Master Thesis Project Proposals: Machine Learning for Weather Forecasting

Joel Oskarsson[1], Fredrik Lindsten[1], and Tomas Landelius[2]

[1] Division of Statistics and Machine Learning, Department of Computer and Information Science, Linköping University
[2]Swedish Meteorological and Hydrological Institute (SMHI)

October 24, 2023

1 Background

The problem of weather forecasting is a great scientific challenge with high relevance for society, industry and individuals. This problem is curretly tackled using Numerical Weather Prediction (NWP) systems, which combine vast amounts of physics knowledge with powerful computational resources in order to accurately model the atmosphere. Performing these computations is however time-consuming, limiting the possibility to model finer resolutions and detect rare events (Astitha et al., 2023).

Recent years have shown great progress in using machine learning models, in particular deep neural networks, for weather forecasting (Ben-Bouallegue et al., 2023). These data-driven models can match or even outperform existing NWP systems, while producing forecasts in a fraction of the time (Lam et al., 2022; Bi et al., 2023; Pathak et al., 2022). This is possible by training these machine learning models on forecasts from existing NWP systems. By training on *reanalysis* datasets, in which forecasts have been "corrected" using measurements (from weather stations, satellites, etc.), it is possible for the models to outperform the original NWP system.

While much work has focused on global weather forecasting, an equally interesting research direction is how to apply these machine learning methods for regional forecasting, building so called Limited Area Models (LAMs). This has been the focus of ongoing research conducted in collaboration between LiU and the Swedish Meteorological and Hydrological Institute (SMHI). See Oskarsson et al. (2023) for an overview of our work so far. We are now offering a number of thesis projects in this exciting and very active research area.

2 Projects

We offer several projects related to Machine Learning for Weather Forecasting, as detailed below. Our ambition is to find suitable candidates for multiple projects. We envision to then set up a collaborative environment for the involved students, to encourage collaboration on generic questions related to working with the data (see below) and getting familiar with the application area.

Note that you can apply to multiple projects at the same time.

2.1 Deep Generative Models for Weather States

As the atmosphere is a chaotic system, there is a great deal of uncertainty in a weather forecast. Because of this it is not enough to just produce a single deterministic forecast, but we are interested in the probability distribution over likely future weather states. Modeling this distribution is however challenging, due to its high dimensionality and spatio-temporal dependencies.

The goal of this project is to investigate the use of generative machine learning models for modeling the distribution of possible weather states. Methods like Generative Adversarial Networks (GANs) (Goodfellow et al., 2014), Normalizing Flows (Dinh et al., 2017) and Diffusion Models (Ho et al., 2020) have proven to be useful for modeling other types of high-dimensional data. Some initial attempts do also exist for utilizing these methods for weather data (Bihlo, 2021; Li et al., 2023). The project can start by training generative models for a single atmospheric variable, continue by adding on more variables and also including conditional generation based on previous states.

Possible research questions

- How can generative machine learning models be used for modeling the distribution of weather states on a limited area?
- What is a suitable way to evaluate the realism and usefulness of generated samples, from a meteorological perspective?
- What are the trade-offs in terms of training time, memory requirements, sample diversity and sample quality?
- How can conditioning on previous weather states be incorporated in such a generative model?

Prerequisites

- Experience with implementing and training deep learning models in relevant software frameworks.
  - Preferably experience with GPU training of such models.
- Sufficient background in statistics and statistical machine learning.
  - Preferably experience with deep generative models.

2.2 Exploring Model Families for Limited Area Modeling

In our existing research we have utilized graph-based machine learning models for LAM forecasting (Oskarsson et al., 2023). For global weather forecasting, the graph-based approach is far from the only one and a diverse set of models and architectures have been successfully applied. It is of large interest to investigate also how other types of machine learning models can be used in the LAM setting. Some architectures worth investigating are Convolutional Neural Networks (CNNs), Transformers (Bi et al., 2023; Chen et al., 2023; Lessig et al., 2023) and Neural Operators (Pathak et al., 2022; Bonev et al., 2023). The intention is mainly to choose one of these to investigate closer (which one depends on the interest of the student). A possible goal of this project is to implement such models as part of our existing open-source code base for LAM machine learning models[1].

Possible research questions

- How can (*CNNs/Transformers/Neural Operators*) be used for the LAM weather forecasting task?
- How does the model compare to the existing graph-based ones in terms of forecast accuracy?
- Are there benefits to this type of model in terms of training time, inference time or memory requirements?

