Environment and Climate Change Canada / Environnement et Changement climatique Canada

Scale Separation diagnostics and the Symmetric Bounded Efficiency for the inter-comparison of precipitation re-analyses

Barbara Casati, Meteorological Research Division, ECCC - barbara.casati@ec.gc.ca
Cristian Lussana, Division for Climate Services, Met Norway
Alice Crespi, Institute of Earth Observation, EURAC, Bolzano, Italy

Talk outline:

- Precipitation data-sets + spatial verification methods
- Scale Separation diagnostics
  - Energy Normalized Bias
  - Skill on separate scales
  - Symmetric and Bounded Efficiency
- Conclusions

Canada

Data

24h accumulated precipitation fields.

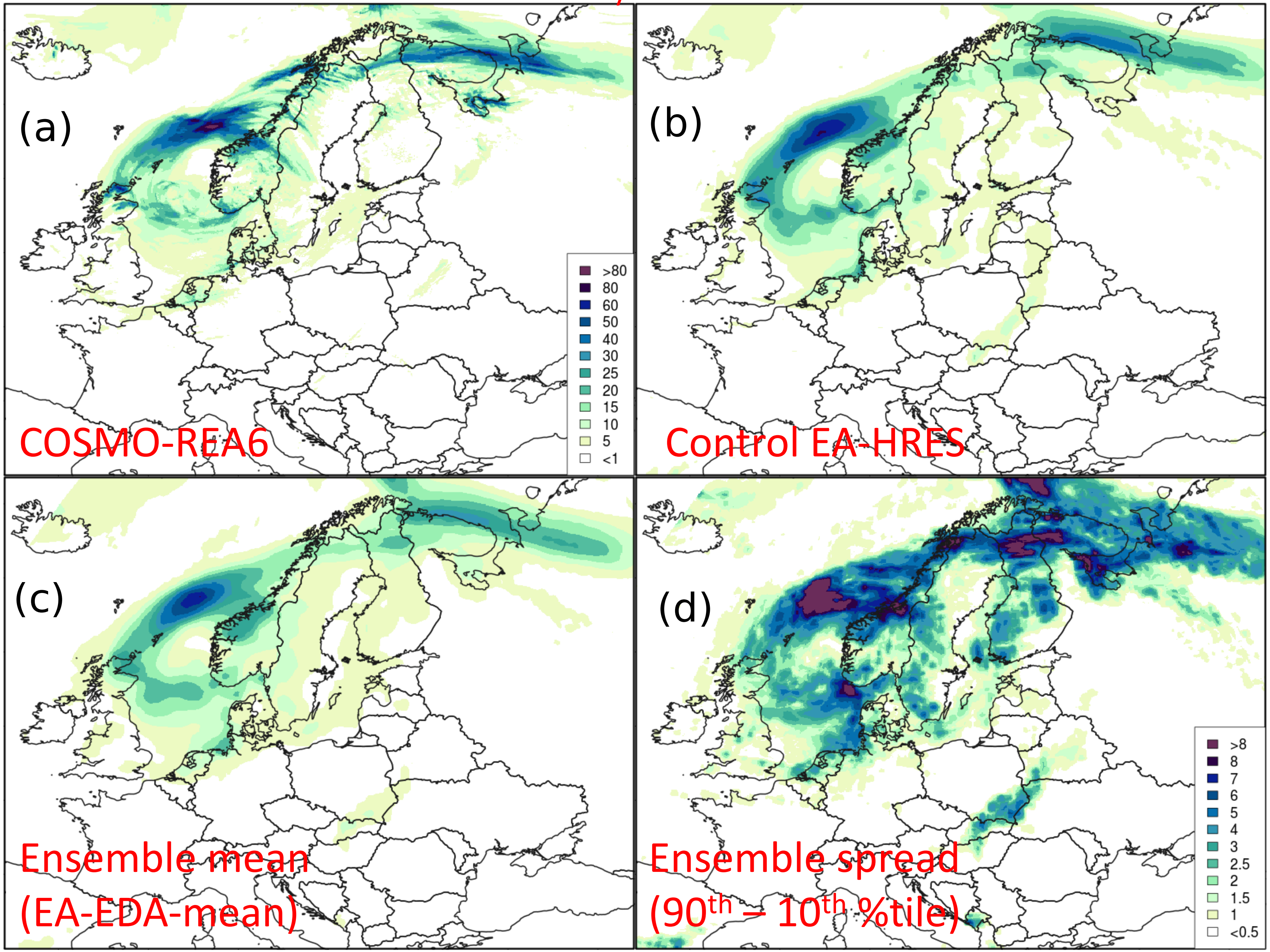
COSMO-REA6: high resolution (6km) regional reanalysis (Bollmeyer et al, 2015, QJRMS)

ECMWF-ERA5 reanalysis:

- High resolution (31km) control (EA-HRES)
- 10 member ensemble (EA-EDA) with reduced resolution

Results aggregated on 50 cases of intense precip in 2010-2014.

Case study 2011-08-29



Spatial verification methods

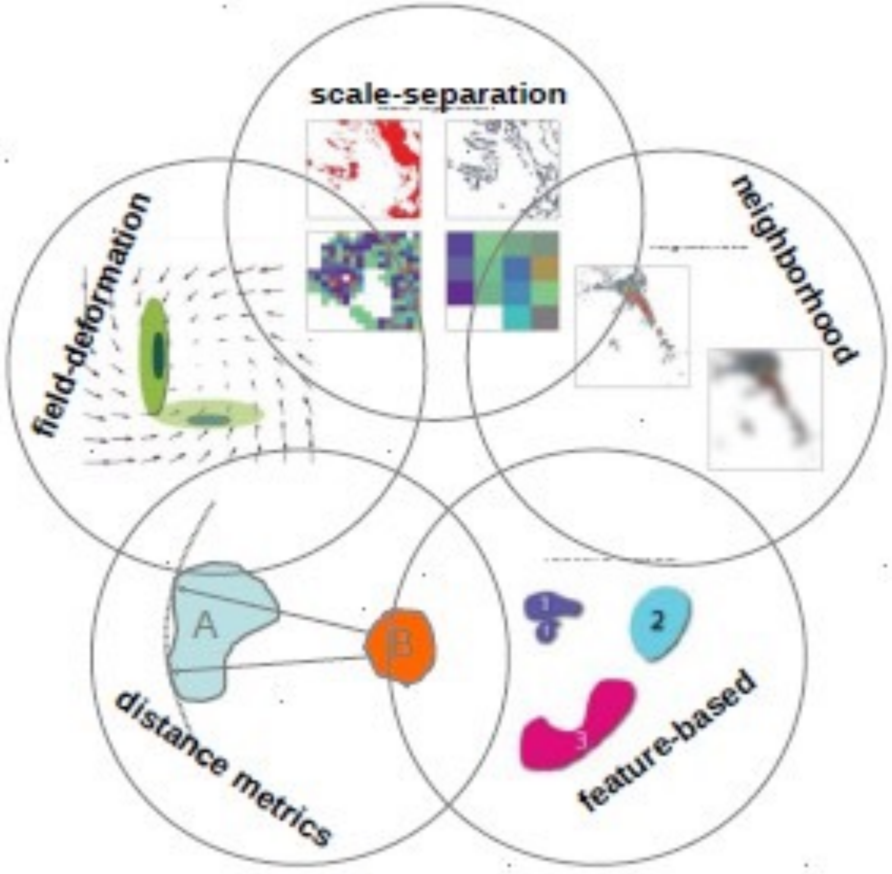
- Account for **coherent spatial structure** and the presence of **features**
- Aim to provide information on **error in physical terms (meaningful verification): e.g.** assess **scale structure** and **displacement error** (separately from **intensity error**)
- Account for **small time-space uncertainties** (avoid **double-penalty** issue)

Spatial method inter-comparisons:

- Spatial Verification Inter-Comparison Project (ICP): Gilleland et al (2010), BAMS
- Mesoscale Verification Intercomparison in complex Terrain (MesoVICT): Dorninger et al (2018), BAMS

http://www.ral.ucar.edu/projects/icp includes an list of more than 200 peer-review articles

Open source community verification tools: R spatialVx package, MET and METplus

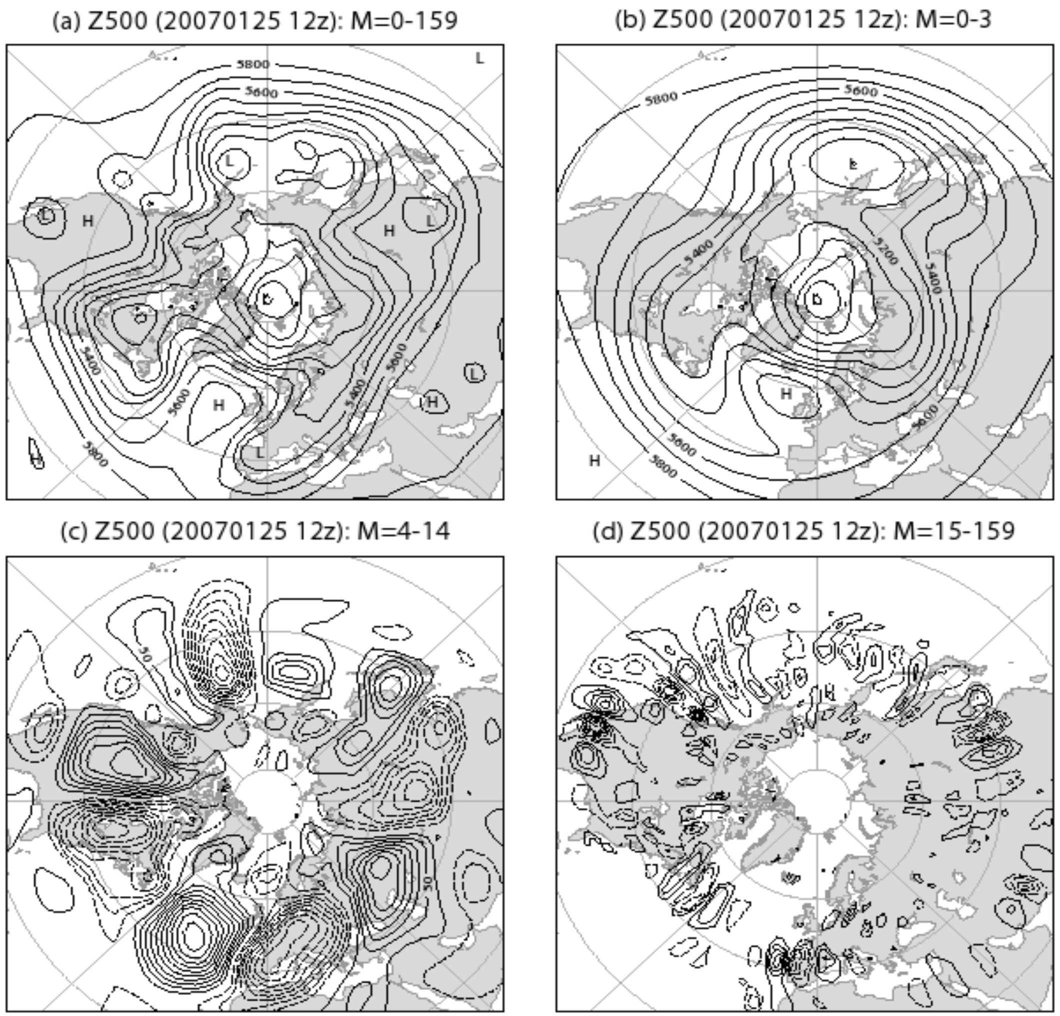


Scale Separation Diagnostics

<u>Rationale:</u> weather phenomena of different spatial scales are governed by different physical processes.

