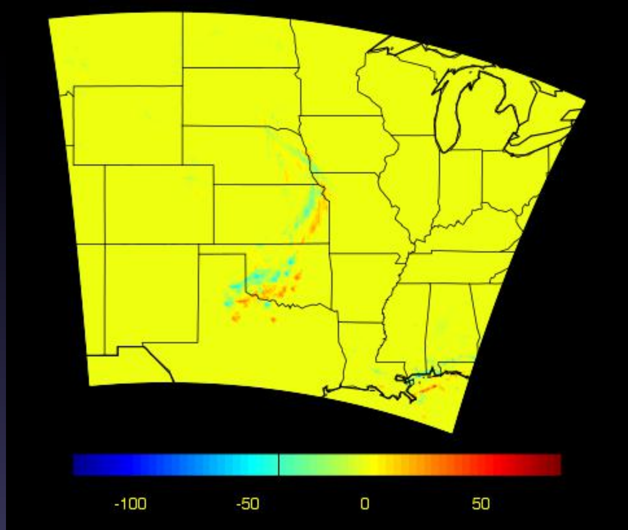
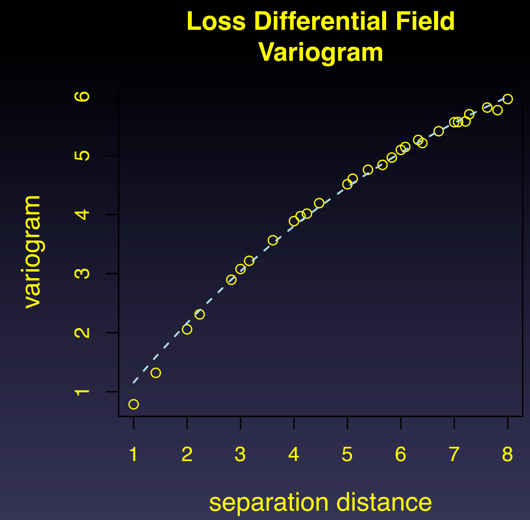
The Spatial Prediction Comparison Test (SPCT)

