IEA Wind Task 36 Session Topic 4: Request for Feedback on Version 1 of the Recommended Practices for Forecast Solution Selection

John W Zack
AWS Truepower, a UL Company
Albany, NY, USA
John.Zack@ul.com

Corinna Möhrlen
WEPROG
Assens, Denmark
com@weprog.com

Caroline Draxl
National Renewable Energy Laboratory (NREL)
Golden, CO, USA
Caroline.Draxl@nrel.gov

***Abstract*—In phase 1 of the IEA Wind Task 36 a group of experts prepared an IEA Recommended Practice on Forecast Solution Selection (RP-FSS), which provides guidance on the process of selecting a new or additional forecast solution, the execution of a forecasting trial or benchmark, and the evaluation metrics and methods used to assess forecast quality.**

The RP-FSS is composed of three documents. This set of documents provides guidance on almost all aspects of the selection of a renewable power forecast solution. 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 second part, "Benchmarks and Trials", of the series offers recommendation on how to best conduct benchmarks and trials in order to evaluate the relative performance and the "fit-for-purpose" of alternative forecasting solutions. The third part, "Forecast Evaluation", provides information and guidelines for the effective evaluation of the performance of forecasts and forecast solutions.

The effectiveness of forecasts in reducing the cost of managing the variability of wind and solar power generation is dependent upon both the accuracy of the forecasts and the ability of users to effectively use the forecast information in application-based decision-making processes. Therefore, there is considerable motivation for stakeholders to try to obtain high quality forecasts that have a format and content that can be effectively used as input to operational processes or market-based transactions. The RP-FSS documents are intended to provide guidance to stakeholders who are seeking to initiate or optimize a forecasting solution that will maximize the benefit for their specific applications.

A key objective of the second phase of the IEA Wind Task 36 is to update the initial version of the RP-FSS documents to maximize their relevance and usefulness to the members of the stakeholder community. The first step towards this objective will be to collect feedback on the first version from a broad sample of the stakeholder community and especially users of operational forecast information. It is envisioned that the knowledge gained from this feedback will guide the update process and result in a second version of the RP-FSS documents that will more effectively address the needs of the stakeholder community. This paper is designed to provide background information and a summary of the RP-FSS for further discussions with the community.

Keywords—forecast solution selection, forecast benchmarks and trials, optimization of forecast value

I. Introduction

The operational use of wind and solar power production forecasts has become widespread in the electric power industry and their benefits for the management of the variability of the generation associated with these renewable energy technologies have been documented in a number of studies (e.g., [1], [2]). However, while the operational use of forecasts has substantially grown over the past decade, there is considerable evidence that the full potential value of the wind and solar forecasts in many applications is often not realized. This relates in many cases to three factors:

(1) The first factor is the specification of the wrong forecast performance objectives in the forecast solution selection process. For example, a user may implicitly or explicitly state that the objective is to minimize the typical or average error of the forecast. However, the user's application may be more sensitive to large errors or errors associated with specific types of events. While it would be ideal to have a system that produces perfect forecasts in all situations, the reality is that the error characteristics of forecasts are linked to the way in which they are optimized. For example, a forecast system that is optimized to minimize the average error will generally not produce the best forecasts of anomalous events.

(2) A second key issue is the use of poorly designed benchmarks or trials to select a forecast solution for the user's application. Poorly designed benchmarks and trials will frequently provide invalid and misleading information to the solution selection process and can result in the selection of a solution that does not provide the best solution for the user's application even though the user thinks it is the best solution based on the data compiled from the benchmark or trial.

(3) A third factor is the use of non-optimal evaluation metrics. A user may correctly specify the performance objective and then conduct a well-designed and executed benchmark or trial but ultimately evaluate the forecasts with metrics that do not measure the performance attributes that are most important to the user's application. This can result in the selection of a solution that is ideal for some other user's application but not for the application of the user conducting the solution selection process.

