

This discussion paper is/has been under review for the journal Atmospheric Measurement Techniques (AMT). Please refer to the corresponding final paper in AMT if available.

Dimensionality reduction in Bayesian estimation algorithms

G. W. Petty

Department of Atmospheric and Oceanic Sciences, University of Wisconsin-Madison, USA

Received: 15 February 2013 – Accepted: 15 February 2013 – Published: 4 March 2013

Correspondence to: G. W. Petty (gwpetty@wisc.edu)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Discussion Paper | Discussion Paper

Discussion Paper | Discussion Paper

Discussion Paper | Discussion Paper

Dimensionality reduction in Bayesian estimation algorithms

G. W. Petty

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Abstract

An idealized synthetic database loosely resembling 3-channel passive microwave observations of precipitation against a variable background is employed to examine the performance of a conventional Bayesian retrieval algorithm. For this dataset, algorithm

5 performance is found to be poor owing to an irreconcilable conflict between the need to find matches in the dependent database versus the need to exclude inappropriate matches. It is argued that the likelihood of such conflicts increases sharply with the dimensionality of the observation space of real satellite sensors, which may utilize 9 to 13 channels to retrieve precipitation, for example.

10 An objective method is described for distilling the relevant information content from N real channels into a much smaller number (M) of pseudochannels while also regularizing the background (geophysical plus instrument) noise component. The pseudochannels are linear combinations of the original N channels obtained via a two-stage principal component analysis of the dependent dataset. Bayesian retrievals based on 15 a single pseudochannel applied to the independent dataset yield striking improvements in overall performance.

The differences between the conventional Bayesian retrieval and reduced-dimensional Bayesian retrieval suggest that a major potential problem with conventional multichannel retrievals – whether Bayesian or not – lies in the common but often inappropriate assumption of diagonal error covariance. The dimensional reduction technique described herein avoids this problem by, in effect, recasting the retrieval problem in a coordinate system in which the desired covariance is lower-dimensional, diagonal, and unit magnitude.

Dimensionality reduction in Bayesian estimation algorithms

G. W. Petty

[Title Page](#)

[Abstract](#)

[Introduction](#)

[Conclusions](#)

[References](#)

[Tables](#)

[Figures](#)

◀

▶

[Back](#)

[Close](#)

[Full Screen / Esc](#)

[Printer-friendly Version](#)

[Interactive Discussion](#)

1 Introduction

Satellite remote sensing entails the indirect determination of a property, or set of properties, of the environment based on measurements of top-of-the-atmosphere radiances at appropriate wavelengths. Conceptual frameworks for undertaking satellite retrievals 5 range from simple ad hoc methods to the iterative inversion of complex physical models of observed radiances.

For remote sensing problems involving highly non-linear and/or difficult-to-model relationships between observation vectors and environmental states, it is increasingly common to rely on so-called Bayesian estimation methods. One of the principal 10 areas of application of Bayesian methods (but by no means the only one) has been in the area of precipitation retrieval from passive and/or active microwave observations (Evans et al., 1995; Olson et al., 1996, 2006; Haddad et al., 1997; Marzano et al., 1999; Bauer et al., 2001; Kummerow et al., 2001, 2011; Tassa et al., 2003; Di Michele et al., 2005; Grecu and Olson, 2006; Chiu and Petty, 2006; Viltard et al., 2006; Seo 15 et al., 2008).

This paper has two purposes: (a) to draw attention to certain practical limitations of Bayesian algorithms as typically implemented, and (b) to describe and demonstrate an objective method of dimensional reduction that substantially improves the robustness of Bayesian retrievals in certain remote sensing applications.

20 The Bayesian methodology is examined here in the context of idealized retrievals of surface precipitation rate. However, the issues raised, and their proposed solution, should have considerably broader applicability.

1.1 Bayesian estimation

25 Bayesian retrieval algorithms purport to obtain estimates of an environmental variable (e.g. rain rate) via application of Bayes' Theorem (Bayes and Price, 1763). In the present context, Bayes' Theorem states that the posterior probability distribution function (PDF) of a desired variable R conditioned on an observation vector x is given

AMTD

6, 2327–2352, 2013

Dimensionality reduction in Bayesian estimation algorithms

G. W. Petty

[Title Page](#)

[Abstract](#)

[Introduction](#)

[Conclusions](#)

[References](#)

[Tables](#)

[Figures](#)

◀

▶

◀

▶

[Back](#)

[Close](#)

[Full Screen / Esc](#)

[Printer-friendly Version](#)

[Interactive Discussion](#)



by

$$P(R|x) \propto P(x|R)P(R) \quad (1)$$

where $P(R)$ is the unconditional (prior) probability density function (PDF) of the scalar variable R to be estimated and $P(x|R)$ is the multidimensional PDF of the observation vector x conditioned on a specific value of R .

With only one known exception (Chiu and Petty, 2006), the prior joint and marginal PDFs are represented not as the continuous functions implied by Eq. (1) but rather via a large database of candidate solutions with associated observed or modeled multi-channel radiances. This variation has been aptly called a Bayesian Monte Carlo (BMC) method (L'Ecuyer and Stephens, 2002), although that more precise terminology does not seem to have achieved wider usage.

Moreover, while a true Bayes Theorem-based retrieval should in principle be able to yield a complete posterior PDF of R as demonstrated by Chiu and Petty (2006), it is typical in practice to extract only an expectation value based on a weighted average of the small set of discrete solution vectors in the database that approximately match the observations.

1.2 Practical limitations

Bayes Theorem has the advantage of providing a rigorous and complete statistical basis for optimal satellite retrievals, provided only that the requisites PDFs are known.

However, this advantage can only be fully realized only under fairly restrictive conditions:

- The prior joint distribution of environmental variables must be well-characterized over the full spectrum of possibilities and with adequate sampling density relative to the assumed observation error (L'Ecuyer and Stephens, 2002).
- The database must be small enough to be efficiently searchable, a requirement that stands in conflict with the previous one.

