Spatial forecast verification

Manfred Dorninger
University of Vienna
Vienna, Austria
manfred.dorninger@univie.ac.at

Thanks to: B. Ebert, B. Casati, C. Keil
7th Verification Tutorial Course, Berlin, 3-6 May, 2017
Motivation:

Model – VERA: RMSE pmsl, 13-24 h

- Aladin (1.76)
- LM (1.80)
- ECMWF (1.33)

RMSE ~9 km

RMSE ~25 km

RMSE ~2 km
Motivation:

→ The „double“ penalty problem

- fine-scale model catches the small-scale trough but at the wrong place (or time)
- gets penalized twice
- increases quadratic measures compared to coarse model
- true for other continuous variables as well (e.g., precipitation, wind speed, etc.)
Traditional spatial verification (grid point wise approach)

Compute statistics on forecast-observation pairs

- Continuous values (e.g., precipitation amount, temperature, NWP variables):
  - mean error, MSE, RMSE, correlation
  - anomaly correlation, S1 score

- Categorical values (e.g., precipitation occurrence):
  - Contingency table statistics (POD, FAR, etc...)

Anomaly correlation

$AC = \frac{\sum (F - C)(O - C)}{\sqrt{\sum (F - C)^2} \sqrt{\sum (O - C)^2}}$ (uncentered)  

or  

$\frac{\sum (F - C) - (F - C)))(O - C) - (O - C))}{\sqrt{\sum (F - C)^2} \sqrt{\sum (O - C)^2}}$ (centered)
Traditional spatial verification using categorical scores

Contingency Table

<table>
<thead>
<tr>
<th>Observed</th>
<th>Predicted</th>
</tr>
</thead>
<tbody>
<tr>
<td>yes</td>
<td>hits</td>
</tr>
<tr>
<td>no</td>
<td>misses</td>
</tr>
</tbody>
</table>

Correct negatives:

\[
\text{Correct Negatives} = \text{misses} + \text{false alarms}
\]

PREDICTED |

\[
\text{FBI} = \frac{\text{hits + false alarms}}{\text{hits + misses}}
\]

\[
\text{POD} = \frac{\text{hits}}{\text{hits + misses}}
\]

\[
\text{FAR} = \frac{\text{false alarms}}{\text{hits + false alarms}}
\]

\[
\text{TS} = \frac{\text{hits}}{\text{hits + misses + false alarms}}
\]

\[
\text{ETS} = \frac{\text{hits} - \text{hits}_{\text{random}}}{\text{hits + misses + false alarms} - \text{hits}_{\text{random}}}
\]
POD=0.39, FAR=0.63, CSI=0.24
Traditional spatial verification

- Requires an exact match between forecasts and observations at every grid point

  - Problem of "double penalty" - event predicted where it did not occur, no event predicted where it did occur

  - Traditional scores do not say very much about the source or nature of the errors

- Hi res forecast
  - RMS ~ 4.7
  - POD=0, FAR=1
  - TS=0

- Low res forecast
  - RMS ~ 2.7
  - POD~1, FAR~0.7
  - TS~0.3
What do traditional scores not tell us?

• Traditional approaches provide overall measures of skill but...

• They don't provide much diagnostic information about the forecast:
  – What went wrong? What went right?
  – How close is the forecast to observation (in terms of spatial thinking)?
  – Does the forecast look realistic?
  – How can I improve this forecast?
  – How can I use it to make a decision?

• Best performance for smooth forecasts!!!

• Some scores are insensitive to the size of the errors...
Spatial forecasts

Weather variables defined over spatial domains have coherent spatial structure and features

New spatial verification techniques aim to:

- account for field spatial structure
- provide information on error in physical terms
- account for uncertainties in location (and timing)
Spatial verification types

