Stay Report

To: Zentralanstalt für Meteorologie und Geodynamik — ZAMG, Vienna, Austria
Period: 11th November – 6th December, 2019
Topic: Work on analog-based post-processing method
Supervisors: Mag. Alexander Kann and Irene Schicker, PhD

Introduction

I stayed at the Zentralanstalt für Meteorologie und Geodynamik (ZAMG) for four weeks during which I was working on the analog-based post-processing method applied to an NWP model output for point and gridded forecasts. This is the continuation of previous work carried out during stay 1 (13/11-09/12, 2017), stay 2(02/02-03/03, 2018) and stay 3 (04/02-01/03, 2019), where the basic algorithm in Python was written and the usability of the analogs method investigated for Austria. Thus, the method was already tested using the AROME deterministic model (1/1/2015-31/08/2017) and corresponding observations from 265 TAWES sites (1/1/2015-31/10/2017). Then, the same method is successfully applied to several configurations of different LAEF forecast:

a) **LAEF_Ws**: raw LAEF wind speed ensemble forecast (17 members)
b) **AnEn_Ws**: LAEF wind speed ensemble forecast used as predictors (17 predictors)
c) **AnEn_Mu**: The means of the LAEF ensemble forecast for the wind speed, direction, temperature (2 m), relative humidity, pressure and precipitation (6 predictors)
d) **AnEn_Std**: The means and the standard deviation of the LAEF ensemble forecast for the wind speed, direction, temperature (2 m), relative humidity, pressure and precipitation (12 predictors)
e) **AnEn_All**: All members of the LAEF ensemble forecast for the wind speed, direction, temperature (2 m), relative humidity, pressure and precipitation (6×17 predictors)

The experiments include 29 sites and the results are provided for January and July 2017 (using the 2015-2016 testing period).

Further development prior to this stay

Several authors in more recent work show that, instead of assigning the same importance to each predictor variable, the brute-force weight optimization can increase the AnEn performance. Even though it is the best possible approach, due to the limited computational resources, not all the possible combinations are tested in this work. The forward selection algorithm is used instead, starting with weight value fixed at 1 for the wind speed parameter. Six ALADIN-LAEF parameters are tested using the forward selection algorithm one after another, in the same order as listed. Five possible weight values (0.00, 0.25, 0.50, 0.75 and 1.00) are investigated for each predictor variable. The predictor weighting strategy is carried out for January and July 2017, using the 2015-2016 period for the training. Therefore, the optimization procedure uses a completely independent dataset from the period for which training, as well as for which forecasting is performed (January and July 2018).

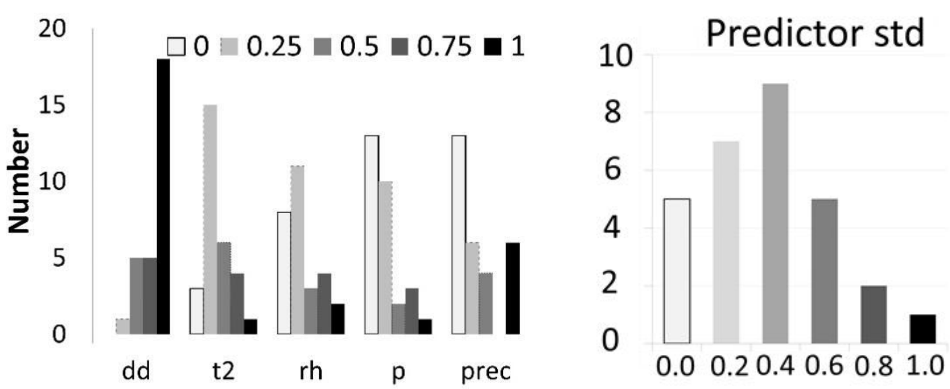

Figure 1. The histogram of the optimized weights for each predictor tested (using the AnEn mean values)(left) and of the optimized weights for standard deviation predictor for different meteorological parameters (right), at 29 stations in Austria in January and July 2017.