[Title Page](#)

[Abstract](#)

[Introduction](#)

[Conclusions](#)

[References](#)

[Tables](#)

[Figures](#)

◀

▶

[Back](#)

[Close](#)

[Full Screen / Esc](#)

[Printer-friendly Version](#)

[Interactive Discussion](#)

Dimensionality reduction in Bayesian estimation algorithms

G. W. Petty

10 This paper is motivated by the observation that all of the above challenges increase exponentially as the dimensionality of the search space increases. For example, imagine that a mere 10^4 solution vectors evenly distributed throughout a 3-dimensional observation space is minimally adequate to characterize the prior distribution of environment states and their associated observation vectors. Depending on interchannel and intervariable correlations, up to 10^{12} candidate solutions might be required to achieve 15 comparable density when Bayesian retrievals are directly based on the radiance observations of, say, a nine-channel instrument like the Tropical Rainfall Measuring Mission (TRMM) Microwave Imager (TMI; Kummerow et al., 1998).

Even if the overall density seems adequate, infrequent combinations of environmental variables – e.g. those associated with a severe storm or hurricane – will still tend to appear as outliers in an inadequately populated corner of channel space, in which case either no suitable match may be found at all or else the match sample may be small and potentially nonrepresentative.

Moreover, it usually proves difficult to confidently specify the optimal match or weighting criteria in a high-dimensional space. Formally, one usually specifies an error covariance that serves as the basis for assessing consistency between an observation vector x and a candidate solution in the database. In practice, the full covariance is rarely known, and only an assumed per-channel error variance is usually specified. This is the approach taken by the current version of the Goddard Profiling (GPROF)

Title Page	
Abstract	Introduction
Conclusions	References
Tables	Figures
◀	▶
◀	▶
Back	Close
Full Screen / Esc	

Printer-friendly Version



algorithm for TRMM (Kummerow et al., 2011), for example, among many other retrieval and assimilation schemes.

Note further that while the formal error covariance follows from an analysis of model and/or instrument error alone, this specification is only useful if the database is densely populated relative to that expected error. When the density is low, as is especially the case for less common scenes, one must often either arbitrarily expand the the search neighborhood until one or more nominal matches are found or else flag the retrieval for that observation as “missing” owing to a failure to find matches within the prescribed tolerance.

Finally, while the neighborhood that constitutes a match in observation space should in general be an arbitrarily oriented hyperellipsoid with appropriate major and minor axes, as specified by the true error covariance matrix, the convenient but arbitrary substitution of a diagonal covariance matrix automatically – and usually inappropriately – implies an ellipsoid or spheroid with principal axes perfectly aligned with channel coordinate axes. The retrieval algorithm may thus include inappropriate candidate solutions and/or exclude appropriate ones, leading to significant retrieval biases under some conditions, especially when the correct solutions lie in close proximity to incorrect solutions. As will be demonstrated below, the inappropriate assumption of diagonal covariance can severely degrade retrieval performance.

20 1.3 Objectives

In this paper, we demonstrate some important practical limitations of the Bayesian method as typically applied to satellite retrievals. Rather than use real satellite data, we construct an idealized synthetic database comprising only three simulated “channels” and associated with a single scene variable, “rain rate.” These channels are subject to considerable variability due to prescribed background noise. As is characteristic of passive microwave observations of precipitation over land, the signal due to “rain rate” is intentionally weak in absolute magnitude relative to the background noise but with a spectral component that is distinct from the background variability.

Dimensionality reduction in Bayesian estimation algorithms

G. W. Petty

[Title Page](#)

[Abstract](#)

[Introduction](#)

[Conclusions](#)

[References](#)

[Tables](#)

[Figures](#)

◀

▶

◀

▶

[Back](#)

[Close](#)

[Full Screen / Esc](#)

[Printer-friendly Version](#)

[Interactive Discussion](#)

We first show that conventional Bayesian retrievals in 3-channel space can lead to nearly useless estimates owing to match failures and/or over-averaging. We then demonstrate a technique for reducing the dimensionality of the database and for obtaining significantly more robust retrievals from the same data and using the same 5 Bayesian framework. In effect, we derive a smaller number of “pseudochannels” – i.e. linear transformations of the original channels or simple functions thereof – that retain most of the desired signal while rejecting a large part of the background noise. What remains of the background noise is decorrelated and scaled to unit variance.

Note that while the benefits of appropriate dimensional reduction should hold in general for any database-type retrieval problem, the particular dimensional reduction algorithm described below is most directly applicable to semi-continuous variables like rain rate or cloud liquid water path for which admissible values are either exactly zero or 10 positive.

2 Synthetic database

15 A Gaussian pseudorandom number generator was used to create 20 000 vectors of three-channel “background brightness temperatures” with prescribed mean $\bar{x} = (220, 240, 250)$ and covariance

$$\mathbf{S} = \begin{pmatrix} 506 & 81 & -205 \\ 81 & 173 & 140 \\ -205 & 140 & 269 \end{pmatrix}. \quad (2)$$

These statistical properties are arbitrary apart from the desire that the synthetic data 20 lie in an oblique 3-D plane with added uncorrelated random noise having unit variance.

For 10 % of these scenes, a non-zero “rain rate” R was assigned. This rain rate obeys a half-Gaussian (positive only) distribution with unit standard deviation. The unique spectral signature of the rain is described by a unit vector $\hat{a} = (0.366, -0.682, 0.633)$,

AMTD

6, 2327–2352, 2013

Dimensionality reduction in Bayesian estimation algorithms

G. W. Petty

[Title Page](#)

[Abstract](#)

[Introduction](#)

[Conclusions](#)

[References](#)

[Tables](#)

[Figures](#)

◀

▶

◀

▶

[Back](#)

[Close](#)

[Full Screen / Esc](#)

[Printer-friendly Version](#)

[Interactive Discussion](#)



so that raining scenes are simply non-raining scenes with an added brightness temperature perturbation given by $R \hat{a}$. This idealized rain signal is of course considerably cleaner than that encountered in real satellite observations and serves our purpose of highlighting retrieval issues associated with the Bayesian framework as as opposed to the physics.

This dataset is split evenly into a “training dataset” (TRAIN) and a “validation dataset” (VAL), each consisting of 10 000 “observations.” Our objective is to utilize the TRAIN data to implement a Bayesian algorithm capable of achieving reasonable performance when applied to the independent VAL data. Because the two data sets are statistically identical, problems we identify will be associated exclusively with issues relating to sampling density and dimensionality, not to representativeness or modeling error.