WRF Users' Workshop

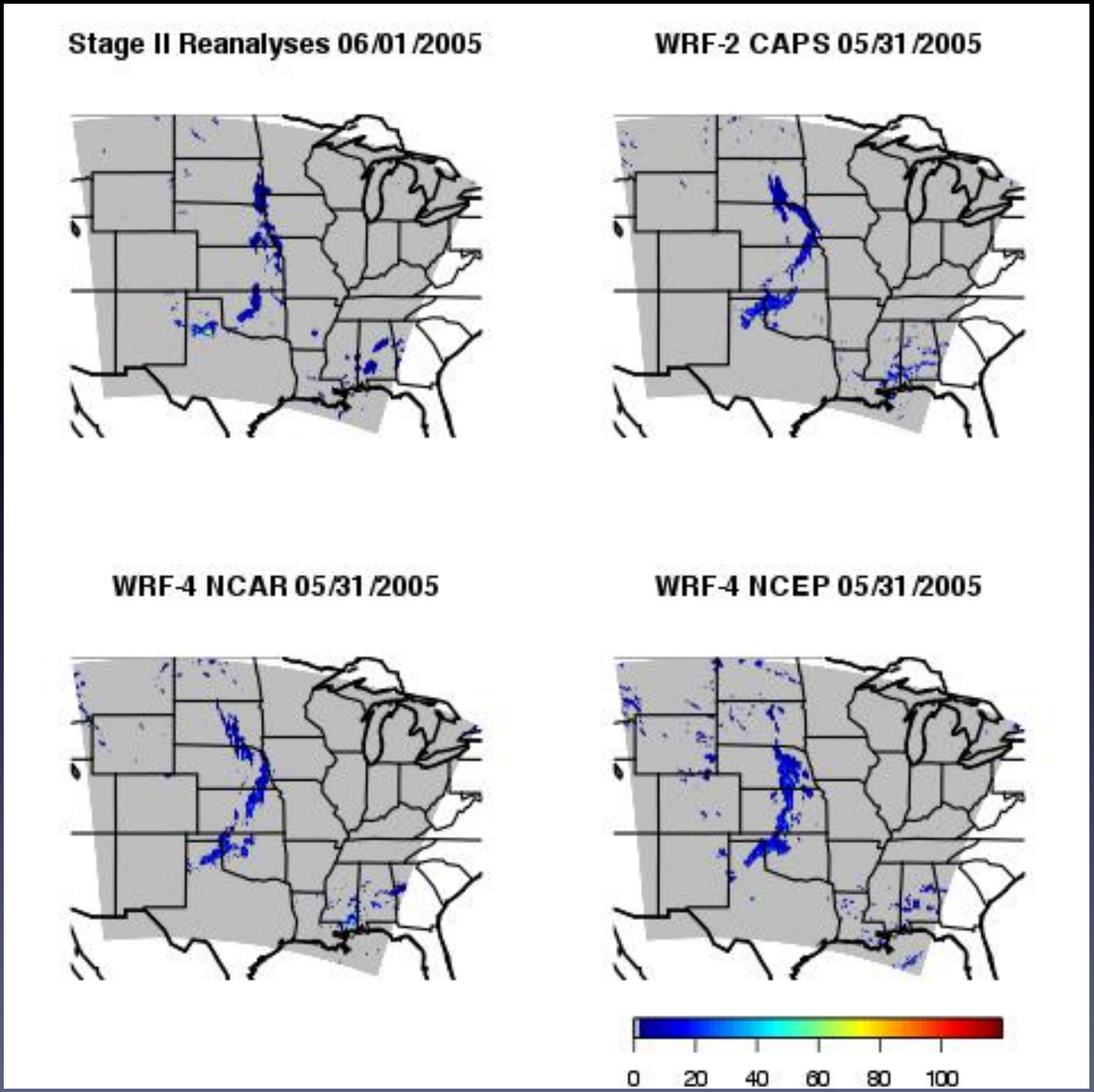
28 June 2013



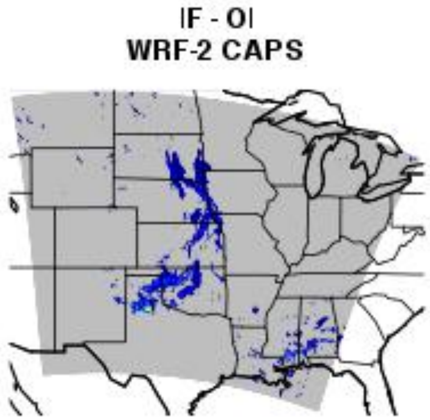


Eric Gilleland
Joint Numerical Testbed (JNT)
Research Applications Laboratory (RAL)
National Center for Atmospheric Research (NCAR)

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


Stage II Reanalyses 06/01/2005
WRF-2 CAPS 05/31/2005
WRF-4 NCAR 05/31/2005
WRF-4 NCEP 05/31/2005
0
20
40
60
80
100

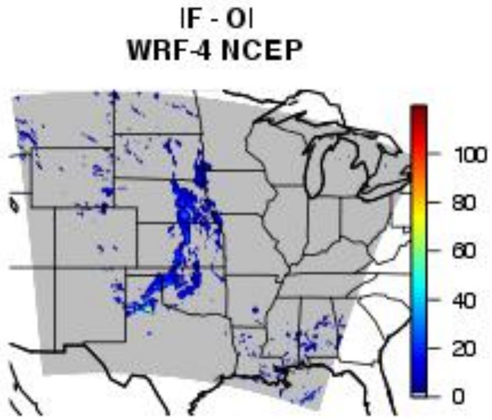
Mean ≈ 2.31 mm/h
Std. Err. ≈ 0.0095 mm/h



Mean ≈ 2.31 mm/h
Std. Err. ≈ 0.0098 mm/h

High Spatial Correlation! Std. Err.'s are too small!

