Forest Fire Aerosol – Weather Feedbacks over Western North America Using a High-Resolution, ~~Fully~~ On-line Coupled, Air-Quality Model

Paul A. Makar[1], Ayodeji Akingunola[1], Jack Chen[1], Balbir Pabla[1], Wanmin Gong[1], Craig Stroud[1], Christopher Sioris[1], Kerry Anderson[2], Philip Cheung[1], Junhua Zhang[1], Jason Milbrandt[3]

[1]Air Quality Research Division, Atmospheric Science and Technology Directorate, Environment and Climate Change Canada, 4905 Dufferin Street, Toronto, M3H 5T4, Canada
[2]Natural Resources Canada (emiritus), Summerland, British Columbia, Canada.
[3]Meteorological Research Division, Atmospheric Science and Technology Directorate, Environment and Climate Change Canada, 2121 Trans Canada Highway, Montreal, Canada

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

Abstract. The influence of both anthropogenic and forest fire emissions, and their and subsequent chemical and physical processing, on the accuracy of weather and air-quality forecasts, was studied using a high resolution, ~~fully~~ on-line coupled air-quality model. Simulations were carried out for the period 4 July through 5 August 2019, at 2.5-km horizontal grid cell size, over a 2250 x 3425 km^2 domain covering western Canada and USA, prior to the use of the forecast system as part of the FIREX-AQ ensemble forecast. Several large forest fires took place in the Canadian portion of the domain during the study period. A feature of the implementation was the incorporation of a new on-line version of the Canadian Forest Fire Emissions Prediction System (CFFEPSv4.0). This inclusion of thermodynamic forest fire plume-rise calculations directly into the on-line air-quality model allowed us to simulate the interactions between forest fire plume development and weather.

Incorporating feedbacks resulted in weather ~~improvements in most metrics of both air-quality and meteorological model~~ forecast performance that exceeded or matched the ~~, through comparison of~~ no-feedback forecast, ~~and feedback simulations with surface, radiosonde, and satellite observations. For the meteorological simulations, these improvements occurred~~ at greater than ~~the~~ 90% confidence, at most times and heights in the atmosphere. The feedback forecast out-performed the feedback forecast at 35 out of 48 statistical evaluation scores, for PM2.5, NO_2 and O_3 ~~level~~. Relative to the climatological cloud condensation nuclei and aerosol optical properties used in the no-feedback simulations, the ~~fully~~ on-line coupled model's aerosol indirect and direct effects were shown to result in feedback loops characterized by ~~increased~~ decreased surface temperatures in regions affected by forest fire plumes, decreases in stability within the smoke plume, increases in stability further aloft ~~, decreased lower troposphere temperatures~~, and increased lower troposphere cloud droplet and raindrop number densities. The aerosol direct and indirect effect reduced oceanic cloud droplet number densities and increased oceanic rain drop number densities, relative to the no-feedback climatological simulation. The aerosol direct and indirect effects were responsible for changes to the near-surface PM2.5 and NO_2 ~~aerosol~~ concentrations at greater than the 90% confidence level near ~~throughout~~ the forest fires, with ~~model domain, and to NO_2 and~~ O_3 changes remaining below the 90% confidence level ~~concentrations within forest fire plumes~~.

Formatted: Subscript

36   The simulations show that incorporating aerosol direct and indirect effect feedbacks can significantly improve the
37   accuracy of weather and air quality forecasts, and that forest fire plume rise calculations within a on-linefully coupled
38   model changes the predicted fire plume dispersion and emissions, the latter through changing the meteorology driving
39   fire intensitybehaviour and fuel consumptiongrowth.

1 Introduction

41   Atmospheric aerosol particles may be emitted (primary particles) or result from the condensation of the products of
42   gas-phase oxidation reactions (secondary aerosol). With increasing transport time from emission sources, the
43   processes of coagulation (colliding particles stick adhere creating larger particles) and condensation (low volatility
44   gases condense to particle surfaces) tend to result in particles which have a greater degree of internal mixing (internal
45   homogeneous mixtures). Primary and near-source particles are more likely to have a single or a smaller number of
46   chemical constituents (external mixtures).
47   Atmospheric particles also modify weather through well-established pathways. Under clear sky conditions, the
48   particles may absorb and/or scatter incoming light, depending on their size, shape, mixing state (internal, external or
49   combinations) and their composition. The presence of the particles themselves may thus affect the radiative budget
50   of the atmosphere, resulting in either positive or negative climate forcing (i.e. the absorption of a greater amount of
51   incoming solar radiation versus increased scattering reflection of that radiation back out into space, a process known
52   as the Aerosol Direct Effect; ADE). Aerosols can also alter the atmospheric radiative balance through interactions
53   with clouds, this influence being referred to as the Aerosol Indirect Effect (AIE). Three broad classes of categories
54   by which cloud/aerosol interactions take place (Oreopoulos *et al.,* 2020) include the first indirect effect, where higher
55   aerosol loadings resulting in increasing numbers of cloud droplets with smaller sizes, hence increasing cloud albedo
56   (Twomey *et al.,* 1977), the second indirect effect, where higher aerosol loadings suppress the collision-coalescence
57   activity of the smaller droplets, reducing precipitation/drizzle, changing cloud heights, and changing cloud lifetime in
58   warm clouds (Albrecht, 1989), and aerosol "invigoration" of storm clouds, where higher aerosol loadings may result
59   in delayed glaciation of cloud droplets, in turn leading to greater latent heat release and stronger convection (Rosenfeld
60   *et al.,* 2018).
61   The uncertainties associated with the ADE and particularly AIE account for a large portion of the uncertainties in
62   current climate model predictions for radiative forcing between 1750 and 2011 (Mhyre *et al.,* 2013). Carbon dioxide
63   is believed to have a positive (warming) global radiative forcing of approximately 1.88 +/- 0.20 Wm^{-2}, while the direct
64   and indirect effects both have nominal values of approximately -0.45 Wm^{-2}, with uncertainty ranges encompassing -
65   0.94 to +0.07 and -1.22 to 0.0 Wm^{-2} respectively. These uncertainties have spurred research designed to better
66   characterize the ADE and AIE, and reduce these uncertainties, through both observations and atmospheric modelling.
67   Observational studies of the ADE have established its large impact; for example, high aerosol loading over Eurasian
68   boreal forests has been found to double the diffuse fraction of global radiation (i.e. increased scattering), a change
69   sufficient to affect plant growth characterized via gross primary production (Ezhova *et al.,* 2018). Aerosol assimilation
70   of Geostationary Ocean Color Imager Aerosol Optical Depth (AOD) observations into a coupled meteorology-
71   chemistry model showed that South Korean AOD values increased by as much as 0.15 with the use of assimilation;

72 these increases corresponded to a local -31.39 W m^{-2} reduction in solar radiation received at the surface, and reductions

73 in planetary boundary layer height, air temperature, and surface wind speed over land, and a deceleration of vertical

74 transport (Jung *et al.,* 2019). Other studies in East Asia have shown ADE decreasing local shortwave reaching the

75 surface by -20 W m^{-2} (Wang *et al.,* 2016), as well as significant changes in surface particulate matter and gas

76 concentrations in response to these radiation changes.

77 However, one commonality amongst the recent studies of the ADE for air-quality models is a tendency towards

78 negative biases in predicted aerosol optical depths, potentially indicating systematic under-predictions in aerosol mass,

79 aerosol size, and/or inaccuracies in the assumptions for shape and/or mixing state. Mallet *et al. (*2017) noted this

80 negative bias for regional climate model AOD predictions associated with large California forest fires compared to

81 OMI and MRIS satellite observations. Palacios-Pena *et al. (*2018) noted that high AOD events associated with forest

82 fires were under-predicted by most models in a study employing a multi-regional-model ensemble. The chosen AOD

83 calculation methodology and mixing state assumptions employed in models also plays a role in systematic biases:

84 Curci *et al. (*2015) compared aerosol optical depths, single scattering albedos, and asymmetry factors at different

85 locations to observations, varying the source model for the aerosol composition, as well as the mixing state

86 assumptions used in generating aerosol optical properties, for Europe and North America. AODs were biased low by

87 a factor of two or more, regardless of model aerosol inputs or mixing state assumptions at 440 nm, single scattering

88 albedos were biased low by up to a factor of two, with the poorest performance for "core-shell" approaches, while

89 asymmetry factor estimates showed no consistent bias relative to observations. However, the assumed mixing state

90 was clearly a controlling factor in the negative biases; the AOD predictions closest to the observations at 440 nm

91 assumed an external mixture with particle sulphate and nitrate assumed to grow hygroscopically as pure sulphuric

92 acid, lowering their refractive index with increasing aerosol size. This mixing state assumption and the different

93 homogeneous mixture assumptions gave the best fit for single scattering albedo relative to observations. While not

94 commenting on aerosol direct effect implications, Takeishi *et al. (*2020) noted that forest fire aerosols increase particle

95 number concentrations but reduce their water uptake (hygroscopicity) relative to anthropogenic aerosols, with the

96 latter effect reducing the resulting cloud droplet numbers by up to 37%. Mixing state and hygroscopicity properties

97 of aerosols were thus shown to have a controlling influence on the ADE.

98 The AIE has often been shown to be locally more important for the radiative balance than ADE in terms of magnitude

99 of the radiative forcing and response of predicted weather to AIE and ADE (Makar *et al.,* 2015(a); Jiang *et al.,* 2015;

100 Nazarenko *et al.,* 2017). Several recent studies have attempted to characterize the relative importance of the AIE with

101 the use of multi-year satellite observations, sometimes making use of models and data assimilation. Saponaro *et al.*

102 (*2017) used MODIS/Aqua linked observations of aerosol optical depth and Ångström exponent to various cloud

103 properties, noting that the cloud fraction, cloud optical thickness, liquid water path, and cloud top height all increased

104 with increasing aerosol loading, while cloud droplet effective radius decreased, with the effects dominating at low

105 levels (between 900 to 700 hPa). Zhao *et al.* (2018) examined 30 years of cloud and aerosol data (1981-2011), and

106 found that increasing aerosol loading up to AOD < 0.08 increased cloud cover fraction and cloud top height, while

107 further increases in aerosol loading (AOD from 0.08 to 0.13) resulted in higher cloud tops, and larger cloud droplets.

108 In polluted environments (AOD > 0.30) cloud droplet effective radius, optical depth and water path; cloud droplet

109 effective radius increased with increasing AOD. The first ADE was most sensitive to AOD in the AOD range 0.13 to
110 0.30; and the reduction of precipitation efficiency associated with the second aerosol indirect effect occurred for AODs
111 between 0.08 and 0.4, in oceanic areas downwind of continental sources.

112 However, sources of uncertainty in AIE estimates persist, in part due to the number of poorly understood processes
113 contributing to the atmospheric response to the presence of aerosols. Nazerenko *et al.* (2017) showed that short-term
114 atmospheric radiative changes were reduced in magnitude when sea-surface temperature and sea-ice coupling was
115 included in climate change simulations. Suzuki *et al.* (2019) showed that the vertical structure of atmospheric aerosols,
116 as well as their composition, had a significant influence on radiative forcing. Penner *et al.* (2018) and Zhu *et al.* (2020)
117 examined the impact of aerosol composition on cirrus clouds via ice nucleation, finding negative forcings for most
118 forms of soot, but a contrary impact of secondary organic aerosols. Rothenburg *et al.* (2018) noted that tests of aerosol
119 activation schemes carried out under current climate conditions had little variability, but had much greater variability
120 for pre-industrial simulations, implying that the available data for evaluation using current conditions may poorly
121 constrain ADE and AIE parameterizations used in simulating in past climates.

122 Forest fires are of key interest for improving the understanding and representation of ADE and AIE in models, due to
123 the large amount of aerosols released during these biomass burning events. Forest fire emissions and interactions with
124 weather are also of interest due to the expectation that the meteorological conditions resulting in forest fires may
125 become more prevalent in the future under climate change (Hoegh-Guldberg *et al.*, 2018). Observations of aerosol
126 optical properties during long-range transport events of North American forest fire plumes to Europe showed 500 nm
127 AOD values of 0.7 to 1.2 over Norway, with Ångström exponents exceeding 1.4 and absorbing angstrom exponents
128 ranging from 1.0 to 1.25, along with single scattering albedos greater than 0.9 at the surface and up to 0.99 in the
129 column over these sites (Markowicz *et al.*, 2016). Biomass burning was shown to have a specific set of optical
130 properties relatively independent of fuel type for three different types of biomass burning in China (cropland), Siberia
131 (mixed forest) and California (needleleaf forest). The increase in upward radiative forcing at the top of the atmosphere
132 due to fires being linearly correlated to AOD (R from 0.48 to 0.68), with slopes covering a relatively small range from
133 20 to 23 W m^{-2} unit AOD^{-1}. O'Neill *et al.* (2001) showed that forest fires have a profound impact on aerosol optical
134 depth in western Canada, accounting for 80% of the summer AOD variability in that region, with a factor of three
135 increase in AOD levels from clear-sky to forest fire plume conditions. O'Neill *et al.* (2001)'s analysis of TOMS
136 AVHRR and GOES imagery suggested that forest fire aerosols increase in size with increasing downwind distance,
137 due to secondary aerosol aging and condensation chemistry. We note here that reanalyzing the data presented in
138 O'Neill *et al.* (2001) results in a linear relationship between fine mode particle effective radius (r_{eff}, μm) and the base
139 10 logarithm of distance from the fires (D, km) of $r_{eff} = 0.0106 \, log_{10}(D) + 0.1163, R^2 = 0.18$). Mallet *et al.*
140 (2017) simulated AODs in the range 1 to 2 for biomass burning events, and also noted changes in direct radiative
141 forcing at the top of the atmosphere from positive to negative in both model results and simulations, with increasing
142 downwind distance from the sources. Lu *et al.* (2017) carried out simulations with 5-km horizontal grid spacings for
143 the eastern Russia forest fires of 2002 assuming an internal mixture for emitted aerosols with the WRF-CHEM model,
144 and noted impacts on cloud formation for two different periods. The first period was characterized by high cloud
145 droplet and small ice nuclei numbers, where the fire plumes reduced cloud rain and snow water content, large scale

146 frontal system dynamics were altered by smoke, and precipitation was delayed by a day. The second period was

147 characterized by high numbers for cloud droplets and ice nuclei, where the fire plumes reduced rain water content,

148 increased snow water content, and precipitation locations changed locally across the simulation domain. Russian

149 forest fire simulations for 2010 with suites of on-linefully coupled air-quality models (Makar *et al.,* 2015; Palacios-

150 Pena *et al.,* 2018; Baro *et al.,* 2017) showed substantial local impacts, such as reductions in average downward

151 shortwave radiation of up to 80Wm^{-2} and temperature of -0.8 ^{o}C (Makar *et al.,* 2015(a)).

152 Given the above developments in direct and indirect parameterizations, and the increasing amount of information

153 available for estimating forest fire emissions, the impact of forest fires on weather, in the context of weather

154 forecasting, is worthy of consideration. Air-quality model predictions of forest fire plumes have been provided to the

155 public under operational forecast conditions of time- and memory-space limited computer resources (e.g. Chen *et al.,*

156 2019; James *et al.,* 2018; Ahmadov *et al.,* 2019, Pan *et al.,* 2017). These simulations make use of satellite retrievals

157 of forest fire hot-spots, climatological data on the extent of area burned by land use type, databases of fuel type linked

158 to emission factors, and an *a priori* weather forecast to provide the meteorological inputs required to predict forest

159 fire plume rise. The latter point is worthy of note in the context of the direct and indirect feedback studies noted above

160 – both climate and weather simulations with prescribed forest fire emissions have consistently resulted in large

161 perturbations of weather patterns in the vicinity of the forest fires. However, theirthe approaches for predicting forest

162 fire plume rise and fire intensity and fuel consumption in operational regional scale forecasts up until now have relied

163 on weather forecast information provided *a priori* and hence lacking those meteorological feedback effects.