The result of these and other flaws in the forecast solution selection process is that the value of renewable energy forecast information is reduced below its full potential for both the specific users and implicitly for a broad range of stakeholders in the energy community since it results in higher integration costs for wind and solar electricity generation and also inhibits a higher penetration level for these generation resources on grid systems

In order to address this issue, an international group of experts has worked under the structure of Task 36 of the International Energy Agency's (IEA) Wind Technology Collaboration Program (known as "IEA Wind") to develop a set of three recommended practices documents to provide guidance on forecast solution selection. The IEA is an independent international entity that is composed of 30 member countries and 8 associate countries. Information about the IEA may be found at https://www.iea.org/about/. IEA Wind is an international co-operation consortium of a subgroup of IEA members that shares information and research activities to advance wind energy research, development and deployment in member countries. There are also a number of other consortia that operate under the auspices of the IEA to address issues associated with other energy technologies such as solar photovoltaic generation and the more traditional fossil fuel based generation technologies. IEA Wind Task 36 is a focused activity that facilitates the interaction of international experts to address issues associated with short-term wind power forecasting. The first phase of the Task 36 activities extended from 2016 through 2018. The second phase began at the start of 2019 and will extend through the end of 2021. Information about the past, current and future activities of Task 36 can be obtained from the task's web portal at ieawindforecasting.dk. Additional information about the activities of Task 36 can also be found via the Research Gate Project web portal [3] and also via the Task 36 YouTube channel [4].

One of the objectives of the second work package in the first phase of Task 36 was to construct the recommended practice guidelines for forecast solution selection to facilitate the use of more optimal wind forecast solutions for a broad range of applications related to the operation of electric power systems. The interim progress of this activity was summarized in papers and presentations at the 2017 [5] and 2018 [6] Wind Integration workshops. An initial version of the three Recommended Practices for Forecast Solution Selection (RP-FSS) documents was completed at the end of the first phase of Task 36. The title pages of these three documents are shown in Fig. 1. The next three sections of this paper provide an overview of the contents and key points addressed in each of these documents. These are followed by a concluding section that provides a summary of the plans to refine these documents during second phase of Task 36 (i.e., 2019-2021).

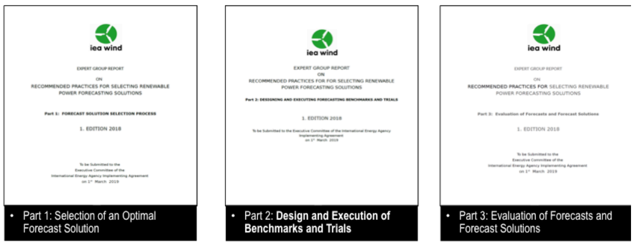

Fig. 1. Title pages of the three IEA Wind Recommended Practices for Forecast Solution Selection (RP-FSS) documents.

II. Part 1: Forecast Solution Selection Process

The first of the three RP-FSS documents addresses the process of selecting an optimal wind forecasting solution for a specific set of applications. This is intended to provide guidance for the design of an economically viable process that will maximize the probability of obtaining an optimal forecast solution for a user's applications. The document is divided into two core sections. The first is a discussion of the "big picture" issues that should be considered before starting the design of a selection procedure. The second is the presentation and discussion of a Decision Support Tool (DST) that steps through the issues that should be considered during the design of a forecast solution selection process. The following two subsections summarize some of the key points in these core components of RP-FSS Part 1.

A. *Initial Considerations*

The first step in the forecast solution selection process is to define the objectives of the forecasting application. For example, very different forecasting strategies are needed for the balancing of supply and demand on a system with a significant penetration of renewable generation versus the selling of generated electricity in the power market. In the first application, 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 to use a forecast whose errors are least correlated with other forecasts in the market.

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. Furthermore, if there is no knowledge in the organization regarding the techniques required to reach the objective, it is recommended to start with a "request for information" (RFI) from a set of forecast providers and thereby gain an understanding and overview of the various existing solutions and their capabilities.

Once the applications objectives are clearly and specifically defined the next step is a detailed specification of the desired outcome of the solution selection process. The key questions to be asked are:

- What specific forecast information is needed for the application?
- What infrastructure and resources does the user have to support in the solution selection process and ultimately the implementation of the forecast solution?
- What criteria will be used to determine which is the best solution for the target application?
- What forecast services are available from solution providers?
- What level of customization is available from solution providers?
- What is the cost range of the available forecasts?
- What is the historical performance level of the available solutions?

