Methods for simulation of weather and climate (mobility).

1st Orfeo Tërkuçi
Science
University of Antwerp
Boechout, Belgium
orfeo.terkuci@student.uantwerpen.be

2nd Jason Liu
Science
University of Antwerp
Lier, Belgium
jason.liu@student.uantwerpen.be

***Abstract*—In this paper we discuss about different methods for simulating weather and climate. We discuss about STEP, OpenWeatherMap, Meteomatics, ClimaX, and GraphCast. We discuss about their advantages and disadvantages, and their motivation.**

I. Introduction

To simulate both mobility and energy use, realistic modeling of the weather is necessary. In this paper, we will discuss ways in which we can do this. Considering the effect of the position of the virtual city, seasons, and the ability to represent specific phenomena (storms, snow, heat waves).

II. Topics

A. STEP (short-term ensemble prediction system) computation algorithm

Ensemble forecasting is a method used in or within numerical weather prediction. Instead of making a single forecast of the most likely weather, a set (or ensemble) of forecasts is produced. This set of forecasts aims to give an indication of the range of possible future states of the atmosphere. Ensemble forecasting is a form of Monte Carlo analysis. The multiple simulations are conducted to account for the two usual sources of uncertainty in forecast models:

- The errors introduced by the use of imperfect initial conditions, amplified by the chaotic nature of the evolution equations of the atmosphere, which is often referred to as sensitive dependence on initial conditions;
- The errors introduced because of imperfections in the model formulation, such as the approximate mathematical methods to solve the equations.

Ideally, the verified future atmospheric state should fall within the predicted ensemble spread, and the amount of spread should be related to the uncertainty (error) of the forecast. In general, this approach can be used to make probabilistic forecasts of any dynamical system, and not just for weather prediction [1].

Today ensemble predictions are commonly made at most of the major operational weather prediction facilities worldwide, including

- National Centers for Environmental Prediction (NCEP of the US)
- European Centre for Medium-Range Weather Forecasts (ECMWF)
- Météo-France

[1]

Key Components of Ensemble Prediction Systems:

- Base Models: Ensemble systems typically consist of multiple base models, each trained independently on the same or different datasets.
- Diversity: The effectiveness of ensemble methods relies on the diversity among the base models. If the models are too similar, the ensemble may not provide significant improvements.
- Combination Method: Ensemble methods employ a combination or aggregation method to merge the predictions of individual models into a single, more accurate prediction.
- Weighting: In weighted averaging, each base model's prediction is assigned a weight, and the final prediction is a weighted sum of individual predictions.
- Training and Validation: Base models are trained on historical data, and the ensemble system is validated and calibrated using separate datasets to ensure its accuracy

An ensemble-based probabilistic precipitation forecasting scheme has been developed that blends an extrapolation nowcast with a downscaled NWP forecast, known as STEPS: Short-Term Ensemble Prediction System. The uncertainties in the motion and evolution of radar-inferred precipitation fields are quantified, and the uncertainty in the evolution of the precipitation pattern is shown to be the more important.

The use of ensembles allows the scheme to be used for applications that require forecasts of the probability density function of areal and temporal averages of precipitation, such as fluvial flood forecasting—a capability that has not been provided by previous probabilistic precipitation nowcast schemes.

The output from a NWP forecast model is downscaled so that the small scales not represented accurately by the model are injected into the forecast using stochastic noise.

This allows the scheme to better represent the distribution of precipitation rate at spatial scales finer than those adequately resolved by operational NWP.

1) Advantages of Ensemble Forecasting:

- Improved Accuracy: Ensemble forecasting often provides more accurate predictions than individual models by leveraging the collective knowledge of diverse models.
- Quantifying Uncertainty: Ensemble systems offer a way to estimate the uncertainty associated with predictions. The spread or variability among ensemble members provides a measure of prediction confidence.
- Reduced Overfitting: By combining multiple models with different training data or parameters, ensemble methods reduce the risk of overfitting to a particular dataset.
- Enhanced Generalization: Ensemble methods can generalize well to different scenarios and datasets, making them versatile for various applications.
- Flexibility: Ensemble systems can incorporate a variety of models and data sources, making them adaptable to different prediction tasks and domains.

