Improving fast radiative transfer model predictions of water vapour line absorption

V. Sherlock

National Institute of Water and Atmospheric Research,
Wellington, New Zealand

Abstract

Accurate description of water vapour absorption has always proved challenging in fast radiative transfer modelling. Despite recent advances, further improvements in modelled water vapour line absorption are needed if advanced infrared sounder data is to be exploited to its full potential. Given that current fast model predictors and fast model error characteristics in water vapour line centres differ, and given the requirement to minimise regression instability due to predictor collinearity, we ask whether it is possible to find an optimal subset of predictors for water vapour line absorption. This paper outlines the predictor selection strategy adopted, and presents the major results from the predictor selection studies. Significant improvements in both forward model and Jacobian accuracy have been made.

Introduction

Accurate prediction of water vapour absorption has been a critical yet challenging problem for all fast radiative transfer (RT) developments to date. Although recent RT developments for the advanced sounders have given marked improvements in fast model accuracy for the 6.7 micron water vapour band [Matricardi et al., 2001, Sherlock et al., 2002], further improvements in the modelling of water vapour absorption, particularly in line centres, are required if data is to be exploited to its full potential.

In this paper we describe the two main reasons for seeking an optimal, minimum set of predictors for water vapour line absorption. We then describe the selection strategy adopted, and present the major results from predictor selection studies.

Motivation for the predictor selection study

Predictor collinearity

Fast model regression schemes typically involve large ($>$10) numbers of predictors, and these predictors are not linearly independent (see Table 1 for example). In this case, predictor collinearity could, in principle, degrade the accuracy of regression coefficient estimates.

Belsley [1991] proposed a method for identifying such cases, based on a decomposition of the error (variance) associated with the coefficient estimates in terms of the singular vectors of the regression matrix. This method has been implemented to examine the impact of predictor collinearity in water vapour line absorption regressions.

In the wings of water vapour lines the variance in the response variate – the effective layer optical depth – is low (errors ϵ in the linear model y=$\mathbf{X}\beta+\epsilon$ are small) and corresponding errors in coefficient estimates are small, mitigating the effects of collinearity. In water vapour line centres the variance in the response variate y is large. Subsets of linearly dependent predictors can be identified which have a large fraction of the uncertainty in their associated coefficient estimates associated with the same singular vector of X. This can give rise to large fluctuations in the value of coefficients for a given predictor from layer to layer, which is not in keeping with any physical process.

As Belsley points out, such collinearity is not damaging *per se*: the subset of collinear predictors taken as a whole may provide stable and useful predictions. Clearly though, the representativity of the training sample or dependent set becomes critical here – where accurate predictions depend on compensation effects, prediction for states which lie outside the space defined by the dependent set is fraught with hazards. In this instance it is preferable to limit the number of linearly dependent predictors in the regression as far as possible.

In addition to improving the stability of the regression scheme, reducing the number of predictors will reduce fast model execution times. Ensuring regression stability and accuracy remains the primary motivation in the predictor selection studies described here.

Fast model error characteristics

Formulations for describing water vapour absorption and dependent regression data used in fast model developments differ. Thus while it is not surprising that forward model and Jacobian error characteristics differ too, it is difficult to isolate the precise reasons for differences. To make some judgement about the merits of different prediction schemes, predictors have to be applied in a consistent manner to the same dependent data set(s).

In Figure 1 fast model error statistics for a model (G00) using H_2O line absorption predictors proposed by Hannon et al. [1996] are compared with equivalent statistics for a model (LRF) using H_2O line absorption predictors proposed by Matricardi et al. [2001]. Analysis shows that maximum errors occur in water vapour line centres for both predictor formulations. Line centre error characteristics differ markedly however. For example, maximum bias in the nadir view radiance simulations is reduced in the Matricardi et al. [2001] formulation, while clustering in the maximum H_2O Jacobian error plot, and to a lesser extent in the standard deviation scatter plot, suggests that there are classes of channels which one or other of the two formulations describes best.