The answers to these questions should play a major role in defining the scope of the selection process. The answers to the first three questions define the requirements and limitations of the desired forecast solution. The last four questions provide information about what is available in the forecasting marketplace. The degree of alignment between the user's requirements and limiting factors and what is available in the marketplace should be the basis for the formulation of a selection process.

For example, a lot of time and resources can be wasted by conducting trials or benchmarks (to determine the best performing solution for the user's application) that are not aligned with the user's requirements and also planned and conducted by personnel who are not experienced with these issues. In order to avoid this, it is recommended that the user compile a "requirements list" at the start of the selection process. An example of a requirements list is presented in RP-FSS Part 1.

In some cases, it can be beneficial to test solutions prior to implementation. The difficulty with this approach lies in the quality of the information from the tests, especially, when they are based on a short time period. In many cases they do not answer the questions an end-user needs to answer. This is because such tests are usually simplified in comparison to the real-time application but still require significant resources to conduct. For such cases, this guideline provides other methods for an evaluation of alternative forecast solutions. The pitfalls and challenges with trials and/or benchmarks are addressed in Part 2.

B. Decision Support Tool

From an end-user perspective, it is a non-trivial task to decide which path to follow in the selection and implementation of a forecasting solution for a specific application. In most user situations there are multiple stakeholders involved in the decision-making process. A relatively straightforward way to decide on the path is to use a decision support

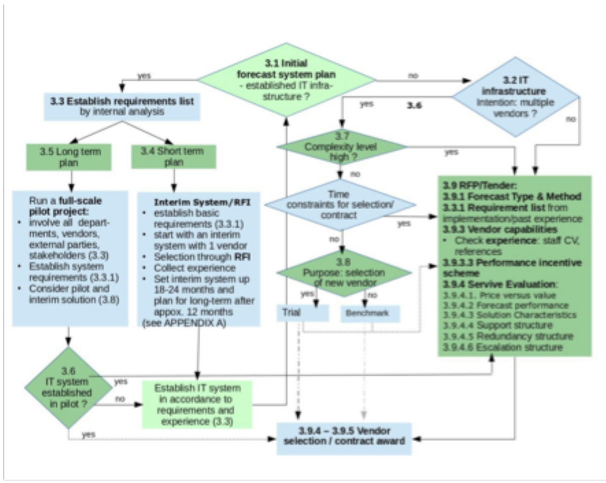

Fig. 2. A decision support tool for the planning and design of a variable generation forecast solution. The numerical citations in the flow chart objects refer to the sections in RP-FSS Part 2 document in which associated with the referenced topic are discussed

tool. Fig. 2 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. The high-level thought construct shown in Fig. 2 is targeted to assist in considering the required resources and involvement of staff in the decision process. The decision tool is constructed to begin with initial considerations to establish a "Forecast System Plan". There are cross-references in the decision tool and referrals to alternative decision streams, depending on the answer at each step of the decision flow.

III. Part 2: Design and Execution of Benchmarks and Trials

The second of the three RP-FSS documents provides guidance for the design and execution of benchmarks or trials (B/T). For the purposes of the RP-FSS documents, a **benchmark** is defined as an exercise conducted to determine the features and quality of renewable energy forecast systems or methods such as those used to produce wind or solar power forecasts. The exercise is normally conducted by an institution or their agent and multiple participants including private industry forecast providers or applied research academics. A **trial** is an exercise conducted to test the features and quality of operational renewable energy forecast solutions. This may include one or more participants and is normally conducted by a private company for commercial purposes. A trial is a subset of benchmarks. A trial may be part of the process to select an initial, replacement or additional forecast solution providers or part of a periodic evaluation process for an existing forecast solution. In any of these cases the fundamental objective of the trial is to determine which solution represents the best value for a user's application.

