PP-MILEPOST: MachIne LEarning-based POST-processing.

Project Duration: *Start 09.2020 – End 08.2022 (two COSMO years, with the possibility of extension)*

Version 2.0, 02.07.2020

Total FTE request: 4.06

Project leader

Andrzej Mazur (IMGW)

The main goal of the Priority Project is to provide COSMO community with new and/or advanced and elaborated methods of post-processing which would allow the best possible approximation of the forecast to the actual future state of the atmosphere.

The project aims at collecting under a unique umbrella the experience available in the Consortium regarding to Machine Learning for post-processing, and to expand them in order to provide useful indications for the other COSMO partners which are interested in Machine-Learning based Post-processing (MLP).

The result of this project would be the examination of the relation between numerical forecasts in terms of Direct Model Output (DMO) and ML-based Post-processing (MLP), including verification against observations, especially and mainly with regards to MLP.

All proposed methods should eventually be delivered to interested parties in the form of software packages that will be used for advanced post-processing.

Introduction

Numerical weather prediction (NWP) has long been a difficult task for meteorologists. Atmospheric dynamics is extremely complicated to model, and chaos theory teaches us that the mathematical equations used to predict the weather are sensitive to initial conditions; that is, slightly perturbed initial conditions could yield very different forecasts.

The importance of accurate forecasts of all necessary weather parameters and fields is obvious. The Direct Model Output variables are not optimal direct estimates of local weather forecasts [Kalnay, 2003] because models have biases; ground surface (in model) is not ideal representation of the actual terrain; etc. etc. The constant need to provide increasingly accurate weather forecasts leads to the question of how to use DMO to prepare such an accurate forecast.

Moreover, not all the models forecast (directly) some required parameters, such as visibility and probability of thunderstorms. Or they do it with a lot of inaccuracy and/or errors. To improve the use of NWP as guidance to human forecasters, it has been customary to use statistical methods to “post-process” the model forecasts and adapt them to produce local forecasts. There is a growing need for schemes that consider the nature of weather parameters – continuous, such as air temperature or atmospheric pressure, and discrete – like precipitation, while being universal enough to be applicable to all forecast elements from a given class.

Studies of neural networks, logistic regression, and genetic algorithms have shown improvements over standard linear regression for precipitation prediction [Applequist *et al*., 2002]. Gagne *et al*. [2014] proposed using multiple machine learning techniques to improve precipitation forecasting. They used Breiman’s random forest technique [Breiman, 2001], which had previously been applied to other areas of meteorology, including aviation turbulence [Williams, 2013], to learn from the CAPS storm-scale ensemble forecast (SSEF) data.

The convolution neural networks highlight the spatial structures and can be useful for convective weather prediction. Zhou *et al.* [2019] use the deep convolution network to predict severe convective weather including short-duration heavy rain, hail, convective gusts and thunderstorms. Shi *et al.* [2015] and Agrawal *et al.* [2019] propose the advanced types of convolution neural networks for precipitation nowcasting problem: the convolutional long-short memory neural network (ConvLSTM) and the so-calledU-Net.

Machine Learning, basically, is an application of Artificial Intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. It focuses on the development of computer programs that can access data and use it learn for themselves. The process of learning begins with observations or data, such as examples, direct experience, or instruction, in order to look for patterns in data and make better decisions in the future based on the examples that we provide. The primary aim is to allow the computers learn automatically without human intervention or assistance and adjust actions accordingly.

Since the resolution of the NWP models increases, the information contained in DMO that could be used in MLP is also becoming more detailed and in high resolution. However, DMO alone cannot be a good enough approximation of the current state of the atmosphere. Therefore, different – and more efficient in terms of forecasts' quality – post-processing methods are required.

Recently, machine learning techniques have started to be applied to NWP. The ML methods can be included into NWP [Krasnopolsky, 2013, Krasnopolsky, 2020], but in this project we consider the ML for post-processing the DMO data.

Motivation

Despite rapid development in the field of machine learning, systems still remain somewhat human-dependent. The very process of designing the system requires that man defines ways of acquiring knowledge and its representation. In addition to the system creation stage, the following problems arise:

- too weak or too strong dependence of the system on the environment in which it is located, which can lead to incomplete data analysis or misinterpretation,
- credibility and correctness of the generated conclusions – knowledge acquired as a result of observation is only implied, and inductive reasoning often cannot be fully proven, but only falsified,
- incomplete or partly contradictory data,
- not defining domain restrictions may lead to comprehensive generalizations and erroneous conclusions.

