

ADVANCED COMPUTATIONAL APPROACHES FOR DATA-DRIVEN AND INTELLIGENT MODELING OF ATMOSPHERIC PROCESSES

Dimitar Makariev, Plamen Trenchev

Space Research and Technology Institute – Bulgarian Academy of Sciences
e-mail: dmakariev@space.bas.bg; ptrenchev@space.bas.bg

Keywords: Earth Observation, Atmospheric Processes, Digital Twins, System Dynamics, Monte Carlo Simulation, Neural Networks, Extreme Learning Machines, Genetic Algorithms, Genetic Programming, Large Language Models

Abstract: A review of contemporary methods for the intelligent modelling and analysis of atmospheric processes has been conducted, with a focus on their application in the study of atmospheric pollution and the extraction of data for simulating atmospheric conditions and scenarios. Earth System Digital Twins (ESDTs) are defined and discussed as dynamic virtual replicas that can simulate and predict the development of atmospheric conditions. Approaches to modelling system dynamics to capture complex interrelationships and processes are presented. Monte Carlo simulations are employed to address uncertainties in 'what if' scenarios and risk assessment. The analysis covers genetic algorithms for model optimization, as well as genetic programming and symbolic regression for extracting interpretable dependencies in atmospheric data. The predictive accuracy and real-time efficiency of neural networks and extreme learning machines are compared. Additionally, the capabilities of large language models (LLMs) and multimodal base models — including those developed by NASA and IBM — for integrating heterogeneous Earth observation data into intelligent atmospheric analysis are examined. Together, these methods support the development of modern approaches to processing and interpreting data from satellite and ground-based observations.

СЪВРЕМЕННИ МЕТОДИ ЗА ИНТЕЛИГЕНТНО МОДЕЛИРАНЕ И АНАЛИЗ НА АТМОСФЕРНИ ПРОЦЕСИ

Димитър Макариев, Пламен Тренчев

Институт за космически изследвания и технологии – Българска академия на науките
e-mail: dmakariev@space.bas.bg; ptrenchev@space.bas.bg

Ключови думи: Наблюдение на Земята, Атмосферни процеси, Дигитални близнаци, Системна динамика, Монте Карло симулация, Невронни мрежи, Машини за екстремно обучение, Генетични алгоритми, Генетично програмиране, Големи езикови модели

Резюме: Извършен е обзор на съвременни методи за интелигентно моделиране и анализ на атмосферни процеси, с фокус върху приложението им в изследването на атмосферното замърсяване и извличането на данни за симулация на атмосферни състояния и сценарии. Дефинирани и дискутирани са Дигиталните близнаци на Земната система (Earth System Digital Twins – ESDTs), разглеждани като динамични виртуални реплики, способни да симулират и прогнозират развитието на атмосферни условия. Представени са подходи за моделиране на динамиката на системи с цел улавяне на сложни взаимовръзки и процеси. За справяне с несигурности в „какво-ако“ сценарии и оценка на рискове са използвани Монте Карло симулации. Анализирани са генетични алгоритми за оптимизация на модели, както и генетично програмиране и символична регресия за извличане на интерпретируеми зависимости в атмосферните данни. Проведен е сравнителен анализ между невронни мрежи и екстремни машини за обучение по отношение на тяхната прогностична точност и ефективност в реално време. Освен това, обсъдени са възможностите на големите езикови модели (LLMs) и мултимодалните основни модели, включително разработените от NASA и IBM, за интегриране на хетерогенни данни от наблюдение на Земята в интелигентния атмосферен анализ. Колективно тези методи подпомагат развитието на съвременни подходи за обработка и интерпретация на данни от сателитни и наземни наблюдения.

Introduction

Atmospheric processes including air quality dynamics, climate variability, and pollutant transport are inherently complex, nonlinear, and subject to significant uncertainties across spatial and temporal scales. Accurately modeling and forecasting these phenomena require advanced computational methodologies that integrate observational data with sophisticated modeling frameworks.

In this review, we synthesize recent progress in intelligent and data-driven approaches for atmospheric modeling, with a particular focus on methods that leverage Earth observation (EO) data from both satellite platforms and ground-based monitoring systems [1]. These methods address critical challenges such as integrating heterogeneous data sources, capturing complex interdependencies within the atmosphere, and dynamically adapting to evolving environmental conditions.