Two-dimensional scatter plots for each possible pair of channels are depicted for the TRAIN data in Fig. 1. The most important feature revealed in these plots is that no two channels are sufficient by themselves to distinguish between raining and non-raining scenes. It cannot even be discerned from these plots whether the raining scenes are in any way separable from the non-raining scenes, let alone whether it is possible to reliably estimate rain rate using all three channels.

In fact, the background noise in this demonstration is confined to a 3-dimensional plane (apart from 1 K Gaussian noise), and the subtler rain “signature” \hat{a} has, by design, a component normal to that plane. There is thus indeed some degree of separation in 3-channel space between the raining and non-raining pixels, though it would be visible in a 2-D plot only under an appropriate rotation relative to the original channels. The question addressed in the next section is whether a conventional Bayesian retrieval algorithm that relies on matching of observations in 3-channel space and employing the usual assumption of diagonal error covariance can successfully pull the relatively weak rain signal out of the much larger-magnitude background noise.

Dimensionality reduction in Bayesian estimation algorithms

G. W. Petty

[Title Page](#)

[Abstract](#)

[Introduction](#)

[Conclusions](#)

[References](#)

[Tables](#)

[Figures](#)

◀

▶

◀

▶

[Back](#)

[Close](#)

[Full Screen / Esc](#)

[Printer-friendly Version](#)

[Interactive Discussion](#)

3 Bayesian retrieval in channel space

3.1 Method

We begin with a straightforward retrieval method conceptually similar to that currently used for TMI and envisaged as well for the future GPM Microwave Imager. The TRAIN data serves as a database that is searched for scenes that match a given observation vector to within a specified tolerance. Specifically, each prospective match is assigned a weight given by $w = \exp(-s)$, where

$$s = \sum_i \left(\frac{x_i - x'_i}{\sigma_i} \right)^2, \quad (3)$$

where x_i is the observation from the i th channel, x'_i is the corresponding value for the database entry, and σ_i is a channel-dependent uncertainty that captures modeling error and/or observation error. As previously noted, a rigorous calculation of s should actually be based on a full error covariance matrix \mathbf{S} , of which σ_i^2 are the diagonal elements, but this is typically not done. For the idealized experiments described herein, we take σ to have the same value for all “channels.”

Consistent with the current implementation of the Goddard Profiling algorithm (GPROF; Kummerow et al., 2001, 2011) for the TRMM Microwave Imager, we admit only matches for which $w > 0.01$. The retrieval for a given observation vector is then given by

$$\hat{R} = \frac{\sum_{j=1}^P w_j R'_j}{\sum_{j=1}^P w_j}, \quad (4)$$

where P is the number of qualifying matches.

For a given observation vector x , we require at least $N = 1$ to have any valid retrieval at all; larger N will improve the statistical representativeness of the retrieval and

[Title Page](#)

[Abstract](#)

[Introduction](#)

[Conclusions](#)

[References](#)

[Tables](#)

[Figures](#)

◀

▶

◀

▶

[Back](#)

[Close](#)

[Full Screen / Esc](#)

[Printer-friendly Version](#)

[Interactive Discussion](#)

may permit error statistics to be derived as well. In principle, we can always increase σ until sufficient matches are found, but the benefits of increased sample size must be weighed against the loss of retrieval quality that results from treating increasingly dissimilar scenes as “matches.”

5 As noted earlier, the problem of finding matches within a sufficiently small neighborhood of the observation is dependent on the local density of the training sample in observation space. The higher-dimensional the observation space, the larger the training database must be to ensure adequate density in any given neighborhood. Moreover, extreme values of the variable to be retrieved will usually occupy the most sparsely populated regions in channel space.

10

3.2 Application to TRAIN data

It is instructive to apply the above algorithm to the same observations stored in the TRAIN database, not only as a sanity check but to illustrate an important consideration in the choice of σ . We know that even an arbitrarily small value of σ will still yield an exact match in every case, because each observation passed to the algorithm is present in the database. Of interest here is what happens when the tolerance is loosened so as to find additional matches.

15 Figure 2 depicts scatter plots of the retrieved versus actual “rain rate” for four different choices of σ . For $\sigma = 0.03$ (Fig. 2a), the agreement is essentially perfect, because for each observation, exactly one match is found, and that match is the observation itself. But for $\sigma > 0.1$ (Fig. 2b–d), there is an increasing tendency toward underestimation of the true rain rate, because now dissimilar scenes are being included in the average, most of which have a significantly different (usually zero) rain rate. For $\sigma = 1.0$ (Fig. 2d), the result is nearly useless on a pixel-by-pixel basis, even though the mean rain rate for all pixels will still be correct.

20

25 The initial conclusion to draw from this comparison is that *even when exact matches to all observations are available in the database, the use of an inappropriately large value for σ will seriously degrade the retrievals.*

[Title Page](#)

[Abstract](#)

[Introduction](#)

[Conclusions](#)

[References](#)

[Tables](#)

[Figures](#)

◀

▶

[Back](#)

[Close](#)

[Full Screen / Esc](#)

[Printer-friendly Version](#)

[Interactive Discussion](#)

3.3 Application to VAL data

We now apply the same algorithm to the independent VAL data. Statistically, these data are identical to those in the TRAIN database, but there likely to be very few exact matches. Consequently, when $\sigma = 0.03$, the match failure rate is an unacceptably large 5 94 % (Fig. 3a). Moreover, the failure rate is highest by far for the non-zero rain rate scenes, since these are more thinly spread in channel space.

When $\sigma = 0.1$, the match failure rate is still 40 %, but at least there is some semblance of skill in the retrieval for those pixels that do have matches (Fig. 3b). Again, the failure rate specifically for precipitating pixels is far higher, so that an average of 10 only the successfully matched observations would yield a severe underestimate of the true mean value of “rain rate”. Increasing σ further eliminates most of the match failures but leads to the same collapse of retrieval skill and systematic low bias previously shown for the dependent (TRAIN) data set. In summary, there is no single value of σ that yields an acceptable tradeoff between match rate and algorithm performance, 15 despite our reliance on a database of 10 000 entries populating a mere 3-dimensional observation space.

The GPROF algorithm, to give one example, recognizes this problem and undertakes multiple passes through the data base. If a match is not found for a given value of σ , the value is doubled and the search repeated. In this way, it is ensured that matches 20 will eventually be found for all observations, albeit with very loose tolerances for the rarest combinations of channel values.

