International Journal of Agriculture and Environmental Research

ISSN: 2455-6939

Volume: 11, Issue: 02 "March-April 2025"

CLIMATE INFORMATICS: CLIMATE PREDICTION AND MACHINE LEARNING

Surabhi Singh

Faculty, Pillai College of Engineering, New Mumbai, India.

DOI: https://doi.org/10.51193/IJAER.2025.11208

Received: 11 Mar. 2025 / Accepted: 20 Mar. 2025 / Published: 25 Mar. 2025

ABSTRACT

One of the important problems in today's world is climate change, and how can we predict this effectively? Climate informatics is a growing technique/field that combines climate science and data science to understand, model, and predict climate change more effectively. It is evolving because of the combination of data analysis, modern visualization tools, and technically advanced computational skills to manage the complicated and problematic data generated by various climate projects. It has been proved that the deterioration of climate started with the advent of industrialization. Since Machine learning has developed in a bigger way in the last few years, the datasets can be used to be predictive for climate information. In machine learning, machines are trained with data to perform specific tasks and deliver accurate results. Fundamentally, Artificial Intelligence is enabling machine learning to analyze with the help of data that is fed into it. We may have noticed the fluctuations in the climate because of global warming, ozone depletion, and various other reasons, the data obtained from the various sources can be impacted positively or negatively by computer science and machine learning. There are multiple datasets available in the market from different sources, here is the introduction of various types of analysis to interpret the data sets.

Keywords: Climate, machine learning, datasets, atmosphere, environment

INTRODUCTION

Climate informatics is a branch that includes research that combines climate science with the addition of statistics, machine learning, and a set of data.

The growing impact of climate change will have a profound effect on the world's social arrangement in the current century and future generations. We can observe the changes in temperature, polar ice, and sea level, which implies that there is a clear need for a better

ISSN: 2455-6939

Volume: 11, Issue: 02 "March-April 2025"

understanding of the climate system. The whole climate data system is a very complicated and complex phenomenon. So far, this system has been wrongly observed and even more wrongly simulated. In this era, there is an ever-growing supply of climate data from satellites and environmental sensors, and the significance of climate data output cannot be interpreted with relatively simple tools. Therefore, a computational approach will be crucial for these analysis challenges. This paper explains how computational approaches are being used to analyze data.

CLIMATE AND INDUSTRIALIZATION

The effect of Global warming has become more pronounced since the 80s-90s. In the developing countries, people are dependent on the natural resources for their daily livelihood. Scientists have experimental confirmation that fossil fuel, coal gas would increase the percentage of CO_2 in the atmosphere and this gas will increase the insulation effect of the earth. As per one of the important sayings, "The warmth of our fields and gardens would pour itself unrequited into space, and the sun would rise upon an island held fast in the iron grip of frost."[1] It is a well-known fact that carbon dioxide would magnify the warming effect because of the other gases present in the atmosphere. If there is a rise in average temperature, it would make the atmosphere absorb more moisture. The moisture is impervious to heat rays, so the warming effect on the atmosphere will increase. People who are residents of cold northern countries or developed countries, are not very much experiencing the heat of the ill atmosphere. Earlier, it was a common assumption that industrialization would cause more pleasant and colder weather, especially in the colder regions of the earth.[2] before industrialization, no one predicted how the use of fossil fuels would affect the world. As industrialization speeded up in the twentieth century, scientists gave the reason that the seas and oceans would absorb excess carbon dioxide. The story behind it is that scientists were very keenly interested in the heat observing properties of gases and their investment in industrialization—these all processes were remaking the earth's atmosphere. [3] Indeed, the climate was not a common subject in the nineteenth century, it was connected to the people who worked in fields like medicine.

A subset of artificial intelligence, machine learning facilitates a system to learn and improve itself using neural networks and deep learning. The scope of this study is to focus on developing and studying statistical algorithms. It can be learned from data and simplified to unseen data. This type of system cannot require explicit instructions and perform tasks by analyzing a big range of data, learning from the insight analysis, and making informed decisions.

