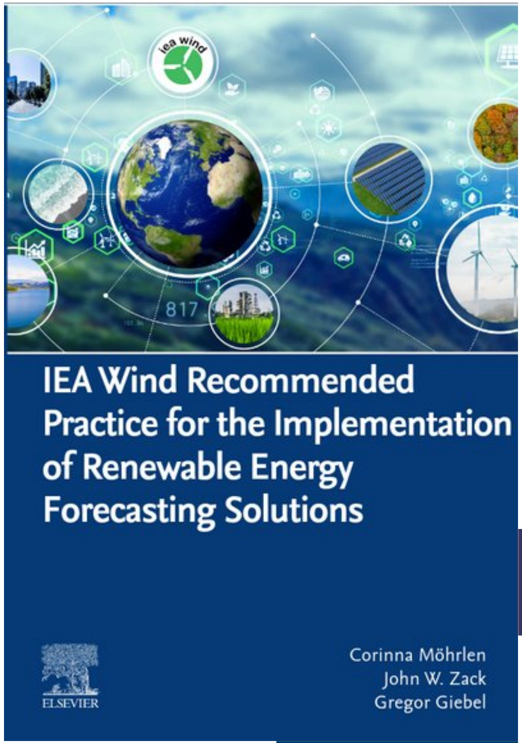
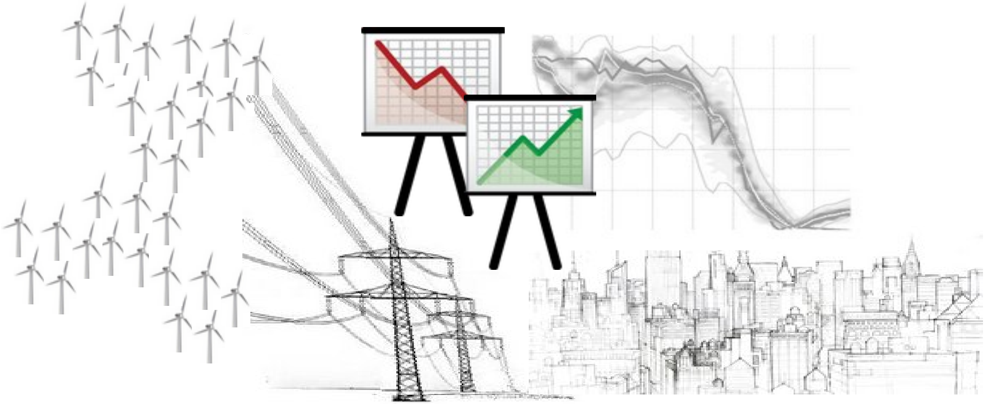
IEA Wind Task 51

IEA Wind Recommended Practice for the Implementation of Renewable Energy Forecasting Solutions





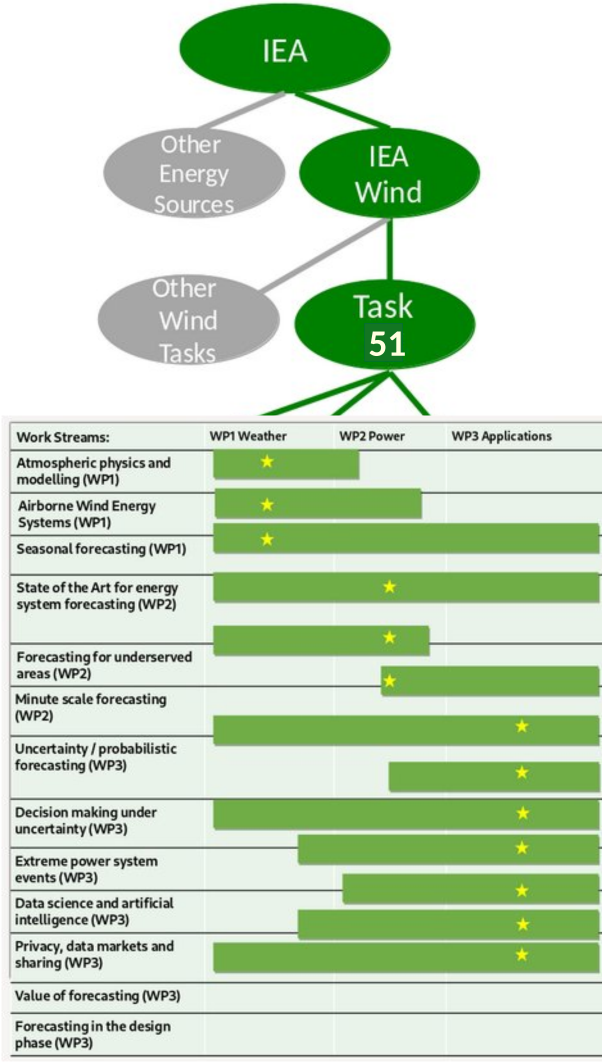
Hands-on examples for the use of the guideline

Mathias Blicher Bjerregaard

Windintegration Workshop
Kgs. Lyngby, Denmark, 25. Sept. *2023*

MESO ATMOSPHERIC RESEARCH FORECASTING · SIMULATIONS
WEPROG
DTU

IEA Wind Task 51: Forecasting for the weather-driven energy system



What is the IEA (International Energy Agency)? (www.iea.org)

- International organization within OECD with 30 members countries and 8 associates
- Promotes global dialogue on energy, providing authoritative analysis through a wide range of publications
- **One activity: convenes panels of experts to address specific topics/issues**

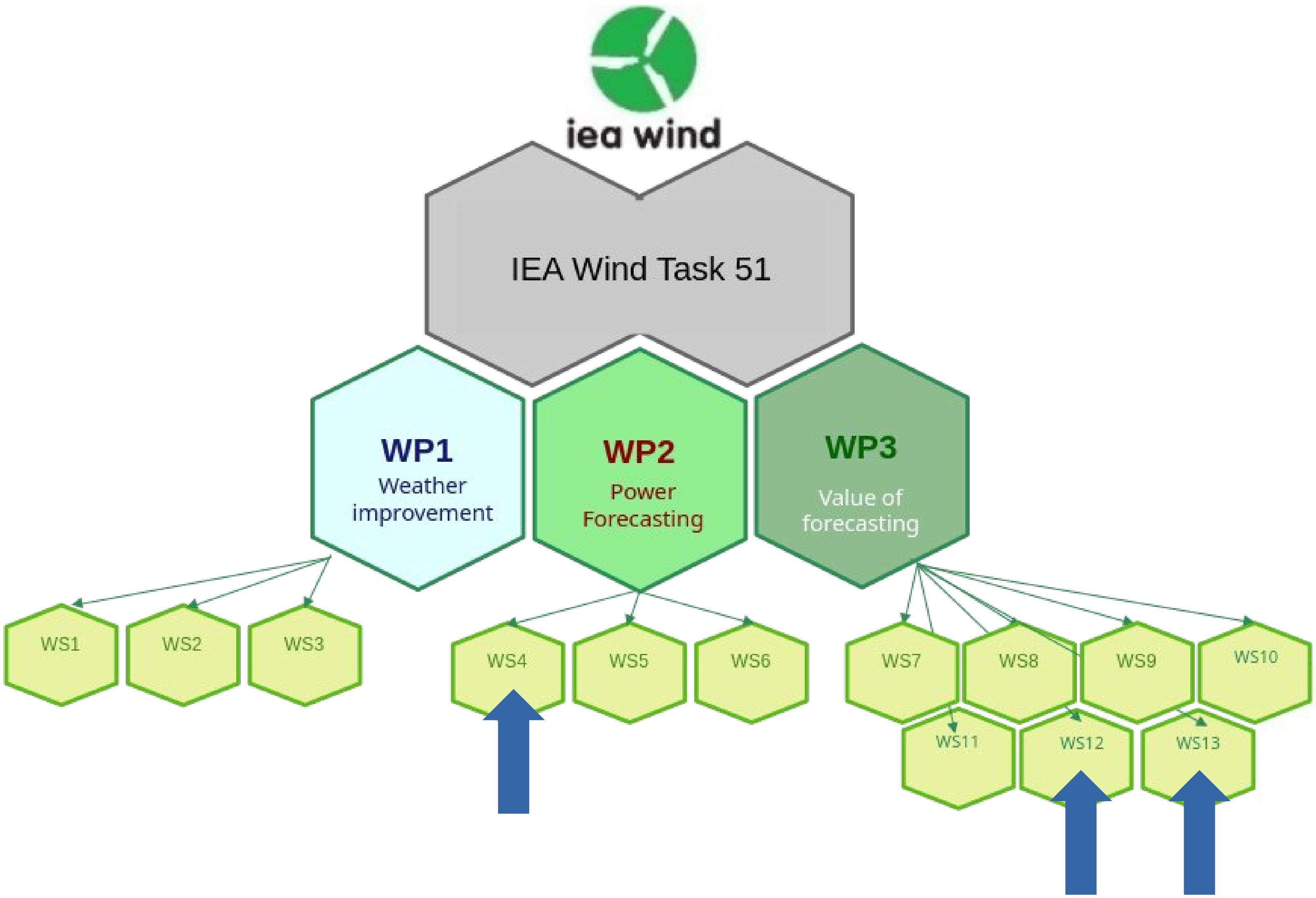
Task 51: Forecasting for the weather driven Energy System:

- One of 17 Tasks of IEA Wind: https://iea-wind.org/
- Task 36: Phase 1: 2016-2018; Phase 2: 2019-2021 **Task 51: Phase 3: 2022-2025**
- Operating Agent: Gregor Giebel of DTU Wind Energy
- Objective: facilitate international collaboration to **improve wind energy forecasts**
- Participants: (1) research organization and projects, (2) forecast providers, (3) policy-makers and (4) end-users & stakeholders

Task 51 Scope: 3 “Work Packages” / 13 “Workstreams”

- WP1: Global Coordination in Forecast Model Improvement
- WP2: Benchmarking, Predictability and Model Uncertainty
- WP3: Optimal Use of Forecasting Solutions

Task homepage: https://iea-wind.org/task51

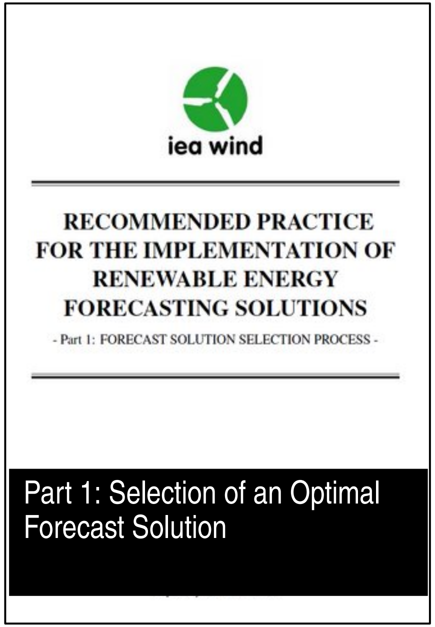


iea wind
IEA Wind Task 51
WP1
Weather improvement
WP2
Power Forecasting
WP3
Value of forecasting
WS1
WS2
WS3
WS4
WS5
WS6
WS7
WS8
WS9
WS10
WS11
WS12
WS13

Overview

- Background: IEA Wind Recommended Practice (RP) for the Implementation of Renewable Energy Forecasting Solutions
  - What it is
  - Where to get it
- Use Case Examples based on Recommendations
  - Wind speed evaluation at a Danish Coastal Site
  - Wind power evaluation at a substation in Ireland
  - Meteorological sensor performance assessment at a site in the German Bight

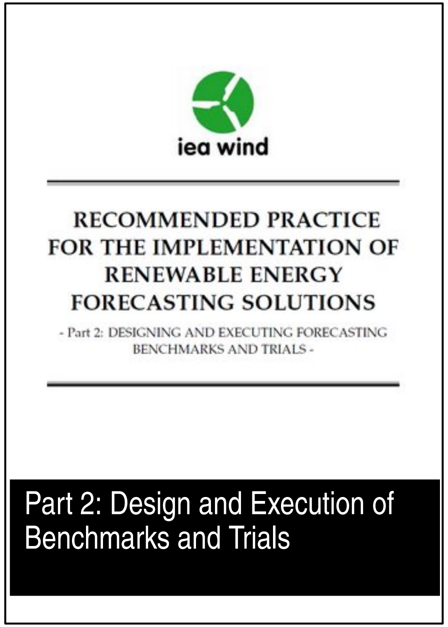
IEA Wind Recommended Practice for the Implementation of Renewable Energy Forecasting Solutions: Set of 4 Parts


iea wind

RECOMMENDED PRACTICE FOR THE IMPLEMENTATION OF RENEWABLE ENERGY FORECASTING SOLUTIONS

- Part 1: FORECAST SOLUTION SELECTION PROCESS -

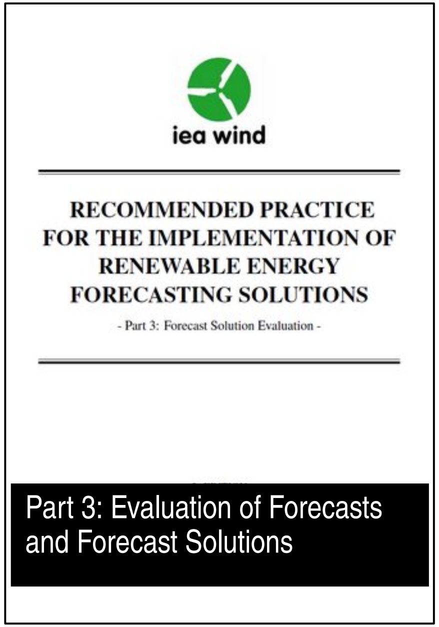
Part 1: Selection of an Optimal Forecast Solution


iea wind

RECOMMENDED PRACTICE FOR THE IMPLEMENTATION OF RENEWABLE ENERGY FORECASTING SOLUTIONS

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

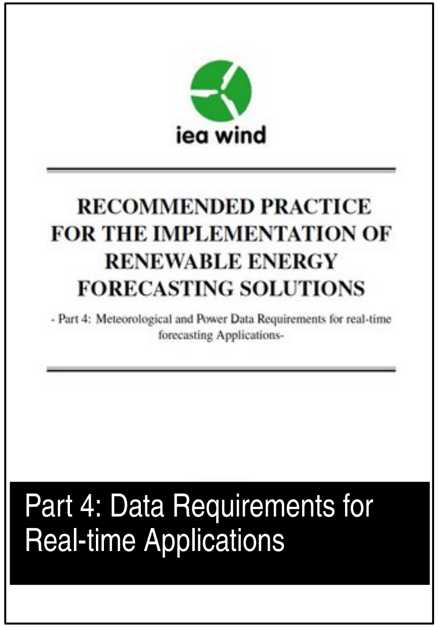
Part 2: Design and Execution of Benchmarks and Trials


iea wind

RECOMMENDED PRACTICE FOR THE IMPLEMENTATION OF RENEWABLE ENERGY FORECASTING SOLUTIONS

- Part 3: Forecast Solution Evaluation -

Part 3: Evaluation of Forecasts and Forecast Solutions


iea wind

RECOMMENDED PRACTICE FOR THE IMPLEMENTATION OF RENEWABLE ENERGY FORECASTING SOLUTIONS

- Part 4: Meteorological and Power Data Requirements for real-time forecasting Applications-

