Existing Verification Techniques

Traditional methods:

1. graphical summary (scatter-plot, box-plot)
2. Continuous scores (RMSE, correlation)
3. Categorical scores from contingency tables (FBI, ETS, POD)

Spatial methods:

1. Scale-decomposition
2. Neighbourhood-based
3. Field-morphing
4. Feature-based
5. Distance metrics for binary images

There is no single technique which fully describes the complex observation-forecast relationship

Key Q: what do we wish to know from our verification ?

Spatial verif inter-comparison: Gilleland et al (2009)

1. Scale-decomposition approaches

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 the sum of spatial components on different scales (wavelets, Fourier, DCT)
2. Perform verification on different scale components, separately (cont. scores; categ. approaches; probability verif. scores)

Account for the field coherent spatial structure:

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

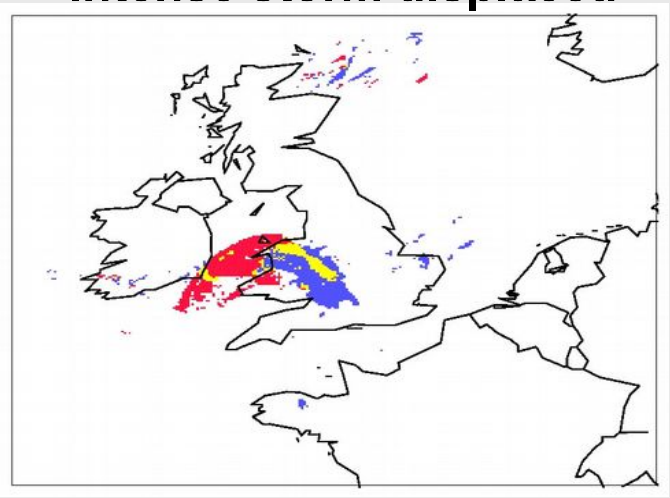
Tech Note: in the scale decomposition approaches the scale is obtained by a single-band pass filter. The scale is associated then to the feature size, provides feedback on physical processes associated to phenomena on different scales.

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

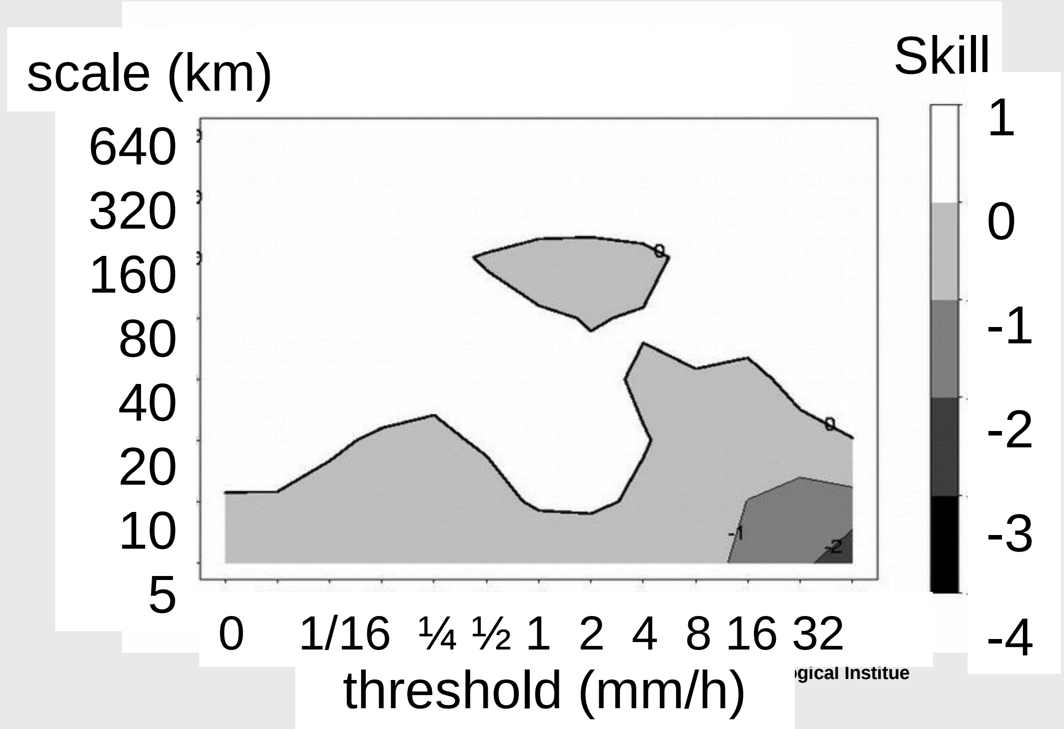
The intensity-scale verification approach measures the skill as function of precipitation intensity and spatial scale of the error

