Supplement of Atmos. Chem. Phys., 21, 10557–10587, 2021
https://doi.org/10.5194/acp-21-10557-2021-supplement

Supplement of

Forest-fire aerosol–weather feedbacks over western North America using a high-resolution, online coupled air-quality model

Paul A. Makar et al.

Correspondence to: Paul A. Makar (paul.makar@canada.ca)

This PDF file includes:

Section S1 / Table S1: Model Evaluation Table S1 and Supplemental Text

Section S2: Supplemental Text: 90% Confidence Interval Methodology

Tables S2, S3, S4, S5: Additional evaluations (referred to in the main paper text).

Section S1: Model Evaluation Statistics Description

Table S1: Model Evaluation Statistics Formulae

Metric and Formula	Range	Ideal Score
Factor of 2 (FO2) = Number of (O_i, M_i) pairs for which $\frac{1}{2} \leq \frac{M_i}{O_i} \leq 2$	[0,1]	1
$Mean\ Bias\ (MB) = \frac{1}{N}\sum(M_i - O_i) = \bar{M} - \bar{O}$		0
$Mean\ Gross\ Error\ (MGE) = \frac{1}{N}\sum\lvert M_i - O_i\rvert$		0
$Normalized\ Mean\ Gross\ Error\ (NMGE) = \frac{\sum\lvert M_i - O_i\rvert}{\sum O_i}$		0
$Pearson\ Correlation\ Coefficient\ (R) = \frac{\sum(M_i - \bar{M})(O_i - \bar{O})}{\sqrt{\sum(M_i - \bar{M})^2\sum(O_i - \bar{O})^2}}$	[-1,1]	1
$Root\ Mean\ Square\ Error\ (RMSE) = \sqrt{\frac{1}{N}\sum(M_i - O_i)^2}$		0
$Coefficient\ of\ Efficiency\ (COE) = 1 - \frac{\sum\lvert M_i - O_i\rvert}{\sum\lvert O_i - \bar{O}\rvert}$	[-∞, 1]	1
$Index\ of\ Agreement\ (IOA) = \begin{cases} 1 - \frac{\sum\lvert M_i - O_i\rvert}{2\sum\lvert O_i - \bar{O}\rvert}, when \sum\lvert M_i - O_i\rvert \leq 2\sum\lvert O_i - \bar{O}\rvert \\ \frac{2\sum\lvert O_i - \bar{O}\rvert}{\sum\lvert M_i - O_i\rvert} - 1, when \sum\lvert M_i - O_i\rvert > 2\sum\lvert O_i - \bar{O}\rvert \end{cases}$	[-1,1]	1

Supplemental Text: Description of each Model Performance Metric

Factor Of 2 (FO2)

The FO2 provides a measure of the scatter of model-measurement pairs. An advantage of this metric is that it is intuitively easy to interpret, and another advantage of this metric is that equally applicable over model and observed values which may vary by several orders of magnitude; it is not unduly influenced by outliers.

Mean Bias (MB)

The mean bias (Ali and Abustan, 2014; Fox, 1981) is the difference between the model and observed means, and thus provides a measure of the ability of the model to represent the average behaviour of the model. This is a desirable performance measure for model species used to estimate cumulative impacts of air quality, such as chronic health exposures and totals of acidifying deposition, where the long-term behaviour of the model has a greater degree of relevance. However, the mean bias does not provide a measure of the ability of the model to capture specific events (if the averaging is over time), or identify specific regions where the model chronically over- or under-predicts the observations (if the averaging is also over space – though the latter may be alleviated by constructing station-specific mean biases). The MB values for comparisons where model and observed values vary by many orders of magnitude will be unduly influenced by the larger numbers in the range of magnitudes. Another disadvantage of MB is the potential for the cancellation of large positive and negative outliers giving a false impression of high model performance.

Mean Gross Error (MGE)

The mean gross error (also known as the mean absolute error) quantifies the magnitude of the deviation between the model and observations as the average of those magnitudes across all model pairs (Ali and Abustan, 2014; Fox, 1981). The MGE removes the disadvantage of the MB of the potential cancellation of the impact of positive and negative deviations. However, the MGE may be unduly influenced by outliers. The MGE values for comparisons where model and observed values vary by many orders of magnitude will be unduly influenced by the larger numbers in the range of magnitudes.

