



IEA Wind Task 36: Forecasting for Wind Energy Workpackage 2.1



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

Wind Integration Workshop 2017
Grid Operation Tools Session 8c
Berlin, 28th October 2017



Corinna Möhrten, John Zack, Craig Collier, Jeff Lerner, Aidan Tuohy

Background of this investigation: IEA Task 36: Forecasting for Wind Energy

Task Objective is to encourage improvements in:

- weather prediction
- power conversion
- use of forecasts

Task Organisation is to encourage international collaboration between:

- Research organisations and projects
- Forecast providers
- Policy Makers
- End-users and stakeholders

Task Work is divided into 3 work packages:

WP1: Weather Prediction Improvements inclusive data assimilation

WP2: Development of a benchmarking platform & best practice guidelines

WP3: Communication of best practice in the use of wind power forecasts

Follow us on our webpage: www.ieawindforecasting.dk





Background of this investigation: IEA Task 36: Forecasting for Wind Energy

Task 2.1 Best Practice

Design of benchmark exercises: best practice.

Aim: Develop an IEA Recommended Practice on forecast benchmarking processes.

The objective of Task 2.1 is to compile a **guidance and a standard for private industry, academics and government in executing a renewable energy forecasting benchmark or trial**. Benchmark and trial exercises can consume a lot of time both by the company conducting it and by the participating forecast providers. **These guidelines and best practices will be based on many years of industry experience** and are intended to achieve maximum benefit and efficiency for all parties involved in the benchmark or trial exercise.

Follow us on our webpage: www.ieawindforecasting.dk
Task page: <http://www.ieawindforecasting.dk/topics/workpackage-2/task-2-1>





Introduction to the work



Difficult task



Long process



expensive

Complex factors

"Trial Trilemma"

fair

diverse

Time limited



Pre -trial /benchmark questions for end users

What most forecast users of a forecast look for:

- (1) Accuracy
- (2) Price
- (3) Ease of use



Is there not something missing ...??

In addition to the above, the following factors should also be considered:

- (4) **Experience and Reliability** (SLAs)
- (5) **Customer service** (e.g. responsiveness, live support)
- (6) **Ability to maintain state-of-the-art performance.**

Be aware of:

A **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.



iea wind



AWS TRUEPOWER





Definitions for Renewable Energy Forecasts*

RFI – Request for information

Collection of written information about vendor’s capabilities

RFP – Request for Proposals

proposal, often made through a bidding process in procurement of a (forecasting) service

Benchmark – forecast exercise

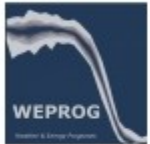
Exercise to determine the features and quality of forecasts with multiple participants

Trial - forecast test

Test the features and quality of a renewable energy forecast of one or more participants for commercial purposes. A trial is a subset of a Renewable Energy Forecast Benchmark.

Live Trial: forecasts are delivered in an operational (automated) mode

Retrospective/backcast Trial: forecasts are provided for a period in the past



AWS TRUEPOWER



* last slide provides long version of definitions



RFP/RFIs versus Benchmark/Trial - a brief history in time -

RFI/RFPs are the most common way to consult services in competition

Wind energy industry's early days (2005+) trials were made due to:

- (1) test new vendor's capabilities to deliver reliably in real-time
- (2) difficult to evaluate methods and quality outside the operational environment

In the past few years **trials** have become very popular in industry, but not so for the forecast vendors due to:

- (A) trials are mostly on the risk of the vendor (no-cost basis)
- (B) often significant resources and expenses are required
- (C) resources are spent on trials rather than development
- (D) less time for real customer support

Questions the WP2.1 group posed:

- why have trials/benchmarks become so popular in industry ?
- are trials/benchmarks superior to RFI/RFP for decision making ?



AWS TRUEPOWER



Comparison of Advantages between RFP/RFI and Benchmark/Trial

Note: for simplicity we do not consider hybrids here



Major advantages of RFP/RFIs:

clear structure of the required internal and external processes

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

evaluation on different criteria is directly comparable

Confidentiality and Buyer Investment can be trusted

Pre-Qualification possible

involvement of all relevant departments

cross-departmental evaluation

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

Comparison of Disadvantages between RFP/RFI and Benchmark/Trial

Note: for simplicity we do not consider hybrids here



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 disadvantages of Benchmark/Trials:

- evaluation in test mode is difficult
- Represent. real-time test environment impossible
- representative period + delivery = long process
- No information control
- No reliability commitment by the vendors
- Deliberate delays from vendors to spoil process
- No price transparency in simplified environments
- Mostly low requirement level for standard services
- Handling many vendors require a lot of resources



iea wind



AWS TRUEPOWER



**How important
is a real-time test for us?
Do we get around a RFP by
doing so, saving time
and money ?
But, what about the length
of a trial
- could that maybe
delay us even more ?**

Hm, so many these days
are carrying out a trial,
i had not expected
that many arguments
against it... still, with a trial
we will see the vendors
performance in real-time ...
It may save cost ..?





iea wind

The difficulty of fair evaluation

For even the fairest, most diverse, deterministic trials with standardized forecast error metrics, **trial evaluation is**

- **deceptively difficult**
- **sustainability cannot be guaranteed beyond the trial period**



AWS TRUEPOWER

2 Public Best Practice and Excellence Examples:

AESO pilot project in 2007-2008

Main results:

- forecasting in Alberta was possible, beneficial for system operation
- None of the 3 forecast providers was best on all scores or on a majority of scores

EPRI trials for CPS and Southern Company in 2015/2016:

Main results:

- Anonymity showed some benefits, but also disadvantages
- No significant difference in performance among the top 3 vendors



The Difficulty of Fair Evaluation: An Experiment

Trial experiment carried out by **DNV-GL***:

Trial 1 Period: 1-Month

- To represent three unique trial participants:
- 3 independent model solutions
- unique parameterizations and initial and boundary condition data
- Forecasts were provided for three actual sites, each separated by > 2000 km
- Forecast target was day-ahead wind power from each facility assuming full availability

Trial 2 Period: 12-month

the independent forecasts were run for a subsequent 12-month period, unchanged from their initial states

* see full experiment outline and results [here](http://ieawindforecasting.dk/publications) or by visiting <http://ieawindforecasting.dk/publications> → C. Collier: Why Do Forecast Trials Often Fail to Answer the Questions for which End-Users Need Answers: A Forecaster's Point of View





The difficulty of fair evaluation: reliability over 1 year

Site 1: Best forecasts by MAPE

TRIAL	Month 1	Month 2	Month 3	Month 4	Month 5	Month 6	Month 7	Month 8	Month 9	Month 10	Month 11	Month 12
	●	●	●	●	●	●	●	●	●	●	●	●

Site 2: Best forecasts by MAPE

TRIAL	Month 1	Month 2	Month 3	Month 4	Month 5	Month 6	Month 7	Month 8	Month 9	Month 10	Month 11	Month 12
	●	○	○	○	○	○	○	○	○	○	○	○

Site 3: Best forecasts by MAPE

TRIAL	Month 1	Month 2	Month 3	Month 4	Month 5	Month 6	Month 7	Month 8	Month 9	Month 10	Month 11	Month 12
	○	○	○	○	○	○	○	○	○	○	○	○

—————12 MONTH CONTRACT TERM—————→

● Vendor 1 ● Vendor 2 ○ Vendor 3

Note:

- The best vendor may change if the forecast was in another month
- Trial selection was repeatable by only 50% for all and 75% per site!
- Looking over a 1-year period, there is no vendor that outperforms on all sites → no significance!



iea wind

The difficulty of fair evaluation with 1-month tests



AWS TRUEPOWER



Site 1

Month 0-1	Month 1-2	Month 2-3	Month 3-4	Month 4-5	Month 5-6	Month 6-7	Month 7-8	Month 8-9	Month 9-10	Month 10-11	Month 11-12
●	●	●	●	●	●	○	●	○	●	●	●

Site 2

Month 0-1	Month 1-2	Month 2-3	Month 3-4	Month 4-5	Month 5-6	Month 6-7	Month 7-8	Month 8-9	Month 9-10	Month 10-11	Month 11-12
○	○	○	○	○	○	○	○	○	○	○	●

Site 3

Month 0-1	Month 1-2	Month 2-3	Month 3-4	Month 4-5	Month 5-6	Month 6-7	Month 7-8	Month 8-9	Month 9-10	Month 10-11	Month 11-12
○	○	○	○	○	○	○	○	○	○	○	●

Chronological Month Pairing →



Note:

- Extending in any period to 2 months increases the repeatability of the result, but...
- Dependent on the choice of the 2-months, the winner may still not be representative
- None of the participants wins at all sites and more than once