The results show that the wind direction is the most important predictor in addition to wind speed, followed by temperature and relative humidity parameters. The pressure and precipitation parameters are often optimized with the 0.00 weight, meaning that they are not carrying additional benefits at certain stations. Additionally, the optimal contribution of the standard deviation predictors is about 40 % of the ensemble mean predictors' contribution in the majority locations tested.

The optimal weights are included in all previously mentioned experiments. For that reason, the testing period is shifted from January and July 2017 to January and July 2018.

New results regarding LAEF input to the analog method

During one of the previous stays, one experiment (Member-by-Member approach **AnEnMem**) conducted contained, unfortunately, a bug. The problem occurred due to unsorted pasting the data within a loading module. In the previous version of Pandas, the sorting was done ascendingly (e.g. at the version used at ZAMG guest machine), but this was changed in a more recent version of pandas. Therefore, the algorithm is updated and this experiment is repeated.

Additionally, another experiment is conducted – the „deterministic“ analog approach using six LAEF meteorological parameters, but only the control member: **AnEnCtrl**. The overview of all results is shown in Tables 1 and 2, and in Figure 2. The results are compared to two ensemble model output statistics (EMOS) experiments:

- the **EMOSws** only uses the last 30 days as training and only the wind speed as an input,
- the **EMOSstd** uses all available training data and all variables including seasonal functions,

which are provided by M. Dabernig/ZAMG.

All six analog-based experiments are able to outperform the **LAEFws** and can reduce all three error sources for the ensemble mean (systematic - bias of the mean or σ; unsystematic – dispersion error). The new and most "simple" experiment in terms of input data, the **AnEnCtrl**, produce a similar result to **AnEnWs.** Also, it successfully removes the systematic errors in the bias and σ bias similar to the **EMOS** approach. Even more successful in removing the predominant dispersion source are the experiments with the additional predictors: **AnEnMu**, **AnEnStd**, and **AnEnAll**. In addition to improving the results for the ensemble mean, the average ensemble spread matches the average RMSE better after any post-processing.

The analog-search pool in the **AnEnMem** experiment is smaller than in other analog experiments since the search is performed dependently for the same ensemble member. Possibly, that is why the **AnEnMem** could not increase the skill of the raw probabilistic input, as one would inherit undesirable properties of the input model, such as under-dispersion and lower resolution issues. Additionally, **AnEnMem** is the most computationally expensive setup. So, even though the results are noticeably better after correcting the algorithm, it can still be considered as the least successful analog experiment. On the other hand, the "deterministic" **AnEnCtrl** experiment can still be considered as successful, since the improvement over the raw model is evident and the results comparable to **EMOS**, while keeping the algorithm simple and reducing the computational costs. Even though the **AnEnCtrl** and the **AnEnMu** use the same number of the meteorological parameters as predictor variables, the **AnEnMu** performs better for both months and at all lead times tested.

The **AnEnMu**, **AnEnStd**, and **AnEnAll** experiments show a nearly similar improvement measured by BSS and CRPS. The better overall results are achieved for July than in January when wind speed and its variance is higher on average. The other post-processing approaches improve LAEF forecast less.

*Table 1. The average values and confidence interval (0.95 sig. level) of several verification measures for the different models at all available stations in Austria and all lead-times during January 2018. The best result among compared forecasts is underlined (the spread is better when closer to the RMSE value). The values significantly different from the **AnEnStd** forecast (0.05 sig. level) are marked with an asterisk sign.*