Mean ≈ 2.34 mm/h
Std. Err. ≈ 0.0092 mm/h



Ignoring grid points where differences are zero.

Loss Differential Fields (|F1 – O| - |F2 – O|)

Mean ≈ 0.13 mm/h
Std. Err. ≈ 4.716 mm/h

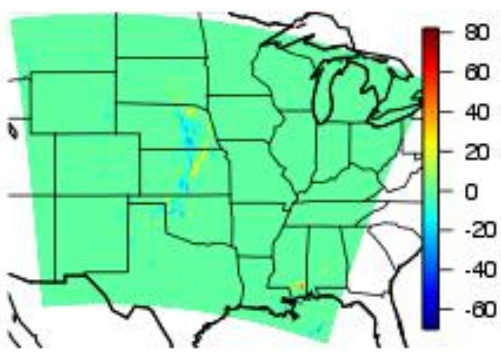
WRF-2 CAPS vs WRF-4 NCAR

WRF-2 CAPS vs WRF-4 NCEP

Mean ≈ -0.41 mm/h
Std. Err. ≈ 4.762 mm/h

High Spatial Correlation! Std. Err.'s are too small!

WRF-4 NCAR vs WRF-4 NCEP



Mean ≈ -0.52 mm/h
Std. Err. ≈ 4.878 mm/h

Ignoring grid points where differences are zero.

The Spatial Prediction Comparison Test (SPCT)

Some Notation

d1 = d(**F1**,**O**) = loss(**F1**,**O**)

d2 = d(**F2**,**O**) = loss(**F2**,**O**)

(e.g., **d1**=|**F1** – **O**|)

Note: **d1** = [d1(x,y)] is calculated at each grid point, so that each grid point has a value that is ultimately aggregated over the entire field.

D = **d1** – **d2**

Note: **D** is a spatial field, called the loss differential field, of the difference in loss function values (again at each grid point) for F1 and F2.

$$\bar{\mathrm{D}} = \frac{1}{N} \sum_{i \in (x,y)} \mathrm{D}_i$$

Average loss differential

Precipitation: calculate only over grid points where at least one field is non-zero.

The Spatial Prediction Comparison Test (SPCT)

Test Statistic

$$S_{\bar{\mathbf{D}}} = \frac{\bar{\mathbf{D}} - \mu}{\mathrm{Var}\left(\bar{\mathbf{D}}\right)} = \frac{\bar{\mathbf{D}}}{\mathrm{Var}\left(\bar{\mathbf{D}}\right)}$$

Null hypothesis is that the test statistic is zero. Two-sided alternative is that it is not equal to zero. Can also perform a one-sided less (or greater) than alternative instead.

Assumption for test is that the test statistic follows a standard normal distribution. That is, that the mean loss differential is normally distributed.

The estimation of

$\mathrm{Var}\left(\bar{\mathrm{D}}\right)$ is the trick.

The Spatial Prediction Comparison Test (SPCT)

(1 – α) 100% Confidence Intervals can also be estimated using:

$$\bar{\mathrm{D}} \pm z_{\alpha/2}\sqrt{\mathrm{Var}\left(\bar{\mathrm{D}}\right)}$$

The Spatial Prediction Comparison Test (SPCT)

Estimate Proposed by Hering and Genton (2011, *Technometrics*, **53**, 414 – 425), which is based on the univariate time series test proposed by Diebold and Mariano (1995, *J. Bus. Econ. Stat.*, **13**, 253 – 263) is:

$$\widehat{\mathrm{Var}}\left(\bar{\mathrm{D}}\right) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\hat{C}\left(h_{ij}\right)$$

h_{ij} are all points separated by the amount |i – j|.

