Forecast verification

….. a few comments

Anna Ghelli

anna.ghelli@ecmwf.int

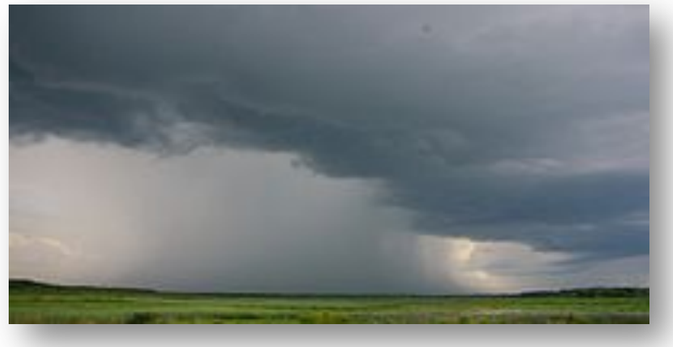
ECMWF

Forecast quality versus forecast value

A forecast has high QUALITY if it predicts the observed conditions well according to some objective or subjective criteria.

Quality but no value

A forecast has VALUE if it helps the user to make a better decision.



Value but no quality

ECMWF

Verification goals and process

What are our goals with forecast evaluation?

- Evaluate usefulness of forecasts
  - In general?
  - For specific users?
- Improve ensemble and modeling system
- Track changes in forecast performance over time

PROCESS

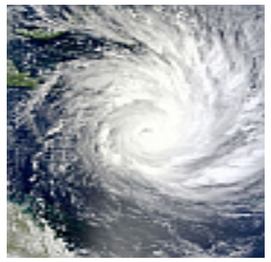
- Start by determining
  - ***What are the questions we want to answer??***

ECMWF

Impact forecasts

Floods

Courtesy Beth Ebert



Tourism

Air travel

Energy

Sports

Agriculture

Roads

Emergency management

ECMWF

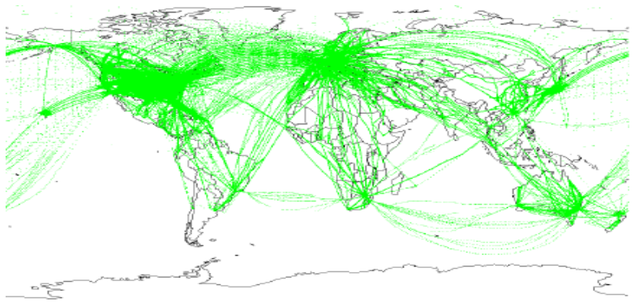
User-relevant verification - Aviation

$$\text{Flight time error (FTE)} = \text{flight\_time}_{obs} - \text{flight\_time}_{fcst}$$

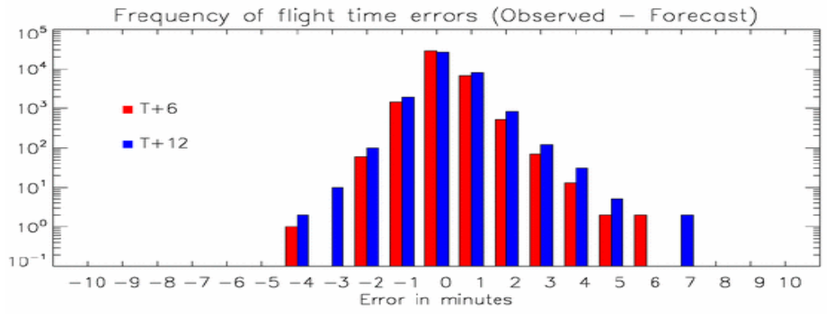
Accurate measure of wind forecast accuracy *directly relevant to airlines*

Calculated using the track that the aircraft actually took

Uses AMDAR observations from real flights rather than model analyses or radiosondes



