IEA Wind Task 36: Forecasting for Wind Energy
Workpackage 2.1

RECOMMENDED PRACTICES FOR THE IMPLEMENTATION OF WIND POWER FORECASTING SOLUTIONS

Part 1: Forecast Solution Selection Process
Part 2: Designing and Executing Benchmarks
Part 3: Evaluation of Forecast Solutions

Wind Integration Workshop 2018
Forecasting Session 2c

Stockholm, 18th October 2018

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

Version 8

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

Task Objective is to encourage improvements in:
1) weather prediction
2) power conversion
3) 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

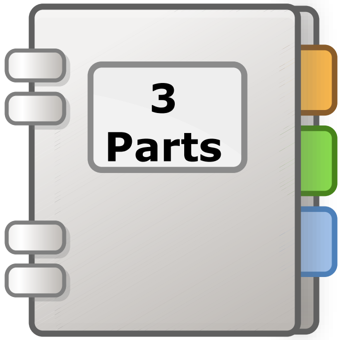
Follow us on our webpage: www.ieawindforecasting.dk
Task page: http://www.ieawindforecasting.dk/topics/workpackage-2/task-2-1

iea wind

About the IEA WIND RECOMMENDED PRACTICE for the implementation of wind power forecasting solutions

Aim: Develop an IEA Wind Recommended Practice

Objective: Compile guidance for the implementation of renewable energy forecasting into system operation.



1) Selection of an Optimal Forecast Solution

2) Design and Execution of Benchmarks and Trials

3) Evaluation of Forecasts and Forecast Solutions

The best practices guidelines are based on many years of industry experience and are intended to achieve maximum benefit for all parties involved in the forecasting area.

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

Starting Point of Solution Selection Process

Establishment of a Requirement List:

Description of the envisioned situation

- current situation
- obtain input from forecast vendor(s)
- list other organizational and statutory requirements

Put as much detail into this part as possible!!!

Engage with the forecast providers

- describe the forecast objective in as much detail as possible
- ask specific questions that are required in the decision process
- ask forecasters to provide information and insights from their experience in other jurisdictions or areas.

Establishment of a Requirement List

Main parts of a requirement list:

1) IT infrastructure

2) Forecast attributes and available methodologies

3) Support and service

4) Performance specifications and incentives

5) Contracting terms

Other secondary aspects to consider:

6) Short-term interim solution

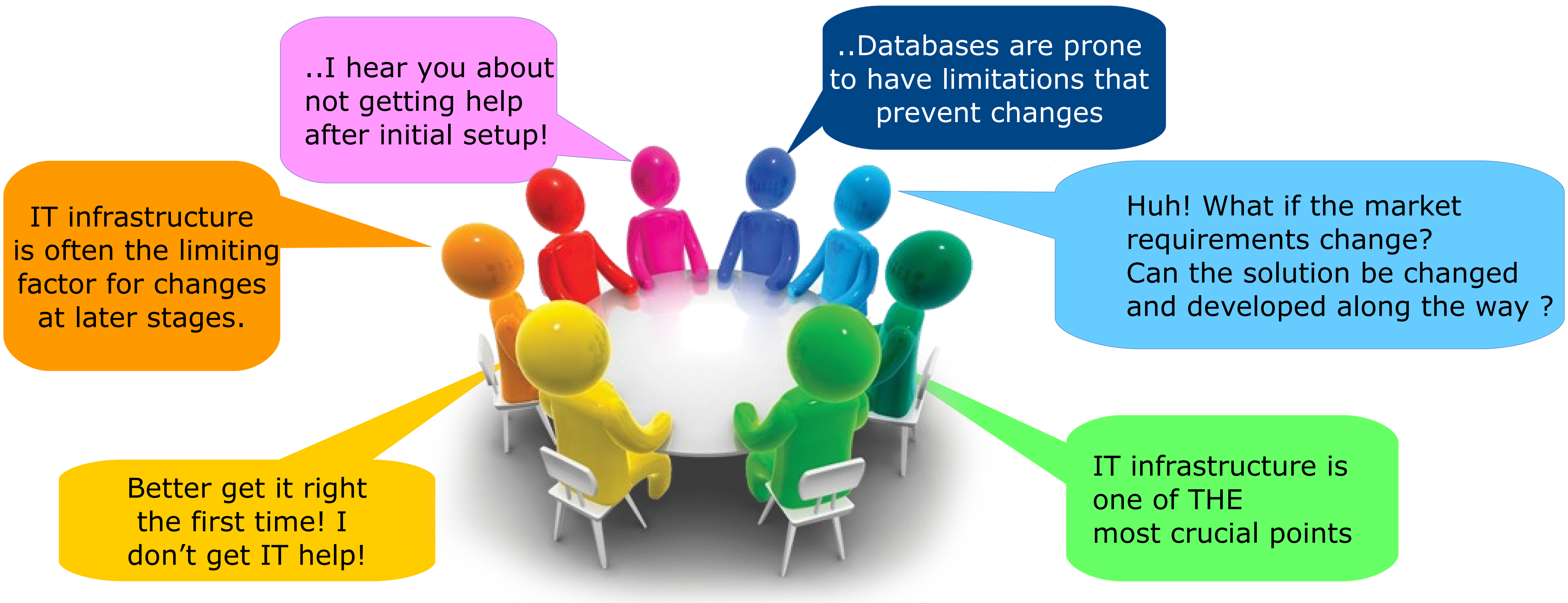
7) Long-term solution

- Involvement of all relevant departments (internal and external parties)
- Analysis of the statutory environment (long-term planning possible ?)
- Establish system requirements
- Pilot maybe used as interim solution

IT infrastructure and forecast solution architecture

Important aspects in the IT infrastructure to be considered are:

