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

Grant Petty and Ke Li

University of Wisconsin-Madison

8 November 2011

- 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.

Issues I

- 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 more 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.

Issues II

Fundamentally, reliance on a high-dimensional solution data base using raw input channels implies

- Very large data base
- Long search times
- Non-robust statistics for infrequent combinations of channel TBs
- A need to rigorously account for highly correlated geophysical noise between channels.

Issues III

Above problems are greatly exacerbated over land

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

Objective I

- Demonstrate that the dimensionality of Bayesian retrieval problem can be radically reduced ($9 \rightarrow 3$) leading to robust, high-quality retrievals using very compact data base and efficient data extraction.
- Method can be readily incorporated into existing physical and statistical retrieval frameworks – e.g., current PMW algorithm. Simply replace large number of actual channels with smaller number of optimally defined pseudochannels.
- Motivated by need for "S0" over-land retrievals, the methods and benefits are demonstrated here most immediately and clearly for over-ocean retrievals (single surface class).
- *Hot off the press:* Preliminary land results too!

Objective II

Claimed benefits:

- Large contributing sample size ($10^2 - 10^6$) for most retrievals despite moderately tight tolerances $\delta \sim 1$.
- Explicit reference to 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.
- Database reduces to small pre-computed lookup table.
- Respectable global performance.

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 and TCWV.

Procedures I

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 **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.*

Procedures II

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 considerably larger than unity. *The added variance is due solely to the influence of precipitation.*
- For these raining pixels, we compute $\mathbf{S}_{y,\mathrm{r}} \equiv \langle \mathbf{y}\mathbf{y}^{\mathsf{T}} \rangle$, with eigenvectors $\mathbf{E}_{y,\mathrm{r}}$ and eigenvalues $\mathbf{\Lambda}_{y,\mathrm{r}}$.
- We define the precipitation *pseudochannels* $\mathbf{z} \equiv \mathbf{y}^{\mathsf{T}}\mathbf{E}_{y,\mathrm{r}}$.
- Outside of precipitation, these 9 pseudochannels still have zero mean and unit uncorrelated variance.

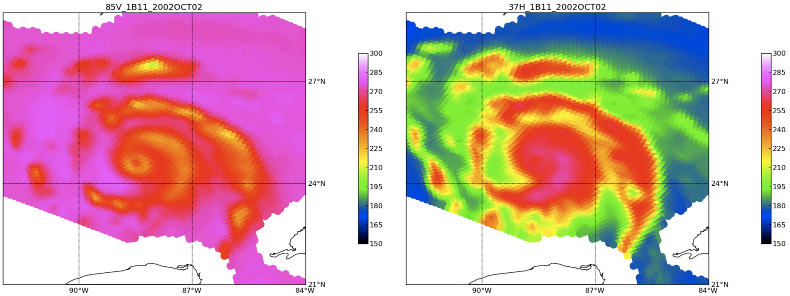
For precipitating scenes, only the first three (z_1, z_2, z_3) have variance $\sigma^2_{z,i}$ significantly greater than unity. We therefore conclude that *these contain virtually all extractable information concerning the properties of the precipitation in the the scene.*

Procedures III

Stage 3:

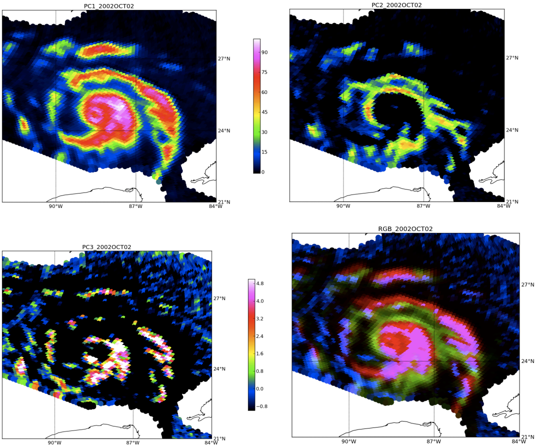
- Apply transformation $T_B \rightarrow \mathbf{z}$ for all ocean scenes ($N = 1.25 \times 10^8$).
- Odd scenes used for retrieval database ; even scenes for validation.
- Aggregate retrieval database into 5-D array: 3 pseudochannels ($\Delta z \approx 1$) plus SST ($\Delta SST = 5$ K) and TCW ($\Delta TCW = 10$ mm).
- Retrieval consists solely of indexing into array and extracting precomputed statistics.
- Extremely fast: One year's worth of global TMI data processed in 2–3 hours.

