Wind verification with DIST method: preliminary results

Maria Stefania Tesini

Offenbach - 6 September 2016

COSMO General Meeting

Motivation

Current wind forecast verification at Arpae is not completely satisfactory

The DIST methodology used operationally to verify precipitation has pointed out some advantages with interesting results

MesoVict project encourages the investigation of the ability of existing or newly developed spatial verification methods to verify fields other than deterministic precipitation forecasts, e.g., wind forecasts

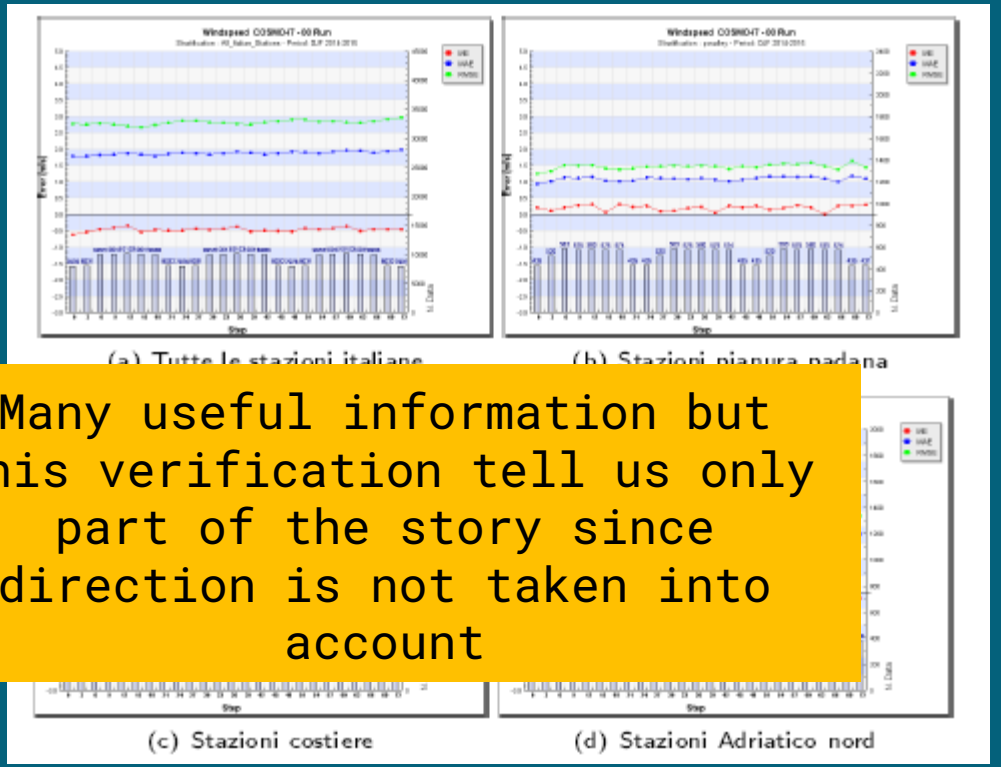
Is it possible to apply DIST to wind forecast?

Outline

- Current state of wind forecast verification at Arpa
- What is DIST and how could be extended to wind both for intensity and direction
- First results with the MesoVict cases (only 3 at the moment) using VERA analysis and COSMO-2 interpolated on the same grid
- Application of the methodology to COSMO-1 model (original grid) against the VERA analysis for MesoVict core case
- Discussion of the results, with doubts on the usefulness of the method and on its implementation, and some ideas for future tests to perform (one case could not be so meaningful)
- *N.B. all the following results are intended to test the methodology, not to assess models performances!*

Current status of wind verification at Arpae

- Wind speed errors (ME, MAE, RMSE) using GTS Synop stations, stratifying the stations according to geographical position (Po valley, coast, north Adriatic coast)

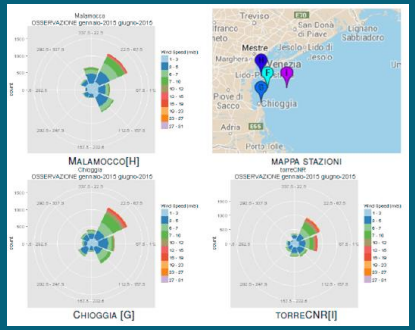


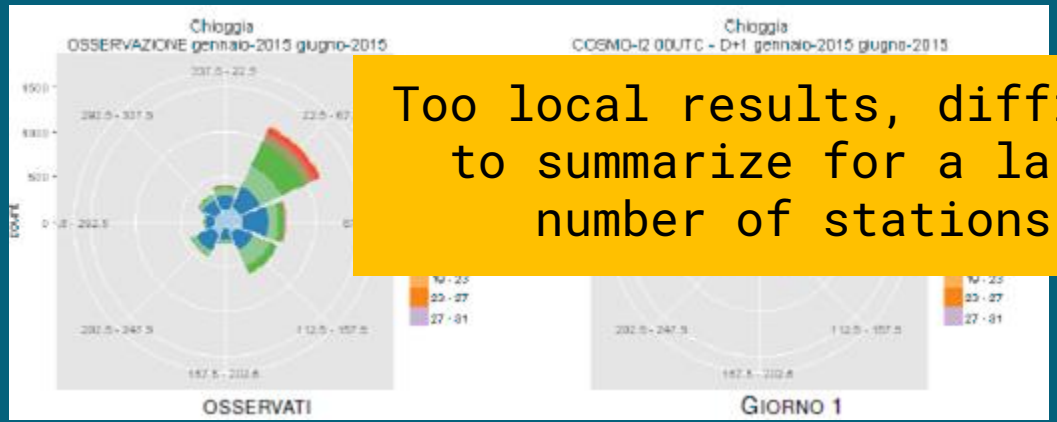
Many useful information but this verification tell us only part of the story since direction is not taken into account

Current status of wind verification at Arpae

User oriented verification:

- comparison of wind rose of both observations and forecast for each station
- contingency table to summarize results for a single station for different thresholds of wind speed





Too local results, difficult to summarize for a large number of stations

COSMO-I2

tutti i casi	PC
	42%

	BIAS SCORE	HIT RATE	FAR
N	1,05	0,27	0,75
NE	1,02	0,60	0,41
E	0,97	0,37	0,62
SE	0,87	0,47	0,45
S	0,97	0,35	0,64
SO	0,77	0,29	0,62
O	1,72	0,36	0,79
NO	0,95	0,25	0,74

ws >= 5 m/s (oss & pre)

casi osservati	PC
36%	69%

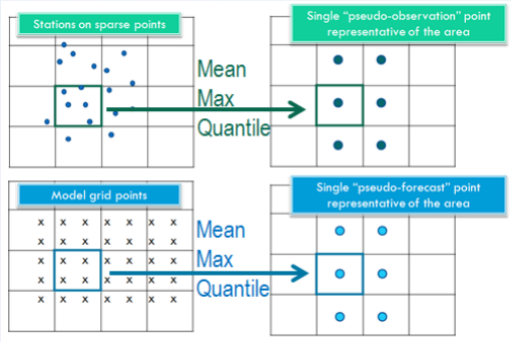
	BIAS SCORE	HIT RATE	FAR
N	1,02	0,31	0,69
NE	1,01	0,82	0,19
E	1,12	0,45	0,60
SE	0,90	0,71	0,21
S	1,42	0,50	0,65
SO	0,62	0,46	0,25
O	1,31	0,46	0,65
NO	0,48	0,24	0,50

casi osservati	PC
6%	81%

	BIAS SCORE	HIT RATE	FAR
N	0,17	0,17	0,00
NE	1,00	0,89	0,11
E	1,29	0,29	0,78
SE	Inf	NaN	1,00
S	NaN	NaN	NaN
SO	NaN	NaN	NaN
O	NaN	NaN	NaN
NO	NaN	NaN	NaN

The “distributional method (DIST)”

- The verification domain is subdivided into a number of “boxes”, each of them containing a certain number of observed and forecast values.
- For each box, several parameters of the distribution of both the observed and forecast values falling in the box can be computed (mean, median, percentiles, maximum).
- Verification is then performed using a categorical approach, by comparing for each box one or more parameters of the forecast distribution against the corresponding parameters of the observed distribution, using a set of indices.



