

Radio Occultation Data: Its Utility in NWP and Climate Fingerprinting

Stephen S. Leroy, Yi Huang, and James G. Anderson

*Harvard School of Engineering and Applied Sciences
Anderson Group, 12 Oxford St., Link Building, Cambridge, Massachusetts, USA*

1 Introduction

My job in the Anderson Group at Harvard is centered about support for the Climate Absolute Radiance and Refractivity Observatory (CLARREO), a climate mission, so I must begin my talk with an apology, that I won't say much if anything about numerical weather prediction and instead focus on climate. I think you will still find some interesting and very recognizable concepts. In this talk I will give a brief overview of climate benchmarking, how one can interpret time series of climate benchmarks by scalar prediction, and how an objective method for formulating mission accuracy requirements naturally falls out. Then I will discuss what makes radio occultation a climate benchmark data type, how retrieval is performed on radio occultation data, and the information a time series of radio occultation data can yield on the climate system. Next I will discuss the climate information content of high spectral resolution infrared spectra and what might be gained from considering radio occultation data and infrared spectral data jointly. Finally I will summarize.

The deployment of space-based climate benchmarks was demanded by the recent decadal review of the U.S. National Aeronautics and Space Administration (NASA) and National Oceanographic and Atmospheric Administration by the National Research Council. In particular, it recommended the CLARREO mission:

CLARREO addresses three key societal objectives: 1) the essential responsibility to present and future generations to put in place a benchmark climate record that is global, accurate in perpetuity, tested against independent strategies that reveal systematic errors, and pinned to international standards; 2) the development of an operational climate forecast that is tested and trusted through a disciplined strategy using state-of-the-art observations with mathematically-rigorous techniques to systematically improve those forecasts to establish credibility; and 3) disciplined decision structures that assimilate accurate data and forecasts into intelligible and specific products that promote international commerce as well as societal stability and security.

A bit of translation is in order. "Accurate in perpetuity" is not a requirement that the present and all future generations of NASA engineers fly CLARREO satellites but rather it is a requirement that the data obtained by CLARREO be useful for measuring climate change to all future generations of climate scientists. "International standards" refers to S.I. traceability, a metrological approach to instrument design that assures an unbroken and testable chain of calibration to the international standards that define the units of measurement. "Tested against independent strategies that reveal systematic error" simply requires that climate benchmark instruments have the ability to obtain its own error bars empirically. Lastly, "mathematically rigorous techniques" refers to Bayesian statistics (in which optimal detection methods are implicit) to infer climate trends underlying long time series of data. My particular specialty is the last.

2 Climate benchmarks

S.I. traceability is described in *The International Vocabulary of Basic and General Terms in Metrology* (ISO 1993): “Traceability [is] a property of the result of a measurement or the value of a standard whereby it can be related to stated references, usually national or international standards, through an unbroken chain of comparisons all having stated uncertainties.” In short, “S.I. traceability” is far more than contemporary scientific marketplace jargon; instead it points toward the measurement practice we learned in grade school that observations demand error estimates based on the overall accuracy of our observing apparatus. The only way error estimates can be obtained is by a documented and reproducible chain of comparisons back to the international standard that defines the units of the observations. Two organizations that maintain such standards are the Bureau des Poids et Mésures in Paris and the U.S. National Institutes for Standards and Technology.

Climate benchmarks must be S.I. traceable. The great advantage gained is that a climate benchmark can be used for observing climate change even in the case of a discontinuous time series of data. The community’s experience in constructing “climate data records” from instruments whose calibrations were deemed “stable” has not been good. Take, for example, the records of microwave brightness temperature constructed from measurements of the NOAA satellites’ Microwave Sounding Units (MSU) and of total solar irradiance (TSI). The time series can only be formed by bias-correcting the records of individual instruments so that they match the records of overlapping preceding and succeeding instruments. These efforts have failed because multiple versions of climate data records based on the same data have yielded different long term trends and because even a minor break in the time series of observations renders most of the record useless. With S.I. traceability, the record survives breaks in the time series of observations. Moreover, independent efforts at obtaining long term trends based on S.I. traceable observations will yield the same trends to within empirically determined error.

Others have written on the design of S.I. traceable instrumentation, and since it is not my specialty, I will not do so here. For reference, see [Pollock et al. \(2000\)](#), [Pollock et al. \(2003\)](#). One such design is given by [Dykema and Anderson \(2006\)](#).

Not every known data type can be obtained by instrumentation of S.I. traceable design, but once we have sorted out those that can be, it is interesting to find out what can be learned about the climate system from long term trends in those data types. *Scalar prediction* is one way to determine the information content in S.I. traceable data types. The goal of any investigation of information content in climate monitoring must be the reduction of uncertainty in climate prediction. With credible time series of information-rich data, one ought to be able to obtain improved accuracy and precision in climate projection. (I have used “projection” to refer to prediction aided by intelligent application of data.) A brief description of scalar prediction follows.

3 Scalar prediction

Scalar prediction is two levels of Bayesian inference applied to long term trends. It is closely related to linear multi-pattern regression ([Hasselmann 1997](#); [Allen and Tett 1999](#)) and optimal detection ([Bell 1986](#); [Hasselmann 1993](#); [North et al. 1995](#)). Scalar prediction, or any method used to extract information from long term trends in climate, requires a model for the data that is at least minimally credible and must account for the natural inter-annual variations of climate as a source of error. One central tenant of climate is that it responds approximately linearly to subtle changes in external forcing—the forcing generally a perturbation to the radiative balance of the system not typically associated with a steady-state climate—which cannot be observed directly because the atmosphere-ocean-cryosphere-biosphere system varies from year to year even in the absence of external forcing. Information content studies should seek to minimize the effects of these natural fluctuations and count the residuals as uncertainty

in the inference of “climate trends”.

In the first level of inference, that of optimal detection, one has a long term trend in data $d\mathbf{d}/dt$ which is linearly related to a climate trend $d\alpha/dt$ in some as yet unnamed (and arbitrary) variable α :

$$\frac{d\mathbf{d}}{dt} = \left(\frac{d\mathbf{g}}{d\alpha} \right)_i \frac{d\alpha}{dt} + \frac{d}{dt} d\mathbf{n} \quad (1)$$

in which $\mathbf{g}(\mathbf{x})$ is the model and data operator, a function of the atmosphere-ocean-cryosphere-biosphere state \mathbf{x} , $(d\mathbf{g}/d\alpha)_i$ the total derivative of that model and data operator with respect to arbitrary variable α . Natural variability enters through the random inter-annual fluctuations $d\mathbf{n}$ and the statistically insignificant trends $d(d\mathbf{n})/dt$ to which they give rise. In the first level of inference, the operator $\mathbf{g}(\mathbf{x})$ is considered linear and its derivative \mathbf{x} , $(d\mathbf{g}/d\alpha)_i$ constant and certain while the climate trend $d\alpha/dt$ is completely unknown. The solution for the most likely value of the climate trend $d\alpha/dt$ is

$$\left(\frac{d\alpha}{dt} \right)_{\text{ml}} = (\mathbf{s}_i^T \Sigma_{d\mathbf{n}/dt}^{-1} \mathbf{s}_i)^{-1} \mathbf{s}_i^T \Sigma_{d\mathbf{n}/dt}^{-1} \left(\frac{d\mathbf{d}}{dt} \right) \quad (2)$$

where $\mathbf{s}_i = (d\mathbf{g}/d\alpha)_i$ and $\Sigma_{d\mathbf{n}/dt}$ is the covariance of the random quantity $d(d\mathbf{n})/dt$ as determined from a steady-state simulation of climate. The posterior uncertainty in the determination of $d\alpha/dt$ is

$$\sigma_{d\alpha/dt}^2 = (\mathbf{s}_i^T \Sigma_{d\mathbf{n}/dt}^{-1} \mathbf{s}_i)^{-1}. \quad (3)$$

Equations 1, 2, and 3 are those of optimal detection, the first level of Bayesian inference.

