



Status of Task 3.1 – Verification of forecasts of intense convective phenomena

Andrzej Mazur, Joanna Linkowska

Institute of Meteorology and Water Management – National Research Institute





1. Introduction

2. Done

3. Examples

4. Conclusions

"Discrete" (1-3) vs. "Continuous" (4) verification

1. SAL (Structure-Amplitude-Location) approach

S – structure – compare the volume of the normalized objects.

A- amplitude – corresponds to the normalized difference of the domain-averaged values

L- location –Combinations of a difference of mass centers of fields and averaged distance between the total mass center and individual objects

The perfect forecast $S = A = L = 0$

2. Fraction Skill Scores (FSS) assessment

Direct comparison of the forecast and of observed fractional coverage of grid-box events in spatial windows of increasing size. Most sensitive to rare events.

FSS = 0 - no correspondence between observations and forecasts

FSS = 1 - perfect match

FSS \geq FSS uniform - „useful” forecast.

3. Contingency tables analysis.

4. ME, MAE, RMSE – which metric is better?

- RMSE has the benefit of penalizing large errors more so can be more appropriate in some cases
- RMSE does not describe average error alone as MAE does
- Distinct advantage of RMSE over MAE – RMSE doesn't use the absolute value – which is good in many mathematical calculations

Observations: lightnings (C2G, C2C) from the Polish lightning detection network PERUN, covering Poland + parts of neighbouring countries

Forecast: CAPE-based FLR (Flash Rates) as follows:

$$W = 0.3 \cdot \sqrt{2 \cdot CAPE}$$
$$FR = \left(\frac{W}{14.66} \right)^{4.54}$$
$$\text{if } CTT > -15^{\circ}\text{C} \quad FR = FR \cdot \left[\max \left(\frac{-CTT}{15}, 0.01 \right) \right]$$
$$\text{if } CBT < -5^{\circ}\text{C} \quad FR = FR \cdot \left[\max \left(\frac{CBT + 15}{10}, 0.01 \right) \right]$$

Archive observations vs. forecasts (2011-2017)

Cases selection:

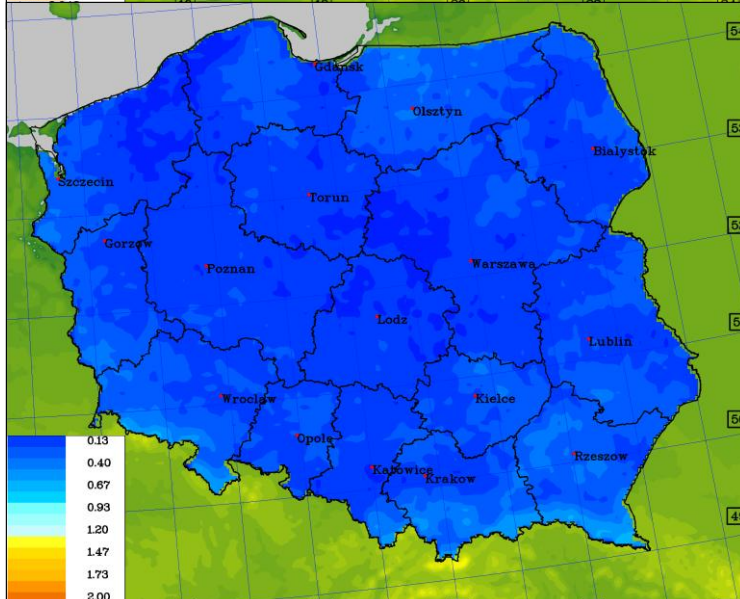
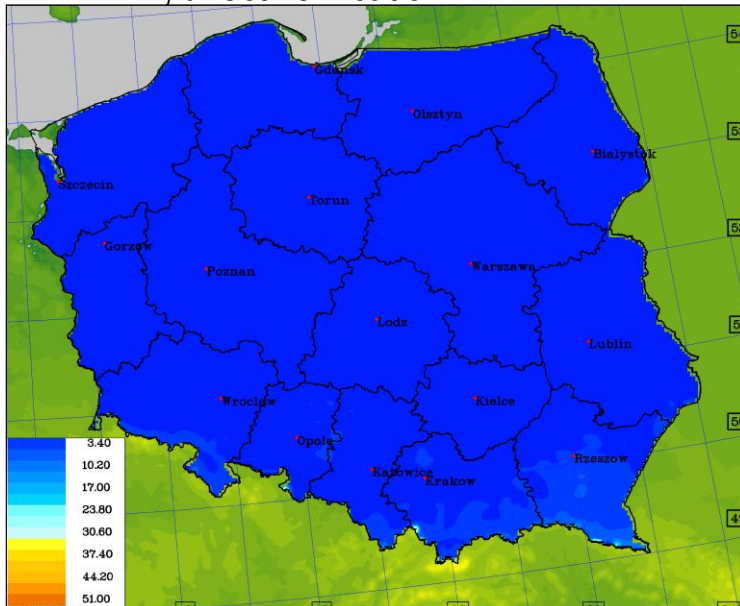
For both observations and forecasts –

