Precipitation extremes in NWP :

- some results from spatial verification in DMI -

Bent Hansen Sass, DMI, September 2021

1. **Motivation for studying \`extremes´**

2. **Monitoring the properties of a LAM NWP system to forecast extreme precipitation : prerequisites, dilemmas and requirements.**

3: **A method for monitoring the forecast quality of high (extreme) precipitation events .**

4: **An example and verification results for precipitation forecasts in Denmark**

5. **Conclusions and outlook**

Motivation for verifying extremes :

A large focus on extremes today, both in climate prediction and in daily weather forecasts, e.g. due to

- **potential large damage from flooding due to extreme precipitation events and related loss of life and properties**

  Example: Huge amount og precipitation resulted in severe floodings 14 July 2021 , affecting Germany, Belgium, Luxembourg and the Netherlands

- **large economical impacts!**

Traditional verification schemes do not verify extremes as a part of daily forecasting !

Then it may be asked <u>how to best measure and monitor the ability of NWP to forecast `extreme´ precipitation.</u>

Precipitation extremes :
Prerequisites and dilemmas for LAMs

Prerequisites:

0: A prerequisite for studying precipitation extremes is the availability of a detailed precipitation analysis (e.g. using radar data) to be compared with a corresponding model forecast.

Dilemmas :

1: `True´ extremes are rare phenomena. It may be necessary to include long verification periods to estimate the model's ability to forecast extremes from a statistical point of view.

2: How large values are needed to qualify as an extreme ? This depends on the application or may be determined in terms of deviation from a `climatology´. If high thresholds are chosen there may be long periods in forecasting without extremes .

3: Should extremes be searched for with some tolerance, e.g. points in a domain withing few percent of the absoute maximum ?

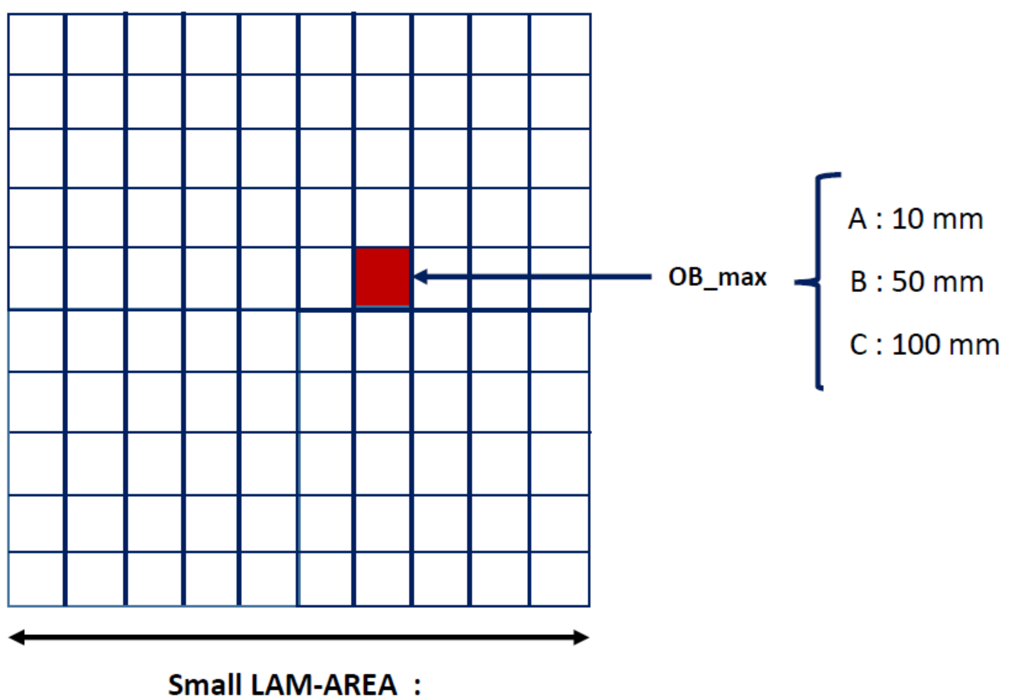
Precipitation extremes : Requirements

1: It is necessary to consider spatial and temporal aspects of predicting the `extremes´ (neighborhood considerations)

2: Verification must be able to diagnose very clearly the extreme precipitation of a relative small subdomain (figure below): Measures such as traditional *bias* and *mae* over the model domain will mask (hide) the ability to forecast extremes and are not adequate in the context of extremes.

