Atmos. Meas. Tech. Discuss., 6, C651–C654, 2013
www.atmos-meas-tech-discuss.net/6/C651/2013/

Atmospheric Measurement Techniques Discussions

Interactive comment on “Dimensionality reduction in Bayesian estimation algorithms” *by* G. W. Petty

Anonymous Referee #4

Received and published: 18 April 2013

Corrections: Section 1.2 (pg 2330, Line 19): "... provided only that the requisites PDFs are known." -> requisite

Conclusions/Discussion (Page 2342): "To mitigate the problem of dimensionality in Bayesian retrievals, we described an algorithm for objectively distilling the relevant information content from N channels into a smaller number (M) pseudochannels while also regularizing the background (geophysical plus instrument) noise component. In the present demonstration, M = 3 and N = 1. In the application of this method to TMI data described by Petty and Li (2013), M = 9 and N = 3."

Shouldn't M < N ? This is confusing. Also, it sounds like you used M=3 in the present demonstration, but M = 1 ... I think M and N need to be switched here.

General Discussion:

Section 1.2

Where do these "candidate solutions" come from? At some point, somewhere in a given retrieval algorithm, there is a modeled relationship between the radiances and the geophysical parameter(s) one is interested in. Whether it's a radar-derived precipitation rate (e.g., a Z-R relationship) co-located with radiance observations, or a CRM database of profiles with forward modeled radiances.

Consequently, the model bias that one is so eager to disconnect themselves from gets buried somewhere or, worse, over-constrains the retrieval problem by under-populating the solution space. Given a (theoretical) perfectly co-located and beam-matched radar observation for each feedhorn on a given sensor, one is still wholly limited by both (i) the sensitivity range (and instrument error) of the radar; and, (ii) the physical relationships between geophysical parameters (gas, precipitation, surface, multiple-scattering, clutter noise, etc.) and the measured reflectivities. Given this idealized scenario, one now has a basis for a "pretty good" retrieval algorithm, _but only for the cases that the radar(s) could observe_. In the case of TRMM, for example, this would mean a very large percentage of precipitation occurrence (e.g., light precipitation) would never be retrieved with skill. Could one improve upon this by performing a similar "dimensionality reduction" on the radar observations (or whatever source observations)? The reflectivity at each range gate is a measurement, although not truly independent of the preceding ones due to path-integrated attenuation and, possibly, multiple-scattering effects (see Battaglia, for example).

Section 3.1 (pg 2335, Equation 3): This example may strengthen your argument: I was playing around with a simple example of equation 3, and noticed that if one simply increases the number of channels – without adjusting sigma_i – "s" also naturally increases. So if one adds additional radiometer channels to the typical "Bayesian" retrieval, the weight (w=exp(-s)) rapidly decreases. The act of adding a single channel will, because of the threshold w > 0.01, will result in potentially worse retrieval quality.

Figure 4: I realize this is still "background" stuff, but what's the deal with the near-zero retrievals when the true rates are as high as 3? It would be interesting to have a color-coding (or shading) to indicate what the sigma value is for each point. Are there cases where the retrieval is near the 1:1 line, but the sigma values are really large – i.e., a good match for the wrong reasons? Figure 5, 6, & 7: It's hard to tell what the actual retrieval skill is on these plots, particularly at low precipitation rates.

Figure 7: Same sentiment as my comment about figure 4.

Other Comments and Recommendations

For a very long time, the community has been recycling poor (statistically) "matching" algorithms, and, we keep putting lipstick on the pig by improving the various bits and pieces without changing the actual framework. Even worse, perhaps, is that the retrievals obtained from these algorithms get propagated into various climate datasets, degrading the potential knowledge obtainable from past and present precipitation retrievals.

The present method here, while not necessarily mathematically new, presents an important (and easy to implement) approach to improving upon this long-standing problem. Future retrieval approaches would be wise to utilize the method presented here to improve upon the dimensionality problem, and isolate the variables to be retrieved – or, alternatively, determine those that cannot be isolated.

A few things I would have liked to seen in this paper: (1) Application to real observations (I realize space considerations are an issue, will this be a subject of a future publication?)

(2) Additional eigenvalues (M > 1) and a physical relationship between values of the M-th eigenvalue and precipitation rate (or whatever variable it's actually sensitive to, that's never clearly stated .. despite matching to precipitation rate in the training/val database. It could be that, for example, cloud ice is strongly correlated with precipitation rate,

and the first eigenvalue happens to be the sensitivity to that. Which is "okay" in the sense that it ultimately gives you what you want, but that limits one to a certain set of microphysical processes in retrievals – i.e., you might miss warm rain altogether).

(3) Retrieval skill. Visually it's easy to discern that at high precipitation rates, the proposed algorithm performs well. At low precipitation rates (what GPM is purportedly designed to retrieve), it's difficult to discern on the figures how well or poorly it is doing. A log scale in precip rate would be an easy step to accommodate this visual inspection, a slightly more involved step would be to assign a skill to the retrieval or clearly denote variance in a different way. I don't have an immediate good idea about how to communicate that clearly.

(4) Dealing with extreme and/or uncommon events. It was mentioned in the beginning, but I didn't notice any additional discussion of this important aspect of retrievals.

(5) A final comment about non-linearity – there's very little discussion of non-linear relationships between the transformed TBs and the precipitation rate. It appears that you are arguing that by reducing the off-diagonal elements of the covariance matrix, that you are mitigating the non-linear response. It's not clear to me that this is what is occurring. Could you discuss the effects of non-linear relationships in the present approach?

Disposition: Accept with minor revisions.

Interactive comment on Atmos. Meas. Tech. Discuss., 6, 2327, 2013.