eISSN 2583-6986
IDEALISTIC JOURNAL OF ADVANCED RESEARCH IN PROGRESSIVE SPECTRUMS (IJARPS)
A MONTHLY, OPEN ACCESS, PEER REVIEWED (REFEREED) INTERNATIONAL JOURNAL
Volume 04, Issue 03, March 2025

Evaluating the performance of climate models in simulating extreme weather Events.

Aminu Ado Kaugama[1], Suleiman Waziri Dahiru[1,] Aliyu Abdul-malik Jibo[3] Musbahu Baba Muhammad[4] ,
[1]Department of environmental management and toxicology Federal University Dutse
[2,4]Department of Geography Federal University of Education (T) Bichi, Kano
[3]Department of Hospitatlity and Tourism Management Federal Uniuversity Wukari

Received: 15 March 2025 Accepted & Reviewed: 25 March 2025, Published: 28 March 2025

Abstract

Accurate simulation of extreme weather events is critical for understanding and mitigating the impacts of climate variability and change. This study evaluates the performance of multiple climate models in capturing the intensity, frequency, and spatial distribution of extreme weather phenomena, including heatwaves, heavy precipitation, and tropical storms.

A comprehensive dataset of observed weather records and reanalysis data was used as a benchmark for validation. Statistical metrics, such as bias, root mean square error (RMSE), and correlation coefficients, were employed to assess the models' ability to replicate observed patterns. Results indicate significant variability in model performance across different geographic regions and event types, highlighting the importance of tailored model selection for regional climate impact assessments. Additionally, the study underscores the need for improving parameterizations of atmospheric processes to enhance predictive accuracy. This evaluation provides valuable insights for policymakers and researchers aiming to develop robust adaptation and mitigation strategies.

Keywords: Extreme weather, climate model evaluation, model bias, predictive accuracy, regional climate impacts, atmospheric processes, adaptation strategies.

Introduction

Climate operates as a highly intricate dynamical system, requiring an interdisciplinary approach to comprehend its complexities. It exhibits variability across all spatial and temporal scales, ranging from interannual fluctuations to planetary lifespans, and from localized differences in mountainous regions to continental disparities.

As with any scientific discipline, climate research is heavily reliant on observations and data. There is no universally accepted suite of metrics for evaluating the performance of climate models or determining their skill in predicting future climate change. Lucarini et al. (2007) emphasized that assessing climate models requires a combination of global metrics and process-oriented evaluations. Similarly, Gleckler et al. (2008) proposed a multidimensional approach that compares the spatiotemporal variability of climatic variables with reference datasets. However, they noted that developing a scalar metric to comprehensively summarize model performance remains a complex challenge.

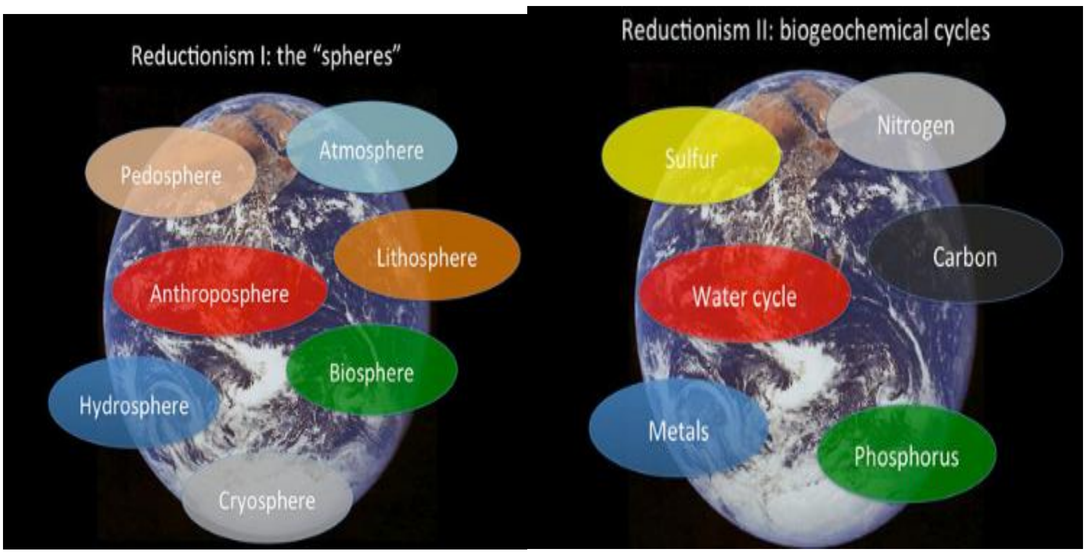
Efforts to address this gap have included combining diagnostic tools and metrics to assess specific climate system features (Eyring et al., 2016, 2020) and testing models' representation of thermodynamic processes in the climate system (Lembo et al., 2019). A scalar metric adhering to mathematical principles, such as those governing Euclidean distance, is particularly desirable. For example, the root-mean-square distance, an L2 metric, satisfies these principles but falls short in fully capturing differences between distribution functions. To address these limitations, Ghil (2015) proposed using the Wasserstein Distance (WD) as a generalized metric in climate science. This approach builds on the foundational work of Dobrushin

