Universidad Politécnica de Madrid

Escuela Técnica Superior de Ingenieros Informáticos

Data Science Seminars

Atmospheric Science Modelling Systems

Authors:

José Antonio Ruiz Heredia

Teacher:

Roberto San José

Date:
May 30, 2025

Contents

1 Introduction

Atmospheric modeling systems are crucial for measuring and predicting climate and weather phenomena over space and time. These models help scientists understand atmospheric processes but require significant computational power, traditionally provided by supercomputers.

Recently, techniques such as big data and cloud computing has offered a flexible, cost-effective approach by delivering scalable resources on demand, simplifying the deployment and management of these complex models. In addition, incorporating data science techniques such as machine learning and statistical modeling significantly improves their accuracy and predictive capabilities [1].

2 Data Sources in Atmospheric Modelling

Modern atmospheric models rely on a vast amount of data inputs collected from diverse and unstructured sources. Satellite remote sensing provides global metrics for variables such as temperature, humidity, aerosol concentration, and greenhouse gases. Ground-based stations offer high-frequency local observations, crucial for efficient model calibration and validation. In addition, tools like radiosondes, aircraft-mounted sensors, and weather balloons play a key role in collecting data for vertical atmospheric profiling.

These diverse data sources present integration challenges due to variations in spatial and temporal resolution, measurement errors, and data formats. Data science offers tools to standardize these datasets, apply quality control, and add missing values to enhance the robustness of the models [2].

3 Modelling Techniques

Atmospheric modeling systems generally fall into two main categories: **physically-based numerical models** and **data-driven models**. Physically-based models simulate atmospheric processes by numerically solving partial differential equations that describe fluid dynamics, thermodynamics, radiation, and chemical reactions. Examples of such models include the *Weather Research and Forecasting (WRF)* model and the *Global Forecast System (GFS)* [3, 4].

These models are grounded in fundamental physical laws, such as the Navier-Stokes equations and the conservation of mass, momentum, and energy [5]. Surface-atmosphere interactions, known as boundary layer dynamics, are commonly modeled using K-theory or more advanced closure schemes [6].

Physically-based models require detailed input data and substantial computational power to achieve high spatial and temporal resolution. They are typically validated and calibrated using observational datasets, including satellite images, radiosonde profiles, and ground-based weather stations.

In contrast, data-driven approaches leverage statistical and machine learning techniques to learn patterns and relationships directly from historical data, without explicit knowledge of the governing physical equations. Models such as *Artificial Neural Networks (ANN)*, *Random Forests*, and *Support Vector Machines (SVMs)* have been increasingly applied to short-term weather forecasting, air quality prediction, and anomaly detection. Advanced architectures, including *Recurrent*

Neural Networks (RNNs) and *Convolutional Neural Networks (CNNs)*, are particularly effective in capturing spatiotemporal dependencies. These models are computationally efficient and offer rapid inference, making them suitable for real-time forecasting applications [7].

Hybrid modelling approaches are emerging as a promising direction, combining the strengths of physical models with the flexibility of machine learning. One prominent strategy involves *data assimilation (DA)*, where observational data are integrated with model outputs to improve initial conditions and overall forecast accuracy. Ensemble-based DA systems, such as the *Hurricane Ensemble Data Assimilation System (HEDAS)*, are employed alongside operational forecast models like *HWRF* to assimilate diverse observational inputs—ranging from satellite radiances to airborne Doppler radar data—thus enhancing the predictability of extreme weather events, particularly tropical cyclones [8].

Furthermore, environmental pollution modelling has increasingly relied on coupled systems that integrate mesoscale meteorological models with *Chemical Transport Models (CTMs)*. These coupled models not only forecast atmospheric conditions but also simulate the dispersion and transformation of pollutants. The integration of observational networks, *Numerical Weather Prediction (NWP)* models, and AI-driven pattern recognition contributes to a more comprehensive understanding of atmospheric processes across scales [9].