The performance of machine learning systems can be improved over time as the system is exposed to a higher range of data. As the system uses more data, the model will get better. Data preprocessing is one of the important steps in machine learning. It includes cleaning the data,

ISSN: 2455-6939

Volume: 11, Issue: 02 "March-April 2025"

handling missing data, and normalizing the data. All these steps improve the data interpretation by the machine learning model correctly.[4]

The last decade has been very enriching for the area of machine learning because it has drawn ideas from different disciplines like artificial intelligence, statistics, and optimization. As a result, this field has grown significantly. Machine learning is being employed in an extensive variety of domains right from Internet applications to scientific problems. These methods are very useful for a variety of things for example predictive modelling as well as exploratory data analysis problems. If one talks about predictive modeling, noticeable developments have been made in linear classification and regression, nonlinear models based on kernels, hierarchical linear models, as well as ensemble methods that combine outputs from different predictors. In the case of exploratory data analysis, clustering and dimensionality reduction have been improved drastically, it also includes nonlinear methods which are used to find out low-dimensional data and manifold structures. Modern machine learning is very much inclined by modern datasets that have scientific, societal, industrial, and commercial applications.

Particularly, the traditional approaches consider tremendously big datasets, running into millions or billions of data points, going up to thousands and thousands or more dimensions; having very intricate statistical dependencies, and violating the independent and identically distribution assumption.

Such properties are very much available in climate datasets, including observations, reanalysis, as well as climate model outputs. These aspects have led to increased emphasis on scalable optimization methods[5], online learning methods[6], and graphical models [7], which can handle large-scale data in high dimensions with statistical dependencies.

1. Understanding climate data

Nowadays climate data are available from various sources, which gives a huge scope for future data analysis and research for machine learning. Here is a brief introduction of varieties of data available. This discussion gives us some interesting problems and suggestions on how they can be used for the predictions.

Here is a brief introduction to every measurement, with a few examples.

1.1 In-Situ Observations In-situ measurements

It refers to fresh measurements of varied climate system properties. It includes temperatures, rainfall, wind- velocity, cloud density, sun radiation, etc., collected from different locations. These locations are chosen either at the surface (for example weather stations) or comprised of

ISSN: 2455-6939

Volume: 11, Issue: 02 "March-April 2025"

atmospheric measurements, sources can be balloons, subsurface ocean data from floats, data from ships, aircraft, and special intensive observing platforms.

The gridded or processed observation is done from a raw in-situ network, the next step is synthesizing those networks into quality-controlled regularly gridded datasets. There are several advantages over the raw data and they are easier to work with, they are more comparable to model output;

1.2 Satellite Retrieval

From 1979 onwards, low-earth orbit and geostationary satellites have been used for global and near-global observations of the Earth's climate. These observations are either recorded by passive radiations emitted directly from the Earth via reflected solar radiation or by active scanning. These recordings are done with the help of lasers or radars. These satellites are mainly operated by U.S. agencies (NOAA, NASA), the European Space Agency, and the Japanese program (JAXA). The data are generally available in near-real-time. There are various levels of data, which range from raw radiances (Level 1), processed data as a function of time (Level 2), and gridded averaged data globally.

1.3 Paleoclimate Proxies

From a vision of the longer term, information about the climate must be extracted from self-styled "proxy" archives, such as ice centers, ocean surfaces, lake deposits, tree-trunk rings, records of pollen movement, caves, or corals. These sources can retain information that is sometimes highly connected to precise climate variables or proceedings [8].

1.4 Reanalysis Products

Several observational data can be integrated to produce the 6-hour forecast. It includes in-situ, remote sensing, etc., which are excellent analyses of the climate status at any one time. However, models improve over time, the time series of weather forecasts can contain trends related only to the change in the model rather than changes in the real world. Thus, many of the weather forecasting groups have undertaken "reanalyses" that use a fixed model to reprocess data from the past to have a consistent view of the real world

1.5 Global Climate Model (GCM) Output

These are physics-based simulations of the climate system, adding components for the atmosphere, ocean, sea ice, land surface, vegetation, ice sheets, atmospheric aerosols and chemistry, and carbon cycles. Simulations can either be transient in response to changing boundary conditions (such as hindcasts of the 20th century), or time slices for periods thought to be relatively stable (such as the mid-Holocene 6,000 years ago). There is a possibility of variations in output depending on the

ISSN: 2455-6939

Volume: 11, Issue: 02 "March-April 2025"

initial systems (because of the disordered nature of the weather), the model used, or variations in the forcing fields.

1.6 Regional Climate Model (RCM) Output

Horizontal resolution is required for global models. If more detailing like high resolution is required at regional topography, global models should be reanalyzed.

