Appendix B

Time Series Additive-Model Analysis

Lead Authors: John Scinocca & John Austin

Co-authors: Trevor Bailey
Luke Oman
David Plummer
David Stephenson
Hamish Struthers

In this appendix we provide a detailed description of the TSAM analysis, focusing on its development and application to CCMVal-1 and CCMVal-2 ozone-related time series in this chapter. This material is complemented by the supplement to Chapter 9 in which a more complete set of TSAM diagnostics is included, along with an analysis of its sensitivity to outliers and a comparison with the simpler 1:2:1 filtering employed by previous studies of CCMVal-1 time series.

B.1 Multi-Model Ensemble Analysis

The REF2 CCMVal-1 experiment (REF-A2) had a specified integration period of 1980-2050, while the current CCMVal-2 experiment (REF-B2) has a specified integration period of 1960-2100. In each inter-comparison project, ensembles of simulations were also requested. Designing a multi-model analysis of REF-A2 (CCMVal-1) and REF-B2 (CCMVal-2) time series for the purpose of making multi-model trend (MMT) estimates represents a significant challenge due to a number of complicating factors. Particularly,

1. The specified periods for the REF-A2 and REF-B2 experiments are not of equal extent. Furthermore, each modelling centre generally provided a subset of the requested data. For example, individual REF-A2 contributions ranged from ensembles of one, extending over the period 2000-2019, to ensembles of three, extending over the expanded period 1960-2100.
2. In general, large inter-model differences in various latitude bands make it difficult to compare directly the model time series of ozone and chlorine indices, as well as to compute multi-model trend estimates.

Here, we introduce a statistical modelling approach that uses nonparametric regression to estimate smooth

trends from the CCMVal raw data. The nonparametric regression uses a set of optimal thin plate splines to represent the trends and can be used to make formal inference (*e.g.*, calculate confidence and prediction intervals). As discussed in Section 9.2, the approach adopted here consists of three distinct steps: estimation of individual model trends (IMT), baseline adjustment of the trends, and the weighted combination of the individual model trends to produce a multi-model trend (MMT) estimate. In this appendix the development and application of this approach will be illustrated using the time series data presented in **Figure B.1**.

This data corresponds to the CCMVal-1 raw time series in Figure 7 of Eyring *et al.* (2007), which includes both REF-A1 and REF-A2 data for several of the models. The top panel (Figure B.1a) presents the March averaged total column ozone in the latitude band 60°N-90°N, while the bottom panel (Figure B.1b) presents the October averaged total column ozone in the latitude band 60°S-90°S.

B.2 Nonparametric estimation of the individual model trends

The time series $y_{jk}(t)$ of an ozone-related index, such as one of those displayed in Figure B.1, is additively modelled as the sum of a smooth unknown model-dependent trend, $h_j(t)$, and irregular normally-distributed noise:

$$y_{jk}(t) = h_j(t) + \varepsilon_{jk}(t), \qquad \text{(B.1)}$$

where the noise field

$$\varepsilon_{jk}(t) \sim N(0, \sigma^2) \qquad \text{(B.2)}$$

is assumed to be an independent normally distributed random variable with zero mean and variance σ^2, and the indices j and k respectively represent model and ensemble-member number. (Here, the ensemble index k extends over both REF-A1 and REF-A2 simulations for some models.) This is a nonparametric regression of the index on time. The regression is nonparametric because the function of time does not have a fixed functional form with explicit parameters. The noise term (Equation B.2), representing natural variability about the trend, is considered to be an independent normally distributed random variable; independent between different times, models, and runs. The variance of the noise is assumed to be constant over all models and runs. By fitting the trend to all the data rather than to each model separately, one can obtain better estimates of the noise variance (referred to as "borrowing strength").

The unknown smooth functions $h_j(t)$ are estimated by fitting the data to a finite set of smooth basis functions having optimal interpolating properties. This was done here by using the **gam()** function in the **mgcv** library of the R language (R Development Core Team, 2008). The default option was used, which fits the data to a set of thin plate regression splines, by maximising penalized likelihood to find the coefficients multiplying the basis functions. The smoothness of the basis functions is controlled by a smoothing parameter, which is chosen using a leave-one-out generalised cross-validation prediction approach (see Woods (2006) for more details). Unlike iterated 1:2:1 smoothing (*e.g.*, see Section 9S.3 of the supplement to