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

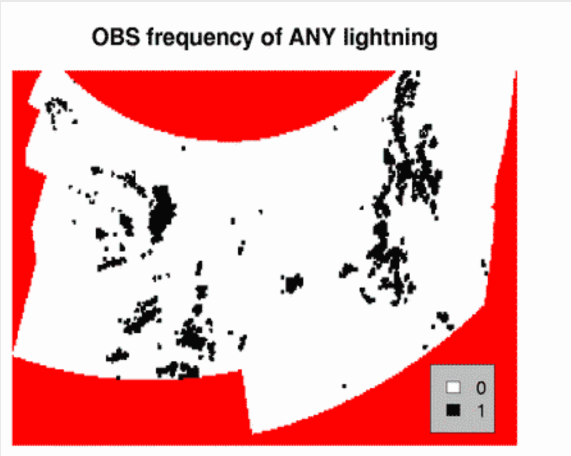
Intense storm displaced

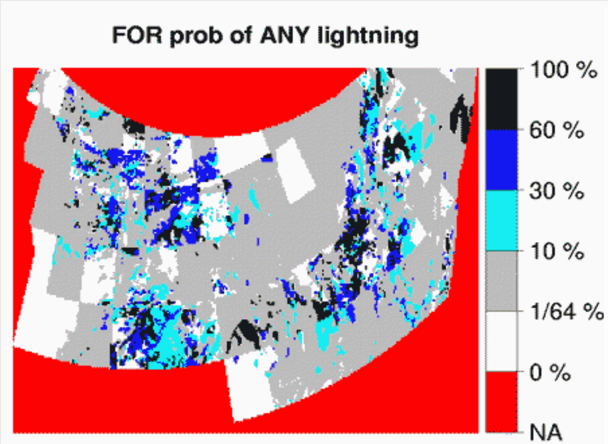


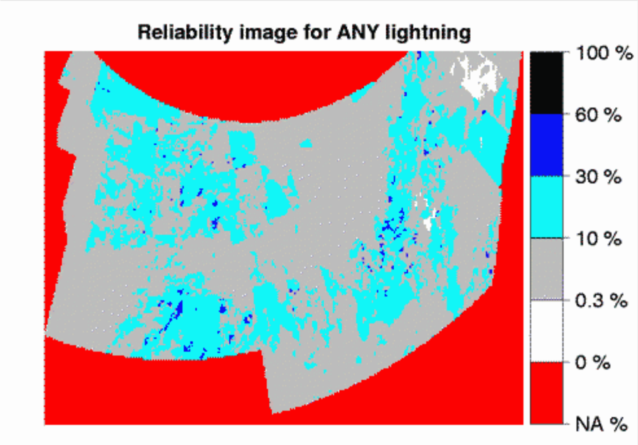
threshold = 1mm/h

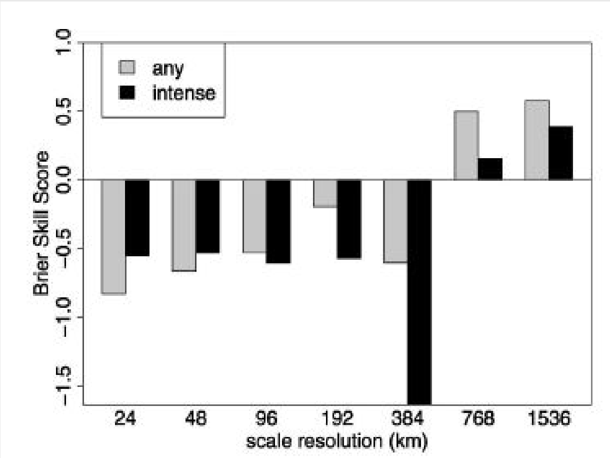


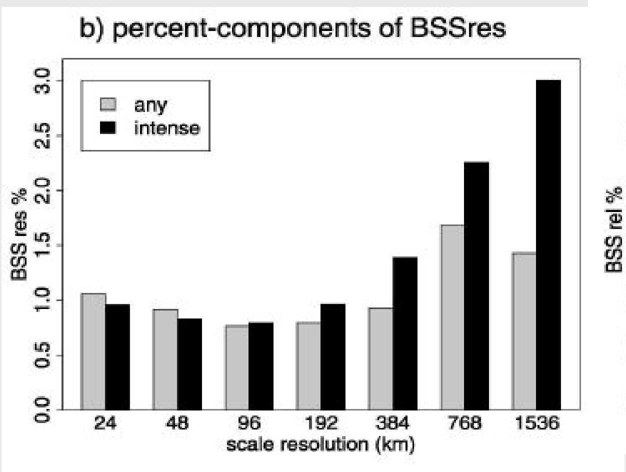
Casati and Wilson (2007) MWR 135

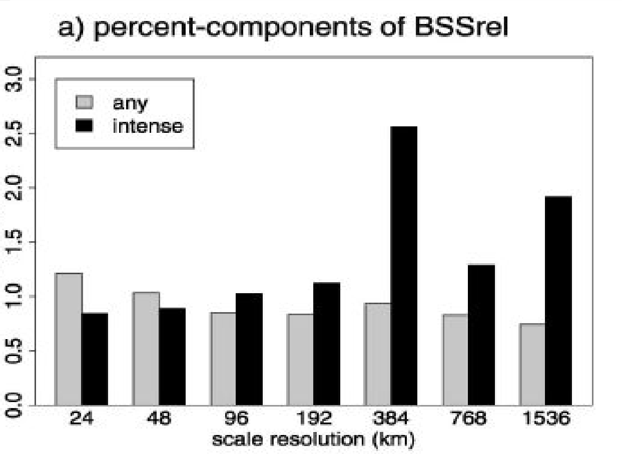










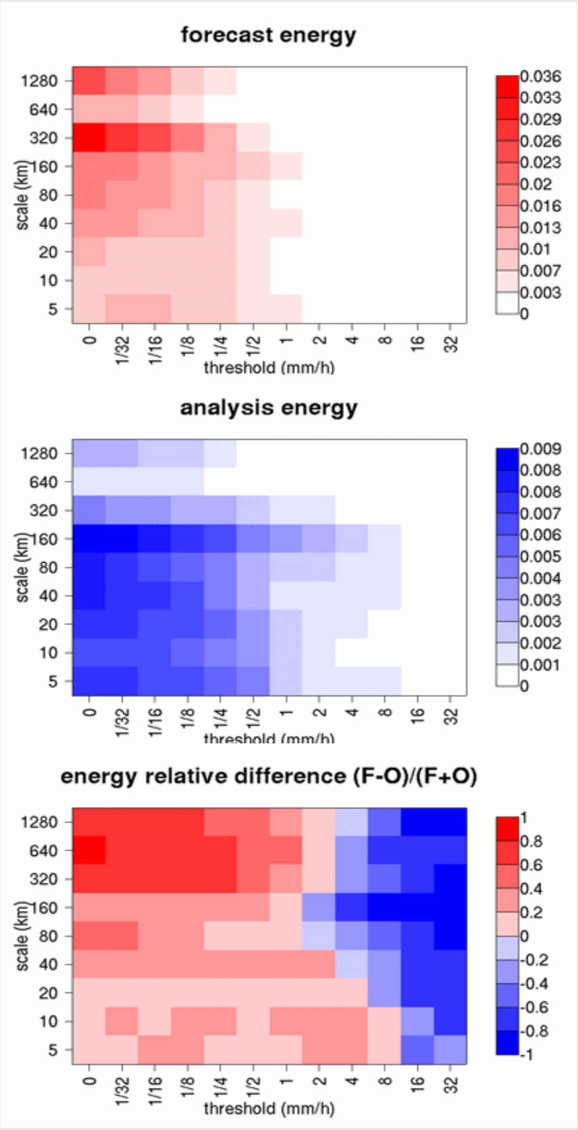


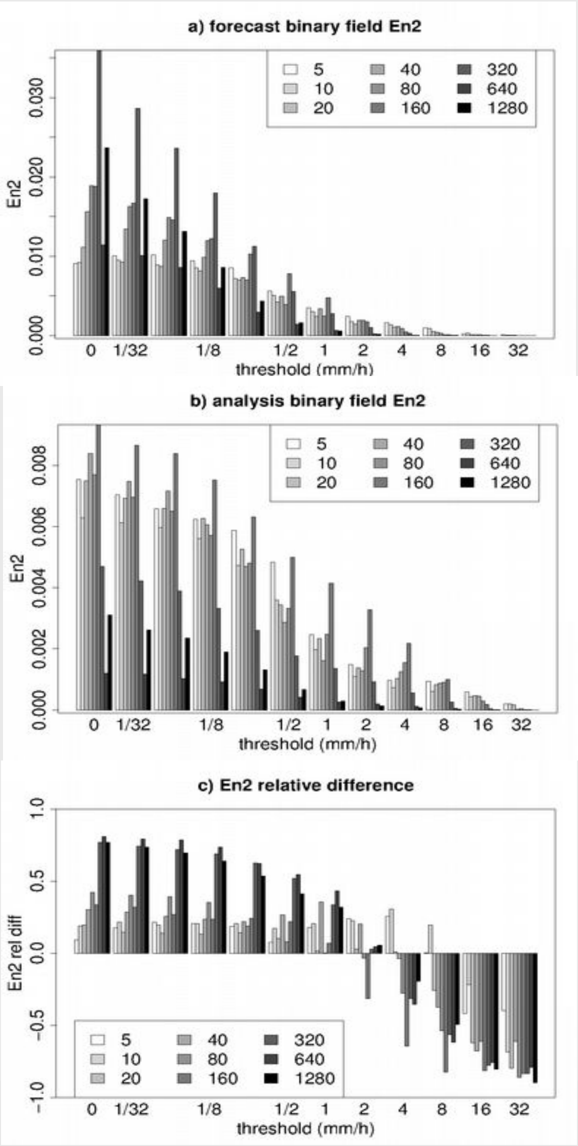
