Verification techniques for spatial forecasts

Barbara Casati

Environnement Canada
Environment Canada

Talk outline:

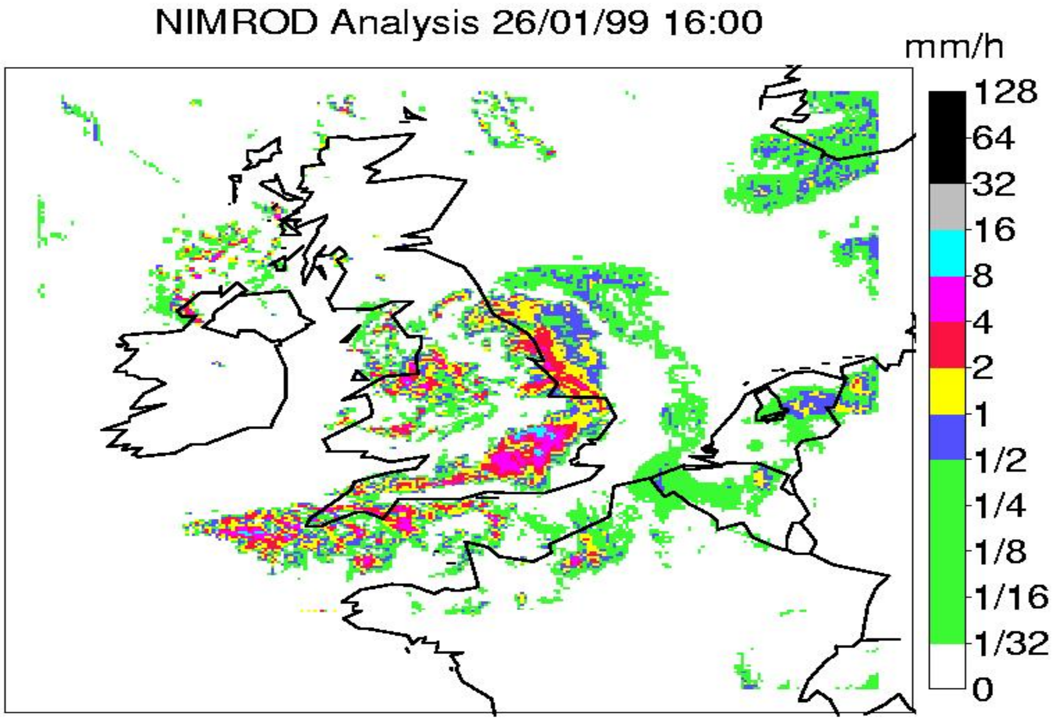
1. Motivation
2. Scale verification
3. Neighborhood-based (fuzzy) verification
4. Error decomposition: displacement + amount
5. Feature-based approaches
6. Hausdorff metrics
7. What about observations ?

Environment Canada Environnement Canada

Canada

Aims and motivation

Weather variables defined over spatial domains: **coherent spatial structure and features** (intrinsic spatial correlation)



Spatial verification techniques **aim** to:

→ account for field spatial structure

→ provide information on error in physical terms

→ account for time-space uncertainties

Verification on different scales

- **Briggs and Levine (1997)** → CONT (MSE, corr)
- **Casati et al. (2004)** → CAT (Heidke SS)
- **Casati and Wilson (accepted)** → PROB (Brier SS)
- **Denis et al. (2003), De Elia et al. (2002)** → CONT (Taylor D)

- **Zepeda-Arce et al. (2000), Harris et al. (2001), Tustison et al. (2003)**

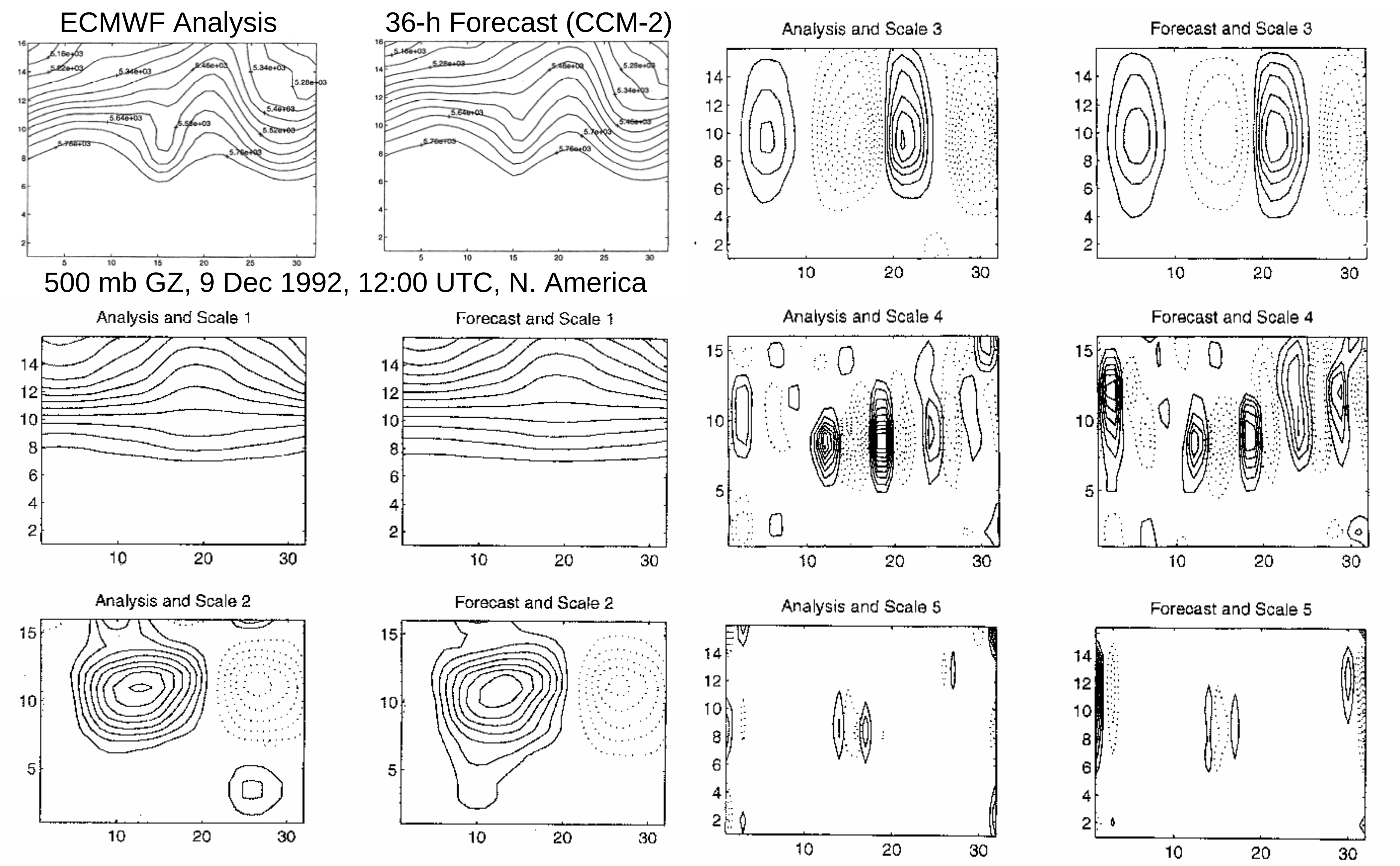
1. Decompose forecast and observation fields into the sum of spatial components on different scales → features of different scales → different physical processes and model parametrizations
   Spatial filters: wavelets, discrete cosine transforms, Fourier, …

2. Perform verification on different scale components, separately (cont. scores; categ. approaches; probability verif. scores)

> - Assess quality and skill on different scales
> - Scale dependency of predictability (no-skill to skill transition scale)
> - Assess the forecast ability of reproducing scale spatial structure of observed precipitation fields

Briggs and Levine 1997

Wavelet scale components

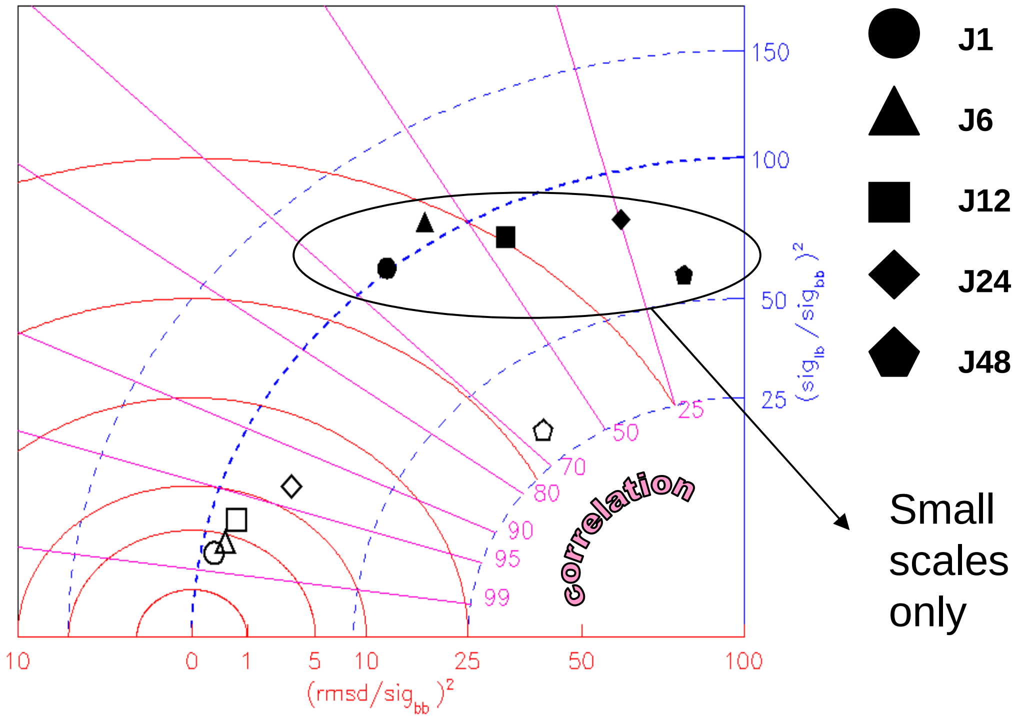


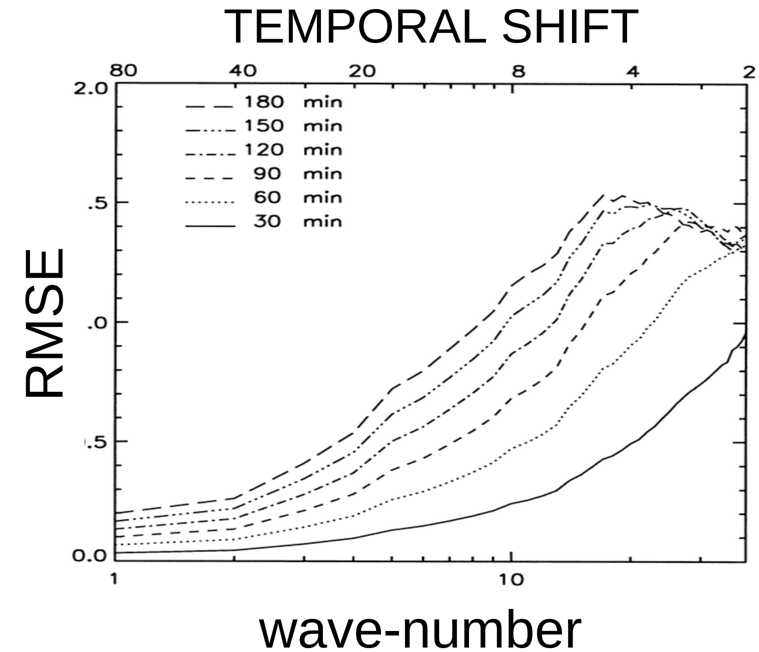
Denis et al. (2003), De Elia et al. (2002)

