Research in
verification methods for high resolution NWP

A spatial verification scheme based on analyzing local extremes.

Bent Hansen Sass
1 October 2018

CONTENTS

1) Background for introducing more spatial verification

2) Verification scheme analyzing local extremes

3) Results of idealized tests

4) Concluding remarks

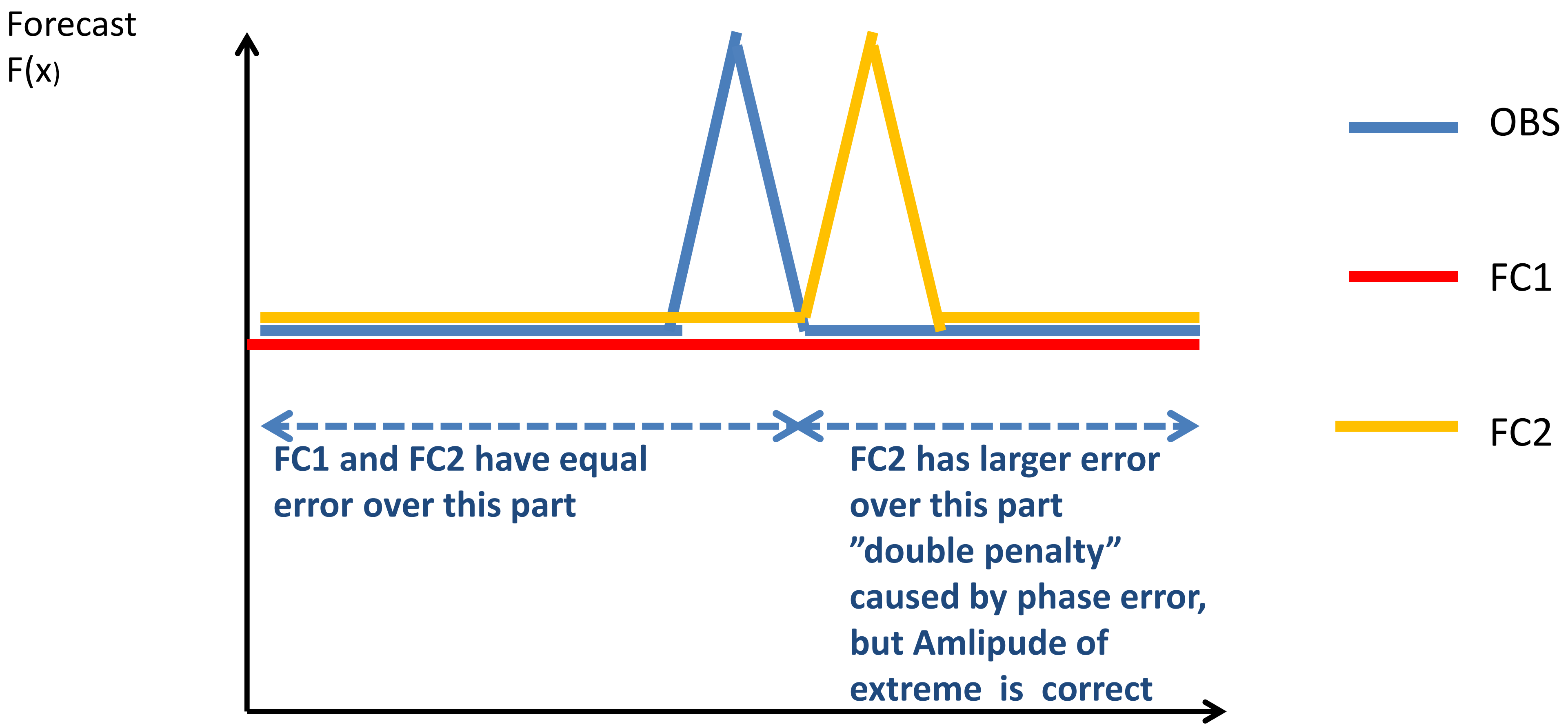
BACKGROUND

It is generally accepted that

- Limitations of verification schemes due to `Double penalty´, in space (and time) should be taken into account in high resolution NWP.

- Previous reasonable balance in 1980s -1990s between number of observation points and model grid points in verification based on point observations has been offset by recent big model resolution increases in the NWP community which makes verification of entire fields a natural choice!

Illustration of `double penalty´ issue



Independent space variable X could alternatively be replaced by time t,
FC2 gets no `reward´ for predicting the amplitude of the fluctuation correctly !

Thoughts behind a new spatial verification scheme in NWP

- Would it be appropriate to construct a scheme which is quite suitable to both model developers and users of NWP compared with most traditional schemes which are often somewhat biased towards the thinking of model developers ?

