote sensing devices, are compact in form and are able to measure the wind speed

Figure 2.6: Ultra-sonic anemometer spinner technology iSpin from ROMO Wind as schematic and also an example of a nacelle top mounted from METEK (bottom right).

upstream of the turbine in several distances in front of the rotor. The devices are typically mounted at the back of the nacelle similar to the classical cup anemometer. Because nacelle lidars measure the wind speed at several points and in several distances upstream of the rotor, they are able to measure wind characteristics such as shear over the whole rotor plane. They are also able to measure the wind before it actually impacts the turbine operation, thus gathering preview information of the wind characteristics. This makes the instrument an interesting device not only for performance measurements, but also for forecasting.

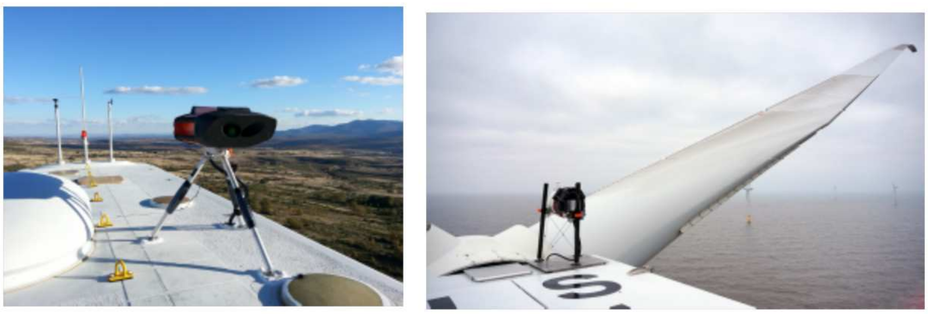


Figure 2.7: Example of nacelle mounted LiDAR WindCube Nacelle (left) and ZX TM (right).

The advantages of nacelle lidars can be summarised as follows:

- compact and small devices

- measurement of turbine inflow at several distances in front of the rotor over the whole rotor plane
- measurement of wind characteristics such as shear and TI
- detection of yaw misalignment
- preview information of the inflow of the turbine

It should be noted however, that the turbulence intensity measured by a lidar is not the same as the one provided by a cup or sonic anemometer. Due to the measurement volume of the lidar, which acts as a low pass filter for the measured wind speed fluctuations, the measured turbulence is smaller. However a recent study shows that "the lidar-based along-wind unfiltered variances with those from a cup anemometer installed on a meteorological mast close to the turbine shows a bias of just 2%" [50].

2.2 Instrumentation for Solar Projects

The “Best Practices Handbook for the Collection and Use of Solar Resource Data for Solar Energy Applications” (referred to also as “PVPS solar resource handbook”) [56] presents detailed descriptions of the relevant meteorological measurements and specific references to even more detailed publications. Here, the focus is on those instrumentations relevant to real-time forecasting purposes as defined in section 1.4.4.

Different applications may require different and specific measurements in order to get relevant irradiation information from a solar site. With the sun’s radiation reaching the earth’s surface

- Direct Normal Irradiation (DNI)
- Diffuse Horizontal Irradiation (DHI)

are linked through the equation for Global Horizontal Irradiation (GHI):

$$GHI = DHI + DNI \times cos(\theta) \tag{2.1}$$

where θ is the solar zenith angle, GHI is the global horizontal irrandiance, DHI is the Diffuse Horizontal Irradiation and DNI is the Direct Normal Irradiation. On a sunny day, it can be assumed that the insolation is 100% GHI with 20% DHI and 80%$DNI \times cos(\theta)$.

There are various components that can and others that are recommended to be measured at a solar plant, where solar energy forecasting is at stake. The most common measurements that are relevant for solar energy forecasting are GHI and DNI, as as DHI can be easily computed from the other two components, but not so easily measured.

1. **System components for Radiation Measurement:**

(i) Data logger
(ii) Steel cabinet with solar power supply and communication system
(iii) Pyranometer
(iv) Pyrheliometer
(v) Optional:
  (a) 2 solar reference cells (horizontal and tilted)
  (b) Pyranometer tilted in same angle as solar module to measure GTI
  (c) Rotating Shadowband Irradiometer to measure GHI, DHI (DNI calculated)
  (d) Delta-T SPN1 Pyranometer to measure GHI, DHI (DNI calculated)

2. **Meteorological representation of a site for forecasting by a "Met Station":**

(i) Ambient temperature
(ii) Wind speed and wind direction
(iii) Air humidity
(iv) Air pressure
(v) Precipitation amount and frequency
(vi) Soiling

The instrumentation and relevant parameters for these components are:

- Met Stations:
  - solar irradiance (various components)
  - Wind Speed anemometers
  - Ambient Temperature and Relative Humidity sensors
  - Atmospheric Pressure sensor
  - Precipitation sensor or gauge
  - Aerosols and Water Vapor
- Radiation Measurements:
  - Spectral Irradiance
  - Ultraviolet Irradiance
  - Soiling
  - All sky imagers
  - Circumsolar Radiation
  - Beam Attenuation Between Heliostats and Receiver in Tower Power Plants

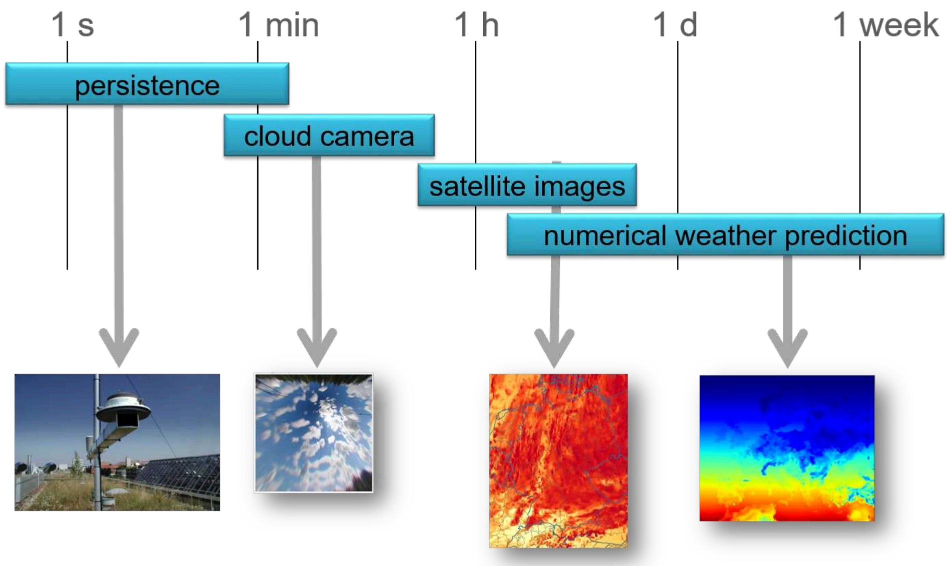


Figure 2.8: Application of different data sources in the frame of short-term PV forecasting. Persistence in this case encompasses time series based statistical methods in general; Cloud camera is a different denomination for all-sky imager.

– Surface Albedo measurement/calculation

For the purpose of short-term solar forecasting, the optimal use of different instruments and data can be viewed as a cascade, as indicated in Figure 2.8. When targeting a particular range of forecast horizons for a given application, this should be taken into account before excessive instrumentation is deployed.

The corresponding instruments are described in more detail in the following sections.

2.2.1 Point Measurements

Most instruments used for solar energy provide point measurements. For example a pyranometer measures the solar irradiance (units W/m^2) incoming from the hemisphere above the sensor plane using a roughly 1cm^2 sensor element. Compared to the size of the solar power plants or pixels of satellite derived irradiance data this area can be described as point like. Pyranometers can be used to measure GHI (global horizontal irradiance) if they are installed horizontally facing the sky or GTI (gloabl tilted irradiance) if installed tilted. GTI is also called POA (plane of array irradiance) if it refers to the plane of the solar collectors. Pyranometers can also be used to measure the DHI (diffuse horizontal irradiance), if they are shaded by a specific shading structure. Such a shading structure can be a shading ball or shading disk tracked according to the solar position by a solar tracker. Also shading rings are used at times, but the accuracy of such systems is lower as they cover not only the sun, but the whole daily solar path as seen from the pyranometer which also excludes part of the diffuse radiation form the measurement. In order to measure the albedo the GHI can be compared to the RHI (reflected horizontal irradiance) which is measured by a horizontally mounted ground facing pyranometer. Reference cells can be used to measure the GTI. Actually the measurement signal of the reference cells is only an approximation of the GTI, but it is helpful as it closer to the power generated by a PV module than the GTI itself. This is because the spectral and incidence angle behaviour of reference cells are quite similar to PV modules of the same PV technology. To measure DNI (direct normal irradiance) a pyrheliometer on a sun tracker can be used. As solar trackers are expensive and as they require a significant maintenance effort, also other instruments exist that allow the determination of the DNI. Such instruments are for example RSIs (Rotating Shadowband Irradiometers) that consist of a horizontally levelled pyranometer and a shadowband, that rotates around the pyranometer at times, e.g. each minute. While the shadowband is beneath the sensor plane the pyranometer measures GHI. During the rotation the shadowband shades the pyranometer at some point and then DHI is measured. The DNI can then be derived using the DHI, GHI and the solar elevation angle. Also, systems of several pyranometers under a specific shading mask can be used to measure DHI and GHI (SPN1). Further approaches to avoid solar trackers e.g. by using sunshine duration sensors also exist.

Soiling measurements can be done by measuring the short circuit current of a pair of reference cells if one of the cells is cleaned and the other not. The approach can also be used with a pair of PV modules. Specific soiling sensors with active light sources or camera systems are also used.

For concentrating systems the measurement of circumsolar radiation can be beneficial. This can be done with rotating shadowband irradiometers, pyrheliometers with different acceptance angles or special camera systems. For solar tower plants further measurements are of interest to provide information on the beam attenuation between the mirrors and the receiver.

The further instrumentation for point measurements with anemometers, temperature sensors, pressure sensors, hygrometer sensors and precipitation sensors has bee described above related to met masts.

2.2.2 All sky imagers

All sky imagers (ASI) are cameras that take photos of the hemisphere above the ground. ASIs are used to provide nowcasts of the solar irradiance for the next 10 to 30min. Most ASIs use fish eye lenses to obtain such photos, but systems using curved mirrors are also used at times. The sky images can be common RGB photos, or infrared images. Clouds can be seen in the photos and they can be detected and even classified by image processing and/or machine learning techniques.

A physical approach may use image series with one photo e.g. every 30 seconds to derive the cloud motion vectors of individual clouds or groups of clouds. Assuming that the clouds will not change their form and velocity their position can be predicted. Some ASI forecasting systems also include estimation of the change of the cloud's form and size. As only a limited area is covered by an ASI the forecast horizon is not long, and the actually reached forecast horizon depends on the cloud speed and cloud height. The effect of the clouds on the irradiance can also be estimated e.g. based on the combination of the ASI photo with co-located radiometers and considered for the nowcast. If more than one ASI is used also the cloud height can be determined using stereophotography. With such a cloud height measurement, alternative measurements or estimations of the cloud heights ASIs also allow to derive spatially resolved maps of the solar irradiance. Such maps can then also be predicted for the forecast horizon.

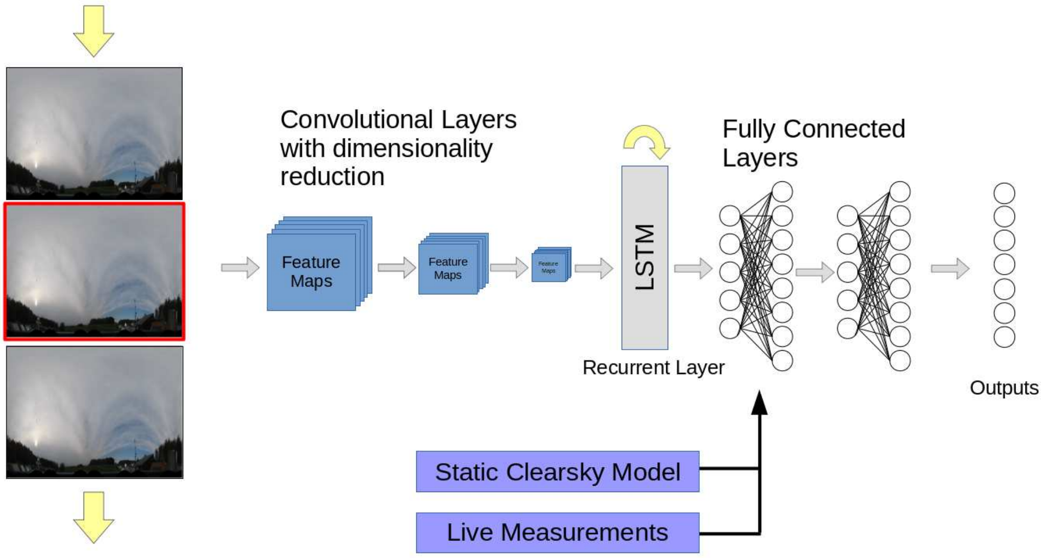


Figure 2.9: Example of a machine learning model for nowcasting solar irradiance based on all sky images [53]. A series of cylindrically transformed ASI images is first compressed via a series of convolutional neural network layers. The resulting features are processed by a recurrent LSTM [27] layer, which remembers the evolution of the cloud field. Some fully connected layers can be used to incorporate other, low dimensional input data and assemble an irradiance or power forecast over the next few minutes.

With only a single ASI available at the site in question, an alternative approach is to use machine learning (ML) for discerning patterns in the cloud field. As a data driven approach, this typically requires several weeks or months of ASI images plus corresponding irradiance and/or PV power measurements. Using image processing (convolutional) neural networks and related techniques – which in recent times received a lot of research attention, e.g. in the contexts of face recognition and autonomous driving – the motion and transformation of the cloud field and its impact on the target quantity can be learned, to a certain degree. Figure 2.9 shows an example of such an ML model. Note however that there are many different ways of setting up such models, and a single best practice can not yet be identified. Still, in view of increasing data volume and computation power, as well as the lower on-site hardware requirements, this approach may be worthwhile to investigate for some projects.

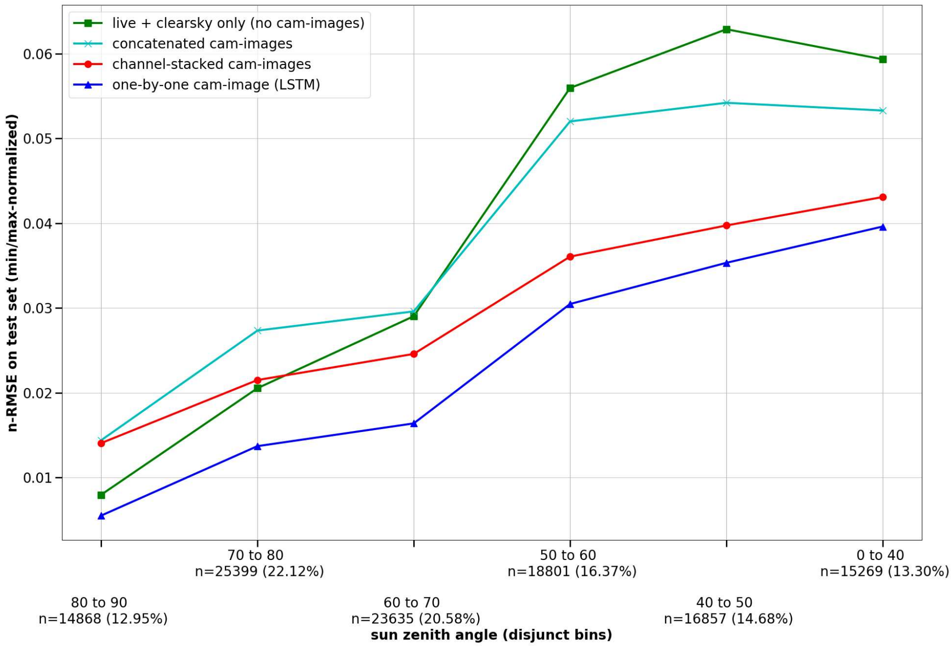


Figure 2.10: Evaluation of the ASI-based 5 minute forecast model from Figure 2.9 (labeled *one-by-one cam-image*) and three other models, which vary image treatment and the internal neural network structure. At high sun elevation, the errors due to moving clouds are most pronounced, as is the improvement gained by the ASI forecast.

In Figure 2.10, part of the evaluation of the described example nowcast model is shown.

To summarize, ASIs can help to operate solar power plants and electricity grids better, as the spatial and temporal variability of the solar irradiance can be captured and predicted. This is for example interesting to avoid high PV power ramps that might cause technical issues in the grid or that might be regulated by law and connected to fines. More information on ASIs and their application can be found in the “PVPS solar resource handbook” [56] including several references for further reading.

2.2.3 Satellite Data

Due to the global coverage of the earth with geostationary weather satellites (except at very high latitudes), it is relatively straightforward to obtain satellite-based solar irradiance data for any given PV project. Both GHI and DNI, plus sometimes other components of the radiation field are available. Surface irradiances are typically produced from multispectral satellite images using one of the many variants of the HELIOSAT method.[e.g. 52] The general idea here is that clouds as seen from the satellite are always brighter than the surface. Thus the amount of cloudiness at a fixed location (i.e. pixel) can be estimated from the difference between the albedo observed by the satellite, and the lowest albedo observed for this location during the course of (typically) a month. Surface irradiance is calculated from a clear-sky model, and then reduced by factoring in the satellite derived cloudiness. Of course, the more recent HELIOSAT methods also take into account changes in atmospheric composition, surface snow and ice, slanted viewing geometries and other factors.

While using satellite data for PV power resource assessment is standard practice, its application to short term forecasting is not quite as common. Except for some rapid scan areas, most of the globe features image update rates of 15 min, but some areas are only provided new images once per hour. Similar to processing ASIs, cloud motion vectors can be calculated from sequential images, and used to provide short term forecasts of cloudiness and hence, surface irradiance. Again, part or all of these calculations can be supplanted with machine learning techniques if sufficient training data are available.

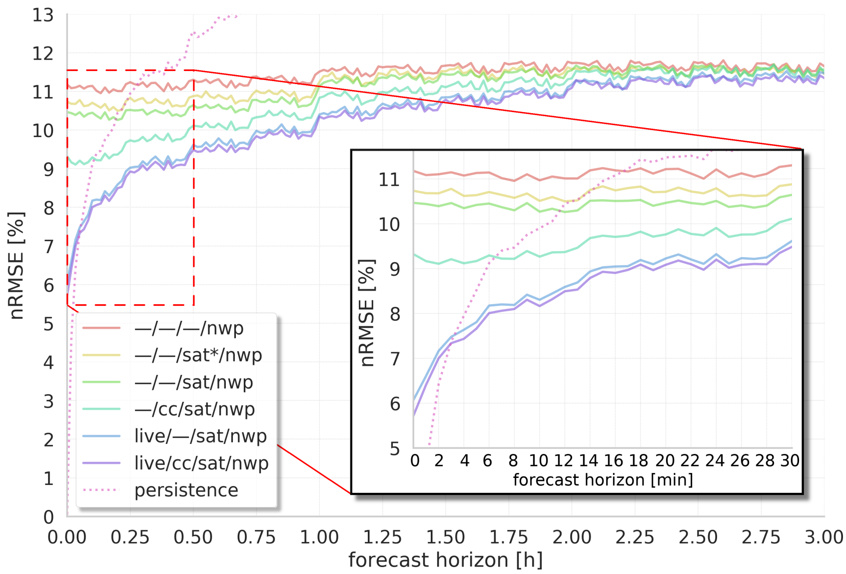


Figure 2.11: Forecast error for different input data combinations in an ML-based irradiance forecast model. *sat* and *sat**: different satellite data setups, *cc*: cloud camera, *live*: live data from the pyranometer.

However, in contrast to ASIs, data from meteorological satellites is constantly being assimilated into weather forecast models. Hence the advantage of using real-time satellite data diminishes within the first few forecasted hours, if NWP forecasts are also utilized. To illustrate this point, Figure 2.11 shows the contributions of different types of sensor and model data in a short-range forecast experiment. Note that this model was not tuned towards very short term accuracy and uses a simplified ASI processing method. Still, the contribution of the inputs towards error reduction is clearly seen, as well as the typical relaxation towards the pure NWP-based forecast at longer forecast lead times.

Chapter 3

Power Measurements for real-time operation

Key Points

- *Live power production data are a valuable input to very short-term forecasting systems, especially those with frequent updates and high temporal granularity*
- *Additional operational data such as plant set-points, operational capacity, and Power Available signals are critical for robust forecasting and forecast evaluation*
- *Real-time data should be collected at the same or higher temporal resolution and frequency than that of the desired forecasts*

Real-time vs non-real-time...

In real-time forecasting systems rely on input data that are highly correlated with future values of the the quantify to be predicted. In the case of wind and solar power, recent measurements of power production are highly correlated with future measurements on very short time scale, those from seconds to minutes and hours ahead. Furthermore, as these are measurements of the target variable itself, they are not subject to errors from data processing, such as mathematical models of the weather-to-power conversion process, which may be applied to recent weather observations for example. Therefore, real-time measurement of power production can be extremely valuable inputs for forecasting systems. Furthermore, as wind and solar plants are subject to partial or complete outages due to maintenance or instructions from network operators, data describing the operational capacity of a plant are also critical for successful forecasting.

3.1 Live power and related measurements

Real-time measurements of power production from wind and solar farms are a valuable input to very short-term forecasting systems, those with lead-times of minutes to hours ahead. This and related data can also assist monitoring of plant availability, which is important for operational forecasting. Real-time power data has the potential to significantly increase the accuracy of very short-term forecasts compared to those based on Numerical Weather Prediction and/or those which incorporate real-time meteorological data. Power measurements have the advantage of being the quantity that is being forecast and therefore not subject to errors when converting wind speed to power, for example. However, power production is affected by non-weather effects, such as control actions from system operators, which should be accounted for by forecasters.

Power production may be measured at multiple points between individual generators (individual wind turbines or solar panels) and the point of connection to the electricity network. In the majority of cases, the energy metered at the connection point and used in settlement of the electricity market is the quantity to be forecast. Accuracy standards for settlement meters are generally high, and live data are typically available to at least the plant and relevant electricity system operator. Power from individual generators or other points in a wind or solar park's internal network will be subject to losses so will not match the settlement meter. This data may however be utilised to improve forecasts.

Temporal resolution is an important factor when considering the use of live data in forecasting systems. Different users may be concerned with power/energy production on different time scales. Market participants' primary concern is energy in a given settlement period, which differ in length between countries/regions, and between financial products for energy and other services, such as reserve products. Settlement periods with 5, 15, 30, 60, 240 minute duration are common. System operators, on the other hand, may be more concerned with shorter periods (one minute or less), or even instantaneous power, for balancing purposes. In practice, instantaneous is difficult to define and measure, and may therefore be be approximated by averaging over a short time period or filtering/smoothing. Real-time data should be collected at the same or higher resolution than the user requires for their forecasts. If the user requires forecast for each 15 minute period, data should be gathered at 15 minute resolution or higher. Higher resolution data also enables more frequent updates to forecasts as new data becomes available more frequently.