The real question of interest then is whether one can identify specific predictors which are responsible for the improved description of water vapour line absorption in these instances ? If so, ultimately we would like to define an optimal subset of predictors from the union of the two predictor sets: *optimal* because neither of the two models considered has superior performance in all channels, *subset* because

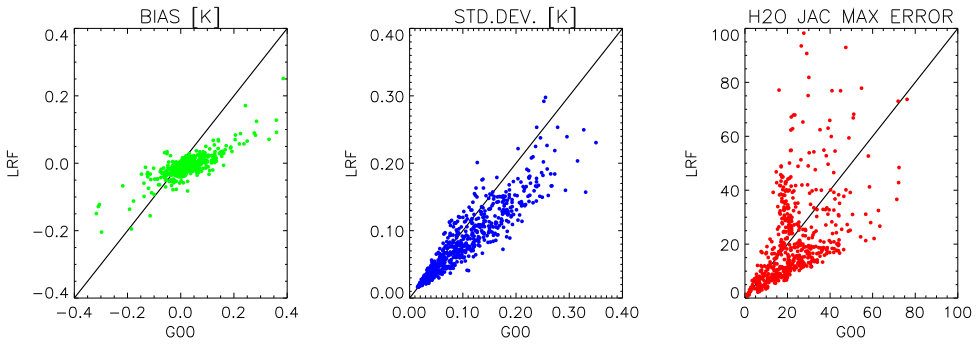

Figure 1: Scatterplots of forward model bias, standard deviation and maximum water vapour Jacobian errors, comparing the Hannon et al. [1996] (G00) and Matricardi et al. [2001] (LRF) water vapour line absorption predictors on the 1200–1650 cm^{-1} interval (602 channels). Bias and standard deviation are estimated based on comparisons with line-by-line RT calculations from kCARTA for the 176 ECMWF diverse profile set (nadir view). Jacobian error estimates, quantified in terms of the Garand measure of fit are estimated based on comparisons with line-by-line RT calculations from kCARTA for the dependent (regression) profile set (UMBC AIRS 48 profile set). See Sherlock et al. (2002) for further details.

we want to minimise the potential error and instability due to collinearity between predictors in the regression by using the smallest number of predictors possible into the regression.

Predictor Selection Strategy

Following from the fast model predictor intercomparison results outlined in the previous section, the union of water vapour line absorption predictors proposed by Hannon et al. [1996] and Matricardi et al. [2001] formed the basis set of predictors in the selection studies. These predictors are tabulated in Table 1 (total number of predictors, N_{union} =15). Regression calculations were performed using the UMBC convolved AIRS transmittance data (48 profiles) data, with water vapour line and continuum absorption modelled/predicted separately [Sherlock et al., 2002].

In order to favour selection of a minimum set of predictors and address potential redundancy problems raised by predictor collinearity specifically, two methods of stepwise regression were explored. The first of these methods uses a forward stepwise selection based on Akaike's Information Criterion (AIC) (basically, the minimisation includes a penalty term for the number of predictors in the regression). It clearly identifies lead predictors in regression, but is often of limited predictive use as the selection criteria is very restrictive (typically ≤ 3 predictors are selected).

The second of these methods, Efroymson's method, uses forward selection and backward elimination to select predictors and test for redundancy with previously selected predictors. This method gives selections with useful predictive skill, but analysis of results (tabulation of selected predictors for subsets of size $N \leq N_{union}$) for large amounts of data is difficult, at least in the context of the fast model layer optical depth prediction problem.

Preliminary studies using Efroymson's method (and the AIC) effectively showed that the lead predictors for modelling water vapour line absorption varied with both channel wavenumber and the pressure of

the absorbing layer. This is of course a natural reflection of different physical processes (e.g. line broadening processes) and spectral absorption regimes (e.g. line centre, line wing). It does however complicate the implementation and analysis of a global (all channels, all layers) predictor selection method. Channel and layer specific predictors are not a feasible or appropriate solution, principally because it is extremely difficult to ensure continuity and consistency in modelled absorption if predictors vary from layer to layer and channel to channel.

In response to this problem a two step selection strategy was devised. In the first stage a representative set of channels and a subset of layers were identified which covered the range of absorption regimes to be modelled. Using the stepwise regression models described above, predictors were ranked based on their frequency of selection using the Akaike Information Criterion and their frequency of selection among the lead 10 predictors using Efroymson's Method. Using this ranking, a minimum set of 'necessary' predictors, a set of 'candidate' predictors and a set of 'redundant' predictors were identified.

Given the set of necessary predictors, candidate predictors were then evaluated and ranked based on forward model and Jacobian errors for the full set of AIRS channels. Three specific measures were considered in order to quantify the accuracy and robustness of the regression relations: 1. forward model errors (bias and standard deviation) for an independent profile set, 2. forward model error inflation on passing from the dependent to the independent profile set and 3. temperature and humidity Jacobian errors, as characterised by the Garand measure of fit, for the dependent profile set.

