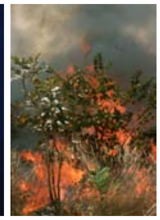
# Uncertainty in Climate Predictions





Douglas Nychka,[1] *National Center for Atmospheric Research (NCAR)*,
Juan M. Restrepo, *University of Arizona*
and Claudia Tebaldi, *Climate Central*.

Carbon dioxide ($CO_2$) and other greenhouse gases, released into the atmosphere from human activities, have altered Earth's natural climate. Naturally produced greenhouse gases have a different chemical signature from the man-made variety. The increasing concentration of $CO_2$ in the atmosphere, correlated to human industrial activities, has been shown through modeling studies and a host of empirical evidence to cause a large fraction of the warming trend recorded in the last decades. Moreover, larger and more rapid future changes are expected in the absence of significant mitigation.

*The uncertainty of climate modeling: all the detailed topography and heterogeneous physical processes in this aerial image from Colorado is reduced to a single grid point in the typical resolution (e.g. 100km×100km) for a global climate model. (NCAR Digital Library)*

Understanding the interactions among different components of the Earth's physical system and the influence of human activities on the environment, has progressed immensely, but it is far from complete. The qualitative assessment of global warming needs to be buttressed by more quantitative and precise estimates of different impacts of climate change on natural, social and economic systems.

*A dramatic cumulonimbus cloud with a rain shaft. This thunder storm in Colorado is "weather". However, based on the sparse vegetation one can infer that the "climate" for this area involves limited rainfall. (NCAR Digital Library)*

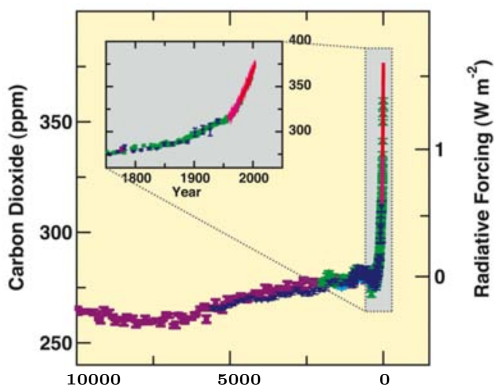
Quantitative assessments of the impacts of future climate change are not straightforward and require new tools in mathematics and statistics providing not only estimates on climate impacts but also companion measures of uncertainty. *The Intergovernmental Panel on Climate Change (IPCC) 2007 report* highlights uncertainty quantification as one of the most pressing research issues in determining our overall ability to propose how human activities can be modified to mitigate our effects on natural climate and how we can adapt to unavoidable changes.

**Climate and weather**

Climate refers to the distribution of the state of the atmosphere, oceans, and biosphere at time scales of decades, centuries, millennia. It can be viewed as a forced/diffusive system, with very complex and long-range interactions between the oceans, the land mass, and the atmosphere. Examples of climatic events are the "Ice Ages", droughts and the current global warming. Weather, on the other hand, refers to the state

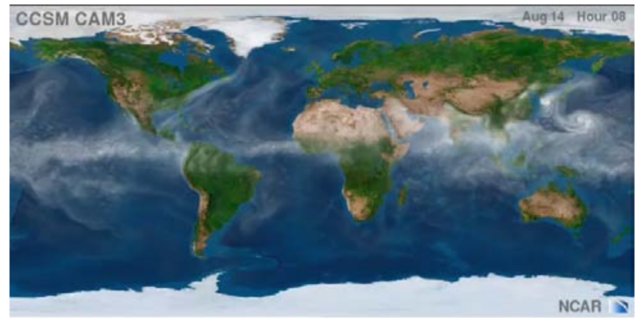
[1]Web links in blue
Photo Credits: E. Nychka and NCAR Digital Library

of the above coupled systems at much shorter spatio-temporal scales. Its behavior can be very sensitive to initial conditions and so is often difficult to predict. Examples of weather events are more familiar to us: storm systems, hurricanes or tornados. The distribution of weather events over a long period of time and at a given location is the climate.