- FLR max_value_{domain} > 20 strikes/hour

The duration of the storm must be > 6 hours

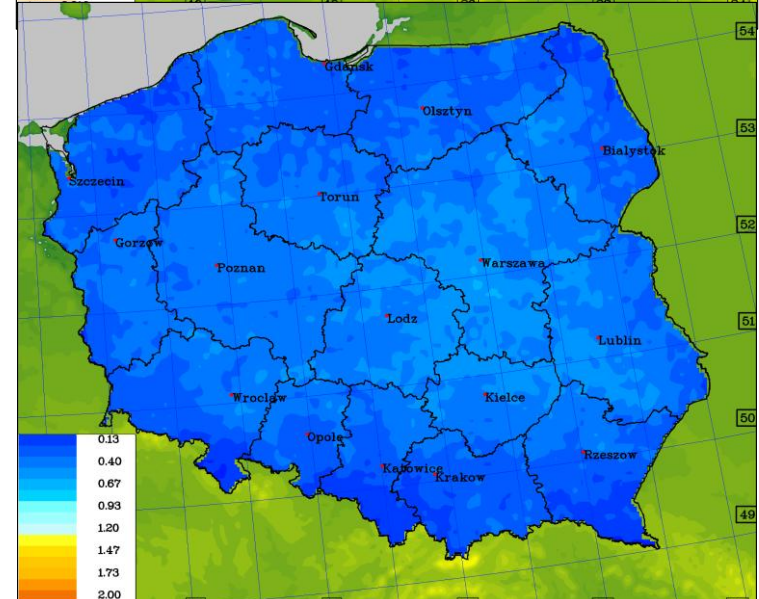
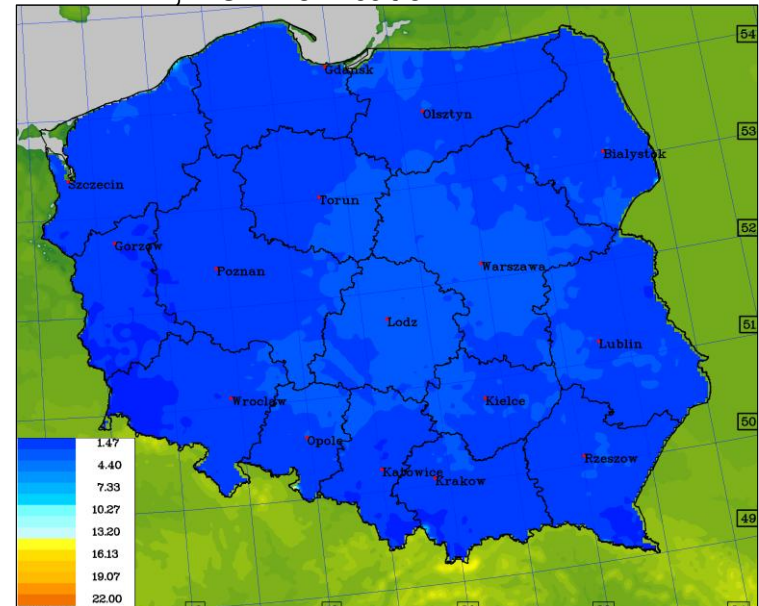
Categorical analysis based on contingency tables

FBI, direct verification



POD, direct verification

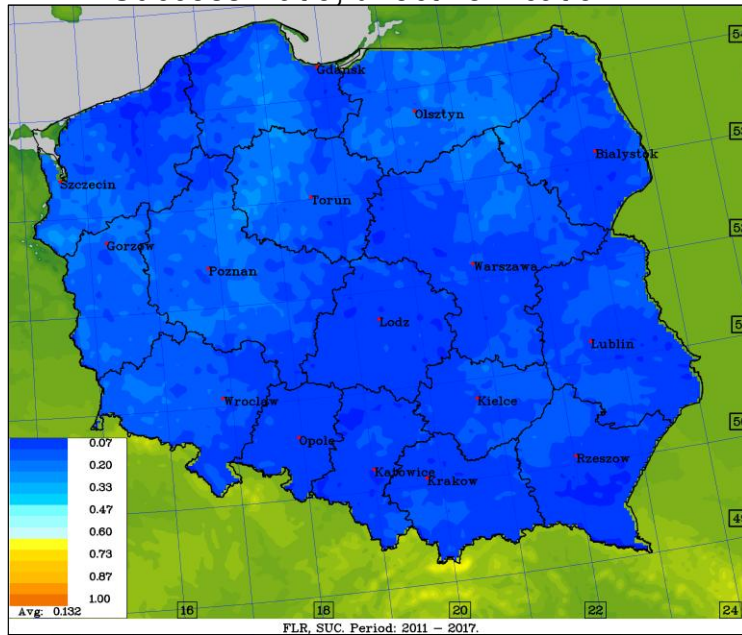
FBI, VOD verification



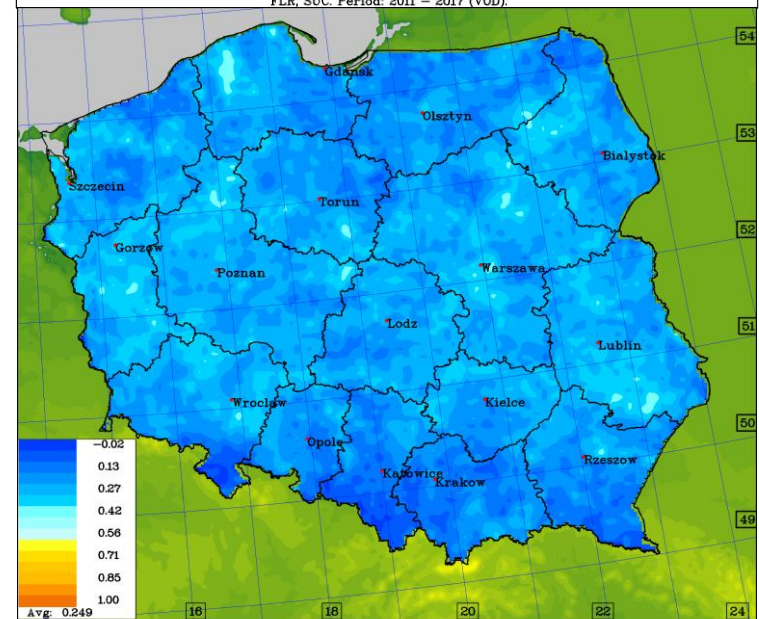
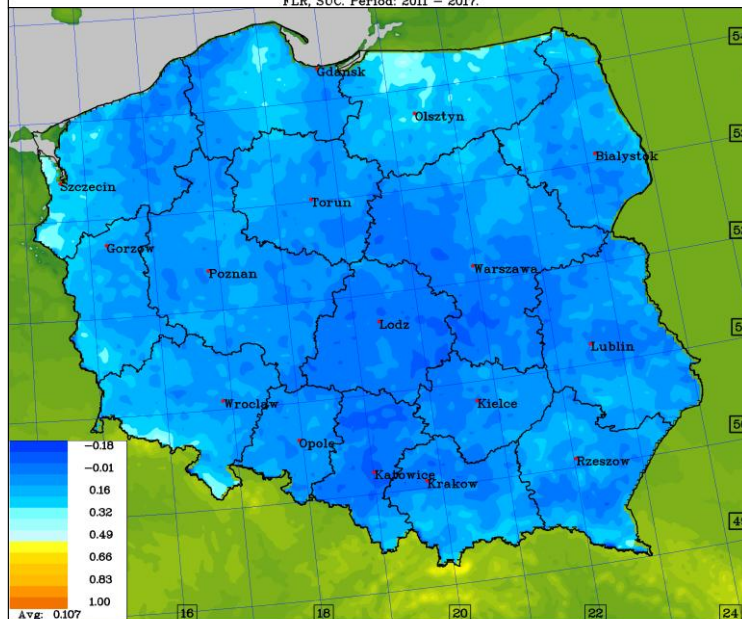
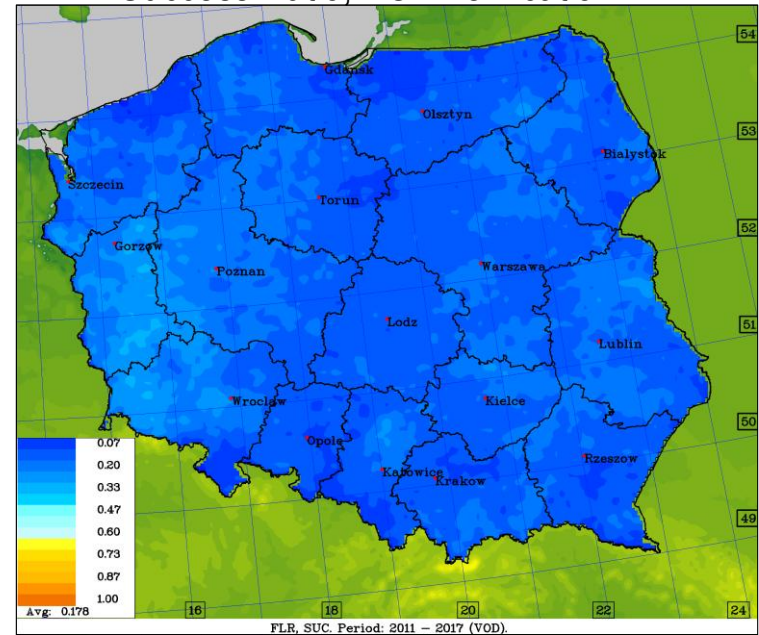
POD, VOD verification