Prerequisites

- Experience with implementing and training deep learning models in relevant software frameworks.
- Solid programming skills, in particular in python.
- Knowledge of and experience with good software engineering practices is beneficial.

[1] `https://github.com/joeloskarsson/neural-lam`

2.3 Utilizing Global Models for Limited Area Forecasting

A few of the machine learning models for global weather forecasting have been framed as foundation models, with the premise that they can be fine-tuned to many different tasks (Lessig et al., 2023; Nguyen et al., 2023). One such task is regional LAM forecasting. Still, there are many crucial details in how such fine-tuning can be performed given a large LAM dataset.

The goal of this project is to investigate how global foundation models best can be fine-tuned for LAM forecasting. To investigate this, a global model should be fine-tuned on our existing data for the Nordic region. A possible next step of the project would be to investigate methods for integrating information from (fine-tuned) global models into local models.

Possible research questions

- How can global foundation models be fine-tuned for weather forecasting in the LAM setting?
- How can LAM data on a local grid be used for fine-tuning global models defined on a different spatial grid?
- How does the amount of data used for fine-tuning a global foundation model impact its forecasting performance?
- How does fine-tuned global models compare to pure LAM models (Oskarsson et al., 2023) in terms of forecasting performance and computational time (for training and inference)?
- How can representations or predictions from global foundation models be used to improve LAM machine learning models?

Prerequisites

- A solid understanding of deep learning models and experience with training such models in practice.
  - Experience with GPU training is beneficial.
- Experience in working with large datasets on Linux systems is beneficial.

2.4 Learning the Data Assimilation Process

Data assimilation is the process of estimating a system state (the state of the atmosphere) by combining known (or learned) dynamics with real-world observations. In weather forecasting this means combining a forecast from an NWP system or machine learning model with measurements from weather stations and satellites, in order to get a more accurate estimate of the state of the atmosphere (Kalnay, 2002). Such estimates of the atmospheric state are important both for analysis of past weather-events and for use as initial conditions for new forecasts. Even more valuable are probabilistic estimates, capturing the distribution of the state.

Just as with NWP forecasting, the data assimilation process is computationally expensive and rely on approximations in the relationships between the state and observations (Coiffier, 2011). This begs the question, can we utilize machine learning methods also for the data assimilation process? The goal of this project would be to investigate how we can use existing data to train machine learning models to efficiently approximate this complex state estimation process. The problem could either be framed as just approximating a mapping (forecast + observations $\rightarrow$ state) using e.g. deep learning or viewed as a more probabilistic Bayesian inference problem. As an extensions it could be investigated if it is possible to surpass the performance of existing data assimilation methods by utilizing machine learning forecasting models.

Possible research questions

- How should the data assimilation problem be framed as a machine learning / Bayesian inference problem?
- How does machine learning based data assimilation compare to traditional methods in terms of accuracy of estimation and computational cost?
- What are the benefits in using machine learning forecasting models over traditional NWP forecasting models in the data assimilation process?

Prerequisites

- A good understanding of machine learning methods and problem formulations.
- A solid mathematical background, in particular in optimization and linear algebra.
- If following the probabilistic route: A solid background in statistics and knowledge of Bayesian statistics.

2.5 Time-Continuous Weather Forecasting

While the physical weather system is continuous in both space and time, we need to perform discretizations in order to build numerical models. The coarseness of such discretizations have a large impact on the computational cost of weather forecasting models. Still, the underlying equations dictating NWP are time-continuous Partial Differential Equations (PDEs) and Ordinary Differential Equations (ODEs).

Existing machine learning models for weather forecasting use discrete time steps of ≥ 1 h (Lam et al., 2022; Pathak et al., 2022; Bi et al., 2023; Oskarsson et al., 2023). It would be valuable to get more fine-grained temporal resolution, with forecasts available also at intermediate time points. One way to achieve this would be to explicitly frame such a model as a discretization of a time-continuous system. Existing graph-based models (Keisler, 2022; Lam et al., 2022; Oskarsson et al., 2023) feature *residual connections* between the layers of the model and between predictions. Such residual connections can be interpreted as a discretization of an ODE solved through the Euler method Behrmann et al. (2019). This leads the way to a possible approach for building time-continuous weather forecasting models where the weather state can be evaluated for arbitrary future time points.

