Why considering only “systematic error” and “random error” (or “accuracy” and “precision”) can be problematic – some MLS-based examples

Nathaniel Livesey, Maggie Johnson, and Amy Braverman
Jet Propulsion Laboratory, California Institute of Technology

Thomas von Clarmann
Karlsruhe Institute of Technology

1 Overview

- **Measurement uncertainty is typically described in terms of “random errors” (often referred to as “precision”) and “systematic errors” (also referred to as “accuracy”).**
- **Although this approach is good enough for many analyses, it has limitations.**
- **In particular, it is less useful when considering quantities such as long term trends that are derived from statistics of many measurements.**
- **Here we discuss these limitations, consider an example, and point towards a solution.**
- **We suggest that the sources of uncertainty that come into play in such cases be collectively known as “interaction bias” (although the informal name of “headache errors” has also been suggested and may be the one that sticks).**

2 High level view of the issue

What is the current practice?

- The typical approach taken when reporting uncertainty in atmospheric measurements (as with many other measurements) is to divide the different uncertainty components into two categories:

  Random errors, often synonymous with “precision”. Typically the largest contributor to these is the noise on the raw radiance measurements made by the instrument.

  Systematic errors, often synonymous with “accuracy”. Examples of these include errors in instrument calibration or uncertainties in spectroscopy.

What are its limitations?

- This approach, while useful in many circumstances, risks oversimplifying the situation, particularly when considering quantities derived from multiple related observations.
- For example, consider the study of decadal trends in stratospheric ozone from a spaceborne instrument such as the Aura Microwave Limb Sounder (MLS).
  - Given a few thousand individual measurements per day globally, any purely random instrumental error terms will be reduced by a significant $1/\sqrt{n}$ factor when considering long-term trends.
  - Further, any systematic errors that manifest purely as constant “biases” will cancel out in a trend calculation.
  - A framework that reported uncertainty solely in those terms would accordingly lead users to conclude that decadal trends estimated from the observations are virtually error free.
  - Such an assessment is clearly unhelpful and represents overly optimistic guidance to users.
- Similar issues arise when computing spatial gradients, seasonal cycles, etc. based on many individual observations.

What is really going on?

- The reality in most situations is a far more complex set of uncertainties that don’t readily fall into such characterization.
  - Even if radiance noise is uncorrelated, its impacts on retrieved states are not.
  - The impact of an error may be a function of atmospheric state (e.g., stronger when temperature is greater).
  - In situ measurements face comparable issues, e.g., biases in temperature sensors on radiosondes give errors that are correlated vertically but typically not temporally or spatially (batch-to-batch issues aside).

What can we do about it?

- Most data providers have computed the sensitivity of their measurements to systematic errors (e.g., uncertainty in spectroscopy) through perturbation studies or similar calculations.
- It is possible to take this a step further and quantify the dependence of the impacts of these errors on the atmospheric state and/or other parameters.
- With this knowledge, the potential impact of such errors on quantities such as trends and spatial gradients can, in principle, be estimated.
- A dilemma is how to communicate this complexity to users when needed, without putting them off using the data by presenting them with an uncertainty reporting framework that is unnecessarily complicated for many applications.

3 An Aura Microwave Limb Sounder (MLS)-based example

MLS background

- MLS, like many similar instruments, measures atmospheric spectra and uses these to deduce temperature and composition profiles.
- The spectral lines are temperature and pressure dependent, and many lines from various atmospheric species overlap with each other.
- Accordingly, the MLS products are all intertwined and systematic errors affect them in coupled ways.

An example of such an error

- For example, an error in ozone spectroscopy will affect not only ozone, but also any other product with nearby spectral lines.
- In all cases, the magnitude of the error introduced will be a function of how much ozone is present in the atmosphere.

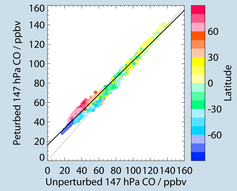

Figure 1: MLS 146 hPa (~10 km) CO retrieved with perturbed O_3 spectroscopy (y-axis), vs. that retrieved with the correct spectroscopy (x-axis) for a simulated day’s worth of data. Black line is a linear fit, thin grey line is the 1:1 line.