Major results from the predictor selection study

Results from predictor classification and candidate predictor ranking studies for water vapour line absorption predictors proposed by Hannon et al. (H1996) and Matricardi et al. (M2001) are presented in Table 1. N denotes necessary predictors, and defines the minimum predictor set for useful predictive accuracy. C denotes candidate predictors, whose rank (or order of subsequent selection) $.n$ is determined from more extensive tests of forward model and Jacobian errors for the entire AIRS spectrum. R denotes redundant predictors; predictors which give no significant improvement in predictive skill, and which are rarely selected in stepwise regression tests.

Identification of a missing lead predictor

The $\sqrt{aW_{\mathrm{r}}}W_r/W_z$ predictor is a lead predictor for water vapour line absorption in the 1400–1650 cm^{-1} interval, but is absent from the set of predictors proposed by Hannon et al. [1996]. Figure 2 compares radiative transfer errors for calculations using the Hannon predictor set (G00) and using the the minimum predictor set identified in this study (LR9), illustrating the role of the $\sqrt{aW_{\mathrm{r}}}W_r/W_z$ predictor on this spectral interval: significant reductions in forward model errors, and in the upper quartile of Jacobian Garand measures of fit are observed in water vapour line centres.

Preferred candidate predictors

Forward model and Jacobian errors on the 1400–1600 cm^{-1} interval are further reduced on introduction of the aW_{z} predictor, as illustrated in the left hand panel of Figure 3. aW_{z} is the lead candidate predictor for all AIRS spectral intervals. $\sqrt{aW_{\mathrm{z}}}$ is the second selected candidate predictor. It's role in improving modelled line absorption in the longwave window region is illustrated in the right hand panel of Figure 3. Note in this instance unmodified maximum forward and Jacobian errors occur in the same channels – these channels should be excluded from a channel subset for use in data assimilation. The $\sqrt{aW_{\mathrm{z}}}$ predictor also gives some small improvements in modelled absorption in the 1400–1650 cm^{-1} interval.

Candidate predictors aW_{r}^2/W_z and aW_{r}^3 play a smaller, often ambiguous role in error reduction. Their inclusion in an optimal minimum predictor subset is debatable, although the aW_{r}^2/W_z predictor does play a distinct role in specific channels in the 1200–1400 cm^{-1} interval (see below).

Caveats for the 1200–1400 cm^{-1} interval

The $\sqrt{aW_{\mathrm{r}}}W_r/W_z$ predictor is not a lead predictor for the 1200-1400 cm^{-1} interval. As illustrated in the left hand graphics of Figure 4, improvements in standard deviation are marginal, and maximum water vapour Jacobian errors tend to be degraded by the presence of this predictor in regressions. This is also true, but to a lesser degree, of window region line absorption. The role of the $\sqrt{aW_{\mathrm{r}}}W_r/W_z$ in the 1400–1650 cm^{-1} interval is so significant, its retention in a minimum set of optimal predictors is favoured. However this will almost certainly have implications for channel selection in the 1200–1400 cm^{-1} interval.

All of the G00, LR9, LR10_M3 (= LR9 + aW_{z}) and LR11_M5 (= LR10_M3 + $\sqrt{aW_{\mathrm{z}}}$) predictor sets exhibit zenith angle dependent model errors (manifest here as a bias in nadir view forward model errors) in channels where there is interferring water vapour and methane line absorption. As illustrated in the right hand graph of Figure 4, this zenith angle dependence is reduced by the aW_{r}^2/W_z, but errors remain significant with respect to instrumental noise. This, combined with representativity errors associated with unmodelled variations in methane concentrations, suggests these channels should be excluded from the subsets selected for use in data assimilation.

Conclusions

Subsets of predictors with stable and robust error characteristics and useful predictive skill have been identified from the the union of predictors for water vapour line absorption proposed by Hannon et al. [1996] and by Matricardi et al. [2001]. Typically 9, 10 or 11 predictors for water vapour line absorption suffice for accurate radiative transfer simulations.

The role of specific predictors has been identified, enabling the sources of different model error characteristics to be fully described. In this context, a particularly significant outcome of the predictor classification study has been the identification of a lead predictor for water vapour line absorption on the 1400–1650 cm^{-1} interval which is absent from the predictor set proposed by Hannon et al. [1996].

Marked improvements in forward model and Jacobian fitting errors are obtained on introduction of this predictor in regressions.