Figure 4 illustrates the results of this procedure applied to our synthetic database. There are now no match failures, but the quality of the retrieval remains poor, with a great many underestimated values of larger “rain rates” and a similar number of non-25 zero retrievals where the true value is zero. Less apparent is that this procedure does not even conserve the ensemble averaged “rain rate” for the entire data set – the average for all 10 000 points is only 39 % of the true value. This is because non-raining scenes typically find matches for low values of σ so that only other non-raining

[Title Page](#)

[Abstract](#)

[Introduction](#)

[Conclusions](#)

[References](#)

[Tables](#)

[Figures](#)

◀

▶

[Back](#)

[Close](#)

[Full Screen / Esc](#)

[Printer-friendly Version](#)

[Interactive Discussion](#)

(or low-raining) scenes are included in the retrieval, while high rain rates in the tail of the distribution typically require large σ with the consequent incorporation of poorer matches (typically lower rain rates) into the retrieval.

3.4 Preliminary assessment

5 In view of the poor performance of the Bayesian scheme in the above demonstration, one might reasonably ask whether the signal-to-noise ratio is simply too poor in the synthetic data set for any retrieval method to yield high-quality results, at least using a solution database comprising only 10 000 entries.

In fact, it will be shown shortly that retrieval performance is markedly improved simply
10 by first applying an operator that retains sensitivity to the rain rate signal while rejecting most of the background noise and by using the single resulting pseudochannel as an index into the TRAIN database.

4 Dimensional reduction

4.1 General goals

15 In the present context, the process of dimensional reduction entails the following:

1. Starting with the original set of sensor channels, find linear transformations that normalize and decorrelate the temporal and spatial background variability (geophysical noise). That is, we want our transformed channels to have unit variance and zero cross-correlation when only scenes containing no rain are considered.
- 20 2. From those first-stage transformed channels, perform a second linear transformation that collects most of the sensitivity to the desired variable (e.g. rain or cloud) into a significantly smaller number of pseudochannels.

AMTD

6, 2327–2352, 2013

Dimensionality reduction in Bayesian estimation algorithms

G. W. Petty

[Title Page](#)

[Abstract](#)

[Introduction](#)

[Conclusions](#)

[References](#)

[Tables](#)

[Figures](#)

◀

▶

◀

▶

[Back](#)

[Close](#)

[Full Screen / Esc](#)

[Printer-friendly Version](#)

[Interactive Discussion](#)



3. Utilize those pseudochannels in place of the larger number of original channels in a lower-dimensional Bayesian retrieval scheme.

Note that while principal component analysis (PCA) is utilized in the first two steps and is, in general, a common method for dimensional reduction, a single-stage PCA decomposition of a data set does not accomplish either of the two steps on its own.

4.2 Details

4.2.1 Stage 1

The following procedure is applied to pixels for which the variable y to be retrieved is zero. In the present case, this condition is satisfied by 90 % of the database, or 9000 observations.

First, we compute the mean $\langle \mathbf{x} \rangle$ and the covariance \mathbf{S}_x for the rain-free pixels. From \mathbf{S}_x , we then compute the eigenvectors \mathbf{E}_x and eigenvalues Λ_x . We define the first-stage transformed channels \mathbf{y} via

$$y_i = \lambda_{x,i}^{-1/2} [(\mathbf{x} - \langle \mathbf{x} \rangle)^T \mathbf{E}_x]_i. \quad (5)$$

That is, we take the projection of $(\mathbf{x} - \langle \mathbf{x} \rangle)$ onto the i th eigenvector and then scale by square root of the eigenvalue to obtain unit variance. With appropriate definitions of the coefficient matrix \mathbf{A} , the above operation reduces to

$$\mathbf{y} = \mathbf{A}(\mathbf{x} - \langle \mathbf{x} \rangle). \quad (6)$$

$\langle \mathbf{y} \rangle = 0$ and $\mathbf{S}_y = \mathbf{I}$ for the set of transformed channels when observing the background only, but they otherwise retain all of the same information as found in the original \mathbf{x} . Thus, we may now conveniently treat the total background noise (instrument plus geophysical) as having unit variance and zero correlation between transformed channels \mathbf{y} .

[Title Page](#)

[Abstract](#)

[Introduction](#)

[Conclusions](#)

[References](#)

[Tables](#)

[Figures](#)

◀

▶

◀

▶

[Back](#)

[Close](#)

[Full Screen / Esc](#)

[Printer-friendly Version](#)

[Interactive Discussion](#)

4.2.2 Stage 2

We now apply Eq. (6) to the 10 000 precipitating scenes \mathbf{y}_r with $R > 0$. Unlike the case for rain-free scenes, the variance in each transformed channel y_i is now larger than unity owing to the added effects of precipitation on top of the background variability.

5 isolate that added variance, we compute $\mathbf{S}_{y,r} \equiv \langle \mathbf{y}_r \mathbf{y}_r^T \rangle$ and compute eigenvectors $\mathbf{E}_{y,r}$ and eigenvalues $\Lambda_{y,r}$.

We now define the precipitation *pseudochannels*

$$z \equiv \mathbf{y}^T \mathbf{E}_{y,r}. \quad (7)$$

Outside of precipitation, these pseudochannels still have zero mean and unit uncorrelated variance.

10 For precipitating scenes, however, the added variability will have been pushed into the first few eigenvectors $\mathbf{E}_{y,r}$.

We may keep the first M elements of z so as to account for at least, say, 95 % of the variance computed from the sums of the eigenvalues. The rest are discarded. For the synthetic dataset discussed above, $M = 1$.

15 The first- and second-stage linear transformation may be combined to give

$$z_{1\dots M} = \mathbf{B}(\mathbf{x} - \langle \mathbf{x} \rangle), \quad (8)$$

where \mathbf{B} is an $M \times N$ matrix of coefficients consistent with Eqs. (5) and (7). Note, by the way, that for the present dataset, $M = 1$ and therefore \mathbf{B} is a $1 \times N$ matrix or, equivalently, a vector corresponding in direction to the third eigenvector of the complete dataset.

20 We could have arrived at \mathbf{B} more directly in this particular case, but the dimensional reduction algorithm described above works well for higher-dimensional dataset, unlike the case for conventional single-stage PCA, which provides no basis for assigning eigenvectors to specific geophysical signatures.

