Forecast verification
Optimizing point forecasts

Karolina Stanisławska, Belgingur

Seychelles, 23 February 2016

Outline

- Introduction to forecast verification
- Matching forecasts with observations
- Verification measures
- The representativeness error
- Improving point forecasts

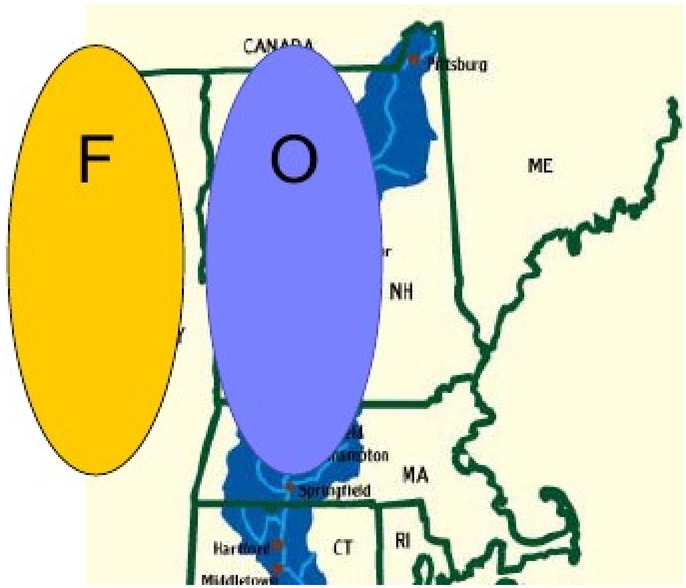
What is verification

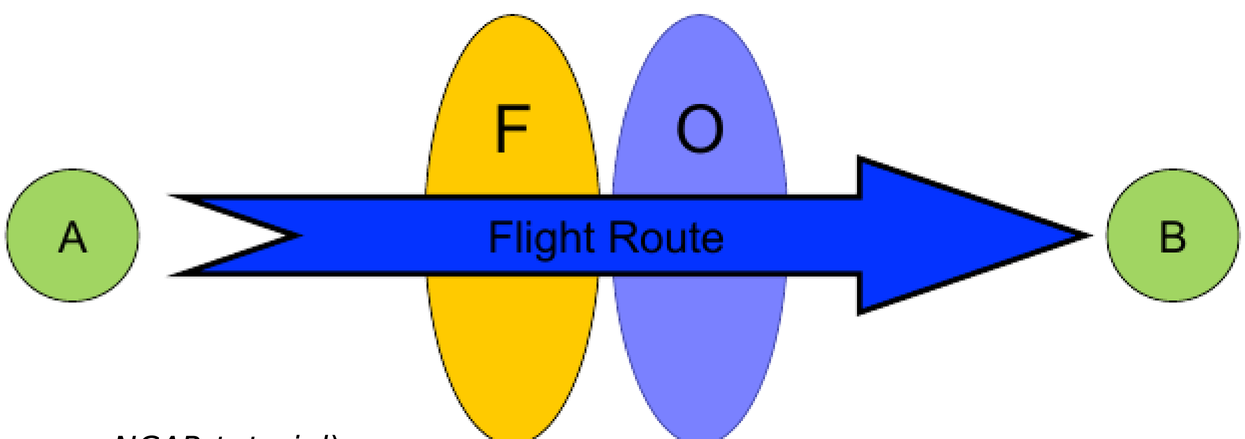
- Evaluation of the quality of the forecast
- Comparison of predicted values with reference values
- Can the models be trusted?

Motivation

- Can the model be trusted on the modeled area?
- How did the new model version improve the forecasts, did it?
- Which parameterizations options to use? What resolution and domain size is best?
- What accuracy would we expect from operational models?

When is a forecast good?