In the second level of inference, one must account for the existence in uncertainty in modeling. The Jacobian $d\mathbf{g}/d\alpha_i$, while minimally credible, is most definitely uncertain. This uncertainty is factored in by weighting each model in an ensemble of climate models according to the quality of its fit to the data. The final result for the most probable trend $(d\alpha/dt)_{\text{mp}}$ and its uncertainty $\sigma_{d\alpha/dt}$ is given by

$$\left(\frac{d\alpha}{dt} \right)_{\text{mp}} = (\bar{\mathbf{s}}^T \Sigma^{-1} \bar{\mathbf{s}})^{-1} \bar{\mathbf{s}}^T \Sigma^{-1} \left(\frac{d\mathbf{d}}{dt} \right) \quad (4)$$

$$\sigma_{d\alpha/dt}^2 = (\bar{\mathbf{s}}^T \Sigma^{-1} \bar{\mathbf{s}})^{-1} \quad (5)$$

$$\Sigma \equiv \Sigma_{d\mathbf{n}/dt} + \left(\frac{d\alpha}{dt} \right)^2 \Sigma_{\delta\mathbf{s}}. \quad (6)$$

The equations are the same as those of optimal detection with the exceptions that $\Sigma_{d\mathbf{n}/dt}$ is replaced by Σ and \mathbf{s}_i by $\bar{\mathbf{s}}$. Both of these new quantities are derived from an ensemble of climate models, the quantity $\bar{\mathbf{s}}$ being the mean \mathbf{s}_i of the ensemble of climate models, and $\Sigma_{\delta\mathbf{s}}$ the covariance of $\delta\mathbf{s} = \mathbf{s}_i - \bar{\mathbf{s}}$ over the ensemble of models. The term $(d\alpha/dt)^2$ appearing in equation 6 is only a prior best guess estimate of the climate trend. Equations 4 through 6 are those of scalar prediction.

There is a clear parallel between the equations of scalar prediction and those of data assimilation in NWP. If one substitutes $d\mathbf{d}/dt$ with the observation increment $\mathbf{d} - \mathbf{y}$, $d\mathbf{g}/d\alpha$ with the observation kernel \mathbf{K} , $d\alpha/dt$ with the analysis increment $\delta\mathbf{x}$ and Σ with the sum of the observation and background error covariances $\mathbf{O} + \mathbf{B}$, one obtains the equations of variational data assimilation.

To see how scalar prediction works, I apply it to the problem of climate trends in Northern Europe. The data space will be a map of Northern Hemisphere surface air temperature, so $d\mathbf{d}/dt$ will be a map of the trend of Northern Hemisphere surface air temperature and $\mathbf{g}(\mathbf{x})$ will be a forward operator that produces maps of surface air temperature from a climate model given the state variable \mathbf{x} . I am interested in the climate trend of surface air temperature in Northern Europe, so I am free to define the completely general variable α as the surface air temperature over the region of Northern Europe. That makes $d\mathbf{g}/d\alpha$ the rate of change of a map of Northern Hemisphere surface air temperature divided by the rate of change of Northern Europe area-averaged surface air temperature. The result is the dimensionless map \mathbf{s}_i and it depends on model i used to simulate it. I use the model output of the World Climate Research

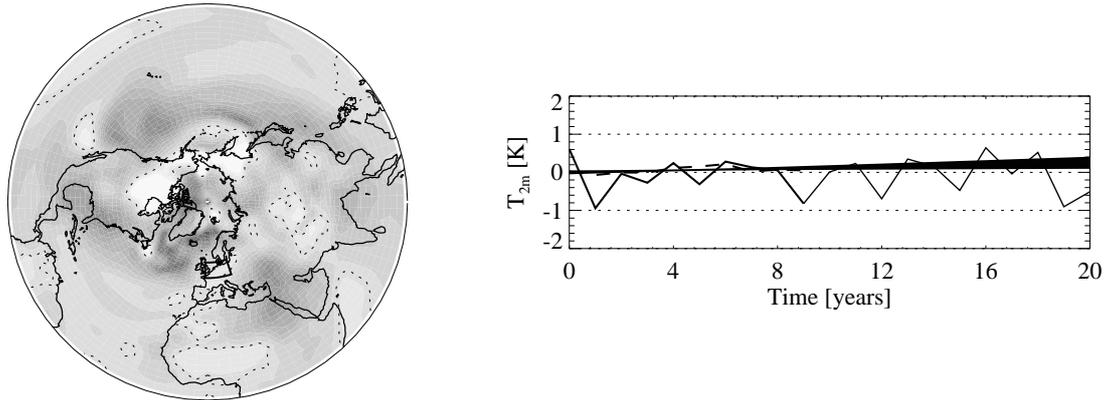


Figure 1: The optimal fingerprint for Northern European surface air temperature climate trend estimation and its inner product with annual average Northern Hemispheric surface air temperature. The plot on the left is the optimal fingerprint \mathbf{f} constructed from the CMIP3 models for determination of climate response of Northern European surface air temperature to SRES A1B forcing given the data space of Northern Hemisphere surface air temperature. The dotted line is the zero-contour. The plot on the right shows the record of Northern Hemisphere surface air temperature (thick curve), the inner product of the optimal fingerprint and annual average Northern Hemisphere surface air temperature $\mathbf{f}^T \mathbf{d}(t)$ for the same period (dashed curve), the one-sigma envelope of the best fit to the latter curve (gray shaded region), and the future evolution of Northern European surface air temperature (thin curve). Scalar prediction is a highly precise estimate of climate trends as illustrated by the narrowness of the shaded region.

Programme's (WCRP's) Coupled Model Intercomparison Project phase 3 (CMIP3) multi-model dataset to generate the many \mathbf{s}_i and then compute $\bar{\mathbf{s}}$ and $\Sigma_{\delta\mathbf{s}}$.