A list of common measurements and their potential uses in forecasting systems is provided in Table 3.1, and are discussed in more detail in the following section. In general, they are collected by the plant's Supervisory Control and Data Acquisition (SCADA) system, or by dedicated metering/monitoring devices that communicate with the electricity system and/or market operator.

Table 3.1: List of power-related quantities measured at wind (W) and solar (S) plants that have a role in forecasting and examples of how they may be used.

Quantity	Wind (W)/ Solar (S)	Location	Use
Active Power	W&S	SCADA or Connection Point	Input/feature in very short-term forecasting models, indicator of plant availability if combined with meteorological measurements
Plant Effective Capacity (live and future)	W&S	SCADA, operator systems (sometimes submitted to TSO/market operator)	Forecast input and quality control
Controller Setpoint	W&S	SCADA	Flagging curtailment and other control actions
Capacity in operation	W&S	SCADA	Scaling forecast based on available capacity
Voltage	W&S	SCADA or Connection Point	Flagging plant unavailability
Status of breakers	W&S		Flagging plant unavailability
Temperature	W&S	Plant	Forecast input: PV efficiency, snow, icing
Wind Speed	W	Turbine SCADA or Met Mast	Forecast input and quality control
Wind Direction	W	Turbine SCADA or Met Mast	Forecast input and quality control
Wind Turbine Yaw Angle	W	Turbine SCADA	To check alignment of wind vane(s) and other data quality control
High Speed Shutdown Flag	W	Turbine/Plant SCADA	Forecast model training and evaluation
Icing Shutdown Flag	W	Turbine/Plant SCADA	Forecast model training and evaluation
Horizontal direct solar radiation	S	Plant	Forecast input and quality control
Panel Tilt	S	Plant	Forecast input and quality control
Sky images	S	Plant	Forecast input

3.2 Measurement systems

3.2.1 Connection-point Meters

The interface between a power plant and the electricity system, the so-called connection-point, is typically metered to a high degree of precision as the data is used in electricity market settlement and as a live feed to plant and system operators. The minimum precision of such measurements is usually dictated by network codes and/or market rules in a given region. For example, international standard IEC 62053-21 and IEC 62053-23 describe classes of active and reactive power meters, respectively. These standards are references in the Great Britain Balancing and Settlement Code, where active power meter errors are required to be less than 1.5% under normal conditions, or 2.5% if the power factor is beyond 0.5 lag or 0.8 lead, or active power is below 20% or rated import/export. It should be noted that these measurements may not match those recorded by plant plant SCADA systems (the topic of the next section) due to differences in equipment or location of measurement devices relative to other balance of plant equipment.

Live power data is often streamed to plant and system operators and made available to control room engineers and their support systems. It is often visualised along side forecasts to track out-turn relative to forecasts continuously. Live data is also a key input to very short-term forecasting systems with high temporal granularity, which is discussed in Section 3.4.2.

Energy is typically metered according to the duration of market settlement periods, which typically range from 5 to 60 minutes. As the volume of energy used in market settlement, predicting the volume of energy measured by these meters is the objective of the majority of forecasting systems, especially those used by market participants.

3.2.2 Wind Power SCADA Systems

Wind turbines routinely measure operational variables for their control and operation. Several of these are valuable in forecasting, particularly on the very short-term horizons if provided in (close to) real-time to forecasting systems. These are listed in Table **??** and include variables for measuring both power production and plant availability. It is important to note the particular characteristics of these, in particular of the nacelle mounted anemometer and wind vane, which are impacted by the up-wind rotor, see Section 2.1.2 for more details. The sampling frequency if also significant. Typically 10-minute mean values are stored and communicated to control control centres, but higher resolution may be of value, particularly if maximising forecast accuracy on lead-times of 10-minutes or less is a high priority. Live feed of instantaneous active power can be valuable in very short-term forecasting, though it is important to treat both instantaneous power and power averaged over different time periods as distinct.

Plant availability is an important factor in forecast production and evaluation. This includes unavailability due to fault or maintenance outage, either of wind turbines or balance of plant equipment, as well as control actions that reduce availability, notably high speed shut down and icing. Live availability alone is useful in forecasting but future changes in availability are also important. Signals indicating future availability are therefore valuable and should be made available to forecast systems.

Finally, data quality is very important. These variables can be a source of error in forecast production if they are not of high enough quality or availability.

(i) **Live active power measurements** are a key input to very short-term forecasting systems and status variables provide important information regarding plant availability.

(ii) **Active power set-point** is the maximum power the wind farm controller will allow to be produced, e.g. if instructed to limit output by the system operator.

(iii) **Current outages and 'effective capacity'** are not necessarily a predictor of continued outage/effective capacity, but it is an important input on very short-time scales and when diagnosing errors in longer-term forecasts. These are:

- Turbine reduced capacity
- Turbine outage
- Strings of turbines out of service
- Voltage and breaker status at connection point

(iv) **Nacelle wind speed and direction measurement** can indicate the operating regime of the turbine, and be used to verify whether it is operating a full or reduced capacity.

(v) **Power available signals** accurately estimate the amount of power that could be produced in the absence of control actions, such as reduced set-points.

(vi) **Turbine yaw angle** indicates the direction that the turbine is facing, and can be used to verify normal operation, e.g. by comparing to the necelle measured wind direction.

3.2.3 Solar Power SCADA Systems

Solar power plants routinely measure operational variables for their control and operation. Several of these are valuable in forecasting, particularly on the very short-term horizons if provided in (close to) real-time to forecasting systems. These are listed in Table **??** and include variables for measuring both power production and plant availability. The sampling frequency is also significant. Typically 10-minute mean values are stored and communicated to control control centres, but higher resolution may be of value, particularly if maximising forecast accuracy on lead-times of 10-minutes or less is a high priority. Live feed of instantaneous active power can be valuable in very short-term forecasting, though it is important to treat both instantaneous power and power averaged over different time periods as distinct.

(i) **Live active power measurements** are a key input to very short-term forecasting systems and status variables provide important information regarding plant availability.

(ii) **Current outages** are not necessarily a predictor of continued outage, but it is an important input on very short-time scales and when diagnosing errors in longer-term forecasts. These may be recorded using SCADA flags as outages of:

- Inverters
- String(s) of panels
- Specific outages at lower levels
- Voltage and breaker status at connection point

(iii) **Irradiance measurements** allow the efficiency of the plant to be determined by comparing to output and can be used in forecast model tuning.

(iv) **Back panel temperature** effect panel efficiency and can be used in forecasting systems directly, and in combination with other variables to determine plant efficiency.

(v) **Panel tilt** if measured, can also be useful since trackers may be miscalibrated or faulty, and this data can be used to aid in cleaning of training data and identifying issues with live forecasts

Meteorology station temperature and wind speed, which has bee described in detail in section 2, **??** can be combined with irradiance measurements to provide

Finally, data quality is very important. These variables can be a source of error in forecast production, if they are not of high enough quality or availability.

3.3 Power available signals

Power Available (PA) signals provide an estimate of the potential power that could be produced by a wind or solar plant if not constrained in any way. Constraints are typically enacted by controller set-points issued by the plant or electricity system operator. PA is equal to active power under normal operation, or the power that could be produced if any constraints were lifted. Modern wind and solar plants typically produce PA signals internally for control purposes and can be made available as SCADA data feeds. In some jurisdictions these are shared with electricity system operators are are subject to accuracy requirements.

Wind and solar power forecasts are typically configured to forecast the power that would be produced in the absence of constraints, which therefore corresponds to PA signals when such actions are in place. Furthermore, forecasts that use live power as an input should switch to taking PA as an input during constraints, or at least flag power data as corrupt during these periods.

3.3.1 Embedded Wind and Solar "behind the meter"

Embedded wind and solar, which is connected 'behind the meter' is in many electricity systems handled as negative demand. This poses a challenge as both the installed capacity, location, and power output of these generators may be unknown.

Various actors require forecasts of net-load with gross load and embedded generation disaggregated. In this case, a combination of net-load metering, meteorological observations, and wind and solar production data all add value to forecasting systems.

Key variables that should be made available to forecasting systems are:

(i) **Installed capacity** of embedded wind and solar generation. National and commercial databases are available but levels of accuracy and detail are variable.

(ii) **Metered production** of embedded wind and solar generation, most likely a sub-set of total installations deemed to be representative of the total in a given region.

(iii) **Geographically distributed meteorological observations** measuring wind speed and direction, solar irradiance, and temperature, as a minimum.

3.4 Live power data in forecasting

Live measurements of the quantity of being forecast are the most important inputs to any forecasting system on very short lead-times. Their recency and locality cannot be matched by modelling due to latency and epistemic uncertainty introduced by modelling, except in the case where live measurements are corrupted. In general, methods based on live data outperform NWP-based forecasts for lead-times shorter than 2–10 hours, depending on the specific target variable and reference NWP. State-of-the-art forecasting systems combine both approaches for optimal performance at all lead-times, and may also include live data from multiple neighbouring measurement locations for added benefit.

In addition to live measurements of the forecast variable, operational data such as plant availability (e.g. proportion of turbines/panels in service) and control actions (e.g. curtailments) are also required as they change the nature of the power measurement. State-of-the-art forecasting systems will adapt to changing operational regimes and must be calibrated to predict the variable of interest to the user: what the actual power production is expected to be in the future, or, the the power production would be expected to be in the future if no control actions were in effect. Forecasting systems must also be robust to missing data and able to either adapt to the loss of a live feed, or impute missing values.

3.4.1 Specifics for producers of forecasts

It is best practice for those who produce forecasts for internal use or to supply to incorporate live data into their forecasting system for both continuous forecast verification (see [2] part 3) and production of very short-term forecasts. As discussed above, live data is a valuable input to very short-term forecasting systems, and depending on the nature of forecast produce/service, may be necessary for post-processing to account for plant availability. Live data may also be included in forecast visualisation so that users may see recent history in the same figure as a forecast.

3.4.2 Specifics for consumers/users of forecasts

Consumers/users of forecasts should be aware of whether their forecast provider is using live data when producing forecasts, especially for very short-term lead-times. Some of the benefit of including live data may be realised by rudimentary methods, such as blending live data with forecasts or simply visualising recent observations alongside forecasts. However, to maximise forecast performance, it is recommended to employ a forecasting system/provider that leverages live data.

3.5 Summary of best practices

In summary, the following are recommended practices for use of power measurements in real-time operation and forecasting of wind and solar power:

1. Live/real-time power measurements should be used in very short-term wind and solar power forecasting, i.e. for lead-times from minutes to hours ahead.

2. Power data should be recorded at the same or higher temporal resolution and frequency than that for the forecasting system.
   - For electricity system operation temporal resolution may be one minute or less, for trading operations this should match the duration of the shortest traded product or settlement period.
   - Forecast update frequency should match the highest frequency with which new information becomes available, within the constraints of IT infrastructure.

3. The power measurement should be taken from the metering point that is most relevant to the forecast user, this will usually be the metering system at the connection point to the electricity network.

4. Live data on plant operating conditions (e.g. control actions the effect maximum output) and availability (e.g. maintenance outages) should be used in forecast production and evaluation. Where direct data is not available, this should be inferred using indirect measurements.

5. Data on future plant availability (e.g. maintenance actions) should be used in forecast production. Typically, power forecasts do not include the effect of control actions issued by electricity system operators.

6. Live Power Available signals should be used in forecast production in place of live power production data during periods when output is curtailed due to control actions

Chapter 4

Measurement Setup and Calibration

Key Points

- *Instrumentation Selection*
  1. *Evaluation of the required data accuracy or uncertainty levels*
  2. *Cost-benefit analysis of instrumentation regarding quality and maintenance requirements*
  3. *Specification on redundancy levels*
- *Representativeness of Measurements*
- *Verification of correctness of installation and calibration*
- *Setup and Calibration Logging*
- *Maintenance Schedules of instrumentation*

4.1 Selection of instrumentation

In this section we will describe best practice for the selection of instrumentation for forecasting applications of wind and solar projects. We focus in this description not on instruments, but rather on the requirements and necessary considerations for the selection of instrument classes that are appropriate for the pre-defined requirements.

The first step in the selection procedure is the definition of requirements for the forecasting application. The following priority regarding the selection process of instrumentation is recommended for both wind and solar forecasting applications [56]:

1. **Accuracy requirements**:

Accuracy requirements need to be defined for the application/project and aligned with the associated levels of effort necessary to operate and maintain the measurement system on under these constraints. An overall cost-performance determination should be carried out to adapt the budget to the accuracy requirements and vice versa.

2. **Reliability requirements**:
Reliability can be achieved with redundant instrumentation and/or high quality instrumentation. Redundancy enhances and ensures confidence in data quality. Selection of multiple instruments need to be aligned with the accuracy needs.

Table 6.1 additionally provides typical forecasting applications and respective appropriate requirement classes or standards.

Table 4.1: Forecast applications and respective recommended requirements for appropriate instrumentation.

Forecast Application	**Requirements for Wind** according to IEC 61400 Part 12-1 [15]	**Requirements for Solar** according to ISO9060: 2018, EC 61724-1:2021 and WMO Guide No.8
System Operation Forecasting	Class A	Class A or B
Utility Scale Forecasting	Class A	Class A or B
Wind/Solar Park control	Class A	Class A
Park / Turbine Monitoring	Class S	Class B or C
Electricity Market Trading	Class B	Class B

The above classification in Table 6.1 is only a categorisation of a best practice, and is at best a rouch guideline as requirement in forecasting applications to ensure high quality measurements to establish the potential for improved forecasting. For those applications, where there are provided 2 classes, the second, lower class should be understood as minimum requirement and the higher class as a requirement, if high quality is required and the respective budges justifiable.

The classes for wind projects are in detail defined in the IEC 61400 Part 12-1 [15] and explained in Annex D of the standard. The classes for solar or PV systems are defined in the ISO 9060:2018 [32], the WMO Guideline No.8 [66] and the IEC 61724-1:2021 [16], where differences, consistency and overlapping areas are explained in detail by the IEA Task 16 PVPS in their "Best Practices Handbook for the Collection and Use of Solar Resource Data for Solar Energy Applications" (PVPS-Handbook) [56] and a few more detailed differences for specific forecasting project types in section 4.1.2.

4.1.1 Selection of instrumentation for wind projects

The selection of instrumentation of meteorological measurements for forecasting applications of operational wind farms should be selected to ensure (1) an independent source of measurement to the power generation and (2) a good fit to the power generation of the wind farm or solar plant.

In other words, there is a need for appropriate measurement equipment in the sense of:

1. Accuracy
   - appropriate quality of instrumentation
   - uncertainty evaluation of instrumentation
2. Reliability
   - correct installation of instrumentation
   - correct calibration of instrumentation
   - if needed, redundant setup of instrumentation and logging
3. Availability
   - if needed, redundant instruments
   - loggers and instruments being fail safe
4. Weather resistance and safety
   - meet local requirements of climate
   - meet local requirements of landscape
   - meet safety requirements (e.g. flight safe in case of met masts at hub height)

This list provides a general guidance of the five most important aspects that need consideration in the selection process of instrumentation for real-time forecasting applications.

It is important for any authority (e.g. TSO, DSO, utility or balance responsible party), assessment party or advisor to go through these aspects, when setting requirements for the instrumentation. Dependent on these aspects, there are standards available that guide the power plant operators to the instrumentation that is appropriate for the requirements set out for the purpose and the siting of the instrumentation.

In this recommended practice guideline, we therefore focus on the procedure and instrument classes rather than the individual instrumentation, as these are handled by available standards (see section 1.6 for a description of the applicable standards).

For wind measurements, there is a principle of three tiers of instrument standard class available that are dependent on the requirements to accuracy and reliability and are also associated with different cost levels. It is therefore recommended defining the requirements first, before requiring a specific instrumentation quality standard.

The "GUIDE TO INSTRUMENTS AND METHODS OF OBSERVATION - VOLUME I"[66] defines important requirements for meteorological instruments to:

1. Simplicity of design which is consistent with requirements
2. Durability

3. Convenience of operation, calibration and maintenance
4. Acceptable cost of instrument, consumables and spare parts
5. Safe for staff and the environment

It is recommended that authorities refer to the classification of instruments in their requirements or provide accuracy limits.

The IEC 61400 standard [14], for example, provides information on anemometers according to class A, class B or class S types, where the classes of instrumentation are dependent on the terrain structure defined in Annex B of the standard. In areas with complex terrain, the class B standard requires e.g. specific setup of instrumentation to accommodate the influence of turbulence from varying terrain etc.

Table 4.2: Instrument classes for various environments.

Environment type	Environment description	Instrument class
Standard terrain	Flat terrain with few obstacles	class A
Complex terrain	varying heights and many obstacles	class B
Offshore	wave and current driven turbulence	
salty air	classe S	

The class S is associated with specific requirements, "where the influence parameter ranges are restricted to allow for the specified accuracy of the anemometer." [14]. Manufacturers of wind instrumentation that sell instruments for wind projects provide information about the class of instrumentation to make it easy for wind farm owners to purchase the appropriate instruments.

Table 6.2 describes the main and applicable types of instrument classes in their respective environments.

4.1.1.1 Components of a Wind measurement system

4.1.2 Selection of instrumentation for solar power plants

In the IEA PVPS handbook for the collection and use of solar resource data [56] the authors present recommendations for the selection of the instruments. These recommendations also apply for the real-time applications. Depending on the solar technology used different radiation components must be measured. The requirements for fixed monofacial PV or thermal collectors, bifacial PV, concentrating systems and tracked non-concentrating PV are different. Besides radiometers also further meteorological parameters must be measured as stated above. For selecting the instrumentation, the user must also evaluate the required data accuracy and consider the effort necessary to operate and maintain the measurement system. Specifically, the most accurate instrumentation should not be purchased, if the project resources cannot support the maintenance required to ensure measurement quality consistent with the radiometer design specifications and manufacturer recommendations. In other words, the accuracy of instruments needs to be considered under the aspect of it's purpose of measuring and the expectation on the accuracy of the measured quantity. For example, if the accuracy of an instrument is ∓ 0.3 and the target is an accuracy of 0.1, neither the instrument nor the measured quantity can provide a representative picture.

Various radiometers of the same type should be considered for big power plants and are also helpful to ensure a higher data quality. The number of required instruments for PV parks of different peak power is given in IEC 61724-1 along to further requirements on instrument accuracy, maintenance and data logging. Multiple radiometers within a project location and/or providing for the measurement of all three solar irradiance components (GHI, DHI, and DNI) can greatly enhance opportunities for post-measurement data quality assessment.

To summarise, the following considerations should be taken in the selection process:

(a) Evaluation of the required parameters, data accuracy or uncertainty levels

(b) Cost-benefit analysis of instrumentation regarding quality and maintenance requirements

(c) Specification of number of instruments

For the selection of pyranometers, the ISO 9060 standard "Solar energy - Specification and classification of instruments for measuring hemispherical solar and direct solar radiation"[32] separates since 2018 between 3 main classes of pyranometers A, B and C and four classes of pyrheliometers (AA, A, B, C).

Furthermore, the characteristics “spectrally flat” and “fast response” were defined and made up the largest change from the 1990 version of the standard to the newer 2018 version, aligning it to the WMO Guide No.8 on *Instruments and Methods of Observation* [66] and from IEC 61724-1[16]; here, the performance verification for temperature dependence and directional response must now be included with every individual instrument of the highest accuracy class A. In general they were added to the class to further specify the instrument’s characteristics.

It should also be noted that the IEC 61724-1 PV system performance monitoring standard requires compliance with the latest ISO 9060:2018 version[32], so that requirements for the highest accuracy, Class A PV monitoring systems, also changed.

The purpose of the WMO Guide No. 8 and IEC 61724-1 standard, specifying Spectrally Flat instruments is to ensure “.. continuity of performance”[66, 16]. In that sense, the newer ISO 9060:2018 specifies since 2018 the spectral error by using a clear sky spectrum and horizontally installed instruments, where the Spectrally Flat category is defined for instruments installed non-horizontal, potentially measuring a different spectrum, using the more stricter WMO Guide, IEC and ISO definition of the so-called spectral selectivity.

These standards state the use of Spectrally Flat instruments for:

(a) reflected solar radiation

(b) albedo

(c) plane of Array (including reflected) irradiance

(d) diffuse irradiance (diffusometers)

Details and more practical explanations on the interception and interplay for the selection of instrumentation according to the three standards WMO Guide No. 8 [66], IEC 61724-1:2021[16] and ISO 9060:2018 [32] can be found in the NREL Handbook for PV [56] or with instrument manufacturer and service providers (e.g./ [7, 25].

The following Table **??** only provides a very brief summary of the applicability of those classes that are considered relevant for solar energy forecasting.

Table 4.3: Pyranometer classes for various requirements.

Requirements	**Instrument class**	**Instrument description**
Meteorological or Energy forecasting/ modelling for utility scale PV systems	Class A Spectrally Flat (ISO9060: 2018) Class A (IEC 61724-1:2021 Class 1 (WMO Guide No.8)	Scientific quality and highest accuracy
Measurements for high level monitoring for large commercial scale PV systems	Class B (ISO9060:2018) Class B (IEC 61724-1:2021) Class 2-3 (WMO Guide No.8)	Good quality
Economic solutions for routine measurements and small PV systems	Class C (ISO9060: 2018) Class C (IEC 61724-1:2021 Class 4-5 (WMO Guide No.8)	Medium quality

For anyone changing requirements from the 1990-standard with the classes Secondary Standard (now Class A), First Class (class B) and Second Class (class C), the most sensible way is to let stakeholders contact the manufacturers or service providers to find out about the costs for upgrading instruments to the new standard. Especially for pyranometers, the newer 2018 ISO 9060 standard defines for each class the required instrument heating, azimuth and tilt angle accuracy. It also defines cleaning and calibration intervals for pyranometers and requirements for measurement of module- and air temperature, wind speed and direction, soiling ratio, and (AC and DC) current and voltage [7].

For the selection of requirements for real-time forecasting purposes, it is crucial to identify the needs regarding the quality of the instrumentation, as costs for the instrumentation, but also for operation and maintenance are significantly different.

For example, compliance with IEC61724-1 Class A pyranometers with directional response and temperature response test results is equivalent to ISO 9060 Spectrally Flat Class A instruments. In some cases, older instruments from the ISO 9060:1990 standard's [31] highest quality "Secondary standard" (now Class A), may be updated, if a pyranometer lacks the required test reports for the performance verification on temperature dependence and directional response (e.g. [7]).

4.1.3 Measurement Characteristics of Different Technologies

4.1.3.1 Measurement Characteristics of Lidars

Measuring Laminar Low Level Jets

An advantage of the remote sensing device's ability to measure the wind profile is it's ability to "measure" low level jets [35] that can have significant influence for the power production forecasting. The low level jets are a meteorological phenomenon that is a well-known and extensively researched topic in meteorology since the 1950s (e.g. [9, 70]) and has been a topic in wind energy research since forecasting started.