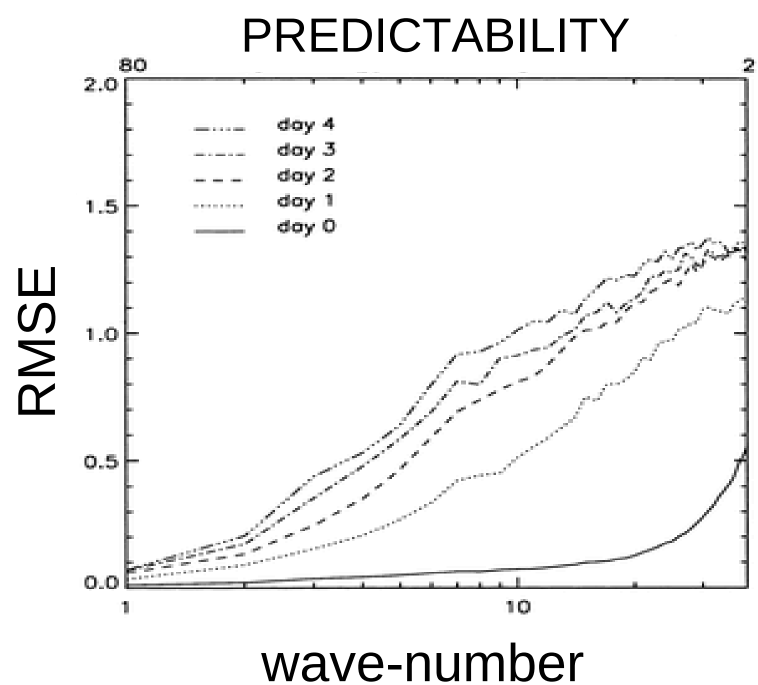
Scale: cosine transforms

Verif: continuous scores

Taylor diagram for precipitation





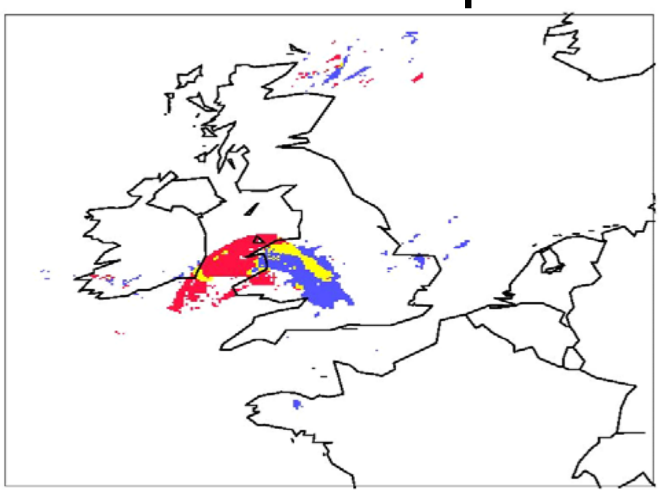


Intensity-scale verification technique
Casati et al. (2004), Met App, vol. 11

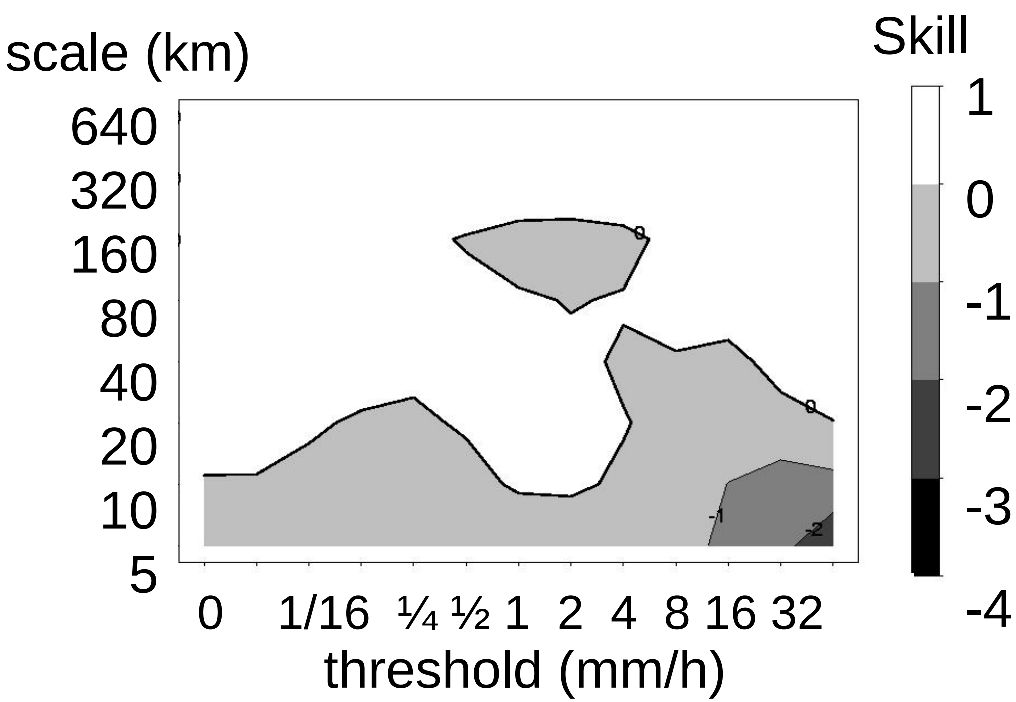
The intensity-scale verification approach measures the skill as function of precipitation intensity and spatial scale of the error

1. Intensity: threshold → Categorical approach
2. Scale: 2D Wavelets decomposition of binary images
3. For each threshold and scale: skill score associated to the MSE of binary images = Heidke Skill Score

Intense storm displaced

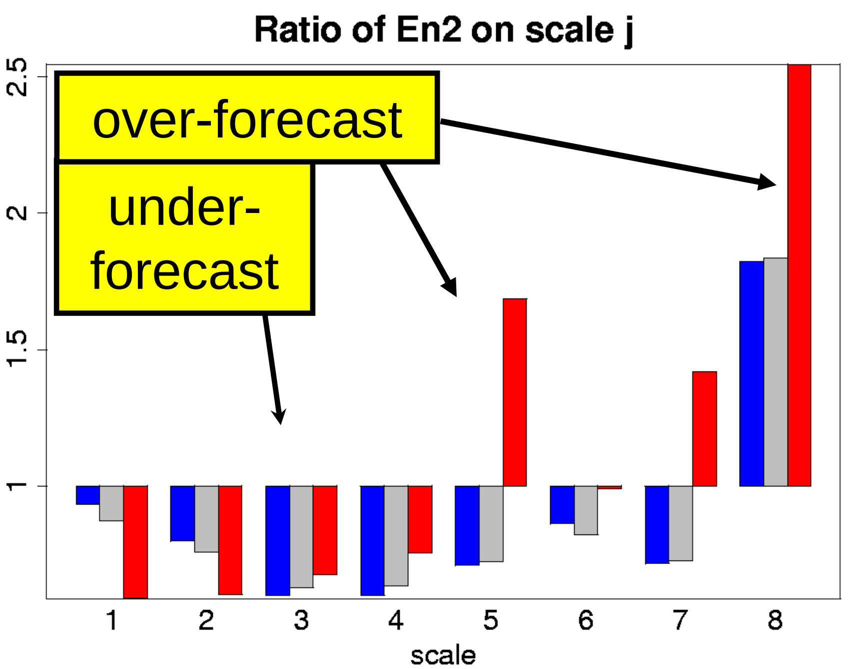


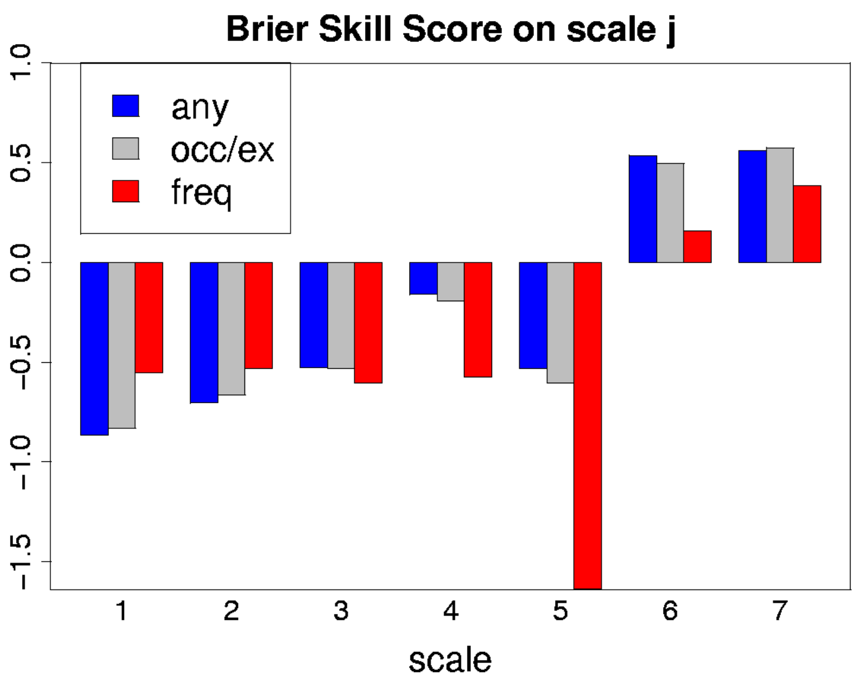
threshold = 1mm/h



Casati and Wilson (MWR accepted)

Scales by wavelets – probabilistic verification





Bias on different scales:

over-forecast of 320 km features for frequent lightning

Skill on different scales:

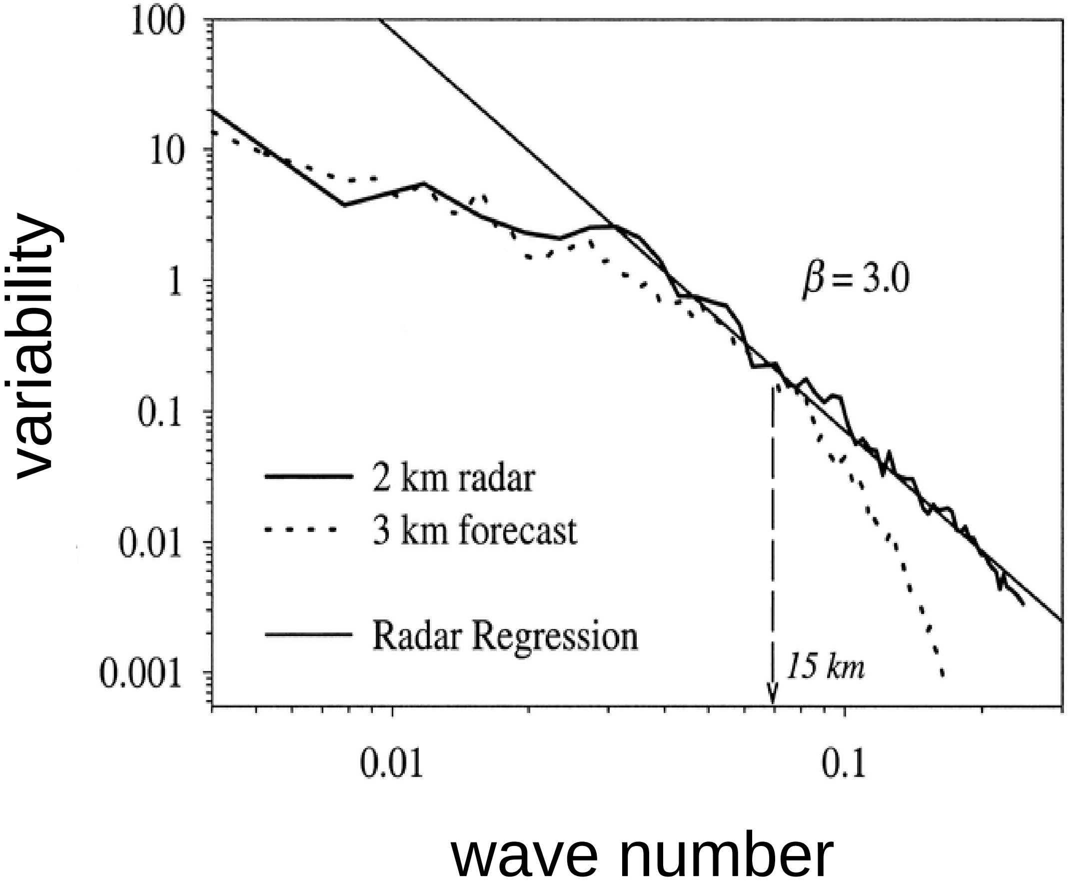
Transition scale ~ 500 km

Very negative skill for 320 km scale features for the frequent lightning

Zepeda-Arce et al. (2000), Harris et al. (2001), Tustison et al. (2003)

Assess ability of reproducing multi-scale spatial structure and space-time dynamics of precipitation fields

Assess scale-invariant parameters related to the scale-to-scale variability and smoothness, feature depth-area-duration and spatio-temporal organization



Neighborhood-based verification

See talk of **E. Ebert: Fuzzy verification**

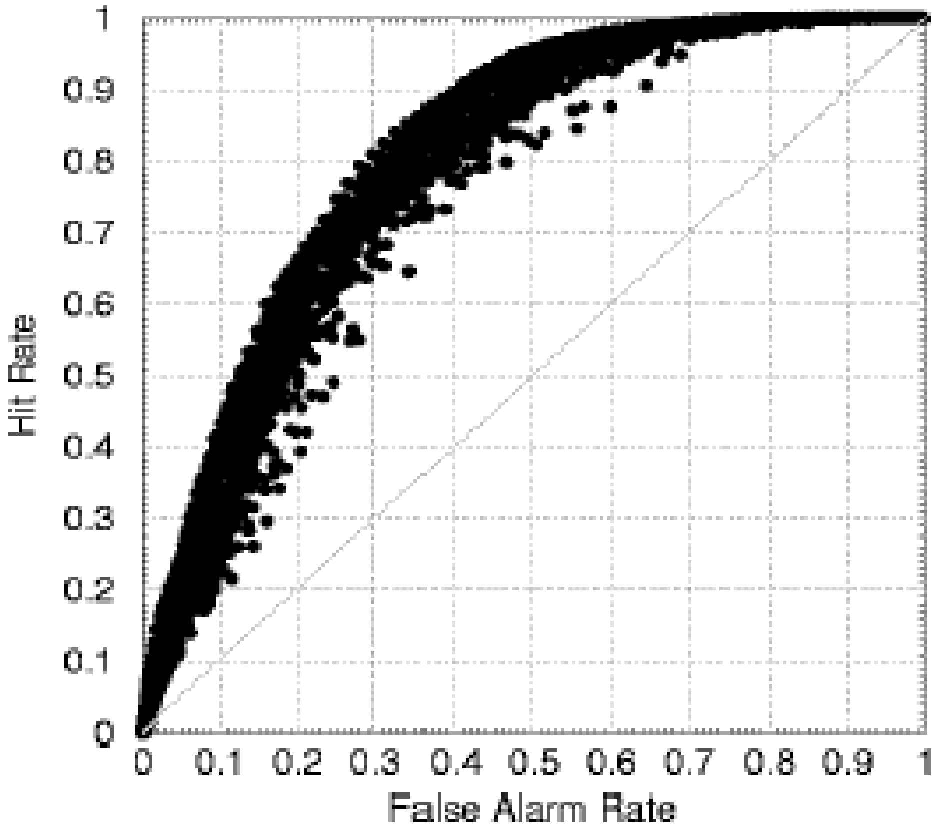
Use neighbor grid-points:

➢ Relax requirements for exact positioning; account of time-space uncertainty; suitable for high resolution models

e.g. Atger (2001) spatial multi-event ROC curve; Rezacova and Sokol (2005), rank RMSE; Tremblay et al. (1996), distance-dependent POD, POFD; Roberts and Lean (2005), Fraction Skill Score;

➢ assess deterministic forecast with probabilistic verification approach

e.g. Theis et al (2005); Marsigli et al (2005, 2006)



Note: scale = neighborhood size (smoothing process → matching requirements more and more relaxed)

Decomposition of forecast error

Hoffmann et al. (1995):

Error = $\boxed{\text{displacement error}}$ + $\boxed{\text{amplitude error}}$ + $\boxed{\text{residual error}}$

displacement error by translating the forecast (e.g. wind field)
amplitude error by applying a scalar geopotential field
until a "best fit criterion" is satisfied (e.g. max correlation)

error measures directly physical quantities
(e.g. displacement in km);
verification easily interpretable (e.g. advection)

