Stay Report

To: Zentralanstalt für Meteorologie und Geodynamik — ZAMG, Vienna, Austria
Period: November 21st – December 16th, 2022
Topic: Work on analog-based post-processing method
Supervisors: Mag. Alexander Kann and Irene Schicker, PhD
Collaborator: Ivan Vujec

The analog-based method application to gridded data post-processing

1. Introduction

I stayed at the Zentralanstalt für Meteorologie und Geodynamik (ZAMG) for four weeks during which I was working on the continuation of the analog-based post-processing method applied to an NWP model output for gridded forecasts. Analogies between, for example, similar past forecasts, measurements, or analyses are a potentially useful tool when the training dataset is long enough, thus enabling an adequate identification of true analogs. Thus, reducing the number of degrees of freedom in the matching procedure makes this method an excellent candidate for point-based post-processing, where NWP input can be deterministic or ensemble forecast. Previously, the point-based analog approach was thoroughly tested as a deterministic approach (Odak Plenkovic et al., 2018) and applied to calibrate the A-LAEF ensemble (Odak Plenkovic et al, 2020).

However, accurate forecasts at remote locations are used to drive many user-specific applications (e.g., road temperature forecasts along an entire roadway, soil temperature forecasts for agriculture, wind speed for windfarms). For that reason, besides the point-based post-processing for the measuring sites, there is also an increasing demand for gridded products. The latter is a direct motivation for the development of tools needed for using an analog-based method to produce gridded output based on an analysis.

The purpose of this report is to present recent developments which I achieved during the ACCORD stay at ZAMG, that took place November 21st – December 16th, 2022, including some practical details. During this stay, I tested and debugged the previously developed algorithms for two analog-based approaches that produce gridded products. After improving algorithm performance, I explored the sensitivity to predictor weighting and normalization procedure, as well as to the training length.

2. Data and algorithm

The algorithms that are previously developed and described (Odak Plenković, 2021) are now debugged and optimized further. After that, several experiments were performed, using two distinctive approaches, Point-by-Point (PbP) and Field-wise (FW) approach. The PbP is an approach in which every grid point is treated as an independent location, like point-based approaches. For FW approach, the distance metric is calculated independently for every forecast lead time and grid point as in PbP, and then the calculated metric is averaged over an entire field to select the best match, instead of each grid point separately.

As input to the analog methods, the control member of the ECMWF ensemble forecast is used as a raw forecast. The gridded INCA analysis fields (Haiden et al., 2011) are used as a “ground truth”, like the observations in the point-based analog approach. For testing the novel gridded analog algorithm within a decent amount of time, the INCA wind speed analyses are bilinearly interpolated onto the ECMWF forecast grid.

When not mentioned otherwise, the experiments mentioned in this report use wind speed, wind direction and temperature variables as predictors. The analog ensemble (AnEn) consists of 10 INCA wind speed values corresponding to the 10 best-matching analogs for the ECMWF control member. The analyzed INCA wind speed values are also used as observed values in the verification procedure. The forecast range is 90 hours (about 4 days).

Since the period used for training and a validation data set in the following experiments sometimes differ, they will be explicitly stated. The maximum size of the training dataset refers to 1.1.2017.-30.11.2018. period. The experiments are tested for one winter month - January of 2019.

For the fairer comparison of the presented experiments, only the wind speed values at ECMWF grid points are used (bilinearly interpolated for FW approach), but one needs to keep in mind that FW approach can produce results with the resolution matching the INCA output.

3. Results

As already mentioned, several different analog-based experiments are executed and compared, using INCA wind speed analysis as observed values. The experiments are compared for one winter month (January 2019) since the previous work shows that it is more sensitive to the settings than one summer month (July 2019), while both producing consistent results. Several sensitivity tests are conducted after algorithms optimization, regarding the sensitivity to training length, predictor weighting and normalization; for both PbP and FW approach.