Categorical analysis based on contingency tables

Success Ratio, direct verification



Success Ratio, VOD verification



Threat Score, direct verification

Threat Score, VOD verification

Categorical analysis based on contingency tables

	EQS		FAR		FBI		PFD	
	<i>Direct</i>	<i>VOD</i>	<i>Direct</i>	<i>VOD</i>	<i>Direct</i>	<i>VOD</i>	<i>Direct</i>	<i>VOD</i>
<i>2012</i>	0.0302	0.0842	0.8832	0.8240	2.7196	2.3366	0.1736	0.1611
<i>2013</i>	0.0773	0.1140	0.8254	0.7920	2.4679	2.1431	0.1483	0.1232
<i>2014</i>	0.0299	0.0671	0.9060	0.8632	3.4946	2.6446	0.1550	0.1258
<i>2015</i>	0.0263	0.1022	0.8785	0.7970	2.1706	1.8439	0.1311	0.1120
<i>2016</i>	0.0555	0.0751	0.8532	0.8370	2.7295	2.4354	0.1592	0.1344
<i>2017</i>	0.0505	0.0954	0.8296	0.7976	1.9107	1.6072	0.1180	0.0978
<i>Mean</i>	0.0420	0.0867	0.8676	0.8221	2.3164	1.9426	0.1499	0.1283
	POD		SUC		THS		TRS	
	<i>Direct</i>	<i>VOD</i>	<i>Direct</i>	<i>VOD</i>	<i>Direct</i>	<i>VOD</i>	<i>Direct</i>	<i>VOD</i>
<i>2012</i>	0.2366	0.4287	0.1169	0.1760	0.0826	0.1398	0.0754	0.2551
<i>2013</i>	0.3245	0.4685	0.1747	0.2081	0.1249	0.1667	0.2012	0.3202
<i>2014</i>	0.2193	0.3863	0.0940	0.1368	0.0681	0.1096	0.0935	0.2313
<i>2015</i>	0.1659	0.3890	0.1215	0.2030	0.0704	0.1543	0.0538	0.2579
<i>2016</i>	0.2644	0.3750	0.1469	0.1630	0.1030	0.1274	0.1299	0.2157
<i>2017</i>	0.1981	0.3433	0.1704	0.2025	0.0925	0.1452	0.1002	0.2253
<i>Mean</i>	0.2349	0.3987	0.1324	0.1779	0.0898	0.1390	0.1066	0.2489

The structure component S investigates the size and shape of event objects. It is defined in the range $[-2 \cdots + 2]$, where negative values correspond to too small and/or too peaked objects, while **positive values indicate too large and/or too flat simulated objects**.

Corresponding to SAL, a value $S = 0$ indicates a perfect structure.

The amplitude component A evaluates the total amount of event occurrence in a predefined region.

The values of A are within $[-2 \cdots + 2]$, where 0 represents again the perfect value.