Neighborhood (fuzzy) verification methods
give credit to "close" forecasts

Scale separation methods
measure scale-dependent error

Features-based methods
evaluate attributes of identifiable features

Field deformation
evaluate phase errors
Spatial verification types


Fig. 1. Schematic representations of the four categories of verification methods reviewed in this paper. (top) The neighborhood and scale-separation methods can both be considered "filtering" approaches while (bottom) the feature-based and field deformation methods fall under the "displacement" category.
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
<th>Method type</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BCETS</td>
<td>Bias-corrected ETS</td>
<td>Traditional</td>
<td>Mesinger (2008)</td>
</tr>
<tr>
<td>Composite</td>
<td>Composite method</td>
<td>Features based*</td>
<td>Ebert and McBride (2000); Ebert and Gallus (2009)</td>
</tr>
<tr>
<td>CRA</td>
<td>Contiguous rain area</td>
<td>Features based</td>
<td></td>
</tr>
<tr>
<td>DIST</td>
<td>Distributional method</td>
<td>Neighborhood</td>
<td>Marsigli et al. (2006)</td>
</tr>
<tr>
<td>FQI</td>
<td>Forecast quality index</td>
<td>Field deformation*</td>
<td>Venugopal et al. (2005)</td>
</tr>
<tr>
<td>FSS</td>
<td>Fractions skill score</td>
<td>Neighborhood</td>
<td>Roberts (2005); Roberts and Lean (2008); Mittermaier and Roberts (2009)</td>
</tr>
<tr>
<td>IS</td>
<td>Intensity scale</td>
<td>Scale separation</td>
<td>Casati et al. (2004); Casati (2009)</td>
</tr>
<tr>
<td>IW</td>
<td>Image warping</td>
<td>Field deformation</td>
<td>E. Gilleland, J. Lindström, and F. Lindgren (2009, unpublished manuscript); Lindström et al. (2009)</td>
</tr>
<tr>
<td>MODE</td>
<td>Method for Object-based Diagnostic Evaluation</td>
<td>Features based</td>
<td>Davis et al. (2006, 2009)</td>
</tr>
<tr>
<td>MSV</td>
<td>Multiscale variability</td>
<td>Scale separation</td>
<td>Zapeda-Arce et al. (2000); Harris et al. (2001); Mittermaier (2006)</td>
</tr>
<tr>
<td>Neighborhood</td>
<td>Neighborhood based methods</td>
<td>Neighborhood</td>
<td>Ebert (2008, 2009)</td>
</tr>
<tr>
<td>Procrustes</td>
<td>Cell identification and Procrustes shape analysis</td>
<td>Features based</td>
<td>Micheas et al. (2007)</td>
</tr>
<tr>
<td>Procrustes2</td>
<td>Multiscale cell identification and Procrustes shape analysis</td>
<td>Scale separation–Features based</td>
<td>Lack et al. (2009)</td>
</tr>
<tr>
<td>SAL</td>
<td>Structure, amplitude, and location</td>
<td>Features based</td>
<td>Wernli et al. (2008, 2009)</td>
</tr>
<tr>
<td>Traditional</td>
<td>Point-based comparison</td>
<td>Point</td>
<td>Jolliffe and Stephenson (2003)</td>
</tr>
<tr>
<td>VGM</td>
<td>Variogram</td>
<td>Scale separation*</td>
<td>Marzban and Sandgathe (2009)</td>
</tr>
</tbody>
</table>

* A method that only loosely belongs to the given method type.
Neighborhood (fuzzy) verification methods

→ give credit to "close" forecasts

“close“

“not close“
Neighborhood verification methods

- Don't require an exact match between forecasts and observations
  - Unpredictable scales
  - Uncertainty in observations

- Look in a space / time neighborhood around the point of interest

- Evaluate using categorical, continuous, probabilistic scores / methods
Neighborhood verification methods

Treatment of forecast data within a window:

- Mean value (upscaling)
- Occurrence of event* somewhere in window
- Frequency of events in window $\rightarrow$ probability
- Distribution of values within window

May also look in a neighborhood of observations

*Event defined as a value exceeding a given threshold, for example, rain exceeding 1 mm/hr
Moving windows

For each combination of neighborhood size and intensity threshold, accumulate scores as windows are moved through the domain.
Neighborhood verification framework

Neighborhood methods use one of two approaches to compare forecasts and observations:

- single observation – neighborhood forecast (SO-NF, user-oriented)
- neighborhood observation – neighborhood forecast (NO-NF, model-oriented)
Different neighborhood verification methods have different decision models for what makes a *useful forecast*