Over the past decade, rapid advancements in machine learning, high-performance computing, and data assimilation have substantially improved the scalability, precision, and interpretability of atmospheric models [2]. Neural networks (NNs), for example, have demonstrated high accuracy in applications such as storm trajectory prediction and cloud classification [3]. Large Language Models (LLMs)—such as NASA's Galileo and IBM's TerraMind now support multimodal EO data interpretation and natural language-based interaction [4, 5]. Meanwhile, Earth System Digital Twins (ESDTs) have emerged as next-generation platforms for multiscale simulation of phenomena such as aerosol transport, planetary boundary layer evolution, and climate impact assessment [1].

In parallel, system dynamics (SD) modeling offers a framework for understanding feedback-driven systems, while Monte Carlo simulations support probabilistic reasoning and uncertainty quantification in risk assessment [6]. For real-time applications, Extreme Learning Machines (ELMs) provide fast and efficient retraining capabilities, making them well-suited for dynamic atmospheric monitoring [7, 8]. Moreover, evolutionary computation techniques such as genetic algorithms (GAs) and genetic programming (GP) are being increasingly used to optimize model architectures and extract interpretable equations from observational data [9, 10].

This review presents a comprehensive overview of these methodologies including digital twins, system dynamics, Monte Carlo simulations, neural networks, ELMs, genetic algorithms, symbolic regression via genetic programming, and large language models highlighting their roles, interactions, and applications in intelligent atmospheric modeling. Conceptual diagrams are included throughout (see Sections 2.1, 2.2, and 7) to illustrate key components and synergies across methods.

Key Methods for Intelligent Atmospheric Modeling

1. Earth System Digital Twins

An **Earth System Digital Twin (ESDT)** is a dynamic and interactive digital representation of the Earth system that integrates observational data and numerical models to replicate past and current states, predict future scenarios, and simulate the impacts of hypothetical conditions [1]. ESDTs offer a unified framework for addressing "What-Now", "What-Next", and "What-If" questions by continuously assimilating data from satellite sensors (e.g., MODIS, AIRS, TROPOMI), ground-based monitoring networks, and model outputs [1].

Key characteristics of ESDTs include:

- **High-fidelity multi-source data integration**, enabling accurate and timely representation of atmospheric phenomena [1].
- **Multiscale modeling capabilities**, supporting applications ranging from global climate simulations to local air quality nowcasting [1].
- **Hybrid modeling approaches**, combining physics-based numerical models with machine learning emulators to enhance computational efficiency without compromising physical realism [1]. In atmospheric applications, ESDTs have proven particularly valuable for modeling:
- **Planetary boundary layer (PBL) dynamics**, which influence pollutant dispersion and cloud formation;
- **Aerosol transport**, essential for understanding transboundary air pollution and radiative forcing;
- **Urban air quality**, especially in complex, fast-changing environments.

Furthermore, ESDTs are increasingly used in support of climate risk assessments, disaster preparedness, and policy analysis, such as evaluating the impacts of wildfire smoke, urban heat island effects, and emission control scenarios. Their modular and federated architectures also allow the interconnection of multiple digital twins (e.g., atmosphere, land, ocean, human systems), enabling more comprehensive Earth system simulations [1].

As such, ESDTs represent a paradigm shift in atmospheric science, bridging observation and simulation while enabling real-time decision support and scenario planning.