Example I



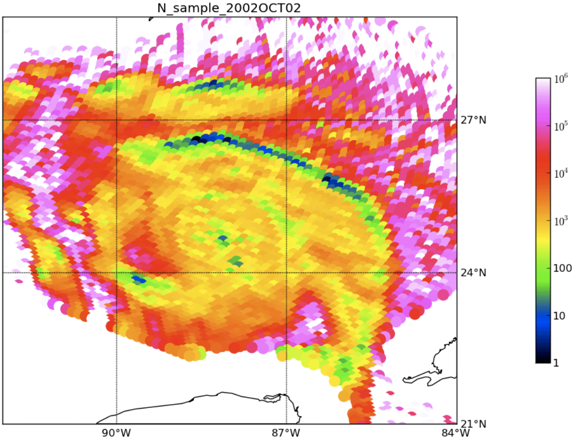
Example: Hurricane Lili, October 2, 2002. T85V (left) and T37H (right).

Example II



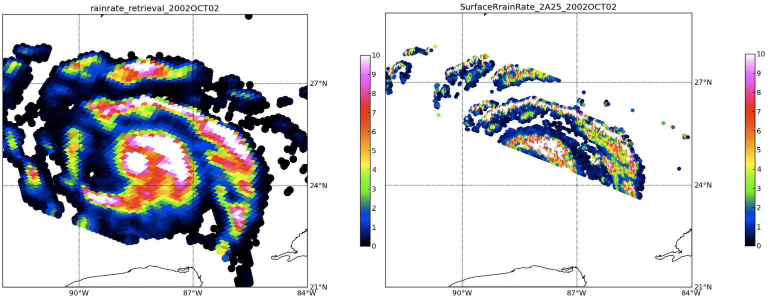
Three precipitation pseudochannels: PC1 (upper left), PC2 (upper right), PC3 (lower left), RGB composite (lower right)

Example III



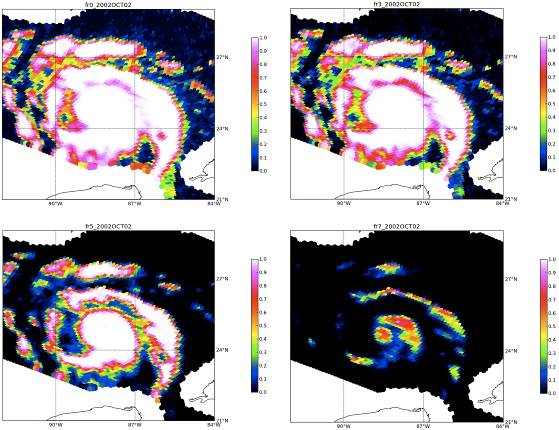
Training sample size contributing to each pixel's retrieval results.

Example IV



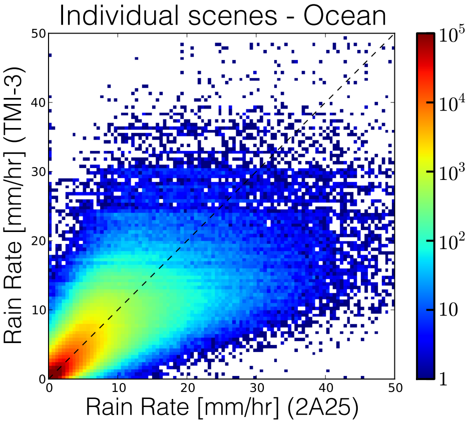
Instantaneous retrieval comparison. Expected value of rain rate from TMI retrieval (left), “true” rain rate from PR (right).