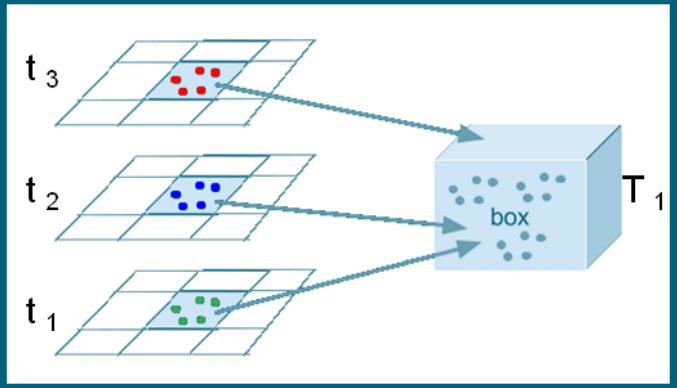
Marsigli, C., Montani, A. and Paccangnella, T. (2008), A spatial verification method applied to the evaluation of high-resolution ensemble forecasts. Met. Apps, 15: 125–143. doi: 10.1002/met.65

DIST is used at Arpae to verify precipitation

- one of the main advantages is that it can be performed using both for sparse point observations and gridded observation against gridded forecast (even if the grids are different)
- the size and even the shape of the “box” can be freely defined (e.g. alert areas for hydrological purposes)
- it provides simple information to forecaster or hydrologist about the performances of models in a single area of interest (e.g. Alert Area) or over the whole model domain aggregating the results of all boxes

DIST implementation for wind

- For precipitation the very simple way to take into account timing errors is to accumulate over longer period.
- Since it does not have sense for wind, all the values of 3 consecutive hourly forecasts (and observations) that belong to the same area are put together



For each TIME step

ORIGINAL STEP	NEW STEP
t1	
t2	T1
t3	
t4	
t5	T2
t6	
t7	
t8	T3
t9	
t10	
t11	T4
t12	

	OBSERVATIONS		FORECASTS	
	SPEED	DIRECTION	SPEED	DIRECTION
BOX 1	os1,os2,os3,…,osn	od1,od2,od3,…,odn	fs1,fs2,fs3,…,fsn	fd1,fd2,fd3,…,fdn
BOX 2	os1,os2,os3,…,osn	od1,od2,od3,…,odn	fs1,fs2,fs3,…,fsn	fd1,fd2,fd3,…,fdn
BOX 3	…	…	…	…
BOX 4	…	…	…	…
BOX 5	…	…	…	…
BOX …	…	…	…	…
BOX n	os1,os2,os3,…,osn	od1,od2,od3,…,odn	fs1,fs2,fs3,…,fsn	fd1,fd2,fd3,…,fdn

- The next step consist in the evaluation of the representative value of each box:

Mean, max or percentiles as for precipitation?

The representative value of the box

Wind speed

- Thinking to a more user-oriented verification we considered:
  - The median (e.g. the value below (or above) which 50% of the data may be found)
  - 90th percentile (e.g. the value below which 90% of the data may be found, or above which 10% of data may be found)

Wind direction

- As a first step the values were binned into 8 category (N, NE, E, SE, S, SW,W,NW)
- Then the most populated category was taken as representative for the direction in the box
- Since the direction for light wind may not be significant, an other representative value has been evaluated considering only the direction for wind with intensity > 3 m/s

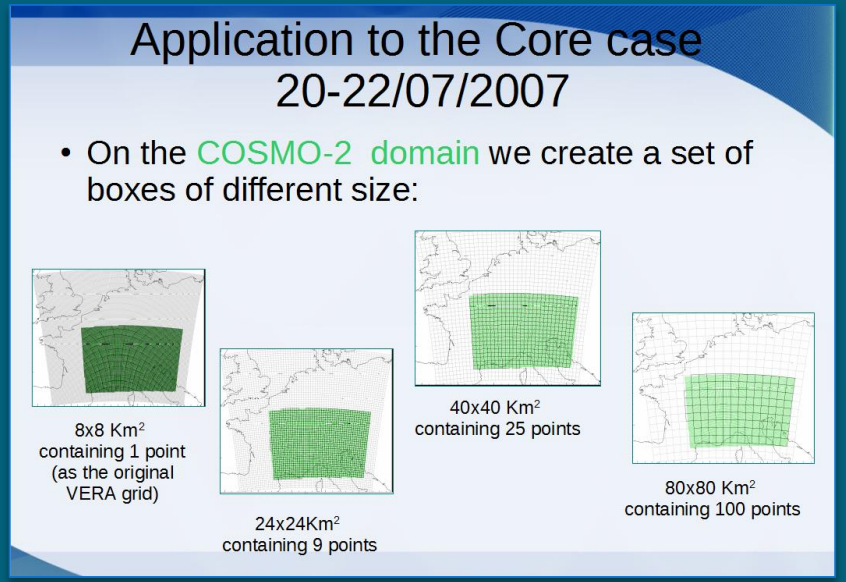
Application to MesoVict cases 1-3

- Available data:

  Observations: VERA analysis (8 Km grid)

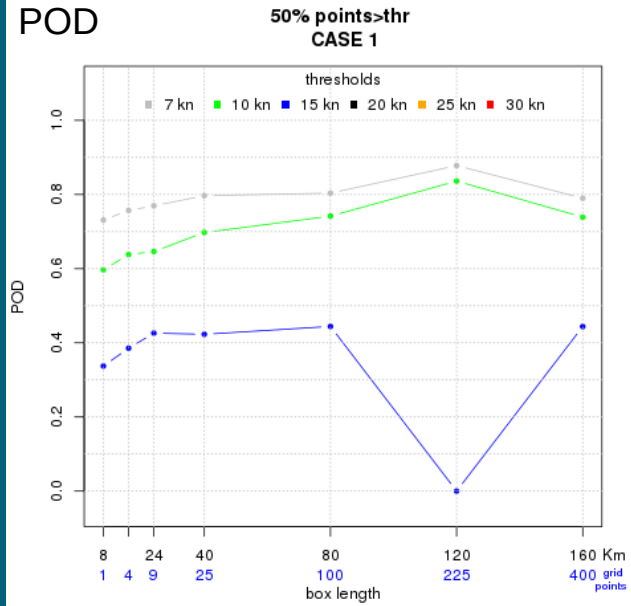
  Forecasts: COSMO-2 interpolated on the VERA grid
- As for the last year tests about precipitation, several set of boxes were created:

Dimension of box in Km	Grid points in the box	Points in the box with time aggregation
8x8	1	3
16x16	4	12
24x24	9	27
40x40	25	75
80x80	100	300
120x120	225	675
160x160	400	1200

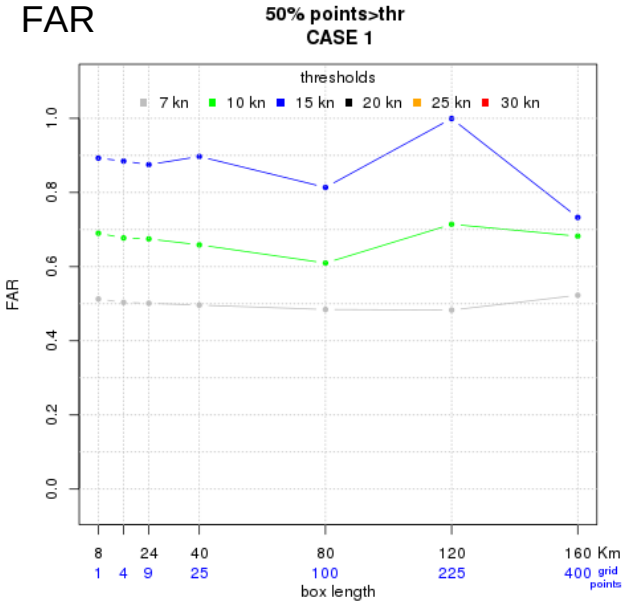


WIND SPEED case 1 – Cosmo-2 50%

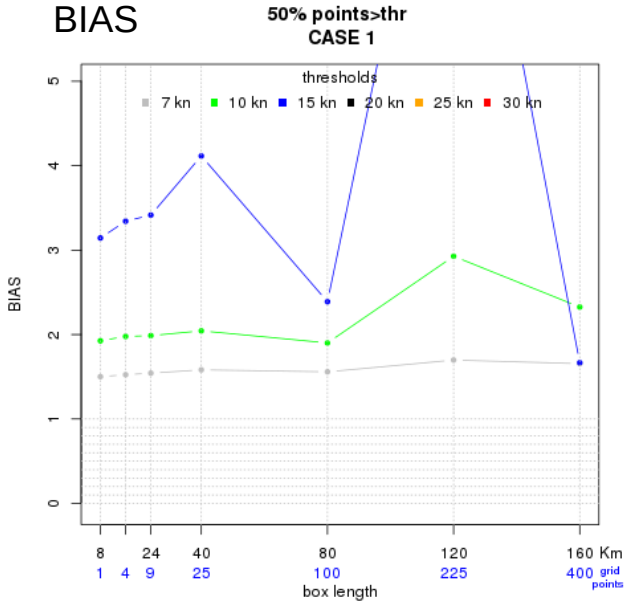
POD



FAR



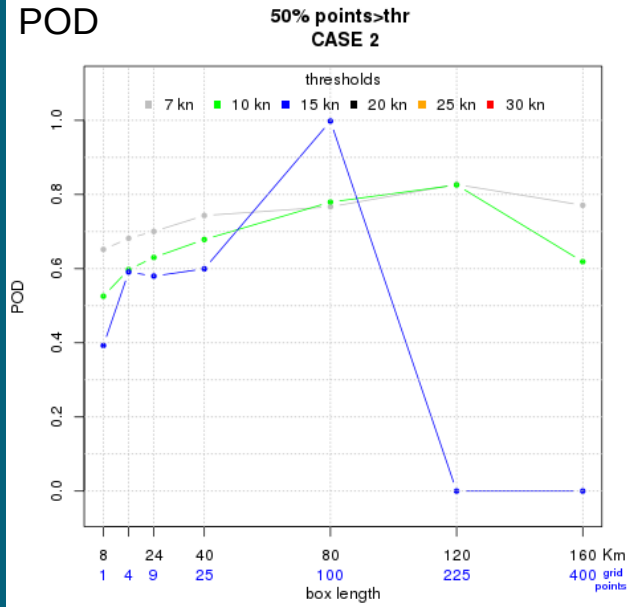
BIAS



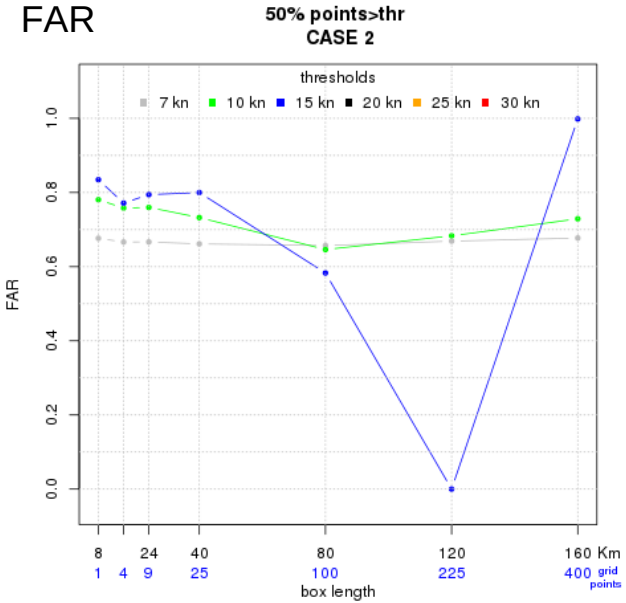
The event is defined as “median exceeding a predefined threshold”
The scores are plotted as a function of the box dimension