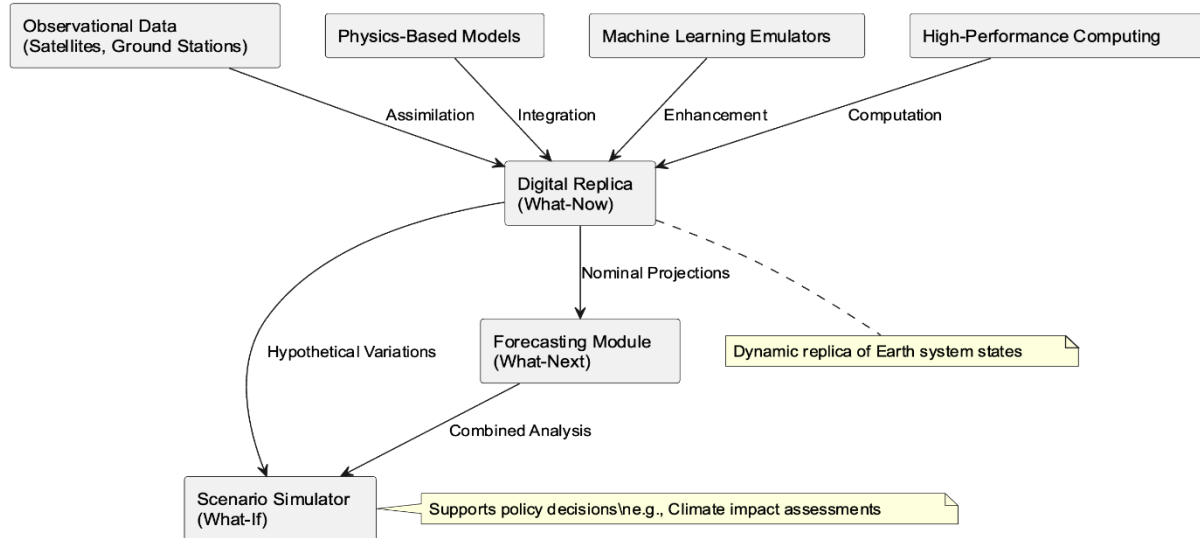


Fig. 1. Component diagram for Earth System Digital Twins, illustrating data assimilation, forecasting, and scenario exploration for atmospheric applications

2. System Dynamics Modeling

System dynamics (SD) modeling is a computational methodology for analyzing and simulating the behavior of complex systems over time through the representation of feedback loops, stocks, flows, and time delays. Originating in control theory and systems thinking, it is particularly well-suited for capturing the nonlinear and interdependent nature of atmospheric and environmental processes [11].

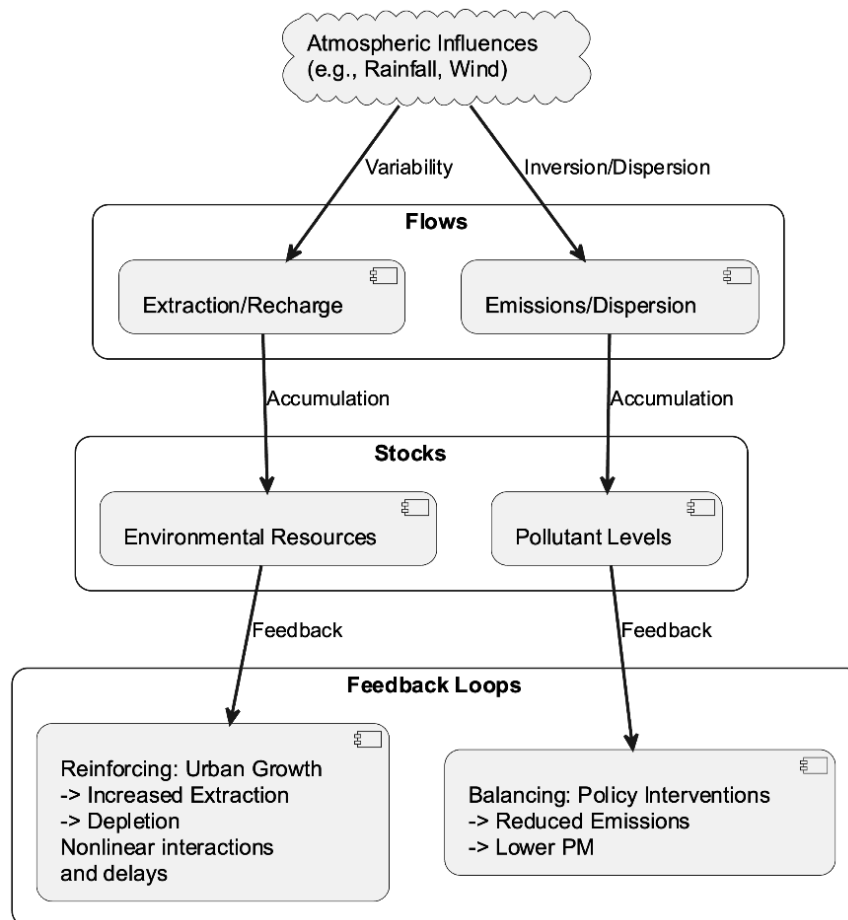


Fig. 2. Causal loop diagram illustrating system dynamics for atmospheric processes, showing stocks, flows, and feedback loops influenced by atmospheric factors

In the context of atmospheric modeling, SD offers a powerful framework for:

- **Characterizing pollutant dispersion**, including accumulation and decay under varying meteorological conditions;
- **Simulating hydrological and biogeochemical cycles**, influenced by atmospheric variables such as temperature, precipitation, and solar radiation;
- **Assessing long-term impacts of interventions**, such as emission control policies or afforestation programs.

