Machine Learning towards a Global Parameterisation of Atmospheric New Particle Formation and Growth

Theodoros Christoudias

Mihalis Nicolaou

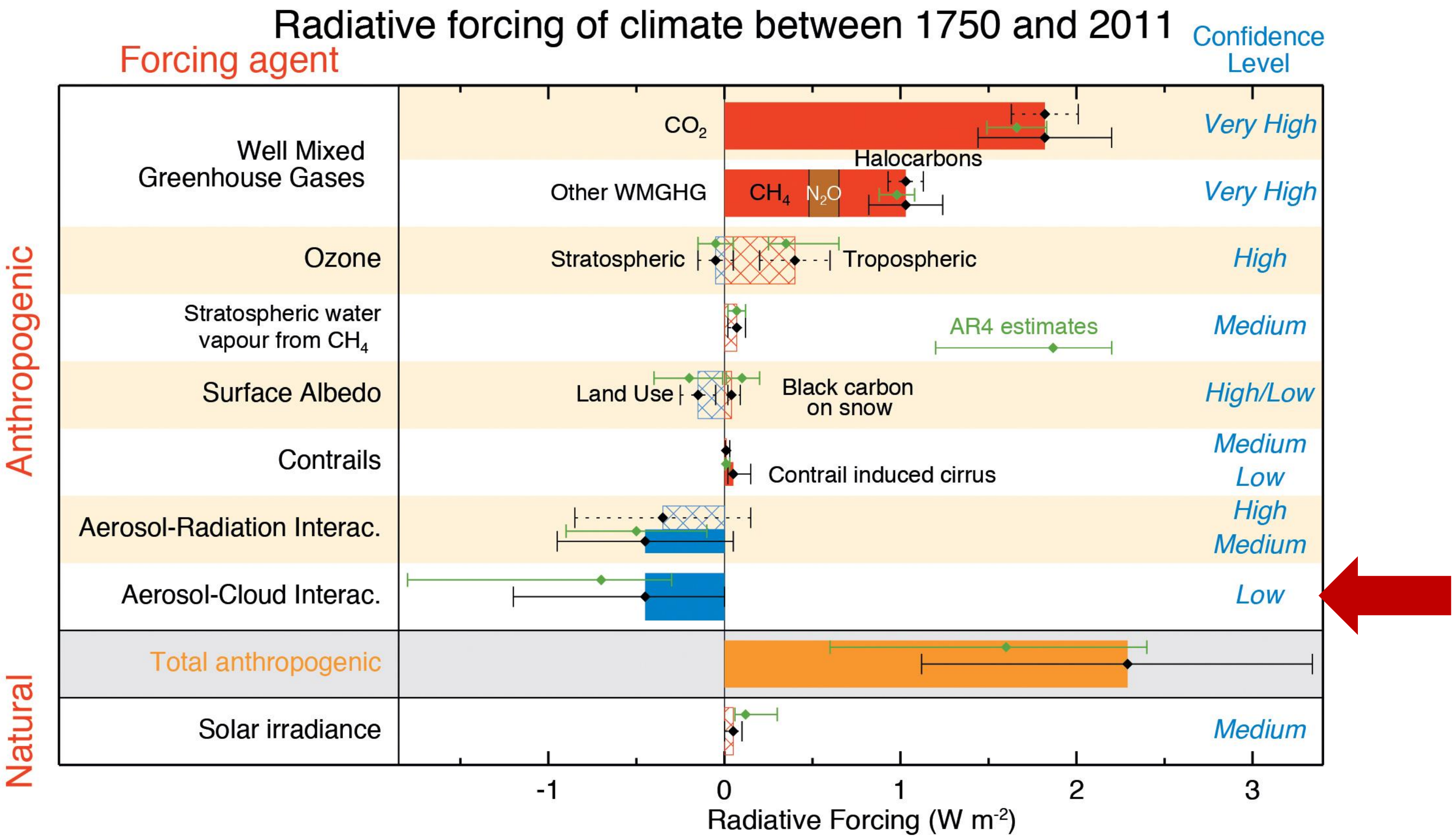
The Cyprus Institute
RESEARCH·TECHNOLOGY·INNOVATION