January	*LAEFw*	*EMOSws*	*EMOSstd*	*AnEnCtrl*	*AnEnWs*	*AnEnMu*	*AnEnStd*	*AnEnAll*	*AnEnMem*
Bias [ms^{-1}]	-0.210* *[-0.232, -0.185]*	-0.053* *[-0.069, -0.039]*	-0.160* *[-0.174, -0.146]*	-0.060* *[-0.072, -0.046]*	-0.036 *[-0.048, -0.022]*	-0.029 *[-0.042, -0.016]*	-0.023 *[-0.035, -0.011]*	-0.061* *[-0.075, -0.048]*	-0.048* *[-0.061, -0.034]*
CC	0.378* *[0.371, 0.385]*	0.831* *[0.826, 0.835]*	0.841* *[0.837, 0.845]*	0.841* *[0.837, 0.845]*	0.845* *[0.841, 0.849]*	0.861* *[0.858, 0.865]*	0.863 *[0.858, 0.865]*	0.863 *[0.860, 0.867]*	0.856* *[0.852, 0.860]*
Disp. Err [ms^{-1}]	2.670* *[2.645, 2.696]*	1.801* *[1.784, 1.826]*	1.705* *[1.681, 1.733]*	1.694* *[1.672, 1.715]*	1.705* *[1.682, 1.727]*	1.613 *[1.593, 1.633]*	1.608 *[1.589, 1.626]*	1.596* *[1.573, 1.618]*	1.634* *[1.612, 1.654]*
σ bias [ms^{-1}]	-1.501* *[-1.545, -1.458]*	-0.322* *[-0.378, -0.278]*	-0.454* *[-0.505, -0.404]*	-0.495* *[-0.546, -0.444]*	-0.391* *[-0.444, -0.340]*	-0.386 *[-0.438, -0.328]*	-0.372 *[-0.433, -0.314]*	-0.405* *[-0.455, -0.352]*	-0.420* *[-0.483, -0.367]*
RMSE [ms^{-1}]	3.070* *[3.029, 3.111]*	1.831* *[1.812, 1.851]*	1.772* *[1.748, 1.795]*	1.766* *[1.743, 1.792]*	1.749* *[1.729, 1.771]*	1.659 *[1.639, 1.677]*	1.650 *[1.632, 1.672]*	1.647 *[1.624, 1.667]*	1.688* *[1.670, 1.707]*
Spread [ms^{-1}]	0.850* *[0.846, 0.854]*	1.611* *[1.599, 1.622]*	1.605* *[1.592, 1.617]*	1.776* *[1.750, 1.779]*	1.663 *[1.650, 1.675]*	1.672 *[1.660, 1.686]*	1.667 *[1.655, 1.679]*	1.641* *[1.629, 1.654]*	1.728* *[1.714, 1.742]*
BSS (>5 ms^{-1})	-0.075* *[-0.093, -0.059]*	0.490* *[0.479, 0.500]*	0.515* *[0.505, 0.524]*	0.520* *[0.510, 0.529]*	0.513* *[0.504, 0.523]*	0.546 *[0.537, 0.555]*	0.549 *[0.541, 0.558]*	0.555 *[0.546, 0.563]*	0.526* *[0.517, 0.535]*
CRPS [ms^{-1}]	1.631* *[1.613, 1.648]*	0.883* *[0.875, 0.892]*	0.823* *[0.815, 0.831]*	0.814* *[0.806, 0.820]*	0.823* *[0.816, 0.831]*	0.777 *[0.770, 0.784]*	0.772 *[0.765, 0.779]*	0.769 *[0.762, 0.776]*	0.816* *[0.809, 0.823]*

*Table 2. The average values and confidence interval (0.95 sig. level) of several verification measures for the different models at all available stations in Austria and all lead-times during July 2018. The best result among compared forecasts is underlined (the spread is better when closer to the RMSE value). The values significantly different from the **AnEnStd** forecast (0.05 sig. level) are marked with an asterisk sign.*