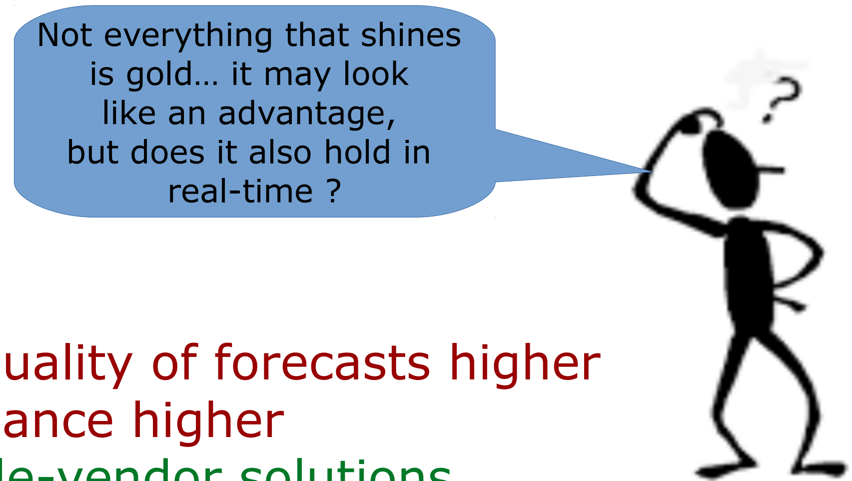
1) Database structure
2) Communication layer
3) Monitoring and error handling
4) Data storage and historic data accessibility
5) Verification and evaluation of forecasts



IT infrastructure and forecast solution architecture

Factors to Consider: Single versus Multiple Vendor Solutions

- Multiple Vendor Solution:
  - infrastructure more complex
  - database requirements are higher
  - need forecast use strategy: blending vs. primary/secondary designation
  - higher total cost
  - possibly higher accuracy
  - possibly higher reliability



- Single Vendor Solution:
  - requirements for reliability and quality of forecasts higher
  - need for monitoring and performance higher
  - less data volume than for multiple-vendor solutions
  - database structure less complex than for multiple-vendor solutions

IT infrastructure and forecast solution architecture

Factors to Consider:
Deterministic versus Probabilistic Forecasts

- **Deterministic Solution**:
  - \- Less information available (uncertainty ignored)
  - \- reduced future compatibility
  - \+ more simple data handling
  - \+ less storage requirements

- **Probabilistic Solution:**
  - \- increased storage requirements
  - \- more complex data handling
  - \+ More information available (forecast variable + uncertainty)

When weighing the pros and cons, look also at your conditions!

Probabilistic forecasts are most beneficial:

- Areas with high penetration (> 30% of energy consumption)
- Areas with high wind speeds (complex terrain, mix of surface types)
- Significant variability to cause strong ramps and high-speed shutdown
- Badly interconnected control zones

iea wind

Evaluation Criteria

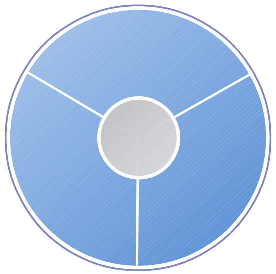
VENDOR INDEPENDENT

Step 1: most important evaluation criteria for a forecast solution to be defined in a tender process:

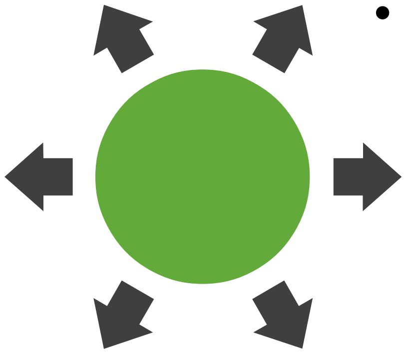
- **Required forecast attributes**
  - Look-ahead period(s) (e.g., hours-, days-, or weeks-ahead)
  - Update frequency
  - Time resolution (forecast interval length)
- **Methodology that fits forecast application**
- **Compliance to technical and contractual requirements**

Evaluation Criteria

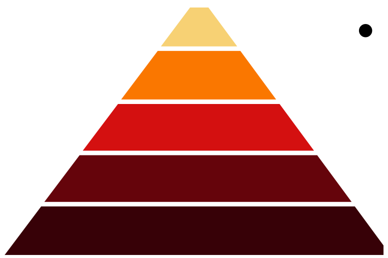
Step 2- most important factors are:



- **Solution Attributes**
  - Deterministic vs. Probabilistic
  - Single vs. multiple forecast providers
  - Forecast Content and Format