1. Decompose forecast and observation fields into spatial scale components (filters: wavelets, Fourier, DCT, …)
2. Perform verification on the different scale components, separately (continuous, categorical, probabilistic scores)

➔ Assess scale structure
➔ Bias, error and skill on different scales
➔ Scale dependency of forecast predictability (no-skill to skill transition scale)



from Jung and Leutbecher (2008)

ERA5 and COSMO-REA6 precipitation fields (Z) are decomposed into the sum of components on different spatial scales ($j=1 \dots J$) by using a 2D Haar discrete wavelet transform:

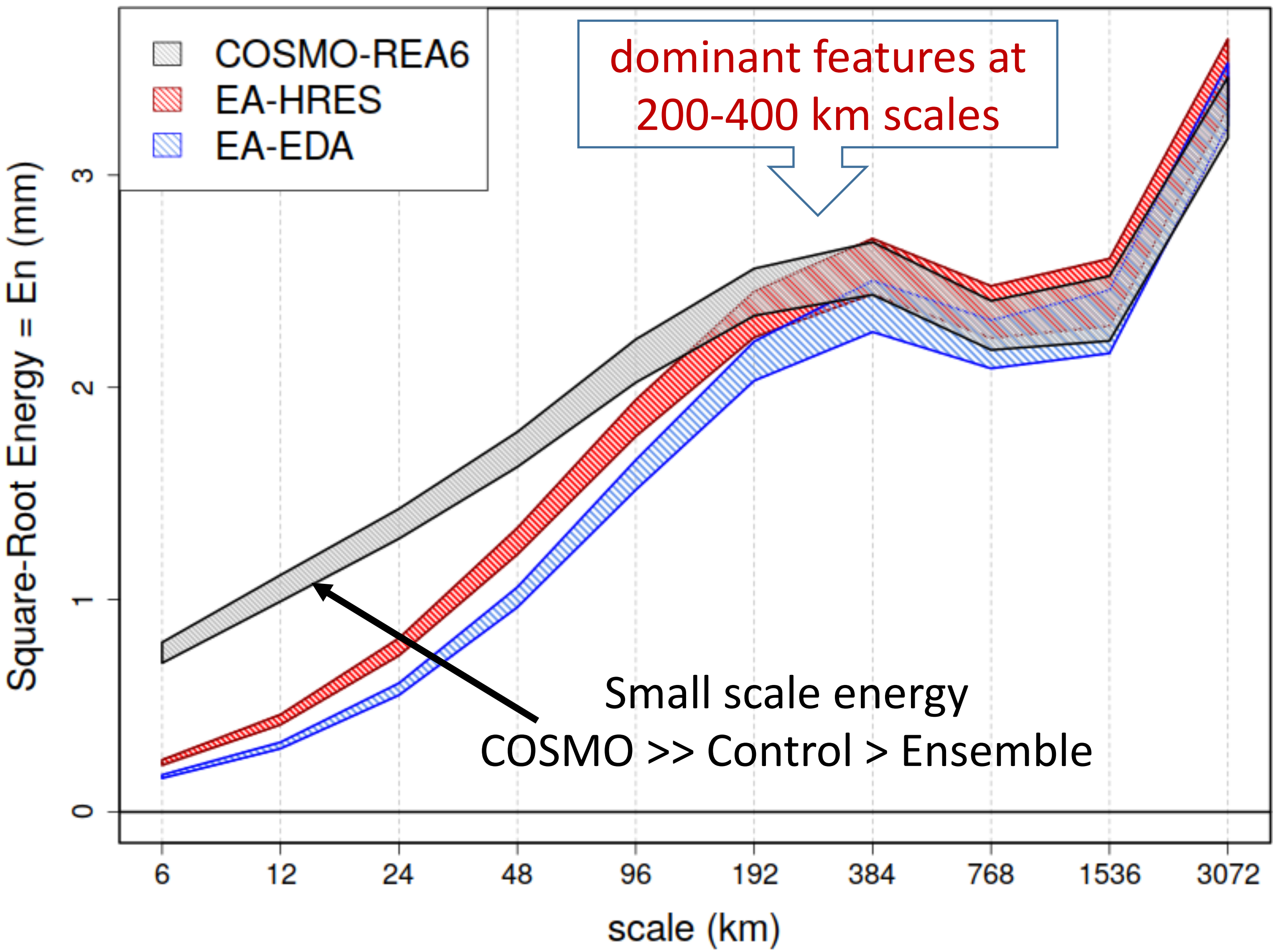
$$Z = \Sigma_{j=1}^{J} Z_j + \bar{Z}$$

Assessment of the (energy) bias on separate scales (1/2)

$$En^2\left(Z_j\right) = \overline{\left[Z_j\right]^2}$$

- The energy is proportional to magnitude and number of features on each scale
- Comparison of energy informs on bias on different scales (and scale structure)

- Wavelet components have zero mean $\mu_{Zj} = 0$; their energy is the field standard deviation $En\left(Z_j\right) = \sigma_{Zj}$; energy for the largest scales is the field spatial average μ_Z



Assessment of the (energy) bias on separate scales (2/2)

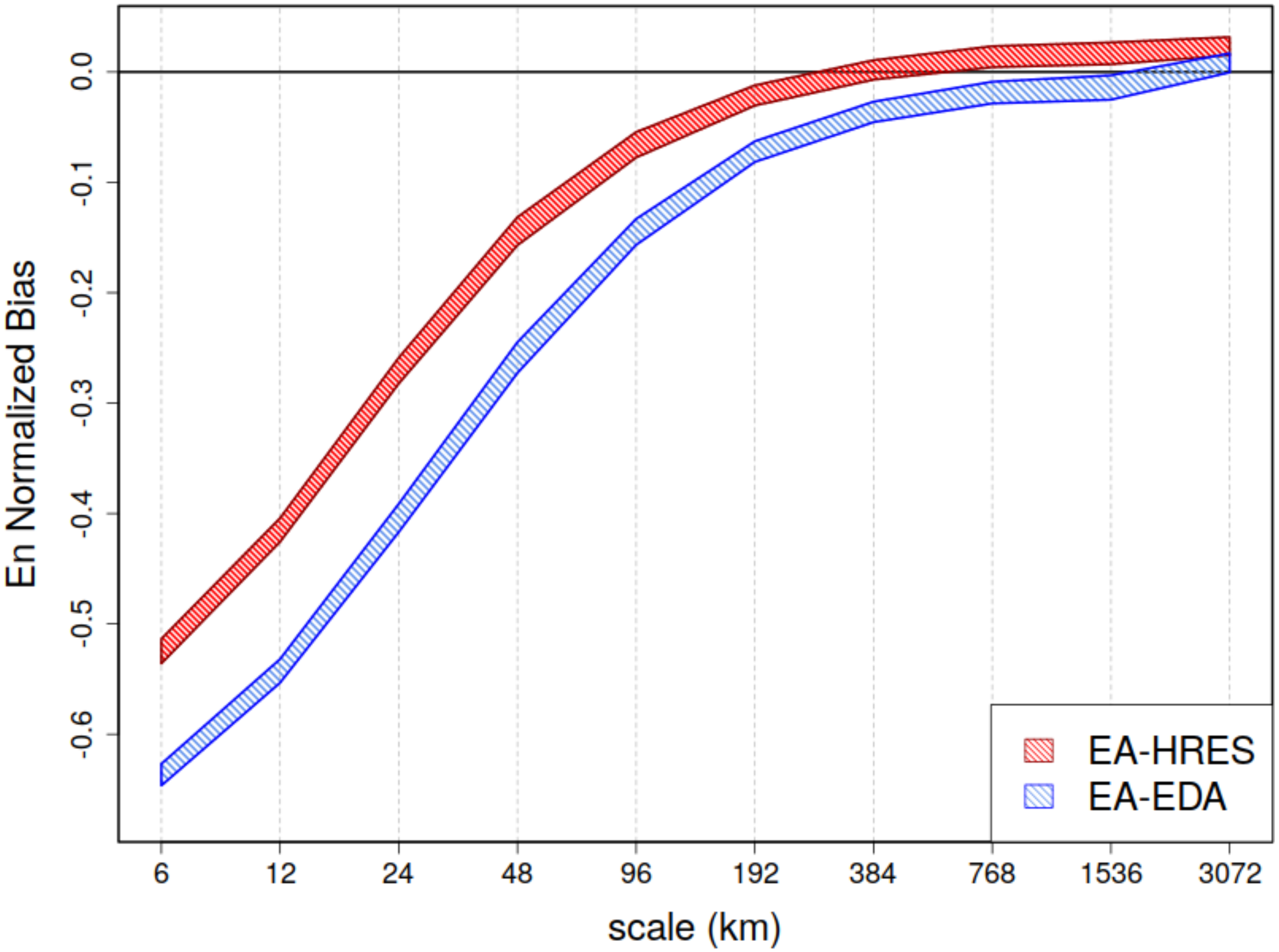
- Energy Normalized Bias (NB_{En}):