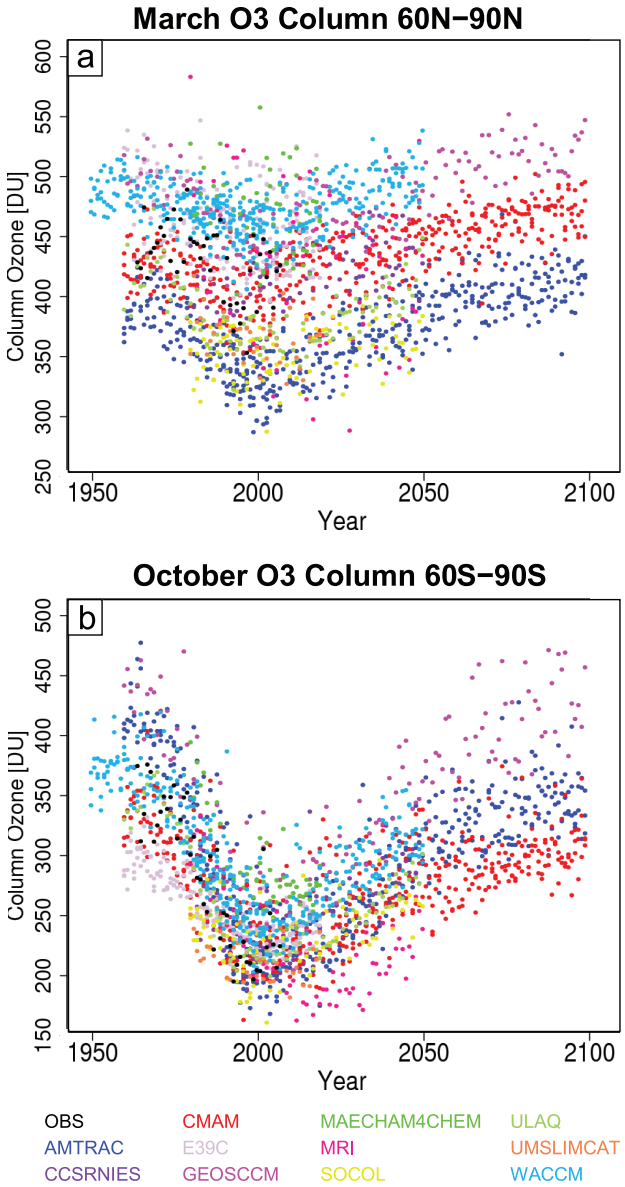

Figure B.1: CCMVal-1 time series of monthly averaged total column ozone in the latitude band 60°N-90°N for March (panel a) and in the latitude band 60°S-90°S for October (panel b). Following Eyring et al. (2007), these time series include REF-A1 data in addition to REF-A2 data for several of the models.