While a B/T may intuitively seem to be the best approach to identify the best forecasting solution for a specific application, in practice, the use of a trial as part of the solution selection process is not always the best option and has a number of limitations to go along with its benefits. The trade-off between the limitations and benefits of a trial should be carefully considered before a decision is made to conduct a trial. The Part 2 document addresses the benefits and limitations of a trial in a range of scenarios.

The structure of Part 2 is based on the three phases of a B/T: (1) preparation, (2) execution and (3) evaluation and decision-making. Some of the key issues in each of these phases are summarized in the following three subsections.

TABLE I. Key Factors in the Design of a B/T

Attribute	Issues
Forecast Time Horizon	• Are forecast horizons less than 6 hours operationally important? If "no", a live data feed may not be necessary. Although there are advantages of a live trial, it is one of the most time consuming and costly aspects of a B/T. • Are forecast lead times greater than "day-ahead" operationally important? If "no", this will reduce the volume of data that needs to be processed saving time and resources. • If many lead times are important, consider that the performance of providers will likely vary across lead times; therefore, ranges of lead times should be evaluated separately.
Anticipated Weather Conditions	• Will the benchmark take place during the likely periods of weather conditions that reflect the organization's difficulties in handling renewable generation, e.g. windy or cloudy periods? If the answer is "No", the trial operator should strongly consider doing a retrospective forecast (a "backcast") that includes the types of conditions that are critical for the user's application.
Availability of Historical Data	• For locations in which there are significant seasonal differences in renewable generation levels and variability, it is best to provide 12 months or more of historical data from the target generation facilities for the purpose of training forecast models. However, if it is not feasible a minimum of 3-6 months of historical observations is required. • Advanced machine learning methods often exhibit significantly greater performance improvement over less sophisticated methods as the training sample size increases. Thus, solutions that employ advanced machine learning prediction tools may not be able to demonstrate their ultimate value if only short historical data sets are provided. • In general it is recommended that the T/B operator should provide a data set of the typical length that is available for the application that is the target of the B/T since variations in the size of training data sets can bias the trial results in favor of particular methods
Target Sites(s) 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).
Evaluation 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. 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.

A. Preparation

The preparation phase is the period before the start of the forecasting activities during which the structure and protocols of the B/T are formulated by the B/T operator and disseminated to the solution providers that will participate in the B/T. The decisions and actions during this phase often have a very large impact on the ultimate quality and therefore the value of the information obtained from the B/T. The use of performance results from a poorly designed B/T is often worse than not conducting a B/T since this information is typically viewed as an objective basis for making a selection of a forecast solution and therefore a set of unrepresentative evaluation data can lead to an incorrect conclusion that has the illusion of reliability and objectivity.

There are a number of key decisions that will determine the complexity and therefore the level of effort and cost of a trial. It will also play a major role in determining the quality of the information produced by the B/T. Table I summarizes the key attributes of a trial that have an impact on both the cost of a B/T and the quality of the information produced by the exercise.

B. Execution

The execution phase is the period during which forecasts are produced by the participating solution providers and submitted to the B/T operator. In a live (or real-time) trial, the providers should receive near-real-time data for the forecast target facilities from the B/T operator and submit forecast data on a prescribed schedule to IT platforms designated and controlled by the B/T operator. In a retrospective trial, the providers should receive a historical set of data for the target facilities (for statistical model training purposes) and produce forecasts for a specified evaluation period (that does not overlap with the historical data sample period).

In a well-designed B/T, most of the communication between the trial operator and the solution providers should be during the preparation phase. However, issues often arise during a trial (especially in live trials). It may be helpful to all B/T participants to establish an open forum during the first part of the live B/T period to provide a way to effectively and uniformly resolve all issues. However, it is strongly recommended that if any attributes of the B/T are changed during the live part of the B/T, the changes should be communicated to all participants immediately as they might require action of the solution providers.