[Title Page](#)

[Abstract](#)

[Introduction](#)

[Conclusions](#)

[References](#)

[Tables](#)

[Figures](#)

◀

▶

[Back](#)

[Close](#)

[Full Screen / Esc](#)

[Printer-friendly Version](#)

[Interactive Discussion](#)

5 Bayesian retrieval in pseudochannel space

Using the pseudochannel transformations derived above, we may undertake Bayesian retrievals in M -dimensional space rather than the original N -dimensional space. The procedure is otherwise identical to that described in Sect. 3. For our synthetic database, 5 $M = 1$.

Figure 5 depicts results of the algorithm applied to the dependent (TRAIN) data for selected values of σ , analogous to Fig. 2 (note that values of σ here cannot be directly compared with the values of σ for the 3-D retrievals, owing to the difference in scaling of the observation vector). We see an increase in retrieval error with larger values of 10 σ , but the degradation is not nearly as severe as was the case for the original retrieval using three channels.

Figure 6 depicts results for the independent (VAL) data, analogous to Fig. 3. The improvement relative to 3-D retrieval is striking. The match failure rate is extremely low for all but the smallest values of σ , and the errors are generally small and unbiased. 15 For larger values of σ , there is a hint of underestimation at the low end of the scale owing to inclusion of nearby zero values.

Finally, Fig. 7 depicts the results obtained when an iterative increase in σ is used to ensure that matches are found for all observations. These results may be directly compared to the results of the same procedure applied to 3-D Bayesian retrievals in 20 Fig. 4. The improvement in retrieval performance is striking.

6 Conclusions and discussion

Starting with an idealized synthetic database that loosely resembles 3-channel passive microwave observations of precipitation against a highly variable background (e.g. heterogeneous land surfaces and/or land-water mixes), we examined the performance of 25 a conventional Bayesian retrieval algorithm that searched for matches in the full three-dimensional channel space. First we showed that even when the algorithm is applied

AMTD

6, 2327–2352, 2013

Dimensionality reduction in Bayesian estimation algorithms

G. W. Petty

[Title Page](#)

[Abstract](#)

[Introduction](#)

[Conclusions](#)

[References](#)

[Tables](#)

[Figures](#)

◀

▶

◀

▶

[Back](#)

[Close](#)

[Full Screen / Esc](#)

[Printer-friendly Version](#)

[Interactive Discussion](#)



to the dependent (TRAIN) data, performance suffers when the match criterion is too loose (i.e. large σ). Conversely, when the same algorithm was applied to the independent (VAL) dataset, the match failure rate was unacceptably high *unless* the match criterion was loose.

5 The net result of both effects was that retrievals were quite poor for the independent data set, even when σ was iteratively adjusted, as shown in Fig. 4. Of course, the need to use large σ to overcome a high match failure rate is a function of the size of the dependent dataset. In the present demonstration, the TRAIN dataset consisted of 10 000 unique entries. If we were to employ a much larger database, the match failure 10 rate would go down, allowing smaller σ and presumably leading to improved overall performance.

But as one moves to more realistic retrieval problems employing more sensor channels – e.g. the nine channels of the TMI or the 13 channels of the GMI – the “curse of dimensionality” (Bellman, 1961) greatly magnifies the size of the database required to 15 adequately populate the observation space and thus to ensure not only the existence of suitable matches for any given observation but also, it would be hoped, a statistically representative *distribution* of such matches.

Of course it might be possible to reduce dimensionality in some cases by simply throwing out channels that are deemed to provide little information. As clearly seen in 20 Fig. 1, this would not have been possible in the present demonstration. With higher-dimensional real satellite data, the decision as to which channels do or not contain useful information is nontrivial and undoubtedly context-dependent.

To mitigate the problem of dimensionality in Bayesian retrievals, we described an algorithm for objectively distilling the relevant information content from N channels into 25 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 $N = 3$.

Dimensionality reduction in Bayesian estimation algorithms

G. W. Petty

[Title Page](#)

[Abstract](#)

[Introduction](#)

[Conclusions](#)

[References](#)

[Tables](#)

[Figures](#)

◀

▶

◀

▶

[Back](#)

[Close](#)

[Full Screen / Esc](#)

[Printer-friendly Version](#)

[Interactive Discussion](#)

5 Bayesian retrievals based on the single “pseudochannel” derived for the synthetic dataset were shown to yield striking improvements in overall performance, as shown in Fig. 7. These empirical results, more than any theoretical arguments, underline the likely benefits of dimensional reduction in Bayesian retrievals relying on a database of multichannel observations as a proxy for the prior joint and marginal PDFs.

10 It must be reiterated that the details of the particular dimensional reduction method given here depend on one being able to stratify the dependent dataset into two subsets: one representing the “pure” background (e.g. rain-free or cloud-free), and the other representing non-zero values of the variable to be retrieved (e.g. raining or cloudy). For variables where this is not possible, another dimensional reduction algorithm would need to be employed; however, the benefits for Bayesian retrievals should be similar.

15 As discussed in the introduction, it might have been possible in principle to achieve the same results for the 3-D channel-based retrieval as for the 1-D pseudochannel-based retrieval, provided that an appropriate covariance matrix were specified for the computation of the weights for candidate matches. In the present demonstration, the covariance in question would have corresponded to a match zone shaped like a highly flattened spheroid oriented exactly parallel to the plane containing most of the background variability in Fig. 1. That is, channel variations orthogonal to the principal plane of background variability would be given far greater weight than variations spectrally 20 consistent with the background variability. In short, the appropriate covariance would be non-diagonal, and retrieval performance would depend strongly on getting it exactly right.

25 From our results and from the above considerations, we conjecture that a major potential problem with conventional multichannel retrievals and assimilation schemes – whether Bayesian or not – lies in the very common but often inappropriate assumption of diagonal error covariance. The dimensional reduction technique described herein avoids this problem by, in effect, recasting the retrieval problem in a coordinate system in which the desired covariance is (a) lower-dimensional, (b) diagonal, and (c) unit magnitude.