In light of the above problems:

- knowledge generated by the systems should be subject to human control and assessment, according to the criteria provided by him,
- the system should be able to provide an explanation in case of a problem,
- knowledge should be understandable to man, i.e. expressive in the description and mental model adopted by him.

In COSMO consortium there have been some individual studies partially related to Machine Learning, especially in terms of Ensemble Prediction Systems (EPS) and forecasts. Yet, this Priority Project seems to be the very first effort to summarize experiences in the subject, to provide results of (partial and case) studies and to evaluate and develop Machine Learning techniques in post-processing to be applied widely over the Consortium, not only connected to EPS results.

The different forecast data from the Machine-Learning based Post-processing training data can be used. The number of forecasts needed to obtain a statistically significant verification signal, the choice of cases need to be considered (e.g. for temperature we might want a mixture of winter/summer and cloudy/clear sky, snow/snow free and anticyclonic/cyclonic etc.).

Cases when the operational model (i.e. DMO) performed poorly may offer the possibility for greater benefit from MLP, but a good mix of cases from different weather situations should be chosen as well.

The selection of verification methods(s) (as defined as WG5's activities and PPs) that would help solve this query to a large extent will depend on the model variable. The following questions can be asked:

- Which variables are in scope? Is there a suggested priority order, e.g. temperature, then precipitation, etc?

- Will the output of MLP be values at the locations of station observations (i.e. site specific) or a 2D field (gridded)?
- Will verification based on post-processed operational model/models be considered in each country (if it exists available and is available)? Determining the quality (or improvement of) the results due to post-processing versus "simple" DMO, one can determine which verification methods are best suited for this task.

Status and available expertise of main participants within COSMO

IMGW-PIB (as the National Weather Service) operates a post-processing system based on the EPS system and the use of an artificial neural network ANN. This ANN itself uses a back-propagation method with a specific sigmoidal activation function (hyperbolic tangent). Teaching and testing of ANN is carried out in the operational mode, using archived measurements, observations and appropriate forecasts, as well as other parameters such as geographical (coordinates, elevation) or related to the operational setup (e.g. forecast start time, forecast lead time, etc.).

The key benefits of ANN are:

- ANN have the ability to learn and model nonlinear and complex relationships, which is really important because in fact many relationships between input and output are nonlinear as well as complex.
- ANN can generalize – after drawing conclusions from the initial input data and their dependencies, it can also infer about relationships that did not occur during the learning process, on (previously) unregistered data. Thus, the model generalizes and predicts also on the basis of previously unnoticed input data.
- Unlike many other forecasting techniques, ANN imposes no restrictions on input variables (such as how they are distributed).

RHM runs a post-processing system based on deterministic ground level DMO of COSMO-Ru model and using the ANN technology (Bykov, 2020). This setup is cross-platform (Windows/Linux), hardware-independent (uses GPU if available) and based on PyTorch framework. The ANN was learned and validated in nonoperational mode and tested in operational mode.

The RHM MLP system uses the DMO with various initial and lead times (including lagged forecasts, but not only) at a considered point. The system can process the data in two modes: using the last SYNOP data and without it.

The key revealed advantages of the ANNs are as follows:

- ANNs are continuous and approximate any continuous dependence
- ANNs techniques are very computationally efficient and well compatible with big datasets

ANNs are built from blocks of several different types and can find the approximation for the hidden variables. The key limitation of the ANNs is the following: the ANNs don't conclude about the significance of the features.

MeteoSwiss is actively developing an operational post-processing system to provide local post-processed forecasts for the whole Swiss domain; that is, at any surface location, with or without observing stations. Further, the goal is to produce a probabilistic output, meaning that forecasts are given in terms of either probabilities or distributions for each parameter. The project will also look into the physical consistency in space and time, as this is of high relevance for impact modelers. Finally, this system should not depend on a particular model covering a particular time horizon, meaning that it should be seamless in time and be able to combine multiple NWP models. A review of such challenges is provided in Vannitsem *et al.* (2020).