164 The connection of the ADE and AIE within a regionalan air-quality and weather forecast model context is referred to

165 as " coupling", with such a model being described in that body of literature as "on-line "fully coupled" (Galmarini *et*

166 *al.,* 2015) or " aerosol-aware" (Grell and Freitas, 2014).". However, several researchers have examined aerosol-

167 radiative coupling along with *fire spread and growth* (as opposed to fire intensity and fuel consumption). The latter

168 work employs we note that in the field of very high -resolution forest fire spread and growth models, and due to their

169 very high resolution, an additional level of coupling, that ofbehaviour modelling, coupling of biomass burning with

170 the atmosphere has also been defined as the interaction of dynamic meteorology with the heat released by the fire,

171 may be included. However, the resolution requirements for these models (and their need for a relatively small

172 computational time step) constrains their application to a relatively small region. A requirement for these approaches

173 is the use of a very high resolution fire growth model imbedded within the air-quality model. At these resolutions,

174 the simulated local-scalewhere the initial meteorology determines fire spread on the landscape, which. This in turn,

175 modifies the temperature and wind fields, in turn affecting future fire spread. The seminal work on this topic was

176 carried out by (Clark *et al.* (, 1996), and Linn *et al.* (, 2002). More recent work includes the development of the

177 WRF-FIRE model (Mandel *et al.,* 2011; Coen *et al.,* 2013), with full chemistry added in the WRFSC model

178 (Kochanski *et al.,* 2016). Examples of the resolution required for these models include inner domain resolutions of

179 444 m with an imbedded fire model mesh of 22.2 m resolution, and a time step of 3.3 seconds (Kochanski *et al.,*

180 2016); 1.33 km with an imbedded fire model mesh of 67.7m, and a time step of 2 seconds (Kochanski *et al.,* 2019),

181 and 222m, with a fire model mesh of 22m and a time step of 2 seconds (Peace *et al.,* 2015). Kochanski *et al* (2016)

182 also noted a 13 to 30 hour computational time requirement to run their high-resolution modelling system. These

183 modelling efforts allow for this additional level of coupling – but at the expense of additional computation time
184 preventing, at the current state of supercomputer processing, their application on synoptic-scale forecast domains
185 combined with a full gas chemistry and size-resolved multi-component particle chemistry representation. Here we
186 explore the effects of fire emissions characterized by fire intensity and fuel consumption modelling on the aerosol
187 direct and indirect effects over synoptic scale domain. Our coupling~~The coupling presented in the present paper~~ refers
188 to that between the aerosols released by fires and other sources to meteorology through the ADE and AIE, with the
189 resulting changes in meteorology in turn influencing fire ~~behaviour (fire~~ intensity and~~fuel, fuel~~ consumption~~., etc.)~~,
190 in turn influencing plume rise, emissions height, and distribution, closing this feedback loop. We do not implement a
191 very high resolution growth model, noting that this is impractical for operational forecasts at the current time, while
192 showing that synoptic scale 2.5km simulations incorporating fire feedbacks may be carried out within an operational
193 window with currently available supercomputers. As shown below, we find that a sufficiently substantial feedback
194 between the aerosol direct and indirect effects can be discerned to change the vertical distribution of emitted pollutants.
195 A key consideration in parameterizing the AIE (via aerosol-cloud interaction) is the manner in which the cloud
196 condensation process is represented in the meteorological component of the modelling system. In numerical weather
197 prediction (NWP) models, clouds and precipitation are represented by a combination of physical parameterizations
198 that are each targeted at a specific subset of moist processes. These include "implicit" (subgrid-scale) clouds generated
199 by the boundary layer and the convection parameterization schemes (e.g Sundqvist, 1988), and "explicit" clouds from
200 the grid-scale condensation scheme (Milbrandt and Yau, 2005(a,b), Morrison and Milbrandt, 2015, Milbrandt and
201 Morrison, 2016). Depending on the model grid these "moist physics" schemes vary in their relative importance.
202 However, regardless of the horizontal grid cell size, the grid-scale condensation scheme plays a crucial role in
203 atmospheric models, though to different degrees and using different methods, depending on the grid spacing and the
204 corresponding relative contributions of the implicit schemes. A grid-scale condensation scheme will in general consist
205 of the following three components: 1) a subgrid cloud fraction parameterization (CF, or cloud "macrophysics"
206 scheme); 2) a microphysics scheme; and 3) a precipitation scheme (Jouan *et al.*, 2020). The cloud fraction (CF) is the
207 percentage of the grid element that is covered by cloud (and is saturated), even though the grid-scale relative humidity
208 may be less than 100%. The microphysics parameterization computes the bulk effects of a complex set of cloud
209 microphysical processes. If precipitating hydrometeors are advected by the model dynamics, the precipitation is said
210 to be *prognostic*; if precipitation is assumed to fall instantly to the surface upon production, it is considered *diagnostic*.
211 The precipitation "scheme" is not a separate component per se, since it simply reflects the level of detail in the
212 microphysics parameterization, but it is a useful concept to facilitate the comparison of different grid-scale
213 condensation parameterizations.
214 With a wide range of grid cell sizes in current NWP models, there is a wide variety of types of condensation schemes
215 and degrees of complexity in their various components. For example, cloud-resolving models (with grid spacing on
216 the order of 1 km or less) have typically used detailed bulk microphysics schemes (BMSs), with prognostic
217 precipitation, and no diagnostic or prognostic CF component (i.e. the CF is either 0 or 1). Large-scale global models
218 use condensation parameterizations, sometimes referred to as "stratiform" cloud schemes, typically with much simpler
219 microphysics and diagnostic precipitation, but with more emphasis on the details of the CF. However, with continually

220 increasing computer resources and decreasing grid spacing (both in research and operational prediction systems), the
221 distinction between schemes designed for specific ranges of model resolutions is disappearing and condensation
222 schemes are being designed or modified to be more versatile and usable across a wider range of model resolutions
223 (e.g. Milbrandt and Morrison, 2016).

224 Aerosol-cloud interactions and feedback mechanisms are difficult to represent in grid-scale condensation schemes
225 with very simple microphysics components. For example, to benefit from the predicted number concentrations of
226 cloud condensation nuclei and ice nuclei, the microphysics needs to be double-moment (predicting both mass and
227 number) for at least cloud droplets and ice crystals, respectively. Until recently, detailed BMSs were only used at
228 cloud resolving scales, hence requiring these relatively high resolutions to be recommended in feedback modelling.
229 In recent years, multi-moment BMSs have been used in operational NWP for model grid spacings of 2-4 km (e.g.
230 Seity *et al.*, 2010, Pinto *et al.*, 2015, Milbrandt *et al.*, 2016). Further, condensation schemes with detailed microphysics
231 are starting to use non-binary CF components (e.g. Chosson *et al.*, 2014, Jouan *et al.*, 2020), thereby allowing detailed
232 microphysics to be used at larger scales, and hence allowing the same indirect feedback parameterizations to be used
233 at multiple scales. Nevertheless, the expectation is that detailed parameterization will provide a more accurate
234 representation of cloud formation at the near cloud-resolving scales, without the complicating aspect of a diagnostic
235 CF, motivating the use of km-scale grid spacing for feedback studies.

236 The formation of secondary aerosols from complex chemical reactions are another key consideration in feedback
237 forecast implementation, given the impact of aerosol composition on aerosol optical and cloud formation properties,
238 as described above.

239 In the sections which follow, we describe our high resolution, on-line~~fully~~ coupled air-quality model with on-line
240 forest fire plume rise calculations, which was created as part of the FIREX-AQ air-quality forecast ensemble
241 (https://www.esrl.noaa.gov/csl/projects/firex-aq/), to address the following questions:

242 (1) Will a on-line~~fully~~ coupled model of this nature provide improved forecasts of *both* weather and air-quality, using
243   standard operational forecast evaluation tools, techniques and metrics of forecast confidence? That is, despite the
244   uncertainties in the literature as described above, are these processes sufficiently well described in our model that
245   their use results in a formal improvement in forecast accuracy?

246 (2) Are the changes in forest fire plume rise associated with implementing this process directly within a on-line~~fully~~
247   coupled model sufficient to result in significant perturbations to weather predictions and to chemistry? What are
248   these perturbations?

249 We employ our on-line~~fully~~ coupled model with 2.5-km grid cell size domain covering most of western North
250 America, and compare model results to surface meteorological and chemical observations, and to vertical column
251 observations of temperature and aerosol optical depth (AOD), in order to quantitatively evaluate the effect of feedback
252 coupling of the ADE and AIE on model performance. We then compare feedback and no-feedback simulations to
253 show the impacts of the ADE and AIE feedbacks on cloud and other meteorological predictions, and on key air quality
254 variables (particulate matter, nitrogen dioxide, and ozone). We begin our analysis with a description of our modelling
255 platform.

256 ## 2 Model Description

257 ### 2.1 GEM-MACH

258 The Global Environmental Multiscale – Modelling Air-quality and CHemistry (GEM-MACH) model in its on-
259 line~~fully~~ coupled configuration has been described elsewhere (Makar *et al.,* 2015a,b; Gong *et al.,* 2015, 2016).
260 The~~Briefly, the~~ model combines the Environment and Climate Change Canada Global Environmental Multiscale
261 weather numerical weather prediction model (GEM, Cote *et al.,,* 1998, Girard *et al.,* 2014) with gas and particle
262 process representation using the on-line paradigm, with options for climatological versus full coupling between
263 meteorology and chemistry. GEM-MACH's main processes~~In GEM-MACH's climatological coupling configuration,~~
264 ~~prescribed, invariant climatological values~~ for ~~the two configurations aerosol optical properties and cloud~~
265 ~~condensation nuclei (CCN) are~~ employed here are described~~within the model's radiative transfer and cloud~~
266 ~~microphysics modules. In the full coupling configuration, the ADE is simulated using GEM-MACH's predicted~~
267 ~~aerosol loading and Mie scattering using a binary water dry aerosol homogeneous mixture assumption, at the 4~~
268 ~~wavelengths employed by GEM's radiative transfer algorithms, and at additional wavelengths for diagnostic purposes.~~
269 ~~The full coupling also includes the AIE by simulating aerosol-cloud interaction via explicit droplet nucleation using~~
270 ~~the algorithm of Abdul-Razzak and Ghan (2002) along with on-line aerosol composition and size (Gong *et al.,* 2015).~~
271 ~~This droplet nucleation replaces the decoupled model's existing droplet nucleation calculation~~ in ~~the Predicted Particle~~
272 ~~Properties (P3) microphysics scheme (Morrison and Milbrandt, 2015, Milbrandt and Morrison, 2016). The latter~~
273 ~~assumes an invariant aerosol population of a single lognormal size distribution (with a geometric mean diameter of~~
274 ~~100 nm and total aerosol number concentration of 300 cm^{-3}, consisting of pure ammonium sulphate; Morrison and~~
275 ~~Grabowski, 2008). The prognostic cloud droplet number and mass mixing ratios from the P3 microphysics are then~~
276 ~~transferred back to the chemistry module for using in cloud processing of gases and aerosols (cloud scavenging and~~
277 ~~chemistry) calculations, completing the AIE feedback process loop (Gong *et al.,* 2015).~~ Table 1.
278 ~~The chemistry modules of GEM-MACH also include process representation for gas-phase chemistry (ADOMII~~
279 ~~mechanism, 42 gas species, Stockwell *et al.,* 1989), cloud processing including aqueous chemistry, scavenging of~~
280 ~~gases and aerosols, below-cloud removal and wet deposition (Gong *et al.,* 2015), particle microphysics employing a~~
281 ~~sectional size distribution and 8 chemical species (Gong *et al.,* 2003), particle inorganic thermodynamics (Makar *et*~~
282 ~~*al.,* 2003), the formation of secondary organic aerosols using a modified yield approach (Stroud *et al.,* 2018), process~~
283 ~~representation for surface fluxes as a boundary condition on the solution for vertical diffusion, and semi-Lagrangian~~
284 ~~advection for transported chemical tracers.~~
285 ~~The specific base model version employed in these simulations is GEMv4.9.8/GEM-MACHv2 (Moran *et al.,* 2018),~~
286 ~~incorporating the following additional improvements in addition to those noted above: (a) the AIE parameterization~~
287 ~~was modified for use with the P3 cloud microphysics scheme; (b) forest-canopy shading and turbulence was~~
288 ~~parameterized following Makar *et al.,* (2017); (c) anthropogenic plume rise was parameterized through calculating~~
289 ~~residual buoyancy of the rising plume (Akingunola *et al.,* 2018); (d) emissions of crustal material undergo a~~
290 ~~meteorological modulation with crustal material emissions being inhibited when the soil water content is predicted to~~
291 ~~be greater than 10%; (e) emissions/deposition of NH$_3$ are implemented using a bidirectional flux parameterization~~
292 ~~(Whaley *et al.,* 2018; Zhang *et al.,* 2003); (f) CH$_4$ is treated as a reactive and emitted tracer; (f) the KPP-generated~~

8

293 ~~RODAS-3 solver (Sandu and Sander, 2006) is used for the solution of ADOMII gas-phase mechanism (Stockwell~~
294 ~~and Lurmann, 1989); (g) MODIS retrievals were used to create monthly leaf area index values for use in the model's~~
295 ~~biogenic emissions, forest canopy shading and turbulence, and deposition algorithms; (h) a parameterization for the~~
296 ~~impacts of vehicle-induced turbulence on vertical diffusive transport was employed (Makar *et al.,* 2020).~~
297 Simulations were carried out with a 2.5-km horizontal grid cell spacing over a 900 x 1370 grid cell domain, covering
298 most of western Canada and the USA (Figure 1). The meteorological boundary conditions for the simulation were a
299 combination of 10-km resolution GEM forecasts updated hourly (themselves originating in data assimilation analyses
300 of real-time weather information; Figure 1(a)),~~,~~ and 2.5-km GEM simulations (Figure 1(c)) employing, in the northern
301 portion of this~~the~~ 2.5-km domain, the Canadian Land Data Assimilation System (Carrera *et al.,* 2015), to better
302 simulate surface conditions. Both "feedback" and "no feedback" simulations were carried out on a 30-hour forecast
303 cycle (Figure 2). Following the usual practice for weather forecasts, the analysis-driven meteorological forecasts at 10
304 km resolution were updated operationally every 24 hours at 12 UT (Figure 2(a)).~~.~~ These 10 km resolution weather
305 forecasts were used to drive a 30-hour, 10-km resolution GEM-MACH forecast (Figure 1(b), Figure 2(b)), which
306 employed ECMWF reanalysis data for North American chemical lateral conditions (Innes *et al.,* 2019). The 10-km
307 resolution weather forecasts were also used to drive~~followed by~~ a 30-hour meteorology-only forecast at 2.5-km
308 resolution on the high resolution domain (~~of~~ Figure 1(c), Figure 2(c)). The~~, the~~ last 24 hours of the 10-km resolution
309 GEM-MACH forecast was also used to provide chemical lateral boundary conditions for the 24-hour 2.5km on-line
310 coupled GEM-MACH simulation (Figure 1(c), Figure 2(d)). The last 24 hours of the 2.5-km GEM simulation~~which~~
311 were used as meteorological initial and boundary conditions for the 24-hour 2.5-km on-line ~~fully~~ coupled GEM-
312 MACH simulation (Figure 1(c), Figure 2(d)). The~~. These~~ two stages of meteorology-only simulations were carried
313 out to prevent chaotic drift from the observed meteorology, and to allow spin-up time for the cloud fields of that
314 meteorology to reach equilibrium (6-hour timeframe). ~~Chemical lateral boundary conditions were taken from~~
315 ~~climatologies based on ECMWF MACC-II global atmospheric chemical composition modelling and reanalysis~~
316 ~~(Inness *et al.,* 2013).~~ Chemical initial concentrations for each consecutive forecast within the 2.5-km GEM-MACH
317 model domain were "rolled over" or "daisy-chained" between subsequent forecasts without chemical data
318 assimilation. Forecast performance scores presented here are for the inner 2.5-km domain from this set of linked 24
319 forecast simulations, mimicking operational forecast conditions.