Part 4: Data Requirements for Real-time Applications

Video Introduction

Introduction: https://www.youtube.com/watch?v=XVO37hLE03M iea wind

IEA Wind Recommended Practice Book

Note

Elsevier Book

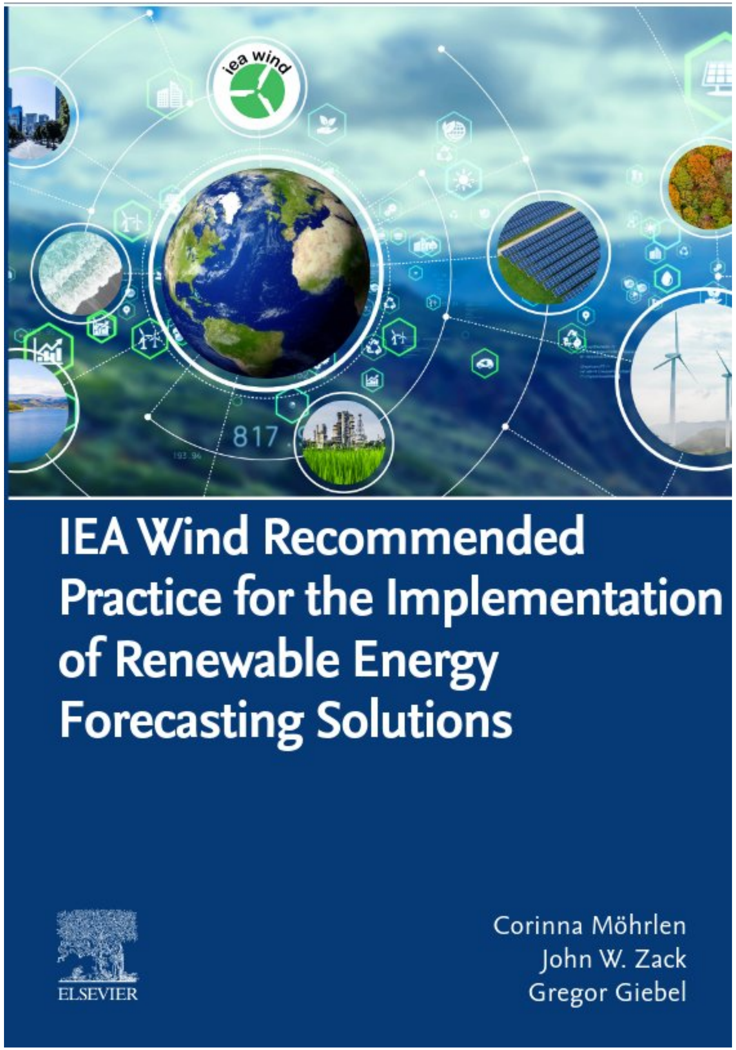
https://www.elsevier.com/books/iea-wind-recommended-practice-for-the-implementation-of-renewable-energy-forecasting-solutions/mohrlen/978-0-443-18681-3

Online OpenAccess:

https://www.sciencedirect.com/book/9780443186813/iea-wind-recommended-practice-for-the-implementation-of-renewable-energy-forecasting-solutions

IEA Wind Task 51 Information

iea-wind.org → Task 51 → Publications → Recommended Practice



IEA Wind Recommended Practice for the Implementation of Renewable Energy Forecast Solutions

Application Areas for the Recommendations

1. System Operation, Balancing and Trading

- Situational awareness in critical weather events
- High-Speed Shutdown events
- Grid related down-regulation or curtailments
- Short-term forecasting with updates from measurements
- Intra-day power plant balancing

2. Wind Turbine, Wind Farm and Solar Plant Operation and Monitoring

- Wind turbine and Power Plant Control
- Condition Monitoring

Companion Evaluation Software: “WE-validate-prob”

Assessment of forecasts with an R-package code

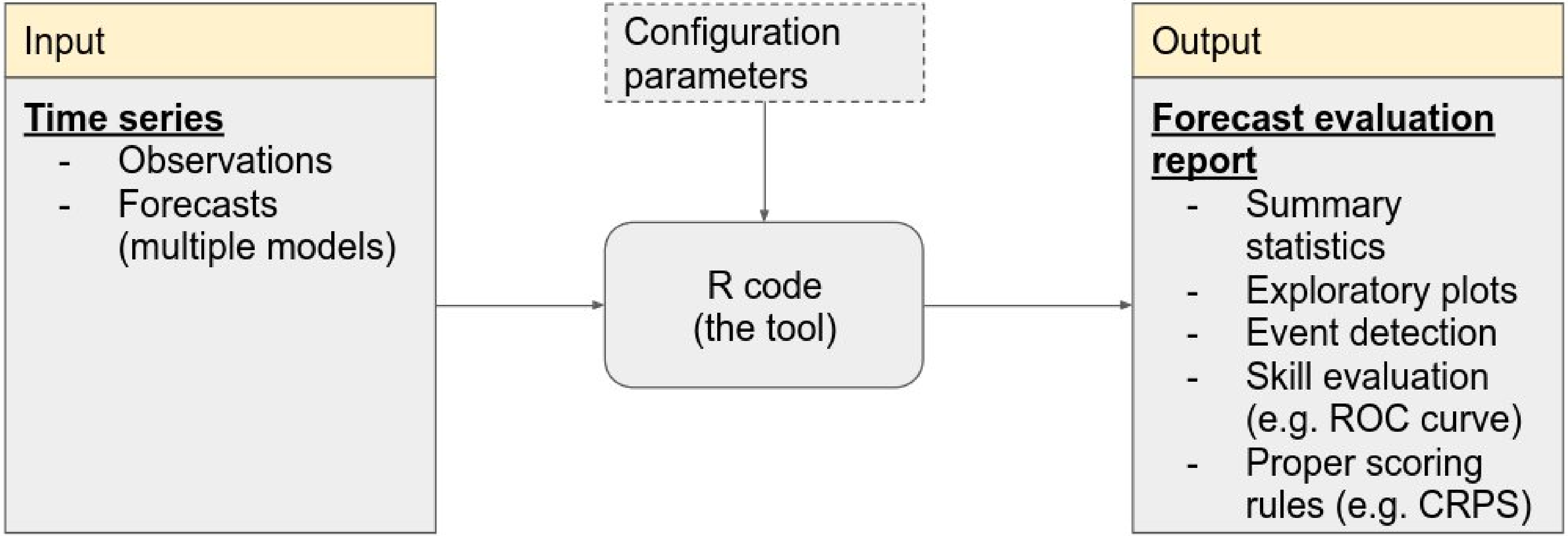
ACADEMIC PRESS

IEA Wind Recommended Practice for the Implementation of Renewable Energy Forecasting Solutions

Wind Energy Engineering

2023, Pages 321-322

Appendix G - Validation and verification code examples



Recommendation: Establish an Evaluation Framework

Key Components

(1) the forecast application
(2) the key forecast time frames
(3) a ranking of the importance of forecast performance attributes

Specify the forecast framework

Define the evaluation sample

(1) Choose a time period likely to produce a representative sample of relevant weather patterns
(2) Choose a sufficient and well-defined evaluation time frame (e.g. 3 months, 1 year, ...)

(1) Strategy to deal with missing or erroneous data & forecasts
(2) Specify evaluation criteria on delivery performance

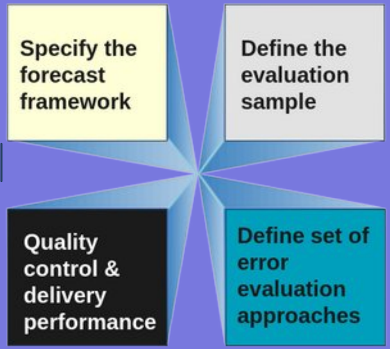
Quality control & delivery performance

Define set of error evaluation approaches

(1) visual inspection
(2) use of more specific metrics: SDE, SDBIAS, StDev, VAR, CORR
(3) use of histogram or box plot for evaluation of outliers
(4) use of contingency tables for specific event analysis
(5) use of improvement scores relative to a relevant reference forecast

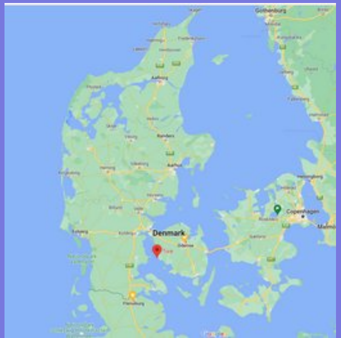
Example 1: Evaluation of Wind Speed at a Danish Coastal Site

Aim:
Verify the high resolution versus the low-resolution setup of an ensemble prediction system and evaluate improvement versus cost