WIND SPEED case 2 – Cosmo-2 50%

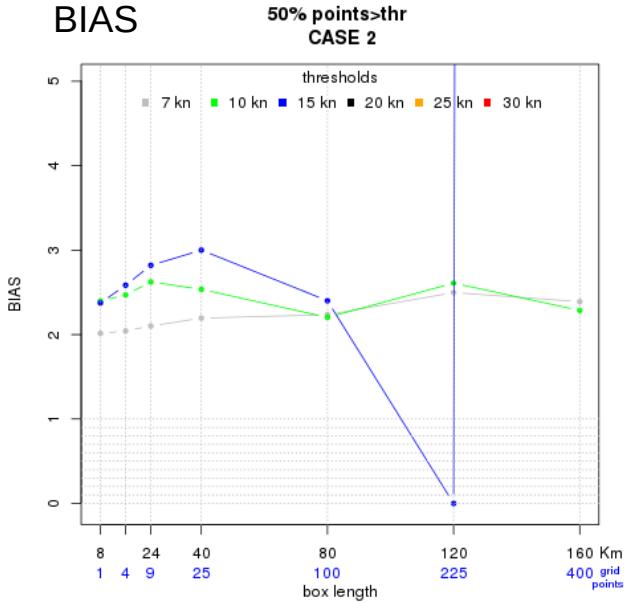
POD



FAR



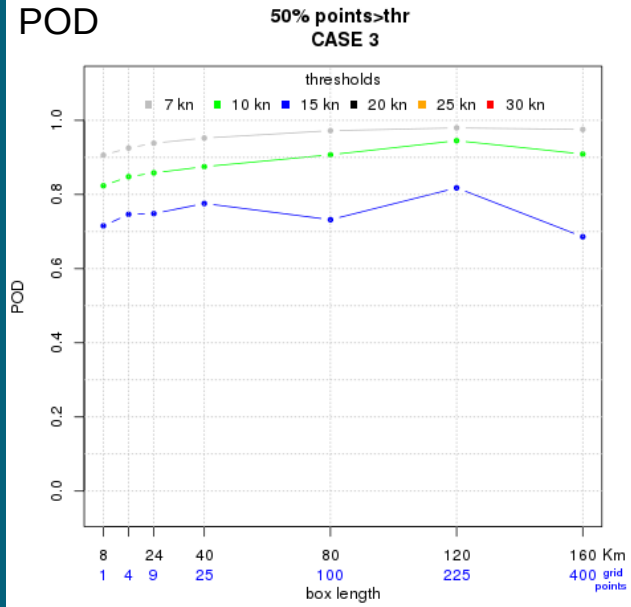
BIAS



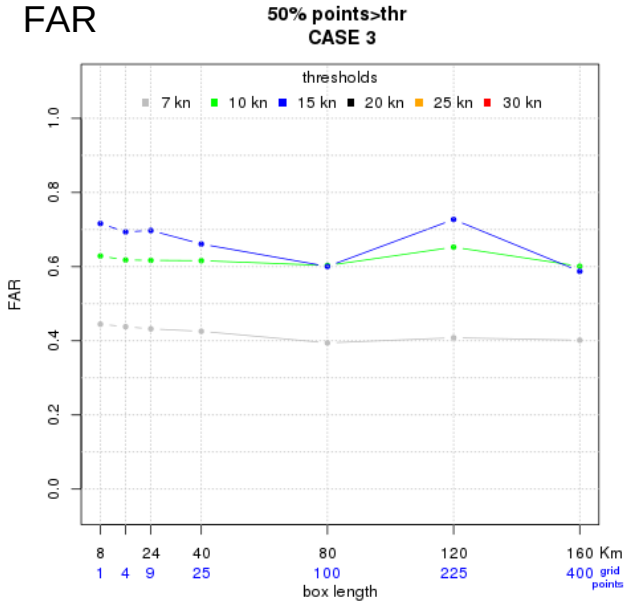
The event is defined as “median exceeding a predefined threshold”
The scores are plotted as a function of the box dimension

