IMGW activities in PP MILEPOST

Andrzej Mazur, Grzegorz Duniec, Joanna Linkowska

Institute of Meteorology and Water Management
National Research Institute

Contents

1. Introduction
2. Database from DWD – preliminary activities in common verification setup
3. Operational ANN for EPS/time-lagged forecasts – some results
4. Weighted MLR – an alternative for ANN – some results
5. Done, to-dos

Introduction

Three main points to be presented:

Subtask 2.1: Set-up and application of ANNs

Subtask 2.2: Set-up and application of other Machine-Learning techniques

Task 3. General ML-based post-processing and verification. Definitions of comparison setup to establish an evaluation framework

Database from DWD (for common verification setup, task 3.2)

758 WMO stations (with WMO no. Provided), mostly in Germany.

Data given in the nearest neighbour grid point of COSMO-DE/2-EPS.

Variables ending with "_MS„/"_LS" – medium/large scale predictors – means of 11*11 resp. 21*21 grid points around the stations.

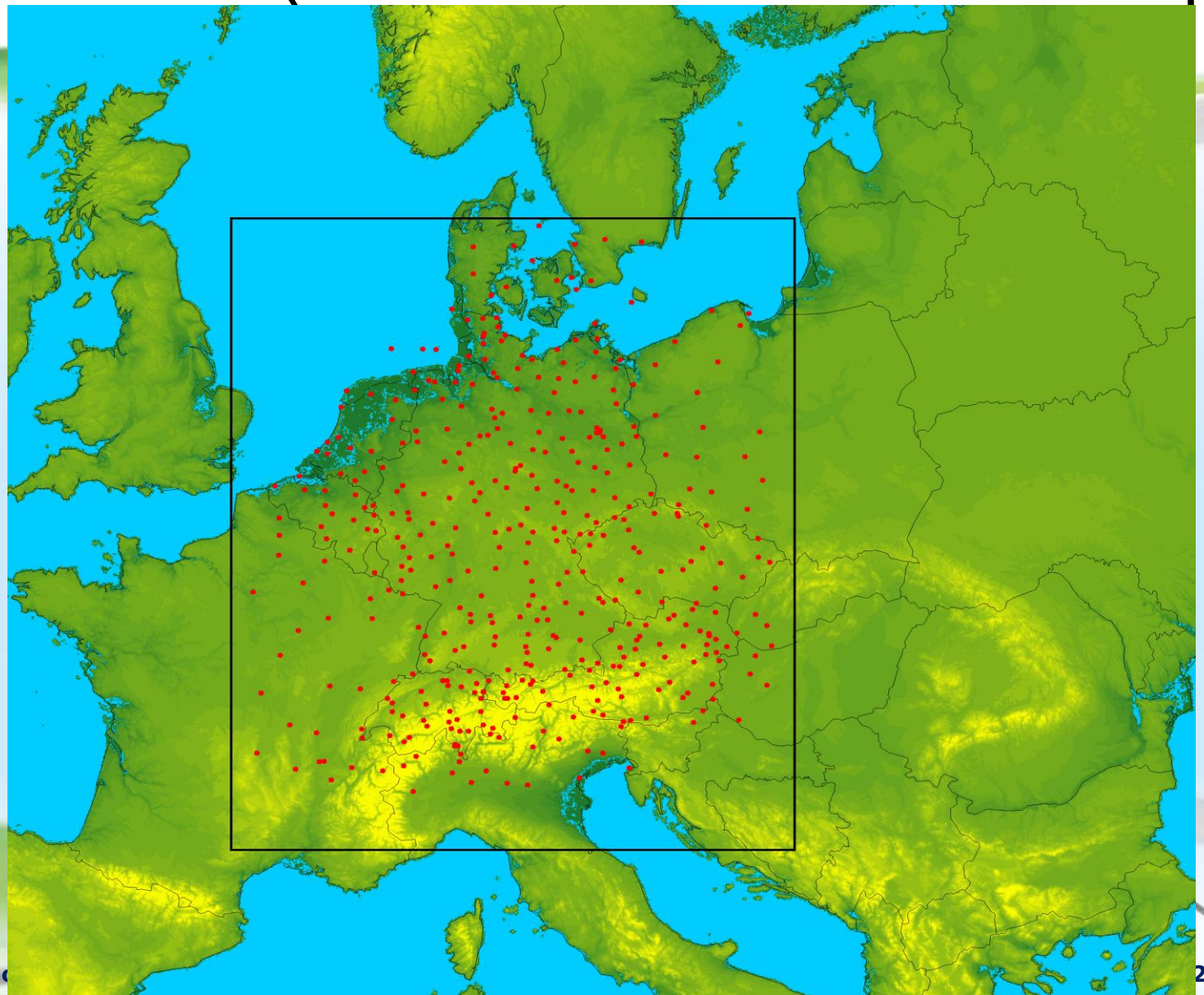
For COSMO-DE – area means of 28*28/ 56*56km.