Example:

Observed precipitation area (in red) :



Precipitation extremes : Example

LAM-AREA :

Dimension e.g. 500 km * 500 km

Precipitation area : 50*50 km with indicated accumulation
Forecasted precipitation = 0 over the ENTIRE GRID , i.e.
the meso-scale precipitation area of max precipitation is missed
Computing traditional grid point based *bias* and *mae* gives:

A: *bias* = - 0.10 kg/m^2, *mae* = 0.10 kg/m^2

B: *bias* = - 0.50 kg/m^2, *mae* = 0.50 kg/m^2

C: *bias* = - 1.00 kg/m^2, *mae* = 1.00 kg/m^2

The small/modest numbers do NOT give an impression of a serious precipitation error over part of the model grid.

The proposed method of computing largest occurring *bias* and *mae* occurring in a sub-area gives

A: *bias* = - 10 kg/m^2, *mae* = 10 kg/m^2

B: *bias* = - 50 kg/m^2, *mae* = 50 kg/m^2

C: *bias* = - 100 kg/m^2, *mae* = 100 kg/m^2

Considering 4 equally sized sub-areas the red area would be 1/25 = 4 % of the sub-domain with size 250 km *250 km

Score documented in :

” A scheme for verifying the spatial structure of extremes in numerical weather prediction: exemplified for precipitation” Link: http://dx.doi.org/10.1002/met.2015

Implementation in DMI based on:

- **Operational spatial analysis, based on radar data and in situ data**
- **Spatial analysis used in combination with HARMONIE-AROME precipitation for the same time intervals, e.g. 3 hour intervals and same spatial resolution**
- **Operated in DMI’s implementation of *harp* .**

The present verification scheme: computes for every forecast its ability to forecast the highest value(s) analysed, becoming the `extreme´ value of the day.

The advantage of this approach is that an assessment in terms of a score can be computed for all forecasts.

A verification of `true extremes´ can be done as a `post processing´ if a computed score is stored together with the actual extremes occurring on that day.

When considering accumulated information over many forecasts the statistics of the score may be computed as a function of the `extreme´ values involved.

Minima are included in the assessment since minima in some cases dominated by (large) precipitation amounts appear as `extremes´ in the context of the actual forecast.

The nomenclature is:

The following four scores are computed for points qualifying as `extremes´:

SLX_{ob_max} : How well does forecast maximum match observed maximum in its neighbourhood ?

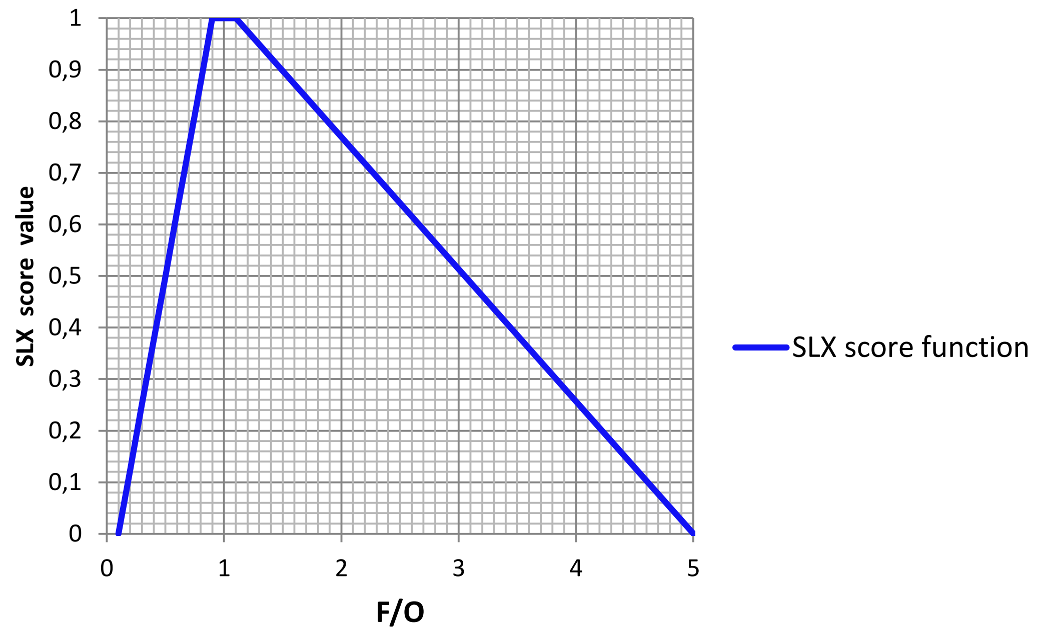
SLX_{ob_min} : How well does forecast minimum match observed minimum in its neighbourhood ?

SLX_{fc_max} : How well does observed maximum match forecasted maximum in its neighbourhood ?