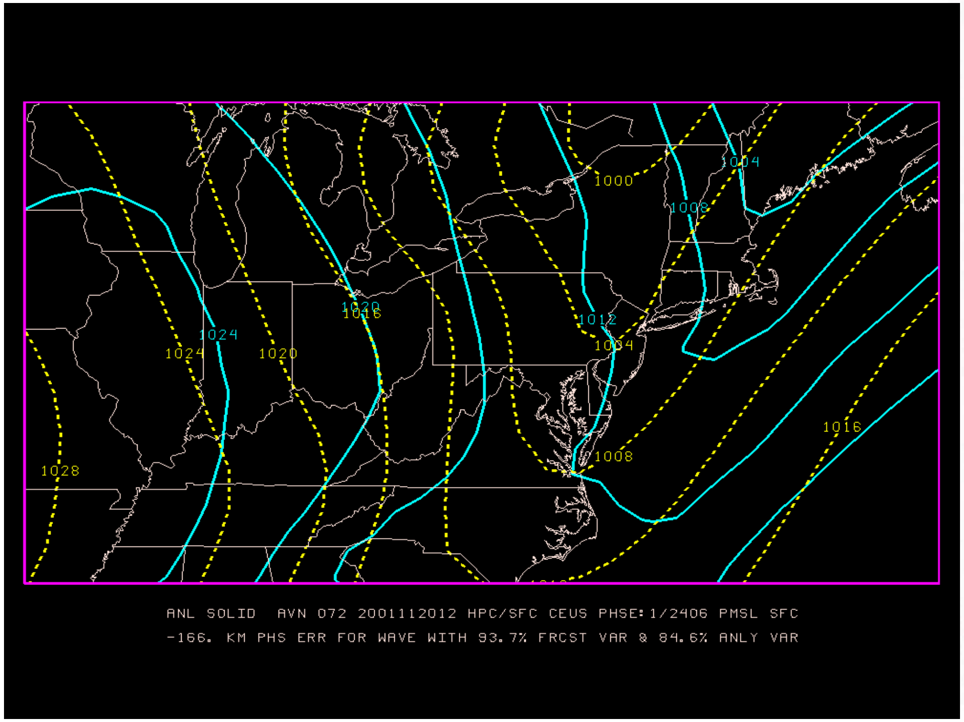
Douglas (2000), Brill (2002), Du et al. (2000), Hoffman and Grassotti (1996), Nehrkorn et al. (2003), Brewster (2003), **Germann and Zawadzki (2002, 2004) I, II and III** Turner, Lee, …

- Error decomposition is performed on different spectral components
- Feedback used in data assimilation/now-casting; **whole field**

Examples

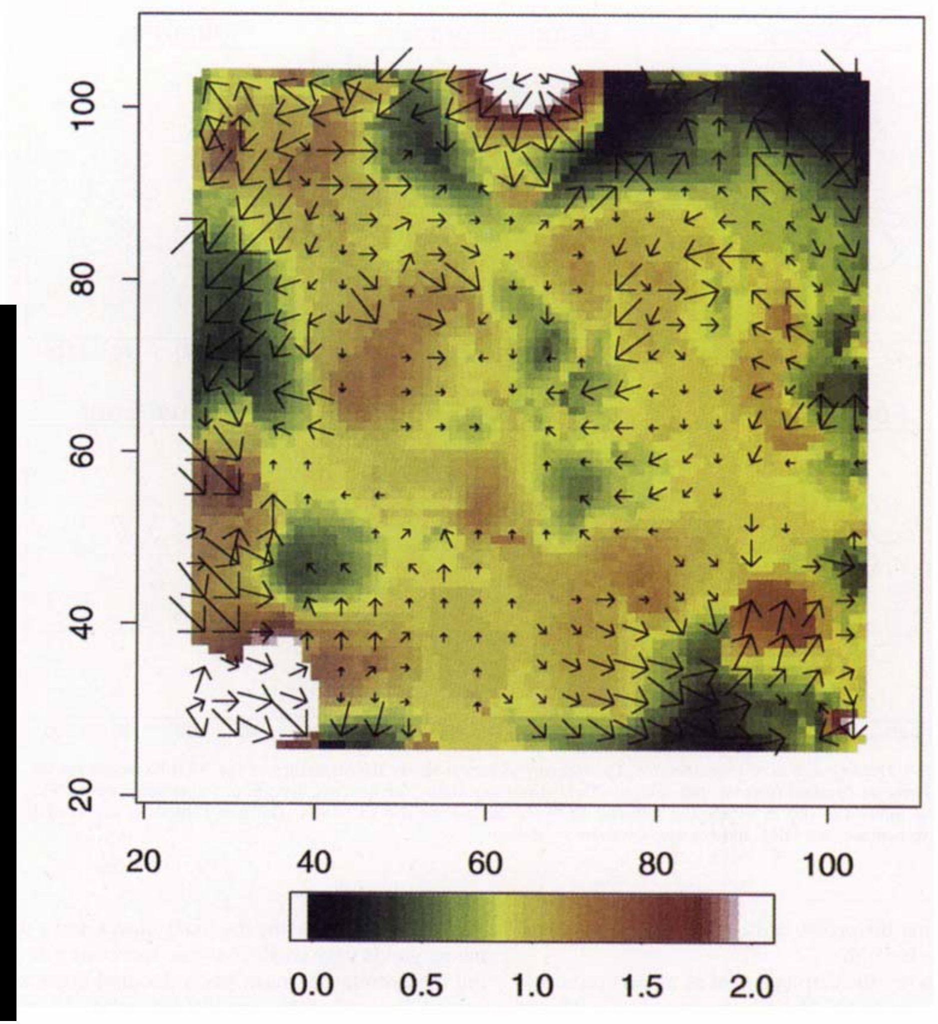
Brill (2002)

Mean sea level pressure
east-west phase error = 166 km



Hoffmann et al (1995)

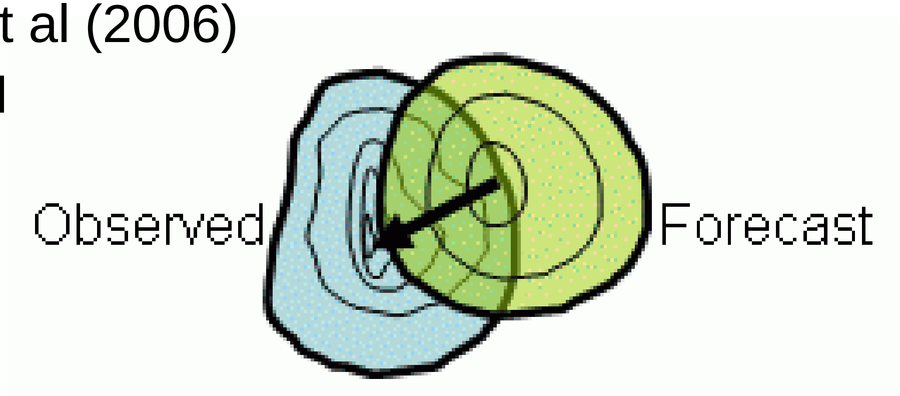
500 hPa GZ: displacement
and amplitude error



Feature-based techniques and decomposition of forecast error

Ebert and McBride (2000), Grams et al (2006)
Davis, Brown, Bullok (2006) I and II
Baldwin et al. (2001)
Nachamkin (2004, 2005)
Marzban and Sandgathe (2006)

Wernli, Paulat, Frei (SAL score)



1. Identify and isolate (precipitation) **features** in forecast and observation fields (thresholding, image processing, composites, cluster analysis)

2. assess displacement and amount error for each pairs of obs and forecast features; identify and verify attributes of object pairs (e.g. intensity, area, centroid location); evaluate feature-distance based contingency tables and categorical scores; verification as function of feature size (scale); classification of mesoscale features; add time dimension → precipitation systems and timing error assessment

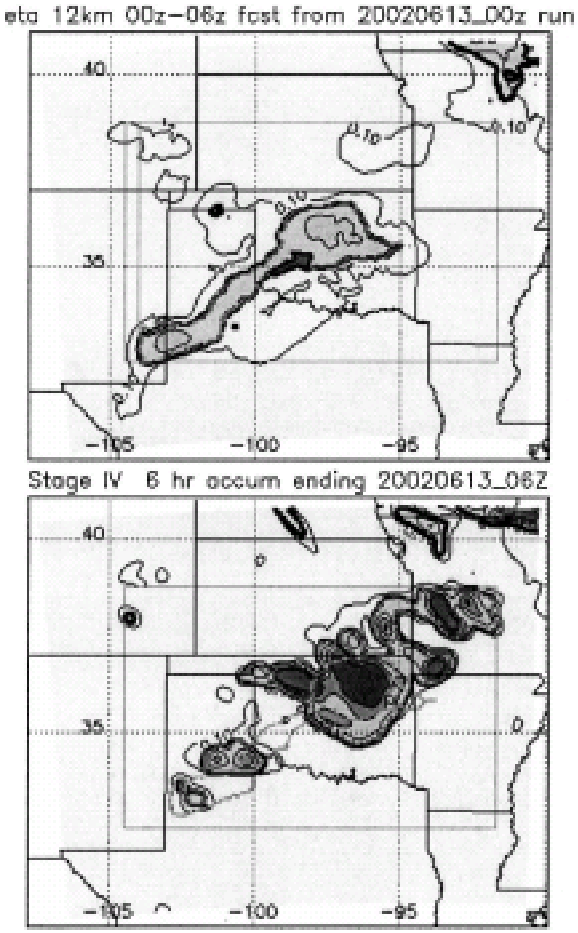
Ebert and McBride (2000), Grams et al (2006)

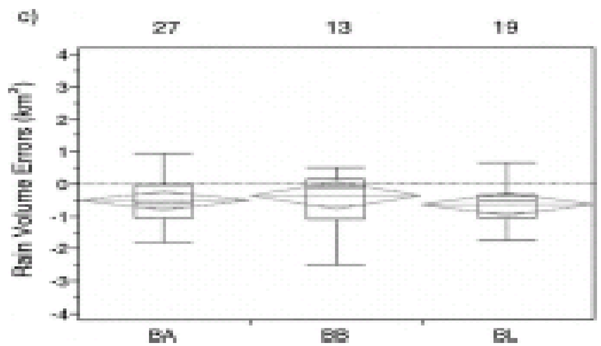
$$MSE = (\bar{x} - \bar{y})^2 + (s_x - s_y)^2 + 2 s_x s_y (1 - r_{trans}) + 2 s_x s_y (r_{trans} - r)$$

bias | pattern | position

CRA

CRA classification





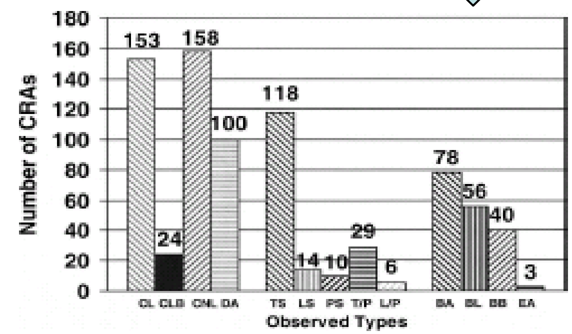
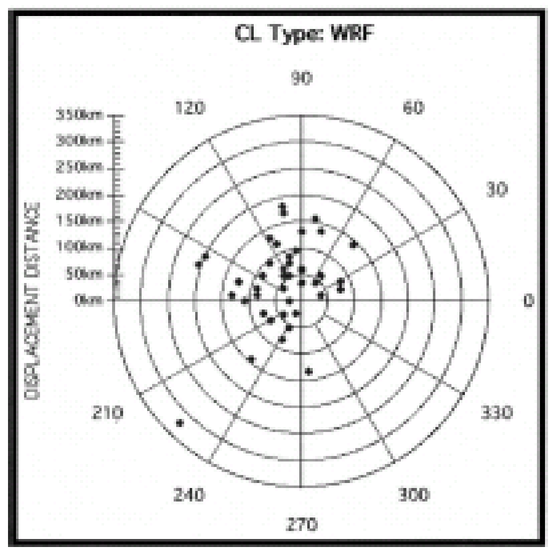


FIG. 6. Histogram of observed systems (general types, stratiform types for linear systems, and development types for linear systems) for all CRAs identified by the EMT applied to the Eta, MM5, and WRF models. Tables 2, 3, and 4 give definitions used for the abbreviation of types.



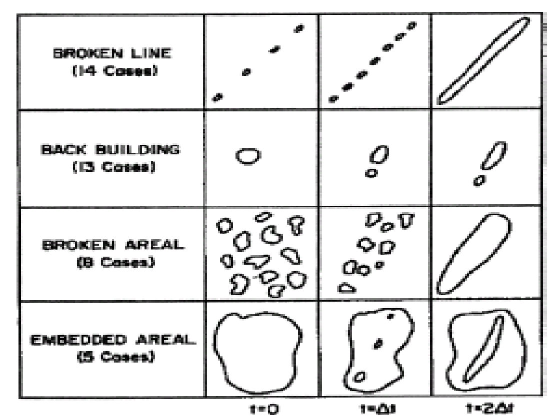
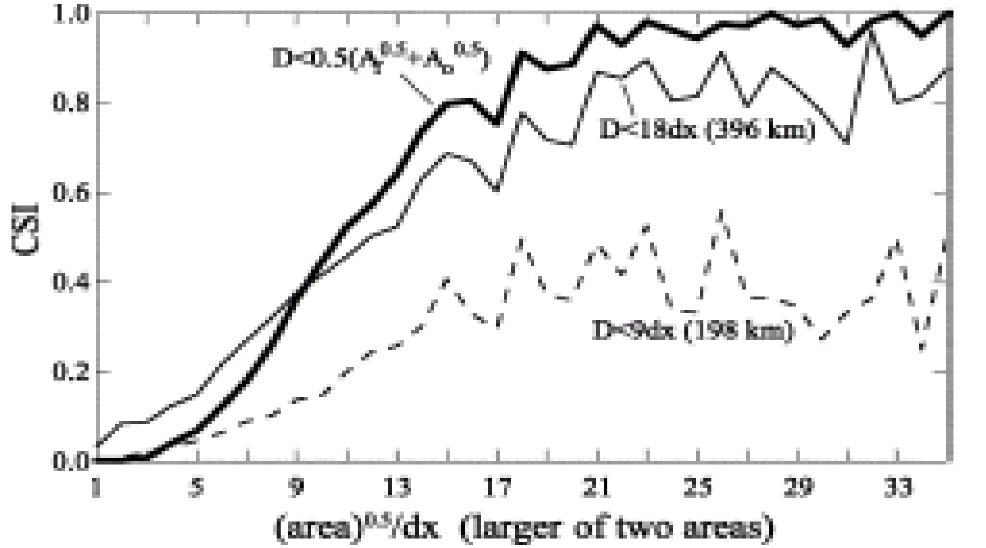


FIG. 4. Development types for linear MCSs; from Bluestein and Jain (1985).