For COSMO-D2 – area means of 22*22/44*44km.

Variables ending with "_MS_S„/"_LS_S" – standard deviations over resp. areas.

Data gathered from COSMO-DE-EPS (Dec. 2010 up to 14. May 2018) and from COSMO-D2-EPS since 15 May 2018 until Dec. 2020.

Database from DWD (for common verification setup, task 3.2)



Database from DWD (for common verification setup, task 3.2)

Variables:

- TMIN_2M, TMAX_2M, T_G;
- VMAX_10M;
- CLCT, CLCL, CLCM, CLCH;
- PMSL;
- U_10M, V_10M;
- T_2M, TD_2M;
- RAIN_GSP, SNOW_GSP, TOT_PREC;
- HBAS_SC, HTOP_SC;
- ASOB_S, ATHB_S, ALB_RAD;
- W_SNOW;

At 500, 700, 850, 950 and 1000 hPa levels:

- Temp, RelHum, Geopot, U/V/Omega

Column-integrated Soil Moisture at **1, 2, 6, 18 and 54cm**.

Operational ANN for EPS/time lagged ANNs

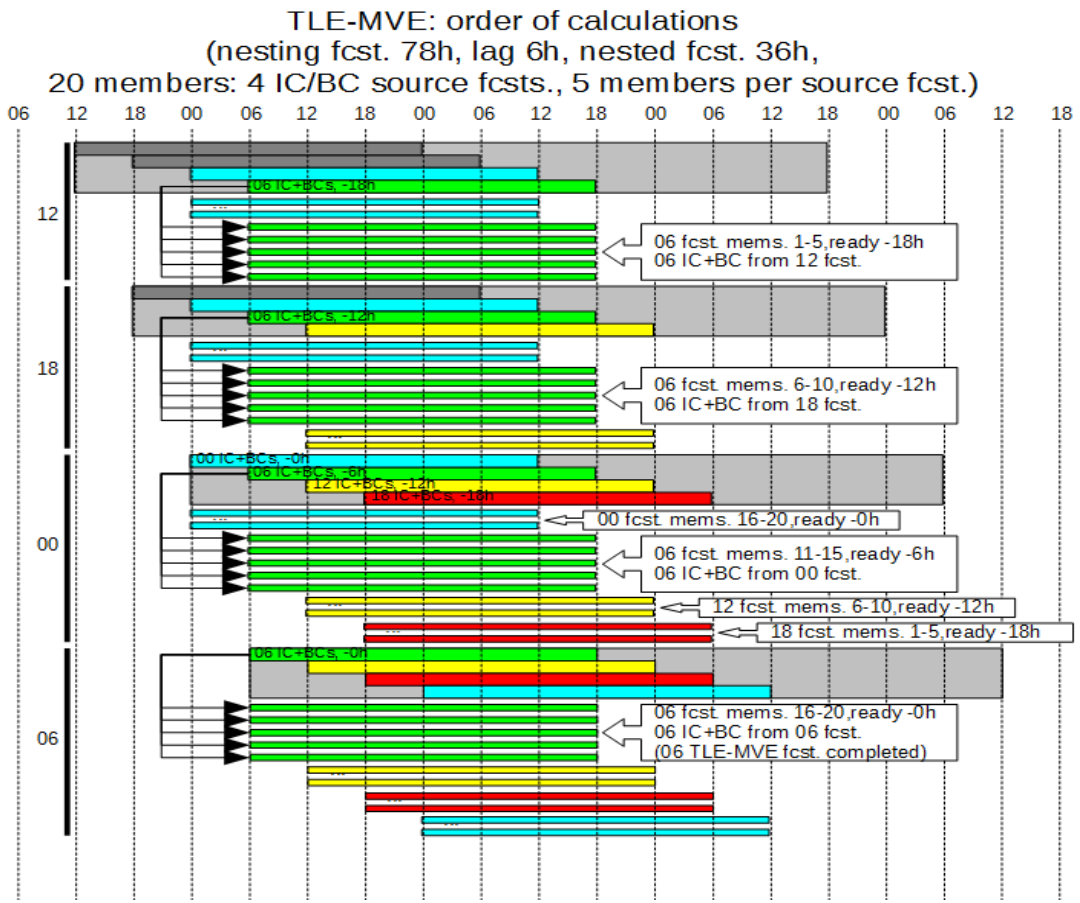
Selecting an appropriate subset of predictors and overall setup of the ANN