Recent progress with the help of wind profilers (SODAR, LiDAR) as well as comparisons with a 120m mast has been made in the Lamar Low Level Jet Program, reported by Kelly et al. [35] over a 1-year period.

4.1.3.2 Lightning effects on instrumentation

It is a know phenomena (e.g. [35]) that lightning can cause high frequency noise contamination to instruments. This is applicable for:

- Remote Sensing instruments (LIDAR, SODAR, RADAR)
- sonic anemometers

In the case of LIDAR or SODAR lightning contaminates the signal processing with echoes. Such echo reflections can make it impossible for the signal software to process the signal correctly and hence the data cannot be used. In resource assessment of research projects, such noise can be corrected and the data re-processed to provide valuable data. In Real-time this is not possible.

In the case of sonic anemometers, high-frequency noise contamination disturbes the sonic signals of wind velocity and temperature and usually makes the data unusable.

Maintenance and mitigation methods These examples shows that the maintenance and upgrades of software to make use of fixes in the signal processing algorithms of the devices are a key technical requirement for real-time use of the devices, or alternatively that the raw data needs to be sent and the processing takes place where the data is used.

Therefore, we can conclude that the reliability of any measurement device in real-time operation requires a maintenance schedule to be a technical requirement in order for it not to deteriorate. If this is done, the remote sensing instruments as well as sonic anemometers have proven to provide reliable time series of wind speeds and gusts in general conditions.

4.2 Location of Measurements

The WMO Report Nr. 55 [1] describes the site and instrumentation selection as a 6-tier problem:

1. selection of a suitable site

2. correct exposure of instruments
3. determination of the area of representativeness of measurements
4. description of the site and its facilities
5. homogeneity of the meteorological/climatological data*
6. access to information*

The last two tiers require international collaboration to ensure measured data can be compared and the precision and accuracy of the measured data follows a common understanding for the use in models and methods to improve or interpret model results. We will not deal with these aspects in this document, but want to highlight these topics to anyone dealing with measurements and that the meteorological community has solved these critical problems by establishing international data networks to provide modellers, researchers, instrument developers and end-users the possibility to progress and exchange information.

The fifth tier deals deals with the lack of international homogeneity on measuring standards that has been a problem in meteorology and can also be observed in the definition of measurement requirements for real-time purposes if wind and solar projects on an international scale (e.g. [64]). If data measuring standards are not international, it is difficult to obtain homogeneity in data and thus models and methods, where such data are used.
The WMO Guide for Meteorological Measurements [66] ("WMO Guide 8") contains in it's *ANNEX 1.D. SITING CLASSIFICATIONS FOR SURFACE OBSERVING STATIONS ON LAND* information about siting of instrumentation to ensure representativeness. The guide covers all meteorological instrumentation that are relevant for wind and solar projects and operation. The WMO Guide 8 states that "..the environmental conditions of a site may influence measurement results. These conditions must be carefully analysed, in addition to assessing characteristics of the instrument itself, so as to avoid distorting the measurement results and affecting their representativeness.." and recognises that "...there are sites that do not respect the recommended exposure rules".
Out of these conditions, the WMO has developed a classification with 5 classes to "help determine the given sites representativeness on a small scale (impact of the surrounding environment)" and help to better take their exposure rules into consideration. The classes are however not clearly defined, but are deviations from a standard class with minimun impact of the surrounding environment. For wind projects, the IEC 61400-12-1 standard classifies 2 classes, (1) simple terrain and (2) complex terrain (see next section 4.2.1), and recommends to adjust any instrumentation requirements for sites that fall outside these well-defined classes.

In the following two sections, we will define and discuss specific requirements for wind and solar projects, respectively. The most relevant and useful rules to consider and to apply for the definition of requirements for both wind and solar projects can be summarised with the following 5 rules:

1. The classification must takes place per instrument, not per site as a whole
2. The rating of each site should be reviewed periodically as environmental circumstances can change over a period of time
3. An update of the site classes should be done at least every five years

4. An uncertainty due to siting has to be added in the uncertainty budget of the measurement
5. The primary objective of a classification should be to document the presence of obstacles close to the measurement site

4.2.1 Location of representative Measurements specific for Wind Projects

The representativeness of instrumentation for a specific purpose are defined and explained in the MEASNET guideline [41] "Evaluation of site-specific wind conditions " section 6.4 and are valid for "conditions which will influence the installation, operation and maintenance (O&M), loading, durability, performance and energy yield of wind turbines installed at a site". This guideline thereby provides "..a traceable basis concerning the conformity of the design parameters with site-specific conditions according to IEC 61400-1 standard [15, 14]" and can hence also be applied for operational real-time measurement campaigns for wind power projects.
For operational purposes, the minimum requirements for representative measurements of a wind project are defined in the guide through 2 factors:

1. height of the primary wind speed measurement level
2. representativeness radius of a measurement

The MEASNET guideline [41] recommends that these two most important factors for the representativeness of measurements are defined by a "site-specific expert analysis", but also provides minimum requirements for the (1) height of measurement levels in order to reduce the "uncertainty of the vertical extrapolation of the wind conditions"; and (2) the radius around a mast or other instrument at which "the wind conditions measured at the position can be extrapolated with tolerable uncertainty" and where "all wind turbines shall be located in the representativeness radius of at least one measurement". The guideline differentiates here between two terrain classes [41]. The United States Environmental Protection Agency (EPA) provides a Meteorological Monitoring Guidance for Regulatory modelling Applications [42] add to these two types the coastal or offshore location:

- **simple terrain class**
  flat terrain with no noticeable terrain elevation variations, where the wind conditions are mainly influenced by homogeneous roughness conditions.
- **complex terrain class**
  complex terrain with considerable orographic variation (relief), significant slopes and non-homogeneous roughness conditions.
- **Coastal and Offshore Locations**
  the unique meteorological conditions associated with offshore and local scale land-sea breeze circulations necessitate special considerations; the latter especially, if the land side is characterised by complex terrain

This ensures the representativeness of measurements to capture the wind conditions reaching the entire wind farm.

Table 4.4 provide these minimum requirements that can be assumed also valid for the operational part of a wind farm, where the purpose is to measure the wind conditions reaching the wind farm.

Table 4.4: Definition of measurement campaign requirements for different terrain classes according to MEASNET "Evaluation of site-specific wind conditions " [41].

Terrain type	**Minimum measurement height**	**Representativeness radius of a measurement** (max. distance of any wind turbine to the next mast)
Simple terrain	2/3 hub height	10 km
Complex terrain	2/3 hub height	2 km

4.2.2 Location of representative Measurements specific for Solar Projects

For solar power plants the location of the instruments is of importance for the representativeness of the measurements as the irradiance can vary within the solar park due to clouds or shading. Soiling, albedo, wind speed and temperature can also vary strongly. For PV plants recommendations on the number and position of the instruments can be found in IEC 61724-1 [16].

The most important rules can be summarised to:

- **Distribution of sensors:**
  Sensors should be distributed over the whole area of the power plant and of course not concentrated at a single site

- **Location of the instruments:**
  The location should be selected as representative for the covered sections of the power plant

- **Radiometers**:

  – Avoid shading for radiometers mounted on the front side of the PV modules
  – For rear side instruments and albedo measurements, the number of required radiometers depends also on the variability of the ground
  – For rear side sensors multiple sensors are recommended at different positions on the module, as the illumination profile is not uniform
  – Specific information on the installation of radiometric stations can be found in [56]

4.3 Maintenance and Inspection Schedules

Data quality is usually dependent on the care taken for routine and preventive maintenance. While this can be a requirement in a real-time environment, it is difficult to control. The alternative to setting maintenance reports on the requirement list, is to routinely quality check the delivered data and set time limits for repair of instrumentation into the requirement list with a recommendation to follow the manufacturers maintenance schedules.

In the section 5 “Assessment of Instrumentation Performance”, the assessment of the data for the monitoring of the data quality provides recommendations that may make it unneccessary to require reports on maintenance schedules. It can in many cases be more straight forward to have a data quality measure in place.

Nevertheless, if such a maintenance and inspection schedule is part of the requirement list, the “Meteorological Monitoring Guidance for Regulatory Modeling Applications” [42] provides useful recommendation for all common meteorological parameters and their respective routine and preventive maintenance schedules. For most modelling applications, a 6 monthly schedule or a schedule according to the manufacturers recommendation, whichever is the shorter time period, is recommended.

The following list provides the most important components that such a protocols and reports should contain. A station checklist may be developed which should include the following components according to the monitoring guidance [42]:

- A List of safety and emergency equipment.
- List of items to be inspected following severe weather
- A checkoff to ensure there is adequate disk space for on-site storage of the raw data
- A checkoff to indicate that backup of data has been completed
- A checkoff to indicate that clocks have been checked and adjusted as necessary
- A checkoff for the cables and guy wires securing the equipment
- All routine and preventive maintenance activities should be recorded in the station log and/or on the appropriate checklist
- The station log and checklist provide the necessary paper trail to support claims of accuracy

In the following section maintenance requirements for specific instrumentation is being described and discussed in case the data quality monitoring requirements should contain maintenance schedule reporting.

4.3.1 Maintenance of Radiometers

The maintenance of the radiometers is often challenging. IEC 61724-1, ASTM G183 and ISO TR9901 recommend maintenance and inspection tasks including

- cleaning
- levelling/tracking

- ventilation
- desiccant
- general instrument status incl. cables
- and the quality control of the acquired data sets.

Also the intervals for the different tasks are given. With modern radiometers, and professional installation of the instruments, the task that has to be performed most frequently at the station itself is the cleaning of the sensors. The recommendations for cleaning are between daily (ISO), daily inspection and at least weekly cleaning (ASTM) and weekly cleaning (IEC). In fact the required maintenance frequency depends on the site, the instruments and the accuracy requirement as discussed in [56]. The quality control of the data must also be performed frequently as discussed above. Further information o the maintenance and inspection of solar measurement stations can be found in [56].

Chapter 5

Assessment of Instrumentation Performance

Key Points

- High measurement data quality requires continuous maintenance
- In real-time environments, insufficient data quality cannot be compensated for
- Uncertainty in measurements need to be taken into account as:
  (a) uncertainty from the calibration (and possibly installation) as part of standing data
  (b) uncertainty from known issues in the operation of instrumentation as part of the data delivery in real-time operation
- Quality assurance (QA) and quality control (QC) require a quality assurance project plan (QAPP), which is the management and documentation of QC and QA procedures and should include:
  (a) Project Description and organisation
  (b) QA procedures and monitoring
  (c) Design of QC procedures

The quality of measured data that is supposed to be used in real-time applications is of immense importance. Modern technologies to assimilate measured data into forecasting models and processes usually have automatic algorithms that blacklist data that is out of range or in other ways suspicious, because bad data is often worse for the forecast model than no data.

While low quality data can be compensated with longer collection periods or in general longer time-series for non real-time applications such as plant assessment and configuration, resource assessment, etc; this is not possible in real-time environments, where the data has to be quality checked at the time of retrieval and processed or dismissed. Quality control and clear assessment criteria are therefore important tools for the success of the application using real-time measurements. In this chapter, we will describe challenges and mitigation strategies for the measurement data processing and quality control.

5.1 Measurement Data Processing

Measurement data processing is a non-trivial task due to the expectation that measurements are to show the real conditions of the measurand. This may be so in laboratory experiments, but is seldom the case in atmospheric monitoring. The uncertainty in the correctness of measurements are due to two aspects:

1. *instrument uncertainty*:
   - incorrect calibration
   - failures/errors in signal processing (e.g. inexact values of constants and other parameters obtained from external sources*)
   - finite instrument resolution or discrimination threshold*
   - inexact values of measurement standards and reference materials*
   - approximations and assumptions incorporated in the measurement method and procedure* j) variations in repeated observations of the measurand under apparently identical conditions.
   - incorrect mounting
   - yaw misalignment
   - limitation of line-of-sight
2. *environmental uncertainty*:
   - obstacles
   - interfering signals
   - wake effects

*[1] There are many more uncertainties in measurements of atmospheric monitoring (see e.g. 3.3 Uncertainty in [6]). Nevertheless, the above list is sufficient evidence to point out how important quality assurance and control as well as the knowledge of the uncertainty coming with instrumentation is for real-time forecasting purposes. The following sections

[1]Items marked with a "*" are items also mentioned in [6].

will describe a number of *known issues* in relation to wind and solar forecasting that are relevant considerations for the end-users to understand possible quality issues in the received data, and for stakeholders to acknowledge the importance to not underestimate the resources that are required for high-quality data collection and processing in forecast models.

In the next two sections, some of the important aspects of uncertainty expression in measurments and *known issues* for measurements associated with real-time forecasting for wind and solar projects will be discussed.

5.2 Uncertainty expression in measurements

The “Guide to the expression of uncertainty in measurement” describes the *result of a measurement only as an approximation or estimate of the value of the measurand and thus is complete only when accompanied by a statement of the uncertainty of that estimate* [6].
In section 4.1 and 4.3 it was described that requesting uncertainty measures from instrumentation in real-time is complex. A first step to integrate a solution, although often too limited, is to provide a standing data value as a percentage that is determined at the setup/calibration of the instrument and provided as part of the standing data. In that way, at least the instrument specific uncertainty could be accounted for in the handling of measurements.

1. **Formal requirement for uncertainty expression in measurements**:
   For a standardised technical requirement on uncertainty expression in measurements, the JCGM guides [20, 59, 6, 60, 61] offer a valuable general source, also applied in meteorology and oceanography. In that way, a harmonisation of "best practices" with these directly related real-time disciplines can be achieved. In fact, the guides do not only consider the measurand as a physical quantity, but also provide guidance to the conceptual design and the theoretical analysis of measurements and methods.
   One way to carry out an uncertainty estimation is with the Monte-Carlo method described in [60] (pp23-33).

2. **Informal practical solution for uncertainty expression in measurements**:
   For wind measurements, a practical solution for the expression of uncertainty is to add a mean uncertainty value to raw measurements, as applied by Pinson and Hagedorn [51] in an experiment over Ireland and Denmark with wind measurements from standard met masts [Pinson and Hagedorn, 2012 p7].
   For solar projects it is well known that maintenance and ambient conditions of the measurement heavily influence the uncertainty (see section 7.2 Measurement Uncertainty in the PVPS Handbook [56] for more details). These influences should be included in the uncertainty estimates. Influences such as the solar position or ambient temperature can be considered automatically. Other influences such as, e.g. the levelling or soiling of instruments is more difficult to include in the uncertainty estimates, but may be added to the standing data as described above or in [6].

Whether the formal or informal way is chosen for the quantification of uncertainty in measurements, it is recommended to take the uncertainty of the allowed instrumentation for a real-time forecasting application into account in the following way:

(a) **Standing data**: uncertainty from the calibration (and possibly installation) should be part of the standing data of the instrumentation. In that way, forecasters can take this uncertainty into account in there quality control and data screening.

(b) **Data Provision Flag**: uncertainty from known issues in the operation of instrumentation should be flagged as part of the data provision in real-time operation, if the instruments provide such a flag. Most atmospheric instruments complying to e.gċlass A and B (IEC 61400[14], PVPS Handbook[56]) provide such flags and only require the IT solution to allow for an additional value.

5.3 Known issues of uncertainty in wind and solar specific instrumentation

This section provides a few *known issues* associated with the specific use of of instrumentation at the nacelle of wind turbines.

5.3.1 Effects of uncertainty in nacelle wind speed measurements and mitigation methods

From a theoretical perspective, it can be expected that wake effects from the rotating blades are strongest at high speeds and low speeds that affect the mean wind flow, but not so much the correlation in the "normal" operating range. When plotting nacelle anemometer measured data in a scatter plot, where the linear relationship is strong in the range 3-12m/s and along the linear part of the power curve, this becomes most apparent. Below cut-in and above the wind speed where the power curve gets flat, the linear relationship does not hold any more. It has also been shown with frequency distributions that met mast anemometers produce an approximate Weibull distribution, where nacelle mounted instruments often produce strong biases at the lower wind speeds affected by wake effects.

This is due to 2 effect:

1. wake effects from rotating blades
2. Yaw misalignment of wind turbine

In order to make use of nacelle mounted instrumentation, corrections are necessary. The problems associated with these known issues are described in the following, together with possible correction methodologies.

5.3.1.1 Wake effects from rotating blades

Drechsel et al. [2012] found out that the wake effects of nacelle measured wind speeds are highest up until the cut-in wind speed and above approximately 10m/s, where the power curve starts getting flat. Allik et al. [2014] found out in a study with nacelle mounted cup anemometers, nacelle mounted sonic anemometers and a reference met mast that the mean of the three measurements did not coincide very well. The nacelle wind speed measurements, due to the wake effects of the blades, have a much lower mean. The correlations however were strong between cup anemometer and met mast anemometer and even stronger between sonic anemometer and met mast readings within the range of 3-12m/s.

Jing et al. [34] also found that blade wakes affect the measurement of nacelle anemometer and result in the inconsistency between nacelle wind speed (NWS) and free stream wind speed, which seriously affects the power forecasting and performance evaluation of wind turbine. They propose a practical way of overcoming such wake effects by developing a site calibration model with a NWP model or CFD model and an independent measurement from a met mast or scanning lidar to have a model calibrated for free wind-stream. This model can then be used to correct for the wake effects. In their data-driven calibration method, reported in [34], a “Relevance Vector Machine” was used to establish a site calibration model between a Lidar wind speed (LWS) and a numerical weather forecasted wind speed (NWS). Once the model is calibrated, the calibrated LWS can be used to replace the free stream wind speed and wake effects on nacelle anemometers.

5.3.1.2 Yaw misalignment of wind turbine for scanning lidars

Held et al. [26] reports that due to the limitation of line-of-sight measurements and the limited number of focus positions of a scanning Lidar, assumptions are necessary to derive useful inflow characteristics at the turbine nacelle. The horizontally homogeneous inflow assumption is violated, if a wake impinges the field of view of one of the Lidar beams. In such situations, the turbine yaw misalignment measurements show large biases which require the detection and correction of these observations.

5.3.2 Application of nacelle wind speeds in Real-time NWP Data Assimilation

The only recorded project to date that carried out dedicated real-time studies with nacelle wind speeds in a real-time forecasting environment so far is the US Department of Energy funded Wind Forecasting Improvement Project (WFIP). The project had a demonstration phase of 1-year and used 410 nacelle wind speeds for the data assimilation of NOAAs models [Wilzcak, 2014, Marquis, 2014].

One of the main findings in the experiment was that the nacelle wind speeds were contaminated too much by wake effects to be useful as individual measurements. Due to the constraints in the data assimilation techniques, it was important to find a strategy that made it possible to use the raw data from the cup anemometers. The research team of NOAA found that the best way to handle the contaminated data was to average the individual turbine data per wind farm and then blacklist those measurements that were outside the range of 2 standard deviations from the mean of the wind farm. This is a reasonable constraint to ensure that measurements contaminated by wake effects will not be passed into the assimilation program.

Additionally, the measurements were averaged over the nearest model grid point in the numerical weather prediction model. By doing this, it was possible to remove systematic biases and make use of the direct outcome of the model at the grid points.

To summarise, the strategy to use all 411 nacelle measured wind speeds at 23 wind farms has been:

- averaging wind speed measurements over each wind farm
- blacklisting measurements that were more than two times a STD from the mean
- interpolating and averaging at the nearest grid point of the NWP model
- BIAS correcting at the model grid points

The advantage of this approach is that wake effects are smoothed out through the averaging within the wind farm and averaging at the NWP model grid points ensures that bias corrections are brought forward to the model result, i.e. the wind power forecast. In this way, it could be demonstrated that nacelle wind speeds can become useful signals seen from a general forecasting perspective.
Pinson and Hagedorn [51] used a different path to reduce uncertainty of the 633 meteorological stations with cup anemometers that they compared to model results. Their assumption was made according to the recorded uncertainty of unbiased state of the art anemometer uncertainty, which is a standard deviation of around 0.5m/s. It was shown that this was a reasonable and valid assumption. However, it is not known how much this assumption is dependent on the number of measurement units and their distribution. Therefore, such assumptions must be considered with care.

5.3.3 Known uncertainty in Radiation Measurements

In the PVPS handbook [56], uncertainty is also recommended to be treated according to the Guide to the Expression of Uncertainty in Measurements (GUM) [6], also reflected in the ISO 2008. Here, a detailed descrption of the "Estimation of Calibration and Field Measurement Uncertainty" and mitigation methods are described, specific for solar applications.

The PVPS handbook describes and details the GUM procedure specific for solar applications in section 7.2.1 in six steps [56]:

1. <u>Define the measurement equation for the calibration and/or measurement system</u>
   The refers to the mathematical description of the relation between sensor voltage as well as any other independent variables and the desired output (calibration response or engineering units for measurements)

2. <u>Determine the sources of uncertainty</u>

   (a) uncertainty from statistical calculations, specifications from manufacturers, and previously published reports on radiometric data uncertainty or professional experience
   (b) uncertainty associated with SZA response, spectral response, nonlinearity, temperature response, thermal loss, data logger accuracy, soiling, and calibration, including the drift of the calibration constant(s)

3. Calculate the standard uncertainty, u
   Calculate an u value for each variable in the measurement equation by using:
   (a) Type A: statistical method
   (b) Type B: uncertainty component from manufacturer specifications, calibration results, and experimental or engineering experience

4. Compute the sensitivity coefficient, c
   The coefficient weighs the various uncertainties of the variables in a measurement equation and provides appropriate contributions of uncertainty of each input factor for the irradiance value.

5. Calculate the combined standard uncertainty, u c.
   This is the combined standard uncertainty using the propagation of errors formula and quadrature (square root sum of squares) method. It is applicable to both Type A and Type B sources of uncertainty.

6. Calculate the expanded uncertainty (U 95).
   The expanded uncertainty is calculated by multiplying the combined standard uncertainty by the coverage factor, typically by applying the Students t-analysis to determine the appropriate value of k (typically 1.96 for 95% and 3 for 98% confidence, respectively, for large data sets assuming a Gaussian distribution

These six steps are considered a cycle when quantifying the uncertainty of (ir-)raiation measurments. Step 6 can in that way be for example used as an input to a calculations of the performance ratio of solar conversion systems: to calculate the ratio of system output/solar input, the expanded uncertainty in Step 6 is used as an input to evaluate the denominator (solar input), and the cycle continues to ultimately quantify the expanded uncertainty of the performance ratio. Further, these steps are applicable to the quantification of the uncertainty in both calibration and field measurements (see section 7.2.1 [56]).

Figure 5.1 provides typical calibration uncertainty for pyrheliometers and pyranometers and shall here provide a guideline and recommendation on how uncertainty of pyranometer and pyrheliometer instruments may be provided in standing data or limited as accuracy requirement.