2.2 CFFEPS Version 4.0: On-line forest-fire plume rise calculations

321 In addition to the above algorithm improvements relative to GEM-MACH implementations, this model system setup
322 has incorporated the first on-line calculation of forest-fire plume-rise by energy balance driven using on-line
323 meteorology, in a new version of the Canadian Forest Fire Emissions Prediction System (CFFEPS). The algorithms
324 of CFFEPSv2.03 are described in detail and evaluated elsewhere (Chen *et al.,* 2019), but will be outlined briefly here,
325 as well as subsequent modifications to this forest fire emissions processing module.
326 CFFEPS combines near-real-time satellite detection of forest fire hotspots with national statistics of burn areas by
327 Canadian province and by specific fuel type across North America. CFFEPS assumes persistence fire growth in the
328 subsequent 24- to 72-hour forecasts with hourly fuel consumed calculated (kg m^{-2}), based on GEM forecast

329   meteorology and predicted fire intensity and fuel consumptionbehaviour in grid cells representing fire locations. The

330   modelled fire fuel consumption is then linked with combustion-phase specific emission factors (g kg^{-1}) for fire specific

331   emissions and chemical speciation. Fire energy associated with the modelled combustion process is also estimated,

332   and is used in conjunction with *a priori* forecasts of meteorology within the column to determine plume rise. In its

333   off-line/non-coupled configuration (Chen *et al*., 2019), CFFEPS carries out residual buoyancy calculations at five

334   preset pressure levels (surface, 850, 700, 500, 250 mb). CFFEPS predicts plume injection heights, which are in turn

335   used to redistribute the mass emissions below the plume top to the model hybrid levels. This approach employed in

336   CFFEPSv2.03 provided a substantial improvement in forecast accuracy relative to the previous approach employing

337   modified Briggs (Briggs, 1965, Pavlovic *et al.,* 2016) plume rise formulae in the offline GEM-MACH forecast system

338   (Chen *et al.,* 2019). A recent evaluation of the plume heights predicted by CFFEPS was carried out utilizing MISR

339   and TROPOMI satellite retrieval data (Griffin et al, 2020). Seventy cases studied using MISR data showed good

340   agreement between satellite and CFFEPS-predicted maximum and mean plume heights (maximum plume height

341   observed versus predicted values and standard deviations: 1.7+0.9 versus 2.0+1.0 km; mean plume height observed

342   versus predicted: 1.3+0.6 versus 1.3+0.4 km). A larger number of case studied using TROPOMI data (671 in total)

343   also showed a reasonable agreement, with CFFEPS showing a small tendency to overpredict heights (maximum

344   observed versus predicted plume heights 2.2+1.6 versus 2.5+1.2 km; mean observed versus predicted plume heights

345   0.7+0.5 versus 1.1+0.6 km).

346   However, other work has shown the substantial impact of large forest fires on regional weather (Makar *et al.,* 2015a;

347   Palacios-Pena *et al.,* 2018, Baro *et al.,* 2017), including changes to the surface radiative balance and atmospheric

348   stability. These findings imply that plume rise calculations employing an *a priori* weather forecast lacking the impact

349   of fire plumes via the ADE and AIE may not accurately predict the weather conditions critical to subsequent forest

350   fire plume rise prediction. In order to study this possibility, and to allow forest fire plumes to influence weather and

351   hence subsequent fire spread/growth, several changes were made to CFFEPS implementation, resulting in version 4.0

352   of CFFEPS, used here. The process flow within CFFEPSv2.03 versus CFFEPSv4.0 are compared in Figure 3. The

353   original C language CFFEPSv2.03 code was converted to FORTRAN90, and following successful off-line

354   comparisons to the original code, was then integrated as an on-line subroutine package within GEM-MACH itself,

355   with the near-real-time satellite hotspot data and location fuel parameters being read into GEM-MACH directly

356   (CFFEPSv4.0 is this new on-line package). A key advantage of the CFFEPSv4.0 subroutine integration within GEM-

357   MACH is that the residual buoyancy calculations for plume injection heights are now carried out over the model

358   hybrid model layers, rather than the five coarse resolution, prescribed pressure levels of CFFEPSv2.03, making

359   complete use of GEM-MACH's detailed vertical structure. Additionally, CFFEPSv4.0 allows plume rise calculations

360   to be updated during model runtime. When GEM-MACH is run in on-line fully coupled mode, the ADE and AIE

361   implementations allow model-generated aerosols to modify the predicted meteorology, in turn influencing predicted

362   fire emissions and plume rise, closing these feedback loops. The on-line implementation of CFFEPSv4.0 thus allows

363   us to investigate the effects of meteorology on subsequent forest fire plume development, the changes to modelled

364   aerosol compositions, and, ultimately, the feedbacks to weather.

365 The formation of particles from forest fires affects meteorology on the larger scale via the ADE and AIE, in turn

366 modifying the regional scale atmospheric features affecting fire growth, such as the temperature profiles below forest

367 fire plumes. However, we note that CFFEPSv4.0 employs forest fire heat to determine plume rise as a subgridscale

368 thermodynamic process parameterization rather than a very high resolution explicit fire growth parameterization; the

369 very~~However, we note that the~~ local scale weather modifications due to the addition of forest fire heat to the

370 atmosphere are not ~~yet~~ incorporated into fire spread or GEM microphysics. Specifically, when the feedback version

371 of GEM-MACH incorporating CFFEPSv4.0 is used in its on-line~~fully~~ coupled configuration, CFFEPSv4.0 uses

372 estimates of~~calculates forest fire plume rise using~~ the heat released to calculate forest fire plume rise. These

373 calculations employ lapse rates at the fire locations, that with feedbacks enabled, ~~meteorological predictions which~~

374 include the ADE and AIE generated by forest fire aerosols on atmospheric stability within~~from~~ the current on-line

375 fully~~coupled model timestep. This is ~~The resulting added aerosol mass due to the fire~~ in contrast to earlier off-line

376 implementations of CFFEPS, which made use of *a priori* non-~~turn affects the meteorology through ADE and AIE,~~

377 ~~closing this~~ feedback weather forecast lapse rates.~~loop.~~ To the best of our knowledge, this is the first implementation

378 of a dynamic forest fire plume injection height scheme incorporated into a on-line~~fully~~ coupled high-resolution,

379 operational air quality forecast modelling system. The impact of this feedback on both weather and air-quality can be

380 substantial, as we show in the following sections.

381 The locations of the daily forest hotspots detected during the study period, and the corresponding magnitude of the

382 daily PM2.5 emissions generated by CFFEPS for each hotspot are shown in Figure 4. Individual hotspots with the

383 highest magnitude emissions are located in the state of Nevada (Figure 4(a), southern boxed region). However, the

384 largest ensemble emissions from a suite of hotspots occurs in northern Alberta (Figure 4(a), northern boxed region).

385 Expanded views of the northern Alberta and Nevada hotspots are shown in Figure 4(b,c) respectively – the use of

386 smaller symbols shows that the Alberta hotspots are groups representing large spreading fires, which overplotted in

387 Figure 4(a), while the Nevada hotspots indicate single fires of small spatial extent and duration rather than larger

388 spreading fires. The Alberta fires are thus the most significant sources of forest fire emissions in the study domain for

389 the period analyzed here.

390

2.2 Feedback and No-Feedback Simulations

392 Two simulations were carried out for the period July 4th through August 5th 2019; a "feedback" (ADE and AIE

393 feedbacks enabled – on-line~~fully~~ coupled model) and a "no-feedback" simulation (ADE and AIE make use of GEM's

394 climatological aerosol radiative and CCN properties – the one-way coupled~~decoupled~~ model). During this period,

395 five large forest fires took place in the northern portion of the modelling domain. The two parallel combined

396 meteorology and air-quality forecasts in the on-line~~fully~~ coupled model with/without ADE and AIE coupling were

397 evaluated for meteorological~~using the US EPA AIRNOW data (https://www.airnow.gov)~~ and air quality

398 variables.~~Environment and Climate Change Canada's EMET and ARCAD operational forecast evaluation systems,~~

399 ~~respectively.~~ Following evaluation, the simulation mean values of hourly meteorological and chemical tracer

400 predictions were compared to analyze the impact of on-linefully coupled ADE and AIE feedbacks on both sets of
401 fields.

3 Model Evaluation

3.1 Meteorology Evaluation

404 Surface meteorological conditions were evaluated at three-hour intervals from the start of both of the two sets of paired
405 24-hour forecasts using standard metrics of weather forecast performance including mean bias (MB), mean absolute
406 error (MAE), root mean square error (RMSE), correlation coefficient (R) and standard deviation (σ).   In all
407 comparisons, a 90% percent confidence level assuming a normal distribution was used to identify statistically different
408 results between forecast simulations. Note that 90% confidence levels are commonly used in meteorological forecast
409 evaluation, with values of 80% to 85% recommended (Pinson and Kariniotakis, 2004) and up to 90% used (Luig *et*
410 *al.,* 2001) for variables such as wind speed, rather than the 95% or 99% confidence levels in other fields, in recognition
411 of the difficulties inherent in prognostic forecasts of the chaotic weather system.   Here, the confidence range
412 formulation of Geer (2014) has been applied using a 90% confidence level in model predictions, with the statistical
413 measures considered different at the 90% confidence level when the 90% confidence ranges do not overlap.   The
414 surface meteorological evaluations shown here only include those variables and metrics where results were
415 significantly different at the 90% confidence level.
416 Several model forecast output variables were evaluated and the surface variables showing statistically significant
417 differences relative to observations at the 90% confidence level included: 2 m temperature, surface pressure, 2 m
418 dewpoint temperature, 10 m wind speed, sea-level pressure, and accumulated precipitation (the latter in 3 different
419 metrics). The comparisons are shown as time series in three-hourly intervals as a function of forecast hour prediction
420 time forward from forecast hour 0, for grid cells corresponding to measurement locations in Figures 5, 6, 7, 8, 9, 10,
421 and 11 for each of these quantities, respectively. Note that these statistics measure domain-wide performance, across
422 all of the reporting stations within the model domain, during the sequence of 24-hour forecasts comprising the
423 simulation period. The duration of the time series in these comparison figures is thus a function of the duration of the
424 contributing forecasts.
425 Figure 5 shows an example analysis for surface temperature bias for the entire model domain. Figure 5(a) shows the
426 average model mean bias (MB) time series across all stations and all forecasts at the given forecast hours, while Figure
427 5(b) shows the corresponding difference in the MB absolute values. The difference plot in Figure 5(b) shows the
428 feedback – no-feedback scores, such that scores below the zero line indicate superior performance of the feedback
429 forecast, while those above the zero line indicate superior performance of the no-feedback forecast. Here, the feedback
430 forecast was statistically superior at forecast hours 3, 6, 15, 18 and 2418 at the 90% confidence level at these forecast
431 hours, and both simulations were at par (differences below the 90% confidence level) at hours 9, 12 and, 21, with the
432 no-feedback forecast being superior at 90% confidence at hour 9. and 24. The feedback forecast thus has superior
433 performance, at greater than 90% confidence, over half of the forecast hours evaluated within the domain, equivalent

434 performance at two hours (hours 12 and 21, both within 90% confidence limits), and inferior performance at one hour
435 (hour 9), during the simulation period.
436 All of the metrics for which surface temperature forecast performance differed at the 90% confidence level are shown
437 in Figure 6. In addition to MB, the scores for MAE, and RMSE showed superior forecast performance for the feedback
438 relative to the no-feedback case at the 90% confidence level for hours 6 through 15 and 18, while the improvement
439 for the correlation coefficient was only reached the 90% confidence level at hour 18hours 6 and 12.
440 The meteorological forecast performance metrics with statistically significant differences for surface pressure,
441 dewpoint temperature, and sea-level pressure are shown in Figures 7, 8, and 9 respectively. The model performance
442 differences in these three Figures show a similar pattern: a degradation in performance with the use of feedbacks at
443 hour 3, with the differences between the two forecasts either dropping below the 90% confidence level, or the feedback
444 forecast showing an improvement by hour 9, followed by several hours in which the feedback forecast has a superior
445 performance, usually at greater than 90% confidence. The duration of this latter period varies between the metrics,
446 from up to 18 hours for MAE for surface pressure (Figure 7(b)) to 3 hours for the correlation coefficient of dew-point
447 temperature (Figure 8(d)).
448 The initial loss of performance for the feedback forecast may represent a form of "model spin-up" that may be unique
449 to on-line coupled models, but may be affected or improved with further adjustments to the forecast cycling setup for
450 the chemical species. As noted earlier (Figure 2), in order to prevent chaotic drift from observed meteorology, we
451 made use of a 30-hour 2.5-km resolution analysis-driven weather forecast to update our on-line coupled model's initial
452 meteorology at hour zero of each 24 hour forecast. The cloud fields provided as initial conditions at hour zero include
453 observation analysis for the 6 hours prior to hour zero - these have reached a quasi-equilibrium in the high-resolution
454 weather forecast (Figures 2(b,e)) by the time they are used as initial and boundary conditions in the on-line coupled
455 model (Figure 2(c,f)). However, the on-line coupled model's *aerosol* fields at hour zero, used to initialize the
456 subsequent forecast (Figure 2, dashed blue arrow), still reflect the locations of aerosol-cloud interactions in the
457 previous on-line coupled simulation. The initial three to six hours of feedback forecast degradation represents the
458 time required for the on-line coupled model to reach a new equilibrium consistent between both its aerosol and the
459 cloud fields.
460 One possible solution for this model spin-up inconsistency would be to eliminate the intermediate driving 2.5-km
461 meteorological simulation in favour of a longer 30-hour on-line coupled forecast with the first six hours removed as
462 spin-up (i.e. extend the duration of steps (c) and (f) in Figure 2 to 30 hours, starting at UT hour 6). The duration of
463 the forecast experiments carried out here was limited to 24 hours due to limited computational resources, and, more
464 importantly, the operational requirement for an on-time forecast delivery for the purpose of the FIREX-AQ field
465 campaign. The 24-hour forecast simulations carried out in Figure 2 (c,f) each required nearly 3 hours of
466 supercomputer processing time; longer simulation periods were not possible within the operational window available
467 for forecasting.
468 Model 10-m windspeed forecasts were also improved with the incorporation of feedbacks for hours 3 and 6, for all
469 metrics (Figure 10).hour ranges between hour 3 and hour 12, depending on the metric, with the longest duration
470 improvement for MB, MAE, and RMSE, and the shortest duration for correlation coefficient and standard deviation

471 ~~(Figure 9).~~ A decrease in MB~~marginal~~ performance ~~degradation of the feedback forecast at close to 90% confidence~~

472 at hours 21 and 24can~~15-18 can~~ also be seen ~~for root mean square error, correlation coefficient, and standard deviation~~

473 in this Figure.

474 Precipitation forecast performance from the two simulations varied depending on the metric chosen (Figure 11). The

475 metrics in this case were~~are~~ based on the number of coincident precipitation "events" versus "non-events" as shown

476 in contingency Table 2~~1~~.

477 The Heidke skill score $\{$ $HSS =$

478 $2\,(AD - BC)/[(A + C)(C + D) + (A + B)(B + D)]$ ~~$(AD - BC)/[(A+C)(C+D) + (A+B)(B+D)$~~ $\}$ measures

479 the fractional improvement of the forecast over the number correct by chance. The Frequency Bias $\{$ $FB =$

480 $(A + B)/(A + C)$ $\}$ measures the frequency of event over-forecasts (FB>1) versus event under-forecasts (FB<1). The

481 Equitable Threat score $\{$ $ETS = (A - \tilde{A})/(A + C + B - \tilde{A})$, where $\tilde{A} = (A + B)(A + C)/(A + B + C + D)$ $\}$

482 measures the observed and/or forecast events that were correctly predicted. Following standard practice at

483 Environment and Climate Change Canada, the HSS is used as a measure of total precipitation accumulated over a 6-

484 hour interval, with no lower limit on the amount of precipitation defining an "event", while FB and ETS define

485 precipitation "events" as being those with greater than 2mm / 6 hours – consequently FB and ETS have a smaller

486 number of data points for comparison than HSS.