Skill on different scales: positive on large scales (> 700 km); negative on small scales (< 350 km); no-skill to skill transition scale ~ 500 km. **Bias on different scales**: overforecast of 400km feature (leads to poor skill). Assess scale structure.

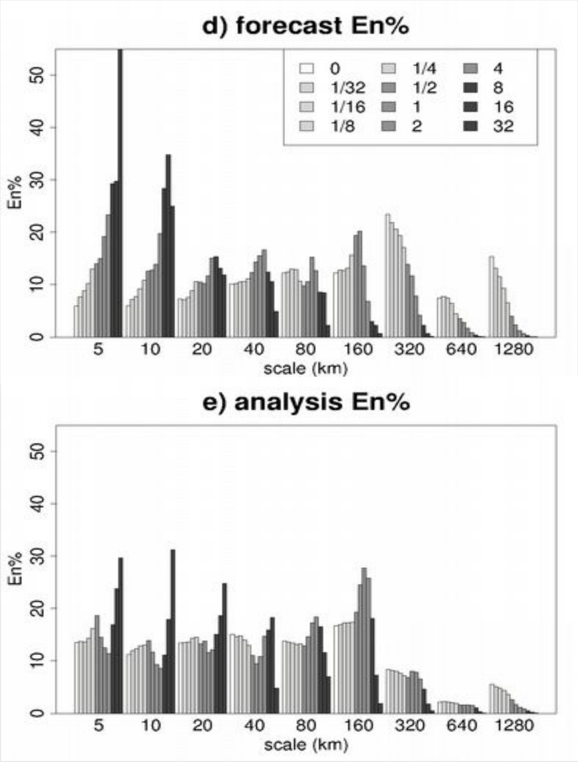
Quantify **contribution of different scales to reliability and resolution** components in the Brier Skill Score (BS=rel-res+unc, BSS=BSSres-BSSrel).