Lessons Learned

Selection of the best period to test different vendors is difficult

→ experiment showed that results can change from month to month and site to site

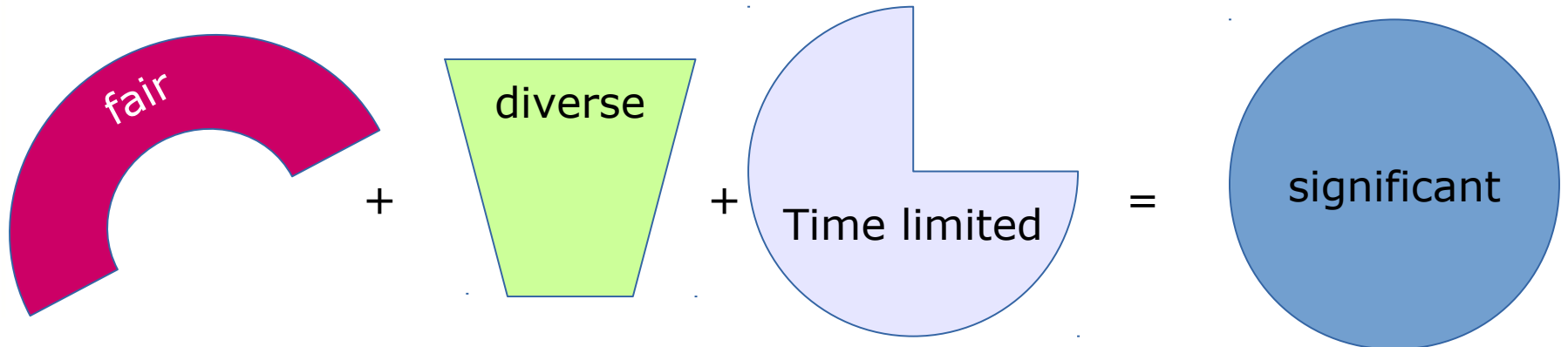
Statistical significance is difficult to gain in a trial

→ mostly the top providers shift rank from month to month

Repeatability is necessary, but difficult to measure

→ climate, technical issues and resources used in a trial can produce unrepresentative results

The "Trial-Trilemma" is a high mountain to climb!



Most Common Pitfalls in Benchmarks and Trials

Incomplete details

Poor communication

Improper time and resource allocation

Over-engineering the problem

Non representative design



The outcome is often truly unrepresentative results that can lead to terrible solutions, not fitting the purpose at all !



iea wind



AWS TRUEPOWER





iea wind

Pre-trial/benchmark questions to avoid such pitfalls

Qualifying questions should be asked ahead of conducting a benchmark to help determine the scope and required resources.

Some examples are:

Are forecast horizons of less than 6 hours operationally important?

Will the benchmark take place during a windy or cloudy period?

Do I have enough historical on-site observation data to feed the forecast provider's statistical methods?

Is the benchmark location representative from a wind climatology perspective of what will be require operationally ?

Are the metrics to evaluate the forecasts meaningful to my project or to the operational reliability?



AWS TRUEPOWER



The next generation forecasting applications

Now that we have discussed pitfalls and practices for deterministic benchmarks and trials....

How about benchmarking probabilistic forecasts ?



