Challenges and best practice in energy forecast evaluation

Dr Jethro Browell
Lecturer and EPSRC Innovation Fellow

University of Strathclyde, Glasgow, UK
jethro.browell@strath.ac.uk

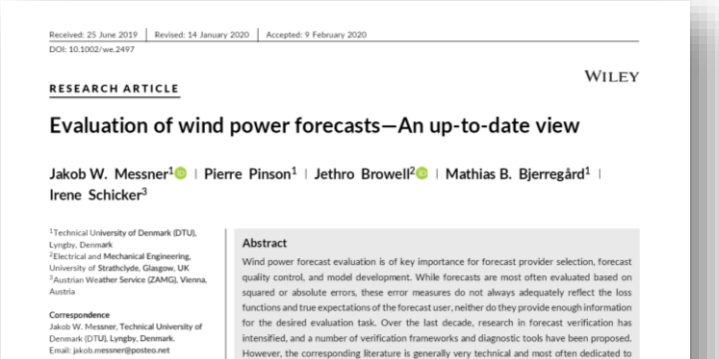
Avon Local Group of the Royal Statistical Society
29 September 2020

Innovation Fellowship EP/R023484/1

Contents

- Energy forecasting: who, what, why?
- Starting point: Choosing the *"best"* forecast/vendor
  - Metrics (briefly)
  - Comparing forecast performance
  - Practical issues
- Challenges and best practice:
  - Running trials (research and commercial)
  - Quality vs Value Example: Hydro Power

Thanks to collaborators! I'll be drawing on joint work:



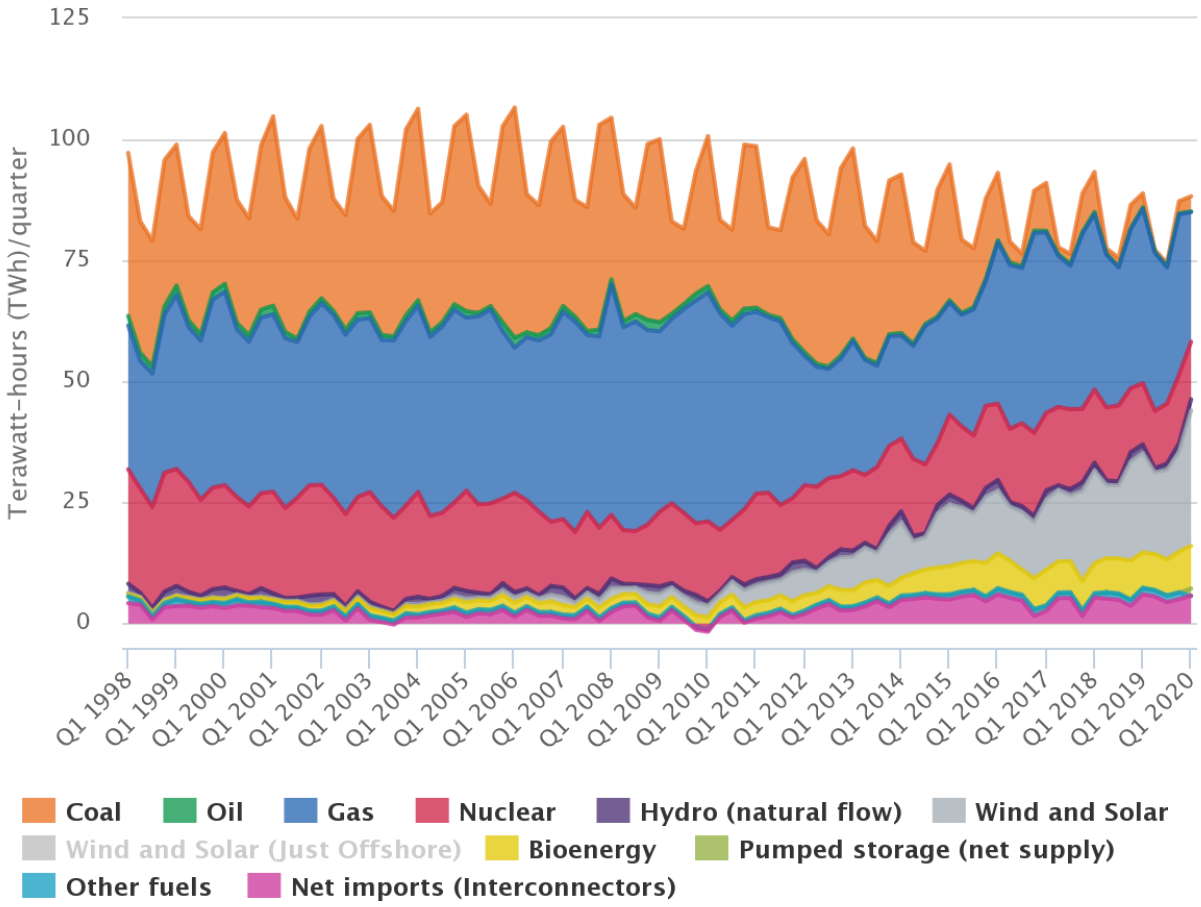
https://www.ieawindforecasting.dk/Publications/RecommendedPractice

https://doi.org/10.1002/we.2497

Energy Forecasting

University of Strathclyde Glasgow

Electricity generation mix by quarter and fuel source (GB)



It is getting much harder to manage!