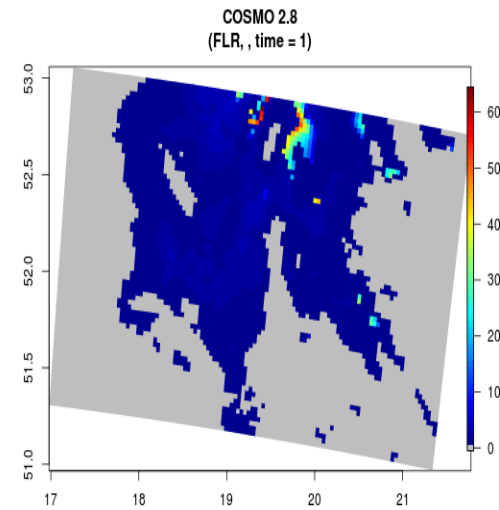
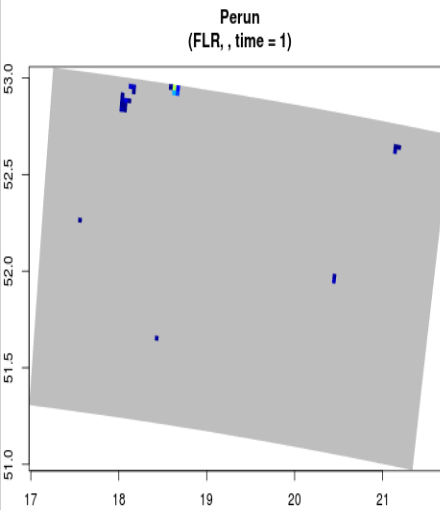
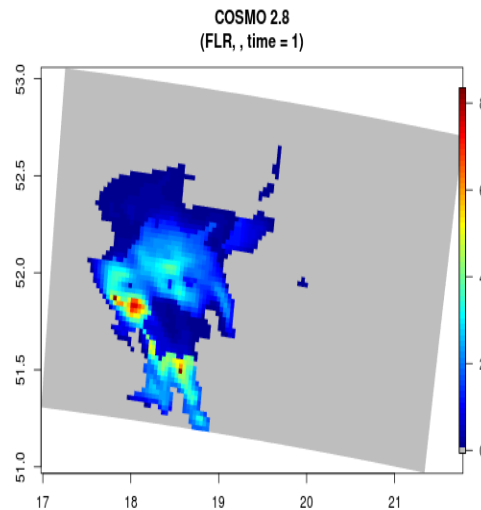
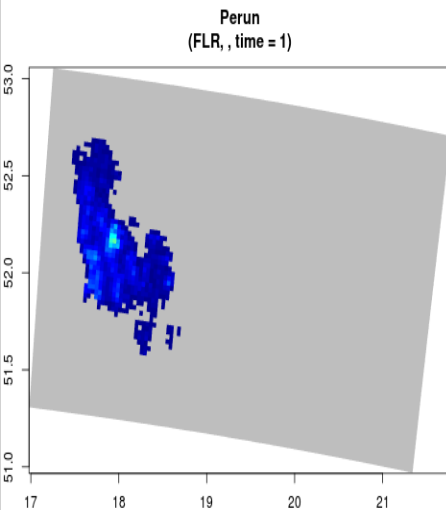
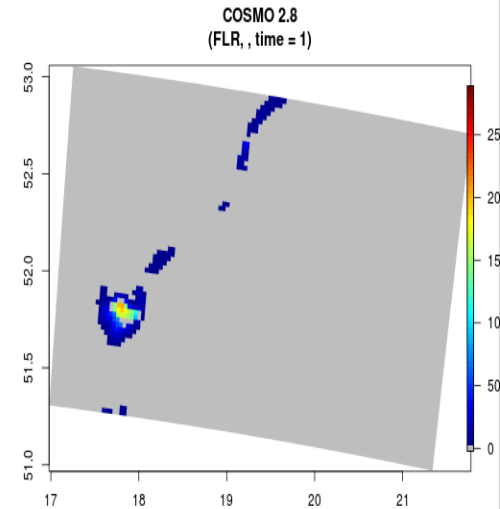
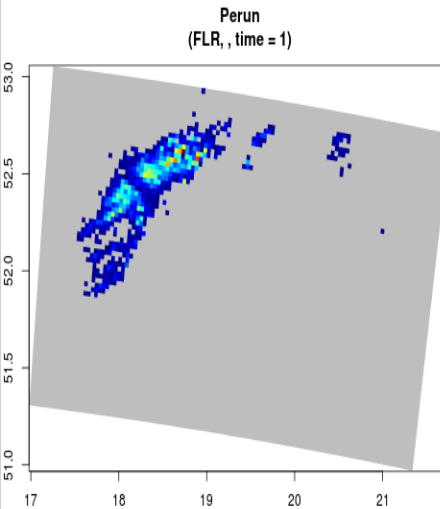
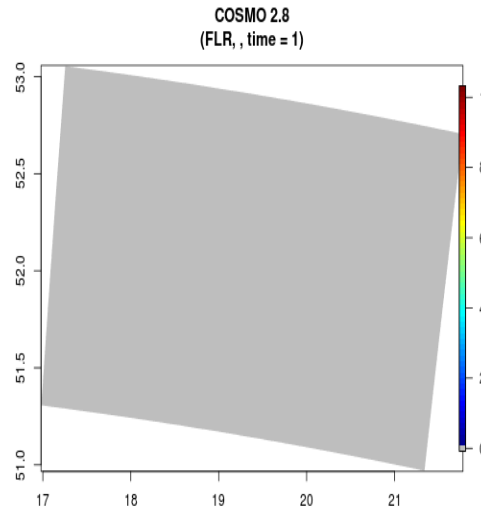
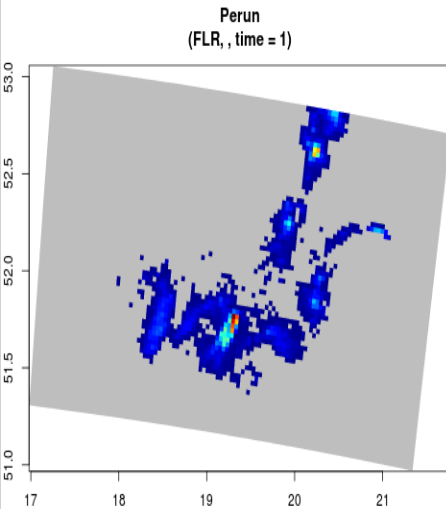
Negative values of A correspond to too little and **positive values to too much predicted event occurrence**, respectively.

The location component L quantifies the displacement of observed and simulated precipitation objects, relative to their overall centers of mass.

The values of L are within $[0 \cdots + 2]$ and also here 0 denotes the perfect value.