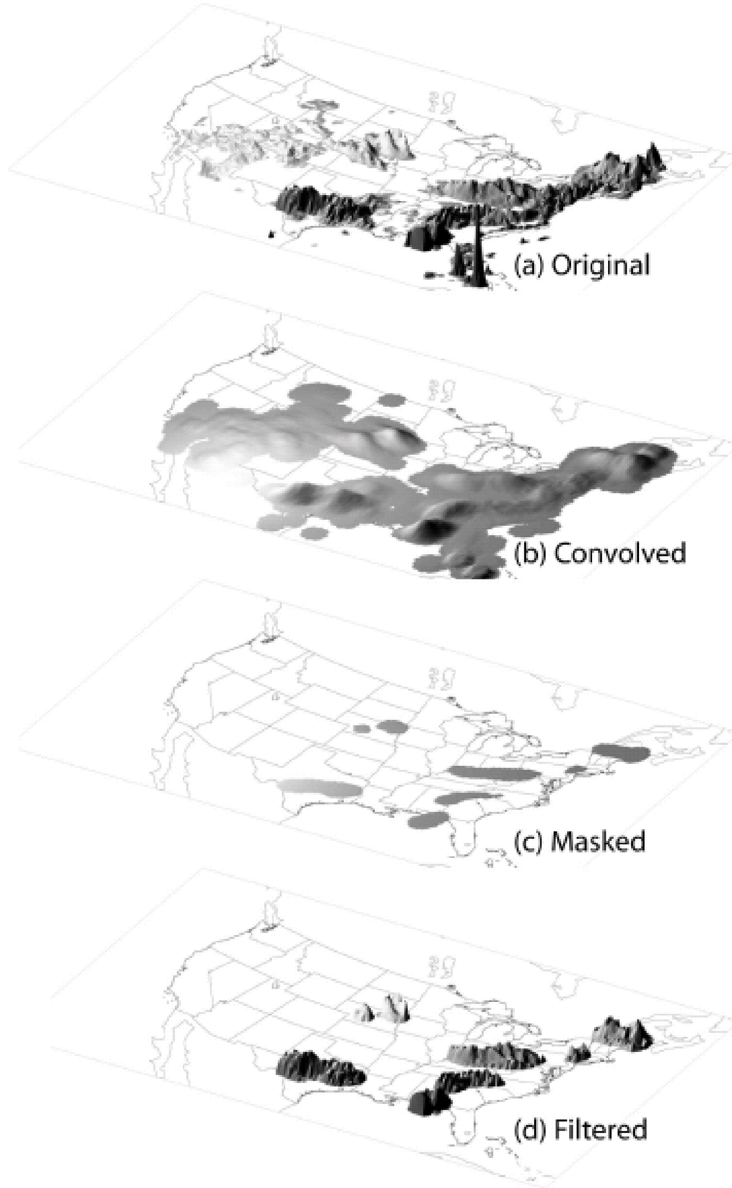
Davis, Brown, Bullok (2006)

Objects categ verif as function of scale



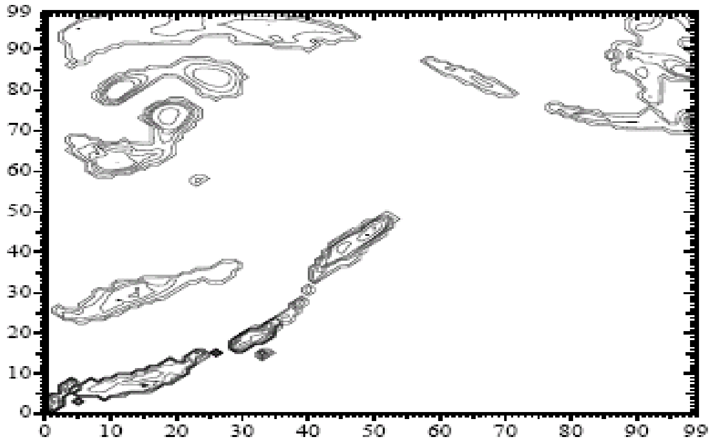
Rain systems

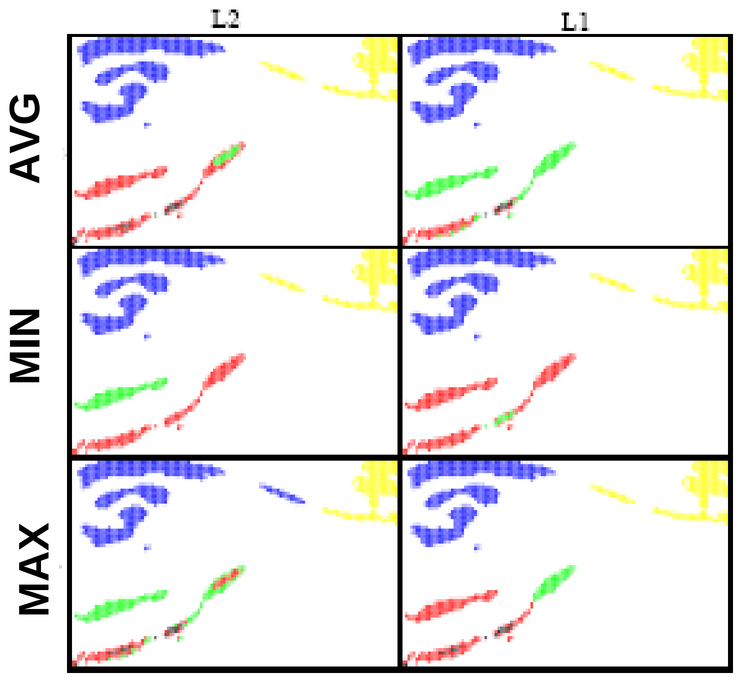
Region	CSI	No.	Timing	*dx*	*dy*	Area	I25	I75
West	0.38	80	1.2	−3.9	9.4	−9	0.8	4.0
Central	0.51	175	1.2	12.4	−2.5	266	0.7	5.2
East	0.51	191	0.8	10.4	19.3	55	−0.3	3.9
Total	0.48	446	1.0	8.8	8.9	127	0.3	4.4

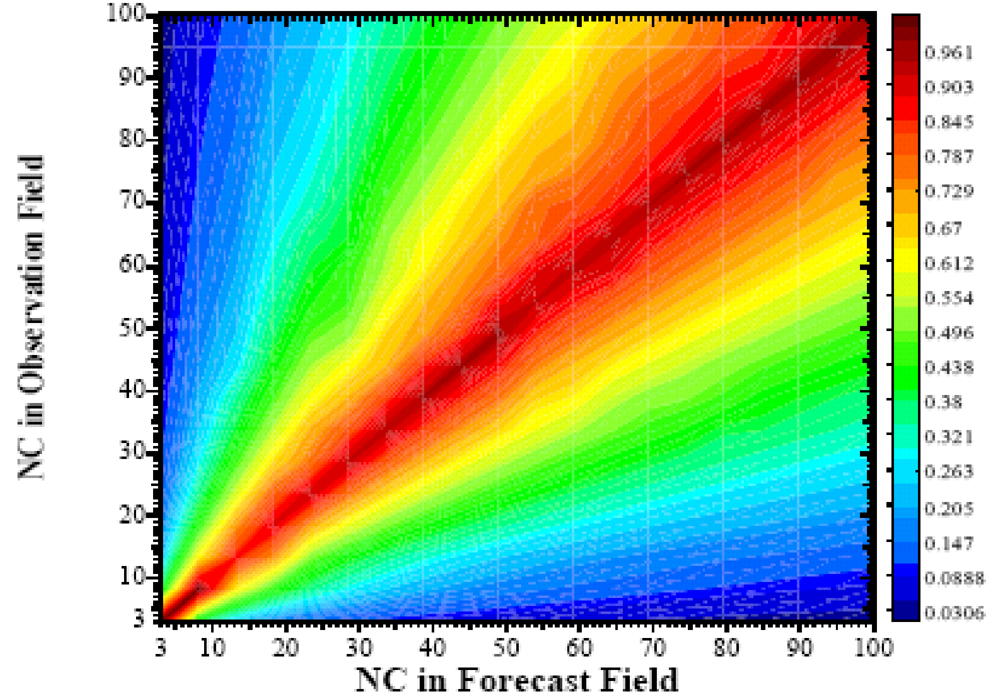


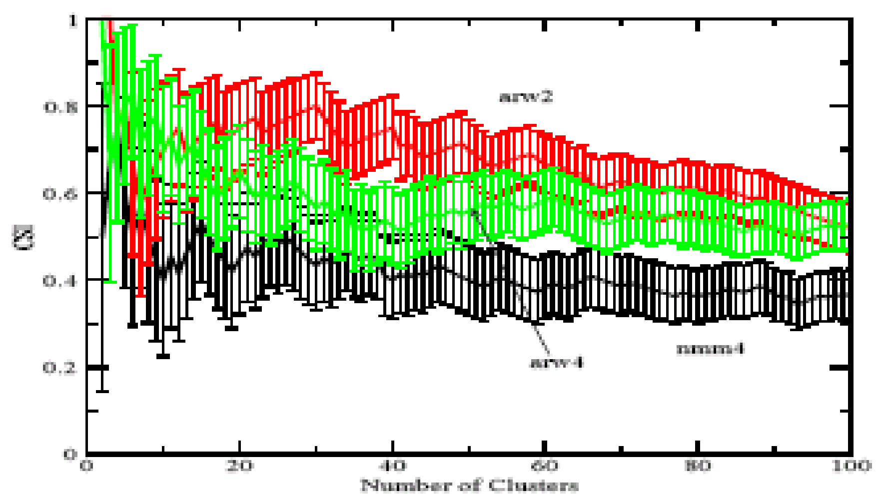
. 2. Example of application of object-identification approach to a particular WRF precipitation ast grid: (a) original precipitation grid, with intensity presented as the vertical dimension; (b) lved grid, after the smoothing operation has been applied; (c) masked grid, following appli- of the intensity threshold; and (d) filtered grid, showing the precipitation intensities inside the ified objects. The grid covers the entire United States.

Marzban and Sandgathe (2006), cluster analysis









Nachamkin (2004)

mistral composite:

collect events from multiple occasions

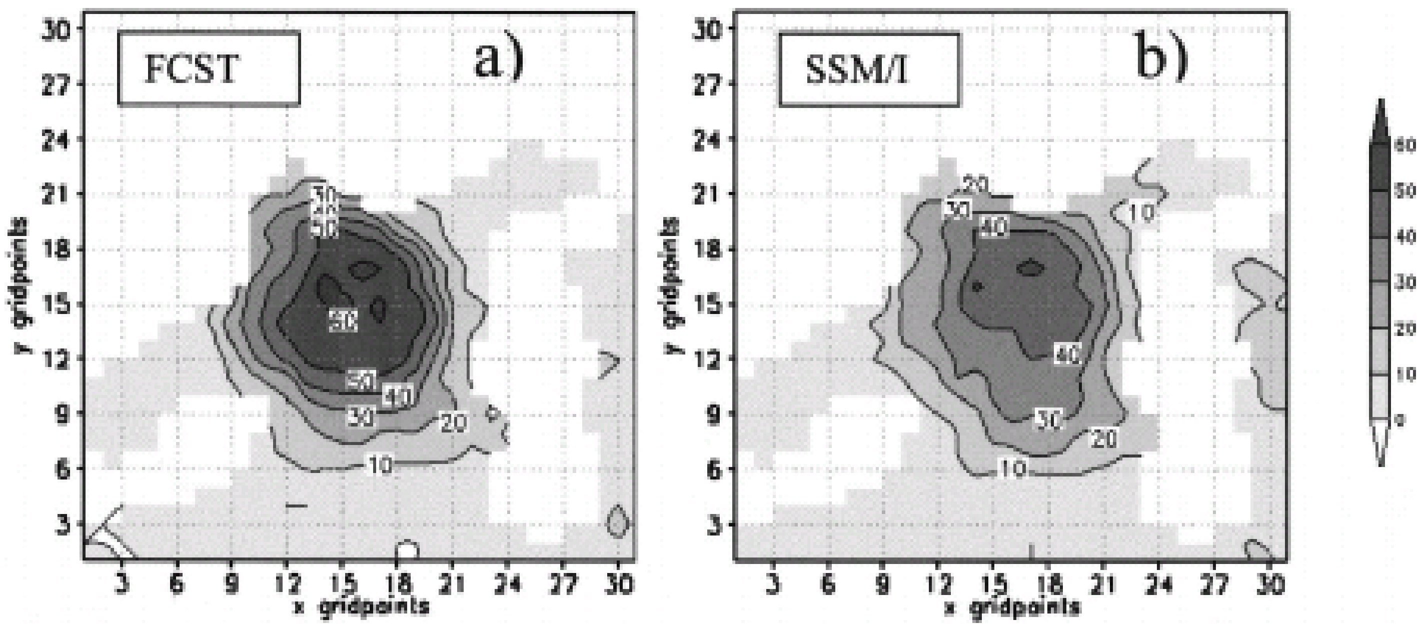
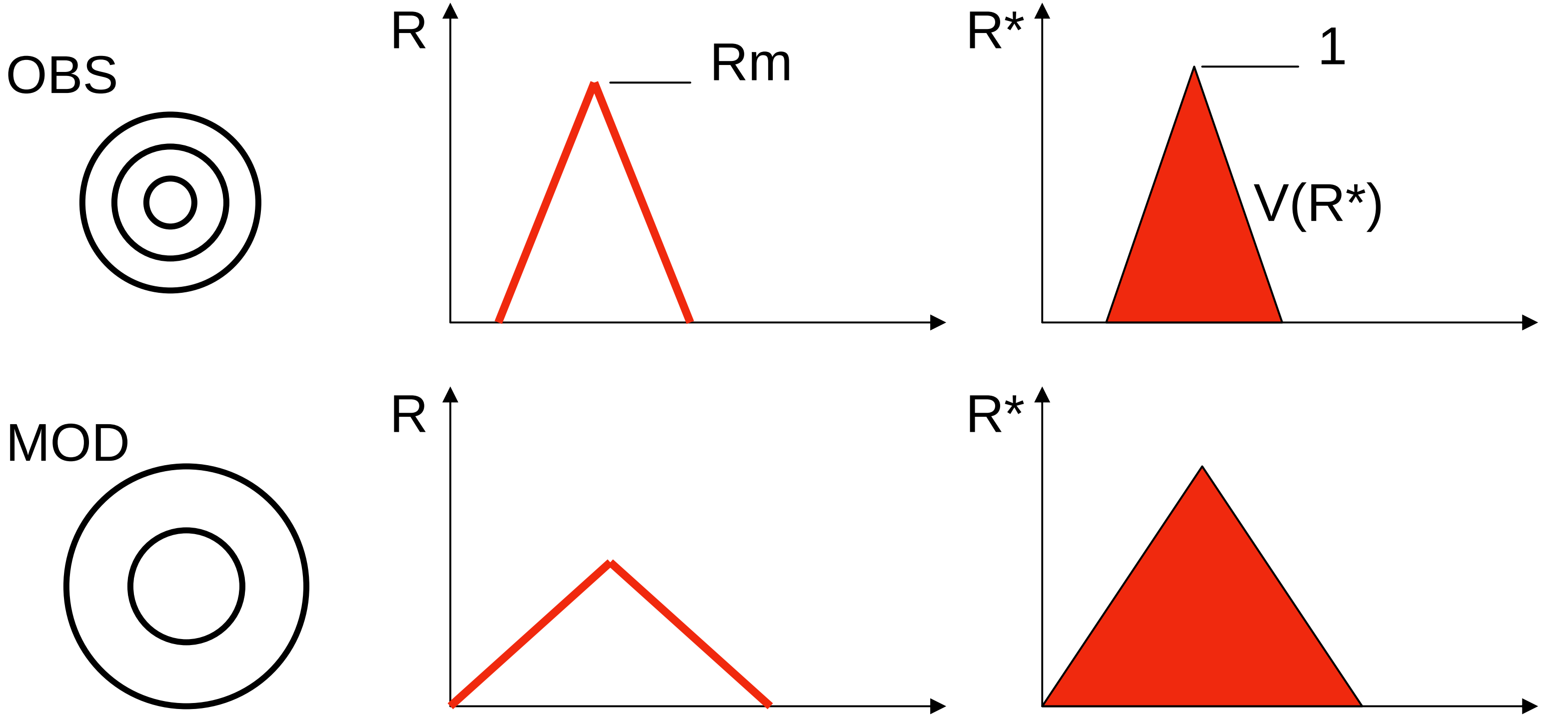


