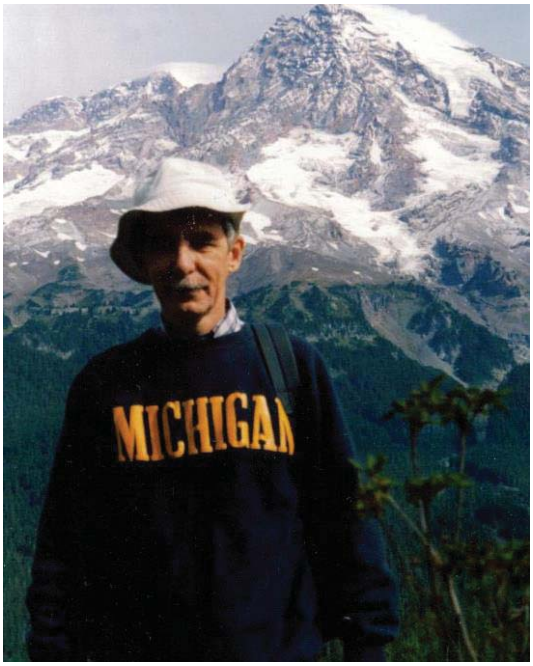
Challenge to Develop and Demonstrate New User-Oriented Forecast Verification Metrics

Barbara Brown (NCAR), Beth Ebert (BOM),
Thomas Haiden (ECMWF), Martin Göber (DWD),
Marion Mittermaier (Met Office), and
Pertti Nurmi (FMI)

HIW Workshop

April 2016

"Forecasts contain no intrinsic value. They acquire value through their ability to influence the decisions made by users of the forecasts."



"Forecast quality is inherently multifaceted in nature… however, forecast verification has tended to focus on one or two aspects of overall forecasting performance such as accuracy and skill."

Allan H. Murphy, *Weather and Forecasting*, **8**, 1993: "What is a good forecast: An essay on the nature of goodness in forecasting"

Relationship between "Quality" and "Value"

Improving the quality of a forecast (i.e., verification scores) does not guarantee an increase in forecast value

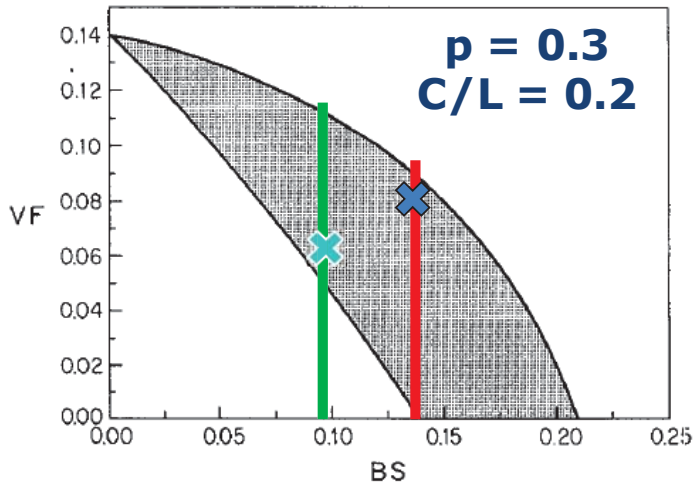


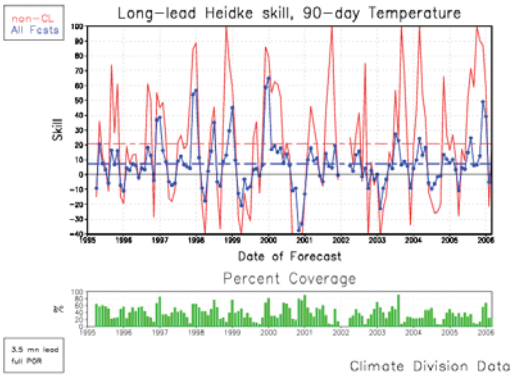
FIG. 4. Relationship between forecast accuracy and forecast value in the cost–loss ratio situation, with climatological probability $\pi = 0.3$ and cost–loss ratio $C/L = 0.2$ (taken from Murphy and Ehrendorfer 1987).