Definition of the Sample:
Danish synoptic meteorological site:
South-west Funen "Assens"

- High-Resolution (HR):
  5km grid cells with 60 vertical levels
- Low resolution (LR):
  15km grid cells with 32 vertical levels

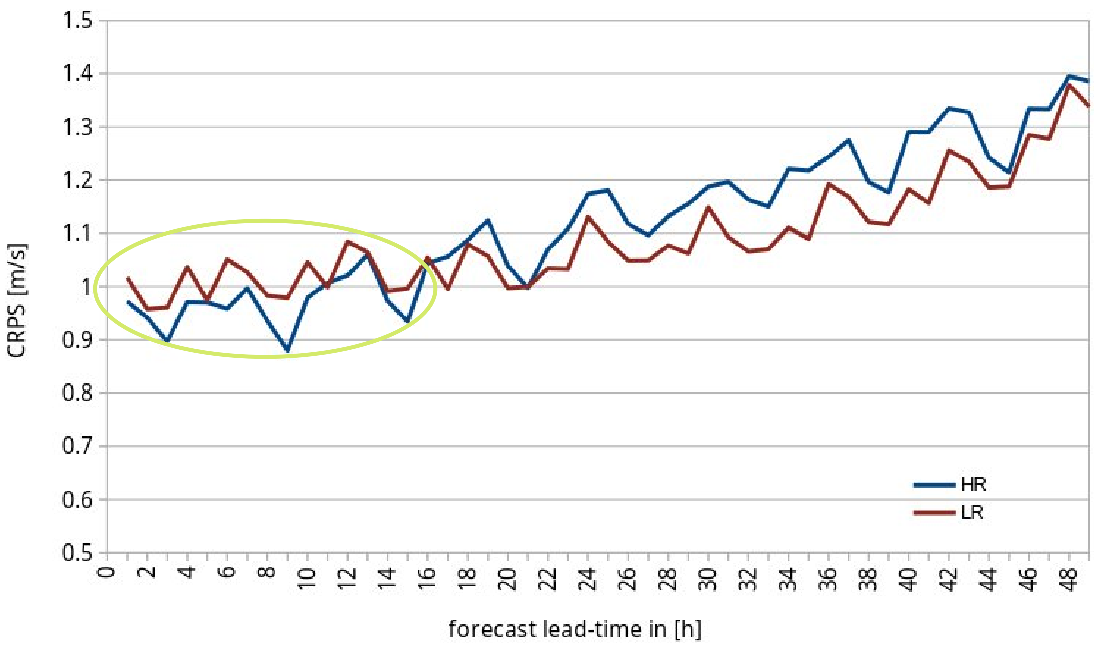


Evaluation Approach:
- CRPS
- CRPS lead-time dependency
- Reliability Diagram

Evaluation of Wind Speed at a Danish coastal site

Assessment of a high-resolution versus low resolution ensemble system

Forecast Type	CRPS	Improvement to Reference [%]
Reference	1.6635	
Lead-time	**6-11h**	
HR	1.140	-31.5
LR	1.159	-30.3
Lead-time	**0-48h**	
HR	1.1236	-32.5
LR	1.0925	-34.3



Result from Test 1:
High-resolution setup has only value in the first 12 hours

Conclusion:
High-resolution setup can be complementary in the intra-day…

Introduction to Probabilistic Forecast Assessment of Ramping Events:
Reliability Diagram CORP approach versus Murphy's approach

Reliability is the degree to which the forecasted probabilities are in agreement with the outcome frequencies

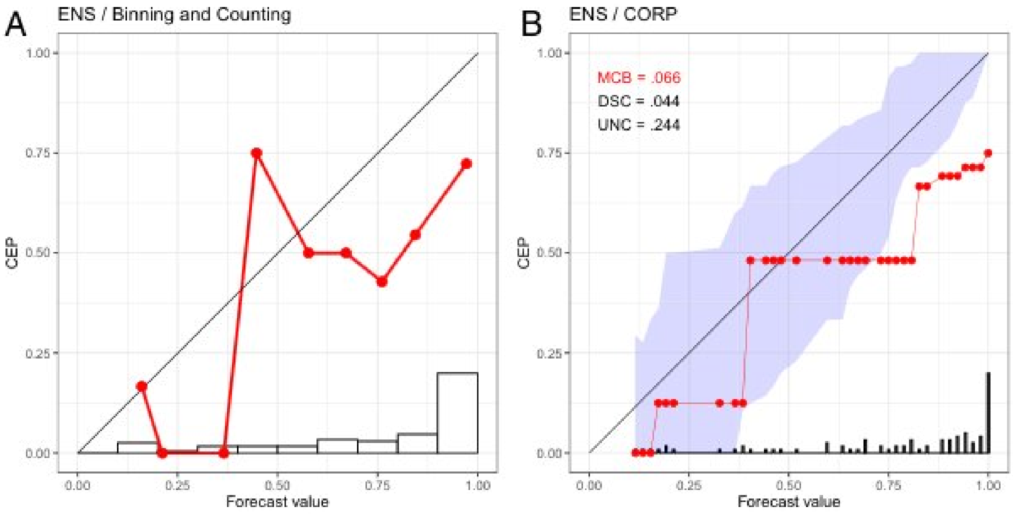


Fig. From documentation (doi:10.1073/pnas.2016191118)

Equidistant binning

non-equidistant binning + 90% consistency band

Reliability Diagram with CORP appraoch:

X-axis: forecasted probabilities

Y-axis: conditional event probabilities (CEP) → the frequency of observed events given the specific forecast probability

Evaluation Criteria Sensitivity:

4 variable to choose: A,B,C,D

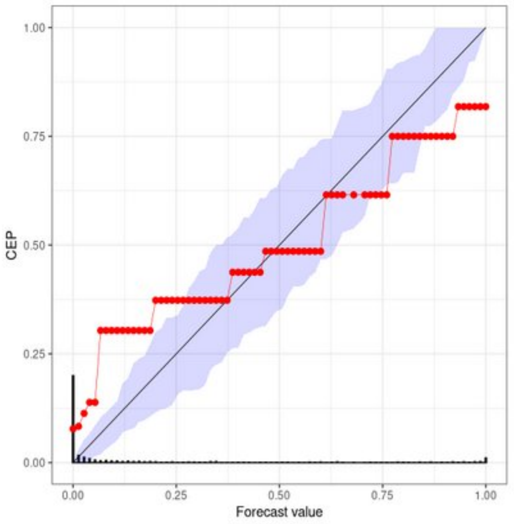
- Threshold: A (a minimum of A "positives" needed for an event)
- Forecast horizon: B1-B2 hours
- Change: C [var unit] over a D [time] window.

Evaluation of Wind Speed at a Danish coastal site

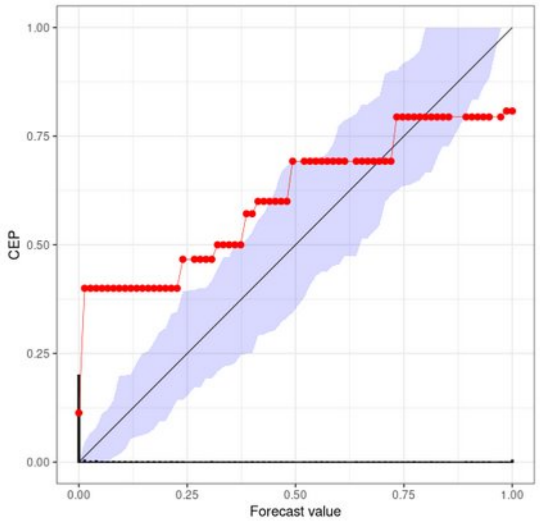
Assessment of a high-resolution versus low resolution ensemble system

Reliability is the degree to which the forecasted probabilities are in agreement with the outcome frequencies

HR-setup



LR-setup



Explanation of Score:
Reliability Diagram
X-axis: forecasted probabilities
Y-axis: conditional event probabilities (CEP) → the frequency of observed events given the specific forecast probability

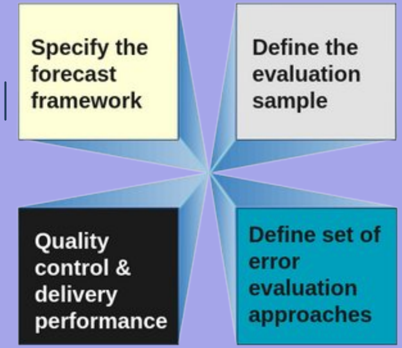
Evaluation Criteria:

- Threshold: 5 (a minimum of 5 "positives" needed for an event)
- Forecast horizon: 6-11 hours
- Change in wind speed: 3m/s over a 3 hour window.

