Can benchmarks and trials help develop new operational tools for balancing wind power ?

Corinna Möhrlen*, Craig Collier†, John Zack‡ and Jeffrey Lerner§

*WEPROG ApS, 5631 Assens, Denmark, Email: com@weprog.com
†DNV GL, San Diego, CA 92123, USA, Email: craig.collier@dnvgl.com
‡UL AWS Truepower, Albany, NY 12205 USA, Email: jzack@awstruepower.com
§Vaisala, Seattle, WA 98121 ,USA, Email: Jeffrey.Lerner@vaisala.com

***Abstract*—Wind power forecasting trials and benchmarking have become more and more popular over the past few years. Numerous similar benchmarking and trials on international levels have been used to evaluate forecast vendors and methods and enhanced price competition. It could be expected that such benchmarking should have led to a high degree of understanding of the importance of forecast accuracy, guidance on forecast evaluation and understanding of error patterns and to what extent such error patterns will remain in the future. However, the contrary seems to have happened. There is a flooded market delivering very similar services at low costs because benchmarking and trials have some notable design flaws and simplistic evaluation methods that don't push the state-of-the-art.**

The IEA Wind Task 36 has therefore dedicated a work package (2.1) to generate a guideline for the power industry that is intended to provide a compendium of information for all those that are designing or renewing their forecasting system tools in the operation of the power system. The guideline will not only discuss how to setup a trial or benchmarking exercise in order to find the best methodology, but also discuss the many pitfalls that leave the end-user with open questions and all participants with a lot of wasted resources. A decision support process will be provided to make the right decisions from the outset and to assist in the design of benchmarks that make the end-user first answer the questions of own requirements and needs instead of following practices that at the end of the day stall necessary evolution.

In this paper, we will present a summary of the findings from our analysis of past and current industry practices and a prospect of the development towards the full compendium.

I. Introduction

The forecasting community and research projects have not paid enough attention to the importance of explaining different types of forecast error and different types of forecasting methods and their applicability. Recent trends in the procurement of forecasting solutions in the power industry show that the same types of benchmarks and trials are conducted with usually a very similar outcome: no contract is executed or the cheapest solution is chosen. The first reason often is due to the fact that other vendors are benchmarked against the incumbent vendor; the latter is often due to the fact that the cheapest vendor may be significantly cheaper and the results (using simplistic metrics) are not significantly different. The common use of MAE or RMSE as the most used metrics for evaluation today is one of the factors that leads to the seemingly low diversity of vendors' performance when compared for a number of sites. Even though the end-users are aware of the fact that high errors cost more than low errors, particularly if they are correlated with other errors. A forecast which is variable and out of phase with reality is more expensive to balance than one with less volatility. Thus, two forecasts with the same MAE can have a very different economic value both on the day-ahead and intra-day horizon as many of the largest economic losses and gains are event driven. If the physical power, and in particular, balancing is taken into account neither MAE nor RMSE would today be used as error metrics because reserve costs and energy prices would be far more important to operations. The forecast error characteristics during volatile energy price events are far more important than bulk MAE and RMSE. The cumulative cost (or benefit) due to VER forecast error becomes much more important to the decision on which forecast provider to contract with.

Nevertheless, the simplest solution and the cheapest vendor has an opportunity to win a contract, though this vendor may be the one falling behind on the latest innovations. Thus, provision of an old simple well-tuned service is from a business perspective clever for the forecast vendor in such cases. However, this strategy does not lead to evolution of tools required to physically handle larger penetration levels. It also means that the integration costs of variable energy resources (VER) increases and the utilization of VER is going in the wrong direction. Typically, more curtailment and higher balancing costs are the result. If such aspects are not taken into consideration, economics on the individual service costs takes over and evolution stalls.

If renewable energy is to displace conventional energy in an effort to curb climate change, we cannot accept that investment in renewable energy forecasting development should stall, as it is a key factor in minimizing the cost of integrating variable renewable energy into grid systems and thereby facilitating higher penetration levels.