In this context, MeteoSwiss is currently experimenting with ANNs to post-process COSMO-E 10m agl. wind forecasts (hourly wind speed, gust, direction, and u v components). The ANN is trained to predict a parametric probability distribution for the model errors by considering the set of predictors available at the point of interest, as in Rasp and Lerch (2018). In addition, predictors include high-resolution topographical descriptors (Schaer 2019), thus allowing the ANN to generalize at unseen locations. The results show an improvement of the quality of the forecasts in terms of CRPS for all lead times. In this application, the ANN can provide a flexible estimator of nonlinear relationships between DMO ensemble statistics, topography, and the forecast error.

As far as the DWD is concerned, there are work on the MFASIS project (L. Scheck) in the version based on ANN. MFASIS (the “official” version is currently based on lookup tables) is an efficient observation operator for all-sky satellite radiances in the visible (VIS) range. Assimilation of VIS radiances is still being developed in DWD under the WG1 umbrella and will be part of the new KENDAscope project.

The bias correction approach was firstly developed by Otkin and Potthast in the context of all-sky infrared satellite radiances for (see KENDAscope project), and has recently been adapted to assimilate 2 m humidity and 2 m temperature observations. This approach is usually called “online conditional non-linear bias correction”, but it can also be called “recursive non-linear regression”.

DWD also submitted a project proposal for estimating observation errors using AI.

Moreover, DWD works with partners as part of a research project that aims to assimilate information from (web) cameras. Work has just begun on the first ANN test towards an observation operator for such images.

As it was stressed before, the results of this PP could/should be used both in statistical- (EPS) and deterministic forecasts. Therefore, any system that is intended to use the results could include both forecasting approaches.

Requirements and actions proposed:

Actions that would be carried out in this PP result from the following facts. First, key issues related to improving forecasts that are included in the project are:

- Which methods and schemes can be used in the MLP process?
- How long should “learning” take to make the results meaningful and significantly better compared to DMO (and the associated problem of sufficient computing resources)?

- What kind of post-processing brings the greatest benefits of among ML applications? (calibration, weighted combination, creating decision trees to define a new product, etc.)
- Even if until now in the COSMO Consortium the experience about ML was limited to ensemble post-processing and to data assimilation, the technique has a much wider scope of application. In this project, though limited to post-processing, applications to both deterministic and to ensemble forecast are planned.

Last but not least, one should be aware of the shortcomings and deficiencies – related to the nature of the ML system(s) – that must be overcome and are as follows:

- Neural networks and other ML systems require relatively high computing power and memory resources: in the ANN, in general artificial neurons are simulated entirely in software. Not every processor is well suited for such calculations, even with the adopted software and/or hardware acceleration.
- Except for technical problems, ML have many limitations due to their nature. The key is their statistical approach – high accuracy requires a large set of examples to learn, and the taught structure will not cope with the effects that did not occur in these examples (gaps in data set).
- Too short learning period can cause erroneous results, even worse than DMO
- In turn, too long a learning period can cause problems related with computing power and too long/too expensive MLP calculations.
- In addition to the above point, learning must be done anew for each significant configuration change of the NWP model, which occurs quite regularly.
- Some compromise is definitely required, enabling correct results to be obtained in a not long time.

Thus, project – to be focused on the following activities and actions – must:

- Analyze and improve existing methods used in MLP (ANN, Recursive Least Squares, Multi-Linear Regression). This action would also include further evaluation of the data and planned set-up to be used for training, testing and calculating. Moreover, a resulting MLP post-processing should be universal enough to operate in both COSMO and ICON-LAM, and also be applicable to both deterministic and statistical (EPS) forecasts.
- Hence, estimation should be made of how the transition from COSMO to ICON in the learning period affects the skill of post-processing.
- Develop new MLP applications that would take into account deterministic, time-lagged forecasts. For example, running the forecasts four times a day, one can get a set of input (predictors), which would consist of even a dozen or so elements – previous forecast results. At the same time, such applications should include – if possible – ways to avoid errors due to lack of input data (cf. second bullet in Shortcomings and Deficiencies above) or bad ICs/ BCs.
- Develop seamless, probabilistic and spatial post-processing ML methods for wind speed, wind direction, wind gusts.
- Create a set of test cases that would be identical for all PP participants to verify their methods (link to WG5).
- Provide, for each method, an indication of its advantages and disadvantages, together with its applicability for the specific purpose (as a guideline for other COSMO partners)