SLX_{fc_min} : How well does observed minimum match forecasted minimum in its neighbourhood ?

$SLX_{\_ave}$: Combined average score = ¼ (SLX_{ob_max} + SLX_{ob_min} + SLX_{fc_max} + SLX_{fc_min}) .

Score function S



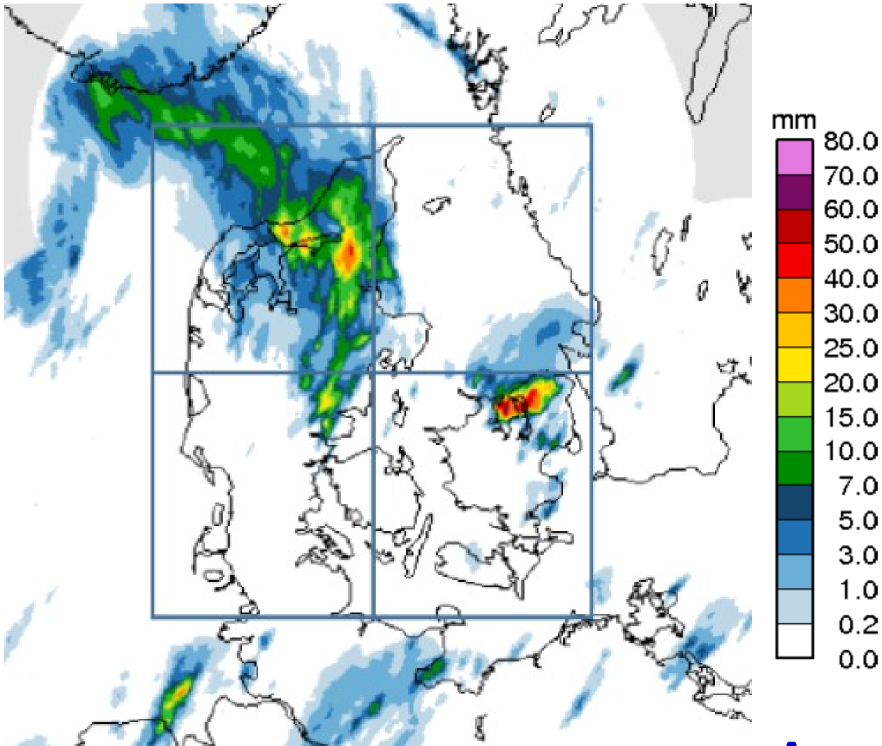
Individual scores are computed as a function S(R) of the ratio R =F/O between forecast (F) and observation (O).
$0 \leq S \leq 1$.

Highest score (perfect match between forecast and observation gives S =1)
Lowest score (poorest match between forecast and observation gives S=0)
Special treatment for O or F close to zero (perfect score assumed for O and F both <= threshold ~ 0.2 Kg/m^2

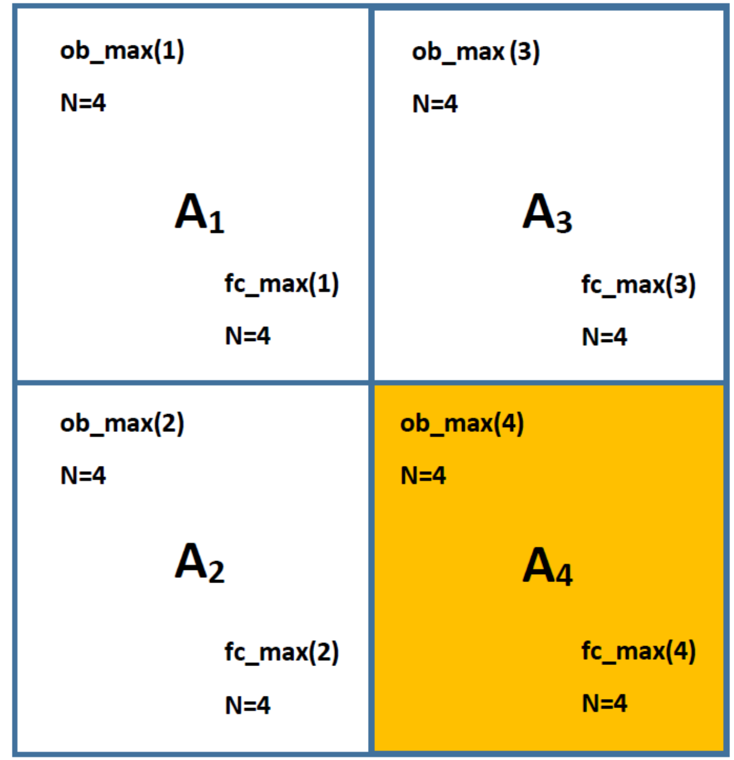
All points qualifying as maxima and minima respectively contribute with
equal weight to a score computation of the fields involved (Separate scores for maxima and minima)

- Consider score computation in a number of subareas, and store additionally the maximum values (forecast and observation) in the subarea with highest maxima (forecasted or observed).