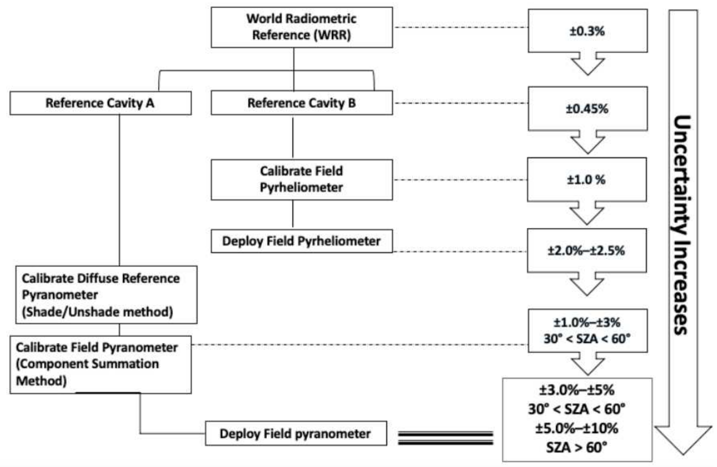


Figure 5.1: Calibration traceability and accumulation of measurement uncertainty for pyrheliometers and pyranometers (coverage factor k = 2), ©PVPS Handbook, Figure 7-2 [56] by NREL

5.4 General data quality control and quality assurance (QCQA)

The following general data quality and assurance process recommendations are based on the *Meteorological Monitoring Guidance for Regulatory Modeling Applications* [42] of the Environmental Protection Agency's (EPA) definition of QC, QA and QCQAPP. We have translated these guidelines for use in real-time wind and solar forecasting with support from meteorological and power measurements.

The underlying principle, in line with EPA's Quality Assurance/Quality Control (QA/QC) procedures, is to ensure that the data collected meet standards of reliability and accuracy.

- **Quality Control (QC)** is defined as those operational procedures that will be routinely followed during the normal operation of the monitoring system to ensure that a measurement process is working properly. These procedures include periodic calibration of the instruments, site inspections, data screening,data validation, and preventive maintenance. The QC procedures should produce quantitative documentation to support claims of accuracy.

- **Quality Assurance (QA)** is defined as those procedures that will be performed on a more occasional basis to provide assurance that the measurement process is producing data that meets the data quality objectives (DQO). These procedures include routine evaluation of how the QC procedures are implemented (system audits) and assessments of instrument performance (performance audits).
  It is important that the person providing the QA be independent of the organisation

responsible for the collection of the data and the maintenance of the measurement systems. Ideally, there should not be any lines of intimidation available to the operators which might be used to influence the QA report and actions.

- **Quality Assurance Project Plan (QAPP)** is the management and documentation of QC and QA procedures and should include the following items (which are a subset of the EPS guideline [42]), relevant for real-time wind and solar projects:

1. **Project Description and organisation**
   - Description of Measurement Application and Usage in the Forecasting Process
   - Validity Support of Measurement Data by the end-user
   - QA objective and documentation for validity claims
   - Documentation of Calibration method (and frequency) requirements
   - Documentation of Data flow - from samples to archived valid values
   - Documentation of Data flow - from samples to archived valid values
   - Development of a Validation and reporting method for Forecasting Process
2. **QA procedures and monitoring**
   - Performance monitoring by end-user
   - Reporting of data validation and verification by independent agent
   - Procedures to implement QA objectives
   - Management support for corrective action and reports
3. **Design of QC procedures**
   - Analysis of time dependency of quality control
   - Analysis of required data quality for forecasting process
   - Analysis of sampling requirements
   - Selection of data screening methods
   - Selection of validation and verification methods

When followed, the QA task reduces to spot checks of performance and examination of records thus providing the best data with the best documentation at the least cost Once the QA and QAQPP are planned and setup, the quality control of measurements needs to be designed in order to ensure that the data quality is sufficient to be used in the forecasting process. The following sections therefore provide recommendations separated for wind and solar forecasting purposes and on two different time horizons:

1. QC in real-time forecasting at the time of receipt

2. QC in historic mode:
   quality control of a specific time interval of data such as a day, a month or one or more years for resource assessment, performance monitoring, model/forecast training/tuning etc.

5.5 Historic Quality Control (QC)

There are different ways of establishing a quality control procedure and management. Dependent on the level of quality that should be achieved, there are practical solutions that the forecasting party can take over or an independent consultant or alternatively the end-users themselves. The decision on that is dependent on the purpose of the quality control, the requirements to the quality of the measurements and transparency.

In the following principles that are also used or recommended QC processes from standards or guidelines in site assessment auditing, resource assessment and meteorological monitoring have been translated for real-time applications.

The recommendation section 6 will distinguish some of these principles with respect to the desire of conformity and relevance to the mentioned quality standards. Note that all principles are selected to ensure the end-user and forecast provider using the meteorological data in the forecasting process with high-quality data for real-time applications.

5.5.1 QC for Wind Forecasting Applications

In most cases the QC methodology should be designed for both long-term analysis of e.gå year and shorter periods such as weekly, monthly or quarterly examination of observational data signals. There are two targets for the validation and quality control:

1. To identify the amount of valid data submitted

2. To produce a comprehensive analysis which will provide the wind farm owner with a description of the root of the detected error(s) in the signals.

3. To limit the time from when an issue with data signals starts until it is diagnosed and solved

4. To avoid cross correlations between measurement from different sources to fill data gaps

The last item is too much challenged by irregular distances of the generating unit's locations and would only produce justifiable results, when done over long verification periods, which is often not desired in real-time applications.

The "ISO Guide to the Expression of Uncertainty in Measurements" [20, 6, 61] and its 2 supplements [[59, 60] from the Joint Committee for Guides in Meteorology (JCGM) distinguish between two types of measurement uncertainty that are to be accounted for in any standardised taken measurement. These guidelines are relevant also for QC of measurements with the purpose of supporting real-time wind and solar forecasting:

1. systematic errors, which are also known as measurement bias, often associated with offsets of the measured quantity

2. random errors, which are associated with the fact that 2 measurements of the same quantity are seldom the same

In section 3.1.2 of the guide, [20, 60] it is stated that "the result of a measurement .. is only an approximation or estimate .. of the value of the measurand and thus is complete only when accompanied by a statement of the uncertainty ... of that estimate". Considering this definition, all measurements should ideally have an uncertainty term associated with it. This is impractical in real-time operations, where the value of the measurements lies in the availability of the data at a given time. Therefore, it is unrealistic to request uncertainty measures. However, it could be a standing data value that is determined at the setup of the instrument and provided as part of the standing data. In that way, the instrument specific uncertainty could be accounted for in the handling of measurements.

The alternative is to carry out an uncertainty estimation with e.g. the Monte-Carlo method described in [60] pp23-33) or a mean uncertainty value must be added to raw measurements, as applied by Pinson and Hagedorn in an experiment over Ireland and Denmark with wind measurements from standard met masts [Pinson and Hagedorn, 2012 p7]. If a more standardised technical requirement is desirable, the JCGM guides offer a valuable general source, also applied in meteorology and oceanography. In that way, a harmonisation of "best practices" with these directly related real-time disciplines can be achieved. In fact, the guides do not only consider the measurand as a physical quantity, but also provide guidance to the conceptual design and the theoretical analysis of measurements and methods.

In the introduction to the Guide [JCGM, 2009], it is stated that ..the principles of this Guide are intended to be applicable to a broad spectrum of measurements, including those required for:

- maintaining quality control and quality assurance in production
- complying with and enforcing laws and regulations
- calibrating standards and instruments and performing tests throughout a national measurement system in order to achieve tractability to national standards
- developing, maintaining, and comparing international and national physical reference standards, including reference materials

To summarise, the handling and integration of wind power into the electric grid is an equally important step to harness the full potential of the energy resource in an efficient and environmentally friendly way.

This requires that measurements are trustworthy and maintained to a quality that allows for their use in forecasting tools in order to produce high quality forecasts and thereby reduce the need of reserves. These guides in combination with the IEC 61400-1 standard provide a good foundation for any grid code technical requirement specifications.

5.5.1.1 Specific Control Procedures

In the following a few specific control procedures for the most common instrumentation are provided.

1. Cup anemometer on a met mast:
   The accuracy of a single value from a cup anemometer is not high without a 10min time average. The purpose of the 10min averaging process is to eliminate the impact of the turbulent motion, which is generated as a result of frictional forces from the terrain on the air as well as the imbalance in the diurnal cycle and the temperature difference between the air and the surface. From a 15 minute data delivery of a noise contaminated signal, it is almost impossible to prove the correctness or falsify the data signals, because some of the values are realistic and others are not.
   In contrast to the nacelle measurements, a single cup anemometer on a mast can be inspected at a much lower cost, today even with a drone. Several anemometers can be mounted on the mast and it is at least possible to submit the data directly without giving the wind farm software the possibility to delay, block or manipulate the data.

2. **Cup anemometer on a wind turbine nacelle:**
   Previous studies [18, 64, 69, 57, 33] agree that the disturbances, whether it be flow induction or wake induced, are of significant size for nacelle mounted measurement units. While there has been progress over the years in applying precautionary measures to reduce the risk of measurement errors and thereby reduce the uncertainty of the measurement signals, the physical aspects leading to the disturbances cannot be resolved for units that are placed behind the rotor. Unfortunately, all studies that looked at high wind speeds (> 15m/s) up to cut-off wind speeds concluded that the relationship to the reference measurements from met masts were no longer linear and deteriorated strongly. This also means that nacelle mounted measurement units are unsuitable in the control room regarding ramping in high speed wind events as well as pitch regulations due to curtailment or safety, because it is throughout these times, where the measurements are most unreliable. At the time of writing, there are limited peer-reviewed studies that would provide any hint of these instruments providing proven and consistent quality in a real-time operation.

3. **Remote sensing devices:**
   Nacelle-mounted lidar systems offer the possibility of remotely sensing the inflow of

wind turbines. Due to the limitation of line-of-sight measurements and the limited number of focus positions, assumptions are necessary to derive useful inflow characteristics.
Typically, horizontally homogeneous inflow is assumed which is well satisfied in flat, homogeneous terrain and over sufficiently large time averages. However, it is violated if a wake impinges the field of view of one of the beams. In such situations, the turbine yaw misalignment measurements show large biases which require the detection and correction of these observations.

Correction Example for complex flow [26] have developed a method to detect wakes in the inflow of turbines using nacelle lidars and developed a correction method. Here, a detection algorithm is proposed based on the spectral broadening of the Doppler spectrum due to turbulence within the probe volume. The small-scale turbulence generated within wake flows will typically lead to a significantly larger broadening than in the ambient flow. Thus, by comparing the spectral widths at several locations,situations, where a wake is impinging the field of view of one or more beams can be identified. The correction method is based on an empirical relationship between the difference in turbulence levels at distinct beams and the difference in wind direction derived from the lidar and the real wind direction. The performance of the algorithm is evaluated in a field experiment identifying all wake situations, and thus, correcting the lidar derived wind direction.

4. **Blade-pressure computed nacelle wind:**
A relatively new development is the nacelle wind speed computed with a method that evaluates signals from blade pressure sensors. In one recent analysis in Ireland [47] the data signals from the computed nacelle wind show a better fit of the met signals to the power production compared to the data signals from met masts. Among the nacelle sourced met signals, such signals often provide the most reliable sourced met data. The better fit of these data has been found due to the calibration of the signals to fit the wind turbine's and/or wind farm's power curve.
There are benefits and drawbacks with this method, which need to be examined before permission is given and controlled regularly: The alignment with the power production is probably the largest benefit, while costs are another. The drawback on the other hand is that the signal is not independent of the wind turbines and is prone to failures under cut-off or other outages of the turbine, where an independent measurement would still provide a representative picture of the weather situation at hand. Special care need to be taken for these cases, i.e. when the blades are pitched and do no longer provide a representative wind measurement. Unless there is a calibrated (against flow induction, wake effects and yaw misalignment) nacelle wind anemometer installed that provides or corrects the signals, when the turbine blades are in some way pitched and indicates with a flag which type of data is provided, it is not recommended using such data for forecasting purposes with a time horizon of minutes or hours. Otherwise,

or maybe in combination with a met mast at reduced (lower than hub height) height, these measurements provide a cost-effective alternative to instruments mounted on met masts or lidars, when hub height wind measurements are necessary.

5.5.1.2 Practical Methodology for quality control of measurement for wind applications

A practical methodology to apply for the validation and quality control of meteorological measurements is a combination of different consistency checks:

- Missing Values and time stamps
- Forecasted wind speed versus measured wind speed
- Forecasted temperature, wind direction against measured values
- Forecasted power versus active power checked with SCADA MW
- Computed active power from measured wind speed versus actual active power
- Comparison against previous years of the same wind farm
- Comparison to the average error level for wind farms in the same period

5.5.1.3 Statistical tests and metrics for the QC process

One practical solution to test measurement signal quality is to use verification methods similar to the verification of forecast errors with the exception that a forecast or a forecast range of the respective variable is used as the reference, because it is the measurement that needs validation. The forecast has a known accuracy level, which is used to find changes in quality in the measurement signals. By using e.g. ensemble forecasts, the uncertainty of the measurement signals can also be quantified (see e.g. [43]).

By validating in different sub periods of the year, it can be shown whether the error pattern has been temporary or on a long-term basis.

By using a variation of different statistical tests as recommended in part 3 of this recommended practice [2], the data basis is large enough to interpret the data accuracy. The following set of statistical metrics are recommended:

1. **BIAS:**
   The BIAS in itself should be low, but is no guarantee of correctness of the data, because a BIAS can be low for the incorrect reason
2. **MAE:** MAE and BIAS together show, if the data has an offset.
3. **RMSE:** There are few extreme errors, if the ratio RMSE/MAE exceeds 1.3.
4. **CORRELATION:**
   The correlation allows for easy detection of constant measurements as well as sign errors.

5. **Frequency distribution:**
   The frequency distribution from e.gȧ one year data set of 15-min mean values of a wind speed should be a smooth curve with decreasing probability of high wind speeds. A temporary instrument fault will be visible as a skewness of the curve. Comparing the frequency distributions of an ensemble mean forecast against measurements is recommended in this case, as a mean smooths outliers in the data set.

   Positive and negative phase errors between a forecast and measured data tend to cancel each other out over a long enough period. Therefore, a high similarity between two independent time series of the same physical variable can be expected.

The formulas of the test metrics can be found in the Appendix B.
A graphical analysis of measurement signals with these metrics can then also be used to define acceptance limits for the meteorological variables that are to be delivered.

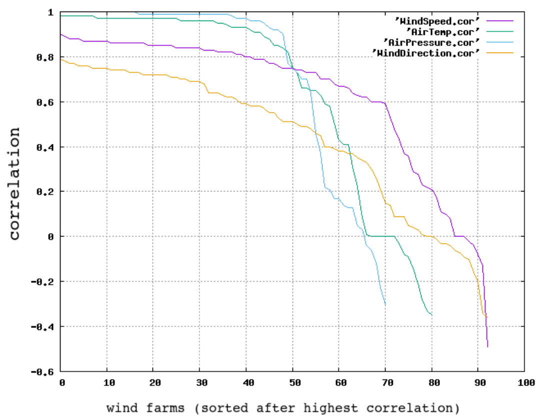


Figure 5.2: *Example of a graphical analysis of met data signals for 4 variables to define acceptance limits. The x-axis shows the percentual amount of wind farms ordered starting from lowest to highest error score (y-axis) .*

Figure 5.2 and 5.3 show examples of results of a historic quality control procedure of data signals in form of CORRELATION, MAE and BIAS for the following 4 variables:

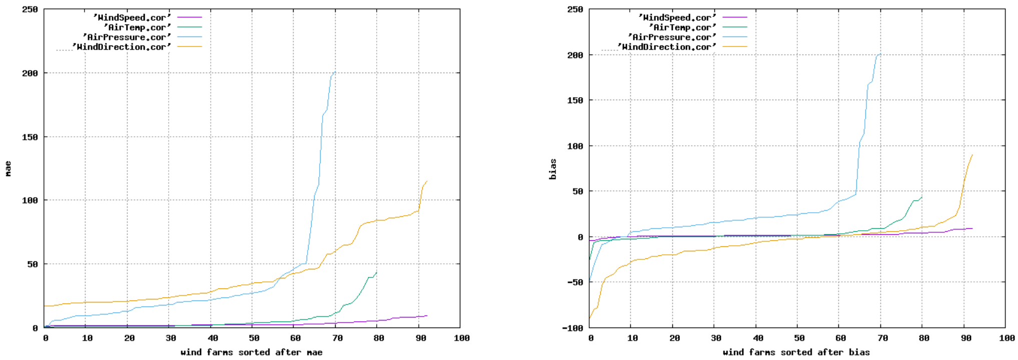
(a) wind speed

(b) wind direction

(c) air temperature

(d) air pressure

Each metric is calculated for one wind farm. The ranking for the quality will always be variable dependent. For example temperature varies slower than wind speed and achieves therefore in general a higher correlation.
With such a graphical analysis it is possible to define acceptance limits for the quality of the data in terms of BIAS, MAE and CORRELATION. This type of validation procedure



(a) Historic MAE QC **(b)** Historic MAE QC

Figure 5.3: Example of a graphical analysis of met data signals for 4 variables to define acceptance limits. The x-axis shows the percentual amount of wind farms ordered starting from lowest to highest error score (y-axis).

has been found [47] to allow for the examination of the wind farm met signal data in an anonymous form suitable to keep overview of the data quality. The graphs directly show the percentage of capacity that lacks quality. Usually, temporary outages show up as spikes in data time series. Although forecasting systems have to have an ad-hoc quality control and must contain a certain fault tolerance for blacklisting spiky data, some of the spikes are weather related and may in certain waether situations no longer be distinguishable from data that is realistic. Extreme errors often occur in individual months, after periods of higher reliability and indicate a weather dependency [47].

5.5.2 QC for solar applications

There are several sets of QC tests for historic radiation data (BSRN, SERI QC, QCRad, MESOR, ENDORSE. The tests in the literature use different limits for individual irradiance components (DHI, GHI, DNI) or parameters derived from these components together with solar position and at times clear sky irradiance. Types of limits are

- physical possible limits
- rare limits
- extremely rare limits

The existing QC tests have been harmonized in the framework of IEA PVPS Task 16 for stations with measurements of all three irradiance components (GHI, DHI, DNI) or two (GHI and DHI or DNI) and are documented in chapter 3 in the section 3.4.2 Data Inspection of the PVPS Handbook [56] in more detail.

The quality control consists of automatic tests and visual inspection by an expert. The QC results in one flag per time stamp and test. The flag's value is either "data seems ok", "data point was detected as problematic" or "test could not be performed". The latter can occur for a missing time stamp/data, or if the test domain was not met and the test could not be applied.
The visual inspection of the data is of importance to detect bad data and manually assign the flag. This also includes the control of the meta data (logbook with comments, calibration information). Visual inspection can also help to determine. if the time stamps refer to the end of the averaging interval (e.g. 1min, 10min or 1h averaging). The correct interpretation of the timestamps is essential. The following tests are defined. All test results are visualized and manual flagging can be used to improve the automatic test results. Some tests are not automated, but purely visual inspections:

- Missing time stamps
- Missing values
- K-Tests[2]
- Baseline Surface Radiation Network (BSRN) QC
  - three component test (if DNI, DHI and GHI are measured, one can calculate GHI from DHI and DNI and the solar elevation angle and compare it to the measured GHI)
  - extremely rare limits test
  - Physically possible limits (PPL) test
- Tracker off test (test that checks for high GHI and low DHI at DNI close to zero)

[2]The test compares three K values that are ratios of the global and direct irradiance components to the extraterrestrial irradiance and the ratio of diffuse to global irradiance. The tests check if these values are within certain common intervals (see details in chapter 3 “Measuring Solar Radiation” of the PVPS handbook[56]).

- Visual inspection

The PVPS handbook [56] provides some information about "gap filling" methods in section *3.4.2.1 Data Quality Control and Assurance* to fill out missing values or time stamps in time series. While this is relevant for the generation of bankable data in resource assessment, it is more important in real-time applications to ensure that data gaps are not counted into the verification, where they can lead to misleading results, if they are e.g. filled with zeros instead of a well defined flag in the databases, where the data is stored for historic evaluations.

For the visual inspection several plots are needed. Heat maps of the irradiance with axes hour of the day (y) and day of year (x) are used. To check for radiometer shading heat maps of the maximum irradiance that occurred in bins of Sun elevation and azimuth angle with axes elevation (y) and azimuth (x) are commonly used. Objects can be seen in this plot as patterns of low irradiance. The deviation of the DNI measurement from the DNI calculated from GHI and DHI should also be inspected (e.g. as heat map or as a plot vs. time). The AM/PM symmetry of the GHI data should be visualized. Calibration changes can be visible in a plot of the clear-sky index (measured GHI/extraterrestrial GHI) vs. time.

To decide, if a data point can be used not only the result of a single flag per timestamp is needed, but all flags for that timestamp and surrounding timestamps must be considered.

For the validation of satellite and model derived radiation data the following procedure is recommended: A data point is usable, if

- all individual QC flags are indicating that the data is ok
- the test could not be performed and all measured radiation components are available

Data should be excluded, if

- 30% or more of the timestamps (daytime, solar elevation >0) from one day are flagged, exclude the entire day
- Intervals between flagged data that are shorter than 60 min are also excluded

For the determination of the length of the interval only timestamps with elevation >0 are considered. Less stringent exclusion rules could be applied for other purposes, such as the determination of yearly sums.

5.6 Real-time Quality Control (QC)

The differences between real-time quality control and historic quality control (QC) is somewhat different. The major difference is that real-time QC is restricted to the data values in real-time and hence the methods used need to relate to the historic validations, calibration results and pre-defined accuracy limits and levels, but not all can be conducted and validated in real-time.

For most general quality control of measurment data, there is no difference, whether the measurements are targetted to wind or solar forecasting applications. For this reason, the

first part of this section describes the recommended quality control procedures that apply for both wind and solar forecasting applications.
A general description of errors in both wind and solar measurements can be summerised to (see e.g. [36, 37]:

(1) **data management errors:**

  (a) data transcription and collection

  (a) errors that occurred during data manipulation (e.g. duplication of data sequences)

  (a) standardization of practices (e.g. measurement units, reference times)

(2) **measurement errors:**

  (a) temporal or spatial consistency in the data

  (a) errors produced at the moment of sampling as a result of :
    – instrumental malfunction
    – calibration
    – exposure problems

The data management errors can in many case be avoided or solved fast, if there is a QCQA project plan (QCQAPP) as described for the historic QC in section 5.4 that contains the neccessary documentation to source the errors.