Fig. 4. Number distributions of (a) predicted and (b) observed mistral occurrence on the 31 × 31 point relative grid. Each labeled interval represents three grid points or 51 km. The samples were derived from the 18-h forecasts and were conditional on the occurrence of a predicted mistral. Grid points with less than 20 SSM/I samples are not plotted.

Structure – Amplitude –Location (SAL)

Wernli, Paulat, Frei



Idealized small-intense obs and large-weak model:

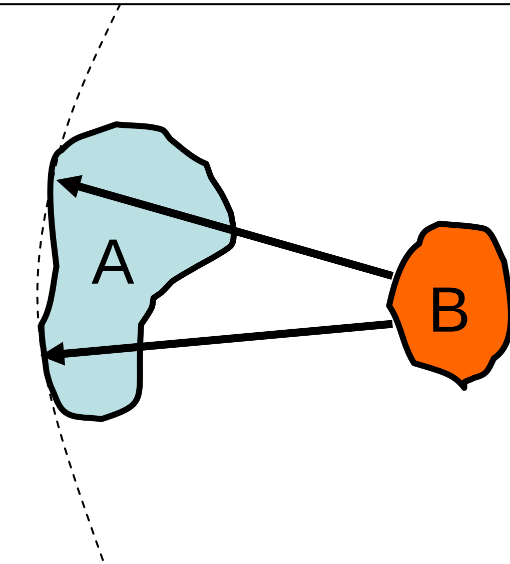
A = 0 (area mean precip error), S > 0 structure is different

Distance measures for binary images

1. Average distance
2. Housdorff metric
3. Baddeley metric
4. Pratts' figure of merit
5. ...

➢ Account for object shape, distance, …

➢ Binary images → alternative to use along with traditional categorical scores



$$H(A,B) = \max\left\{ \sup_{a\in A}\left(d\left(a,B\right)\right), \sup_{b\in B}\left(d\left(b,A\right)\right) \right\}$$

Venugopal et al. (2005); Gilleland et al.(2006)

Summary

1. **Motivation**: coherent spatial structure and features
2. **Scale verification**: features of different scales, assess different physical processes and model parametrizations (scale structure, predictability)
3. **Neighborhood-based (fuzzy) verification**: relax time-space matching requirements; probabilistic approaches
4. **Error decomposition: displacement + amount**
5. **Feature-based approaches: error measured by physical quantities**
6. Distance metrics for binary images

Spatial verification techniques need observations over spatial domains

1. Spatial observations: satellites, radars, …
2. Point observations: radiosondes, gauges, …
3. Analysis ← background model (can be incestuous)
4. Block kriging, Cressman analysis, Barnes analysis, …

Canadian precipitation analysis relies heavily on forecast background model; radar measurements suffer still of several uncertainty in QPE; radar network covers only southern Canada; satellite and radar are not (yet) assimilated for precipitation. What remains ? **GAUGES** …

Environment Canada Environnement Canada

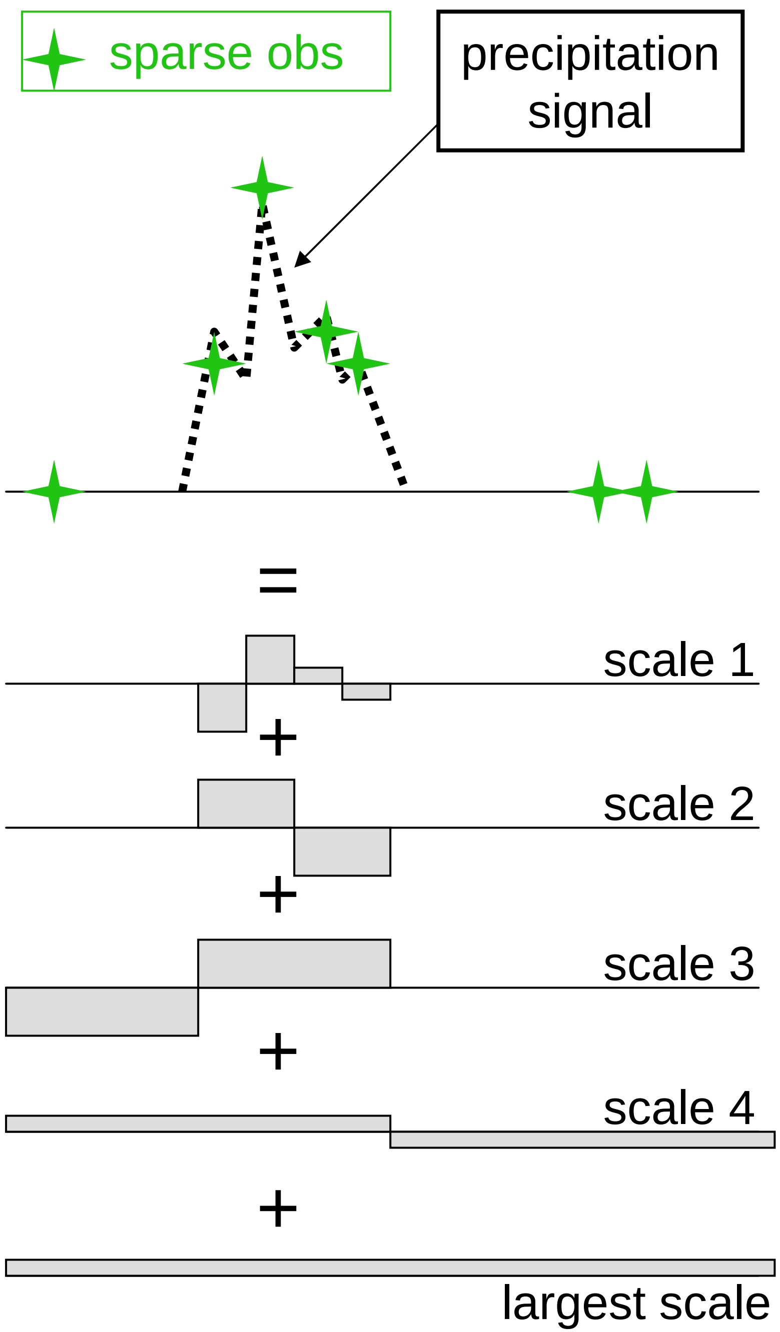
Canada

Reconstruction of a precipitation field from sparse gauges obs by using 2D Haar wavelets

Background idea: any real function can be expressed as linear combination of wavelets (i.e. sum of components on different scales)

1. Compute wavelet coefficients from sparse gauge obs
2. Reconstruct field as sum of components on different scales

NOTE: no gauges = missing obs, no dense gauge network = no information on small scales, large scales only !

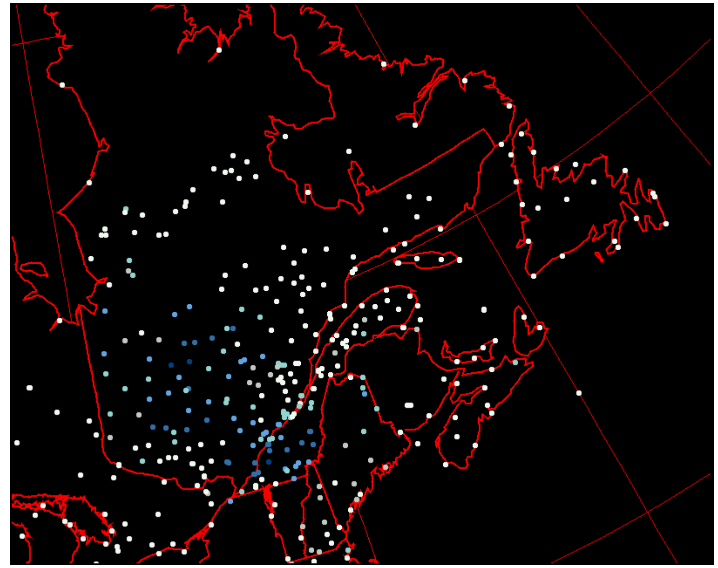


Example: 6h acc (mm)

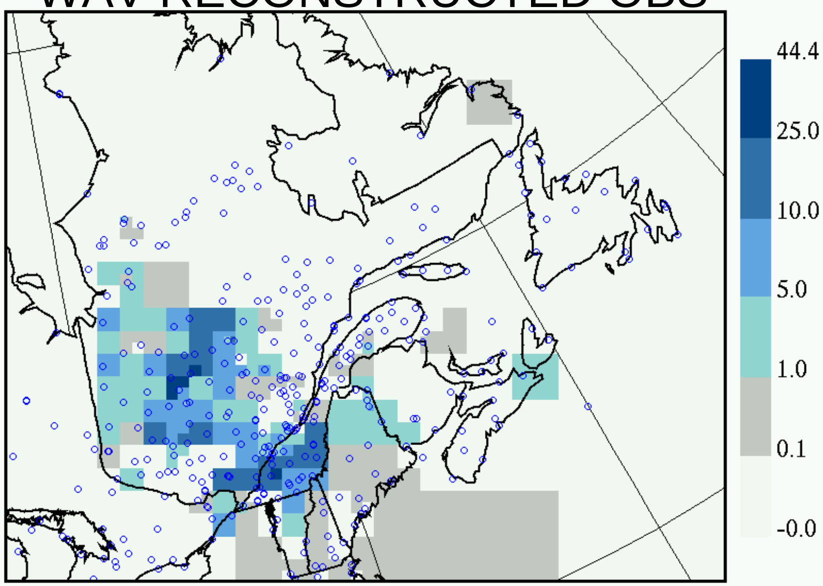
27th Aug 2003, 6:00 UTC

1. Account for existence spatial structures on different scales
2. Account for gauge network density
3. Value at station location = to gauge value