NeurIPS 2020 Workshop

Earth Changing Energy Budget Uncertainty: Aerosols

Clouds and aerosols continue to **contribute the largest uncertainty** to estimates of the Earth's changing energy budget

Atmospheric aerosols have direct and indirect effects **on Earth's climate and impacts on public health**



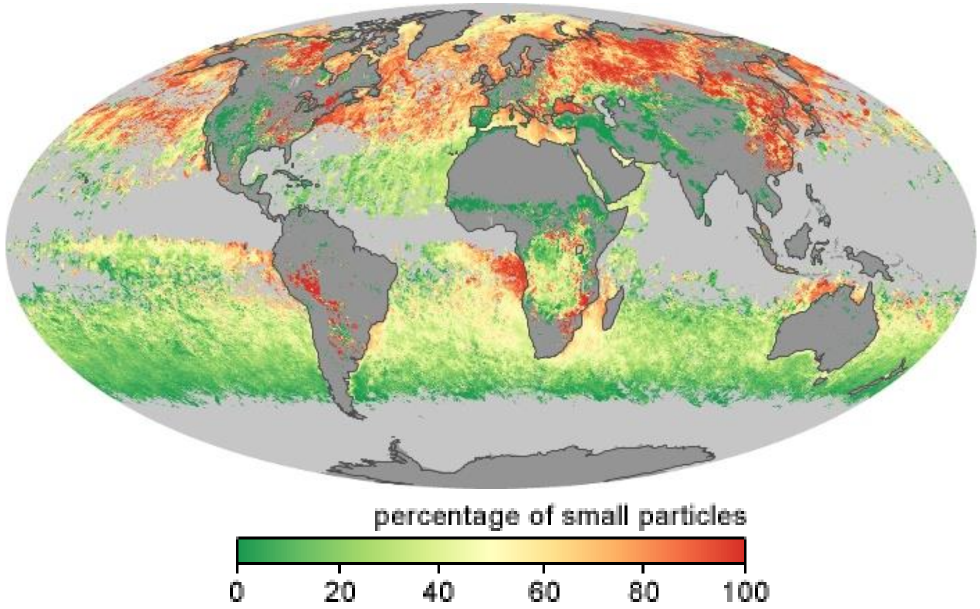
IPCC AR5

New Particle Formation (NPF) and Growth

Aerosols (atmospheric particulate matter) originate from several natural and anthropogenic sources.

New particle formation (NPF), gas-to-particle conversion of atmospheric vapours:

- Major source of aerosols that are **cloud condensation nuclei (CCN)** and further affect the climate
- First step of complex process leading to formation of 40–70% CCN globally
- Observed in boreal forests, coastal, agricultural, and urban areas, including polluted megacities
- Profoundly affects climate, weather, air quality, and human health



earthobservatory.nasa.gov

SoA Models

New Particle Formation (Nucleation):

- Presently, atmospheric models rely on simple parameterisations:
  - Typically polynomial fits to measured NPF rates as a function of vapour concentration (and airborne ions)
  - They are only valid for the environments and conditions that match each observation site

Thermodynamics:

- Most commonly used thermodynamic models (ISORROPIA, EQSAM, E-AIM) use simplifications over the parameter space
- More detailed theoretical models including additional species (e.g. nano-Köhler theory) were also shown to have limitations and are computationally expensive

The Challenge

Despite a wealth of available observations, the NPF parameterisations in regional and global models of the atmosphere are lacking:

> For computational efficiency, process-based models rely on simple parameterisations or fits to disparate single experiments