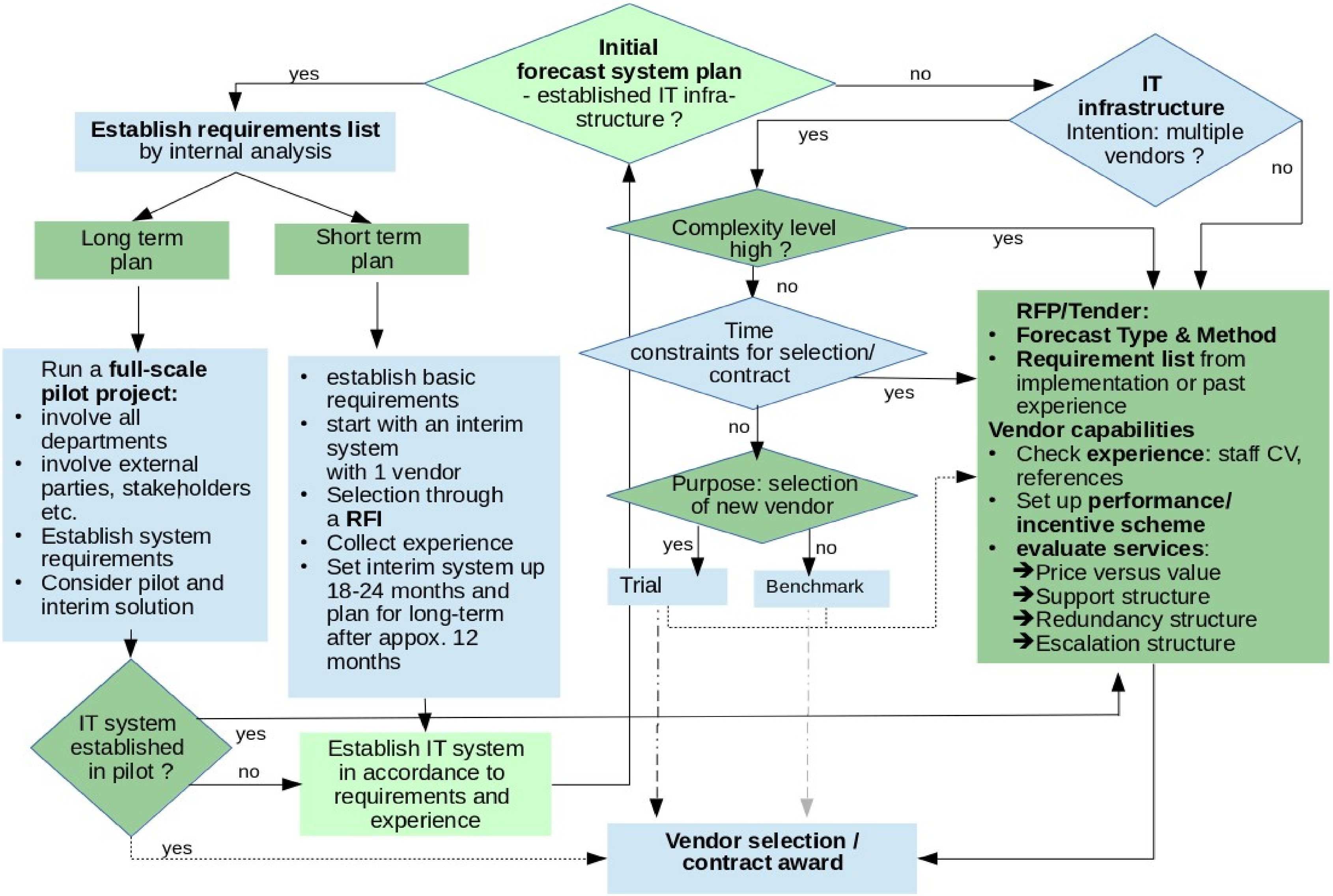


- **Vendor Capabilities**
  - Experience and reliability
  - Ability to maintain state-of-the-art performance
  - Performance incentive Schemes
  - Evaluation of services
  - Price versus Value



- **Vendor Service Structure**
  - Support maintenance service Structure
  - Redundancy Structure
  - Escalation Structure

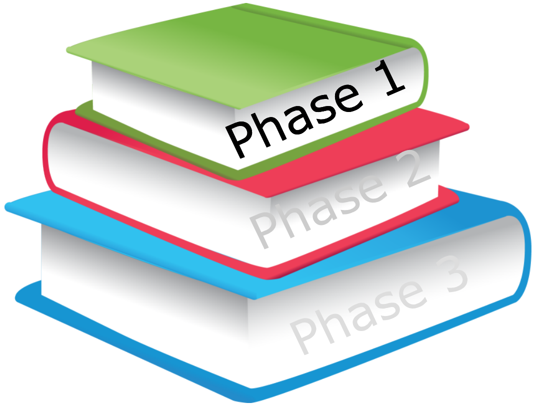
Decision Support Tool for the selection process of a forecasting solution



Takeaway: The most cost effective way to select a forecast solution depends on the user's situation - this DST provides guidance based on the user's situation

Benchmarks and Trials

The 3 Phases of a Benchmarking Process



Preparation Phase:
Determining the Scope and Focus of the Performance Evaluation

Forecast Horizons

Available Historical Data

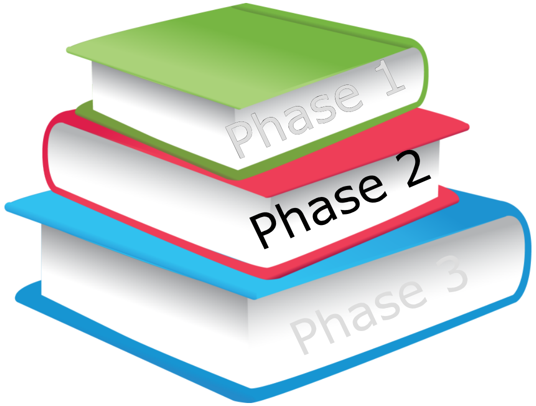
Appropriate length of Benchmark

Are conditions during Benchmark **representative?**

Meaningful evaluation metrics

Think of what is most important, when you make a big or long term purchase (e.g. home, car, forecasting system)?

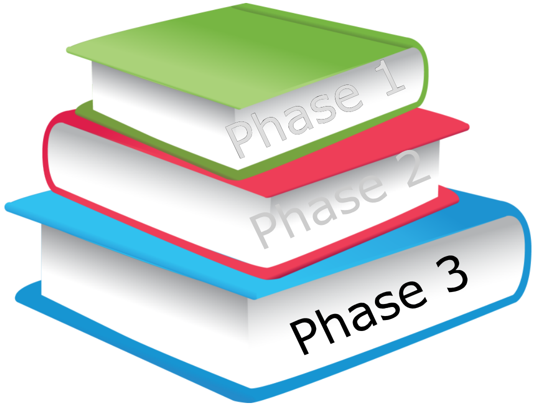
The 3 Phases of a Benchmarking Process



During the Benchmark:
Ensuring a Fair and Representative Process

- Data monitoring (forecasts and observations)
- For fairness and transparency test accuracy and delivery performance.
- Monitor forecast receipt (reliability)
- Sample should be normalized (all forecasters evaluated for same period)
- Develop and refine your own evaluation scripts

The 3 Phases of a Benchmarking Process



Evaluation Phase:
Compiling a Comprehensive and Relevant Assessment

- **Critical Evaluation Criteria:**
  - Application-relevant accuracy of the forecasts
  - Performance in the timely delivery of forecasts
  - Ease of working with the forecast provider

Benchmark and Trial Pitfalls to Avoid

Poor communication

E.g. All FSPs should receive the same information. Answers to questions should be shared with all FSPs.

Unreliable validation results

E.g. Comparing forecasts from two different power plants or from different time periods.

Bad design

E.g. One month trial length during a low-wind month. No on-site observations shared with forecast providers. Hour ahead forecasts initiated from once a day data update.

Details missing or not communicated

E.g. time zone changes, whether data is interval beginning or ending, plant capacity of historical data differs from present.

Allowing the possibility of cheating or "gaming" the process

E.g. no check on forecast delivery time allowing "after the fact" forecast delivery; no penalty for missing forecasts allowing providers to skip difficult forecast situations

Evaluation of Forecasts and Forecast Solutions

EVALUATION OF FORECASTS AND FORECAST SOLUTIONS

Most crucial requirements for an evaluation are:

1) Fairness

2) Transparency

3) Representativeness

(significance and repeatability)

Verification has an inherent uncertainty that comes from the selection/size of data set to be Verified…

Beware: every evaluation or verification is highly subjective!!

