NCAR

A spatial prediction comparison test for competing models

RAL Retreat, 8 December 2014

Center Green 1 Auditorium: Center

Eric Gilleland

Joint Numerical Testbed

National Center for Atmospheric Research

Univariate Setting

Diebold-Mariano test

Consider three series:

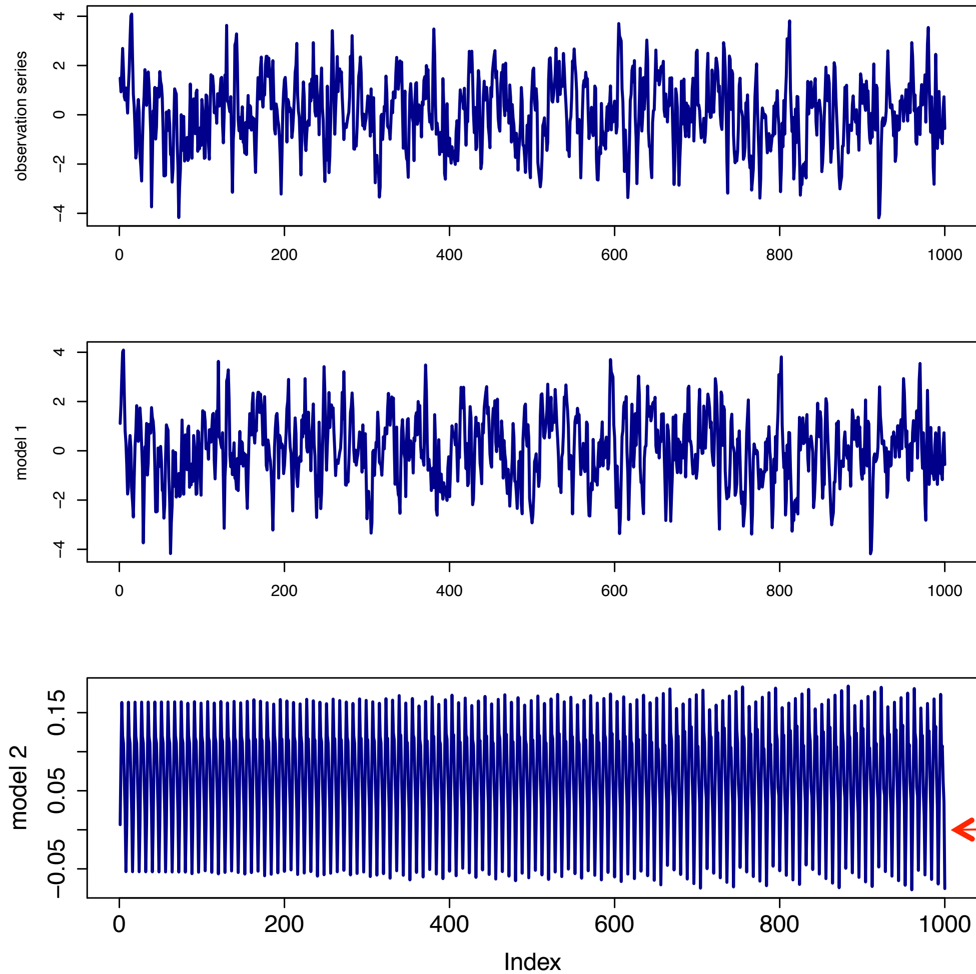
0. Observation: AR(2) time series given by
   $X_t = 0.8 * X_{t-1} - 0.2 * X_{t-2}$

1. Model 1: Same series as above, but shifted ten places to the left.
   $[\text{model 1}]_t = [\text{Observation}]_{t+10}$

2. Model 2: A smooth Fourier series fit to the observations in 0 above.
   $[\text{model 2}] \approx 0.07 - 0.01 * \cos(2\pi * t / 2) + \sin(2\pi * t / 2) + \ldots + \sin(2\pi * t / 8)$