Dimensionality reduction in Bayesian estimation algorithms

G. W. Petty

[Title Page](#)

[Abstract](#)

[Introduction](#)

[Conclusions](#)

[References](#)

[Tables](#)

[Figures](#)

◀

▶

◀

▶

[Back](#)

[Close](#)

[Full Screen / Esc](#)

[Printer-friendly Version](#)

[Interactive Discussion](#)

Acknowledgements. Helpful comments by Tristan L'Ecuyer and an anonymous reviewer led to improvements in the manuscript. This research was supported by NASA Grant NNX10AGAH69G.

AMTD

6, 2327–2352, 2013

References

5 Bauer, P., Amayenc, P., Kummerow, C., and Smith, E.: Over-ocean rainfall retrieval from multi-sensor data of the Tropical Rainfall Measuring Mission. Part II: Algorithm implementation, *J. Atmos. Ocean Tech.*, 18, 1838–1855, 2001. 2329

10 Bayes, T. and Price, R.: An essay towards solving a problem in the doctrine of chance, by the late Rev., Mr. Bayes, F. R. S. communicated by Mr. Prices, in a letter to John Canton, A. M. F. R. S., *Philos. T. R. Soc. Lond.*, 53, 370–418, 1763. 2329

15 Bellman, R.: *Adaptive Control Processes: a Guided Tour*, Princeton University Press, 1961. 2342

20 Chiu, J.-Y. and Petty, G.: Bayesian retrieval of complete posterior PDFs of oceanic rain rate from microwave observations, *J. Appl. Meteorol. Clim.*, 45, 1073–1095, 2006. 2329, 2330

25 Di Michele, S., Tassa, A., Mugnai, A., Marzano, F., Bauer, P., and Baptista, J.: Bayesian algorithm for microwave-based precipitation retrieval: Description and application to TMI measurements over ocean, *IEEE T. Geosci. Remote*, 43, 778–791, 2005. 2329

30 Evans, K., Turk, J., Wong, J., and Stephens, T.: A Bayesian approach to microwave precipitation profile retrieval, *J. Appl. Meteorol.*, 34, 260–279, 1995. 2329

35 Grecu, M. and Olson, W.: Bayesian estimation of precipitation from satellite passive microwave observations using combined radar-radiometer retrievals, *J. Appl. Meteorol. Clim.*, 45, 416–433, 2006. 2329

40 Haddad, Z., Smith, E., Kummerow, C., Iguchi, T., Farrar, M., Durden, S., Alves, M., and Olson, W.: The TRMM “day-1” radar/radiometer combined rain-profiling algorithm, *J. Meteorol. Soc. Jpn.*, 75, 799–809, 1997. 2329

45 Kummerow, C., Barnes, W., Kozu, T., Shiue, J., and Simpson, J.: The tropical rainfall measuring mission (TRMM) sensor package, *J. Atmos. Ocean Tech.*, 15, 809–817, 1998. 2331

50 Kummerow, C., Hong, Y., Olson, W., Yang, S., Adler, R., McCollum, J., Ferraro, R., Petty, G., Shin, D., and Wilheit, T.: The evolution of the Goddard profiling algorithm (GPROF) for rainfall

[Title Page](#)

[Abstract](#) [Introduction](#)

[Conclusions](#) [References](#)

[Tables](#) [Figures](#)

◀

▶

◀

▶

[Back](#)

[Close](#)

[Full Screen / Esc](#)

[Printer-friendly Version](#)

[Interactive Discussion](#)

Dimensionality reduction in Bayesian estimation algorithms

G. W. Petty

[Title Page](#)

[Abstract](#)

[Introduction](#)

[Conclusions](#)

[References](#)

[Tables](#)

[Figures](#)

◀

▶

[Back](#)

[Close](#)

[Full Screen / Esc](#)

[Printer-friendly Version](#)

[Interactive Discussion](#)



Fig. 1. Two-channel scatter plots of an idealized, stochastically generated database consisting of three-channel “brightness temperatures”. Gray circular markers are used for “non-raining” scenes; black crosses indicate “raining” scenes. While no separation between raining and non-raining scenes is evident in these 2-D plots, there is in fact partial separation in three dimensions.

Dimensionality reduction in Bayesian estimation algorithms

G. W. Petty

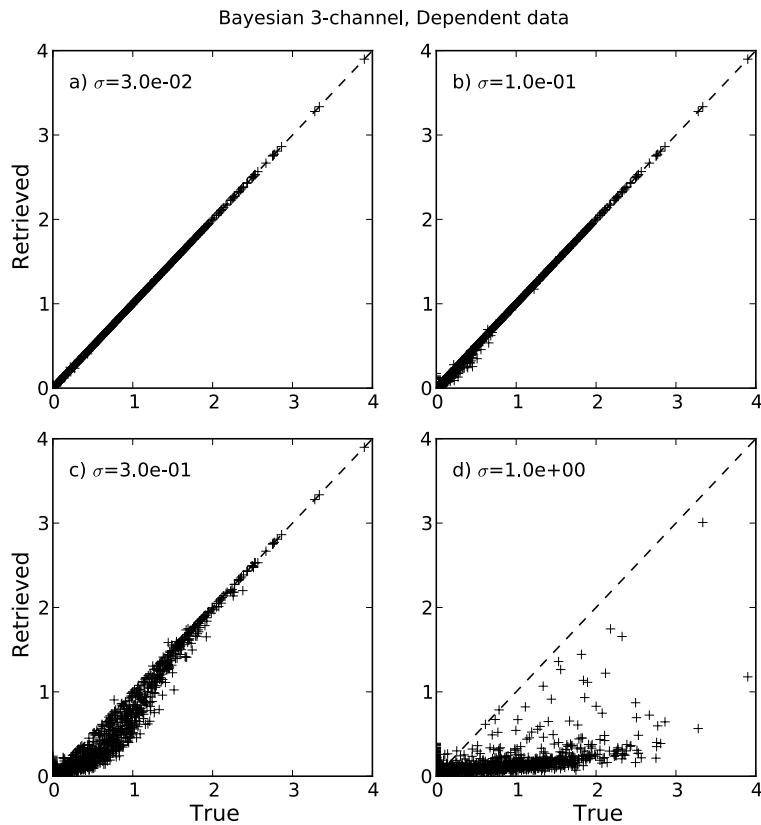


Fig. 2. Results of a Bayesian retrieval algorithm applied to the three-channel dependent (TRAIN) dataset for four different values of the error parameter σ .