Caveats do apply however, as this predictor is not a lead predictor for other spectral intervals. Results suggest that it's presence in regressions may degrade fitting accuracy, particularly on the 1200–1400 cm^{-1} interval. The impact of improvements in model performance in the 1400–1650 cm^{-1} interval and possible degradations in the 1200–1400 cm^{-1} interval will therefore be quantified in terms of measurement information content before a final, optimal subset of predictors is defined. By corollary, predictor subset selection will almost certainly have implications for channel selection in the context of data assimilation.

Nonetheless, with the optimal 11 predictor subset identified in this study, we obtain errors in radiative transfer calculations in line centres which are significantly lower than the AIRS instrumental noise specification at the scene temperature *for the first time*. As such, this study has resulted in a major step forward in fast radiative transfer model accuracy.

Acknowledgements

We thank Scott Hannon and Sergio De Souza Machado (UMBC) and Xiaogu Zheng (NIWA) for assistance and discussions during the course of this work. This work was funded under the NSOF contract PDQ023.

References

D. A. Belsley. *Conditioning diagnostics. Collinearity and weak data in regression*. John Wiley and Sons, 1991.

S. Hannon, L. Larrabee Strow, and W. W. McMillan. Atmospheric infrared fast transmittance models: a comparison of two approaches. In *Optical Spectroscopic Techniques and Instrumentation for Atmospheric and Space Research*, pages 94–105, 1996.

M. Matricardi, F. Chevallier, and S. Tjemkes. An improved general fast radiative transfer model for the assimilation of radiance observations. Technical Report 345, European Centre for Medium Range Weather Forecasts, Reading, United Kingdom, 2001.

V. Sherlock, A. Collard, and R. Saunders. Development and validation of Gastropod, a fast radiative transfer model for the advanced sounders. In *Technical Proceedings of the Twelfth International ATOVS Study Conference, Lorne, Australia, 26 February - 05 March, 2002*, 2002.

Predictor	Source		Class	Predictor	Source		Class
	H1996	M2001			H1996	M2001	
aW_{r}	√	√	N	aW_{z}	√		C.1
aW_{r}^2	√	√	N	$\sqrt{aW_{\mathrm{z}}}$	√		C.2
aW_{r}^3	√	√	C.4	aW_{z}^2	√	√	N
aW_{r}^4		√	R	aW_{z}^4		√	R
$\sqrt{aW_{\mathrm{r}}}$	√	√	N	$\sqrt{aW_{\mathrm{r}}}W_r/W_z$		√	N
$\sqrt[4]{aW_{\mathrm{r}}}$	√	√	N	aW_{r}^2/W_z		√	C.3
$aW_{\mathrm{r}}dT$	√	√	N				
$aW_{\mathrm{r}}dT \mid dT \mid$	√	√	N				
$\sqrt{aW_{\mathrm{r}}}dT$	√	√	N				

Table 1: Results from predictor classification studies. For reference, predictors are defined as follows: a is the secant of the satellite zenith angle; W_r is the ratio of the layer mean mixing ratio to the layer mean mixing ratio of a reference profile; W_z is the ratio of the pressure weighted sum of the layer mean mixing ratios to the corresponding weighted sum of the reference profile; dT is the difference between the layer mean temperature and the layer mean temperature of a reference profile.

