Using R for forecast verification and the spatial prediction comparison test

Eric Gilleland
Research Applications Laboratory
National Center for Atmospheric Research
SPECS fourth General Assembly
Norrköping, Sweden, 16 September 2015
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

- 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
  \[ H_0: E[d_t] = 0 \]
Simple loss for these series: \( \text{mean}(d) \approx -0.2 \)

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

Diebold-Mariano test

Test Statistic:

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

Key is in estimating \( \text{se}(\text{mean}(d)) \)


Interest is generally in \( \mu_d = 0 \).
Univariate Setting

Diebold-Mariano test

Test Statistic:

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

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

Assumption: \( S \to N(0,1) \) as \( n \to \infty \)
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

\[ D = D_1 - D_2 \]

Accounting for Location Errors

This loss function did not seem to yield a powerful testing procedure.

Binary fields obtained via setting all values below 5 mm to zero.

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

Accounting for Location Errors

Image warping

Figure from Johan Lindström

Accounting for Location Errors

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

mean(d)

Test appears to be powerful, and agrees with expert opinion and other results

Summary of SPCT and SPCT + Warping

- hypothesis test (or confidence intervals) for competing forecast models.
- Accounts for spatial correlation.
- Does not require a distributional assumption about the underlying fields (only the test statistic, S).
- Works for any loss function (though some work better than others).
- powerful test
- R software package SpatialVx conducts the test.
- Image Warping loss allows one to also account for location and small scale errors.
Future Work

- Compare with variance inflation factor and block bootstrap methods.
- Add image warping to `SpatialVx`, and create an image warping package for R.
- 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*).
R package: verification

• Written by Matt Pocernich with contributions from many within the weather forecast verification community
• Now maintained by me
• Primary function is called “verify”
• Most traditional univariate verification methods are included.
• Some newer cutting edge methods
• SpatialVx includes most spatial methods, but is still a work in progress
R package: verification

verify(obs, pred = NULL, p = NULL, baseline = NULL,
       frcst.type = "prob", obs.type = "binary",
       thresholds = seq(0, 1, 0.1), show = TRUE,
       bins = TRUE, fudge = 0.01, ...)

verify <- function(x, … ) {

    UseMethod("verify")

}
R package: verification

predcomp.test(x, xhat1, xhat2,
    alternative = c("two.sided", "less", "greater"),
    lossfun = "losserr", lossfun.args = NULL,
    test = c("DM", "HG"), ...)