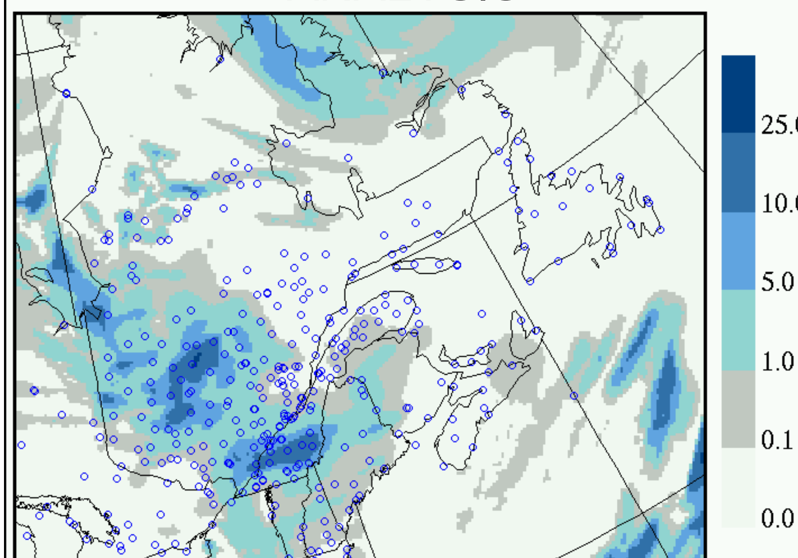
GAUGES OBSERVATIONS

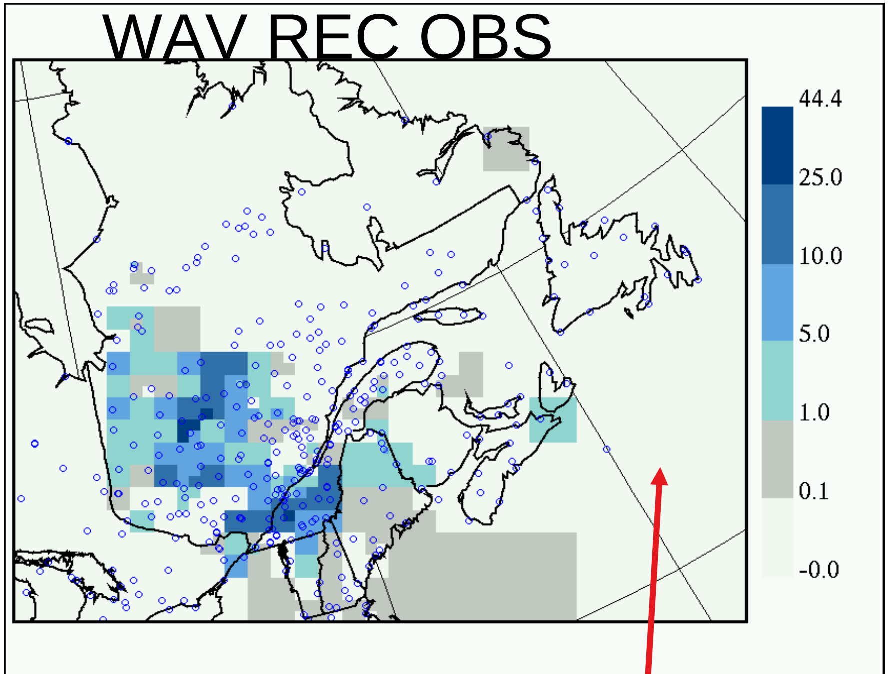


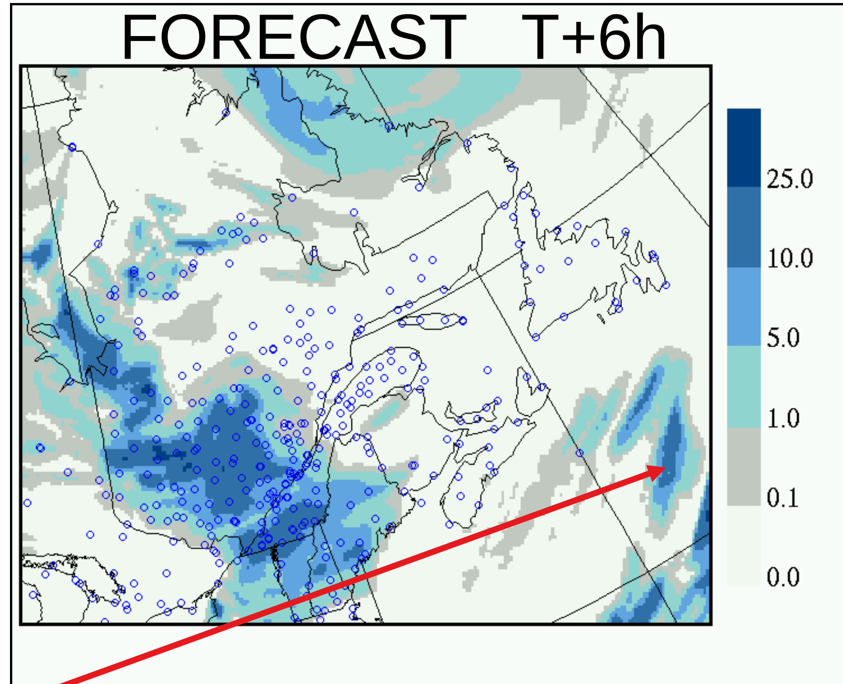
WAV RECONSTRUCTED OBS



ANALYSIS

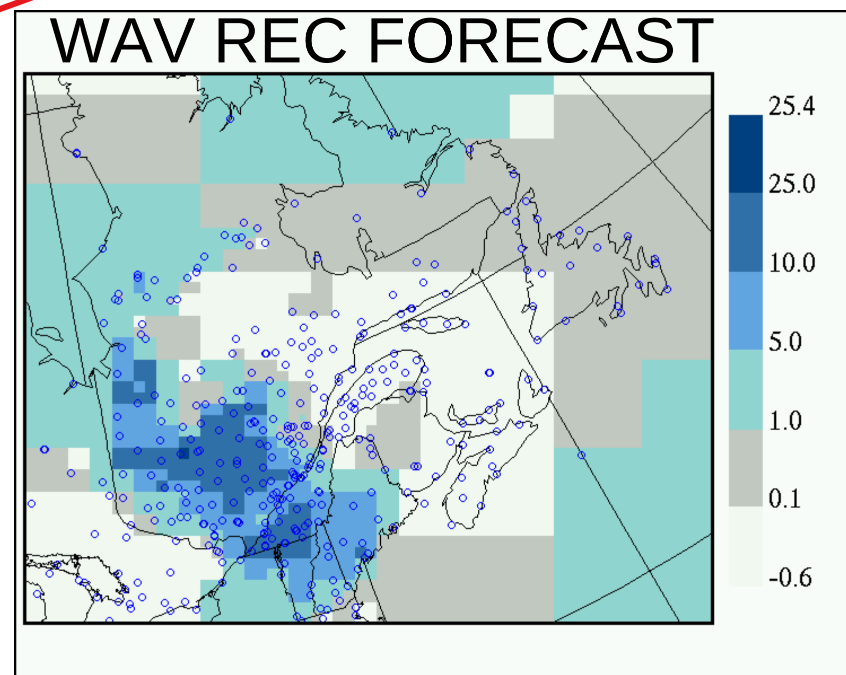


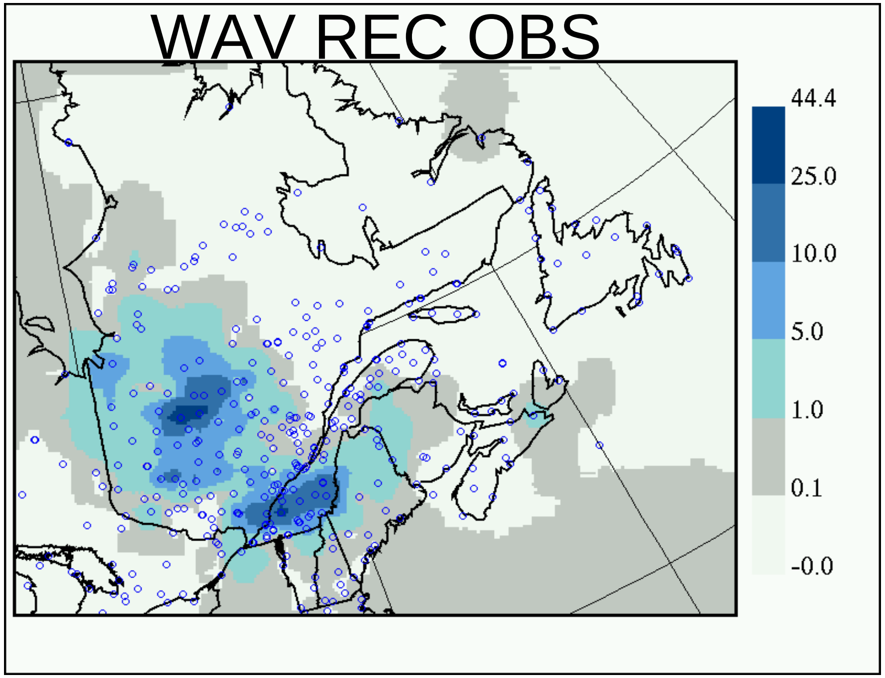


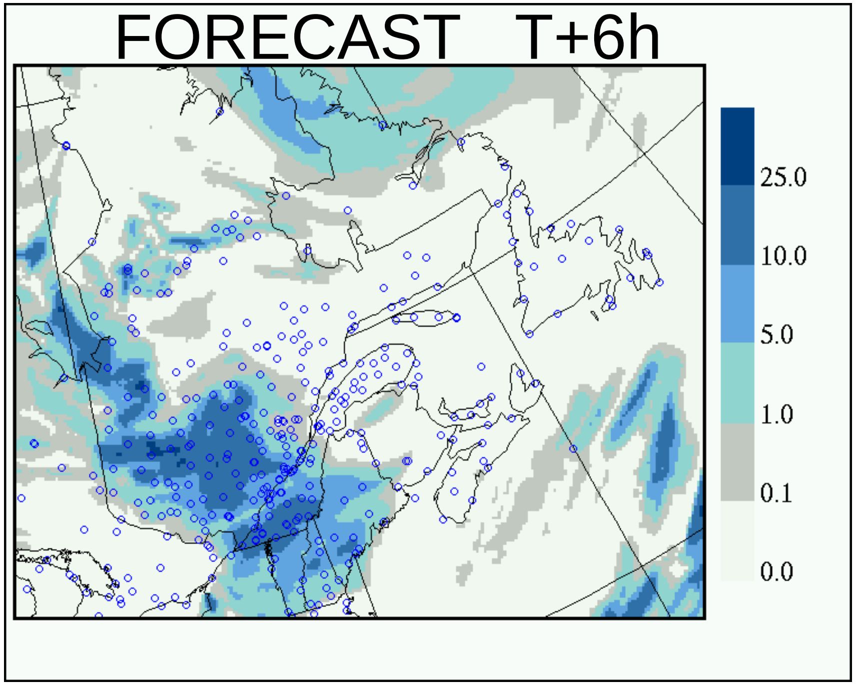


No gauges = missing obs,
but forecast has features!

1. Decompose forecast with wavelets
2. Set to NA wavelet coefficients where no obs
3. Reconstruct forecast field



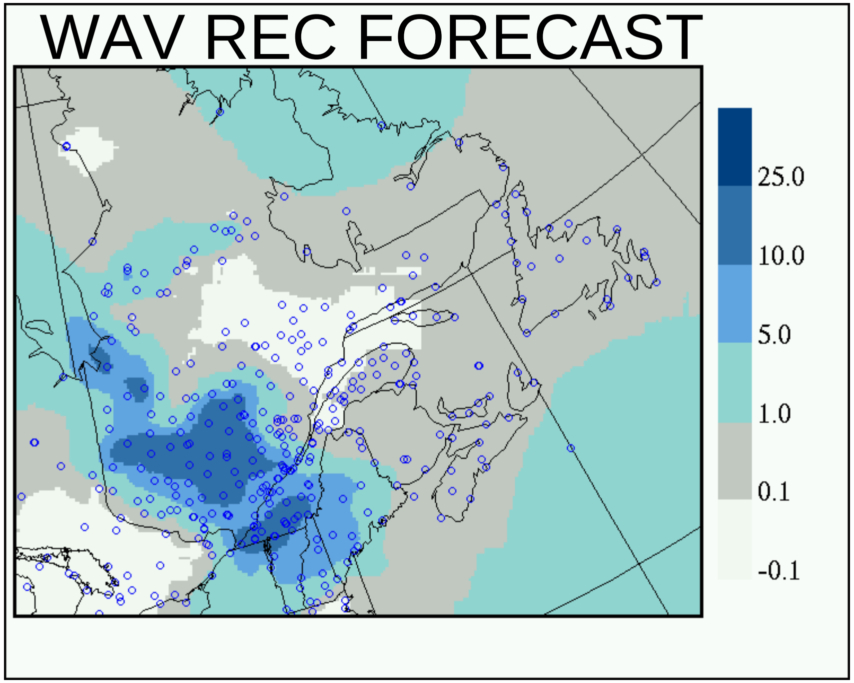


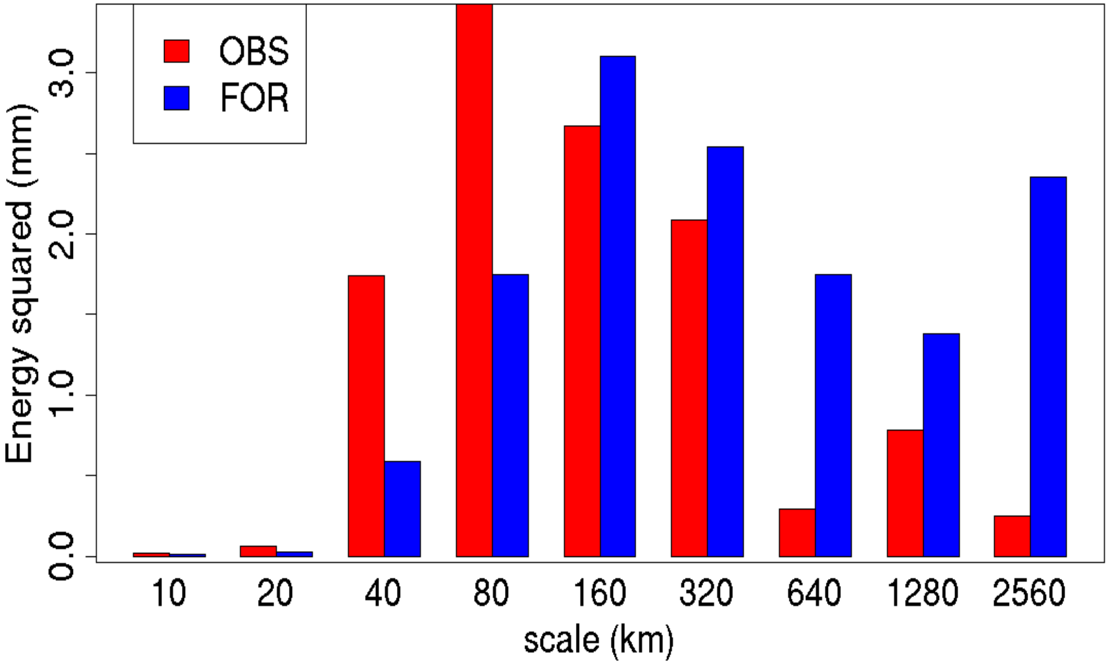


Discrete wavelets = squared areas with fix location; these are not always representative

Eliminate discrete effect by **dithering** the wavelet support and averaging (100 random)

→ Continuous wavelets





Verification

on different scales, but only **where obs are available**

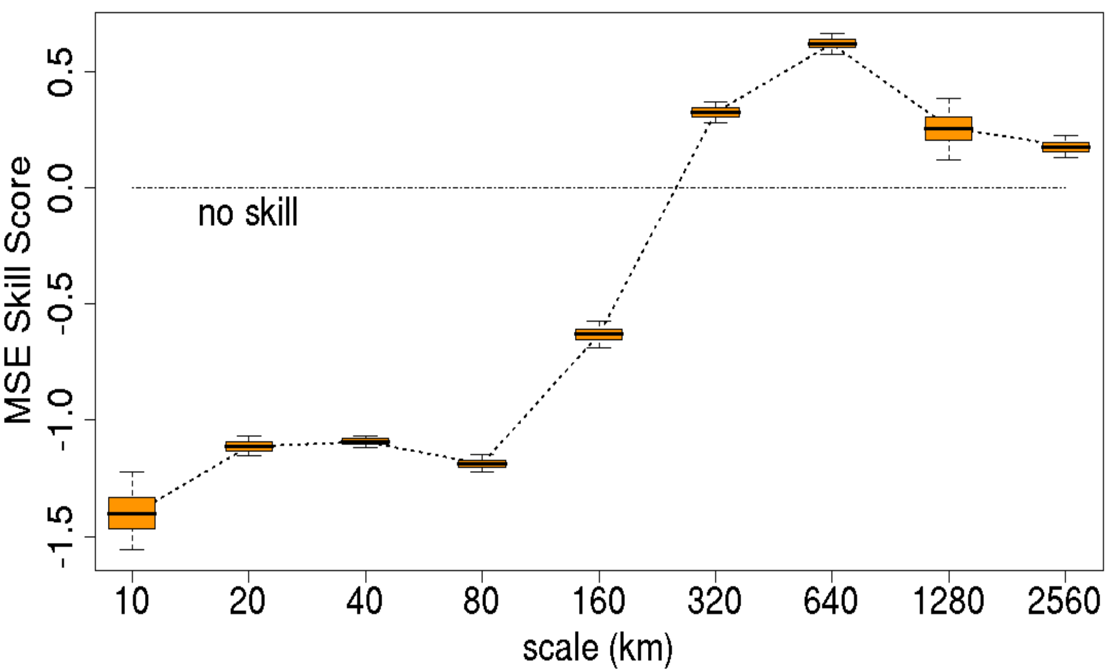
1. Energy squared:

$$En^2(X) = \overline{X^2}$$

Measures the quantity of events and their intensity at each scale => BIAS, scale structure

2. MSE Skill Score:

$$1 - \frac{2MSE(Y,X)}{En^2(X) + En^2(Y)}$$



Summary

Wavelet-based approach to reconstruct a precipitation **field from** sparse **gauge** observations:

- Account of existence of features and field coherent spatial structure + scales
- Account of gauge network density
- Preserve gauge precip. value at its location

Verification on different scales/resolution, but only **where obs are available**

Future work: uncertainty mask

Acknowledgments:

E.Ebert, B.Brown, L.Wilson, C.Marzban, V.Fortin – help, inputs

WMO – support

Thank you!