Links to other Priority Projects

The project will use the results and data obtained in the AWARE Priority Project, specifically the ones from sub-task 4.2. After launching PP MILEPOST, task 4.2 of PP AWARE will be completed. This will allow, on the one hand, to avoid duplication of activities, and on the other hand, for generalization of results pertaining to both deterministic and statistical approaches.
The project will also make use of the experience and results obtained in PPs COTEKINO, SPRED and APSU in the field of preparing input data and post-processing results from EPS.

Risks

- Availability of appropriate computing resources for training/learning of Machine Learning Post-Processing systems;
- availability of datasets needed for learning, especially in case of transition from COSMO to ICON-LM;
- inappropriate (e.g. redundant or, on the contrary, too concise or restrained) selection of parameters (predictors) to setup the MLP system;
- missing input data for teaching, testing and/or validation.

Description of individual tasks

Task 0. Administrative Tasks

Due to a wide range of issues included in the project, administrative activities will be previewed, to keep a good collaboration/information flow between all participants (web conferences, workshops, etc.).

Deliverables

Project coordination, meetings, preparation of plans/reports, workshops and regular web-conferences organization. Preparation of the final PP report.

Contributors

IMGW, Andrzej Mazur, 0.12 FTE/COSMO year

Also in the preparation of final report (all parties involved, 03.2022-08.2022):

RHM – Philipp Bykov, Anastasia Bundel, 0.05 FTE
MCH – Daniel Cattani, Daniele Nerini, 0.05 FTE

Estimated needed resources: 0.34 FTE

Start: 09.2020; **End:** 08.2022

Task 1. General survey of Machine Learning.

This task will include a more detailed review of the history of ML methods and state-of-the-art in the world, taking into account the limitations, but also the advantages of using ML. This survey should justify further research. This task should use the results of sub-task 4.2, carried out under PP AWARE, as described above. However, planned activities in this task will be significantly expanded compared to PP AWARE. This is due to the fact that they will cover – in addition to existing issues related to intense convection phenomena – also "ordinary" meteorological elements. Moreover, other methods used in ANN/MLR/RLS should be checked to see if they can be used for general post-processing. For example, so far only the back-propagation method has been tested in ANN. MLR (Multi-Linear Regression) was limited to calculations not taking into account the weights (spatial and/or time) of individual predictors. RLS (Recursive Least-Squares) is definitely worth exploring in a similar direction.

Work steps:

- Review of literature on the history of use and MLP methods for continuous and discrete meteorological elements (temperature, wind speed, pressure, as well as precipitation, cloudiness, wind gusts).
- Create a catalogue of ongoing and planned developments in all member parties regarding ML and post-processing. This will allow the establishment of a platform for exchange of experiments that should be a benefit for all.
- Select the ML methods which should be investigated and document them.
- Formalization of the ML task: select the loss functions and metrics for each meteorological element.

Deliverables:

Reports on literature review and case studies; guidelines and suggestions for further research directions.

Contributors:

MCH – Daniel Cattani, 0.05 FTE
RHM – Philipp Bykov, 0.05 FTE
IMGW – Andrzej Mazur, 0.02 FTE

Estimated needed resources: 0.12 FTE/ COSMO year

Start: 09.2020; **End:** 02.2021

Task 2. Set-up and applications of Machine Learning Techniques for Post-processing

This task should make use of results of Task 1 in terms of both existing and suggested ML techniques to be introduced and developed. Previous work (basically in PP AWARE, Subtask 4.2, as well as operational post-processing tested at IMGW for the ensemble system applied for high-resolution COSMO model) suggested that the best direction of research will be to focus on ANN. However, other methods should not be neglected and should create some alternative to neural networking. Since MLR/RLS methods require shorter data set for learning (especially with small forgetting factor of RLS) they might be used, for example, during the COSMO-to-ICON transition period. Then, the proper ANN-learning data (from ICON results) are only being collected and the possibility of using the neural network will appear after some time.