The optimal fingerprint $\mathbf{f} = \Sigma^{-1} \bar{\mathbf{s}} (\bar{\mathbf{s}}^T \Sigma^{-1} \bar{\mathbf{s}})^{-1}$ is the vector/map by which to multiply $d\mathbf{d}/dt$ to obtain the most probable trend $(d\alpha/dt)_{\text{mp}}$ of the Northern Europe temperature trend associated with climate change. We show the optimal fingerprint and its inner product with maps of Northern Hemisphere surface air temperature produced by an independent climate model subjected to SRES A1B external forcing in Figure 1. Astonishingly, scalar prediction is able to determine the climate response of Northern European surface air temperature when subjected to SRES A1B forcing and given a 20-year record of Northern Hemispheric surface air temperature to within $0.1 \text{ K decade}^{-1}$ when the actual Northern European temperature record by itself shows a trend with uncertainty $0.7 \text{ K decade}^{-1}$. From this example, one can conclude that scalar prediction is a method of determining climate response of any variable of the climate system, including regional average quantities, from arbitrary data sets. See Leroy and Anderson (2010) for a more in depth explanation.

4 Accuracy requirements

It is possible to derive an objective method for determining accuracy requirements from the equations of scalar prediction. In the above derivation, I have not considered observation error, but it plainly belongs as an extra term Σ_{obs} in the Σ of equation 6. A political goal of any climate benchmarking system must be to delay as little as possible the positive detection of trends or refinement of climate projection. Already the natural variability of the climate system places lower bounds on detection times (because of $\Sigma_{d\mathbf{n}/dt}$). It would be politically damaging to substantially increase any time-to-detection by imposing an observation error Σ_{obs} that is comparable in magnitude to natural variability. After accounting for

coherence times of natural variability τ_n and mission duration τ_{obs} , one arrives at the requirement that

$$\sigma_{\text{obs}}^2 \tau_{\text{obs}} \ll \sigma_n^2 \tau_n. \quad (7)$$

Here I have defined σ_{obs} as the one-sigma observation error, σ_n^2 the natural variability and τ_n its coherence time. Astonishingly, the requirements on overall error of a climate benchmarking instrument in space depends on the lifetime of the mission! The error of the benchmark is the root-sum-square of the sampling error and instrument accuracy. See [Leroy et al. \(2008\)](#) for a more expansive derivation.

5 GNSS radio occultation

Radio occultation originated in the planetary sciences and has generated a large catalogue of temperature profiles of planetary atmospheres, especially Venus's. A radio occultation occurs when a planetary atmosphere occults a microwave radio beacon synchronized to an ultra-stable oscillator (USO) as viewed by a receiver outside the occulting atmosphere. In planetary missions, the occulted transmitter is the inter-planetary spacecraft (e.g., Pioneer Venus, Magellan, Voyager 1, Voyager 2, Mariner 10), the occulting atmosphere that of the planet, and the receiver one on the Earth's surface. The same can be done for Earth's atmosphere if one uses the transmitters of the Global Navigation Satellite Systems (GNSS), the Earth's atmosphere, and one or more GNSS receivers in low Earth orbit (LEO). The lone GNSS currently available is the Global Positioning System (GPS).

In radio occultation, the occulting atmosphere refracts the radio signal that transects the atmosphere, and the bending of the signal shifts the frequency of the radio signal as obtained by the LEO receiver. The dimension of the observable in radio occultation is inverse time. The demand of S.I. traceability in this case is that the transmitter and receiver of the radio signal be calibrated by a chain of comparison to the international definition of the second, which is the time required for 9,192,631,770 cycles the hyperfine splitting of the ground state of the Cs¹³³ atom. Traceability is established at the GNSS transmitter by synchronizing the radio transmissions to an ensemble of cesium-rubidium clocks. Traceability at the receiver can be done by other synchronizing the receiver to an on-board USO or by calibrating a poorly calibrated receiver clock to non-occulted GNSS transmitters' signals. If the GNSS transmitters' clocks themselves have questionable accuracy, they in turn can be calibrated by observing their signals with receivers on the ground synchronized to better calibrated clocks. This process is commonly called "double differencing" ([Hardy et al. 1994](#)) and is illustrated in figure 2.

The relationship between Doppler shift and the angle ε through which the occulting atmosphere bends the ray is

$$\lambda \Delta\nu = v_{\text{GPS}} \cos \phi_{\text{GPS}} + v_{\text{LEO}} \cos \phi_{\text{LEO}}. \quad (8)$$

The transmit and receive angles $\phi_{\text{GPS}}, \phi_{\text{LEO}}$, transmitter and receiver velocities $v_{\text{GPS}}, v_{\text{LEO}}$ are illustrated in Figure 3; λ is the vacuum carrier wavelength of the transmitted signal and $\Delta\nu$ the measured Doppler-shift of the radio signal's frequency. An assumption of local spherical symmetry in the atmosphere together with S.I. traceable knowledge of the positions and velocity of the transmit and receiver satellites allows a determination of the bending angle ε as a function of impact parameter p during a GNSS radio occultation. With spherical symmetry, the impact parameter of the radio signal is the same on both the transmit and receive sides of the occultation. Inversion ([Fjeldbo et al. 1971](#)) of radio occultation obtains vertical profile of the microwave index of refraction by way of an Abelian transform:

$$n(p) = \frac{1}{\pi} \int_p^\infty \frac{\varepsilon(p') dp'}{\sqrt{p'^2 - p^2}} \quad (9)$$

where the independent coordinate $p = nr$, r the distance from the atmosphere's local center of curvature, which to first order is the Earth's radius. The index of refraction is related to atmospheric density and

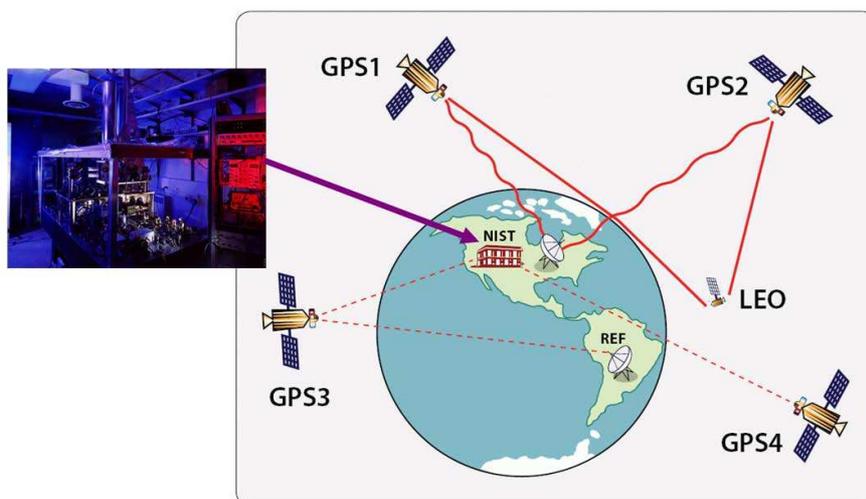


Figure 2: Double differencing. The LEO to GPS1 link is the occulted link. A poorly calibrated LEO clock can be corrected by a link to the “reference” non-occulted satellite GPS2. If the GPS clocks’ calibrations are questioned, both GPS1 and GPS2 can be calibrated to a ground-based clock, in this case located at NIST, which is calibrated by extremely accurate clocks. NIST in Boulder, Colorado, USA, hosts a cesium fountain clock, accurate to 10^{-15} 1-second Allen variance.