Result: tendency to lie on top of the diagonal for LR; Hr only in lower bins → indicates a negative BIAS and\or a slight mis-calibration
Conclusion: HR setup has a better balance between resolution and calibration, staying mostly within the blue 90% consistency band.

Example 2: Wind Power Evaluation at a Substation in the North-west of Ireland

Aim:
Verify forecasts from 2 different ensemble prediction systems and use a set of scores for decision-making regarding which of the forecasts provide better value



Definition of the Sample:
Sub station in North-west of Ireland:
A number of wind farms are feeding into the substation (wind farm cluster).
<u>Forecast type:</u>
Ramp forecasts
- High-Resolution (HR): 5km grid cells with 60 vertical levels
- Low resolution (LR): 15km grid cells with 32 vertical levels

Evaluation Approach:
- CRPS
- Brier Score
- ROC

Wind Power Evaluation at a Substation in the North-west of Ireland

Probabilistic Forecast Assessment of Ramping Events: **CRPS & Brier Scores**

CRPS score

overall performance of prob. forecast

Forecast	CRPS [MW]	CRPS [% inst. cap]
HR	10.5	5.8
LR	10.9	6.0
Reference	20.6	11.5

No significance!

Explanation of the score:

- CRPS is the probabilistic analogue to the Mean absolute error (MAE) for a deterministic forecast.
- Lower CRPS values indicate smaller error and therefore better performance.
- CRPS scores for each forecast over the 3-month test period

BRIER score

overall accuracy of a probabilistic *event* forecast

Fore-cast	20MW 1hour	30MW 3 hours	40MW 3 hours	60MW 3 hours
HR	0.0501	0.089	0.0513	0.021
LR	0.0459	0.084	0.0464	0.018
$\Delta(HR - LR)$	0.0043	0.0053	0.0049	0.0028

Large sensitivity to event choice!

Explanation of the score:

- BS is the probabilistic analogue to mean squared error (MSE/RMSE) of deterministic forecast
- BS measures the mean squared difference (MSE/RMSE) between the forecasted probability (e.g., 0 to 1) and the actual outcome (e.g., 0 or 1).
- The BS values range between 0 and 1 with lower values indicating better performance.

Wind Power Evaluation at a Substation in the North-west of Ireland

Probabilistic Forecast Assessment of Ramping Events: **Brier Score Decomposition**

Decomposition of BRIER Scores

Fore-cast	MS	CAL	DSC (RES)	UNC
Limit:	30MW/3h			
HR	0.0892	0.0105	0.0274	0.106
LR	0.0839	0.0062	0.0283	0.106
Limit:	40MW/3h			
HR	0.0513	0.0074	0.0153	0.0592
LR	0.0464	0.0029	0.0157	0.0592
Limit:	60MW/3h			
HR	0.0210	0.0018	0.0024	0.0217
LR	0.0182	0.0010	0.0045	0.0217
Limit:	20MW/1h			
HR	0.0501	0.00494	0.00457	0.0498
LR	0.0459	0.00248	0.00639	0.0498

Explanation of the Scores:

- **Mean Score (MS)** measure the overall predictive event performance
- **Calibration/reliability (CAL)**: measures the agreement of forecasted probability with frequency of event occurrence given the forecasted probability (conditional event probability)
- **Discrimination/resolution (DSC/RES)**: measures the ability of forecasts to correctly distinguish differences in probabilities among the cases.
  → higher values contribute to lower BS, i.e. indicate better performance.
- **Uncertainty (UNC):** measures inherent uncertainty in the event and is related to the event frequency in the sample.
  → lower values contribute to lower BS, max. UNC for 50% events in sample

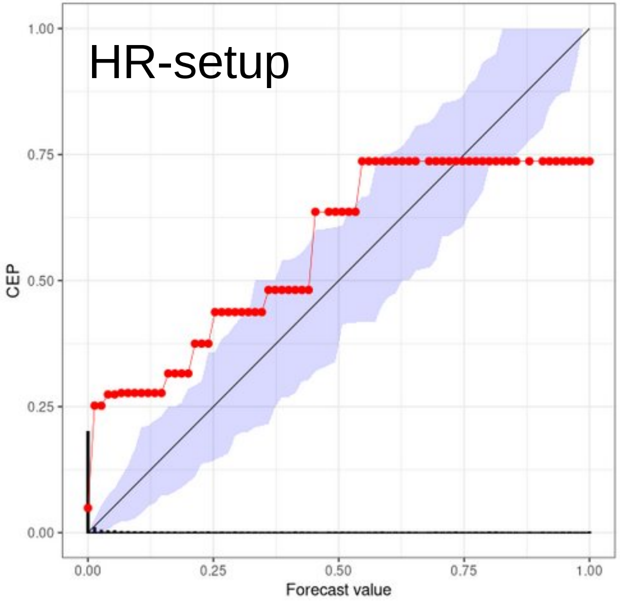
Result: The difference between HR and LR insignificant overall (MS), but quite significant for some components and sensitive to the thresholds and classifiers: the calibration (CAL) in the 40MW/3h class and the discrimination (DSC) in the 60MW/3h class is significantly better for the LR setup...

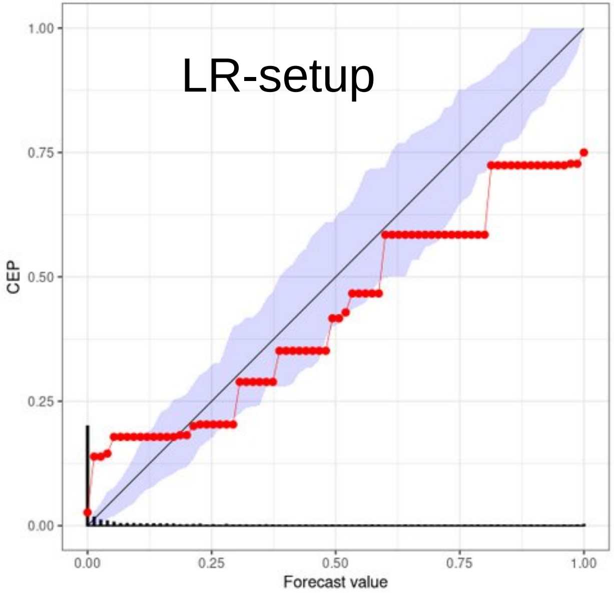
Conclusion: Decomposition of the Brier score is important, as it reveals differences in the forecast's skill related to distinguish events and to match occurrence with probabilities.

Wind Power Evaluation at a Substation in the North-west of Ireland

Probabilistic Forecast Assessment of Ramping Events: **Reliability Diagram**

Evaluation Criteria: Threshold: 5 - Forecast horizon: 6-11 hours - Change: 30MW over a 3 hour window.





