Testing Climate Models with GPS Radio Occultation Measurements

Stephen S. Leroy and James G. Anderson

Anderson Group, School of Engineering and Applied Sciences
Harvard University, Cambridge, MA 02138, USA
leroy@huarp.harvard.edu

One way of encapsulating the science of climate change as it is relevant to society is in the form of three questions: (1) Is the atmosphere warming? (2) Are humans responsible for any part of this warming? and (3) Can we predict future climate? The first two questions have been settled in large part by the third and fourth scientific assessments of the Intergovernmental Panel on Climate Change (IPCC), albeit the first more solidly than the second. The Fourth IPCC Scientific Assessment shows (IPCC AR4), however, that we are little better at predicting climate change, even when radiative forcing of the climate system is prescribed. Indeed, it has been argued that the IPCC AR4 *underestimated* the uncertainty in predicting the climate with global climate models. When a model's parameters are varied with physically reasonable uncertainties, the spread in climate sensitivity produced by the perturbed physics ensemble is noticeably larger than that produced by the models contributing to the IPCC AR4.

For the practical benefit of society, then, it is mandatory that the climate research community produce a tool that can predict climate with satisfactory accuracy and useful precision. Conventionally, the path to improving climate models' predictive capability has been the improvement by way of increased sophistication of climate models' physical parameterization. Models, because of their poor spatial resolution dictated by computing limitations, are required to approximate important small scale physical processes empirically in order to produce a realistic climate system. Those processes are complex, and hence it is argued that their parameterizations should be complex. In fact, though, the increasing sophistication in parameterizations leads to increased uncertainty in climate prediction. Another argument is made that climate models will be improved by Moore's Law, that the 18-monthly doubling of computing power will eventually lead us to a land of resolved physical processes and the extinction of parameterizations. Uncertainty as we know it will no longer exist. Even if this were to happen, though, the empirical nature of the scientific method has been lost: what observational evidence can we have that any theory of climate, in the form of a model, is any more valid than another? The climate research community is accruing vast amounts of satellite data on the climate system, but how can any of this data be used to lend credence to any multi-decadal prediction of a climate model?

Richard Goody gave us some ideas on how this might be done ("Testing Climate Models: An Approach", in *Bulletin of the American Meteorological Society*, 1998). One can apply the statistical mechanical fluctuation dissipation theorem to the climate:

$$\delta\mathbf{u}(t) = \int_{-\infty}^{t} \mathbf{U}(t - t')\,\mathbf{U}^{-1}(0)\,\delta f(t') \tag{1}$$

where **u** is the state of the climate system, $\mathbf{U}(\tau)$ is the statistical time-lagged covariance of the state vector, and δf is an externally imposed perturbation to the climate system (Leith, *Journal of Climate*, 1975). A burgeoning literature has attempted to exploit this relationship between second-moments of the climate system—the term on the right—to the sensitivity of the climate system, but with meager results. In my opinion, the problem has been that the covariance matrix **U**(τ) is a very large matrix and no one knows how to simplify it, let alone take its inverse at zero lag, so that it still retains the necessary information for predicting long-term change. Instead, one can infer from this equation a significant lesson, despite its practical shortcomings. The only sound quantitative relationship relevant to testing climate models to lend confidence to their predictive power is between lagged covariance of the climate system and trends in the climate system. There is no indication that reproduction of the mean state of the climate system by a climate model should be a cause for confidence in that climate model's predictions of future trends. Little has come of efforts at relating climate sensitivity and lagged covariance of the climate system, so we take the approach of testing climate models by comparing their trends to observed trends of the climate system.