$$NB_\sigma = \frac{\sigma_{Yj} - \sigma_{Xj}}{\sigma_{Yj} + \sigma_{Xj}}$$

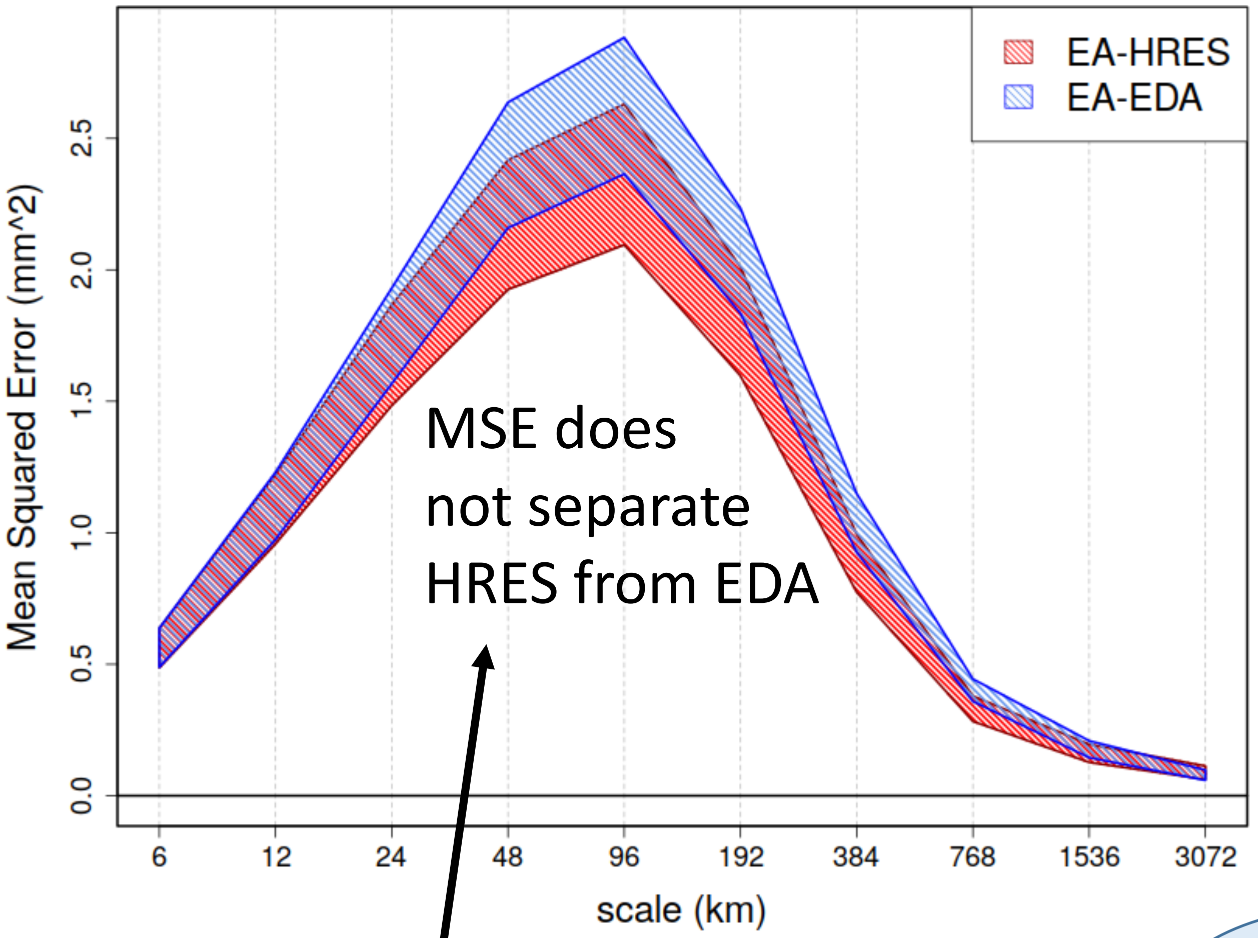
$$NB_\mu = \frac{\mu_Y - \mu_X}{\mu_Y + \mu_X}$$

for the largest scale

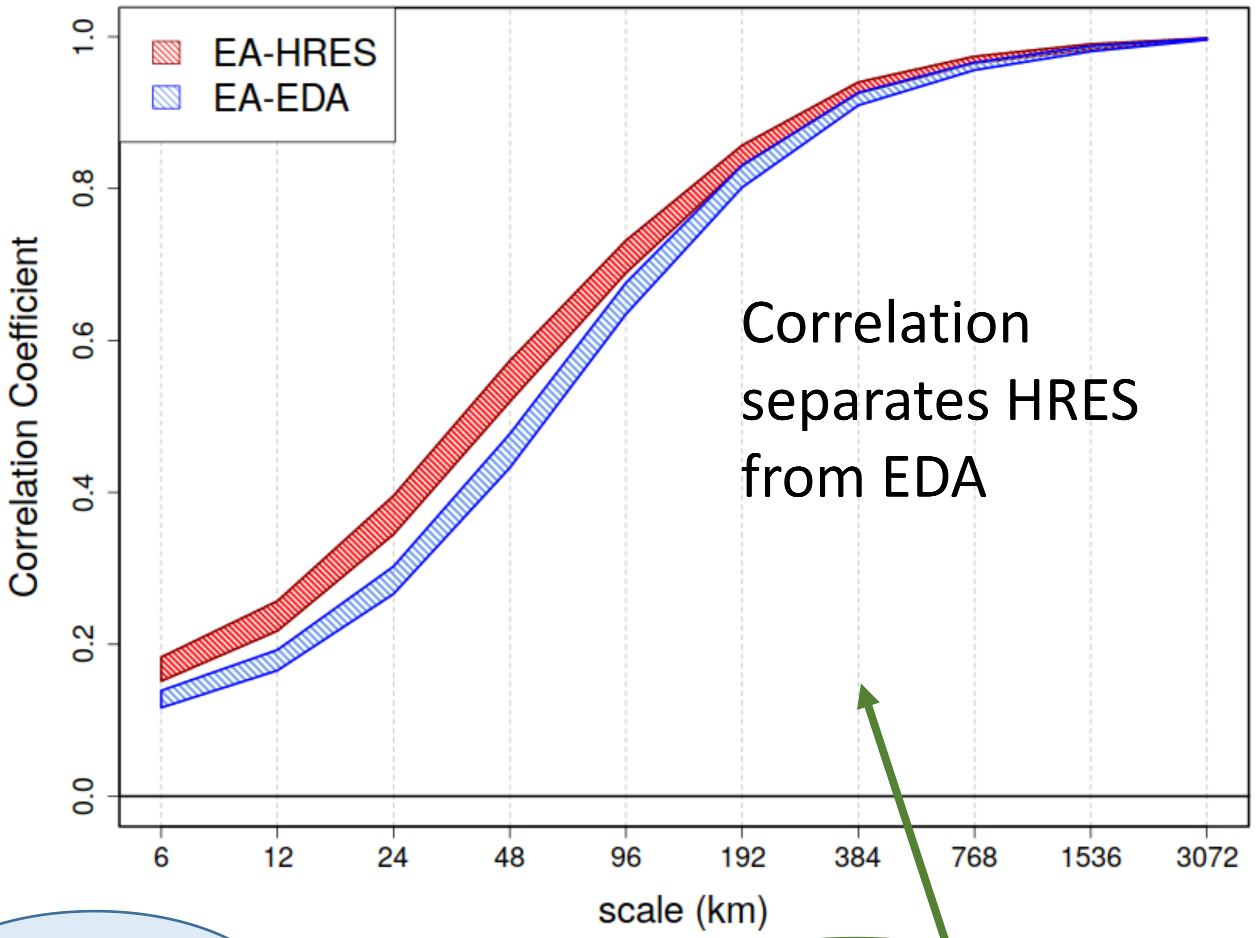
- Why normalizing? Additive bias would be small for small energies, even if their relative difference (e.g. as assessed by the ratio) could be large.
- Bounded: range [-1,1]; Symmetric (these are key properties for defining the Symmetric and Bounded Efficiency).



Measures of error and association



MSE does not separate HRES from EDA



Correlation separates HRES from EDA

$$MSE(Y,X) = \overline{[Y-X]^2} = (\mu_Y - \mu_X)^2 + \boxed{\sigma_Y^2 + \sigma_X^2} - 2\sigma_Y\sigma_X r_{Y,X}$$

Bias^2 Variability Covariance

Assessment of the skill on separate scales (1/2)

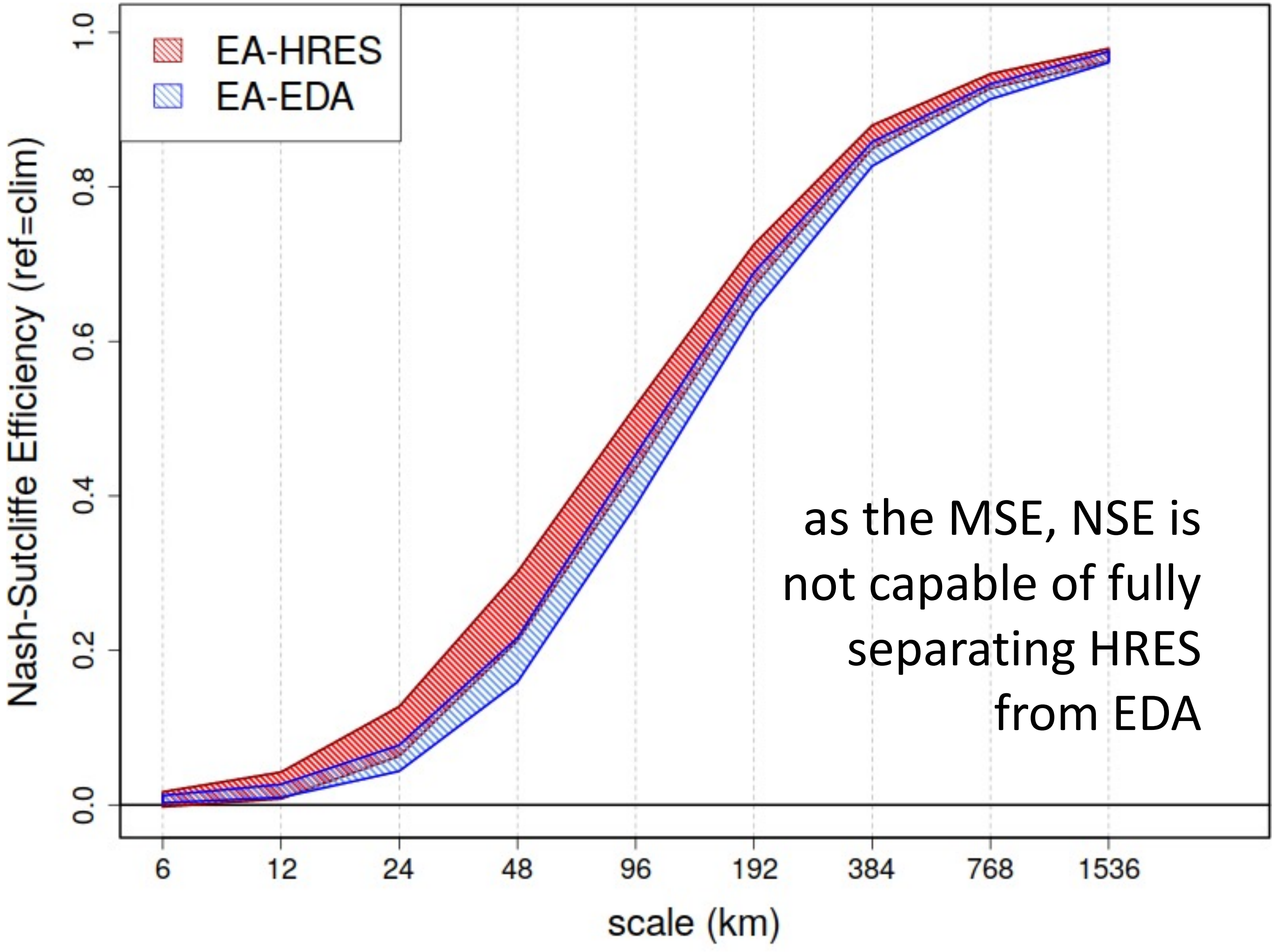
Skill: evaluate the performance against a reference (benchmark) forecast