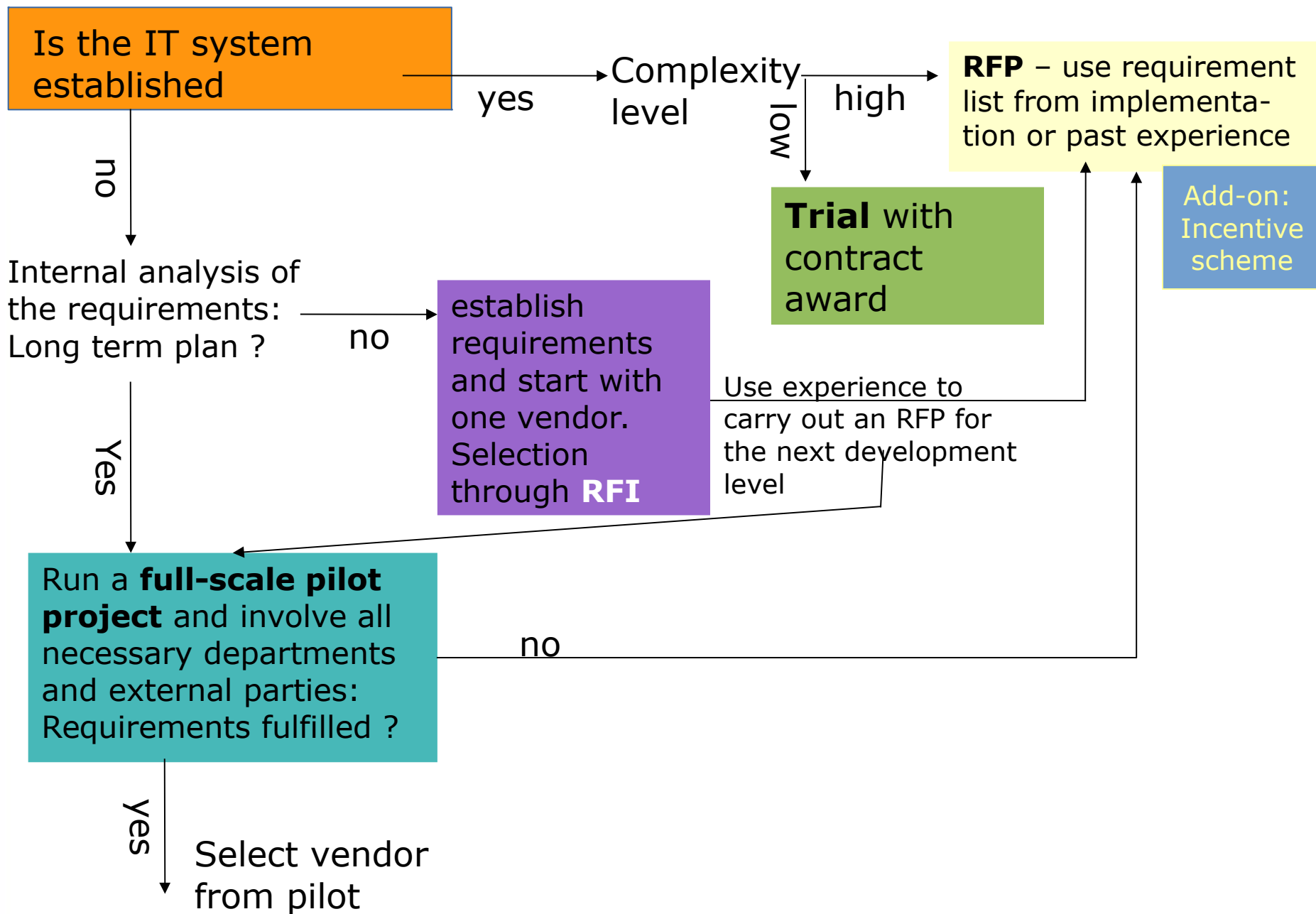
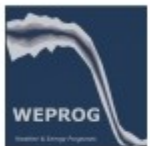
Simple statistical metrics are no longer useful

Mixing probabilistic forecasts is more difficult

Effort for trials even larger for provider and end-user

Maybe what is most needed is a decision support tool that visualizes the impact of our decisions...?

A decision support tool for choosing a path for the implementation of the *best* forecasting solutions.



Wrap up and key take-aways



iea wind



AWS TRUEPOWER



Don't go shopping without a shopping list !



Don't think you can take a shortcut by

→ only setting up a trial/benchmark

(you may end up with a terrible solution!)

→ going out there without knowing what you need

(too many 'interesting' offers to keep focus on what you need)

Analyse in detail what you really need

→ use your experience from current system or trial

→ evaluate the complexity level

Find the key requirements and look for a solution that fulfills these

Think about the future

→ a system solution should be future compatible

→ your service provider should have an incentive to improve over time



A shopping list helps to stay focused



Thank you for your attention !

Discuss with us challenges, experiences & solutions....
and follow our work @ ieawindforecasting.dk

Project webpage

<http://www.ieawindforecasting.dk/>

Task-page:

<http://www.ieawindforecasting.dk/topics/workpackage-2/task-2-1>

Contact:

Task Leaders

Corinna Möhrten: com@weprog.com

Jeff Lerner: Jeffrey.Lerner@vaisala.com

Co-Authors and Task Partners:

John Zack: jzack@awstruepower.com

Craig Collier: craig.collier@dvngl.com

Aidan Tuohy: atuohy@epri.com



Definitions for Renewable Energy Forecasting

RFI – Request for information

is a standard business process whose purpose is to collect written information about the capabilities of various suppliers. Normally it follows a format that can be used for comparative purposes.

RFP – Request for Proposals

is a document that solicits proposal, often made through a bidding process, by an agency or company interested in procurement of a (forecasting) service to potential suppliers to submit business proposals.

Benchmark – forecast exercise

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.

Trial – forecast test

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.

Realtime or Live Trial: most common form of a trial whereby forecasts are delivered in an operational (automated) mode from one or more forecast providers.

Retrospective/backcast Trial: less common form of a trial whereby forecasts for a period in the past are provided.



iea wind



AWS TRUEPOWER