Example V



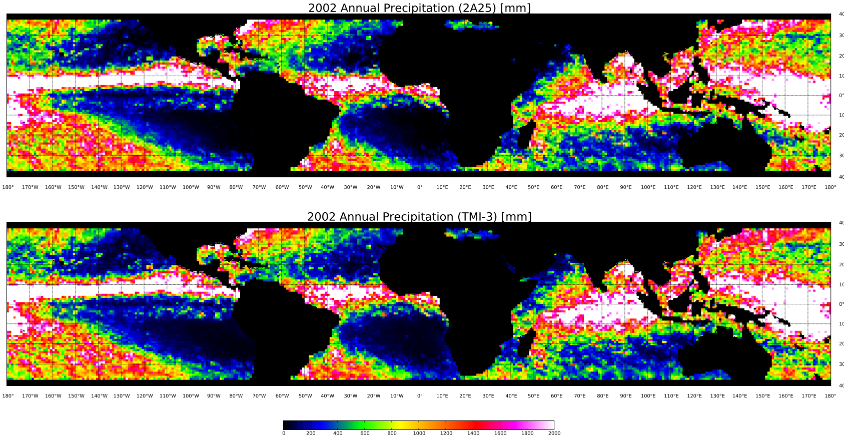
Bayesian probability of rain rate exceeding a specified threshold: $R > 0$ (upper left), $R > 0.1$ (upper right), $R > 1$ (lower left), $R > 10$ (lower right)

Results (Ocean) I



Pixel-by-pixel comparisons between 2A25 and TMI retrievals over ocean (independent data only) for 2002.

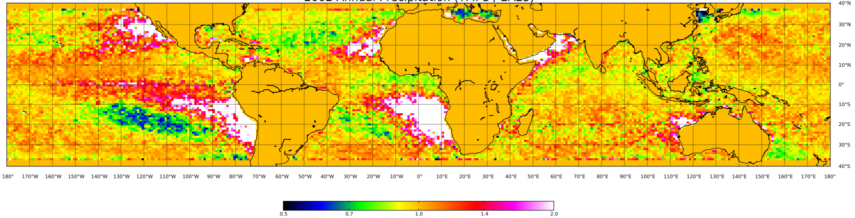
Results (Ocean) II



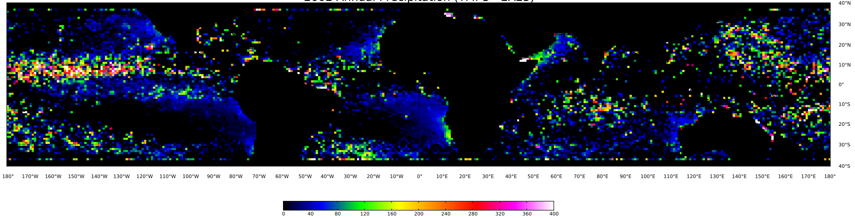
Annual total precipitation (independent data only) for 2002 from 2A25 (top) and from TMI Bayesian algorithm (bottom).