e.g. Reduction of Variance

$$\frac{MSE - MSE_{clim}}{MSE_{perf} - MSE_{clim}} = 1 - \frac{MSE}{\sigma_X^2}$$

Use as reference forecast the sample climatology μ_X.

Also known as Nash-Sutcliffe Efficiency (1970) *J. of Hydrology*



as the MSE, NSE is not capable of fully separating HRES from EDA

Assessment of skill on separate scales (2/2)

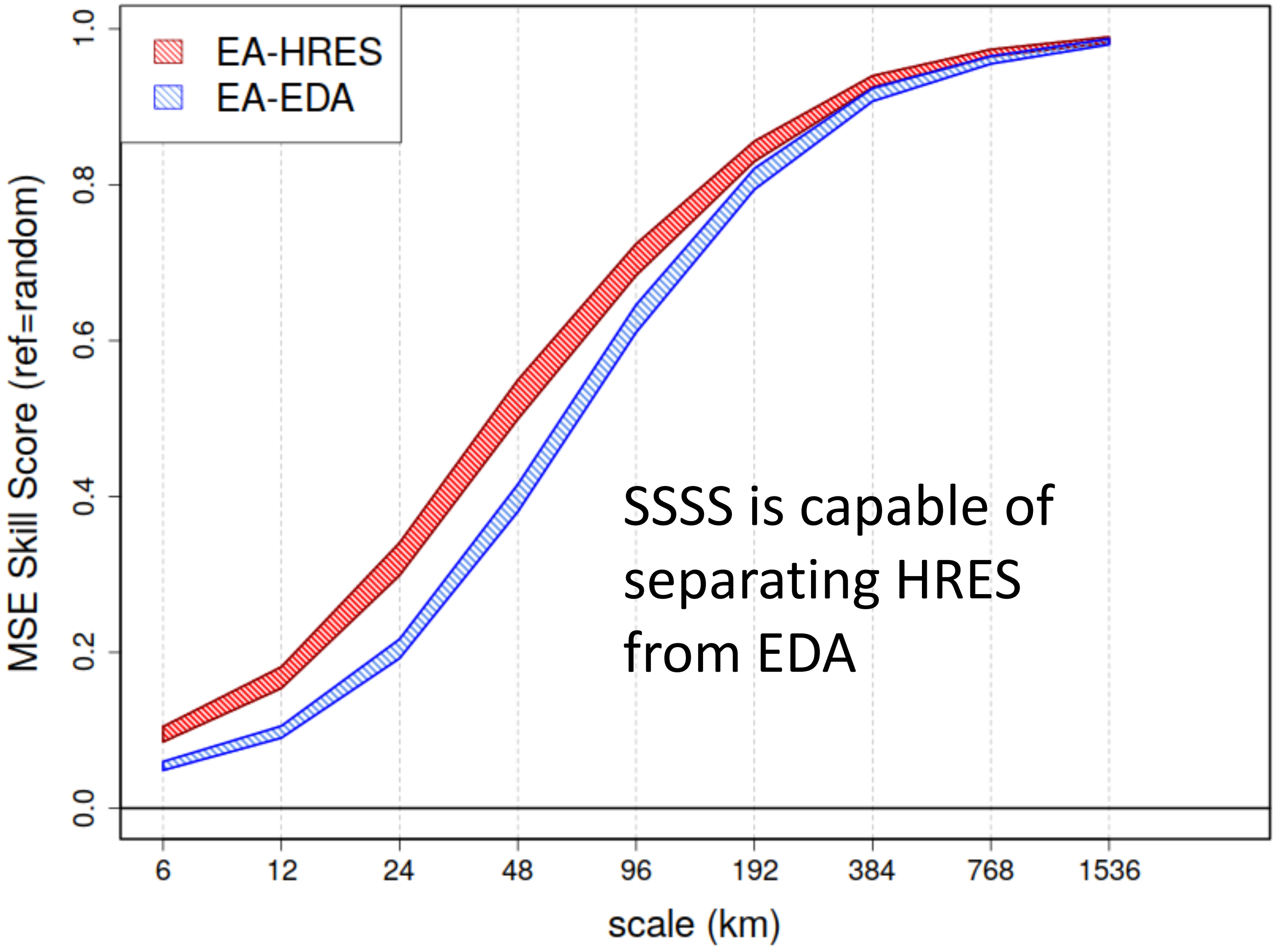
The Scale Separation Skill Score (SSSS) consider random chance* as reference forecast

$$SSSSj = 1 - \frac{MSEj}{\boxed{\sigma_{Yj}^2 + \sigma_{Xj}^2}}$$

Normalize with the sum of both obs and forecast variability

$$SSSSj = \frac{2\sigma_{Yj}\sigma_{Xj}r_{Yj,Xj}}{\sigma_{Yj}^2 + \sigma_{Xj}^2}$$

Is negative if forecast and obs are decorrelated



* Proof: $MSE_{random}(Y_j, X_j) = (\mu_{Yj} - \mu_{Xj})^2 + \sigma_{Yj}^2 + \sigma_{Xj}^2 - 2\sigma_{Yj}\sigma_{Xj}r_{Yj,Xj}$

0 0 0

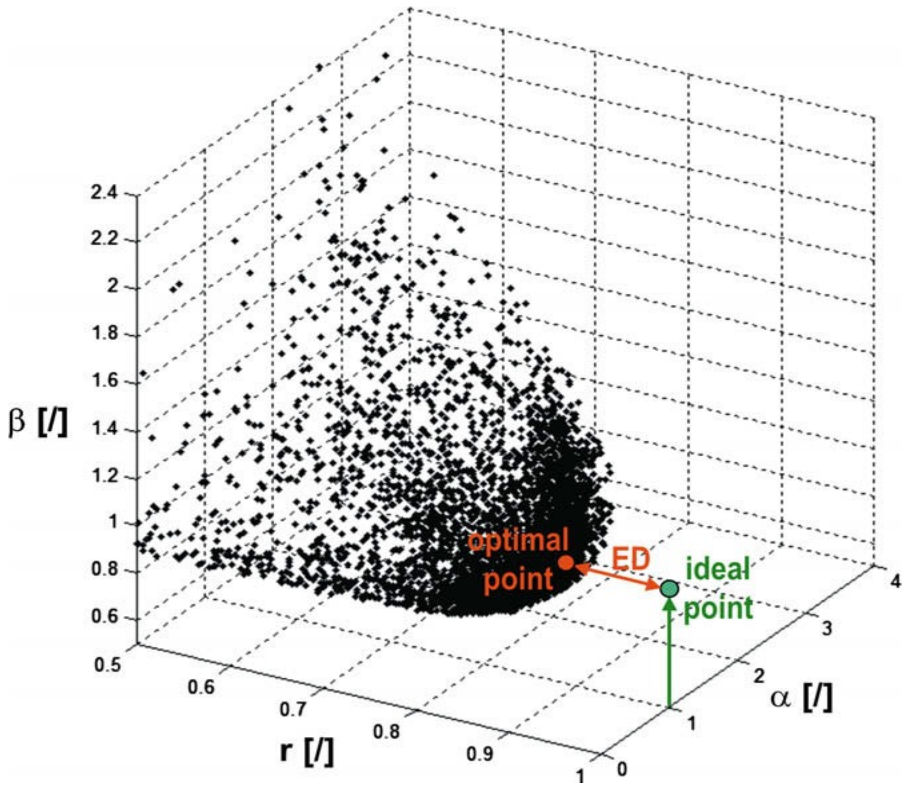
MSE, corr and Skill: summary of the key messages

- MSE depends on variability of both obs and forecast: historically, the MSE tends overly penalize high resolution products wrt coarser/smoother ones.
- NSE normalizes by the obs stdev only: forecasts with higher variability are still penalized!
- SSSS, on the other hand, normalizes by both variabilities, giving high resolution forecasts a chance! SSSS enables a fairer comparison of products with different resolutions, and the assessment of the added value of increasing resolution.