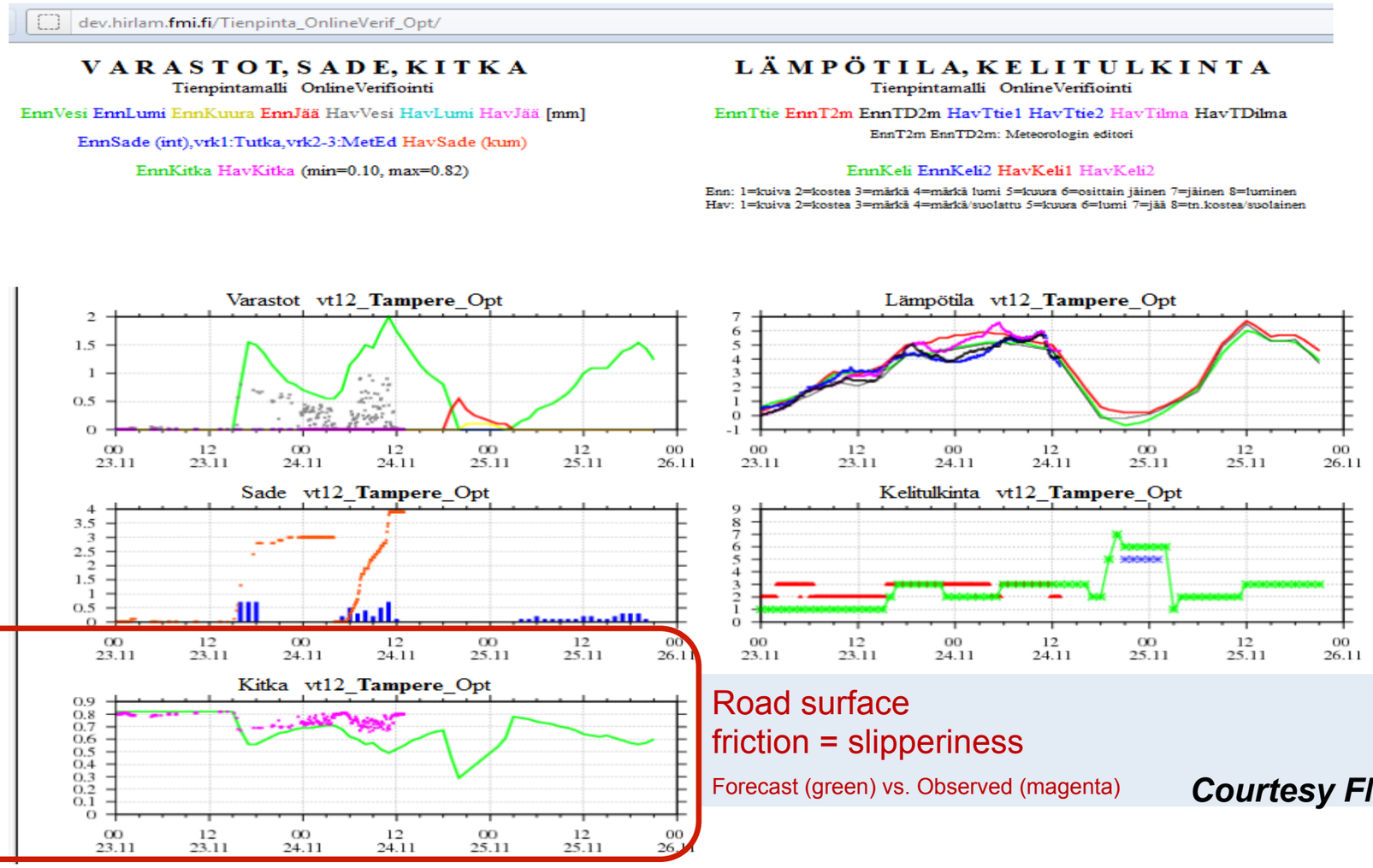
AMDAR – 1-7 Feb 2010



Courtesy Phil Gill

ECMWF

FMI



Road surface
friction = slipperiness

Forecast (green) vs. Observed (magenta)

Courtesy FMI

ECMWF

Uncertainty in observations

As models improve, can no longer ignore observation error!