SAL input

Totally missed forecast



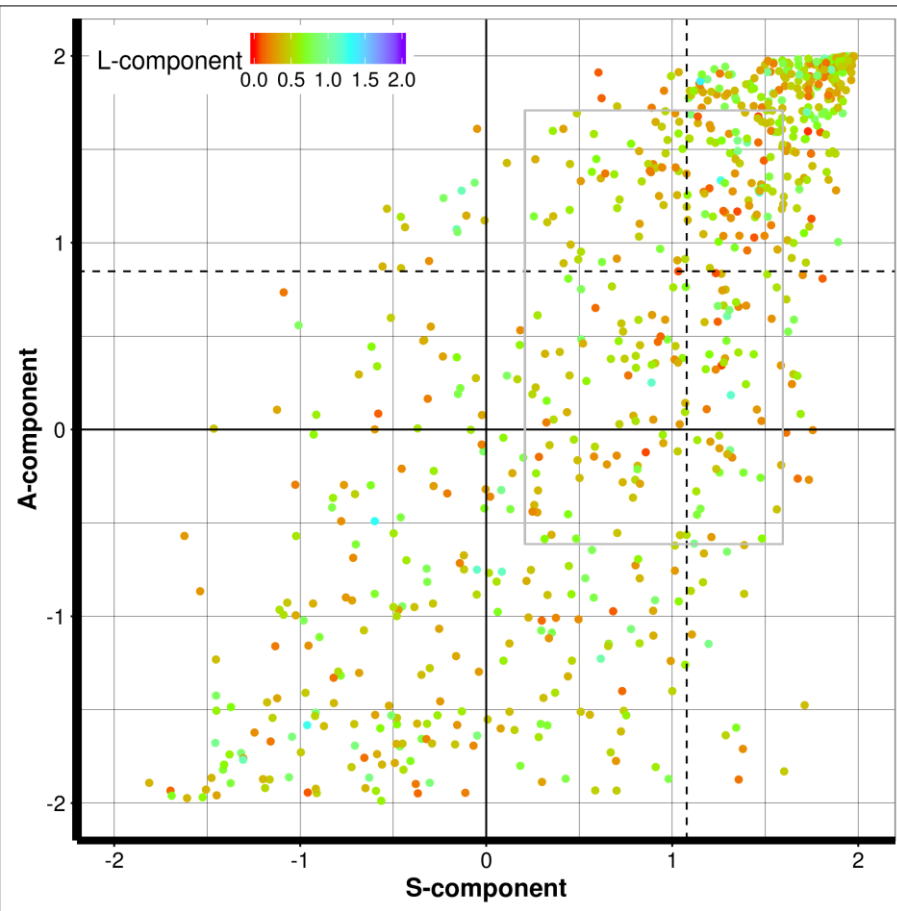
Overestimated forecast

Underestimated forecast

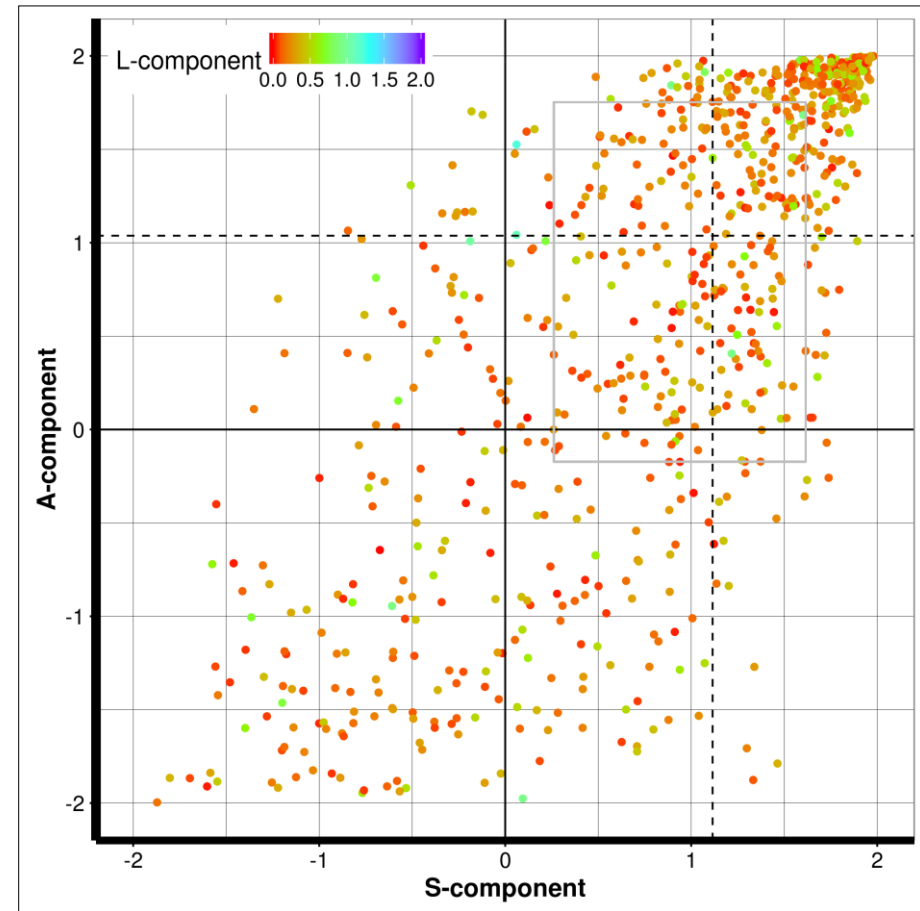
Totally false alarm

Examples (1)

SAL



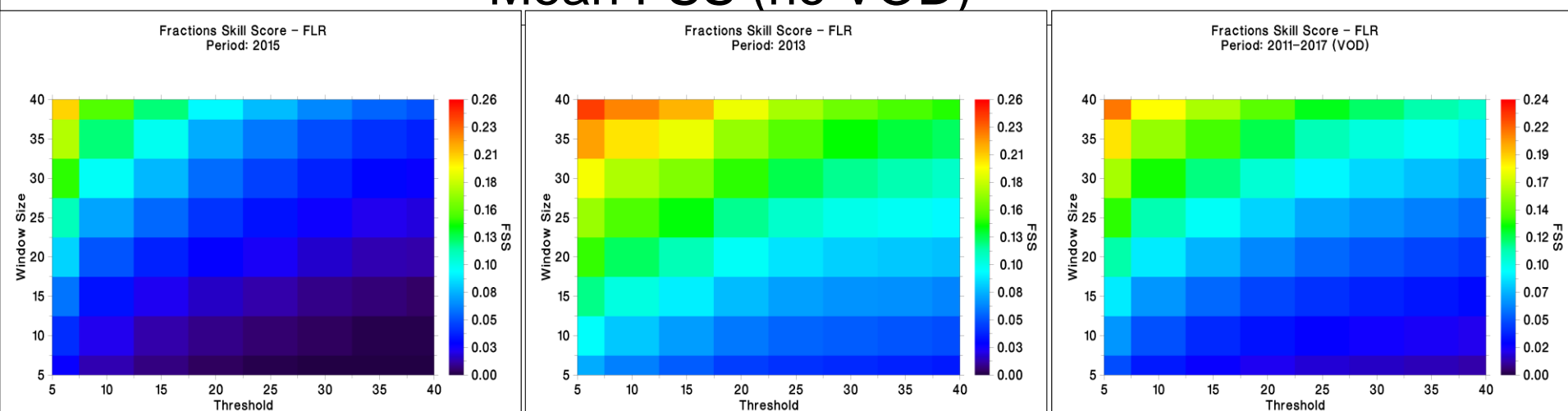
SAL with cross-correlation



All selected cases (2011-2017)

Dotted lines denote the median Structure- and Amplitude-component scores, resp.
The box corresponds to the 25 and 75 quartiles of S (x-axis) and A (y-axis) components.