Intensity-scale verification technique

Casati (2010) Wea & For, vol. 25





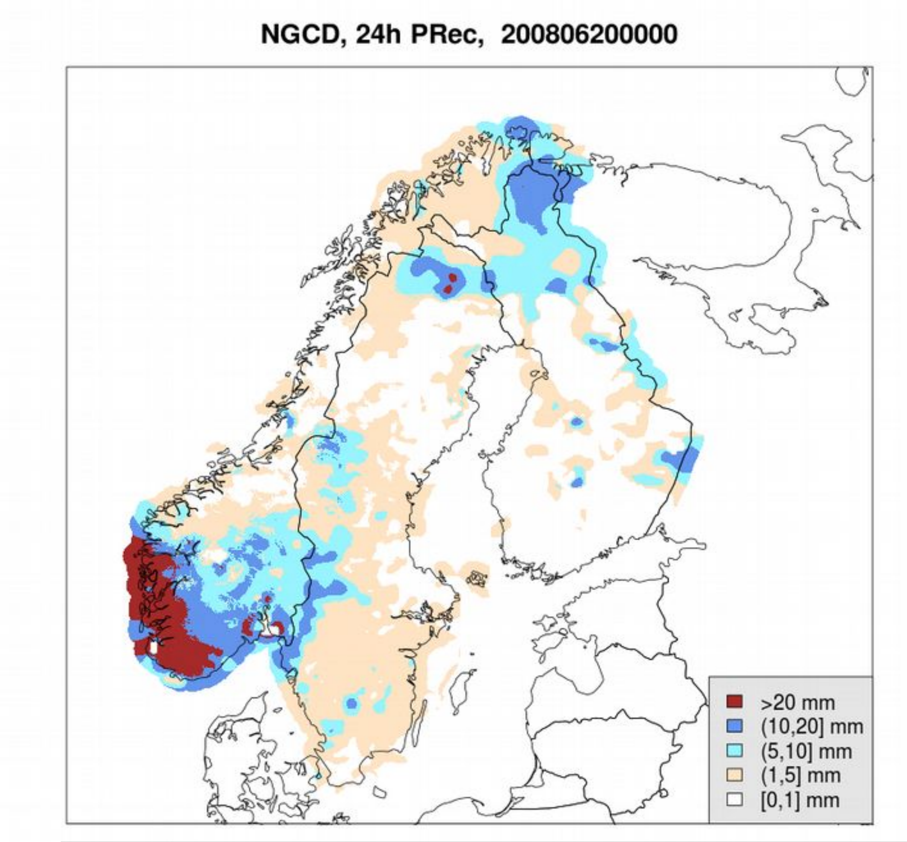


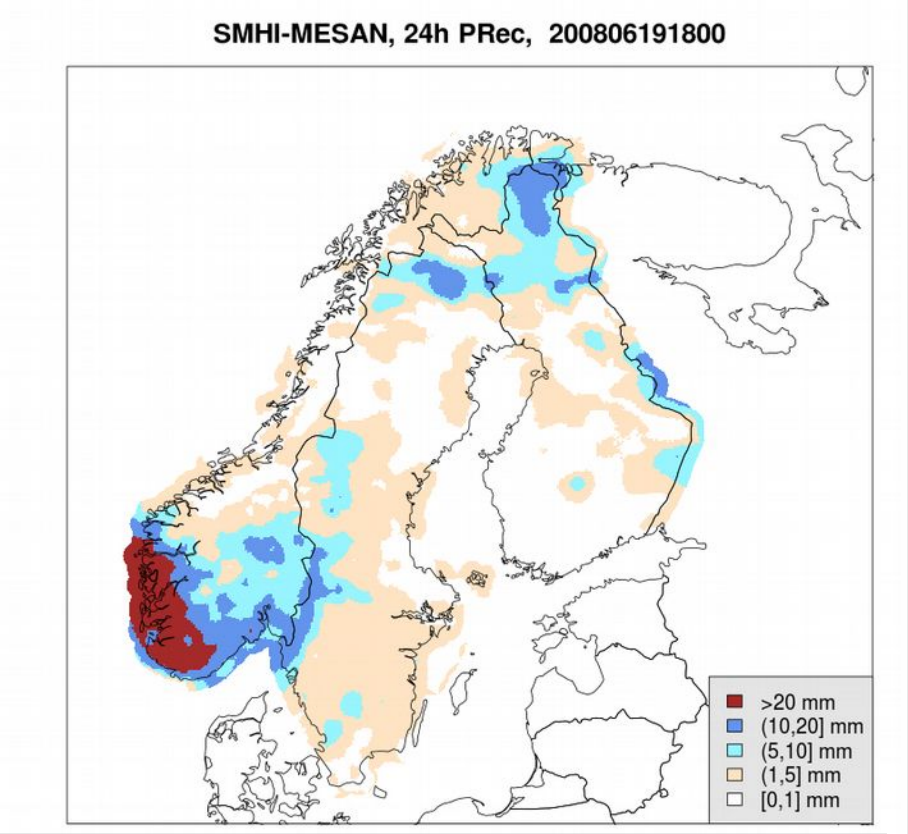
<u>For each threshold and scale:</u> **energy** informs on the amount of events; energy relative difference measures the **bias**; energy % assess the **scale structure**.

UERRA: Scale-Separation verification technique

- **RRA**: EURO4M MESAN data – European high-res surface reanalysis
- **Ref:** NGCD
- **Var:** RR, daily prec
- Time period considered: JJA 2008