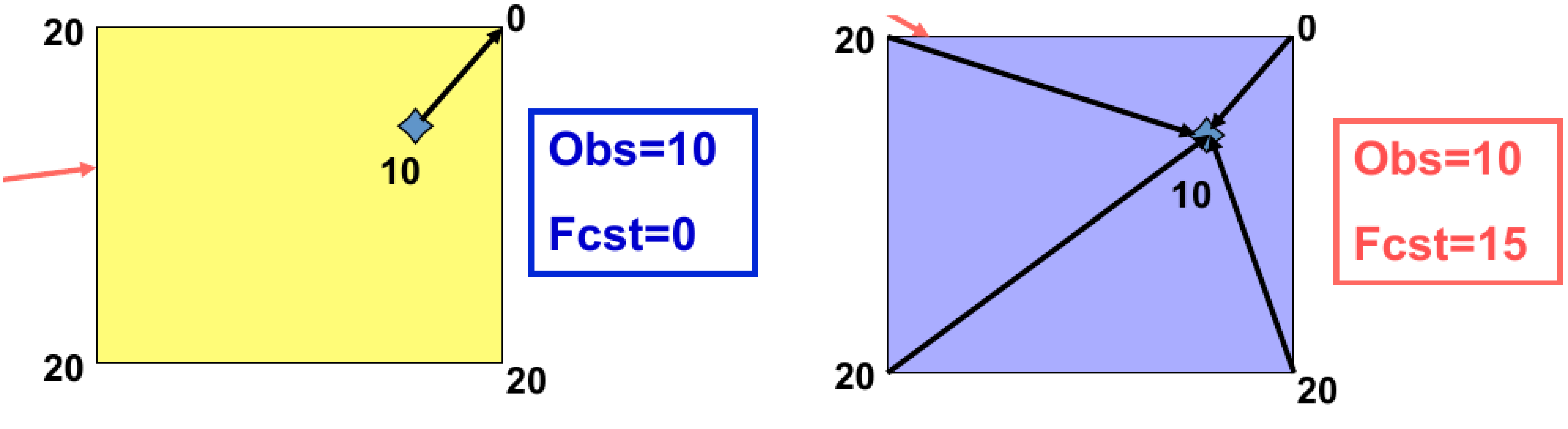
(source: NCAR tutorial)

What to use for comparison

- Observations
  - which kind of observations?
  - what is the quality of them?
  - do we have enough of them?
- Analysis, reanalysis
  - which model is the analysis from? What is the resolution?
  - analysis is still a model result
  - temporal resolution of reanalysis (6 hours)
- Reference forecast

Matching forecasts with observations

- Which grid point or points should be used to make a point forecast for that observation?