Results (Ocean) III

2002 Annual Precipitation (TMI-3 / 2A25)

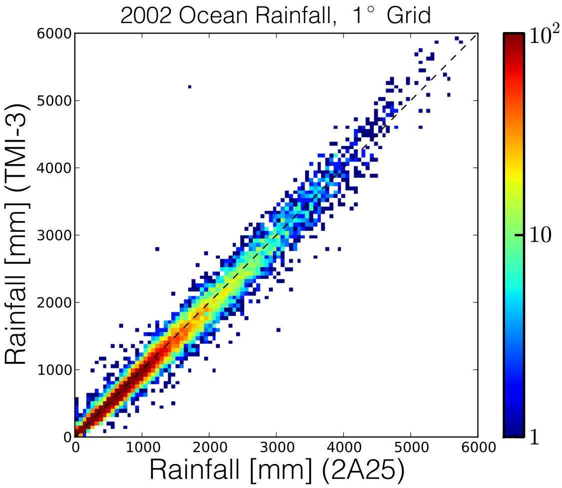


2002 Annual Precipitation (TMI-3 - 2A25)

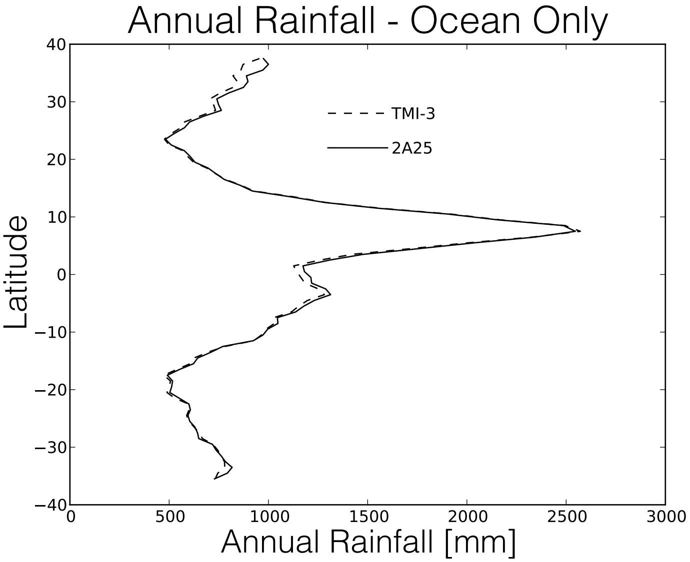


Ratio (top) and difference (bottom) of TMI Bayesian retrieval with respect to 2A25 “truth” for 2002.

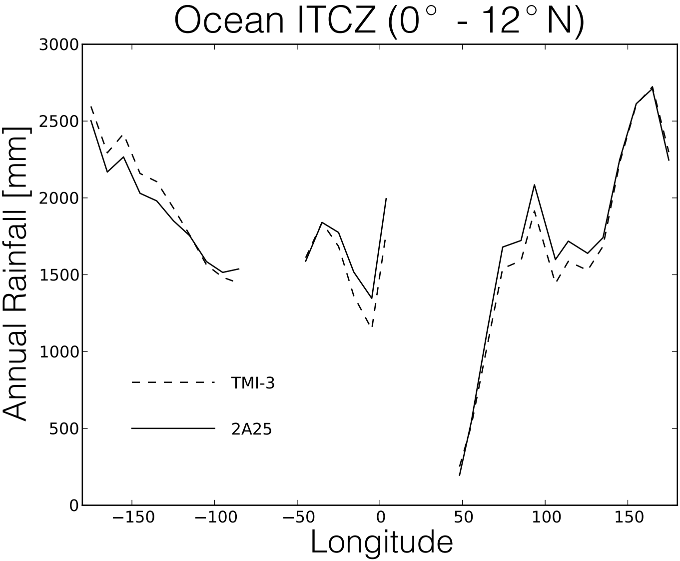
Results (Ocean) IV



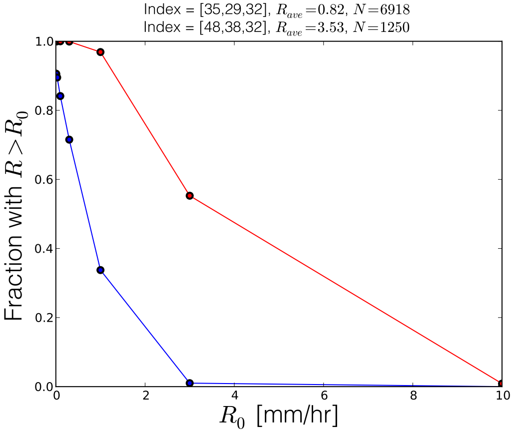
Results (Ocean) V



Results (Ocean) VI



Results (Ocean) VII



Two examples of the ability to infer not only the expected rain rate for a specific scene but also the *probability* of rain rate exceeding specified thresholds.

