Stay Report

To: Zentralanstalt für Meteorologie und Geodynamik — ZAMG, Vienna, Austria

Period: 13th November - 9th December, 2017

Topic: Work on analog-based post-processing method

Supervisors: Mag. Alexander Kann and Irene Schicker, PhD

Summary

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 a NWP model output. We agreed on rewriting all the algorithms from Matlab to Python programming language. Since it is widely used and freeware, we agreed that this is a great opportunity to write the entire analog-based scheme consistently, as well as for me to learn using Python.

During the first week, I showed an overview of my previous work in a short introductory presentation entitled "Analog-based wind speed predictions (in the complex terrain)". The idea was to present the methodology and previous experiences on analog-based method application to NWP output. The analog-based method can be used to improve the deterministic NWP output and to estimate the probability distribution of the forecast. The method uses historical data within the specified analog training period for which both the deterministic NWP (starting model) and the verifying observation are available. The analog-based method uses one consistent grid-point. It is usually the closest one to the measurement station if there is not some extreme differences. For example, it might be situated over the water surface. Then, one can choose another, more representative one. The best-matching historical forecasts to the current prediction (analogs) may originate in any past date within the training period. The quality of the analog is evaluated by the following metric:

$$\|F_t A_{t'}\| = \sum_{i=1}^{N_A} \frac{w_i}{\sigma_{fi}} \sqrt{\sum_{j=-\tilde{t}}^{\tilde{t}} \left(F_{i,t+j} - A_{i,t'+j}\right)^2}, \tag{1}$$

where F_t is the current NWP deterministic forecast at given location, valid at the future time t, whereas $A_{t'}$ is an analog at given location with the same forecast lead time, but valid at a past time t'. The N_A is the number of predictors used in the search for analogs, w_i are the weights corresponding to particular predictor, normalized with the standard deviation of the time series of past forecasts of a given variable at the same location is σ_{fi}. The $\tilde{t}$ is equal to half the number of additional times over which the metric is computed (the half of the time window of any specified width), therefore $F_{i,t+j}$ and $A_{i,t'+j}$ are the values of the forecast and the analog in the time window for a given variable, respectively. Analogs are found independently for every forecast time and location, narrowing the search around particular time of a day by a time window. The verifying observations of the best-matching analogs are the members of the analog ensemble (AnEn). The assumption is that the errors of the good (quality) analog forecasts are likely to be similar to the error of the current forecast (Delle Monache et al., 2011). Once the AnEn is formed, it can be used to produce the deterministic analog-based prediction (i.e. forecasting the mean of the AnEn), as well as the probabilistic forecast (e.g., to estimate the probability of a predefined event).

The AN_t forecast for the future time t at a given location is an average (weighted, if $\gamma \neq 1/N$) of the observations O_i corresponding to N most similar analogs $A_{t'}$ (measured by metrics previously defined in equation 1):

$$AN_t = \frac{1}{N} \sum_{i=1}^{N} \gamma O_i(A_{t',i}). \tag{2}$$

In another words, the AN_t is a (weighted) mean of N-sized AnEn for a (future) time t. Several authors, such as Delle Monache et al. (2013), state that the AnEn rank histograms are uniform. Every member of the AnEn is thus equally probable outcome, even though some analogs are closer to the current forecast than the others (measured by previously defined metrics). Hence, the value assigned to the weights γ is 1.

After the introductory presentation, we decided that the analog-based method should be tested by using (deterministic) AROME NWP data and corresponding observations (OBS) first. During the first and second week:

- We were working on the installation and the preparation of the software needed
- The datasets (deterministic NWP) are provided:
    - AROME (1/1/2015-31/08/2017)
    - ALARO (1/1/2015-31/10/2017)
    - ECMWF (1/1/2015-31/08/2017),

  as well as the corresponding OBS (1/1/2015-31/10/2017)
- Also, I was getting more familiar with the Python programming language.

During the third week I was dealing with the data - building a database for the training period. This is computationally the most demanding part of the entire analog-based scheme. Therefore, the idea is to be able to build the database directly from the (historical) model outputs (for all the locations), but not for every analog forecast. For instance, it can be updated once every month, once every few months or even once a year. I started with the AROME NWP (and corresponding OBS), but I was writing the algorithms so that they can easily be modified and used tor the other NWPs as well. During this phase, I managed to match the inconsistencies in the AROME and OBS files (such as a change in file naming, differently marked missing data etc.). However, due to the lack of RAM available, I was not able to load the entire database. This part will be continued during my next stay when more computational power will be provided to my user. However, at this point, I loaded only two months in order to continue developing the algorithms.

I was developing the basic scripts for analog method application (the AN forecasting) during the last week. The algorithms were written so that they would use pre-prepared database and "current" AROME NWP. This way the algorithm might easily be implemented to run operationally. Due to the lack of the data in the training period, we were not able to provide any further optimization or meaningful testing. We agreed to test and verify the algorithm for the two months selected (summer and winter month) by using the AROME data during my next stay. This will show the calibration potential and I shall get even more familiar with the data (both forecasts and the measured data). This is the most important task for my next stay (05/02/2018-03/03/2018) and an important step before application to ensemble forecasts.

The algorithms can be modified for some other inputs, such as other deterministic forecasts and finally the ALADIN-LAEF ensemble forecast in a long-term plan. After my next stay the continuation of this work can include the NWP ensemble calibration comparison with the ensemble generated directly from deterministic model by analog-based method.

References

Delle Monache, L., Nipen, T., Liu, Y., Roux, G., Stull, R., 2011: Kalman filter and analog schemes to post-process numerical weather predictions. Monthly Weather Review 139, 3554-3570.

Delle Monache, L., Eckel, T., Rife, D., Nagarajan, B., 2013: Probabilistic weather prediction with an analog ensemble. Monthly Weather Review 141, 3498-3516.