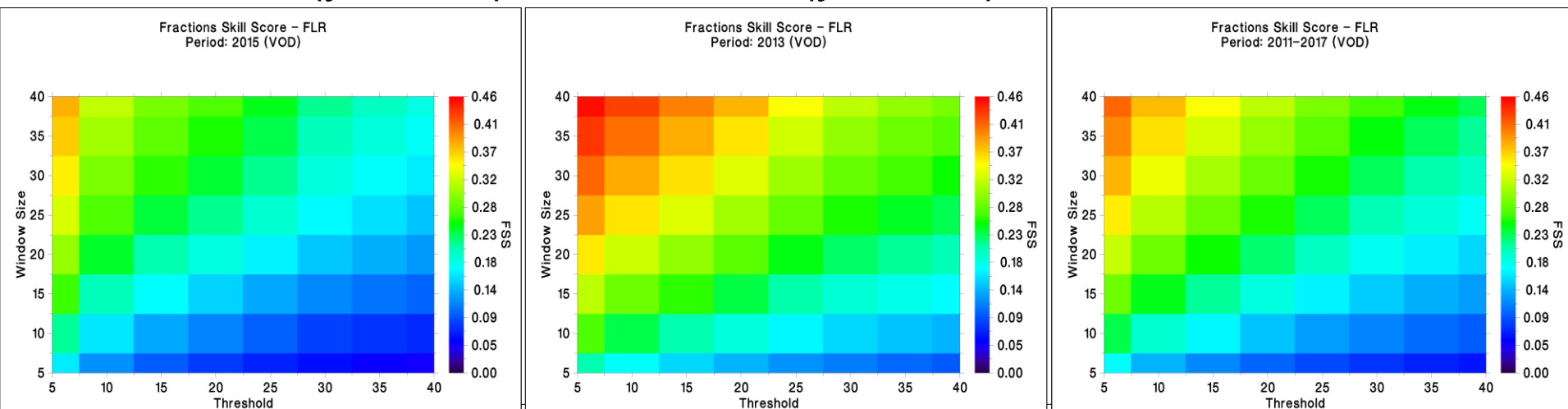
Mean FSS (no VOD)



Worst case (y. 2015)

Best case (y. 2013)

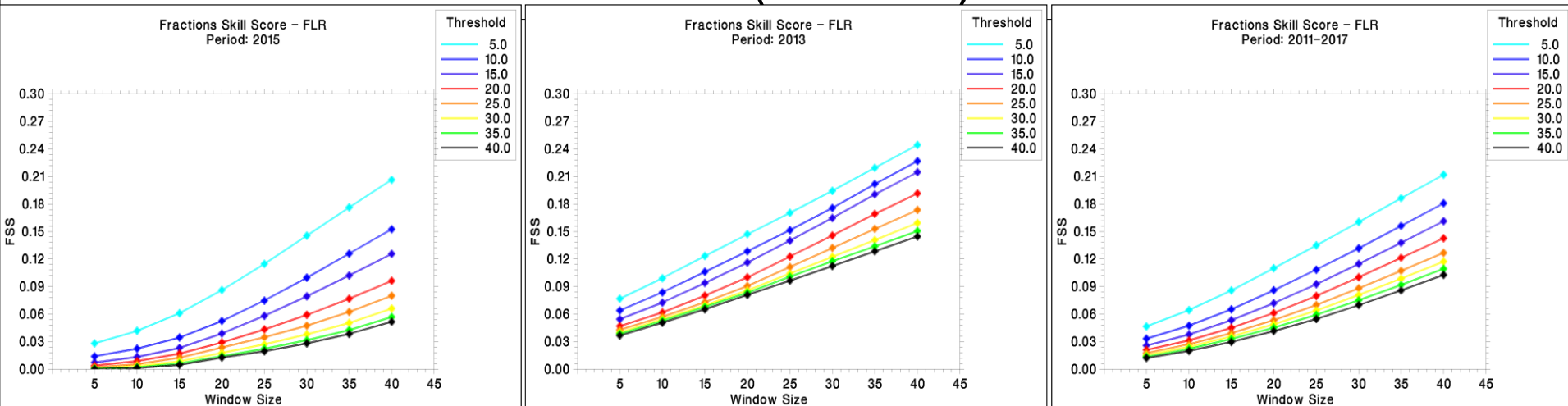
Mean for 2011-17



Mean FSS with VOD

Examples (1)

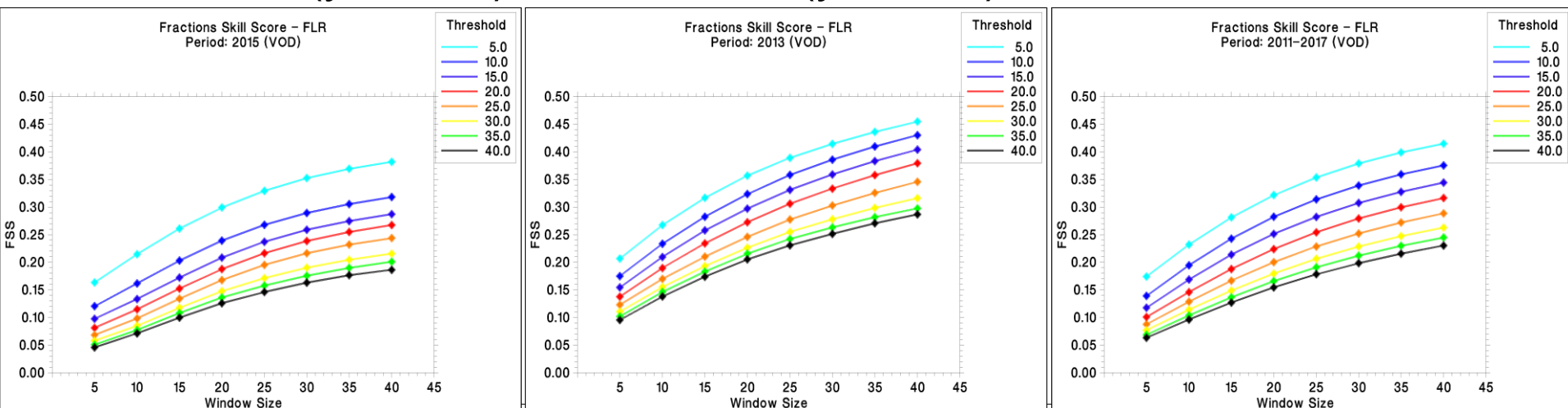
Mean FSS (no VOD)



Worst case (y. 2015)