*Atmospheric concentrations of Carbon Dioxide over the last 10,000 years (large panel) and since 1750 (inset panel) showing the rapid increase due to increased consumption of fossil fuels. (See IPCC Fourth Assessment Report)*

Projections of future climate are based on the output of atmosphere/ocean general circulation models and are used to simulate conditions in the future based on projected levels of greenhouse gases. These models are physically based, computer codes that couple the dynamics among the ocean, the atmosphere, sea ice and land along with biogeochemical processes that affect concentrations of $CO_2$. It may come as a surprise that the essential physics for these models is rarely debated, drawing on classical fluid dynamics and thermodynamics. However, due to the finite resolution of the models the exact physical processes must be approximated. Due to their size and complexity, most climate system models can not achieve a resolution much smaller than regions of 100 kilometers square. The result is that the model must account for physical processes that are not resolved, i.e. explicitly simulated, such as cloud formation and complex topography. These unresolved processes still have a strong influence on climate at large scales. The technique of representing unresolved processes is termed a *parameterization* and is an active area of climate research. Not surprisingly, parameterizations are also a key factor in model uncertainty.

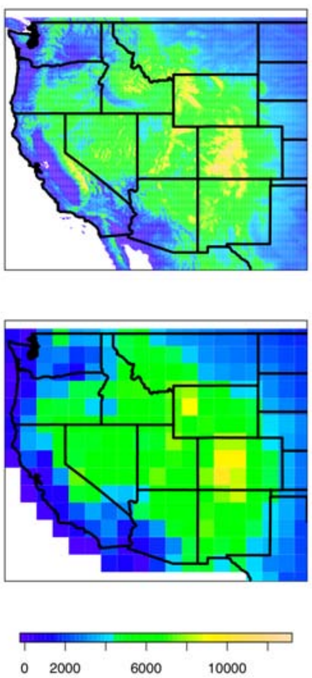



*Snapshot of a high resolution (∼40km) simulation of the global atmosphere using the NCAR Community Atmosphere Model (CAM)* **animation**

**Dynamical systems**

To an applied mathematician, a climate model can be viewed as a discrete dynamical system with some additional variables entering as external forces. The state of the atmosphere, ocean and other components at a time $t$ are represented by a large state vector, say $\boldsymbol{x}_t$, of three dimensional fields and the external "forces", collected together as another vector $\boldsymbol{y}_t$, including the changes in $CO_2$ over time. A mapping $F$ takes the state from time $t$ to $t+1$. It uses both the current state of the climate system, $\boldsymbol{x}_t$ and the values of the external forces, $\boldsymbol{y}_t$ to produce a dynamical result for the next time step. Concisely, $\boldsymbol{x}_{t+1} = F(\boldsymbol{x}_t, \boldsymbol{y}_t)$. Here the individual time steps of the system can be interpreted as generating weather while long term averages or long term probability distributions of the state vector reflect the climate of the system. Suppose that a more accurate physical description of $\boldsymbol{x}_{t+1}$ involves a more detailed or higher resolution state, say $\boldsymbol{x}_t^*$, and accordingly for some $G$, $\boldsymbol{x}_{t+1} = G(\boldsymbol{x}_t^*, \boldsymbol{y}_t)$ gives a better forecast at $t+1$. The parameterization problem from the geosciences is to express this dynamical relationship in terms of the $\boldsymbol{x}_t$ vector. This process is also known as closure and presents interesting new mathematical problems for geophysical systems.

At the same time as climate models have become ever more complex, efforts are made to produce simpler dynamical systems that are more amenable to mathematical analysis. For example, energy balance models are used to gain an understanding of the basic dynamics climate scientists and mathematicians are also involved in proposing ways to capture important unresolved physics in a consistent way: up-scaling and closure techniques are used to produce coarse models that can convey the unresolved physics in a consistent way. Recently, stochastic techniques are being applied to capture (or close) unresolved or poorly understood phenomena in such a way that the models have a statistical behavior that is consistent with observations.