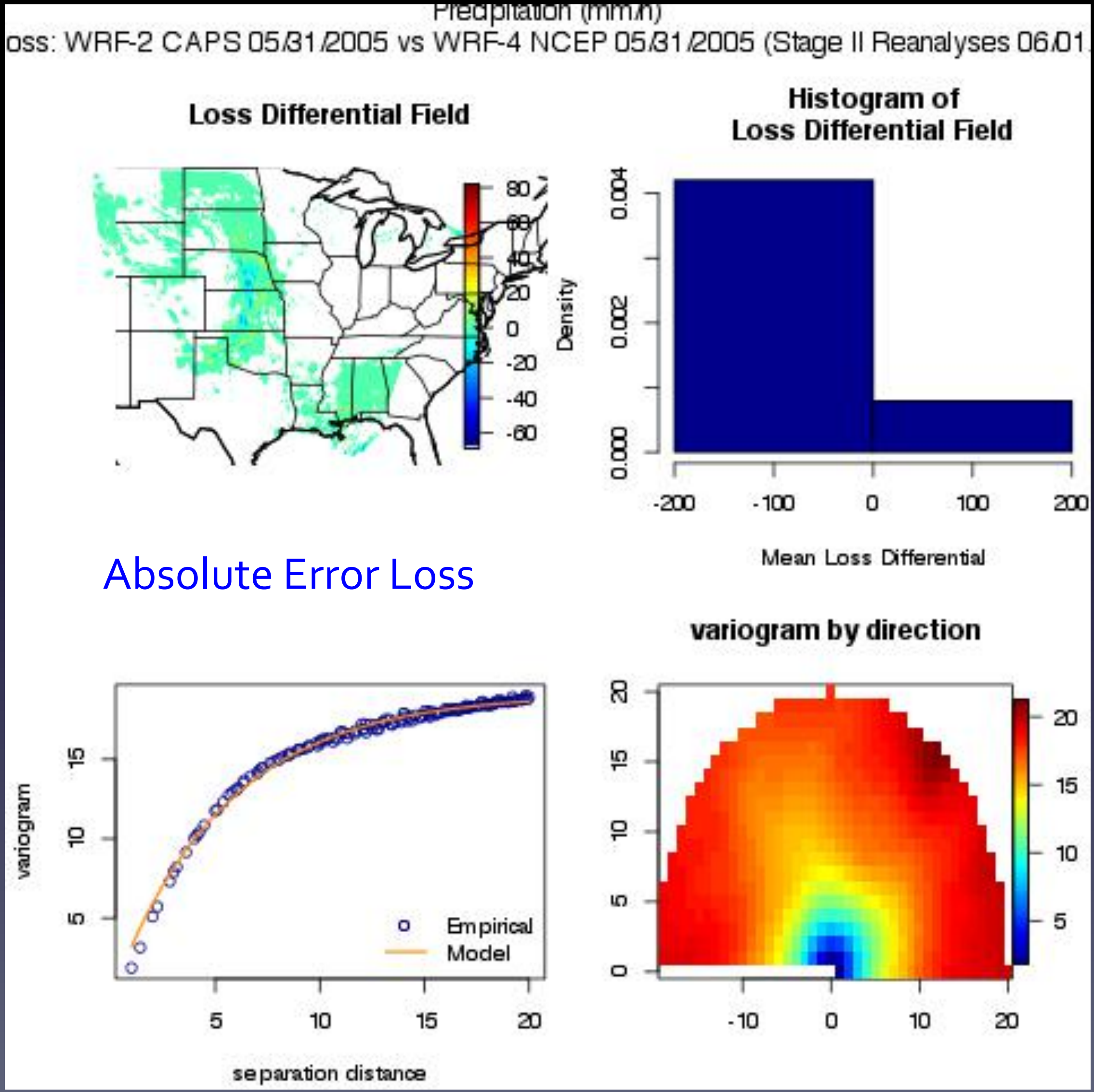
$$\hat{C}\left(h_{ij}\right) = \hat{\gamma}\left(\infty\right) - \hat{\gamma}\left(h_{ij}\right)$$

$$\gamma\left(h_{ij}\right) = \frac{1}{2\left|N\left(h_{ij}\right)\right|}\sum_{N\left(h_{ij}\right)}\left[\mathrm{D}\left(x_i, y_i\right) - \mathrm{D}\left(x_j, y_j\right)\right]^2$$