From Murphy 1993

Levels of user focus

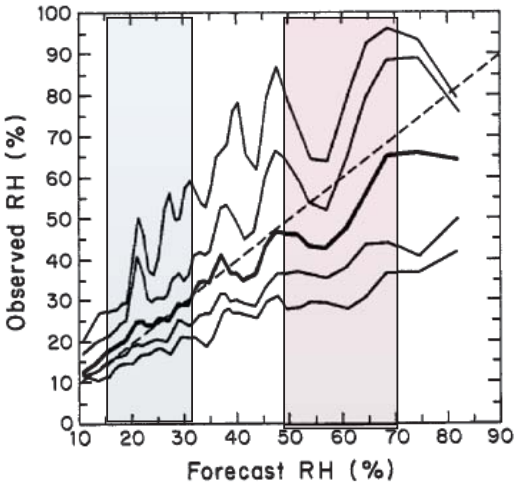
Level 0: Conventional *measures-based approaches*

- Best for administrative purposes



Level 1: Broad *diagnostic approaches*

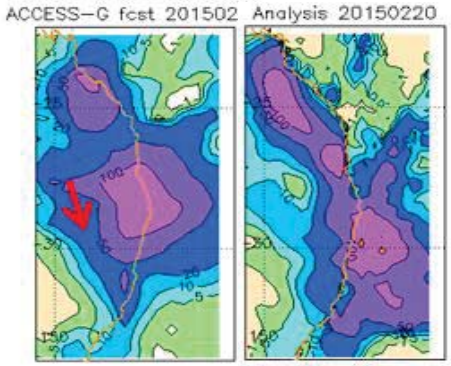
- Evaluate variables of interest to users
- User-selectable information (stratifications, thresholds)
- Often graphical
- Confidence intervals



Levels of user focus

Level 2: *Features-based and enhanced diagnostic* approaches applied

- Evaluation of **multiple attributes** of broad interest to users



Ebert-McBride example

Level 3: *User-specific* approaches and measures

- Interact closely with users to determine meaningful approaches and measures
- May include specialized datasets that are user-specific

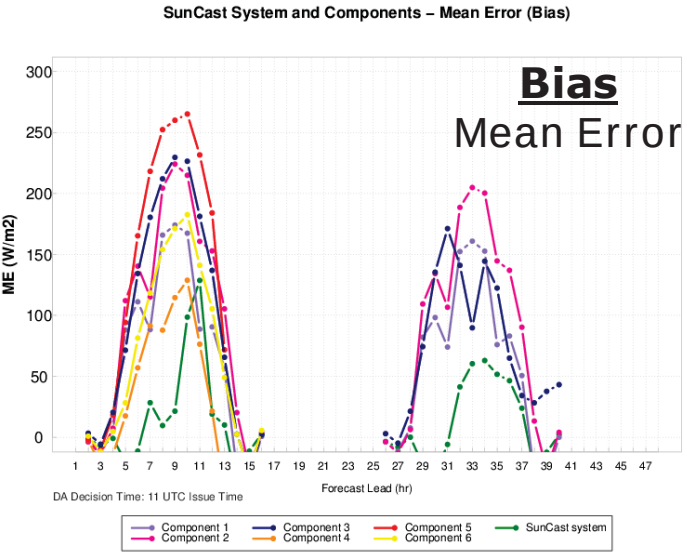
Level 4: *Forecast value* estimated, making use of user-focused verification information

- Close interaction with users
- Deep understanding of users' decision-making and applications of forecasts

Current status

- Typical verification is at Levels 0 or 1
- Level 2 - features based – is becoming more common as people see the advantages of these approaches
- Levels 3 and 4 are very limited and still often hard to reach
- **Hence**: The User-oriented verification method **Challenge**!
  - An opportunity for clever people to move this field forward in significant ways
- First some examples…