2. Methodology for applying ML in Climate Prediction

Machine learning in climate prediction is a very fast-growing area. It offers new ways of explaining and predicting complex climate systems.

The methodology is as follows;

2.1 Data Acquisition and Preprocessing:

Diverse Data Sources: Data can be collected from various sources which comprise of

- Observation of satellites.
- Various weather stations.
- signals and floats of oceans.
- model simulations of Climate (e.g., General Circulation Models - GCMs).
- Reanalysis datasets (combining observations and model outputs).

Data Preprocessing: The step involves:

- missing data are handled
- Spatial and temporal interpolation.
- standardization and normalization of Data
- Creating new relevant variables from existing ones: Feature Engineering

2.2 Model Selection and Training:

ML Techniques: Various ML algorithms are employed, including:

- Neural networks (deep learning): Excellent for capturing complex, non-linear relationships.
- Random forests: Effective for handling high-dimensional data.
- Support vector machines (SVMs): Useful for classification and regression tasks.
- Ensemble models: Combining multiple models to improve accuracy.

ISSN: 2455-6939

Volume: 11, Issue: 02 "March-April 2025"

Model Training:

- To train the model on historical climate data, allowing it to learn patterns and relationships.
- To ensure the model's generalizability, techniques like cross-validation prevent overfitting.

2.3 Model Evaluation and Validation:

Evaluation Metrics: Assessment of the Model is done by using metrics like:

- Mean absolute error (MAE).
- Root mean squared error (RMSE).
- Skill scores.

Validation:

- to ensure reliability, models are tested on independent datasets
- to compare model predictions with observed climate data.
- To evaluate the model's ability to forecast events.

Bias Correction:

- To address systematic errors in model simulations to expand accuracy.

2.4 Application in Climate Prediction:

Specific Applications:

- To predict temperature and precipitation patterns.
- To forecast extreme weather events (e.g., hurricanes, droughts).
- To downscale climate model outputs to regional scales.
- To analyze climate variability and change.
- To improve climate model parameterizations.

Key features;

- **Data Quality:** The quality and completeness of the input data explain the accuracy of ML predictions.
- **Model Interpretability:** To build trust and gain scientific insights, understanding how ML models arrive at their predictions is crucial.
- **Uncertainty Quantification:** there are uncertainties in climate predictions, and ML models should be able to quantify these uncertainties.

ISSN: 2455-6939

Volume: 11, Issue: 02 "March-April 2025"

- **Computational Resources:** To train complex ML models, significant computational power is required.

CONCLUSION

This study gives an insightful glimpse of the informative role of machine learning in predicting climate change. Climate scientists are observing different sets of problems whenever computational techniques are involved.

This study also suggests that implementation should not be considered as the last word, because the study suggests future research directions. It creates a roadmap for multidisciplinary work.

In summary, applying ML in climate prediction involves rigorous data handling, model development, and evaluation. This approach has immense potential to increase our understanding of the climate system and improve our ability to predict further climate conditions.

REFERENCES

[1]. J.F. Donges, Y. Zou, N. Marwan, and J. Kurths. The backbone of the climate network. Eur. Phys. Lett., 87(4): 48007, 2007.

[2]. R. Donner, S. Barbosa, J. Kurths, and N. Marwan. Understanding the earth as a complex system—Recent advances in data analysis and modeling in earth sciences. Eur. Phys. J. Special Topics, 174: 1–9, 2009

[3]. M.N. Evans, A. Kaplan, and M.A. Cane. Pacific sea surface temperature field reconstruction from coral δ18O data using reduced space objective analysis. Paleoceanography, 17: 7, 10.1029/2000PA000590, 2002.

[4]. T. Hastie, R. Tibshirani, and J. Friedman. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer, New York, 2001

[5]. S. Sra, S. Nowozin, and S. Wright. Optimization for Machine Learning. MIT Press, 2011

[6]. N. Cesa-Bianchi and G. Lugosi. Prediction, Learning and Games. Cambridge (UK) and New York: Cambridge University Press, 2006.

[7]. D. Koller and N. Friedman. Probabilistic Graphical Models: Principles and Techniques. MIT Press, Cambridge, MA, 2009.

[8]. P.D. Jones, K.R. Briffa, T.J. Osborn et al. High-resolution palaeoclimatology of the last millennium: A review of current status and future prospects. The Holocene, 19: 3–49, 2009