Subtask 2.1 Set-up and application of ANNs

Work steps:

- For all participants involved: description of method(s) of interest
- Development and test of new methodologies for selecting an appropriate subset of predictors for the ANN; Including the classification (integer-valued) predictors such as station index, vegetation type, soil type (via embedding layers, see Rasp and Lerch, 2018) and the local predictors (Taillardat and Mestre, 2020)
- Collect and exchange the historical datasets of the selected predictors and the target values
- Time-lagged ANNs – necessary or sufficient condition to have a proper network?
- Deterministic vs. EPS-based ANNs
- ANN performance optimizations for MLP problem
  - Optimizations for the GPU
  - Optimize the parameters of the learning strategy
  - Examine the low and mixed precision ANN learning and inference in MLP problem (Micikevicius *et al.* 2017; Jacob *et al.* 2018).
- Tuning ANN's hyper-parameters for the MLP problem
- Determine the relations between the dataset volume, the number of features and the complexity of the optimal ANN for the MLP problem
- The methods to work around missing values
- Appraisal of the calculation time, computing resources etc. needed for learning and then for the operational phase.
- Develop MLP suggested package(s) to be shared among participating members

Deliverables:

Intermediate reports, contribution to the final report, list of basic verification scores, guidelines and suggestions for further research directions, package(s) to be disseminated, publication in JCR journal.

Contributors:

IMGW – Grzegorz Duniec, 0.15 FTE/COSMO year
MCH – Daniele Nerini, 0.4 FTE/COSMO year
RHM – Philipp Bykov, Gdaly Rivin, 0.3 FTE/COSMO year

Estimated needed resources: 0.85 FTE/COSMO year

Start: 09.2020; **End:** 08.2022

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

Work steps:

- Description of method(s) of interest
- Tuning MLR-based post-processing – parameters, weights, predictors, learning- vs. training and testing period
- Tuning RLS-based post-processing – parameters, weights, predictors, learning- vs. training and testing period (in terms of forgetting factor),
- Appraisal of the calculation time, computing resources etc. needed for learning and then for the operational phase.
- Development and testing of new method, development of package to be distributed among interested parties

Deliverables:

Intermediate reports and contribution to the final report, guidelines and suggestions for use, optionally: package(s) to be disseminated.

Contributors:

IMGW – Andrzej Mazur, Grzegorz Duniec, 0.1 FTE/COSMO year

Estimated needed resources: 0.1 FTE/COSMO year

Start: 09.2020; **End:** 09.2022

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

This task will be complementary to tasks 2.1 and 2.2 in terms of verification of test results. A special attention will be paid to the preparation of common sets of test data. A contribution in this task should be a definition and setup of the common frame and reference dataset, in order that various research groups can experiment and compare the performance of the various techniques in a transparent and reproducible way.

Work steps

- Verification of “added value” of ML vs. observations, DMO, Kalman-filter filter (e.g. Monache, 2011), spatial statistics (e.g. Hewson, 2020), etc.
- Focus on probabilistic scores and on stations seen/not seen during the learning.
- Estimation of the volume of data needed, specifically the time period on which one has to make the learning,
- Selection of methods available/applicable to the data type
- Selection of individual datasets to be used for training/learning, testing and verification
- Setting up and conducting experiments
- Dissemination and comparison of results to assess a performance and quality of various methods used.

Deliverables:

Intermediate reports, common verification dataset to be prepared and disseminated, common verification results scores for various elements/setups etc., contribution to the final report.

Contributors:

MCH – Daniele Nerini, Daniel Cattani, 0.2 FTE/COSMO year
RHM – Philipp Bykov, 0.3 FTE/COSMO year
IMGW – Joanna Linkowska, Grzegorz Duniec, Andrzej Mazur, 0.35 FTE/COSMO year

Estimated needed resources: 0.85 FTE/COSMO year

Start: 09.2020; **End:** 08.2022

Summary for tasks 0-3:

Total Resources needed: 4.06 FTEs

Project Participants

IMGW-PIB: Andrzej Mazur, Grzegorz Duniec, Joanna Linkowska
RHM: Philipp Bykov, Gdaly Rivin, Anastasia Bundel
MCH: Daniel Cattani, Daniele Nerini

References

S. Applequist, G. Gahrs, R. Peffer and X. Niu; Comparison of Methodologies for Probabilistic Quantitative Precipitation Forecasting. Weather and Forecasting, Volume 17, pages 783-799, 2002.