Setup:

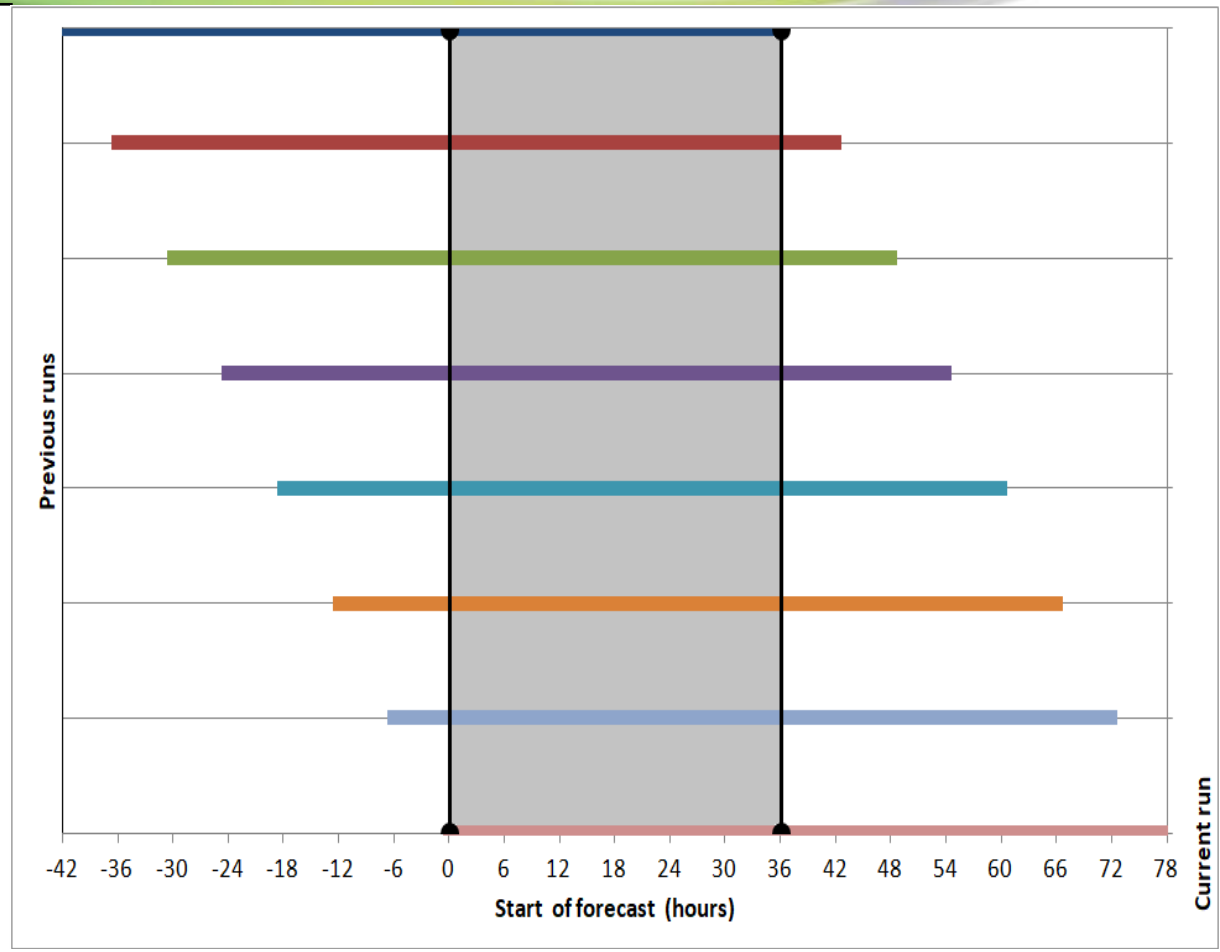
1. EPS-ANN: 25 input neurons (20 members + λ,φ,h,t_s); 5 neurons in a single hidden layer (four blocks of TL-ICs/BCs and spatio-temporal coordinates – blocked).
2. det-ANN: 12 input neurons (8 members – 36 hours + λ,φ,h,t_s); 2 neurons in a single hidden layer (referring to a single block of det-DMOs and spatio-temporal coordinates – blocked).
3. Every forecast (temperature, wind speed, pressure, etc.) treated independently.
4. Activation function: hyperbolic tangent (symmetric with respect to 0,0).
5. Training method: backward propagation of errors (back-prop).
6. Optimization: gradient descent.