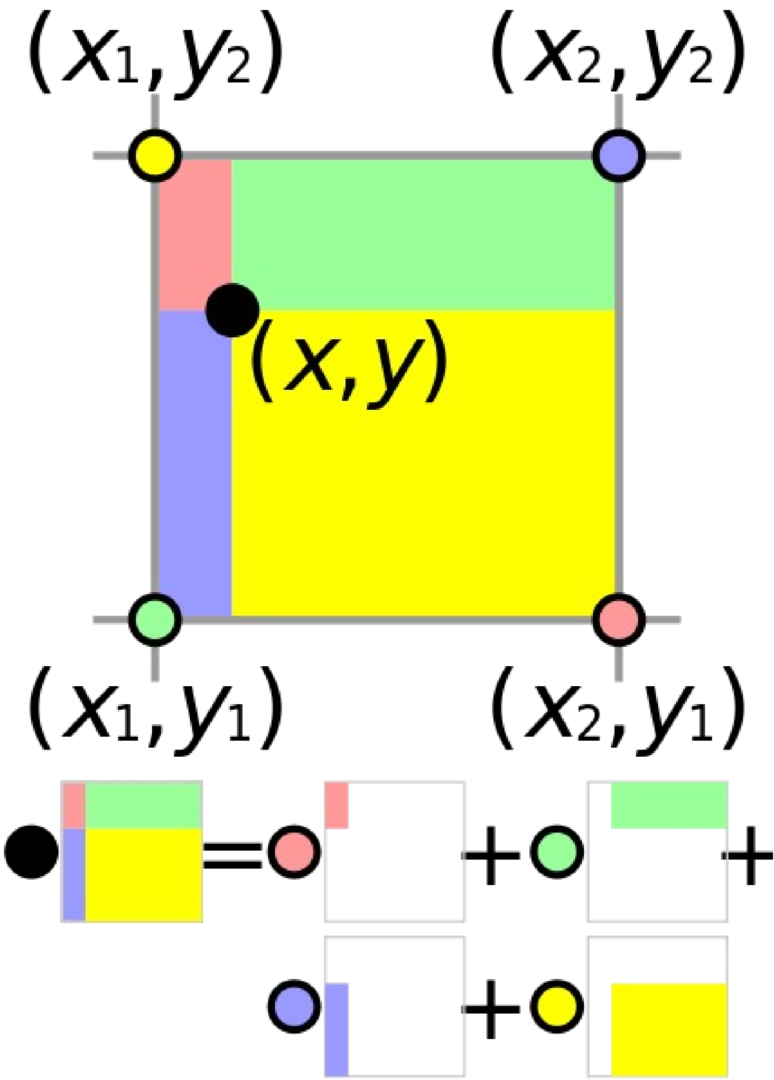


(source: NCAR tutorial)

Matching forecasts with observations

- Usual approaches:
  - nearest grid point
  - bilinear interpolation

- Alternative approaches?



Verification measures

- Depends on the nature of the variable
- Continuous vs. dichotomous
  - time series of wind speed, temperature, humidity
  - occurrence of precipitation, thunderstorm
- Multi-category
  - types of precipitation

Mean Absolute Error

- Average magnitude of forecast error
- *Average of absolute errors*
- Disregards direction of the error
- Each error has the same weights (linear score)

$$MAE = \frac{1}{n}\sum_{k=1}^{n} \left| x_k - o_k \right| ,$$

Mean Square(d) Error

- MSE
  - more sensitive to large errors, because the errors are squared
  - good if the intention is to penalize large errors
  - sensitive to large variance
- RMSE
  - same dimensions as the forecast and observations

$$MSE = \frac{1}{n}\sum_{k=1}^{n}(x_k - o_k)^2 . \qquad RMSE = \sqrt{MSE}$$

Correlation

- Measures any linear association between a series of observations and measurements
- No information about the linear relation (inclination of the regression line)

$$r_{XY} = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \bar{y}\right)\left(x_i - \bar{x}\right)}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2 \cdot \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}} = \frac{\mathrm{cov}(Y,X)}{s_Y s_X}$$