Before a long time series of data is ever obtained, it is well worthwhile to estimate what trends in the climate system are likely to be found with a given data type. Any time series will contain a trend, but of greatest interest are trends in the climate system that underlie the naturally occurring inter-annual variability of the climate system. The first significant detection of an underlying climate trend will be in a sub-space of a data type that is associated with a large signal amplitude and low natural variability. As time progresses, more component sub-spaces will yield significant trends, and those sub-spaces that already yielded detectable trends will do so with increasing confidence. This is the motivation for a simulated optimal detection study. We estimate an emerging signal as $\mathbf{s}\Delta t$ and the fluctuations of natural variability we call $\delta\mathbf{n}$ so that a difference in two climate benchmarks separated by time interval Δt as

$$\delta\mathbf{d} = \alpha\mathbf{s}\Delta t + \delta\mathbf{n} \tag{2}$$

We use α to scale the signal, as it might emerge more rapidly with time than we anticipate. Detection in one corner of the space of the data type is generally strongly correlated with other corners of the space because natural variability is strongly correlated across the space. On inter-annual timescales, for example, the entire tropics fluctuates nearly coherently, a consequence of ENSO events. Likewise, fluctuations at high northern (southern) latitude are correlated over the entire region as a consequence of the Northern (Southern) Annular Mode. On long timescales, it is reasonable to assume these fluctuations are roughly normally distributed, so that we can write

$$\delta\mathbf{n} = \delta\mathbf{d} - \alpha\mathbf{s}\Delta t \sim N(\mathbf{0}, \mathbf{\Sigma}_\nu) \tag{3}$$

with $\mathbf{\Sigma}_\nu$ the covariance matrix of fluctuations $\delta\mathbf{n}$. The solution for the most likely α and its associated uncertainty is

$$\begin{aligned} \alpha &= (\mathbf{s}^T \mathbf{\Sigma}_\nu^{-1} \mathbf{s})^{-1} \mathbf{s}^T \mathbf{\Sigma}_\nu^{-1} \left(\frac{\delta\mathbf{d}}{\Delta t} \right) \\ \delta\alpha &= (\mathbf{s}^T \mathbf{\Sigma}_\nu^{-1} \mathbf{s})^{-1/2} / \Delta t \end{aligned} \tag{4}$$

Because of the computational limitations in estimating $\mathbf{\Sigma}_\nu$, it becomes necessary to retain just a small number of its eigenvalues and eigenvectors in forming its pseudo-inverse.

For radio occultation, it is highly preferable to use downward integrated refractivity, or dry pressure, for purposes of climate monitoring. For one, dry pressure, as opposed to refractivity, is better suited to analyzing trends in dynamical forcing of the atmosphere. Secondly, when discretized on a vertical grid, it should more accurately retain information on precipitable water in the atmosphere. Thirdly, trends in its logarithm can be very easily interpreted as thermal expansion of the atmospheric layers beneath a given level and thus temperature. If refractivity were used instead, it would become difficult to infer temperature and pressure changes. If dry temperature were used instead, all information on pressure is lost.

When detecting trends in a simulated data set, we have found that the first signal to emerge due to increasing well-mixed greenhouse gases should be at 40º to 50º latitude in both hemispheres. The effect is barotropic in the troposphere. If dry pressures were to increase only in these bands, it would represent poleward migration of the zonal wind maximum in both hemispheres, consistent with the universal prediction of climate models that mid-latitude climate regimes will all migrate toward the poles. The second signal to emerge above natural variability should be in the tropics. This signal, too, is barotropic in the troposphere. This represents thermal expansion of the tropical troposphere, the dominant natural mode of which is El Niño/Southern Oscillation (ENSO). See Leroy et al., *Journal of Geophysical Research,* 2006, for details.

Estimating time scales to detection before a data set is obtained is a worthy activity, but it is unclear how this is useful for testing climate models. By itself, it isn't. Instead, one expects that normal linear regression will be performed and statistical significance evaluated. Linear discriminant analysis is ideally suited to this task. In order to test a climate model, then, one would compare the trends produced by transient runs of that climate model to observed trends. Agreement is highly desirable, of course, but disagreement would beg the question of its causes. It is possible that an expert climate modeler would have the intuition to find the source of climate model error responsible, but there is no guarantee of the uniqueness of a solution. For this reason, we have not pursued this specific approach to testing climate models to date. We have instead directed our efforts toward climate prediction.