WIND SPEED case 3 – Cosmo-2 50%

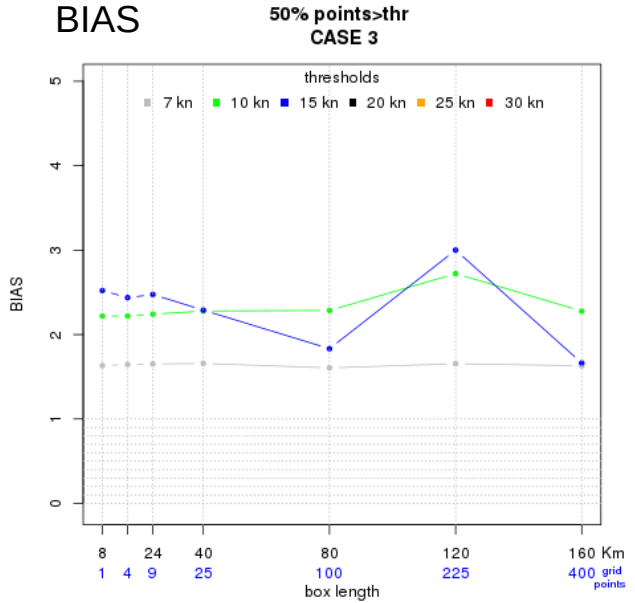
POD



FAR

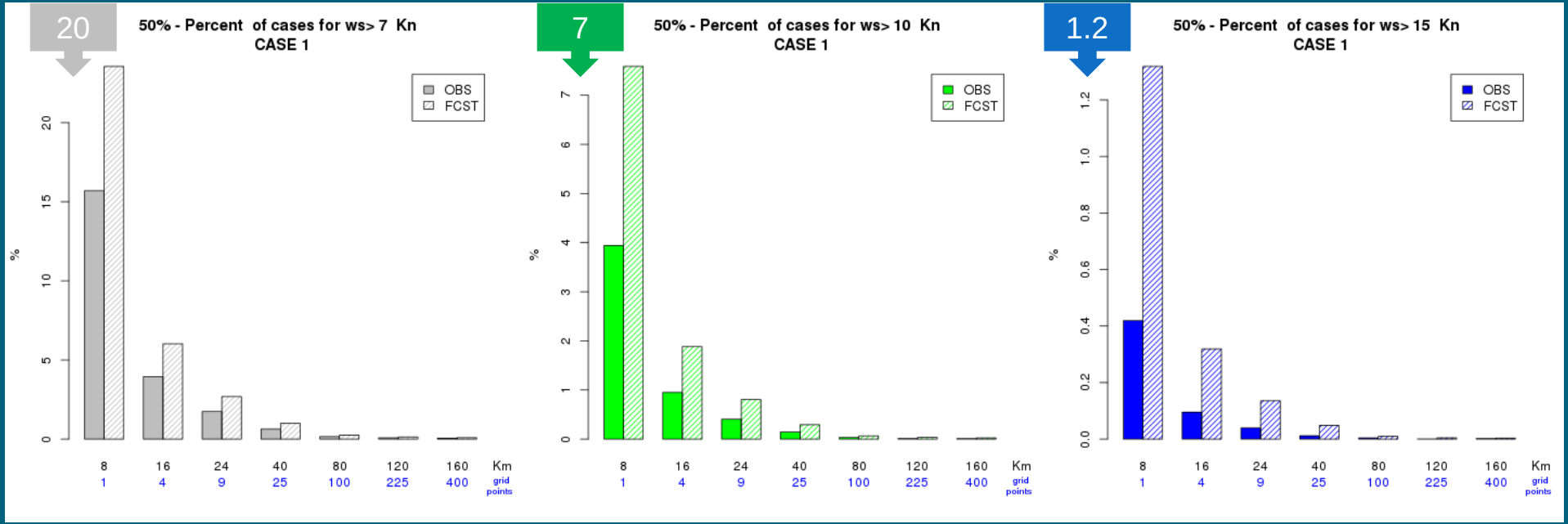


BIAS



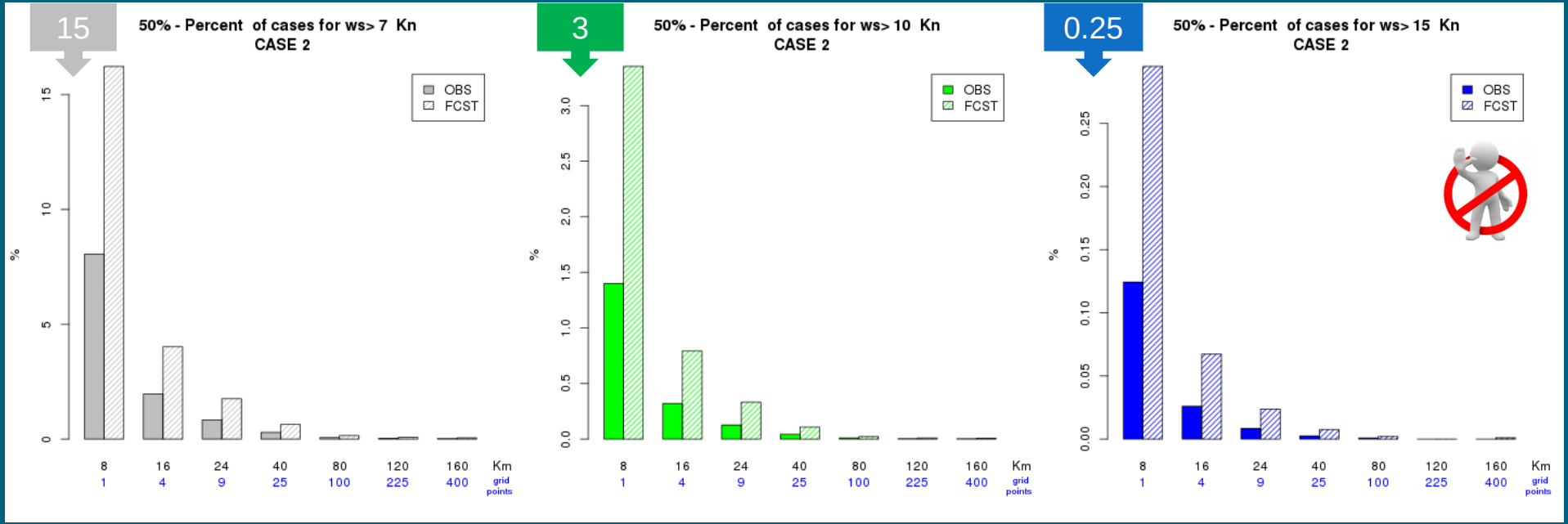
The event is defined as “median exceeding a predefined threshold”
The scores are plotted as a function of the box dimension

WIND SPEED case 1 – Cosmo-2 50%



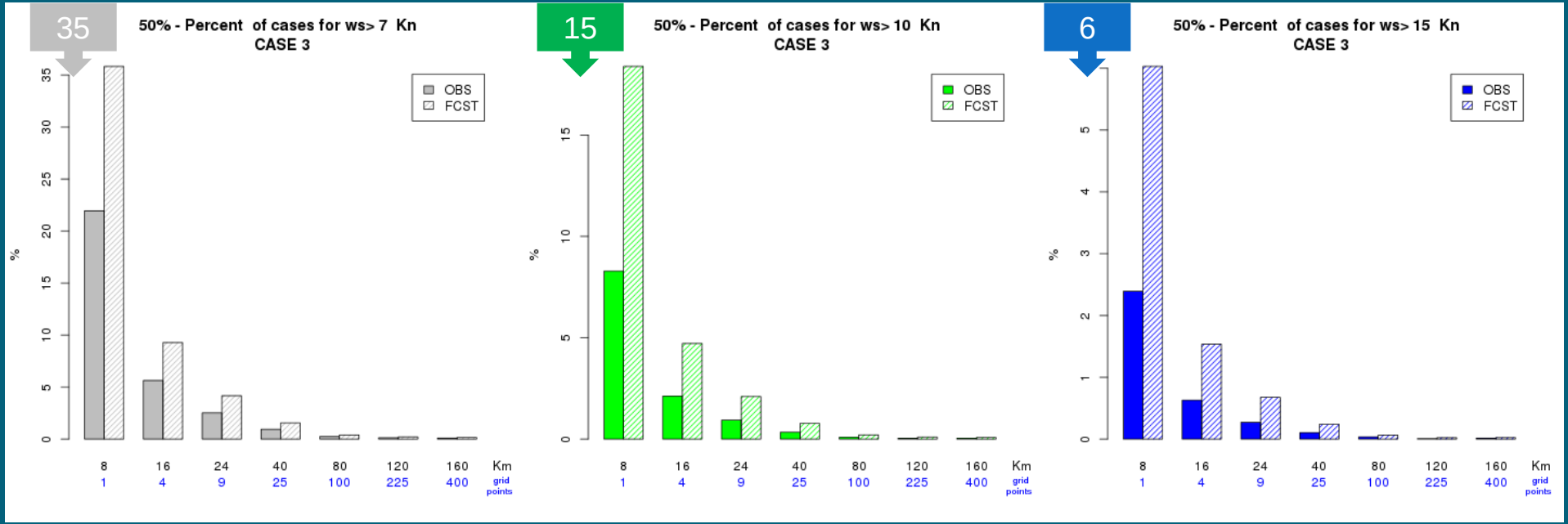
Percent of event for each thresholds respect to the total number of event (e.g. *observed yes/ total* and *forecast yes/ total)*

WIND SPEED case 2 – Cosmo-2 50%



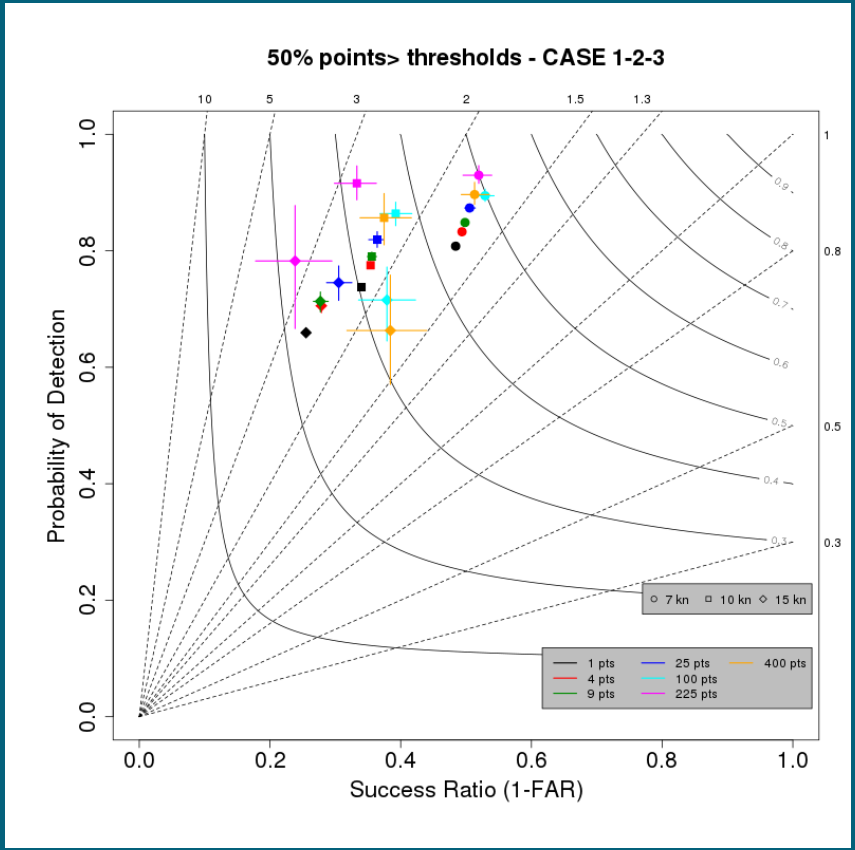
Percent of event for each thresholds respect to the total number of event (e.g. *observed yes/ total* and *forecast yes/ total)*

WIND SPEED case 3 – Cosmo-2 50%



Percent of event for each thresholds respect to the total number of event (e.g. *observed yes/ total* and *forecast yes/ total)*

WIND SPEED – case 1+2+3 Cosmo-2 50%

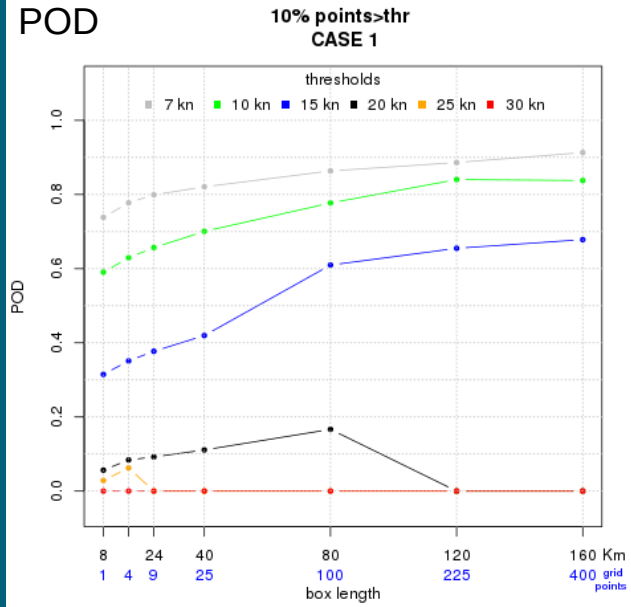


- In general, enlarging the box increases BIAS, but POD and FAR are slightly better
- the scores move more or less linearly from 1 to 25 points in the box, but not from 100 to 400 pts