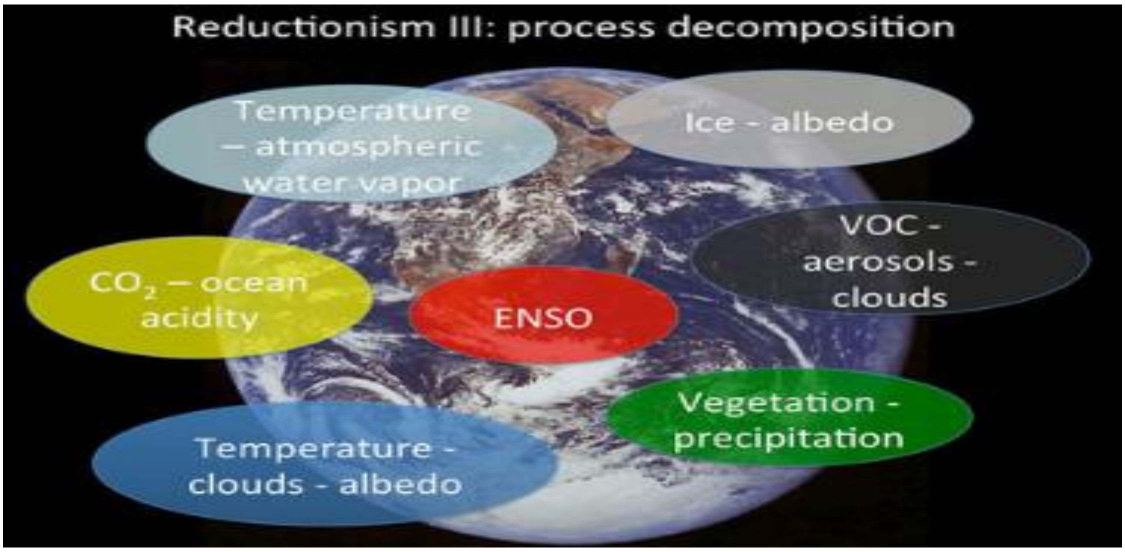
(1970), Kantorovich (2006), and Villani (2009) and seeks to extend the concept of equilibrium climate sensitivity under time-dependent forcings, such as seasonal or anthropogenic changes (Ghil & Lucarini, 2020). Robin et al. (2017) demonstrated the utility of WD by computing differences between snapshot attractors in the Lorenz (1984) model under varying forcings, linking nonautonomous dynamical systems and optimal transport theories. Similarly, Vissio and Lucarini (2018) employed WD to evaluate the effectiveness of stochastic parameterizations in fast-slow systems, while Ning et al. (2014) applied it to quantify errors in variational data assimilation, particularly for advection-diffusion dynamics with systematic parameter errors. Despite its advantages, WD has notable drawbacks.