Certainly, all climate models are "wrong" in that it is always possible to find a comparison between model and data that is unfavorable to a model. This makes it difficult to determine which model is more "right" than the others. To complicate matters, different models can have different strengths: one model might predict trends in ENSO frequency and intensity well but fail at predicting trends in the North Atlantic circulation while another model does exactly the opposite. One cannot say that one model is better or worse than the other. Instead, what we really want to know is how a model can be evaluated for its strengths and weaknesses in predicting trends in a specific variable (or scalar) and how that can be done with data. If we assume that trends are more reliable indicators of climate predictive capability than means, one can form a joint probability distribution of trends in an arbitrary data set $d\mathbf{d}/dt$ and trends in a scalar of interest $d\alpha/dt$. The joint probability distribution $P(d\alpha/dt \mid d\mathbf{d}/dt)$ should cover all sources of uncertainty in climate prediction. What is generally required is a perturbed physics ensemble (PPE) of runs of a climate model subjected to historical radiative forcing. Such a PPE must be incredibly large because the joint probability distribution can have a very large number of degrees of freedom.

In our approach, we instead assume that, while climate models produce a wide range of uncertainty in the sensitivity of the climate system, they nevertheless show broad agreement in the patterns of climate change. Our linear model is

$$\frac{d\mathbf{d}}{dt} = \left(\frac{d\mathbf{d}}{d\alpha}\right)_i \frac{d\alpha}{dt} + \frac{d}{dt}\delta\mathbf{n} \tag{5}$$

in which $(d\mathbf{d}/d\alpha)_i$ is the amount of change $d\mathbf{d}$ in an arbitrary data type for a corresponding change $d\alpha$ in an arbitrary scalar as produced by climate model *i*. For many data types, climate models do yield very similar patterns of change $(d\mathbf{d}/d\alpha)$, but there are differences between models. Those differences must be taken into consideration. Given an ensemble of models, the most likely estimate for the scalar trend $(d\alpha / dt)$ is given by

$$\frac{d\alpha}{dt} = \mathbf{f}^T \frac{d\mathbf{d}}{dt} \qquad (6)$$
$$\mathbf{f} = \mathbf{\Sigma}^{-1}\bar{\mathbf{s}}\left(\bar{\mathbf{s}}^T \mathbf{\Sigma}^{-1}\bar{\mathbf{s}}\right)^{-1}$$

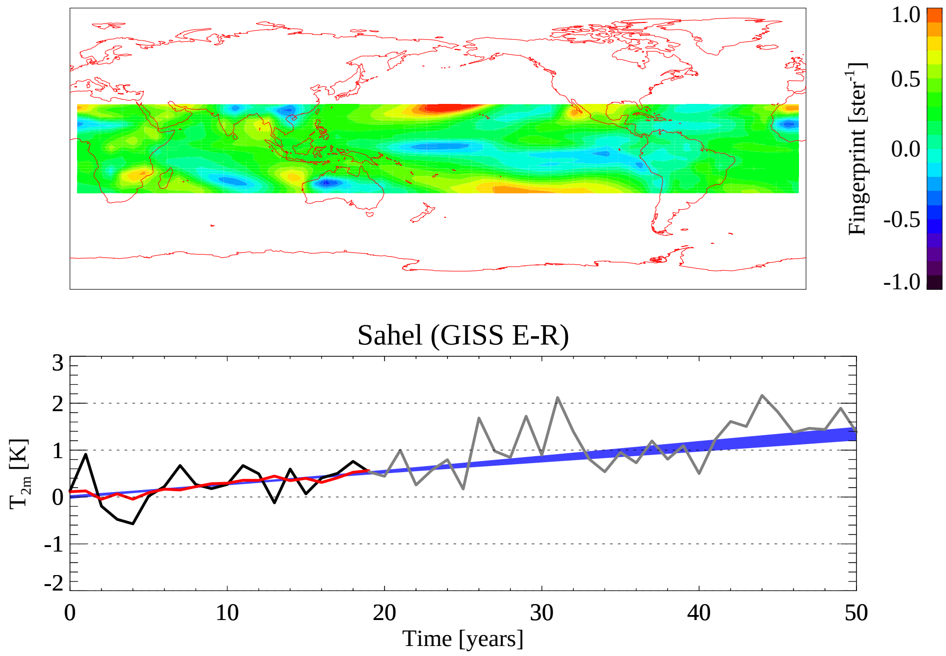

Figure 1. Generalized scalar prediction for the surface air temperature in the Sahel. The top figure shows the contravariant fingerprint when gridded tropical surface air temperature is the data field and Sahel surface air temperature is the target scalar. The model output of the SRES-A1b runs of the IPCC AR4 models were used to "train" the contravariant fingerprint. The lower plot shows the Sahel surface air temperature for the first twenty years in black, the indicator time series in red, a forecast shown as the blue envelope, and the future evolution of Sahel surface air temperature in gray.