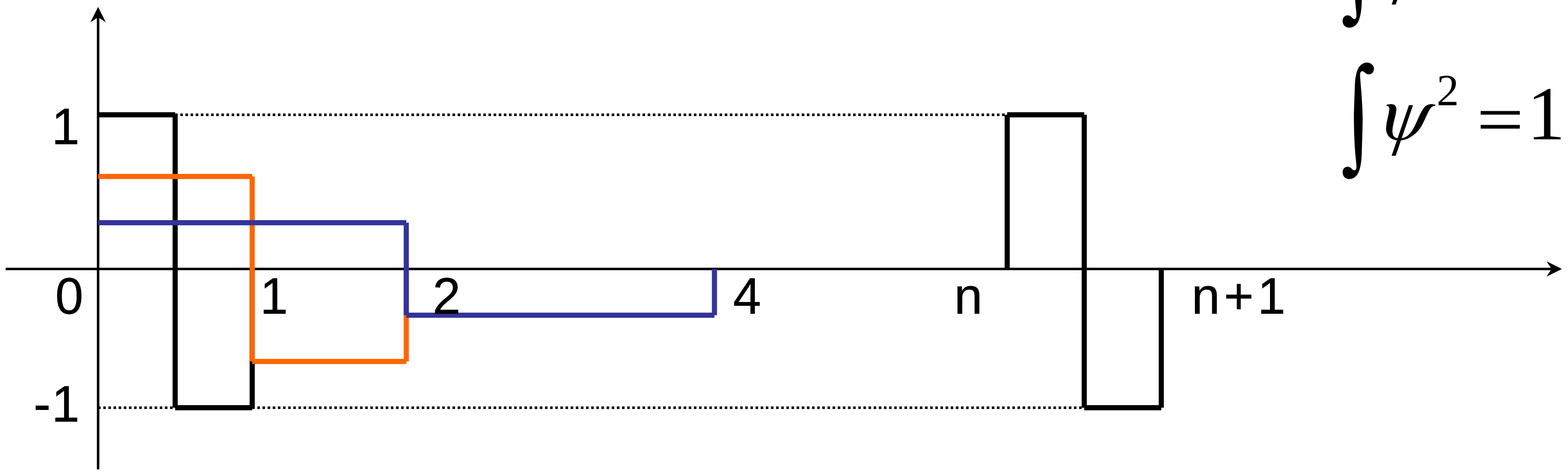
barbara.casati@ec.gc.ca

Wavelets

Haar mother wavelet ψ

$$\int \psi = 0$$

$$\int \psi^2 = 1$$

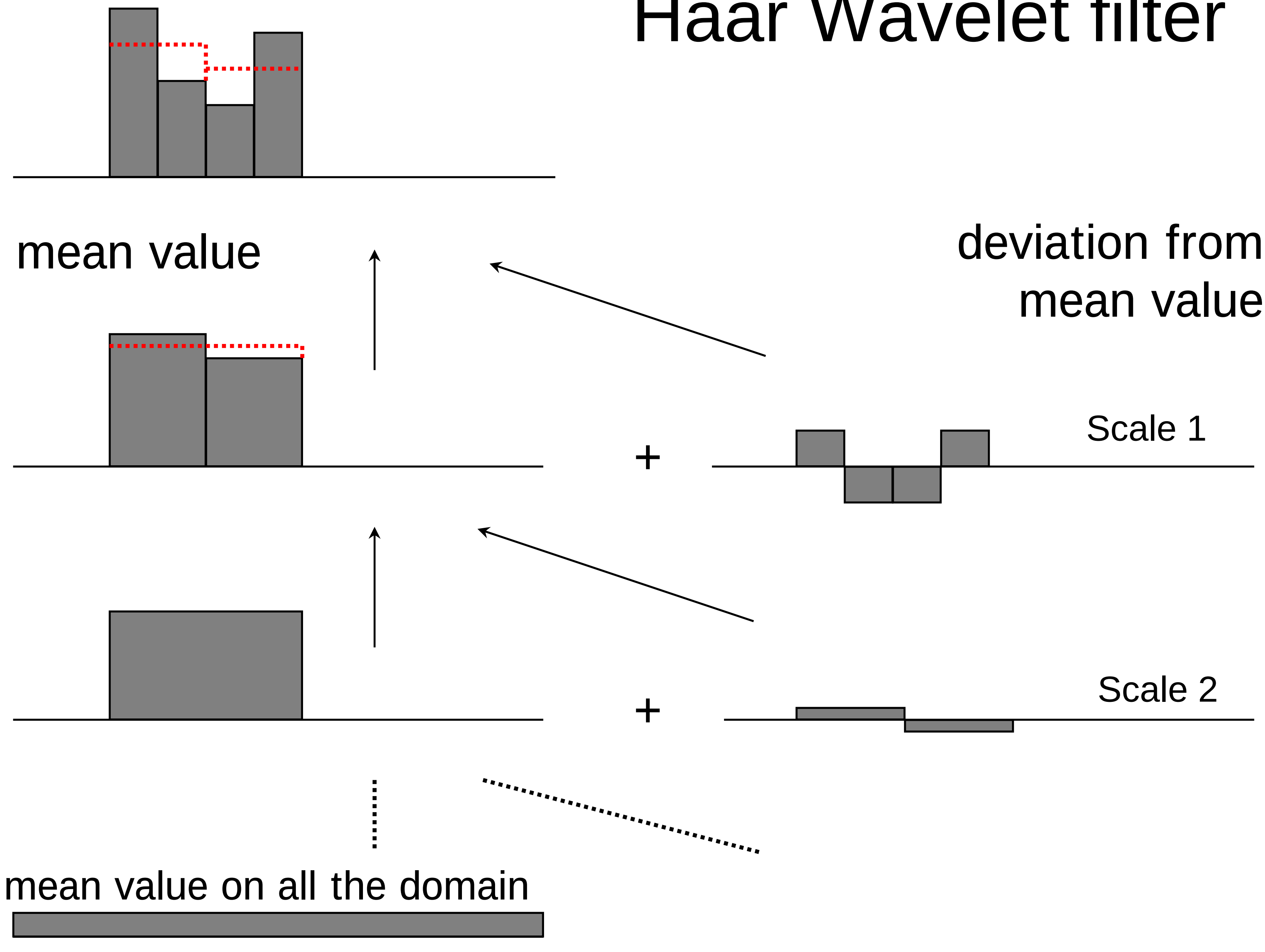


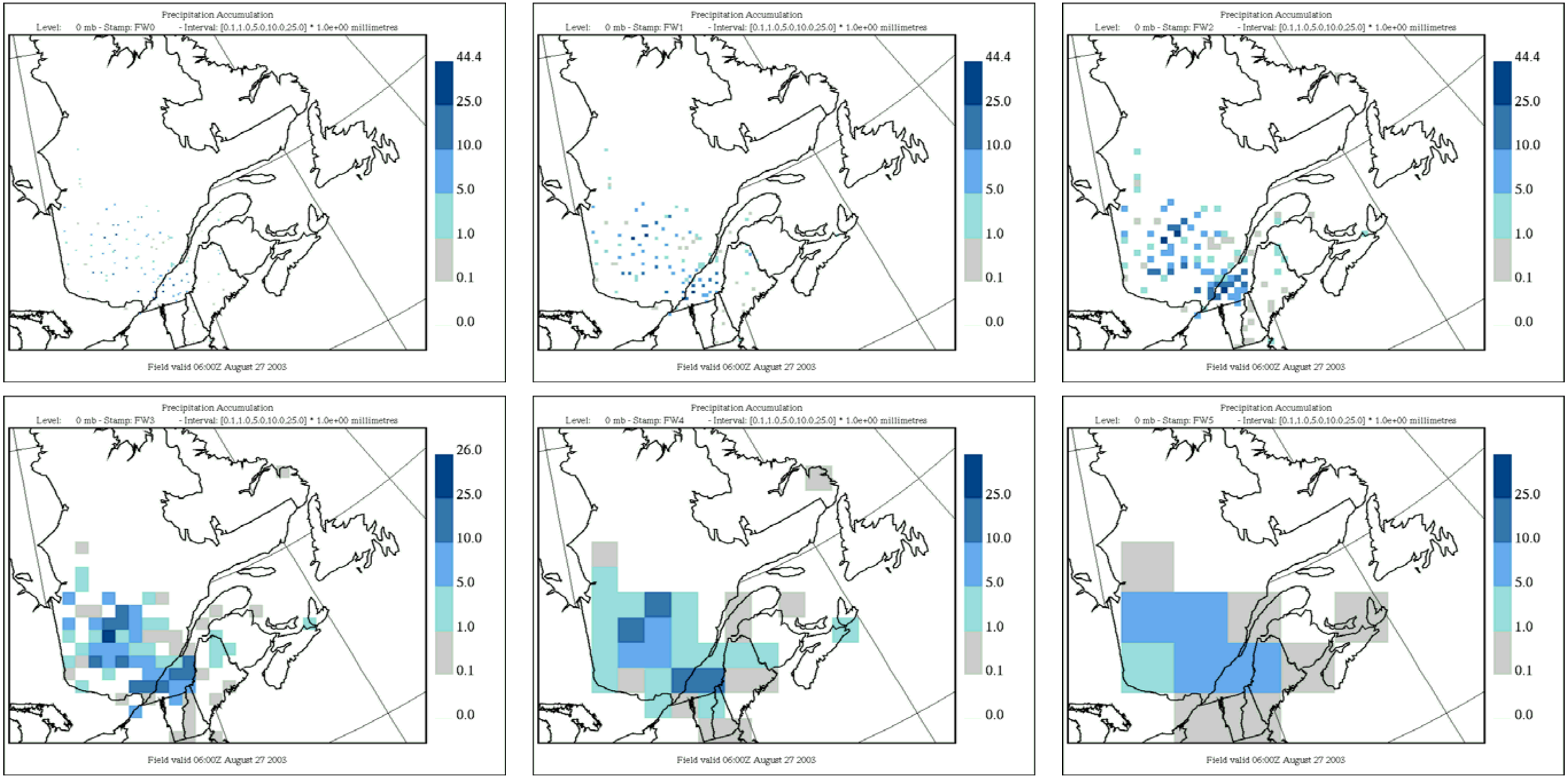
➢ Wavelets are locally defined real functions characterised by a **location** and a **spatial scale**.

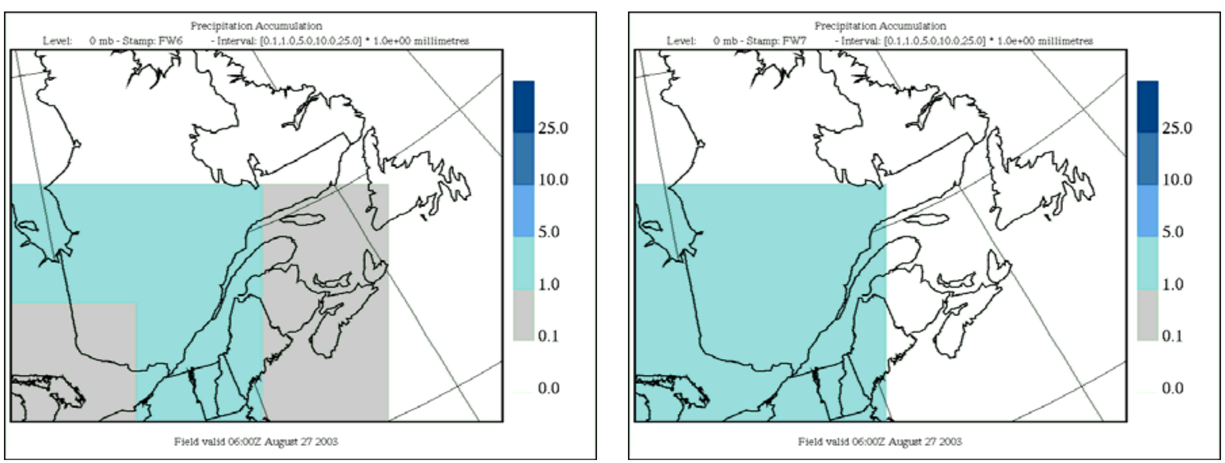
➢ Wavelets are a basis: Any real function can be expressed as a linear combination of wavelets, i.e. as a sum of components with different spatial scales.