Average errors are mostly too simplistic to reflect the quality and value of a forecast solution for the user's specific applications

Assessment of a forecast system

A forecast with **high skill** can have **no value**

A forecast with **poor skill** can be **valuable**

Objective Measures: *Metrics*
- MAE|RMSE|BIAS|StDV…
- Brior Skill Scores, Rank Histograms
- Spread skill, CRPS…

Needs: Correct Selection for purpose

Subjective Measures:
- Goodness-of-fit for purpose
- Solves end-users problem
- Reduces costs
- Increases confidence

Needs: Evaluation Report

SKILL
A forecast has skill, if it predicts the observed conditions well according to some objective criteria

VALUE
A forecast has value, if it helps the user to make better decisions than without knowledge of the forecast

Forecast Verification

Forecast verification is the practice of comparing forecasts to observations.

Forecast verification serves to
1) **monitor forecast quality**
2) **compare the quality** of different forecasting systems
3) is a first step towards **forecast improvement**

Visual inspection
→ Rapid identificatin of periods of poor performance or some types of systematic error

- Average error metrics
- Dichotomous (yes/no) evaluation

--→ beware of the "forecasters dilemma"

... in a hit rate scanario, the forecaster is extremely successful, when always predicting an "event", unless, hit rate <u>and</u> false alarm are verified!

Significance of Results

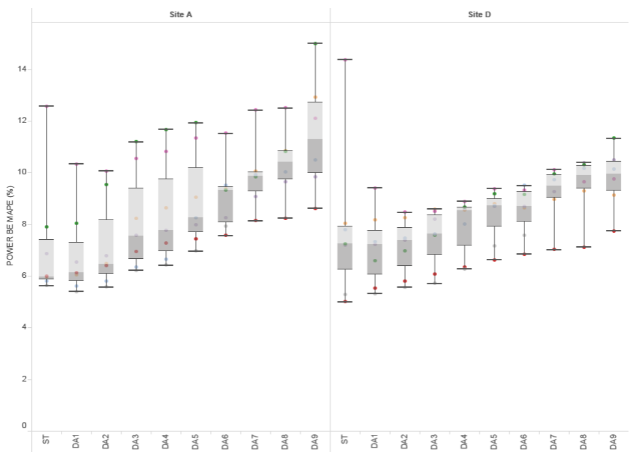
Take uncertainty into account when verification is used to rank forecasts

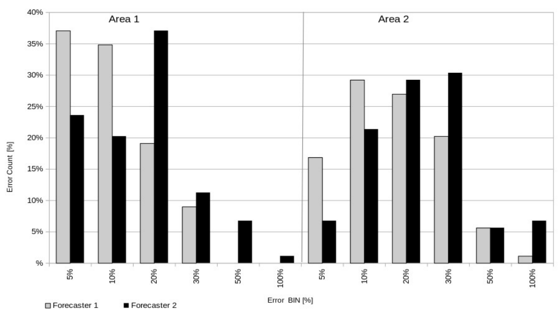
Use metrics that operate on the individual error measures BEFORE averaging

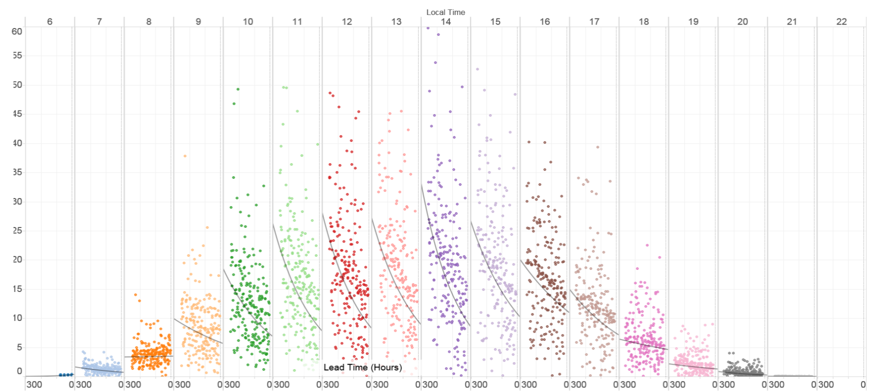
- Parametric test framework to estimate the significance of score differences.
- Non-parametric bootstrapping method to compare error structures

Develop a set of measures to ANALYSE the errors

- Frequency distributions
- Box- and Wiskers Plots
- Scatterplots of errors at different time scales or forecast horizons