Care must be taken in the construction of $\bar{\mathbf{s}}$ and $\mathbf{\Sigma}$. The former is the mean signal form over all models, each producing $\mathbf{s}_i = (d\mathbf{d}/d\alpha)_i$. The latter is the sum of the natural variability covariance and signal uncertainty covariance $\mathbf{\Sigma} = \mathbf{\Sigma}_v + \mathbf{\Sigma}_u$ where

$$\mathbf{\Sigma}_u = \left\langle (\mathbf{s}_i - \bar{\mathbf{s}})(\mathbf{s}_i - \bar{\mathbf{s}})^T \right\rangle \, (7)$$

where the ensemble average is taken over all models of the PPE.

The mathematics should be familiar to those familiar with optimal estimation or variational data assimilation, but the construction of the various terms makes this a uniquely "climate" problem. This derivation makes no demand on the type of data **d**, on the scalar in question *α*, or that there needs to be an obvious relationship between the two. The scalar can be highly localized, and the data can be a gridded global field. For this reason, generalized scalar prediction is an analysis method which allows very large scale phenomena to be useful for attributing trends to global scale phenomena. The IPCC AR4 recognized the lack of a statistical method to apply the high confidence of detection of anthropogenic climate change at large scales to trends at much smaller scales. Generalized scalar prediction fills that gap.

Generalized scalar prediction works by searching the data space for information on the scalar that is generally agreed upon between all model realizations in the PPE and where inter-annual variability is relatively small. If $\bar{\mathbf{s}}$ is the fingerprint for the scalar detection, **f** is the contravariant fingerprint. The contravariant fingerprint shows what physical phenomena contain reliable and sensitive information to pick out the underlying trend of the scalar. In Fig. 1 we show the contravariant fingerprint when we take tropical surface air temperature as the data field and surface air temperature in the Sahel as the target scalar. In a zonal average sense, the higher latitudes in the tropics are positive and the equatorial tropics are negative, showing that the difference in surface air temperature between higher and lower latitudes is a powerful indicator for temperature change in the Sahel. We expect the meridional gradient in surface air temperature to change due to expansion of the Hadley circulation. That Sahel temperature is related to Hadley cell expansion was anticipated subjectively by the IPCC AR4. Generalized scalar prediction bears this out objectively and quantitatively.