28% Wind + Solar in Q1-2020!!!


BBC bob News Sport Weather iPlayer Sounds

NEWS

Home Coronavirus US Election UK World Business Politics Tech Science

Science & Environment

UK can be 'Saudi Arabia of wind power' - PM

By Paul Rincon
Science editor, BBC News website

24 September 2020 Share

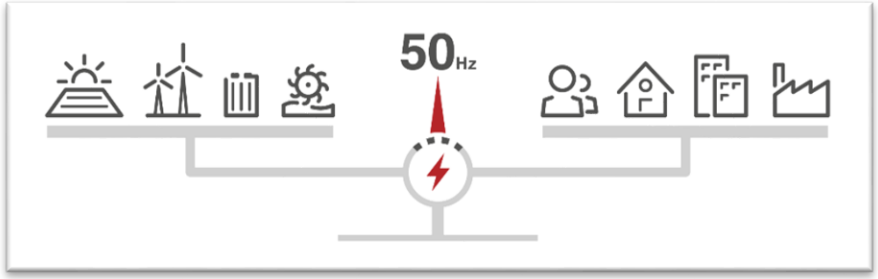
https://www.ofgem.gov.uk/data-portal/electricity-generation-mix-quarter-and-fuel-source-gb

Energy Forecasting

Electricity is a *special* commodity!

- Supply must meet demand in *real time*
- Delivery subject to physical restrictions
- Storage is uneconomic (today…)

Slightly different picture for heat and gas…



Source: blog.ebl.ch/durch-regelenergie-zur-sicheren-stromversorgung/

Who’s who: Forecast Users

- Generators of energy
  - Selling wholesale energy

Who's who: Forecast Users

- Generators of energy
  - Selling wholesale energy
- Suppliers/Retailers
  - Purchasing wholesale energy

Who's who: Forecast Users

- Generators of energy
  - Selling wholesale energy
- Suppliers/Retailers
  - Purchasing wholesale energy
- Network operators
  - Maintain security of supply: technical constraints on delivery

Who's who: Forecast Users

- Generators of energy
  - Selling wholesale energy
- Suppliers/Retailers
  - Purchasing wholesale energy
- Network operators
  - Maintain security of supply: technical constraints on delivery
- Others
  - Aggregators
  - Energy market speculators

Who's who: Forecasters

University of Strathclyde Glasgow

- **National and private weather centres** produce global and regional numerical weather prediction (NWP)

Weather Forecasts

- **Forecast vendors** produce and sell specific weather and energy forecasts

Specialised Weather and **Power** Forecasts

Software tools for interacting with forecasts

- **Forecast users** procure weather and/or power forecast to present to decision-makers on trading desks and in control rooms

Wide range of models from "in-house vendors" to complete dependency on service providers

Choosing the *"best"* forecast

Motivation

> best
>
> /bɛst/\
> *adjective*
> 1.of the most excellent or desirable type or **quality**.
> "the best forecast leads to the better decisions"

Need to be objective:

- Commercial integrity: tenders, financial reward/penalty
- Regulatory compliance
- Research integrity
- ➔**Metrics**

Google's English dictionary, provided by Oxford Languages.

Choosing the *"best"* forecast

Motivation

> best
>
> /bɛst/\
> *adjective*
> 1.of the most excellent or desirable **type** ~~or~~ **and quality**.
> "the best forecast leads to the better decisions"

Need to be appropriate:

- Forecast matched to end-use
- Quantitative or qualitative use?
- Quantify uncertainty?
- Dependency with other quantities?
- Minimise cost/risk/**metric**?

Need to be objective:

- Commercial integrity: tenders, financial reward/penalty
- Regulatory compliance
- Research integrity
- ➔**Metrics**

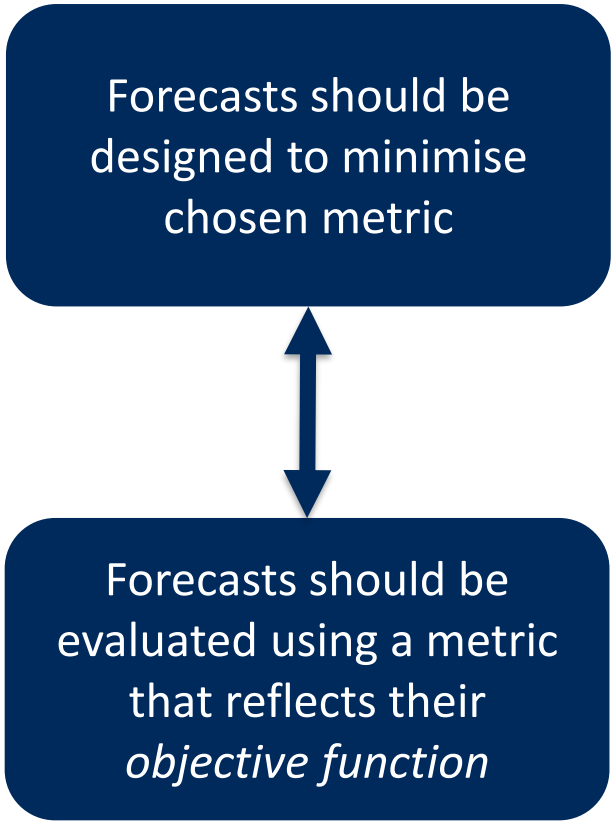
Google's English dictionary, provided by Oxford Languages.