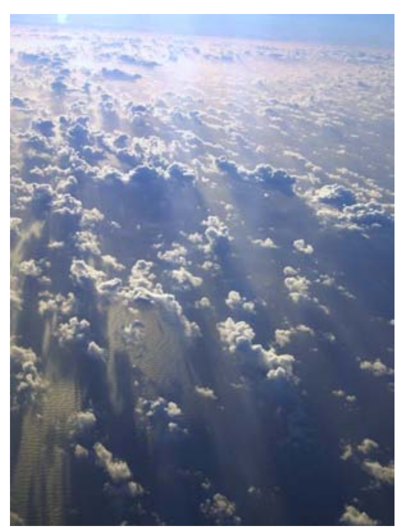

*Elevations for the Western US at 4km and 128km resolution. The coarser resolution shown is a typical scale for climate model experiments and the finer scale is at the limits of historical observational climate data sets.*

**Numerical simulations**

A numerical experiment with a climate model involves specifying the initial state, $\boldsymbol{x}_1$, and the forcing series $\boldsymbol{y}_t$ and then stepping the system forward many time steps. The intriguing aspect of climate modeling is that although the model is formulated as a differential or short term mapping, the climate produced by the model is the long term distribution (including averages) of the state vector. The scientific challenge in climate modeling is to represent the physical processes of the atmosphere and ocean at time scales on the order of minutes but produce realistic simulations of the Earths climate when variables, such as surface temperature or rainfall, are averaged over 20 to 30 years. Typically a state-of-the-art climate model when run on a super computer can simulate several “model” years for each “wall clock” day, thus runs of a model extending 100 years or more are limited by available computer resources. The computational operations increase roughly by a factor of 16 when the grid size is cut in half and this limit places practical constraints on any significant increase in climate model resolution. For example, moving from a 128km grid to a cloud resolving scale of 4km would increase the amount of computations by a factor of more than $10^6$.

**Climate model uncertainties**

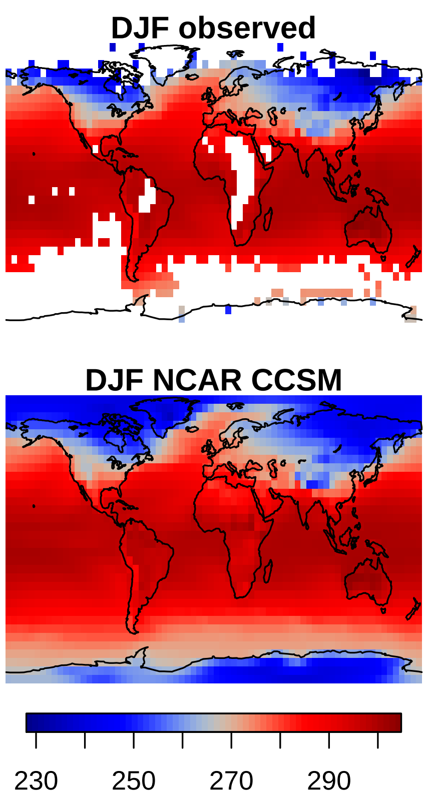
There exist large sources of uncertainty in constructing and applying climate models. These can be classified into initial conditions, boundary conditions, parameterizations and model structure. Initial conditions are often dismissed as a source of uncertainty in climate projections since the results from model runs are usually averaged over decades. However, there still may be some uncertainty introduced for shorter climate predictions due to the initial state of the deeper layers of the oceans and the slow time scale parts of the carbon cycle.



*Cumulus clouds over the South Pacific, (NCAR Digital Library)*

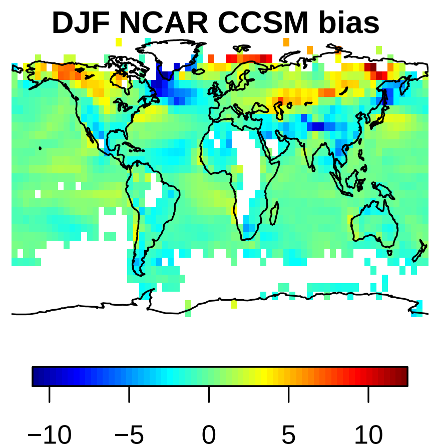
*Uncertain forcings* Boundary condition uncertainty arises because the model experiments, which are otherwise self-contained, rely on inputs (or forcings) that represent external influences. Natural external influences are the solar cycle or volcano eruptions. Man-made influences for climate change studies are primarily greenhouse gas emissions, whose future behavior depend on unpredictable socio-economic and technological factors. Given the uncertainty for future human activities, simulations of future climate are termed *projections* rather than predictions to denote their dependence on a particular scenario of human activity.