<table>
<thead>
<tr>
<th>Neighborhood method</th>
<th>Matching strategy*</th>
<th>Decision model for useful forecast</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upscaling (Zepeda-Arce et al. 2000; Weygandt et al. 2004)</td>
<td>NO-NF</td>
<td>Resembles obs when averaged to coarser scales</td>
</tr>
<tr>
<td>Minimum coverage (Damrath 2004)</td>
<td>NO-NF</td>
<td>Predicts event over minimum fraction of region</td>
</tr>
<tr>
<td>Fuzzy logic (Damrath 2004), joint probability (Ebert 2002)</td>
<td>NO-NF</td>
<td>More correct than incorrect</td>
</tr>
<tr>
<td>Fractions skill score (Roberts and Lean 2008)</td>
<td>NO-NF</td>
<td>Similar frequency of forecast and observed events</td>
</tr>
<tr>
<td>Area-related RMSE (Rezacova et al. 2006)</td>
<td>NO-NF</td>
<td>Similar intensity distribution as observed</td>
</tr>
<tr>
<td>Pragmatic (Theis et al. 2005)</td>
<td>SO-NF</td>
<td>Can distinguish events and non-events</td>
</tr>
<tr>
<td>CSRR (Germann and Zawadzki 2004)</td>
<td>SO-NF</td>
<td>High probability of matching observed value</td>
</tr>
<tr>
<td>Multi-event contingency table (Atger 2001)</td>
<td>SO-NF</td>
<td>Predicts at least one event close to observed event</td>
</tr>
<tr>
<td>Practically perfect hindcast (Brooks et al. 1998)</td>
<td>SO-NF</td>
<td>Resembles forecast based on perfect knowledge of observations</td>
</tr>
</tbody>
</table>

*NO-NF = neighborhood observation-neighborhood forecast, SO-NF = single observation-neighborhood forecast*

Detailed description of Fraction Skill Score (FSS) (Roberts and Lean, 2008)

We want to know

- How forecast skill varies with neighborhood size
- The smallest neighborhood size that can be used to give sufficiently accurate forecasts
- Does higher resolution NWP provide more accurate forecasts on scales of interest (e.g., river catchments)

Step 1: FC and Observation/Analysis have to be on the same grid.
Step 2: Choose suitable thresholds $q$ (e.g.: 0.5, 1, 2, 4 mm)
Step 3: Convert FC/AN fields to binary fields $I_O$ and $I_M$ according to threshold

$$I_O = \begin{cases} 
1 & O_r \geq q \\
0 & O_r < q \end{cases} \quad \text{and} \quad I_M = \begin{cases} 
1 & M_r \geq q \\
0 & M_r < q \end{cases}$$
Detailed description of Fraction Skill Score (FSS) (Roberts and Lean, 2008)

Step 4: Generate fractions for all thresholds:

\[ P_{\text{obs}} \quad 1\times1 \]
\[ P_{\text{fcst}} \quad 1\times1 \]

\[ P_{\text{obs}} \quad 35\times35 \]
\[ P_{\text{fcst}} \quad 35\times35 \]

\[ P_{\text{obs}} = \text{fraction of obs grid points} \geq \text{threshold} \]
\[ P_{\text{fcst}} = \text{fraction of fcst grid points} \geq \text{threshold} \]
Detailed description of Fraction Skill Score (FSS) (Roberts and Lean, 2008)

Step 5: Compute fraction skill score for all thresholds:

\[
FSS = 1 - \frac{1}{N} \sum_{i=1}^{N} (P_{\text{fcst}} - P_{\text{obs}})^2
\]

\[
FSS(n) = \frac{\text{MSE}(n) - \text{MSE}(n)_{\text{ref}}}{\text{MSE}(n)_{\text{perfect}} - \text{MSE}(n)_{\text{ref}}} = 1 - \frac{\text{MSE}(n)}{\text{MSE}(n)_{\text{ref}}}
\]

Maximum estimation (low-skill reference) of MSE:

\[
(P_{\text{fcst}} - P_{\text{obs}})^2 = P_{\text{fcst}}^2 - 2P_{\text{fcst}}P_{\text{obs}} + P_{\text{obs}}^2 \sim P_{\text{fcst}}^2 + P_{\text{obs}}^2 = \text{MSE}_{\text{ref}}
\]
Detailed description of Fraction Skill Score (FSS) (Roberts and Lean, 2008)