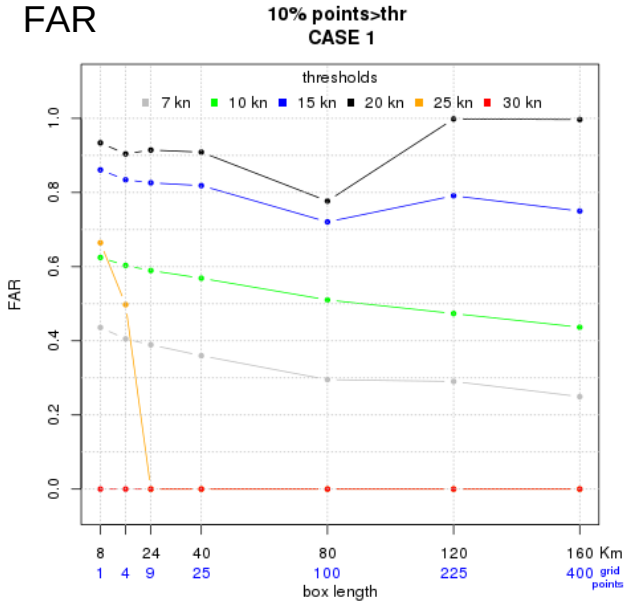
does this behavior depends on some scale of the considered phenomena or the data are not enough to produce consistent statistics?
Need more investigation…

WIND SPEED case 1 – Cosmo-2 90%

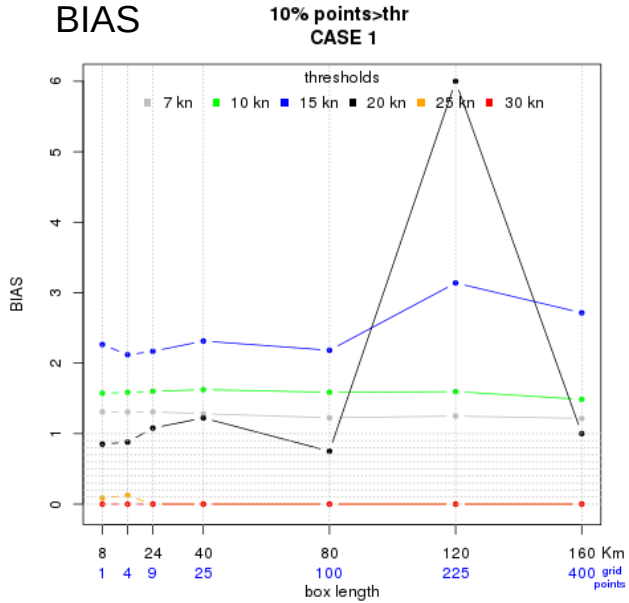
POD



FAR



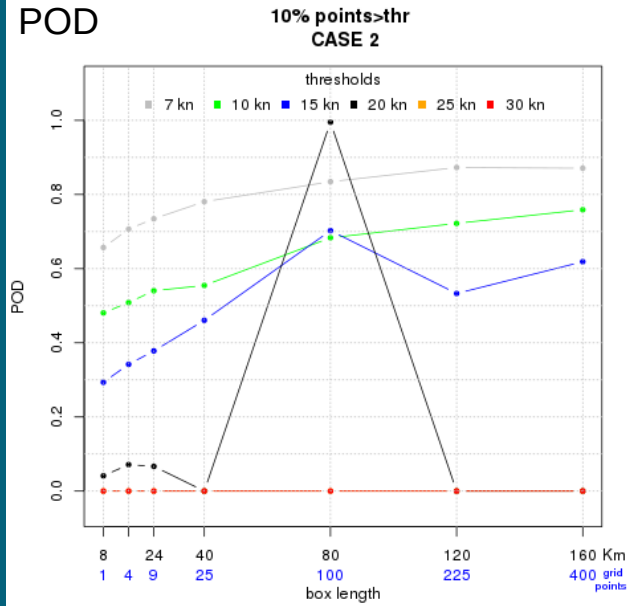
BIAS



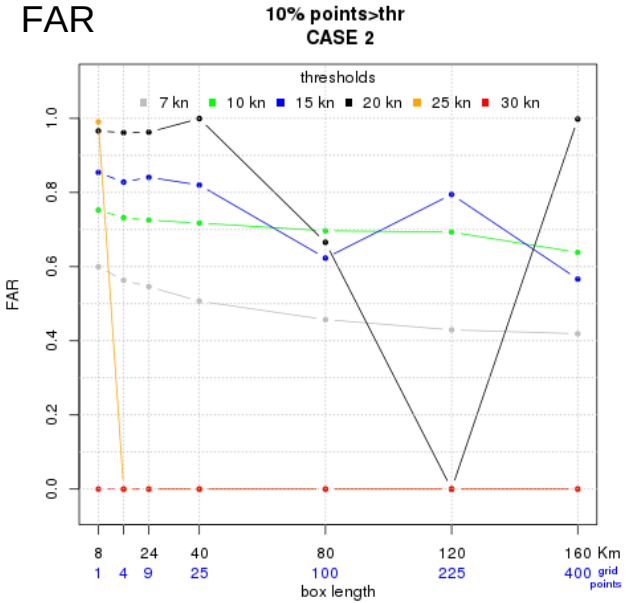
The event is defined as “10% of points exceeding a predefined threshold”
The scores are plotted as a function of the box dimension

WIND SPEED case 2 – Cosmo-2 90%

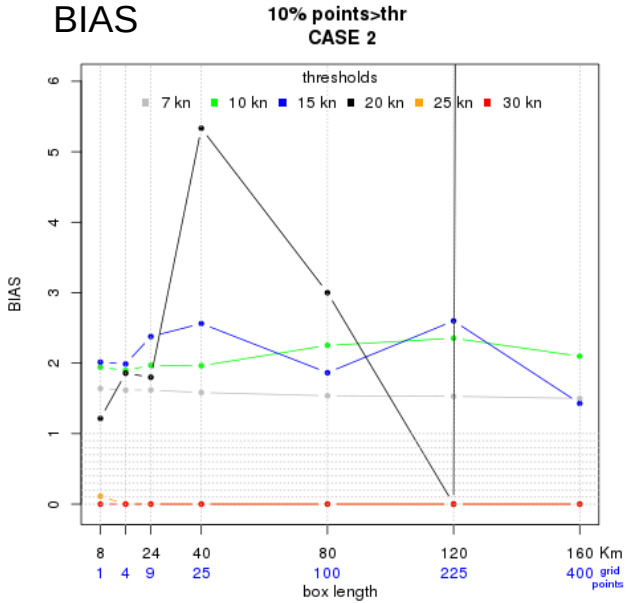
POD



FAR



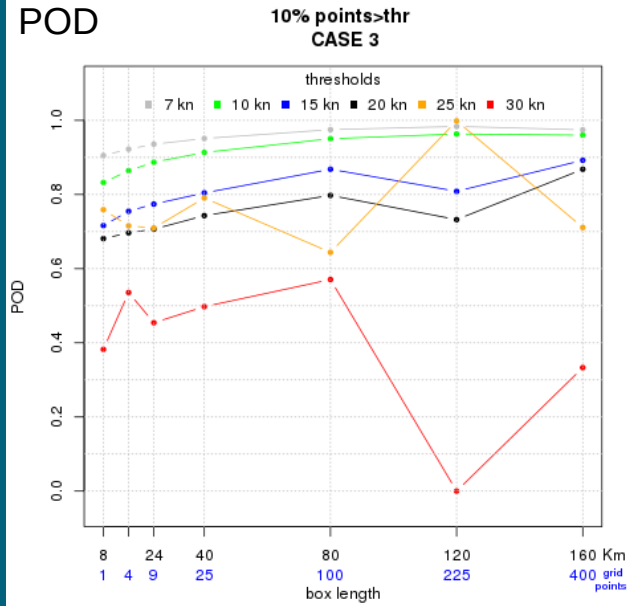
BIAS



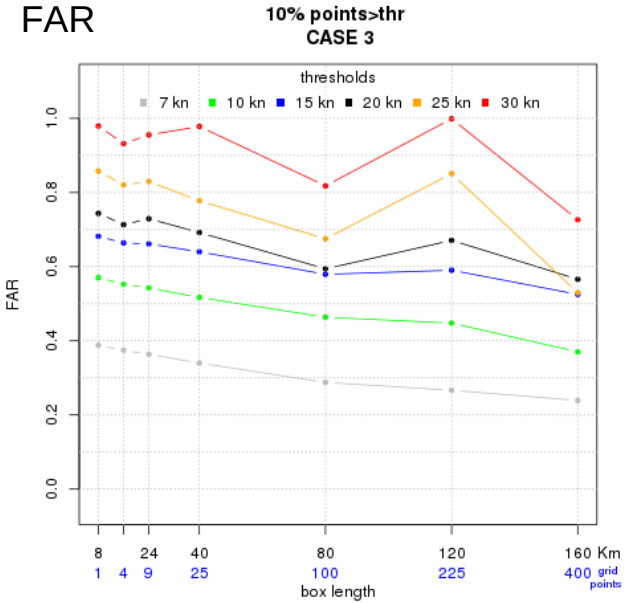
The event is defined as “10% of points exceeding a predefined threshold”
The scores are plotted as a function of the box dimension