*Subgrid scale uncertainty* Parameter and structural uncertainties are intrinsic to numerical climate models. Because of limited computing power, climate processes and their interactions can be represented in computer models only up to a certain spatial and temporal scales (the two being naturally linked to each other).




*Observed mean winter surface temperatures and those simulated from the NCAR Community Climate System Model (CCSM). These data are available as part of CMIP3.*

The influence of processes existing at finer scales than the model grid cells depend on the technique of parametrization or mathematical closure described above.

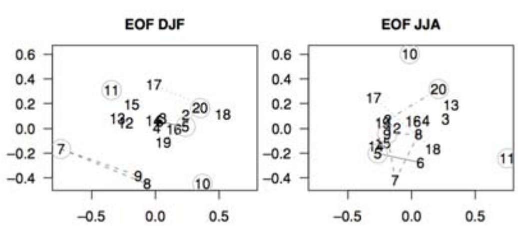



*Climate model bias: the CCSM simulated climate minus the observations.*

For example, the percentage of cloud cover and the type of clouds within a model grid cell are responsible for feedbacks on the large scale quantities, determining how much heat penetrates the lower layers of the atmosphere and how much is reflected back to space. However, the formation of clouds is an extremely complex process that cannot be measured empirically as a function of other large scale observable (and/or representable in a model) quantities. Thus the choice of the parameters controlling the formation of clouds within the model cells has to be dictated by a mixture of expert knowledge and goodness-of-fit arguments.

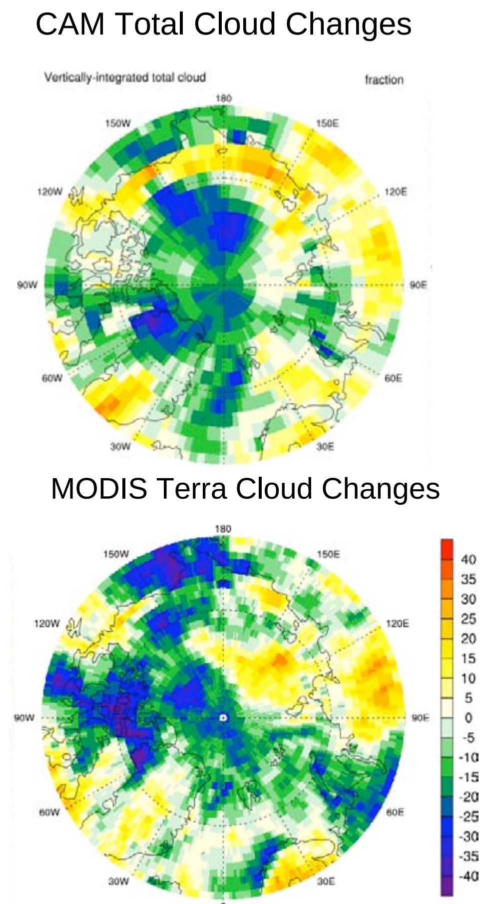
Parameter uncertainties have started to be systematically explored and quantified by so-called perturbed physics ensembles (PPE), sets of simulations with a single model but different choices for various parameters. The design of these experiments and their analysis provide fertile ground for the application of statistical modeling. Because climate model runs are computationally intensive, part of the challenge is to vary several model parameters simultaneously using a limited set of model runs. (See climateprediction.net for a large and particularly innovative PPE based on individuals running parts of an experiment on personal computers.) Bayesian statistical approaches are often preferred because they can also incorporate expert judgment in the prior formulation.