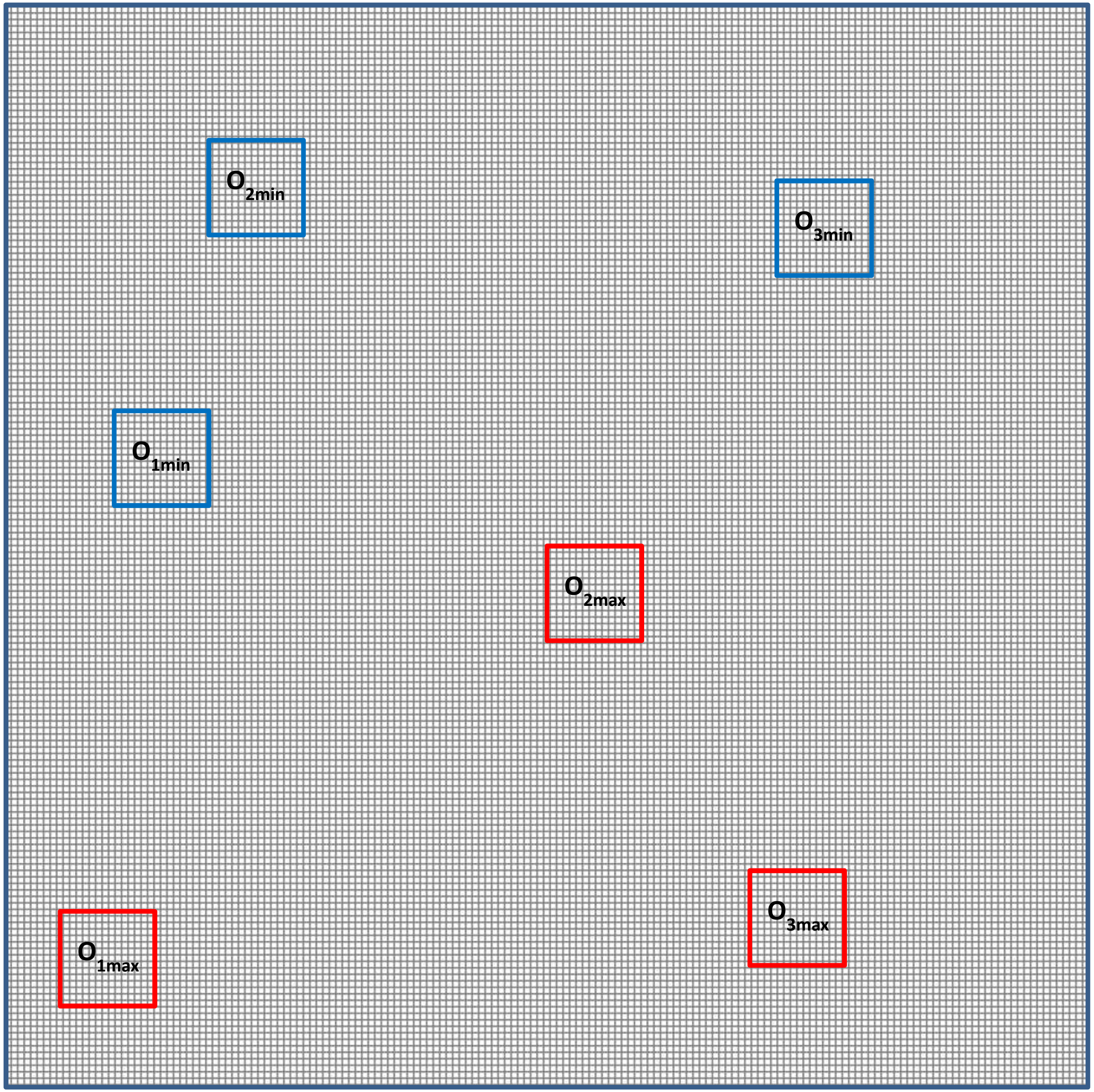
- Meteorological (duty) forecasters seem to have a strong need to focus on predicting extremes (high or low values) of some parameter, and in this context where and when such values occur. To help verifying model behavior from this perspective it seems desirable to design a spatial verification scheme which is able to verify model's ability to forecast local extremes.

- Would it be reasonable to make a scheme general to allow that a basic score function used in the verification can be chosen to differ from a default function if the application in some context indicates such need ?

Decision on design requirements

The new verifcation scheme should

- be based on spatial verification in order to treat double penalty issue
- diagnose quality based on an analysis of how well forecast and analysis match in areas of local extremes.
- assign a value between 0 and 1 measuring quality of any forecast field compared with an analysis field, 0 being poorest assignment and 1 representing a perfect forecast.
- in the first implementation focus on verifying precipitation
- be able to *verify any possible fields of forecast and analysis*. This implies, - unlike Fractions Skill Score, that zero fields of both analysis and forecast will be handled as fields that match perfectly in a default setup. Random points are chosen for verification in the case of entire fields being zero (In R-programming this is done by means of a ”sample”-command).



GENERALIZATION of a related scheme developed in DMI

An early version of the present idea for spatial verification named

SWS (Significant Weather Score)

has been used operationally for several years: It is based on highest and lowest values of synoptic surface observations and not on analyzed field (SWS scheme by Sass and Yang 2012)