Solar Power forecasts:
Adapting standard metrics to meet user needs



Intra Hour

Day Ahead

Comp1
Comp2
Comp3
Comp4
Comp5
Comp6
SunCast

Preferences for end users:
Raw Values (Non-normalized)
Normalized by Capacity/Clear Sky
Normalized by Actuals

User: Energy Trader
Sum of Errors over entire day more informative

Sum of Errors over Forecast Period

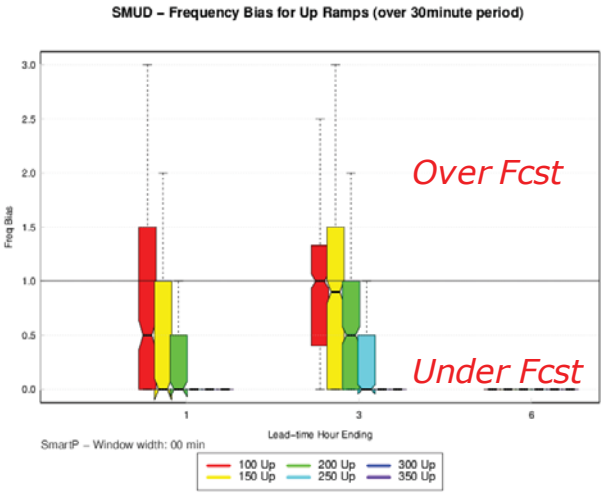
	Intra Hour	Day Ahead
Comp1	960 (-76 / 1036)	858 (-79 / 937)
Comp2	1418 (-43 / 1461)	1295 (-43 / 1339)
Comp3	1433 (-33 / 1466)	1106 (-5 / 1111)
Comp4	510 (-84 / 594)	n/a
Comp5	1748 (-12 / 1760)	n/a
Comp6	1004 (-40 / 1044)	n/a
Blended Model	226 (-92 / 318)	178 (-110 / 288)