These include significant computational demands that scale with the number of data points used to construct empirical distributions and the "curse of dimensionality," wherein the data required for accurate analysis increases exponentially with dimensionality. Vissio and Lucarini (2018) addressed the first challenge by introducing data binning techniques, significantly reducing computational overhead. For the second challenge, WD can be computed in reduced phase spaces focusing on key physical variables, making it a flexible option for evaluating climate models. The WD-based approach can complement existing evaluation methods, such as ranking model performance based on the root-mean-square error of ensemble medians (Flato et al., 2013) and weighted ensemble averaging, which adjusts for discrepancies between model outputs and observational data (Knutti et al., 2017).

Today, the availability of extensive, high-resolution, and precise datasets—sourced from satellites and a widespread network of ground-based monitoring stations—has significantly advanced the field. However, the sheer volume of these observational datasets presents challenges in storage, access, and dissemination. Ensuring that this information is effectively managed and made available to scientists, policymakers, and other stakeholders is a critical task.

To address these challenges, international initiatives such as the Global Earth Observation System of Systems (GEOSS), coordinated by the Group on Earth Observations (GEO), have been established. These programs aim to facilitate universal access to Earth observation data, involving over 90 governments and numerous international organizations (GEO, n.d.).





Different reductionist approaches for disentangling the complexity of the climate system

Types of Climate Models

Climate models are computational tools designed to simulate the Earth's climate system. They vary in complexity, scope, and purpose, with each type providing insights into specific aspects of the climate system. Below are the major types of climate models: **General Circulation Models (GCMs)** are comprehensive models that simulate the entire climate system, including atmospheric dynamics, ocean circulation, and land processes. These models are built on fundamental physical principles such as fluid dynamics, thermodynamics, and radiative transfer**. Strengths this** Provides a global perspective of the climate system and are essential for understanding large-scale climate processes. **Limitations** typically operate at coarse spatial resolutions, limiting their ability to represent local-scale phenomena (Randall et al., 2007). **Regional Climate Models (RCMs)** downscale outputs from GCMs to provide finer spatial resolution over specific regions. They are widely used for studying localized climate impacts, such as extreme weather events or regional temperature trends.

Capture regional variability and are critical for climate impact assessments. Dependent on the accuracy of GCM inputs and may introduce their biases during downscaling (Giorgi, 2019). **Earth System Models (ESMs)** extend GCMs by incorporating additional components, such as the carbon cycle, vegetation dynamics, and biogeochemical processes. They are used to explore interactions between the climate and Earth's ecosystems. Suitable for assessing long-term feedbacks and scenarios involving greenhouse gas emissions.Greater computational demands and added uncertainties due to increased complexity (Collins et al., 2011). **Simple Climate Models (SCMs)** are less complex models focusing on specific aspects of the climate system, such as global temperature response to greenhouse gas emissions, Lack detailed spatial and temporal resolution (Meinshausen et al., 2011). **Coupled Atmosphere-Ocean Model** These models couple atmospheric and oceanic components to study interactions between the two, which are crucial for understanding phenomena like El Niño-Southern Oscillation (ENSO). Improved representation of ocean-atmosphere feedbacks. Challenges in accurately modeling small-scale oceanic processes (Delworth et al., 2006). Hybrid models combine physical principles with statistical methods, while purely statistical models rely on historical climate data for predictions. Suitable for short-term forecasting and regions with limited computational resources. Depend heavily on the quality of historical data and may struggle with non-linear climate dynamics (Wilks, 2011).

Criteria for Evaluating Climate Models

Evaluating climate models is essential to ensure their reliability in simulating the complex interactions within the Earth's climate system. Criteria for evaluation focus on how well models replicate observed data, predict future scenarios, and represent key physical processes. Models are evaluated based on their ability to reproduce historical climate patterns, including temperature trends, precipitation, and extreme weather events. Historical simulation is compared against observational data and reanalysis datasets. **Example:** Studies have shown that models capturing key ocean-atmosphere interactions, such as ENSO, are more reliable in long-term projections (Flato et al., 2013).