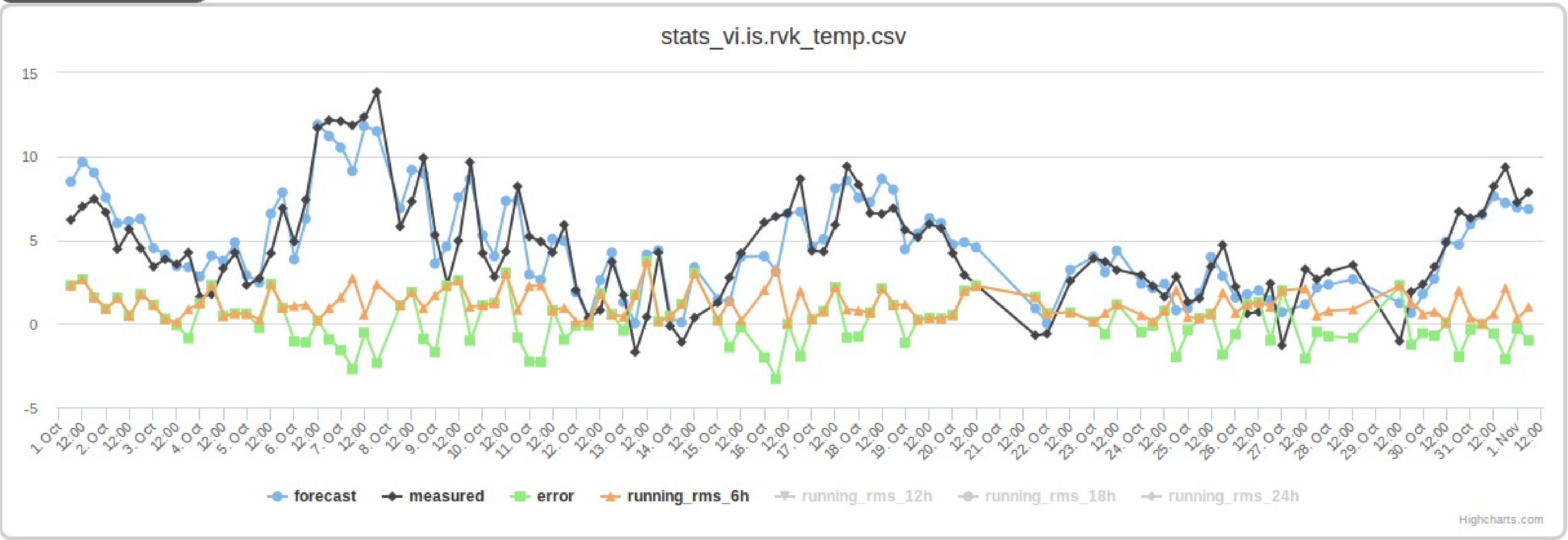
Errors on a timeseries

- Forecast error – series of error between forecast and observed values
- Running RMSE – RMS calculated over a time window of *n* hours

- Inspection which periods were well forecast and which have problems

Errors on a timeseries

Choose a file...



Representativeness error

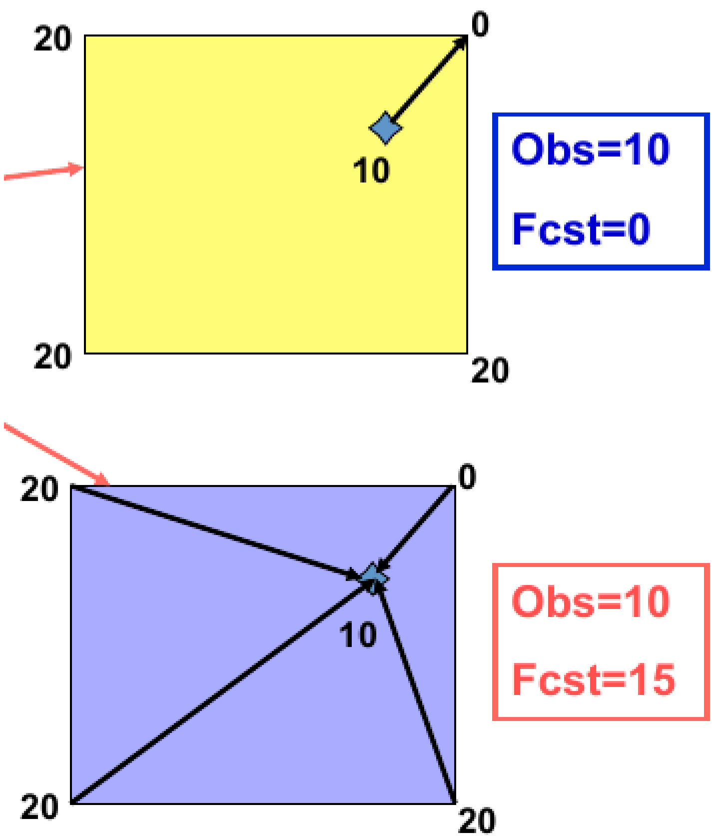
- Mismatch between spatial and temporal scales of forecasts and observations
- A value for a grid point is in fact an average of an area, while an observation is made in a point
- Not possible to compensate

Point forecast

- Not only validation of area forecasts is important
- We want to produce point forecasts at the same point where our observations are collected
- We want to produce point forecasts for some points of interest

Creating point forecasts

- How to get a point forecast from a grid
  - Nearest point?
  - All the surrounding points?
- What weights should they have?