A parametric variogram is fit to the empirical one, and it is the sums of the lags of the parametric model that are used to estimate the variance of the mean loss differential.

The Spatial Prediction Comparison Test (SPCT)

p-value for one-sided less than hypothesis is about 0.479 (fail to reject null hypothesis that mean loss differential is equal to zero).



Mean loss differential is about -0.21 with 95% CI about (-8.12, 7.69)

The Spatial Prediction Comparison Test (SPCT)

Examples of Loss Functions:

- Simple loss: F – O
- Square Error Loss: $(F - O)^2$
- Absolute Error Loss: $|F - O|$
- Correlation Skill

$$\left(O - \bar{O}\right)\left(F - \bar{F}\right)$$

The Spatial Prediction Comparison Test (SPCT)

Summary and Conclusions

- Test is for competing forecasts.
- Assumption is that the test statistic follows a standard normal distribution.
- No assumption about the distributions of O, F1 or F2.
- Works for any loss function.
- Accounts for spatial correlation.
- Does not require a gridded spatial field.
- Computationally efficient for gridded spatial fields and other spatial fields that do not have a lot of locations.
- Cannot always find a good fitting variogram.
- Does not account for location errors or accumulation of numerous small-scale errors.

The Spatial Prediction Comparison Test (SPCT)

Warping + Absolute Error Loss

The Spatial Prediction Comparison Test (SPCT)