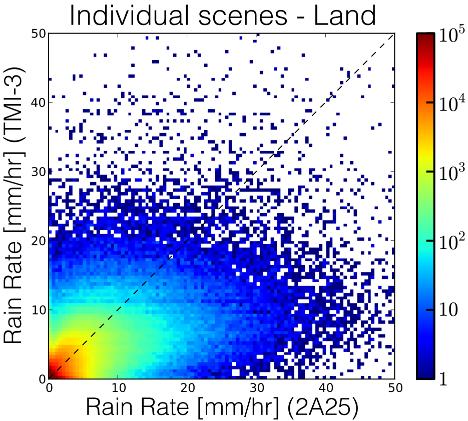
Some key facts

- Although three pseudochannels are utilized, the third one adds relatively little value. *Retrieval performance only slightly degraded using only two.*
- Ocean data base consists of 176K entries occupying only 19MB but representing 114 million scenes in the training set – *an average of 650 training scenes per data base entry.*
- In the event of no match, graceful recovery (i.e., valid Bayesian retrieval with PDF) by tossing 3d, and if necessary, 2nd pseudochannel.

Preliminary Experiments Over Land

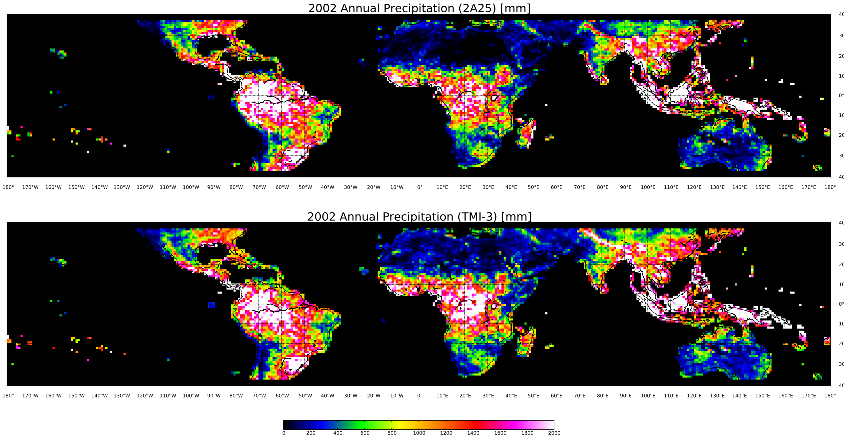
- Throw all non-ocean matchups into one single class.
  - vegetated
  - desert
  - coastal
  - inland water
- Exclude (for now) $T_s < 275$ K (potential snow surfaces)
- Redefine transformed brightness temperatures: $\mathbf{x} = \mathbf{T_B}/T_s$
- Three global "pseudochannels" again used to isolate precipitation signature against variable land background.
- Much poorer signal-to-noise ratio than over water.

Results (Land) I



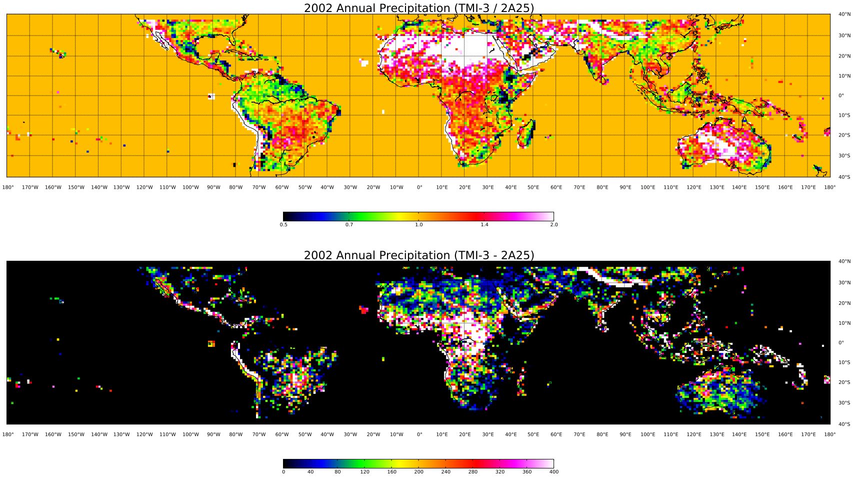
Pixel-by-pixel comparisons between 2A25 and TMI retrievals over ocean (independent data only) for 2002.