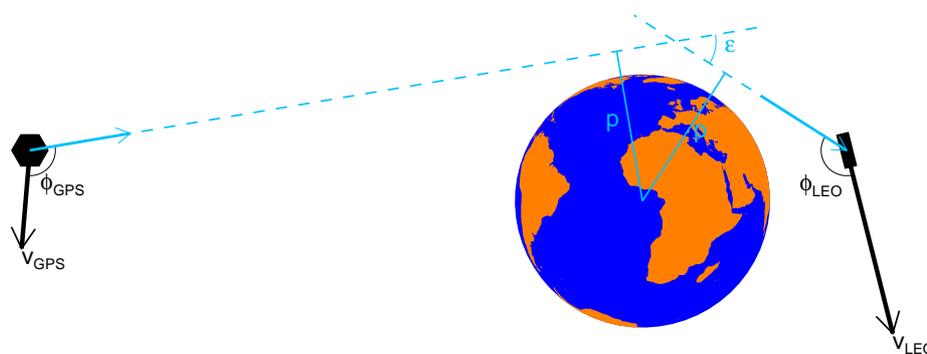


Figure 3: The geometry of a GPS radio occultation. Doppler shifting of the radio signal is governed by the angle between transmitter velocity and the direction of the transmitted ray ϕ_{GPS} and the angle between the received ray and the velocity of the LEO receiver ϕ_{LEO} . They can be inferred from the geometry of the occultation, the measured Doppler shift Δv , and an assumption of local spherical symmetry of the atmosphere. The bending angle ϵ , the angle ϵ through which the occulting atmosphere bends the ray, and the impact parameter p , the asymptotic miss distance of the vacuum ray paths from the center of curvature, can be calculated given ϕ_{GPS} and ϕ_{LEO} by simple trigonometry.

water vapor by an empirically determined relationship. At this point, suffice it to say that the index of refraction is simply a linear function of atmospheric density throughout the atmosphere and the contribution of water vapor is a significant contributor only in the lower troposphere when the temperature exceeds ≈ 250 K.

To be sure, exercising the equations given above require care. The integral of the inversion equation (9) to infinity is noisy and ill-determined. The fundamental measurement equation (8) requires relativistic corrections. For details, see, for example, [Hajj et al. \(2002\)](#). Moreover, the complex signal dynamics associated with diffraction and multipath are best undone by physical optics techniques ([Gorbunov et al. 2004](#)). An authoritative description and error analysis of radio occultation is given in [Kursinski et al. \(1997\)](#). Lastly, the independent coordinate r can be converted to geopotential height—to ease interpretation to atmospheric scientists—by application of a gravity model ([Leroy 1997](#)). The important points here are that GNSS radio occultation is S.I. traceable by calibration against atomic clocks and its observables are bending angle as a function of impact parameter $\varepsilon(p)$ or index of refraction as a function of geopotential height $n(h)$.

6 Verifying a benchmark

Implementing a climate benchmark offers the advantage of empirical verification. There are two qualities of a climate benchmark, reproducibility of standards and reproducibility of trends, that can be tested using its own observations. Reproducibility of standards is the quality wherein the international standard used to calibrate the observation can be reproduced anytime and anywhere. Reproducibility of trends is the quality wherein trends of geophysical variables obtained by climate benchmarks can be obtained accurately independent of the retrieval algorithm used.

The qualities of reproducibility of standards and of trends should apply to any climate benchmark, and they can be checked using GNSS radio occultation data which already exists. The quality of reproducibility of standards for GNSS radio occultation has been checked by comparing co-located data obtained by different GNSS radio occultation missions by [Hajj et al. \(2004\)](#). For many co-located soundings, a random component due to differing view geometries and times was found to dominate. Most importantly, though, no bias was found in the ensemble of co-located soundings throughout the troposphere. In the stratosphere, however, an unexpected bias was found which remains unexplained. My view is that local multi-path or erroneous spacecraft attitude information in either one of the two missions would create the bias pattern seen in Figure 22 of [Hajj et al. \(2004\)](#). The quality of reproducibility of trends can be checked by comparing inter-annual trends in the index of refraction obtained by independent retrieval algorithms. This was done in [Ho et al. \(2009\)](#). Four (semi-independent) algorithms were used to process CHAMP data and retrieve the refractive index in the vertical region around the tropopause. There was no statistically significant difference in trends found over the lifetime of CHAMP data, but only after an accounting for sampling error. In GNSS radio occultation, though, sampling error is problematic because there is no single objective quality control method that can qualify individual soundings. As a result, different subsets of CHAMP soundings are retained by the different algorithms thus resulting in different sampling errors. Objective quality control, therefore, must be given careful attention when handling climate benchmark data.

7 Information in GNSS radio occultation

A climate benchmark requires S.I. traceable data types; GNSS radio occultation is S.I. traceable; GNSS radio occultation is a climate benchmark data type, and two empirical tests have demonstrated as much. What information does a decadal scale time series of radio occultation have on the climate system? First

I will show what statistically significant signal will emerge first in a long time series and then how radio occultation is useful in constraining global surface air temperature trends.

I predict the first statistically significant climate signal using optimal detection, the first inference in scalar prediction described earlier. The first inference demands only one climate model to determine the signal $\mathbf{s} = d\mathbf{g}/d\alpha$. Consequently, the statistical significance of $(d\alpha/dt)_{\text{ml}}$ with a time series of length Δt is independent of the definition of α . The optimal fingerprint determined using optimal detection, then, is the pattern of climate change that will be first detected significantly. This was done for radio occultation in Leroy et al. (2006), and I will synopsise the results here.

In Earth radio occultation, it is common to speak in terms of “refractivity” rather than “index of refraction”. The refractivity N is related to the index of refraction n through $N = (n - 1) \times 10^6$, the index of refraction less one in parts per million. The refractivity is related to pressure p , temperature T , and water vapor partial pressure p_w through $N = (77.6 \text{ K hPa}^{-1})(p/T) + (3.73 \times 10^5 \text{ K}^2 \text{ hPa}^{-1})(p_w/T^2)$. The radio occultation research community has minted new meteorological variables, “dry pressure” among them. It is a convenient quantity to use in climate signal detection studies because it is easily interpreted. Dry pressure is the downward integral of refractivity in height multiplied by a constant factor. If water vapor did not contribute at all to refractivity—hence *dry* pressure—this integral would simply be pressure. Precisely, dry pressure p_N is

$$p_N = p + (7730 \text{ K}) \int_0^p \frac{q(p') dp'}{T(p')} \quad (10)$$

where q is specific humidity. Dry pressure p_N is the same as pressure p except where the second term on the right of equation 10 is large. Trends in the log of dry pressure are simply interpreted because, above the lower troposphere, variability in its global average is the same as variability in tropospheric thickness. A positive trend in dry pressure near the tropopause is thermal expansion of the troposphere.