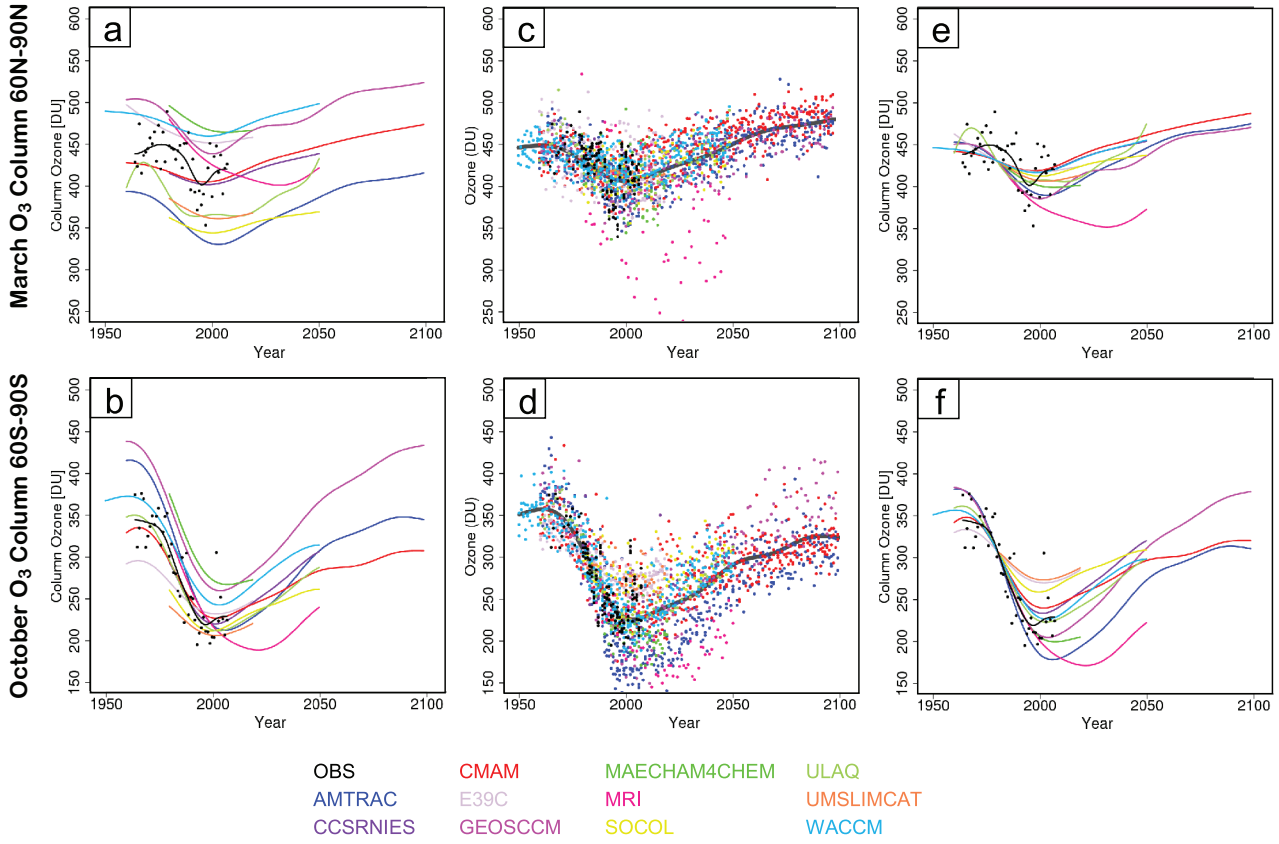

Figure B.2: Panels a and b: The initial estimate of the individual model trends $h_j(t)$ for the raw time series displayed in Figure B.1. This represents the first step in the TSAM analysis. Panels c and d: the 1980 baseline-adjusted time series data y'_{jk} following from (Equation B.7) with t_0 =1980. Panels e and f: The 1980 baseline-adjusted trend estimate $h'_j(t)$. This represents the second step in the TSAM analysis. The thick grey line in panels c and d represents the trend estimate $g'(t)$ for the simpler nonparametric additive model (9.9). For reference, following Eyring et al. (2007) smooth fits to the observations in these plots have been created by 30 iterations of a 1:2:1 filter (black lines).

Chapter 9), the thin plate splines are guaranteed to give smooth trend estimates and do not alter their properties at the ends of the series.

The first step in the TSAM analysis is to apply the nonparametric regression (Equation B.1) to the raw time series data. This is illustrated in panels a and b of **Figure B.2** by the IMT estimates $h_j(t)$ of the CCMVal-1 March 60°N-90°N and October 60°S-90°S total column ozone displayed in Figure B.1. (Note that, while the smooth trend estimates $h_j(t)$ extend over the full period (1950-2100), in Figure B.2a,b we have elected to display the $h_j(t)$ only over the period where data exists for each model.)

B.3 Baseline-adjustment of the trend estimates

The initial IMT estimates $h_j(t)$ in Figure B.2a, b reveal significant differences in the background values of column ozone - particularly in the Arctic (panel a). To facilitate a comparison of the trends across models, anomaly time series are constructed relative to a pre-ozone-hole baseline value of the index. While this is analogous to the procedure employed by Eyring *et al.* (2007), the smoothness of $h_j(t)$ allows a more robust definition of the baseline at a particular time t_0 (*i.e.*, $h_j(t_0)$), rather than from the average over some period about t_0. This results in the anomaly time series:

$$y_{jk}(t) - h_j(t_0). \qquad \text{(B.3)}$$

By construction, the anomaly time series (Equation B.3) is centred on a baseline value of zero at the time t_0. Here, we chose to have this baseline changed from zero to the multi-model mean of $h_j(t_0)$ resulting in the "t_0 baseline-adjusted time series":

$$y'_{jk}(t) = y_{jk} - h_j(t_0) + h(t_0) \qquad \text{(B.4)}$$

where

$$h(t_0) = \text{mean}_j[h_j(t_0)]. \qquad \text{(B.5)}$$

As discussed in Section 9.3, since the multi-model average of the IMT estimates $h(t_0)$ is a close approximation to the final multi-model trend estimate (MMT) derived in the third step of the TSAM analysis, the baseline adjustment may be viewed simply as forcing the anomaly time series to go roughly through the final MMT estimate at the reference date t_0.

The time series (B.4) contains all the information of (B.3) plus the multi-model average $h(t_0)$, which can be compared with observations. In the comparison of CCMVal-1 and CCMVal-2 we have used the baseline t_0 = 1980. Following (B.4), the 1980 baseline-adjusted time series data, y'_{jk} for the CCMVal-1 March 60°N-90°N and October 60°S-90°S total column ozone are displayed in Figure B.2c and d respectively. The corresponding 1980 baseline-adjusted non-parametric IMT estimates $h'_j(t)$ are presented in Figure B.2e and f. Following (B.1) and (B.4) the 1980 baseline-adjusted non-parametric smooth trend in our model is:

$$h'_j(t) = h_j(t) - h_j(t_0) + h(t_0) \qquad \text{(B.6)}$$

with

$$y'_{jk}(t) = h'_j(t) + \varepsilon_{jk}(t). \qquad \text{(B.7)}$$

Before moving on to the third step in the TSAM, we may ask if (9.8) represents one of the simplest models that satises the assumptions of our statistical model (*e.g.*, that the noise term $\varepsilon_{jk}(t)$ is independent from year-to-year, is normally distributed, and is drawn from the same underlying distribution with zero mean and similar variance). For example, we could have chosen the simpler nonparametric model:

$$y'_{jk}(t) = g'(t) + \hat{\varepsilon}_{jk}(t), \qquad \text{(B.8)}$$

where one trend estimate is made for all time series data instead of individual trend estimates for each model (B.7). This implicitly defines a different random noise component $\hat{\varepsilon}_{jk}(t)$. The nonparametric trend estimate $g'(t)$ is displayed as the thick grey line in panels c and d of Figure B.2. If (9.9) were a reasonable model for the data then, in addition to being an IMT, $g'(t)$ could also serve as the MMT thereby eliminating the need for the third step of the TSAM. Visual inspection of the smooth estimate $g'(t)$ to the 1980 baseline-adjusted time series y'_{jk} in Figure B.2c, d would suggest a reasonable fit. However, because we have built the analysis on a probabilistic model, the goodness of the $g'(t)$ and $h'_j(t)$ fits may be tested against the model's underlying assumptions.

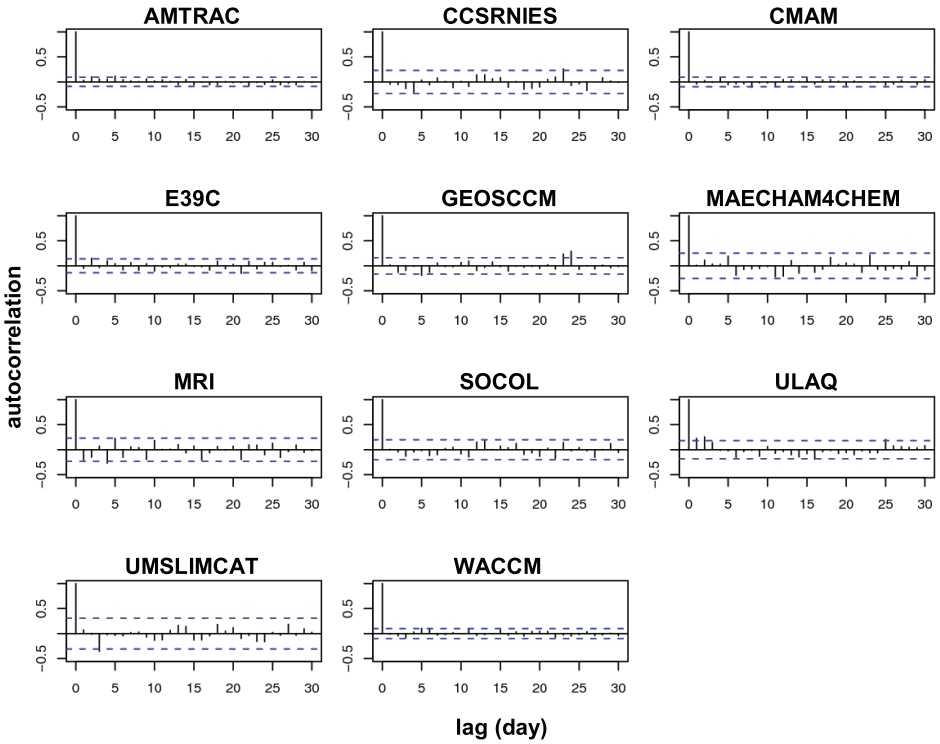

Figure B.3: Individual model autocorrelation functions for the residuals $\varepsilon_{jk}(t)$ for CCMVal-1 October total column ozone in the latitude band 60°S-90°S. This noise corresponds to the nonparametric model (9.8) with 1980 baseline trend estimates $h_j(t)$ displayed in Figure B.2f. The blue dashed lines represent 96% confidence limits for the sample autocorrelation function. This suggests that the assumption of year-to-year independence is a good one for the (B.7) model.

The year-to-year independence of the model noise term may be tested by calculating its autocorrelation function. In **Figure B.3** the autocorrelation function for the noise term $\varepsilon_{jk}(t)$ is displayed for each model for the nonparametric fit (B.7) to the CCMVal-1 October 60°S-90°S column ozone. The dashed blue lines in this figure represent 95% confidence limits. Lines that extend beyond these limits are considered to be sample correlations that are significantly different from zero. Inspection of all the models reveals that the assumption of year-to-year independence is a good one for the model (B.7). This is not, however, the case for the simpler model (B.8). The autocorrelation of the noise term $\hat{\varepsilon}_{jk}(t)$ is displayed in **Figure B.4** and displays significant violations of the assumption of year-to-year independence for several of the models.

