The benefits of dimensionality reduction in Bayesian retrievals of rain rate from passive microwave observations. Application to TMI observations over ocean

Grant W. Petty and Ke Li

Atmospheric and Oceanic Sciences Department
University of Wisconsin-Madison, 1225 W. Dayton St., Madison, Wisconsin 53706

gwpetty@wisc.edu

1. Background

- The PMM Passive Microwave Algorithm Working Group has settled on a Bayesian framework for retrieval of global precipitation from the GPM Microwave Imager (GMI).
- Method entails querying a globally representative data base of matched GMI observations and independently-determined rain rates and structures provided by the Dual-frequency Precipitation Radar.
- Prototype is under development at CSU using data from TRMM Microwave Imager (TMI) and Precipitation Radar (PR).
- In its current form for TMI, algorithm attempts to finds matches to all 9 channels simultaneously, albeit with varying channel weights.

2. Issues

- Data base entries are "raw" TBs and thus encompass variability due to a variety of sources in addition to precipitation.
- High-dimensional search space is difficult to populate with a sufficiently dense, diverse and statistically representative set of observations.
- Large tolerances must sometimes be allowed to ensure a reasonable number of matches.
- For many pixels, retrieval will be determined by a very small number of loosely matching data base entries.

Fundamentally, reliance on a high-dimensional solution data base implies

- Need for a very large data base
- Long search times
- Non-robust statistics for rarer combinations of channel TBs
- Need to account for highly correlated geophysical noise between channels

Above problems are greatly exacerbated over land

- Heterogeneous background types
- Poorer signal-to-noise ratio
- Much smaller training sample for given surface classification

3. Objective

Demonstrate that the dimensionality of Bayesian retrieval poblem can be radically reduced (from 9 to only 2 or 3) without impairing retrieval performance.

Although ultimate goal is to adapt these methods to over-land retrievals, the benefits are illustrated here in the context of over-ocean retrievals.

Why?

- Clearest demonstration with fewest complicating variables.
- Ability to evaluate results in context of past performance by other established over-water algorithms (e.g., GPROF).
- Signal-to-noise ratios and the precipitation information content of all TMI channels presumed to be superior over open water. Therefore, the consequences of inadvertently throwing away "good" information should be more readily detectable.

Benefits:

- Large contributing sample size ($10^2 - 10^6$) for most retrievals despite moderately tight tolerances.
- Explicit allowance for background noise budget in setting tolerance.
- In addition to a single rain rate estimate for each pixel, robust PDFs (e.g., percent likelihood of $R > R_0$).
- Graceful handling of rare non-matches.
- Insight into "true" useful information content of passive microwave channels with respect to retrievable rain cloud properties.
- Data base reduces to small (1.5 MB!) pre-computed lookup table
- With extremely little R&D (to date), comparable global performance to current GPROF.

4. Data

- Matched TMI brightness temperatures and PR (2A25) surface rain rates
- (De-)convolved to 19 GHz channel resolution
- One calendar year (2002) global data
- ERA-Interim analysis 6-hourly SST

5. Procedures

Stage 1:

- Transform raw TBs: $\mathbf{x} = \log(T_S - \mathbf{T_B})$
- Compute global mean $\langle \mathbf{x} \rangle$ and covariance $\mathbf{S}_x$ for all **non-precipitating** scenes.
- Compute eigenvectors $\mathbf{E}_x$, eigenvalues $\mathbf{\Lambda}_x$ of $\mathbf{S}_x$
- Define transformed channels $y_i = [(\mathbf{x} - \langle \mathbf{x} \rangle)^{\mathsf{T}} \mathbf{E}_x]_i / \lambda_{x,i}^{1/2}$
- By design, $\langle \mathbf{y} \rangle = 0$ and $\mathbf{S}_y = \mathbf{I}$ *outside of precipitation*

Summary: The 9 transformed channels $\mathbf{y}$ retain *all* information found in the original T_B, BUT they have been completely decoupled AND have they have an uncorrelated total noise variance (instrument plus geophysical) of unity *outside of precipitation.*

Stage 2:

- We now apply the same transformation to precipitating scenes with $R > 1$ mm/hr. The variance in each transformed channel y_i is now considerable larger than unity. The added variance is due solely to the influence of precipitation.
- For these raining pixels, we compute $\mathbf{S}_{y,r} \equiv \langle \mathbf{y}\mathbf{y}^{\mathsf{T}} \rangle$, with eigenvectors $\mathbf{E}_{y,r}$ and eigenvalues $\mathbf{\Lambda}_{y,r}$.
- We define the precipitation *pseudochannels* $\mathbf{z} \equiv \mathbf{y}^{\mathsf{T}} \mathbf{E}_{y,r}$.
- Outside of precipitation, these 9 pseudochannels still have zero mean and unit uncorrelated variance.