  - NSE is not capable of separating HRES from EDA, whereas SSSS does (correlation does too, but less remarkably): SSSS factors in the differences/similarities in variability

- **SSSS sign = correlation sign**: can be negative if forecast and obs anticorrelate (less likely for re-analyses than for forecasts, e.g. on small scales and long lead times).
- **NSE is not symmetric, whereas SSSS is symmetric** (invariant wrt the order of comparison)

The Kling-Gupta and the Symmetric Bounded Efficiency



Gupta et al (2009) *J. of Hydrology*:

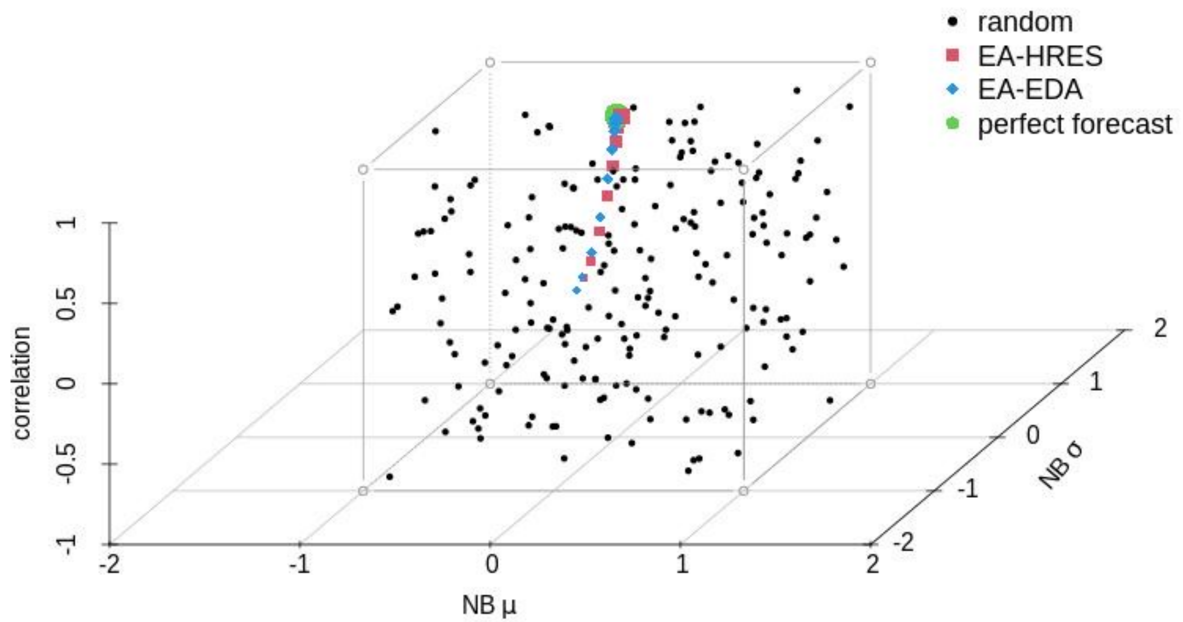
$$KGE = 1 - \sqrt{(r-1)^2 + \left(\frac{\sigma_Y}{\sigma_X} - 1\right)^2 + \left(\frac{\mu_Y}{\mu_X} - 1\right)^2}$$

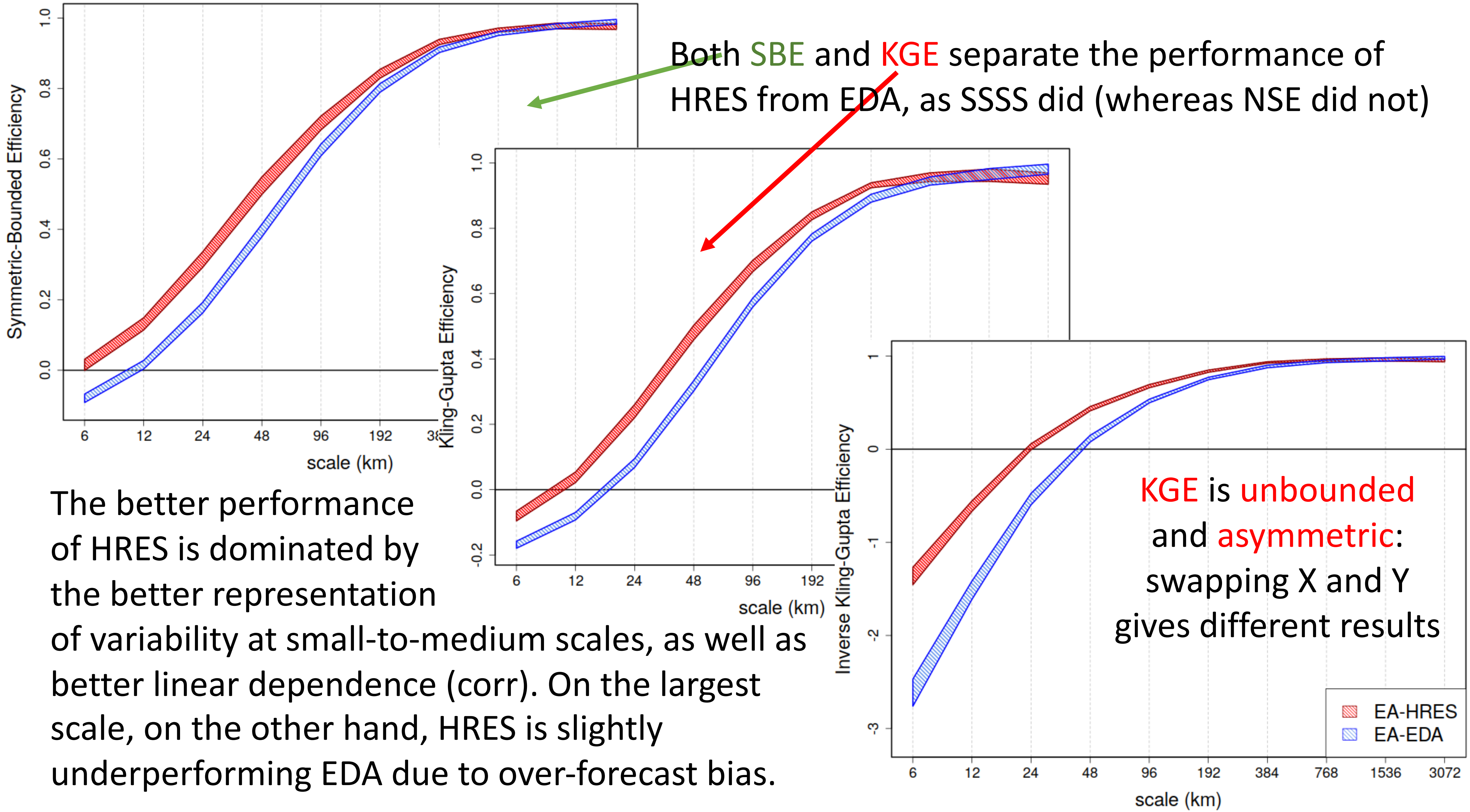
Corr Variability Bias

The SEB is defined as the KGE, but the ratio-1 is replaced by the Normalized Bias:

$$SEB = 1 - \sqrt{(r-1)^2 + \left(\frac{\sigma_{Yj} - \sigma_{Xj}}{\sigma_{Yj} + \sigma_{Xj}}\right)^2 + \left(\frac{\mu_{Yj} - \mu_{Xj}}{\mu_{Yj} + \mu_{Xj}}\right)^2}$$

Corr Variability Bias





Both SBE and KGE separate the performance of HRES from EDA, as SSSS did (whereas NSE did not)

KGE is unbounded and asymmetric: swapping X and Y gives different results

The better performance of HRES is dominated by the better representation of variability at small-to-medium scales, as well as better linear dependence (corr). On the largest scale, on the other hand, HRES is slightly underperforming EDA due to over-forecast bias.

Summary and Conclusions

- Scale separation diagnostics assess the bias, error and skill on separate scales: they enable indepth analysis of forecast performance in association with its scale structure.
- MSE as NSE are proportional to the forecast variability and hence tend to overly penalize high resolution products wrt coarser/smoother ones); The SSSS (ref=random) enables a fairer comparison of products with different resolutions.
- The efficiencies summarize the overall performance (accounting for correlation, variability and average value). SBE is bounded and symmetric, suitable for assessing reanalyses, since invariant wrt the order of comparison.
    - The Energy NB, SSSS, SBE separate HRES from EDA (whereas MSE and NSE did not).
    - The better performance of HRES is dominated by better representation of variability at small-to-medium scales, as well as better linear dependence (corr). On the largest scale, on the other hand, HRES is slightly underperforming EDA due to over-forecast bias.

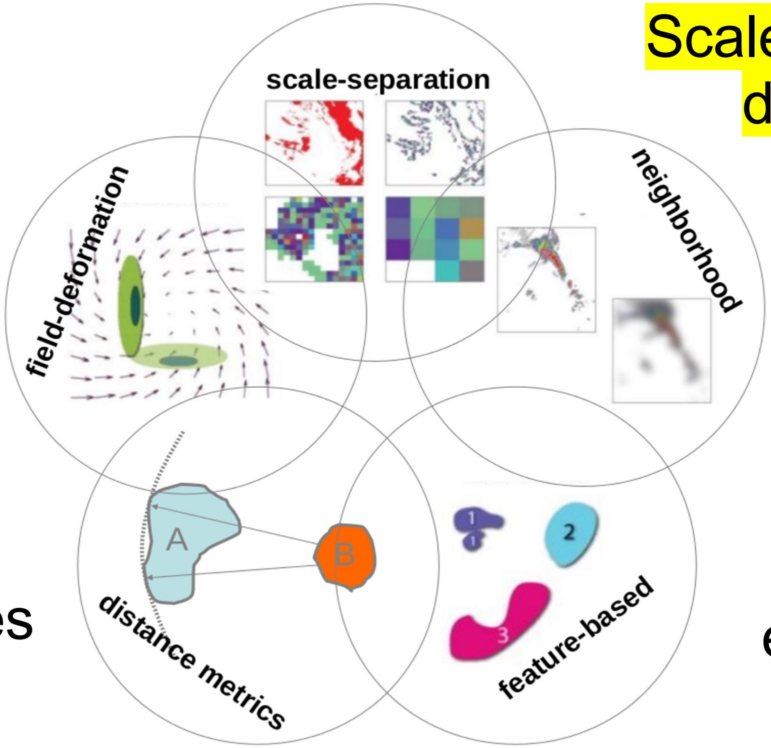
THANK YOU!