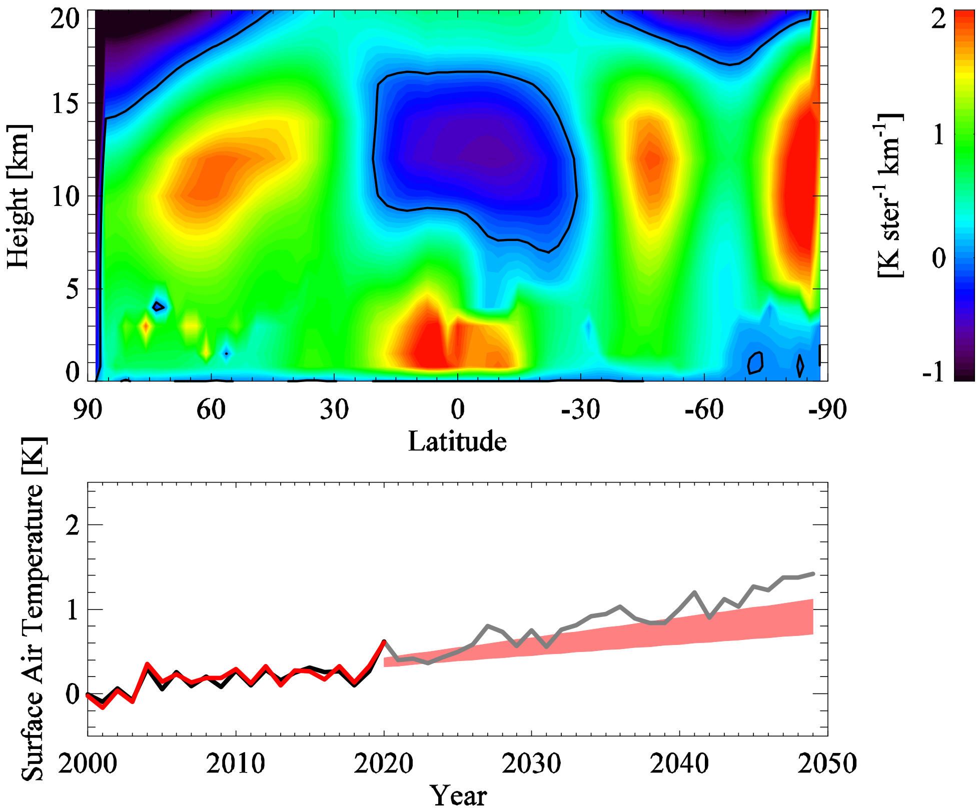

Figure 2. Generalized scalar prediction applied to upper air log-dry pressure with global surface air temperature as the target scalar. The scheme is the same as in Fig. 1 but with the 1 prediction envelope in light red.

In Fig.1 we also show "indicator" time series. The indicators are the results of multiplying the contravariant fingerprint **f** by the data field for each time interval. The slope of the indicator time series is the same as the optimally estimated slope in Eq. 6. The dimensions of the indicators are the same as those of the target scalar, which in this case is surface air temperature in the Sahel. To show the performance of generalized scalar prediction, we compare the indicator time series to the actual surface air time series for the Sahel, at least when applied to the output of a climate model (not used in formulating the contravariant fingerprint). The indicator time series has the same slope as the actual time series of surface air temperature in the Sahel but with most of the inter-annual variability removed. This we call "optimization", and it comes about because the data field extends beyond the domain that defines the scalar and thus can add physical information that isn't available in the Sahel alone. Most importantly, the precision of the prediction is ~0.1 K 30 years into the future. It happens to be accurate as well, as the future evolution simulated by the model bears out.

One problem in optimal fingerprinting techniques is overcome and another problem is introduced by generalized scalar prediction. In optimal fingerprinting, it is necessary to truncate the eigenmodes of the natural variability covariance matrix when forming its inverse because of the existence of unrealistically small eigenvalues in most formulations of the matrix. Depending on how the truncation is performed, it is possible to arrive at very different estimates of scalar trends. Others have pointed out that the problem can be addressed by demanding that post-fit residuals be consistent with the prescription of natural variability, but even this prescription isn't necessarily a good one. The essence of the problem is that optimal fingerprinting searches in subspaces where natural variability is small in comparison with the signal, but often the signal in those subspaces is highly uncertain. The importance of signal shape uncertainty was pointed out over 20 years ago. It "fuzzes" out the subspaces of uncertainty in signal shape and thus prevents optimization in those subspaces. On the other hand, in generalized scalar prediction at present there is no method for determining consistency of the contravariant fingerprint and the actual data.

We apply generalized scalar prediction to the zonal average of the logarithm of dry pressure as might be produced by GPS radio occultation. See Fig. 2. The signal of zonal average log-dry pressure is of expansion of the tropical troposphere with some poleward expansion of climate regimes. In Fig. 2 we show the contravariant fingerprint and indicator time series as in Fig. 1. The target scalar is global average surface air temperature. The contravariant fingerprint subtracts thermal expansion of the tropical troposphere from increased humidity of the lower tropical troposphere and adds in some poleward migration of the maximum wind location in mid-latitudes. GPS radio occultation clearly contains enough information to track global average surface air temperature. There is no optimization, though, possibly because of a one-to-one relationship between the dynamical state of the upper atmosphere and the distribution of surface air temperature.

Generalized scalar prediction makes it possible to relate GPS radio occultation dry pressure to any scalar of the climate system. With a simulation study, we can find out how much GPS RO can contribute to climate forecasting to any variable of interest. We can also find out what physical developments in climate change are most relevant to long-term trends of that variable of interest.