Remove observation *bias* errors where possible

Effects of *random* obs error on verification

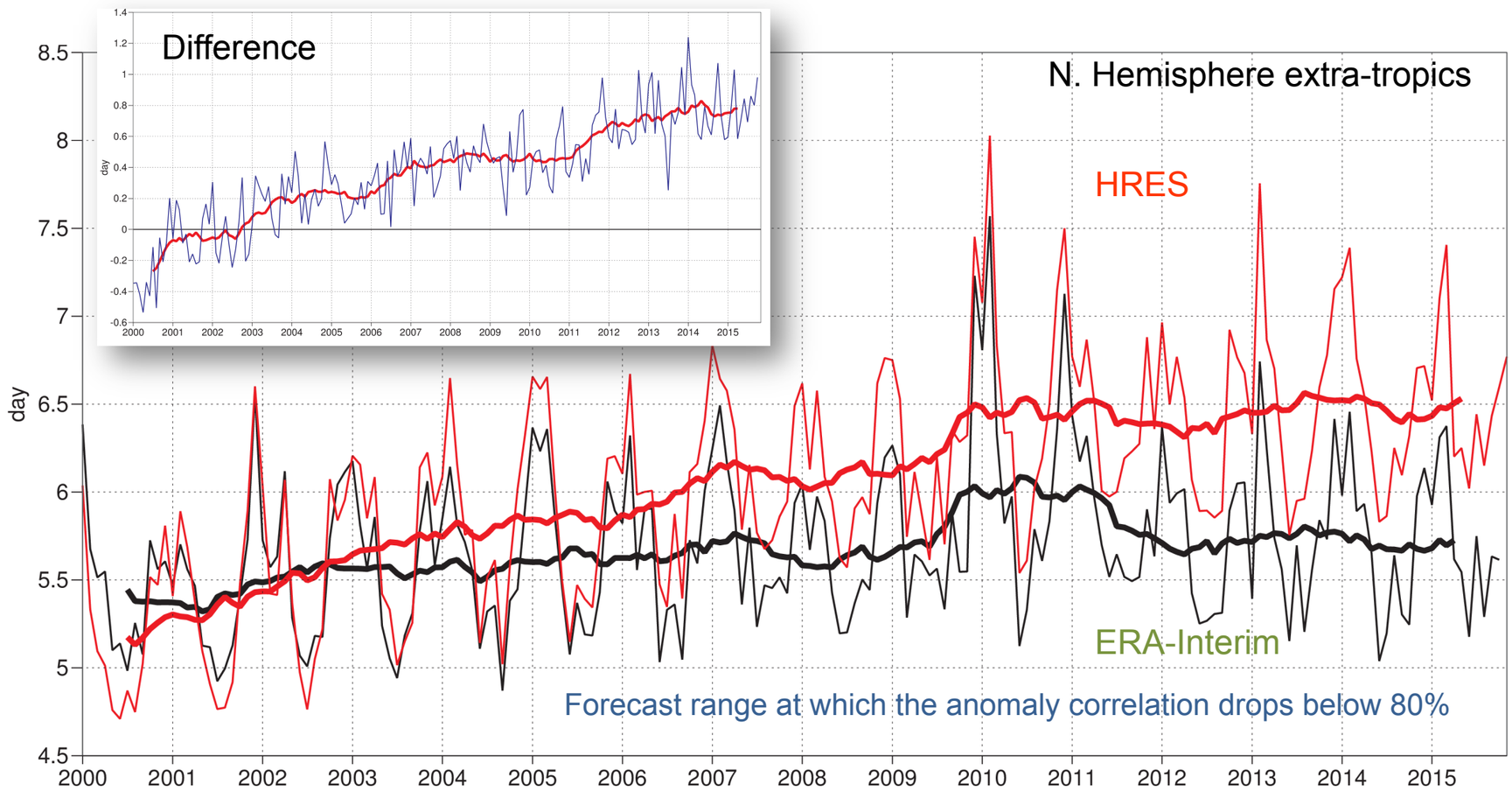
- “Noise” leads to poorer scores for deterministic forecasts
- Ensemble forecasts have poorer reliability & ROC

What can we do?

- Error bars in scatter plots
- Quantitative reference to “gold standard”
  - Correct for systematic error in observations
  - RMSE – Ciach & Krajewski (*Adv. Water Res.,*1999)
  - Categorical scores – Briggs et al. (*MWR*, 2005), Bowler (*MWR*, 2006)
- Multiple observation sources