487 Figure 11 shows improvements to the on-line~~fully~~ coupled precipitation forecast at the 90% confidence level were

488 seen for the HSS 6-hour accumulated metric at~~by~~ hours 12~~18~~ and 24, while the frequency bias index of 6-hour

489 accumulated precipitation showed degradation at hours 6 and improved performance at hour 12~~24~~, and the equitable

490 threat score of 6-hour accumulated precipitation showed significant differences at 90% confidence between the two

491 simulations.~~degradation at hour 24.~~ As is noted above, the latter two metrics employed a minimum 6-hour

492 precipitation threshold of 2 mm prior to comparisons (this is the reason for the reduced number of points available for

493 comparison in Figure 11(b,c) relative to Figure 11(a)). These findings suggest that the on-line~~fully~~ coupled model's

494 improvements for ~~trend towards improved~~ total precipitation ~~over time~~ (Figure 11(a)) are~~is~~ the result of slightly

495 improved performance for relatively light~~low-level~~ precipitation events (< 2mm 6hr^{-1}). ~~.), offsetting a degradation of~~

496 ~~performance for higher level precipitation events. Precipitation events have thus become more frequent, but "lighter"~~

497 ~~with the use of the feedback parameterizations.~~

498 ~~The meteorological forecast performance metrics with statistically significant differences for surface pressure,~~

499 ~~dewpoint temperature, and sea level pressure are shown in Figures 7, 8, and 10 respectively. The model performance~~

500 ~~differences in these three Figures show a similar pattern: a degradation in performance with the use of feedbacks at~~

501 ~~hour 3, with the differences between the two forecasts either dropping below the 90% confidence level, or the feedback~~

502 ~~forecast showing an improvement by hour 6, followed by several hours in which the feedback forecast has a superior~~

503 ~~performance. The duration of this latter period varies between the metrics, from up to 18 hours for MAE for surface~~

504 ~~pressure (Figure 7(b)) to 6 hours for the standard deviation of dew-point temperature (Figure 8(d)).~~

505 ~~We believe that the initial loss of performance for the feedback forecast may represent a form of "model spin-up" that~~

506 ~~may be unique to fully coupled models, but may be affected or improved with further adjustments to the forecast~~

507 ~~cycling setup for the chemical species. As noted earlier (Figure 2), in order to prevent chaotic drift from observed~~

14

508 meteorology, we made use of a 30-hour 2.5-km resolution analysis-driven weather forecast to update our fully coupled
509 model's initial meteorology at hour zero of each 24 hour forecast. The cloud fields provided as initial conditions at
510 hour zero include observation analysis for the 6 hours prior to hour zero – these have reached a quasi-equilibrium in
511 the high-resolution weather forecast (Figures 2(b,e)) by the time they are used as initial and boundary conditions in
512 the fully coupled model (Figure 2(c,f)). However, the fully coupled model's *aerosol* fields at hour zero, used to
513 initialize the subsequent forecast (Figure 2, dashed blue arrow), still reflect the locations of aerosol-cloud interactions
514 in the previous fully coupled simulation. We believe that the initial three to six hours of feedback forecast degradation
515 represents the time required for the fully coupled model to reach a new equilibrium consistent between both its aerosol
516 and the cloud fields.
517 One possible solution for this model spin-up inconsistency would be to eliminate the intermediate driving 2.5 km
518 meteorological simulation in favour of a longer 30-hour fully coupled forecast with the first six hours removed as
519 spin-up (i.e. extend the duration of steps (e) and (f) in Figure 2 to 30 hours, starting at UT hour 6). The duration of
520 the forecast experiments carried out here was limited to 24 hours due to limited computational resources, and, more
521 importantly, the operational requirement for an on-time forecast delivery for the purpose of the FIREX-AQ field
522 campaign. The 24-hour forecast simulations carried out in Figure 2 (e,f) each required nearly 3 hours of
523 supercomputer processing time; longer simulation periods were not possible within the operational window available
524 for forecasting.
525 The amalgamated observations and model pairs of vertical temperature profile data from 39 radiosonde sites in western
526 North America are shown in Figures 12 and 13. Improvements in the forecasted temperature vertical profile with
527 increasing forecast time are evident at 250, 300, 400, 500, and 850 hPa in the 12th hour forecast, with degradations at
528 200 and 700 hPa (Figure 12). Improvements) and at 300, 925 and 1000 hPa may be seen in the 24th hour (Figure 13)
529 forecast; it is. The forecast simulation with aerosol feedbacks enabled also worth noting the entire region at and below
530 300showed slight improvements in the 10, and 50 hPa 12th hour temperatures and 50 hPa has improved temperature
531 forecasts (mean values to the left of the vertical line), albeit not always at >90% confidence. 24th hour temperatures,
532 while 500 hPa 24th hour temperature performance degraded slightly. There are larger differences between the 1000
533 hPa forecasts, though these also have the least number of contributing stations (i.e. only those located close to sea-
534 level contribute to the lowest level temperature biases). Other levels of the atmosphere showed no statistically
535 significant change at the 90% confidence level in temperature profile forecast performance with the use of feedbacks.

536 **3.2 Chemistry Evaluation**

537 Improvements to airChemistry forecast quality model performanceis usually evaluated using standard statistical
538 metrics have been a focus for research since the 1980's starting with dispersion model evaluation (Fox, 1981), and the
539 identification of mean bias and normalized mean square error as potentially useful metrics to complement the Pearson
540 correlation coefficient (Hanna, 1988). More recently, the Pearson correlation coefficient has been noted as being
541 capably of producing high values for relatively poor model results (Krause *et al*., 2005), as well as being unable to
542 distinguish systematic model underestimation (Yu *et al*., 2006), unable to provide information on whether data series
543 have a similar magnitude and capable of providing a false sense of relationship where none exists due to outliers

544 (Duveiller *et al.*, 2016) and clusters of model-observation pairs (Aggarwal and Ranganathan, 2016). More recently,
545 model evaluation has focused on metrics which do not have the tendency to weight the higher magnitude values unduly
546 (a particularly useful property with air-quality variables which may vary by several orders of magnitude), which are
547 dimensionless (allowing a comparison across different evaluated variables), and which are bounded and symmetric
548 (properties allowing comparisons to be made and equally valued across the entire range of possible concentrations;
549 e.g. Yu *et al.* (2006)). Metrics such as the modified coefficient of efficiency (Legates and McCabe, 1999) and the
550 more recent incarnations of the Index Of Agreement (Willmott *et al.*, 2012) are examples of the more recent metrics
551 used for air-quality model evaluation. Here, we have made use of a range of metrics spanning the literature on this
552 topic, with the understanding that the properties of different metrics vary, that no single metric provides a perfect
553 means of evaluating model performance, and that a variety of metrics should be applied. The metrics used here span
554 the variety that have appeared in the literature since the early 1980's, and include Factor of 2, Mean Bias, Mean Gross
555 Error, Normalized Mean Gross Error, Correlation Coefficient, Root Mean Square Error, Coefficient of Efficiency,
556 and Index of Agreement. The formulae for these metrics and a brief description of their relative advantages and
557 disadvantages appears in Appendix A (Supplemental Information).
558
559 ~~against hourly observations collected from surface measurement stations.~~ Both simulations' performance for ozone
560 (O_3), nitrogen dioxide (NO_2) and particulate matter with diameters less than 2.5 μm(PM2.5) were evaluated using the
561 above metrics, employing hourly AIRNOW data (USA: AQS network: https://www.epa.gov/aqs; Canada: NAPS
562 network: http://maps-cartes.ec.gc.ca/rnspa-naps/data.aspx) and the *openair* package (Carslaw and Ropkins, 2012).~~)~~
563 The summary performance metric scores for the two simulations grouped, according to contributing measurement
564 network, are shown in Table 3~~2~~, with boldface values indicating the better score for the given simulation case. With
565 respect to this table, we note that:
566 (a) The feedback simulation generally outperforms the no-feedback simulation (more bold-face scores in the
567   " feedback" rows, for 35 out of 48 metric comparisons~~columns, with a few notable exceptions, discussed below~~).
568 (b) Feedback forecast score improvements occurred were more noticeable for PM2.5 (usually first to second digit),
569   followed by O_3, with the NO_2 scores often being the same for the first few digits.~~In some evaluation metrics, the~~
570   ~~feedback simulation showed substantial quantitative improvements over the no-feedback simulation (e.g.~~
571   ~~feedback PM2.5 MB is reduced by over a factor of 3 relative to its no-feedback counterpart over Western Canada,~~
572   ~~the region of wildfire activity).~~
573 (c) We note that the boundary conditions employed for our 2.5km simulations had a strong impact on model air-
574   quality performance. As described above, these boundary conditions originated in a 10-km resolution simulation
575   making use of ECMWF global reanalysis values on its own lateral boundaries. The magnitudes of the statistics
576   of Table 3 may be compared to the magnitudes of the statistics from our initial ACPD submission (which made
577   use of a MOZART 2009 reanalysis for chemical lateral boundary conditions for the 2.5km GEM-MACH
578   domain); these earlier results are shown in the S.I., Table A2~~)~~. The use of feedbacks had a similar relative impact
579   on forecast performance (34 out of 48 statistics improving in the feedback forecast in the initial simulation,
580   compared to 35 out of 48 statistics in the current work). However, the net impact of the ECMWF-driven 10-km

581 GEM-MACH values being used for chemical lateral boundary conditions, rather than the MOZART climatology,

582 was a degradation of performance: comparing the equivalent entries in Table 3 and Table A2, it can be seen that

583 71 out of 96 scores were better with the earlier use of the MOZART reanalysis. As we show below, however,

584 the revised boundary conditions led to improvements in model aerosol optical depth performance relative to

585 observations.

586

587 (c) For cases when the no-feedback simulation outperforms the feedback simulation, the relative magnitude of the

588 performance difference is smaller than the feedback simulation's improvements (e.g. Western USA PM2.5 and

589 Western Canada NO$_2$ mean bias degradations of 9.0 and 27.8 percent relative to the PM2.5 improvement of a

590 factor of 3 noted above).

591 (d) Both simulations have negative mean biases for O$_3$ of -3.5 to -3.7 ppbv throughout the model domain, negative

592 biases for PM2.5 in the Western USA, positive biases for PM2.5 in Western Canada, negative biases for NO$_2$ in

593 Western Canada and positive biases for NO$_2$ in the Western USA.

594 One possible cause for the overall model biases noted in (d) may reside in inadequate chemical boundary conditions

595 used for the forecasting setup. An unprecedented large high-resolution (2.5-km) model domain and on-time delivery

596 of forecasts in support of FIREX-AQ put a significant constraint on available computational resources. 2.5 km

597 chemical boundary conditions for these simulations were taken from seasonal climatologies from ECMWF global

598 model analyses, rather than coarse resolution model simulations. During the time simulated, large forest fires occurred

599 both within the domain (in northern Alberta and Saskatchewan in Canada), and outside of this model domain (in

600 Alaska north of the Panhandle). These out-of-domain sources were thus not available as lateral boundary conditions

601 on the 2.5-km domain, with possible impacts on model performance for the three species thus resulting in overall

602 forecast biases in both the feedback and the no-feedback simulations, particularly in Western Canada.

603 The impact of lateral boundary conditions on model predictions can be seen when comparing MODIS retrievals of

604 aerosol optical depth (AOD) with model predictions (Figure 14). AOD is a function of both the particle's abundance

605 and optical properties, integrated throughout the vertical column. However, direct comparisons between satellite and

606 model-predicted AOD values must be undertaken with some care, due to the nature of the satellite retrievalretrievals

607 quality assurance and control procedures, the motion of the orbiting spacecraft, and the scan time of the instrument.

608 The manner in which AOD is calculated introduces additional uncertainty due to the range of values which may be

609 derived from the same aerosol speciation using different methodologies (Curci *et al.*, 2015). For a polar-orbiting

610 instrument such as MODIS, the time at which overpasses occur varies with location, and valid satellite retrievals may

611 not occur when the location being scanned is obscured by clouds. Observed averages may be built up over multiple

612 valid scans over time, but the number of valid scans contributing to the local average at any given location will vary,

613 due to the time and space variation in cloud cover. Here, individual valid Collection 6.1 MODIS/Aqua (MYD04_L2

614 AOD_550_Dark_Target_Deep_Blue_Combined) 10 km resolution 550 nm AODs were matched in time and space to

615 the nearest model 2.5-km grid cell and output frequency hour. Levy *et al.,* (2013) contains details on the MODIS

616 combined AOD product. No averaging was employed in our comparison (Figure 14); all satellite overpass AOD

617 pixels and matching model AOD pixels are shown.. Noting that the Both model and observed values were then locally

618 averaged across the simulation period, to a common resolution 0.1° latitude-longitude grid co-located with the model

619 grid, in order to generate the comparison shown in Figure 14. Individual pixels of this image thus incorporate a

620 spatially varying number of values in the local averages, but these local averages are matched in time, space, and local

621 averaging period. The region over which comparisons were made thus corresponds to the high resolution model

622 domain, and white areas within the images correspond to regions where satellite data used in the averaging were

623 excluded due to QA/QC constraints, such as the presence of clouds, surface ice, etc.

624 The MODIS-observed local average AOD values of Figure 14 (a,b) are generally higher than the model predictions

625 (c,d), with the exception of the region co-located with large forest fires in Northern Alberta and Saskatchewan (Figure

626 14 (d), yellow regions). In addition, as noted above, other satellite imagery has shown the presence of several large

627 forest fires occurring in Alaska, outside of the modelling domain, with smoke plumes extending from these sources

628 down through the northern and western coastal boundaries of the model domain (yellow regions in Figure 14(a)). It

629 is likely that at least some of the biases in PM2.5 and NO_x and consequently in the production of secondary ozone,

630 reflect the absence of these sources. Scatterplots of all paired AOD values in Figure 15 (a,b) and for the northern

631 portion of Alberta (Figure 15(c,d)) show that the overall negative bias is due to a large number of underestimated

632 "background" values of AOD in the model simulations, while in the immediate vicinity of the Alberta / Saskatchewan

633 forest fires, model values are considerably higher than observations.

634 The local model positive biases in AOD colour scale is logarithmic, the model simulation driven using the ECMWF

635 + 10-km resolution GEM-MACH for boundary conditions (Figure 14(b)) is a much better match to observations

636 (Figure 14(a)) than the model simulation driven by MOZART climatological boundary conditions (Figure 14(c)). The

637 slope of the linear best fit line between all observation and model pairs in each case mirrors this finding, with the

638 original (MOZART climatology) boundary conditions having a slope of 0.15 and R^2 of 0.0382, and the revised

639 ECMWF + GEM-MACH 10-km boundary conditions having a slope of 0.56 and an R^2 of 0.067.

640

641 within Northern Alberta might partly be attributed to overestimates in emitted particulate matter mass in the CFFEPS

642 module, or from inaccurate treatment of fire plume centerline dispersion downwind of primary emissions over large

643 fire sources. Previous work with CFFEPS by Chen *et al.* (2019) for the 2017 fire season has shown similar $PM_{2.5}$

644 positive biases for western Canada, with MB of +5.8 µg m^{-3} (88 stations) and for Western USA with MB of +8.6 µg

645 m^{-3} (221 stations). These positive biases (Chen *et al.,* 2019) were higher specific to sub-regions closer to areas of

646 active fires (MB of +12 µg m^{-3} for the sub-region including the provinces of Alberta and British Columbia, and +29

647 µg m^{-3} for the sub-region comprising the states of Idaho, Montana, Oregon and Washington, respectively). At least

648 part of the positive biases may be due to 10km GEM-MACH forest fire emissions occurring in the state of Alaska

649 being overestimated during the study period. However, the ECMWF reanalysis also captures significant particulate

650 mass crossing the Bering Strait from fires in Siberia during this period, so the relative contributions of fires within the

651 low resolution GEM-MACH domain and the ECMWF boundary conditions driving that domain are combined, and

652 can't be separated in the runs carried out Our analysis here and these earlier results suggest a positive bias in CFFEPS'

653 PM2.5 emissions or insufficient dilution/vertical extent of the predicted fire plumes.

654 The local AOD positive biases associated with fires could also be the result of the mixing state assumptions of the

655 Mie code used here for generating aerosol optical properties. These assumptions may also account for negative AOD

656 biases over much of the remainder of the model domain. ~~We have used a mass-weighted homogeneous mixture~~

657 ~~approach, with the complex refractive index values for the 8 particle species being calculated for pure water-dry~~

658 ~~component homogeneous mixtures at each of the 12 particle size bins, followed by mass weighting to generate values~~

659 ~~for each of the model components.~~ As noted earlier, this overall negative bias of AOD predictions (both boundary

660 condition configurations result in observation:model slopes less than unity) is a common problem in air-quality

661 models, and may be due to assumptions regarding the model mixing state (Curci *et al.*, 2015). That comparison of

662 multiple mixing state assumptions on AOD with observations for European and North American modelling domains

663 (Curci *et al.*, 2015), showed a typical factor of two model under-prediction of 440 nm North American AOD across

664 all mixing state assumptions, with European AOD negative biases ranging from unbiased to a factor of 2. ~~For the~~

665 ~~latter group, those models employing an assumption of external mixing, with hygroscopic growth factors for sulphate~~

666 ~~and nitrate assumed to be similar to those of sulphuric acid, had the highest AODs and hence closest values compared~~

667 ~~to observations at 440nm. However, in that investigation, the latter method also sometimes resulted in AOD over-~~

668 ~~predictions by a factor of 2 at 870 nm.~~ These earlier findings along with overestimates at forest fire plumes with our

669 current homogeneous mixture approach at 550nm suggest that the hygroscopic growth may be overestimated for forest

670 fire particles, in turn overestimating forest fire AODs locally, while external mixing assumptions may be required to

671 improve model AOD performance elsewhere in the model domain.