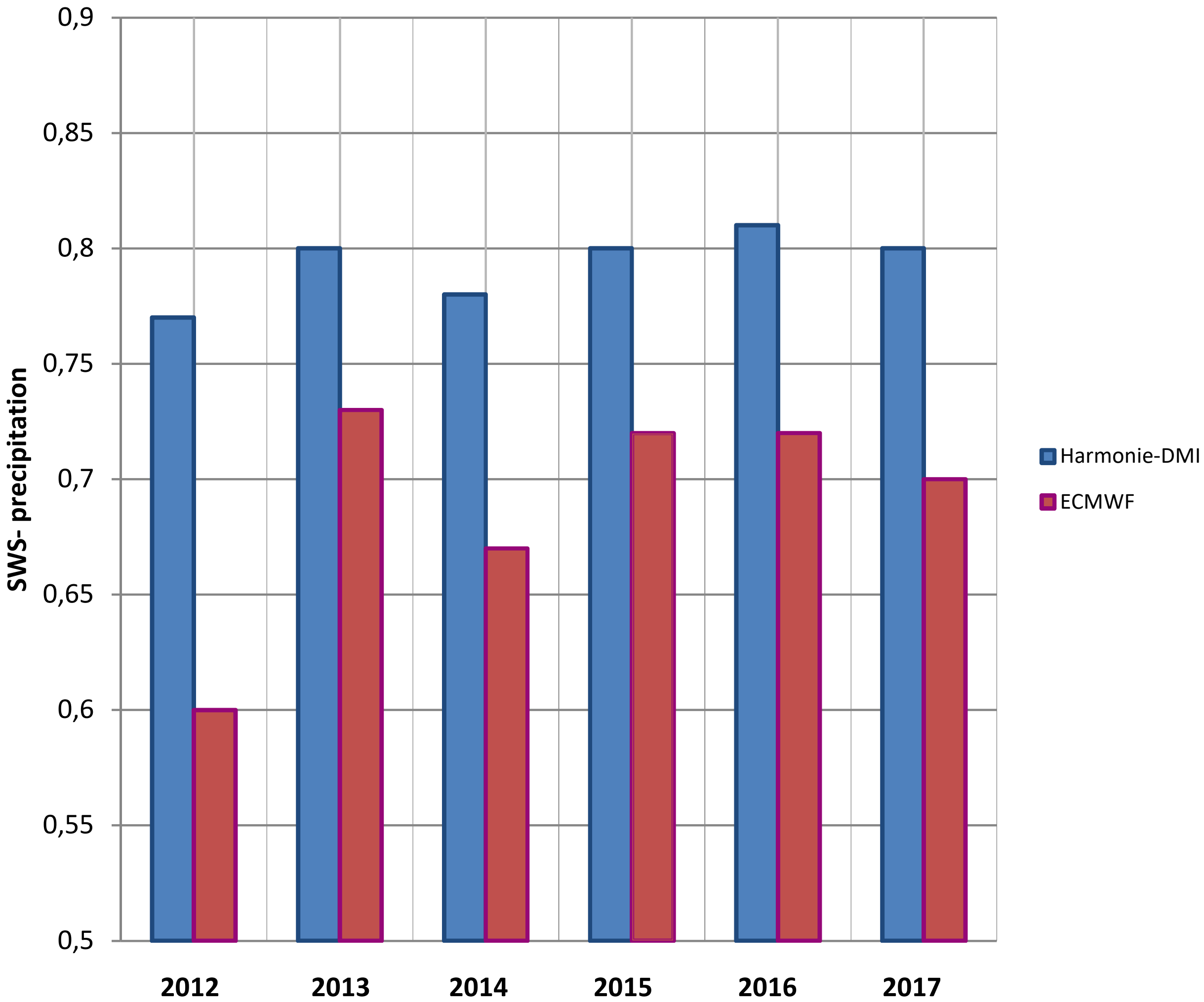
The figure shows an example of simple spatial verification scheme looking at observed extremes in synoptiic observation points and the ability of the spatial forecast field to match observed extremes

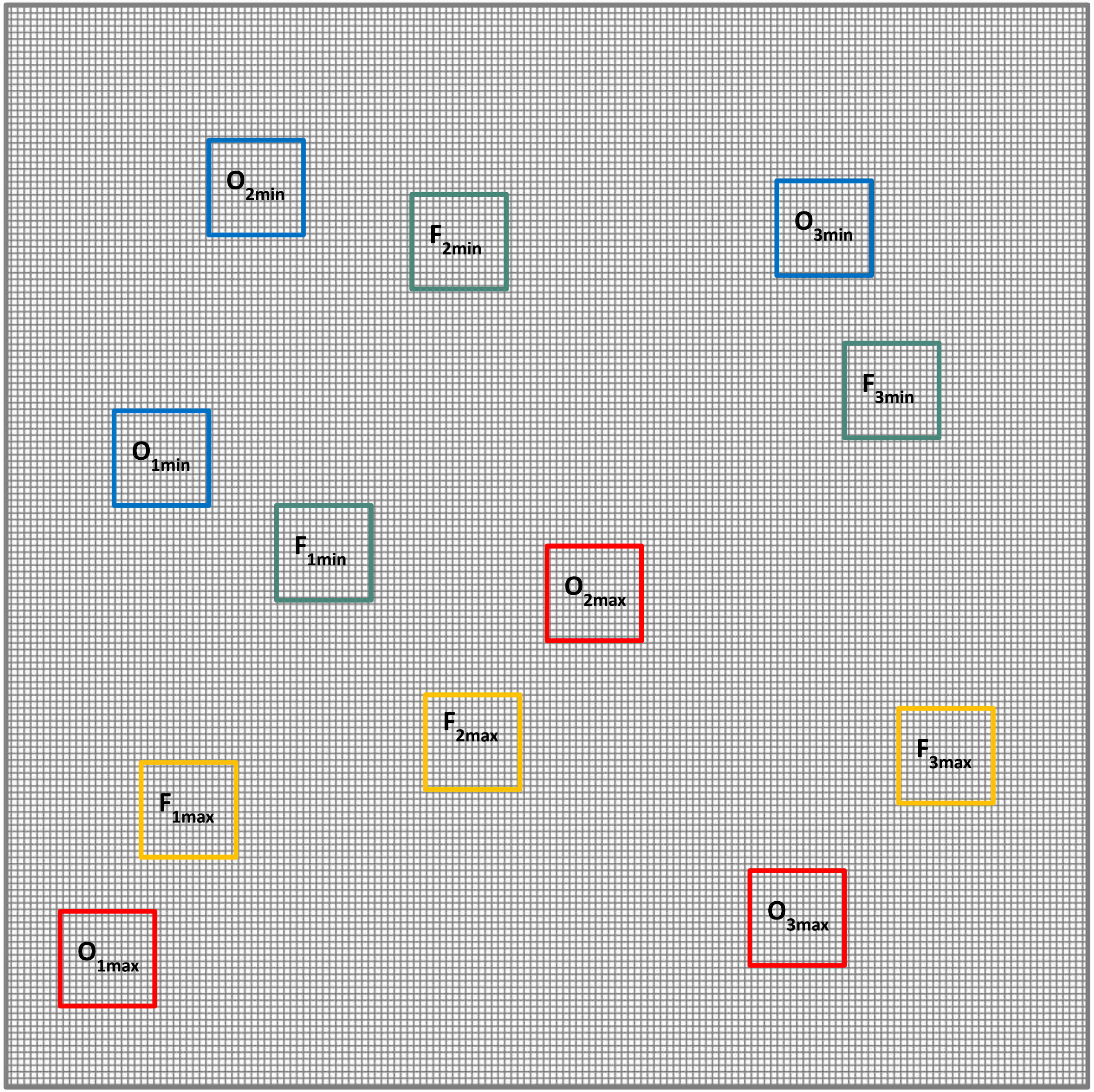
The figure illustrates the 3 highest and the 3 lowest observations in the area.