A key strength of SD modeling lies in its ability to represent causal feedback loops, which govern the evolution of atmospheric systems. For instance, increased urbanization may amplify local temperatures through the urban heat island effect, which in turn affects energy consumption patterns and air quality dynamics that can be explicitly modeled through system dynamics diagrams [11].

SD models also facilitate scenario analysis by allowing users to vary structural assumptions, such as the intensity of feedback mechanisms or the implementation of policy levers, thus enabling comparative exploration of policy or environmental interventions [11].

By capturing both the structural complexity and temporal evolution of atmospheric systems, system dynamics modeling complements data-driven approaches and is particularly valuable for long-term planning, integrated assessments, and the communication of systemic risks to decision-makers [11].

3. Monte Carlo Simulation

Monte Carlo simulation is a statistical method that uses random sampling and repeated numerical experimentation to model complex systems characterized by uncertainty. In atmospheric science, it plays a crucial role in quantifying variability, propagating uncertainty, and assessing probabilistic outcomes in forecasting and environmental risk analysis [6].

In practical terms, Monte Carlo simulations involve generating large ensembles of scenarios by sampling input variables such as wind speed, temperature, emission rates, or humidity from specified probability distributions. These simulations are then used to estimate statistical properties of model outputs, including means, variances, confidence intervals, and exceedance probabilities.

Key applications in atmospheric modeling include:

- **Uncertainty quantification** in air quality forecasts, where Monte Carlo approaches assess the likelihood of exceeding pollution thresholds (e.g., $\text{PM}_{2.5}$ or ozone concentrations);
- **Probabilistic climate projections**, where future atmospheric states are simulated under varied emission trajectories and model parameterizations;
- **Risk assessment of extreme events**, such as heatwaves, storm surges, or wildfire smoke dispersion, where probabilistic modeling helps estimate impact likelihoods and inform emergency preparedness;
- **Error propagation analysis** in satellite retrievals, improving the reliability of remote sensing products by accounting for uncertainties in instrument calibration, cloud masking, or radiative transfer models.

A notable strength of the Monte Carlo approach is its flexibility and model-agnostic nature, allowing it to be integrated with both physics-based and machine learning models. For example, a neural network trained on atmospheric variables can be subjected to Monte Carlo sampling of its inputs to generate uncertainty-aware forecasts.

While computationally intensive, especially when high-resolution simulations are required, the method remains a cornerstone of probabilistic modeling in atmospheric sciences, providing essential support for robust decision-making under uncertainty [6].

Neural Networks and Extreme Learning Machines

1. Neural Networks for Atmospheric Modeling

Neural Networks (NNs) are foundational tools in intelligent atmospheric modeling due to their ability to learn complex, nonlinear relationships from large datasets. Their architectures particularly Convolutional Neural Networks (CNNs) and Long Short-Term Memory networks (LSTMs) are widely applied across spatial and temporal domains [3].

- **CNNs** are highly effective for analyzing satellite imagery (e.g., from MODIS or Sentinel-5P), enabling tasks like cloud classification, aerosol plume detection, and storm system identification. Their layered structure captures spatial patterns critical for high-resolution atmospheric analysis.
- **LSTMs**, a subclass of Recurrent Neural Networks, are adept at modeling sequential data, such as pollution trends, temperature variation, or wind speed dynamics, enhancing short-term forecasting of meteorological conditions.

- **Physics-Informed Neural Networks (PINNs)** integrate domain-specific equations (e.g., radiative transfer or advection-diffusion equations), improving generalization in data-sparse conditions while maintaining physical realism.
While NNs deliver high predictive accuracy, their training is computationally intensive, especially for deep architectures, which can limit their use in real-time applications.

2. Extreme Learning Machines: Speed and Efficiency

Extreme Learning Machines (ELMs) are single-hidden-layer feedforward networks known for fast, non-iterative training, making them well-suited for time-sensitive atmospheric tasks [7, 8].

Unlike traditional NNs, ELMs randomly initialize input weights and analytically compute output weights, often achieving 100–1000× faster training times [8]. This efficiency enables applications such as:

- **Near real-time air quality forecasting**, using streaming satellite data (e.g., PM_{2.5} estimation from TROPOMI);
- **Anomaly detection** in atmospheric variables (e.g., thermospheric density or visibility degradation) [4];
- **Onboard satellite data processing**, where computational resources are constrained.

Recent ELM variants incorporate kernel methods and sparsity constraints, improving robustness against noise and scalability to large Earth observation datasets. While ELMs may underperform in highly complex feature extraction compared to deep NNs, their simplicity and speed offer significant advantages in operational settings.