The goal of this project is to investigate time-continuous formulations of machine learning models for weather forecasting. Starting from existing models, a first step is to explicitly show how these can be formulated as ODE discretizations. With insight from this, we can start to take smaller time steps and evaluate if also these intermediate steps yield useful predictions of the weather state. In extensions it is possible to fully combine existing models with *Neural ODE* components (Chen et al., 2018; Rubanova et al., 2019) and investigate methods for integrating these with the graph architectures (Poli et al., 2021).

Possible research questions

- How can existing machine learning models for weather forecasting explicitly be formulated as discretized solutions to ODEs?
- Are there benefits to a finer temporal discretization, taking many smaller time steps when producing a weather forecast?
- Does intermediate time steps still represent useful weather states? Does the solution yield a meteorologically feasible sequence of states?
- How can time-continuous components be combined with latent graph architectures (Lam et al., 2022; Oskarsson et al., 2023)?
- What are the trade-offs in terms of computational time and memory requirements for choosing the lenght of time steps and ODE solver?

Prerequisites

- A solid mathematical background, preferably including some experience with numerical ODE solving.
- An interest in mathematical modeling and numerical analysis.
- Experience with deep learning models.

3 Data

The main dataset, that can be used for all these projects, comes from the MetCoOp Ensemble Prediction System (MEPS) (Müller et al., 2017). MEPS is a LAM NWP system producing forecasts for the nordic region. This dataset contains roughly 10 weather forecasts per day from a two year period. Each such forecast is a 66 h time series of weather states with 1 h time steps. The weather states contain values for 17 different variables, all on a 238×268 grid (corresponding to a spatial resolution of 10 km). This is the same dataset used in Oskarsson et al. (2023).

The project described in section 2.4 will make use of additional data containing real-world measurements. Such data exists archived at SMHI and will be provided to the student working on this project.

Apart from this data there is also the possibility to use the WeatherBench dataset (Rasp et al., 2020), which contains 40 years of global weather states but at a much coarser spatial resolution. Generally, the details of the dataset format used will be tailored to the specific project and methodology, for example by choosing only a subset of the variables, spatially sub-sampling the weather states or using longer time steps.

4 How to apply

Apply by sending

- CV
- A transcript with grades
- Which project(s) you are interested in
- A brief motivation on why you are interested in the project(s)
- Pointers to any past experience that makes you a particularly suitable candidate (e.g. in relation to the *prerequisites* of each project)

to `joel.oskarsson@liu.se`. Write "*Thesis Project Application: ML for Weather Forecasting*" in the subject line. The last date to apply is November 15. If you have any questions about the projects or application feel free to reach out.

References

M. Astitha, L. Magnusson, and E. Nikolopoulos. Weather forecasting (subchapter 1.2). In *Extreme Weather Forecasting*, pages 7–14. Elsevier, 2023.

J. Behrmann, W. Grathwohl, R. T. Chen, D. Duvenaud, and J.-H. Jacobsen. Invertible residual networks. In *International conference on machine learning*, pages 573–582. PMLR, 2019.

Z. Ben-Bouallegue, M. C. A. Clare, L. Magnusson, E. Gascon, M. Maier-Gerber, M. Janousek, M. Rodwell, F. Pinault, J. S. Dramsch, S. T. K. Lang, B. Raoult, F. Rabier, M. Chevallier, I. Sandu, P. Dueben, M. Chantry, and F. Pappenberger. The rise of data-driven weather forecasting. *arXiv preprint arXiv:2307.10128*, 2023.

K. Bi, L. Xie, H. Zhang, X. Chen, X. Gu, and Q. Tian. Accurate medium-range global weather forecasting with 3d neural networks. *Nature*, pages 1–6, 2023.

A. Bihlo. A generative adversarial network approach to (ensemble) weather prediction. *Neural Networks*, 139:1–16, 2021. ISSN 0893-6080. doi: 10.1016/j.neunet.2021.02.003. URL `https://www.sciencedirect.com/science/article/pii/S0893608021000459`.

B. Bonev, T. Kurth, C. Hundt, J. Pathak, M. Baust, K. Kashinath, and A. Anandkumar. Spherical Fourier neural operators: Learning stable dynamics on the sphere. In *Proceedings of the 40th International Conference on Machine Learning*, pages 2806–2823, 2023.