Biases arise when models systematically overestimate or underestimate climate variables, such as temperatures or precipitation. Reducing these biases is crucial for improving model accuracy. Bias is often quantified using mean absolute error (MAE) or mean bias error (MBE) (Gleckler et al., 2008). **Skill Scores** Skill scores compare model outputs to reference data, providing a quantitative measure of model performance. **Example Brier Score:** Measures the reliability of probabilistic forecasts. **Nash-Sutcliffe Efficiency (NSE):** Indicates how well model predictions match observed data. **Taylor Diagrams:** Visualize correlation, standard deviation, and root mean square error (Taylor, 2001). **Temporal and Spatial Resolution** Higher-resolution models are better at capturing localized weather patterns and extreme events. However, coarse-resolution models can still provide valuable insights at a global scale.

Improved spatial resolution in RCMs has led to more accurate predictions of regional climate phenomena, such as monsoon systems (Kendon et al., 2017). **Representation of Extreme Events** Extreme weather events like heatwaves, floods, and hurricanes test a model's ability to simulate rare but high-impact phenomena. Metrics like the Extreme Climate Index (ECI) and maximum consecutive dry days (CDD) are commonly used (Zhang et al., 2011). **Reproducibility and Robustness** A model's robustness is tested by running it under varying initial conditions or parameters. Consistent outputs under different scenarios enhance confidence in its predictive ability. **Example:** Ensembles of multiple model runs help address uncertainty and ensure robust projections (Tebaldi & Knutti, 2007). **Uncertainty Quantification.** Models must account for uncertainties arising from input data, parameterizations, and future scenarios. Uncertainty quantification often involves using multimodel ensembles and statistical techniques to derive confidence intervals (Knutti et al., 2010).

Regional and Global Performance of Climate Models

Assessing the performance of climate models at both regional and global scales is a critical step in understanding their accuracy and reliability in simulating climate systems. The evaluation process considers how models replicate observed data, reproduce known climate variability, and predict future climate changes under various scenarios.

Global Performance

Global climate models (GCMs) aim to represent the Earth's climate system comprehensively by including atmospheric, oceanic, land surface, and cryosphere components. Evaluating their global performance involves comparing simulated variables, such as temperature, precipitation, and sea surface temperatures, against historical observations and reanalysis datasets (Meehl et al., 2020). For instance, global temperature trends are often used as a benchmark due to their strong correlation with anthropogenic greenhouse gas emissions.

eISSN 2583-6986
IDEALISTIC JOURNAL OF ADVANCED RESEARCH IN PROGRESSIVE SPECTRUMS (IJARPS)
A MONTHLY, OPEN ACCESS, PEER REVIEWED (REFEREED) INTERNATIONAL JOURNAL
Volume 04, Issue 03, March 2025

A critical global evaluation metric is the model's ability to replicate the large-scale circulation of the atmosphere and oceans, such as the Hadley Cell and thermohaline circulation, which are integral to climate dynamics (IPCC, 2021). Biases in simulating these phenomena can propagate errors into regional scales, impacting predictions of extreme events and long-term trends.

Regional Performance

While GCMs provide a broad overview of climate systems, their spatial resolution is often too coarse to capture localized phenomena such as urban heat islands, monsoons, or localized precipitation extremes (Zhou et al., 2021). Regional climate models (RCMs) are therefore used to downscale GCM outputs, offering higher resolution insights into specific areas.

Evaluation of regional performance focuses on the model's ability to reproduce regional climatological features, such as seasonal rainfall patterns in monsoon regions or the diurnal temperature range in arid zones (Wilby et al., 2021). For instance, models simulating the African Sahel's precipitation patterns must account for the interaction between atmospheric dynamics and land surface processes to accurately predict droughts and wet periods (Nicholson, 2013).

