A non-linear method for IASI channel selection

Alison Fowler* | Peter Jan van Leeuwen

Introduction

The IASI instrument measures top of the atmosphere (TOA) radiances in 8641 channels. In many cases it is difficult to transmit, store and assimilate such a large amount of data [Collard, 2007]. A practical solution is to select a few hundred channels based on those with the highest information content.

Mutual information (MI) measures the information content of observations, **y**, as the change in entropy when the observations are assimilated.

$$\mathrm{MI} = \int p(\mathbf{y}) \int p(\mathbf{x}\,|\,\mathbf{y}) \ln \frac{p(\mathbf{x}\,|\,\mathbf{y})}{p(\mathbf{x})} \mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y}, \qquad (1)$$

where **x** represents the state vector.

The posterior distribution, $p(\mathbf{x}|\mathbf{y})$, is not expected to be Gaussian due to the non-linear relationship between the state vector, **x** (profiles of temperature and humidity), and the TOA radiances, **y**.

In this study we make use of RTTOV to model the relationship between **x** and **y**. The size of our state vector is n=102.

Estimating Mutual information

Linear approximation

When the linear/Gaussian approximation holds MI can be shown to be:

$$\mathrm{MI}^{\mathrm{lin}} = \frac{1}{2} \ln |\mathbf{B}\mathbf{P}_a^{-1}|, \qquad (2)$$

where **B** and $\mathbf{P}_a$ are the prior and analysis error covariance matrices respectively. This approximation has been used in previous studies of channel selection (e.g. Rabier et al. 2002).

Non-linear approximation

When the relationship between **x** and **y** is non-linear, as is the case for satellite observations, eqn (2) may no longer hold. Instead we can estimate MI by sampling from both the prior and likelihood distributions.

$$\mathrm{MI}^{\mathrm{samp}} = \sum_{i=1}^{M} p(\mathbf{y}_i) \sum_{j=1}^{N} w_{i,j} \ln w_{i,j} / \widetilde{w}_{i,j} \qquad (3)$$

where $w_{i,j}$ are the posterior weights (proportional to the likelihood) and $\widetilde{w}_{i,j}$ are the prior weights.