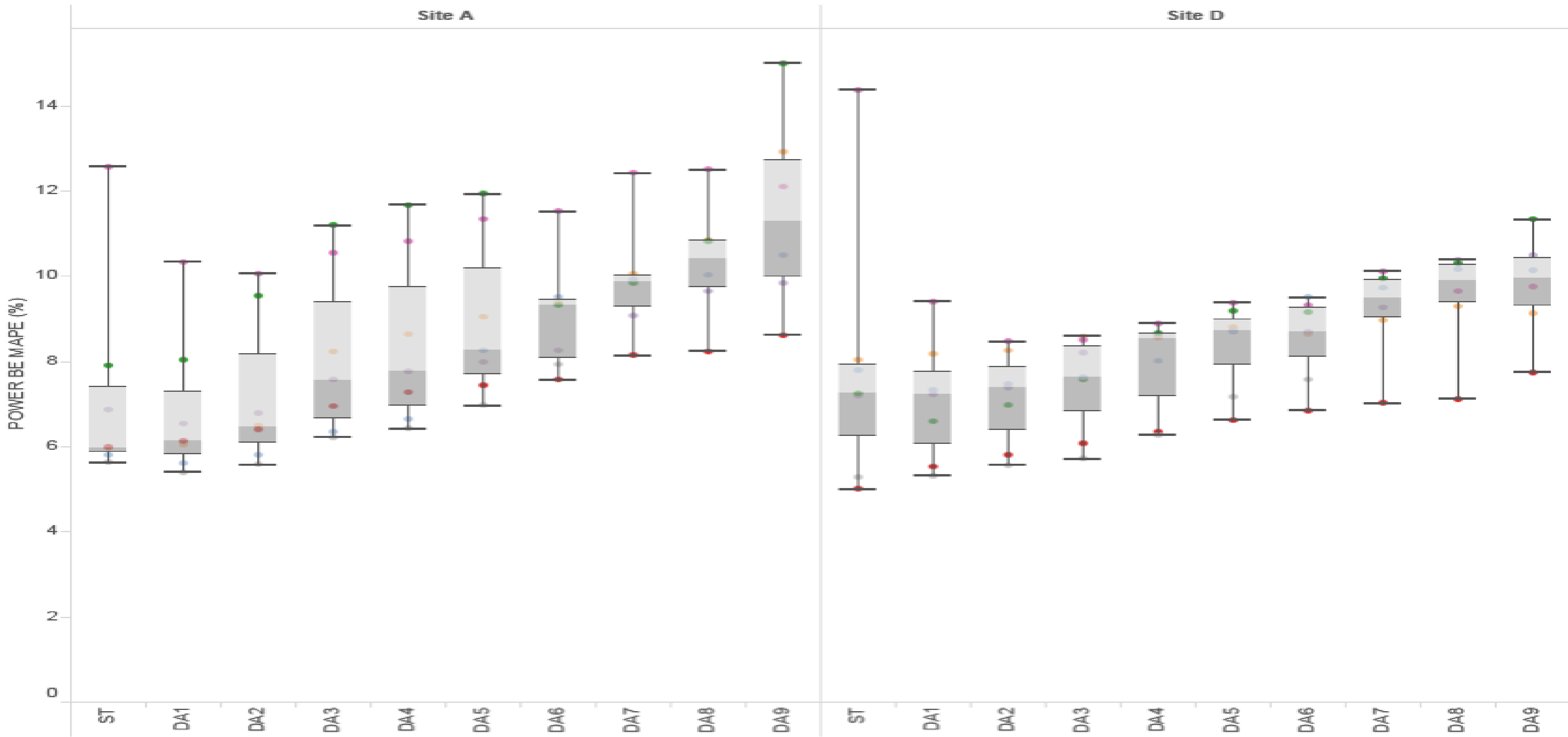






Significance of Results – Looking at Box plots

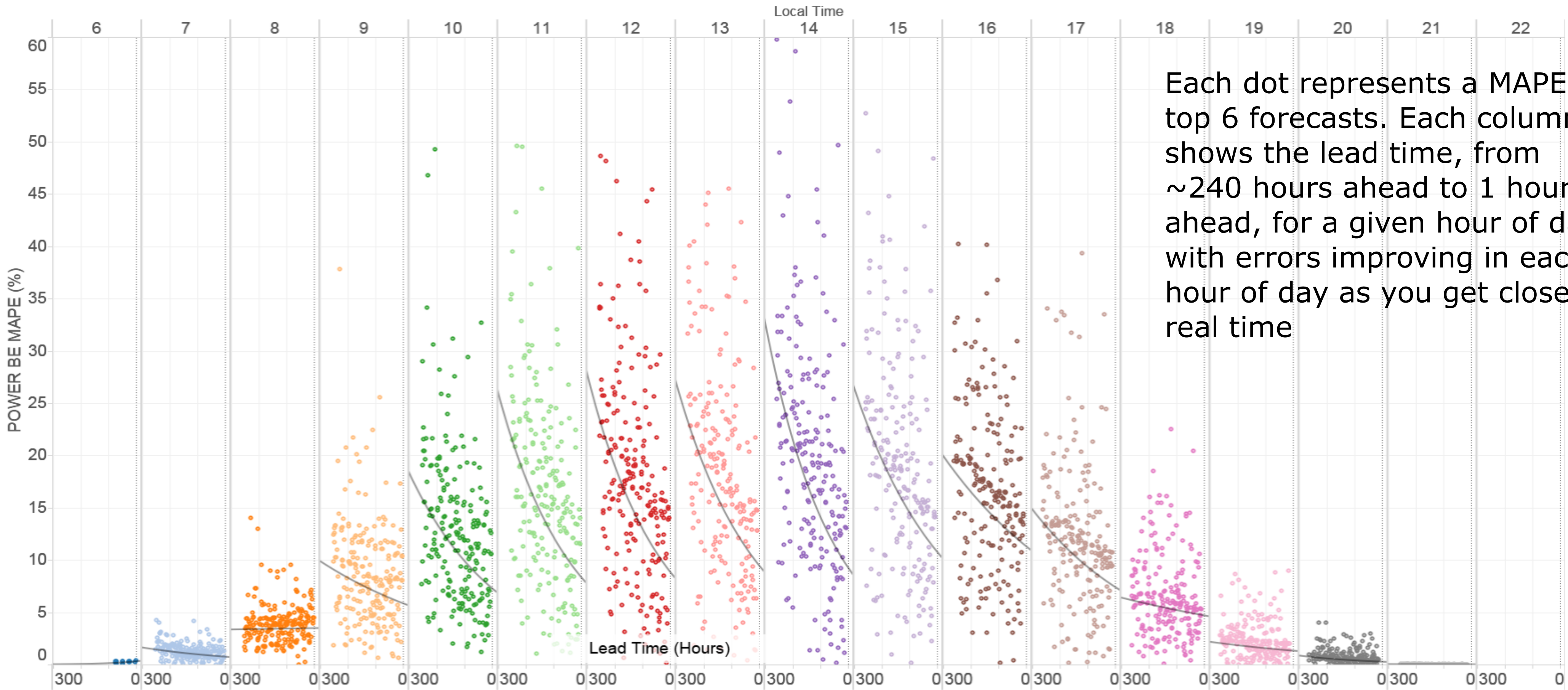
Visualization of the spread in average MAPE of 5 forecasts of different day-ahead time periods (each column) at two different sites. 6 months of data used – each color is for a different forecaster



The box-and-wiskers plots visualises performance of forecasters in different forecast horizons and provide a better overview of single outliers, how much and where, forecast errors deviate and how correlated errors are.