672 We note that the combined use of the ECMWF global reanalysis and a 10km resolution GEM-MACH simulation to

673 provide boundary conditions for our 2.5km domain resulted in a degradation of model performance for surface PM2.5,

674 O_3, and NO_2, for 71 out of 96 statistical scores, compared to the use of a MOZART2009 reanalysis for 2.5km domain

675 boundary conditions. The improvement associated with the use of feedbacks was maintained, showing that the impact

676 of feedbacks is a robust finding. However, the performance degradation associated with the change of boundary

677 conditions is a source of concern.

678 We also note that, while AOD performance has improved with the use of the ECMWF + GEM-MACH 10km boundary

679 conditions in Figure 14, significant overestimates of AOD occur with the use of these boundary conditions, in several

680 regions in the USA. This may be seen by comparing Figure 14 (a) and Figure 14 (b) for the states of Montana,

681 Wyoming, Nevada, New Mexico and Utah, where the dark blue colours in the observations (14(a)) indicate observed

682 AODs less than 0.01, whereas the ECMWF+GEM-MACH 10km driven simulation (14(b)) indicate values of 0.03 to

683 0.05). This is consistent with the increase in positive surface bias in PM2.5 associated with the use of the ECMWF +

684 GEM-MACH 10km boundary conditions (e.g., Feedback runs having Western Canada and Western USA positive

685 biases of PM2.5 of 4.578 and 1.805 $\mu g\ m^{-3}$ (Table 1), compared to the MOZART2009 reanalysis driven run values of

686 0.236 and -1.786 $\mu g\ m^{-3}$, Table A2, respectively). The boundary condition setup thus accounts for a substantial

687 increase in overall surface PM2.5 mass, with the use of ECMWF+GEM-MACH 10km increasing mean PM2.5 by

688 4.34 and 3.59 $\mu g\ m^{-3}$ in western Canada and USA respectively relative to the use of MOZART2009.

689 The reduction in performance may thus be due to two possible causes (or their combination): (1) the domain within

690 the GEM-MACH 10km simulation might be sufficiently large, and the emissions in the regions between the 10km

691 boundaries and the 2.5km domain sufficiently in error, that the 2.5km simulation accuracy is adversely affected; (2)
692 the ECMWF reanalysis employed on the outermost boundary of the 10km domain contributes sufficient PM2.5, NO_2
693 and O_3 to the simulations that the innermost domain model performance at the surface is adversely affected. That is,
694 the degradation in performance may be associated with the GEM-MACH 10km simulation, the ECMWF reanalysis,
695 or a combination of both factors.
696 In order to examine the potential impact of the ECMWF reanalysis used as the outermost domain boundary conditions,
697 we evaluated the ECMWF reanalysis using the same observation data, station locations, and performance metrics as
698 was used for our 2.5km simulations– the results of this analysis are shown in Supplemental Information Tables A3
699 (PM2.5), A4 (O_3) and A5 (NO_2). This additional analysis shows that the ECMWF reanalysis values during the study
700 period have higher positive biases for PM2.5 and O_3, and lower negative biases for NO_2, than the corresponding 2.5km
701 GEM-MACH simulations. A similar pattern occurs for the other statistical metrics, with the ECMWF reanalysis
702 usually having a reduced performance for most statistical scores for PM2.5, O_3, and NO_2 over the study region in
703 comparison to the high resolution model simulations carried out here. The ECMWF reanalysis was also evaluated for
704 the same time period for the entire North American domain, in Tables A3 to A5; the scores for this last analysis
705 suggest that the performance of the reanalysis relative to observations is similar over the continent, and is not limited
706 to our study area. We note that the ECMWF reanalysis has relatively low spatial and time resolution compared to the
707 2.5km simulations carried out here (0.75^ox0.75^o versus 2.5 km x 2.5km, 3 hourly output values compared to 1 hour
708 output values), and relatively coarse resolution for particle sizes (e.g. 3 size bins compared to the 12 bins used within
709 GEM-MACH; these issues may factor into the performance scores. The comparison does not rule out the possibility
710 that GEM-MACH's 10km resolution simulations may *also* contribute adversely to our 2.5km model performance.
711 However, our analysis suggests that our use of the ECMWF reanalysis for boundary conditions on our outermost
712 domain likely accounts for at least some of the performance degradation in our modelling system, compared to the
713 MOZART 2009 boundary condition simulation carried out earlier.

714 **3.3 Model Evaluation Summary**

715 Overall, the incorporation of feedbacks in this study has resulted in improvements in weather and air-quality forecast
716 accuracy, *albeit with some caveats*. Weather forecast variables showed improvements at the 90% confidence level
717 for several fields, and vertical profiles showed a matching performance or improvements at most levels, ~~particularly~~
718 ~~close to the surface,~~ and times. ~~with increasing forecast lead time.~~ Total precipitation scores also showed minor
719 improvements or matching performance at the 90% confidence level. ~~improved.~~ A previously unexpected spin-up
720 issue specific to on-line ~~fully~~ coupled models was noted: the impact of on-line ~~fully~~ coupled particulate matter on cloud
721 variables was sufficiently strong that cloud field adjustment in the first 6 hours of the forecast was required prior to
722 some weather forecast variable improvements to be apparent (surface pressure, dewpoint temperature, sea-level
723 pressure). While the current forecast cycling duration was constrained by operational requirements, this suggests that
724 forecast cycling should include both air-quality and meteorological variables during on-line ~~fully~~ coupled forecast
725 spin-up periods. That is, the model tracer concentrations 6 hours prior to the current forecast start-up could also be
726 used during the initial meteorological spin-up period, thus allowing chemistry and cloud formation to spin-up

20

727 simultaneously. Scores for surface PM2.5, NO$_2$, and O$_3$ also generally improved with the incorporation of feedbacks
728 (35 out of 48 comparisons showed improvements). The choice of lateral boundary conditions was shown to have a
729 significant impact on chemical performance within the model domain., with some metrics showing large
730 improvements. In comparison to satellite-based AOD values, the current model's AOD values were generally biased
731 low, with smaller magnitude biasesexceptions being associated with the ECMWF + 10-km GEM-MACH boundary
732 conditions.in the regions of Alberta and Saskatchewan with active forest fires where AOD was biased high. The latter
733 comparison also showed that large fires off-domain in Alaska and Siberia likely had a large impact on AODs in the
734 eastern and northern section of the model domain, through comparison with our initial due to missing boundary
735 condition contributions. These sources were missing due to operational limitations in the model simulations. shown
736 here.
737
738 **4 Effects of Feedbacks on Selected Simulation-Period Average Variables**
739 In this section, we compare time averages of the entire study period for the two simulations, both at the surface and in
740 vertical cross-sections through the model domain, to illustrate some of the changes in both weather and air-quality
741 associated with the incorporation of feedbacks. We have found differences at greater than 90% confidence between
742 the predicted meteorological and chemical forecasts in the vicinity of the Alberta/Saskatchewan forest fires, as well
743 as in contrasting changes between land and sea. We note again here that the "no-feedback" simulation makes use of
744 time and spatially invariant aerosol CCN and optical properties, within the meteorological portion of the model. The
745 comparisons thus show the differences associated with the use of climatological constant aerosol properties, and the
746 on-linefully coupled model-generated aerosols.
747 As in the meteorological evaluation, we have made use of 90% confidence levels in order to gauge the level of
748 significance of the differences between the feedback and no-feedback simulations in the following analysis. At each
749 model grid cell the values of the standard deviation about the mean for each respective simulation was calculated.
750 The difference between the means becomes significant at a given confidence level c if the regions defined by $M_f \pm$
751 $z^* \frac{\sigma_f}{\sqrt{N}}$ an$M_{nf} \pm z^* \frac{\sigma_{nf}}{\sqrt{N}}$d do not overlap, where N is the number of points averaged, M_f is the feedback mean value, M_{nf}
752 is the no-feedback mean value, σ_f and σ_{nf} are the corresponding standard deviation, and z^* is the value of the $\sqrt{c}$
753 percentile point for the fractional confidence interval c of the normal distribution ($z^* = 1.645$ at $c = 0.90$). Grid cell
754 values where this overlap does not occur (i.e. where the mean values differ at or above the 90% confidence level) may
755 be defined via the following equation[1]:

756
$$\frac{|M_{nf} - M_f|}{\frac{z^*}{\sqrt{N}}(\sigma_f + \sigma_{nf})} > 1 \hspace{3cm} (1)$$

---

[1] 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
$\left(M_{nf} - M_f\right) / \left(\frac{z^*}{\sqrt{N}}(\sigma_{nf} + \sigma_f)\right) > 1$. Similarly, for cases where $M_f > M_{nf}$, significance at the confidence level
associated with z^* occurs when $\left(M_f - M_{nf}\right) / \left(\frac{z^*}{\sqrt{N}}(\sigma_{nf} + \sigma_f)\right) > 1$. Equation (1) may thus be used to describe both
conditions.

757 The approach for representing model grid value 90% confidence levels is described in detail in SI Appendix A2. The
758 differences in the mean grid cell values between the simulations for which the above quantity is greater than unity
759 ~~thus~~ differ at or greater than the 90% confidence level. Differences in the mean values, as well as the value of the
760 above ratio, are thus reported in the following section.

4.1 Effects of Feedbacks on Time-Averaged Meteorology

The feedback – no-feedback differences in the simulation-period average cloud droplet number density (number kg^{-1} of air) and mass density (g water kg^{-1} of air) along centred cross-sections spanning the length and width of the 2.5-km resolution model domain are shown in Figure 15 (cross-section ~~16 (the~~ locations ~~of the cross-sections~~ are shown in Figure 1). The "Ocean", "Land", and "Forest Fire" regions identified are with reference to the approximate locations of these features along these cross-sections. Figure 15 ~~16~~ also shows the confidence ratio values as described above – regions where the predicted mean values differ at or above the 90% confidence level are shown in red, while those differences below the 90% confidence interval are shown in blue. Feedbacks increase the cloud droplet number density over the northern part of the domain, including the region impacted by the Alberta/Saskatchewan forest fires, from the surface up to about 500 ~~600~~ mb (roughly equivalent to hybrid level 0.500 ~~600~~), and decrease at higher elevations further to the south ~~aloft~~ and along the length of the model domain into the western USA (Figure 15 ~~16~~(a)). Cloud droplet numbers also decrease over the ocean, but increase eastwards over the land (Figure 15 ~~16~~(b)). The latter is unrelated to the forest fires; this is an indication that the modelled aerosol number concentration over the ocean is much lower than the single climatological aerosol population assumed in the no-feedback run, resulting in lower cloud droplet number concentrations. The changes ~~In both cases, the differences~~ are significant at the 90% confidence level from the surface up to hybrid level 0.60 in the northern region which is most impacted by forest fire smoke, ~~87~~ and in isolated regions further aloft ~~at hybrid level 0.550~~ along the south to north cross-section (Figure 15 ~~16~~(c)), and over the regions of the ocean in the west to east cross-section (Figure 15 ~~16~~(d)). Higher-than-climatology aerosol loadings, a large portion of which are due to the forest fires, resulted increased cloud droplet number densities in the lower troposphere, while decreasing them in the mid-to-upper troposphere (Figure 15(a)). This impact of feedbacks is in accord with the satellite observations of Saponaro *et al.* (2017), and was also seen in Takeishi *et al.* (2020). In contrast, cloud droplet mass density (i.e. cloud liquid water content) largely decreases across the domain along the north-south cross-section (Figure 15 ~~16~~(e)), as well as over the ocean, with a varying pattern over the land in the east-west cross-section (Figure 15 ~~16~~(f)). The magnitudes and significance levels for the average change in cloud droplet mass are lower than for cloud droplet number, with the most significant differences occurring over the ocean (Figure 15 ~~16~~(g,h)). Consistent with the cloud droplet number changes, rain droplet numbers and mass mixing ratios increase aloft with the feedback simulation, over both the forest region impacted by the forest fires (Figure 16 ~~17~~(a,e)) and over the ocean (Figure 16 ~~17~~(b,f)), with a varying impact over the land and more distant from the forest fire sources (Figure 16 ~~17~~(f)). The changes are significant at the 90% confidence level for rain droplet number in these regions (compare Figure 16 ~~17~~(a) with 16 ~~17~~(c); 16 ~~17~~(b) with 16 ~~17~~(d)), while the rain droplet mass changes sometimes reach ~~approach~~ but are usually below the 90% confidence level (Figure 16(g,h)).

22

792 These results suggest that relative to the no-feedback simulation, which employs climatological aerosol CCN

793 properties, the AIE in the feedback simulation is causing significant change in hydrometeor numbers, and a less

794 significant increase in hydrometeor mass. In the forest fire-impacted region, the ADE and AIE in the feedback

795 simulation significantly increase the number of cloud droplets near the surface and throughout~~decrease~~ the ~~number of~~

796 ~~cloud droplets in the~~ middle to upper troposphere (Figure 15~~16~~(a,c)). The rain drop number in the middle troposphere

797 (Figure 16~~17~~(a,c)) also increases significantly between hybrid levels 0.90 to 0.70 (Figure 16~~17~~(e,g)). Near-surface

798 rain drop number and rain drop mass differences throughout the cross sections (Figure 16~~17~~(e,f)) fall below the 90%

799 confidence level (Figure 16~~17~~(g,h).

800 Over the oceans, water droplet number and mass both decrease (Figure 15~~16~~(b,f)), and raindrop number and mass

801 increase (Figure 16~~17~~(b,f)); more atmospheric water is converted to rain drops as a result of the feedbacks, relative to

802 the climatology in the no-feedback simulation. However, these changes are more significant aloft than at the surface,

803 with the difference in both rain drop number and mass falling below the 90% confidence level near the surface. We

804 interpret these changes as a shift in over-ocean liquid hydrometeor numbers and to a lesser degree the water mass aloft

805 from cloud droplets to rain drops due to the AIE in the feedback setup relative to the climatology of the no-feedback

806 simulation. The changes occur at the 90% confidence level aloft, but the near-surface changes are smaller and are

807 usually below the 90% confidence level.

808 Differences in the average surface precipitation flux and the confidence ratio values are shown in Figure 17~~18~~.

809 Changes in average precipitation (Figure 17~~18~~(a)) appear random, though locally these differences are significant at

810 the 90% confidence level (Figure 17~~18~~(b)). Both the magnitude of the differences and the frequency in their reaching

811 the 90% confidence level increase south-westwards. Given the local and episodic nature of rainfall events, the high

812 level of significance in this case probably results from the presence or absence of individual rainfall events between

813 the two simulations affecting the local average and standard deviations.

814 Several systematic changes in the average values of the model's meteorological output fields were noted due to the

815 use of feedbacks relative to aerosol property climatologies (Figure 18~~19~~), although all fall below the 90% confidence

816 level for the difference in the mean values between the two simulations (Figure 19). Specific humidity~~20). Surface air~~

817 ~~temperature generally~~ increased ~~(Figure 19(b)) though less so~~ in the region most affected by ~~forest~~ fires (Figure 18(a).

