DI

TN

MI

BW

MP

ML

CI

FS

MACHINE LEARNING FOR ERROR QUANTIFICATION IN SIMULATING THE CLIMATE IMPACTS OF ATMOSPHERIC AEROSOLS

Allocation: Innovation and Exploration/200 Knh
PI: Nicole Riemer[1]

[1]University of Illinois at Urbana–Champaign

EXECUTIVE SUMMARY

Atmospheric aerosol particles influence the large-scale dynamics of the atmosphere and climate because they interact with solar radiation, both directly by scattering and absorbing sunlight and indirectly by forming clouds. Current climate simulations use highly approximate aerosol models with large and unquantified errors.

To address this uncertainty, the PI applied a new ultrahigh-detail spatially resolved and particle-resolved aerosol model, WRF–PartMC, which tracks size and composition information on a per-particle basis. The scientists then used its model output to train machine learning models to predict errors owing to a simplified representation of the aerosol. This allows error predictions merely on the basis of output from the simplified model without running a computationally expensive particle-resolved benchmark case. The scientific impact of this work is the development of a new, innovative method that changes the way aerosol impacts on climate are quantified in current regional and global climate models.

RESEARCH CHALLENGE

One of the largest uncertainties in global climate prediction involves aerosols and their impacts on the radiative budget, a topic of great societal relevance [1]. Aerosol interactions are influenced by both the size and composition of individual particles. Models provide important insights in the study of aerosols but experience a trade-off between the representation of physical detail and spatial resolution. State-of-the-art 3D weather- and climate-scale models focus on large-scale transport but assume a crudely simplified aerosol representation. Commonly applied simplifications assume that all particles look very similar within a particle population, which is typically not representative of reality. In contrast, current-generation box models capture the small-scale features of aerosol physics and chemistry but cannot resolve spatial heterogeneities of aerosol populations. As a result, simulating the evolution of aerosols and predicting their impacts remains a challenge owing to the multiscale nature of the system.

METHODS & CODES

To address these research challenges, the PI has developed the particle-resolved aerosol model WRF–PartMC, which is the first model explicitly resolving the evolution of individual aerosol particles within the grid cells of a state-of-the-art atmospheric fluid/meteorology model. This model makes no simplifying assumption in regard to aerosol composition. Therefore, WRF–PartMC is uniquely suited to fill the role of a benchmark model for simulating atmospheric aerosol composition. The particle composition of hundreds of billions of particles in the atmosphere is represented at a given time, and each particle's composition evolves over time owing to coagulation with other particles and condensation of gas vapors.

WRF–PartMC simulations provide a wealth of data on aerosol mixing state and aerosol aging under different environmental conditions. The data generated give one training sample per grid cell per timestep, consisting of the global model state variables in that grid cell and the computed error for the grid cell. The PI uses these data combined with state-of-the-art machine learning techniques to train models for predicting errors in climate-relevant quantities such as optical properties and cloud-forming abilities, which occur when using simplified aerosol treatments. This will allow researchers to make error predictions merely on the basis of output from the simplified model without running a computationally expensive particle-resolved benchmark case.

RESULTS & IMPACT

The PI conducted several particle-resolved simulations for the domain of northern California. The output will serve as training data for the machine learning portion of the project. These stochastic simulations utilize realistic source-resolved emissions, capable of modeling different emission sectors such as diesel vehicles, gasoline vehicles, and power plants, which all have complex aerosol composition. Each simulation was initialized for June 17, 2010, and simulated 24 hours, utilizing 6,656 cores and 12 wall-clock hours. Simulations consisted of 5,000 computational particles per grid cell with a domain size of 170 × 160 and 40 vertical levels. Each simulation models the complex aerosol dynamics and chemistry for on the order of 5 billion individual particles, where each particle is represented as a vector of masses of 20 aerosol species. Simulations of this size were previously unfeasible. Now that such simulations can be conducted, errors owing to aerosol representation can be quantified.