(source: NCAR tutorial)

Sources of problems

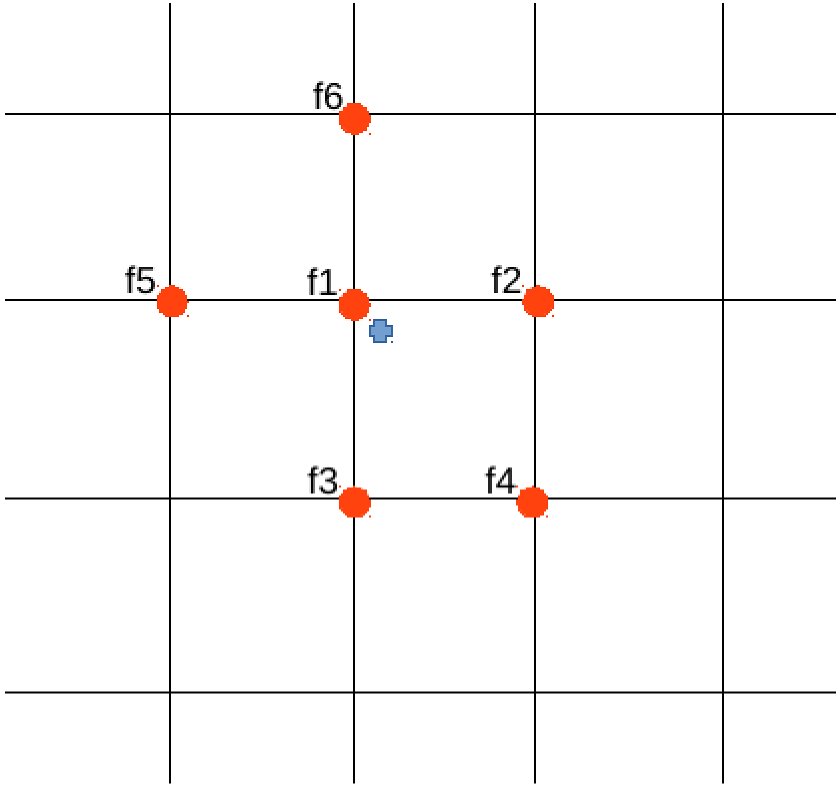
- The grid point closest to the point we approximate is at completely different conditions than the point we approximate
- The closest grid point is at sea, while the point we approximate is on the coast
- Valleys or mountain slopes

- *The weights should not be based exclusively on a distance to the point, they should take into account the complexity of the terrain*

Our solution

- Identification of the best combination of nearby grid points to minimize the mean square difference of the forecast value with the observations.
- Perform multiple linear regression over a timeseries of forecasts for *n* neighbouring grid points

Linear Regression method



$$f = 0.3 * f1 + 0.25 * f3 + 0.2 * f2 + 0.15 * f6 + 0.05 * f4 + 0.04 * f5 + 0.01$$

Linear Regression method

- The statistical method decides about the importance of particular grid points to the prediction of the forecast for a point
- The weights produced by linear regression are used to produce a point forecast as a weighted average of values at the *n* points