Significance of Results –
Looking at time dependent scatter plots

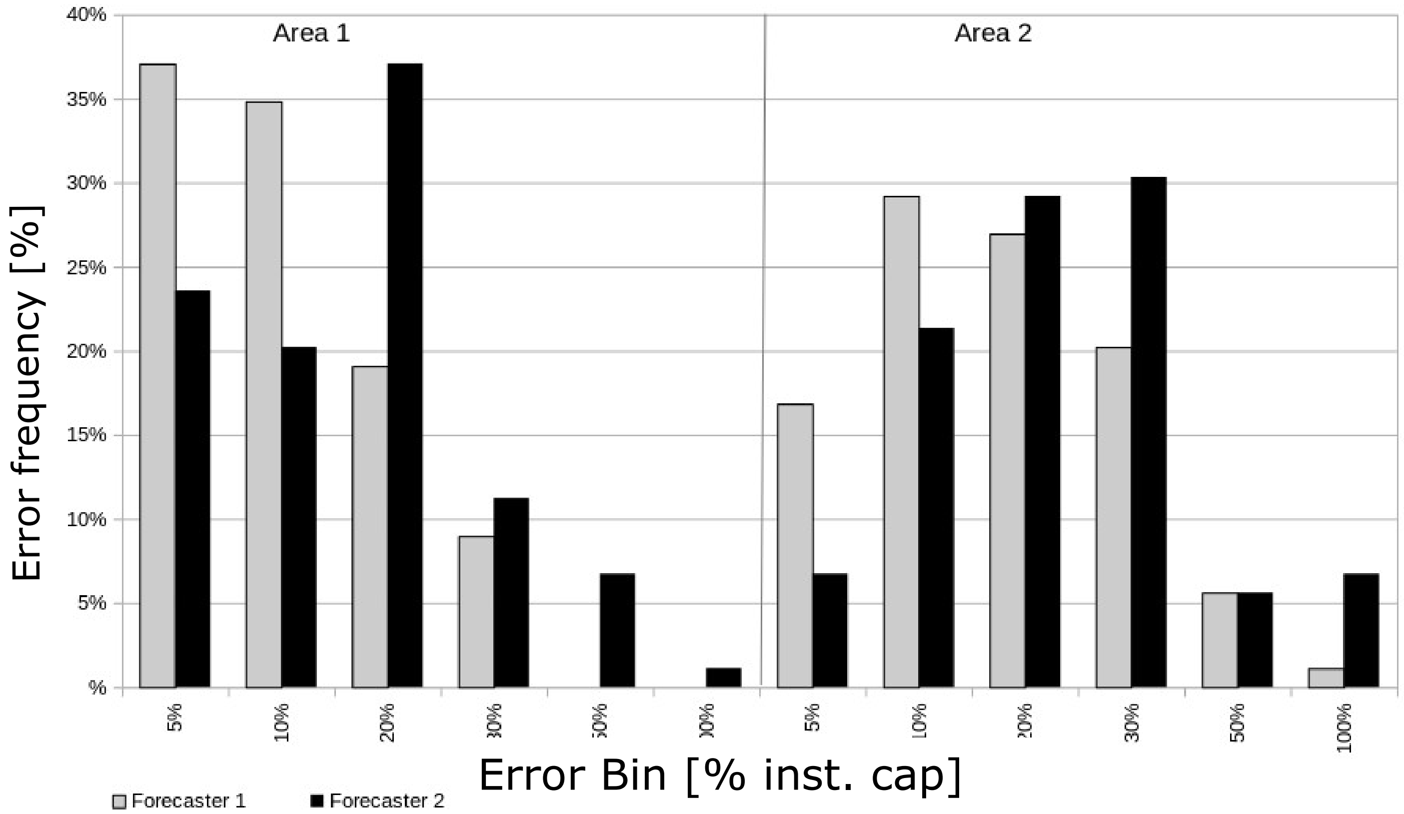
Average of MAPE of best estimate forecast from the top 6 forecasters by hour of day (outside columns) and lead time (within each column).



Each dot represents a MAPE for top 6 forecasts. Each column shows the lead time, from ~240 hours ahead to 1 hour ahead, for a given hour of day, with errors improving in each hour of day as you get close to real time

Example illustrates a very large spread in errors during certain times of the day and Example shows that using MAPE over the entire day may hide some of the story

Significance of Results
- Analysis with frequency distribution of Errors -



Errors have different impact on costs

Average error measures do not show **cost profiles**!

Separating errors into bins helps under-standing the impact of "system critical errors" and enable to give incentives for improvements

MAE/RMSE: FC1 =~ FC2

Frequency distribution:

FC1 >> FC2 errors in bins 5-10%
FC2 > FC1 errors in bins 20-100%
FC2 [20% error] >>> FC1
FC1 [50-100% error] = 0%
FC2 [50-100% error] = 10%

Forecaster 1: much more small errors – how do these contribute to costs ?

Forecaster 2: more large errors
Is large amount of 20% errors significant in costs ?

CASE EXAMPLE: INTRA-DAY FORECASTS FOR HAWAII ELECTRIC LIGHT CO (HELCO):

Background: HELCO operates an island grid with no interconnections and a high penetration of wind and solar generation.

- Net Load: 90 to 180 MW
- Wind Capacity: 31 MW
- Solar Capacity: 90 MW (all BTM)

- **Daily Critical HELCO Decision-making Time Frames:**
  - 0500 HST: preparation for morning net load peak and midday net load minimum
  - 1000 HST: preparation for midday net load ramps
  - 1300 HST: preparation for evening net load peak