Choosing the *"best"* forecast

Metrics – deterministic

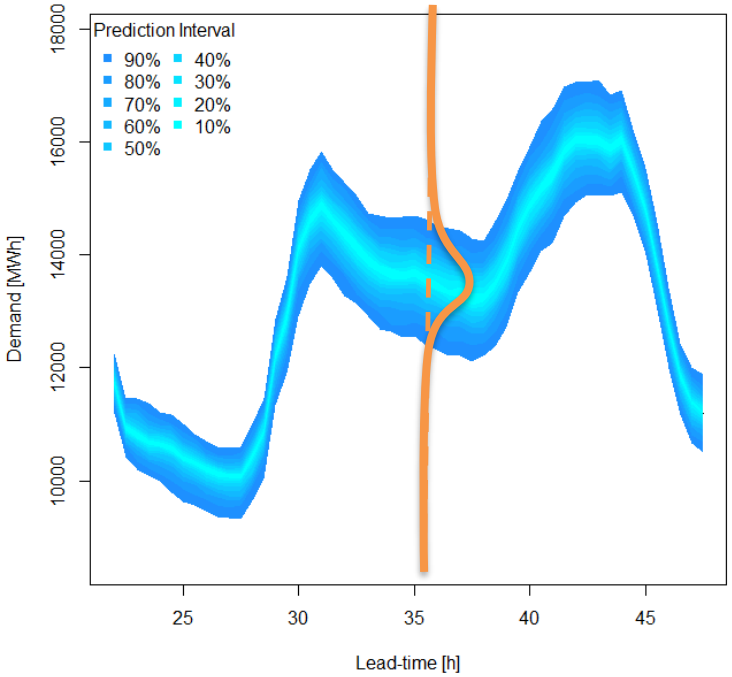
Lots of metrics based on:

- Squared error
  - Exact solution to OLS – nice!
  - Conditional mean
  - Disproportionality penalises large errors
- Absolute error
  - Numerical solutions – fine!
  - Conditional median
  - Reflects costs proportional to error
- Percentage error
  - Many issues: zero/∞, asymmetric penalty of –ve/+ve errors
  - Not recommended!



Choosing the *"best"* forecast

Metrics – probabilistic



A few metrics:

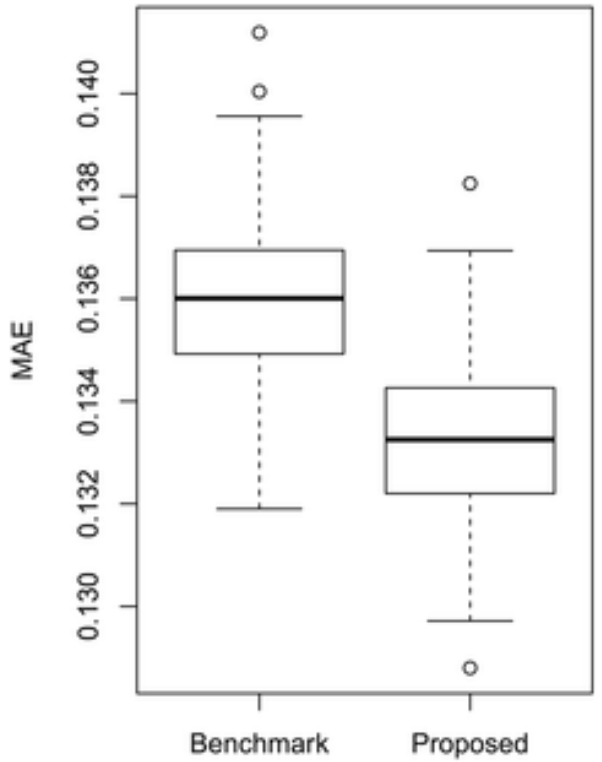
- Reliability/calibration
  - Necessary property
- Pinball/CRPS
  - Strictly proper
  - Same objective as fitting quntile regression
- Log Score
  - Strictly proper
  - VERY sensitive to tails
- Multivariate extensions…

Principle of *"sharpness subject to calibration."*

Choosing the *"best"* forecast

Comparing Forecasts

- Metrics make comparison possible, but we still need to take care
  - Are small differences significant?
- We have a few options:
  - Diebold-Mariano Test
  - Bootstrap Resampling
- Example: two wind power forecast providers…