➢ Training data length

The inconsistencies in the intercomparison of PbP and FW approach regarding different training lengths, which were the consequence of long execution time for PbP are finally resolved by improving the algorithm. Thus, it is now possible to test the sensitivity to training length for both PnP and FW approaches. Three training lengths are tested: 2 years, 1 year and only 4 months. The experiment with a one-year-long training data set is also referred to as experiment 1.

For the testing of the effect of training data length, the weights for all predictor variables are set to 1, whereas the normalization factor (assessed by standard deviation) is calculated separately for each location point and lead time (it is different for every grid point and every lead time).

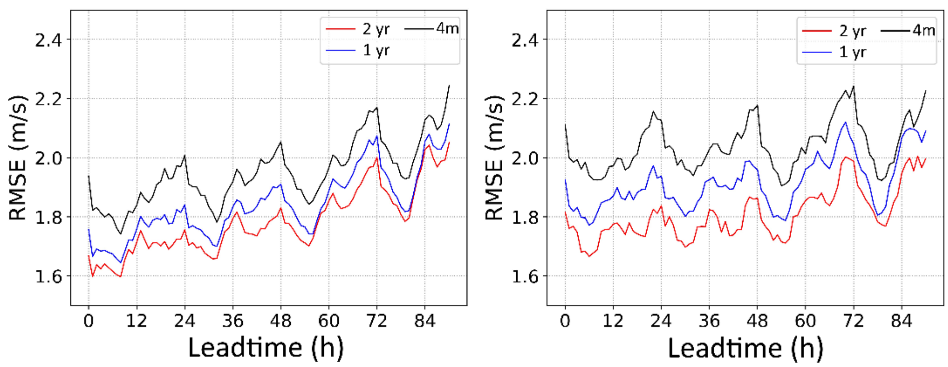

Figure 1 Sensitivity to training length measured by RMSE for PbP (left) and FW (right) approach. The used training lengths are 2 years, 1 year and 4 months. The results are for January 2019.

Measured by RMSE, it can easily be seen that the result improves notably by increasing the training length (Figure 1). The differences are more pronounced for FW than PbP approach. That makes sense since it is harder to find a similar match for the entire field within the training dataset than for one independent location. For that reason, one might consider an even longer training period for FW approach. On the other hand, for PbP approach the impact of prolonged training length reduces, especially towards the end of the forecast period. In this case, the cost-lost ratio might be considered, if the prolonging of training, which is computationally more demanding, is beneficial enough. The results for CRPSS are consistent with the previous discussion (Figure 2).

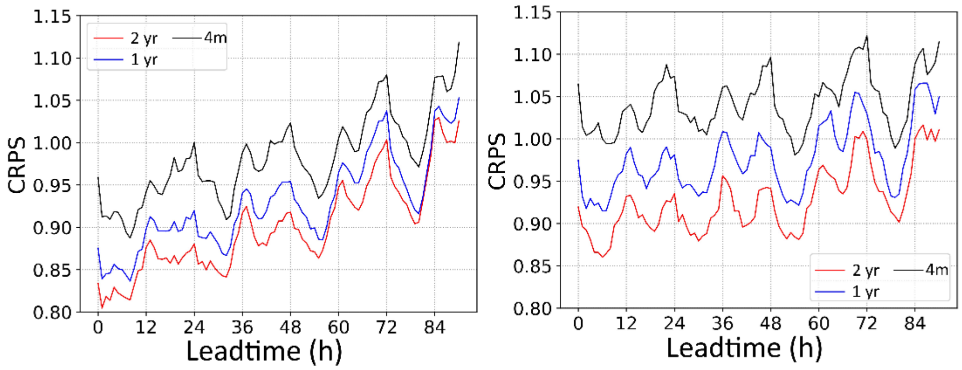

Figure 2 Sensitivity to training length measured by CRPSS for PbP (left) and FW (right) approach. The used training lengths are 2 years, 1 year and 4 months. The results are for January 2019.