Extras

Spatial verification approaches

- account for **coherent spatial structure** and the presence of **features**
- provide information on **error in physical terms (meaningful verification)**, e.g. assess **location and timing errors** (separate from **intensity error**)
- account for **small time-space uncertainties** (avoid **double-penalty** issue)

Field-deformation: use a vector and scalar field to morph forecast into obs



Scale-separation: analyse scale-dependency of forecast error

Neighborhood: relax requirement of exact space-time matching

Distance metrics: evaluate distances for all grid-points

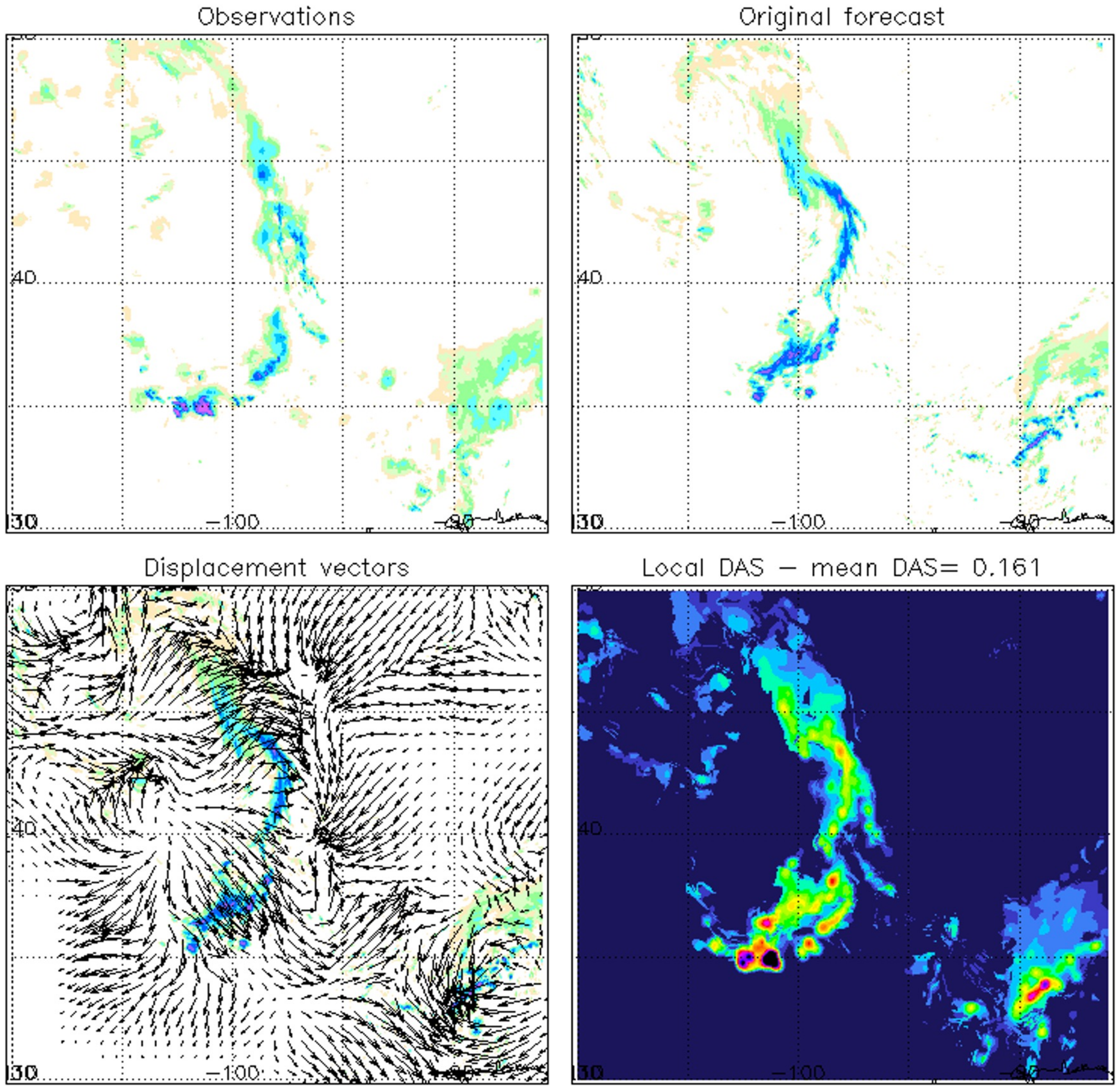
Feature-based: evaluate attributes of isolated features

From Dorninger et al 2018, BAMS

Field-deformation methods

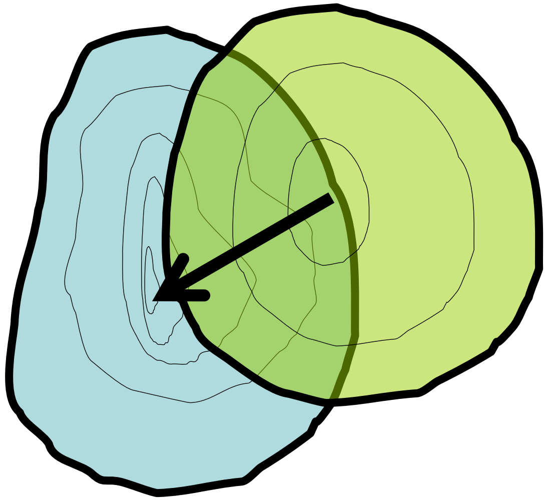
Hoffmann et al (1995); Hoffman and Grassotti (1996), Nehrkorn et al. (2003);
Brill (2002); Germann and Zawadzki (2002, 2004);
Keil and Craig (2007, 2009) DAS;
Marzbar and Sandgathe (2010) optical flow;
Alexander et al (1999), Gilleland et al (2010) image warping

1. Use a vector (wind) field to deform the forecast field towards the obs field
2. Use an amplitude field to correct intensities of (deformed) forecast field to those of the obs field



- Vector and amplitude fields provide physically meaningful diagnostic information: feedback for data assimilation and now-casting.
- Error decomposition is performed on different spectral components: directly inform about small scales uncertainty versus large scale errors.

Feature-based techniques

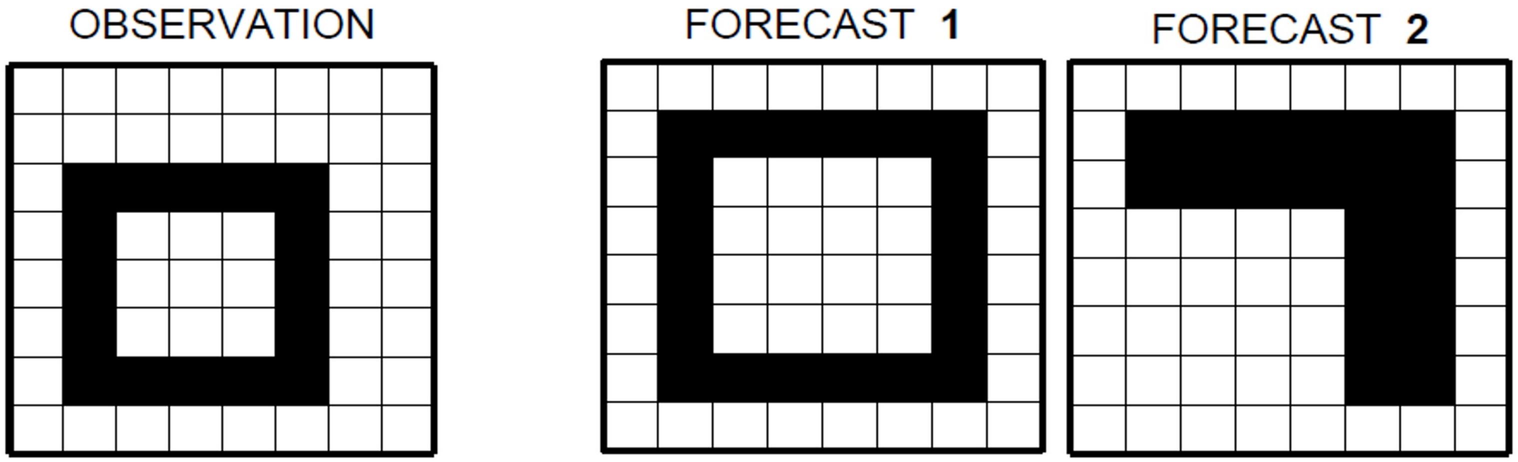


Ebert and McBride (2000), Grams et al (2006), Ebert and Gallus (2009): CRA
Davis, Brown, Bullok (2006) I and II, Davis et al (2009): MODE
Wernli, Paulat, Frei (2008): SAL score
Nachamkin (2004, 2005): composites
Marzban and Sandgathe (2006): cluster
Lack et al (2010): procrustes

1. Identify and isolate (precipitation) **features** in forecast and observation fields (thresholding, image processing, composites, cluster analysis)
2. assess displacement and amount (extent and intensity) error for each pairs of obs and forecast features; identify and verify attributes of object pairs (e.g. intensity, area, centroid location); evaluate distance-based contingency tables and categorical scores; perform verification as function of feature size (scale); add time dimension for the assessement of the timing error of precipitation systems.

Distance metrics for binary images

- Average distance
- K-mean
- Fréchet distance
- **Hausdorff metric**
  - **Modified Hausdorff**
  - **Partial Hausdorff**
  - **Baddeley metric**
- Pratts' figure of merit



Δ= 0.5625	Δ= 0.96875
hits = 9; false alarms = 11; misses = 7; nils = 37	

→ Evaluate distances for all grid-points

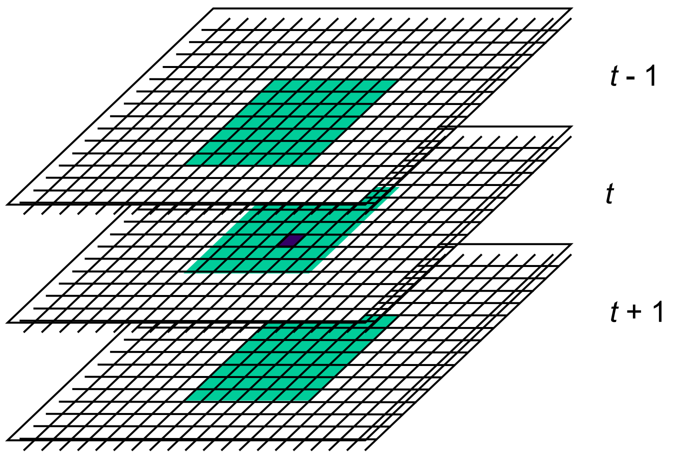
→ Account for distance between objects, and similarity in shapes.

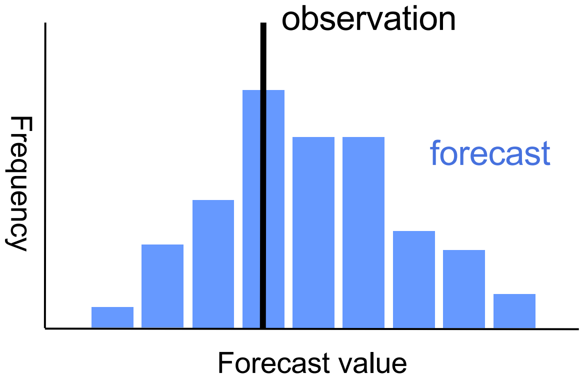
→ Compare binary images: alternative metrics to be used along with traditional categorical scores

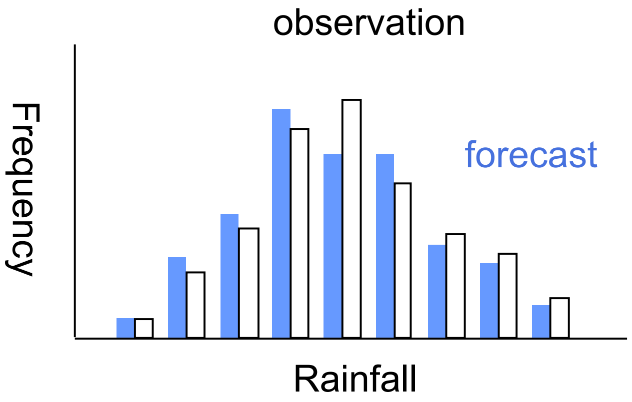
Image Processing: **Baddeley (1992); Dubuisson and Jain (1994).** Precipitation: **Gilleland et al.(2008, 2011)**, Schwedler & Baldwin (2011), Venugopal et al. (2005); Zhu et al (2010), Aghakouchak et al (2011). **Brunet and Sills (2015).** Sea-ice: Heinrichs et al (2006); **Dukhovskoy et al (2015)**, Hebert et al (2015).