Univariate Setting

Diebold-Mariano test



Simulated AR(2) series

Above shifted 10 places to the left

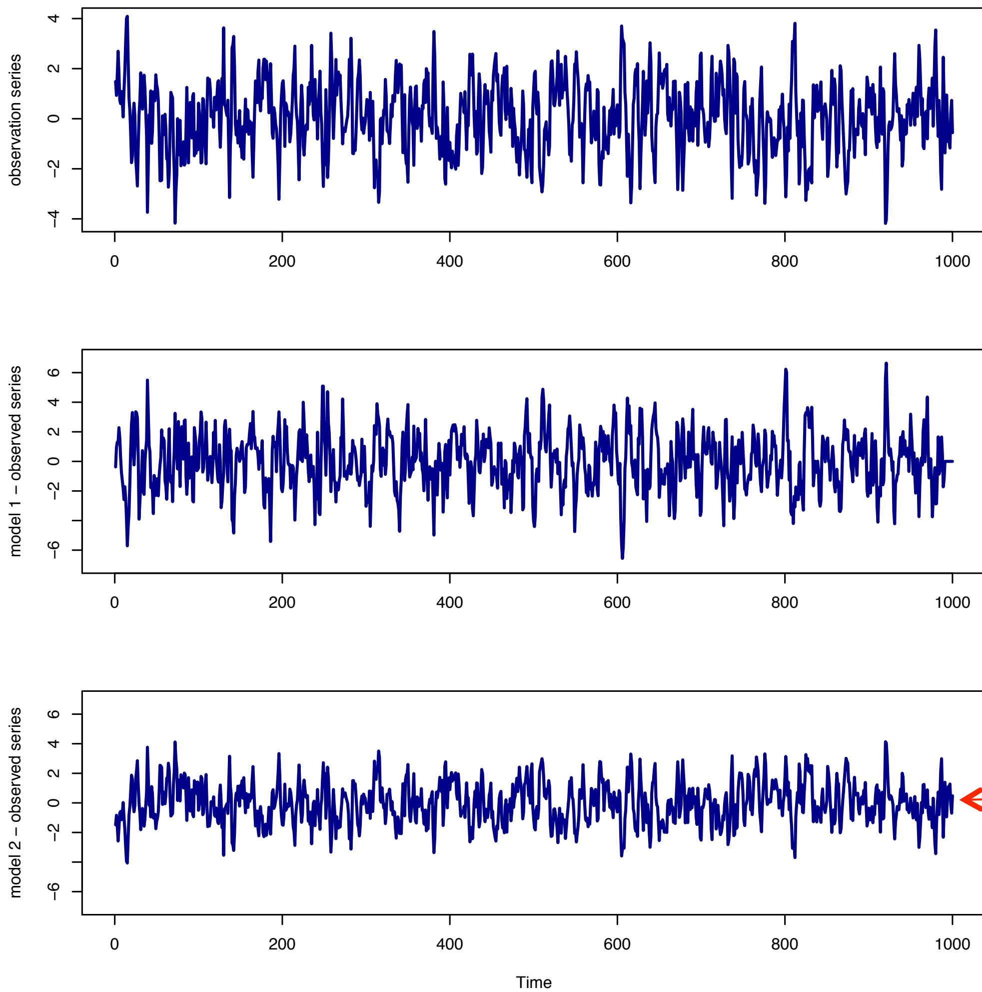
Fourier series fit to observed series

Univariate Setting

Diebold-Mariano test

Simple loss

F - O



Simulated AR(2) series

Above shifted 10 places to the left

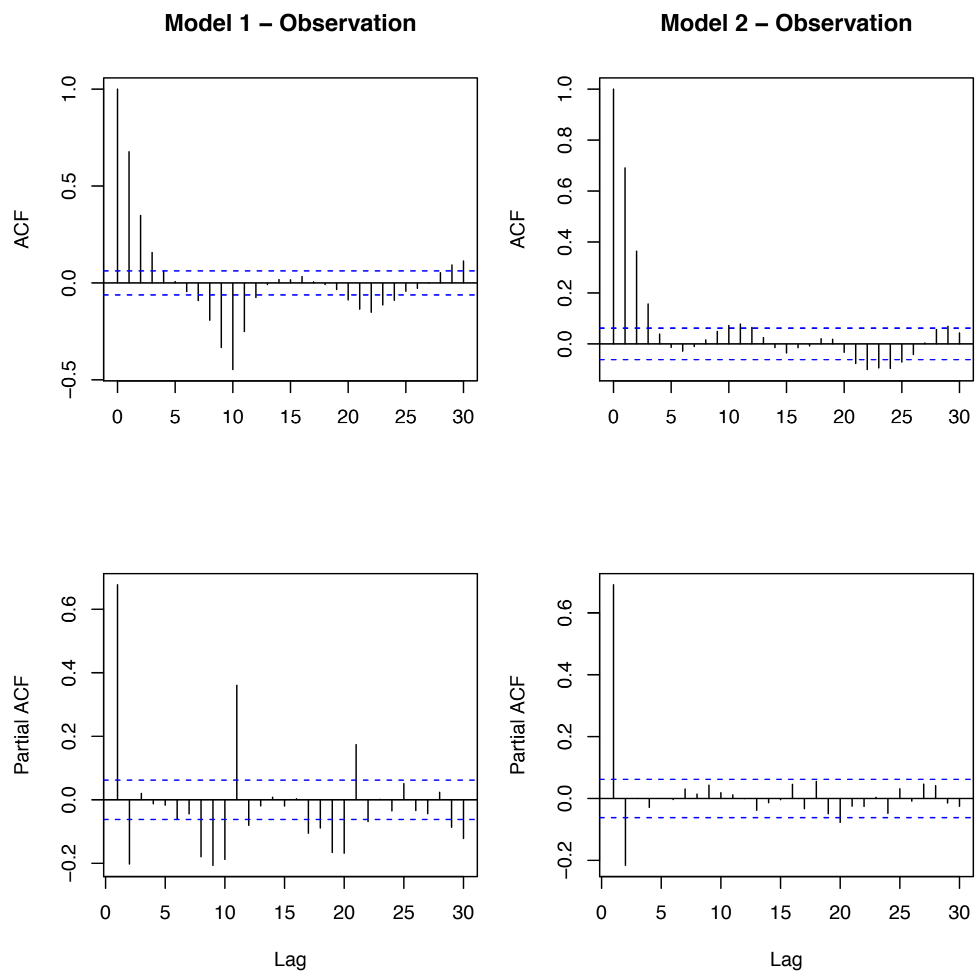
Fourier series fit to observed series

Univariate Setting

Diebold-Mariano test

Simple loss

F - O



Univariate Setting

Diebold-Mariano test

- Let $\mathbf{x} = x_1, \ldots, x_n$ be an observed time series.
- Let $\mathbf{y} = y_1, \ldots, y_n$ and $\mathbf{z} = z_1, \ldots, z_n$ 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:

$$H_0: E[\mathbf{g}(\mathbf{x}, \mathbf{y})] = E[\mathbf{g}(\mathbf{x}, \mathbf{z})]$$