2) Disadvantages of Ensemble Forecasting:

- Difficulty in Model Selection: Selecting appropriate models for the ensemble requires careful consideration, and the effectiveness of the ensemble may be sensitive to the choice of models.
- Potential for Redundancy: If the base models in the ensemble are too similar, there might be limited diversity, reducing the effectiveness of the ensemble approach.
- Overemphasis on Certain Models: In some cases, if a particular model consistently outperforms others, that dominant model might heavily influence the ensemble's performance.
- Increased Training Time: Training multiple models requires additional time and computational resources compared to training a single model.

3) Motivation: Ensemble forecasting is the prevailing method used for weather forecasting worldwide today, by most of the major operational weather prediction facilities. It is widely used because it is more accurate than individual models, and it provides a measure of prediction confidence.

B. Openweather API

OpenWeatherMap is a weather API that provides weather data for any location on the globe. It uses machine learning (ML) to significantly advance both the accuracy and computing speed of global assemble forecasting models, a practice that was impossible only a few years ago [3].

OpenWeatherMap offers a variety of APIs, including the One Call API 3.0, which provides current weather and forecasts, minute forecast for 1 hour, hourly forecast for 48 hours, daily forecast for 8 days, and government weather alerts [3]. The API also provides weather data for any timestamp for 40+ years historical archive and 4 days ahead forecast, daily aggregation of weather data for 40+ years archive and 1.5 years ahead forecast, hourly forecast for 4 days, 16 days forecast, and climatic forecast for 30 days [3].

In addition, OpenWeatherMap provides beautiful multi-layer maps that create the visual perception of weather. You can choose from a set of OpenWeather Model layers such as wind, temperature, pressure, and others, or select radar data for a detailed precipitation picture [3].

1) Data: Openweathermap makes use of different soureces of data, such as:

- **ECMWF** - The European Centre for Medium-Range Weather Forecasts (ECMWF) is an independent intergovernmental organisation supported by most of the nations of Europe and is based at Shinfield Park, Reading, United Kingdom. The center's operational forecasts are produced from its "Forecast System" (sometimes informally known in the United States as the "European model") which is run every twelve hours and forecasts out to ten days [6]. ECMWF is also developing their own digital twin, to predict weather and climate change [7].
- **NOAA** - The National Oceanic and Atmospheric Administration (NOAA) is an American scientific agency within the United States Department of Commerce that focuses on the conditions of the oceans, major waterways, and the atmosphere [8].
- **More** - Openweathermap also makes use of other sources of data, such as: GFS, NEMS, ICON, AROME, and others [4].

Because of these data sources Openweathermap is able to provide accurate weather data for any location on the globe.

2) Advantages: Here are some of the advantages of OpenWeatherMap

- **Accuracy and speed** - OpenWeatherMap uses machine learning to significantly advance both the accuracy and computing speed of global assemble forecasting models, a practice that was impossible only a few years ago [3].
- **Global coverage** - OpenWeatherMap provides weather data for any location on the globe [3].
- **Historical data** - OpenWeatherMap provides weather data for any timestamp for 40+ years historical archive and 4 days ahead forecast, daily aggregation of weather data for 40+ years archive and 1.5 years ahead forecast, hourly forecast for 4 days, 16 days forecast, and climatic forecast for 30 days [3].
- **Multi-layer maps** - OpenWeatherMap provides beautiful multi-layer maps that create the visual perception of weather [3].

3) Disadvantages: Here are some of the disadvantages of OpenWeatherMap

- **Limited free plan** - OpenWeatherMap offers a free plan that allows 60 calls per minute, one million calls per month, and 5-day forecast, but it does not include historical data [3].

4) Motivation: OpenWeatherMap is a weather API that provides accurate weather data for any location on the globe.

Making use of the STEP algorithm explained previously. Combining this algorithm with machine learning (ML) to significantly advance both the accuracy and computing speed.

Also with the free plan we get up to 60 calls per minute, one million calls per month, and 5-day forecast, which is more than enough for our application. Although the free plan doesn't include historical data, we can save the data and make our own historical data. The Downside of this approach is that we have to wait for the data to be collected.

In case we want forecast data for more than 5 days, we have to implement our own model to predict the weather (making use of the collected historical data), or upgrade to a paid plan.

C. MeteoMatics

Meteomatics is a weather API that provides weather data for any location on the globe. It works in a similar way to OpenWeatherMap, but it has some different advantages and disadvantages.

1) vs Openweathermap: Because it works in a similar way to OpenWeatherMap, we will only mention the differences between the two. Meteomatics offers more features in its free plan, such as historical data of the past 24 hours, but the free plan is limited to 500 calls per day.