A Score function between 0 and 1 applied for each local extreme is based on the estimated value of the forecast to the users.

Results of OPERATIONAL application of the Annual SWS index verification at DMI comparing HARMONIE against ECMWF , 2012-2017 (Precipitation , Danish station list)

A benefit of Harmonie –Arome relative to ECMWF using a 15 km box size is seen





Extension of scheme by using fields and a distinction between observed and forecasted local extremes

The figure illustrates the 3 highest and the 3 lowest observations as well as forecasts in the area.

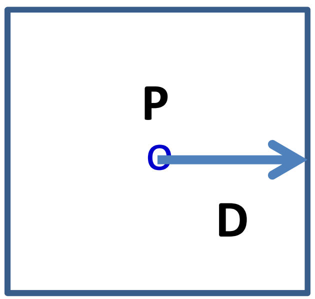
New spatial verification scheme

$\mathcal{S}_{loce}$ =Structure of local extremes

Computational procedure for precipitation:

Decide on the number of local extremes to look for, N_{min} for minima and N_{max} for maxima. A good choice is to make $N_{min} = N_{max}$. The scheme first looks for the this number of field values (orders the field values according to minima and maxima). But It is not certain that this number of points can be accteped finally due to constraints imposed by the size of neighbohoods (explained later). N_{min} and N_{max} should then never exceed half the number of grid points in the domain.

1) Associated indexes and related values of analysis and forecast respectively are determined from automatic algorithmic procedures (R programming)
2) Select the size D (M points), representing the maximum permitted distance > 0 to the point P under consideration for quadratic neighborhoods in verification procedure - If P is an analysis value the verification procedure will search for the lowest/highest forecast value respectively in the neighborhood around P , (2M +1) * (2M+1) points in neighborhood area.
3) A basic Score function S is then a function of the absolute difference between the analyzed extreme value and the the forecasted extreme value based on all points in the neighborhood.



New spatial verification scheme

$\mathbb{S}_{loce}$ =Structure of local extremes

$$\mathbb{S}_{loce} = \tfrac{1}{2}\left(S_A + S_F \right)$$

$$S_A = \frac{1}{2N_A}\left(\sum_{k=1}^{N_A} S(A_{min}(k), F_{min}(k)) + \sum_{k=1}^{N_A} S(A_{max}(k), F_{max}(k)) \right)$$

$$S_F = \frac{1}{2N_F}\left(\sum_{j=1}^{N_F} S(F_{min}(j), A_{min}(j)) + \sum_{j=1}^{N_F} S(F_{max}(j), A_{max}(j)) \right)$$

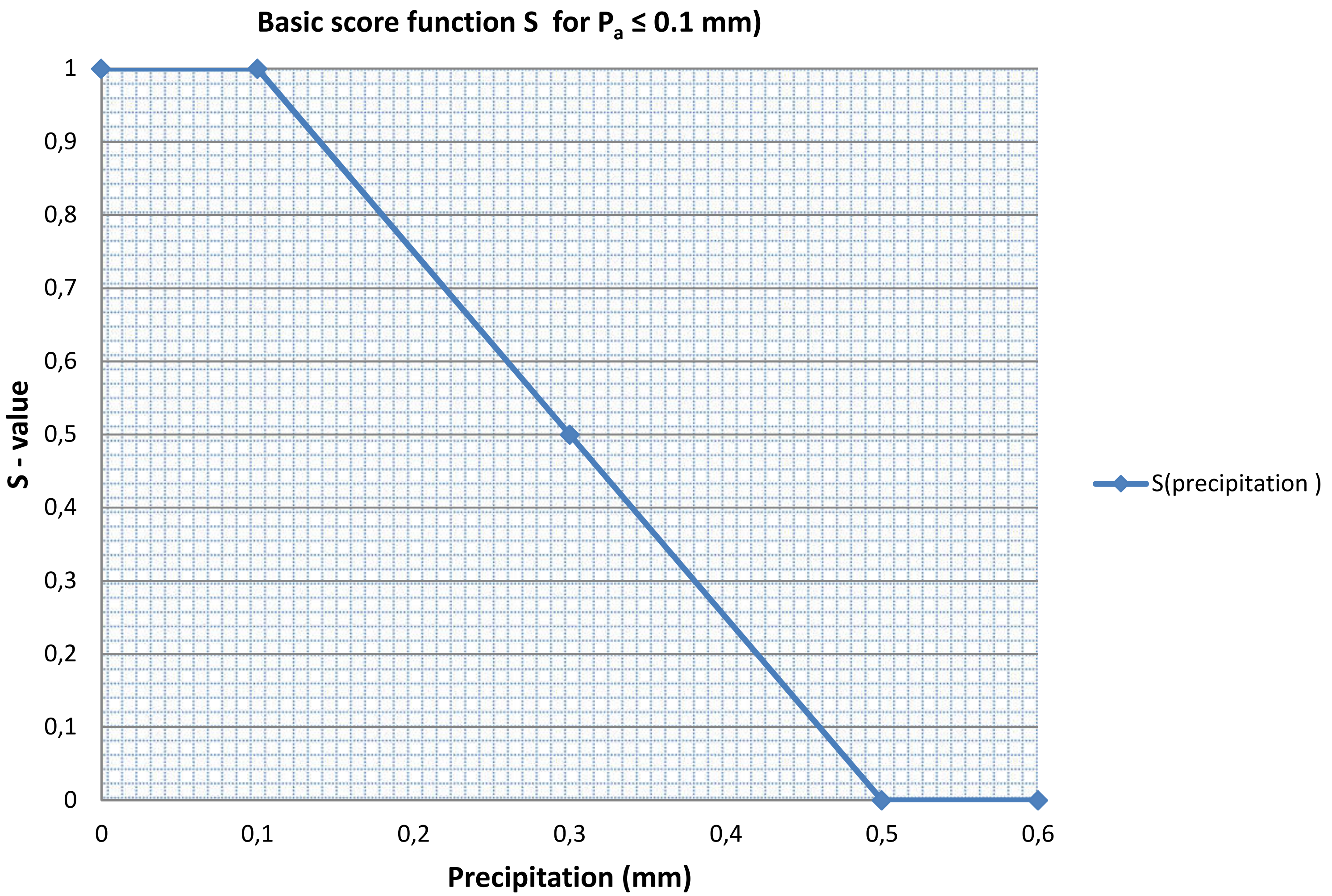
S_A is based on neighborhoods around local analysis extremes
S_F is basen on neighborhoods around local forecast extremes