For precipitating scenes, only the first three (z_1, z_2, z_3) have variance $\sigma^2_{z,i}$ significantly greater than unity. We therefore conclude that *these contain virtually all extractable information concerning the properties of the precipitation in the the scene.*

Stage 3:

- Apply transformation $T_B \rightarrow \mathbf{z}$ for all ocean scenes ($N = 1.25 \times 10^8$).
- Odd scenes used for retrieval database ; even scenes for validation.
- Aggregate database into 4-D array.
- Retrieval consists solely of indexing into array with three pseudochannels z_i ($\Delta z \approx 1$) and SST (ΔSST$= 5$ K).
- No other ancillary data required.

6. Results

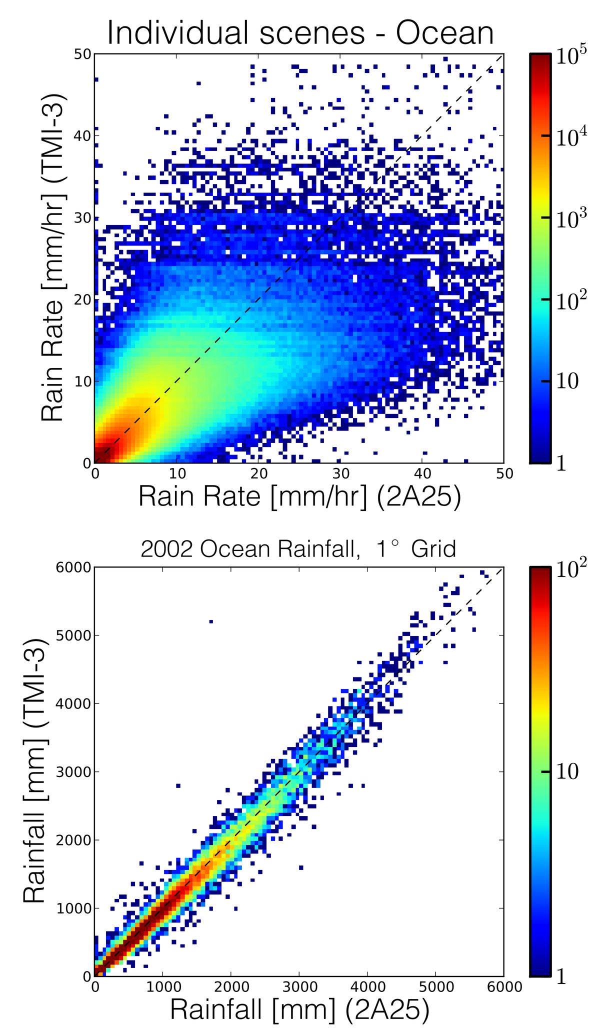

Figure 1: *Validation of TMI retrievals against independent 2A25 (PR) rain rates over one calendar year (2002). Top: Pixel-by-pixel comparison. Bottom: Gridded annual totals at 1° resolution.*

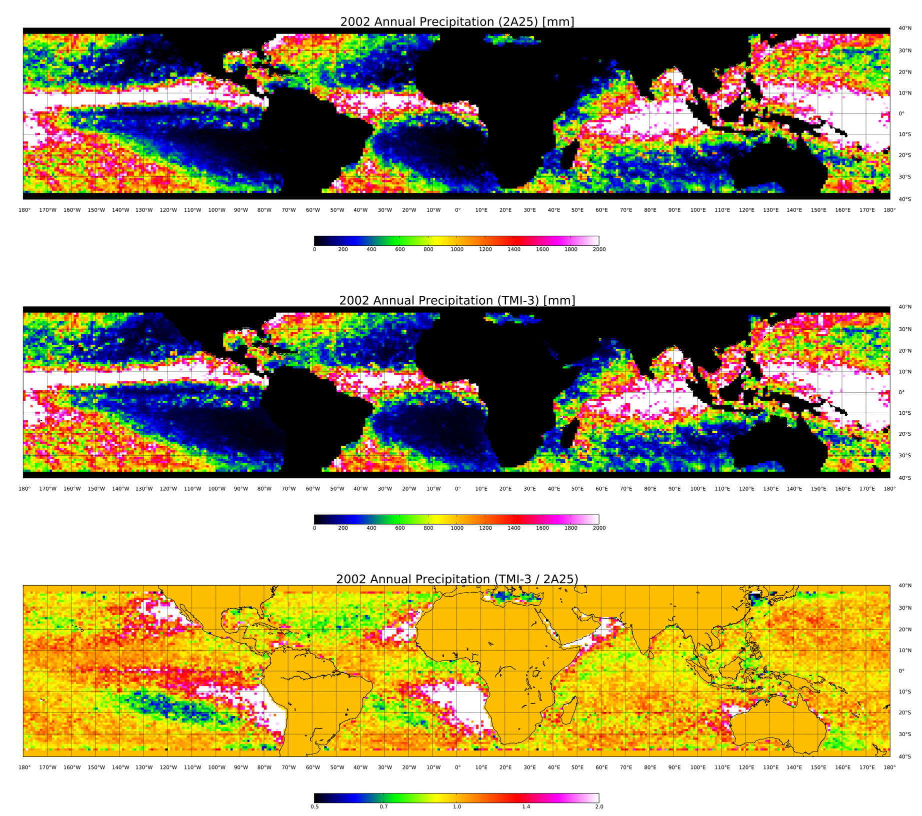

Figure 2: *Maps of annual total precipitation from PR (top) and TMI (middle). Also shown is the ratio (bottom). PR data used in this comparison were from the independent data set.*