Normalized Mean Gross Error (NMGE)

The normalized mean gross error is a refinement of the MGE, in which the latter is scaled by the observed mean. Unlike the MGE, it has the advantage of being dimensionless, allowing performance comparisons across different model variables. However, it may still be unduly influenced by outliers, and particularly by the larger numbers when model and observed values vary by many orders of magnitude.

Correlation Coefficient (R)

The Pearson product-moment correlation coefficient measures the degree of linear dependence between two variables, with a value of zero indicating no linear dependence, and a value of -1 or 1 indicating a perfect linear dependence (negative values indicating a negative dependence). While having the favourable properties of being dimensionless and bounded between -1 and 1. The squared terms in R also result in an undesirable influence of outliers on the resulting metric values, potentially producing a false sense of relationship. Another disadvantage of R is that while measuring the strength of the relationship between model and observations, it gives no indication of whether the two data series have a similar magnitude (Duveiller *et al*., 2016). Other potential disadvantages of R may include its assumption that all observations are independent (which may result in erroneous relationship if identical measurements are reported at co-located monitoring stations), that two clusters of model-observation pairs may provide a false sense of relationship, and that false relationships resulting from data on a non-continuous scale (e.g. comparisons between air-quality health index values which depend on exponential variation are better estimated by a Spearman's rank correlation method) (Aggarwal and Ranganathan, 2016).

Root Mean Square Error (RMSE)

The root mean square error quantifies the dispersion between model and observations as the standard deviation of the residuals; the deviations between the best-fit line of the linear regression between model and observations. The RMSE thus gives an estimate of the typical separation between the best-fit line and the model values. While a common metric for estimating the scatter of the model values about the assumed linear relationship between model and observations, RMSE, like the correlation coefficient, has the disadvantage of being potentially unduly influenced by outliers and by the larger magnitude values when both model and observations range over multiple orders of magnitude.

Coefficient of Efficiency (COE)

The COE defined in Table S1 is the E_1 modified coefficient of efficiency as described by Legates and McCabe (1999), and provides a measure of the deviation of model values to the observed mean relative to the deviation of the observed values to the observed mean An advantage of the COE as defined in Table S1 is that it does not have a dependence on squared terms – this reduces the sensitivity of the metric to outliers which increases the sensitivity of the metric to outliers. COE (and IOA, which follows) are also sensitive to additive and proportional differences between model predictions and observations. The reduced sensitivity to outliers and increased sensitivity to additive and proportional differences are advantages of COE and IOA relative to correlation-based metrics such as R. A disadvantage of the COE is that it is unbounded in the negative direction, which makes poorly performing models more difficult to directly compare.

Index of Agreement (IOA)

The IOA defined in Table S1 follows Willmott *et al*. (2012), is a revision of Willmott's earlier work, and provides a measure of deviations between the model-predicted and observed deviations about the observed mean values which is not influenced by extreme outliers in the data. This revised form has the advantage of having a finite lower bound, which allows a more uniform assessment of poorly performing models. IOA may be interpreted as the ratio of two sums; that of the magnitudes of the differences between the model-predicted and observed deviations about the observed mean, to that of the perfect model and observed deviations about the observed mean. As the deviation between the model-predicted and observed deviations about the observed mean become large, IOA approaches -1, as they become small, IOA approaches 1. A second advantage of this IOA implementation is that the effects of extrema are not exacerbated due to the use of absolute values, rather than squaring, in the numerator and denominator.

References for Model Performance Metrics:

Ali, M.H., and Abustan, I., A new novel index for evaluating model performance, J. Nat. Res. and Dev.,4, 1-9, 2014.

Aggarwal, R., and Ranganathan, P., Common pitfalls in statistical analysis: the use of correlation techniques, Perspect. Clin. Res., 187-190, 2016.

Duveiller, G., Fasbender, D, and Meroni, M., Revisiting the concept of a symmetric index of agreement for continuous datasets, Sci. Rep., 6, 19401; doi: 10.1038/srep19401, 2016.

Fox, D.G., Judging air quality model performance – a summary of the AMS workshop on dispersion model performance, Bull. Am. Met. Soc., 62, 599-609, 1981.

Legates, D.R., and McCabe, G.J. Jr, Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation, Water Resources Res., 35, 233-241, 1999.

Willmott, C.J., Robeson, S.M., Matsuura, K., A refined index of model performance, Int. J. Climatol., 32, 2088-2094, 2012.