Results (Land) II



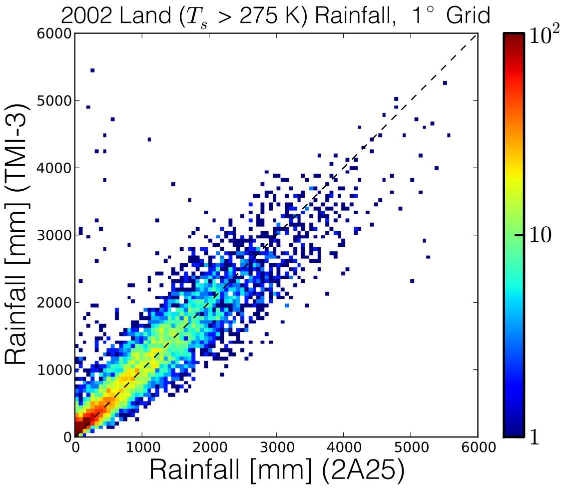
Annual total precipitation (independent data only) for 2002 from 2A25 (top) and from TMI Bayesian algorithm (bottom).

Results (Land) III

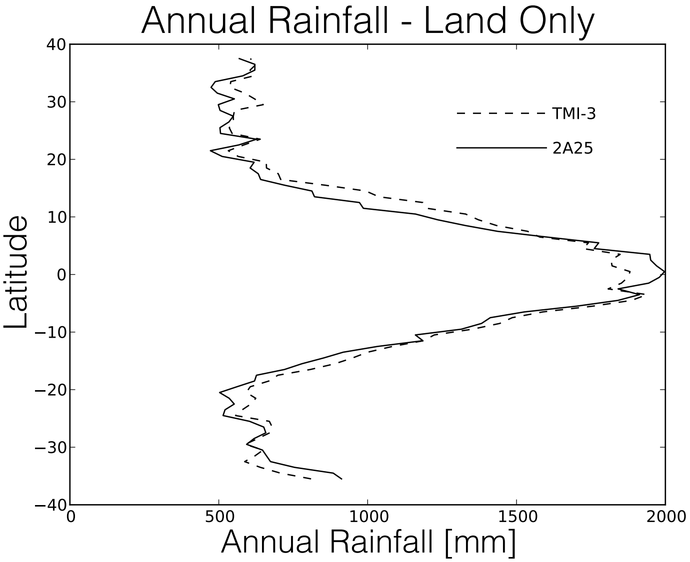


Ratio (top) and difference (bottom) of TMI Bayesian retrieval with respect to 2A25 “truth” for 2002.

Results (Land) IV



Results (Land) V



Next Steps

- Current ocean version of PMW algorithm can be easily modified to utilize 3 pseudochannels instead of 9 brightness temperatures.
  - Should greatly improve number of valid matches while also using more rigorous match criteria.
  - Likely improvement in light rain detection skill.
  - More robust posterior PDFs of rain properties.
- With stratification of land regions into a small number of self-similar classes, derive pseudochannels and data bases for each class.
  - Land surfaces with $T_s < 275$ K will be a separate class.
  - $\rightarrow$ "S0" retrieval scheme for "unknown" surface emissivities.

Acknowledgements

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