Comparison of linear and non-linear approximation to MI

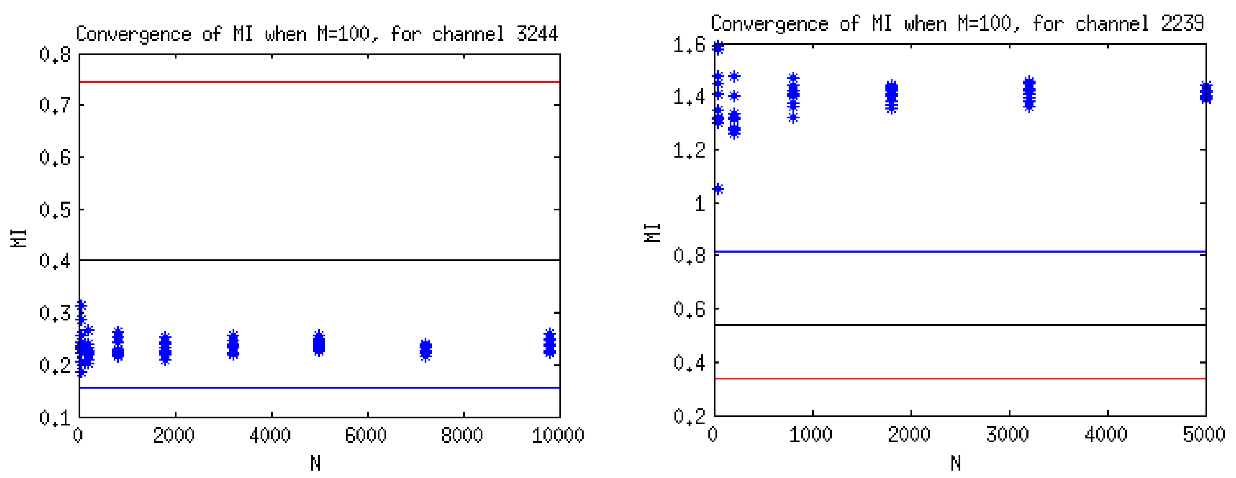

Figure 1. Linear estimate of MI (solid lines, black=linearised about the true state, $\mathbf{x}_t$, red=linearised about $\mathbf{x}_t$-$\boldsymbol{\sigma}_b$, blue=linearised about $\mathbf{x}_t$+$\boldsymbol{\sigma}_b$, where σ_b is the error standard deviation of the prior estimate) and sample estimate (stars) of MI for 2 different IASI channels. For a standard (no cloud) state.

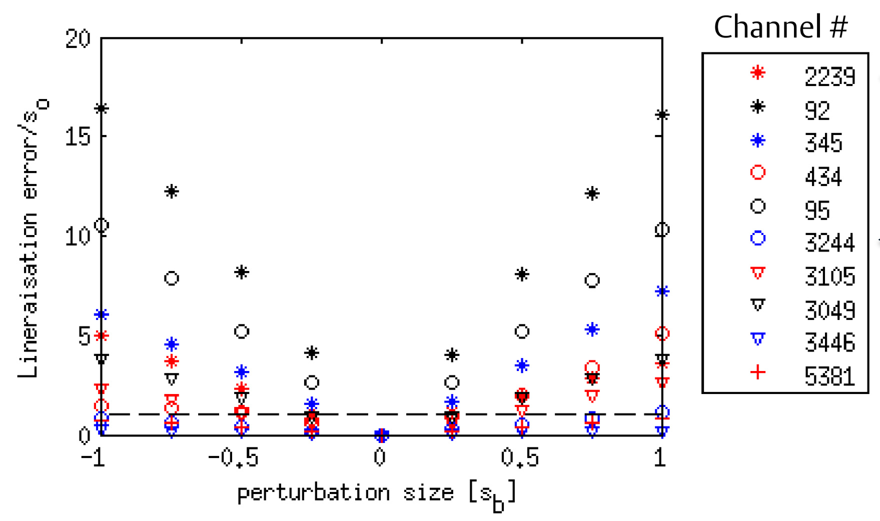

Figure 2. Linearisation error normalised by the error standard deviation of the channel, $\boldsymbol{\sigma}_o$, as a function of perturbation size (a fraction of $\boldsymbol{\sigma}_b$).

Linearisation error is defined as $\mathbf{H}\delta\mathbf{x} + h(\mathbf{x}) - h(\mathbf{x}+\delta\mathbf{x})$, where h is the non-linear observation operator and **H** is the linearised observation operator.

Channel selection algorithm

1. Sample N times from prior distribution, $\mathbf{x}_j$~$N(\mathbf{x}_t,\mathbf{B})$ for j=1:N, and set $\widetilde{w}_{i,j} = 1/N$ for j=1:N and i=1:M.
2. Sample M times from likelihood, $\mathbf{y}_i$~$N(h(\mathbf{x}_t),\mathbf{R})$ for i=1:M.
3. Transform prior sample to observation space using RTTOV, $h(\mathbf{x}_j)$ for j=1:N.
4. Update the weights of the prior sample given the observations (one channel, c, at a time)
   $w^c_{i,j}$=const*$\widetilde{w}_{i,j}$exp[-0.5(y^c_i-$h(\mathbf{x}_j)^c$)$\mathbf{R}^{-1}$(y^c_i-$h(\mathbf{x}_j)^c$)]
5. Approximate the marginal distribution before normalising the weights.
   $$p(y_i^c) = \sum_{j=1}^{N} w_{i,j}^c$$
6. Calculate MI for each channel using eqn (3).
7. Select channel with highest MI and update prior given the information from this selected channel i.e. $\widetilde{w}_{i,j} = w_{i,j}^{selected}$
8. Repeat steps 4 to 7 for the remaining channels until desired number of channels have been selected.

An example of the channel selection algorithm for the 10 channels given in fig. 2 is shown in fig. 3.