Section S2. Supplemental Text: 90% Confidence Interval Methodology

At each model grid cell the values of the standard deviation about the mean for each respective simulation was calculated. The difference between the means becomes significant at a given confidence level c if the regions defined by $M_f \pm z^* \frac{\sigma_f}{\sqrt{N}}$ and $M_{nf} \pm z^* \frac{\sigma_{nf}}{\sqrt{N}}$d do not overlap, where N is the number of points averaged, M_f is the feedback mean value, M_{nf} is the no-feedback mean value, σ_f and σ_{nf} are the corresponding standard deviation, and z^* is the value of the $\sqrt{c}$ percentile point for the fractional confidence interval c of the normal distribution ($z^* = 1.645$ at $c = 0.90$). Grid cell values where this overlap does not occur (i.e. where the mean values differ at or above the 90% confidence level) may be defined via the following equation:

$$\frac{|M_{nf} - M_f|}{\frac{z^*}{\sqrt{N}}(\sigma_f + \sigma_{nf})} > 1 \qquad (1)$$

The differences in the mean grid cell values between the simulations for which the above quantity is greater than unity thus differ at or greater than the 90% confidence level.

Note that for cases where $M_{nf} > M_f$, significance at the confidence level associated with z^* occurs when the range of standard errors about the mean do not overlap, ie. $M_{nf} - z^* \frac{\sigma_{nf}}{\sqrt{N}} > M_f + z^* \frac{\sigma_f}{\sqrt{N}}$, or $\left(M_{nf} - M_f\right)/\left(\frac{z^*}{\sqrt{N}}\left(\sigma_{nf} + \sigma_f\right)\right) > 1$. Similarly, for cases where $M_f > M_{nf}$, significance at the confidence level associated with z* occurs when $\left(M_f - M_{nf}\right)/\left(\frac{z^*}{\sqrt{N}}\left(\sigma_{nf} + \sigma_f\right)\right) > 1$. Equation (1) may thus be used to describe both conditions.

Table S2: Summary performance metrics for ozone, nitrogen dioxide, and PM2.5, MOZART2009 boundary conditions. Bold-face indicates the simulation with the better performance score for the given metric, chemical species and sub-region, italics indicate a tied score, and regular ont the simulation with the lower performance score. FO2: fraction of scores within a factor of 2. MB: Mean Bias. MGE: Mean Gross Error. R: Correlation Coefficient. RMSE: Root Mean Square Error. COE: Coefficient of Error. IOA: Index of Agreement.

Chemical	Region	Simulation	FO2	MB	MGE	NMGE	R	RMSE	COE	IOA
PM2.5	Western Canada	No Feedback	0.451	0.777	4.335	0.878	**0.278**	7.642	-0.669	0.165
		Feedback	**0.453**	**0.236**	**4.215**	**0.829**	0.219	**6.976**	**-0.534**	**0.233**
	Western USA	No Feedback	**0.536**	**-1.639**	3.793	0.605	0.358	6.061	-0.166	0.417
		Feedback	0.524	-1.786	**3.773**	**0.602**	**0.361**	**5.978**	**-0.162**	**0.419**
O_3	Western Canada	No Feedback	0.773	-3.683	7.76	0.347	0.634	9.996	0.138	0.569
		Feedback	**0.78**	**-3.553**	**7.693**	**0.344**	**0.635**	**9.854**	**0.145**	**0.573**
	Western USA	No Feedback	0.879	-3.584	9.667	0.257	*0.763*	12.534	0.319	0.66
		Feedback	**0.881**	**-3.456**	**9.607**	**0.256**	*0.763*	**12.458**	**0.322**	**0.661**
NO_2	Western Canada	No Feedback	**0.528**	**-0.417**	2.29	0.594	**0.519**	3.353	0.053	*0.527*
		Feedback	0.52	-0.533	**2.274**	**0.591**	0.513	**3.314**	**0.055**	*0.527*
	Western USA	No Feedback	**0.428**	0.669	2.278	0.759	**0.585**	3.917	-0.161	0.419
		Feedback	0.427	**0.578**	**2.24**	**0.746**	0.581	**3.858**	**-0.141**	**0.429**

Table S3 : Comparison of summary performance metrics for PM2.5, for GEM-MACH 2.5km simulations and ECMWF reanalysis files (Western Canada, Western USA), and over all of North America (ECMWF Reanalysis Only).