Another advantage is that Meteomatics provides a lott of connectors to different program languages, for instance Python (that is free of charge) [9]. This makes it easier to implement the API in our application.

Also, the response time of Meteomatics is faster than OpenWeatherMap. While it is not directly stated how fast Openweathermap response time is, it is known that it takes at most 1s, while Meteomatics is around 30ms.

2) Motivation: Meteomatics is a weather API that provides accurate and fast weather data. It has a free plan that offers more features than OpenWeatherMap, but it is limited to 500 calls per day. So in case we want to make more than 500 calls per day, we have to upgrade to a paid plan or choose for OpenWeatherMap.

It also provides a lot of connectors to different program languages, which makes it easier to implement the API in our application, because we are split in teams and each team might have a different programming language this is a great advantage.

D. ClimaX

ClimaX is the first foundation model designed to perform a wide variety of weather and climate modeling tasks. For weather, these tasks include standard forecasting tasks of relevant weather variables like temperature, humidity, etc. with various lead-times at various resolutions, both globally and regionally. For climate, ClimaX can help to make better long-term projections, or to downscale lower resolution model outputs to higher resolutions. At its core, ClimaX is a multi-dimensional image-to-image translation architecture based on Vision Transformers (ViT).
ViT-based architectures are especially well suited for modeling weather and climate phenomena since they naturally tokenize the spatial nature of multiscale data akin to different spatial-temporal inputs. Additionally, they offer the opportunity to extend tokenization towards a wide range of multichannel features [10].

1) Results highlights: Forecasting the future values of key weather variables at different temporal horizons is critical to ensuring the safety of communities and infrastructure around the world.
ERA5 is the latest climate reanalysis produced by ECMWF, providing hourly data on many atmospheric, land-surface and sea-state parameters together with estimates of uncertainty [11].
ERA5 reanalysis data from the ECMWF underlies as the key source of data for training and evaluating machine learning models on this task with performance of Operation IFS being the current state-of-the art numerical weather prediction baseline.
ClimaX when fine-tuned on the same ERA5 data, even at medium resolutions 1.40625°already performs comparably, if not better than IFS on short and medium-range predictions, while being substantially better at longer horizon predictions [10].

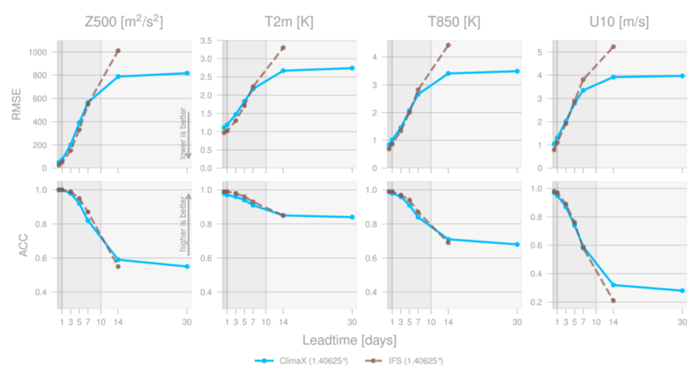

Fig. 1. ClimaX vs IFS on global forecasting of key weather variables at different lead time horizons

In the graphs above, we compare the performance of ClimaX vs IFS on global forecasting of key weather variables at different lead time horizons:

- Temperature T2M (2m above ground)
- Temperature T850 (850hPa)
- Wind speed U10M (10m above ground)
- Geo-potential height Z500 (500hPa)

The graphs on the first row show the RMSE (root mean squared error) of the predictions, while the graphs on the second row show the ACC (accuracy) of the predictions. The x-axis shows the lead time in days. As we can see, for example, on the Z500 graph, both models start at 100%

accuracy and 0 RMSE.
As the lead time increases, the accuracy of both models decreases, and the error increases.
What a decrease in accuracy means is that the model is less confident in its predictions.
What an increase error means that the model is less accurate and less trustworthy.

For example, on the Z500 graphs, we can see that the accuracy of both models starts at 1.0 (100%), and decreases to around 0.6 (60%) at 14-days lead time.
At the same time, the RMSE of both models starts at 0 and increases to around 800 for ClimaX and 1000 for IFS at 14 days lead time.