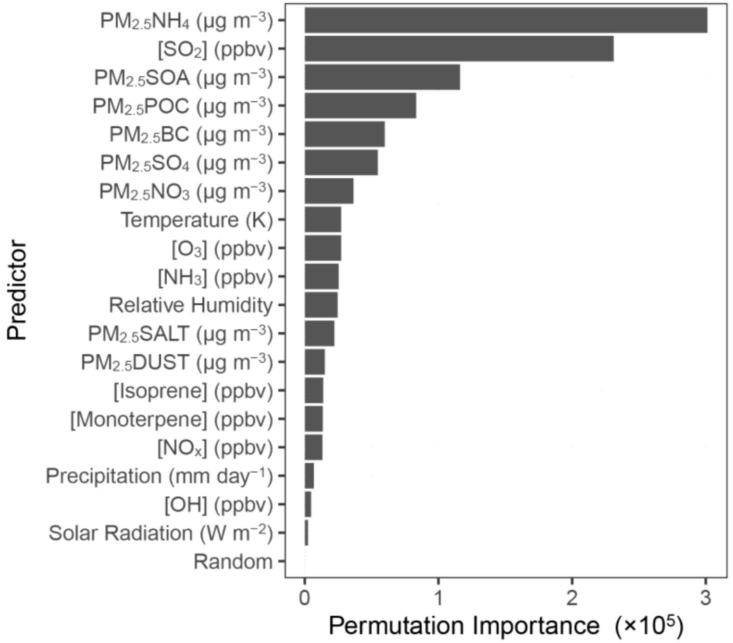
Understanding and **improving modelling of NPF is imperative** to:

1. Reduce uncertainties in climate projections and
2. Tackle urban air quality problems

Machine Learning Proposal: A consistent, NPF parameterisation for atmospheric modelling, combining both predictive capacity throughout the atmosphere ***and*** computational efficiency

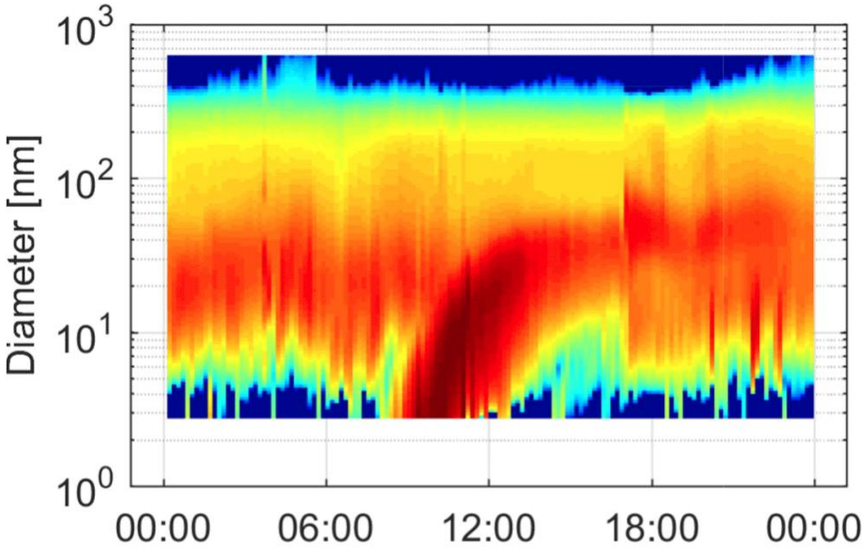
Related Work

The introduction of machine learning methods in this field is limited to:



Using random forest regression of atmospheric model data to a-posteriori derive measured CCN.

Atmos. Chem. Phys., 20, 12853–12869, 2020



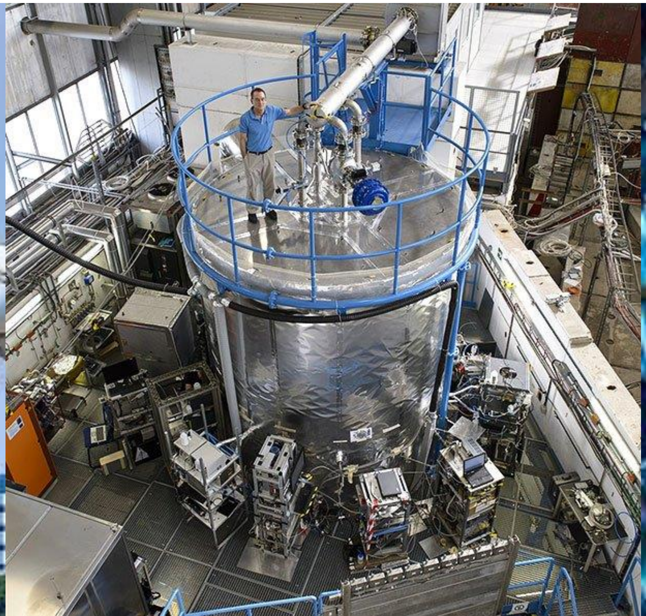
Automating the manual process of observed event identification based only on particle size distributions, with no inference or additional insights.