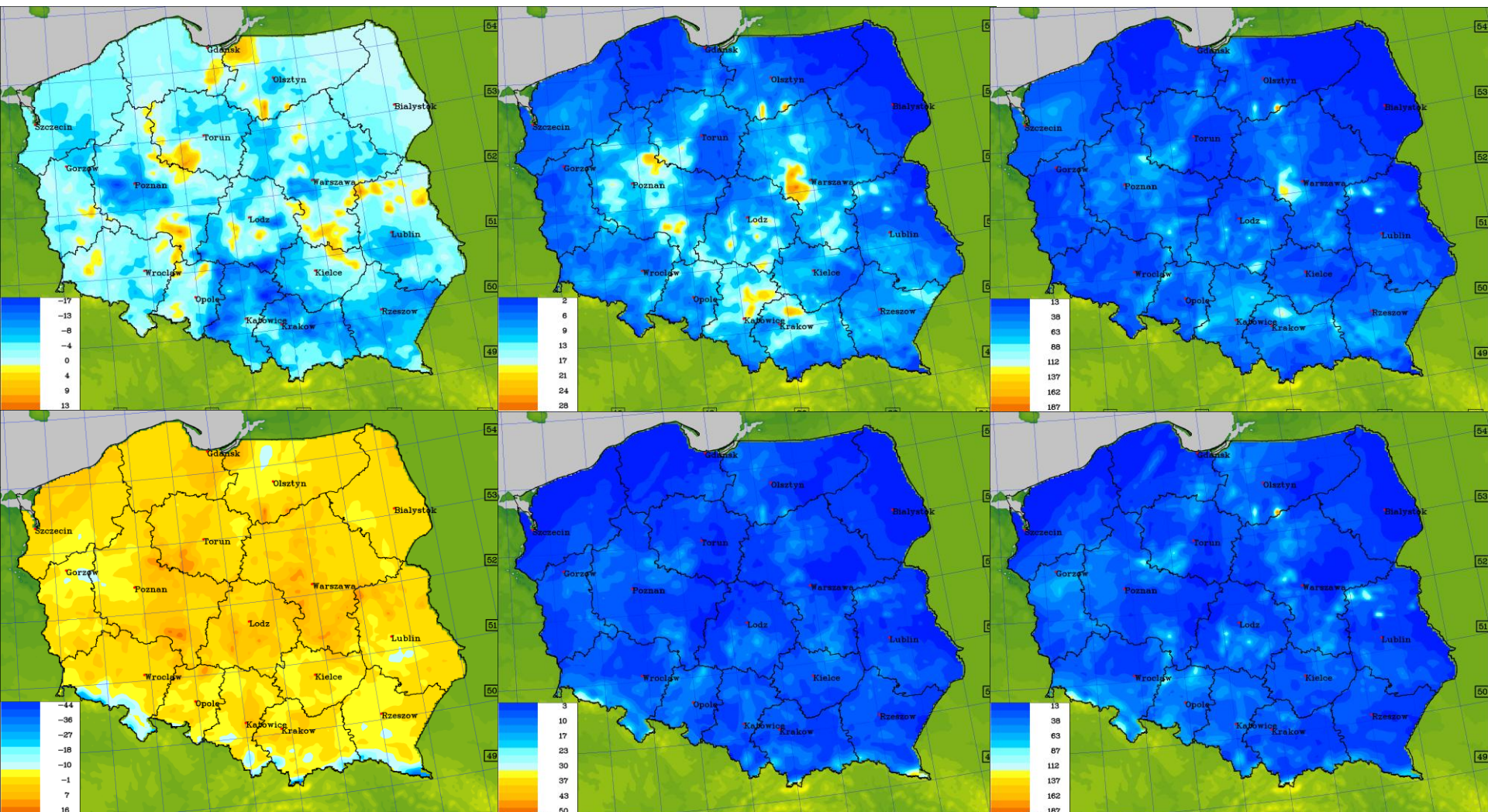
Best case (y. 2013)

Mean for 2011-17



Mean FSS with VOD

ME/MAE/RMSE 2013 (direct – upper, VOD – lower)