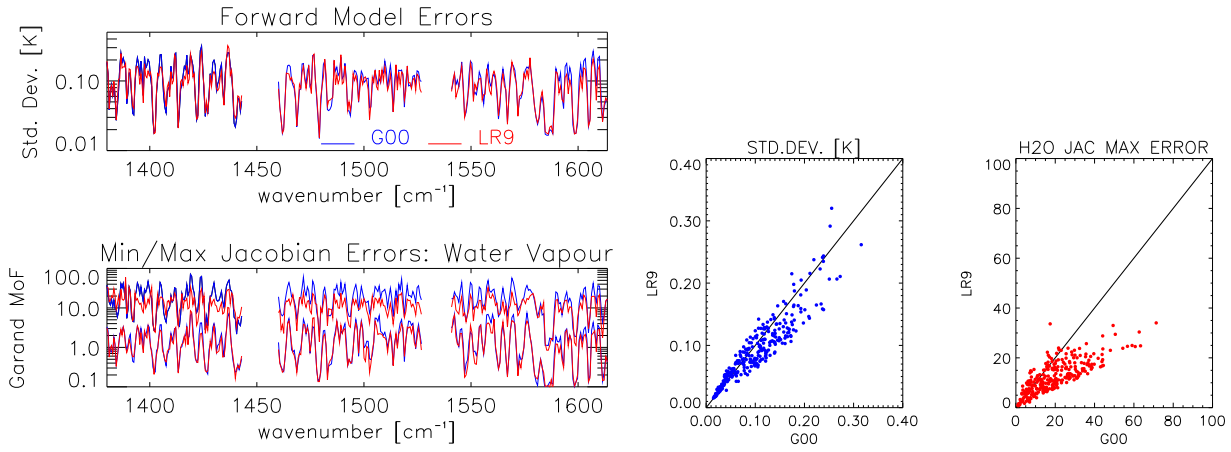

Figure 2: Comparison of forward model and min/max water vapour Jacobian errors (Garand measure of fit) for radiative transfer calculations on the 1400–1650 cm^{-1} interval using the predictors proposed by Hannon et al. [1996] (G00) and the minimum necessary predictor set identified in this study (LR9). Corresponding scatterplots are illustrated for reference.

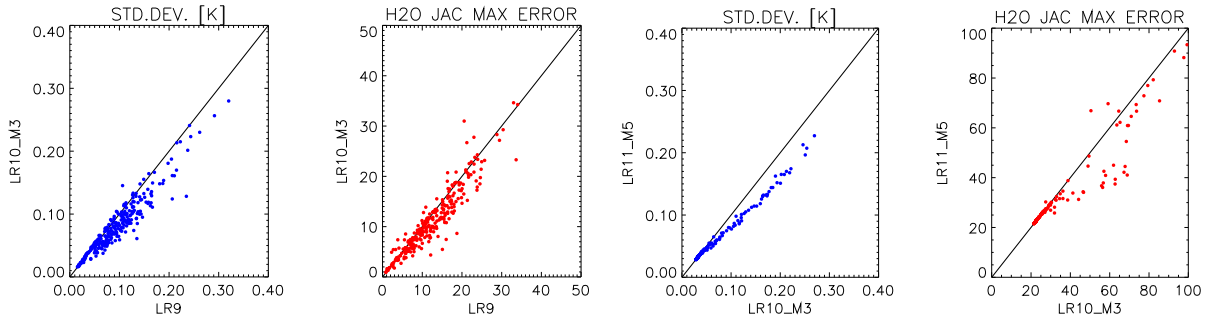

Figure 3: Left: comparison of forward model and maximum water vapour Jacobian errors for radiative transfer calculations on the 1400–1650 cm^{-1} interval on addition of the aW_z predictor to the minimum necessary predictor set (LR10, LR9 respectively). Right: comparison of forward model and maximum water vapour Jacobian errors for radiative transfer calculations in the longwave window region on addition of the $\sqrt{aW_z}$ predictor to the LR10_M3 set.

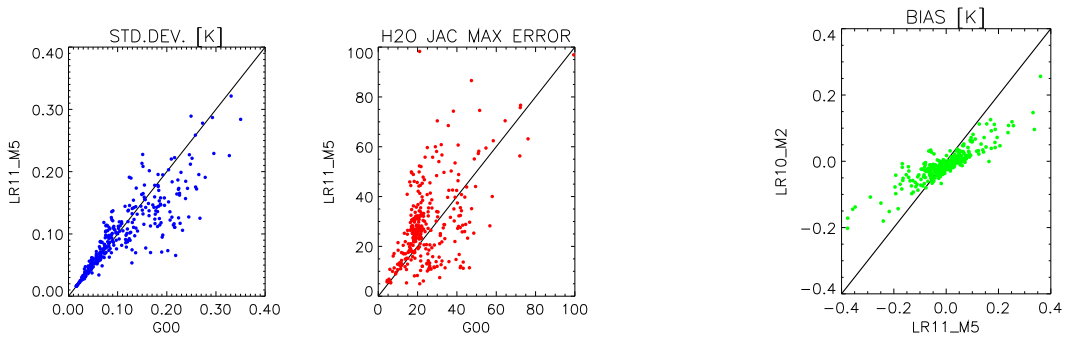

Figure 4: Left: comparison of forward model and maximum water vapour Jacobian errors for radiative transfer calculations on the 1200–1400 cm^{-1} interval using the predictors proposed by Hannon et al. [1996] (G00) and using the LR11_M5 predictor set (= LR9 + aW_z + $\sqrt{aW_z}$) identified in this study. Right hand graphic: comparison of forward model bias for radiative transfer calculations on the 1200–1400 cm^{-1} interval for the LR11_M5 predictor set, and for the LR9 + aW_r^2/W_z (LR10_M2) predictor set.