July	***LAEFws***	***EMOSws***	***EMOSstd***	***AnEnCtrl***	***AnEnWs***	***AnEnMu***	***AnEnStd***	***AnEnAll***	***AnEnMem***
Bias [ms^{-1}]	-0.229* *[-0.242, -0.215]*	<u>-0.001</u>* *[-0.008, -0.010]*	-0.119* *[-0.129, -0.111]*	-0.012 *[-0.021, -0.001]*	-0.090* *[-0.099, -0.080]*	-0.055 *[-0.063, -0.046]*	-0.063 *[-0.072, -0.054]*	-0.088* *[-0098, -0.080]*	-0.043* *[-0.053, -0.033]*
CC	0.415* *[0.406, 0.422]*	0.750* *[0.745, 0.754]*	0.764* *[0.759, 0.768]*	0.752* *[0.748, 0.757]*	0.739* *[0.735, 0.744]*	0.770* *[0.766, 0.774]*	<u>0.774</u> *[0.769, 0.778]*	<u>0.774</u> *[0.770, 0.778]*	0.759* *[0.754, 0.763]*
Disp. Err [ms^{-1}]	1.602* *[1.589, 1.616]*	1.229* *[1.215, 1.240]*	<u>1.144</u>* *[1.132, 1.154]*	1.229* *[1.216, 1.241]*	1.262* *[1.250, 1.273]*	1.156* *[1.144, 1.167]*	1.145 *[1.136, 1.157]*	1.148* *[1.138, 1.159]*	1.183* *[1.172, 1.194]*
σ bias [ms^{-1}]	-0.773* *[-0.794, -0.754]*	-0.344* *[-0.368, -0.325]*	-0.474* *[-0.494, -0.452]*	-0.344* *[-0.364, -0.323]*	<u>-0.331*</u> *[-0.353, -0.308]*	-0.400* *[-0.418, -0.377]*	-0.409 *[-0.429, -0.387]*	-0.396* *[-0.416, -0.375]*	-0.403* *[-0.423, -0.383]*
RMSE [ms^{-1}]	1.794* *[1.775, 1.813]*	1.276* *[1.262, 1.288]*	1.244* *[1.234, 1.256]*	1.272* *[1.261, 1.284]*	1.307* *[1.294, 1.321]*	1.225 *[1.213, 1.237]*	1.219 *[1.208, 1.229]*	<u>1.218</u> *[1.206, 1.228]*	1.251* *[1.238, 1.262]*
Spread [ms^{-1}]	0.651* *[0.648, 0.654]*	1.170* *[1.164, 1.176]*	1.138* *[1.133, 1.144]*	1.318* *[1.311, 1.326]*	1.256* *[1.248, 1.263]*	1.253 *[1.246, 1.261]*	1.244 *[1.236, 1.250]*	<u>1.190*</u> *[1.184, 1.197]*	1.301* *[1.294, 1.308]*
BSS (>5 ms^{-1})	0.032* *[0.009, 0.055]*	0.329* *[0.314, 0.345]*	0.337 *[0.322, 0.353]*	0.329* *[0.313, 0.344]*	0.319* *[0.303, 0.335]*	0.349 *[0.334, 0.365]*	<u>0.355</u> *[0.341, 0.369]*	0.353 *[0.338, 0.369]*	0.325* *[0.310, 0.340]*
CRPS [ms^{-1}]	1.032* *[1.022, 1.042]*	0.648* *[0.643, 0.653]*	0.624* *[0.619, 0.629]*	0.636* *[0.631, 0.641*	0.650* *[0.645, 0.656]*	0.613 *[0.608, 0.618]*	<u>0.610</u> *[0.605, 0.615]*	0.612 *[0.606, 0.617]*	0.635* *[0.630, 0.640]*