L. Breiman; Random Forests. Machine Learning, Volume 45, pages 5-32,2001.

D. Gagne, A. McGovern and M. Xue; Machine Learning Enhancement of Storm-Scale Ensemble Probabilistic Quantitative Precipitation Forecasts. Weather and Forecasting, Volume 29, pages1024-1043, 2014.

E. Kalnay, Atmospheric modeling, data assimilation and predictability. Cambridge University Press, 2003.

M. Tracton, and E. Kalnay; Operational Ensemble Prediction at the National Meteorological Center: Practical Aspects. Weather and Fore-casting, Volume 8, pages 379-398, 1993.

S. Rasp and S. Lerch; Neural Networks for Postprocessing Ensemble Weather Forecasts, Mon. Wea. Rev. Volume 146, pages 3885-3900, 2018.

M. Schaer: Downscaling of gridded wind fields using Deep Learning methods (Master's thesis), 2019.

S. Vannitsem, and Coauthors: Statistical Postprocessing for Weather Forecasts – Review, Challenges and Avenues in a Big Data World. arXiv preprint arXiv:2004.06582, 2020.

J. Williams; Using random forecasts to diagnose aviation turbulence. Machine Learning, Volume 95, pages 51-70, 2013.

V. M. Krasnopolsky. The Application of Neural Networks in the Earth System Sciences, 2013. 10.1007/978-94-007-6073-8_6.

S-A. Boukabara, V. Krasnopolsky, J.Q. Stewart, E.S. Maddy, N. Shahroudi, R.N. Hoffman. Leveraging Modern Artificial Intelligence for Remote Sensing and NWP Benefits and Challenges // *Bull. Amer. Meteor. Soc.* (2019) 100 (12): ES473–ES491. https://doi.org/10.1175/BAMS-D-18-0324.1

X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W. Wong, W. Woo. Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting, arXiv preprint arXiv:1506.04214, 2015

K. H. Zhou, Y. G. Zheng, B. Li, *et al.*, 2019: Forecasting different types of convective weather: A deep learning approach. J. Meteor. Res., 33(5), 797–809, doi: 10.1007/s13351-019-8162-6.

S. Agrawal, L. Barrington, C. Bromberg, J.R. Burge, C. Gazen, J. Hickey (2019). Machine Learning for Precipitation Nowcasting from Radar Images. arXiv preprint arXiv:*abs/1912.12132*.

R. Isaksson, 'Reduction of Temperature Forecast Errors with Deep Neural Networks', Dissertation, 2018.

Taillardat, M., & Mestre, O. (2020). From research to applications–Examples of operational ensemble post-processing in France using machine learning. *Nonlinear Processes in Geophysics Discussions*, 1-27.

Jacob, B., Kligys, S., Chen, B., Zhu, M., Tang, M., Howard, A., Kalenichenko, D. (2018). Quantization and training of neural networks for efficient integer-arithmetic-only inference. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2704-2713).

Micikevicius, P., Narang, S., Alben, J., Diamos, G., Elsen, E., Garcia, D., Wu, H. (2017). Mixed precision training. *arXiv preprint arXiv:1710.03740*.

Delle Monache, L., T. Nipen, Y. Liu, G. Roux, and R. Stull, 2011: Kalman Filter and Analog Schemes to Postprocess Numerical Weather Predictions. *Mon. Wea. Rev.*, **139**, 3554–3570, https://doi.org/10.1175/2011MWR3653.1.

T. Hewson, F. Pillosu. A new low-cost technique improves weather forecasts across the world. arXiv preprint arXiv:2003.14397, 2020.

Bykov, F.L. Statistical Correction of the COSMO Model Weather Forecasts Based on Neural Networks. Russ. Meteorol. Hydrol. 45, 141–152 (2020). https://doi.org/10.3103/S1068373920030012

Rasp, S. and S. Lerch, 2018: Neural Networks for Postprocessing Ensemble Weather Forecasts. Mon. Wea. Rev., 146, 3885–3900, https://doi.org/10.1175/MWR-D-18-0187.1

Rasp, S., Dueben, P. D., Scher, S., Weyn, J. A., Mouatadid, S., & Thuerey, N. (2020). WeatherBench: A benchmark dataset for data-driven weather forecasting. arXiv preprint arXiv:2002.00469.