Step 6: Graphical presentation for each threshold and spatial scale:

Interpretation:
• Skill increases with spatial scale
• The smaller the displacement error the faster the skill increases with increasing spatial scale
• When the length of the moving window is smaller or equal the displacement error there is no skill and FSS=0
Detailed description of Fraction Skill Score (FSS) (Roberts and Lean, 2008)

Q: What happens if size of moving window is equal to domain size?
Q: What are useful (skillfull) numbers of FSS?

\[ f_0 = \text{domain obs fraction on the grid scale} \]

For \( f_0 = 0.2 \) (20%), the target skill is:

\[ \text{FSS} = 0.5 + \frac{0.2}{2} = 0.6 \]
Detailed description of Fraction Skill Score (FSS) (Roberts and Lean, 2008)
Scale separation methods

→ scale-dependent error

1. Which spatial scales are well represented and which scales have error?
2. How does the skill depend on the precipitation intensity?

NOTE: **scale** = single band spatial filter → features of different scales → feedback on different physical processes and model parameterizations

In the neighborhood based (fuzzy) verification, the **scale** is the neighborhood size (low band pass filter): as the scale increases the exact positioning requirements are more and more relaxed
What is the difference between neighborhood and scale separation approaches?

• Neighborhood verification methods
  - Get scale information by *filtering out higher resolution scales*

• Scale separation methods
  - Get scale information by *isolating scales of interest*
Nimrod case study: intense storm displaced

Step 1: Gridded data, square domain with dimension $2^n$
It can be applied to any meteorological field ... however, it was specifically designed for spatial precipitation forecasts ...
Step 2: Intensity: threshold to obtain binary images (categorical approach)
Step 3: Scale $\rightarrow$ wavelet decomposition of binary error

mean (1280 km)

Scale $l=8$ (640 km)

Scale $l=7$ (320 km)

Scale $l=6$ (160 km)

Scale $l=5$ (80 km)

Scale $l=4$ (40 km)

Scale $l=3$ (20 km)

Scale $l=2$ (10 km)

Scale $l=1$ (5 km)

$$E_u = \sum_{l=1}^{L} E_{u,l} \quad \text{MSE}_u = \sum_{l=1}^{L} \text{MSE}_{u,l}$$
Step 4: MSE skill score for each threshold and scale component

$$SS_{u,l} = \frac{MSE_{u,l} - MSE_{u,l,random}}{MSE_{u,l,best} - MSE_{u,l,random}} = 1 - \frac{MSE_{u,l}}{2\varepsilon(1-\varepsilon)/L}$$

Sample climatology (base rate)
**Strengths**

Categorical approach → robust and resistant

Wavelets → cope with spatially discontinuous fields characterized by the presence of few sparse non-zero features

→ suitable for spatial precipitation forecasts

**Weaknesses**

need gridded data on a square domain with dimension $2^n$
Figure 3: Intensity-scale skill score (ISS) for 6 h accumulated precipitation forecasts and up to 18 h lead time. Forecast ranges: 0-6 h, 6-12 h and 12-18 h; period: Jun-Nov 2007; x-axes: scale (1=8 km, 2=16 km, 4=32 km, 8=64 km, 16=128 km, 32=256 km, 64=512 km); y-axes: precipitation thresholds in mm/6 h.
Features-based methods

→ evaluate attributes of features
Feature-based approach (CRA)

Ebert and McBride, J. Hydrol., 2000

• Define entities using (user defined) threshold (Contiguous Rain Areas)

• Horizontally translate the forecast until a pattern matching criterion is met:
  – minimum total squared error between forecast and observations
  – maximum correlation
  – maximum overlap

• The displacement is the vector difference between the original and final locations of the forecast.
CRA error decomposition

Total mean squared error (MSE) before shifting

$$MSE_{total} = MSE_{displacement} + MSE_{volume} + MSE_{pattern}$$

The displacement error is the difference between the mean square error before and after shifting

$$MSE_{displacement} = MSE_{total} - MSE_{shifted}$$

The volume error is the bias in mean intensity

$$MSE_{volume} = (\bar{F} - \bar{X})^2$$

where $\bar{F}$ and $\bar{X}$ are the mean forecast and observed values after shifting.