C. Evaluation and Decision-making

Intuitively, one might expect the evaluation and decision-making phase to begin after all the forecast data has been gathered from the solution providers at the end of the live or retrospective B/T periods. However, in a well-designed B/T that should not be the case. The forecast evaluation process should begin soon after the first forecasts have been received from the solution providers. This will enable the B/T operator to assess its evaluation design and results production protocols (e.g. software to calculate error metrics, displays to view results, etc.) before the end of the B/T execution period and possibly make adjustments to the evaluation or forecast submission process to mitigate issues that may compromise the quality of the information. It is also recommended that the B/T operator provide at least one interim report that summarizes their evaluation of forecast performance to the solution providers during a live trial. This provides an opportunity for any discrepancies between the evaluations methods of the B/T operator and expectations of the solution providers to be resolved before the end of the trial.

If an interim report was provided during the B/T, 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 B/T operator shares the anonymized forecast results from each solution provider at the time-interval frequency that forecasts were being made at (e.g., hourly). This will help solution providers discover where forecasts are similar or different from the competition which may spawn improved methodologies.

D. Evaluation and Decision-making

Forecast service providers who have participated in numerous trials over the past decade have indicated that there are a number of design and execution problems that have repeatedly appeared in trials during this period. The consequences of errors and omissions in trials are often underestimated. However, if results are not representative, the efforts that have gone into a trial 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. A list of significant issues that have frequently been encountered is presented in Table II.

TABLE II. FREQUENTLY ENCOUNTERED B/T DESIGN AND EXECUTION PROBLEMS

Problem Label	Description
Poor Operator-Provider Communication	All solution providers do not receive the same information due to one-on-one answers to questions via email or phone conversations or previous knowledge (e.g. incumbents).
Unrepresentative Forecast Performance Comparisons	Performance of forecasts for two different power plants or different time periods are compared. Forecast performance can vary substantially among locations and time periods due to variations in forecast difficulty
Bad Design	Bad design examples: (1) a trial with 1-month length during a low-wind month (2) no on-site observations shared with forecast providers (3) hour-ahead forecasts based on once a day data update (4) forecast data only processed in batches or at end of real-time trial
Critical Information Not Provided	Critical information not provided examples: (1) handling of daylight savings time changes in data not specified, (2) time stamp definition for data intervals (beginning/ending) not specified, (3) plant capacity of historical data differs from present capacity, (4) data about curtailment and maintenance outages not provided
Submission Protocol Creates Possibility for Cheating	Examples of protocols that enable cheating: (1) Missing forecasts: Forecast Solution Providers (FSPs) that do not submit forecasts in "difficult situations" are often not penalized. However, missing data may bias "average" forecast metrics, potentially resulting in the formulation of incorrect conclusions. *Recommendation*: remove data for dates for all FSPs in cases in which forecasts are missing for one FSP. (2) 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 forecasted, FSPs can potentially renew forecasts with higher accuracy due to updated information being available. *Recommendation*: This type of situation should not be underestimated and care taken in the evaluation.

IV. PART 3: FORECAST EVALUATION

Part 3 of the document series provides guidance on the evaluation of forecasts. The evaluation process is a large component of the forecast solution selection process, if a benchmark or trial is conducted as part of the process. An evaluation is also an important component of an ongoing performance assessment program.

The Part 3 document is composed of four core sections. These provide (1) a description of the general factors that determine the evaluation uncertainty, (2) an overview of the uncertainty associated with the data from the forecast target site, (3) a discussion of the importance of choosing appropriate metrics to evaluate forecast performance and (4) a compilation of the recommended best practices for forecast evaluation. The following subsections present a summary of the key points in each of these core components of RP-FSS Part 3.

A. *Overview of Evaluation Uncertainty*

The first component of RP-FSS Part 3 provides an overview of the general factors that differentiate the level of uncertainty among forecast evaluation exercises. The focus is on three key points:

- All evaluations of forecast solutions have a degree of uncertainty, which is associated with the three core attributes of the evaluation process: (1) representativeness, (2) significance and (3) relevance.
- A carefully designed and implemented evaluation process that considers these three attributes can minimize the uncertainty and yield the most meaningful results.
- Disregarding these issues can lead to uncertainty that is so high that the conclusions of the evaluation are meaningless and no valid information is available for decision-making.

Representativeness can be defined as 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. There are many factors that influence the ability of the evaluation to be a good predictor of future 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, (4) availability of a complete and consistent set of evaluation procedure information.