In the following we will therefore try to shed some light into the pitfalls that have been observed over the past years and provide answers to the most important questions on how to choose the forecasting solution that most effectively solves the problems of an individual end-user.

II. Purpose of benchmarks and Trials

Even though forecasting benchmarks and especially trials began to be employed by forecast users more than a decade ago, it has only become a routine industry practice during the past 5 years. In the early days of the new millennium trials were often conducted to ensure that the forecasting vendor was in fact able to deliver forecasts in real-time with a reliability that was expected and required for the real-time applications of the system operators. In subsequent years, a number of countries introduced trading of wind and solar power by the facility owners or corresponding balance responsible parties (BRP) at the power exchanges. Denmark was the first to "privatize" the trading of wind power by introducing a bonus system for the imbalances that now were handled by the BRPs and no longer the system operator. Spain and Germany followed that principle in a slightly different way. Nevertheless, the moving the trading and balancing to the BRPs, in the sudden opening of opportunities for forecast vendors to enter the forecasting market. While there was far less doubt in the quality of the forecasting and a natural expectation that any vendor winning a contract would improve over time due to more information being collected, it was reliability in supply of the forecast that had a high focus. To the authors knowledge, the obligation of Germany's energy commissioner to use at least 2-3 forecasts in order to prevent speculation against the system operators in the market, was unique when introduced in the renewable energy law in 2005. Nevertheless, it was at that time that business opportunities opened for forecast vendors to serve this market. With the introduction of a market bonus system in 2012, the amount of forecasting service business in the German market rose to new heights. With more vendors entering the business however, some BRPs found that it was better to employ more than one vendor in order to ensure (1) reliable forecast supply and (2) higher accuracy. It was immediately recognized that the blending forecasts, even by just computing an average of two uncorrelated forecasts, can increase the accuracy of the forecast. As the amount of forecasting clients and vendors increased, testing forecasts was enforced by the vendors entering the market, but also by the forecast clients having difficulties evaluating the different approaches. For many, it seemed easiest to consume forecasts in real-time and compare them or blend them with their running system in order to test the value of a new forecast.

Typically benchmarks and trials attempt to assess the relative accuracy of forecasts according to some pre-defined metric (e.g. MAE, RMSE, economic value). However, they should be attempting to define the true value not just the accuracy as measured by an arbitrary metric. In the next section we will therefore analyze and discuss the difference of using benchmarks or trials together with or versus using only request for proposals (RFP) or request for information (RFI).

III. Analysis of RFP/RFI and benchmark/trial processes

In business practices of all kinds, RFPs and RFIs have become the predominant way of selecting vendors or consultants to deliver specific products or to carry out specific tasks for an organization. Dependent on the size of the task, such RFP processes are more or less cumbersome. If the target is a product, it is often relatively straight forward to evaluate the quality by requesting some samples. When purchasing services that depend on many other factors, the evaluation process is no longer trivial and often requires a substantial amount of resources in order to find the best solution.

As noted in the previous section benchmarks and trials initially entered the power industry to test reliability and in some cases to find out the level of performance that can be expected from a forecast solution for a specific area and application, where there were no similar known cases. With changing practices and requirements, benchmarks and trials have entered into the industry practices before and supplementary to RFPs. There are a number of factors responsible for that have lead to this change in practice, the two most common are (1) increasing competition and (2) difficulties to evaluate methods and quality outside the operational environment.

For simplicity, we do not mix the hybrid of carrying out an RFP in conjunction with a benchmark into the following discussion, even though this practice has increased recently.

Both, the RFP/RFI and benchmark/trial method have advantages and disadvantages. We have collected the most obvious and important points in the following lists. Note, that we consider benchmarks and trials with the objective of entering into a business relationship between the operator of the benchmark/trial and the vendor. There are of course also benchmarks that have the sole objective to compare approaches and methodologies on a scientific level. These types of benchmarks do not target any specific end-user or business application. For the purpose of simplicity these types of benchmarks are not considered here.