Model assumptions related to the noise term may be further investigated by "notched box-and-whisker" plots. These are displayed for $\hat{\varepsilon}_{jk}(t)$ and $\varepsilon_{jk}(t)$ respectively in panels a and b of **Figure B.5** again for the CCMVal-1 October 60°S-90°S column ozone (see caption for details). From panel b we can see that the noise term $\varepsilon_{jk}(t)$ has a similar location and scale for each model, validating the model assumption that the residuals were drawn from the same distribution with zero mean and roughly the same variance. Again, the same cannot be said for the $\varepsilon_{jk}(t)$ residuals (panel a) suggesting that $g'(t)$ in (B.9) is not a good estimate of the trend.

We conclude, therefore, that (B.7) represents one of the simplest nonparametric additive models that is satisfied by the ozone indices considered in the two examples. (The same is basically true for the remainder of ozone-related indices analysed in Chapter 9).

B.4 Multi-model trend estimates

The final step of the TSAM analysis involves combining the IMT estimates $h'_j(t)$ to arrive at an MMT estimate:

$$h'(t) = \sum_j w_j(t) h'_j(t), \tag{B.9}$$

where the weights $w_j(t)$ have the properties

$$w_j(t) \geq 0 \text{ and } \sum_j w_j(t) = 1. \tag{B.10}$$

If the weights are assumed to be non-random, and the errors in the individual trends are assumed to be independent, then the standard error of the weighted sum is given by:

$$s^2_h(t) = \sum_j w^2_j(t) s^2_j(t), \tag{B.11}$$

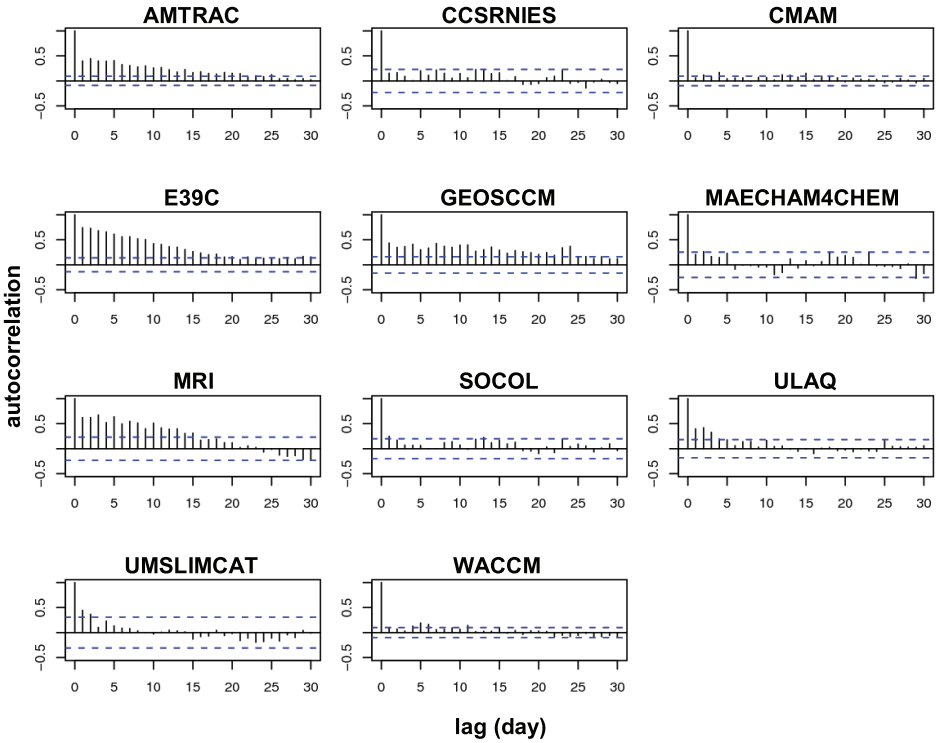

Figure B.4: Individual model autocorrelation functions for the noise term $\hat{\varepsilon}_{jk}(t)$ for CCMVal-1 October total column ozone in the latitude band 60°S-90°S. This noise corresponds to the simpler nonparametric model(9.9) with a 1980 baseline trend estimate g(t) displayed in Figure B.2d. The lines extending past the blue-dashed lines for several models indicates that the assumption of year-to-year independence is not well satisfied for the (B.8) model.