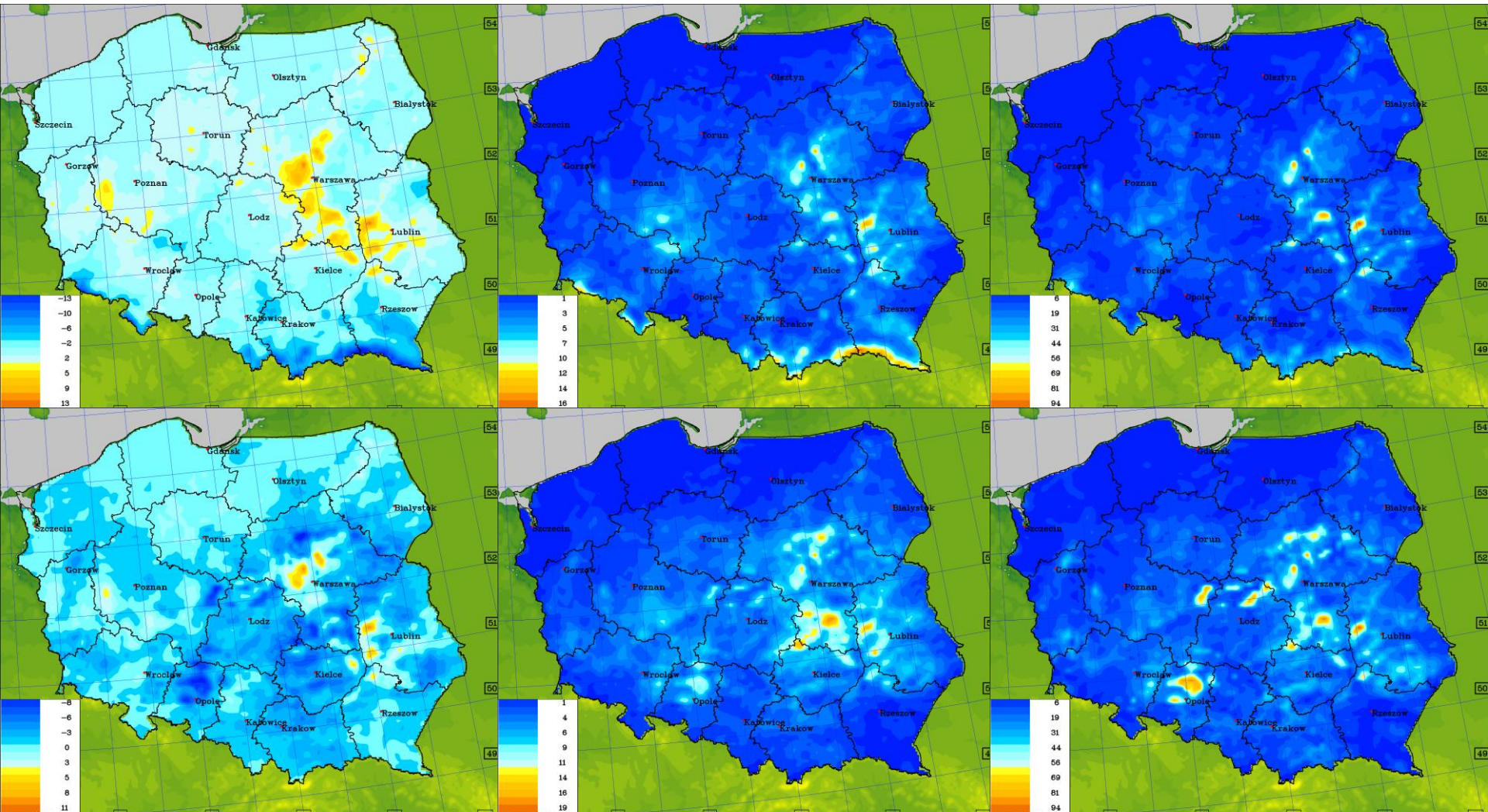


Mean Error

Mean Absolute Error

Root Mean Square Error

ME/MAE/RMSE 2017 (direct – upper, VOD – lower)



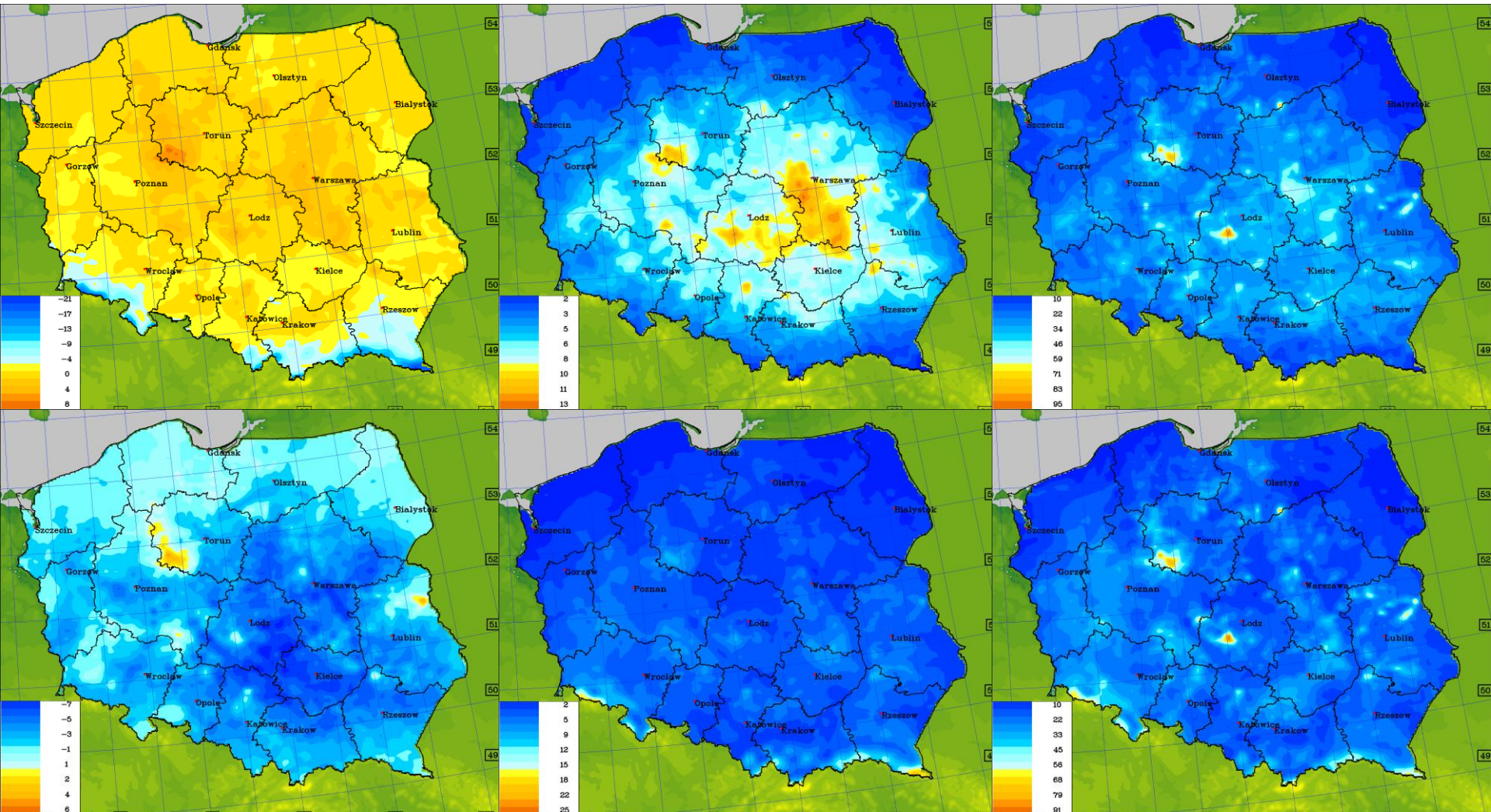
Mean Error

Mean Absolute Error

Root Mean Square Error

Examples (3)

ME/MAE/RMSE 2011-2017 (direct – upper, VOD – lower)



Mean Error

Mean Absolute Error

Root Mean Square Error

ME/MAE/RMSE with vs. w/out cross-correlation

	Direct			VOD		
Year	ME	MAE	RMSE	ME	MAE	RMSE
2011	2.128	4.712	18.904	1.887	4.213	18.051
2012	-2.811	5.913	18.866	-3.681	5.027	17.482
2013	-3.674	2.184	10.556	1.078	1.949	9.970
2014	-3.712	1.516	9.186	-2.192	1.374	8.960
2015	-2.023	2.025	11.871	-3.722	1.819	11.391
2016	-2.291	3.360	14.695	-0.699	2.950	13.904
2017	-1.286	2.817	12.761	-0.176	2.015	11.879
2011-2017	-1.953	3.218	13.834	-1.071	2.764	13.091

1. VOD improves categorical predictands (FBI, POD, THS...) by ~10 upto 45%.
2. SAL – VOD forces some improvement in L-component and (to some extent) in A-component. S-component to a large extent remains unchanged. Forecasts are evidently overestimated. Smaller domain (SAL is more effective) and more cases selected – no significant improvement...
3. FSS – results are not very impressive. VOD, however, **significantly** improves it.
4. MAE/RMSE (direct comparison) – The worst values in mountainous regions – hard(er) to predict thunderstorms?
5. MAE/RMSE w. cross-correlation – slight improvement compared to direct verification, maxima moved towards domain centre.
6. Discrete vs. continuous verification?