Major advantages of RFP/RFIs:

- clear structure of the required internal and external processes
- involvement of all relevant departments
- cross-departmental evaluation
- evaluation on different criteria is directly comparable
- reliability, service and quality of vendor can be verified through references
- real requirements are defined and internal processes established
- vendor's methods can be compared and verified on future compatibility
- Confidentiality and Buyer Investment can be trusted
- Pre-Qualification possible

Major disadvantages of RFP/RFIs:

- expensive process for the organization
- quality of forecast in real-time cannot be tested
- Complex system design can reduce amount of possible vendors

Major advantages of benchmarks and trials:

- reliability and real-time quality can be tested
- methods can be compared in a real-time environment
- cost savings by testing standard services
- process more simple, if only a fraction is tested
- many vendors can be invited

Major disadvantages of benchmarks and trials:

- evaluation in test mode difficult, one extreme event can decide on the winner without being statistically significant over longer time
- real-time setup for a full-scale test is expensive
- impossible to setup a representative test environment to evaluate real-time performance

- Can become a long process to capture representative period and have all vendors deliver
- No information control - often anonymity is in the way of discussing errors or difficult cases in the test period
- No commitment by the vendors on reliability, if it's on a no-cost basis
- Deliberate delays from vendors to spoil process
- No price transparency in simplified test environments
- Usually low requirement level by testing standard services
- many vendors require a lot of resources and increase costs

Even though this is probably only a selection of advantages and disadvantages of the two methodologies in a procurement situation, it provides an overview of the major difficulties associated with each approach. It can also be seen that some advantages are also associated with disadvantages. This is particularly so in the cases, where a simplified service is tested. Such a simplification can be a cost saving, but only, if a pre-qualification has been taking place. Otherwise a large amount of participants can easily increase the costs and resources over the savings.

In order to ensure that a selected solution also fulfills the future requirements of an organisation, it is important to note that benchmarks and trials by themselves often lack the possibility of a structured analysis and the development of an optimal solution for the specific problem to be solved for the target application.

Saying this, there are also established processes and practices that indeed are so well-defined on all aspects of forecasting and implementation that carrying out a benchmark exercise or a trial is answering the questions that are relevant for the end-user to choose the best solution.

IV. Decision making: what's the best way forward ?

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 a number of departments involved in the decision-making process. A relatively straight forward way to decide for or against a trial or benchmark exercise is to use a decision support tool. Figure 1 shows an example of a decision support tool that can help on high level decisions.

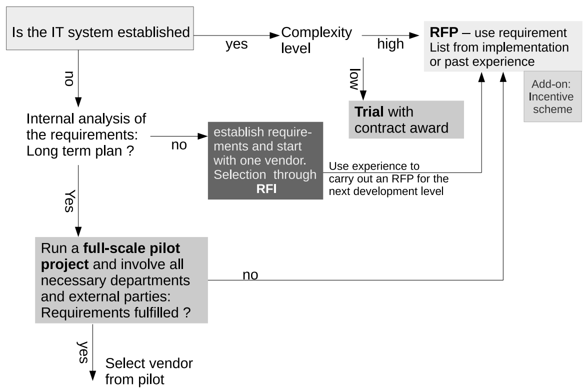

Fig. 1. A simple decision support tool for choosing a path for implementation of forecasting solutions.

The Trial Trilemma

Constructing a proper forecasting trial forces the extremely difficult task of satisfying three needs, none of which can be omitted. First a trial must be considered *fair* – by which is meant unbiased and standardized. If any training data is offered by the trial operator, the same set of training data should be offered to all participants at the same time. Further, each participant should have access to the same real-time operator data as all other participants and at the same frequency. The trial operator should not choose a forecasting site or set of sites for which he has received any previous forecasting data from any of the participants. In other words, if the trial operator has provided any past data support or feedback to a participant on a particular site, he has unwittingly rigged the trial for that participant, despite efforts for fairness within the trial itself.