➢ Wavelet are local => deal better than Fourier with discontinuous, on/off fields with features (e.g. precipitation)

Haar Wavelet filter

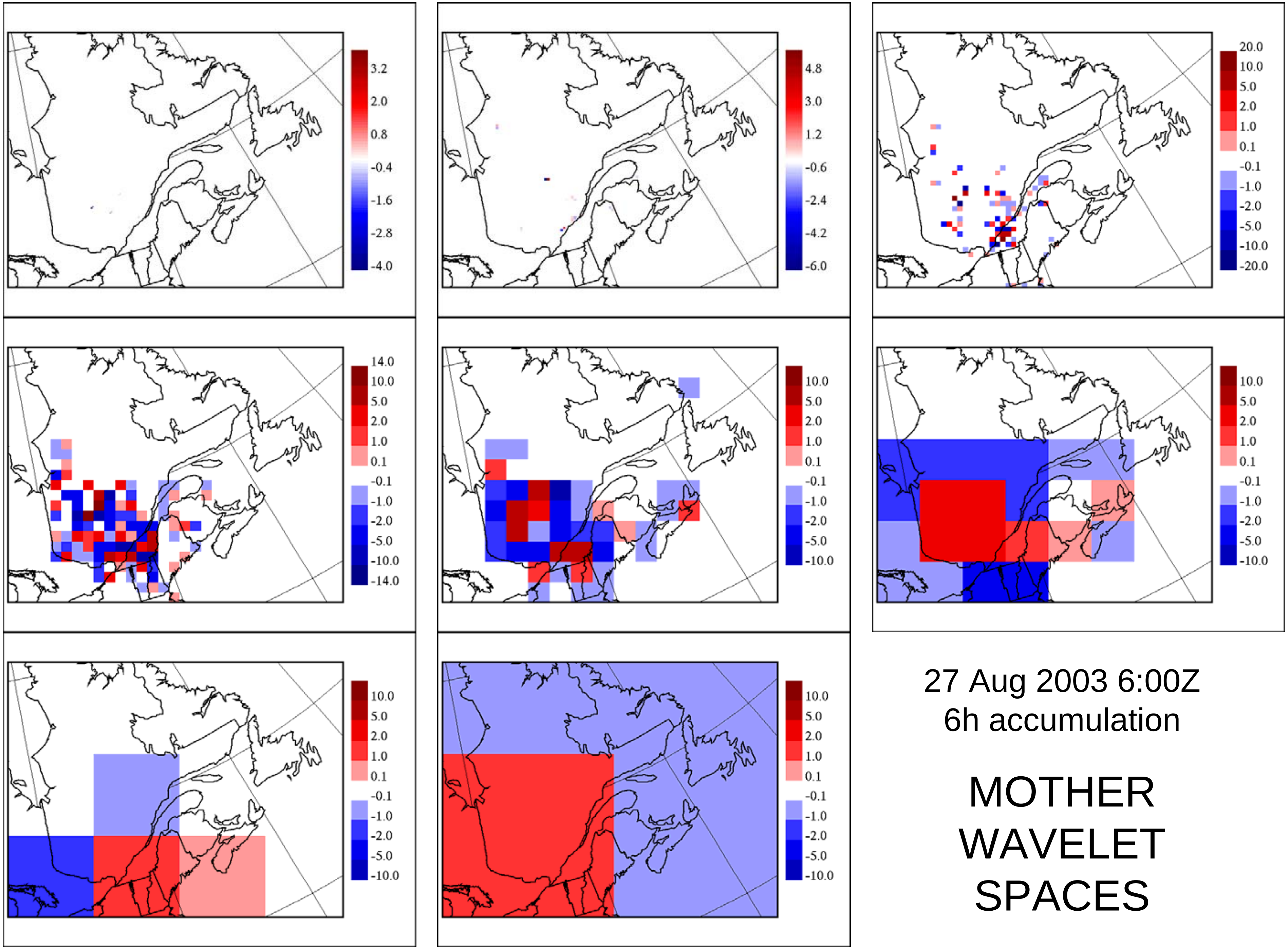






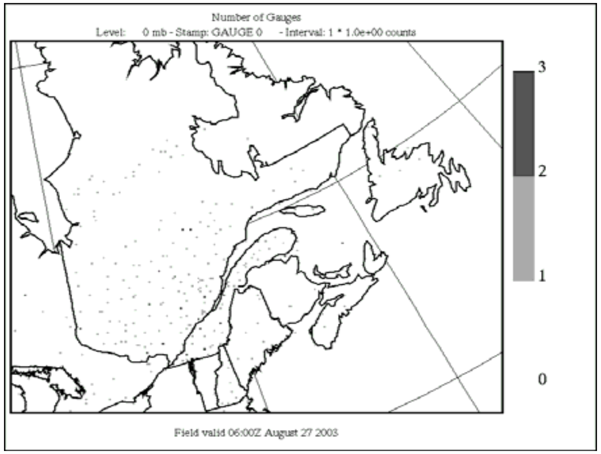
27 Aug 2003 6:00Z
6h accumulation

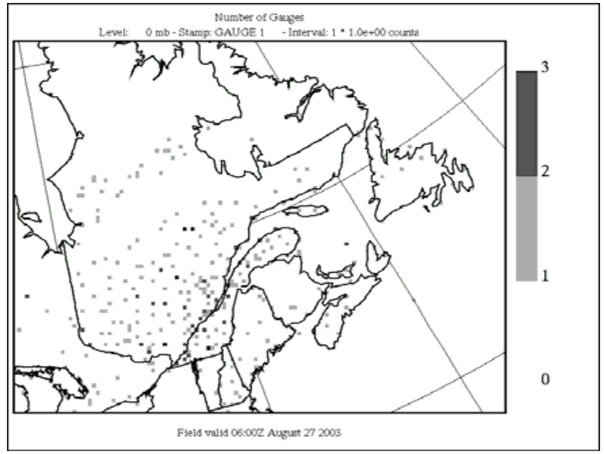
FATHER WAVELET SPACES

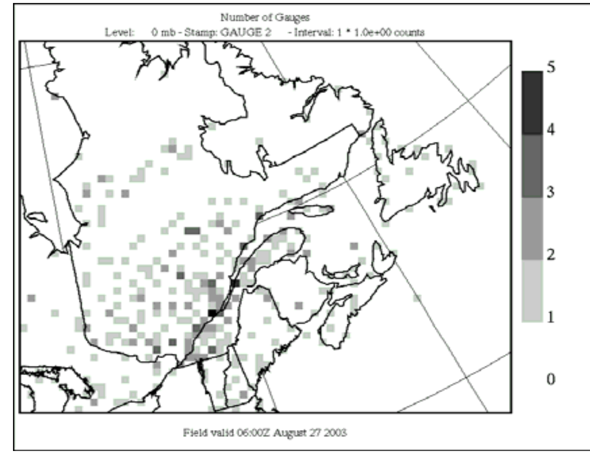


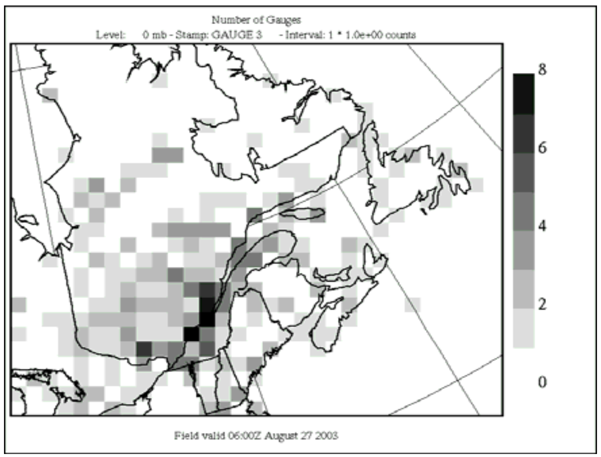
27 Aug 2003 6:00Z
6h accumulation

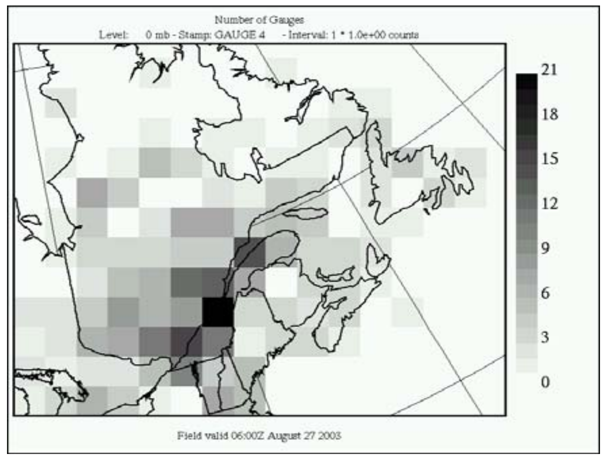
MOTHER
WAVELET
SPACES

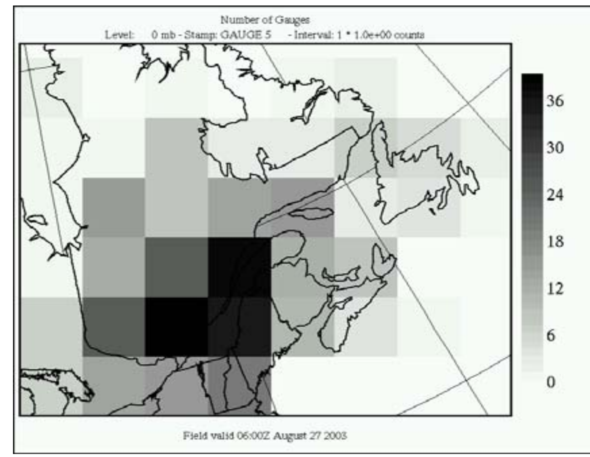


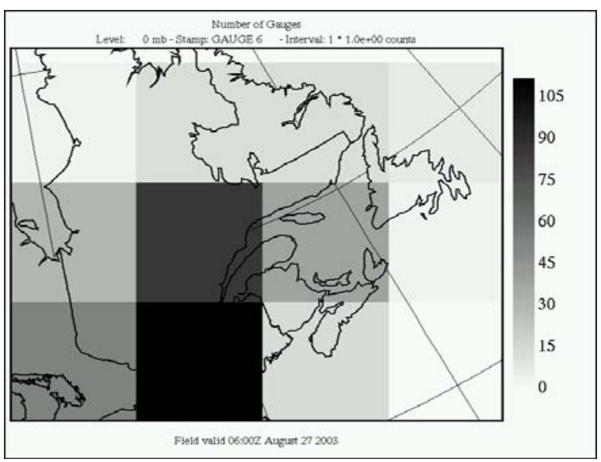


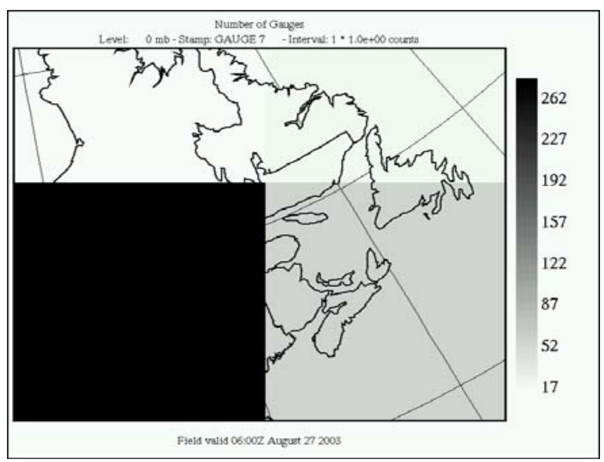












27 Aug 2003 6:00Z
6h accumulation

GAUGES NUMBER

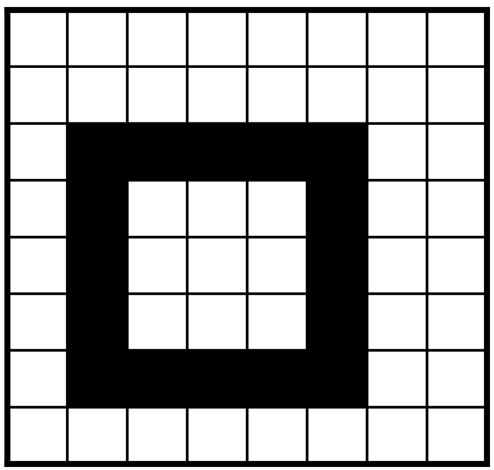
Hausdorff metrics, Baddeley Δ metric

Measure distance between binary images

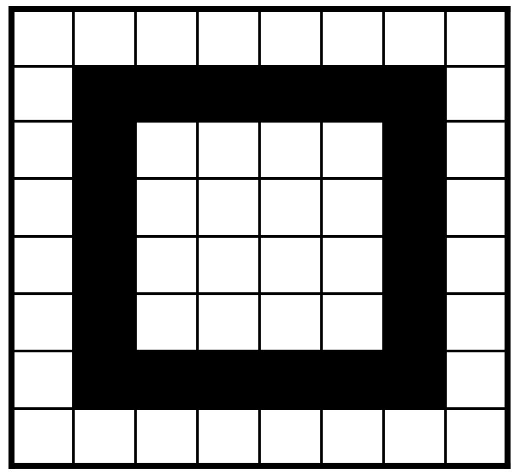
Account for object shape, distance, …

Alternative to use along with traditional categorical scores

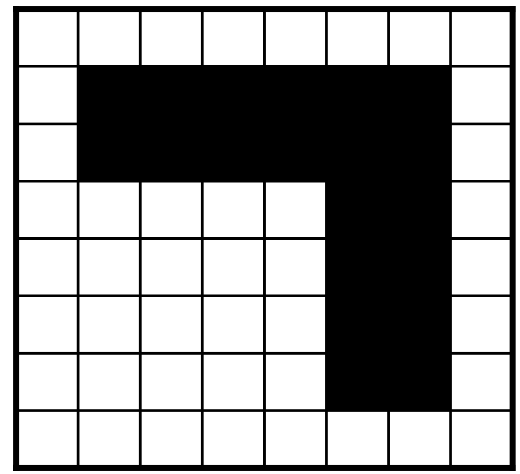
OBSERVATION



FORECAST **1**



FORECAST **2**



Δ = 0.5625	Δ = 0.96875
hits = 9; false alarms = 11; misses = 7; corr.rej. = 37	

Baddeley (1992);

Venugopal et al. (2005);

Gilleland et al.(2006)