*Uncertainty due to model construction* Structural uncertainty is introduced by scientific choices of model design and development. There are sources of variation across model output that cannot be captured by parameter perturbations, and can introduce a larger degree of inter-model variability than any PPE experiment. The model output collected under the Climate Model Intercomparison Project (CMIP) 3 has provided the basis for all the future projections featured in the last IPCC report. Formal statistical analysis of model output, either coming from multi-model ensembles such as CMIP3 or from PPE experiments is crucial to quantifying the uncertainties caused by models approximations. Unfortunately, the most popular use of these ensembles is to compute simple descriptive statistics: model means, medians and ranges. These may not be appropriate because the given ensemble may not be representative of the super-population of all possible models. For example, models that are currently derived from different climate groups may share a common origin and so inherit subtle structural similarities. Building mathematical tools that can recognize these dependencies among models is a difficult but challenging problem. Closely related to this issue is that some models may be "better" than others in simulating climate leading to methods for weighting models differently when synthesizing their future projections.




*A cluster analysis of 19 climate models based on their spatial biases in simulating winter (DJF) and summer (JJA) mean surface temperature. Points joined by dashed lines are models from the same group (e.g. 7,8,9 are from NASA/Goddard Institute for Space Studies.) The five circled points are a "space-filling" subset of the 19 models. (Courtesy of M. Jun, Texas A&M)*

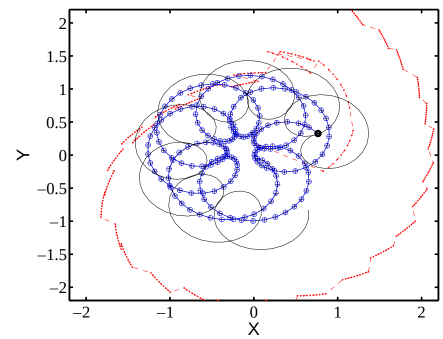
**Data assimilation**
The blending of data and models has been used since the early days of weather forecasting. In the last 25 years estimation and statistical methods have been introduced to weather prediction producing significantly better forecasts by handling the inherent error of models and data in a more consistent way.




*The difference in cloud cover over the Earth from a North pole projection between July 2007 and July 2006. Between these two years there was a large decrease in the sea ice extent. There is a corresponding change in the cloud cover over open ocean in the observations (MODIS) but the model (CAM) results do not produce this change in clouds and suggest some problems with cloud formation over polar oceans. To make this comparison, observational data must be assimilitated into CAM to match the model state to a particular period of observation. (Research and figure (Jen Kay, NCAR)*

These methods, known collectively as data assimilation, are especially useful when the data is sparse, both in space and in time. A new area in climate model development is diagnosing problems in their formulation based on how well they predict weather. This is in contrast to more traditional methods where long term averages or other statistics from a model simulation are compared to the statistics of observations. Specifically, $F$ is used to make a short term prediction and the results are compared to what was actually observed. The differences can then be used to check or modify parameterizations or other components of the model. Another strategy is to assimilate observations with a model over a period of time and then compare the model state to other atmospheric measurements withheld from the assimilation process. The figure given above considering cloud cover over open ocean and its relationship to sea ice is an example of this method.

The blurring of the distinction between climate models and weather forecasting models is a new opportunity for theory on the relationship between short term behavior of $F$ and its long term average properties. Climate and weather are inherently nonlinear, and as a result also have non-Gaussian distributions. Assimilation of data and models with strong nonlinearities and non-Gaussianity pose special mathematical challenges as they differ from the standard linear formulation of the Kalman Filter. Nonlinear/non-Gaussian methods of data assimilation are intended to provide better forecasts that track complex geophysical processes, such as cloud cover or sea ice extent. These can improve the initial conditions and boundary conditions used as a starting point in climate model simulations and aid in model development.




*Predictions of a particle trajectory in a nonlinear/non-Gaussian test problem. Benchmark particle filter (blue) and the Diffusion Kernel Filter (blue circles) and the traditional data assimilation method, the extended Kalman Filter (red). True path in black. (Research and figure, P. Krause and J. M. Restrepo, U. Arizona)*