Secondly, the trial needs to be *diverse* – by which it should measure if performance is extensible to more than one site or wind climate and will be sufficient. It is important that the trial operator, where possible, test performance of participants at more than one site, in diverse geographies or wind climates. As an example, two sites for which balancing costs demonstrate maximum sensitivity to forecast error may be chosen as prioritized sites. In this way, performance is measured for its sufficiency in solving an operator's problem. Another useful way of measuring extendibility is to test performance at a single site in different seasons. If a time lapse is involved, it would be more practical, if the operator collects back-cast data from participants. Such a trial in real-time would be overly time consuming. The drawback and pitfall for the operator when deciding for back-cast data is that this creates the possibility of implicit cheating, i.e. using the actual weather conditions to estimate what is happening at the forecast location, even if the forecaster doesn't have the actual outcome data for the facility.

Thirdly, the trial needs to be *time limited* - or fast. In most current trials, forecasters must "pay to play" – as contracts are only given to users who provide trial participation or data to the user at no cost. Such a situation consumes valuable human and machine resources that could otherwise be used for advancing modeling efforts and serving other customers. As such, a trial should not continue indefinitely. It should be time-ordered, with proper milestones and deadlines, and such a time line should be provided to trial participants to support a decision to participate. There should be a deadline for the final evaluation to be completed and for a contract negotiation to commence. Trials should conclude with a contract offered or otherwise clearly indicated that this is not the purpose of the trial.

V. Pre-trial/benchmark Questions for End Users

There are three main elements that a user of a forecast service looks for in selecting a forecast provider: (1)

Accuracy, (2) Price, and (3) Ease of use. A benchmark normally addresses accuracy and price, but not necessarily how easy the forecast service is to use (e.g., graphical tools, retrieving historical data). This is especially the case in which many forecasts are being evaluated for accuracy and price is being compared. Unfortunately, it is not uncommon for the individual(s) conducting the benchmark to not be the end user. For example, a wind asset manager may task a data or IT analyst to conduct a forecast trial, but the end user is a power marketer. Several qualifying questions should be asked ahead of conducting a benchmark which will help determine the scope, which in turn affects the required resources.

This would typically be (1) Accuracy, (2) Price, and (3) Ease of use. Depending how things are defined, there are at least 3 more factors that should be considered: (4) customer service, which asks whether the provider is responsive when issues arise and will the provider give live support in critical situations, (5) reliability and experience and (6) the ability to maintain state-of-the-art performance. The benchmark only provides a snapshot of performance at the present time, but does not show, whether the provider engages in ongoing method refinement/development and forecast improvement activities.

Other aspects and questions to be asked at the outset can be summarized to:

1) Are forecast horizons of less than 6 hours operationally important? If the answer here is ”no”, establishing a live data feed may not be necessary. This is usually a time consuming component of the benchmark for the operator.
2) Will the benchmark take place during a windy or cloudy period? The trial operator wants to make sure the answer here is ”Yes” to insure the sample size of harder to forecast events is sufficient. If the answer here is ”No”, trial operator should strongly consider doing a retrospective forecast (also known as ”back-cast”), if possible.
3) Do I have enough historical on-site observation data to feed the forecast provider’s statistical methods? If the answer is ”No”, operator might consider another location.
4) Is the benchmark location representative from a wind-climatology perspective of what operator will require contractually?
5) Are the metrics that I’ll be using to evaluate the forecasts meaningful to the bottom line of my project or to the operational reliability?