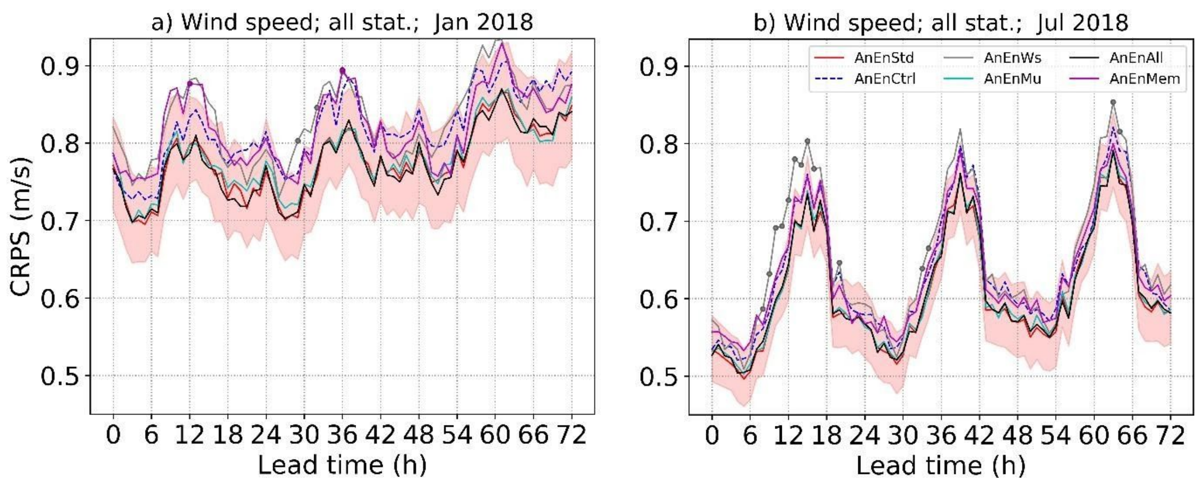

Figure 2. Continuous rank probability score (CRPS) depending on lead-time for five different analog-based ensemble experiments during January (left) and July (right) 2018 at 29 stations in Austria. The markers are set for the results significantly different from the AnEnStd forecast (95 % confidence level), while the red shaded area represents the AnEnStd 95% confidence interval calculated by the bootstrap percentile method [Jolliffe, 2007].

The wind speed increases towards the northeastern part of Austria (Pannonian plate) for both January and July, which also suggests a spatial pattern in forecast performance. The value for the **LAEFws** monthly mean CRPS is following the climatological wind speed pattern, having higher values at the stations prone to higher winds (Figure 3). The error is reduced for the analog experiments compared to the **LAEFws** following a similar pattern. The difference among all analog experiments, including the recently added **AnEnCtrl** and **AnEnMem**, is barely noticeable.

The added experiments produce worse results than the **AnEnMu** experiment (which is very similar to the **AnEnStd** and **AnEnAll**), but better results than **AnEnWs** for the high wind speed BSS (Figure 4). This underlines the potential benefits of using more than one meteorological parameter as an input to the analog method.