http://exporter.nsc.liu.se/620eed0cb2c74c859f7d6db81742e114/



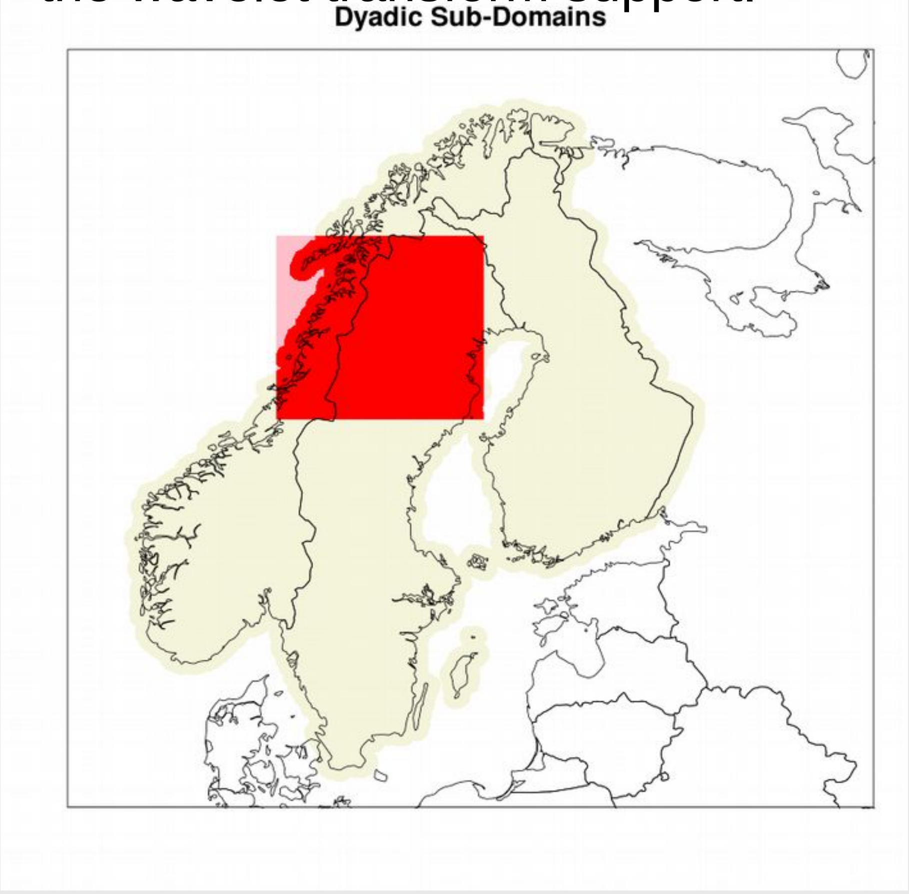


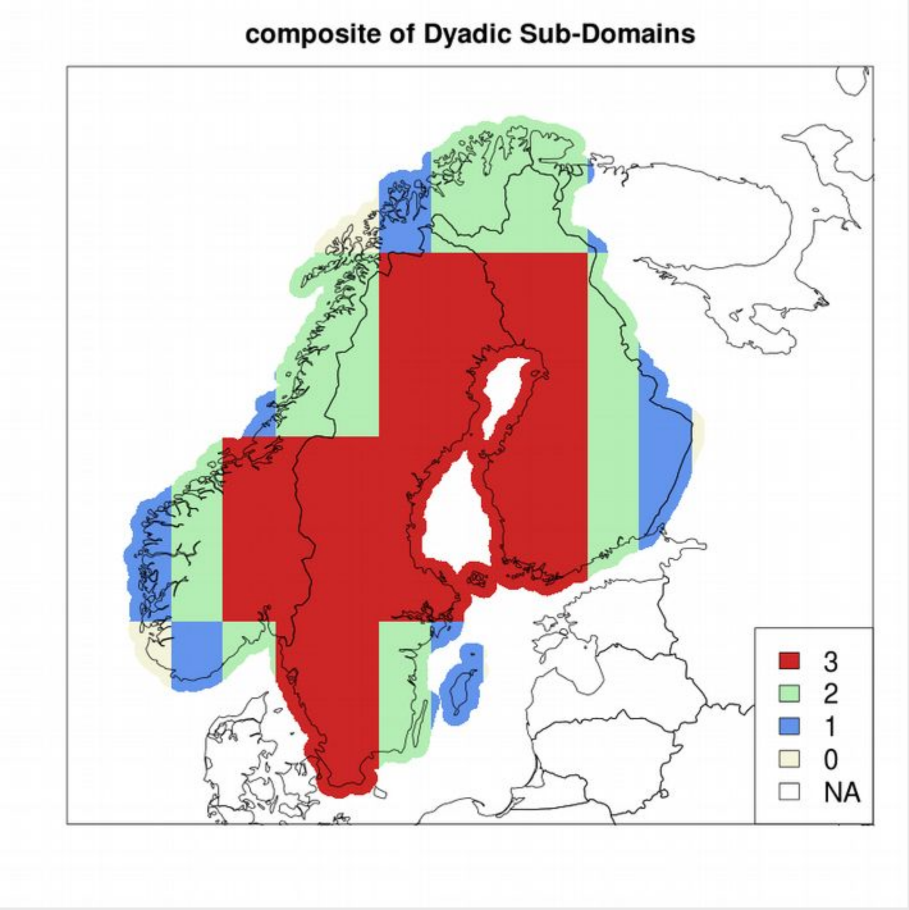
!! NGCD and SMHI-MESAN are both based on ECA&D data !!
VERIFICATION of NOT INDEPENDENT datasets

Change of Support problem: reprojection/regridding...

- SMHI-MESAN has been reprojected on the NGCD grid (nearest neighbor interpolation)
- Given that: NGCD covers land only & discrete wavelet transform need 2**n x 2**n domains (i.e. dyadic domain)

---> verification over several overlapping dyadic sub-domains 512 Km x 512 Km. *Tiling approach*: smoothes out the effect due to discreteness of the wavelet transform support.





case studies: RR for 1 day in summer 2008

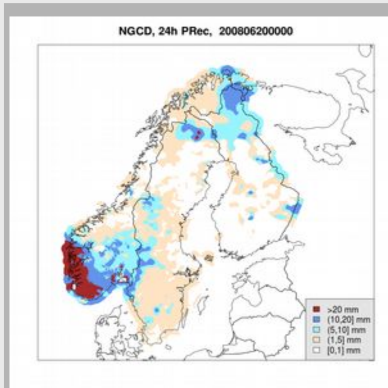
Energies are evaluated for each scale component

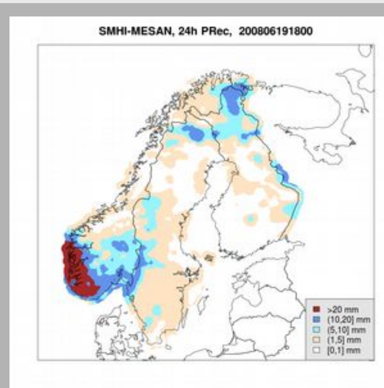
$$\mathrm{x} = \sum_{l=1}^{L} W_l^m(\mathrm{x}) + W_L^f(\mathrm{x})$$

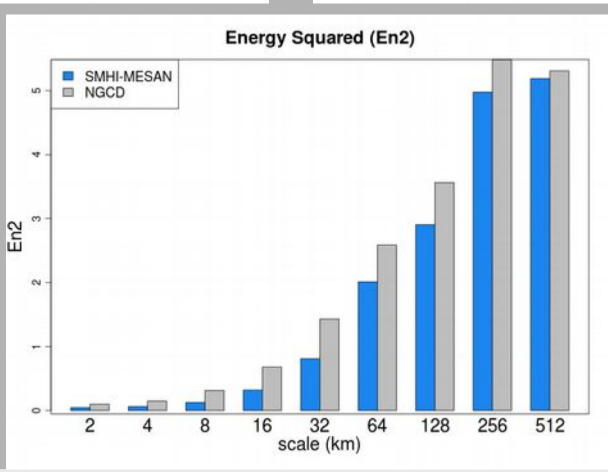
$$\mathrm{En}^2(\mathrm{x}(t)) = \overline{\mathrm{x}(t)^2}$$

$$\mathrm{En}^2(\mathrm{x}(t)) = \sum_{l=1}^{L} \mathrm{En}^2\left[W_l^m(\mathrm{x}(t))\right] + \mathrm{En}^2\left[W_L^f(\mathrm{x}(t))\right]$$