Effective sample size

The effective sample size is given by $ess_i = 1/\sum_{j=1}^{N} w_{i,j}^2$. This measures the number of samples that have any significance in representing the posterior distribution. If ess_i is small (~n then the sample estimate of MI will be poor.

In figure 3, we see that ess decreases with each subsequent channel selection as the region of high probability becomes more and more focused.

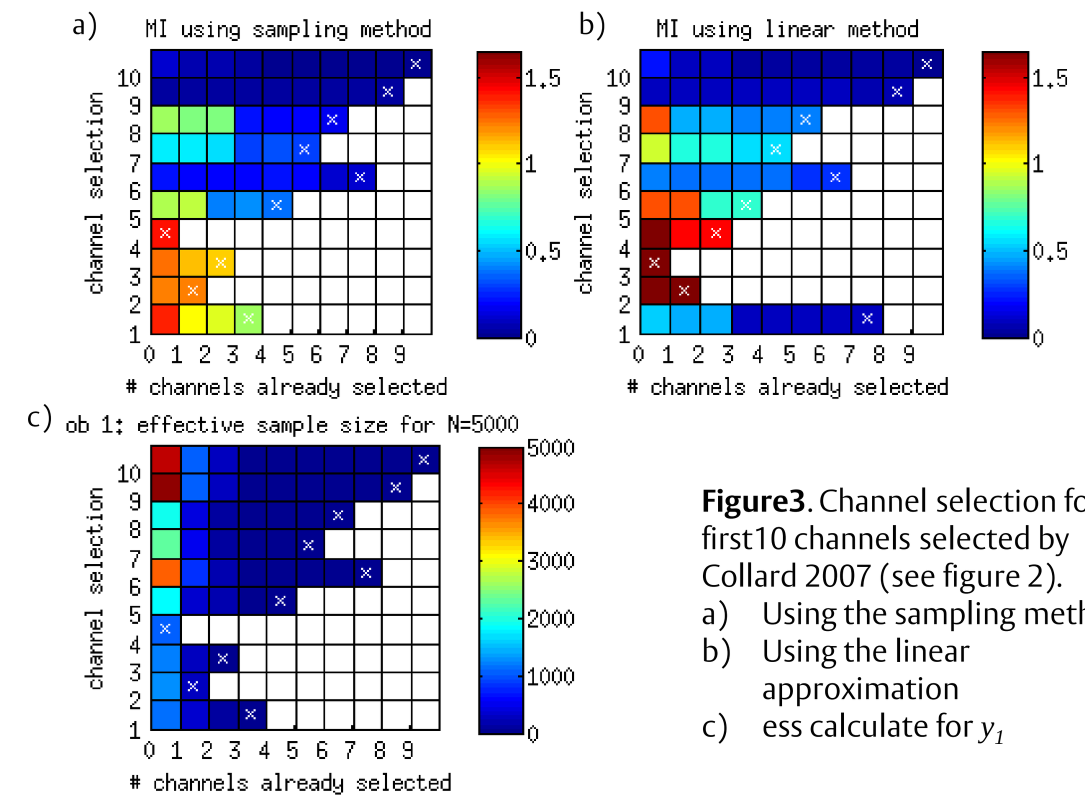

Figure3. Channel selection for the first10 channels selected by Collard 2007 (see figure 2).
a) Using the sampling method
b) Using the linear approximation
c) ess calculate for y_1

Gaussian mixture resampling

In order to have control over the effective sample size we have modified the channel selection algorithm to resample from the posterior distribution after each channel selection. We wish to preserve any non-Gaussian structure and so we fit a Gaussian mixture (GM) to the distribution with the number of components specified by the sample characteristics. Resampling from this distributions resets ess to N.

An example of GM resampling is shown in fig. 4 and the effect on channel selection is given in fig. 5.

*Contact information

- School of Mathematical and Physical sciences, University of Reading, Whiteknights, RG6 6AH
- Email: a.m.fowler@reading.ac.uk Web: www.reading.ac.uk/gj901587

References

1. Collard, 2007: Selection of IASI channels for use in NWP. *Q. J. R. Met. Soc.*, **133**, 1977-1991.
2. Rabier et al., 2002: Channel selection methods for IASI radiances. *Q. J. R. Met. Soc.*, **128**, 1011-1027.