Operational ANN for EPS/time lagged ANNs

Selecting an appropriate subset of predictors and overall setup of the ANN



EPS – ANN



det – ANN

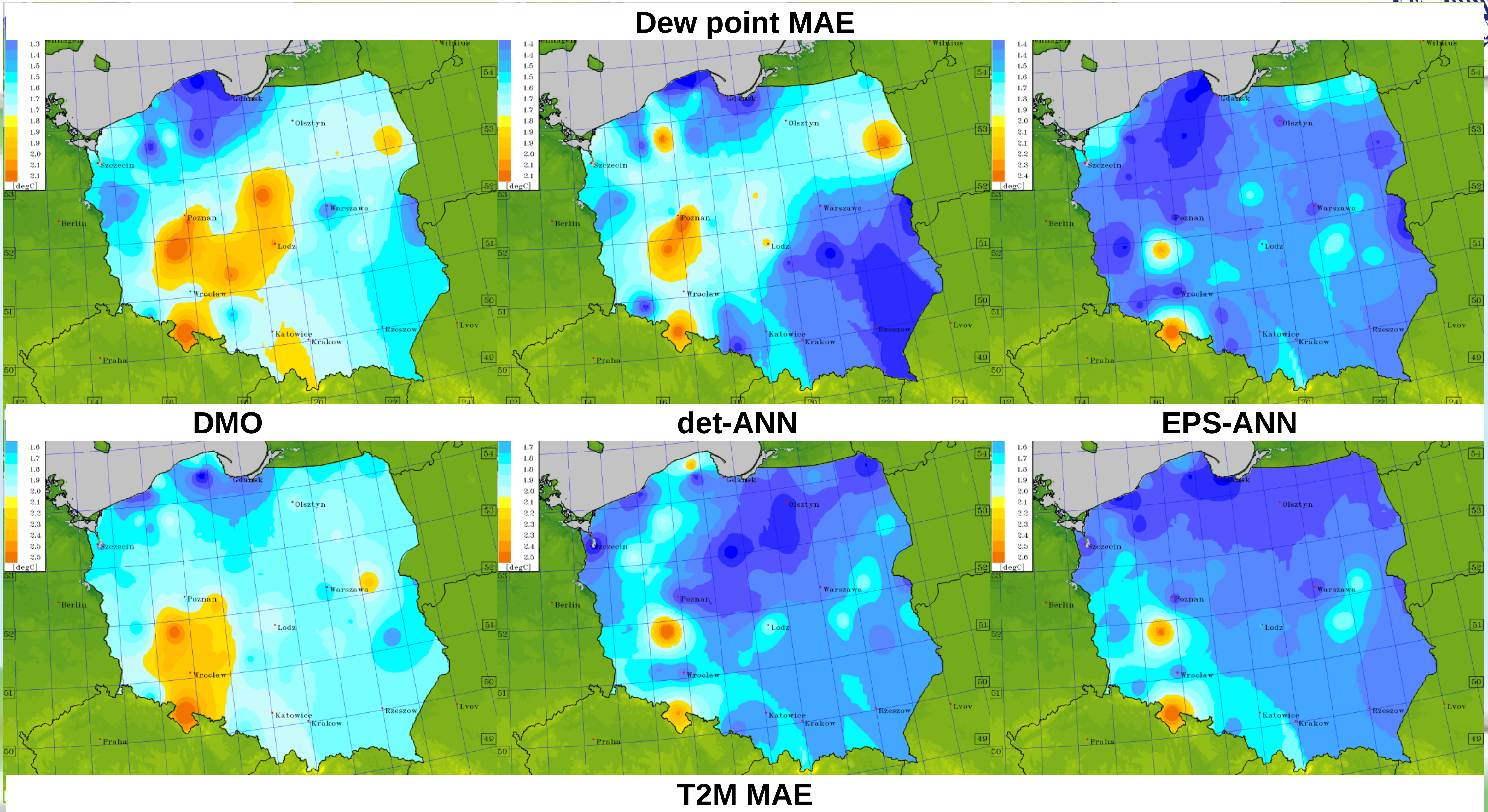
Operational ANN for EPS/time lagged ANNs

Time-lagged ANNs – necessary/sufficient condition for a proper network? Deterministic vs. EPS-based ANNs.