WIND SPEED case 3 – Cosmo-2 90%

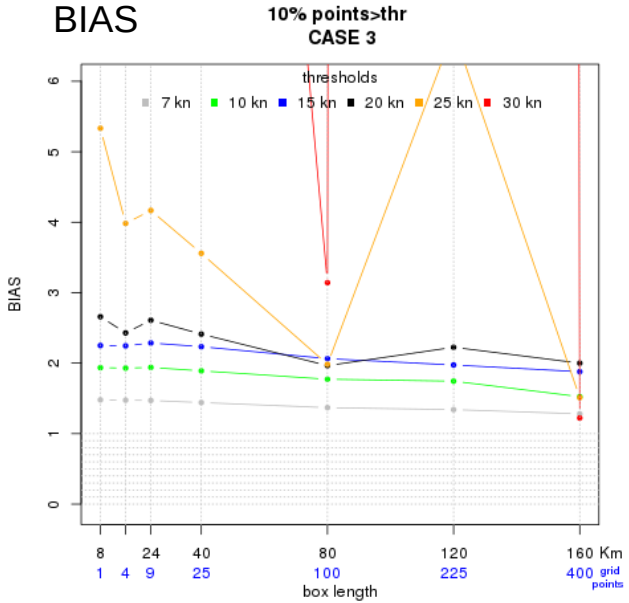
POD



FAR

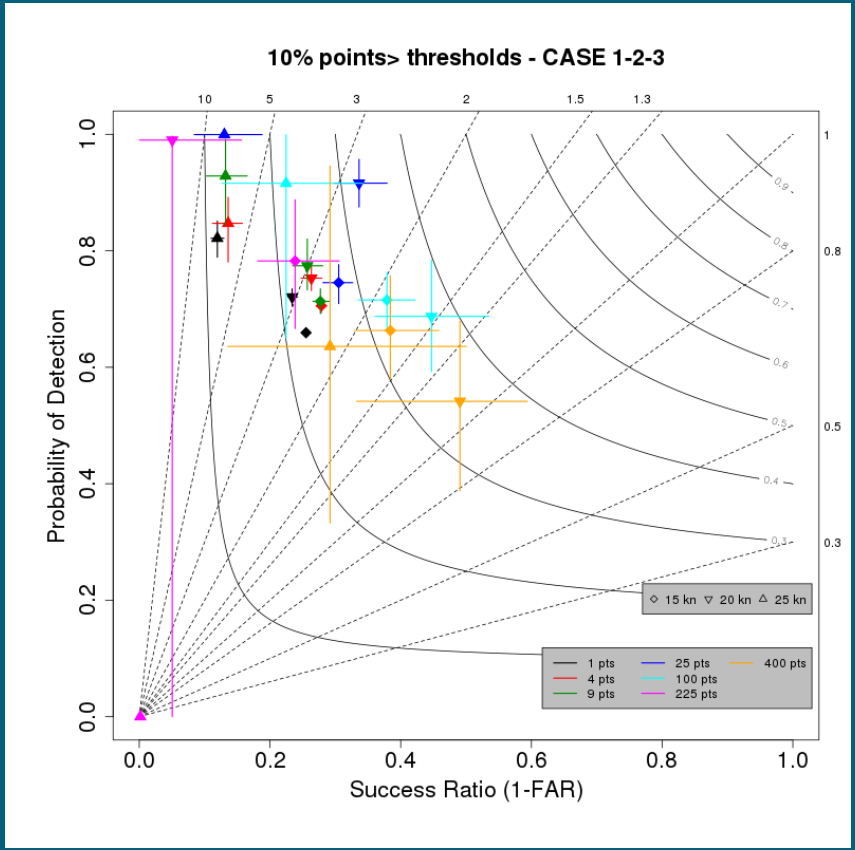


BIAS



The event is defined as "10% of points exceeding a predefined threshold"
The scores are plotted as a function of the box dimension

WIND SPEED – case 1+2+3 Cosmo-2 90%



- The number of false alarm is in general quite high, lower for 100 and 400 points grids (but also POD is smaller). Error bars for these cases are longer, meaning that the data are not enough for a more consistent statistics

Why the 225 grid is not in the group with 100 and 400? It is only by chance or we are looking to different events?

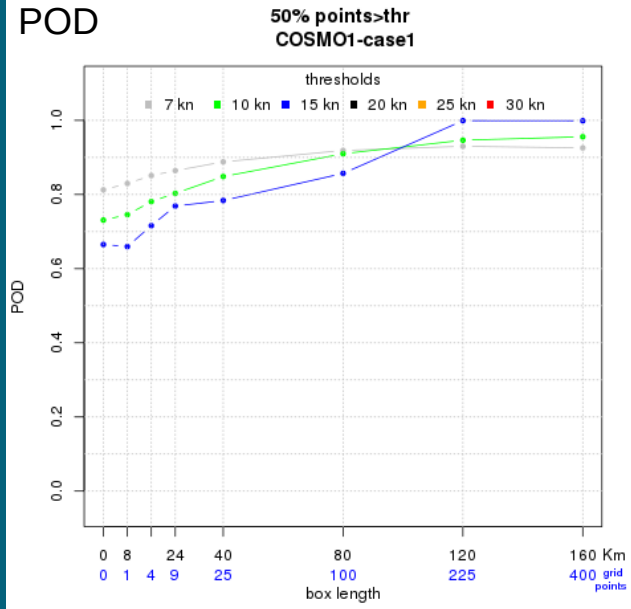
Application to MesoVict core case using COSMO-1 data (thanks to Meteo CH)

- COSMO-1 data are considered on their original grid, so the number of points of observations and forecasts in each box are very different
- Scores are also performed considering the nearest COSMO-1 grid point to all the VERA points

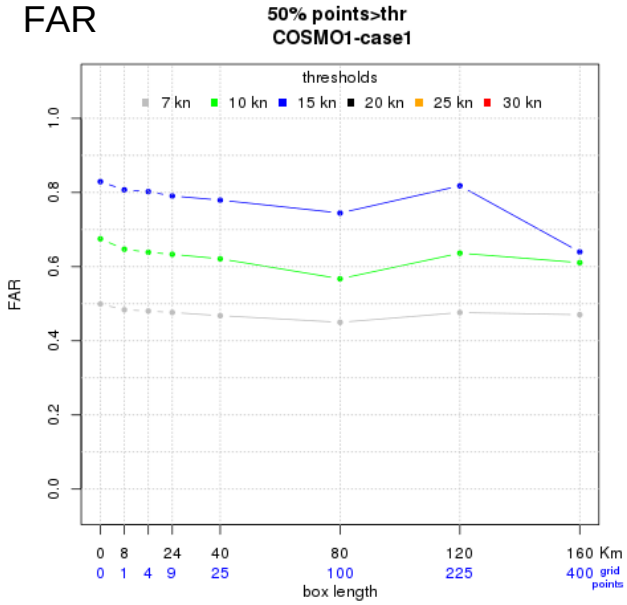
Dimension of box in Km	VERA grid points in the box	COSMO-1 grid point in the box
0	-	1
8x8	1	~50
16x16	4	~230
24x24	9	~500
40x40	25	~1400
80x80	100	~6000
120x120	225	~14000
160x160	400	~25000
	In fact the numbers are tripled because of the time aggregation	