5.6.1 Data screening in real-time Wind and Solar Forecast Applications

General rules for data sceening in real-time forecasting applications apply for both, wind and solar. The most common data errors that a quality control system needs to catch are:

- repeated values
- values out of range limits (see 6)
- formatting of data file/decoding error
- consistency of time stamps/chronologically
- data sampling and time averaging definitions (see 5.6.2
- site coordinates
- instrumentation identification

5.6.2 Data Sampling Thresholds in real-time Wind and Solar Forecast Applications

Section *4.5 Sampling Rates* of the EPA guidance *on Meteorological Monitoring for Regulatory Modeling Applications* deals with sampling of data signals and averaging of data. These recommendations are to a large extend relaevant for the wind and solar real-time forcasting applications in the energy industry, as the recommendations are directed to the data quality required by numerical weather prediction models to add value. All wind and solar forecasting processes above a few minutes forecast horizon are based on NWP models; hence, our recommendation to follow these guidelines.
The following recommendations are adjusted for wind and solar forecasting applications;

1. Estimates of means should be based on at least 60 samples (one sample per minute for an hourly mean)
2. Estimates of the variance (standard deviation to within 5 or 10%) should be based on at least 360 samples (six samples per minute for an hourly variance).
3. To compute the mean wind direction:
    - Single-pass processing wind direction sensor (wind vane): samling once per second to insure that consecutive values do not differ by more than 180 degrees
    - Multi point analog recorders: sampling at least once per minute

5.6.3 Real-time QC for Wind Applications

5.6.3.1 Data Screening

The recommended data screening criteria for wind and solar projects as a practical approach to data quality control and management follow the EPA guideline [42] and are suggested with some limits that are easy to control. Table 6.6 provides some useful and practical limits that can be recommended for the data screening task.

Table 5.1: Practical data screening criteria for data quality control of wind projects

Variable	Screening Criteria: Flag data if the value
Wind Speed	- is less than zero or greater than 50 m/s - does not vary by more than 0.1 m/s for 3 consecutive hours - does not vary by more than 0.5 m/s for 12 consecutive hours
Wind Direction	- is less than zero or greater than 360 degrees - does not vary by more than 1 degree for more than 3 consecutive hours - does not vary by more than 10 degrees for 18 consecutive hours

Variable	**Screening Criteria: Flag data if the value**
Temperature	- is greater than the local record high - is less than the local record low (limit could be applied on a monthly basis) - is greater than a 5°C change from the previous hour - does not vary by more than 0.5° for 12 consecutive hours
Temperature Difference	- is greater than 0.1°C/m during the daytime - is less than -0.1°CC/m during the night time - is greater than 5.0°C or less than -3.0°C
Dew Point Temperature	- is greater than the ambient temperature for the given time period - is greater than a 5°C change from the previous hour - does not vary by more than 0.5°C for 12 consecutive hours - equals the ambient temperature for 12 consecutive hours
Pressure	- is greater than 1060 mb (sea level) - is less than 940 mb (sea level) (values should be adjusted for elevations other than sea level) - changes by more than 6 mb in three hours
Radiation	- is greater than zero at night - is greater than the maximum possible for the date and latitude

5.6.4 Real-time QC for Solar Forecasting Applications

For the real-time QC of solar radiation data, the QC methods for historic data can only be partly applied. QC methods requiring visual inspection cannot be applied automatically. Such methods can be used to analyse errors that are detected differently.

The following tests can also be applied in a real-time environment (more details are provided in section 5.5.2):

- Missing time stamps
- Missing values
- K-Tests[3]
- Baseline Surface Radiation Netork (BSRN) QC
  - three component test
  - extremely rare limits test
  - physically possible limits (PPL) test

QC results indicating that the result is suspicious such as the above described tests for rare limits or some versions of the three component tests might lead to too many data gaps,

[3]The tests compares three K values that are ratios of the global and direct irradiance components to the extraterrestrial irradiance and the ratio of diffuse to global irradiance. The tests check if these values are within certain common intervals (see details in chapter 3 “Measuring Solar Radiation” of the PVPS handbook[56]).

depending on the real-time application. If this is the case, these strict QCs should be adapted with less strict limits or used only for the further analysis of the data once another QC test was not passed.

Limits should always be defined based on the required data quality and data availability. A site dependence of these limits should be considered. For different instrument options different limits might be required, as e.g. cosine errors of pyranometers vary strongly between different models. Currently no standardised QC for real-time radiation data exist.

Chapter 6

Best Practice Recommendations

The Application Areas for the Recommendations:
The following application areas and associated applications are covered in this recommended practice guideline:

(A) *System Operation, Balancing and Trading*
- *Situational awareness in critical weather events*
- *High-Speed Shutdown events*
- *Grid related down-regulation or curtailments*
- *Short-term Forecasting with updates from measurements*
- *Intra-day Power plant balancing*

(B) *Wind Turbine, Wind Farm and Solar Plant Operation and Monitoring*
- *Wind turbine and Power Plant Control*
- *Condition Monitoring*

The information behind the recommendations
The recommendation's underlying information have been structured to account for a decision making processes for the use of real-time meteorological and power measurements for real-time wind and solar forecasting purposes in the following order:

- *Description of relevant meteorological instrumentation (section 2)*
- *Description of relevant power measurements (section 3)*
- *Setup and Calibration of relevant instrumentation (section 4)*
- *Quality Assurance and Quality Control (section 5)*

Principles for the Recommendations of Accuracy and Reliability:
The recommendations for the selection of instrumentation in this section are based on the following set of principles:

1. ***Accuracy requirements**:*
Accuracy requirements need to be defined for the application/project and aligned with the associated levels of effort necessary to operate and maintain the measurement system on under these constraints. An overall cost-performance determination should therefore always be carried out to adapt the budget to the accuracy requirements and vice versa.

2. ***Reliability requirements**:*
Reliability can be achieved with redundant instrumentation and/or high quality instrumentation. Redundancy enhances and ensures confidence in data quality. Selection of multiple instruments need to be aligned with the accuracy needs.

In the remainder of this chapter, the principles developed in these four previous sections are brought to the application level. In other words, the somewhat theoretical considerations from the previous chapters are now applied to real-world problems.

6.1 Definitions

This recommended practice guideline is targeted to the use of real-time meteorological measurements for real-time forecasting applications for wind and solar projects. We define real-time applications in the following way (see 1.4):

The aim of **real-time applications** is to have reliable information about the current weather situation; gaps (missing data) are more critical as for these times, forecasts or plant control parameters cannot be generated or adapted with the help of measurements. Forecasts and other control options with lower quality must then be used. Temporarily erroneous data is more critical as the forecasts or plant control might lead to energetic and economic inefficiencies, system security issues etc.

6.2 Instrumentation

In this recommended practice guideline, we have been investigating the usefulness and challenges associated with the use of the following instrumentation for real-time forecasting purposes:

1. **Instrumentation for Wind projects** (section 2.1)

   (a) Remote Sensing Instrumentation for wind farms (section 2.1.1, 6.4.1.2)
   (b) Nacelle instrumentation and measurements (section 2.1.2, 5.3.1, 5.3.2)
   (c) Cup anemometers (section 2.1.3)
   (d) Sonic and ultra-sonic anemometers (section 2.1.4)
   (e) Horizontally mounted nacelle LiDAR (section 2.1.5

2. **Instrumentation for Solar Projects** (2.2)
   (a) Point Measurements (section 2.2.1)
   (b) All sky imagers (section 2.2.2)
   (c) Satellite Data (section 2.2.3)

3. **Power Measurements for real-time operation** (section 3)
   (a) Live power and related measurements (section 3.4.1)
   (b) Power available signals (section 3.3)
   (c) Connection-point Meters (section 3.2.1)
   (d) SCADA Systems (section 3.1, 3.2.3)

The purpose of describing these instrumentation types is to enable end-users to verify and specify which order of selection to follow, (1) which resolution and/or accuracy ranges limit the use of certain instrumentation for pre-defined applications, or (2) which instrumentation can be applied for pre-defined criteria.

6.3 Recommendations for real-time measurements by Application Type

The choice of instruments is recommended to be made according to a standardized classification when the choice of requirements has lead to the definition of quality criteria by application type.

Table 6.1 provides a guideline to a best practice categorisation for the development of requirements in real-time forecasting applications to ensure high-quality measurements. For those applications, where there are two classes provided, the second, lower class should be understood as minimum requirement and the higher class as a requirement, if high quality is required and the respective budges justifiable.

The classes for wind projects are in detail defined in the IEC 61400 Part 12-1 [15] and explained in Annex D of the standard. Any further detail to the quality of the instrumentation should be referred to this standard. Some more detailed explanation can be found in section **??**.

The classes for solar or PV systems are defined in the ISO 9060:2018 [32], the WMO Guideline No.8 [66] and the IEC 61724-1:2021 [16], where differences, consistency and overlapping areas are explained in detail by the IEA Task 16 PVPS in their "Best Practices Handbook for the Collection and Use of Solar Resource Data for Solar Energy Applications" (PVPS-Handbook) [56] and a few more detailed differences for specific forecasting project types in section 4.1.2.

Table 6.1: Forecast applications and respective recommended requirements for appropriate instrumentation.

Forecast Application	**Requirements for Wind** according to IEC 61400 Part 12-1 [15]	**Requirements for Solar** according to ISO9060: 2018, EC 61724-1:2021 and WMO Guide No.8
System Operation Forecasting	Class A	Class A or B
Utility Scale Forecasting	Class A	Class A or B
Wind/Solar Park control	Class A	Class A
Park / Turbine Monitoring	Class S	Class B or C
Electricity Market Trading	Class B	Class B

Resolution and accuracy for the respective applications and parameters are provided by the standard's categories for the instrumentation. Manufacturers of such instrumentation follow these classifications and provide a straight forward decision tool for requirements for wind measurements and radiation measurements.

The drawback of this selection process is that other meteorological variables are not covered under these standards and need to be defined according to the resolution and accuracy method for the different applications in the following sections 6.4, 6.5, **??** and 6.6.

6.4 Recommendations for real-time Measurements for Power Grid and Utility-scale Operation

6.4.1 Recommendations on Quality Requirements

The choice of instrumentation should always take into account the need for appropriate measurement equipment in the sense of (see details in section 4.1):

(a) Accuracy and Resolution

(b) Reliability

(c) Availability

(d) Weather resistance and safety

6.4.1.1 Requirements for Wind Forecasting Applications according to environment

Table 6.2 describes the main and applicable types of instrument classes in their respective environments (see section 2.1 for details:

Table 6.2: Instrument classes for various environments.

Environment type	Environment description	Instrument class
Standard terrain	Flat terrain with few obstacles	class A
Complex terrain	varying heights and many obstacles	class B
Offshore	wave and current driven turbulence salty air	class S

The class S is associated with specific requirements, "where the influence parameter ranges are restricted to allow for the specified accuracy of the anemometer." [14]. Manufacturers of wind instrumentation that sell instruments for wind projects provide information about the class of instrumentation to make it easy for wind farm owners to purchase the appropriate instruments.

6.4.1.2 Wind Measurement Alternatives to Met Masts

There are various alternatives to met masts for the collection of meteorological measurements, such as lidars, sodars, radars, nacelle anemometer, nacelle lidars, sonic anemometer, etc (see more details in section 2.1. However, most of these have either a high measurement uncertainty due to e.g. weather conditions, obstacles, that limit these instrumentations to the more non-critical, or non-real-time use. Some may simply not have been tested for the purpose yet.
The two alternatives to met masts, that are most common, tested and recommended at present for grid operation and utility-scale operation (see detailed information in section 2.1.1 and 2.1.2 and 5.5.1.1) are:

1. Remote Sensing Devices: Lidars
   Lidars and Sodars can be an acceptable source of wind measurement in the real-time power system or utility scale operation, if the requirements for delivery rate and validity/uncertainty of the measured value can be kept within specific limits.

   The recommended technical requirements for lidars is to ensure high-quality data in long-term real-time operation and as alternatives to met mast measurements are:

   - measurements must be raw or technical requirements must include maintenance and software updates
   - lightning protection and recovery strategy after lightning measurements should be taken at a height appropriate for the wind farm, either at one of preferable at both hub height and around 30m

- instruments must be serviced and maintained by skilled staff
- version control must be maintained for signal processing
- wind characteristics data must be on wind turbine level
- Lidars and Sodars in complex terrain require special consideration

Details for these requirements can be found in section 2.1.1.

2. Blade-pressure computed wind speed at the nacelle (see also :
This method can be recommended for the use of real-time measurements for forecasting purposes under the following conditions:

   - the portfolio of wind farms contain a minimum of 30% well distributed wind measurements from other independent instrumentation such as met masts or from Lidars/Sodars
   - a well calibrated transition from blade-pressure computations to anemometer measurements at the nacelle in high-speed wind ranges, or down-regulation situations (pitched blades) (see section 5.3.1)
   - a combination of a met mast with instrumentation at lower levels, e.g. 10m, 20m and 30m is used together with the blade-pressure method to ensure coverage over the full wind range from 0-40m/s

   If these measures are taken into consideration, the blade-pressure based wind observations provide a cost-effective alternative to instruments mounted on met masts or lidars, when hub height wind measurements are a requirement. See more details about this method in section 5.5.1.1.

6.4.1.3 Recommendations for Solar Forecasting Applications

the following considerations should be taken in the selection process:

(a) Evaluation of the required parameters, data accuracy or uncertainty levels

(b) Cost-benefit analysis of instrumentation regarding quality and maintenance requirements

(c) Specification of number of instruments

Table **??** provides a very brief summary of the applicability of those classes that are considered relevant for solar energy forecasting at power grid level or utility-scale operation. Details can be found in section 4.1.2.

For the selection of requirements for real-time forecasting purposes, it is crucial to identify the needs regarding the quality of the instrumentations, as costs for the instrumentation, but also for operation and maintenance are significantly different. Details and decision support can be found in section 4.1.2.

Table 6.3: Pyranometer classes for various requirements.

Requirements	*Instrument class*	*Instrument description*
Meteorological or Energy forecasting/ modelling for utility scale PV systems	Class A Spectrally Flat (ISO9060: 2018) Class A (IEC 61724-1:2021 Class 1 (WMO Guide No.8)	Scientific quality and highest accuracy

For example, compliance with IEC61724-1 Class A pyranometers with directional response and temperature response test results is equivalent to ISO 9060 Spectrally Flat Class A instruments. In some cases, older instruments from the ISO 9060:1990 standard's [31] highest quality "Secondary standard" (now Class A), may be updated, if a pyranometer lacks the required test reports for the performance verification on temperature dependence and directional response (e.g. [7]).

6.4.1.4 Recommendations for Power Measurements for real-time Wind and Solar Forecasting

In summary, the following are recommended practices for use of power measurements in real-time operation and forecasting of wind and solar power:

1. Live/real-time power measurements should be used in very short-term wind and solar power forecasting, i.e. for lead-times from minutes to hours ahead.
2. Power data should be recorded at the same or higher temporal resolution and frequency than that for the forecasting system.
   - For electricity system operation temporal resolution may be one minute or less, for trading operations this should match the duration of the shortest traded product or settlement period.
   - Forecast update frequency should match the highest frequency with which new information becomes available, within the constraints of IT infrastructure.
3. The power measurement should be taken from the metering point that is most relevant to the forecast user, this will usually be the metering system at the connection point to the electricity network.
4. Live data on plant operating conditions (e.g. control actions the effect maximum output) and availability (e.g. maintenance outages) should be used in forecast production and evaluation.

- Where direct data is not available, this should be inferred using indirect measurements.

5. Data on future plant availability (e.g. maintenance actions) should be used in forecast production. Typically, power forecasts do not include the effect of control actions issued by electricity system operators.

6. Live Power Available signals should be used in forecast production in place of live power production data during periods when output is curtailed due to control actions

6.4.2 Accuracy and Resolution Recommendations

Table 6.4 and 6.5 show the accuracy and resolution recommendation for system operation tasks on grid or utility-scale in the energy area. The limits are partially adapted from EPA's *Guidance on Meteorological Monitoring for Regulatory Modeling Applications*, section 4.5 Sampling Rates and from the *Example Performance Specifications* of 8.1 Instrumentation Procurement [42], the WMO Guide No.8 "Guide to Instruments and Methods of Observation" [66] and the heights from IEC 61400-12-1 [15] and system operators that have gained experience with these and other relevant guidelines and standards adopted from meteorological monitoring for real-time applications.

It is also recommended for the instrumentation accuracy levels or limits to consider the uncertainty of the instrumentation. Section 5.2 provides a number of useful measures how to define and request instrumentation to comply or to provide such uncertainty quantification from the calibration of instruments as part of the standing data.

Table 6.4: Recommendations for system accuracy and measurement resolution for real-time wind forecasting applications in the power grid and utility-scale operation.

Wind Aggregated Generating Facility Meteorological Data Requirements						
Measurement Type	**Units**	**Precision**	**Range**	**Accuracy**	**Height of instrument**	
					Met Mast	**Alternative**
Wind Speed	Meters /Second (m/s)	0.1 m/s	0 to 50	± 1m/s	At Hub Height + lower blade tip height (±10% of rotor diameter)	At 35m Meters + Nacelle or Lidar

Wind Aggregated Generating Facility Meteorological Data Requirements						
Measurement Type	**Units**	**Precision**	**Range**	**Accuracy**	**Height of instrument**	
					Met Mast	**Alternative**
Wind Direction	Degrees from True North	1 degree	0 to 360	± 5°	At Hub Height + lower blade tip height (±10% of rotor diameter)	At 35m
Barometric Pressure	Hecto Pascals (HPa)	1 hPa	800 to 1000	± 1.0 hPa at −20 to 50 °C, and ± 1.5 hPa at below -20° C	at or corerected to hub height	
Ambient Temperature	Degree Celsius (řC)	0.1ř C	-50 to +50	± 0.2°C	within 10m of hub height	
Dewpoint	Degrees Celsius (°C)	0.1ř C	-50 to +50	± 0.2°C	within 10m of hub height	At 2m or 10m
Relative Humidity	Percentage (%)	1.00%	0 to 100 %	± 2%	at hub height	
Ice-up Parameter	Scale 0.0 to1.0	0.1	0 to 1	n/a	within 10m of hub height	At 35m
Precipitation	(mm/min)	0.1	0 to11	2% up to 0.417 mm/mon 3% over 0.417 mm/min	at 2m or 10m	at 2m or 10m

Table 6.5 shows the accuracy recommendation that we recommend for system operation tasks in the energy area. The limits are partially adapted from EPA's *Guidance on Meteorological Monitoring for Regulatory Modeling Applications*, section 4.5 Sampling Rates [42] and system operators that have gained experience with these and other relevant recommendations from the WMO Guide No.8 [66].

Table 6.5: Recommendations of Accuracy and Resolution requirements for real-time forecasts of solar projects

Meteorological Data Requirements for Solar Projects					
Measurement	**Units**	**Precision**	**Range**	**Accuracy**	**Height of instrument**
Wind Speed	Meters /Second (m/s)	0.1 m/s	0 to 50	± 1m/s	Between 2-10 meters
Wind Direction	Degrees from True North	1 degree	0 to 360	± 5°	Between 2-10 meters
Barometric Pressure	Hecto Pascals (HPa)	1 hPa	800 to 1000	± 1.0 hPa at −20 to 50 °C, and ± 1.5 hPa below −20°C	Between 2-10 meters
Ambient Temperature	Degree Celsius (řC)	0.1°C	-50 to +50	± 0.2°C	Between 2-10 meters
Dewpoint	Degrees Celsius (řC)	0.1ř C	-50 to +50	± 0.2°C	Between 2-10 meters
Relative Humidity	Percentage (%)	1.00%	0 to 100 %	± 2%	Between 2-10 meters
Ice-up Parameter	Scale 0.0 to1.0	0.1	0 to 1	n/a	Between 2-10 meters
Precipitation	Millimetres /minute (mm/min)	0.1	0 to11	2% up to 0.417 mm/mon 3% over 0.417 mm/min	Between 2-10 meters
Backpanel Temperature	Degree Celsius (°C)	0.1°C	-50 to +50	± 0.1°C at −27 to +50°C, and ± 0.2°C at below −27°C	Between 2-10 meters
Global Horizontal Irradiance	Watts/Square Metre (W/m2)	0.1	0 to 4000	± 3%	Between 2-10 meters
Diffused Horizontal Irradiance	Watts/Square Metre (W/m2)	0.1	0 to 4000	± 3%	Between 2-10 meters
Direct Normal Irradiance	Watts/Square Metre (W/m2)	0.1	0 to 2000	± 3%	Between 2-10 meters

Measurement	Units	Precision	Range	Accuracy	Height of instrument
Sunshine Duration	V	0.1	0 to 1	90.00%	Between 2-10 meters

6.4.3 Validation and Verification Recommendations

The quality of measured data that is supposed to be used in real-time applications is of immense importance. Modern technologies to assimilate measured data into forecasting models and processes usually have automatic algorithms that blacklist data that is out of range or in other ways suspicious, because bad data is often worse for the forecast model and processes than no data.
To avoid bad quality data to take precedence, quality control and clear assessment criteria for the data are therefore important tools for the success of any application using real-time measurements.

Section 5.4 details both the real-time data handling (section 5.6.1) and the post real-time data quality assessment (section **??**), necessary to keep a high quality data standard for the real-time applications and to make sure that instrument failures are found, reported and fixed.

The following principle are recommended for the quality assurance and control of real-time measurements for forecasting applications in power grid and utility-scale operations:

- High measurement data quality requires continuous maintenance
- In real-time environments, insufficient data quality cannot be compensated for
- Uncertainty in measurements need to be taken into account (**??**
- Quality assurance (QA) and quality control (QC) require a quality assurance project plan (QAPP), which is the management and documentation of QC and QA procedures and should include (see section 5.4):

  (a) Project Description and organisation
  (b) QA procedures and monitoring
  (c) Design of QC procedures

6.4.3.1 Practical Methodology for historic quality control of measurement for wind applications

A practical methodology to apply for the validation and quality control of meteorological measurements is a combination of different consistency checks:

- Missing Values and time stamps
- Forecasted wind speed versus measured wind speed
- Forecasted temperature, wind direction against measured values

- Forecasted power versus active power checked with SCADA MW
- Computed active power from measured wind speed versus actual active power
- Comparison against previous years of the same wind farm
- Comparison to the average error level for wind farms in the same period

6.4.3.2 Data Screening in Real-time Environment

The recommended data screening criteria for wind and solar projects as a practical approach to data quality control and management follow the EPA guideline [42] and are suggested with some limits that are easy to control. Table 6.6 provides some useful and practical limits that can be recommended for the data screening task.

Table 6.6: Practical data screening criteria for data quality control of wind projects

Variable	Screening Criteria: Flag data if the value
Wind Speed	- is less than zero or greater than 50 m/s - does not vary by more than 0.1 m/s for 3 consecutive hours - does not vary by more than 0.5 m/s for 12 consecutive hours
Wind Direction	- is less than zero or greater than 360 degrees - does not vary by more than 1 degree for more than 3 consecutive hours - does not vary by more than 10 degrees for 18 consecutive hours
Temperature	- is greater than the local record high - is less than the local record low (limit could be applied on a monthly basis) - is greater than a 5°C change from the previous hour - does not vary by more than 0.5° for 12 consecutive hours
Temperature Difference	- is greater than 0.1°C/m during the daytime - is less than -0.1°CC/m during the night time - is greater than 5.0°C or less than -3.0°C
Dew Point Temperature	- is greater than the ambient temperature for the given time period - is greater than a 5°C change from the previous hour - does not vary by more than 0.5°C for 12 consecutive hours - equals the ambient temperature for 12 consecutive hours
Pressure	- is greater than 1060 mb (sea level) - is less than 940 mb (sea level) (values should be adjusted for elevations other than sea level) - changes by more than 6 mb in three hours
Radiation	- is greater than zero at night - is greater than the maximum possible for the date and latitude

6.5 Recommendations for real-time measurements for power plant operation and monitoring

The key applications for wind plant operation, where real-time meteorological measurements are required, are:

- **Wind turbine control**
  Due to wake effects on nacelle anemometer, independent site data from a met mast or LIDAR can assist the turbine controller to work more safe and efficient. Preview information of the turbine inflow from a nacelle-based lidar system can be used to induce blade pitch action and thus reduce loads and improve turbine power performance.