VI. Best Practices and Worst Cases

A. *What characterizes a good benchmark/trial ?*

Although there are many different ways that trials and benchmarks have been conducted, there have been some common threads of successful trials that provide the operator with the best forecast solution and the participants either with winning the business or knowledge of where their forecast ranks amongst the competition. Here are a few:

- Winning criteria are spelled out by end-user
- Accuracy metrics are defined in advance
- Time line is clear
- Question and answer period before the benchmark period begins
- A week for testing the transfer of data between participant and operator
- Frequent communication to all participants regarding any changes or answers to questions that arise
- Forecast output example file sent to participants
- Same historical and project metadata is provided to all participants
- independent evaluation
- there is a budget and sufficient resources allocated

Apart from these threads, it has been observed that the last thread, the budget and allocated resources are usually key to successful projects. In fact, if the operator of the benchmark or trial is looking upon the task as a project, allocating resources, time and structure, the chances are that the outcome is to the benefit of all.

Cases of Excellence

One of the best known wind power forecasting benchmarking exercises is the AESO pilot project [1], carried out by the Alberta Electric System Operator. The project was setup as a 1-year benchmarking of 3 forecast vendors that were selected via an RFP. There was a project steering team consisting of the client (AESO), an independent consultant for the evaluation of the forecast data, a group of stakeholders, representing the wind farms in the control area. The project had a project manager that followed the project closely, stayed in contact with all vendors and collected examples, where the forecasts failed or did very well in more or less extreme conditions, typically for the area. These events were analyzed in great detail, where the vendors had to explain in detail why their methodology did well or failed in the respective case. These events were also documented. At the end of the project a report from the independent consultant was provided together with a report from each provider, analyzing their own methodology and describing the experience as well as their understanding of what can be achieved in Alberta in relation to forecast quality and performance.

Two more recent examples of well designed and executed forecasting benchmark projects are the solar forecasting trials conducted by the Electric Power Research Institute (EPRI) on behalf of two individual utilities. Both projects were structured in a similar manner although the later project implemented some refinements based on what was learned in the earlier project. The first project epri2015 evaluated forecasts for solar generation facilities on the CPS Energy system in Texas . The second project epri2017 assessed the performance of forecasts for solar facilities that supply power to the Southern Company, The structure of these projects was different from that of the AESO pilot project in several ways. First the trial was designed and operated by an independent 3rd party, not the ultimate forecast user. Second, a system was implemented that enabled each forecast provider to be anonymous to all other entities associated with the trial including the trial operator (EPRI) and the ultimate forecast users (CPS Energy and Southern Company). At the end of the trial the potential users had the option of using the performance results to contact one or more providers via an anonymous ID to pursue negotiations for potential services. Third, the trials also had a substantial research component with EPRI conducting an extensive analysis of forecast error patterns.

Typical Pitfalls

Based on over a decade of industry experience in benchmark and trial participation, there are several recurring pitfalls which a trial end-user (and participant) could easily avoid thus minimizing wasted resources. We highlight a few here that are the most common.

- Incomplete details - often important details are missing that can cause confusion or delays (e.g., plant capacity constraints, time zone of the historical training data).
- Non representative design - oftentimes the benchmark operator has a strict deadline to finish a trial and thus may choose a period that is not representative of the full range of conditions that will be encountered in the target application (e.g. too short or during a season in which forecasting systems aren't tested under more difficult-to-predict weather regimes).
- Poor communication - up front and frequent communication to all participants is vital to efficiently executed and productive benchmarks.
- Proper time resource allocation - benchmarks can be a labor intensive undertaking for the operator and, if time is not allocated adequately, short cuts and less representative outcomes might be the result.
- Over-engineering the problem - there have been many instances in which a benchmark operator goes through a lot more effort than is required to get the answer to a simpler problem. The most common example is the configuration of a live datafeed for a benchmark in which only the day-ahead forecast performance is being evaluated.

The IEA Task 36 Work Package 2.1 team will produce a document and publicly accessible web site with template forms that, when completed, will provide the guidance that can save the operator and participant time and result in an outcome that provides a more optimal solution for the target application.