	Benchmark	Proposed
Normalised Mean Absolute Error	0.136	0.133

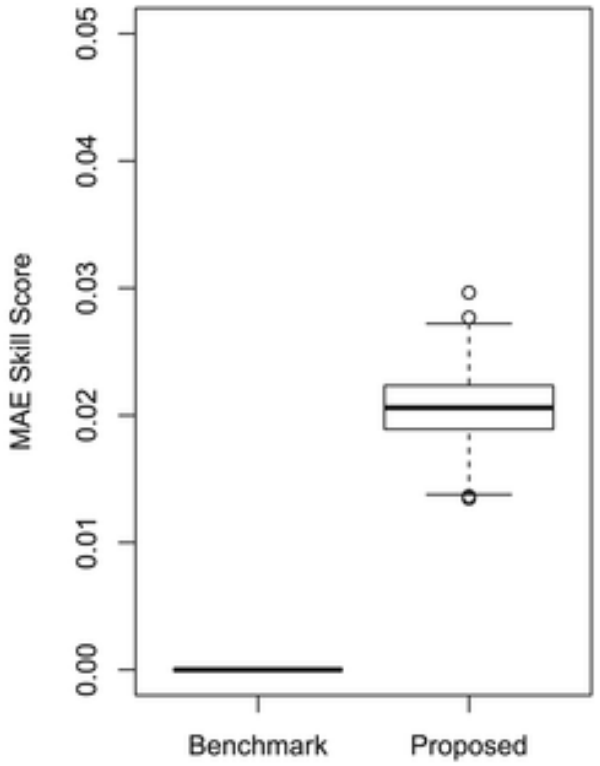
Choosing the *"best"* forecast

Comparing Forecasts

- Skill scores can provide better discrimination:

- $\text{Skill Score} = \frac{S_{ref} - S}{S_{ref} - S_{perf}}$

  - S Candidate forecast
  - S_{ref} Reference/benchmark
  - S_{perf} Perfect score (often 0)



Choosing the *"best"* forecast

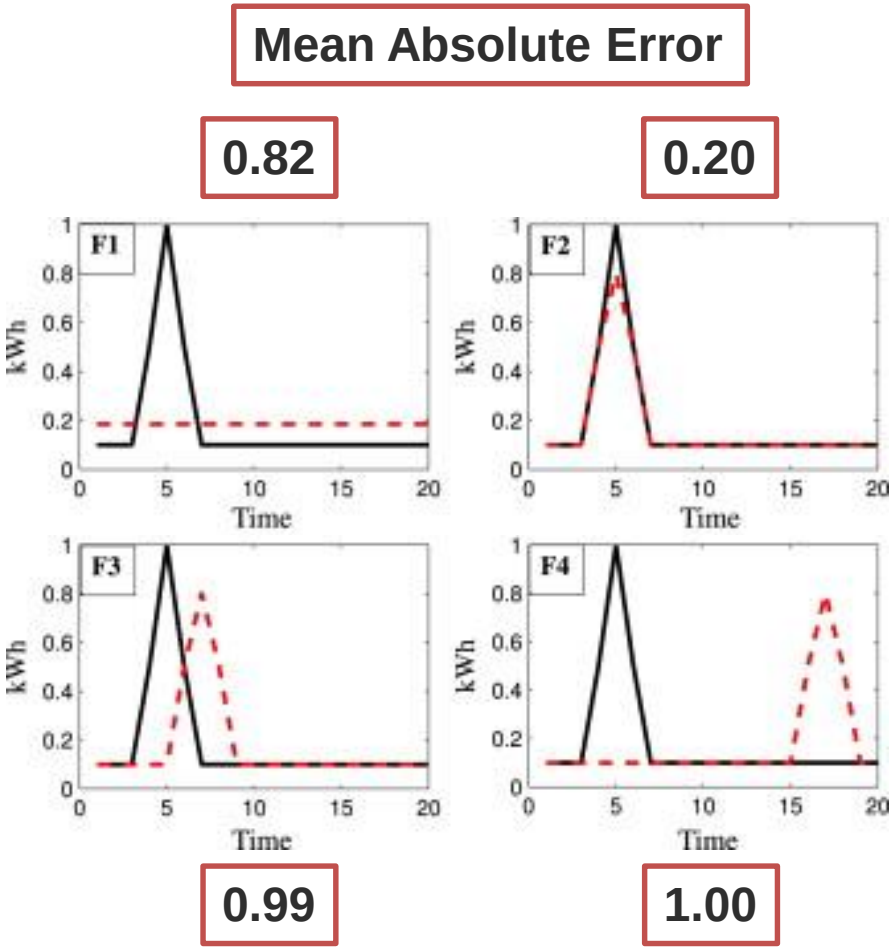
Challenges

Matching metrics with end-use:

- Double penalty for space/time errors (right)
- Discrimination between error types poor/difficult to interpret
- User must encode trade-off between error types…

Solution 1

- Forecast full multidimensional density
- Use *(Strictly) Proper* scoring rules
- …complex and decision support require for practical use



Source: Haben et al. "A new error measure for forecasts of household level, high resolution electrical energy consumption," DOI: 10.1016/j.ijforecast.2013.08.002.