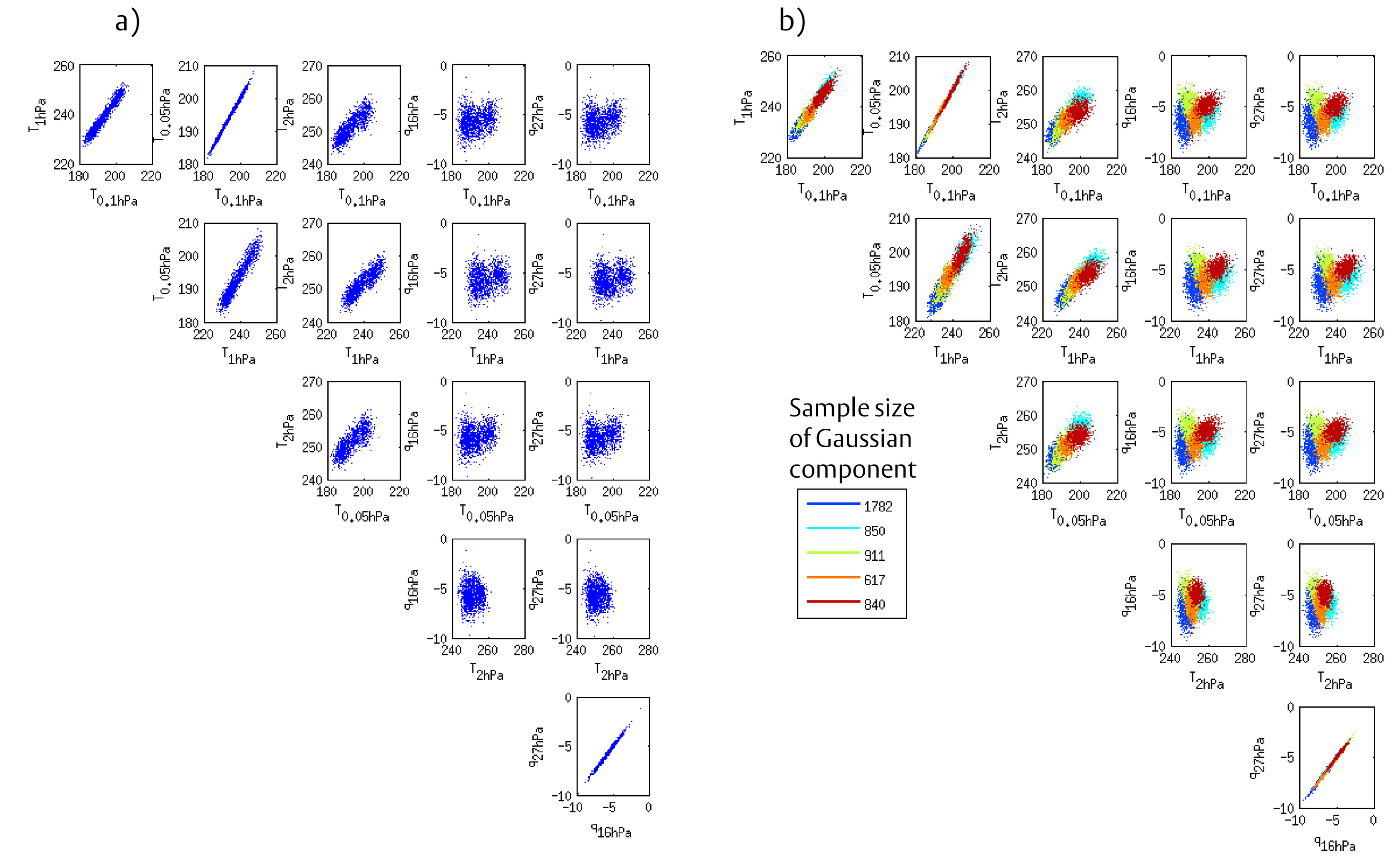

Figure4. a) Example of some joint distributions from the posterior distribution after the 8th channel is selected (ess=472).
b) GM resampling of same distributions.