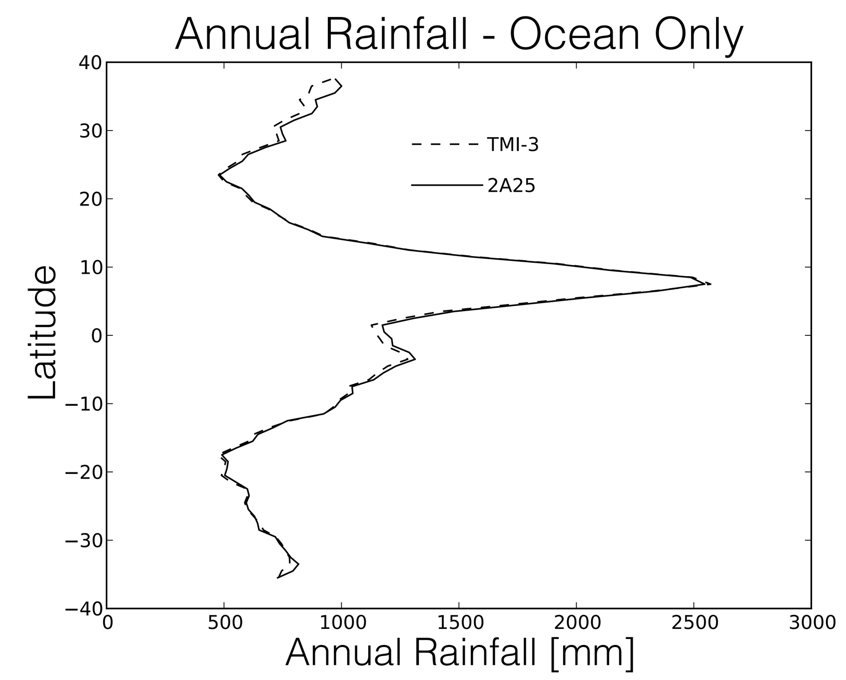

Figure 3: *Zonally averaged PR and TMI annual total ocean precipitation for 2002.*

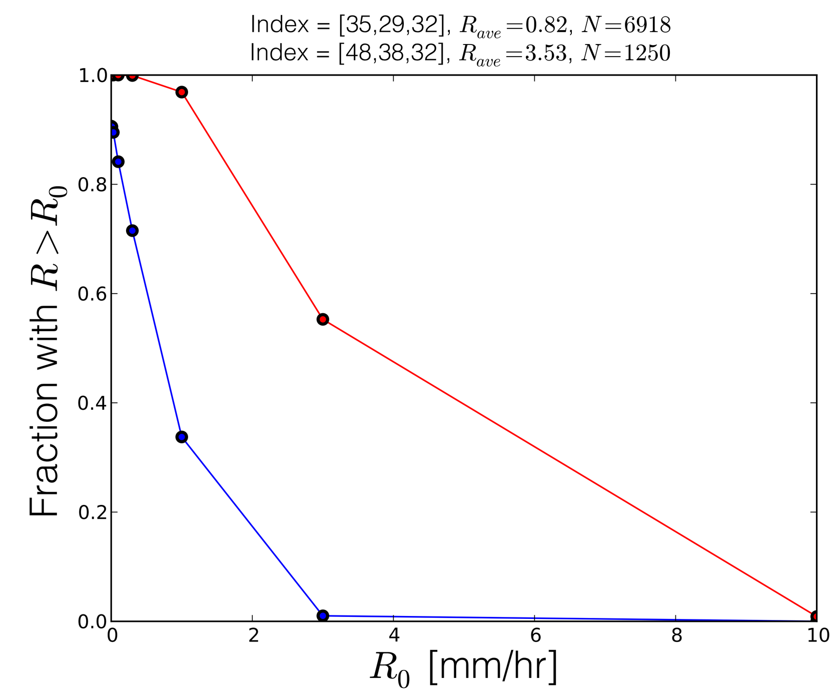

Figure 4: *Because of the typically very high density of samples in the low-dimensional database, it is possible to extract not only a mean (or expected) rain rate for a scene but also robust statistics concerning probability of precipitation exceeding a given intensity.*

7. Acknowledgments

C. Kummerow, D. Randel, and S. Rungerud provided deconvolution coefficients and access to ERA-Interim analysis. This work is supported by NASA Grant NNX10AGAH69G