818 surface air temperature decreased below the smoke plumes while increasing further south (Figure 18(b)), while~~

819 dewpoint temperature decreased (Figure 18~~19~~(c))~~.))~~ implying a decrease in relative humidity with feedbacks. Surface

820 pressure increased over the land (mostly east of the Rockies),~~,~~ particularly in the region downwind of the Alberta /

821 Saskatchewan fires while decreasing over the ocean (Figure 18~~19~~(d)). Planetary boundary layer height increased over

822 the land (Figure 18~~19~~(e)) except in the immediate vicinity of the Alberta/Saskatchewan fires,~~)),~~ consistent with

823 decreased atmospheric stability in the lowest part of the atmosphere.~~.~~ The friction velocity also increased with the use

824 of feedbacks (Figure 18~~19~~(f)); this is consistent with a decrease in stability and an increase in turbulent energy The

825 air temperature increases occur at the surface south of the forest-fire impacted region~~are limited to the lowest part of~~

826 ~~the atmosphere (Figure 21 (a,b)), usually below hybrid level 0.914 (approximately 1km above the surface). The~~

827 ~~feedbacks decrease temperatures between hybrid levels 0.914~~ and above roughly 750 mb, decreasing temperatures

828 from the surface in the forest-fire impacted region up to 750 mb (Figure 20 (a,b)).~~0.721.~~ Feedbacks thus increase

23

829 near-surface ~~or do not affect~~ temperatures, relative to the no-feedback meteorological model's simple aerosol
830 climatology, in regions far from the fires, decreasing them near the fires,~~surface, and~~ decrease temperatures in the
831 lower free Troposphere, and increase temperatures further aloft. All.~~However, all~~ of these differences between
832 feedback and no-feedback simulations~~features~~, despite their large geographic range, fall below the local 90%
833 confidence ratio. However, when the differences in air temperature resulting associated with feedback and no-
834 feedback forecasts are compared to observations across the entire domain (as opposed to at gridpoint locations as in
835 Figures 18 and 19)~~90% confidence level, reflecting the large variability in surface temperatures contained within each~~
836 ~~simulation. Longer time simulations than carried out here are required in order to improve confidence in the~~
837 ~~temperature predictions across all forecast hours. However, these results, particularly for surface temperature, may~~
838 ~~be contrasted with Figures 6(a), 12 and 13, which suggests that temperature differences rise above~~ the 90% confidence
839 level is exceeded both at the surface ~~interval~~ at specific forecast times (Figure 6(a)), and at multiple heights aloft at
840 the 12th and 24th forecast hours (Figures 12, 13). ~~at surface and upper atmosphere measurement sites.~~

841 **4.2 Effects of Feedbacks on Time-Averaged Chemistry**

842 In the previous meteorological impacts section, changes in aerosol loading relative to the climatology, dominated by
843 forest fires, were shown to have a significant impact on cloud formation and atmospheric temperatures through ADE
844 and AIE. These might be expected to in turn influence and be influenced by particulate matter emitted by the forest
845 fires, with the plume rise of the forest fires dependent on the meteorological changes. Air temperatures increase
846 slightly in the model surface layer south of the fires (Figure 18~~19~~(b), +0.01 to +0.05 ^{o}C) but decrease at greater
847 magnitudes through the rest of the lower Troposphere (surface near the fires to hybrid level 0.749~~980 up to 0.667~~,
848 Figure 20(a)),~~21)~~, with a maximum decrease of -0.5^{o}C between hybrid levels 0.893~~871~~ and 0.848~~824~~. The
849 reduction~~increase~~ in temperatures between hybrid levels 0.90 to 0.70 from the impact of the smoke plumes is similar
850 to the findings of Saponaro *et al.* (2017). These changes air temperatures ~~near the surface and decreases aloft (a~~
851 ~~stronger negative vertical gradient in temperature)~~ implies a decrease in near-surface atmospheric stability associated
852 with feedbacks, given that the overall temperature gradient from the surface has become more negative (that is, the
853 ambient lapse rate has increased). Rising air parcels will follow an adiabatic lapse rate; these increases in the ambient
854 lapse rate imply that rising air parcels will have an increasing tendency to be warmer than their environment.
855 Feedbacks have thus reduced atmospheric stability within the forest fire smoke in the lowest part of the atmosphere;
856 the atmosphere there has become more unstable. Meanwhile, the feedbacks decrease the environmental lapse rate
857 further aloft above the forest fire smoke, between hybrid levels 0.848 and 0.339. Rising air parcels in this region
858 following an adiabatic lapse rate will thus have an increasing tendency to be colder than their environment – the
859 atmosphere above the smoke plumes has become more stable.~~.~~ This is echoed by the response of the concentration
860 fields to the near-surface stability change, as can be seen through comparisons of the PM2.5, NO$_2$ and O$_3$ surface
861 concentrations changes (Figure 21~~22~~) and as vertical cross-sections (Figures 22, 23, 24,~~25~~), respectively.
862 Changes~~For all three surface fields, changes~~ above the 90% confidence level for PM2.5 and NO$_2$ occur near the forest
863 fires themselves (red regions, near top of model domain, Figure 21~~22~~(a,b)), though remain below,~~c)). The differences~~
864 ~~in particulate matter concentrations are also significant at the~~ 90% confidence for O3~~level throughout the model~~

24

865 ~~domain (Figure 21(c)).~~ 22(a)). ~~Note that while the PM2.5 mean values are significantly different at the 90%~~

866 ~~confidence level throughout the model domain, the magnitude of those differences are sometimes small, particularly~~

867 ~~in the upper atmosphere, where the aerosol concentrations are relatively small. However, the regions with the larger~~

868 ~~magnitude regional differences in PM2.5 concentrations also occur at greater than the 90% confidence level (compare~~

869 ~~spatial locations of coloured regions in Figure 22(a,b) to red regions in Figure 22(c,d)).~~

870 Feedbacks result in near~~Near~~-surface PM2.5 decreases in the regions downwind of the forest fires (Figure 21~~22~~(a),

871 Figure 22~~23~~(a), note the large blue region and more intense blue region near surface in Figure 22~~23~~(a)), suggesting

872 less PM2.5 mass is present near the surface due to the feedbacks. Given the increase in near-surface stability below

873 the fire plumes noted above, this ~~This could reflect a~~ change in the vertical distribution ~~probably reflects a~~ ~~injection~~

874 ~~height of the plumes in addition to other transport changes associated with the~~ decrease in downward diffusive mixing

875 of the forest fire plumes once aloft – the feedbacks thus have a tendency to increase the smoke plume concentrations

876 aloft, by preventing the downward mixing of smoke injected by the fires. These ~~atmospheric stability. Lower~~

877 ~~troposphere decreases in~~ PM2.5 concentration effects rise above ~~over the ocean and increases over the land at or greater~~

878 ~~than~~ the 90% confidence level within the region closest to the fires. ~~may also be seen (Figure 23(b,d)).~~

879 Feedbacks result in an increase in near-surface NO_2 in several inland urban centers and less NO_2 at surface level

880 downwind (Figure 21~~22~~(b), though these differences are only significant at the 90% confidence level within the forest

881 fire plumes (Figure 21(e), Figure 23(24~~(a,~~c)). Ocean versus land NO_2 differences remain below the 90% confidence

882 level.

883 Feedbacks decreased lower Troposphere~~surface~~ O_3 near the forest fires (Figure 21~~22~~(c), Figure 24~~25~~(a)), while

884 ~~decreasing~~ increasing O_3 near above hybrid level 0.383.~~aloft.~~ The forest fires are also the only area where the

885 differences in between mean ozone forecasts approach~~have greater than the~~ 90% confidence.

886 Overall, the most significant effects of the feedbacks were: (1) increases~~changes~~ in PM2.5 aloft and decreases

887 near~~concentrations throughout~~ the surface in areas impacted by the fires~~model domain~~, and (2) increases~~changes~~ in

888 NO_2 aloft and decreases near the surface near the fires, to lesser extent than PM2.5, and (3) decreases in lower

889 troposphere ~~and~~ O_3, particularly near the surface in the region impacted by the fires ~~within the forest fire plumes~~.

890 The feedback-induced changes in primary and secondary pollutants in the forest fire regions are consistent with the

891 decrease in atmospheric stability noted above – a greater proportion of the primary particulate matter and NO_2 resulting

892 from near-surface forest fire emissions of NO remain aloft~~are carried upwards~~ with the addition of feedbacks. The

893 decrease in surface ozone and increase further aloft in the fire region (Figure 24~~25~~(a)) spatially matches the

894 decrease~~changes~~ in surface NO_2 (Figure 22~~23~~(a)). Chemically,)) this may imply~~implies~~ that the changes associated

895 with feedbacks occur in NOx-limited environments, i.e., with relatively high VOC/NOx ratios, since in these. ~~In such~~

896 environments, decreases in NOx emissions may lead to decreases in the rate of secondary O_3 formation. Alternatively,

897 the reduction in near-surface O_3 concentrations may reflect a decrease in light levels reaching the surface due to cloud

898 attenuation (aerosol indirect effect), with the resulting lower photolysis rates resulting in a reduction in surface

899 photochemical ozone production.

900 Our analysis thus suggests a net enhanced upward transport occurs in forest fire plumes due to feedbacks, and that this

901 transport is linked to feedback-induced:

902   (1) Increases ~~decreases~~ in local near-surface atmospheric stability, reducing downward mixing of particulate
903 plumes once aloft (Figure 22~~21~~(a));
904   (2) Increases~~increases~~ in cloud droplet numbers throughout~~near~~ the lower troposphere~~surface~~ (Figure
905 15~~16~~(a)); and
906   (3) Increases~~increases~~ in rain drop numbers aloft (Figure 16~~17~~(a)).
907   This combination suggests the presence of an AIE feedback loop – increased lower atmosphere~~decreased~~
908 stability results a greater proportion of particulate matter remaining aloft~~in higher forest fire plume rise~~, in turn
909 resulting in~~lofting proportionally~~ more particles remaining at~~to~~ higher levels in the atmosphere where they may act as
910 cloud condensation nuclei, increasing cloud droplets aloft (Figure 15~~16~~(a)). This in turn results in increased lower
911 middle troposphere cooling, through the 1st AIE (increase in cloud droplet numbers aloft leading to increased cloud
912 albedo and cooling of the atmosphere below the cloud tops) while the corresponding decreases in particles and cloud
913 condensation nuclei at lower levels results in a smaller near-surface impact on the AIE and ADE, hence relatively
914 minor changes on near-surface temperatures (Figure 20~~21~~(a)). This combination maintains a feedback-induced near-
915 surface unstable~~less stable~~ temperature gradient, relative to the no-feedback simulation employing aerosol property
916 climatologies. We acknowledge that these changes in temperature fall below the 90% confidence level for the
917 averages over all times, though note that differences in mean bias relative to observations for the two simulations
918 became significantly different at specific times of day in the forecasts (Figure 6(a), hours 3, 6, 15 and 18,
919 corresponding to 15, 18, 3 and 6 UT, or 9 AM, 12 noon, 9 PM, and midnight MDT), implying that the temperature
920 changes at these specific times reach a higher level of significance. Similarly, Figures 12 and 13 show reductions in
921 the near-surface temperature biases with the use of feedbacks.
922 ~~Similarly, over the oceans, the feedback-induced decrease in surface PM2.5 (Figure 23(a)) is accompanied by lower~~
923 ~~middle troposphere cooling (Figure 21(b), note the dark blue band over ocean between hybrid levels 0.914 and 0.824),~~
924 ~~implying a decrease in stability. While aerosol increases aloft are not large, the lower level PM2.5 act as CCN,~~
925 ~~resulting in increased cloud formation and convection, in turn increasing cloud droplet and rain drop numbers aloft,~~
926 ~~which through the 1st AIE, maintains the slightly less stable temperature profile. We acknowledge that these changes~~
927 ~~in temperature fall below the 90% confidence level for the averages over all times, though note that differences in~~
928 ~~mean bias relative to observations for the two simulations became significantly different at specific times of day in~~
929 ~~the forecasts (Figure 6(a), hours 3, 6, 15 and 18, corresponding to 15, 18, 3 and 6 UT, or 9 AM, 12 noon, 9 PM, and~~
930 ~~midnight MDT), implying that the temperature changes at these specific times reach a higher level of significance.~~

931 **4.3 Summary, Differences in Forecast Simulation-Period Averages**

932 Relative to the no-feedback simulation employing an aerosol climatology, the AIE feedback as simulated here is
933 associated with increases~~decreases~~ in near-surface stability over both ocean and forest-fire influenced land areas. Over
934 oceans, near-surface particulate matter is removed as cloud condensation nuclei, resulting in increased cloud droplet
935 numbers, maintaining the temperature gradient through the 1st aerosol indirect effect. In the vicinity of forest fires,
936 increases~~decreases~~ in near-surface stability result in more~~increased transport of~~ PM2.5 remaining aloft, increasing the
937 availability of cloud condensation nuclei aloft, increasing cloud droplet numbers aloft, hence also maintaining the less

26

Formatted: Indent: First line: 1.27 cm

938 stable near-surface~~increased~~ temperature gradient through the 1st aerosol indirect effect. We note that the ADE may
939 also play a weak role, particularly in the southern part of the domain, where lower atmosphere temperature gradient
940 increases are not accompanied by significant changes in cloud droplet numbers (Figure 15~~16~~(a), southern half of the
941 cross-section), but are accompanied by significant though small magnitude increases in PM2.5 in the lower atmosphere
942 (Figure 22~~23~~(a), southern half of cross-section), and temperature profile changes (Figure 20~~21~~) below the 90%
943 confidence level.

944

945 **5 Conclusions**
946 The work carried out here suggests that the answers to our two research questions ("Can on-line~~fully~~ coupled models
947 improve both air-quality and meteorological forecasts?" and "Are the changes in forest fire forecasts associated with
948 implementing forest fire emissions within a on-line~~fully~~ coupled model sufficient to significantly perturb weather and
949 chemistry~~meteorology~~?") are both a *qualified* "yes". Within
950 ~~The simulations analyzed here were conducted in preparation for an experimental forecast carried out as part of the~~
951 ~~FIREX-AQ campaign, and hence were limited by operational time constraints to a sequence of nested 24-hour~~
952 ~~forecasts. However,~~ the high resolution domain size employed here, ~~was sufficiently large to result in~~ improvements
953 or matching~~in~~ weather forecast performance was seen for most times~~both surface~~ and heights in the atmosphere, ~~profile~~
954 ~~variables~~ at greater than~~or above the~~ 90% confidence~~level~~. Improvements in model performance for surface PM2.5,
955 NO$_2$ and O$_3$ were also found, across most statistical measures (35 out of 48 statistical evaluation scores showed
956 improvements). Comparing average vertical cross-sections, the chemical concentration changes associated with
957 feedbacks were . ~~The differences between feedback and no-feedback simulations occurred at or above~~ the most
958 significant close to ~~90% confidence level throughout the model domain for PM2.5, though were limited at the 90%~~
959 ~~confidence level for NO$_2$ and O$_3$ to the immediate vicinity of~~ the forest fires in the northern portion of the domain.~~.~~
960 There, increased net vertical transport associated with decreased near-surface stability ~~feedbacks~~ lowered near-surface
961 PM2.5 and NO$_2$ concentrations and increased them aloft, and resulted~~resulting~~ in reduced surface O$_3$. ~~in the NOx-~~
962 ~~limited regions of the forest fire plume.~~
963 Our~~The~~ simulations suggest that aerosol optical depth in the region, as well as the overall chemical performance of
964 the model, was strongly influenced by upwind boundary conditions. AODs were biased low despite PM2.5 positive
965 biases, suggesting that the homogeneous mixture approach for aerosol optical properties results in a general under-
966 prediction of aerosol optical depths, in accord with Curci *et al. (2015), and that obtaining better data for forest fire
967 aerosol optical properties should be a priority for future study, as well as an examination of*~~(2015), and~~ external
968 mixture approaches. Positive ~~are recommended in further study. However, this general negative bias in simulated~~
969 AOD ~~is locally offset by positive~~ biases in the region affected by ~~vicinity of forest~~ fires. ~~This~~ suggests that forest fire
970 plumes have significantly different optical properties, and may be less hygroscopic, than industrial aerosols of
971 comparable size. Special / separate treatment of forest fire CCN and optical properties are therefore also recommended
972 in future work.
973 On-line~~Fully~~ coupling forest fire plume rise calculations with the weather parameters was shown to have a significant
974 impact on the height of primary pollutants reached by forest fires, the formation of near-surface ozone near the forest

fires, and on particulate matter. ~~throughout much of the three-dimensional model domain.~~ These changes were largely driven by the AIE, which maintains an increased lapse rate (decreased near-surface ~~temperature gradient (reduced~~ stability) over the forest-fire-influenced and oceanic portions of the region studied. Weak evidence for the influence of the ADE was shown in the southern part of the domain, where increases in particulate matter were also accompanied by decreases in stability between the surface and the lower-middle troposphere (the differences were at a lower than 90% confidence level for these comparisons of temperatures averaged over all model times).