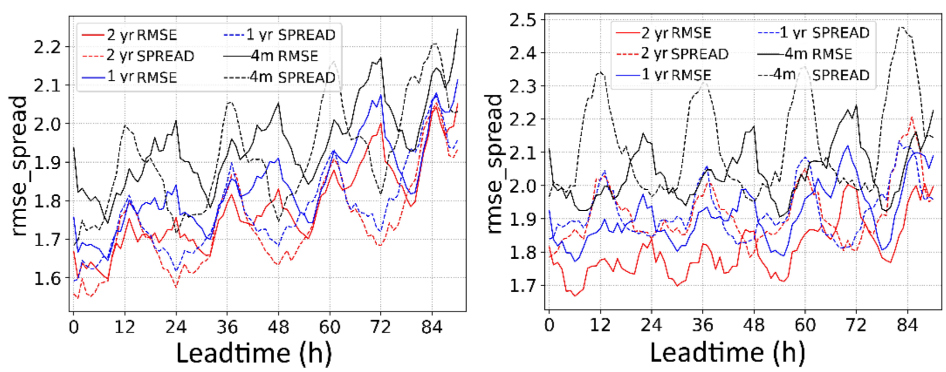

Figure 3 Sensitivity to training length measured by RMSE-SPREAD diagram for PbP (left) and FW (right) approach. The used training lengths are 2 years, 1 year and 4 months. The results are for January 2019.

Finally, to determine if the uncertainty is adequately represented in the analog method, the ensemble spread is compared to RMSE (Figure 3). SPREAD can be considered quite close to RMSE for PbP approach, with some overspread about noon and underspread about midnight. For the FW approach overspread is more pronounced, especially about noon. SPREAD seems to be the most adequate when only a one-year-long training dataset is used.

➢ Normalization by standard deviation optimization

The next step is to decide how to normalize different predictor variables by standard deviation. For this purpose, a one-year-long training dataset is used for all experiments in this chapter. Additionally, predictor weights are used, where values are estimated as average values used in previous work (optimization for operational wind speed post-processing in Croatia): 1.00 for wind speed, 0.70 for wind direction, and 0.25 for temperature predictor variable.

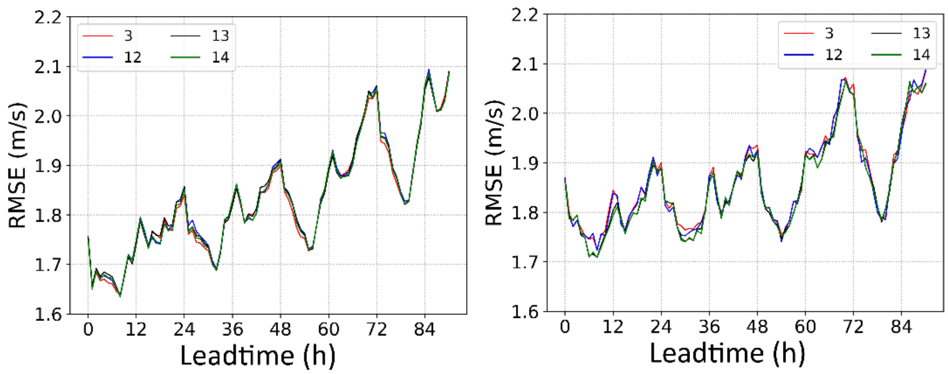

Figure 4 Sensitivity to normalization factor (determined by standard deviation), measured by RMSE for PbP (left) and FW (right) approach. The results are for January 2019. The experiments are: 3 – normalization factor depends on the (grid) point location and leadtime; 12 - normalization factor depends on the (grid) point location and not on the loadtime; 13 – it depends on leadtime, and not on the exact (grid) point location; 14 – it is a constant value for each predictor variable.