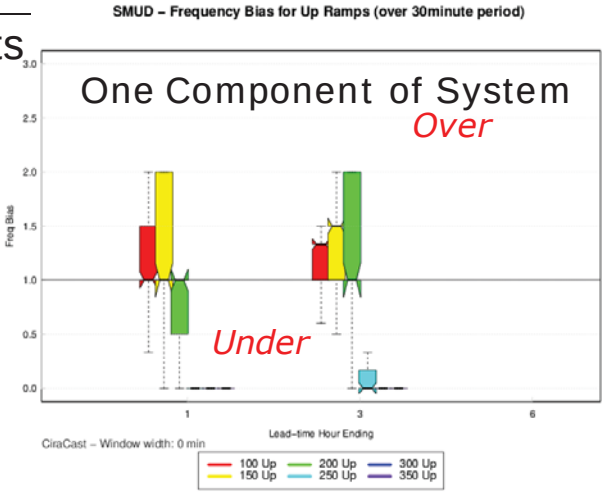
Courtesy T. Jensen

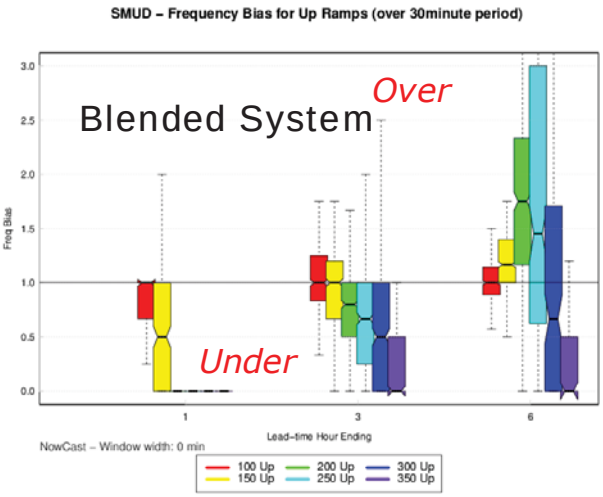
Capturing Energy Ramps

Frequency Bias = $\frac{\text{\# of Fcst Events}}{\text{\# Observed Events}}$

Baseline







100 Up
150 Up
200 Up
250 Up
300 Up
350 Up

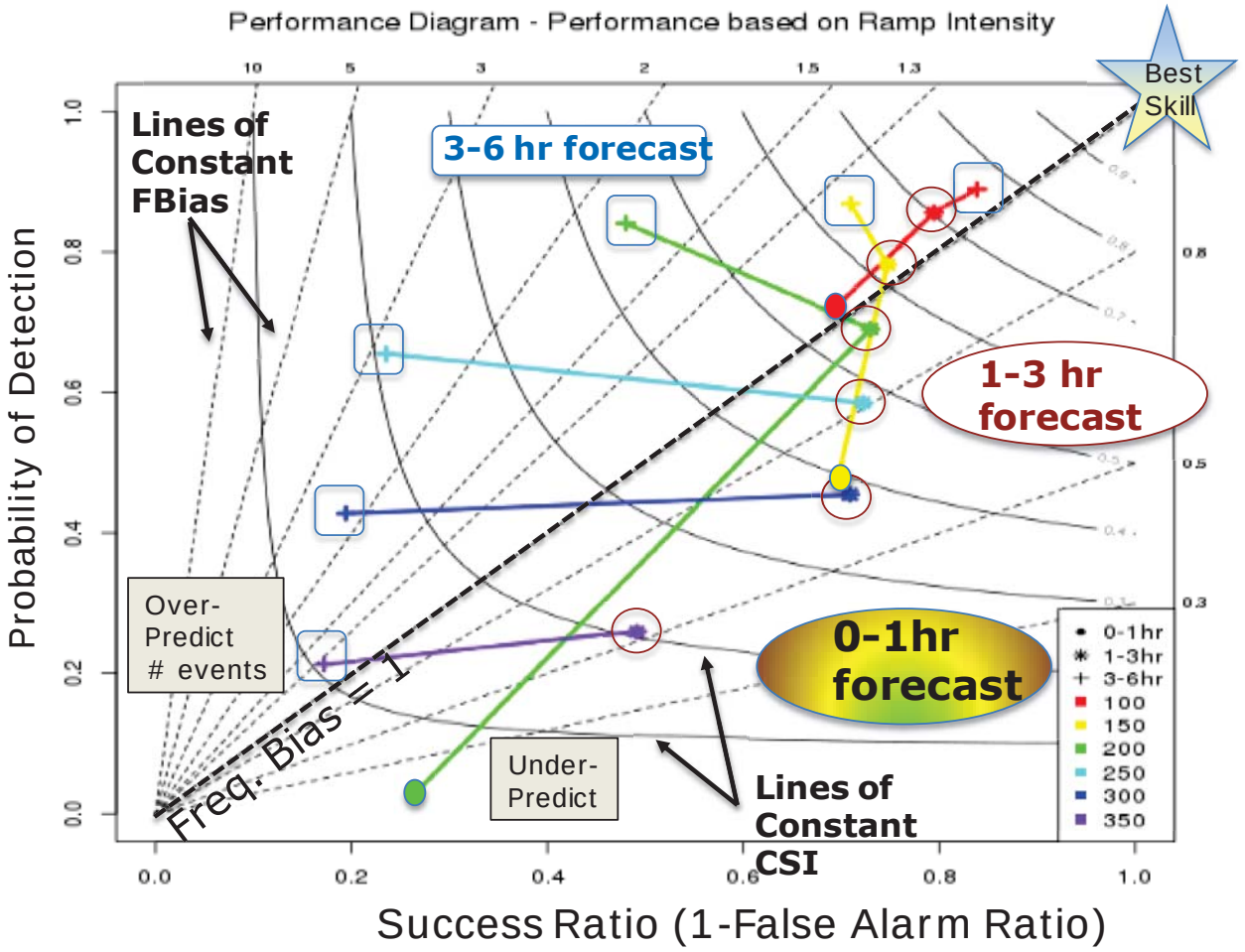
Credit: T. Jensen

Performance Diagram

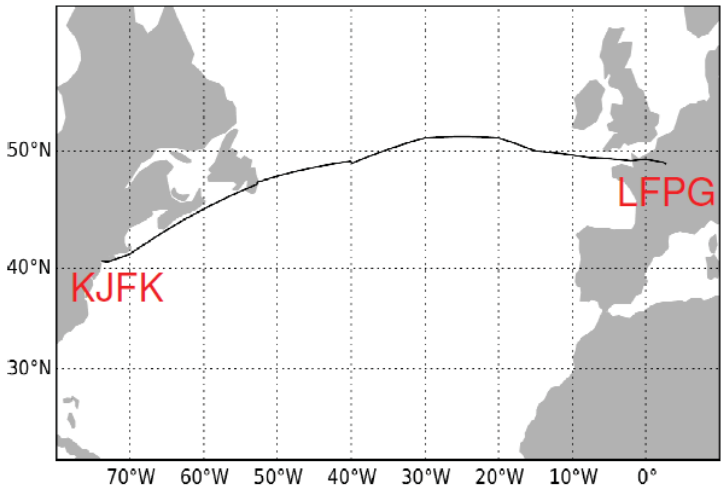
User: Forecaster

Easy way to display 4 Categorical Metrics at one time and assess overall skill

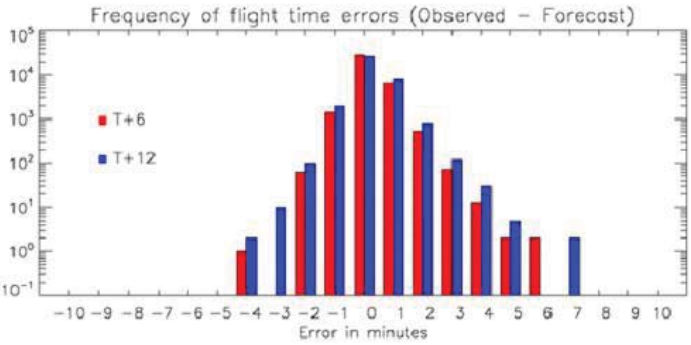
This example shows stratification by ramp intensity for 3 short range forecast bands