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 likely performance in a long-term operational application of a solution. Real performance differences 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 but both noise minimization and awareness of the uncertainty are crucial for reliable decision-making.

Relevance is defined as the degree of alignment between the evaluation metrics used for an evaluation and the true sensitivity of a user's 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.

B. *Measurement Uncertainty*

The second section of Part 3 provides an overview of the factors that contribute to measurement uncertainty, which is a part of the representativeness attribute. The key points are:

- 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 from multiple sources are noted in RP-FSS Part 3

- 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 of Part 3.

C. *Targeted Evaluation of Forecast Performance*

The third component of the Part 3 document addresses the importance of employing an appropriate set of metrics in the evaluation process. A number of publications have compiled lists of metrics for the evaluation of a broad range of attributes of wind and solar power generation forecasts and have provided some examples of their application (e.g., [7]). However, there is little guidance available for the selection of the most relevant set of metrics for a specific application.

The relevance of different aspects of forecast performance depends on the user's application(s). For instance, one user may be concerned with the size of typical forecast errors, while another may only be concerned with the size and frequency of particularly large errors. This component of RP-FSS Part 3 provides a description of the key issues in evaluating specific attributes of forecast performance with a focus on: (1) the relationship between forecast performance attributes and widely-used error metrics and (2) metric-based forecast optimization (i.e., configuring the forecast system for the best performance for a specific metric).

An example of how one's selection of a performance metric and therefore what forecast performance attribute is measured can determine one's perspective on what is considered the best forecast is presented in Fig. 3. Three forecasts for a wind ramp event are depicted. Despite being the only forecast (the left plot) that correctly predicts the ramp rate and duration, the forecast with a phase error has the largest MAE. Thus, this could be considered the worst (i.e., highest MAE) or best forecast (i.e., lowest ramp rate or duration forecast error) depending on one's perspective.

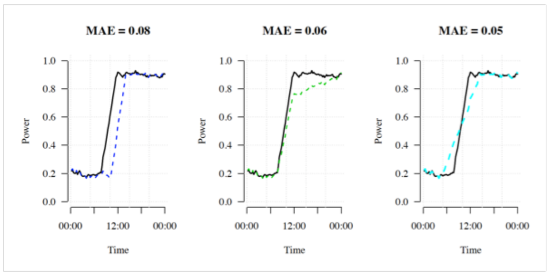

Fig. 3. Examples of 3 types of ramp forecast error. Actual power is shown as solid black lines, forecasts as colored dashed lines. Three types of ramp event forecast errors are depicted: (1) phase or timing error, (2) amplitude error and (3) ramp rate error. The associated MAE (fraction of capacity) is shown at the top of each chart [3].

D. *Best Practices for Forecast Evaluation*

The fourth component of the RP-FSS Part 3 document provides an overview of the recommended practices for forecast evaluation. The key points of this component are:

- Forecast verification is subjective; it is important to understand its limitations.

- Verification has an inherent uncertainty due to its dependence on the evaluation data set

- Evaluation should contain a set of metrics to measure a range of forecast performance attributes

- Evaluation should include a "cost function" i.e. the metric set should assess the value of the solution to the application.

V. PLANS FOR RP-FSS VERSION 2

As noted earlier, the first version of the series of three RP-FSS documents was the culmination of a 3-year effort under Phase 1 of IEA Wind Task 36 and that this series of documents is in the final stages of preparation as a formal IEA report, which is expected to be available in the second part of 2019. Plans are already in place for a continuation of the work on RP-FSS under the second phase of Task 36. The second phase of work on RP-FSS is planned to have two major components: (1) a campaign to obtain feedback from forecast users and other power system stakeholders to identify areas in which the RP-FSS can be improved and (2) the preparation of a second version of the RP-

FSS that addresses the issues identified in the feedback from stakeholders and also expands and refines the scope of the documents. All parties interested in the RP-FSS are strongly encouraged to become part of this international collaboration to improve the value of wind power forecasting.