- Interest is in the "loss differential"

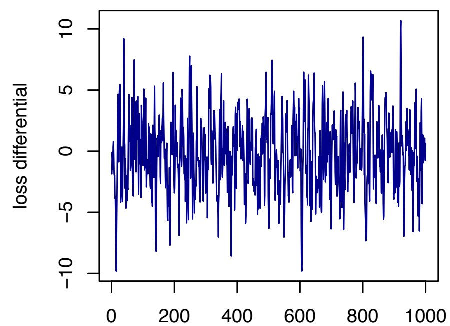
$$\mathbf{d} = \mathbf{g}(\mathbf{x}, \mathbf{y}) - \mathbf{g}(\mathbf{x}, \mathbf{z})$$

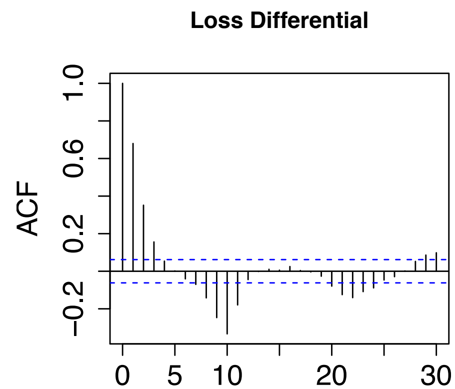
OR

$$H_0: E[d_t] = 0$$

Univariate Setting

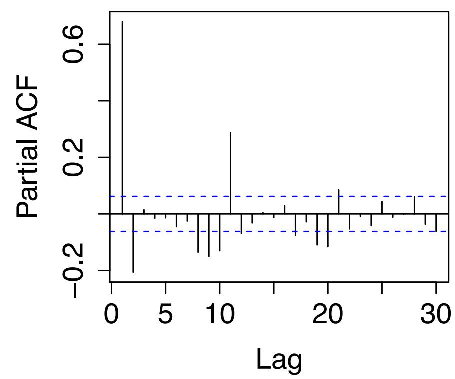
Diebold-Mariano test





Simple loss for these series:
mean(**d**) ≈ -0.2

Absolute error loss for these series:
mean(**d**) ≈ 7.5



Univariate Setting

NCAR

Diebold-Mariano test

Test Statistic:

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

Interest is generally in $\mu_d = 0$.

Key is in estimating $s_d(0)$

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

Hering and Genton (2011, *Technometrics*, **53**, (4): 414—425) suggest fitting a parametric autocovariance model to the sample autocovariances first.

Univariate Setting

Diebold-Mariano test

Test Statistic:

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

Key is in estimating $s_d(0)$

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

Univariate Setting

Diebold-Mariano test

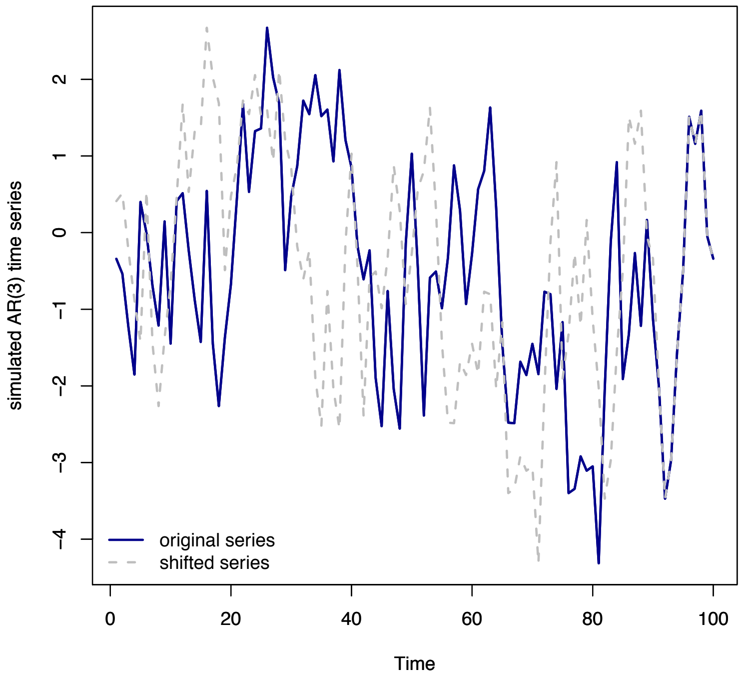
Our example:

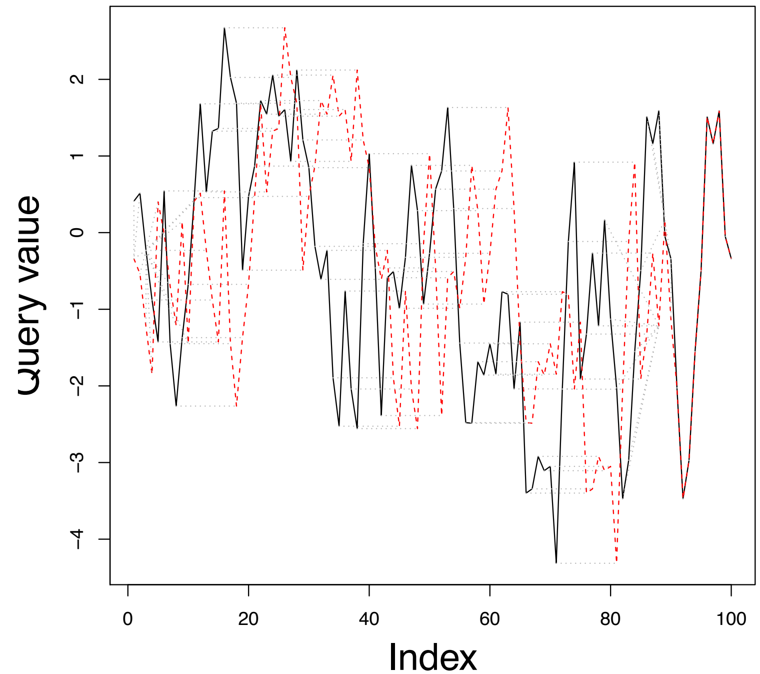
Simple loss: mean(**d**) ≈ -0.2 and p-value ≈ 0.8 (not significant)

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

Univariate Setting

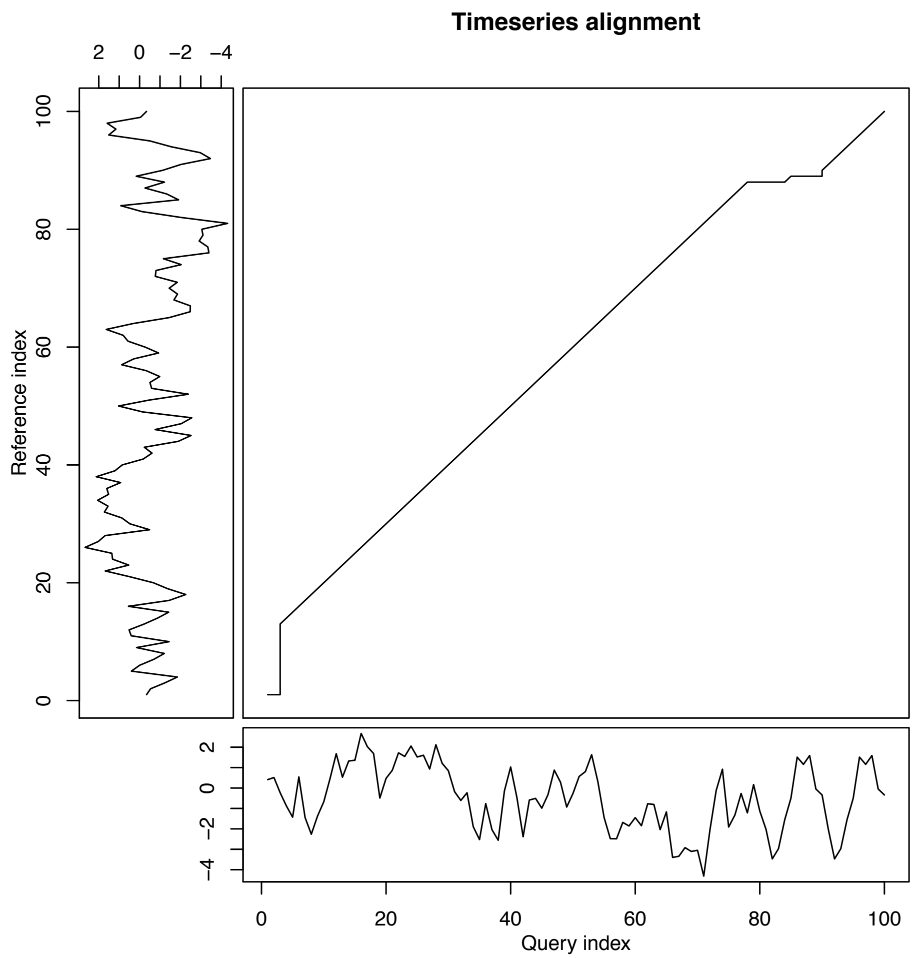
Dynamic Time Warping (DTW)





Univariate Setting

Dynamic Time Warping (DTW)



Univariate Setting

Dynamic Time Warping (DTW)

G. and Roux (2014, accepted to *Meteorol. Appl.*)

introduce loss function based on DTW:

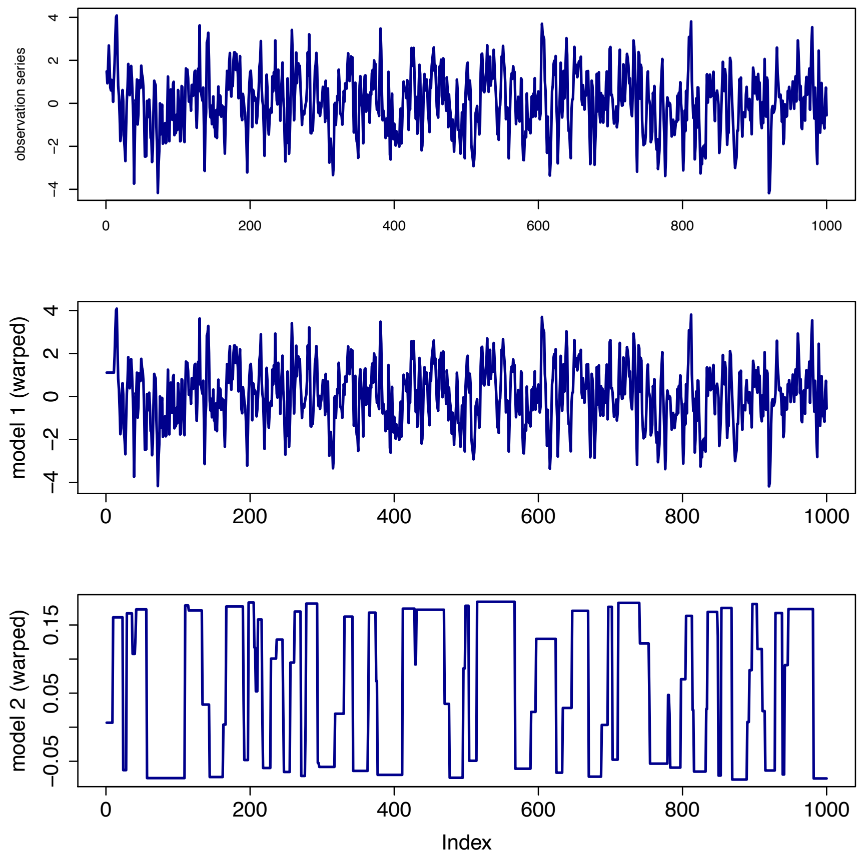
$$g(x_t, y_t) = f(t, w(t)) + h(x_t, y_{w(t)})$$

distance traveled in time → $f(t, w(t))$

Usual loss function → $h(x_t, y_{w(t)})$

Univariate Setting

Dynamic Time Warping (DTW)