Performance of the LR method

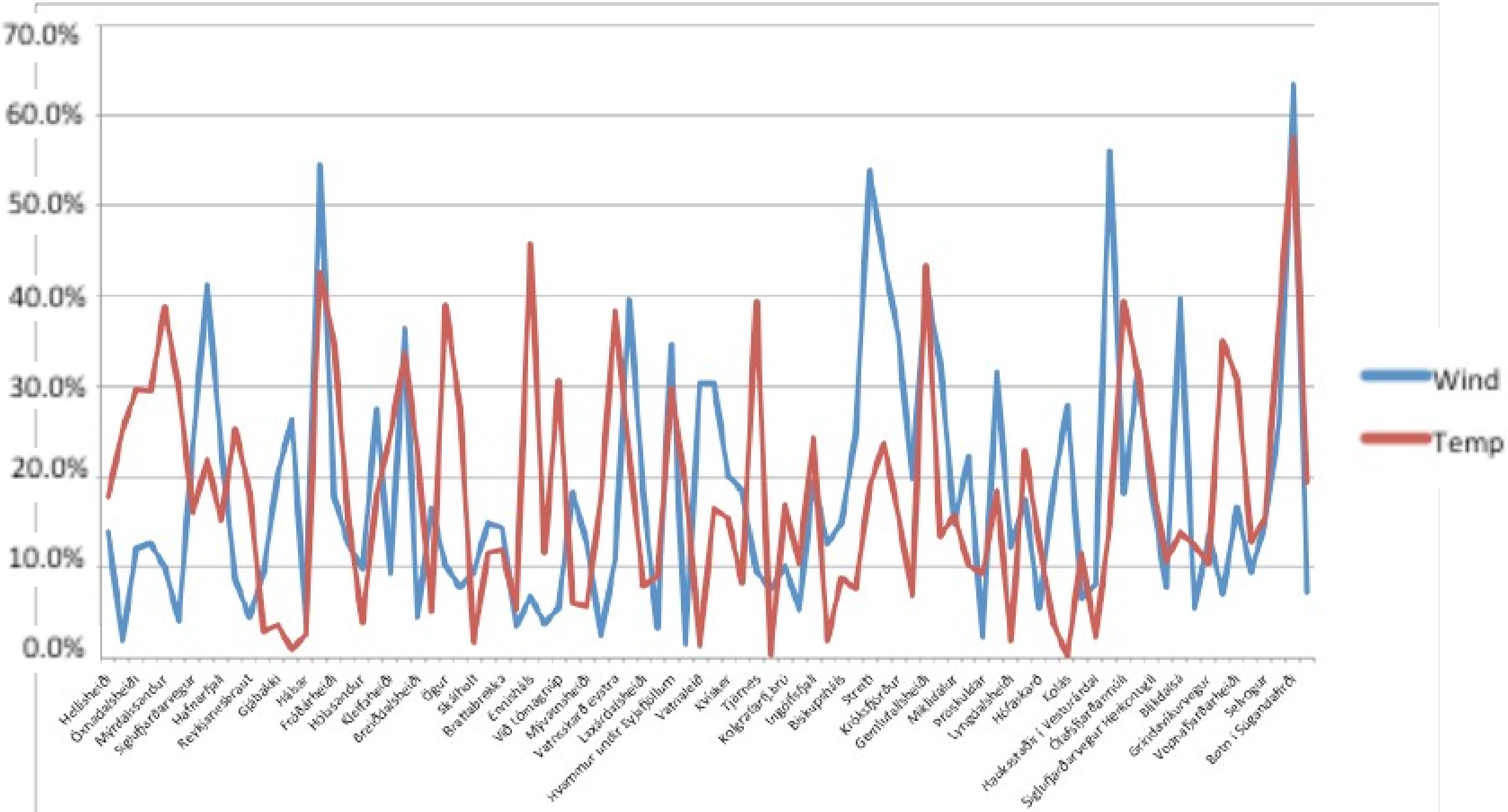


Figure 2: Relative improvement [%] compared to bilinear interpolated raw model output for wind speed (blue line) and temperature (red line) for the 86 observational sites (horizontal axis, note that not all sites are marked).