The pattern error, computed as a residual, accounts for differences in the fine structure,

$$MSE_{pattern} = MSE_{shifted} - MSE_{volume}$$
Example: CRA verification of precipitation forecast over USA

1. What is the location error of the forecast?
2. How do the forecast and observed rain areas compare? Average values? Maximum values?
3. How do the displacement, volume, and pattern errors contribute to the total error?
wrf2 fcst 20050601 hour 00-24

CRA 20050601

Predicted rainfall (shifted)

Analyzed rainfall

wrf2 24h fcst 20050601  n=8423
(33.49°, -102.28°) to (37.77°, -96.00°)
Verif. grid=0.042°  CRA threshold=1.0 mm/h

<table>
<thead>
<tr>
<th>Analysed</th>
<th>Forecast</th>
</tr>
</thead>
<tbody>
<tr>
<td># gridpoints ≥1 mm/h</td>
<td>3304</td>
</tr>
<tr>
<td>Average rainrate (mm/h)</td>
<td>3.58</td>
</tr>
<tr>
<td>Maximum rain (mm/h)</td>
<td>119.63</td>
</tr>
<tr>
<td>Rain volume (km³)</td>
<td>0.51</td>
</tr>
</tbody>
</table>

Displacement (E,N) = [2.20°, 1.92°] max.corr matching

RMS error (mm/d)  
Original: 12.81  
Shifted: 10.24

Correlation coefficient
Original: -0.167  
Shifted: 0.305

Error Decomposition:
Displacement error: 36.1%
Volume error: 0.0%
Pattern error: 63.9%
wrf2 fcst 20050601 hour 00–24

Analysis 20050601

wrf2 24h fcst 20050601 n=11007
(37.52°,−101.29°) to (45.29°,−94.65°)
Verif. grid=0.042° CRA threshold=1.0 mm/h

<table>
<thead>
<tr>
<th></th>
<th>Analysed</th>
<th>Forecast</th>
</tr>
</thead>
<tbody>
<tr>
<td># gridpoints ≥1 mm/h</td>
<td>4840</td>
<td>5699</td>
</tr>
<tr>
<td>Average rainrate (mm/h)</td>
<td>1.52</td>
<td>2.68</td>
</tr>
<tr>
<td>Maximum rain (mm/h)</td>
<td>21.08</td>
<td>27.69</td>
</tr>
<tr>
<td>Rain volume (km³)</td>
<td>0.26</td>
<td>0.46</td>
</tr>
<tr>
<td>Displacement (E,N) = [0.52°,−0.84°]</td>
<td>max corr matching</td>
<td></td>
</tr>
<tr>
<td>RMS error (mm/d)</td>
<td>Original: 5.11</td>
<td>Shifted: 4.65</td>
</tr>
<tr>
<td>Correlation coefficient</td>
<td>−0.040</td>
<td>0.193</td>
</tr>
</tbody>
</table>

Error Decomposition:
- Displacement error: 18.7%
- Volume error: 4.9%
- Pattern error: 76.4%
Sensitivity to rain threshold

- 1 mm h\(^{-1}\)
- 5 mm h\(^{-1}\)
- 10 mm h\(^{-1}\)
Strengths of CRA

The entity-based CRA verification method has a number of attractive features:

(a) It is intuitive, quantifying what we can see by eye
(b) it estimates the location error in the forecast,
(c) the total error can be decomposed into contributions from location, intensity, and pattern,
(d) forecast events can be categorized as hits, misses, etc. These descriptions could prove a useful tool for monitoring forecast performance over time.
Weaknesses of CRA

There are also some drawbacks to this approach:

(a) Pattern matching: it must be possible to associate entities in the forecast with entities in the observations. This means that the forecast must be halfway decent. The verification results for a large number of CRAs will be biased toward the "decent" forecasts, i.e., those for which location and intensity errors could reliably be determined.

(b) The user must choose the pattern matching method as well as the isoline used to define the entities. The verification results will be somewhat dependent on these choices (subjective).