Choosing the *"best"* forecast

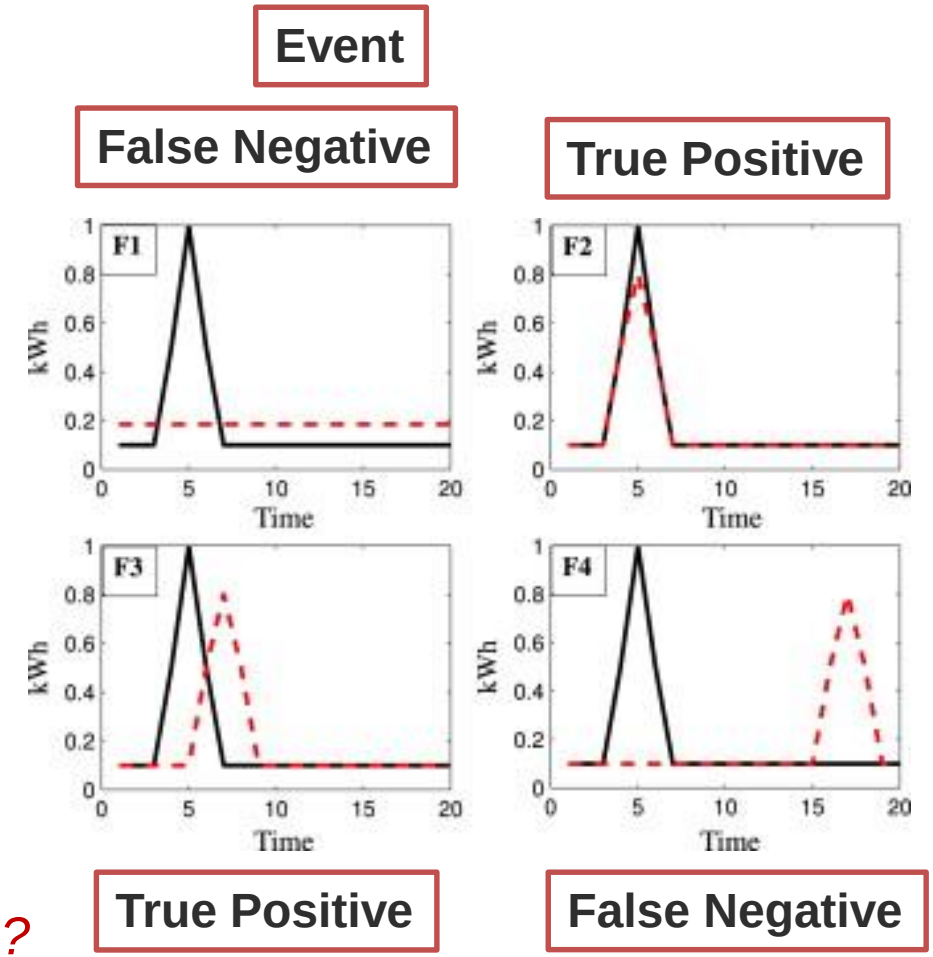
Challenges

Matching metrics with end-use:

- Double penalty for space/time errors (right)
- Discrimination between error types poor/difficult to interpret
- User must encode trade-off between error types…

Solution 2

- Event-based forecasts
- Confusion matrices and associated metrics
- E.g. *Spike between t=0 and t=10?*



Source: Haben et al. "A new error measure for forecasts of household level, high resolution electrical energy consumption," DOI: 10.1016/j.ijforecast.2013.08.002.

Choosing the *"best"* forecast

Practical Issues

- Cheating is easy, and can be accidental!
- Multiple uses for a single forecast product – can it be *"optimal"* for them all?
- Factors such as customer service and availability guarantees are important for users

Best practice: careful and considered study/trial design

Trials and Research Studies

Challenges

- Designing a study or trial:
  - Temporal and spatial coverage
  - Metric/Evaluation framework
  - Scope: operational realities including data quality issues
- Running a study or trial:
  - Resource to set-up and monitor
  - Expertise to monitor and evaluate

IEA Wind Recommendations

Trials should be:

- Representative
- Significant
- Relevant

Trials and Research Studies

Best Practice

Challenge	Best Practice
Temporal and spatial coverage	• Cross-validation for model development • Testing on withheld test data • **Representative** seasonal and situational coverage • Large enough for **significant** results • R&D: Use open data, enter competitions • Commercial: Live trial
Metrics and evaluation	• Framework comprising a set of metrics and error analysis **relevant to use-case** • Competitive benchmarks required for context
Scope	• Define at outset, adjust evaluation framework accordingly • Must be **relevant to use-case** or research question
Resources	• Make use of open tools and training material • Consult experts • Check IT requirements as part of design phase!

Quality vs Value

Example: Extended-range Hydro Inflow Forecasting

Context:

- Range of weather forecasts extending all of the time
- Hydro plant incentivised to operate flexibly

Opportunity?

- Can be forecast inflow beyond one week-ahead?
- What is the value?