Courtesy Beth Ebert

Model performance: HRES relative to ERA-I

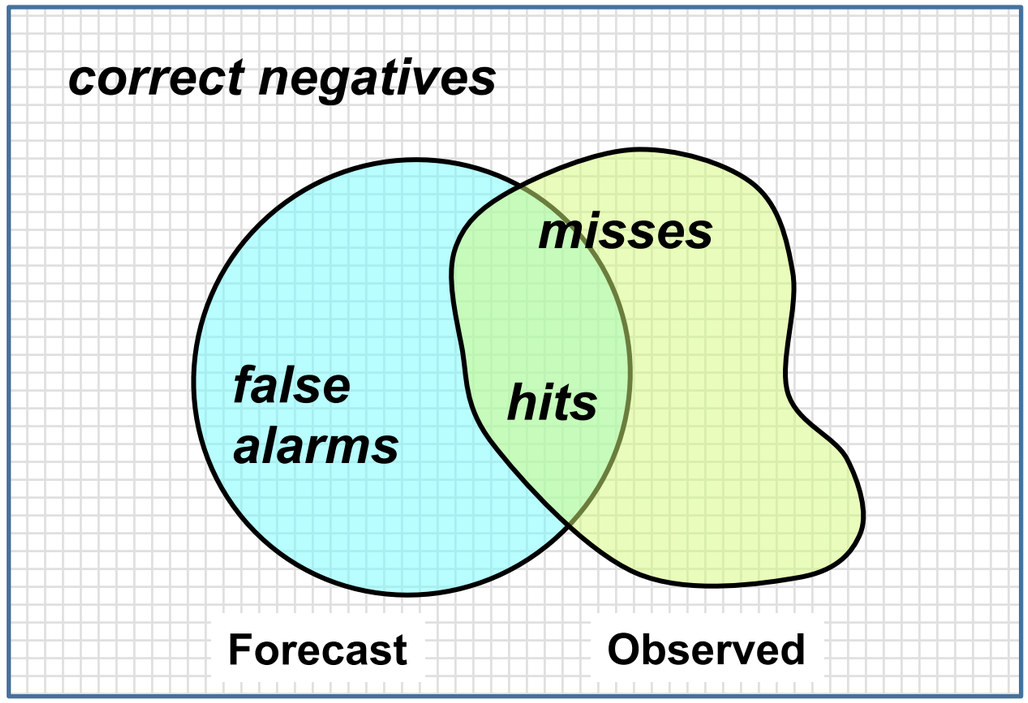


EXTRA

ECMWF

Measure	Attribute evaluated	Comments
Probability forecasts		
Brier score	Accuracy	Based on squared error
Resolution	Resolution (resolving different categories)	Compares forecast category climatologies to overall climatology
Reliability	Calibration	
Skill score	Skill	Skill involves *comparison* of forecasts
Sharpness measure	Sharpness	Only considers distribution of forecasts
ROC	Discrimination	Ignores calibration
C/L Value	Value	Ignores calibration
Ensemble distribution		
Rank histogram	Calibration	Can be misleading
Spread-skill	Calibration	Difficult to achieve
CRPS	Accuracy	Squared difference between forecast and observed distributions Analogous to MAE in limit
IGN score	Accuracy	Local score, rewards for correct category; infinite if observed category has 0 density

Traditional spatial verification



Predicted \ Observed	yes	no
yes	*hits*	*false alarms*
no	*misses*	*correct negatives*

$$POD = \frac{hits}{hits + misses}$$

$$FAR = \frac{false\ alarms}{hits + false\ alarms}$$

$$TS = \frac{hits}{hits + misses + false\ alarms}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(F_i - O_i)^2}$$

ECMWF

Verifying rare extreme values

Categorical scores

- Metrics should reward hits, penalise misses and false alarms
- For rare events, traditional categorical scores like TS → 0
- New extremal dependence scores:

Event forecast	Event observed		
	Yes	No	Marginal total
Yes	a	b	a + b
No	c	d	c + d
Marginal total	a + c	b + d	a + b + c + d =n

H = a / (a+c), hit rate

F = b / (b+d), false alarm rate

p = (a+c) / n, base rate

q = (a+b) / n, relative frequency of forecasted events

$$EDS = \frac{\log p - \log H}{\log p + \log H} \qquad SEDS = \frac{\log q - \log H}{\log p + \log H}$$

$$EDI = \frac{\log F - \log H}{\log F + \log H} \qquad SEDI = \frac{\log F - \log H - \log(1-F) + \log(1-H)}{\log F + \log H + \log(1-F) + \log(1-H)}$$

Ferro & Stephenson, *Weather & Forecasting*, 2011

ECMWF

Scores

Root Mean Square Error:

$$E = \sqrt{\overline{(fc - an)^2}}$$

Measures accuracy
Range: 0 to infinity perfect score = 0

Bias:

$$BIAS = \overline{FC - OBS}$$

Measures bias
Range: -infinity to +infinity
perfect score = 0

Mean Absolute Error :

$$MAE = \overline{|FC - OBS|}$$

Measures accuracy
Range: 0 to infinity perfect score = 0

Anomaly Correlation:

$$ACC = \frac{\overline{(fc - c)(an - c)}}{\sqrt{A_{fc} A_{an}}}$$

$$A_{fc} = \overline{(fc - c)^2}$$

$$A_{an} = \overline{(an - c)^2}$$

Measures accuracy
Range: -100% to 100%
perfect score = 100%

ECMWF