Warping + Absolute Error Loss

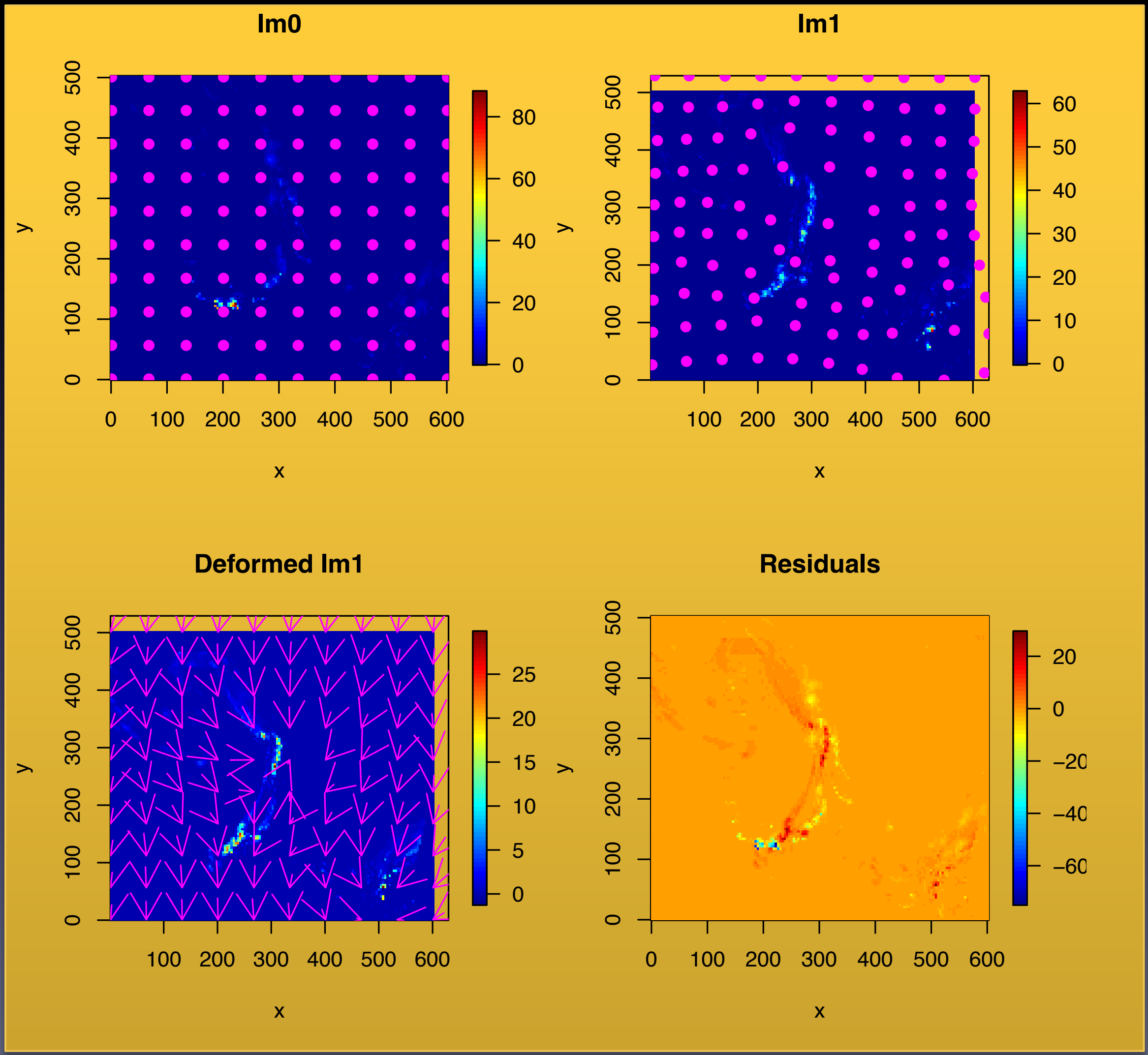
$$\text{loss}\Big(\text{F}\big(W(x_i, y_i)\big),\ \text{O}(x_i, y_i)\Big) + \text{loss}\Big((x_i, y_i), W(x_i, y_i)\Big)$$

$$\Big|\text{F}\big(W(x_i, y_i)\big) - \text{O}(x_i, y_i)\Big| + \sqrt{\big(x_i - W_x(x_i, y_i)\big)^2 + \big(y_i - W_y(x_i, y_i)\big)^2}$$

W represents a vector-valued (warping) function that yields the new coordinates of the points after applying the (optimal) warp.

The Spatial Prediction Comparison Test (SPCT)

Warping + Absolute Error Loss



The Spatial Prediction Comparison Test (SPCT)

Warping + Absolute Error Loss

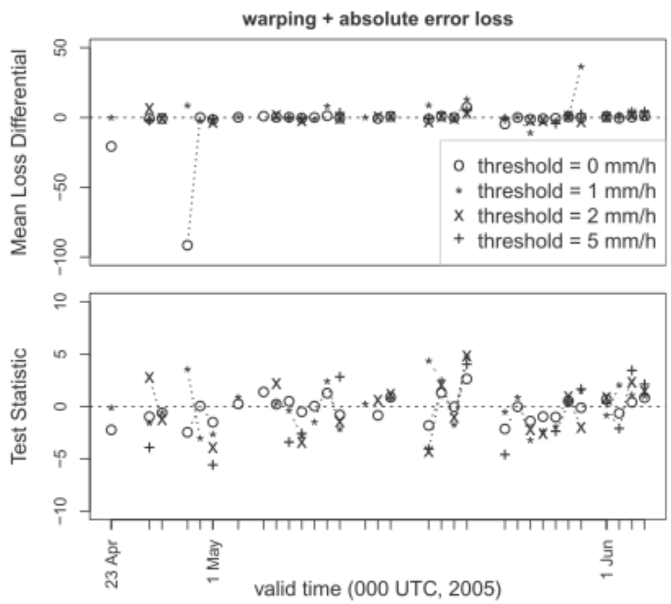


FIG. 3. Results for (top) mean differential $\overline{D}$ based on warping plus AE loss, and (bottom) the associated test statistic. Dotted lines indicate contiguously available scores. Negative (positive) values imply that ARW-WRF (NMM) is better on average in terms of warping loss.