N_A is the number of analyzed and accepted local maxima $> \epsilon \approx 0.1$ mm
N_F is the number of forecasted and accepted local maxima
A_{min} Analyzed local minimum value
F_{min} Forecasted local minimum value
A_{max} Analyzed local maximum value
F_{max} Forecasted local maximum value
S Score function defined as a function local analysis- and forecast extremes

The formulas above have been written for equal number of local maxima and minima

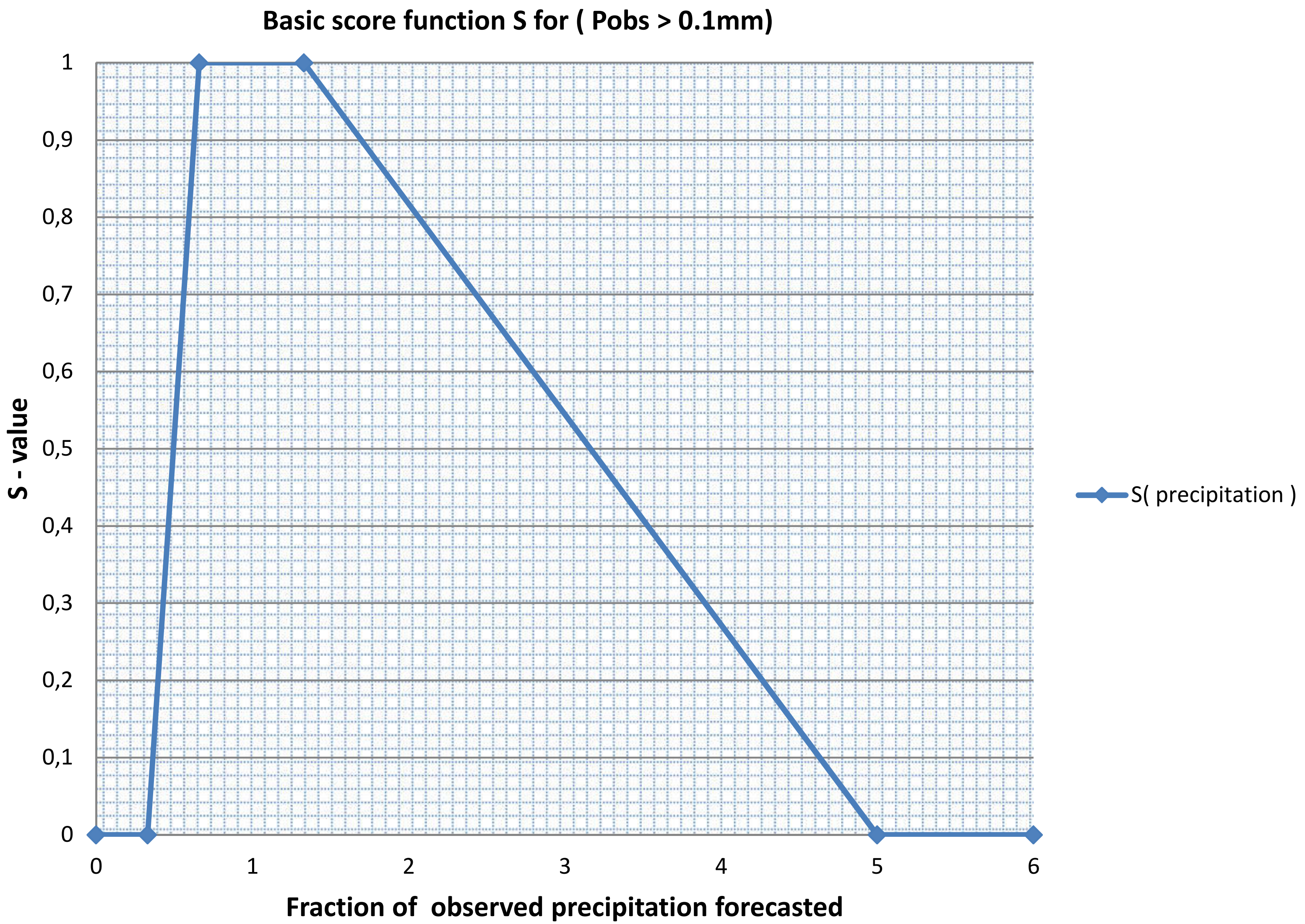
New spatial verification scheme

Precipitation score function **S** ($P_a \leq 0.1$ mm)