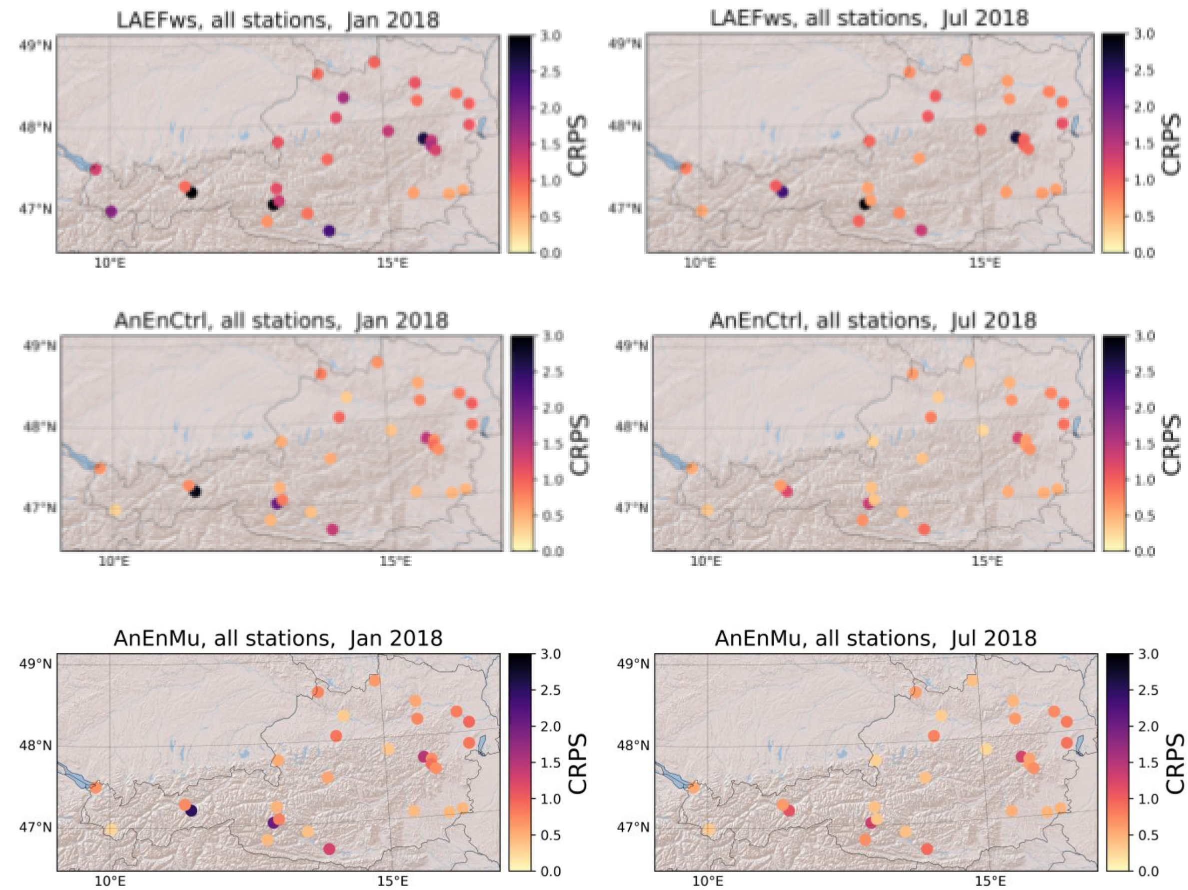

*Figure 3. The spatial distribution of the monthly mean continuous rank probability score for the **LAEFws, AnEnCtrl and AnEnMem** forecasts for January (left) and July (right) 2018.*

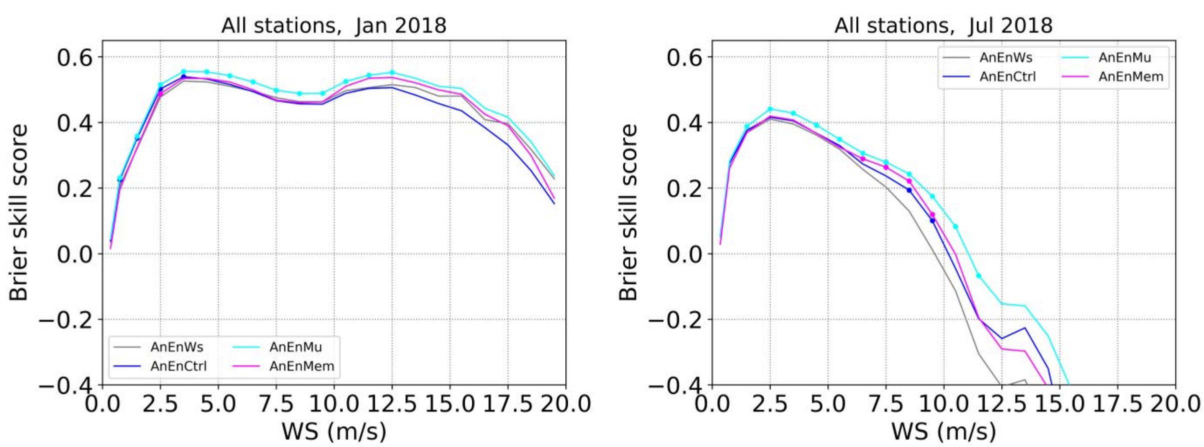

*Figure 4. Brier skill score (top) and relative frequency (bottom) depending on a wind speed threshold. The analog probabilistic forecasts shown for January (left) and July (right) 2018 at 29 stations in Austria. The markers are set for the BSS results significantly different from the **AnEnWs** forecast (95 % confidence level).*

The analog method for the gridded input

During this stay, I started working on developing the algorithms that would allow testing the implementation of the analog method on the gridded forecasts. The first step is to get familiar with the available literature regarding gridded forecast post-processing. The main issue is how to compare historical forecasts on the grid with the current one. It is possible to use a point-by-point approach (as done before) on every grid point. However, even though this is a good starting point, it is to be expected that it will produce noisy forecasts (as in Frediani et al., 2017). Alternatively, one can compare an average error on the entire field and use the mean value to choose the most similar (entire) fields. More complicated approaches include additional methods in order to simplify the information. For instance, one can identify objects (also in Frediani et al., 2017), use canonical correlation analysis (CCA; as in Fernandez and Saenz, 2003), use principal components (PC, as in Xavier and Goswami, 2007) or empirical orthogonal functions (EOF; similarly as point-based application in Barnett and Preisendorfer, 1978; also Zorita and Storch, 1998). In addition to these papers, methods such as quantile mapping and rank-weighted best-member dressing (Hamill and Scheuerer, 2018) or Schaake shuffle (as in Scheuerer and Hamill, 2018) can also be considered.