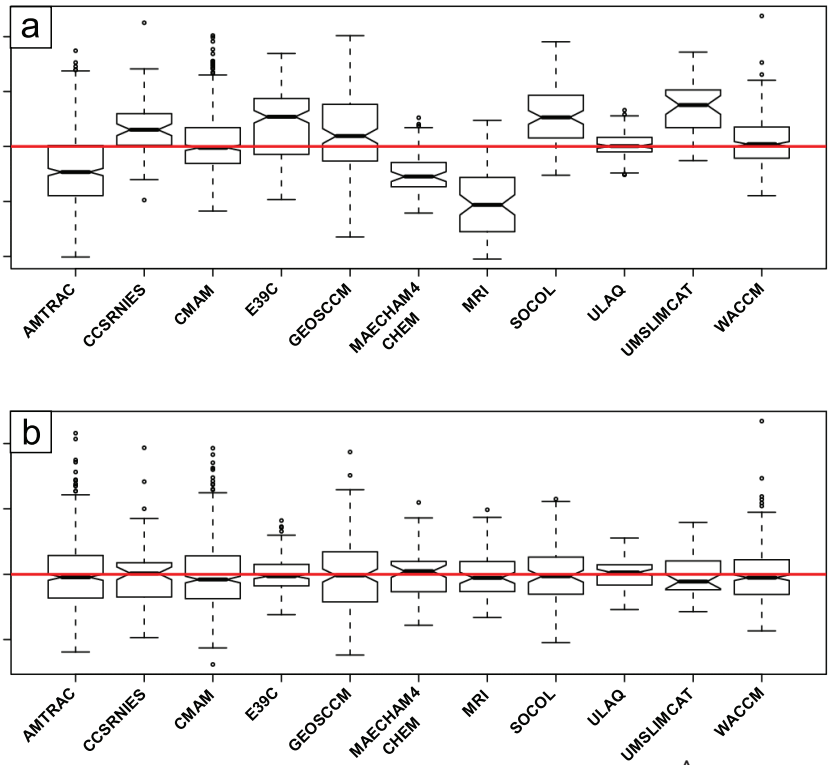

Figure B.5: Individual model notched box-and-whisker plots for the noise term $\hat{\varepsilon}_{jk}(t)$ corresponding to the simpler nonparametric additive model (B.8) (panel a) and for the noise term $\varepsilon_{jk}(t)$ corresponding to the nonparametric additive model (B.7). These apply to the CCMVal-1 October total column ozone in the latitude band 60°S-90°S. In these plots the central black line represents the median, the extent of the notches away from the median line indicates the 95% confidence interval of the median, the top and bottom of the boxes respectively represent the upper and lower quartiles, and the top and bottom whiskers extend out to 1.5 times the distance from the first to third quartiles. For the noise term $\varepsilon_{jk}(t)$ (panel b) the medians of all models fall within the notches and are close to zero. Also, the similar height of the boxes indicates that all models have a similar amount of variance away from the estimated trend $h'_j(t)$. For the noise term $\hat{\varepsilon}_{jk}(t)$, the means are significantly different and the inter-model variance is larger suggesting that (B.8) is not a suitable model for this data.

where $s_j(t)$ is the standard error of the trend estimate $h'_j(t)$, which can be calculated using standard expressions from linear regression (Woods, 2006). The standard error (B.11) can then be used to estimate the confidence and prediction intervals respectively as:

$$[h'(t) - 1.96s_h(t), h'(t) + 1.96s_h(t)] \qquad \text{(B.12)}$$

and

$$[h'(t) - 1.96\sqrt{(s^2_h(t) + s^2_\varepsilon)}, h'(t) + \sqrt{(s^2_h(t) + s^2_\varepsilon)}]. \qquad \text{(B.13)}$$

The 95% confidence interval in the trend gives the uncertainty in the trend estimate. In other words, there is a 95% chance that this interval will overlap the true trend. The interval is point-wise (rather than simultaneous) in that it represents the uncertainty in the trend at each year rather than being an interval for all probable trend curves over the whole period. The 95% prediction gives an idea of how much uncertainty their might be in a predicted index value for a particular year. In other words, there is a 95% chance that a particular index value on a specific year will lie in this interval. This interval is the combination of uncertainty in the trend estimate and the uncertainty due to natural interannual variability about the trend.

The specific choice of weights in (B.9) remains open. In general, we decide to base the construction of the weights on a statistical probability model with testable assumptions. Here, we have chosen a "random-effects" model to determine the weights. This model assumes that the trends for individual models $h'_j(t)$ are random samples from a "true" trend $\sim h'_j(t)$:

$$h'_j(t) = \sim h'_j(t) + \eta(t) \qquad \text{(B.14)}$$

where

$$\eta(t) \sim N(0, \lambda^2). \qquad \text{(B.15)}$$

The quantity λ^2 is included to account for additional variance between model trends that cannot be accounted for merely by sampling the uncertainty s^2_j. Using this random