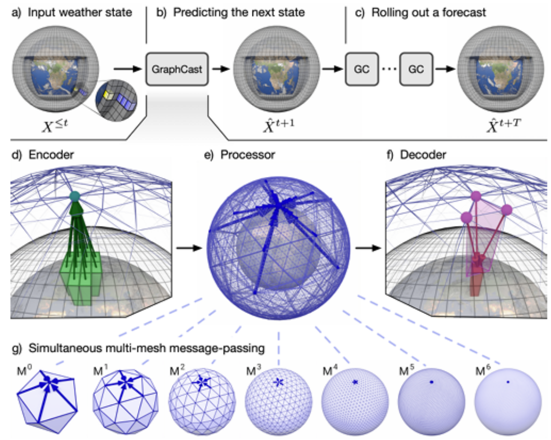

Fig. 2. Steps of GraphCast

But as we can see, ClimaX performs somewhat comparably to IFS on short and medium-range predictions, while being substantially better at longer horizon predictions (in most of these graphs, 14 days and above).

2) Motivation: ClimaX provides a variety of different spatial-temporal resolutions and input channels, which can be used for a wide variety of weather and climate modeling tasks. Being benchmarked against other state-of-the-art models, it holds its own and even outperforms them in some cases.

E. GraphCast

GraphCast is a machine learning-based method developed by Google DeepMind for medium-range global weather forecasting. It is an autoregressive model based on graph neural networks and a novel high-resolution multiscale mesh representation. It is a mesh representation that is used to represent the Earth's surface and atmosphere. The mesh is a grid of points that are connected by lines to form triangles. The mesh is multiscale, meaning that it has different resolutions at different levels. The mesh is high-resolution, meaning that it has a high density of points, which allows for more accurate predictions. GraphCast is trained on historical weather data from the European Centre for Medium-Range Weather Forecasts (ECMWF)'s ERA5 reanalysis archive [12].

It starts with the current state of Earth's weather and data about the weather six hours ago. Then, it makes a prediction about what the weather will look like six hours from now.
GraphCast then feeds those predictions back into the model, performs the same calculation, and spits out longer-term forecasts [12].

GraphCast's multiscale mesh representation.
a) First we insert the data into the model. b) Then we perform GraphCast to predict data, we feed this data back into the model. c) We perform GraphCast again to predict data, we feed this data back into the model and do this repeatedly. d) The encoder component maps local regions of the input (the green boxes) into nodes of the multi-mesh graph representation. e) The processor component performs a series of graph neural network (GNN) message passing steps to update the node features. f) The decoder component maps the updated node features back to the output (the purple boxes). g) The multi-mesh is a set of icosahedral meshes of increasing resolution, from the base mesh (M^0, 12 nodes) to the finest resolution (M^6 , 40, 962 nodes), which has uniform resolution across the globe. Each node belongs to a particular mesh resolution, and is connected to all neighboring nodes at the same resolution, as well as higher resolutions. The learned message-passing over the different meshes' edges happens simultaneously, so that each node is updated by all of its incoming edges.

The package contains example code to run and train GraphCast. It also provides three pretrained models: GraphCast, the high-resolution model used in the GraphCast paper (0.25 degree resolution, 37 pressure levels), trained on ERA5 data from 1979 to 2017, Grap_Cast_small, a smaller, low-resolution version of GraphCast (1 degree resolution, 13 pressure levels, and a smaller mesh), trained on ERA5 data from 1979 to 2015, useful to run a model with lower memory and compute constraints, GraphCast_operational, a high-resolution model (0.25 degree resolution, 13 pressure levels) pretrained on ERA5 data from 1979 to 2017 and fine-tuned on HRES data from 2016 to 2021 [13].

1) Advantages: Here are some of the advantages of GraphCast:

- **Accuracy** - GraphCast outperforms the most accurate previous ML-based weather forecasting model on percent 99.2 of the 252 targets it reported on [12].
- **Speed** - GraphCast can generate a 10-day forecast (35 gigabytes of data) in under 60 seconds on Cloud TPU v4 hardware [12].

2) Disadvantages: Despite the advancement, GraphCast has limitations.

- **Accuracy** - GraphCast is not yet able to predict extreme

weather events, such as hurricanes, tornadoes, and floods. At least not accurately. It did not outperform conventional models in all scenarios, such as the sudden intensification of Hurricane Otis, which hit Acapulco with minimal warning on October 25, 2023 [15].
- **Transparency** - GraphCast is a black box model, meaning that it is not possible or hard to understand how it works.

3) Motivation: GraphCast is a fast and accurate weather forecasting model, that outperforms conventional models in most scenarios.
It also provides pretrained models, which makes it easier to implement in our application. No need to train our own model.