3. Comparison of Neural Networks and Extreme Learning Machines

A comparative analysis of neural networks (NNs) and extreme learning machines (ELMs) reveals trade-offs in predictive accuracy and real-time efficiency, particularly relevant for atmospheric forecasting tasks involving satellite data [3, 7, 8]. NNs generally excel in handling highly complex, nonlinear patterns, achieving superior accuracy in benchmarks such as storm trajectory prediction (e.g., RMSE < 0.05). However, their iterative backpropagation training can take hours to days, hindering real-time deployment. In contrast, ELMs offer comparable or occasionally superior accuracy in time-series applications like wave energy forecasting (closely analogous to atmospheric wave propagation), with training speeds 10–100 times faster, enabling efficient processing of streaming Earth observation data. ELMs also demonstrate strengths in simplicity and robustness for large datasets, though NNs may outperform in scenarios requiring deep architectures for intricate feature extraction, such as multispectral image analysis [3].

The following Table 1 summarizes key metrics from recent studies:

Table 1. Comparison of Neural Networks and Extreme Learning Machines

Aspect	Neural Networks (NNs)	Extreme Learning Machines (ELMs)
Predictive Accuracy	High (e.g., outperforms ELMs in deep feature tasks; RMSE ~0.05 for nonlinear dynamics [3])	Comparable or better in fast time-series (e.g., superior in wave forecasting benchmarks [7])
Training Time	Slow (iterative; hours–days for large datasets) [3]	Fast (analytical; seconds–minutes, 1000x speedup over backprop [8])
Real-Time Efficiency	Limited by computation; suitable for offline [3]	High; ideal for streaming/onboard processing [7, 8]
Suitability for Atmospheric Applications	Complex simulations (e.g., radiative transfer) [3]	Operational nowcasting (e.g., aerosol visibility) [7]

This comparison highlights the complementary strengths of NNs and ELMs. NNs are ideal for deep, accuracy-critical modeling, while ELMs offer practical advantages for real-time and resource-constrained applications [3, 7, 8].

Genetic Algorithms

Genetic Algorithms (GAs) are adaptive, population-based optimization techniques inspired by natural selection. They operate by evolving a population of candidate solutions through selection, crossover, and mutation, with the goal of maximizing or minimizing a fitness function [9]. In the context of atmospheric modeling, GAs serve as powerful tools for hyper-parameter optimization, structural tuning, and decision support under uncertainty.

1. Optimization of Atmospheric Models

GAs are particularly valuable for optimizing the parameters of data-driven and hybrid models such as:

- **Neural Networks and Extreme Learning Machines**, where they can adjust learning rates, hidden layer sizes, and activation functions to improve forecast accuracy;
- **System Dynamics simulations**, where they calibrate feedback coefficients and delay parameters to better represent real-world atmospheric interactions.

These optimization tasks often involve high-dimensional, non-convex search spaces, where traditional gradient-based methods may fail or become inefficient. GAs efficiently explore such spaces using stochastic search and maintain population diversity, thereby avoiding local optima.

2. Practical Applications in Atmospheric Science

Genetic Algorithms have been applied in a variety of atmospheric and Earth observation contexts, including:

- **Satellite scheduling** for atmospheric monitoring: optimizing the timing and path of satellite sensors to improve coverage during dynamic events such as hurricanes or dust storms;
- **Air quality modeling**: calibrating dispersion models to match observed $\text{PM}_{2.5}$ levels in urban areas, using both ground station and satellite-derived data;
- **Model coupling**: facilitating integration between data-driven models and physical process models (e.g., coupling NN-based plume predictions with chemical transport models).

By enabling robust parameter tuning and model calibration, GAs enhance both the accuracy and reliability of atmospheric simulations, particularly under variable and uncertain conditions

3. Genetic Programming and Symbolic Regression

Genetic Programming (GP) is an evolutionary algorithm that extends Genetic Algorithms by evolving computer programs or mathematical expressions, typically represented as tree structures [10]. A specialized form of GP Symbolic Regression aims to discover explicit, interpretable equations that best describe the relationships within a given dataset, without predefining the functional form.

4. Interpretability and Model Discovery

Unlike conventional black-box models such as neural networks, GP and symbolic regression produce transparent, human-readable expressions. This makes them particularly valuable in atmospheric science, where physical interpretability is essential for understanding causal mechanisms and integrating findings into existing scientific knowledge or regulatory models.