The campaign to obtain stakeholder feedback will have several components. One of these will be "feedback" workshops at several geographically diverse venues. These will be specifically designed to facilitate stakeholder feedback about the content and format of the initial version of the RP-FSS documents. A second approach will be via presentation at meetings of stakeholder groups and organizations. A third mode will be via a feedback capability that will be established on the IEA Wind Task 36 web portal.

A small group of stakeholders have already provided some valuable feedback on the initial version of the RP-FSS. One key issue that has been raised is that the first version of the RP-FSS is heavily focused on the evaluation and use of deterministic forecast solutions and that very little information is provided about the evaluation and selection of probabilistic forecast solutions, even though probabilistic solution are often better choices for many applications.

Some stakeholders have also identified a need for background information about the sources of uncertainty in wind power forecasts and the relative magnitude of those sources. This would allow forecast users to better understand the role of the different parts of a forecast system and the limitations on forecast performance.

A third issue in stakeholder feedback was that it would be valuable to provide recommendations for the use of experienced third parties to design and execute benchmarks or trials. This can be a more effective approach in cases in which the forecast user does not have sufficient knowledge or experience to conduct a satisfactory B/T. However, if this path it chosen, it raises the question of how to identify a qualified and independent third party.

ACKNOWLEDGMENTS

Significant contributions to some or all of the three parts of the three RP-FSS documents were made by: Jeff Lerner, Jakob Messner, Jethro Browell, Craig Collier, Aidan Tuohy, Justin Sharp, Mikkel Westenholz and Gegor Giebel. In addition, Gregor Giebel, the operating agent for IEA Wind Task 36 provided abundant support and guidance in the preparation of the documents and the navigation of the IEA report publication process. The contribution of Corinna Möhrlen is partially supported by the Danish EUDP project EUDP18-II IEA Wind Task 36 Phase II Danish Consortium under the contract no. 2018-64018-0515.

REFERENCES

[1] Hodge, B-M, A. Florita, A., J. Sharp, M. Margulis, and D. Mcreavy, 2015: The Value of Improved Short-Term Wind Power Forecasting. Technical Report NREL/TP-5D00-63175. National Renewable Energy Laboratory, Golden, Colorado, USA. Available from http://www.osti.gov/scitech.

[2] Orwig, K., B.-M. Hodge, G. Brinkman, E. Ela, M. Milligan, V. Banunarayanan, S. Nasir, and J. Freedman, 2012: "Economic evaluation of short-term wind power forecasts in ERCOT: Preliminary results," Proceedings of the 11th International Workshop on Large-Scale Integration of Wind Power into Power Systems. Energynautics GmbH, Darmstadt Germany

[3] Research Gate URL: https://www.researchgate.net/pro-ject/IEA-Wind-Task-36-Wind-Power-Forecasting

[4] IEA Wind Task 36 Youtube Channel URL: https://www.youtube.com/channel/UCsP1rLoutSXP0ECZKicczXg

[5] Giebel, G., J. Cline, H. Frank, W. Shaw, B.-M. Hodge, C. Draxl, P. Pinson,J. Messner, G. Kariniotakis, C. Möhrlen, 2017: IEA Wind Task 36 Forecasting: Proceedings of the 16th Interntiaonal Workshop on Large-Scale Integration of Wind Power into Power Systems (Berlin, Germany). Energynautics GmbH, Darmstadt Germany, ISBN: 978-3-9816549-6-7.

[6] Möhrlen, C., J. Lerner , J. W. Messner , J. Browell, A. Tuohy, J. Zack, C. Collier, and G. Giebel, 2018: IEA Wind Recommended Practices for the Implementation of Wind Power Forecasting Solutions – Part 2 & 3: Designing and Executing Forecasting Benchmarks and Evaluation of Forecast Solutions. Proceedings of the 17th International Workshop on Large-Scale Integration of Wind Power into Power Systems (Vienna, Austria). Energynautics GmbH, Darmstadt Germany.

[7] Zhang, J., B.-M. Hodge, A. Florita, S. Lu and H. F. Hamann, V. Banunarayanan, 2013: Metrics for Evaluating the Accuracy of Solar Power Forecasting. Proceedings of the 3rd International Workshop on Integration of Solar Power into Power Systems (London, England). Energynautics, Darmstadt Germany.