Region	FO2	MB	MGE	NMGE	R	RMSE	COE	IOA
Statistics for Western Canada portion of 2.5km domain:								
GEM-MACH 2.5km, Feedback Forecast, B.C.: MOZART 2009	0.453	0.236	4.215	0.829	0.219	6.976	-0.534	0.233
GEM-MACH, 2.5km, Feedback Forecast, B.C.: ECMWF + GEM-MACH 10km	0.414	4.578	6.531	1.291	0.238	9.803	-1.418	-0.173
ECMWF Reanalysis	0.231	11.170	11.701	2.522	0.213	18.674	-3.761	-0.580
Statistics for Western USA portion of 2.5km domain:								
GEM-MACH 2.5km, Feedback Forecast, B.C.: MOZART 2009	0.524	-1.786	3.773	0.602	0.361	5.978	-0.162	0.419
GEM-MACH, 2.5km, Feedback Forecast, B.C.: ECMWF + GEM-MACH 10km	0.556	1.805	5.287	0.813	0.252	8.443	-0.520	0.240
ECMWF Reanalysis	0.430	4.421	8.072	1.172	0.036	20.430	-1.070	-0.034
Statistics for North America (ECMWF Reanalysis Only):								
ECMWF Reanalysis	0.462	7.684	10.279	1.181	0.221	22.287	-1.176	-0.081

Table S4 : Comparison of summary performance metrics for O_3, for GEM-MACH 2.5km simulations and ECMWF reanalysis files (Western Canada, Western USA), and over all of North America (ECMWF Reanalysis Only).

Region	FO2	MB	MGE	NMGE	R	RMSE	COE	IOA
Statistics for Western Canada portion of 2.5km domain:								
GEM-MACH 2.5km, Feedback Forecast, B.C.: MOZART 2009	0.780	-3.553	7.693	0.344	0.635	9.854	0.145	0.573
GEM-MACH, 2.5km, Feedback Forecast, B.C.: ECMWF + GEM-MACH 10km	0.745	5.891	10.969	0.490	0.527	15.268	-0.210	0.395
ECMWF Reanalysis	0.754	8.487	10.838	0.493	0.607	13.148	-0.244	0.378
Statistics for Western USA portion of 2.5km domain:								
GEM-MACH 2.5km, Feedback Forecast, B.C.: MOZART 2009	0.881	-3.456	9.607	0.256	0.763	12.458	0.322	0.661
GEM-MACH, 2.5km, Feedback Forecast, B.C.: ECMWF + GEM-MACH 10km	0.866	1.770	10.663	0.284	0.694	14.225	0.252	0.626
ECMWF Reanalysis	0.835	4.590	12.045	0.326	0.637	14.972	0.120	0.560
Statistics for North America (ECMWF Reanalysis only):								
ECMWF Reanalysis	0.792	9.308	13.118	0.425	0.685	16.054	-0.010	0.495

Table S5: Comparison of summary performance metrics for NO_2, for GEM-MACH 2.5km simulations and ECMWF reanalysis files (Western Canada, Western USA), and over all of North America (ECMWF Reanalysis Only).

Region	FO2	MB	MGE	NMGE	R	RMSE	COE	IOA
Statistics for Western Canada portion of 2.5km domain:								
GEM-MACH 2.5km, Feedback Forecast, B.C.: MOZART 2009	0.520	-0.533	2.274	0.591	0.513	3.314	0.055	0.527
GEM-MACH, 2.5km, Feedback Forecast, B.C.: ECMWF + GEM-MACH 10km	0.429	-1.037	2.758	0.595	0.565	3.936	0.154	0.577
ECMWF Reanalysis	0.310	-1.333	2.876	0.700	0.405	3.981	-0.080	0.460
Statistics for Western USA portion of 2.5km domain:								
GEM-MACH 2.5km, Feedback Forecast, B.C.: MOZART 2009	0.427	0.578	2.24	0.746	0.581	3.858	-0.141	0.429
GEM-MACH, 2.5km, Feedback Forecast, B.C.: ECMWF + GEM-MACH 10km	0.483	-0.427	2.332	0.570	0.651	3.657	0.180	0.590
ECMWF Reanalysis	0.379	-0.393	3.184	0.765	0.354	4.865	-0.065	0.467
Statistics for North America (ECMWF Reanalysis only):								
ECMWF Reanalysis	0.384	-1.210	4.197	0.707	0.455	6.572	0.099	0.549