Explanation of Plots:
X-axis: forecasted probabilities
Y-axis: conditional event probabilities (CEP) → frequency of observed events given the specific forecast probability
Band: 90% consistency band

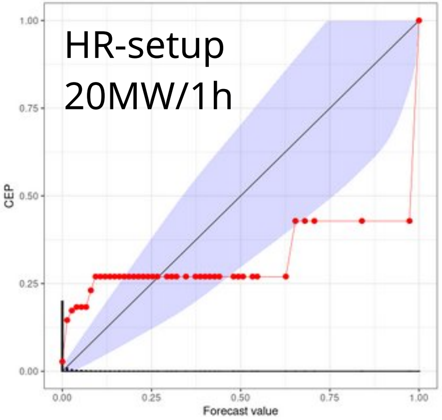
Result: tendency to lie on top the diagonal for HR; LR tendency to lie below diagonal
→ indicates a negative BIAS for LR and positive BIAS for HR … and/or a slight mis-calibration

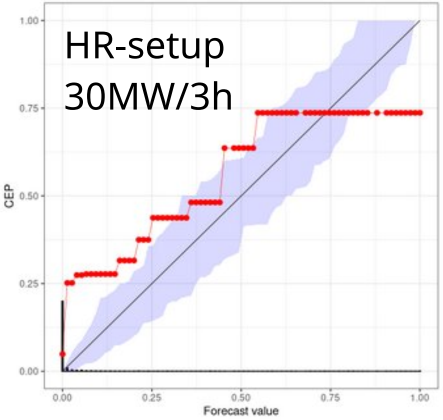
Conclusion: LR setup seems to be in better balance between resolution and calibration, staying mostly within the blue 90% consistency band – consistent with Brier score decomposition results….

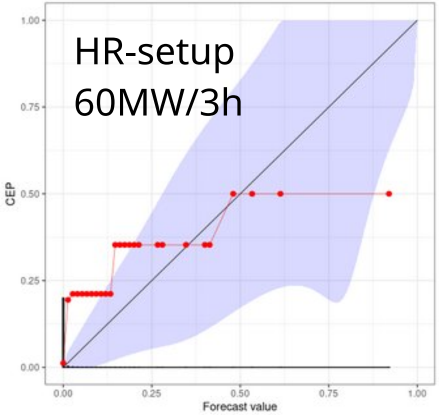
Wind Power Evaluation at a substation in the north-west of Ireland

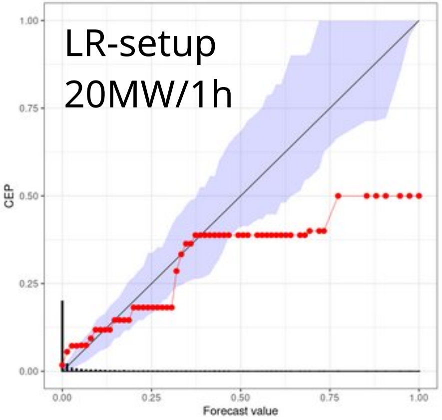
Probabilistic Forecast Assessment of forecasted Ramping Events: **Reliability Diagram**

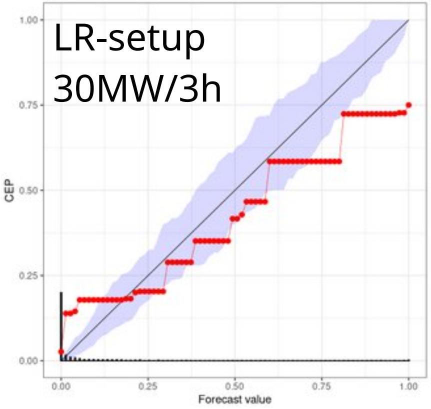
Demonstration of threshold selection sensitivity

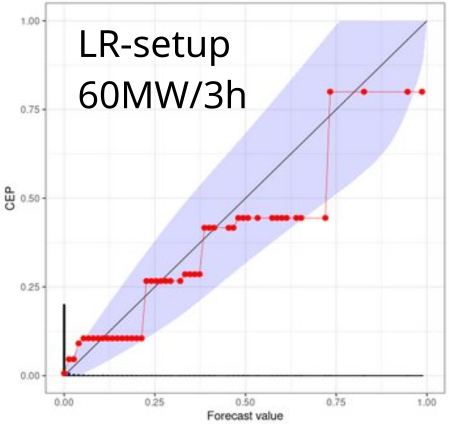












Evaluation of Wind Power at a substation in north-west of Ireland

Probabilistic Forecast Assessment of Ramping Events: **Continency table**

Contingency table + HitRate (HR) and False Alarm rate (FAR)

Fore-cast	Hits	Misses	False Alarms	Correct Neg.	HR	FAR
Limit:	**30MW**	**window:**	**3h**			
HR	149	145	153	1990	0.507	0.071
LR	204	90	393	1750	0.694	0.183
Limit:	**40MW**	**window:**	**3h**			
HR	82	72	91	2192	0.532	0.04
LR	112	42	262	2021	0.727	0.115
Limit:	**60MW**	**window:**	**3h**			
HR	10	44	31	2352	0.185	0.013
LR	30	24	102	2281	0.556	0.043
Limit:	**20MW**	**window:**	**1h**			
HR	37	91	101	2208	0.289	0.044
LR	74	54	302	2007	0.578	0.131

Result:

LR forecasts have much higher number of "hits"

LR forecasts have much more "false alarms"

most extreme example of this pattern is for the 60MW/3hr threshold

→ requires to look into costs for misses versus false alarms…

Explanation of the Score:

The Contingency table lists: absolute number of "hits", "misses", "false alarms" and "correct negatives" in the forecast sample

lists the "hit rate" (HiR)
→ the hits per total number of forecasts

"false alarm rate" (FAR)
→ the false alarms per total number of forecasts.

Conclusion:

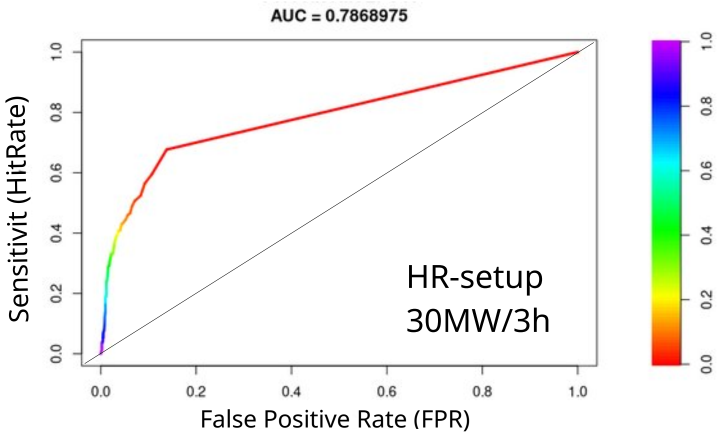
Beware of the threshold selection sensitivity in selection process and when analysing and evaluating the results

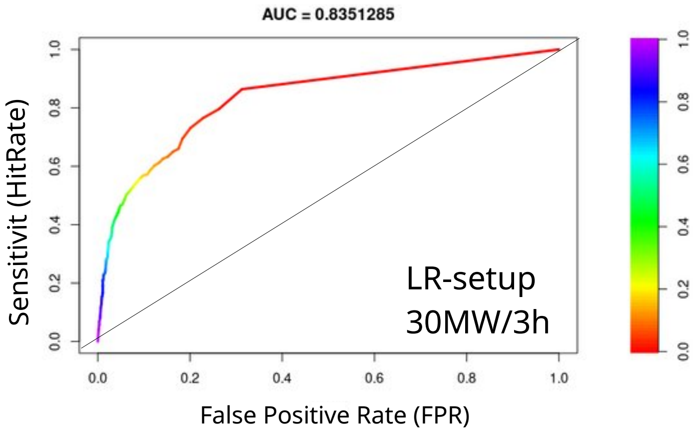
Fair evaluation comparison requires to provide the thresholds in advance

Wind Power Evaluation at a substation in the north-west of Ireland

Probabilistic Forecast Assessment of forecasted Ramping Events: **ROC Curve**

Receiver Operating Characteristics (ROC) curve measures the ability to discriminate between events and non-events and depicts the performance of forecasts at different probability thresholds





"Area Under the Curve" (AUC) for different ramping limits and time windows

Limit Window	20MW 1h	30MW 3h	40MW 3h	60MW 3h
HR	0.7201	0.7869	0.7916	0.7241
LR	0.7899	0.8351	0.8584	0.8380
$\Delta(HR - LR)$	-0.0043	-0.0053	-0.0049	-0.0028

Result:
Both forecast setups perform OK with a AUC > 0.7.