Neighbourhood verification

1. Define neighbourhood of grid-points: relax requirements for exact positioning (mitigate double penalty: suitable for high resolution models); account for forecast and obs time-space uncertainty.







2. Perform verification over neighbourhoods of different sizes: verify deterministic forecast with probabilistic approach

Yates (2006), upscaling, cont&cat scores; **Tremblay et al. (1996)**, distance-dependent POD, POFD; **Rezacova and Sokol (2005)**, rank RMSE; **Roberts and Lean (2008)** Fraction Skill Score; **Theis et al (2005)**; pragmatical approach; **Atger (2001),** spatial multi-event ROC curve; **Marsigli et al (2005, 2006)** probabilistic approach.

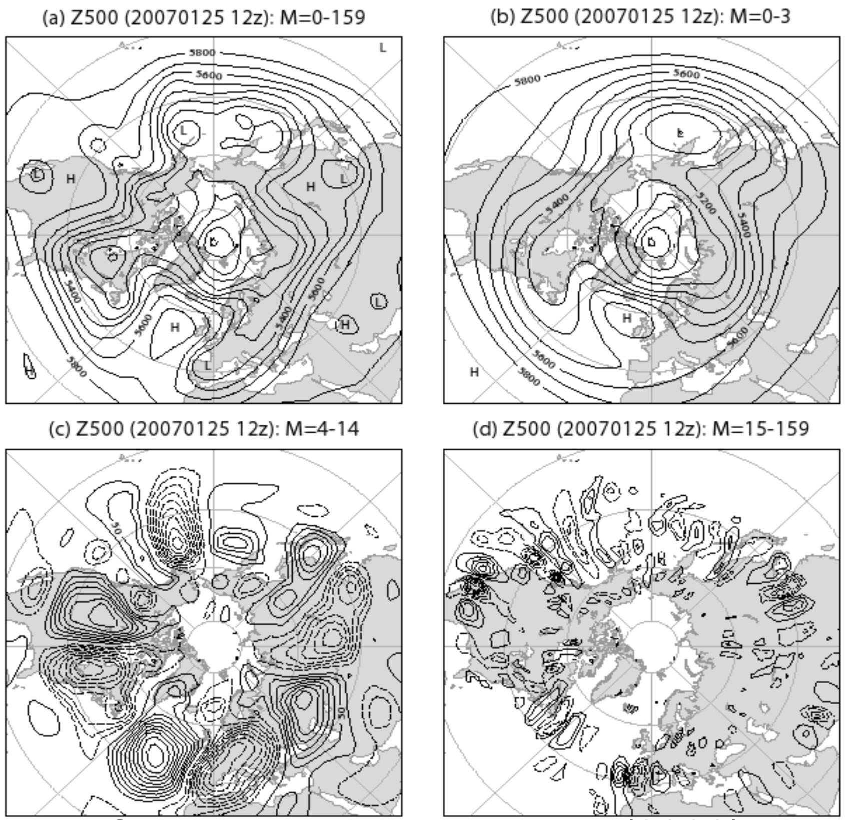
Scale Separation methods

Briggs and Levine (1997), wavelet cont (MSE, corr); **Casati et al. (2004), Casati (2010)**, wavelet cat (HSS, FBI, scale structure); **Zepeda-Arce et al. (2000), Harris et al. (2001), Tustison et al. (2003),** scale invariants parameters; **Casati and Wilson (2007)**, wavelet prob (BSS=BSSres-BSSrel, En2 bias, scale structure); **Jung and Leutbecher (2008)**, spherical harmonics, prob (EPS spread-error, BSS, RPSS); **Denis et al. (2002,2003), De Elia et al. (2002)**, discrete cosine transform, taylor diag; **Livina et al (2008),** wavelet coefficient score. **De Sales and Xue (2010)**

1. Decompose forecast and observation fields into spatial scale components (filters: wavelets, Fourier, DCT, …)
2. Perform verification on different scale components, separately (continuous, categorical, probabilistic scores)

<u>Rationale:</u> weather phenomena of different spatial scales are governed by different physical processes

➔ Assess scale structure
➔ Bias, error and skill on different scales
➔ Scale dependency of forecast predictability (no-skill to skill transition scale)

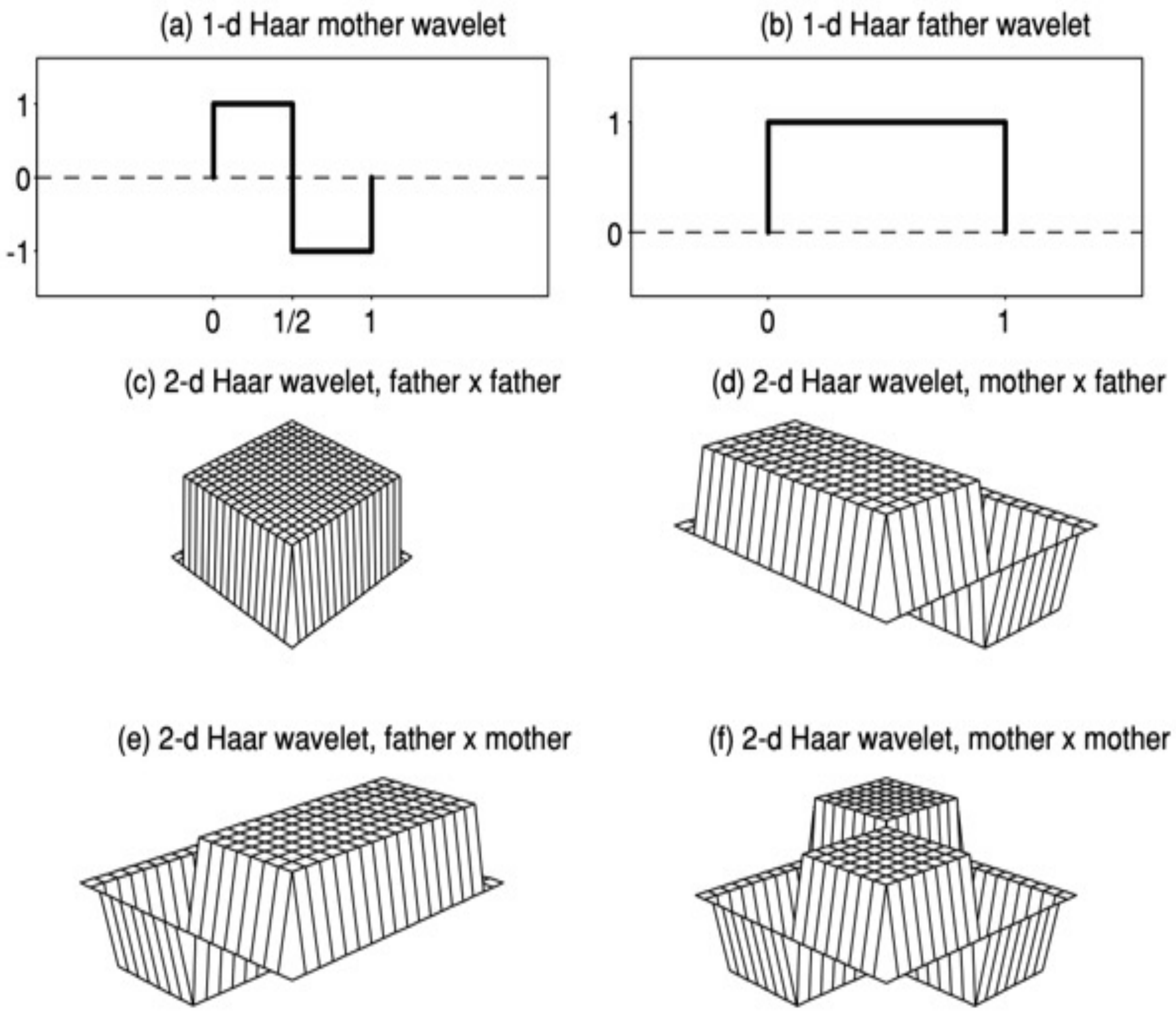


from Jung and Leutbecher (2008)

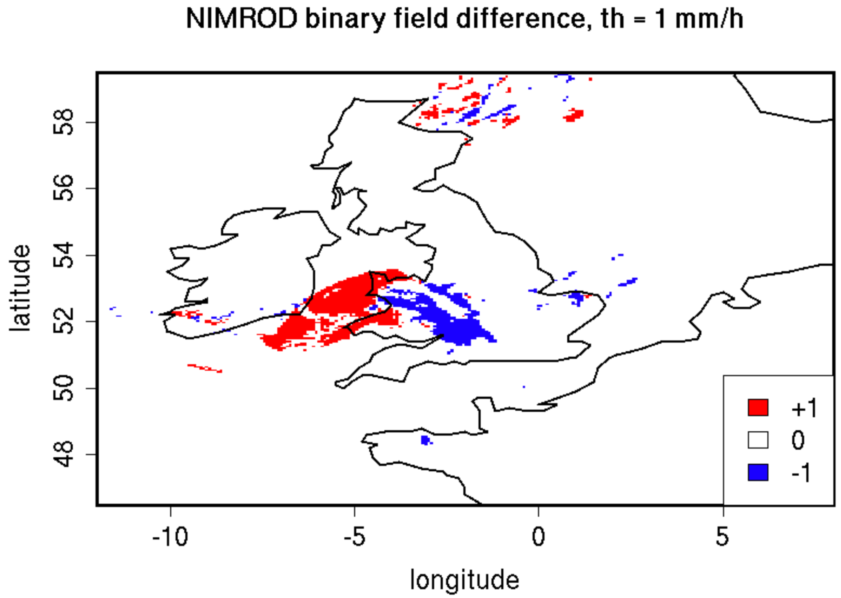
The Haar wavelet filter

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

Similar to the sine and cosine for the Fourier Transforms, discrete wavelets form an orthonormal basis of $L^2(\mathbb{R})$, so that any real function (e.g. a 2D field) can be expressed as a linear combination of wavelets, i.e. as a sum of components with different spatial scales.