Challenges in Regional and Global Evaluations

Despite advancements, discrepancies remain in simulating both regional and global climates. Globally, the representation of cloud processes and aerosols continues to be a significant source of uncertainty. Clouds have a dual role in the climate system, both cooling the surface by reflecting sunlight and trapping heat through the greenhouse effect (Boucher et al., 2013).

Regionally, models often struggle with extreme weather events, such as tropical cyclones or heatwaves, due to their reliance on parameterizations for processes below the model's resolution scale (Wehner et al., 2020). For instance, capturing the intensity and frequency of hurricanes in the Atlantic Ocean requires high-resolution modeling and accurate initial conditions.

Improvements in Model Performance

Efforts to enhance both regional and global model performance include integrating better physical representations, improving data assimilation techniques, and increasing computational power. Ensemble modeling approaches, which combine outputs from multiple models, have also been adopted to reduce uncertainties and improve predictions (Eyring et al., 2019).

Regionally, the development of Earth System Models of Intermediate Complexity (EMICs) has allowed for a more detailed study of localized processes, while advances in artificial intelligence and machine learning are being employed to refine model outputs further (Reichstein et al., 2019).

Conclusion

Assessing the ability of climate models to simulate extreme weather events is essential but complex. It requires evaluating their skill in reproducing observed extremes and forecasting future events under varying climate conditions. While global climate models (GCMs) are effective at capturing broad-scale atmospheric and oceanic patterns, their limited resolution often hampers accurate representation of localized events, highlighting the need for regional climate models (RCMs) for detailed analyses. Advances in evaluation techniques, such as the use of Wasserstein Distance metrics and ensemble modeling, have enhanced model assessments. However, challenges remain, particularly in accurately simulating clouds, aerosols, and extreme phenomena like hurricanes. Ongoing improvements in physical modeling, data integration, and machine learning will be crucial for increasing model reliability and supporting climate resilience efforts.

References-

- Boucher, O., Randall, D., Artaxo, P., et al. (2013). Clouds and Aerosols. In *Climate Change 2013: The Physical Science Basis* (pp. 571–658). Cambridge University Press.
- Eyring, V., Cox, P., Flato, G., Gleckler, P., et al. (2019). Taking climate model evaluation to the next level. *Nature Climate Change, 9*(2), 102–110.
- Intergovernmental Panel on Climate Change (IPCC). (2021). *Sixth Assessment Report: The Physical Science Basis*. Cambridge University Press.
- Meehl, G. A., Covey, C., Taylor, K. E., et al. (2020). An overview of coupled climate model performance. *Bulletin of the American Meteorological Society*, *101*(8), E1337–E1361.
- Nicholson, S. E. (2013). The West African Sahel: A review of recent studies on the rainfall regime and its interannual variability. *ISRN Meteorology*, *2013*, 1–32.
- Reichstein, M., Camps-Valls, G., Stevens, B., et al. (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, *566*(7743), 195–204.
- Wehner, M. F., Arnold, J. R., Knutti, R., et al. (2020). Climate extremes and their representation in climate models. *Nature Reviews Earth & Environment*, *1*(11), 614–626.
- Wilby, R. L., & Dessai, S. (2021). Robust adaptation to climate change. *Weather, Climate, and Society, 13*(4), 877–889.
- Zhou, T., Wu, B., & Wang, B. (2021). Regional climate modeling: Progress, challenges, and prospects. *Advances in Atmospheric Sciences, 38*(5), 750–768.
- Flato, G., et al. (2013). Evaluation of Climate Models. In *IPCC AR5 WG1*.
- Gleckler, P. J., et al. (2008). Performance metrics for climate models. *Journal of Geophysical Research: Atmospheres*, 113(D6).
- Kendon, E. J., et al. (2017). Do convection-permitting regional climate models improve projections of future precipitation change? *Bulletin of the American Meteorological Society*, 98(1), 79-93.
- Knutti, R., et al. (2010). Challenges in combining projections from multiple climate models. *Journal of Climate*, 23(10), 2739-2758.
- Taylor, K. E. (2001). Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research: Atmospheres*, 106(D7), 7183-7192.
- Tebaldi, C., & Knutti, R. (2007). The use of the multi-model ensemble in probabilistic climate projections. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 365(1857), 2053-2075.
- Zhang, X., et al. (2011). Indices for monitoring changes in extremes based on daily temperature and precipitation data. *Wiley Interdisciplinary Reviews: Climate Change*, 2(6), 851-870.
- Collins, W. D., et al. (2011). The Community Climate System Model: A Framework for Collaborative Research. *Journal of Advances in Modeling Earth Systems*, 3(4).
- Delworth, T. L., et al. (2006). GFDL's CM2 Global Coupled Climate Models. Part I: Formulation and Simulation Characteristics. *Journal of Climate*, 19(5), 643–674.
- Giorgi, F. (2019). Thirty years of regional climate modeling: Where are we and where are we going next? *Journal of Geophysical Research: Atmospheres*, 124(10), 5696–5723.