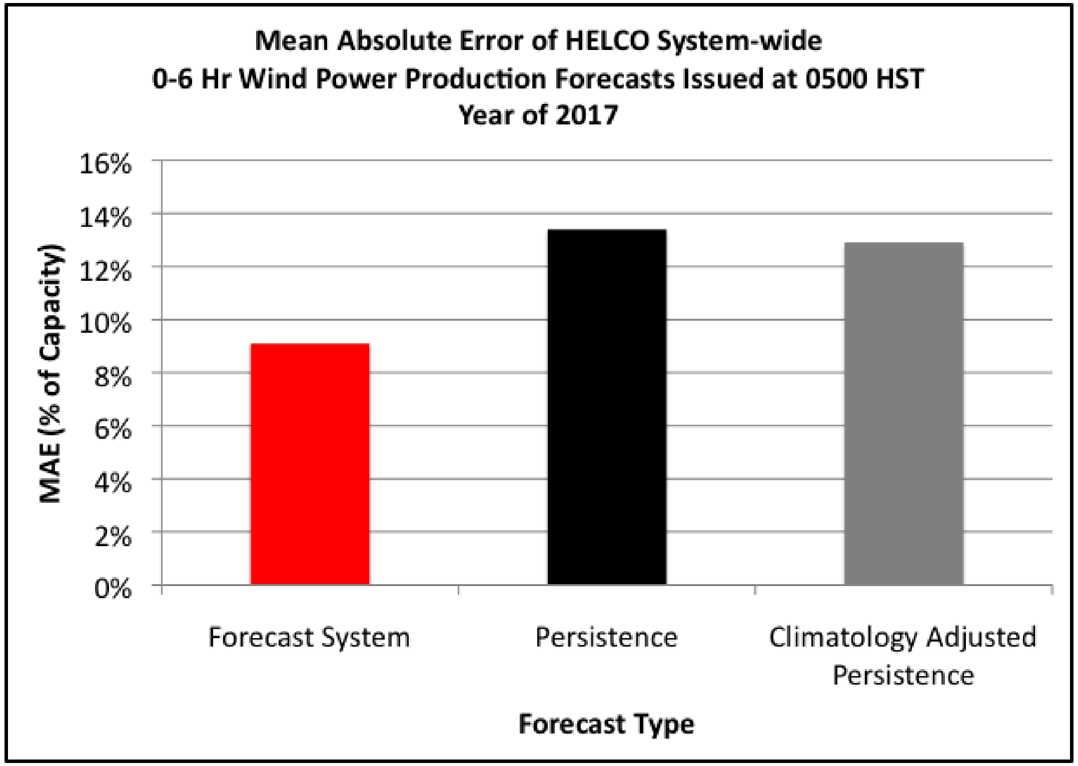
Typical 0500 HST Operational Issues

- Will midday net loads be low enough to shutdown a unit after the morning peak?
  - If YES: simple cycle CT that has no start/stop restrictions but less efficient can be used to serve the morning peak then shutdown when no longer needed.
  - IF NO: more efficient combine cycle (CC) CT will be used all day
  - One CC plant has a start/stop restriction: no multiple starts in a calendar day
- Will an excess energy situation occur due to high "as available" generation?
  - Determine whether curtailment of the as-available renewable generation or taking a unit offline will best address the situation

ISSUE: HOW TO MEASURE THE VALUE OF FORECASTS TO THE OPERATIONAL DECISION MAKING PROCESS?

A TRADITIONAL and TYPICAL APPROACH.. MEAN ABSOLUTE ERROR (MAE)

12-month MAE of 0-6 hour Wind Power Forecasts
Forecast Time Increment: 15 minutes
Issued Time: 0500 HST



- MAE of wind power forecasts is 9.1% of capacity and 32,1% lower than persistence MAE over the 0-6 hr time frame !

This looks pretty good but what information does it provide about the value in the 0500 HST decision-making process?

ISSUE: HOW TO MEASURE THE VALUE OF FORECASTS TO THE OPERATIONAL DECISION-MAKING PROCESS?

iea wind

CUSTOMIZED CATEGORY-BASED METRIC BASED on DECECION-MAKING NEEDS

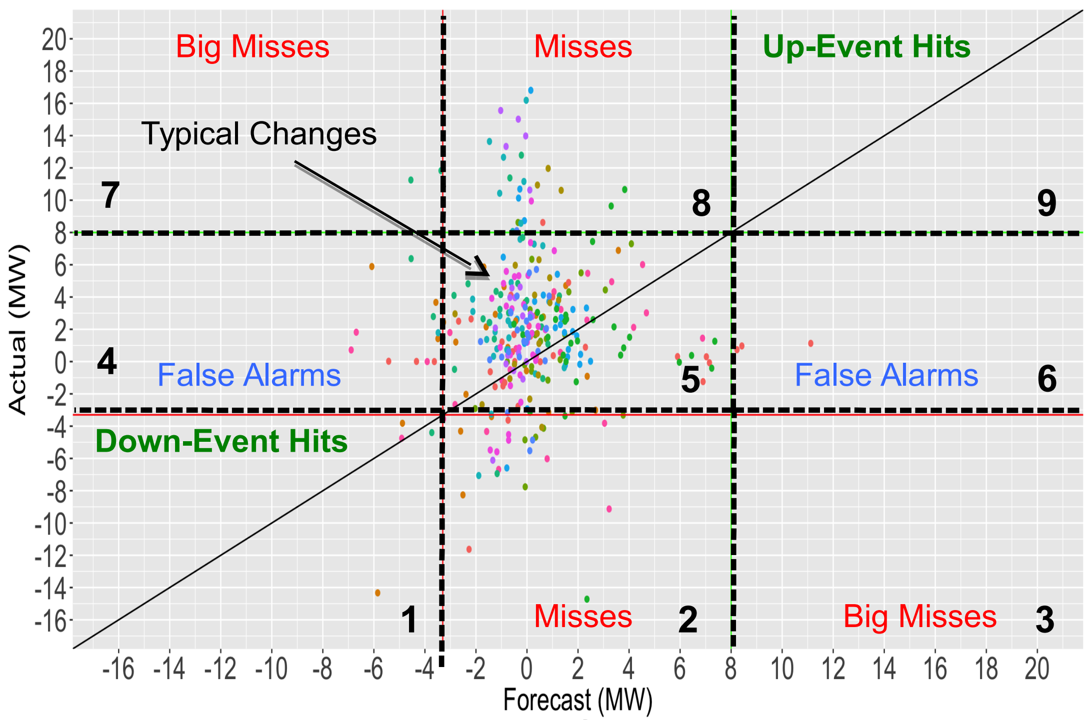
WHAT IS NEEDED FOR 0500 DECISION-MAKING?

- **Will wind significantly increase or decrease from its level at 0500 HST?**
- Will midday solar production be much higher or lower than its typical range?