TABLE 2. Warping plus AE loss results for the 32 test cases. A dash (—) indicates that a good fitting variogram model to $D(\mathbf{s})$ was not found, thus no test was performed. Negative (positive) values imply ARW-WRF performs better (worse) than NMM. Note that there is no stage II precipitation ≥5 mm h^{-1} on 29 Apr 2005.

Valid date (0000 UTC 2005)	$\overline{D}$ Threshold (mm h^{-1}) 0	1	2	5
23 Apr	13.35[a]	−9.83[b]	17.67[c]	−34.73[c]
26 Apr	—	—	21.83[c]	−20.81[c]
27 Apr	7.80	−69.97[b]	9.63[a]	8.05[a]
29 Apr	3.42	31.92[c]	—	—
30 Apr	—	−145.51[c]	—	—
1 May	−36.82[a]	−42.87[b]	−36.92[c]	−100.93[c]
3 May	—	24.59[a]	—	—
5 May	2.69	—	—	—
6 May	1.92	0.48	—	—
7 May	—	−54.48[b]	—	−19.65[c]
8 May	−6.64	−25.21[c]	−26.78[c]	−10.55[b]
9 May	—	−18.40[a]	—	—
10 May	—	20.52[c]	8.31[a]	—
11 May	—	−17.98[c]	−19.32[c]	26.23[b]
13 May	—	12.27[a]	−9.14[a]	9.74[b]
14 May	−3.70	—	−2.07	—
15 May	7.97	25.18[a]	8.69	−10.02[a]
18 May	−12.42[a]	−26.24[c]	−50.74[c]	−28.66[c]
19 May	14.58[a]	109.14[c]	10.15[c]	21.83[c]
20 May	—	−41.91[c]	—	—
21 May	13.73[a]	−34.54[c]	−43.12[c]	−1.76
24 May	−10.43[a]	56.70[c]	−7.30[c]	−30.66[c]
25 May	−1.51	—	—	—
26 May	−7.82[a]	—	−36.10[b]	−12.60[c]
27 May	−15.03	−23.77[c]	0.71	3.62
28 May	−14.09	−30.33[a]	3.16	2.60[a]
29 May	—	−9.29[a]	0.75	—
30 May	−2.06	8.76	−59.34[a]	−3.99
1 Jun	−1.47	20.27	—	—
2 Jun	—	1.91	−11.37[b]	−46.96[c]
3 Jun	3.30	3.43	5.27[a]	—
4 Jun	5.81	—	10.05	12.10*

[a] Significance at the 5% level.
[b] Significance at the 1% level.
[c] Significance at the <1% level.

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

The Spatial Prediction Comparison Test (SPCT)

Summary and Conclusions

- Applying image warping first results in a test that accounts for location errors as well as spatial correlation.
- Optimizing the warp function takes time, but is not terribly inefficient either.
- Can be applied to non-gridded fields, but perhaps trickier.
- R image warping package on its way.

Future Work

Additional uncertainty introduced because of uncertainty associated with fitting the warp function to the fields. Can this be incorporated into the test?

It is possible to extend this to a test for spatio-temporal fields, but how exactly?

The Spatial Prediction Comparison Test (SPCT)

`SpatialVx`: R package for implementing spatial forecast verification methods. Available on CRAN, but still incomplete and not heavily tested.

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

Includes functions for implementing the spatial prediction comparison test. See help files for `spatMLD, fit.spatMLD` as well as their summary and plot method functions. Image warping + AE loss will be included once the warping package is finished and uploaded.