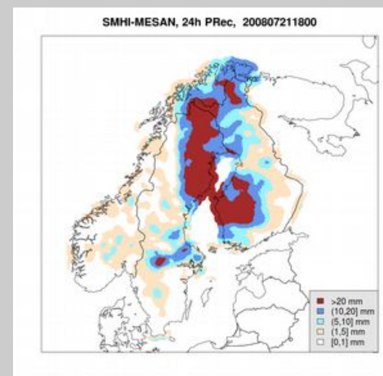
An intensity-scale skill score to assess the added value of enhanced resolution
B. Casati, A. Glazer, J. Milbrandt, and V. Fortin
http://presentations.copernicus.org/EMS2015-250_presentation.pdf

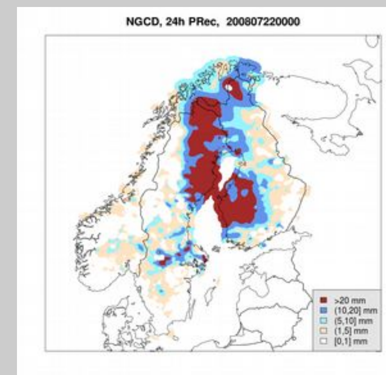




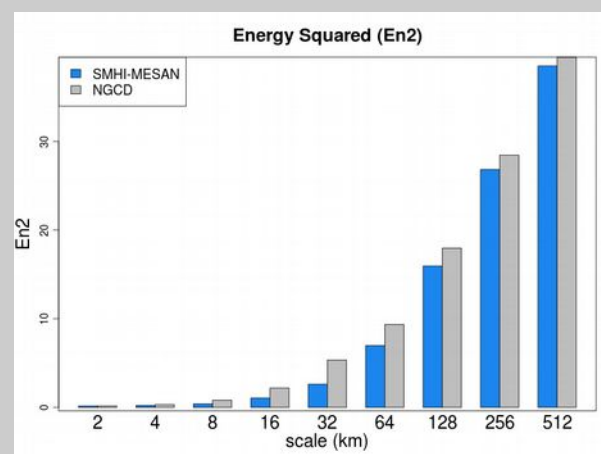


Case 1





Case 2

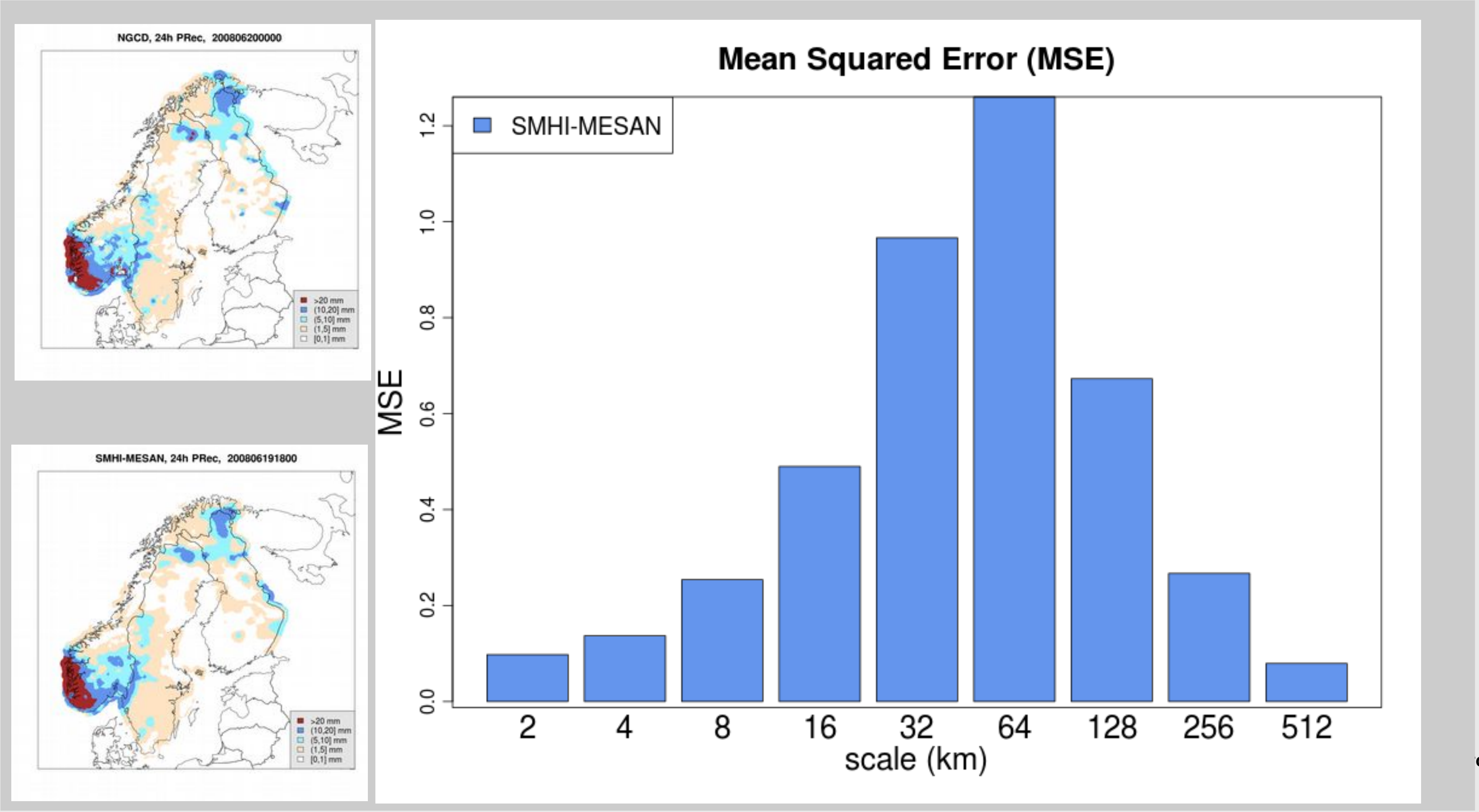


case studies: RR for 1 day in summer 2008

Mean Squared Error(MSE) for each scale component

$$\mathrm{MSE}\left(\mathrm{x}^{\mathrm{rra}},\mathrm{x}^{\mathrm{ref}}\right)\equiv\mathrm{En}^2\left(\mathrm{x}^{\mathrm{rra}}-\mathrm{x}^{\mathrm{ref}}\right)=\overline{\left(\mathrm{x}^{\mathrm{rra}}-\mathrm{x}^{\mathrm{ref}}\right)^2}$$

$$\mathrm{MSE}_l\left(\mathrm{x}^{\mathrm{rra}},\mathrm{x}^{\mathrm{ref}}\right)=\begin{cases}\left\langle\mathrm{En}^2\left[W_l^m\left(\mathrm{x}^{\mathrm{rra}}(t)-\mathrm{x}^{\mathrm{ref}}(t)\right)\right]\right\rangle & l=1,\ldots,L\\ \left\langle\mathrm{En}^2\left[W_L^f\left(\mathrm{x}^{\mathrm{rra}}(t)-\mathrm{x}^{\mathrm{ref}}(t)\right)\right]\right\rangle & l=L+1\end{cases}$$