A method for monitoring the forecast quality of extreme precipitation events



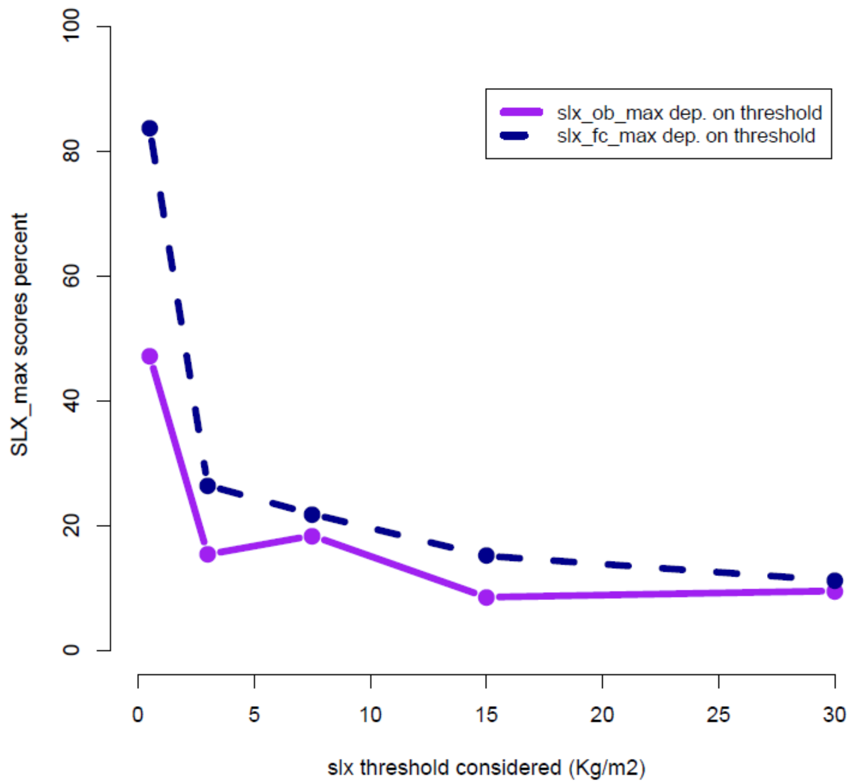
A



B

Verification setup: A: Example of analyzed 3h acc. precipitation 2021-08-07, 15 UTC .
DMI implementation is currently on a 750m* 750m using 4 subareas
B : ob_max(i) is observed maximum of subarea i , fc_max (i) is forecast maximum of subarea i.
Store values in area with largest maximum (observed or analyzed) , e.g. area A_4

Verification of HARMONIE-AROME at DMI
(Scores of maxima as a function of the precipitation thresholds involved)



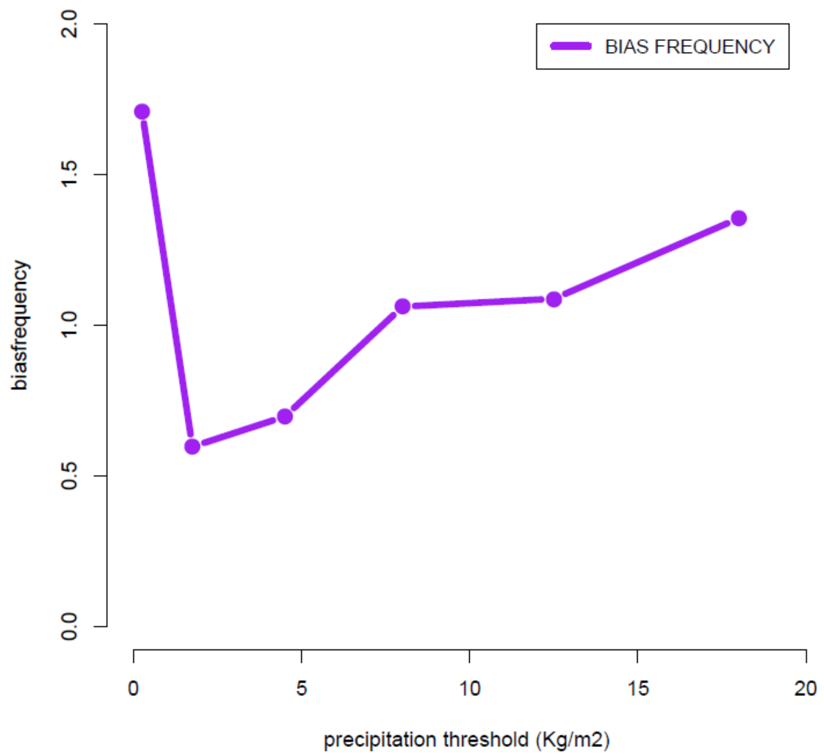
- SLX_{ob_max} is computed statistically (purple curve) as a function of observed precipitation threshold in areas with selected maxima (all forecasts used in May – July 2021). Similarly SLX_{fc_max} (dashed blue curve) is computed statistically as a function of the stored precipitation maximum.