Operational results (2011 – current)

Means	ME	MAE	RMSE	MinE	MaxE
Dew point					
DMO	0.299	2.618	4.021	-	-
det-ANN	-0.412	2.214	3.618	-16.2	17.1
EPS-ANN	-0.271	2.101	3.263	-14.1	16.9
Air temp.					
DMO	0.953	2.953	4.619	-	-
det-ANN	0.651	2.740	3.921	-17.1	19.1
EPS-ANN	0.219	2.603	3.682	-15.2	17.8
Windspeed					
DMO	-0.837	2.023	3.150	-	-
det-ANN	0.351	1.759	2.719	-9.4	16.0
EPS-ANN	-0.425	1.572	2.236	-8.8	14.1

IMGW activities in PP MILEPOST



WMLR – an alternative for ANN

$$f(x) = y = a \cdot x + b$$

$$\chi^2(a,b) = \sum_{i=1}^{n} \frac{(y_i - a \cdot x_i - b)^2}{\sigma_i^2}$$

$$w_i = 1/\sigma_i^2$$

$$\frac{\partial \chi^2}{\partial a} = 0 = -2 \sum_{i=1}^{n} w_i \, x_i (y_i - a \cdot x_i - b)$$

$$\frac{\partial \chi^2}{\partial b} = 0 = -2 \sum_{i=1}^{n} w_i \, (y_i - a \cdot x_i - b)$$

$$S = \sum_{i=1}^{n} w_i \qquad S_x = \sum_{i=1}^{n} w_i x_i \qquad S_y = \sum_{i=1}^{n} w_i y_i \qquad S_{xy} = \sum_{i=1}^{n} w_i x_i y_i \qquad S_{xx} = \sum_{i=1}^{n} w_i x_i^2$$

$$\Delta = S \cdot S_{xx} - (S_x)^2$$

$$a = \frac{S \cdot S_{xy} - S_x \cdot S_y}{\Delta}$$

$$b = \frac{S_{xx} \cdot S_y - S_x \cdot S_{xy}}{\Delta}$$

”The method of least squares is used not because we consider it mathematically certain, but because no-one has suggested a better method so far.”

WMLR – an alternative for ANN

Initial weights – sigma values calculated from archive measurements vs. forecasts for initial period.
From the ”zero” approximation – compute the value of sum of (squares of) residues.

$$RES = \sum_{i=1}^{n} (y_i - a \cdot x_i - b)^2$$

Set an arbitrary threshold. If RSS is above it:

$w_i = 1/\sigma_i^2$	$w_i = w_i(t)$
Recompute weights *via* extending period back in time*)	Recompute weights *via* assuming a time dependency*)

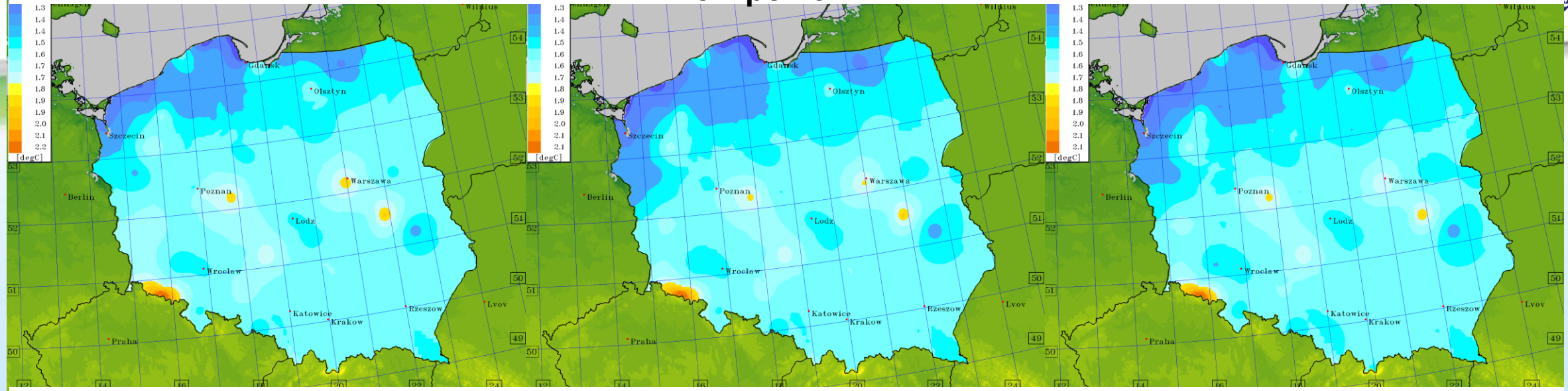
WMLS recalculation must be performed with changed weights to achieve the assumed convergence criterion