Example: flight time error (FTE) in aviation

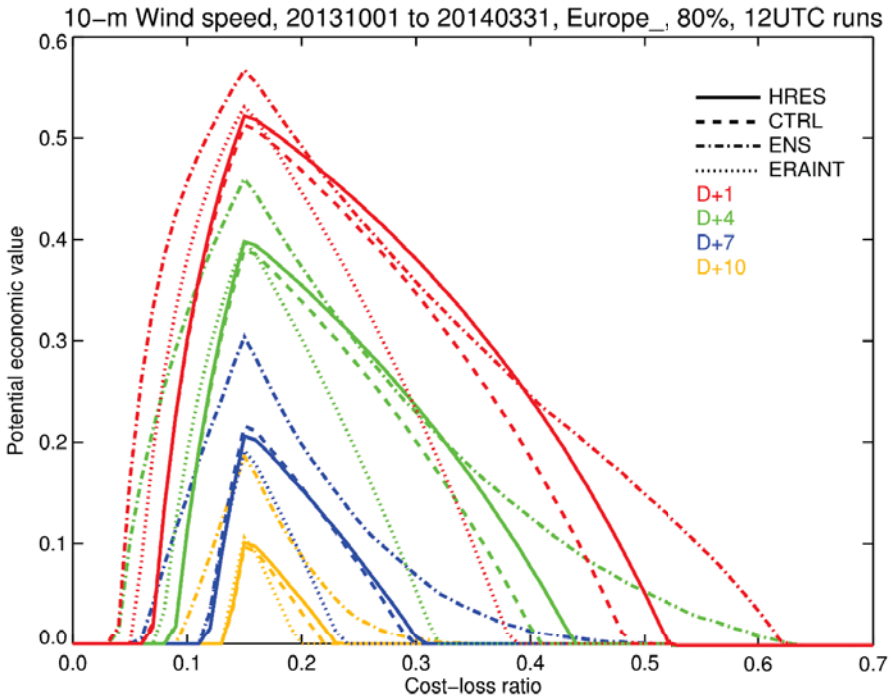


(Source: UK Met Office)



FTE model – equivalent to a non-linear 4-D weighting of forecast wind speeds along a trajectory

Example: cost/loss framework



PEV of forecasts of wind speed >80th percentile (Source: ECMWF)

← different users →

Other user-oriented verification methods

- Conditional verification tailored to the user (e.g. wind speed error when there is snowfall, for snow drift forecasts)
- Object-oriented verification of the onset and cessation of events (e.g. for aviation)
- . . .