Figure 4 shows the trends in log-dry pressure as simulated by twelve CMIP3 models. The dominant signal is a broad maximum centered at approximately 20 km stretching across the low to mid-latitudes. Much of the maximum can be explained as thermal expansion of the tropical troposphere, but not all. Extension of the maximum into mid-latitudes can only be explained by something like expansion of the tropics and a general poleward migration of baroclinic zones. The covariance matrix $\Sigma_{d\mathbf{n}/dt}$ appearing in the equations of optimal detection, 2 and 3, is difficult to invert because it is ill-conditioned. Instead, one must compute a “pseudo-inverse” by truncating the data space to those of the dominant eigenvectors of $\Sigma_{d\mathbf{n}/dt}$. The eigenvectors of $\Sigma_{d\mathbf{n}/dt}$ are typically called empirical orthogonal functions (EOF). It turns out that the EOFs that contribute most dominantly to optimal detection are those that correspond to poleward migration of baroclinic zones and thermal expansion of the tropical troposphere. The former is probably closely associated with poleward migration of the jet stream and the latter with the El Niño-Southern Oscillation (ENSO). It is independent of the CMIP3 model used to determine the signal and the model used to approximate natural variability. Given SRES A1B forcing in reality, it should take only 7 to 13 years to detect human influence with 95% confidence. This level of confidence is obtained with $\simeq 7$ m of thermal expansion of the troposphere.

In the application of scalar prediction to radio occultation data, I choose to use trends in zonal average log-dry pressure to determine the optimal fingerprint of change in global average surface air temperature associated with SRES A1B forcing. To do so, I apply equations 4 through 6, defining α to be global average surface air temperature and \mathbf{g} to be the zonal average log-dry pressure. For each model i , the fingerprint \mathbf{s}_i is computed by dividing the 40-yr trend in zonal average log-dry pressure ($d\mathbf{g}/dt$) by the 40-yr trend in surface air temperature ($d\alpha/dt$) to get $\mathbf{s} = d\mathbf{g}/d\alpha$. The optimal fingerprint \mathbf{f} is shown in Figure 5. It clearly looks for poleward migration of the jet streams and increasing boundary layer water vapor in the tropics. It is the optimal weighting for inferring global average surface air temperature trends from radio occultation trends.

Figure 5 also shows time series of $\mathbf{f}^T \mathbf{d}(t)$ with the data $\mathbf{d}(t)$ taken from an independent climate model.

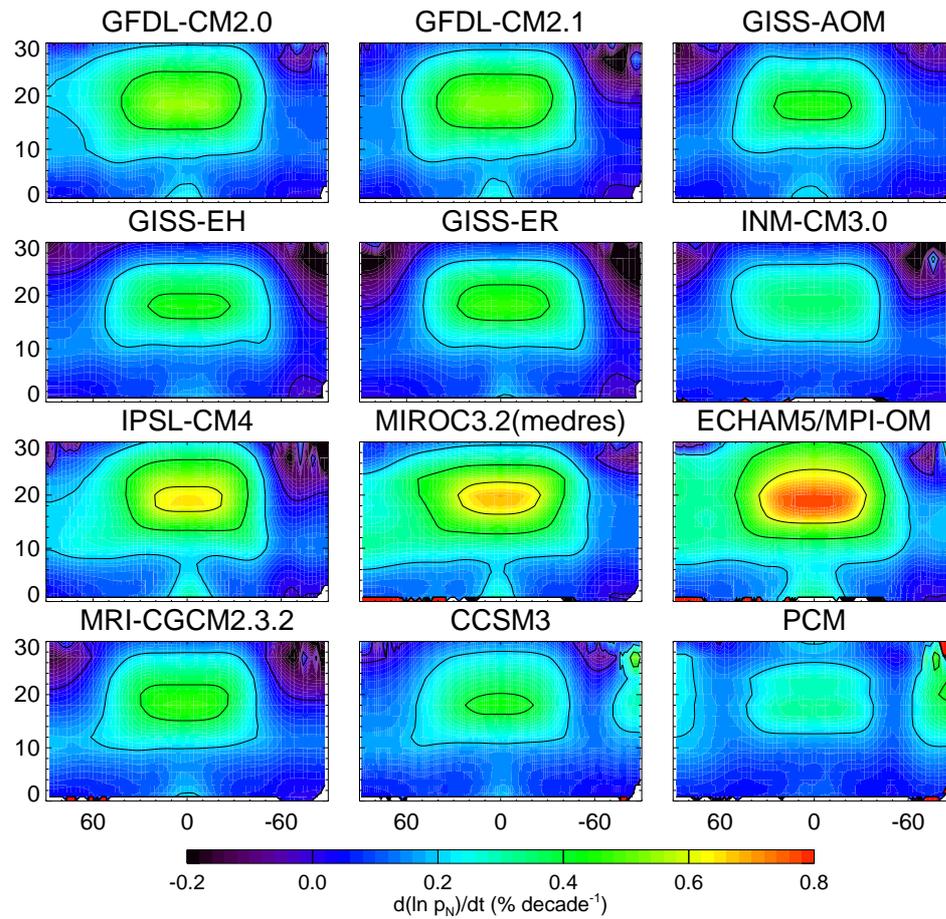


Figure 4: Trends in zonal average log-dry pressure as simulated by CMIP3 models. The zonal average log-dry pressure was computed using equation 10 for the named models subjected to SRES A1B forcing. The trend was determined by linear regression of the first 40 yrs of output. The ordinate is geopotential height (km), and the abscissa is latitude from north to south. Taken from Leroy et al. (2006).

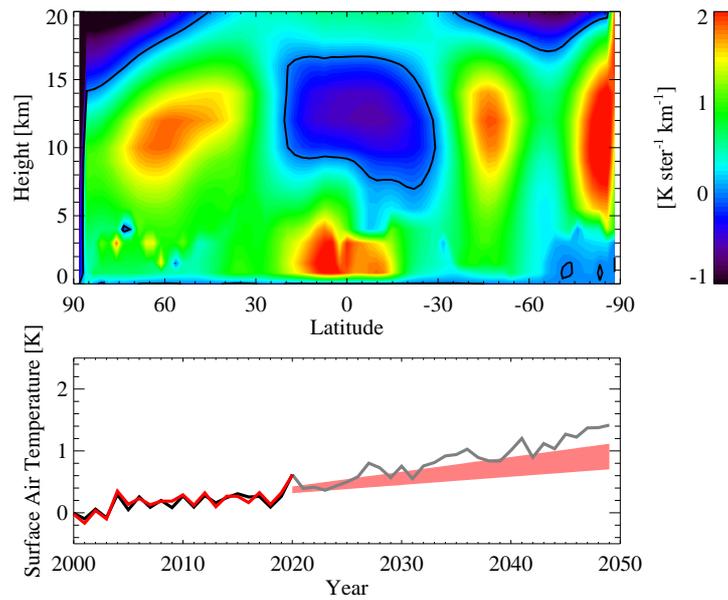


Figure 5: Scalar prediction for surface air temperature trends using radio occultation zonal average log-dry pressure trends. The top plot shows the optimal fingerprint for global surface air temperature change in the space of zonal average log-dry pressure. The lower plot shows the inner product of the optimal fingerprint on annual and zonal average log-dry pressure as a function of time $\mathbf{f}^T \mathbf{d}(t)$ with data $\mathbf{d}(t)$ produced by an independent climate model subjected to SRES A1B forcing (red). It directly represents global average surface air temperature to within an additive constant (black). The red shaded region is the $\pm 1\sigma$ extrapolation of $d(\mathbf{f}^T \mathbf{d})/dt$ into the future, and the gray curve the evolution of global average surface air temperature.