- **Wind farm control**
  Wake measurements of scanning lidars or nacelle-based lidars that measure the wake of turbines, can assist to redirect those wakes and thus reduce loads on downstream turbines and increase their power production.

- **Condition Monitoring**
  Knowledge about the wind conditions that affect a wind turbine or wind farm help to estimate the load budget that the turbines have experience during their lifetime. This information can be used for lifetime extension measures thus increasing the energy yield of a wind farm and its profit.

Since measuring wind and other variables at power plant level is not always connected to power forecasting applications, the following recommendations are basic recommendations for high-quality measurements of these variables translated from the available standards. More detailed information can be found in the IEC 61400-12-1 [15], the MEASNET recommendations [41] for wind and the PVPS handbook [56].

It is important to recognise that the value of measured meteorological data at a wind or solar plant has a relevance, not only for the plant operation and the monitoring of it's performance, but also to ensure that the generated electricity can be fed efficiently and environmentally friendly into the power grid.
The principles of ***accuracy*** and ***reliability,*** as described in the key points to this section and the definitions in section 6.1 and section 6.4 for the generation of reliable information about the current weather situation, are all important corner stones for an efficient energy transition. Measurement campaigns are expensive in resources and costly. The benefits of measurements however can only be harvested, if the quality of the measured data is sufficient to be used in modelling, forecasting and verification of performance.
It is for this reason that we recommend any wind farm control and monitoring design process to be considered together and not separate from possible or potential benefits that can help integrate the generated energy best possible into the power grid, by either designing the instrumentation to also be used for grid integration, balancing and/or trading purposes and in that way get higher benefits and lower overall costs.

6.5.1 Quality Recommendations

6.5.1.1 Requirements for Wind Farms

Table 6.7 provides instrument classes for plant operation applications to ensure high-quality measurements. For those applications, where there are two classes provided, the second, lower class should be understood as minimum requirement and the higher class as a requirement, if high quality is required and the respective budgets justifiable.

The classes for wind projects are in detail defined in the IEC 61400 Part 12-1 [15] and explained in Annex D of the standard. Any further detail to the quality of the instrumentation should be referred to this standard.

Table 6.7: Forecast applications and respective recommended requirements for appropriate instrumentation.

Forecast Application	**Requirements for Wind** according to IEC 61400 Part 12-1 [15]	**Requirements for Solar** according to ISO9060: 2018, EC 61724-1:2021 and WMO Guide No.8
Wind/Solar Park control	Class A	Class A
Park / Turbine Monitoring	Class S	Class B or C

Apart from classic met masts at hub height, the so-called iSpin technology, where sonic anemometers are mounted at the tip of the turbine, are an alternative to consider. The instrumentation is described in section 2.1.4. Using new technologies, such as the iSpin technology may or may not lead to better results. It is therefore recommended that pilot projects are carried out on e.g. one turbine or by analysing data from turbines that carry such instrumentation. There are open source data and information available for this technology in [63, 58].

6.5.1.2 Requirements for wind farms using lidars

The recommended technical requirements for lidars is to ensure high-quality data in long-term real-time operation and as alternatives to met mast measurements are:

- measurements must be raw or technical requirements must include maintenance and software updates
- lightning protection and recovery strategy after lightning measurements should be taken at a height appropriate for the wind farm, either at one of preferable at both hub height and around 30m
- instruments must be serviced and maintained by skilled staff
- version control must be maintained for signal processing
- wind characteristics data must be on wind turbine level

- Lidars and Sodars in complex terrain require special consideration

6.5.1.3 Requirements for Solar Plants

The recommended technical requirements for pyranometer for solar plant performance monitoring are provided in table 6.8.

Table 6.8: Pyranometer classes for various requirements.

Requirements	**Instrument class**	**Instrument description**
Measurements for high-level monitoring for large commercial scale PV systems	Class B (ISO9060:2018) Class B (IEC 61724-1:2021) Class 2-3 (WMO Guide No.8)	Good quality
Economic solutions for routine measurements and small PV systems	Class C (ISO9060: 2018) Class C (IEC 61724-1:2021 Class 4-5 (WMO Guide No.8)	Medium quality

6.5.2 Validation and Verification

In the generally recommended validation of instrumentation as defined in the "ISO Guide to the Expression of Uncertainty in Measurements" [20, 6, 61] and its 2 supplements [[59, 60] from the Joint Committee for Guides in Meteorology (JCGM) two types of measurement uncertainty need to be accounted for in any standardised taken measurement:

1. systematic errors, which are also known as measurement bias, often associated with offsets of the measured quantity

2. random errors, which are associated with the fact that 2 measurements of the same quantity are seldom the same

In most cases the validation and verification of instruments and the respective data should be designed for both long-term analysis of e.gå year and shorter periods such as daily, weekly, monthly or quarterly examination of observational data signals. There are a few important targets for the validation and quality control:

1. To identify the amount of valid data submitted

2. To produce a comprehensive analysis which will provide the wind farm owner with a description of the root of the detected error(s) in the signals.

3. To limit the time from when an issue with data signals starts until it is diagnosed and solved

4. To avoid cross correlations between measurement from different sources to fill data gaps

The last item may be cancelled from the list above, if there exists a short distance of the generating unit's locations to another generating unit.

A practical methodology to apply for the validation and quality control of meteorological measurements is a combination of different consistency checks:

- Missing Values and time stamps
- Forecasted wind speed versus measured wind speed
- Forecasted temperature, wind direction against measured values
- Forecasted power versus active power checked with SCADA MW
- Computed active power from measured wind speed versus actual active power
- Comparison against previous years of the same wind farm
- Comparison to the average error level for wind farms in the same period

6.5.2.1 Statistical tests and metrics for the QC process

A practical and recommended solution for plant monitoring is to use verification methods to test measurement signal quality similar to the verification of forecast errors, with the exception that a forecast or a forecast range of the respective variable is used as the reference, because it is the measurement that needs validation. The forecast has a known accuracy level, which is used to find changes in quality in the measurement signals. By using e.g. ensemble forecasts, the uncertainty of the measurement signals can also be quantified (see e.g. [43]).

By validating in different sub periods of the year, it can be shown whether the error pattern has been temporary or on a long-term basis.

By using a variation of different statistical tests as recommended in part 3 of this recommended practice [2], the data basis is large enough to interpret the data accuracy. The following set of statistical metrics are recommended (details can be found in section 5.5.1.3:

1. **BIAS:** The BIAS alone does not provide a guarantee or correctness.
2. **MAE:** MAE and BIAS together show, if the data has an offset.
3. **RMSE:** There are few extreme errors, if the ratio RMSE/MAE exceeds 1.3.
4. **CORRELATION:** The correlation allows for easy detection of constant measurements as well as sign errors.

5. **Frequency distribution:** The frequency distribution from e.gȧ one year data set of 15-min mean values of a wind speed should be a smooth curve with decreasing probability of high wind speeds. A temporary instrument fault will be visible as a skewness of the curve. Comparing the frequency distributions of an ensemble mean forecast against measurements is recommended in this case, as a mean smooths outliers in the data set.

   Positive and negative phase errors between a forecast and measured data tend to cancel each other out over a long enough period. Therefore, a high similarity between two independent time series of the same physical variable can be expected.

The formulas of the test metrics can be found in the Appendix B.
A graphical analysis of measurement signals with these metrics can then also be used to define acceptance limits for the meteorological variables or to see long-term trends. Examples of such graphical analysis can be found in section 5.5.1.3.

6.5.2.2 Solar specific Validation

Specific validation for solar projects can be found in the PVPS handbook [56] in chapter 9.5 *Solar Resource Data for Plant Operations*. In this section, a variety of approaches for monitoring and measuring the performance of an existing solar power plant are described. Here the handbook says that "..the system is directly linked to the meteorological conditions. For flat-plate thermal collectors and PV, the production is roughly proportional to the incident GTI; for concentrating technologies, the incident DNI is the driving input. In all cases, additional meteorological variables need to be monitored because they play a modulating role."

The handbook concludes that real-time monitoring of the meteorological conditions at the systems location are important for the following applications:

(a) Evaluate a performance guarantee (acceptance testing)

(b) Assess the power plants performance to improve yield predictions and to gain knowledge toward improvements in future plants

(c) Identify conditions of poor performance, including evidence of soiling, shading, hardware malfunction, or degradation, which could lead to warranty replacement, etc.

6.5.2.3 Performance Control for hardware and manufacturer production guarantees

Performance control of wind and solar plants and turbine/panels are important for financial aspects of a project, but also for the predictability of these energy sources, whether this is for system integration, balancing or power trading. In this way, the power performance is directly connected to the efficient usage of the generated power.
And, as described in the previous section, it is a direct recommendation to connect quality control to performance monitoring by designing instrumentation setup and control mechanisms

for the performance control to a high standard of data quality as defined and recommended for power grid applications in section 6.4.

The following recommendations are minimum requirements and recommendations, if the data is only used for performance control.

1. ***Wind Power Performance Control:***
   Performance control of wind farms and wind turbines is best conducted in 3-4 steps:
   (a) ***Measuring basic meteorological parameters that can be used to compute a power generation output:***
       - Wind speed and direction
       - air temperature
       - barometric pressure
       - relative humidity

       It is recommended to measure these parameters such that they represent the weather conditions at the wind turbine rotor centre. At least 2 measurement heights per wind farm with the following measuring heights ate recommended:
       (a) hub height or within 10m of hub height
       (b) at approx. 30-35m above ground (typically first model level of NWP models)
       (c) at 10m above ground (standard meteorological measuring height)

       If (a) is not possible, a lower height can be measured and corrected for according to ISO 2355 (see chap. 6 [15]).
   (b) ***Conversion of the meteorological parameters into a power output:***
       The best and recommended way is to follow the IEC 61400-12-1 standard on power performance measurments and apply a physical formula (equ. 2, chapter 8 [15]).
   (c) ***Comparison power output with measured and forecasted input variables:***
       Additionally, it is recommended to use forecasts from an independent forecasting model that uses power generation output over ideally 1 year, but a minimum of 3 months of various wind conditions and compare these results with the derived measured power output. The improvement from measured parameters should be approximately 3-5% when measured over a minimum of 3 months (see e.g. [43, 47].
   (d) ***Visual Inspection with Ensemble generated Percentiles:***
       By using a physical ensemble forecasting system of type b.1 (physical NWP parameterization schemes) [17] or of type "Physical Methods" (sec. 4.2.1. in[8])"Multi-Scheme Ensemble" (section 3.2.1.4 and Table 3.1 in [44]), for example, 10 percentiles can be computed and compared with the computed and measured power output. Changes in performance can in that way easily be seen by visual inspection. This is a straight forward methodology to avoid surprises and be able to prove, with a physical methodology, hardware malfunctioning.

2. ***Solar Power Performance control***[1]
   If monitoring of a PV plant's output comprises continuous comparisons to the expected output based on actual meteorological conditions leads to 2 main benefits:

   - PV arrays can be cleaned as a function of meteorological conditions
   - Errors and malfunctions of equipment can be detected more quickly

   The essential parameters for CST (concentrating solar technologies) plants are [56]:

   (a) Real-time DNI
   (b) Wind
   (c) Temperature

   The best choice and recommendations for instrumentation are[56]:

   - **Maintenance issues**: a well-maintained reference cell in the POA [2]
   - **Performance assessment**: well-maintained, and regularly cleaned POA pyranometer

   Visual Inspection with Ensemble generated Percentiles:
   By using a physical ensemble forecasting system of type b.1 (physical NWP parameterization schemes) [17] or of type "Physical Methods" (sec. 4.2.1. in[8]) or "Multi-Scheme Ensemble" (section 3.2.1.4 and Table 3.1 in [44]), for example, 10 percentiles can be computed and compared with the computed and measured power output. Changes in performance can in that way easily be seen by visual inspection. This is a straight forward methodology to avoid surprises and be able to prove, with a physical methodology, hardware malfunctioning.

6.6 Recommendations for real-time measurements for power trading in electricity markets

Recommendations for real-time measurements for power trading and balancing in electricity markets are very much dependent on the type of trading that is performed. For day-ahead trading, historic measurements are sufficient, and hence also a cleaning strategy to clean out bad data similar to resource assessments, described in the IEC [14, 15] and covered in section 6.4.1.
The following recommendations therefore refer to intra-day trading or teading in rolling markets, where real-time measurements are of immense importance.
To define the need and background for the measurement data quality, the first step is to understand the possibilities trading strategies with the use of real-time measurements provide.

[1] see details in the PVPS Handbook chapter 9.5.2 [56]
[2] Plane of array, referring to *global tiltet irradiance* (GTI) measurements.

6.6.1 Trading Strategies with real-time Measurements

In [44] it is reported that "..frequent short-term forecasts based on online data add value not only because they are more accurate than the day-ahead horizon.", but also because it enables the trader to trade intelligently within the physical uncertainties of both forecasts and measurements (see also section 5.2).
Nevertheless, trading on intra-day or rolling markets with short-term forecasts that have been adapted with real-time measurements should take place with care, because they also increase the trading volume and thereby the loss compared to the spot market price. This situation can occur, if the short-term forecast errors have opposite sign of the day-ahead error, which statistically occurs for up to 50% likelihood, if the error of the day-ahead is small[44].

In most power exchanges there still exists a significant loss on wind and solar power traidng on the intra-day or rolling markets. This loss is often associated with the higher expenses when buying short-term power for balancing purposes. Selling surplus wind or solar power generated by a BIAS-free day-ahead forecast is still more beneficial.
However, investigations of this pattern have shown that an efficient trading system should neither deal with imbalances that are traded multiple times, nor that one and the same megawatt (MW) is charged several times with reserve costs [46, 45, 44]. This can be accomplished by using an approach that uses uncertainty factors, illustrated in Figure 6.1, and that limits the trading of imbalances according to the expected volumes that lie outside a pre-calculated and with short-term forecasts updated uncertainty bands around a day-ahead forecast as demonstrated in [46] and explained in more detail in section *3.8.2 The Balancing Challenges on the Intra-day Horizon* [44].

6.6.2 Quality Recommendations

The quality recommendations follow to a large extent the recommendations for power grid in section 6.4 as the ideal solution, and the instrument classes described in section 4.1 and defined in Table 6.9 for electricity market purposes.

Table 6.9: Forecast applications and respective recommended requirements for appropriate instrumentation.

Forecast Application	**Requirements for Wind** according to IEC 61400 Part 12-1 [15]	**Requirements for Solar** according to ISO9060: 2018, EC 61724-1:2021 and WMO Guide No.8
Electricity Market Trading	Class B	Class B

Table 6.10 provide the recommended instrument classes for solar projects.

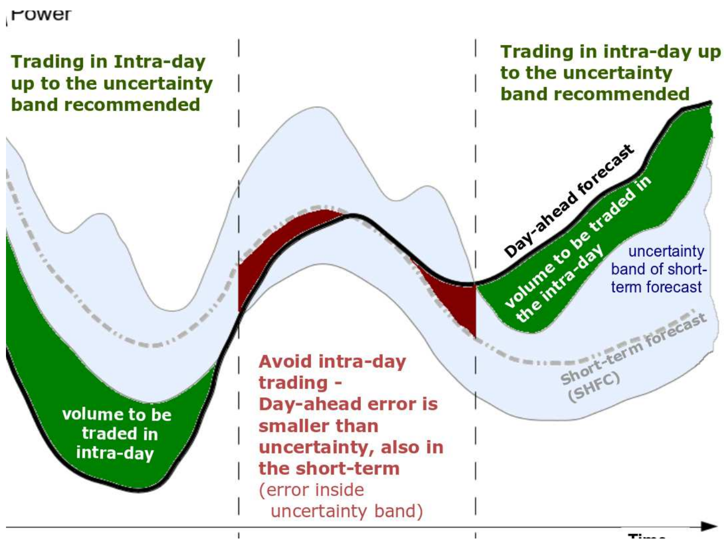


Figure 6.1: Trading principle when the uncertainty band is used for the determination of the volume that is to be traded in an intra-day or rolling market. The dashed grey line is the short-term forecast (SFC), the black line is the day ahead forecast (DFC) and the grey lines are the uncertainty forecast with the upper and lower limit ©Möhrlen et al. [44].

6.6.3 Accuracy and Resolution requirements

It is important to recognise that the value of measured meteorological data at wind or solar plants has a relevance, not only for trading the produced energy from wind and solar plants. It is equally important to also ensure that the generated electricity can be fed efficiently and environmentally friendly into the power grid and that imbalances are avoided, where possible, and not become subject to speculations for pure financial purposes.

The principles of ***accuracy*** and ***reliability,*** as described in the key points to this section and the definitions in section 6.1 and section 6.4 for the generation of reliable information about the current weather situation, are all important corner stones for an efficient energy transition. Measurement campaigns are expensive in resources and costly. The benefits of measurements however can only be harvested, if the quality of the measured data is sufficient to be used in modelling, forecasting and validation of generation performance.

It is for this reason that we recommend also for trading applications that, if this is possible, the design process for real-time measurement setup is not separate from possible or potential coordination with plant operations, monitoring and grid integration requirements. Only, if measurement strategies are aligned, it is possible that all 4 tiers can benefit from the measurements and share the costs. When making use of these synergies, the integration of the generated energy into the power grid will be cheaper and more efficient overall.

Table 6.10: Pyranometer classes for various requirements.

Requirements	**Instrument class**	**Instrument description**
Measurements for high-level monitoring for large commercial scale PV systems	Class B (ISO9060:2018) Class B (IEC 61724-1:2021) Class 2-3 (WMO Guide No.8)	Good quality
Economic solutions for routine measurements and small PV systems	Class C (ISO9060: 2018) Class C (IEC 61724-1:2021 Class 4-5 (WMO Guide No.8)	Medium quality

Bibliography

[1] World Meteorological Organisation (WMO). *Part I – Siting and exposure of meteorological instruments*. Tech. rep. 55. Tech. Rep. WMO/TD- No. 589, IOM Report, 2018.

[2] IEA Wind Task 36. *IEA Wind Task 36 Recommended Practice on Renewable Energy Forecast Solution Selection*. Tech. rep. International Energy Agency Technical Cooperation Programm Wind, 2019. URL: `https://iea-wind.org/task-36/task-36-publications/recommended-practice/`.

[3] Clifton A. et al. *Remote Sensing of Complex Flows by Doppler Wind Lidar: Issues and Preliminary Recommendations*. National Renewable Energy Laboratory, Dec. 2015. URL: `https://www.nrel.gov/docs/fy16osti/64634.pdf`.

[4] Romo Wind A/S. *Performance Transparency Project (PTP)*. Tech. rep. 2021. URL: `https://www.ispin-ptp.com/`.

[5] Axel Albers et al. "Ground-based remote sensor uncertainty – a case study for a wind lidar". In: (Jan. 2013).

[6] *An introduction to the 'Guide to the expression of uncertainty in measurement and related documents*. Version JCGM 104:2009. Joint Committee for Guides in Metrology, Evaluation of measurement data, 2009. URL: `https://en.wikipedia.org/wiki/Joint_Committee_for_Guides_in_Metrology`.

[7] Hukseflux Thermal Sensors B.V. *PV monitoring and meteorological industries prepare for revised pyranometer standard ISO 9060:2018*. Tech. rep. 2018. URL: `https://www.hukseflux.com/uploads/inline/how_to_prepare_for_the_revised_pyranometer_standard_iso_9060-2018_v1812.pdf` (visited on 09/29/2021).

[8] R. J. Bessa et al. "'Good or 'bad wind power forecasts: a relative concept". In: *Wind Energy* 14.5 (2010), pp. 625–636. DOI: `10.1002/we.444`.

[9] Alfred K Blackadar. "Boundary layer wind maxima and their significance for the growth of nocturnal inversions". In: *Bulletin of the American Meteorological Society* 38.5 (1957), pp. 283–290.

[10] Stuart Bradley, Alexander Strehz, and Stefan Emeis. "Remote sensing winds in complex terrain – a review". In: *Meteorologische Zeitschrift* (Nov. 5, 2015). Publisher: Schweizerbartche Verlagsbuchhandlung, pp. 547–555. DOI: 10.1127/metz/2015/0640. URL: https://www.schweizerbart.de/papers/metz/detail/24/84734/Remote_sensing_winds_in_complex_terrain__a_review?l=FR (visited on 09/20/2021).

[11] Stuart Bradley et al. "Corrections for Wind-Speed Errors from Sodar and Lidar in Complex Terrain". In: *Boundary-Layer Meteorology* 143.1 (Apr. 1, 2012), pp. 37–48. ISSN: 1573-1472. DOI: 10.1007/s10546-012-9702-0. URL: https://doi.org/10.1007/s10546-012-9702-0 (visited on 09/20/2021).

[12] Andrew Clifton et al. *A Review of Guidance for Using Ground-Based Vertically-Profiling Wind Lidar For Wind Resource Assessment*. Zenodo, Aug. 7, 2020. DOI: 10.5281/zenodo.3862384. URL: https://zenodo.org/record/3862384 (visited on 08/23/2021).

[13] Andrew Clifton et al. "IEA Wind Task 32: Wind Lidar Identifying and Mitigating Barriers to the Adoption of Wind Lidar". In: *Remote Sensing* 10.3 (2018). ISSN: 2072-4292. DOI: 10.3390/rs10030406. URL: https://www.mdpi.com/2072-4292/10/3/406.

[14] International Electrotechnical Commission. *IEC Standard 61400-12-1:2005 Power performance measurements of electricity producing wind turbines*. 2005.

[15] International Electrotechnical Commission. *IEC Standard 61400-12-1:2017 Power performance measurements of electricity producing wind turbines*. 2017.

[16] International Electrotechnical Commission. *IEC Standard 61724-1:2021 Photovoltaic system performance - Part 1: Monitoring*. 2021.

[17] Jan Dobschinski et al. "Uncertainty Forecasting in a Nutshell: Prediction Models Designed to Prevent Significant Errors". In: *IEEE Power and Energy Magazine* 15.6 (2017), pp. 40–49. DOI: 10.1109/MPE.2017.2729100.

[18] Susanne Drechsel et al. "Wind Speeds at Heights Crucial for Wind Energy: Measurements and Verification of Forecasts". In: *Journal of Applied Meteorology and Climatology* 51.9 (2012), pp. 1602–1617. DOI: 10.1175/JAMC-D-11-0247.1. URL: https://journals.ametsoc.org/view/journals/apme/51/9/jamc-d-11-0247.1.xml.