New spatial verification scheme

Precipitation score function **S** (P > 0.1 mm)



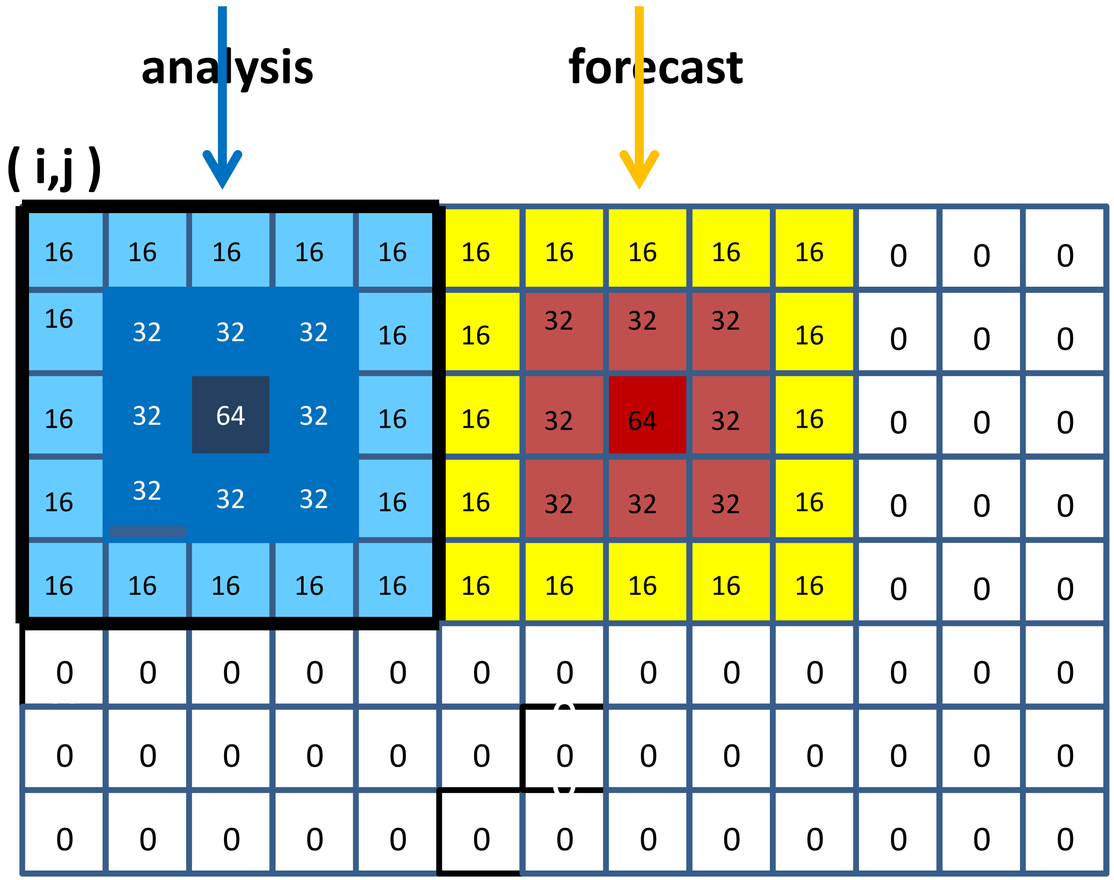
Idealized Test case 1 : CONSTANT fields

- With the default setting of the score function *S* the value of $\mathbb{S}_{loce}$ becomes 1 for constant and equal valued fields of analysis and forecast

  A special case of this is that analyzed completely dry periods tend to get high scores with a reasonable forecast model also forecasting dry conditions. This could in principle be changed if the score function is changed for these conditions.

- In the normal case where analysis and forecast are different but constant over the domain the selection of points for the verification will result in values of $\mathbb{S}_{loce}$ that are dictated by the score function *S* , e.g.

  analyzed value = 5mm and forecast value =10 mm gives $\mathbb{S}_{loce}$ =0.82 and
  analyzed value = 5mm and forecast value =15 mm gives $\mathbb{S}_{loce}$ =0.55



Test case 2 : Constraints related to the selection of local extremes:

Looking at analyzed local extreme of 64 mm and the local search neighborhood in black frame: Forecast is zero in this frame implying a score value of zero. The score increases for larger frames around 64 mm maximum. Score reaches 1 for frames reaching out to forecasted 64 mm.

NB: A frame for second local maximum that will embrace (reach) the first highest extreme selected causes inconsistency and should be avoided.