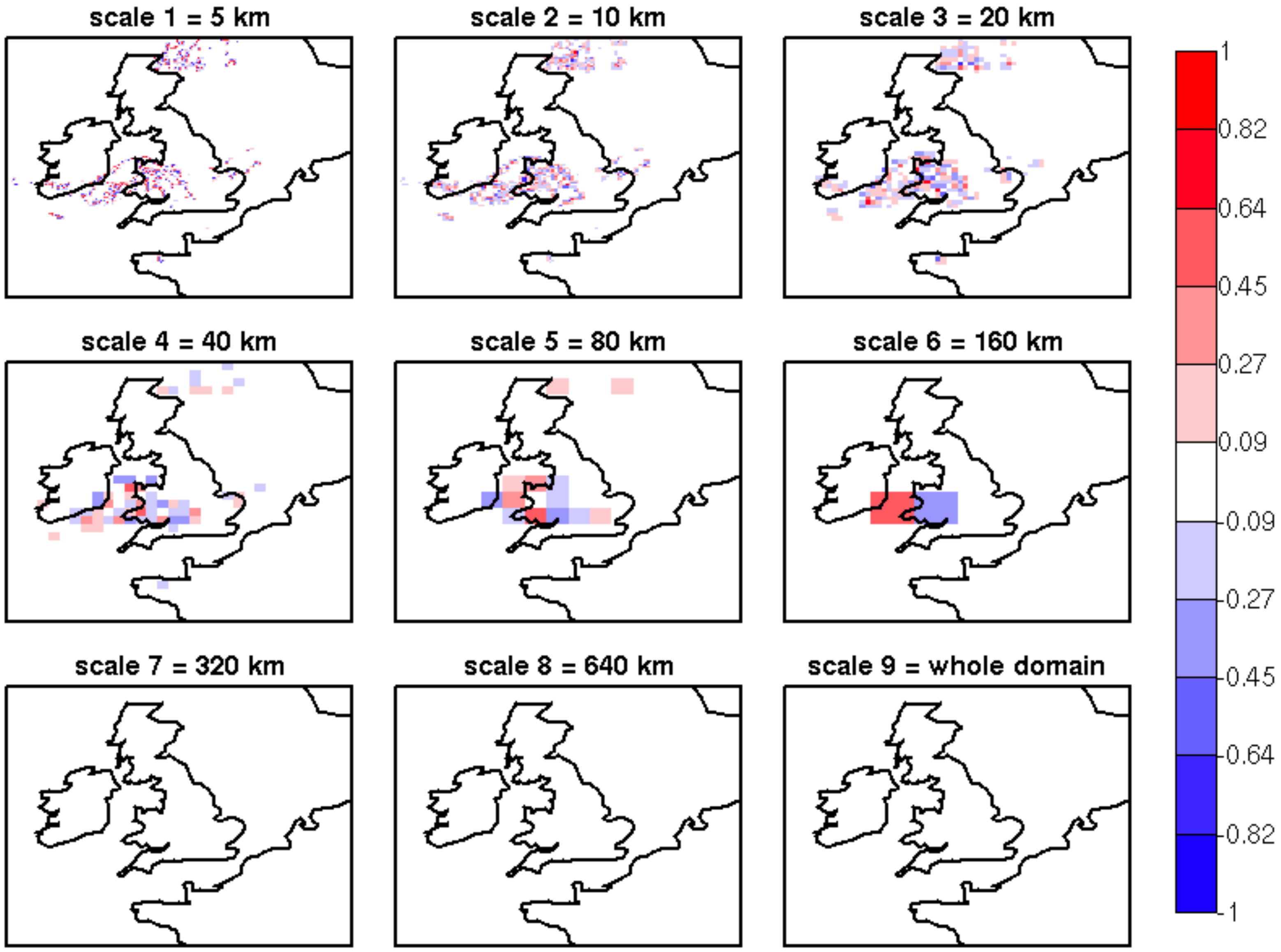
2D Haar discrete wavelet decomposition

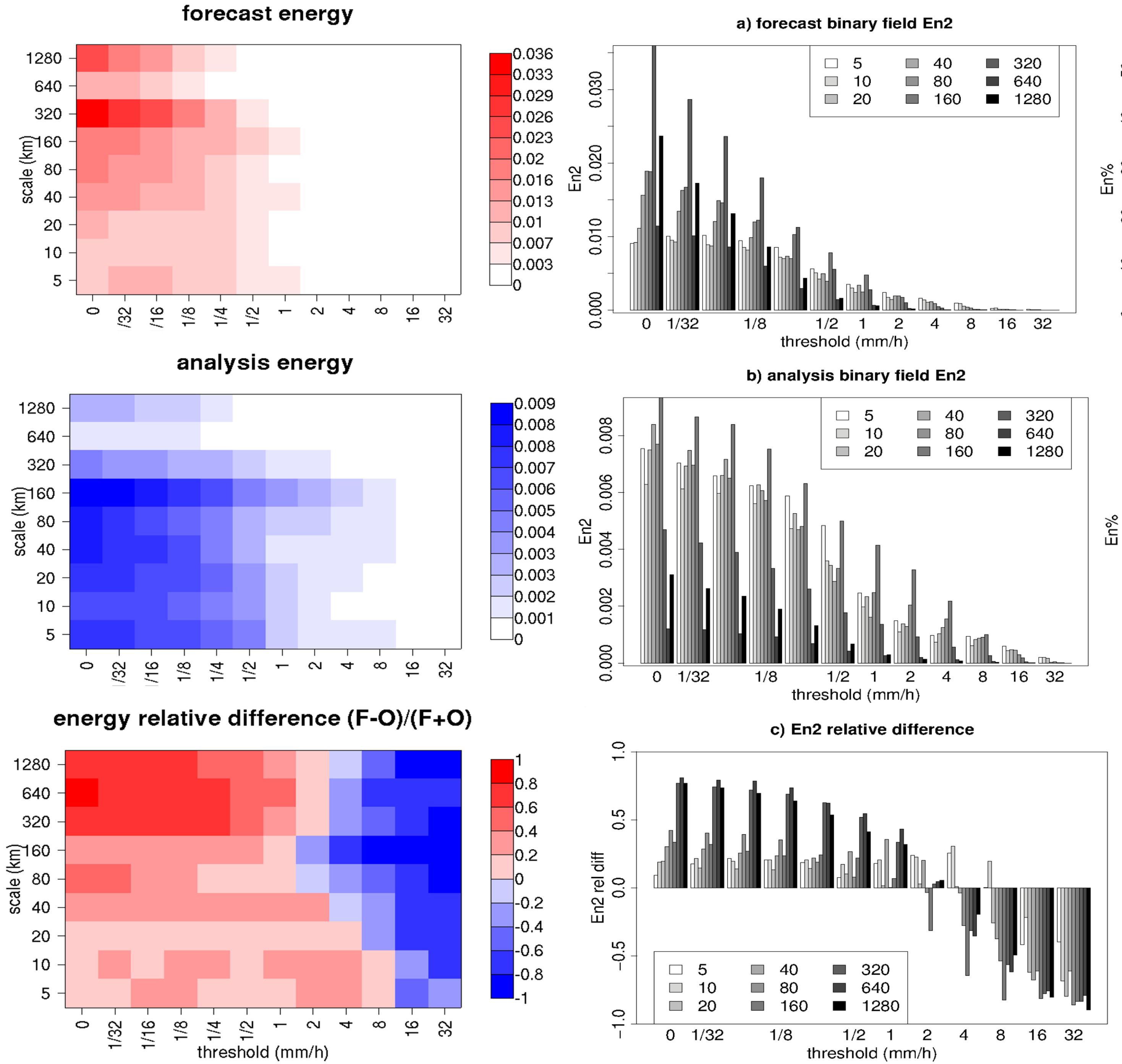


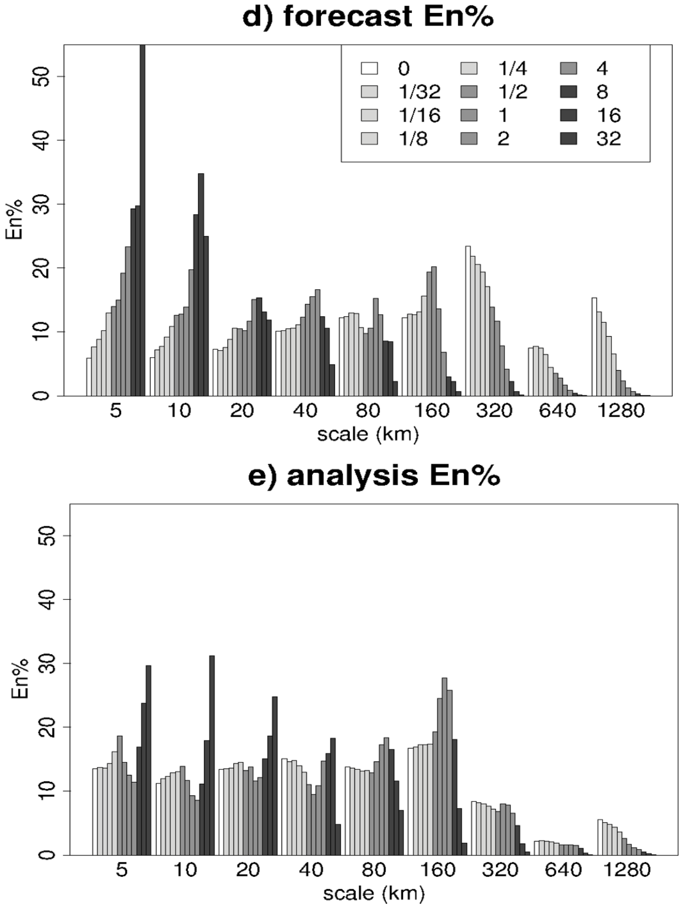
The field is decomposed into the sum of different components on different spatial scales

$$Z = \Sigma_{j=1}^{J} Z_j + \bar{Z}$$

The largest component is the field average.







Casati (2010) Wea & For, vol. 25

Intensity-scale verification technique

For each threshold and scale:

- **energy** informs on the amount of events;
- energy relative diff. measures the **bias**;
- energy % assess the **scale structure**.