- [Title Page](#)
- [Abstract](#) [Introduction](#)
- [Conclusions](#) [References](#)
- [Tables](#) [Figures](#)
- [◀](#) [▶](#)
- [◀](#) [▶](#)
- [Back](#) [Close](#)
- [Full Screen / Esc](#)

[Printer-friendly Version](#)

[Interactive Discussion](#)

Dimensionality reduction in Bayesian estimation algorithms

G. W. Petty

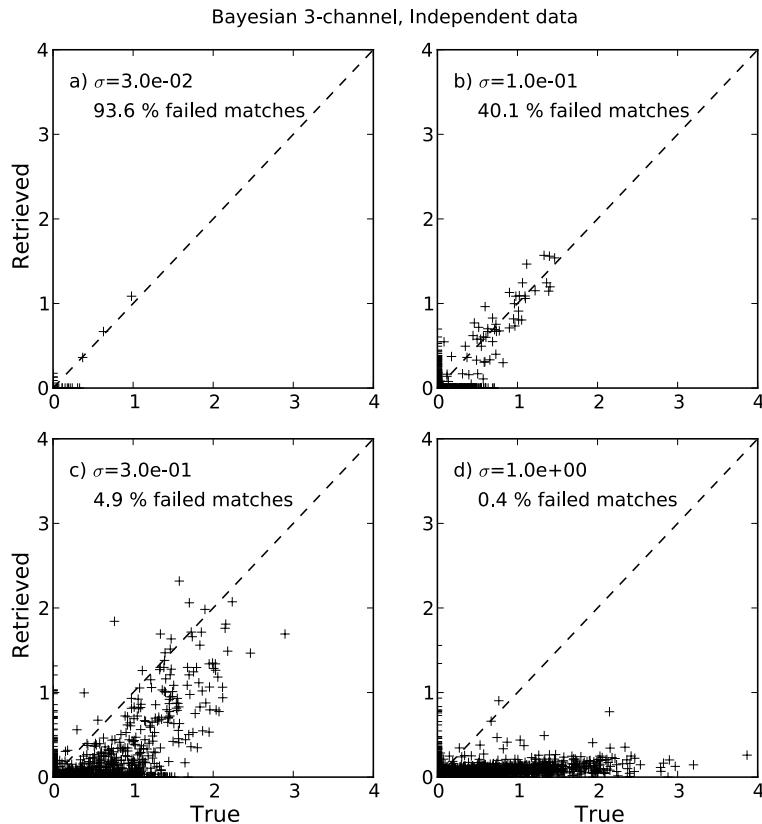


Fig. 3. Same as Fig. 2, but the Bayesian algorithm is applied to the independent (VAL) dataset. Also indicated is the percentage of “observations” for which no match could be found for the given value of σ .

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

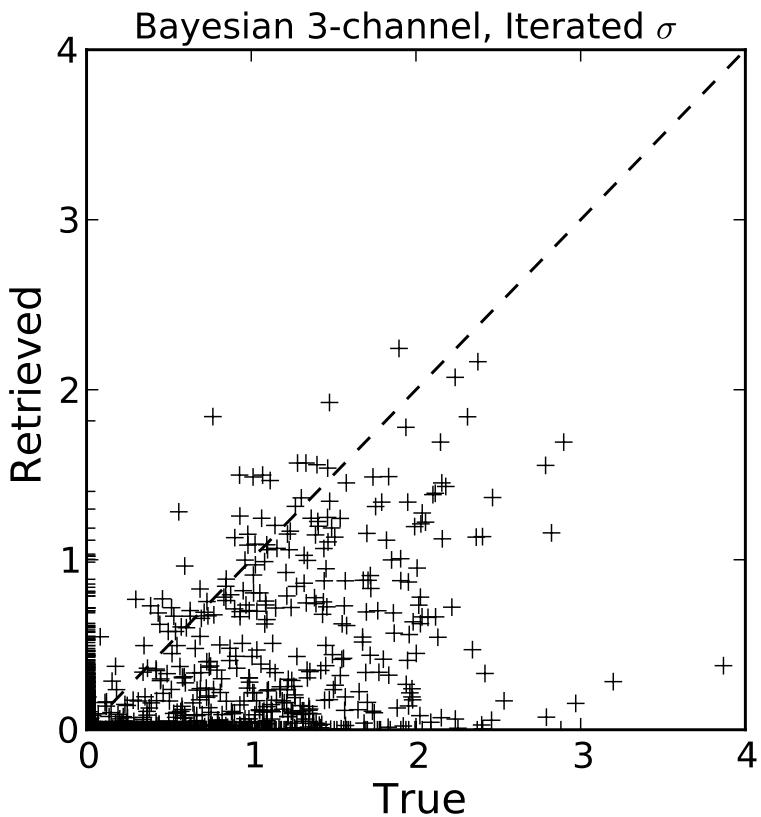


Fig. 4. Results of the Bayesian algorithm applied to the independent (VAL) dataset, using iterative doubling of σ to ensure that matches are found for all “observations”.