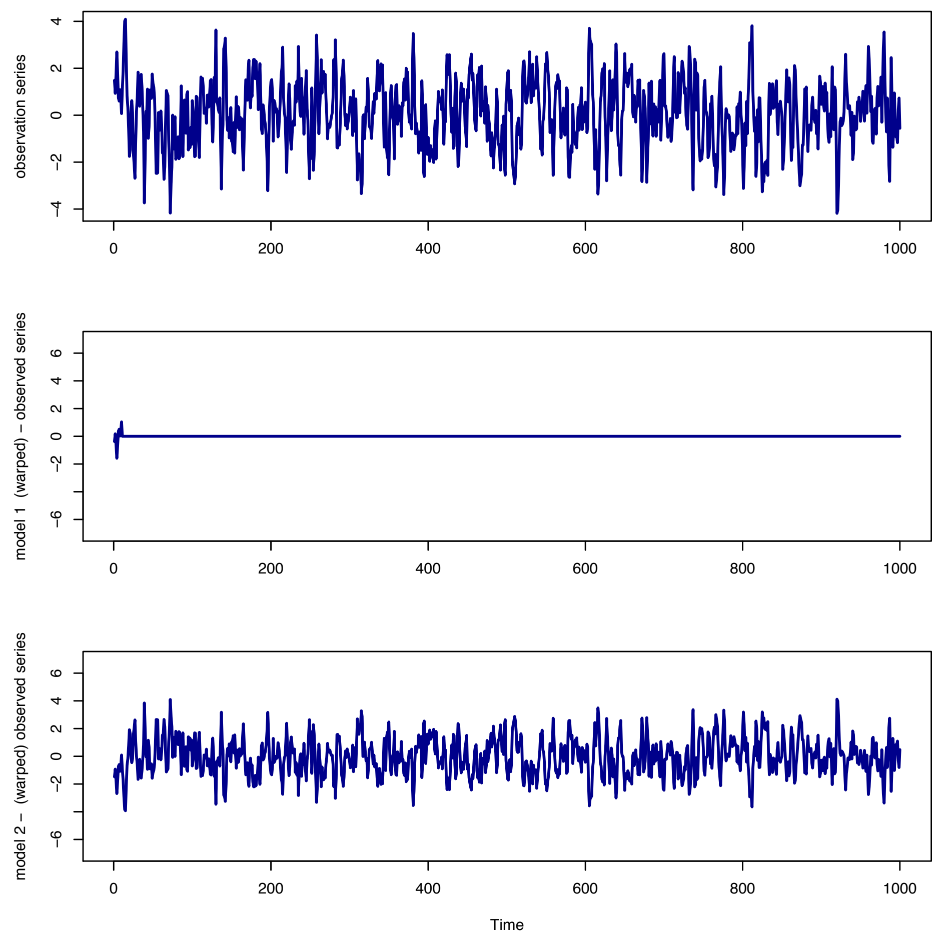


Univariate Setting

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

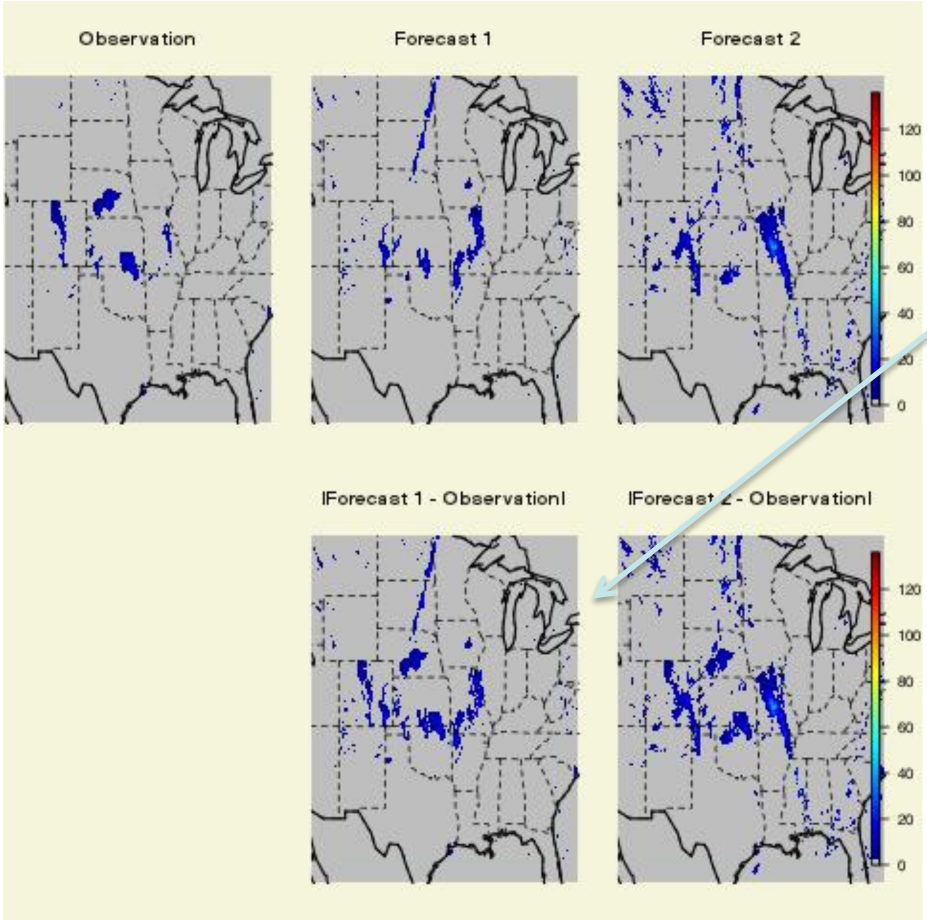
NCAR

DM Test and Dynamic Time Warping (DTW)

Summary of Univariate Setting

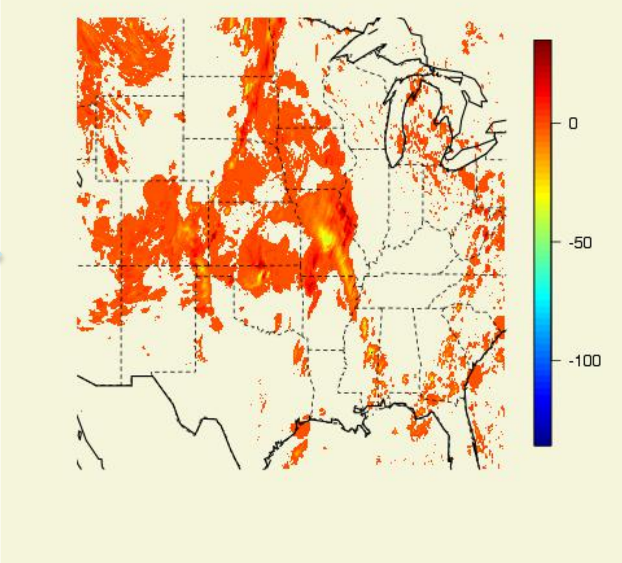
- Diebold-Mariano (DM) test gives an hypothesis test for competing forecasts (which forecast is better in terms of a loss (skill) function).
- Can also get confidence intervals instead of hypothesis test.
- Test accounts directly for temporal correlation.
- Robust to contemporaneous correlation (Hering 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`

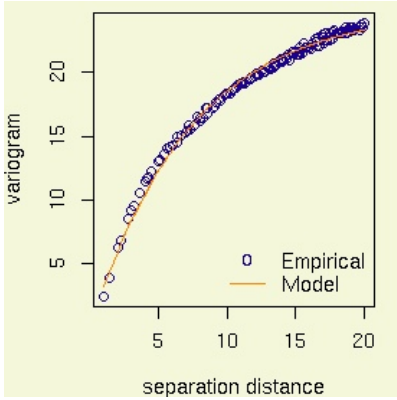
Spatial Prediction Comparison Test



D1 D2

D = D1 – D2





Spatial Prediction Comparison Test

Introduced by Hering and Genton (2011)

Extension of the time series version introduced by Diebold and Mariano (1995).

Spatial Prediction Comparison Test

NCAR

$$S = \frac{\bar{D}}{\sqrt{\operatorname{var}\left(\bar{D}\right)}} \longrightarrow \text{N}(0,\, 1) \text{ as } L \longrightarrow \infty, \text{ where}$$

$$\operatorname{var}\left(\bar{D}\right) = \frac{1}{L^2} \sum_{i=1}^{L} \sum_{j=1}^{L} C\left(h_{ij}\right)$$