Interestingly, the product $\mathbf{f}^T \mathbf{d}(t)$ almost exactly recreates the inter-annual fluctuations of global average surface air temperature without optimization. This can only come about if most of the variance in upper air meteorological quantities is explained by global average surface air temperature. While jet stream migration itself can be positively detected sooner than global warming, monitoring upper air temperature gives no additional information on surface air temperature response to anthropogenic forcing than just the global average surface air temperature by itself.

One interesting consequence of climate benchmarks is that, while retrieved quantities are not themselves necessarily S.I. traceable, change in those quantities can be trusted. For that reason, in radio occultation, a change in a retrieved quantity such as dry pressure can be trusted even though it is traceable to no international standard. Moreover, detection times obtained by optimal methods should be independent of the variable used, either the calibrated observed quantity or a retrieved one. Just above are the results for optimal detection and scalar prediction using a retrieved quantity of radio occultation. [Ringer and Healy \(2008\)](#) performed a sensitivity study for radio occultation working in the space of bending angle ϵ , a quantity more closely linked to the calibrated observed quantity than dry pressure. As expected, they found detection times very similar to those found in [Leroy et al. \(2006\)](#).

8 Information in infrared spectra

The infrared spectrum is rich in information content—a prime motivator for deployment of the operational Atmospheric Infrared Sounder and Infrared Atmospheric Sounding Interferometer (IASI)—but until recently has not been traceable to an infrared standard with accuracy sufficient for climate monitoring. Recent improvements in the standard for infrared radiance, including phase-change blackbodies for calibrating temperature and quantum cascade lasers for calibrating blackbody emissivity, have demonstrated a standard that is accurate to 0.03 K in brightness temperature at room temperature ([Gero et al. 2008](#); [Gero et al. 2009](#)).

Infrared retrieval is common for operational sounders, but what can be learned from trends in the infrared spectrum over decadal time scales? I apply scalar prediction to the infrared spectrum as the data type. To narrow the selection of the scalar α , I note that the infrared spectrum is a special case for climate monitoring because it observes in the space of outgoing longwave radiation (OLR), one of the fundamental quantities for radiative balance of the climate system. Climate’s response to radiative forcing is uncertain in part because the radiative feedbacks of the climate system are so difficult to constrain. Because of its richness of information content, however, monitoring the infrared spectrum ought to yield strong constraints on the longwave feedbacks of the climate system. So I have chosen to apply scalar prediction to the infrared spectrum with the scalars of interest being the longwave feedbacks.

The spectral signal associated with each longwave feedback can be determined by partial radiative perturbation ([Wetherald and Manabe 1988](#)). Partial radiative perturbation (PRP) has been used to diagnose the feedbacks inherent in climate models in a broadband sense ([Bony et al. 2006](#)). There is no reason that PRP cannot also be applied in the spectral sense as well ([Leroy et al. 2008](#); [Huang et al. 2010](#)). In PRP, a climate model is run twice, once subjected to radiative forcing (the “forced” run) and once not (the “control” run). The outgoing infrared spectrum is computed for both runs, the difference being the response of the climate system to radiative forcing as seen in the infrared spectrum. One can obtain the spectral response corresponding to an individual feedback by suppressing the change corresponding to the variable designated by the feedback. For example, to obtain the water vapor feedback, I simulate the change in the infrared spectrum from the forced run but with the output of water vapor taken from the control run. The difference between this spectrum and the simulation for the forced run gives the spectral change corresponding to the longwave-water vapor feedback.

Figure 6 shows the spectral radiance fingerprints for clear-sky radiance simulations integrated over the

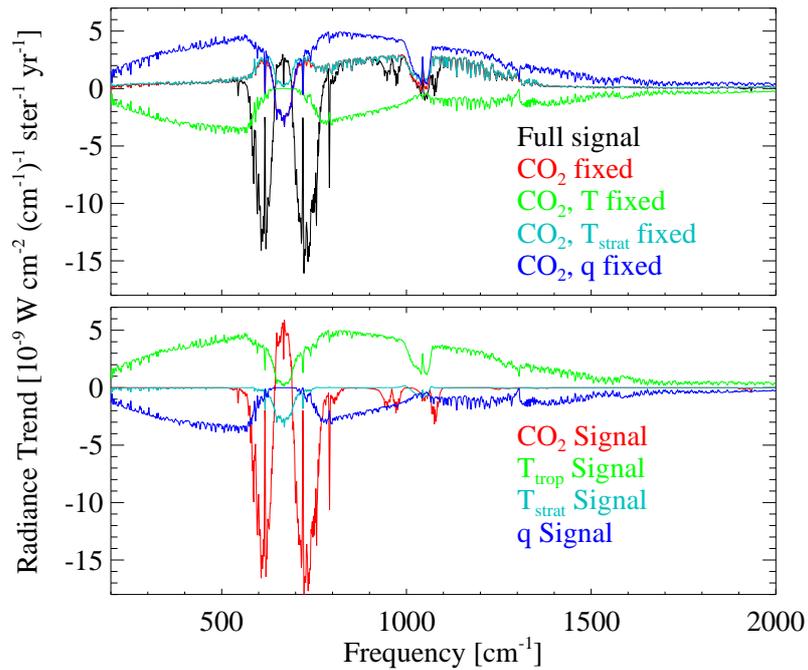


Figure 6: Clear-sky spectral infrared fingerprints. The top plot shows the intermediate results of partial radiative perturbation: “full signal” is the complete spectral response of the tropics (in clear skies) to SRES A1B forcing, and, for example, “CO₂-fixed” is the spectral response of the tropics with the exception of CO₂ being held constant. The lower plot shows the fingerprint radiance signals after subtracting the PRP signals from the full signal. The lower signals are used in scalar prediction. Taken from Leroy et al. (2008).

tropics using the CMIP3 ensemble of climate models. Even though the water vapor and upper air temperature spectral fingerprints look similar if opposite in sign, when model uncertainty in the spectral fingerprints is taken into consideration, they are still unique enough to separate the water vapor and lapse rate longwave feedbacks. When cloudy skies are considered, though, over decadal timescales some of the cloud signals are not easily distinguished from upper air temperature and water vapor signals (Huang et al. 2010). The reason for the ambiguities is that the radiance signals take substantially different forms depending on the background climate in which they appear. For example, a mid-latitude cloud signal can look markedly different than a tropical cloud feedback, and so the uncertainty in the fingerprint is quite large. This certainly inhibits the formulation of a single optimal fingerprint \mathbf{f} for mapping that cloud feedback.