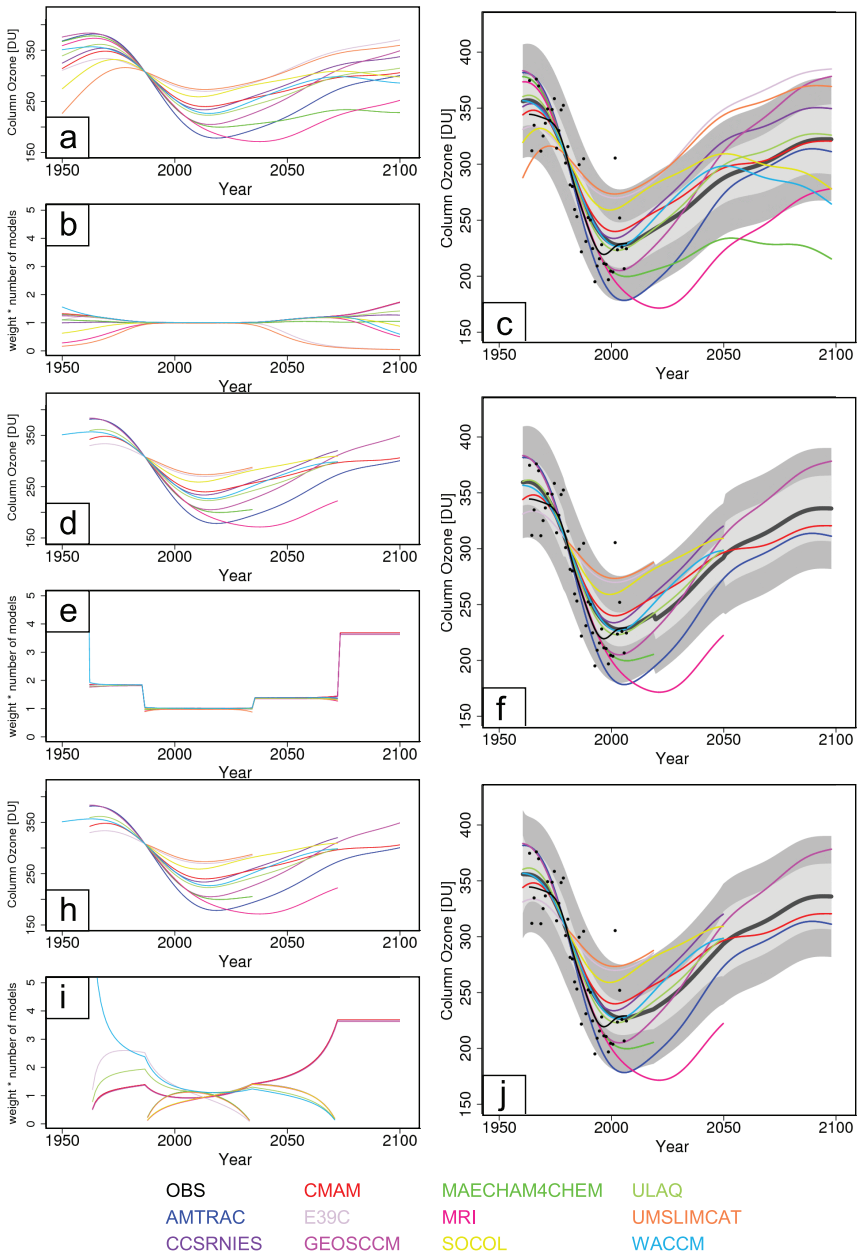

Figure B.6: For time series of CCMVal-1 October total column ozone in the latitude band 60°S-90°S are presented the individual model fits (panels a, d, and h), weights (panels b, e, and i), and trend (MMT) estimate (thick grey line in panels c, f, and j) for three approaches to determining the weights. Results from the "random-effects" model (B.17) are shown in panels a-c. One problem with this approach is that models can contribute to the final MMT estimate at times when no data exists of that model (i.e., in regions where $h'_j(t)$ represents an extrapolation). The introduction of prior weights (9.21) can help mitigate this problem. Results from the use of a simple on/offset of prior weights (having values of one where there is model data and zero where there is none) are presented in panels d-f. One artifact of this approach is that it causes discontinuities in the final MMT estimate. Finally, results from set of prior weights used for the present chapter, which employ a smoother quadratic taper from a value of 1 where time series data exists to a value of 0 where it is absent, is displayed in panels h-j.