(c) When a forecast and/or observed entity extends across the boundary of the domain it is not possible to be sure whether the pattern match is optimal. If the CRA has a reasonably large area still within the domain then the probability of a good match is high. Ebert and McBride (2000) suggest applying a minimum area criterion to address this issue.
Structure-Amplitude-Location (SAL)

- Verification of rain forecasts in a defined domain
- No match of objects

SAL consists of three components:

- S  structure
- A  amplitude
- L  location

Perfect forecast: S=A=L=0
Step 1: Definition of precipitation objects
S A L – definition of the S-component

\[ S = V(R_{mod}^*) - V(R_{obs}^*) \]

\( V(\ldots) \) denotes the mean (weighted) volume of all scaled precipitation objects structure error in the chosen area.

precipitation „field“

scaling for each object: \( R^* = \frac{R}{R_{max}} \); \( R^* \in \left[ \frac{R_{thresh}}{R_{max}}, \ldots, 1 \right] \)
small intense vs. large weak objects

OBS

MOD

A ≈ 0

S > 0
4.2.1 SAL

intense vs. weak objects with same size

\[ A < 0 \]

\[ S = 0 \]
sensitivity to the object structure

OBS

MOD

\[ V(R^*) \]

\[ R^* \]

\[ X \]

\[ S > 0 \]

\[ S < 0 \]
S A L – definition of the A-component

\[ A = D(R_{mod}) - D(R_{obs}) = 0 \]

\[ D(R) = \frac{1}{N} \sum_{(i,j) \in D} R_{ij}, \]

R denotes precipitation
D(…) denotes the area mean in the chosen area
amplitude error
S A L – definition of the L-component

\[ L = |r(R_{mod}) - r(R_{obs})| = 0 \]

\( r(...) \) denotes the precipitation center of mass in the chosen area
displacement error of the center of mass
S A L – definition of the L-component

\[ L = |r(R_{mod}) - r(R_{obs})| > 0 \]

\( r(\ldots) \) denotes the precipitation center of mass in the chosen area displacement error of the center of mass
S A L – definition of the L-component

$L = |r(R_{mod}) - r(R_{obs})| = 0 !?$

$r(...)$ denotes the precipitation center of mass in the chosen area displacement error of the center of mass.
**S A L – definition of the L-component**

\[ L = |r(R_{mod}) - r(R_{obs})| + |d(r_{mod}) - d(r_{obs})| > 0 \]

- `r(…)` denotes the precipitation center of mass in the chosen area.
- `d(…)` mean (weighted) distance between objects- and area-center of mass.
- Displacement error with impact of object distance to center of mass.

**Model**

\[ d(r_{mod}) = 0 \]

**Observations**

\[ d(r_{obs}) = \frac{R(o_1) \cdot d_1 + R(o_2) \cdot d_2}{R(o_1 + o_2)} \]
\[ S = \frac{(V(R_{mod}^*) - V(R_{obs}^*))}{0.5*(V(R_{mod}^*) + V(R_{obs}^*))} \]

\( V(...) \) denotes the averaged volume of all scaled precipitation objects

\( S \in [-2, ..., 0, ..., +2] \)

\[ A = \frac{(D(R_{mod}) - D(R_{obs}))}{0.5*(D(R_{mod}) + D(R_{obs}))} \]

\( D(...) \) denotes the area mean in a catchment

\( A \in [-2, ..., 0, ..., +2] \)

\[ L = \frac{(|r(R_{mod}) - r(R_{obs})| + 2 \cdot |d(r_{mod}) - d(r_{obs})|)}{dist_{max}(area)} \]

\( r(...) \) denotes the precipitation center of mass in a catchment

\( d(...) \) mean (weighted) distance between objects- and area-center of mass

\( L \in [0, ..., 2] \)
Q: Look at precip fields. What do you expect for S, A and L?

A: $S=A=L=0$; SAL is invariant against pure rotation.
precipitation objects are...