We decided to start with benchmark experiments – grid point-by-point approach and field-wise comparison. It has been noticed that using the Python implementation of the *SQL* database is much slower, only *h5* format is used. The *h5* files are prepared using ECMWF ensemble control (deterministic) run out of *grb* files (`IOP_load_ECMWF.py`). Only the wind speed variable is used as a predictor. Similarly, the INCA files are prepared. There are two distinctive variations: the first one includes all the grid points (`IOP_load_INCA_field.py`; for the field-based comparison), while the other one includes only the values interpolated at the ECMWF grid points (`IOP_load_INCA_grid.py`; for the point-by-point comparison). The training includes the 2017-2019 period. The testing will include January and July 2019.

The algorithms for the analog search are modified in order to use the grid-point values (`ANEN_grid.py` and `ANENm_grid.py`) or the field comparison (`ANEN_field.py` and `ANENm_grid.py`). Even though these algorithms work at the moment, they are extremely slow, making it impossible to verify the results before further optimization. Moreover, it was necessary to split the process in order to run it without memory issues (e.g. memory error, segmentation-fault, etc.). At this moment, instead of the observed value, the algorithm is actually 'forecasting' the timestamp of the most similar forecasts. The second part should then be used to load the values from INCA files. In the future work, probably during the next stay, the algorithms will be further optimized (e.g. use *numba* or *dask* Python module) in order to produce viable results. Additionally to these benchmark experiments, at least one field 'simplification' method will be tested as an addition to the analog approach.

References

Barnett, T.P. and R.W. Preisendorfer, 1978: Multifield Analog Prediction of Short-Term Climate Fluctuations Using a Climate State vector. J. Atmos. Sci., 35, 1771–1787, https://doi.org/10.1175/1520-0469(1978)035<1771:MAPOST>2.0.CO;2

Fernandez, J. and J. Saenz, 2003: Improved Field Reconstruction with the Analog Method: Searching the CCA Space. Climate Research, Vol. 24, No. 3, pp. 199-213. doi:10.3354/cr024199

Frediani, M.E., T.M. Hopson, J.P. Hacker, E.N. Anagnostou, L. Delle Monache, and F. Vandenberghe, 2017: Object-Based Analog Forecasts for Surface Wind Speed. Mon. Wea. Rev., 145, 5083–5102, https://doi.org/10.1175/MWR-D-17-0012.1

Hamill, T.M. and M. Scheuerer, 2018: Probabilistic Precipitation Forecast Postprocessing Using Quantile Mapping and Rank-Weighted Best-Member Dressing. Mon. Wea. Rev., 146, 4079–4098, https://doi.org/10.1175/MWR-D-18-0147.1

Jolliffe, I. T., 2007: Uncertainty and inference for verification measures. Weather Forecasting, 22, 637–650.

Scheuerer, M. and T.M. Hamill, 2018: Generating Calibrated Ensembles of Physically Realistic, High-Resolution Precipitation Forecast Fields Based on GEFS Model Output. J. Hydrometeor., 19, 1651–1670, https://doi.org/10.1175/JHM-D-18-0067.1

Xavier, P.K. and B.N. Goswami, 2007: An Analog Method for Real-Time Forecasting of Summer Monsoon Subseasonal Variability. Mon. Wea. Rev., 135, 4149–4160, https://doi.org/10.1175/2007MWR1854.1

Zorita, E. and H. von Storch, 1999: The Analog Method as a Simple Statistical Downscaling Technique: Comparison with More Complicated Methods. J. Climate, 12, 2474–2489, https://doi.org/10.1175/1520-0442(1999)012<2474:TAMAAS>2.0.CO;2