Key advantages include:

- **Discovery of governing equations**: GP can infer new formulations of physical processes, such as pollutant dispersion, turbulent mixing, or atmospheric decay rates.
- **Parameterization of subgrid processes**: In climate models or weather simulations, GP-derived equations can serve as data-driven alternatives to empirical parameterizations.
- **Model refinement**: Symbolic regression can improve or replace empirical relationships used in satellite retrieval algorithms or air quality models.

5. Applications in Atmospheric Science

Recent applications of GP in the atmospheric domain include:

- Deriving equations for urban pollutant decay based on real-time satellite and ground sensor data;
- Modeling aerosol dynamics and cloud formation processes using symbolic expressions trained on remote sensing outputs;
- Creating hybrid physical-ML models where GP provides analytical corrections to numerical simulations based on observed discrepancies.

In these contexts, GP fosters hypothesis generation, model simplification, and uncertainty reduction, especially when observational data is sparse or noisy [10].

Large Language Models

Large Language Models (LLMs) and their multimodal extensions—such as Vision-Language Models (VLMs) represent a transformative class of AI systems capable of processing and generating natural language, imagery, and other data modalities [2, 4, 5]. In the context of Earth observation and atmospheric science, these foundation models enable natural language querying, semantic interpretation, and generative modeling across heterogeneous geospatial datasets.

1. Capabilities and Roles in Atmospheric Modeling

Recent advances have positioned LLMs as integral components in intelligent atmospheric modeling pipelines. Key functionalities include:

- **Natural language interaction:** LLMs enable intuitive access to complex atmospheric datasets (e.g., “Identify regions with high PM_{2.5} concentrations using Sentinel-5P imagery”), allowing non-experts to extract insights from EO archives.
- **Multimodal data fusion:** Vision-Language Models can jointly process satellite imagery, sensor readings, and textual metadata to support cross-modal reasoning.
- **Generative scenario modeling:** LLMs can simulate plausible atmospheric outcomes (e.g., aerosol dispersion under hypothetical conditions) based on learned relationships across data modalities.
- **Zero-shot and few-shot learning:** These models can generalize to new atmospheric tasks with minimal or no retraining, enabling fast adaptation to emerging events or datasets.

2. Recent Developments and Applications

Several domain-specific LLMs and VLMs have been developed to support atmospheric and environmental applications:

- **IBM TerraMind** is a generative multimodal model trained on satellite, elevation, and geospatial data. Its “*Thinking-in-Modalities*” framework enables pollutant transport simulations and urban air quality assessments based on textual prompts. Dual-scale representations allow it to operate effectively in both zero-shot and few-shot regimes [4].
- **NASA Galileo**, developed in collaboration with the Allen Institute for AI, focuses on real-time EO data analysis. It supports tasks such as anomaly detection, storm tracking, and PM_{2.5} hotspot identification using Sentinel-5P and other sources. As an open-source platform, Galileo is designed for further fine-tuning on atmospheric datasets [5].

3. Scientific and Operational Value

LLMs and VLMs offer substantial benefits for:

- **Enhancing situational awareness** during environmental crises (e.g., wildfires, extreme pollution episodes);
- **Supporting decision-makers** with natural-language summaries and predictive scenario outputs;
- **Augmenting traditional models** by generating hypotheses or refining model outputs with contextual information.

By bridging the gap between technical complexity and human understanding, LLMs significantly expand the accessibility and interpretability of atmospheric science tools.

4. Interrelationships and Combinations of Methods

The intelligent atmospheric modeling approaches discussed throughout this review exhibit strong synergies that, when combined, enhance predictive accuracy, computational efficiency, and practical applicability. The integration of these methods enables hybrid frameworks capable of addressing the multifaceted challenges inherent in atmospheric science.

Key interrelationships include:

- **Genetic Algorithms (GAs)** effectively optimize parameters and architectures within Neural Networks, Extreme Learning Machines, and System Dynamics models, leading to improved forecasting performance and model robustness [9].
- **Monte Carlo Simulations** complement deterministic models such as Digital Twins and System Dynamics by providing rigorous uncertainty quantification, essential for risk assessment and scenario evaluation under environmental variability [6].
- **Large Language Models (LLMs)** serve as integrative platforms, fusing outputs from Digital Twins and machine learning models to facilitate natural language querying, cross-modal data interpretation, and user-friendly interaction [2, 4, 5].
- **Genetic Programming (GP)** generates interpretable symbolic expressions that enhance physics-informed modeling and improve the parameterization of Digital Twins, bridging the gap between data-driven insights and mechanistic understanding [10].
- **Extreme Learning Machines (ELMs)** offer rapid retraining and efficient real-time forecasting, making them well-suited for dynamic integration within Digital Twin frameworks where timely updates are critical [7, 8].