Work with Robert Graham &

Image: https://www.sserenewables.com/hydro/great-glen/

Quality vs Value

Example: Extended-range Hydro Inflow Forecasting

Quality vs Value

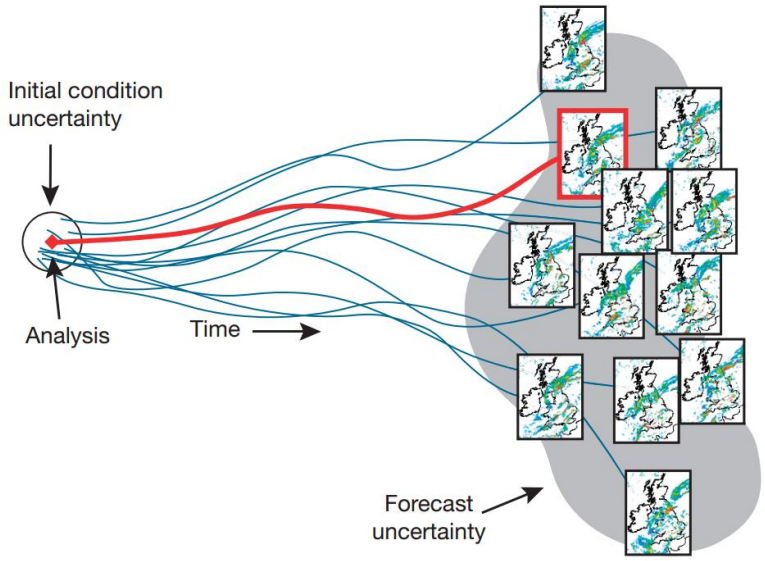
Example: Extended-range Hydro Inflow Forecasting

University of Strathclyde Glasgow

Weather Forecast → Inflow & Water Level Forecast → Operational Decisions

- Probabilistic forecasts of precipitation and temperature
- Produced by Ensemble Numerical Weather Prediction

- Evaluation: Skill score (**relative to model climatology**)
- Weekly totals 1 to 6 weeks ahead



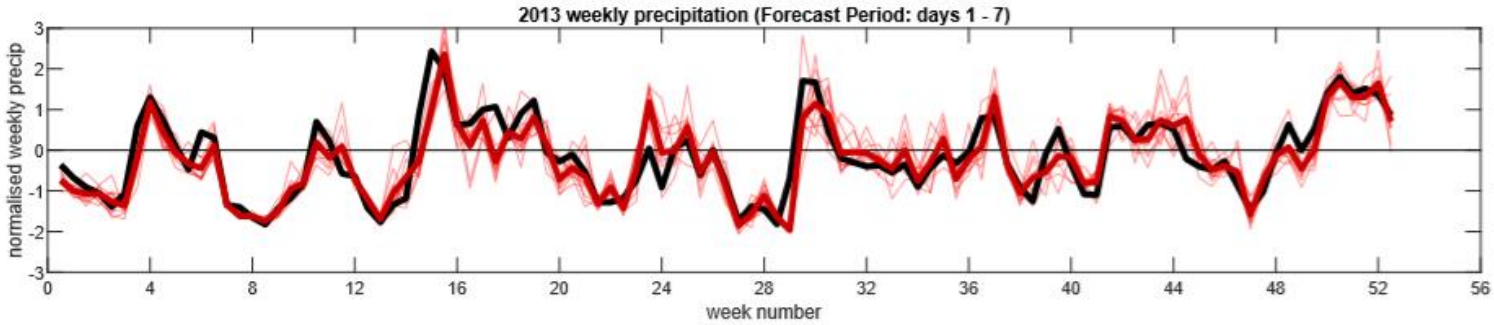
Source: Bauer, Thorpe and Brunet, *"The quiet revolution of numerical weather prediction,"* Nature, 2015