Relative to the no-feedback aerosol climatology for CCN and aerosol optical properties, the simulations carried out here suggested that in the vicinity of forest fires feedbacks significantly increase cloud droplet ~~near the surface, increase cloud droplet~~ number densities near the surface and aloft, and significantly increase rain drop number densities aloft, relative to forecasts driven by climatological aerosol properties. Over the oceans, feedbacks decreased cloud droplet number density and increased rain drop number density aloft, relative to the simulation employing invariant CCN properties. Oceanic cloud. ~~Cloud~~ droplet mass increased to a lesser degree (with smaller regions above the 90% confidence level), as did rain drop mass (the mean differences for which for the most part remained below the 90% confidence level). This provides some evidence for a shift in atmospheric water mass associated with feedbacks from cloud water to rain over the oceans relative to the no-feedback climatology, though this shift occurred largely within the variability of the cloud fields within each simulation. Longer simulations may be needed to achieve higher confidence in this finding.

Data Availability

The datasets used here for model evaluation are available from the publicly accessible websites (AQS network) https://www.epa.gov/aqs and (NAPS network) http://maps-cartes.ec.gc.ca/rnspa-naps/data.aspx.

Author Contribution

PAM: experiment design, conceptualization, analysis, writing of manuscript drafts; AA: model code and run script design and implementation, statistical analysis of model results, model analysis graphics; JC: forest fire emissions processing system design and coding, manuscript contributions, draft review and assistance; BP: forecast system simulations and design; WG~~Wanmin Gong~~[1]: indirect effect updates, advice on P3 implementation~~implmentation~~, manuscript contributions, manuscript review; CSC~~Stroud~~: code version contributions, manuscript review; CSC~~Sioris~~: AOD analysis, manuscript contributions and review; KS: forest fire emissions processing system design and coding, manuscript contributions and review; PC: forecast system simulations and design; JZ: emissions processing and input field assistance, manuscript review and contributions; J.M.: indirect effect updates and advice on implementing AIE in the P3 scheme, manuscript review and contributions.

Competing Interests

The authors declare that they have no conflict of interest.

References

Abdul-Razzak, H., and Ghan, S.J.: A parameterization of aerosol activation. 3. Sectional representation, J. Geophys. Res Atm., 107 (D3), http://dx.doi.org/10.1029/2001JS000483, 6 pp, 2002.

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

Ahmadov, R., James, E., Grell, G., Alexander, C., Benjamin. S., McKeen, S., Pereira, G., Freitas,S., Csiszar, I., Tsidulko, M., Kondragunta, S., Xu, C., Wong, K.Y., and Albers, S.: Forecasting smoke, visibility and smoke-weather interactions using a coupled meteorology-chemistry modeling system: Rapid Refresh and High-Resolution Rapid Refresh coupled with Smoke (RAP/HRRR-Smoke) , EGU 2019 Conference presentation, Geophysical Research Abstracts, Vol. 21, EGU2019-18605, 2019. https://meetingorganizer.copernicus.org/EGU2019/EGU2019-18605.pdf, 2019.

Akingunola, A., Makar, P.A., Zhang, J., Darlington, A., Li, S.-M., Gordon, M., Moran, M.D., and Zheng, Q.: A chemical transport model study of plume-rise and particle size distribution for the Athabasca oil sands, Atm. Chem. Phys., 18, 8667-8688, 2018.

Albrecht, B.A.: Aerosols, cloud microphysics, and fractional cloudiness, Science, 245(4923), 1227-1230, https://doi.org/10.1126/science.245.4923.1227, 1989.

Baro, R., Palacios-Pena, L., Baklanov, A., Balzarini, A., Brunner, D., Forkel, R., Hirtl, M., Honzak, L., Perez, J.L., Pirovano, G., San Jose, R., Schroder, W., Werhahn, J., Wolke, R., Zabhar, R., Jimenez-Guerrero, P.: Regional effects of atmospheric models on temperature: an evaluation of an ensemble of online coupled models, Atm. Chem., Phys., 17, 9677-9696, 2017.

Briggs, G.A.: A plume rise model compared with observations, J. Air. Poll. Cont. Assoc., 15, 433-438, https://doi.org/10.1080/000022470.1965.10468404, 1965.

Briggs, G.A.: Plume rise and buoyancy effects, atmospheric sciences and power production, in: DOE/TIC-27601(DE84005177), edited by Randerson, D., Technical Information Center, U.S. Dept. of Energy, Oak Ridge, TN, USA, 327-366, 1984.

Carrera, M.L., Belair, S., Bilodeau, B.: The Canadian Land Data Assimilation System (CalDAS): description and synthetic evaluation study, J. Hydrometeor., 16, 1293-1314, 2015.

Carslaw, D.C., and Ropkins, K., openair – An R package for air quality data analysis, Env. Mod. Soft., vol. 27-28, pp 52-61, 2012.

Chen, J., Anderson, K., Pavlovic, R., Moran, M.D., Englefield, P., Thompson, D.K., Munoz-Alpizar, R., and Landry, H.: The FireWork v2.0 air quality forecast system with biomass burning emissions from the Canadian Forest Fire Emissions Prediction System v2.03, Geosci. Model Dev., 12, 3283–3310, 2019.

Clark, T.L., Coen, J. and Latham, D.: Description of a coupled atmosphere–fire model. International Journal of Wildland Fire, 13(1), pp.49-63, 2004.

Clark, T.L., Jenkins, M.A., Coen, J.L., Packham, D.R.: A coupled atmosphere-fire model: role of convective froude number and dynamic engineering at the fireline, Int. J. Wildland Fire, 6, 177-190, 1996.

29

Coen, J. Cameron, M., Michalakes, J., Patton, E., Riggan, P., and Yedinak, K., WRF-Fire: coupled weather-wildfire modelling with the Weather Research and Forecasting Model, J. Appl. Meteorol. Climatol., 52, 16-38, 2013.

Côté, J., Gravel, S., Méthot, A., Patoine, A., Roch, M., Staniforth, A.: The operational CMC/MRB global environmental multiscale (GEM) model. Part 1: design considerations and formulation. Mon. Wea. Rev., 126, 1373-1395, 1998.

Curci, G., Hogrefe, C., Bianconi, R., Im, U., Balzarini, A., Baro, R., Brunner, D., Forkel, R., Giordano, L., Hirtl, M., Honzak, L., Jimenez-Guerrero, P., Knote, C., Langer, M., Makar, P.A., Pirovano, G., Perez, J.L., San Jose, R., Syrakov, D., Tuccella, P., Werhahn, J., Wolke, R., Zabkar, R., Zhang, J., and Galmarini, S.: Uncertainties of simulated aerosol optical properties induced by assumptions on aerosol physical and chemical properties: an AQMEII2 perspective, Atm. Env., 115, 541-522, 2015.

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.

Ezhova, E., Ylivinkka, I., Kuusk, J., Komsaare, K., Vana, M., Krasnova, A., Noe, S., Archinov, M., Belan, B., Park, S-B., Lavric, J.V., Heimann, M., Petaja, T., Vesala, T., Mammarella, I., Kolari, P., Back, J., Rannik, U., Kerminen, V.-M., and Kulmala, M.: Direct effect of aerosols on solar radiation and gross primary production in boreal and hemiboreal forests, Atm. Chem. Phys., 18, 17863-17881, 2018.

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.

Galmarini, S., Hogrefe, C., Brunner, D., Makar, P., Baklanov, A., Preface, Atm. Env., 115, 340-344, 2015.

Geer, A.J.: Significance of changes in medium-range forecast scores, Tellus A., 68, 30229, http://dx.doi.org/10.3402/tellusa.v68.30229, 2016.

Girard, C., Plante, A., Desgagne, M., McTaggart-Cowan, R., Cote, J., Charron, M., Gravel, S., Lee, V., Patoine, A., Qaddouri, A., Roch, M., Spacek, L., Tanguay, M., Vaillancourt, P.A., and Zadra, A.: Staggered vertical discretization of the Canadian Environmental Multiscale (GEM) model using a coordinate of the log-hydrostatic-pressure type. *Mon. Wea. Rev.*, 142, 1183–1196, 2014.

Gong, S.L. Barrie, L.A., Blanchet, J.-P., von Salzen, K., Lohmann, U., Lesins, G., Spacek, L., Zhang, L.M., Girard, E., Lin, H., Leaitch, R., Leighton, H., Chylek, P., Huang, P." Canadian Aerosol Module: a size-segregated simulation of atmospheric aerosol processes for climate and air quality models. 1. Module development. J. Geophys. Res., 108, 4007, http://dx.doi.org/10.1029/201JG002002, 2003.

Gong, W., Beagley, S.R., Zhang, J., Cousineau, S., Chen, J., Sassi, M., Munoz-Alpizar, R., Morrison, H., Lyons, L., and Bellavance, P.: Modelling regional air quality in the Canadian Arctic: impact of North American wildfire and Arctic shipping emissions, in: Air Pollution Modelling and Its Application, XXIV, edited by Steyn, D.G. and Chaumerliac, N., Springer, Switzerland, 301-306, 2016.

Gong, W., Makar, P.A., Zhang, J., Milbrandt, J, Gravel, S., Hayden, K.L., Macdonald, A.M., Leaitch, W.R.: Modelling aerosol-cloud-meteorology interaction: a case study with a fully coupled air quality model (GEM-MACH), Atm. Env., 115, 695-715, 2015.

1082 Grell, G.A., and Freitas, S.R., A scale and aerosol aware stochastic convective parameterization for weather and air
1083   quality modelling, Atmos. Chem. Phys., 14, 5233-5250, 2014.

1084 Griffin, D., Sioris, C., Chen, J., Dickson, N., Kovachik, A., De Graaf, M., Nanda, S., Veefkind, P., Dammers, E.,
1085   McLinden, C.A. and Makar, P., The 2018 fire season in North America as seen by TROPOMI: aerosol layer
1086   height intercomparisons and evaluation of model-derived plume heights. Atmospheric Measurement
1087   Techniques, 13(3): 1427-1427, 2020.

1088 Hanna, S.R., Air quality model evaluation and uncertainty, J. Air Poll. Control Assoc., 38, 406-412, 1988.

1089 Hoegh-Guldberg, O., Jacob, D., Taylor, M., Bindi, Brown, S., Camilloni, I., Diedhiou, A., Djalante, R., Ebi, K.L.,
1090   Engelbrecht, F., Guiot, J., Hijioka, Y., Mehrotra, S., Payne, A., Seneviratne, S.I., Thomas, A., Warren, R., and
1091   Zhou, G.: Impacts of 1.5°C Global Warming on Natural and Human Systems. In: *Global Warming of 1.5°C. An*
1092   *IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global*
1093   *greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate*
1094   *change, sustainable development, and efforts to eradicate poverty* [Masson-Delmotte, V., Zhai, P., Pörtner, H.-
1095   O., Roberts, D., Skea, J., Shukla, P.R., Pirani, A., Moufouma-Okia, W., Péan, C., Pidcock, R., Connors, S.,
1096   Matthews, J.B.R., Chen, Y., Zhou, X., Gomis, M.I., Lonnoy, E., Maycock, T., Tignor, M., and Waterfield, T.
1097   (eds.)]. In Press, 2018.

1098 Inness, A., Ades, M., Agusti-Panareda~~Baier, F., Benedetti~~, A., Barre, J., Benedictow, A., Blechschmidt, A.-M.,
1099   Dominguez, J.J~~Bouarar, I., Chabrillat, S., Clark, H., Clerbaux, C., Coheur, P.~~., Engelen, R.., Eskes, H.J., Errera,
1100   Q., Flemming, J., ~~George, M., Granier, C., Hadji-Lazaro, J.,~~ Huijnen, V., ~~Hurtmans, D.,~~ Jones, L., Kipling, Z.,
1101   Massart, S., Parrington, M., Peuch, V-H~~Kaiser, J.W., Kapsomenakis, J., Lefever, K., Leitao, J~~., Razinger, M.,
1102   Remy, S., Schulz, M., and~~Richter, A., Schultz, M.G., Simmons, A.J.,~~ Suttie, M., ~~Stien, O., Thepaut, J.-N.,~~
1103   ~~Thouret, V., Vrekoussis, M., Zerefos, C.,~~ The CAMS~~MACC team: The MACC~~ reanalysis~~: an 8-yr data set~~ of
1104   atmospheric composition, Atmos~~Atm~~. Chem. Phys., 19, 3515-3556, 2019~~13, 4073-4109,~~
1105   ~~http://dx.doi.org/10.5194/acp-13-4703-2013, 2013~~.

1106 James, E., Ahmadov, R., and Grell, G.: Realtime Wildfire Smoke Prediction in the United States: the HRRR-Smoke
1107   Model, in: EGU General Assembly Conference Abstracts, vol. 20 of EGU General Assembly Conference
1108   Abstracts, p. 19526, 2018

1109 Jiang, Y., Yang, X-Q., and Liu, X., Seasonality in anthropogenic aerosol effects on East Asian climate simulated
1110   with CAM5, J. Geophys. Res. Atmos., 120, 10837-10861, doi:10.1002/2015JD023451, 2015.

1111 Jouan, C., J.A. Milbrandt, P. Vaillancourt, F. Chosson, H. Morrison: Adaptation of the Predicted Particles Properties
1112   (P3) microphysics scheme for large-scale numerical weather prediction. *Wea. Forecasting* (under review),
1113   2020.

1114 Jung, J., Souri, A.H., Wong, D., Lee, S., Jeon, W., Kim, J., Choi, Y.: The impact of the direct effect of aerosols on
1115   meteorology and air quality using aerosol optical depth simulation during the KORUS-AQ campaign, J.
1116   Geophys. Res. Atm., 124, 8303-8319, https://doi.org/10.1029/2019JD030641, 2019.

1117 Kochanski, A.K., Jenkins, M.A., Yedinak, K., Mandel, J., Beezley, J., and Lamb, B., Toward an integrated system
1118   for fire, smoke and air-quality simulations. Int. J. Wildland Fire, 25, 534-568, 2016.

31

Kochanski, A.K., Malia, D.V., Fearon, M.G., Mandel, J., Souri, A.H., and Brown, T., Modelling wildfire smoke feedback mechanisms using a coupled fire-atmosphere model with a radiatively active aerosol scheme, J. Geophys. Res. Atm., 124, 9099-9116, 2019.

Krause, P, Boyle, D.P., Base, F., Comparison of different efficiency criteria for hydrological model assessment, Adv., Geosci., 5, 89-97, 2005.

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.

Linn, R., Reisner, J., Colman, J.J. and Winterkamp, J.: Studying wildfire behavior using FIRETEC. International journal of wildland fire, 11(4), pp.233-246, 2002.

Lu, Z., and Sokolik, I.N.: Examining the impact of smoke on frontal clouds and precipitation during the 2002 Yakutsk wildfires using the WRF-Chem-SMOKE model and satellite data, J. Geophys, Res. Atm., 122, 12765-12785, https://doi.org/10.1002/2017JD027001, 2017.

Luig, A., Bofinger, S., Beyer, H.G.: Analysis of confidence intervals for the prediction of regional wind power output, Proc. European Wind Energy Conference, Copenhagen, pp 725-728, 2001. Available at https://www.researchgate.net/publication/228793040_Analysis_of_confidence_intervals_for_the_prediction_of_regional_wind_power_output, last accessed August 23, 2020.