- As expected: Lower scores related with maxima tend to occur for high precipitation max-values, observed – or forecasted, that is : It is increasingly difficult to get high slx- scores if larger maximum precipitation values (observed or forecasted) are involved. All 3h forecasts are considered.

APPLICATION :

This type of plot may be useful to reveal which model version is best at predicting extreme precipitation when curves are computed separately for each model.

Verification of HARMONIE-AROME – NEA at DMI
(Frequency bias of stored maximum precipitation)

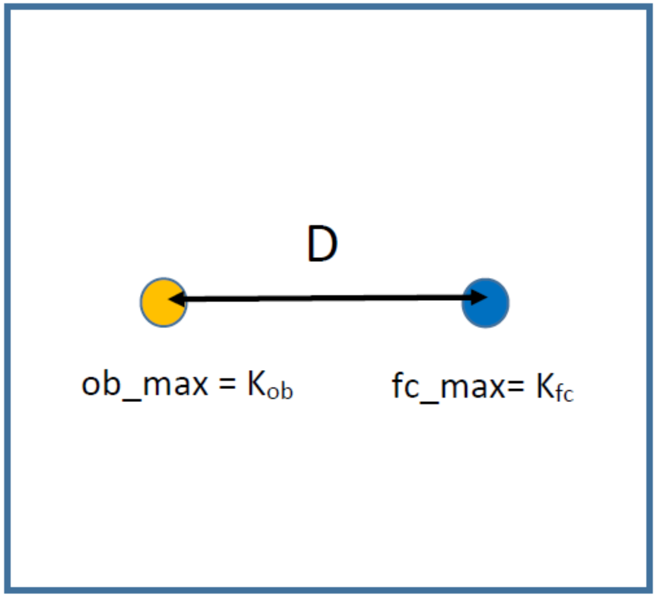


Frequency bias of the forecasted versus observed 3h max-precipitation occuring in one of the four sub-domains. Period : May-July 2021

Cloud burst example and verification results for Denmark.

Results are based on HARMONIE-AROME in DMI at a grid size of 2.5 km, operated every 3 hours.

When predicting localized strong convection it is relevant to consider how the scores behave in the case of a misplaced `shower´ in an otherwise precipitation free environment:

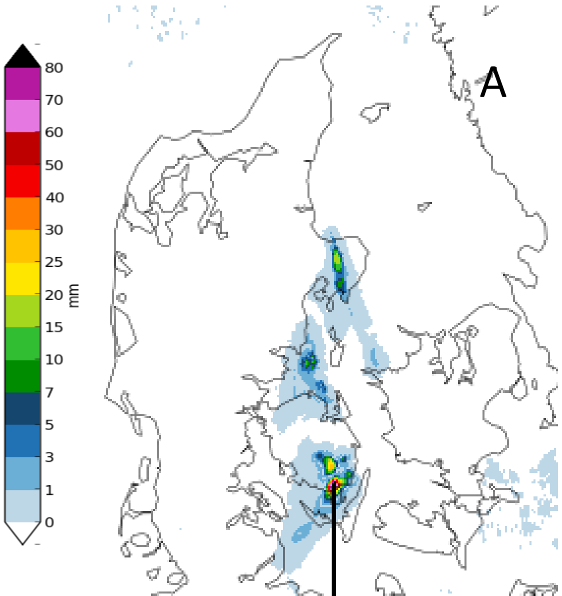


If displacement D between observed and forecasted maxima = 0 and $K_{ob} = K_{fc}$ then $SLX_{ob\_max} = SLX_{fc\_max} = 1$ (perfect score).