These hybridizations not only capitalize on the individual strengths of each method but also address their limitations, fostering more resilient and adaptable atmospheric models. For instance, optimizing deep neural network structures with genetic algorithms can reduce overfitting and

computational costs, while LLMs enable intuitive access to complex model outputs, broadening the scope of stakeholder engagement [2, 4, 5].

Overall, the collaborative deployment of these intelligent modeling techniques forms a versatile and powerful toolkit for advancing atmospheric science, enabling more accurate predictions and responsive environmental management strategies.

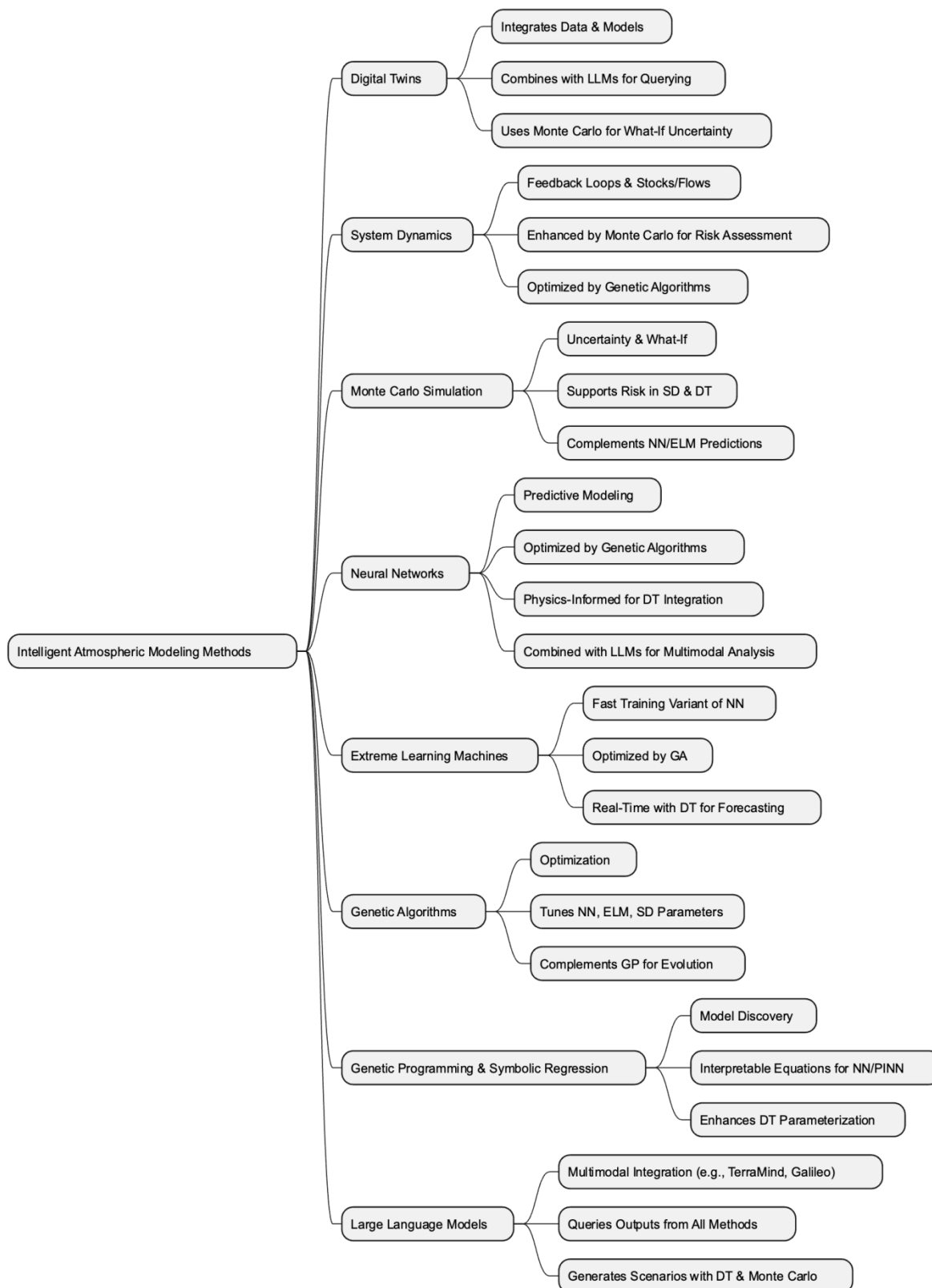


Fig. 3. Hierarchical classification diagram illustrating synergies among intelligent atmospheric modeling methods, emphasizing hybrid approaches