distance between two spatial locations

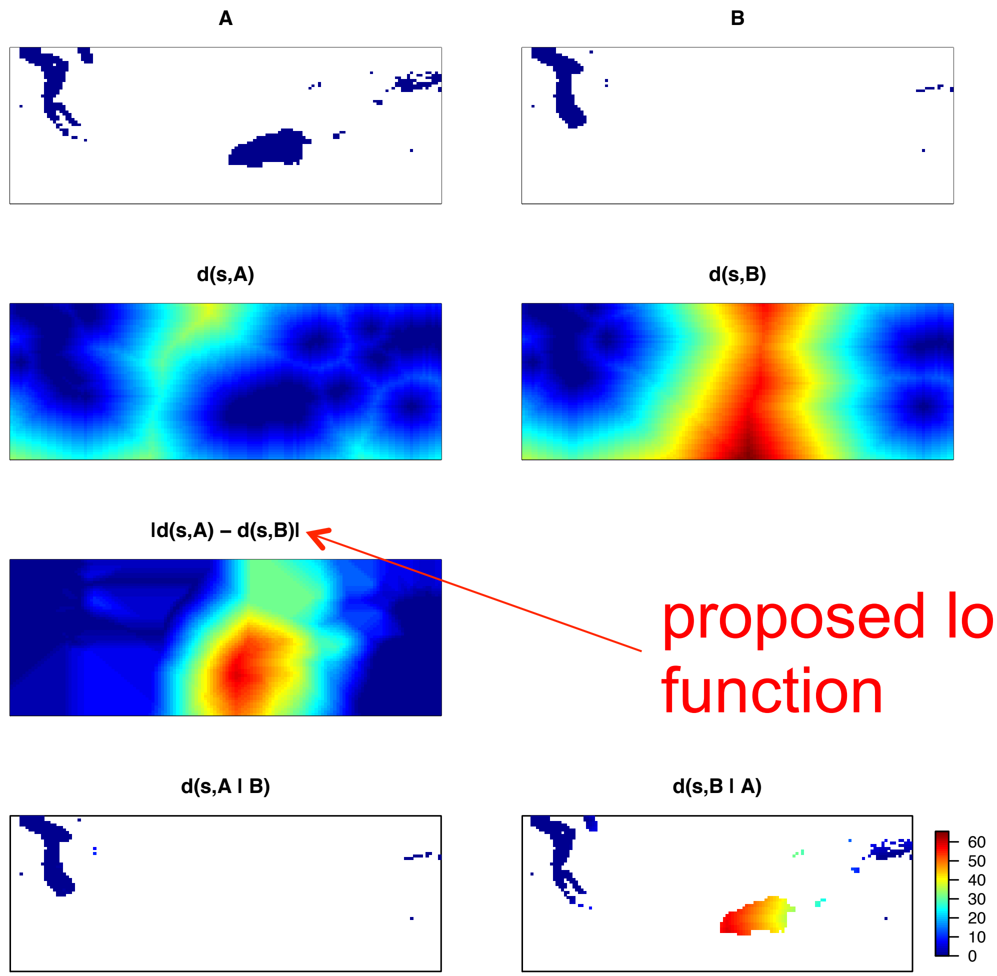
Covariance function for the loss differential's spatial dependence structure (need to replace with an estimate)

Accounting for Location Errors

NCAR

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.