Slightly better, but little (insignificant) difference in the AUC scores for the LR forecasts

Explanation of the Score:

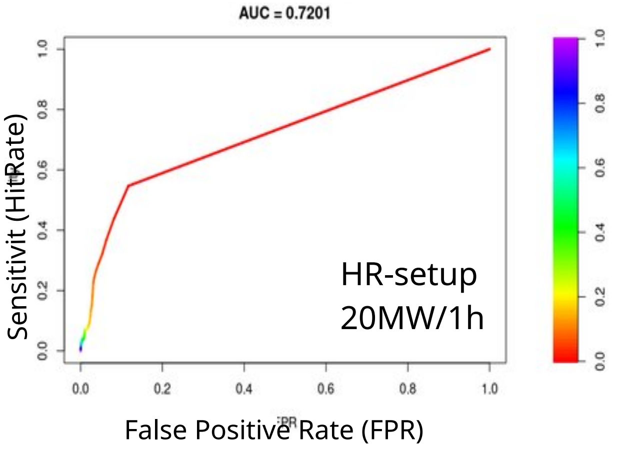
- The ROC curve ascends vertically at FAR=0.0 and horizontally at a sensitivity (hit rate) value of 1.0
- The color scale indicates classification thresholds yielding the points on the curve
- AUC= 1.0 for every forecast is a hit and no false alarms, 0.5 for random classifiers, i.e. forecasts with no skill (diagonal in graph)

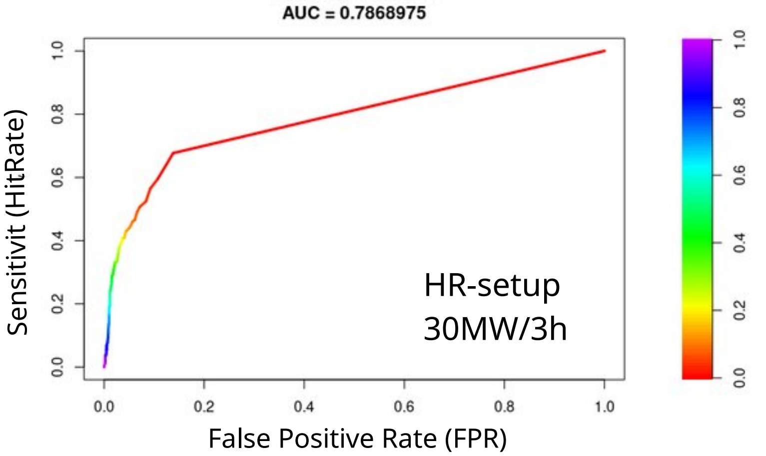
Conclusion: the ROC curve confirms the results from the Brier Scores and indicates that the difference is not due to a mis-calibration.

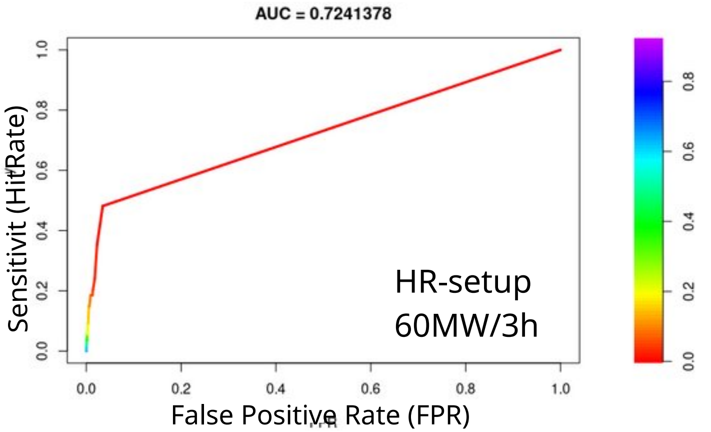
Wind Power Evaluation at a substation in the north-west of Ireland

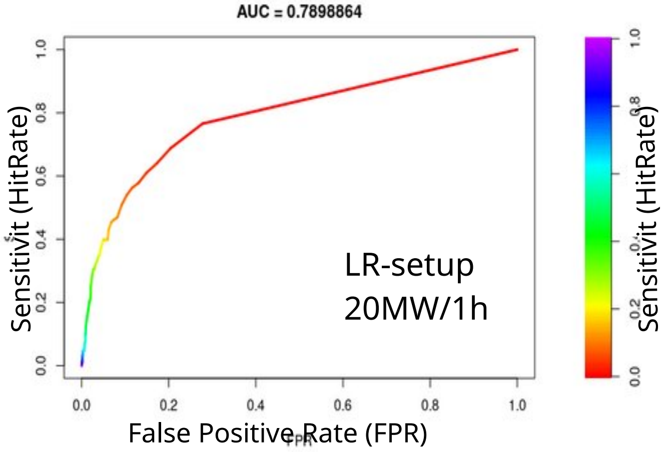
Probabilistic Forecast Assessment of forecasted Ramping Events: **ROC Curve**

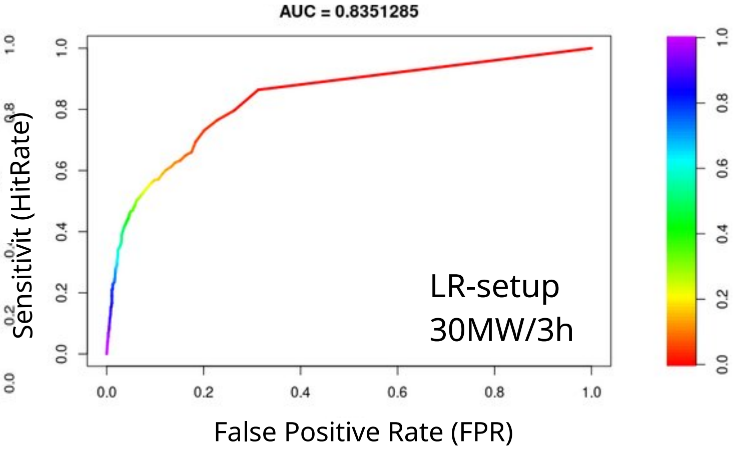
Demonstration of threshold selection sensitivity

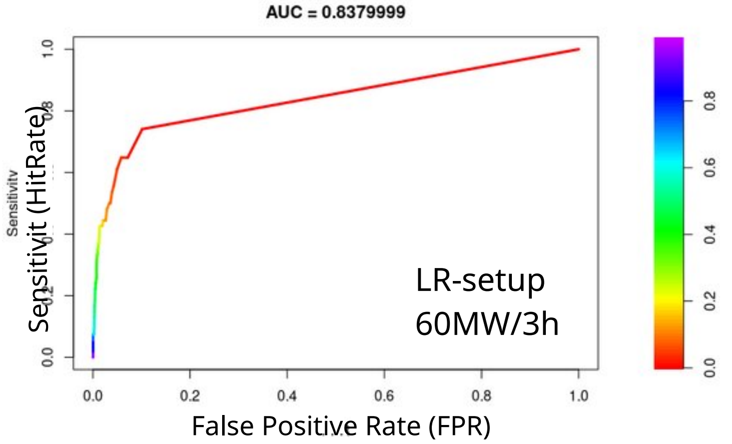












Evaluation of Wind Power at a Substation in North-west of Ireland

Probabilistic Forecast Assessment of Ramping Events: **Composite Performance Metric**

Assessment of the Forecast Error Scores:

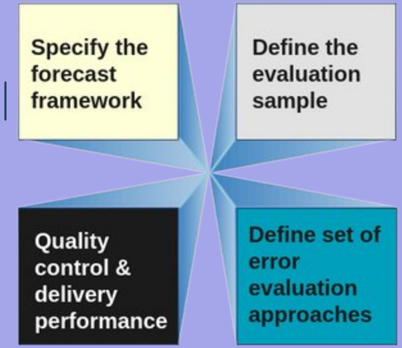
Score	HR	LR	IF weight	HR Final Score	LR Final Score
CRPS	1	0	3	3	0
CRPS leadtime	1	0	4	4	0
BrierScores	0	1	2	0	2
Hit Rate	0	1	1	0	1
False Alarm rate	1	0	2	2	0
Mean Score	0	1	1	0	1
CAL	0	1	1	0	1
DSC	0	1	1	0	1
UNC	-	-	1	-	-
AUC	0	1	1	0	1
SUM	3	6		9	7

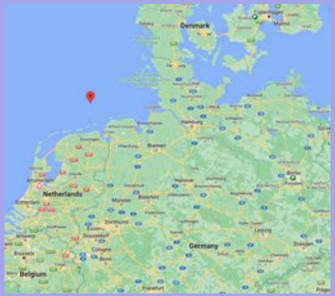
- For the raw (unweighted) scoring, the high-resolution (HR) setup has a lower composite score (is “worse”) than the low resolution (LR) setup
- If weights are applied according to specific targets of an application, the resulting assessment of the error metrics may change!
  In our example, we consider shorter lead-times (<12h) important and false alarms have high costs, which results in the HR being a better choice.