Much left to be explored by creative minds!

Challenge for best new user-relevant verification method

- Sponsored by **WMO Joint Working Group on Forecast Verification Research and WWRP High Impact Weather (HIW) project; joined by S2S and Polar Prediction**
- **Focus:**
  - Applications: all applications of meteorological and hydrological forecasts
  - Metrics can be quantitative scores or diagnostics (e.g., diagrams)
- **Criteria for being selected as "best":**
  - Originality, user relevance, intuitiveness, simplicity and ease of computing, robustness, and resistance to hedging.
  - Desirable characteristics: (i) *Clear statistical foundation*; (ii) *Applicability to a broader set of problems*
- **Dates**:
  - Formal announcement: Sept 2015 (here and now!)
  - Deadline for submission: 31 Oct 2016
- **Prize**: **Invited keynote talk at the 7th International Verification Methods Workshop in 2017**

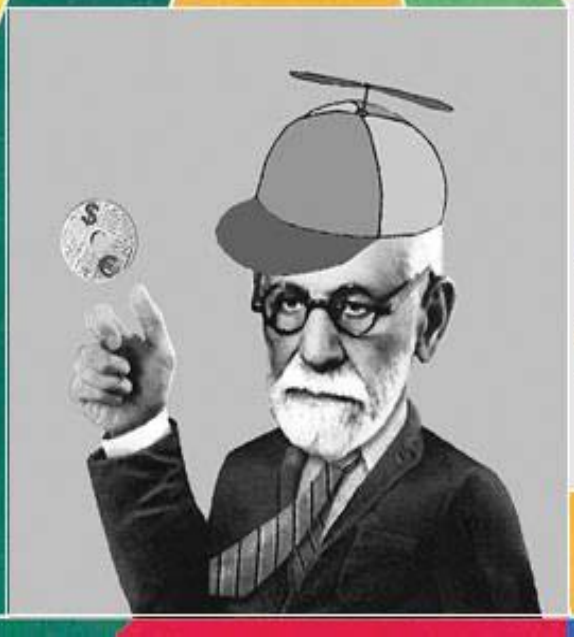
Contact verifchallenge@ucar.edu for more information

See website at

http://www.wmo.int/pages/prog/arep/wwrp/new/FcstVerChallenge.html

Develop and Demonstrate the Best New User-Oriented Forecast Verification Metric

Challenge



Contest run by WMO Joint Working Group on Forecast Verification Research in support of the WWRP High Impact Weather Project

Aim: Promote user-oriented verification, that is, quantitative assessment of forecast quality in terms that are meaningful to particular kinds of forecast users

Scope:

- All applications of meteorological and hydrological forecasts
- Users include industry, emergency management, public, … many possibilities!
- Metrics may be scores or diagrams, must be *new*
- Anyone with a good idea (individuals, teams) can enter

Prize: Paid attendance and keynote talk at next International Verification Methods Workshop in 2017. All participants will be invited to submit an abstract to the workshop.

How to enter:

www.wmo.int/pages/prog/arep/wwrp/new/FcstVerChallenge.html

Timeline:

- **Challenge begins: September 2015**
  - **Deadline for entries: 31 October 2016**
    - **Announcement of winner: : Jan 2017**

Further information
verifchallenge@ucar.edu

Joint Working Group on
Forecast Verification Research

Develop and Demonstrate the Best New User-Oriented Forecast Verification Metric

Please correspond by 31 October 2016

verifchallenge@ucar.edu