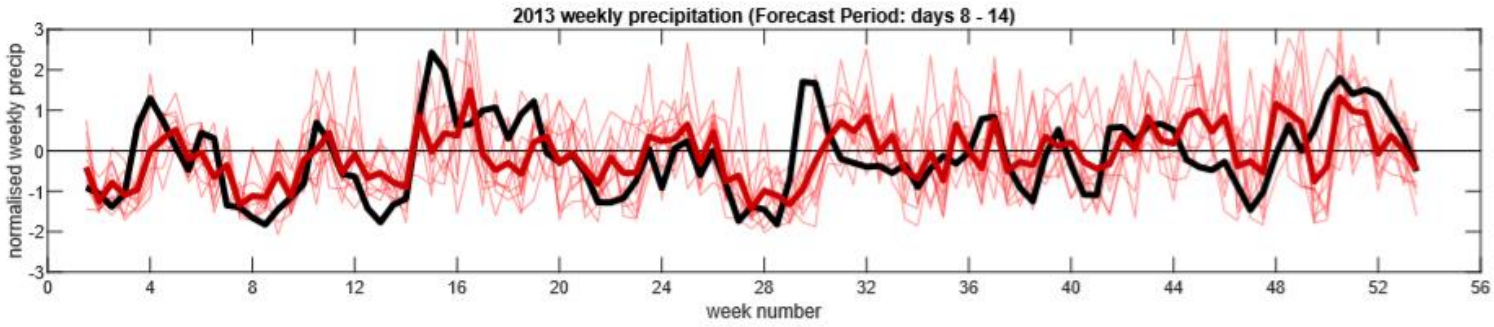
Quality vs Value

Example: Extended-range Hydro Inflow Forecasting

University of Strathclyde Glasgow

Weather Forecast → Inflow & Water Level Forecast → Operational Decisions





Forecast **Mean** and Ensemble Members

Actual Precipitation

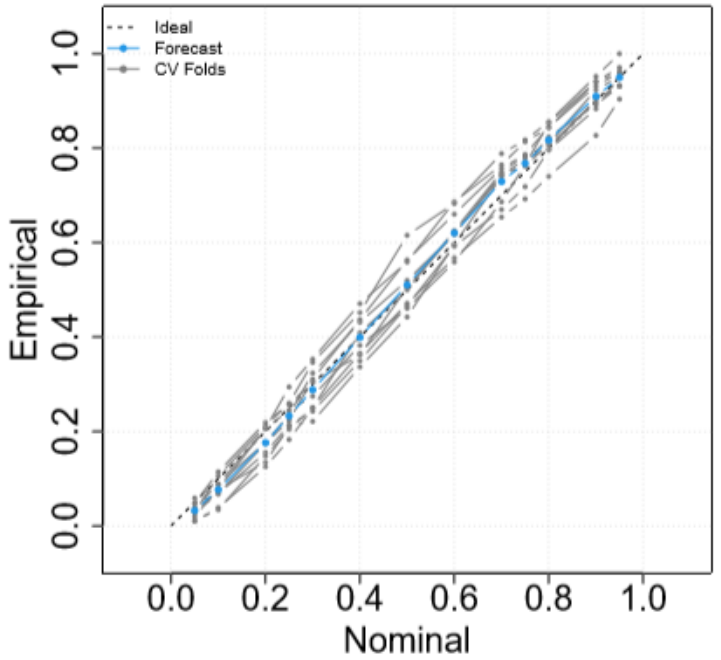
Quality vs Value

Example: Extended-range Hydro Inflow Forecasting

University of Strathclyde Glasgow

Weather Forecast → **Inflow & Water Level Forecast** → Operational Decisions

- Probabilistic forecast of inflow at specific location
- Post-processing required: *gamlss* model with Gamma distribution
- Evaluation relative to observations: calibration and skill



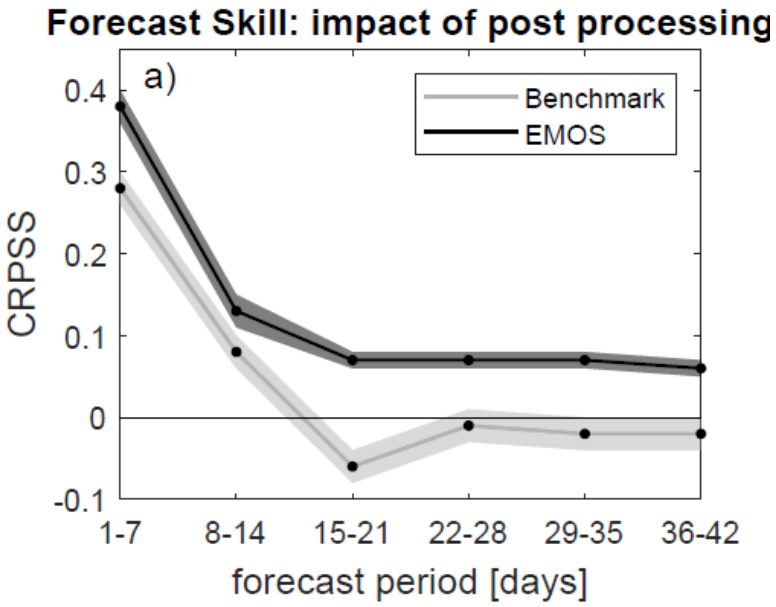
Quality vs Value

Example: Extended-range Hydro Inflow Forecasting

University of Strathclyde Glasgow

Weather Forecast → **Inflow & Water Level Forecast** → Operational Decisions

- Probabilistic forecast of inflow at specific location
- Post-processing required: *gamlss* model with Gamma distribution
- Evaluation relative to observations: calibration and skill



Quality vs Value

Example: Extended-range Hydro Inflow Forecasting

Need to capture how forecast inform decisions, working with end users:

1. Day-to-day: Change generation schedule to manage water level – converting water to energy at peak or off-peak price
   - Costs associated with changed generation schedule (based on forecast)
   - Costs associated with resulting inflow and water level (based on actuals)
2. Large inflow events: Avoid bad situations – water spill, safe limits on water levels and down-stream flow

Quality vs Value

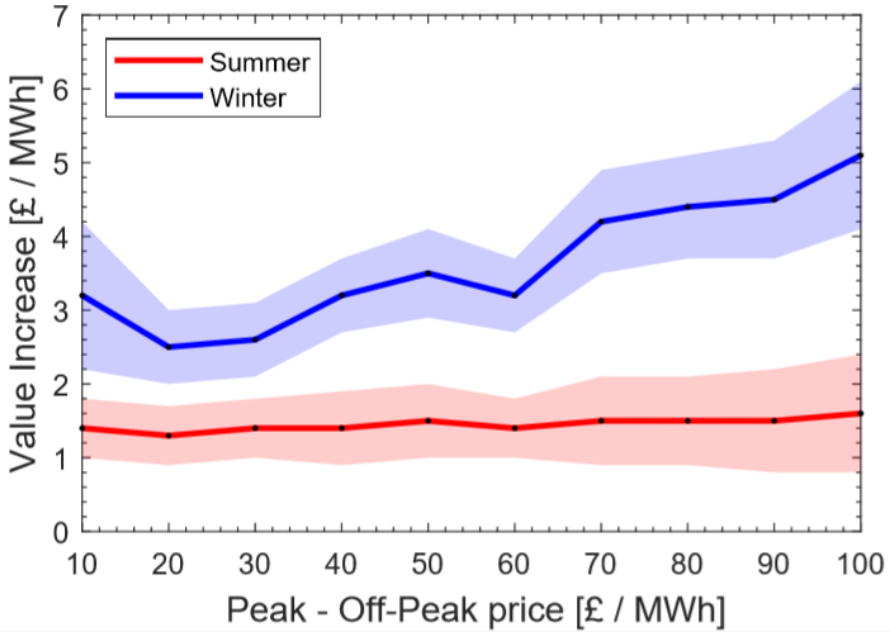
Example: Extended-range Hydro Inflow Forecasting