VII. The difficulty of fair evaluation

For even the fairest, most diverse, and deterministic trials, trial evaluation is deceptively difficult. Even in cases where standardized forecast error metrics like mean absolute percent error (MAPE) are used as the primary quality metric, the sustainability of the forecast quality of the most qualified participant (or trial victor) cannot be guaranteed beyond the trial period itself. This was proven by setting up a simple trial experiment, in which three independent model solutions, with unique parameterizations and initial and boundary condition data, were used to represent three unique trial participants. Forecasts were provided for three actual sites, each separated by greater than 2000 km, where participants were expected to forecast wind power from each facility over a one month period assuming full availability. In order to test various assumptions made by the trial operator, we allowed the independent forecasts to be run for a subsequent 12-month period, unchanged from their initial states during the one-month trial period.

Figure 2 shows graphically the vendor with lowest MAPE per site per month for the initial trial month and for the subsequent 12 months comprising a typical contract term. Surprisingly, timing matters. In the trial month, Vendor 1 exhibited lowest error and the greatest sufficiency of performance across the test sites, but if the trial had been delayed one or two months, Vendor 3 would have been chosen for out-performing its competitors by the same standards. If delayed nine months, Vendor 2 would have been selected. For this portfolio, trial selection was repeatable less than 50% of the time. Interestingly, trial selection was repeatable 75-80% of the time for a single site. This finding supports the axiom that having more than one forecast solution maximizes the value of forecasting generally.

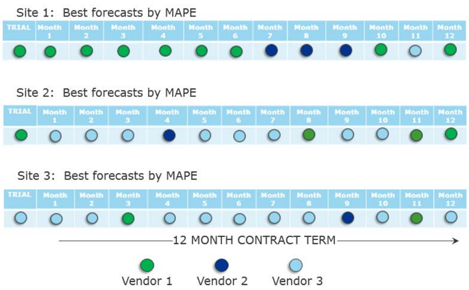

Fig. 2. Table indicating trial participant with lowest MAPE per site per month. The Trial month is given in the far left column.

In a second experiment, we measured the effect of trial duration on outcome, using the same set of sites and identical "participants". As shown in Figure 3, for an initial trial month 0 used in the original experiment (represented analogously in 2), extension of the trial by a single month changes the trial outcome entirely. In the original experiment, Vendor 1 out-performed other vendors on MAPE and on the sufficiency of performance across all sites. However, in the second experiment, Vendor 3 out-performed other vendors on the same standards when the trial period was extended by an additional 30 days. Furthermore, the trial selection at the end of two consecutive months demonstrated a doubling of repeatability for the portfolio. While this limited result is encouraging for the reliability of trial outcome, a trial of two months' duration bears considerable costs to participants and delays contracting decisions for the forecast user, both of which should be carefully considered. It is worth noting that trial duration has little effect on the repeatability or reliability of the trial outcome for a single site. Indeed, Vendor 3 would not have been the obvious selection for site 1, for a trial of either 1-month or 2-months duration.

Another question that could be asked in this respect is whether the trial's outcome can at all be defined as statistically significant or not. From a theoretical perspective, the outcome of cases with statistical significant should lead to a somewhat higher repeatability. The statistical significance could occur in different ways - a large difference in a small sample or a smaller difference in a larger sample. However, to decide that the outcome should have some measure of how meaningful the differences are. Implicitly this would lead to the result that a larger sample is required in order to have a higher probability of a statistically significant difference - and if you can't get that the outcome can really not be considered to provide a meaningful difference among the top providers, leading us back to the "trilemma" of benchmarks and trials.

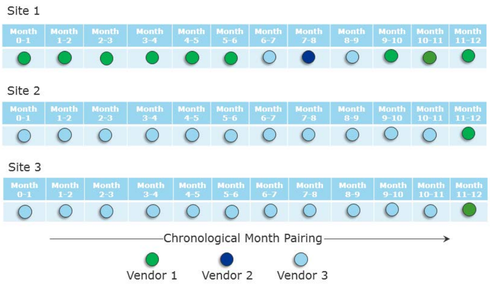

Fig. 3. Table indicating trial participant with lowest MAPE per site per consecutive month pair.