Scalar prediction is a powerful methodology, though, so the results of the all-sky detection problem performed in Huang et al. (2010) can be improved with the addition of a data type that is independent of clouds. We call it joint fingerprinting because multiple data types can be considered jointly in the data vector \mathbf{d} of equation 4. Joint fingerprinting succeeds in resolving most of the cloud ambiguities with two exceptions: it cannot resolve mid-tropospheric and high clouds unambiguously nor surface temperature and low clouds. Figure 7 shows the true upper cloud-longwave feedback and what might be obtained from infrared-only scalar prediction and from joint scalar prediction after a doubling of CO₂. Joint scalar prediction resolves problems with ambiguity especially in polar regions and to a lesser degree in the tropics.

Regarding the information content in infrared spectra, much more work needs to be done. The data type is inherently complex because its dimensionality is both spectral and spatial. Work so far has addressed the spectral dimension, and because there is little to no variability in the spectral dimension

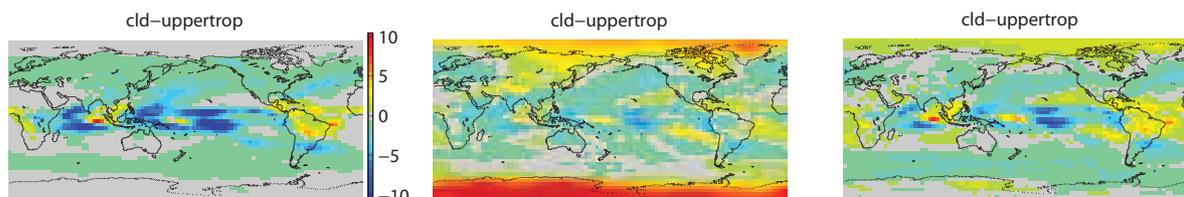


Figure 7: Mapping the upper cloud-longwave feedback. The left plot shows a map of the OLR perturbation associated with the upper cloud-longwave feedback after a doubling of CO_2 , as diagnosed from the output of a model of the Cloud Feedback Model Intercomparison Project. The center shows the upper cloud-longwave feedback as would be obtained by monitoring only the infrared spectrum and applying scalar prediction. The right shows the upper cloud-longwave feedback as would be obtained by considering the infrared spectrum and radio occultation dry pressure and applying scalar prediction.

associated with a process other than the feedbacks being sought, there is little optimization to be had. On the other hand, the spatial dimension is likely to behave quite differently because there is a lot of variability in longwave fluctuations in the spatial coordinate associated with processes other than long-term trends. The problem, though, is that the prior on the spatial structure of some of the feedbacks is frighteningly weak. Optimal detection and scalar prediction require some moderate prior knowledge of the form of the signal in the chosen dimension to be of use. The information ought to be sufficient to distinguish between different signals in the multi-pattern problem or to distinguish between the signal and the gravest modes of natural variability. In the case of the cloud feedbacks, it is unclear the degree to which models can be used to prescribe their patterns in space. Moreover, that climate models typically do not generate the output necessary to simulate all-sky radiances just complicates matters. In the end, though, an information content study must answer the question, at least to first order, of how long a climate monitoring data set of the infrared spectrum must be before climate models can be tested.

Colman (2003) and Bony et al. (2006) have diagnosed the radiative feedbacks of various ensembles of climate models, the latter having diagnosed those of the CMIP3 ensemble contributed to the IPCC Fourth Assessment Report. In order to test such ensembles, it will be necessary to use trend data to estimate the actual radiative feedbacks empirically. Above, I have shown the first steps of a methodological pathway that can be used to do so.

9 Summary

First, the implementation of climate benchmarks as the foundation of a climate observing systems is necessary to prevent sensitivity to breaks in time series of data. The hallmark of a climate benchmark is S.I. traceability, a chain of calibration with demonstrable accuracy to the international standard that defines the units of the fundamental observable. There are two necessary tests that prove the bona fides of a climate benchmark system: reproducibility of standards and reproducibility of trends.

Second, an application of Bayesian inference can be applied in theoretical studies to learn how a climate benchmark data type can be used to test climate models. Scalar prediction, a second level of Bayesian inference built upon the commonly used method of optimal detection, serves this purpose. Scalar prediction can be applied to any potential climate benchmark data type to learn about any arbitrarily chosen quantity (or prediction) of the climate. The outcome is that the quantity is either ambiguously or unambiguously constrained, unambiguously if the data type is sufficiently sensitive to the quantity in question. In that case, another outcome of scalar prediction is the duration of the time series necessary to gain

precise knowledge of the quantity in question.

Third, radio occultation using the Global Navigation Satellite Systems (GNSS) is, in fact, a climate benchmark data type, and its observable can be used to measure thermal expansion of the troposphere caused by global warming, poleward migration of mid-latitude baroclinic zones, and to infer global surface air temperature trends. GNSS radio occultation is S.I. traceable by virtue of calibration of its observable, frequency shifts, to the international definition of the second, realized by atomic clocks. Nonexistent bias in the troposphere between co-located soundings of independent GNSS radio occultation missions has shown that GNSS radio occultation can successfully reproduce its traceable international standard. Also, nonexistent difference between the trends derived by different retrieval algorithms has shown that GNSS radio occultation yields a data type that enables reproducibility in trends. Both tests validate GNSS radio occultation as a climate benchmark data type. The most obvious climate signal in GNSS radio occultation data is thermal expansion of the troposphere, and the first signal to be significantly detected should be poleward migration of baroclinic zones, including poleward shifts of the mid-latitude jet streams. The detection should be 95% confident in 7 to 13 years. Scalar prediction has shown that GNSS radio occultation can be used to infer trends in global average surface air temperature almost perfectly but without optimization.

Fourth, measurement of the outgoing thermal infrared spectrum can be a climate benchmark data type, and it can be used to constrain the longwave radiative feedbacks of the climate system. Improvements in the development of an infrared radiance standard have enabled measurement of the thermal infrared spectrum from space as a climate benchmark. The spectrum itself is a decomposition of outgoing longwave radiation, which is itself a primary regulator of the radiative balance of the climate system. Feedbacks in radiation govern the sensitivity of the climate system, and the spectrum of outgoing longwave radiation can be expected to contain information to resolve the different climate feedbacks. Scalar prediction has been used to show that this is in fact the case with the exception of cloud feedbacks which tend to be ambiguous with upper air and surface temperature. GNSS radio occultation, though, is insensitive to clouds and thus is useful for resolving the ambiguities inherent to the infrared spectrum as a climate benchmark data type. By considering a retrieved quantity of GNSS radio occultation jointly with the infrared spectrum in scalar prediction, it indeed is possible to resolve most of the cloud-longwave feedbacks. The only exceptions are the ambiguity between mid- and upper tropospheric cloud feedbacks and the ambiguity between low cloud-longwave feedback and surface temperature response.