- Figure 1 shows ozone spectroscopy error impacts on MLS 146 hPa CO.
- As would be expected, the perturbations from this error are largest at higher latitudes, where ozone abundances are larger with greater variability (and CO abundances are smaller).
- No simple statement of bias or random error can properly describe this error.

An example where this matters

- An MLS data user contacted us concerning differences between the MLS and IASI estimates of ~150 hPa CO in the Asian Monsoon region.

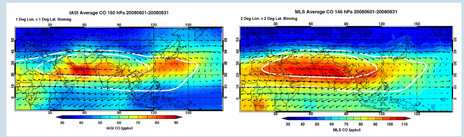

Figure 2: Left: MLS JJA 2008 average 146 hPa CO. Right: IASI JJA 2008 average 150 hPa CO. (Figures courtesy of Laura Pan.)

- One concern was the clear difference in morphology between the two datasets (note the differences in color scale).
- Unsurprisingly, perhaps, the main cause of this discrepancy appears to be related to the limited vertical resolution of IASI.
  - For example, the enhanced IASI CO in the western Pacific probably reflects an enhancement at lower altitudes.
- However, the enquiry did lead to some investigation related to the robustness of features seen in the MLS map.
- Let us examine how morphology in ozone can, through a potential error in ozone spectroscopy, lead to a spatially-dependent bias in CO.

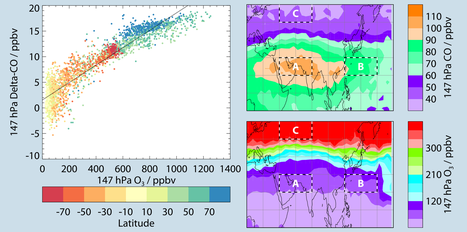

Figure 3: Left: Ozone spectroscopy-induced perturbation in 146 hPa CO (as in Figure 1) shown versus 146 hPa O_3. Right upper: JJA 2008 MLS CO at 146 hPa. Right lower: JJA 2008 MLS O_3 at 146 hPa. Selected regions on the maps are highlighted with yellow/black boxes.

- The left-hand plot in Figure 3 shows how the CO error induced by perturbed ozone spectroscopy relates to the ozone abundance (taken from the same simulations discussed previously).
- A simple linear fit (effectively parameterizing $\mathbf{H}_i$ in the discussion below) can be used to estimate the potential impact of ozone spectroscopy error on 146 hPa MLS JJA-average CO (maps in Figure 3).
- How much might CO gradients (e.g., box-to-box differences) be impacted by ozone spectroscopy error?
  - The A/B difference reduces from 18.8 to 17.6 ppbv CO, a ~6% change.
  - The A/C difference reduces from 61.3 to 55.5 ppbv CO, a ~10% change.

4 Formalizing the situation mathematically

Errors on individual measurements

The typical formalism for describing satellite retrievals is set up as follows:

- We define a “state vector”, $\mathbf{x}$, which describes the atmospheric information we’re seeking to estimate, such as ozone profiles, temperature, etc.
- We seek an estimate of $\mathbf{x}$ (typically referred to as $\hat{\mathbf{x}}$) based on a collection of radiance observations contained in the “measurement vector” $\mathbf{y}$.
- Given a perturbation in some parameter (e.g., spectroscopy), $\Delta\mathbf{b}$, the impact on the retrieved state is given by

$$\Delta\hat{\mathbf{x}} = \frac{\partial\hat{\mathbf{x}}}{\partial\mathbf{y}}\frac{\partial\mathbf{y}}{\partial\mathbf{b}}\Delta\mathbf{b} = \mathbf{G}\mathbf{K}_\mathbf{b}\Delta\mathbf{b} = \mathbf{H}\Delta\mathbf{b} \qquad (1)$$

(assuming linearity over $\Delta\mathbf{b}$), where

$\mathbf{G} = \partial\hat{\mathbf{x}}/\partial\mathbf{y}$, the “retrieval gain matrix”,
$\mathbf{K}_\mathbf{b} = \partial\mathbf{y}/\partial\mathbf{b}$, the “forward model parameter sensitivity matrix”, and
$\mathbf{H} = \mathbf{G}\mathbf{K}_\mathbf{b}$, the product of the two terms above, defined to simplify the algebra that follows.