EVALUATION OF WIND FORECASTS FROM THIS PERSPECTIVE

- Focus on significant changes from 0500 to the 0800-1100 period
- **Event: top 20% or bottom 20% of the daily change in this period**
- **Metric: CSI**
  - Ratio of Hits to Hits (H), Misses (M) and False Alarms (FA)

2017 Daily Forecast vs. Actual of Change in Wind Generation from 0500 to 0800-1100 Average



- **12-month CSI for 2017=5.0%**

Ratios	H	M	FA	Total	CSI
DOWN	4	29	15	48	8.3%
UP	0	29	3	32	0.0%
Total	4	58	18	80	5.0%

State-of-the-art MAE conceals the fact that the forecasts did not have much skill in predicting the (infrequent) significant change events

EXAMPLE CASE: HELCO* ALTERNATIVE FORECAST EVALUATION

(a) Key operating decisions and time frames were identified

(b) A customized categorical forecast evaluation scheme was formulated

(c) A customized categorical forecast performance metric was designed

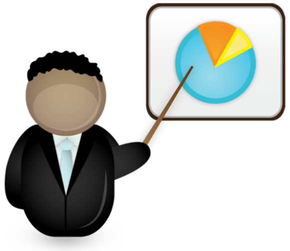
Results

- **Traditional forecast metrics** achieved state-of-the-art forecast performance
- **Customized Category-based metrics** missing some skill in forecasting atypical conditions

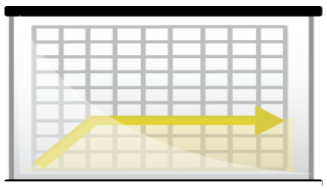
Next steps: Optimize forecast systems to achieve performance for predicting atypical conditions (customized category based evaluation)

* Hawaii Electric Light Co

Evaluation Paradigm



- **Verification is subjective**
  it is important to understand the limitations of a chosen metric

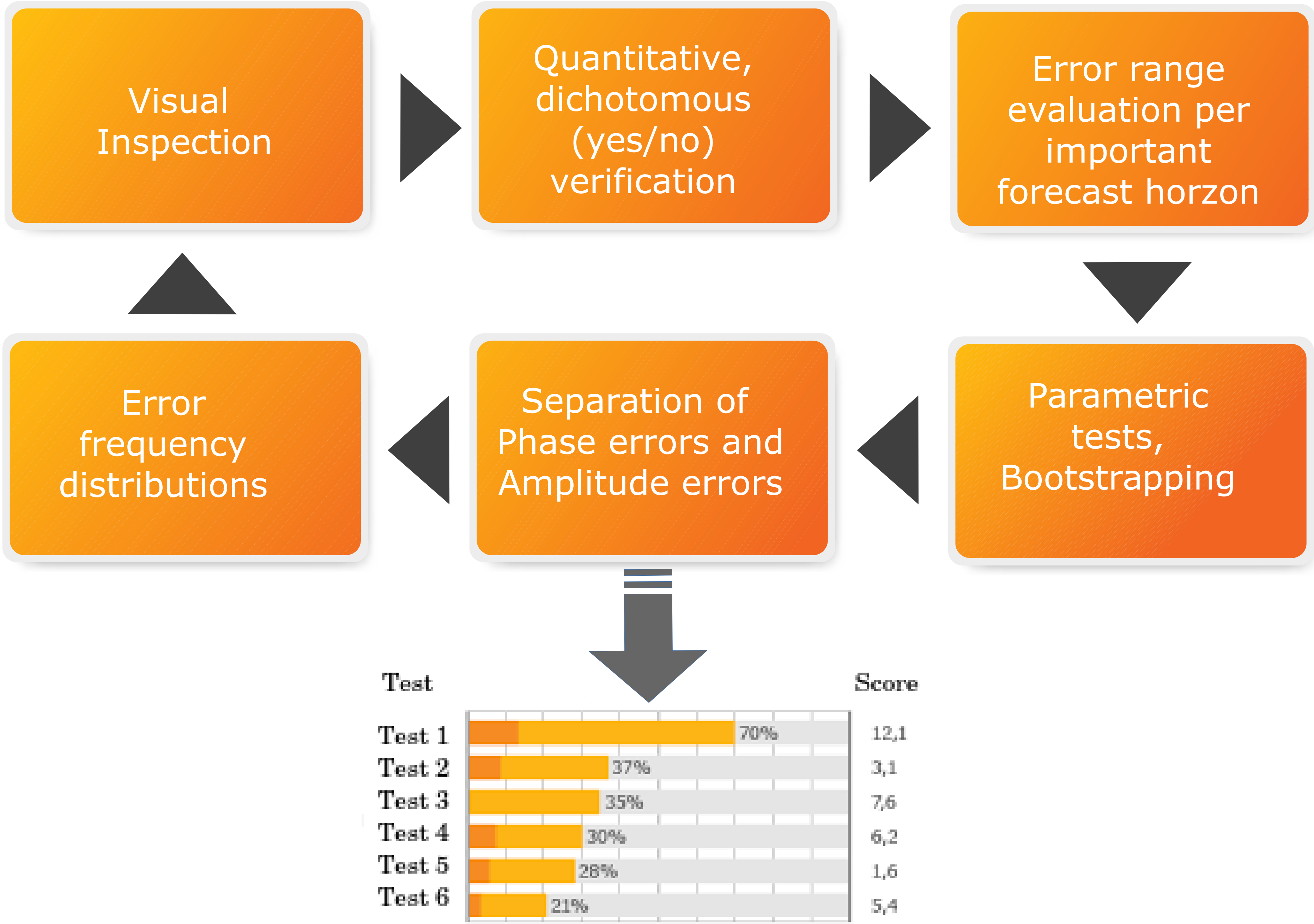


- **Verification has an inherent uncertainty**
  due to its dependence on the evaluation data set

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

- **Evaluation should reflect a "cost function"**
  i.e. the metric combinations should provide value of solution

Evaluation with Verification Methods – development of "cost functions" -



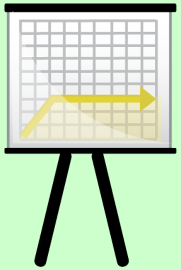
Key Takeaways…

Solution selection process:

Use a *Decision Support Tool* to establish a procedure

Benchmarking:

Setup a representative, transparent and fair test with good user-provider communication



Evaluation:

Develop a "cost function" or use an "evaluation matrix" of different scores according to their importance

Thank you for your attention !

Follow us:

Project webpage

http://www.ieawindforecasting.dk/

Task-page:

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

Contact:

Task Leaders and Presenters

Corinna Möhrlen: com@weprog.com

Jeff Lerner: Jeffrey.Lerner@vaisala.com

John Zack: jzack@awstruepower.com