Idealized Test case 3a:
Extreme 2-GRID noisy analysis and forecast

64	0	64	0	64
0	64	0	64	0
64	0	64	0	64
0	64	0	64	0
64	0	64	0	64

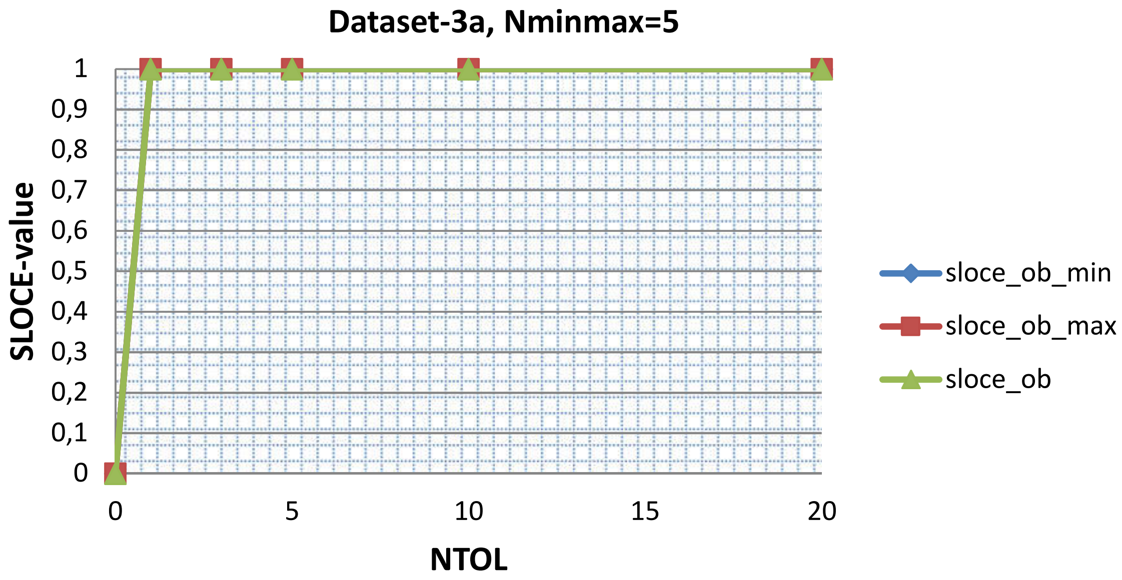
analysis

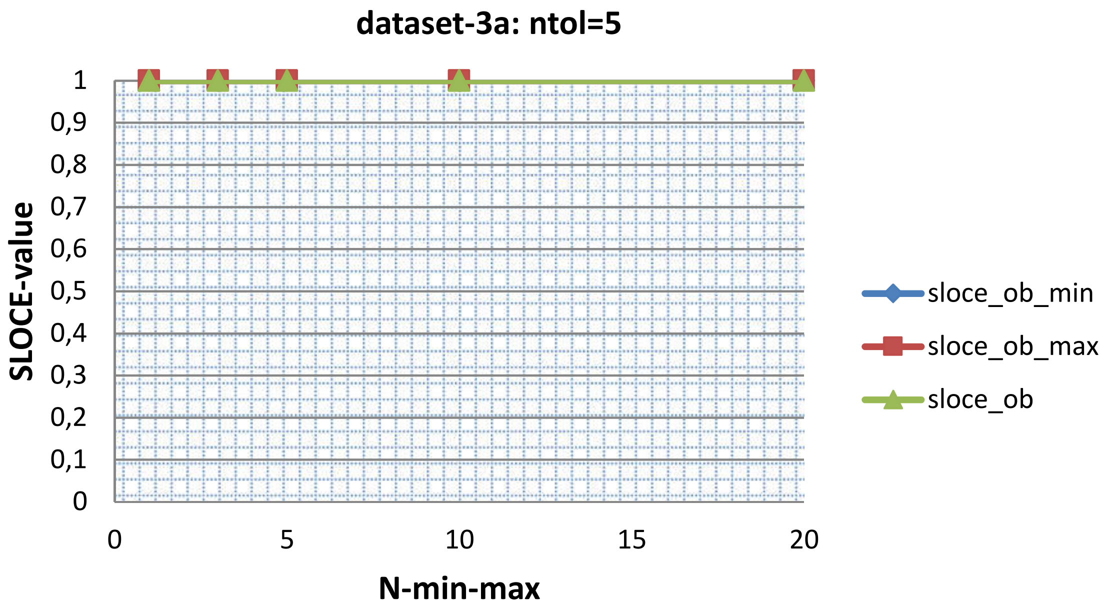
0	64	0	64	0
64	0	64	0	64
0	64	0	64	0
64	0	64	0	64
0	64	0	64	0

forecast

EXTREME 2-GRID NOISY analysis and forecast
with correct values and amplitude of oscillation
But sfifted one grid point implying that
<u>no values are correct on the grid scale</u>

Idealized Test case 3a : EXTREME 2-GRID NOISY analysis and forecast





EXTREME 2-GRID NOISY analysis and forecast with correct values and amplitude of oscillation but sfifted one grid point implying that <u>no values are correct on the grid scale</u>