Performance of the LR method

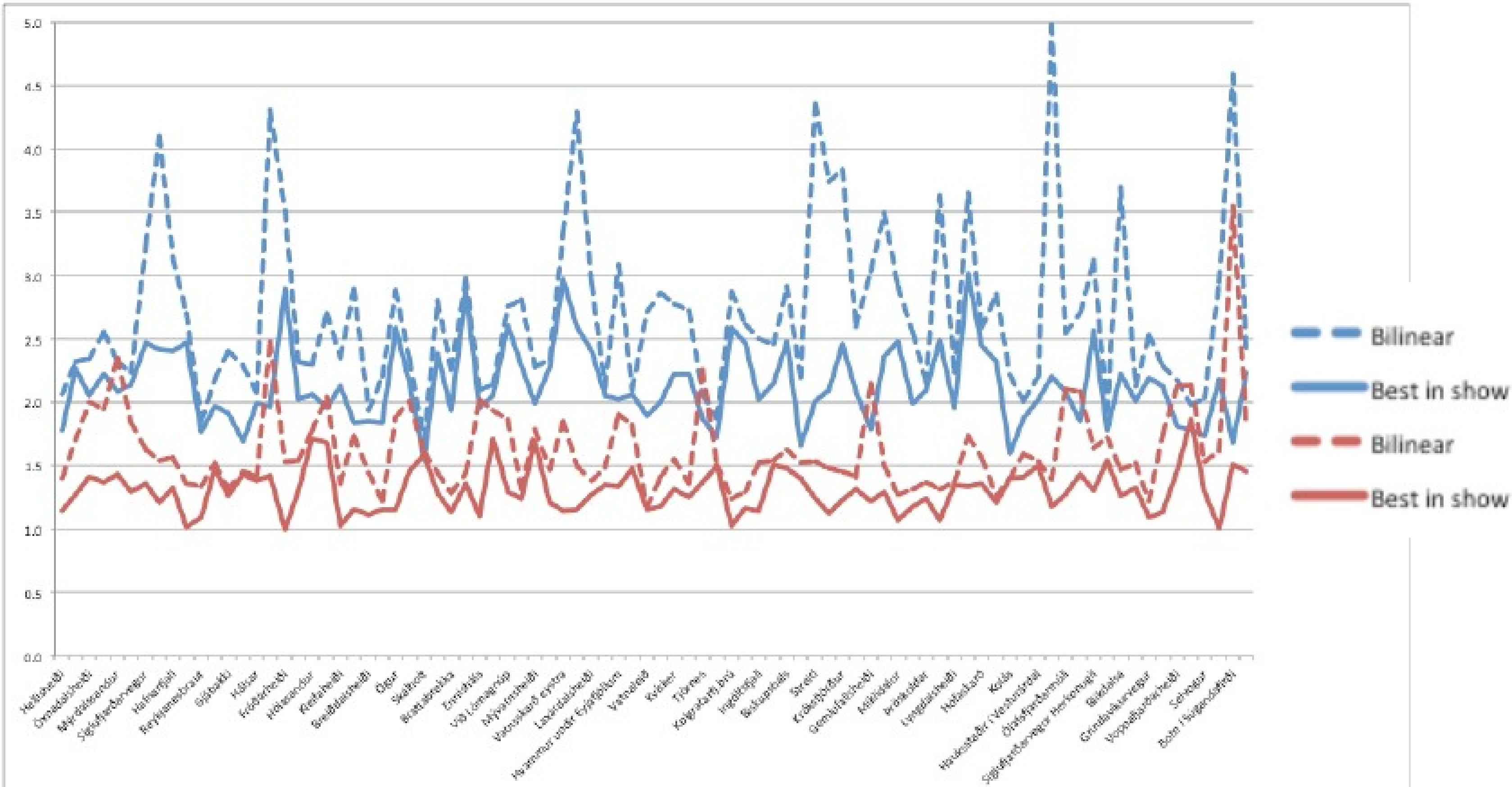


Figure 3: Root Mean Square (RMS) error of forecasts of wind speed (blue lines) and temperature (red lines) when using bilinear interpolated raw model output (dashed lines) and the best regression forecast (solid lines).

Planned extensions

- Create separate linear regression weights for different weather conditions and/or time of day
  - Vertical stability is different at night than in the afternoon
- Intelligent methods beyond the linear regression
  - Precipitation may depend on wind direction
- Downscaling
  - Use dynamical downscaling to produce data to create the weights instead of forecasts

Thank you!