[19] Stefan Emeis, Michael Harris, and Robert M. Banta. "Boundary-layer anemometry by optical remote sensing for wind energy applications". In: *Meteorologische Zeitschrift* 16.4 (Aug. 2007). Place: Stuttgart, Germany Publisher: Schweizerbart Science Publishers, pp. 337–347. DOI: 10.1127/0941-2948/2007/0225. URL: http://dx.doi.org/10.1127/0941-2948/2007/0225.

[20] *Evaluation of measurement data – Guide to the expression of uncertainty in measurement*. Version JCGM 100:2008. Joint Committee for Guides in Metrology, Evaluation of measurement data, 2008.

[21] *EWeLINE project: Development of innovative weather and power forecast models for the grid integration of weather dependent energy sources.* 2011. URL: http://www.projekt-eweline.de/en/project.html.

[22] Lars Falbe-Hansen. *Review of the Spinner anemometer from ROMO Wind iSpin*. Tech. rep. DNV GL, 2015. URL: https://cdn2.hubspot.net/hubfs/2828570/RomoWind%5C%20August2018/Pdf/113605-DKAR-R-01-Rev-3-2015-03-06.pdf.

[23] Troels Friis Pedersen and Giorgio Demurtas. *Calibration procedure for spinner anemometer yaw error measurements*. English. DTU Wind Energy I 0082(EN). This is an internal report and therefore not available in full text. Please contact the author or the director of author institute for further information. Denmark: DTU Wind Energy, 2013.

[24] Troels Friis Pedersen and Paula Gómez Arranz. *Spinner anemometer – best practice. Version 2*. English. DTU Wind Energy E 165. Denmark: DTU Wind Energy, 2018.

[25] Ammonit GmbH. *Solar Measurement Knowledge*. URL: https://www.ammonit.com/en/customer-support/knowledge/solar-measurement-knowledge/.

[26] Dominique Philipp Held and Jakob Mann. "Detection of wakes in the inflow of turbines using nacelle lidars". English. In: *Wind Energy Science* 4.3 (2019), pp. 407–420. ISSN: 2366-7443. DOI: 10.5194/wes-4-407-2019.

[27] Sepp Hochreiter and Jürgen Schmidhuber. "LSTM can solve hard long time lag problems". In: *Advances in neural information processing systems*. 1997, pp. 473–479.

[28] Martin HofsäSS, Andrew Clifton, and Po Wen Cheng. "Reducing the Uncertainty of Lidar Measurements in Complex Terrain Using a Linear Model Approach". In: *Remote Sensing* 10.9 (2018). ISSN: 2072-4292. DOI: 10.3390/rs10091465. URL: https://www.mdpi.com/2072-4292/10/9/1465.

[29] Harald Hohlen. "PERFORMANCE MONITORING USING SPINNER ANEMOMETRY". In: Bremen, Germany, Oct. 17, 2017. URL: https://www.ispin-ptp.com/wp-content/uploads/2017/08/DEWEK2017_ConferenceProceeding_056_PerformanceMonitoringUsingSpinnerAnemometry.pdf.

[30] Harald Hohlen and Markus Rühlmann. "ACHIEVING PERFORMANCE TRANSPARENCY USING SPINNER ANEMOMETRY". In: Amsterdam, The Netherlands, Nov. 28, 2017. URL: https://www.ispin-ptp.com/wp-content/uploads/2017/08/WindEurope2017_ConferenceProceeding_P156_AchievingPerformanceTransparencyUsin pdf.

[31] *ISO 9060:1990 Solar energy – Specification and classification of instruments for measuring hemispherical solar and direct solar radiation*. Version 2. International Organization for Standardization, ISO/TC 180/SC 1, Climate - Measurement and data, 1990.

[32] *ISO 9060:2018 Solar energy Specification and classification of instruments for measuring hemispherical solar and direct solar radiation*. Version 2. International Organization for Standardization, ISO/TC 180/SC 1, Climate - Measurement and data, 2018.

[33] Dahlberg J.-Å., Pedersen T.F., and Busche P. *ACCUWIND – Methods for Classification of Cup Anemometers*. Version Risø-R-1555(EN). DTU Wind Energy, May 2006.

[34] Bo Jing et al. "Data-Dirven Method for Wake Effect Analysis on Nacelle Anemometer". In: *IOP Conference Series: Earth and Environmental Science* 555 (Aug. 2020), p. 012117. DOI: 10.1088/1755-1315/555/1/012117. URL: https://doi.org/10.1088/1755-1315/555/1/012117.

[35] N Kelley et al. *Lamar low-level jet program interim report*. Tech. rep. National Renewable Energy Lab., Golden, CO.(US), 2004.

[36] E. E. Lucio-Eceiza et al. "QQuality Control of Surface Wind Observations in Northeastern North America. Part I: Data Management Issues". In: *Journal of Atmospheric and Oceanic Technology* 35(1) (2018). DOI: https://journals.ametsoc.org/view/journals/atot/35/1/jtech-d-16-0205.1.xml.

[37] E. E. Lucio-Eceiza et al. "Quality Control of Surface Wind Observations in Northeastern North America. Part II: Measurement Errors". In: *Journal of Atmospheric and Oceanic Technology* 35(1) (2018). DOI: https://journals.ametsoc.org/view/journals/atot/35/1/jtech-d-16-0205.1.xml.

[38] J. Mann et al. "Complex terrain experiments in the New European Wind Atlas". In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 375.2091 (Apr. 13, 2017). Publisher: Royal Society, p. 20160101. DOI: 10.1098/rsta.2016.0101. URL: https://royalsocietypublishing.org/doi/full/10.1098/rsta.2016.0101 (visited on 09/20/2021).

[39] M. Marquis et al. *Wind Forecasting Improvement Project (WFIP) final report*. Tech. rep. National centre of Oceanographic and Atmospheric Administration (NOAA), May 15, 2014. URL: http://energy.gov/sites/prod/files/2014/05/f15/wfipandnoaafinalreport.pdf.

[40] C. Matzler. *Thermal Microwave Radiation: Applications for Remote Sensing*. Vol. Chapter 1. The Institution of Engineering and Technology, London, 2006.

[41] MEASNET. *Evaluation of site-specifc wind conditions V2.0. Procedure*. MEASNET, Apr. 2016. URL: http://www.measnet.com/wp-content/uploads/2016/05/MeasnetSiteAssessment%20V2.0.pdf.

[42] *Meteorological Monitoring Guidance for Regulatory modelling Applications*. Doc. EPA-454/R-99-005. 2000. URL: https://www3.epa.gov/scram001/guidance/met/mmgrma.pdf.

[43] C. Möhrlen et al. "EIRGRID Met Mast and Alternatives Study". In: *IET Renew. Power Gener.* (2021). Submitted Aug. 2021.

[44] Corinna Möhrlen and J.U. Jørgensen. "Integration of Large-Scale Renewable Energy into Bulk Power Systems. From Planning to Operation". In: ed. by Aidan Tuohy Pengwei Du Ross Baldick. Springer International Publishing AG, Cham, Switzerland: Springer Nature, 2008. Chap. The role of ensemble forecasting in integrating renewables into power systems: From theory to real-time applications, pp. 79–134. ISBN: Hardcover 978-3-319-55579-9., eBook 978-3-319-55581-2. DOI: 10.1007/978-3-319-55581-2.

[45] Corinna Möhrlen, Markus Pahlow, and Jess U. Jørgensen. *Author English Translation of (Untersuchung verschiedener Handelsstrategien für Wind- und Solarenergie unter Berücksichtigung der EEG 2012 Novellierung / Investigation of various trading strategies for wind and solar power developed for the new EEG 2012 rules*. URL: http://http://download.weprog.com/WEPROG_Trading_strategies_EEG2012_ZEFE_71-2012-01_en.pdf.

[46] Corinna Möhrlen, Markus Pahlow, and Jess U. Jørgensen. "Untersuchung verschiedener Handelsstrategien für Wind- und Solarenergie unter Berücksichtigung der EEG 2012 Novellierung". In: *Zeitschrift für Energiewirtschaft* 36.1 (Mar. 2012), pp. 9–25. ISSN: 1866–2765. DOI: 10.1007/s12398-011-0071-z. URL: https://doi.org/10.1007/s12398-011-0071-z.

[47] Corinna Möhrlen and Ulrik Vestergaard. *EIRGRID Met Mast and Alternatives Study*. Tech. rep. Version 2. 2019. URL: http://www.eirgridgroup.com/site-files/library/EirGrid/EIRGRID-Met-Mast-and-Alternatives-Study-Version-2.pdf.

[48] S. Monserrat and A. J. Thorpe. "Gravity-Wave Observations Using an Array of Microbarographs In the Alearic Islands". In: *Q.J.R. Meteorol. Soc.* 118 (1992), pp. 259–282. DOI: 10.1002/qj.49711850405.

[49] VR Morris. *Ceilometer Instrument Handbook, DOE/SC-ARM-TR-020*. 2016. URL: http://www.arm.gov/publications/tech_reports/handbooks/ceil_handbook.pdf.

[50] A. Peña, J. Mann, and N. Dimitrov. "Turbulence characterization from a forward-looking nacelle lidar". In: *Wind Energy Science* 2.1 (2017), pp. 133–152. DOI: 10.5194/wes-2-133-2017. URL: https://wes.copernicus.org/articles/2/133/2017/.

[51] Pierre Pinson and Renate Hagedorn. "Verification of the ECMWF ensemble forecasts of wind speed against analyses and observations". In: *Meteorological Applications* 19.4 (2012), pp. 484–500. DOI: https://doi.org/10.1002/met.283. eprint: https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/met.283. URL: https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/met.283.

[52] Christelle Rigollier, Mireille Lefèvre, and Lucien Wald. "The method Heliosat-2 for deriving shortwave solar radiation from satellite images". In: *Solar Energy* 77.2 (2004), pp. 159–169. URL: https://hal.archives-ouvertes.fr/hal-00361364/document.

[53] L. Schröder et al. "PV-Kürzestfristvorhersage mit Satellitendaten und Wolkenkamera". 5. Fachtagung Energiemeteorologie, 5.–7. Juni 2018, Goslar, Germany. 2018. URL: https://uol.de/physik/forschung/ehf/energiemeteorologie/veranstaltungen/fachtagung-energiemeteorologie.

[54] M. Sengupta et al. *Best Practices Handbook for the Collection and Use of Solar Resource Data for Solar Energy Applications: First Edition*. Tech. rep. NREL/TP-5D00-63112. National Renewable Energy Laboratory, Golden, Colorado, USA, 2015. URL: %5Curl%7Bhttps://www.nrel.gov/docs/fy15osti/63112.pdf%7D.

[55] Manajit Sengupta et al. *Best Practices Handbook for the Collection and Use of Solar Resource Data for Solar Energy Applications: Second Edition*. Tech. rep. NREL/TP-5D00-68886. National Renewable Energy Laboratory, Golden, Colorado, USA, 2017.

[56] Manajit Sengupta et al. *Best Practices Handbook for the Collection and Use of Solar Resource Data for Solar Energy Applications: Third Edition*. Tech. rep. National Renewable Energy Laboratory, Golden, Colorado, USA, IEA PVPS, 2021. URL: https://iea-pvps.org/key-topics/best-practices-handbook-for-the-collection-and-use-of-solar-resource-data-for-solar-energy-applications-third-edition/.

[57] DA Smith et al. "Wind lidar evaluation at the Danish wind test site in Høvsøre". In: *Wind Energy* 9 (2006), pp. 87–93.

[58] *Spinner Anemometry – Uncertainty Analysis*. Tech. rep. Version E-0166. DTU Wind Energy, 2018. URL: https://www.ispin-ptp.com/wp-content/uploads/2018/04/DTU-E-0165-Spinner-Anemometry-Best-Practice.pdf.

[59] *Supplement 1 to the "Guide to the expression of uncertainty in measurement – Propagation of distributions using a Monte Carlo method*. Version JCGM 101:2008. Joint Committee for Guides in Metrology, Evaluation of measurement data, 2008.

[60] *The role of measurement uncertainty in conformity assessment*. Version JCGM 106:2011. Joint Committee for Guides in Metrology, Evaluation of measurement data, 2011.

[61] *The role of measurement uncertainty in conformity assessment*. Version JCGM 106:2012. Joint Committee for Guides in Metrology, Evaluation of measurement data, 2012.

[62] F. T. Ulaby, R. K. Moore, and A. K. Fung. *Microwave Remote SensingActive and Passive*. Vol. Volume I: Microwave Remote Sensing Fundamentals and Radiometry. Reading, Massachusetts: Addison-Wesley, 1981.

[63] J. W. Wagenaar et al. "Evaluation of the ROMO Wind iSpin Guardian approach". In: (2016). Publisher: ECN. URL: https://repository.tno.nl/islandora/object/uuid%5C%3Abc429508-83c4-4eae-9078-f76510d1b9d9 (visited on 09/24/2021).

[64] James Wilczak et al. "THE WIND FORECAST IMPROVEMENT PROJECT (WFIP): A Public–Private Partnership Addressing Wind Energy Forecast Needs". In: *Bulletin of the American Meteorological Society* 96.10 (2015), pp. 1699–1718. ISSN: 00030007, 15200477. URL: http://www.jstor.org/stable/26224979.

[65] IEA Wind. *Ground-Based Vertically-Profling Remote Sensing For Wind Resource Assessment. Recommended Practice RP 15*. International Energy Agency TCP Wind, Jan. 2013. URL: https://github.com/IEA-Wind-Task-32/RP15-Ground%20-%20Based%20-%20Remote%20-%20Sensing%20-%20for%20-%20Wind%20-%20Resource%20-%20Assessment/releases/tag/1.0.

[66] WMO. *WMO Guide to Meteorological Instruments and Methods of Observation*. Tech. rep. Tech. Rep. WMO, 2018.

[67] Ines Würth et al. "How far do we see? Analysis of the measurement range of long-range lidar data for wind power forecasting". In: (2017). Accepted: 2018-03-15T17:48:48Z. DOI: 10.18419/opus-9693. URL: http://elib.uni-stuttgart.de/handle/11682/9710 (visited on 05/09/2021).

[68] Ines Würth et al. "Minute-Scale Forecasting of Wind PowerResults from the Collaborative Workshop of IEA Wind Task 32 and 36". In: *Energies* 12.4 (2019). ISSN: 1996-1073. DOI: 10.3390/en12040712. URL: https://www.mdpi.com/1996-1073/12/4/712.

[69] Frederik Zahle and Niels N. Sørensen. "Characterization of the unsteady flow in the nacelle region of a modern wind turbine". In: *Wind Energy* 14.2 (2011), pp. 271–283. DOI: https://doi.org/10.1002/we.418. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/we.418. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/we.418.

[70] Shiyuan Zhong, Jerome D Fast, and Xindi Bian. "A case study of the Great Plains low-level jet using wind profiler network data and a high-resolution mesoscale model". In: *Monthly Weather Review* 124.5 (1996), pp. 785–806.

Bibliography

[1] World Meteorological Organisation (WMO). *Part I – Siting and exposure of meteorological instruments*. Tech. rep. 55. Tech. Rep. WMO/TD- No. 589, IOM Report, 2018.

[2] IEA Wind Task 36. *IEA Wind Task 36 Recommended Practice on Renewable Energy Forecast Solution Selection*. Tech. rep. International Energy Agency Technical Cooperation Programm Wind, 2019. URL: `https://iea-wind.org/task-36/task-36-publications/recommended-practice/`.

[3] Clifton A. et al. *Remote Sensing of Complex Flows by Doppler Wind Lidar: Issues and Preliminary Recommendations*. National Renewable Energy Laboratory, Dec. 2015. URL: `https://www.nrel.gov/docs/fy16osti/64634.pdf`.

[4] Romo Wind A/S. *Performance Transparency Project (PTP)*. Tech. rep. 2021. URL: `https://www.ispin-ptp.com/`.

[5] Axel Albers et al. "Ground-based remote sensor uncertainty – a case study for a wind lidar". In: (Jan. 2013).

[6] *An introduction to the 'Guide to the expression of uncertainty in measurement and related documents*. Version JCGM 104:2009. Joint Committee for Guides in Metrology, Evaluation of measurement data, 2009. URL: `https://en.wikipedia.org/wiki/Joint_Committee_for_Guides_in_Metrology`.

[7] Hukseflux Thermal Sensors B.V. *PV monitoring and meteorological industries prepare for revised pyranometer standard ISO 9060:2018*. Tech. rep. 2018. URL: `https://www.hukseflux.com/uploads/inline/how_to_prepare_for_the_revised_pyranometer_standard_iso_9060-2018_v1812.pdf` (visited on 09/29/2021).

[8] R. J. Bessa et al. "'Good or 'bad wind power forecasts: a relative concept". In: *Wind Energy* 14.5 (2010), pp. 625–636. DOI: `10.1002/we.444`.

[9] Alfred K Blackadar. "Boundary layer wind maxima and their significance for the growth of nocturnal inversions". In: *Bulletin of the American Meteorological Society* 38.5 (1957), pp. 283–290.

[10] Stuart Bradley, Alexander Strehz, and Stefan Emeis. "Remote sensing winds in complex terrain – a review". In: *Meteorologische Zeitschrift* (Nov. 5, 2015). Publisher: Schweizerbartche Verlagsbuchhandlung, pp. 547–555. DOI: 10.1127/metz/2015/0640. URL: https://www.schweizerbart.de/papers/metz/detail/24/84734/Remote_sensing_winds_in_complex_terrain__a_review?l=FR (visited on 09/20/2021).

[11] Stuart Bradley et al. "Corrections for Wind-Speed Errors from Sodar and Lidar in Complex Terrain". In: *Boundary-Layer Meteorology* 143.1 (Apr. 1, 2012), pp. 37–48. ISSN: 1573-1472. DOI: 10.1007/s10546-012-9702-0. URL: https://doi.org/10.1007/s10546-012-9702-0 (visited on 09/20/2021).

[12] Andrew Clifton et al. *A Review of Guidance for Using Ground-Based Vertically-Profiling Wind Lidar For Wind Resource Assessment*. Zenodo, Aug. 7, 2020. DOI: 10.5281/zenodo.3862384. URL: https://zenodo.org/record/3862384 (visited on 08/23/2021).

[13] Andrew Clifton et al. "IEA Wind Task 32: Wind Lidar Identifying and Mitigating Barriers to the Adoption of Wind Lidar". In: *Remote Sensing* 10.3 (2018). ISSN: 2072-4292. DOI: 10.3390/rs10030406. URL: https://www.mdpi.com/2072-4292/10/3/406.

[14] International Electrotechnical Commission. *IEC Standard 61400-12-1:2005 Power performance measurements of electricity producing wind turbines*. 2005.

[15] International Electrotechnical Commission. *IEC Standard 61400-12-1:2017 Power performance measurements of electricity producing wind turbines*. 2017.

[16] International Electrotechnical Commission. *IEC Standard 61724-1:2021 Photovoltaic system performance - Part 1: Monitoring*. 2021.

[17] Jan Dobschinski et al. "Uncertainty Forecasting in a Nutshell: Prediction Models Designed to Prevent Significant Errors". In: *IEEE Power and Energy Magazine* 15.6 (2017), pp. 40–49. DOI: 10.1109/MPE.2017.2729100.

[18] Susanne Drechsel et al. "Wind Speeds at Heights Crucial for Wind Energy: Measurements and Verification of Forecasts". In: *Journal of Applied Meteorology and Climatology* 51.9 (2012), pp. 1602–1617. DOI: 10.1175/JAMC-D-11-0247.1. URL: https://journals.ametsoc.org/view/journals/apme/51/9/jamc-d-11-0247.1.xml.

[19] Stefan Emeis, Michael Harris, and Robert M. Banta. "Boundary-layer anemometry by optical remote sensing for wind energy applications". In: *Meteorologische Zeitschrift* 16.4 (Aug. 2007). Place: Stuttgart, Germany Publisher: Schweizerbart Science Publishers, pp. 337–347. DOI: 10.1127/0941-2948/2007/0225. URL: http://dx.doi.org/10.1127/0941-2948/2007/0225.

[20] *Evaluation of measurement data – Guide to the expression of uncertainty in measurement*. Version JCGM 100:2008. Joint Committee for Guides in Metrology, Evaluation of measurement data, 2008.

[21] *EWeLINE project: Development of innovative weather and power forecast models for the grid integration of weather dependent energy sources.* 2011. URL: http://www.projekt-eweline.de/en/project.html.

[22] Lars Falbe-Hansen. *Review of the Spinner anemometer from ROMO Wind iSpin*. Tech. rep. DNV GL, 2015. URL: https://cdn2.hubspot.net/hubfs/2828570/RomoWind%5C%20August2018/Pdf/113605-DKAR-R-01-Rev-3-2015-03-06.pdf.

[23] Troels Friis Pedersen and Giorgio Demurtas. *Calibration procedure for spinner anemometer yaw error measurements*. English. DTU Wind Energy I 0082(EN). This is an internal report and therefore not available in full text. Please contact the author or the director of author institute for further information. Denmark: DTU Wind Energy, 2013.

[24] Troels Friis Pedersen and Paula Gómez Arranz. *Spinner anemometer – best practice. Version 2*. English. DTU Wind Energy E 165. Denmark: DTU Wind Energy, 2018.

[25] Ammonit GmbH. *Solar Measurement Knowledge*. URL: https://www.ammonit.com/en/customer-support/knowledge/solar-measurement-knowledge/.

[26] Dominique Philipp Held and Jakob Mann. "Detection of wakes in the inflow of turbines using nacelle lidars". English. In: *Wind Energy Science* 4.3 (2019), pp. 407–420. ISSN: 2366-7443. DOI: 10.5194/wes-4-407-2019.

[27] Sepp Hochreiter and Jürgen Schmidhuber. "LSTM can solve hard long time lag problems". In: *Advances in neural information processing systems*. 1997, pp. 473–479.

[28] Martin HofsäSS, Andrew Clifton, and Po Wen Cheng. "Reducing the Uncertainty of Lidar Measurements in Complex Terrain Using a Linear Model Approach". In: *Remote Sensing* 10.9 (2018). ISSN: 2072-4292. DOI: 10.3390/rs10091465. URL: https://www.mdpi.com/2072-4292/10/9/1465.

[29] Harald Hohlen. "PERFORMANCE MONITORING USING SPINNER ANEMOMETRY". In: Bremen, Germany, Oct. 17, 2017. URL: https://www.ispin-ptp.com/wp-content/uploads/2017/08/DEWEK2017_ConferenceProceeding_056_PerformanceMonitoringUsingSpinnerAnemometry.pdf.

[30] Harald Hohlen and Markus Rühlmann. "ACHIEVING PERFORMANCE TRANSPARENCY USING SPINNER ANEMOMETRY". In: Amsterdam, The Netherlands, Nov. 28, 2017. URL: https://www.ispin-ptp.com/wp-content/uploads/2017/08/WindEurope2017_ConferenceProceeding_P156_AchievingPerformanceTransparencyUsin pdf.

[31] *ISO 9060:1990 Solar energy – Specification and classification of instruments for measuring hemispherical solar and direct solar radiation*. Version 2. International Organization for Standardization, ISO/TC 180/SC 1, Climate - Measurement and data, 1990.

