Competing Forecasts and Dependence

**Observation**

**Model 1**

**Model 2**
Competing Forecasts and Dependence

Observation

Model 1 - Observation

Model 2 - Observation

Simple loss
Competing Forecasts and Dependence

d = (Model 1 – Observation) – (Model 2 – Observation)

Simple loss: mean(d) ≈ -0.2

Absolute error loss: mean(d) ≈ 7.5
Competing Forecasts and Dependence

d = (Model 1 – Observation) – (Model 2 – Observation)

• Hypothesis test of interest is $H_0: \mathbb{E}[d_t] = 0$

• Two ways we can be wrong:
  • Reject null hypothesis ($H_0$) when it is actually true
  • Fail to reject null hypothesis when it is actually false

• $\Pr\{\text{type I error}\} = \text{size of the test} = \alpha = \text{significance level of the test}$

• $1 - \Pr\{\text{type II error}\} = \text{power of the test} = \text{probability of detecting a difference when one really exists}$. 
d = (Model 1 – Observation) – (Model 2 – Observation)

Test Statistic:

\[ S = \frac{\text{mean}(d) - \mu_d}{2\pi \cdot s_d(0)} \]

Assumption: \( S \xrightarrow{\text{as } n \to \infty} N(0,1) \)

From this assumption, it is possible to calculate the probability of observing an \( S \) at least as large as the one observed, and if this probability is \( < \alpha \), then we reject the null hypothesis.

Alternatively, can obtain confidence limits as mean\( (d) \pm 2\pi \cdot s_d(0) \cdot z(\alpha/2) \)
Competing Forecasts and Dependence

d = (Model 1 – Observation) – (Model 2 – Observation)

Test Statistic:

\[ S = (\text{mean}(d) - \mu_d) / (2\pi * s_d(0)) \]

Key is in estimating \( s_d(0) \)

Obtained through a weighted sum of sample autocovariances

Competing Forecasts and Dependence

d = (Model 1 – Observation) – (Model 2 – Observation)

Our example:

Simple loss: mean(d) ≈ -0.2 and p-value ≈ 0.8 (not significant)

Absolute Error loss: mean(d) ≈ 7.5 and p-value ≈ 0 (significant)
DM Test

Summary of Univariate Setting

• Diebold-Mariano (DM) test gives an hypothesis test for competing forecasts (which forecast is better in terms of a loss (skill) function).
• Can also get confidence intervals instead of hypothesis test.
• Test accounts directly for temporal correlation.
• Robust to contemporaneous correlation (Hering and Genton, 2011).
• Works for any loss/skill function.
• No distributional assumptions for underlying series (only on the mean of the loss differential).
• powerful test (Hering and Genton, 2011).
• Dynamic Time Warping (DTW) allows for analyzing forecast performance while accounting for timing errors.
• R software package verification
• See G. and Roux (accepted to Meteorol. Appl.) for more details.
Competing Forecasts and Dependence

\[ D = D_1 - D_2 \]
Other Spatial Issues

Above Figure from Beth Ebert
Methods Overview

Fig. 2 from G. et al. (2010, BAMS, 91 (10), 1365 – 1373)
Filter Methods: Smoothing/ Neighborhood

Smooth one or both of Observation and model fields or indicator fields.

Calculate statistics at different levels of smoothing (different sized neighborhoods)

Filter Methods: Smoothing/ Neighborhood

Fractions Skill Score (FSS)

\[ \text{FSS} = 1 - \frac{\sum_{s=1}^{n} (\hat{p}_s - p_s)^2}{\sum_{s=1}^{n} \hat{p}_s^2 + \sum_{s=1}^{n} p_s^2} \]

Filter Methods: Scale Separation
Accounting for Location Errors

Binary fields obtained via setting all values below a threshold to zero.

Distance maps can be computed efficiently, and many summary measures are based on them.

Displacement Methods

Image Metrics/Measures

Baddeley’s $\Delta$

From G. (2011, WAF, 26, 409 - 415)
Displacement Methods

Image Metrics/Measures

AghaKouchak et al. (2011, J. Hyrdometeorology, 12, 274-285)
Displacement Methods

Field deformation

Reduction in RMSE is over 50% after applying the (space-time) warp.
Results show (1) forecasts have some skill in capturing these events and (2) in which aspects the forecasts need improvement.

Ex: 90th percentile of precipitation; storm placement/timing.
Accounting for location and small-scale errors in the spatial prediction comparison test

Loss function =

Distance from original location of each point to warped location

Loss (e.g., square error, absolute error) at each point between observation and warped value

Univariate Setting

Dynamic Time Warping (DTW)
Univariate Setting

Dynamic Time Warping (DTW)

Absolute error loss:
mean(d) ≈ -0.97
p-value ≈ 0.17 (not significant)

Recall that without warping: Absolute error loss for these series:
mean(d) ≈ 7.5 and p-value ≈ 0 (significant)
Accounting for location and small-scale errors in the spatial prediction comparison test

32 test cases (NSSL/SPC Spring 2005 Experiment). ARW-WRF vs NMM

Final Remarks

- R software package: SpatialVx (not yet ready for prime time)
- Mesoscale Verification Intercomparison in Complex Terrain (MesoVICT) [http://www.ral.ucar.edu/projects/icp](http://www.ral.ucar.edu/projects/icp)

- List of references relevant to spatial forecast verification
  - See, e.g., review papers:
    - Ahijevych et al. (2009),
    - G. et al. (2010, 2010 BAMS),
    - Brown et al. (2012),
  - MesoVICT test cases available (geometric, perturbed and real)
  - Sign up to receive emails about the MesoVICT