Which Forecast is Better?

Big Data & Environment Workshop
Buenos Aires, Argentina
10 November 2015
Eric Gilleland
Research Applications Laboratory
Univariate Setting

Diebold-Mariano test

Observation

Model 1

Model 2
Univariate Setting

Diebold-Mariano test

Simple loss

F - O

Observation

Model 1 - Observation

Model 2 - Observation
Univariate Setting

Diebold-Mariano test

AE loss

\[ |F - O| \]

| Model 1 – Observation |
| Model 2 – Observation |

| Model 1 – Observation |
| Model 2 – Observation |
Univariate Setting

Diebold-Mariano test

- Let $x = x_1, \ldots, x_n$ be an observed time series.
- Let $y = y_1, \ldots, y_n$ and $z = z_1, \ldots, z_n$ be two competing forecast models for $x$.
- Let $g(x, y)$ and $g(x, z)$ be the loss (or skill) function between the modeled and observed time series (defined at each time point!).
- Null hypothesis of interest is:
  \[ H_0: E[g(x, y)] = E[g(x, z)] \]

- Interest is in the “loss differential”
  \[ d = g(x, y) - g(x, z) \]
  OR
  \[ H_0: E[d_t] = 0 \]
Univariate Setting

Diebold-Mariano test

Simple loss for these series:
\[ \text{mean}(d) \approx -0.2 \]

Absolute error loss for these series:
\[ \text{mean}(d) \approx 7.5 \]

![Graphs showing loss differentials and ACFs for simple and absolute error losses.](attachment:graphs.png)
Univariate Setting

Diebold-Mariano test

Test Statistic:

$$S = \frac{\text{mean}(d) - \mu_d}{\text{se}(\text{mean}(d))}$$

Key is in estimating $$\text{se}(\text{mean}(d)) = (2\pi * s_d(0))$$

$$s_d(0)$$ obtained through a weighted sum of sample autocovariances (Diebold and Mariano, 1995, *J. Bus. Econ. Stat.*, 13: 253—263)


Interest is generally in $$\mu_d = 0$$. 
Univariate Setting

Diebold-Mariano test

Test Statistic:

\[ S = \frac{\text{mean}(d) - \mu_d}{\text{se}(|\text{mean}(d)|)} \]

\[ \text{se}(|\text{mean}(d)|) = (2\pi \cdot s_d(0)) \]

Assumption: \( S \rightarrow N(0,1) \) as \( n \rightarrow \infty \)
Univariate Setting

Diebold-Mariano test

Our example:

Simple loss: $\text{mean}(d) \approx -0.2$ and $p$-value $\approx 0.8$ (not significant)

Absolute Error loss: $\text{mean}(d) \approx 7.5$ and $p$-value $\approx 0$ (significant)
Univariate Setting

Dynamic Time Warping (DTW)
Univariate Setting

Dynamic Time Warping (DTW)


introduce loss function based on DTW:

\[ g(x_t, y_t) = f(t, w(t)) + h(x_t, y_{w(t)}) \]

- distance traveled in time
- Usual loss function after having warped through time
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)
Univariate Setting

DM Test and Dynamic Time Warping (DTW)

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
Spatial Prediction Comparison Test

\[ D = D_1 - D_2 \]

Spatial Forecast Verification

Motivation

Fig. 1 and Table 2 from Ahijevych et al. (2009, WAF, 24, 1485 – 1497)
Spatial Forecast Verification

Motivation

Graphic from Beth Ebert
Spatial Forecast Verification

Overview of Methods

Mesoscale spatial forecast Verification Intercomparison over Complex Terrain (MesoVICT)

http://www.ral.ucar.edu/projects/icp

Fig. 2 from G. et al. (2010, BAMS, 91 (10), 1365 – 1373)
Spatial Forecast Verification

Location measures/metrics

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.

Spatial Forecast Verification

Location measures/metrics

Image Metrics/Measures

Baddeley’s $\Delta$

Spatial Forecast Verification

Image Warping

Forecast Image \((F(s))\)

Observed Image \((O(s))\)

Warped Image \((F(W(s)))\)

Graphic by Johan Lindström
Spatial Forecast Verification

Image Warping

Reduction in error ≈ 32.4%

Graphic by students from the 2010 Industrial Mathematical and Statistical Modeling Workshops for Graduate Students.

G. et al., 2010, NCAR Tech Note TN-482+STR, doi:10.5065/D62805JJ
Spatial Forecast Verification

Image Warping

Thin-plate Spline (TPS) Warp function

\[ W(s; p^O, p^F) = B(s, p^O) p^F \]

- \( B \) a matrix of *pre-calculated* radial basis functions
- \( s \) is a matrix of all possible spatial locations
- \( p^O \) is a subset of \( s \) giving the pre-determined control points in the 0-energy field (i.e., the observation field)
- \( p^F \) is a subset of \( s \) giving the positions of the 1-energy (i.e., forecast field) control points (precise locations determined by optimizing an objective function)
Spatial Forecast Verification

Image Warping

Optimize the log-likelihood:

\[ \ell(p^F | O, F, p^O) = \log p(O | F, p^F, p^O) + \log p(p^F | p^O) + \log p(\theta) \]
Using a thin-plate spline warp function, assuming Gaussian errors, and a Markov Random Field for the control point differences (i.e., $p^F - p^O$), the following objective function can be optimized (to optimize the likelihood):

$$Q(p^F) = \frac{(O(s) - F(W(s))))^T(O(s) - F(W(s)))}{(2\sigma_\varepsilon^2)} +$$

$$\frac{((p_x^F - p_x^O)^T(I - C)(p_x^F - p_x^O))}{(2\sigma_\Delta^2)} +$$

$$\frac{(p_y^F - p_y^O)(I - C)(p_y^F - p_y^O))}{(2\sigma_\Delta^2)}$$
Spatial Forecast Verification

Image Warping

Forcast  
Observation  
Deformed forecast

MSE 471.32  
MSE 0.27
Spatial Forecast Verification

Image Warping

MSE before warping ≈ 17,508
MSE after warping ≈ 9,316

≈ 47% error reduction from warped field
Spatial Forecast Verification

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

Future Work

• Compare HG-test against z-test with variance inflation factor, block bootstrap methods, and others.
• Add image warping to the R package SpatialVx.
• Space-Time Prediction Comparison Test?
  ▪ Challenge is to make simulations with known spatiotemporal correlation structures.
  ▪ Test whether a space-time separable covariance can be used even in the case of non-separability.
  ▪ Is it just overkill?
  ▪ Image warping can be done in space and time together (see, e.g., G. et al., 2010, *NCAR Technical Note*, TN-482+STR).