proposed loss function

G. (2013, *MWR*, **141** (1), 340 – 355)

Accounting for Location Errors

NCAR

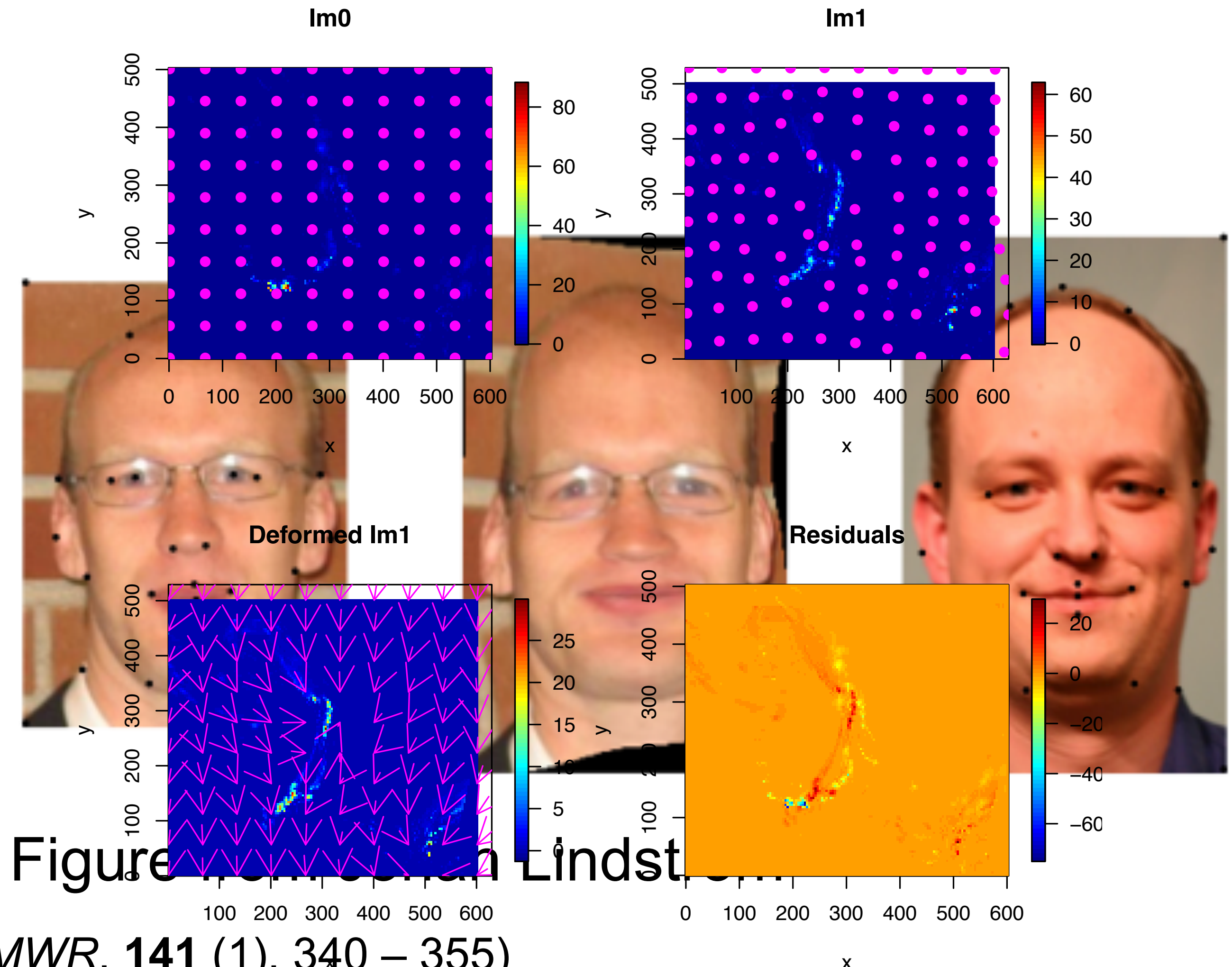


Figure [illegible] Lindst[illegible]
G. (2013, *MWR*, **141** (1), 340 – 355)

Accounting for Location Errors

NCAR

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

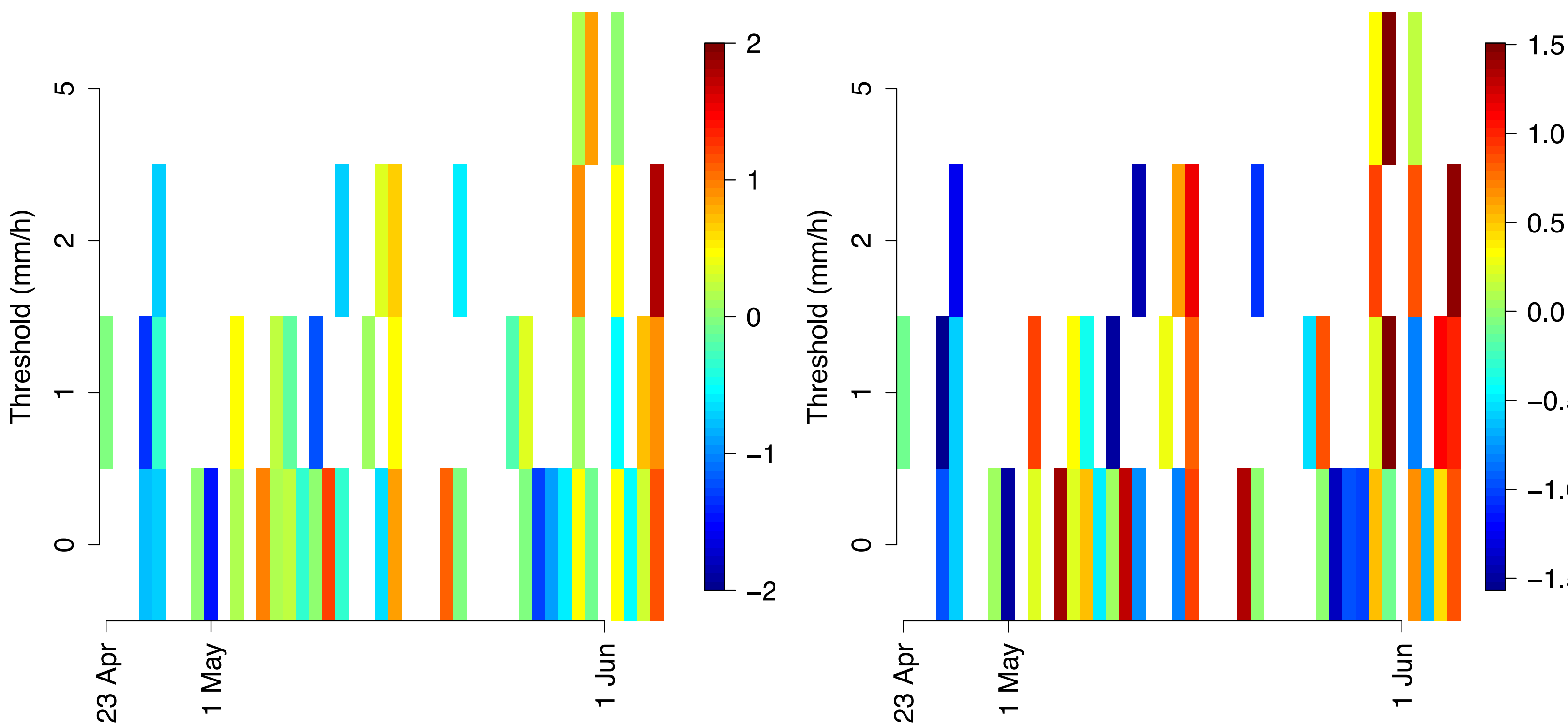
G. (2013, *MWR*, **141** (1), 340 – 355)

Accounting for Location Errors

NCAR

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

mean(**d**) S



G. (2013, *MWR*, **141** (1), 340 – 355)

Summary of SPCT and SPCT + Warping

NCAR

- 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

NCAR

- 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).