The normalization factor is often approximated by the standard deviation for each predictor variable. The same idea is followed here, but there are several ways to implement it. For instance, the value can be calculated only once using an entire training dataset, thus having a constant value for all grid point locations and all lead times (experiment 14). Otherwise, it can depend only on lead time (experiment 13), taking into account diurnal variation, or only on exact grid point location (experiment 12), which would take into account spatial variability. Finally, the normalization factor can be calculated separately for each grid point and each lead time (experiment 3, the same as previous results when using 1-year-long training).

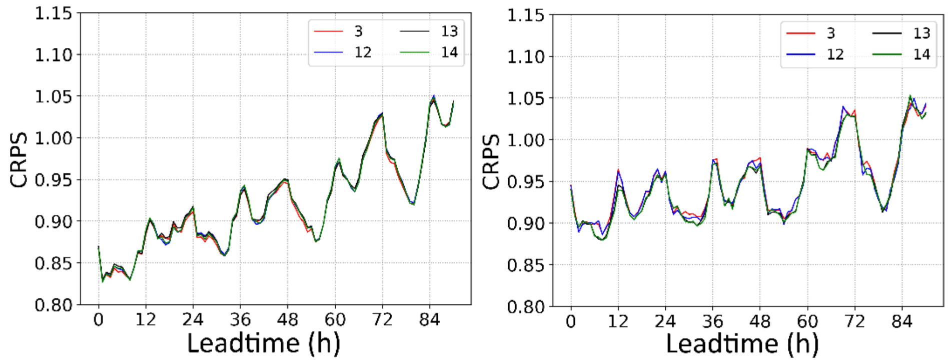

Figure 5 Sensitivity to normalization factor (determined by standard deviation), measured by CRPSS for PbP (left) and FW (right) approach. The results are for January 2019. The experiments are: 3 – normalization factor depends on the (grid) point location and leadtime; 12 - normalization factor depends on the (grid) point location and not on the loadtime; 13 – it depends on leadtime, and not on the exact (grid) point location; 14 – it is a constant value for each predictor variable.

The results show that the sensitivity to the normalization factor calculation procedure is rather small. Measured by both RMSE (Figure 4) and CRPSS (Figure 5), calculations for every grid point and every lead time separately led to only slight improvements for PbP approach. On the other hand, the best for the FW approach is to use a constant value for each predictor variable or include the dependency on lead time. Spatial variability does not seem beneficial, probably since the minimization process of the differences among fields already includes spatial variability.

Finally, for the PbP approach, SPREAD reduces similarly to RMSE (e.g., experiment 3) (Figure 6). For the FW approach, the differences are slightly more pronounced but still rather small and inconsistent, which makes it difficult to draw general conclusions.

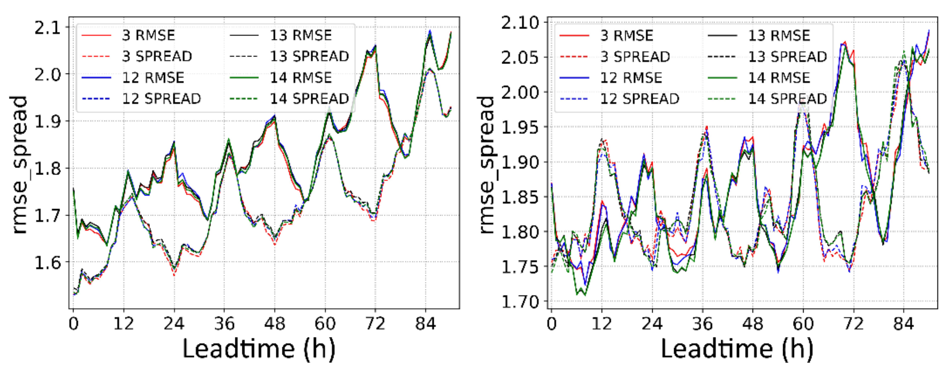

Figure 6 Sensitivity to normalization factor (determined by standard deviation), measured by RMSE-SPREAD diagram for PbP (left) and FW (right) approach. The results are for January 2019. The experiments are: 3 – normalization factor depends on the (grid) point location and leadtime; 12 - normalization factor depends on the (grid) point location and not on the lead time; 13 – it depends on leadtime, and not on the exact (grid) point location; 14 – it is a constant value for each predictor variable.