The mixing state parameter χ, as described in [2], quantifies the extent to which the particle population is internally mixed by examining how complex individual particles are and how similar the particles are within the population. The mixing state parameter varies from 0% to 100%, ranging from all particles containing a single species to 100%, where all particles are identical in composition. Fig. 1 shows relative error in CCN (cloud condensation nuclei) number concentrations as a function of mixing state parameter. Particle populations are projected to fully internally mixed populations, a common representation in other models. When particle populations are further from the model assumption of internally mixed (χ = 100%), models typically overestimate the number of CCN available for cloud formation. This has implications for cloud radiative properties that depend on droplet number and size.

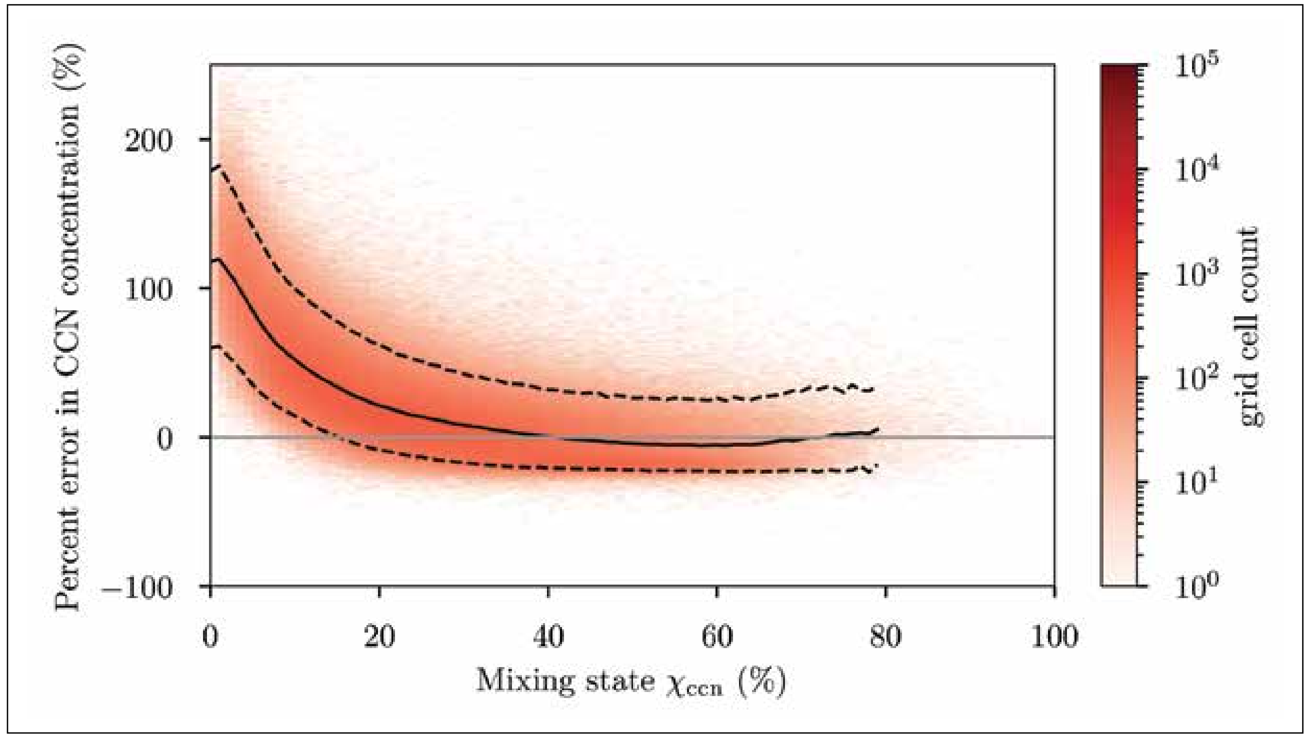

Figure 1: Relative error in cloud condensation number concentrations when assuming an internally mixed aerosol as a function of mixing state parameter χ_{CCN}. The mixing state parameter is 100% for completely internal mixtures and 0% for completely external mixtures.

WHY BLUE WATERS

Access to the computational power and storage space on Blue Waters allows for running simulations to produce data for machine learning. Simulations rely on sufficient memory per core to produce statistically powerful particle populations. Working alongside Blue Waters' staff has alleviated a large portion of the performance issues regarding output by removing tiny writes of data and paying careful attention to what information is output. Reducing the cost of output allows more frequent output, which in turn provides more data for machine learning, as it requires the entire particle state.