WIND SPEED case 1 – Cosmo-1 50%

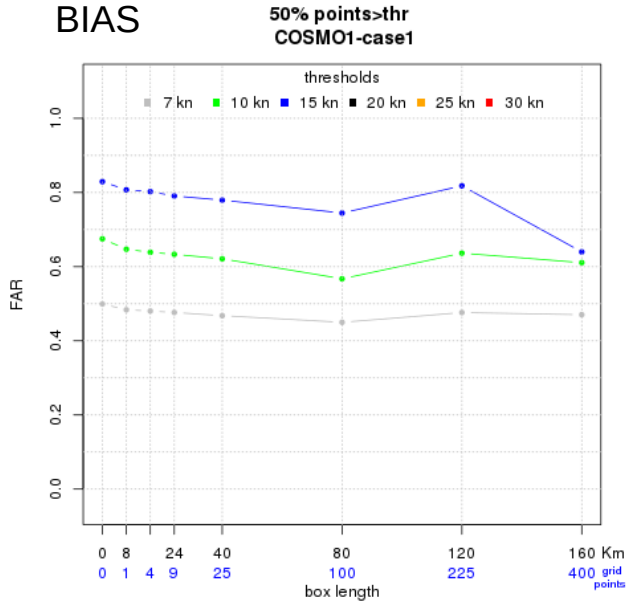
POD



FAR



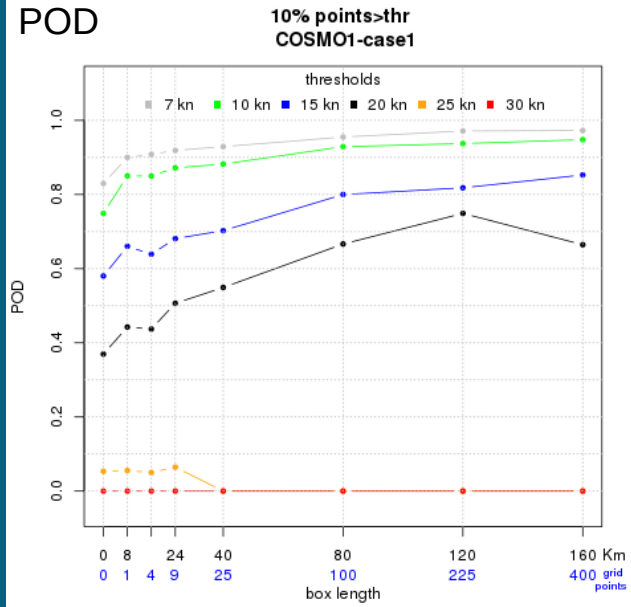
BIAS



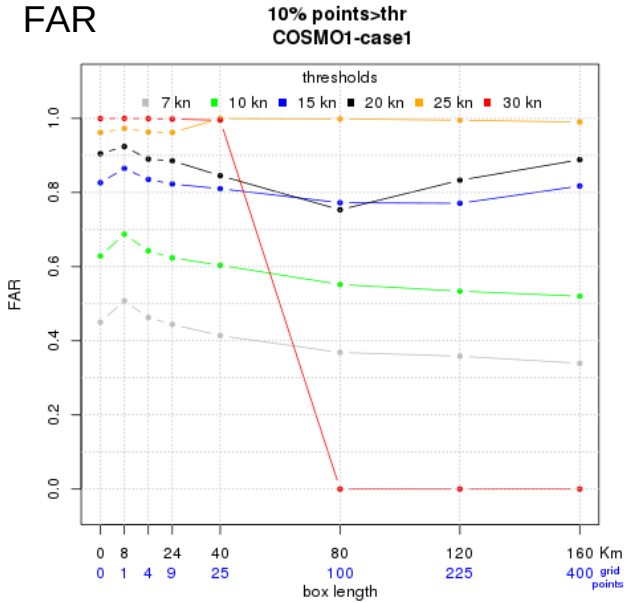
The event is defined as “median exceeding a predefined threshold”
The scores are plotted as a function of the box dimension

WIND SPEED case 1 – Cosmo-1 90%

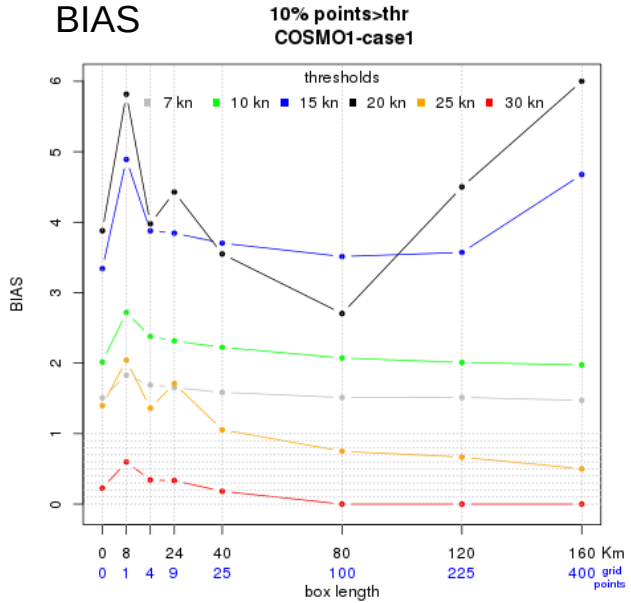
POD



FAR

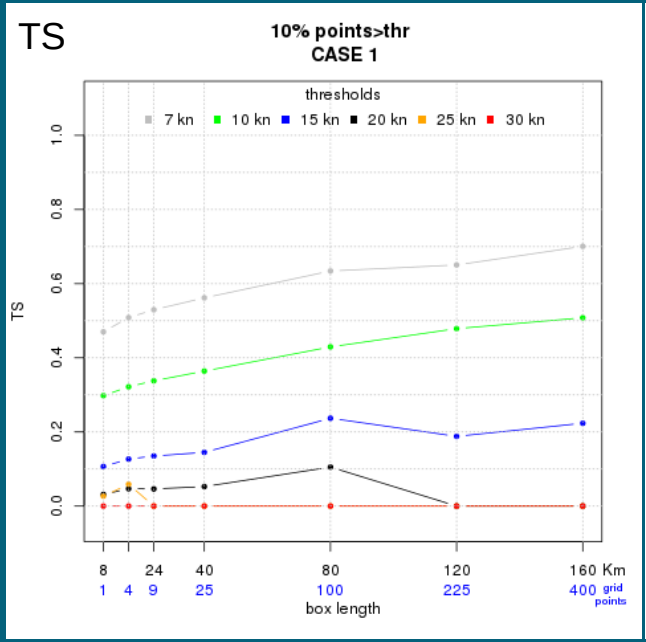


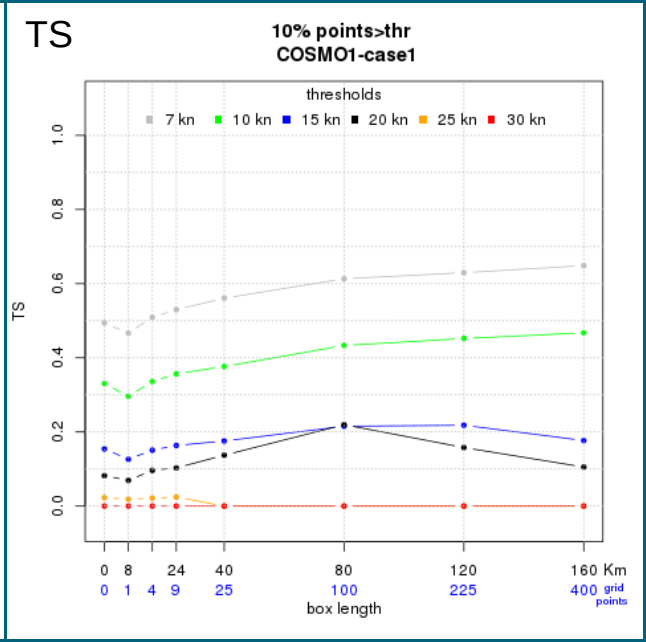
BIAS



The event is defined as “10% of points exceeding a predefined threshold”
The scores are plotted as a function of the box dimension

WIND SPEED case 1 Cosmo-2/Cosmo-1





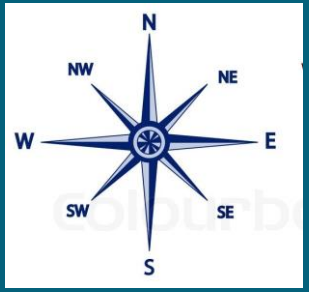
- Scores (and trend of scores) considering boxes are nearly the same for the two models
- COSMO-1 nearest grid point perform better than aggregation on 8 Km box

Can it be explained with wind field characteristic? Is only an unlucky case? Maybe the choice of this percentile is not useful as I thought initially…

The representative value of the box

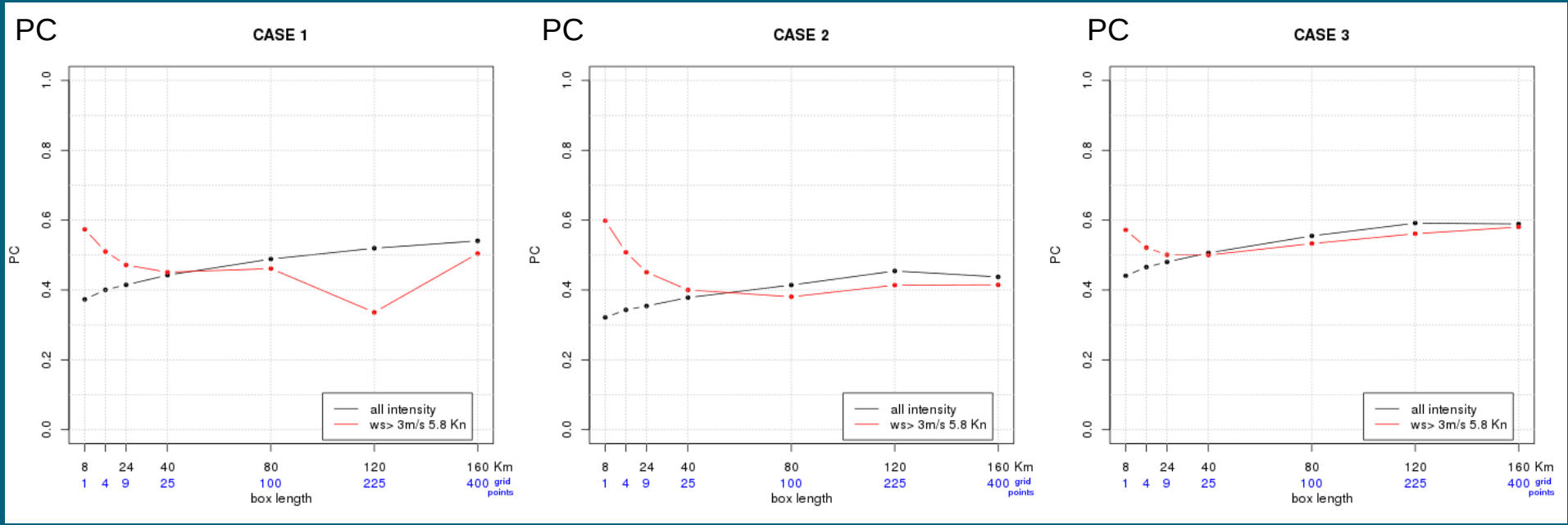
Wind direction

- As a first step the values were binned into 8 category (N, NE, E, SE, S, SW,W,NW)
- Then the most populated category was taken as representative for the direction in the box
- Since the direction for light wind may not be significant, an other representative value has been evaluated considering only the direction for wind with intensity greater than 3 m/s