Conclusions

- Satellite observations are non-linear functions of the atmospheric state variables of interest. As such a linear approximation to their information content may be misleading when used for channel selection.
- We have demonstrated a sampling approximation to mutual information which is free from assumptions about linearity. This shows that for some channels the linear approximation is indeed poor.
- Although this estimate is free from assumptions about the linearity it does suffer from the effects of undersampling when the region of high probability is small.
- Resampling from the posterior distribution after each channel is selected is shown to alleviate this problem.

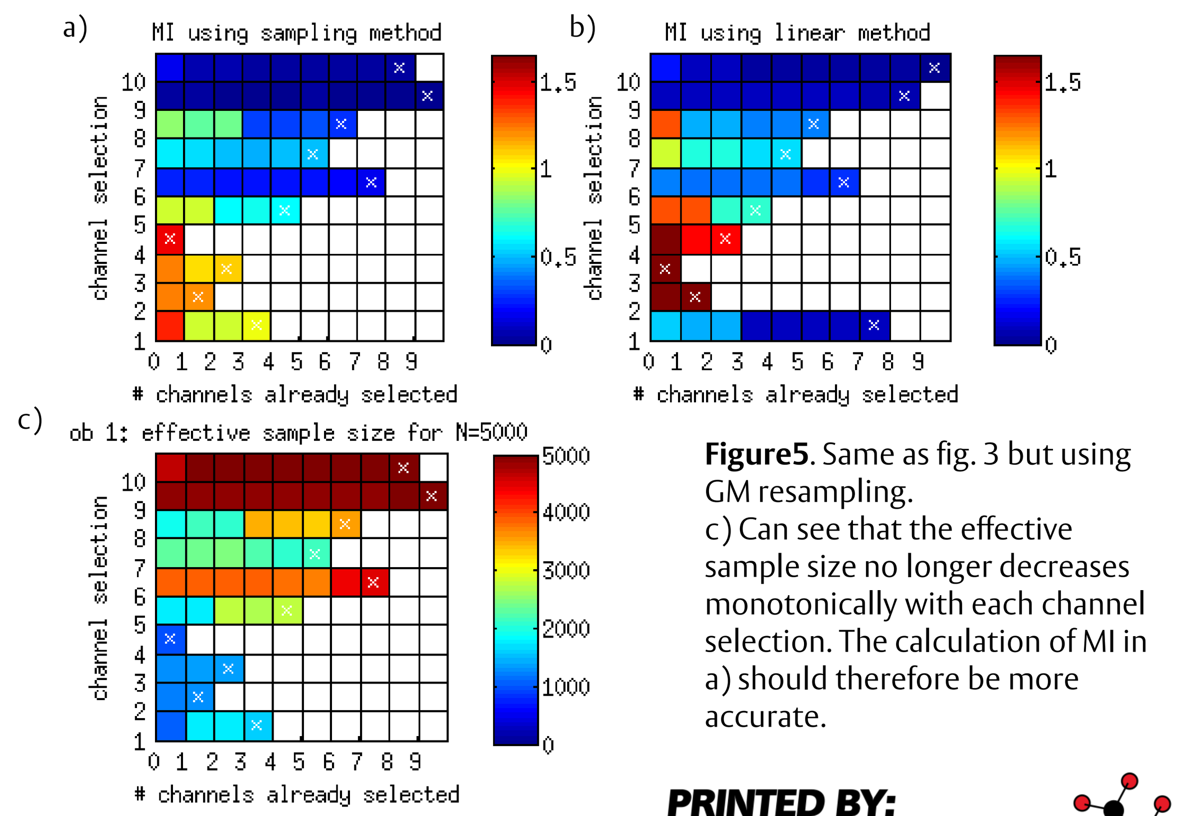

Figure5. Same as fig. 3 but using GM resampling.
c) Can see that the effective sample size no longer decreases monotonically with each channel selection. The calculation of MI in a) should therefore be more accurate.