See also recommendations in chapter 15 of IEA Wind Recommended Practice book

Example 3: Wind measurement Evaluation at an Offshore site in the North Sea

Aim:
Verify performance and quality check of wind measurements with the help of ensemble forecasts





Definition of the Sample:
Wind and power measurements from Offshore platform FINO and Alpha Ventus wind farm, located ~45 kilometres to the north of the island Borkum in the North Sea

<u>Forecast type:</u>
75 Ensemble forecasts from MSEPS
15km grid cells with 32 vertical levels

Evaluation Approach:
- MEAN, BIAS, MAE, RMSE, CORRELATION
- Improvement over forecast
- Delivery Rate

Chapter 21, Section 21.5.1.3. Statistical tests and metrics for the QC process

Quality control of meteorological measurements in the real-time environment:

Recommended Principles for Wind Power Performace Control

Performance control of wind farms and wind turbines is best conducted in the following 3–4 steps:

a) Measuring basic meteorological parameters that can be used to compute power generation output

- wind speed and direction
- air temperature
- barometric pressure
- relative humidity

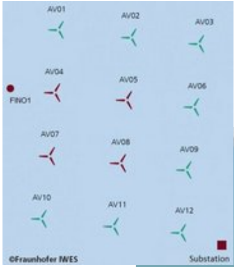
b) Conversion of the meteorological parameters into a power output

The best and recommended way is the IEC 61400-12-1 standard on power performance measurements, which is based on a physical formula (Equ. 2, chapter 8 [142])

c) Comparison of power output with measured and forecasted input variables

d) Visual Inspection with Ensemble generated Percentiles

Example Alpha Ventus +Fino1: Quality control of meteorological measurements in the real-time environment: Recommended test for met data performance control



Pictures taken from RAVE Flyer, © by Fraunhofer IWES

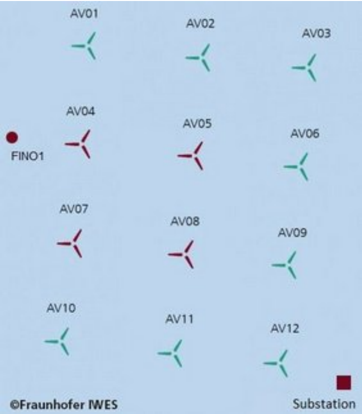
	STATISTICS								
ID Period	Data provision PART 1 (ws,T2m,wdir,ps)	WindSPEED (bias, rmse,corr Realistic values)	Temperature (bias, rmse,corr Realistic values)	WindDIR (bias, rmse,corr Realistic values)	Pressure (bias, rmse,corr Realistic values)	Installed Capacity [MW]	Improve-ment over Forecast >5%	Delivery Rate [%]	BIT MASK
		Good DATA							
2021q3 WAVUWT001	1111	1111	1111	1111	1111	60.0	2.19	99.8	15
capacity						60			
		Bad DATA || MiSSING DATA + Delivery < 98.5%							
2021q2 WAVM8T001	1001	1111	0001	0001	1111	5.0	6.57	10.6	9
WAVM7T001	1001	1111	0001	0001	1111	5.0	6.14	11.4	9
capacity						0			
		Bad Data || Missing data + Requirement 2: Improvement < 5%							
2021q1 WAVM7T001	0101	0111	1111	1001	1111	5.0	0	47.7	10
capacity						5.0			
	1=yes, 0=no								

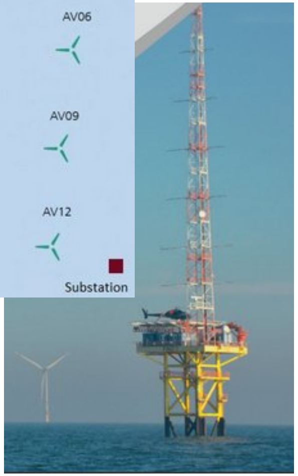
Explanation of BITMASK Available/missing Variables:	
0 or -	bad/missing
1	windspeed (ws)
2	temperature (T)
3	ws+temperature
4	wind direction (wd)
5	ws+wdir
6	wd + T
8	pressure (ps)
9	ws+ps
10	T+ps
11	ws+T+ps
12	wd+ps
13	ws+wd+ps
14	T+wd+ps
15	all variables delivered
1=ok, 0=bad, "-"=missing	

Explanation of columns WS||WDIR||TEMP||PS		
column	1	BIAS
	2	RMSE
	3	CORR
	4	data delivery of realistic values
1=ok, 0=bad, "-"=missing		

Example Alpha Ventus + Fino1: Quality control of meteorological measurements in the real-time environment

We reverse verification: measurement versus forecasts!





Pictures taken from RAVE Flyer, © by Fraunhofer IWES

Variable list and their threshold error limits

Var Number	Variable Name	Mininum Correlation	Maximum |Bias|	Maximum MAE
1	WindSpeed	0.65	3.0	3.0
2	AirTemp	0.75	2.0	2.5
3	WindDirection	0.55	13.0	20.0
4	AirPressure	0.9	50.0	85.0

Exemplary results from the Quality analysis of 6 Turbines & UW

Statis-tic rank	Windfarm ID	Test: ws|temp|wd|ps	wind speed WS	temp-erature T	wind direction WD	surface pressure PS	Description
1	AV07	1111	111	111	111	111	all tests ok
2	AV08	1111	111	111	111	111	all tests ok
3	UW	1110	111	111	111	000	PS fails all tests
4	AV09	1101	111	111	100	111	WD fails, except for WD(BIAS) OK
5	AV10	1101	111	111	101	111	WD fails, except for WD(MAE) OK
6	AV11	1010	111	000	111	110	T fails on all
7	AV12	1001	111	000	101	111	T fails and WD(MAE) fails

Criteria for “goodness” of data

Variable	unit	lower Limit	upper Limit
Wind speed (WS)	m/s	0	40
Wind direction (WD)	deg	0	360
Temperature (T)	°C	-40	40
Surface pressure (PS)	hPa	800	1100

Fino data: Wind, Temperature and Pressure
Turbines/UW: Wind & Power

Lessons Learned and Take-away

Forecast Evaluation is subjective… remember the 4 corner stones for meaningful evaluation

- select carefully what you verify
- use more than one score/metric
- check consistency by applying various tests
- evaluate the significance of your results

THANK YOU FOR YOUR ATTENTION

Questions

Follow us:

Project webpage: http://iea-wind.org/task51

Publications: https://iea-wind.org/task51/task51-publications

RP-page: https://iea-wind.org/task51/task51-publications/task51-recommended-practices/

Contact us…

Presenter:

Dr. John W. Zack
WP2 Leader
MESO, Inc, USA
jzack@meso.com

MESO
ATMOSPHERIC RESEARCH
FORECASTING • SIMULATIONS

Co-authors:

Dr. Corinna Möhrlen
WP3 Leader
WEPROG, DE & DK
com@weprog.com

WEPROG

Dr. Mathias Blicher B.
DTU Compute
Denmark
matbb@dtu.dk

DTU

Dr. Gregor Giebel
Operating Agent
DTU Wind, Denmark
grgi@dtu.dk

DTU