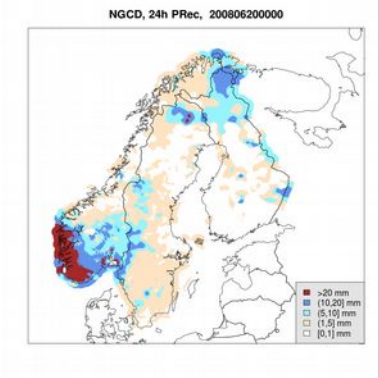


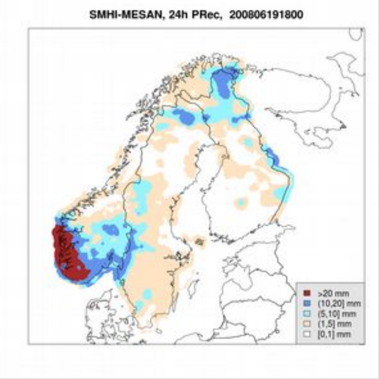
ological Institue

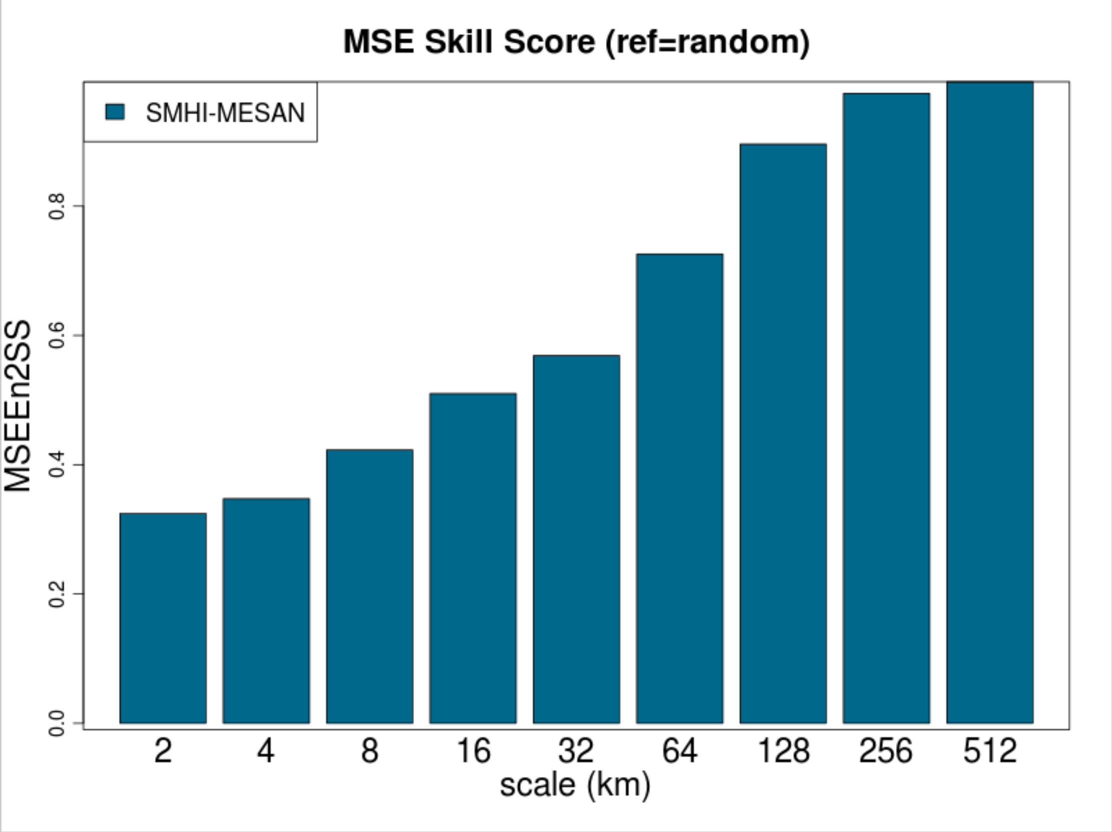
case studies: RR for 1 day in summer 2008

MSE skill-score for each scale component

$$SS_l \equiv \frac{\mathrm{MSE}_l\left(\mathrm{x}^{\mathrm{rra}},\mathrm{x}^{\mathrm{ref}}\right) - \left[\mathrm{MSE}_l\left(\mathrm{x}^{\mathrm{rra}},\mathrm{x}^{\mathrm{ref}}\right)\right]_{\mathrm{random}}}{\left[\mathrm{MSE}_l\left(\mathrm{x}^{\mathrm{rra}},\mathrm{x}^{\mathrm{ref}}\right)\right]_{\mathrm{best}} - \left[\mathrm{MSE}_l\left(\mathrm{x}^{\mathrm{rra}},\mathrm{x}^{\mathrm{ref}}\right)\right]_{\mathrm{random}}}$$