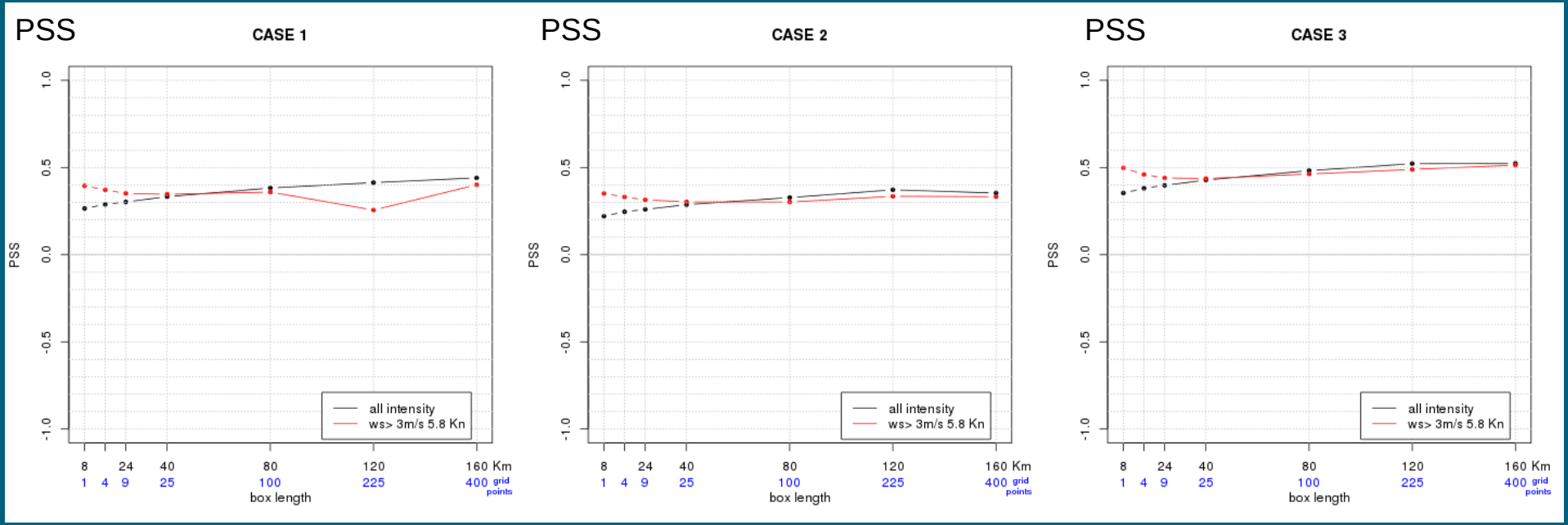
During the score interpretation I realized that cases of no prevalent direction (e.g. same number of cases in different categories) were not correctly managed. I don't know how this can influence the results. Due to lack of time I wasn't able to rerun all the scores…

WIND DIRECTION case 1-2-3 Cosmo-2



What fraction of the forecast were in the correct category?
The scores are plotted as a function of the box dimension

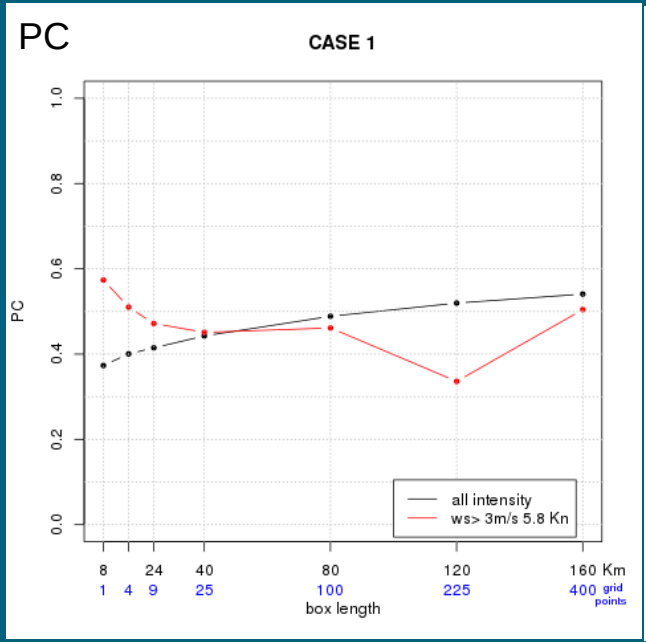
WIND DIRECTION case 1-2-3 Cosmo-2

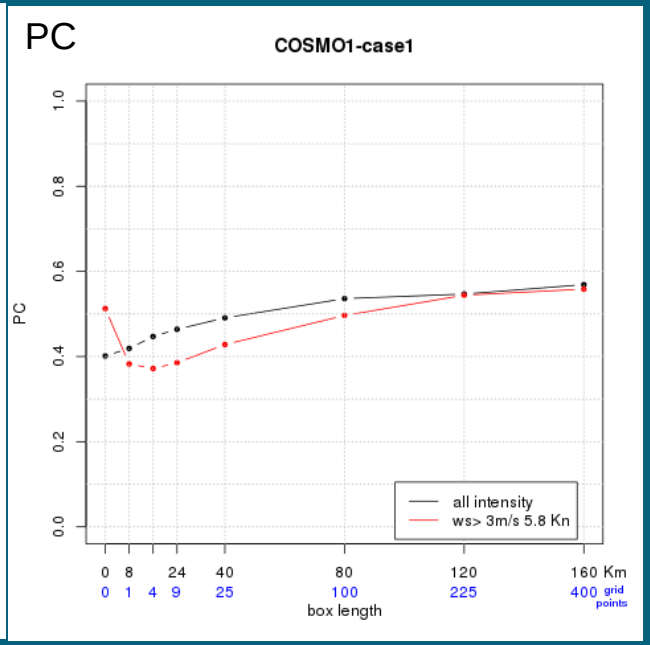


Peirce's skill score: What was the accuracy of the forecast in predicting the correct category, relative to that of random chance?

The scores are plotted as a function of the box dimension

WIND DIRECTION case 1 Cosmo-1

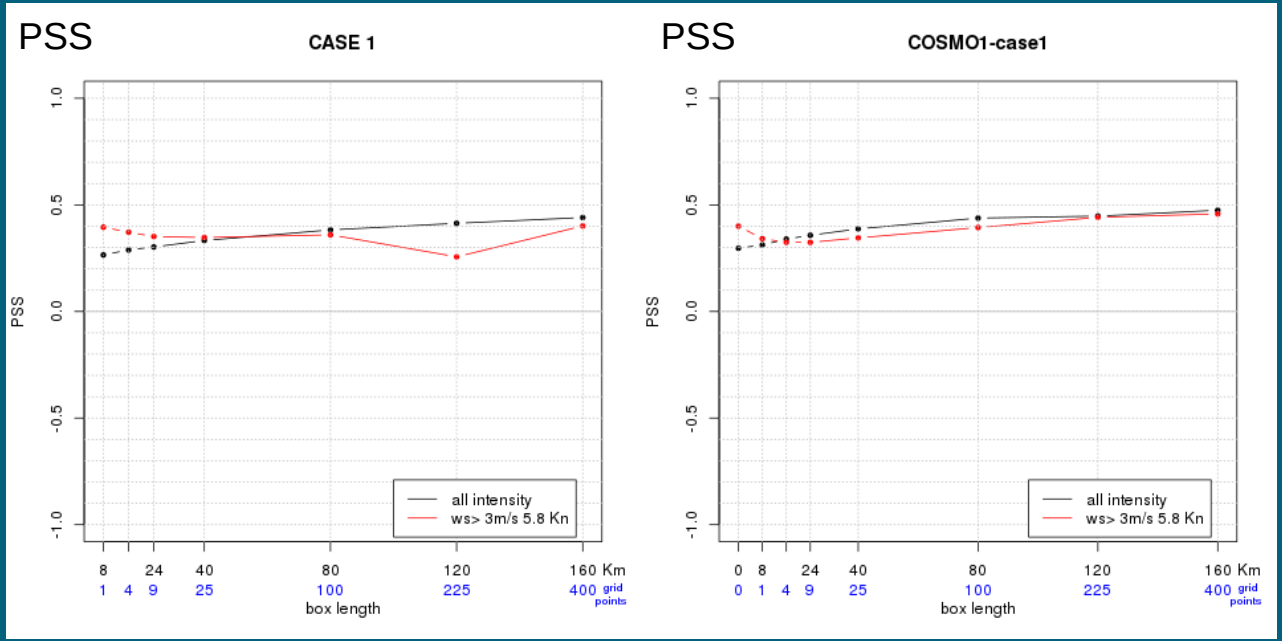




- Considering all cases with no restriction on wind intensity the accuracy increases with enlarging the box size
- But with only significant wind the accuracy decreases

What fraction of the forecast were in the correct category?
The scores are plotted as a function of the box dimension

WIND DIRECTION case 1 Cosmo-1



At larger scale all the local information are filtered out
With boxes of smaller size the information about local winds should be predominant…but is VERA analysis able to reproduce the very local features?

What was the accuracy of the forecast in predicting the correct category, relative to that of random chance?

The scores are plotted as a function of the box dimension

Conclusion

- First results on DIST application to wind are not very satisfactory:
  - Maybe representative values should be chosen differently
  - The verification period was very short
  - maybe wind is too local and the aggregation has is benefit only if the “box” were chosen differently, maybe taking into account of the orography (valley,…)
- But before giving up some other test are needed:
  - For other MesoVict cases
  - Looking to the geographical distribution of the scores
  - Using the JDC original observation (one of the main advantages of DIST was to deal with sparse point obervation…)

Thanks for your attention and suggestion!