*) It means both SD and points to LS equations. Multi-dimensional problem would require time-consuming calculations. Some constraints may become necessary since the procedure is to be repeated on a regular basis.

*) The form of the time dependency should be determined arbitrarily, but with good justification.
As the first approximation, a linear relationship was adopted (the older the data, the less important – ”forgetting” model)

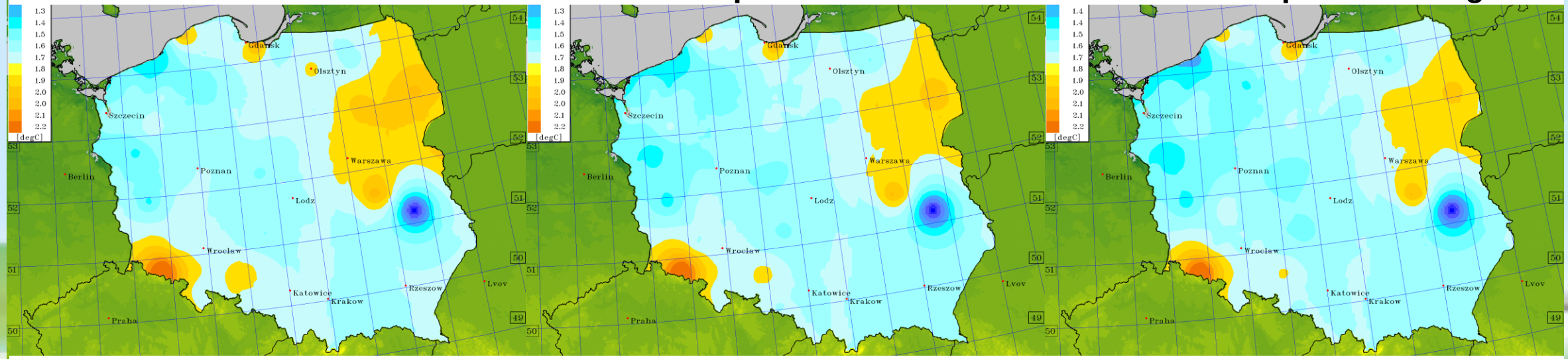
WMLR – an alternative for ANN – some results

Dew point MAE



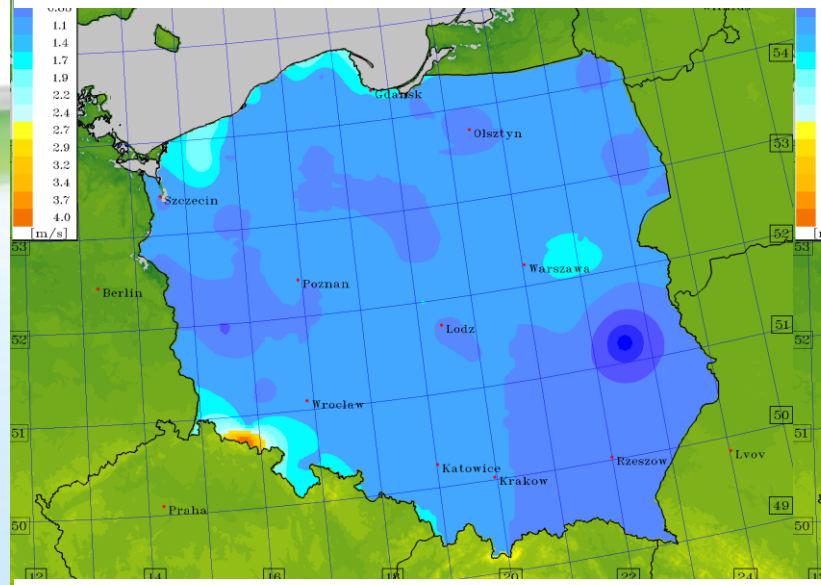
DMO | time period extended | time-dependent weights