University of Strathclyde Glasgow

Weather Forecast → Inflow & Water Level Forecast → **Operational Decisions**

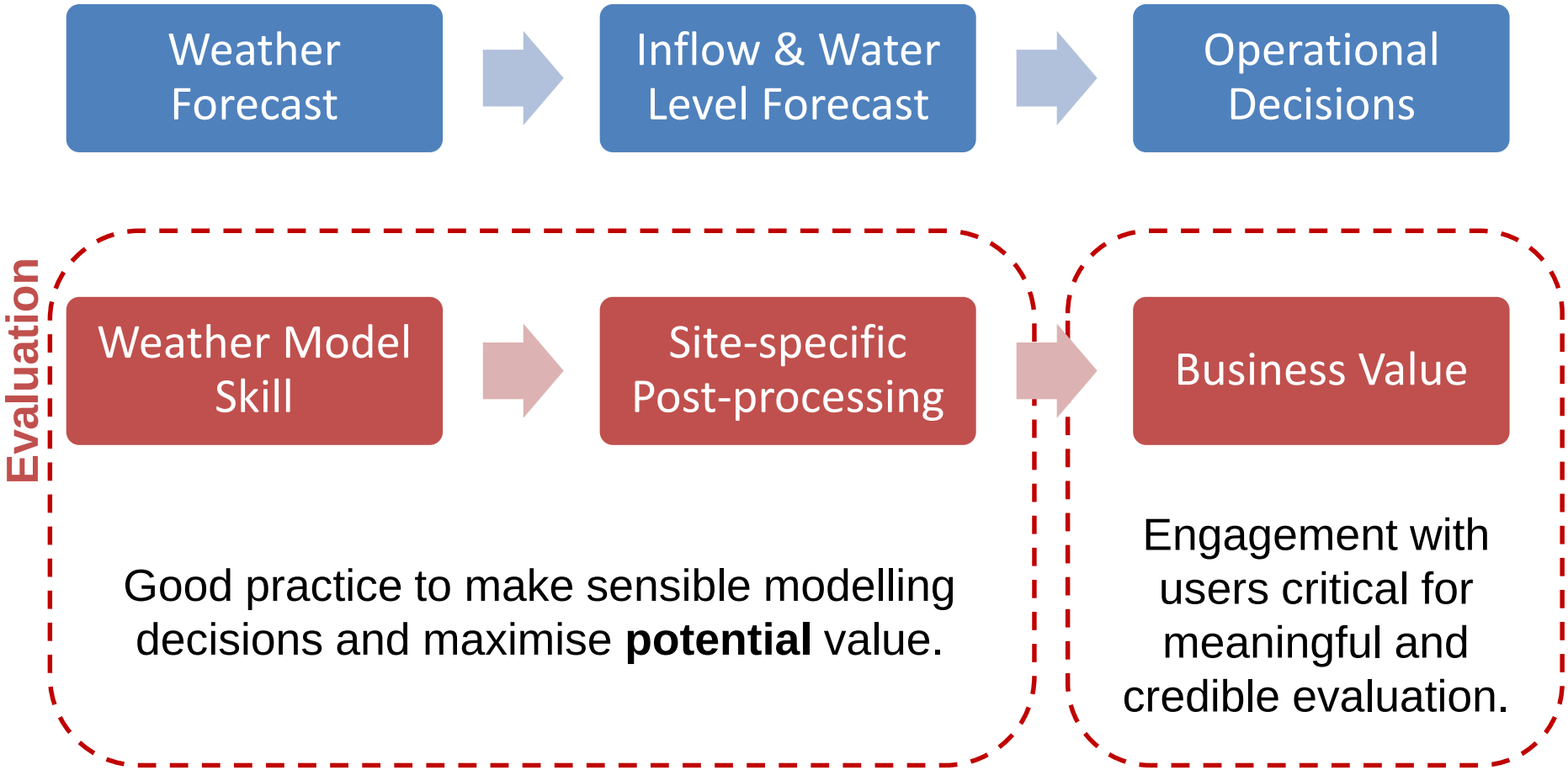
Compare “value of water” based on using climatology vs forecast:

- Stylised model, but indicative numbers
- Significant value!
- Likely that some of this value already realised
- Is value sufficient to justify new product and change in practice?



Quality vs Value

Example: Extended-range Hydro Inflow Forecasting



Summary

Energy Forecast Evaluation

University of Strathclyde Glasgow

- Energy systems are changing fast – forecast quality has never been more important!
- We have tools for (comparative) evaluation, but need to apply them with care
- Evaluation of forecast products should be:
  - Representative
  - Significant
  - Relevant
- Collaboration between forecasters and end-users is invaluable!

Thanks! Questions?

Papers and more at jethrobrowell.com

University of Strathclyde Glasgow