4 Applications

Atmospheric modelling systems support a wide range of applications, from weather prediction and air quality monitoring to climate change projections. Real-time weather forecasts are critical for disaster prevention and agriculture, while long-term simulations inform policy decisions on emissions and sustainability.

Air quality models, such as *CMAQ* and *GEOS-Chem*, are used to simulate the dispersion of atmospheric pollutants and assess their potential health and environmental impacts. These models are essential tools for urban planning, ensuring regulatory compliance, and advancing environmental justice initiatives. In climate science, coupled models that integrate the atmosphere with oceanic, terrestrial, and cryospheric systems enable researchers to simulate future climate scenarios under a range of emission pathways [10, 11].

5 Challenges

Despite significant advancements, atmospheric science modeling continues to face multiple challenges. High-resolution models demand high computational power, which can limit their practical application, especially for long-term or global-scale simulations. Additionally, the chaotic nature of the atmosphere introduces uncertainties in model predictions. Moreover, simplifications and parameterizations necessary for modeling complex atmospheric processes can further contribute to inaccuracies, potentially impacting decision-making based on model outputs.

Machine learning models offer computational efficiency and the ability to learn from *Bug Data*. However, they often lack transparency and physical interpretability. This limitation raises concerns when these models are applied outside the conditions represented in their training data, leading to difficulties in extrapolation and reduced reliability in unfamiliar scenarios.

Future research is expected to emphasize physics-informed machine learning approaches, which incorporate established scientific principles into data-driven models to enhance their robustness, accuracy, and interpretability. Furthermore, the integration of real-time data streams with edge computing technologies could result in models capable of delivering localized and timely forecasts, particularly in regions with constrained computational resources.

In conclusion, atmospheric science modeling systems are advancing rapidly through the integration of data science techniques. This interdisciplinary collaboration is opening up powerful new ways to understand the atmosphere and address critical global issues such as climate change, extreme weather events, and environmental crisis [12].

References

[1] ScienceDirect (2025). "Atmospheric Model". https://www.sciencedirect.com/topics/earth-and-planetary-sciences/atmospheric-model

[2] National Centers for Environmental Information. "Satellite Atmospheric Products." https://www.ncei.noaa.gov/products/satellite/atmosphere

[3] National Center for Atmospheric Research (2025). "Weather Research and Forecasting Model (WRF)". https://www.mmm.ucar.edu/models/wrf

[4] National Centers for Environmental Information (2025). "Global Forecast System (GFS)". https://www.ncei.noaa.gov/products/weather-climate-models/global-forecast

[5] Jacobson, M. Z. (2005). "Fundamentals of Atmospheric Modeling." Cambridge University Press. ISBN: 0-521-63717-X.

[6] Stull, R. B. (1988). "An Introduction to Boundary Layer Meteorology." Kluwer Academic Publishers. ISBN: 90-277-2769-4.

[7] MDPI (2021). "The Development and Application of Machine Learning in Atmospheric Environment Studies". https://www.mdpi.com/2072-4292/13/23/4839

[8] NOAA Hurricane Research Division. "Data Assimilation." https://www.aoml.noaa.gov/hrd/themes/0/

[9] San Jose, R., & Brebbia, C. A. (1999). "Measuring and Modelling in Environmental Pollution." Computational Mechanics Publications, Southampton. ISBN: 1-85312-461-3.

[10] United States Environmental Protection Agency (2025). "Community Multiscale Air Quality (CMAQ) Modeling System". https://www.epa.gov/cmaq

[11] GEOS-Chem Community (2025). "GEOS-Chem: A Global 3-D Model of Atmospheric Chemistry". https://geoschem.github.io/

[12] Frontiers in Earth Science (2013). "Atmospheric Science: From Classical Meteorology to Earth System Science." https://www.frontiersin.org/journals/earth-science/articles/10.3389/feart.2013.00001/full