➢ Predictor weight optimization

The experiments with different predictor weights have already been mentioned in this report. For experiment 1, the weights are set to 1.00 for wind speed, wind direction and the temperature predictor variable. In other words, they are all equally relevant in the search for similar situations in the training period. That is the simplest way to treat predictor variable, but, of course, not the best one. The analog-based method can, in a certain sense, be overfitted. In context of analogs, that would mean that algorithm does find the best statistical match when all the predictor variables are considered, but it does not necessarily mean that it is also the closest match for the parameter we are trying to predict. For instance, if there are three predictor variables (wind speed, wind direction and temperature) and they are treated as equally important when predicting temperature, by choosing the most similar situation we would often find ourselves in the wrong season completely, simply because the pressure system and thus winds are similar. Then, the error can be quite large. This is the consequence of limiting the degrees of freedom in order to be able to use the method. The way of avoiding this type of situation is to optimize weight. In the ideal case, one would minimize the error by trying all the possible combination of weights for every predictor. However, that is computationally a very demanding task, even for point-based setup. The alternative is to use forward searching algorithm. Such an algorithm has been tested and implemented for point-based wind speed predictions, for instance in Croatia for the operational setup. Here, the optimization procedure is even more expensive, which is a limiting factor to trying out a lot of combinations. Thus, we used the (approximately) average values that are calculated for operational purposes in Croatia as a first guess (experiment 3). The exact values of the weights are: 1.00 for wind speed, 0.70 for wind direction and 0.25 for temperature. Then, we tried to exclude the temperature variable (experiment 4), using weight 1.00 for wind speed, 0.70 for wind direction and 0.00 for temperature. Finally, we tried to exclude the wind direction predictor (experiment 5), using the weight 1.00 for wind speed, 0.00 for wind direction and 0.25 for temperature.

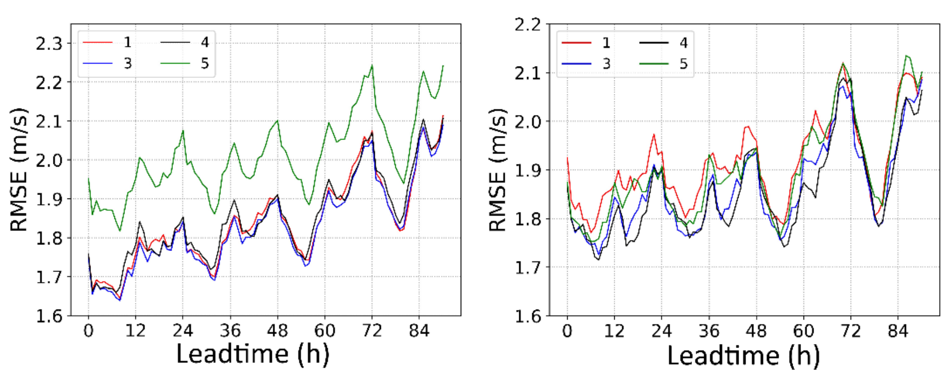

Figure 7 Sensitivity to predictor weights, measured by RMSE for PbP (left) and FW (right) approach. The results are for January 2019. The experiments are: 1 – the weights are all set to 1.00; 3 – the weight for wind speed is 1.00, for wind direction 0.70, for temperature 0.25; 4 – the weight for wind speed is 1.00, for wind direction 0.70, whereas the temperature predictor is not included; 5 – the weight for wind speed is 1.00, for temperature 0.25, whereas wind direction predictor is not included.