Conclusion

The methods reviewed herein comprising Earth System Digital Twins, system dynamics modeling, Monte Carlo simulations, neural networks, extreme learning machines, genetic algorithms, genetic programming, and large language models—constitute a comprehensive and synergistic framework for intelligent atmospheric modeling. By integrating heterogeneous Earth observation data from satellite and in situ sources, these approaches enable detailed characterization and forecasting of complex atmospheric phenomena across multiple spatial and temporal scales [1, 2, 11].

These methodologies demonstrate distinct strengths in capturing nonlinear dynamics, quantifying uncertainties, and delivering interpretable or near real-time predictive capabilities essential for operational atmospheric monitoring and risk assessment. Earth System Digital Twins facilitate high-fidelity replication of atmospheric states and scenario analysis [1]; neural networks and extreme learning machines provide robust data-driven forecasting and anomaly detection [3, 7, 8, 12]; genetic algorithms and programming optimize model parameters and derive interpretable relationships [9, 10]; and large language models enhance data fusion and user interaction through natural language processing and generative modeling [2, 4, 5].

Despite their considerable potential, challenges remain regarding computational demands, model integration complexity, and the necessity for rigorous validation and calibration to ensure reliability [3, 7, 8]. Balancing predictive accuracy with computational efficiency is critical, particularly for real-time applications.

The integration and hybridization of these methods further amplify their effectiveness. For instance, genetic algorithms can be employed to optimize neural network architectures [9], while large language models can interface with digital twins to facilitate advanced decision support [1, 2, 4, 5]. Such combinations enhance predictive performance and enable timely responses to environmental hazards, including air pollution events, wildfire smoke dispersion, and extreme weather occurrences.

Beyond advancing scientific understanding, these intelligent atmospheric modeling frameworks provide invaluable tools for policy formulation, environmental management, and public health protection [1, 6, 11]. They underpin the assessment of climate change impacts, support disaster risk reduction strategies, and inform regulatory and urban planning decisions.

In summary, the continued advancement, integration, and operational deployment of these sophisticated modeling approaches are imperative for improving our capacity to accurately model and manage atmospheric processes. These efforts will equip researchers, policymakers, and stakeholders with enhanced capabilities to address pressing environmental challenges and foster a more resilient and sustainable future [1, 6, 11].

References:

1. NASA Earth Science Technology Office. (2023). Earth System Digital Twins Workshop Report. https://esto.nasa.gov/files/ESDT_Workshop_Report.pdf
2. Xiao, A., et al. (2024). Foundation Models for Remote Sensing and Earth Observation: A Survey. arXiv:2410.16602
3. Zucker, S., et al. (2024). Physics-informed neural networks for atmospheric radiative transfer. *Journal of Quantitative Spectroscopy and Radiative Transfer*
4. Jakubik, J., et al. (2025). TerraMind: Large-Scale Generative Multimodality for Earth Observation. arXiv:2504.11171
5. Allen Institute for AI. (2025). Introducing Galileo: An Open-Source Model for Earth Observation. NASA Harvest Collaboration
6. Trees, V. J. H., et al. (2025). MONKI: A three-dimensional Monte Carlo simulator. Copernicus Publications
7. Ali, M., et al. (2021). Advanced extreme learning machines vs. deep learning models for peak wave energy period forecasting: A case study in Queensland, Australia. *Renewable Energy*, 178, 350–362
8. Huang, G.-B., et al. (2006). Extreme learning machine: A new learning scheme of feedforward neural networks. *Proceedings of the International Joint Conference on Neural Networks*, 2004
9. Pan, Y., Bélanger S. (2024). Genetic Algorithm for Atmospheric Correction. *Remote Sensing of Environment*
10. McGovern, J., et al. (2025). Potential for equation discovery with AI in the climate sciences. *Earth System Dynamics*
11. Izadbakhsh, H., et al. (2024). A Comprehensive System Dynamics Model for Air Pollution Management in Metropolitan Cities. *Proceedings of the International Conference of the System Dynamics Society*, 1–17. <https://proceedings.systemdynamics.org/2024/papers/P1444.pdf>
12. Zhang, L., et al. (2025). Interpretable Machine Learning for Thermospheric Mass Density. *Space Weather*.