A-component

S-component

II  S<0, A>0
...too small and/or too peaked
...overestimated

S>0, A>0
...too large and/or too flat
...overestimated

III S<0, A<0
...too small and/or too peaked
...underestimated

S>0, A<0
...too large and/or too flat
...underestimated
SAL Examples

SAL case 3, 20070807-20070808, 01h acc., co7

SAL case 4, 20070718-20070721, 01h acc., co2
Field deformation

[arrow] evaluate phase errors
DAS in a nutshell

- **Displacement and Amplitude Score DAS**
- constitutes a spatial measure belonging to the *field verification technique*
- is based on an areal image matcher using classical *optical flow technique*
- has *two* components: DIS and AMP (normalized with characteristic values)
- is applied in both *observation* and *forecast space* (to account for misses and false alarms)
- has been used in *deterministic* mode
- is coded in *python* and freely available

*(Keil and Craig, WAF 2009)*
Optical flow algorithm: Pyramidal Matching

1. **Project** observed and simulated images to same grid

2. **Coarse-grain** both images by averaging of $2^f$ pixels onto one pixel element

3. Compute a displacement vector field that *minimizes the RMSE* within the range of +/- 2 pixel elements

4. Repeat step 2 at successively **finer scales**

5. **Displacement vector** for every pixel results from the sum over all scales

*(Mannstein et al., 2002)*
Pyramidal Image Matching

Step 1: projection on same grid

observation

forecast
Pyramidal Image Matching

Step 2: coarse graining $F=4$

observation

forecast
Pyramidal Image Matching

Step 3: compute displacement vector by minimizing RMSE
Pyramidal Image Matching

Step 3: compute displacement vector by minimizing RMSE
Pyramidal Image Matching

Step 3: compute displacement vector by minimizing RMSE

observation

forecast
Pyramidal Image Matching

Step 4: cycle on finer scales using morphed image

observation

morphed forecast
Pyramidal Image Matching

Step 4: cycle on finer scales using morphed image

observation

morphed forecast
Pyramidal Image Matching

Step 5: sum over all scales

observation

morphed forecast
Pyramidal Image Matching

Step 5: sum over all scales

observation

morphed forecast
Pyramidal Image Matching

Step 5: sum over all scales

observation  morphed forecast
Displacement error field DIS
Displacement error field DIS and Amplitude error field AMP
DAS field: combined DIS and AMP fields
Displacement and Amplitude Score DAS

DAS has two components:

1. displacement error (of observed and forecast imagery)

\[
\overline{DIS} = \frac{1}{n} \sum_{A} DIS(x, y)
\]

2. amplitude error (RMSE of observed and morphed forecast imagery)

\[
\overline{AMP} = \sqrt{\frac{1}{n} \sum_{A} AMP^2(x, y)}^{1/2}
\]

• DAS is applied in observation and forecast space:

\[
DIS = \left( \frac{1}{n_{obs} + n_{fct}} \right) \left( n_{obs} \overline{DIS}_{obs} + n_{fct} \overline{DIS}_{fct} \right)
\]

(Keil and Craig, WAF 2009)
Displacement and Amplitude Score DAS

- **DAS shall be a single valued measure** of forecast quality:
  - underlying principle: complete miss = 100% AMP error

\[
DAS = \frac{DIS}{D_{\text{max}}} + \frac{AMP}{I_0}
\]

- **\(D_{\text{max}}\)**: maximum search distance
- **\(I_0\)**: characteristic intensity chosen to be typical of amplitude of the observed features

*(Keil and Craig, WAF 2009)*
Conclusions

• What method should you use for spatial verification?
  – Depends what question(s) you would like to address

• Many spatial verification approaches
  – Neighborhood – credit for "close" forecasts
  – Scale separation – scale-dependent error
  – Features-based – attributes of features
  – Field deformation – phase and amplitude errors
What method(s) could you use to verify

Wind forecast (sea breeze)

- Neighborhood – credit for "close" forecasts
- Scale separation – scale-dependent error
- Features-based – attributes of features
- Field deformation – phase and amplitude errors
What method(s) could you use to verify Cloud forecast

- Neighborhood – credit for "close" forecasts
- Scale separation – scale-dependent error
- Features-based – attributes of features
- Field deformation – phase and amplitude errors
What method(s) could you use to verify Mean sea level pressure forecast

- **Neighborhood** – credit for "close" forecasts
- **Scale separation** – scale-dependent error
- **Features-based** – attributes of features
- **Field deformation** – phase and amplitude errors

5-day forecast

Analysis
What method(s) could you use to verify

Tropical cyclone forecast

Observed

3-day forecast

**Neighborhood** – credit for "close" forecasts

**Scale separation** – scale-dependent error

**Features-based** – attributes of features

**Field deformation** – phase and amplitude errors