The theoretical studies presented here are useful for inferring the information content of climate benchmark data types, yet I have little expectation that scalar prediction will be used when suitably long time series of climate benchmark data becomes available. It is probably much more likely that atmospheric reanalysis systems will become sophisticated enough to take full advantage of the unprecedented accuracy of these data to gain truly accurate reconstructions of the state of the climate system. Impressive steps have already been taken here at ECMWF in this direction, and we in the CLARREO project are excited to see this and encourage its further development. With accurate reanalyses enabled by climate benchmarks, I fully expect that the output of these reanalyses will be perfectly well suited to the testing of climate models by trend analysis. Thank you.

Acknowledgements

We acknowledge the modeling groups, the Program for Climate Model Diagnosis and Intercomparison (PCMDI) and the WCRP's Working Group on Coupled Modelling (WGCM) for their roles in making available the WCRP CMIP3 multi-model dataset. Support of this dataset is provided by the Office of Science, U.S. Department of Energy. We wish to thank Richard Goody for many useful conversations on the topic of testing climate models. This work was supported by grant ATM-0755099 of the National Science Foundation.

References

- Allen, M. and S. Tett (1999). Checking for model consistency in optimal fingerprinting. *Climate Dyn.* 15(6), 419–434.
- Bell, T. (1986). Theory of optimal weighting to detect climate change. *J. Atmos. Sci.* 43, 1694–1710.
- Bony, S., R. Colman, V. Kattsov, R. Allan, C. Bretherton, J. Dufresne, A. Hall, S. Hallegatte, M. Holland, W. Ingram, D. Randall, B. Soden, G. Tselioudis, and M. Webb (2006). How well do we understand and evaluate climate change feedback processes? *J. Climate* 19, 3445–3482.
- Colman, R. (2003). A comparison of climate feedbacks in general circulation models. *Climate Dyn.* 20, 865–873.
- Dykema, J. and J. Anderson (2006). A methodology for obtaining on-orbit SI-traceable spectral radiance measurements in the thermal infrared. *Metrologia* 43, 287–293.
- Fjeldbo, G., A. Kliore, and V. Eshleman (1971). Neutral atmosphere of Venus as studied with Mariner-V radio occultation experiments. *Astronom. J.* 76(2), 123–140.
- Gero, P., J. Dykema, and J. Anderson (2008). A blackbody design for SI-traceable radiometry for Earth observation. *J. Atmos. Ocean. Tech.* 25(11), 2046–2054.
- Gero, P., J. Dykema, and J. Anderson (2009). A quantum cascade laser-based reflectometer for on-orbit blackbody cavity monitoring. *J. Atmos. Ocean. Tech.* 26(8), 1596–1604.
- Gorbunov, M., H. Benzon, A. Jensen, M. Lohmann, and A. Nielsen (2004). Comparative analysis of radio occultation processing approaches based on Fourier integral operators. *Radio Sci.* 39(6), doi:10.1029/2003RS002916.
- Hajj, G., C. Ao, B. Iijima, D. Kuang, E. Kursinski, A. Mannucci, T. Meehan, L. Romans, M. Juarez, and T. Yunck (2004). CHAMP and SAC-C atmospheric occultation results and intercomparisons. *J. Geophys. Res.* 109(D06109), doi:10.1029/2003JD003909.
- Hajj, G., E. Kursinski, L. Romans, W. Bertiger, and S. Leroy (2002). A technical description of atmospheric sounding by gps occultation. *J. Atmos. Solar Terr. Phys.* 64(4), 451–469.
- Hardy, K., G. Hajj, and E. Kursinski (1994). Accuracies of atmospheric profiles obtained from GPS occultations. *Int. J. Sat. Comm.* 12(5), 463–473.
- Hasselmann, K. (1993). Optimal fingerprints for the detection of time-dependent climate change. *J. Climate* 6(10), 1957–1971.
- Hasselmann, K. (1997). Multi-pattern fingerprint method for detection and attribution of climate change. *Climate Dyn.* 13(9), 601–611.
- Ho, S., G. Kirchengast, S. Leroy, J. Wickert, A. Mannucci, A. Steiner, D. Hunt, W. Schreiner, S. Sokolovskiy, C. Ao, M. Borsche, A. von Engeln, U. Foelsche, S. Heise, B. Iijima, Y. Kuo, R. Kursinski, B. Pirscher, M. Ringer, C. Rocket, and T. Schmidt (2009). Estimating the uncertainty of GPS radio occultation data for climate monitoring: Intercomparison of CHAMP refractivity climate records from 2002 to 2006 from different data centers. *J. Geophys. Res.* 114(D23107), doi:10.1029/2009JD011969.
- Huang, Y., S. Leroy, P. Gero, J. Dykema, and J. Anderson (2010). Separation of longwave climate feedbacks from spectral observations. *J. Geophys. Res.* In Press.
- Kursinski, E., G. Hajj, J. Schofield, R. Linfield, and K. Hardy (1997). Observing Earth's atmosphere with radio occultation measurements using the Global Positioning System. *J. Geophys. Res.* 102(D19), 23429–23465.
- Leroy, S. (1997). Measurement of geopotential heights by GPS radio occultation. *J. Geophys. Res.* 102(D6), 6971–6986.

- Leroy, S. and J. Anderson (2010). Optimal detection of regional trends using global data. *J. Climate* Submitted.
- Leroy, S., J. Anderson, and J. Dykema (2006). Testing climate models using GPS radio occultation: A sensitivity analysis. *J. Geophys. Res.* *111*, D17105, doi:10.1029/2005JD006145.
- Leroy, S., J. Anderson, J. Dykema, and R. Goody (2008). Testing climate models using thermal infrared spectra. *J. Climate* *21*, 1863–1875.
- Leroy, S., J. Anderson, and G. Ohring (2008). Climate signal detection times and constraints on climate benchmark accuracy requirements. *J. Climate* *21*(4), 841–846.
- North, G., K. Kim, S. Shen, and J. Hardin (1995). Detection of forced climate signals: I. Filter theory. *J. Climate* *8*(3), 401–408.
- Pollock, D., T. Murdock, R. Datla, and A. Thompson (2000). Radiometric standards in space: The next step. *Metrologia* *37*(5), 403–406.
- Pollock, D., T. Murdock, R. Datla, and A. Thompson (2003). Data uncertainty traced to SI units. Results reported in the International System of Units. *Int. J. Rem. Sensing* *24*(2), 225–235.
- Ringer, M. and S. Healy (2008). Monitoring twenty-first century climate using GPS radio occultation bending angles. *Geophys. Res. Lett.* *35*(5), 10.1029/2007GL032462.
- Wetherald, R. and S. Manabe (1988). Cloud feedback processes in a general circulation model. *J. Atmos. Sci.* *45*(8), 1397–1415.