Veldkamp, S. (2020). Statistical postprocessing of windspeed forecasts using convolutional neural networks (Master's thesis).

Baran, Á., Lerch, S., Ayari, M. E., & Baran, S. (2020). Machine learning for total cloud cover prediction. arXiv preprint arXiv:2001.05948.

Boukabara, S., V. Krasnopolsky, J.Q. Stewart, E.S. Maddy, N. Shahroudi, and R.N. Hoffman, 2019: Leveraging Modern Artificial Intelligence for Remote Sensing and NWP: Benefits and Challenges. Bull. Amer. Meteor. Soc., 100, ES473–ES491, https://doi.org/10.1175/BAMS-D-18-0324.1

Burke, A., N. Snook, D.J. Gagne II, S. McCorkle, and A. McGovern, 2020: Calibration of Machine Learning–Based Probabilistic Hail Predictions for Operational Forecasting. Wea. Forecasting, 35, 149–168, https://doi.org/10.1175/WAF-D-19-0105.1

Bremnes, J.B., 2020: Ensemble Postprocessing Using Quantile Function Regression Based on Neural Networks and Bernstein Polynomials. Mon. Wea. Rev., 148, 403–414, https://doi.org/10.1175/MWR-D-19-0227.1

Foresti, L., I.V. Sideris, D. Nerini, L. Beusch, and U. Germann, 2019: Using a 10-Year Radar Archive for Nowcasting Precipitation Growth and Decay: A Probabilistic Machine Learning Approach. Wea. Forecasting, 34, 1547–1569, https://doi.org/10.1175/WAF-D-18-0206.1

Cloud, K.A., B.J. Reich, C.M. Rozoff, S. Alessandrini, W.E. Lewis, and L. Delle Monache, 2019: A Feed Forward Neural Network Based on Model Output Statistics for Short-Term Hurricane Intensity Prediction. Wea. Forecasting, 34, 985–997, https://doi.org/10.1175/WAF-D-18-0173.1

Taillardat, M., Fougères, A. L., Naveau, P., & Mestre, O. (2019). Forest-based and semiparametric methods for the postprocessing of rainfall ensemble forecasting. Weather and Forecasting, 34(3), 617-634.

Appendix: Task table

FTE-y 2020 (from Sep 2020 to Aug 2021): 2.04
FTE-y 2021 (from Sep 2021 to Aug 2022): 2.02

FTE-y total (from Sep 2020 to Aug 2022): 4.06

Task	Sub task	Contributing scientist(s)	FTE Total	FTE 2021	FTE 2022	Start	Deliverables	Date of delivery
0 **total 0.34**		A. Mazur, D. Cattani, D. Lerini, A. Bundel, Ph. Bykov	0.24 0.05 0.05	0.12	0.12 0.05 0.05	Sep 2020	Project coordination, meetings, preparation of plans and reports, organization of workshops and conferences. Preparation of the final report.	Aug 2022
1 **total 0.12**		D. Cattani Ph. Bykov A. Mazur	0.05 0.05 0.02	0.05 0.05 0.02		Sep 2020	Reports on literature review and case studies; guidelines and suggestions for further research directions.	Feb 2021
2 **total 1.90**	**2.1**	G. Duniec D. Nerini Ph. Bykov G. Rivin	0.3 0.8 0.4 0.2	0.15 0.4 0.2 0.1	0.15 0.4 0.2 0.1	Sep 2020	Intermediate reports, contribution to the final report, list of basic verification scores, guidelines and suggestions for further research directions, package(s) to be disseminated.	Aug 2022
	2.2	A. Mazur, G. Duniec	0.2	0.1	0.1	Sep 2020	Intermediate reports and contribution to the final report, guidelines and suggestions for use, optionally: package(s) to be disseminated.	Aug 2022
3 **total 1.70**		D. Nerini, D. Cattani Ph. Bykov J. Linkowska, A. Mazur, G. Duniec	0.4 0.6 0.7	0.2 0.3 0.35	0.2 0.3 0.35	Sep 2020	Intermediate reports, common verification dataset to be prepared and disseminated, common verification results scores for various elements, setups etc., contribution to the final report.	Aug 2022