K. Chen, T. Han, J. Gong, L. Bai, F. Ling, J.-J. Luo, X. Chen, L. Ma, T. Zhang, R. Su, Y. Ci, B. Li, X. Yang, and W. Ouyang. FengWu: Pushing the skillful global medium-range weather forecast beyond 10 days lead. *arXiv preprint arXiv:2304.02948*, 2023.

R. T. Chen, Y. Rubanova, J. Bettencourt, and D. K. Duvenaud. Neural ordinary differential equations. In *Advances in neural information processing systems*, volume 31, 2018.

J. Coiffier. *Fundamentals of numerical weather prediction.* Cambridge University Press, 2011.

L. Dinh, J. Sohl-Dickstein, and S. Bengio. Density estimation using real NVP. In *International Conference on Learning Representations*, 2017. URL `https://openreview.net/forum?id=HkpbnH9lx`.

I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014.

J. Ho, A. Jain, and P. Abbeel. Denoising diffusion probabilistic models. In H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 6840–6851. Curran Associates, Inc., 2020.

E. Kalnay. *Atmospheric Modeling, Data Assimilation and Predictability.* Cambridge University Press, 2002.

R. Keisler. Forecasting global weather with graph neural networks. *arXiv preprint arXiv:2202.07575*, 2022.

R. Lam, A. Sanchez-Gonzalez, M. Willson, P. Wirnsberger, M. Fortunato, A. Pritzel, S. Ravuri, T. Ewalds, F. Alet, Z. Eaton-Rosen, W. Hu, A. Merose, S. Hoyer, G. Holland, J. Stott, O. Vinyals, S. Mohamed, and P. Battaglia. GraphCast: Learning skillful medium-range global weather forecasting. *arXiv preprint arXiv:2212.12794*, 2022.

C. Lessig, I. Luise, B. Gong, M. Langguth, S. Stadler, and M. Schultz. AtmoRep: A stochastic model of atmosphere dynamics using large scale representation learning. *arXiv preprint arXiv:2308.13280*, 2023.

L. Li, R. Carver, I. Lopez-Gomez, F. Sha, and J. Anderson. SEEDS: Emulation of weather forecast ensembles with diffusion models, 2023. URL `http://arxiv.org/abs/2306.14066`.

M. Müller, M. Homleid, K.-I. Ivarsson, M. A. Ø. Køltzow, M. Lindskog, K. H. Midtbø, U. Andrae, T. Aspelien, L. Berggren, D. Bjørge, P. Dahlgren, J. Kristiansen, R. Randriamampianina, M. Ridal, and O. Vignes. AROME-MetCoOp: A nordic convective-scale operational weather prediction model. *Weather and Forecasting*, 32(2):609–627, 2017.

T. Nguyen, J. Brandstetter, A. Kapoor, J. K. Gupta, and A. Grover. ClimaX: A foundation model for weather and climate. In *Proceedings of the 40th International Conference on Machine Learning*, pages 25904–25938, 2023.

J. Oskarsson, T. Landelius, and F. Lindsten. Graph-based neural weather prediction for limited area modeling, 2023. URL `http://arxiv.org/abs/2309.17370`.

J. Pathak, S. Subramanian, P. Harrington, S. Raja, A. Chattopadhyay, M. Mardani, T. Kurth, D. Hall, Z. Li, K. Azizzadenesheli, P. Hassanzadeh, K. Kashinath, and A. Anandkumar. Fourcastnet: A global data-driven high-resolution weather model using adaptive fourier neural operators. *arXiv preprint arXiv:2202.11214*, 2022.

M. Poli, S. Massaroli, C. M. Rabideau, J. Park, A. Yamashita, H. Asama, and J. Park. Continuous-depth neural models for dynamic graph prediction. *arXiv preprint arXiv:2106.11581*, 2021.

S. Rasp, P. D. Dueben, S. Scher, J. A. Weyn, S. Mouatadid, and N. Thuerey. WeatherBench: A benchmark dataset for data-driven weather forecasting. *Journal of Advances in Modeling Earth Systems*, 12(11), 2020. ISSN 1942-2466, 1942-2466. doi: 10.1029/2020MS002203.

Y. Rubanova, R. T. Q. Chen, and D. K. Duvenaud. Latent ordinary differential equations for irregularly-sampled time series. In *Advances in Neural Information Processing Systems*, volume 32, 2019.