Idealized Test case 3b:
EXTREME 2-GRID NOISY analysis and forecast

64	32	64	32	64
32	64	32	64	32
64	32	64	32	64
32	64	32	64	32
64	32	64	32	64

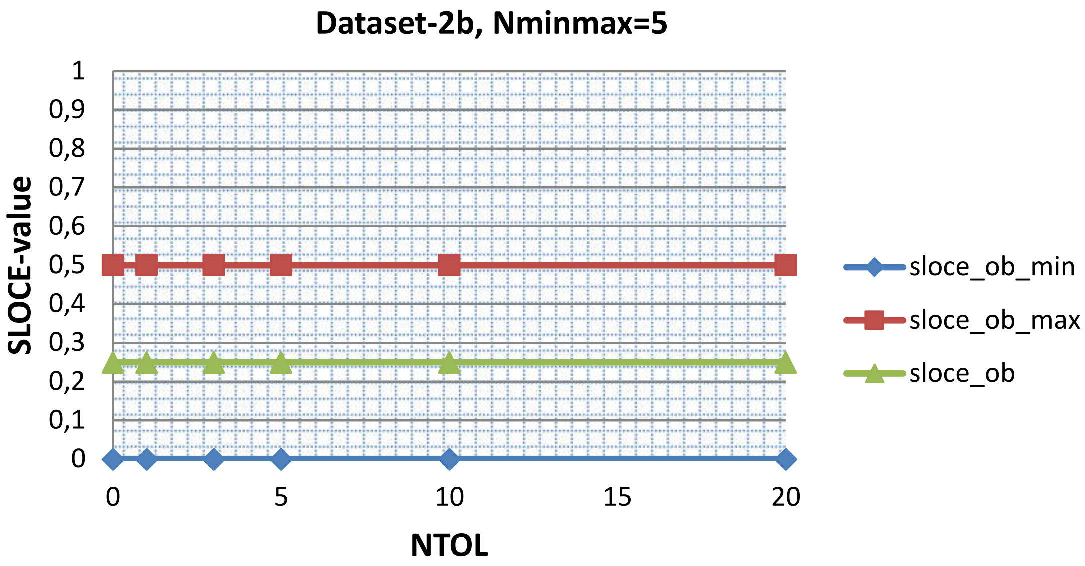
analysis

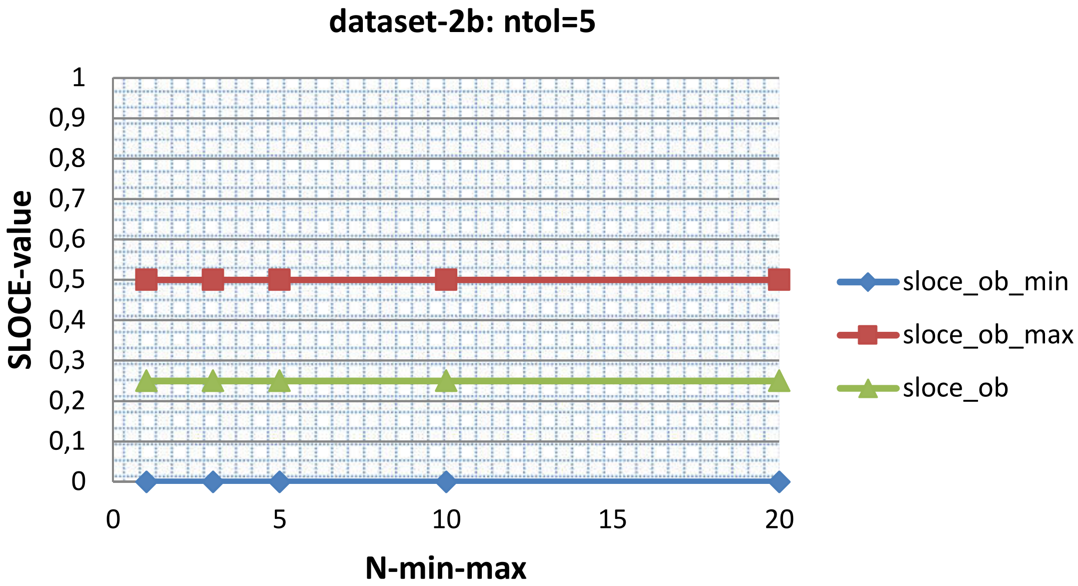
32	0	32	0	32
0	32	0	32	0
32	0	32	0	32
0	32	0	32	0
32	0	32	0	32

forecast

EXTREME 2-GRID NOISY analysis and forecast
<u>with correct amplitude of oscillation but</u>
<u>with half maxmima values of the analyzed ones</u>

Idealized Test case 3b : EXTREME 2-GRID NOISY analysis and forecast





EXTREME 2-GRID NOISY analysis and forecast
Forecast <u>with correct amplitude of oscillation but with half maxmima values of the analyzed ones</u>

<u>NB: The negative forecast bias and zero values implies much poorer scores compared with case 3a</u>

Is the scheme easy to `HEDGE´ ?

The normal answer is NO for the following reasons:

1) **For a variable climate with frequenly occurring precipitation events it is very difficult to do `hedging', because the only way to get a high value of $\mathbb{S}_{loce}$ is that forecast and analysis agree well around the local extremes. This requires spatial skill.**

2) **Dry climates with rare precipitation events is an exception with the default setting of the verification to reward zero values of analysis + forecast . This can be changed by modifying the score function *S* if considered appropriate**

Concluding remarks

- A flexible spatial verification scheme has been defined based on verifying the agreement between analysis and forecast in areas of local extremes. The scheme is a generalization of a spatial scheme that has been used sucessfully in operations, with observed extremes determined from synoptic observations.
- The scheme is so far designed for verification of precipitation but can easily be applied to other parameters.
- The scheme provides a framework for fruitful interaction between developers and users since the verification can be tailored to specific applications.
- More experimentation is planned e.g. related to the effect of choosing neighborhood size in combination with the number of extreme points looked for
- The verification softwate is currently written in R. For small domain sizes the computation time is not an issue. Further algorithmic checks and optimizations will be considered before operational use. FORTRAN could be used as an alternative if fast execution becomes an issue.

Thanks for your attention !

REFERENCES

[1] http://www.cawcr.gov.au/projects/verification/
Theme: Forecast Verification methods Across Time and Space Scales , 7th International Verification Methods Workshop (*WWRP/WGNE Joint Working Group on Forecast Verification Research*)

[2] Diebold, F.X. and Mariano, R., 1995:
"Comparing Predictive Accuracy,"Journal of Business and Economic Statistics, 13, 253-265
,
[3] Gilleland, E., Ahijevych, D.A., Brown, B.G., Ebert, E., 2010: Verifying forecasts spatially, BAMS 1365-1373

[4] Murphy, A.H., 1993: What is a good forecast? An essay on the nature of goodness in weather forecasting. *Wea. Forecasting*, **8**, 281-293.

[5] **Roberts , N.M., and Lean, H.W., 2008:**
Scale –selective verification of rainfall accumulations from high-resolution forecasts of convective events . Mon. Wea. Rev., 136, 78-97

[6] Sass, B.H., and Yang, X, 2012: A verification score for high resolution NWP: Idealized and pre-operational tests , HIRLAM Tech. Rep. no. 69, 28pp [available from www.hirlam.org]

[7] Wernli, H. , Paulat, M., Hage, M. and Frei, C., 2008:
SAL- A Novel Quality Measure for the Verification of Quantitative Precipitation Forecasts , Mon. Wea. Rev. , 136, 4470 – 4487