[32] *ISO 9060:2018 Solar energy Specification and classification of instruments for measuring hemispherical solar and direct solar radiation*. Version 2. International Organization for Standardization, ISO/TC 180/SC 1, Climate - Measurement and data, 2018.

[33] Dahlberg J.-Å., Pedersen T.F., and Busche P. *ACCUWIND – Methods for Classification of Cup Anemometers*. Version Risø-R-1555(EN). DTU Wind Energy, May 2006.

[34] Bo Jing et al. "Data-Dirven Method for Wake Effect Analysis on Nacelle Anemometer". In: *IOP Conference Series: Earth and Environmental Science* 555 (Aug. 2020), p. 012117. DOI: 10.1088/1755-1315/555/1/012117. URL: https://doi.org/10.1088/1755-1315/555/1/012117.

[35] N Kelley et al. *Lamar low-level jet program interim report*. Tech. rep. National Renewable Energy Lab., Golden, CO.(US), 2004.

[36] E. E. Lucio-Eceiza et al. "QQuality Control of Surface Wind Observations in Northeastern North America. Part I: Data Management Issues". In: *Journal of Atmospheric and Oceanic Technology* 35(1) (2018). DOI: https://journals.ametsoc.org/view/journals/atot/35/1/jtech-d-16-0205.1.xml.

[37] E. E. Lucio-Eceiza et al. "Quality Control of Surface Wind Observations in Northeastern North America. Part II: Measurement Errors". In: *Journal of Atmospheric and Oceanic Technology* 35(1) (2018). DOI: https://journals.ametsoc.org/view/journals/atot/35/1/jtech-d-16-0205.1.xml.

[38] J. Mann et al. "Complex terrain experiments in the New European Wind Atlas". In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 375.2091 (Apr. 13, 2017). Publisher: Royal Society, p. 20160101. DOI: 10.1098/rsta.2016.0101. URL: https://royalsocietypublishing.org/doi/full/10.1098/rsta.2016.0101 (visited on 09/20/2021).

[39] M. Marquis et al. *Wind Forecasting Improvement Project (WFIP) final report*. Tech. rep. National centre of Oceanographic and Atmospheric Administration (NOAA), May 15, 2014. URL: http://energy.gov/sites/prod/files/2014/05/f15/wfipandnoaafinalreport.pdf.

[40] C. Matzler. *Thermal Microwave Radiation: Applications for Remote Sensing*. Vol. Chapter 1. The Institution of Engineering and Technology, London, 2006.

[41] MEASNET. *Evaluation of site-specifc wind conditions V2.0. Procedure*. MEASNET, Apr. 2016. URL: http://www.measnet.com/wp-content/uploads/2016/05/MeasnetSiteAssessment%20V2.0.pdf.

[42] *Meteorological Monitoring Guidance for Regulatory modelling Applications*. Doc. EPA-454/R-99-005. 2000. URL: https://www3.epa.gov/scram001/guidance/met/mmgrma.pdf.

[43] C. Möhrlen et al. "EIRGRID Met Mast and Alternatives Study". In: *IET Renew. Power Gener.* (2021). Submitted Aug. 2021.

[44] Corinna Möhrlen and J.U. Jørgensen. "Integration of Large-Scale Renewable Energy into Bulk Power Systems. From Planning to Operation". In: ed. by Aidan Tuohy Pengwei Du Ross Baldick. Springer International Publishing AG, Cham, Switzerland: Springer Nature, 2008. Chap. The role of ensemble forecasting in integrating renewables into power systems: From theory to real-time applications, pp. 79–134. ISBN: Hardcover 978-3-319-55579-9., eBook 978-3-319-55581-2. DOI: 10.1007/978-3-319-55581-2.

[45] Corinna Möhrlen, Markus Pahlow, and Jess U. Jørgensen. *Author English Translation of (Untersuchung verschiedener Handelsstrategien für Wind- und Solarenergie unter Berücksichtigung der EEG 2012 Novellierung / Investigation of various trading strategies for wind and solar power developed for the new EEG 2012 rules*. URL: http://http://download.weprog.com/WEPROG_Trading_strategies_EEG2012_ZEFE_71-2012-01_en.pdf.

[46] Corinna Möhrlen, Markus Pahlow, and Jess U. Jørgensen. "Untersuchung verschiedener Handelsstrategien für Wind- und Solarenergie unter Berücksichtigung der EEG 2012 Novellierung". In: *Zeitschrift für Energiewirtschaft* 36.1 (Mar. 2012), pp. 9–25. ISSN: 1866–2765. DOI: 10.1007/s12398-011-0071-z. URL: https://doi.org/10.1007/s12398-011-0071-z.

[47] Corinna Möhrlen and Ulrik Vestergaard. *EIRGRID Met Mast and Alternatives Study*. Tech. rep. Version 2. 2019. URL: http://www.eirgridgroup.com/site-files/library/EirGrid/EIRGRID-Met-Mast-and-Alternatives-Study-Version-2.pdf.

[48] S. Monserrat and A. J. Thorpe. "Gravity-Wave Observations Using an Array of Microbarographs In the Alearic Islands". In: *Q.J.R. Meteorol. Soc.* 118 (1992), pp. 259–282. DOI: 10.1002/qj.49711850405.

[49] VR Morris. *Ceilometer Instrument Handbook, DOE/SC-ARM-TR-020*. 2016. URL: http://www.arm.gov/publications/tech_reports/handbooks/ceil_handbook.pdf.

[50] A. Peña, J. Mann, and N. Dimitrov. "Turbulence characterization from a forward-looking nacelle lidar". In: *Wind Energy Science* 2.1 (2017), pp. 133–152. DOI: 10.5194/wes-2-133-2017. URL: https://wes.copernicus.org/articles/2/133/2017/.

[51] Pierre Pinson and Renate Hagedorn. “Verification of the ECMWF ensemble forecasts of wind speed against analyses and observations”. In: *Meteorological Applications* 19.4 (2012), pp. 484–500. DOI: https://doi.org/10.1002/met.283. eprint: https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/met.283. URL: https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/met.283.

[52] Christelle Rigollier, Mireille Lefèvre, and Lucien Wald. “The method Heliosat-2 for deriving shortwave solar radiation from satellite images”. In: *Solar Energy* 77.2 (2004), pp. 159–169. URL: https://hal.archives-ouvertes.fr/hal-00361364/document.

[53] L. Schröder et al. “PV-Kürzestfristvorhersage mit Satellitendaten und Wolkenkamera”. 5. Fachtagung Energiemeteorologie, 5.–7. Juni 2018, Goslar, Germany. 2018. URL: https://uol.de/physik/forschung/ehf/energiemeteorologie/veranstaltungen/fachtagung-energiemeteorologie.

[54] M. Sengupta et al. *Best Practices Handbook for the Collection and Use of Solar Resource Data for Solar Energy Applications: First Edition*. Tech. rep. NREL/TP-5D00-63112. National Renewable Energy Laboratory, Golden, Colorado, USA, 2015. URL: %5Curl%7Bhttps://www.nrel.gov/docs/fy15osti/63112.pdf%7D.

[55] Manajit Sengupta et al. *Best Practices Handbook for the Collection and Use of Solar Resource Data for Solar Energy Applications: Second Edition*. Tech. rep. NREL/TP-5D00-68886. National Renewable Energy Laboratory, Golden, Colorado, USA, 2017.

[56] Manajit Sengupta et al. *Best Practices Handbook for the Collection and Use of Solar Resource Data for Solar Energy Applications: Third Edition*. Tech. rep. National Renewable Energy Laboratory, Golden, Colorado, USA, IEA PVPS, 2021. URL: https://iea-pvps.org/key-topics/best-practices-handbook-for-the-collection-and-use-of-solar-resource-data-for-solar-energy-applications-third-edition/.

[57] DA Smith et al. “Wind lidar evaluation at the Danish wind test site in Høvsøre”. In: *Wind Energy* 9 (2006), pp. 87–93.

[58] *Spinner Anemometry – Uncertainty Analysis*. Tech. rep. Version E-0166. DTU Wind Energy, 2018. URL: https://www.ispin-ptp.com/wp-content/uploads/2018/04/DTU-E-0165-Spinner-Anemometry-Best-Practice.pdf.

[59] *Supplement 1 to the "Guide to the expression of uncertainty in measurement – Propagation of distributions using a Monte Carlo method*. Version JCGM 101:2008. Joint Committee for Guides in Metrology, Evaluation of measurement data, 2008.

[60] *The role of measurement uncertainty in conformity assessment*. Version JCGM 106:2011. Joint Committee for Guides in Metrology, Evaluation of measurement data, 2011.

[61] *The role of measurement uncertainty in conformity assessment*. Version JCGM 106:2012. Joint Committee for Guides in Metrology, Evaluation of measurement data, 2012.

[62] F. T. Ulaby, R. K. Moore, and A. K. Fung. *Microwave Remote SensingActive and Passive*. Vol. Volume I: Microwave Remote Sensing Fundamentals and Radiometry. Reading, Massachusetts: Addison-Wesley, 1981.

[63] J. W. Wagenaar et al. "Evaluation of the ROMO Wind iSpin Guardian approach". In: (2016). Publisher: ECN. URL: https://repository.tno.nl/islandora/object/uuid%5C%3Abc429508-83c4-4eae-9078-f76510d1b9d9 (visited on 09/24/2021).

[64] James Wilczak et al. "THE WIND FORECAST IMPROVEMENT PROJECT (WFIP): A Public–Private Partnership Addressing Wind Energy Forecast Needs". In: *Bulletin of the American Meteorological Society* 96.10 (2015), pp. 1699–1718. ISSN: 00030007, 15200477. URL: http://www.jstor.org/stable/26224979.

[65] IEA Wind. *Ground-Based Vertically-Profling Remote Sensing For Wind Resource Assessment. Recommended Practice RP 15*. International Energy Agency TCP Wind, Jan. 2013. URL: https://github.com/IEA-Wind-Task-32/RP15-Ground%20-%20Based%20-%20Remote%20-%20Sensing%20-%20for%20-%20Wind%20-%20Resource%20-%20Assessment/releases/tag/1.0.

[66] WMO. *WMO Guide to Meteorological Instruments and Methods of Observation*. Tech. rep. Tech. Rep. WMO, 2018.

[67] Ines Würth et al. "How far do we see? Analysis of the measurement range of long-range lidar data for wind power forecasting". In: (2017). Accepted: 2018-03-15T17:48:48Z. DOI: 10.18419/opus-9693. URL: http://elib.uni-stuttgart.de/handle/11682/9710 (visited on 05/09/2021).

[68] Ines Würth et al. "Minute-Scale Forecasting of Wind PowerResults from the Collaborative Workshop of IEA Wind Task 32 and 36". In: *Energies* 12.4 (2019). ISSN: 1996-1073. DOI: 10.3390/en12040712. URL: https://www.mdpi.com/1996-1073/12/4/712.

[69] Frederik Zahle and Niels N. Sørensen. "Characterization of the unsteady flow in the nacelle region of a modern wind turbine". In: *Wind Energy* 14.2 (2011), pp. 271–283. DOI: https://doi.org/10.1002/we.418. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/we.418. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/we.418.

[70] Shiyuan Zhong, Jerome D Fast, and Xindi Bian. "A case study of the Great Plains low-level jet using wind profiler network data and a high-resolution mesoscale model". In: *Monthly Weather Review* 124.5 (1996), pp. 785–806.

Appendix A

Examples of System Operator Met Measurement Requirements

A.1 Comparison of Requirements in various jurisdictions

- **AESO**: Alberta Electric System Operator in Calgary, Alberta, Canada [AESO,2011]
- **CAISO**: California Independent System Operator [CAISO, 2014, 2016]
- **BPA**: Bonneville Power Administration in Portland, Oregon, USA [BPA, 2015]
- **ERCOT**: Electricity Council of Texas in Austin, Texas, USA [ERCOT, 2012]
- **NYISO**: New York Independent System Operator in Rensselaer,NY,USA [NYISO, 2016]
- **PJM**: Independent System Operator in Audubon, PA, USA [PJM, 2016]
- **HECO**: Hawaiian Electric Company, Maoui, Hawaii [HECO, 2016]
- **LitGrid**: in Vilnius, Lithuania [LiTGRID, 2010]

A.2 Met Measurement Example from Califorina Independent System Operator in USA

The following tables are examples from the California Independent System Operator taken from the Appendix Q of their *Eligible Intermittent Resources Protocol* (EIRP) from December 2020.

Element	Device(s) Needed	Units	Accuracy
Wind Speed (Meter / Second)	Anemometer, wind vane and wind mast	m/s	± 2m/s
Air Temperature (Degrees Celsius)	Temperature probe & shield for ambient temp	°C	± 1°
Barometric Pressure (hecto Pascals)	Barometer	hPA	± 60 hPa

Figure A.1: *Wind Eligible Intermittent Resources Telemetry Data Points*

Element	Device(s) Needed	Units	Accuracy
Wind Speed (Meter / Second)	Anemometer, wind vane and wind mast	m/s	± 2m/s
Wind Direction (Degrees - Zero North 90CW)	Anemometer, wind vane and wind mast	Degrees	± 5°
Air Temperature (Degrees Celsius)	Temperature probe & shield for ambient temp	°C	± 1°
Barometric Pressure (hecto Pascals)	Barometer	hPA	± 60 hPa
Back Panel Temperature (Degree C)	Temperature probe for back panel temperature	°C	± 1°
Plane of Array Irradiance Watts\Meter Sq.	Pyranometer or Equivalent	W/m²	± 25 W/m²
Global Horizontal Irradiance Watts\Meter Sq.	Pyranometer or Equivalent	W/m²	± 25 W/m²
Direct Irradiance Watts\Meter Sq.	Pyranometer or Equivalent	W/m²	± 25 W/m²

Figure A.2: *Solar Eligible Intermittent Resources Telemetry Data Points*

A.3 Met Measurement Example from Irish System Operator EIRGRID Group

Examples from the Met Data Requirement document of the Irish transmission system operator EIRGRID, which is part of the grid code for wind generation units under section WFPS 1.7.1.6:

Time Delays and Data Quality

Digital signal changes from the Controllable WFPS shall be relayed to the TSO Telecommunication Interface Cabinet within 1 second of the associated change of state event. Analogue signal changes shall be relayed within 5 seconds and with an error of 0.5% or less, with the exception of the Meteorological Data required as per PPM 1.7.1.2.1, which shall be updated within 5 seconds and with an error of 2.5% or less.

System accuracies and Measurement resolution

The meteorological data signals provided shall be as detailed in Table 1: Meteorological data signal accuracy and resolution and Table 2: Meteorological data variable and their error threshold limit. The WFPS shall provide an updated signal every 1 minute.

	Unit	Range	System Accuracy	Measurement Resolution
Wind speed	m/s	0 - 70	≥ 5% improvement	0.1m/s
Wind direction	deg	0 - 360	Statistical and Variance test within acceptable limits as per Table 2	1.0 deg
Air temperature	deg C	-40 - 70		0.1 deg C
Air pressure	mBar	735 - 1060		1 mBar

Figure A.3: *Meteorological data signal accuracy and resolution*

	Maximum Bias	Maximum MAE	Minimum Correlation	Measurement Unit
Wind speed	3.00	3.00	0.65	m/s
Wind direction	13.00	20.0	0.55	deg
Air temperature	2.00	2.50	0.75	degC
Air pressure	50.00	85.0	0.90	mBar

Figure A.4: *Meteorological data variable and their error threshold limit for statistical tests*

A.4 Met Measurement Example from Alberta Electric System Operator in Canada

Requirements for Solar Grid Connected Plant Meteorological Data at the Alberta Electric System Operator in Canada.

Reference: `https://www.aeso.ca/rules-standards-and-tariff/iso-rules/section-304-9-wind-and-solar-aggregated-generating-facility-forecasting/`

Section 304.9 - Wind and Solar Aggregated Generating Facility Forecasting

ISO Rules Part 300 on System Reliability and Operations Division 304 Routine Operations Section 304.9 Wind and Solar Aggregated Generating Facility Forecasting

The rule states on the *Applicability* that "..Section 304.9 applies to: (a)the legal ownerof a wind or solar aggregated generating facility connected to the interconnected electric system or an electric system within the service area of the City of Medicine Hat, including a wind or solar aggregated generating facility situated within an industrial complex that is directly connected to the interconnected electric systemor to an electric system within the service area of the City of Medicine Hatand that has a gross real power capability equal to or greater than 5 MW; and (b)the ISO."

In section of the rule 304.9 the following requirements apply:

(4) The legal ownerof a wind aggregated generating facility must, inresponse to a request by the ISO under subsection 8(3),providethe following facility data:

(a) meteorologicaltower data collection heightin meters (m), with a precision for instantaneousmeasurements to the nearest 1m;

(b) turbinemodelname

(c) turbinemodel capacity in megawatts (MW), with a precision to the nearest 0.1MW

(d) turbinewind speed cut-in in meters per second (m/s), with a precision to the nearest 0.1 m/s

(e) turbinewindspeedcut-outin meters per second (m/s), with a precision to the nearest 0.1 m/s

(f) turbine temperaturecut-out lowerin degrees Celsius (C), with a precision for instantaneousmeasurements to the nearest 1Cand an indicator is required to confirm that the numbers are ambient temperature within the rotor or air temperature

(g) turbinetemperaturecut-out upper in degrees Celsius (C), with a precision for instantaneousmeasurements to the nearest 1Candan indicator is required to confirm that the numbers are ambient temperature within the rotor or air temperature

(h) site latitude and longitude in degrees; and(i)turbine power curves.

Equivalently, the technical rule states that "..the legal owner of a solaraggregated generating facilitymustin response to a request by the ISO under subsection 8(3), provide the following solar arraydata and records, including:

(a) site latitude and longitude in degrees

(b) direct current (DC)real power rating

(c) alternating current (AC) real power rating

(d) inverter manufacturer and model

(e) mounting height from ground in meters(m)

(f) tilt angle or range of tilt angles to horizontal plane in degrees

(g) azimuth angle in degrees;(h)alternating current (AC) real power capacity per solar array in megawatts (MW)

(h) mounting type, tracking (fixed, single or dual axis)

(i) module type (crystalline, thin-film etc.).

Table A.1: Alberta Electric System Operators's Wind Aggregated Generating Facility Meteorological Data Requirements Technical Rule 304.9

Wind Aggregated Generating Facility Meteorological Data Requirements						
Measurement	**Units**	**Precision**	**Range**	**Accuracy**	Height of instrument	
Type					**Set-1**	**Set-2**
Wind Speed	Meters /Second (m/s)	0.1 m/s	0 to 50	± 1m/s	At Hub Height	At 35m Meters
Wind Direction	Degrees from True North	1 degree	0 to 360	± 5°	At Hub Height	At 35m Meters
Barometric Pressure	Hecto Pascals (HPa)	1 hPa	800 to 1000	± 1.0 hPa at −20 to 50 °C, and ± 1.5 hPa at below -20° C	At Convenient location	At Convenient location
Ambient Temperature	Degree Celsius (řC)	0.1ř C	-50 to +50	± 0.2°C	At Hub Height	At 35m Meters
Dewpoint	Degrees Celsius (°C)	0.1ř C	-50 to +50	± 0.2°C	At Convenient location	At Convenient location
Relative Humidity	Percentage (%)	1.00%	0 to 100 %	± 2%	At Convenient location	At Convenient location
Ice-up Parameter	Scale 0.0 to1.0	0.1	0 to 1	n/a	At Convenient location	At Convenient location
Precipitation	Millimetres /minute (mm/min)	0.1	0 to11	2% up to 0.417 mm/mon 3% over 0.417 mm/min	At Convenient location	At Convenient location

Table A.2: Alberta Electric System Operators's Solar Aggregated Generating Facility Meteorological Data Requirements Technical Rule 304.9

Solar Aggragated Generating Facility Meteorological Data Requirements					
Measurement	**Units**	**Precision**	**Range**	**Accuracy**	Height of instrument

Type					**Set-1**	**Set-1**
Wind Speed	Meters /Second (m/s)	0.1 m/s	0 to 50	± 1m/s	Between 2-10 meters	Between 2-10 meters
Wind Direction	Degrees from True North	1 degree	0 to 360	± 5°	Between 2-10 meters	Between 2-10 meters
Barometric Pressure	Hecto Pascals (HPa)	1 hPa	800 to 1000	± 1.0 hPa at −20 to 50 °C, and ± 1.5 hPa below −20°C	Between 2-10 meters	Between 2-10 meters
Ambient Temperature	Degree Celsius (řC)	0.1°C	-50 to +50	± 0.2°C ă	Between 2-10 meters	Between 2-10 meters
Dewpoint	Degrees Celsius (řC)	0.1ř C	-50 to +50	± 0.2°C	Between 2-10 meters	Between 2-10 meters
Relative Humidity	Percentage (%)	1.00%	0 to 100 %	± 2%	Between 2-10 meters	Between 2-10 meters
Ice-up Parameter	Scale 0.0 to1.0	0.1	0 to 1	n/a	Between 2-10 meters	Between 2-10 meters
Precipitation	Millimetres /minute (mm/min)	0.1	0 to11	2% up to 0.417 mm/mon 3% over 0.417 mm/min	Between 2-10 meters	Between 2-10 meters
Backpanel Temperature	Degree Celsius (°C)	0.1°C	-50 to +50	± 0.1°C at −27 to +50°C, and ± 0.2°C at below −27°C	Between 2-10 meters	Between 2-10 meters
Global Horizontal Irradiance	Watts/Square Metre (W/m2)	0.1	0 to 4000	± 3%	Between 2-10 meters	Between 2-10 meters
Diffused Horizontal Irradiance	Watts/Square Metre (W/m2)	0.1	0 to 4000	± 3%	Between 2-10 meters	Between 2-10 meters
Direct Normal Irradiance	Watts/Square Metre (W/m2)	0.1	0 to 2000	± 3%	Between 2-10 meters	Between 2-10 meters
Sunshine Duration	V	0.1	0 to 1	90.00%	Between 2-10 meters	Between 2-10 meters

Appendix B

Statistical Metrics

BIAS: *Indicates whether the model is systematically under- or over-forecasting*

$$BIAS=\frac{1}{n}\sum_{i=1}^{n}(f_i-m_i)$$

where f is the forecast and m the measurement.

Mean Absolute Error (MAE): *The average of all absolute errors for each forecast interval. Measures the average accuracy of forecasts without considering error direction.*

$$MAE=\frac{1}{n}\sum_{i=1}^{n}(|f_i-m_i|)$$

where f is the forecast and m the measurement.

Mean Absolute Percent Error (MAPE): *This is the same as MAE except it is normalized by the capacity of the facility.*

Correlation: Correlation is a statistical technique that is used to measure and describe the STRENGTH and DIRECTION of the relationship between two variables.

$$r(x,y)=\frac{COV(x,y)}{STD_x\cdot STD_y}=\frac{\sum(x-\bar{x})\cdot(y-\bar{y})}{N\cdot STD_x\cdot STDy}$$

where f are the forecasted values, m are the measurements, COV is the covariance, STD is the standard deviation.