  - Meinshausen, M., et al. (2011). The RCP Greenhouse Gas Concentrations and their Extensions from 1765 to 2300. *Climatic Change*, 109(1), 213–241.
  - Randall, D. A., et al. (2007). Climate Models and Their Evaluation. In *IPCC AR4 WG1*.
  - Wilks, D. S. (2011). *Statistical Methods in the Atmospheric Sciences*. Academic Press.
- Dobrushin, R. L. (1970). Definition of a system of random variables by conditional distributions. *Theory of Probability & Its Applications*, *15*(3), 458–486.
- Eyring, V., et al. (2016). ESMValTool (v1.0) – a community diagnostic and performance metrics tool for routine evaluation of Earth system models in CMIP. *Geoscientific Model Development*, *9*(5), 1747–1802.
- Eyring, V., et al. (2020). Earth system model evaluation tool (ESMValTool) v2.0 - An extended set of large-scale diagnostics for quasi-operational and comprehensive evaluation of Earth system models in CMIP. *Geoscientific Model Development*, *13*(7), 3383–3438.
- Flato, G., et al. (2013). Evaluation of climate models. In *Climate Change 2013: The Physical Science Basis* (pp. 741–866). Cambridge University Press.
- Ghil, M. (2015). A mathematical framework for climate sensitivity and variability. *Earth and Space Science*, *2*(10), 300–325.
- Ghil, M., & Lucarini, V. (2020). The physics of climate variability and climate change. *Reviews of Modern Physics*, *92*(3), 035002.
- Gleckler, P. J., et al. (2008). Performance metrics for climate models. *Journal of Geophysical Research: Atmospheres*, *113*(D6), D06104.
- Kantorovich, L. V. (2006). On the translocation of masses. *Journal of Mathematical Sciences*, *133*(4), 1381–1385.
- Knutti, R., et al. (2017). Beyond mean climate models: Quantifying uncertainty in projections of future climate. *Nature Climate Change*, *7*(7), 442–445.
- Lembo, V., et al. (2019). Toward multi-metric climate model evaluation: Application of a process-oriented approach. *Geophysical Research Letters*, *46*(19), 11035–11044.
- Lucarini, V., et al. (2007). Validation and tuning of climate models using statistical mechanics. *Quarterly Journal of the Royal Meteorological Society*, *133*(626), 929–947.
- Ning, L., et al. (2014). Application of the Wasserstein distance in data assimilation: An advection-diffusion example. *Quarterly Journal of the Royal Meteorological Society*, *140*(680), 1453–1463.
- Robin, G., et al. (2017). Nonautonomous attractors in climate dynamics: The Lorenz model under time-dependent forcing. *Physica D: Nonlinear Phenomena*, *349*, 63–77.
- Villani, C. (2009). *Optimal Transport: Old and New*. Springer.
- Vissio, G., & Lucarini, V. (2018). Evaluating stochastic parametrizations in climate models using the Wasserstein distance. *Physical Review E*, *98*(1), 012140.

Group on Earth Observations (GEO). (n.d.). *Global Earth Observation System of Systems (GEOSS)*. Retrieved from http://www.earthobservations.org/index.shtml.