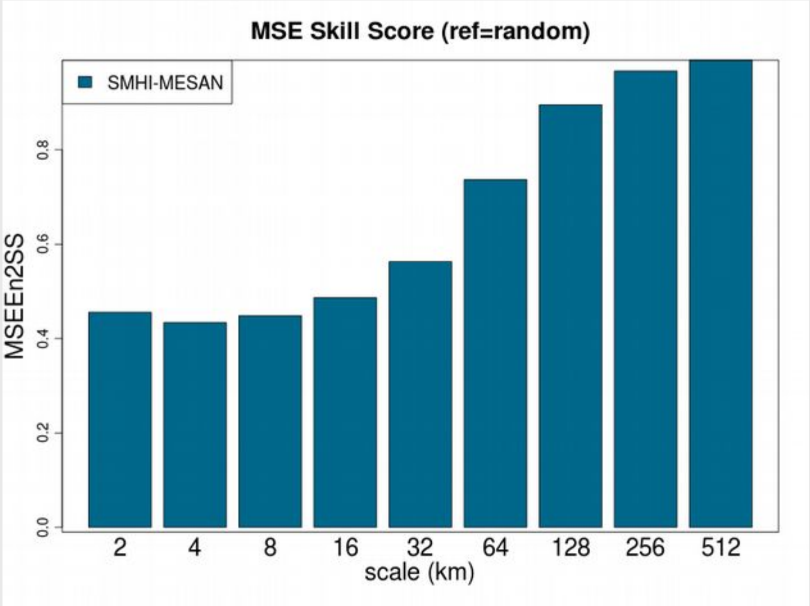


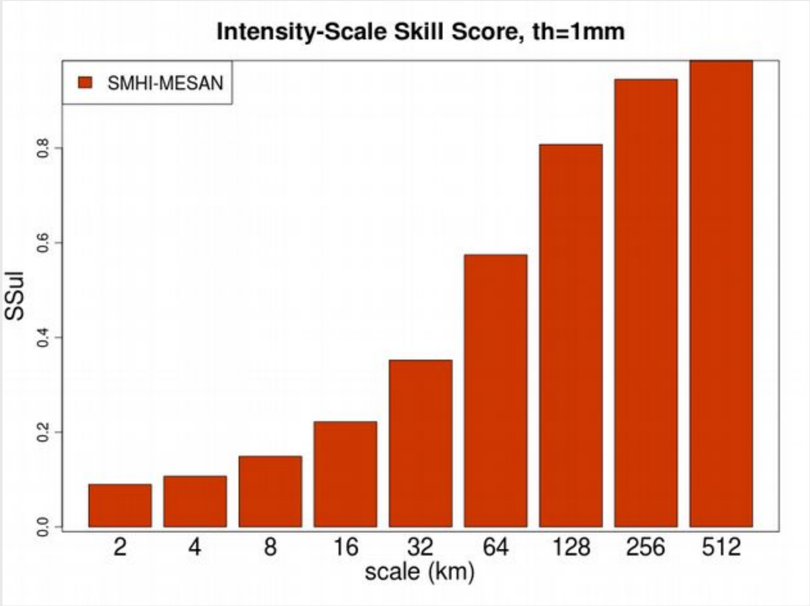


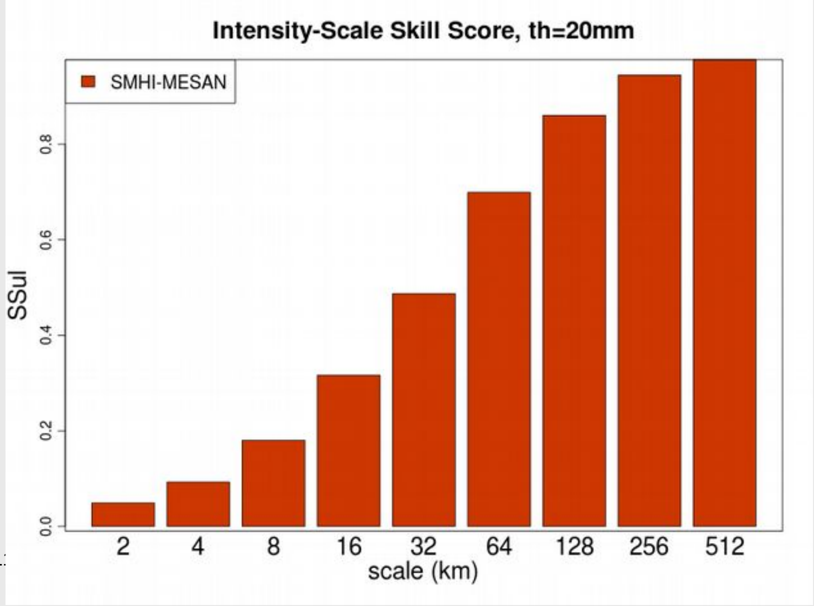


Scale-Separation verification technique

aggregation over JJA 2008







Conclusions. Wavelet based scale-separation MSE skill-score and scale-separation statistics are:
- informative on RRA bias, error and skill on different scales
- suitable for comparing models with different resolutions
- RRA performances for specific intensity events (thresholded binary fields)

events (thresholded binary fields)

UERRA-EVA: Evaluation software tools
https://github.com/UERRA-EVA

a. Assess the added value of enhanced resolution in RRA:
wavelet-based scale-separation MSE skill score
→ *to be included in EVA_gridobs*
b. Scale-decomposition of the Brier score for for the verification of probabilistic RRA...

EVA_gridobs

UERRA common evaluation procedure: assessing uncertainties in reanalysis by evaluation against gridded observational datasets

Reanalysis and gridded observations are assumed to be: on the same coordinate reference system and grid; same temporal aggregation.

Tip: use fimex for regridding (https://wiki.met.no/fimex/start)

List of Skill-Scores/Tests

Developed and tested for daily precipitation

- [x] Probability Density Functions (PDFs) related skill-scores (PDFs are approximated by comparing discrete histograms):
  - [x] difference between modes: mode(reanalysis) - mode(observation)
  - [x] relative precipitation biases = (mode(reanalysis) / mode(obsevration) - 1) * 100%
  - [x] overlapping skill-score (see [1], Eq.(1))
- [x] two-sample Kolmogorov-Smirnov (K-S) test, or Smirnov test (see [2], Eqs (5.17-18)
- [ ] Fractional skill-score
- [ ] Optical-flow

[1] Mayer, S. et al. (2015). Identifying added value in high-resolution climate simulations over Scandinavia. Tellus A

[2] Wilks, D. S. (2011). Statistical methods in the atmospheric sciences (Vol. 100). Academic press.

Thanks for your attention!

cristianl@met.no