- The issue we’re focused on originates from the dependence of $\mathbf{H}$ on $\mathbf{x}$.
  - Which derives from the dependence of $\mathbf{K}_\mathbf{b}$ and $\mathbf{K}$ ($= \partial\mathbf{y}/\partial\mathbf{x}$) on $\mathbf{x}$ in most cases.
  - As the $\mathbf{K}$ term appears dictates $\mathbf{G}$, the latter also typically depends on $\mathbf{x}$.
- This means that $\Delta\mathbf{x}$ is not constant for all states, even though $\Delta\mathbf{b}$ likely is.

Quantities derived from multiple measurements

- Now consider some new (vector or scalar) quantity $\mathbf{z}$, that is formed from an ensemble of N atmospheric states $\mathbf{x}_1 \ldots \mathbf{x}_N$.
  - $\mathbf{z}$ could be, for example, a long-term trend or a difference in the averages between two regions (e.g., inside and outside the polar vortex).
  - We’ll consider linear operations for now (sums, differences, averages, trends etc.), though more general representations are feasible.
- Let $\mathbf{z}$ thus be formed by:

$$\mathbf{z} = \begin{bmatrix}\mathbf{W}_1 & \mathbf{W}_2 & \cdots & \mathbf{W}_N\end{bmatrix}\begin{bmatrix}\mathbf{x}_1\\ \mathbf{x}_2\\ \vdots\\ \mathbf{x}_N\end{bmatrix}, \qquad (2)$$

where the twofold roles of the $(n_z \times n_z)$ $\mathbf{W}_i$ matrices are to:

1. Select/combine elements of interest from each $\mathbf{x}_i$ (e.g., compute the ozone column between 100 hPa and 10 hPa for each profile), and
2. modulate each $\mathbf{x}_i$ contribution to $\mathbf{z}$ to give the desired metric (e.g., $W_i \sim 1/N$ to give a regional average, or $1/N_{\text{north}}$ for northern hemisphere and $-1/N_{\text{south}}$ for southern hemisphere to give a NH/SH difference, etc.).

Errors in such quantities

- Asumming that individual retrievals are linear in $\Delta\mathbf{b}$ then the $\Delta\mathbf{b}$-induced error in $\hat{\mathbf{z}}$ is given by:

$$\Delta\mathbf{z} = \begin{bmatrix}\mathbf{W}_1 & \mathbf{W}_2 & \cdots & \mathbf{W}_N\end{bmatrix}\begin{bmatrix}\mathbf{H}_1\\ \mathbf{H}_2\\ \vdots\\ \mathbf{H}_N\end{bmatrix}\Delta\mathbf{b}. \qquad (3)$$

- The expectation of this term is given by:

$$\mathrm{E}(\Delta\mathbf{z}) = \sum_i \mathbf{W}_i\mathbf{H}_i\,\mathrm{E}(\Delta\mathbf{b}). \qquad (4)$$

- If we assume that our spectroscopy etc. is the “best” choice, then $\mathrm{E}(\Delta\mathbf{b}) = 0$, thus $\mathrm{E}(\Delta\mathbf{z}) = 0$.
- The covariance of $\Delta\mathbf{z}$ is given by:

$$\mathrm{cov}(\Delta\mathbf{z}) = \sum_i\sum_j \mathbf{W}_i\mathbf{H}_i\,\mathrm{cov}(\mathbf{b})\,\mathbf{H}_j^T\mathbf{W}_j^T. \qquad (5)$$

- This equation makes it clear why such errors are very challenging to handle.
- Firstly, there is the problem of communication.
  - Only the user knows they want in $\mathbf{z}$, and thus what each $\mathbf{W}_i$ is.
  - Whereas, only the instrument teams are in a position to provide the $\mathbf{H}_i$ matrices for each profile, and $\mathrm{cov}(\mathbf{b})$, needed to compute the inner term.
- Secondly, these are complex and time-consuming calculations.
  - Full computation of every $\mathbf{H}_i$ for each profile and considering all possible error sources would be a very costly process.
  - The calculation in Equation 5, which users would need to perform, isn’t exactly trivial either.
- In principal, the data providers could produce an $\mathbf{H}_i$ matrix for each retrieved state, along with the $\mathrm{cov}(\mathbf{b})$ matrix.
- Some of parameterization of $\mathbf{H}$ as a function of $\mathbf{x}$ may be more viable.
- Such approaches are under investigation for MLS.

© 2019. All rights reserved.