T2M MAE

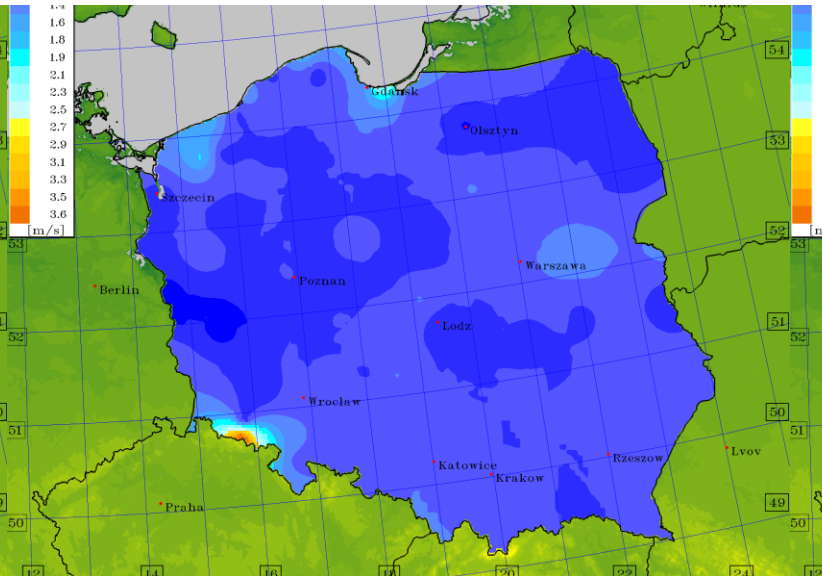


WMLR – an alternative for ANN – some results

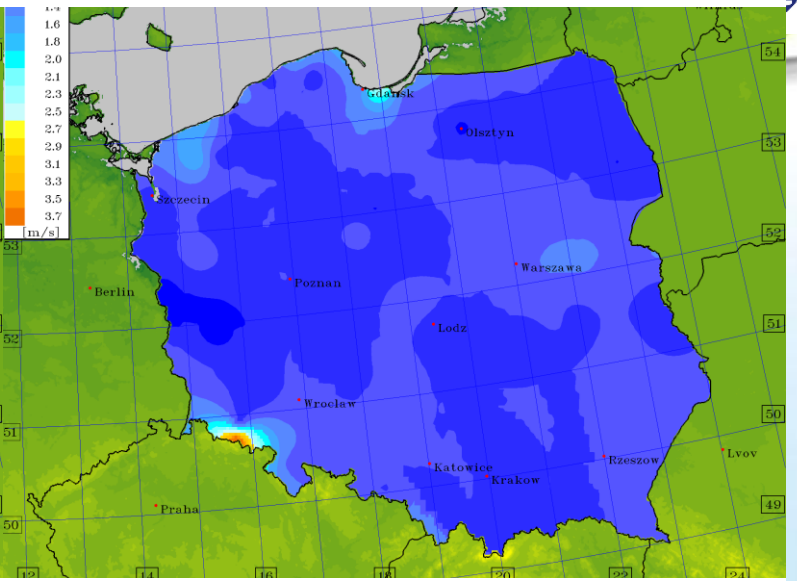
U10M MAE



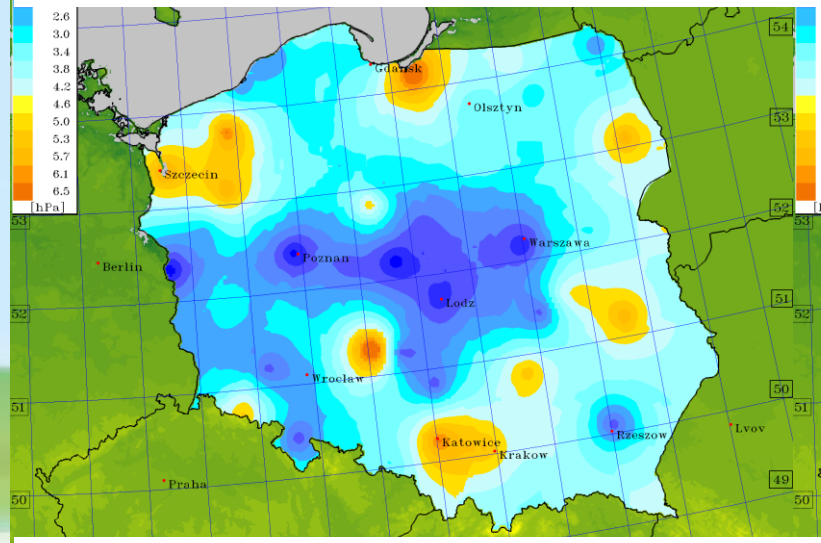
DMO

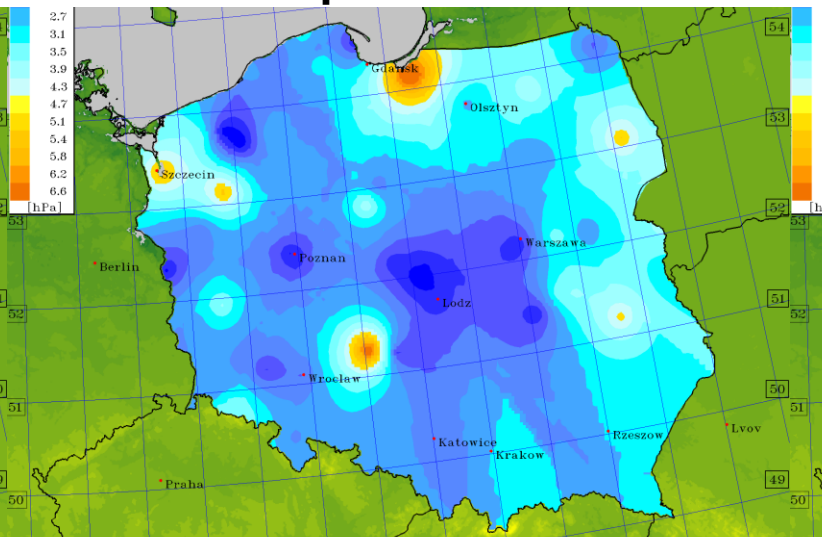


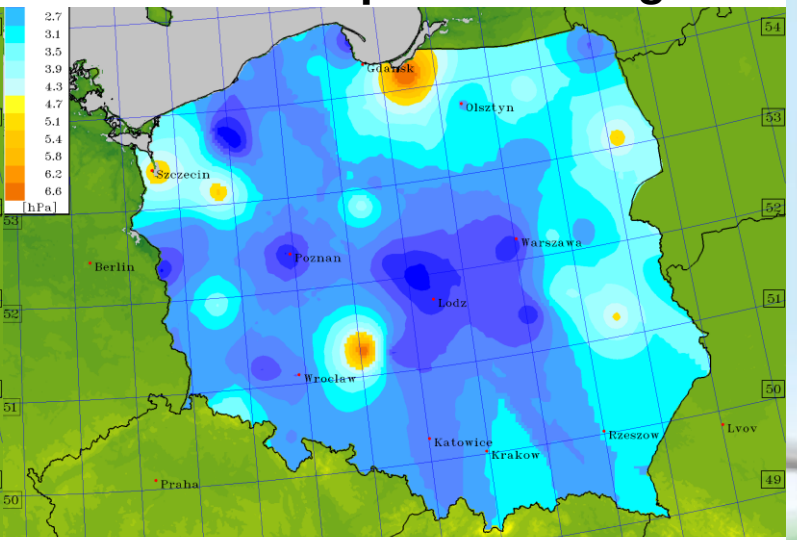
time period extended



time-dependent weights







PS MAE

Done, to-dos

1. Constant work on operational ANNs

2. Alternative solution – RLS is still ”in progress” – time-dependent weights a bit better then sigma-based ones?

3. EPS vs. deterministic ANNs – EPS-based ANN produces more precise results. Perhaps it’s connected with the greater number of members, which, on the other hand, translates into longer times of learning and testing of network.

4. Common verification and testing – suggested area/data/period – Central Europe, DWD data available for 2011-2020 – ready to use!


ONE FLAG. ONE WORLD.
take required precautions while
stepping out of your nest .