Atmos. Chem. Phys., 18, 9597–9615, 2018

Data Aggregation

- **Measurement campaigns** of NPF and growth collocated with ambient conditions measurements include in situ **ground station, tower, and aircraft observations**
- Additional multi-component nucleation measurement data are available from **chamber experiments**






Data Sources

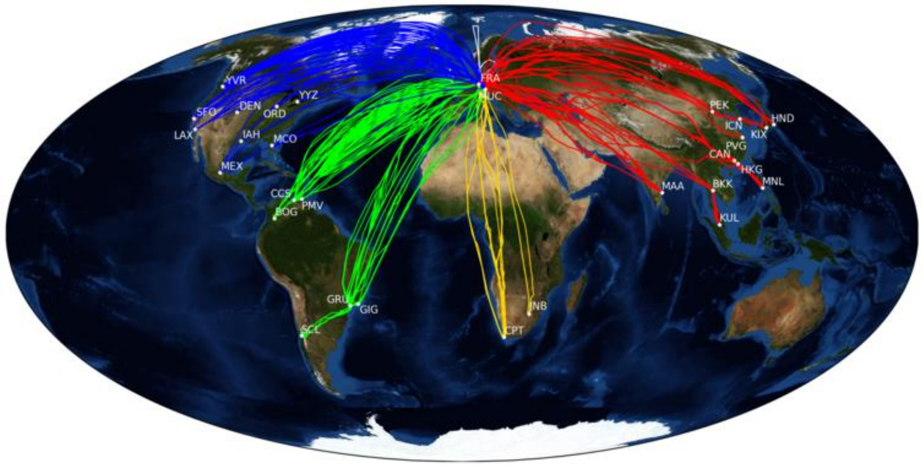
i. **In situ** condensation particle counters (CPCs) for 22 **ground station** locations from the EBAS database (1972–2009)

ii. AGOS **CARIBIC** long distance flights deploying airfreight container with automated scientific apparatus. Using passenger Airbus A340-600 from Lufthansa (more than 550 flights)

iii. **NASA** DC-8 aircraft Atmospheric Tomography Missions (ATom): 0.2-12km altitude, 4 seasons, 4 years

iv. Aerosol, Cloud, Precipitation, and Radiation Interactions and Dynamics of Convective Cloud Systems (ACRIDICON) dataset by German **DLR** High Altitude and Long Range Aircraft (HALO)

v. **Chamber measurements**, in particular the CERN CLOUD experiment



Machine Learning for NPF: Properties

The ML solution should exhibit properties such as:

- Applicability throughout the atmosphere (from lower troposphere to higher levels of the stratosphere),
- Robustness in forecasting under noise and missing/corrupted data,
- Fusion: Ingest data arising from experiments under different conditions (chamber, in-situ, aircraft) that describe the same underlying physical process
- Integrate process-based models with machine learning methods
- Interpretability/explainabilty: Discover insights into the NPF process to improve understanding and guide future capaigns.
- Computational efficiency

Machine Learning for NPF: Proposed Solutions

- Tree-based models ((deep) random forests, decision trees) can provide accurate results while also producing interpretable structures
- Tensor-based methods (e.g., exponential machines) can capture high-order multiplicative interactions between features quickly.
  - Can provide insights in terms of the *interdependencies* of species concentrations and ambient conditions.
- Transfer learning and domain adaptation: transfer knowledge between experiments conducted under different conditions
  - Compensate for covariate shifts, learn common ’shared’ representations
  - ‘Learn’ from process-based models (additional observations, side-information)
- Data imputation methods can be used for robustness under missing measurements

Outlook

- Machine Learning methods can overcome the long-standing challenges in understanding and simulating aerosol nucleation and growth

- Can ingest the data from diverse sources into a unified, global, multi-component parameterisation, valid throughout the atmosphere

- This in turn will **decrease the largest uncertainty in climate projections** and provide a **tool to effectively tackle air quality problems** caused by urbanisation and population growth