effects model, (B.11) then generalises to:

$$s^2_h(t) = \sum_j w^2_j(t)(\lambda^2 + s^2_j(t)), \qquad \text{(B.16)}$$

which is used here to calculate intervals. Assuming this model is valid, a least-squares estimate of may be obtained from (B.9) employing the weights:

$$w_j(t) = w(t)/(\lambda^2 + s^2_j(t)) \qquad \text{(B.17)}$$

where

$$w^{-1}(t) = \sum_j (\lambda^2 + s^2_j(t))^{-1}. \qquad \text{(B.18)}$$

Specification of the weights $w_j(t)$ from (B.17) requires an estimate of the parameter λ^2. For this we have used the following iterative approach: An initial estimate of the true trend is obtained by calculating $h'_{\lambda=0}(t)$. Then, an iterative Newton-Raphson algorithm is employed to determine the λ that gives scaled residuals that have unit variance as is expected from (B.14):

$$\mathrm{var}\left(\frac{h'_j(t) - h'_{\lambda=0}(t)}{\sqrt{\lambda^2 + s^2_j(t)}} \right) = 1. \qquad \text{(B.19)}$$

Employing this model for the weights produces the MMT estimate $h'(t)$ for the 1980 baseline CCMVal-1 October 60°S-90°S column ozone displayed in **Figure B.6c**. The associated individual model trend estimates $h'_j(t)$ and weights $w_j(t)$ are respectively displayed in panels a and b of this figure. In this figure, the weights are scaled by the number of models so that a scaled weight of 1 implies a proportional contribution of that model to the MMT estimate.

While this formulation of weights provides a smooth final trend estimate $h'(t)$, for this example it highlights a potential problem - the individual model weights $w_j(t)$ are very insensitive to the absence of data in the original time series. For example, the time series for the MAECHAM4CHEM model (green) extends only over the period 1980-2019 (see Figure B.2). Its scaled weight, however, has a value of roughly 1 over the entire period 1960-2100 suggesting significant contributions of its trend estimate $h'_j(t)$ at times when there are no model data. The original idea behind this model for the weights was that the natural increase in standard errors $s^2_j(t)$ in the region where $h'_j(t)$ is extrapolated beyond the model data would cause the weights to decrease naturally towards zero. While Figure B.6b indicates that there is some tendency for the weights to display this behaviour, it clearly remains unphysical.

To correct this unphysical behaviour, we introduce the concept of prior weights $w^p_j(t)$ into the formulation such that the final weights now have the form:

$$w^p_j(t) = \frac{w^p_j(t) w_j(t)}{\sum_{j'} w^p_{j'}(t) w_{j'}(t)}, \qquad \text{(B.20)}$$

(with $w'_j(t)$ implicitly replacing $w_j(t)$ in expressions (B.11) and (B.17)). An example set of prior weights would be the "on/off" set: $w^p_j(t) = 1$ at times t when raw time series data exist for model j and $w^p_j(t) = 0$ otherwise. This prescription is illustrated in panels d-f of Figure B.6. It corrects the unphysical behaviour identified when $w_j(t)$ of (B.17) is used alone. However, this on/off prescription is still problematic in that it causes discontinuities in the MMT estimate Figure B.6f. The set of prior weights used for the Chapter 9 employs a smoother quadratic taper, from a value of 1 where time series data exists to a value of 0 where it is absent:

$$w^p_j(t) = \begin{cases} 1 - z^2 & \text{if } 0 \le z^2 \le 1 \\ 0 & \text{otherwise} \end{cases}, \qquad \text{(B.21)}$$

where

$$z = -1 + 2(t - t_{j,min})/(t_{j,max} - t_{j,min}), \qquad \text{(B.22)}$$

and where $[t_{j,min}, t_{j,max}]$ defines the period within which data exist for model j. This scheme is illustrated in panels h-j of Figure B.6.

Finally, the formulation of prior weights (B.20) allows a natural entry point for the specification of prior, time-independent, model weights based on performance metrics. Such metric based weights would take on values in the range [0, 1] and simply multiply $w^p_j(t)$ in the expression (B.20).

References

Eyring, V., D. W. Waugh, G. E. Bodecker, E. Cordero, H. Akiyoshi, J. Austin, S. R. Beagley, B. A. Boville, P. Braesicke, C. Brhl, N. Butchart, M. P. Chippereld, M. Dameris, R. Deckert, M. Deushi, S. M. Frith, R. R. Garcia, A. Gettelman, M. A. Giorgetta, D. E. Kinnison, E. Mancini, E. Manzini, D. R. Marsh, S. Matthes, T. Nagashima, P. A. Newman, J. E. Nielsen, S. Pawson, G. Pitari, D. A. Plummer, E. Rozanov, M. Schraner, J. F. Scinocca, K. Semeniuk, T. G. Shepherd, K. Shibata, B. Steil, R. S. Stolaarski, W. Tian, and M. Yoshiki, 2007: Multi-model projections of ozone recovery in the 21st century, 2007. *J. Geophys. Res.*, **112**, doi:10.1029/2006JD008332.

R Development Core Team, 2008. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Woods, S.N., 2006. Generalized Additive Models: An Introduction with R, Chapman & Hall/CRC, London.