Dimensionality reduction in Bayesian estimation algorithms

G. W. Petty

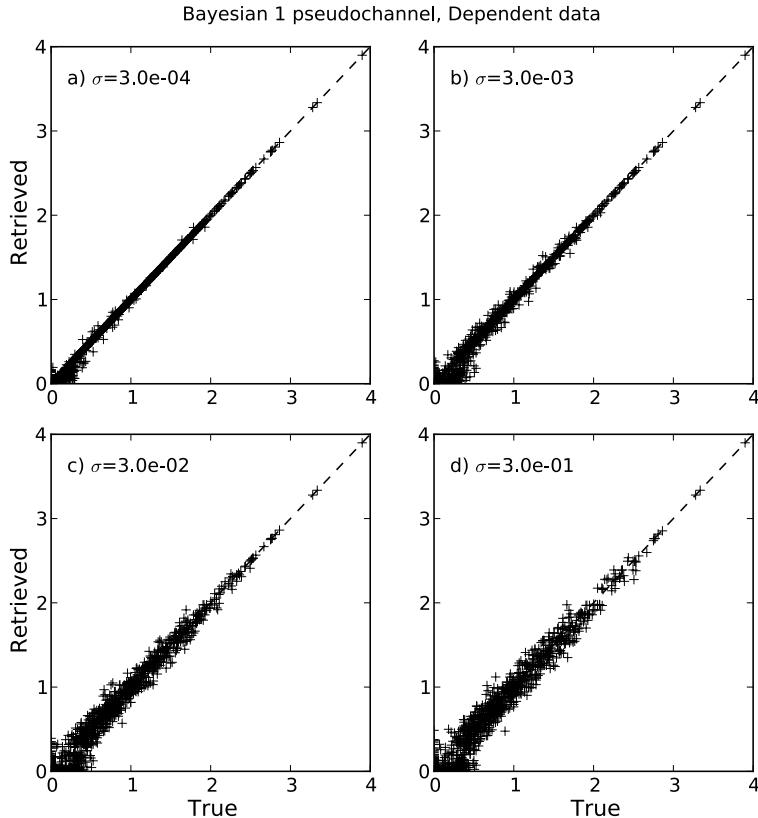


Fig. 5. Similar to Fig. 2, but the Bayesian retrieval applied to the dependent (TRAIN) dataset is based on a one-dimensional “pseudochannel” rather than the original three-dimensional “brightness temperatures”.

[Title Page](#)

[Abstract](#)

[Introduction](#)

[Conclusions](#)

[References](#)

[Tables](#)

[Figures](#)

◀

▶

◀

▶

[Back](#)

[Close](#)

[Full Screen / Esc](#)

[Printer-friendly Version](#)

[Interactive Discussion](#)

Dimensionality reduction in Bayesian estimation algorithms

G. W. Petty

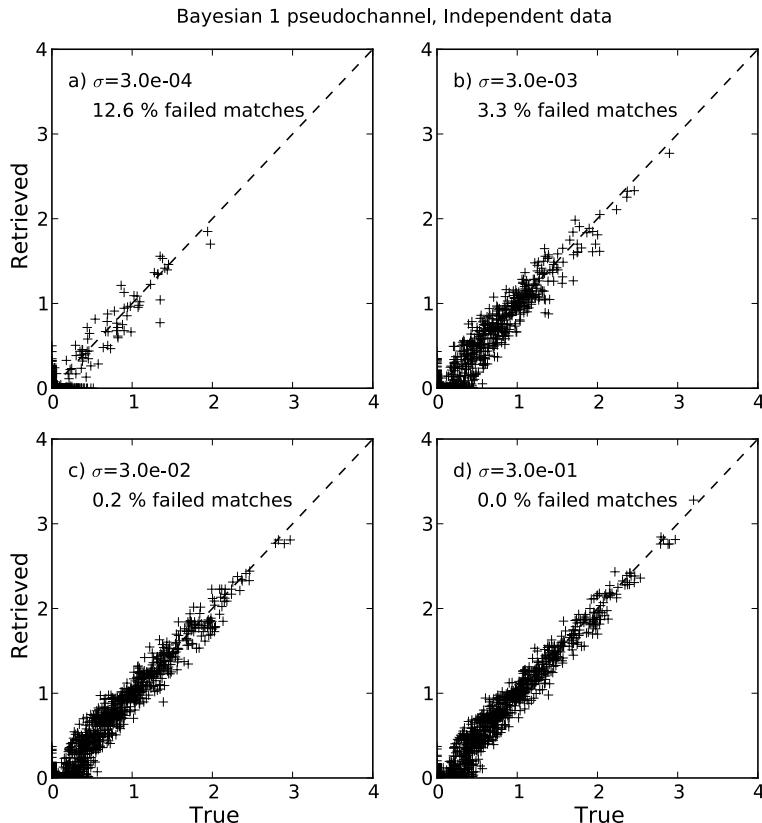


Fig. 6. Same as Fig. 5, but the Bayesian 1-D retrieval is applied to the independent (VAL) data.

[Title Page](#)

[Abstract](#)

[Introduction](#)

[Conclusions](#)

[References](#)

[Tables](#)

[Figures](#)

◀

▶

◀

▶

[Back](#)

[Close](#)

[Full Screen / Esc](#)

[Printer-friendly Version](#)

[Interactive Discussion](#)

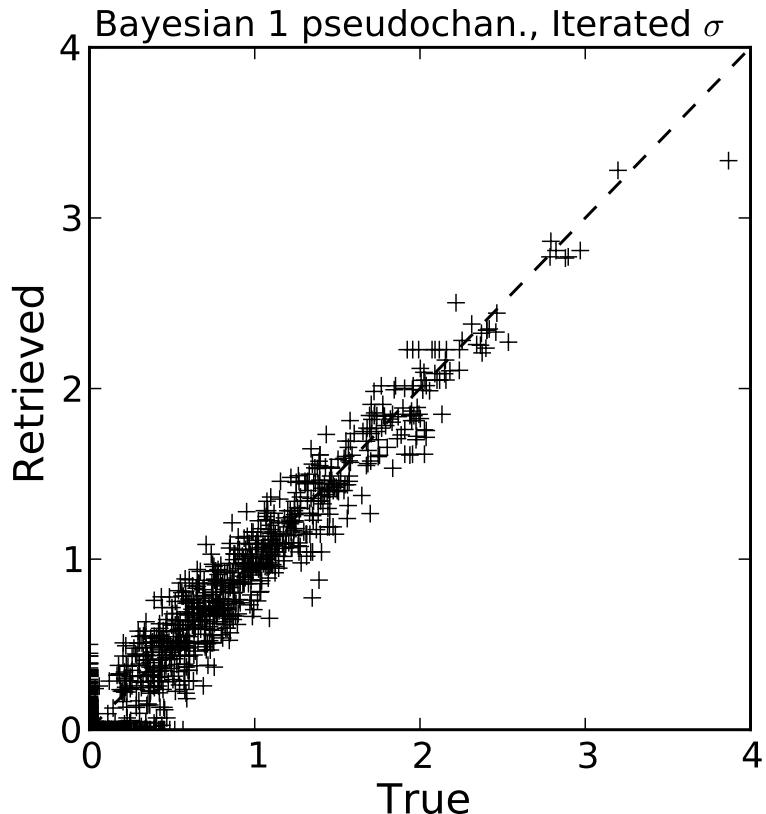


Fig. 7. Results of the 1-D Bayesian algorithm applied to the independent (VAL) dataset, using iterative doubling of σ to ensure that matches are found for all “observations”. Compare with the 3-D retrievals depicted in Fig. 4.