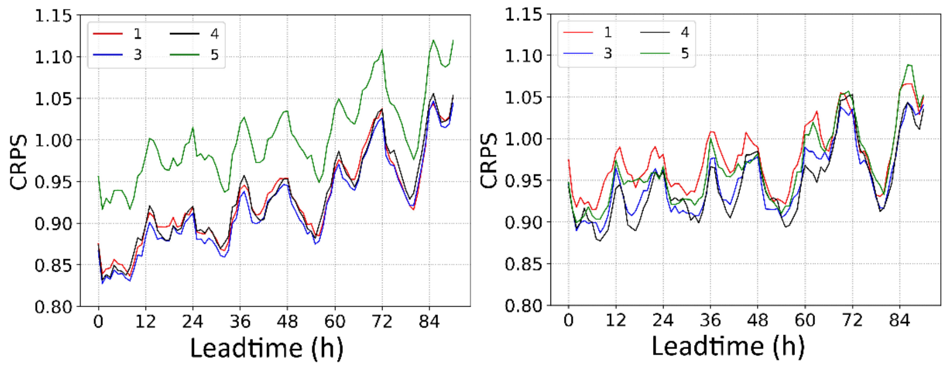

Figure 8 Sensitivity to predictor weights, measured by CRPSS for PbP (left) and FW (right) approach. The results are for January 2019. The experiments are: 1 – the weights are all set to 1.00; 3 – the weight for wind speed is 1.00, for wind direction 0.70, for temperature 0.25; 4 – the weight for wind speed is 1.00, for wind direction 0.70, whereas the temperature predictor is not included; 5 – the weight for wind speed is 1.00, for temperature 0.25, whereas wind direction predictor is not included.

Both RMSE (Figure 7) and CRPSS (Figure 8) results show certain sensitivity to the choice of predictor weights, as expected. For PbP approach, it can be seen that experiment 5 underperformed, which means that wind direction is an extremely valuable predictor. The results for other experiments are similar for PbP approach, and experiment 3 seems to be the most successful. That is expected, since the used weights are especially calculated for point-based approach, and PbP can be considered as an extended version of point-based approach.

For the FW approach, the results are different. The distinction between experiments is more pronounced at the beginning of the forecast period and seems to reduce with lead time. Experiment 4 is the most successful, followed by experiment 3. Experiment 1, with equal and constant predictor weights seems to be the least successful.

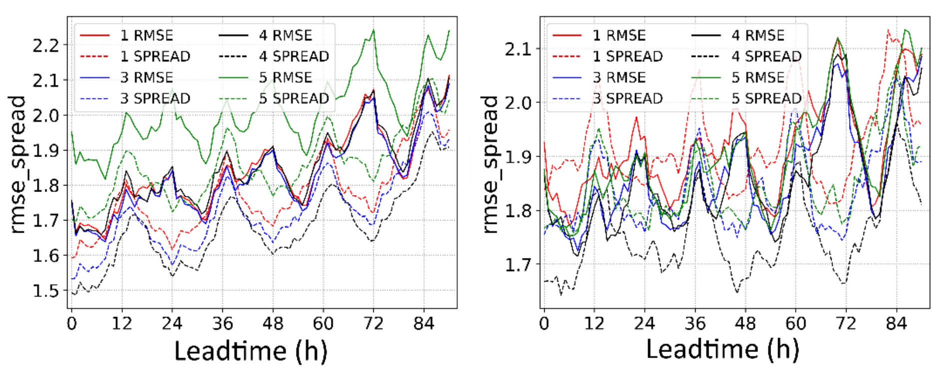

Figure 9 Sensitivity to predictor weights, measured by RMSE-SPREAD diagram for PbP (left) and FW (right) approach. The results are for January 2019. The experiments are: 1 – the weights are all set to 1.00; 3 – the weight for wind speed is 1.00, for wind direction 0.70, for temperature 0.25; 4 – the weight for wind speed is 1.00, for wind direction 0.70, whereas the temperature predictor is not included; 5 – the weight for wind speed is 1.00, for temperature 0.25, whereas wind direction predictor is not included

Finally, if SPREAD is compared to RMSE (Figure 9), it can be seen that there is an underspread for experiments 4 and 5 in PbP approach, whereas for experiments 1 and 3 underspread is only ponounced in nighttime. For FW approach, the SPREAD generally matches RMSE better than for PbP approach. Here, there is some overspread in the daytime, and underspread in the nighttime. Among experiments, it seems that the SPREAD for experiments 3 and 5 matches RMSE better for experiments 1 and 4.