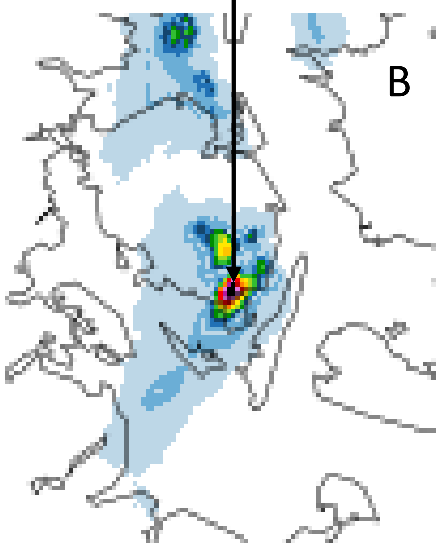
The neighborhood width for these type of fields needs to be ≥ D in order to get perfect score.

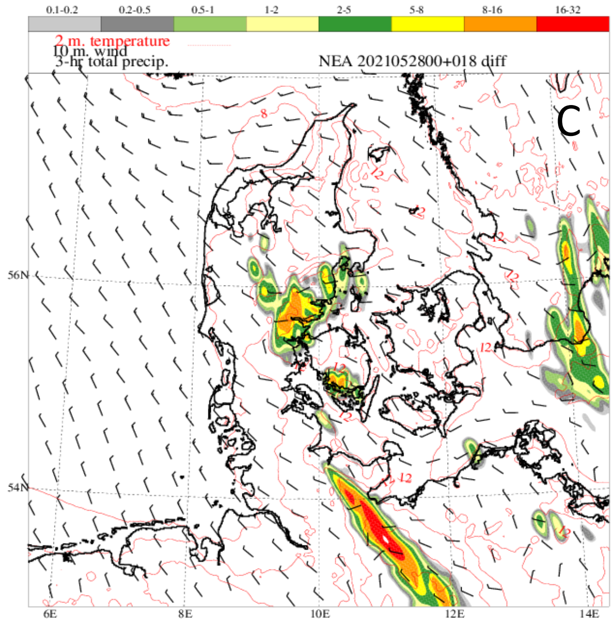
In general the ratio between K_{ob} and K_{fc} determines the size of the score. Values much different from 1 means low scores.

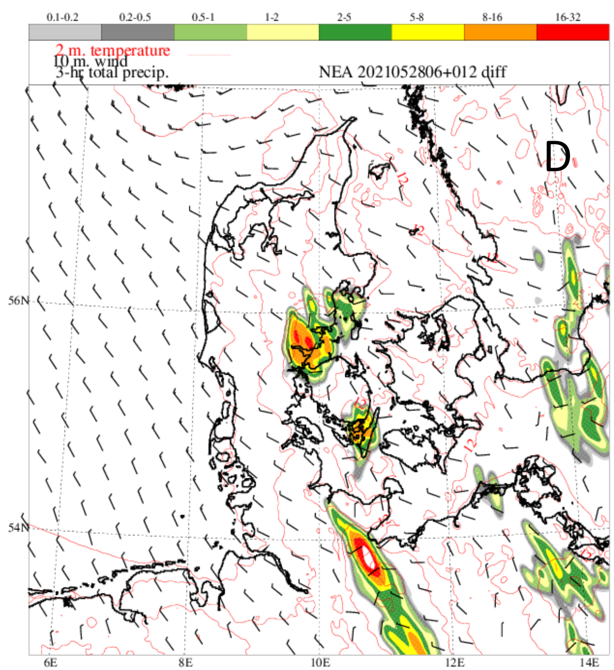
Cloud burst example:



Analyzed 3h accumulated precipitation 28/5 18 UTC 2021







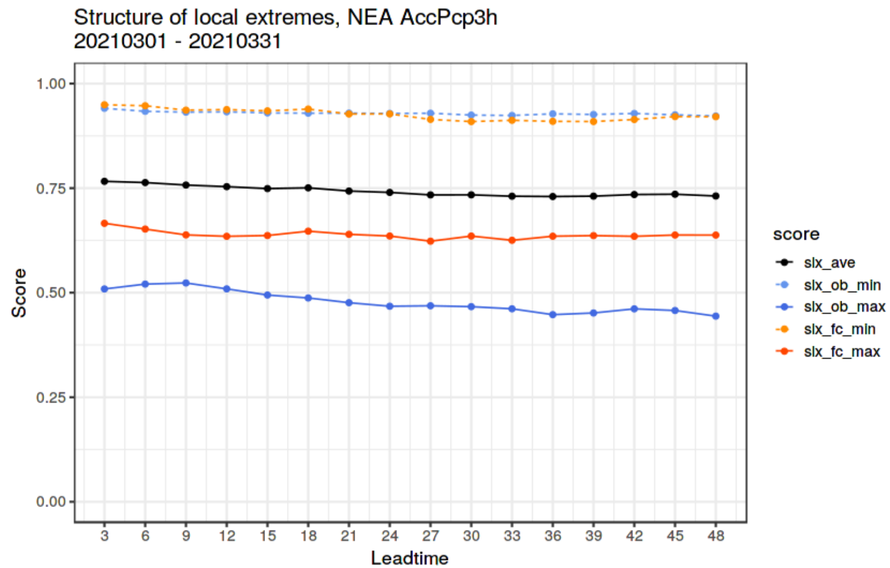
An extreme cloud burst episode happened on 28 May 2021 in southern part of Denmark (location near the city of Svendborg). The accumulated precipitaion in an area of size less than 10*10 km was up to ~ 80 Kg/m^2 in 3 hours , between 15 and 18 UTC, see A and B. The convection evolved in a convectively unstable atmosphere, with week surface winds and some warm air advection in the lower troposphere. Figure C shows the 3 hour accumulated precipitation at 18 hours, with initial time 00 UTC. In D is shown the corresponding result of a 12 hour forecast from 6 UTC .

At first sight both forecasts appear reasonable since they represent the observed cloud burst in the corect region of the country. However the accumulation is about 15-20 % of the observed one. The forecast from 00 UTC has some localization error. The SLX scores related with maxima are therefore small using the score SLX function considering only value of extreme. With neighborhood size =0 the score is zero at some forecast origine times.

Note the " false alarms " of the forecast maxima in eastern parts of Jutland occuring at locations with no observed precipitation.

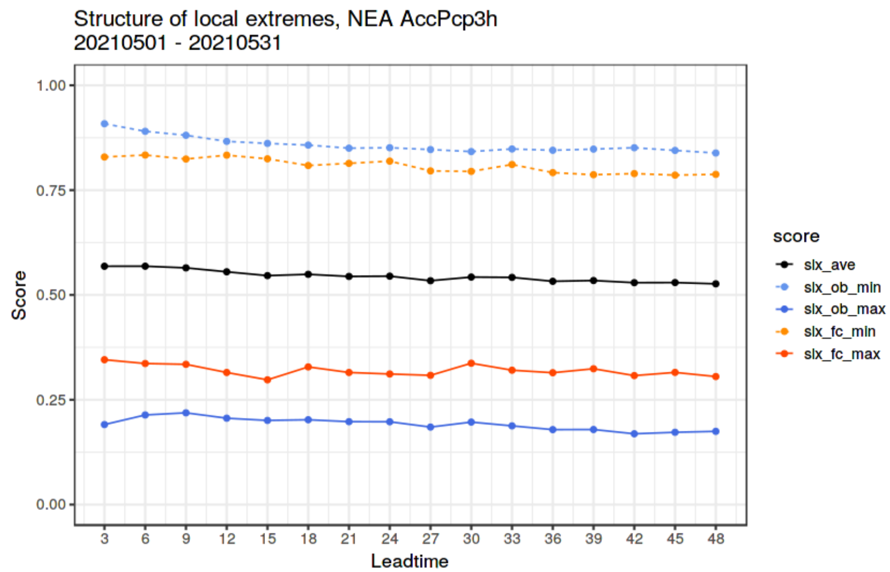
Conclusion: Small scale cloud burst events are very challenging !
The SLX scheme provides a low score as it should in case of inaccurate precipitation amount and location.

Comparing SLX components for different periods: Substantial difference between March 2021 and May 2021.



Verification results of all SLX-score components during **March 2021**. All components and the average (slx_ave) are computed according to forecast lenght. Average scores are computed using all forecasts available during the month.

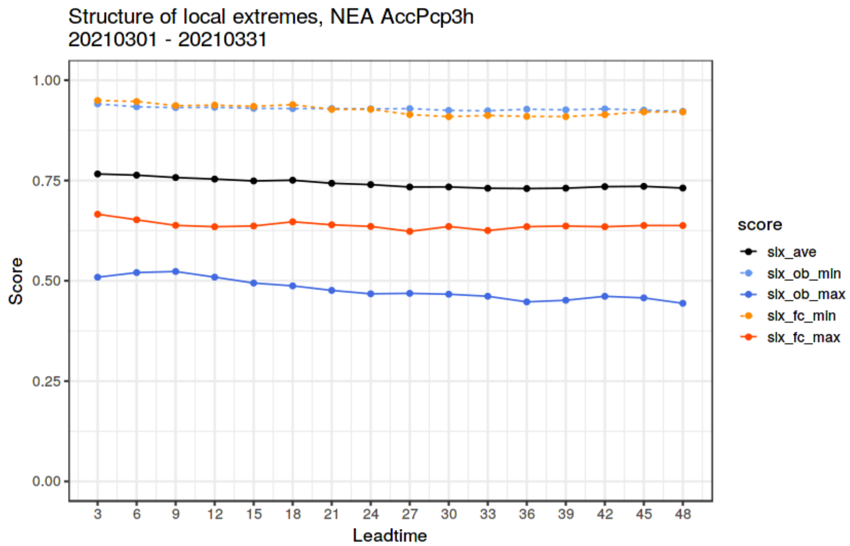
(Neighborhood size = 0)
[points used in comparison =1 point]