It also makes use of the ECMWF's ERA5 reanalysis archive, which is the same data source that OpenWeatherMap uses. It is known to be accurate.

III. Conclusion

We have looked at a variety of different methods for simulating weather and climate. We have looked at STEP, OpenWeatherMap, Meteomatics, ClimaX, and GraphCast.

There are many available methods for simulating weather and climate.
Some of them are more accurate than others, some of them are faster than others, and some of them are easier to use than others.
This includes trained models, APIs, and algorithms.

STEP is an algorithm for computing ensemble forecasts. It is the standard method used in the industry for computing ensemble forecasts by a good deal of different weather forecasting agencies, including the European Centre for Medium-Range Weather Forecasts (ECMWF).

OpenWeatherMap and Meteomatics are APIs for accessing weather data. They are both easy to use, and provide a lot of different current and historical data.
Both having their pros and cons, but overall, they both have a free plan, which makes them both viable options for our application.

ClimaX and GraphCast are both machine learning models for weather and climate forecasting. They are both very accurate, and very fast.
They both provide pretrained models, which makes them easier to implement in our application.

In general, these methods are all viable options for our application and knowing which one to use depends on the specific use case. In case we want to be fully in control of the data, we can use STEP.
In case we want to use machine learning, we can use ClimaX or GraphCast, and make use of the pretrained models.
In case we want to use an API, we can use OpenWeatherMap or Meteomatics.

References

[1] Ensemble forecasting (2023, 17 August) In Wikipedia https://en.wikipedia.org/wiki/Ensemble_forecasting [Accessed 13 November 2023]
[2] STEPS: A probabilistic precipitation forecasting scheme which merges an extrapolation nowcast with downscaled NWP By NEILL E. BOWLER (Met Office, Exeter, UK), CLIVE E. PIERCE (Met Office, Wallingford, UK) and ALAN W. SEED (BMRC, Bureau of Meteorology, Melbourne, Australia) (Received 1 July 2004; revised 18 April 2006) [Accessed 13 November 2023]
[3] Openweathermap.org (2012 - 2023). OpenWeatherMap. [Online] Available at: https://openweathermap.org/api [Accessed 13 November 2023].
[4] Openweathermap.org (2012 - 2023). OpenWeatherMap. [Online] Available at: https://docs.openweather.co.uk/technology [Accessed 13 November 2023].
[5] Openweathermap.org (2012 - 2023). OpenWeatherMap. [Online] Available at: https://openweathermap.org/guide [Accessed 13 November 2023].
[6] ECMWF (2023, 9 September) In Wikipedia https://en.wikipedia.org/wiki/European_Centre_for_Medium-Range_Weather_Forecasts [Accessed 18 November 2023]
[7] ECMWF [Online] https://stories.ecmwf.int/destination-earth/index.html [Accessed 18 November 2023]
[8] National Oceanic and Atmospheric Administration (2023, 15 November) In Wikipedia https://en.wikipedia.org/wiki/National_Oceanic_and_Atmospheric_Administration [Accessed 18 November 2023]
[9] Meteomatics [Online] https://www.meteomatics.com [Accessed 18 November 2023]
[10] Introduction to ClimaX: The first foundation model for weather and climate [Online] https://www.microsoft.com/en-us/research/group/autonomous-systems-group-robotics/articles/introducing-climax-the-first-foundation-model-for-weather-and-climate/
[11] ECMWF ERA5 Climate reanalysis [Online] https://climate.copernicus.eu/climate-reanalysis
[12] GraphCast: Learning skillful medium-range global weather forecasting (2022, 22 December) In DeepAI https://deepai.org/publication/graphcast-learning-skillful-medium-range-global-weather-forecasting [Accessed 18 November 2023]
[13] Graphcast (2022, 22 December) In GitHub https://github.com/google-deepmind/graphcast [Accessed 18 November 2023]
[14] Google GraphCast uses AI and machine learning to predict weather (2023, 14 November) In BGR https://bgr.com/tech/google-graphcast-uses-ai-and-machine-learning-to-predict-weather/ [Accessed 18 November 2023]
[15] AI outperforms conventional weather forecasting for the first time: Google study (2023, 14 November) In Ars Technica https://arstechnica.com/science/2023/11/ai-outperforms-conventional-weather-forecasting-for-the-first-time-google-study/ [Accessed 18 November 2023]