VIII. How to benchmark the next generation forecasting applications

Until now, most of our discussion has assumed that the end-user deals with one deterministic forecast, or maybe a number of deterministic forecasts. However, the trend in the industry with increased penetration levels of VER, especially the mixture of wind power and solar power calls for new methodologies including uncertainty forecasts from probabilistic forecasting tools. When dealing with probabilistic forecasts, the concept of multiple suppliers becomes redundant and in the best case very difficult to implement. Due to the nature of probabilistic forecasts that cannot just be added up and averaged, any mixing of probability distributions are no longer possible. This leads to a new dilemma in the decision making process of current business solutions, as many business solutions today are built up upon multiple providers, where the individual provider may not need to have highest scores on all aspects of the forecast process such as performance, service quality, reliability etc. On the other hand, it opens up a lot of opportunities to streamline processes, if one provider delivers the consistent weather related input to all related processes in an organization.

To this date, there is no best practice that could be recommended, except for that end-users should revert back to their own challenges and requirements when designing their next generation forecasting system, rather than trying to "do what every body else is doing". For probabilistic forecasting to enter the business practices in the power industry a major paradigm shift is required. How this will impact the use of benchmarks and trial is unknown and little predictable at this stage. The IEA task 36 wind power forecasting[1] is dealing with all aspects of probabilistic forecasting in work package 3 and may shed some light into the unknown (see also Bessa et al, [4]).

IX. Conclusion

The IEA Task 36 Work Package 2.1 team is developing an open source compendium that aims to inform renewable energy benchmark and trial operators on best practices. This includes specific recommendations and guidelines to insure that the operator receives the best value forecast service to minimize their integration cost of renewables. The forecast providers, in turn, will have meaningful feedback on their standing amongst the competition so that improved and meaningful forecast accuracy standards are pursued. This goal is different than the current practice of chasing a metric that may not give the forecast user the solution that yields the least cost (or maximum revenue) on power marketing and plant/grid operations.

We have shown that the selection of an optimal forecasting solution for a specific target application is a difficult task that encompasses a number of complex factors that, if improperly assessed, can result in a much less than optimal solution. There are five fundamental factors that are typically critical to the selection of an optimal solution: (1) current accuracy, (2) cost, (3) ease of use, (4) customer service, trying and support, (5) commitment to maintain a state-of-the-art solution

Of the critical factors the current accuracy of a specific solution is the most difficult to reliably assess since it has a high degree of variability associated with location, wind climate regime and other factors. Benchmarks or trials are a frequently used approach to assess forecast accuracy, but they must be intelligently and carefully designed, executed and interpreted in order to yield meaningful results that are appropriate for input into a decision making process.

To conclude, this paper provides specific examples of best practices and pitfalls to avoid based on years of industry benchmark and trial experience. While no two benchmarks or trials are exactly alike, we highlight the common elements of successfully run exercises while pointing out inefficiencies and poor design decisions to avoid.

Acknowledgment

This work is a result of the international collaboration IEA Wind Task 36 Forecasting for Wind Power. The authors would also like to thank Jakob Messner, Gregor Giebel, Justin Sharp, Aidan Tuohy, Iben Holfort Voxen for their input to the project and valuable discussions. The work of Corinna Möhrlen is supported by the Danish EUDP project *Wind Task 36 Forecasting Danish Consortium* under the contract no. 2015-II-499149.

References

[1] M. Hu, 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.

[2] 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

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

[4] R. Bessa, C. Möhrlen, V. Fundel, M. Siefert, J. Browell, S. Haglund El Gaidi, B.-M. Hodge, U. Cali, G. Kariniotakis (2017), *Towards Improved Understanding of the Applicability of Uncertainty Forecasts in the Electric Power Industry*, Energies 2017, 10(9), 1402; doi:10.3390/en10091402, 2017.

[1]see http://www.ieawindforecasting.dk