➢ Overall intercomparison and future work

Finally, in order to compare FW and PbP approaches directly, four experiments are shown. The example of poorly performance is shown through experiment 1 (constant predictor weights, normalization calculations for every grid point and lead time, and 4 months of training), whereas an example of (reasonably) better performing setup is shown using experiment 3 (used average values of optimized predictor weights calculated on different dataset, normalization calculations for every grid point and lead time, and 1 year of training).

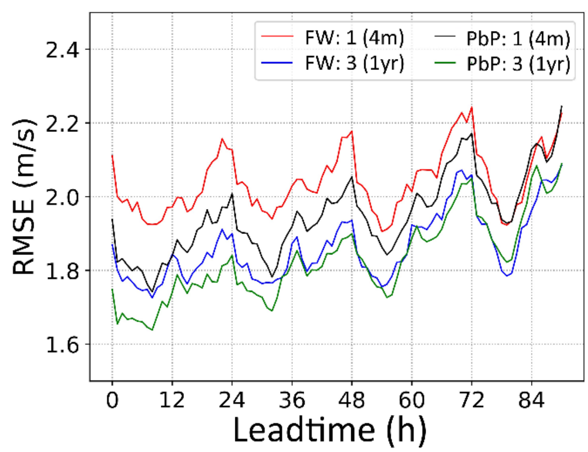

Figure 10 The intercomparison of the experiments for PbP and FW approaches for January 2019 – experiment 3 with one-year-long training dataset and experiment 1 with four-months-long training dataset.

The result shows that PbP produces better results than FW in the beginning of the forecasting period, and the difference reduces towards the end of forecasting period for both experiment 1 (4 months training) and 3 (1 year training) (Figure 10). The sensitivity to proper setup also reduces with lead time for both PbP and FW approaches but is present throughout the forecasting period. Thus, it is important to have long enough training (at least 1 year for PbP and 2 or more for FW approach) and proper predictor weights. The particularities of the normalization procedure are not as relevant, it is probably enough to normalize each predictor variable with average standard deviation for the entire training dataset.

In addition to these experiments, future work might include an additional method for choosing the best analogs to simplify the information. Moreover, the first steps to develop an approach that uses EOFs are already taken. The training data is used to calculate EOFs, and training is thus saved as EOFs and principal components (PCs) timeseries. The raw NWP that needs to be post-processed by analog-based algorithm should then be translated into “current” PCs using pre-calculated EOFs. The “similarity” between different analogs can then be calculated using PCs as different predictors in the algorithm.

4. Acknowledgments

The authors would like to thank our colleagues in the ACCORD and RC-LACE consortiums for their support in our work.

5. References

- Haiden, T., Kann, A., Wittmann, C., Pistotnik, G., Bica, B., Gurber, C., 2011: The Integrated Nowcasting through Comprehensive Analysis (INCA) System and Its Validation over the Eastern Alpine Region. Weather and Forecasting. 26. 166-183. https://doi.org/10.1175/2010WAF2222451.1.

- Odak Plenković I, Delle Monache L, Horvath K, Hrastinski M (2018) Deterministic Wind Speed Predictions with Analog-Based Methods over Complex Topography. Journal of applied meteorology and climatology 57:2047-2070.

- Odak Plenković I, Schicker I, Dabernig M, Horvath K, Keresturi E (2020) Analog-based post-processing of the ALADIN-LAEF ensemble predictions in complex terrain. Quarterly Journal of the Royal Meteorological Society 146:1842– 1860.

- Odak Plenković, I (2021): Stay report - The analog-based method application to gridded data post-processing