Verification results of all SLX-score components during **May 2021**. All components and the average (slx_ave) are computed according to forecast lenght. Average scores are computed using all forecasts available during the month.

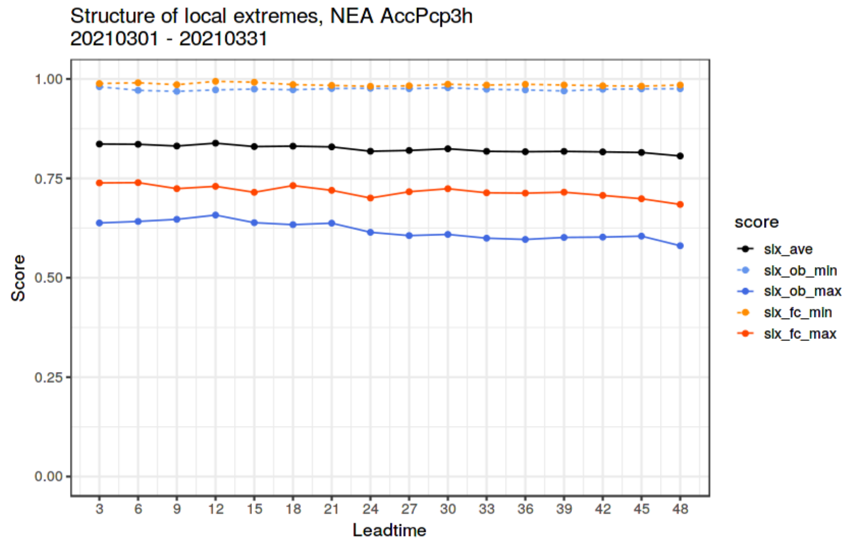
(Neighborhood size = 0)
[points used in comparison =1 point]

Impact of neighborhood treatment (Period: March 2021).



Verification results of all SLX-score components during **March 2021**. All components and the average (slx_ave) are computed according to forecast lenght. Average scores are computed using all forecasts available during the month.

(Neighborhood width = 0 points)
[points used in comparison =1 point]



Verification results of all SLX-score components during **March 2021**. All components and the average (slx_ave) are computed according to forecast lenght. Average scores are computed using all forecasts available during the month.

(Neighborhood width = 15 points)
[points used in comparison = (15+1) *(15+1) points].

Conclusions and outlook

- **Operational use of SLX scheme in combination with storage of maxima in subareas enables conditional verification of scores depending on the level of maxima. When studying the impact of a new model version it is possible to compare if large values of extremes are predicted with higher scores.**

- **A further level of detail may be obtained by studying histograms of SLX components or alternatively absolute errors of forecasts. This provides insight to the fraction of the largest errors occurring in a given model version**

- **Generalization of SLX to ensembles is rather straight forward, e.g. the histograms will change.**

Contact information and References

Contact:

Bent Hansen Sass, Danish Meteorological Institute , Lyngbyvej 100, 2100 , Copenhagen ,
E-mail: bhs@dmi.dk , Tel. +45 50 93 38 23

Acknowledgments :
Colleagues at DMI : Henrik Feddersen for collaboration on code implementation , Rashpal Gill and the radar group for implementation precipitation analysis and making data available. Bjarne Amstrup for collaboration on data from cloud burst case.

References

Sass, Bent H., (2021) : A scheme for verifying the spatial structure of extremes in numerical weather prediction: exemplified for precipitation, *Meteorological Applications* , **http://dx.doi.org/10.1002/met.2015**

Sass, B.H. 2020: Forecasting spatial structure of local precipitation extremes : International Verification Methods Workshop Online (2020-IVMW-O), Nov.2020, https://jwgfvr.univie.ac.at

Sass, B.H. and Yang, X. 2012: A verification score for high resolution NWP: Idealized and preoperational tests. HIRLAM Tech,Rep. No 69, 28 pp , Dec. 2012 , [Available from www.hirlam.org]