Makar, P.A., Gong, W., Milbrandt, J., Hogrefe, C., Zhang, Y., Curci, G., Zabkar, R., Im, U., Balzarini, A., Baro, R., Bianconi, R., Cheung, P., Forkel, R., Gravel, S., Hirtl, H., Honzak, L., Hou, A., Jimenez-Guerrero, P., Langer, M., Moran, M.D., Pabla, B., Perez, J.L., Pirovano, G., San Jose, R., Tuccella, P., Werhahn, J., Zhang, J., Galmarini, S.: Feedbacks between air pollution and weather, part 1: Effects on weather. Atmospheric Environment, 115, 442-469, 2015a.

Makar, P.A., Gong, W., Hogrefe, C., Zhang, Y., Curci, G., Zabkar, R., Milbrandt, J., Im, U., Balzarini, A., Baro, R., Bianconi, R., Cheung, P., Forkel, R., Gravel, S., Hirtl, H., Honzak, L., Hou, A., Jimenez-Guerrero, P., Langer, M., Moran, M.D., Pabla, B., Perez, J.L., Pirovano, G., San Jose, R., Tuccella, P., Werhahn, J., Zhang, J., Galmarini, S.: Feedbacks between air pollution and weather, part 2: Effects on chemistry. Atmospheric Environment, 115, 499-526, 2015b.

Makar, P.A., Moran, M.D., Schultz, M.T., Taylor, A.: Speciation of volatile organic compound emissions for regional air quality modeling of particulate matter and ozone, J. Geophys. Res., 108,D2, 4041, doi:1029/2001JD000797, 2003.

Makar, P.A., Staebler, R.M., Akingunola, A., Zhang, J., McLinden, C., Kharol, S.K., Pabla, B., Cheung, P. and Zheng, Q.: The effects of forest canopy shading and turbulence on boundary layer ozone, Nature Communications, 8, art. no. 15243, doi: 10.1038/ncomms15243, 2017.

Makar, P.A., Stroud, C., Akingunola, A., Zhang, J., Ren, S., Cheung, P., and Zheng, Q.: Vehicle induced turbulence and atmospheric pollution, Proceedings of the National Academy of Sciences of the USA (under review), 2020.

Mallet, M., Solmon, F., Roblou, L., Peers, F., Turquety, S., Waquet, F., Jethva, H., Torres, O.: Simulation of optical properties and direct and indirect radiative effects of smoke aerosols over marine stratocumulus clouds during

summer 2008 in California with the regional climate model RegCM, J. Geophys. Res. Atm., 122, 10312-10337, https://doi.org/10.1002/2107JD026905, 2017.

Mandel, J., Beezley, J.D., Kochanski, A.K., Coupled atmosphere-wildland fire modelling with WRF 3.3 and SFIRE 2011, Geosci. Model Dev., 4, 591-610, 2011.

Markowicz, K.M., Pakszys, R., Ritter, C., Zielensi, T., Udisti, R., Cappelletti, D., Mazzola, M., Shobara, M., Xian, P., Zawadzka, O., Lisok, J., Petelski, T., Makuch, P., Karasinksi, G.: Impact of North American intense fires on aerosol optical properties measured over the European Arctic in July 2015, J. Geophys. Res. Atm., 121, 14487-14512, doi:10.1002/2016JD025310, 2016.

Milbrandt, J.A. and Morrison, H.: Parameterization of Cloud Microphysics Based on the Prediction of Bulk Ice Particle Properties. Part III: Introduction of Multiple Free Categories. J. Atmos. Sci., 73, 975-995, doi: https://doi.org/10.1175/JAS-D-15-0204.1, 2016.

Milbrandt, J. A., S. Bélair, S., Faucher, M., Vallée, M., Carrera, M.L., and Glazer, A.:The pan-Canadian High Resolution (2.5 km) Deterministic Prediction System. Wea. Forecasting, 31, 1791-1816, 2016.

Milbrandt, J.A., and Yau, M.K: A multimoment bulk microphysics parameterization. Part 1: analysis of the role of the spectral shape parameter, J. Atmos. Sci., 62,3051-3064, 2005a.

Milbrandt, J.A., and Yau, M.K: A multimoment bulk microphysics parameterization. Part II: a proposed three-moment closure and scheme, J. Atmos. Sci., 62, 3065-3081, 2005b

Moran, M.D., Pavlovic, R., and Anselmo, D.: Regional air quality deterministic prediction system (RAQDPS): update from version 019 to version 020, Montreal, available at: http://collaboration.cmc.ec.gc.ca/cmc/CMOI/product_guide/docs/tech_note/technote_raqdps-v2.0_20180918_e.pdf, last access: 1 December 2018.

Morrison, H., and Milbrandt, J.A.: Parameterization of Cloud Microphysics Based on the Prediction of Bulk Ice Particle Properties. Part I: Scheme Description and Idealized Tests. J. Atmos. Sci., 72, 287–311, doi: https://doi.org/10.1175/JAS-D-14-0065.1, 2015.

Morrison, H. and W. W. Grabowski, W.W.: Modeling supersaturation and subgrid-scale mixing with two-moment warm bulk microphysics. J. Atmos. Sci., 65, 792–812, doi: https://doi.org/10.1175/2007JAS2374.1, 2008.

Myhre, G., Shindell, D., Bréon, F.-M., Collins, W., Fuglestvedt, J., Huang, J., Koch, D., Lamarque, J.-F., Lee, D. , Mendoza, B. , Nakajima, T., Robock, A., Stephens, G., Takemura, T. and Zhang, H.: Anthropogenic and Natural Radiative Forcing. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds)). Cambridge University Press, Cambridge, United Kingdom and New York, USA, 2013.

Nazarenko, L., Rind, D., Tsigaridis, K, Del Genio A.D, Kelly, M., and Tausnev, N.: Interactive nature of climate change and aerosol forcing, J. Geophys. Res. Atm., 122, 3457-3480, doi:10.1002/2016JD025809, 2017.

Oeropoulis, L., Cho, N., and Lee, D.: A global survey of apparent aerosol-cloud interaction signals, J. Geophys. Res. Atm., 125, e2019JD031287, https://doi.org/10.1029/2019JD031287, 21pp, 2020.

1191   O'Neill, N.T., Eck, T.G., Holben, B.N., Smirnov, A., Royer, A., and Li, Z.: Optical properties of boreal forest fire
1192   smoke derived from Sun photometry, J. Goephys. Res., 107, 4125, doi 10.1029/2001JD000877, 21 pp., 2002

1193   Palacios-Pena, L., Baro, R., Baklanov, A., Balzarini, A., Brunner, D., Forkel, R., Hirtl, M., Honzak, L., Lopez-
1194   Romero, J.M., Motavez, J.P., Perez, J.L., Pirovano, G., San Jose, R., Schroder, W., Werhahn, J., Wolke, R.,
1195   Zabkar, R., and Jimenez-Guerrero, P.: An assessment of aerosol optical properties from remote-sensing
1196   observations and regional chemistry-climate coupled models over Europe, Atm. Chem. Phys., 18, 5021-5043,
1197   2018.

1198   Pan, L., Kim, H. C., Lee, P., Saylor, R., Tang, Y., Tong, D., Baker, B., Kondragunta, S., Xu, C., Ruminski, M. G.,
1199   Chen, W., Mcqueen, J., and Stajner, I.: Evaluating a fire smoke simulation algorithm in the National Air Quality
1200   Forecast Capability (NAQFC) by using multiple observation data sets during the Southeast Nexus (SENEX)
1201   field campaign, Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-207, 2017.

1202   Pavlovic, R., Chen, J., Anderson, K., Moran, M.D., Beaulieu, P.A., Davignon, D. and Cousineau, S.: The FireWork
1203   air quality forecast system with near-real-time biomass burning emissions: Recent developments and evaluation
1204   of performance for the 2015 North American wildfire season. Journal of the Air & Waste Management
1205   Association, 66(9), 819-841, 2016.

1206   Peace, M., Mattner, T., Mills, G., Kepert, J., and McCaw, L., Fire-modified meteorology in a coupled fire-
1207   atmosphere model, J. App. Met. and Clim., 54, 704-720, 2015.

1208   Penner, J.E, Zhou, C., Garnier, A., Mitchell, D.L.: Anthropogenic aerosol indirect effects in cirrus clouds, J.
1209   Geophys. Res. Atm., 123, 11652-11677, https://doi.org/10.1029/2018JD029204, 2018.

1210   Pinson, P., and Kariniotakis, G.: On-line adaptation of confidence intervals based on weather stability for wind power
1211   forecasting, in Proc. Of the Global Wind Power Conference, Chicago, USA, 2004. Available at https://hal-
1212   mines-paristech.archives-ouvertes.fr/hal-00529488/document, last accessed August 23, 2020.

1213   Pinto, J. O., Grim, J.A., Steinter, M.: Assessment of the High-Resolution Rapid Refresh model's ability to predict
1214   mesoscale convective systems using object-based evaluation. J. Atmos. Sci., 30, 892-913, 2015.

1215   Rothenberg, D., Avramov, A., Wang, C.: On the representation of aerosol activation and its influence on model-
1216   derived estimates of the aerosol indirect effect, Atm. Chem. Phys., 18, 7961-7983, 2018.

1217   Sandu, A. and Sander, R.: Technical note: Simulating chemical systems in Fortran90 and Matlab with the Kinetic
1218   PreProcessor KPP-2.1, Atmos. Chem. Phys., 6, 187–195, https://doi.org/10.5194/acp-6-187-2006, 2006

1219   Saponaro, G., Kolmonen, P., Sogacheva, L., Rodriguez, E., Virtanen, T., and de Leeuw, G.: Estimates of the aerosol
1220   indirect effect over the Baltic Sea region derived from 12 years of MODIS observations, Atm. Chem. Phys., 17,
1221   3133-3143, 2017.

1222   Seity, Y., Brousseau, P.,Malardel, S., Hello, G., Bénard, P., Bouttier, F., Lac, C., and Masson, V.: The AROME-
1223   France convective-scale operational model. Mon. Wea. Rev., 139, 876-913, 2010.

1224   Stockwell, W.R. and Lurmann, F.W.: Intercomparison of the ADOM and RADM gas-phase chemical mechanisms,
1225   Electric Power Institute Topical Report, Electric Power Institute, Palo Alto, California, 323pp, 1989.

1226   Stroud, C.A., Makar, P.A., Zhang, J., Moran, M.D., Akingunola, A., Li, S.-M., Leithead, A., Hayden, K, Siu, M.:
1227   Improving air quality model predictions of organic species using measurement-derived organic gaseous and

particle emissions in a petrochemical-dominated region. *Atmospheric Chemistry and Physics*, 18 (18), pp. 13531-13545, 2018.

Sundqvist, H.: Parameterization of condensation and associated clouds in models for weather prediction and general circulation simulation. In: Schlesinger M.E. (eds) Physically-Based Modelling and Simulation of Climate and Climatic Change. NATO ASI Series (Series C: Mathematical and Physical Sciences), vol 243. Springer, Dordrecht, 1988.

Suzuki, K, and Takemura, T.: Perturbations to global energy budget due to absorbing and scattering aerosols, J. Geophys. Res. Atm., 124, 2194-2209, https://doi.org/10.1029/2018JD029808, 2019.

Takeishi, A., Strorelvmo, T., and Fierce, L.: Disentangling the microphysical effects of fire particles on convective clouds through a case study, J. Geophys. Res. Atm., doi:10.1029/2019JD031890, 2020.

Twomey, S.: The influence of polution on the shortwave albedo of clouds, J. Atmos. Sci, 34, 1149-1152, 1977.

Wang, J., Allen, D.J., Pickering, K.E., Li, Z., and He, H.: Impact of aerosol direct effect on East Asian air quality during the EAST-AIRE campaign, J. Geophys. Res. Atm., 121, 6534-6554, doi:10.1002/2016JD025108, 2016.

Whaley, C.H., Makar, P.A., Shephard, M.W., Zhang, L., Zhang, J., Zheng, Q., Akingunola, A., Wentworth, G.R., Murphy, J.G., Kharol, S.K., and Cady-Periera, K.E.: Contributions of natural and anthropogenic sources to ambient ammonia in the Athabasca oil sands and north-western Canada, Atm. Chem. Phys., 18, 2011-2034, 2018.

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

Yu, S., Eder, B., Dennis, R., Chu, S.-H., Schwartz, S.E., New unbiased metrics for evaluation of air quality models, Atmos. Sci. Let., 7, 26-34, 2006.

Zhang, L., Brook, J.R., and Vet., R.: A revise parameterization for gaseous dry deposition in air-quality models, Atm. Chem. Phys., 3, 2067-2082, https://doi.org/10.5194/acp-3-2067-2003, 2003.

Zhao, X., Liu, Y., Yu, F., and Heidinger, A.K.:Using long-term satellite observations to identify sensitive regimes and active regions of aerosol indirect effects for liquid clouds over global oceans, J. Geophys. Res. Atm., 123, 457-472, https://doi.org/10.1002/2017JD027187, 2018.

Zhu, J., and Penner, J.E.: Indirect effects of secondary organic aerosol on cirrus clouds, J. Geophys., Res. Atm., 125, e2019JD032233, https://doi.org/10.1029/2019JD032233, 19 pp., 2020.

1258 **Tables:**

1259

Model Process or Configuration Component	Description	Reference (where applicable)
Base weather forecast model	Global Environmental Multiscale (GEM), v4.9.8	Cote *et al.* (1998), Girard *et al.* (2014)
Base air-quality model	Global Environmental Multiscale – Modelling Air-quality and Chemistry (GEM-MACH) v2	Moran *et al.* (2018)
Aerosol Direct Effect	Feedback simulations: GEM-MACH's predicted aerosol loading and Mie scattering using a binary water-dry aerosol homogeneous mixture assumption, at 4 wavelengths employed by GEM's radiative transfer algorithms, and at additional wavelengths for diagnostic purposes. No-Feedback simulations: invariant climatological values for aerosol optical properties are used.	Makar *et al.* (2015a,b)
Aerosol Indirect Effect	Feedback simulations: Modified P3 cloud microphysics scheme, driven by an aerosol size and speciation specific nucleation scheme (Abdul-Razzak and Ghan, 2002). No-feedback implementation: P3 scheme driven by an invariant aerosol population of a single lognormal size distribution (with a geometric mean diameter of 100 nm and total aerosol number of 300 cm^{-3} consisting of pure ammonium sulphate). The prognostic cloud droplet number and mass mixing ratios from the P3 microphysics are then transferred back to the chemistry module for using in cloud processing of gases and aerosols (cloud scavenging and chemistry) calculations, completing the AIE feedback process loop in the case of the feedback implementation (Gong *et al.*, 2015).	Gong *et al.* (2015), Abdul-Razzak and Ghan (2002), Morrison and Milbrandt (2015), Milbrandt and Morrison (2016), Morrison and Grabowski (2008).
Forest fire plume rise	CFFEPSv4.0 (see text)	

Gas-phase chemistry mechanism	ADOMII mechanism, 42 gas species.	Stockwell *et al.* (1989)
Gas-Phase chemistry solver	KPP-generated RODAS3 solver	Sandu and Sander (2006)
Cloud processing of aerosols	Aqueous chemistry, scavenging of gases and aerosols, below-cloud removal and wet deposition.	Gong *et al.* (2015)
Particle microphysics	Sectional size distribution and 8 chemical species.	Gong *et al.* (2003)
Particle inorganic thermodynamics	Local equilibrium subdomain approach	Makar *et al.* (2003)
Secondary organic aerosol formation	Modified yield approach	Stroud *et al.* (2018)
Vertical diffusion	Fully implicit approach, with surface fluxes as a boundary condition	
Advection	Semi-Lagrangian approach, 3-shell mass conservation correction (ILMC approach)	
Forest canopy shading and turbulence.	Light attenuation within forest canopies and turbulence reductions due to vegetation applied to thermal coefficients of diffusivity.	Makar *et al.* (2017)
Anthropogenic plume rise	Parameterization calculating residual buoyancy of the rising plume.	Akingunola *et al.* (2018).
Meteorological modulation of aerosol crustal material	Aerosol crustal material is inhibited when the soil water content is > 10%.	
Ammonia emissions and deposition	Bi-directional flux parameterization employed.	Whaley *et al.* (2018), Zhang *et al.* (2003).
Methane treatment	Reactive, emitted and transported tracer	
Leaf Area Index data	MODIS retrievals used to create monthly LAI values for biogenic emissions, forest canopy shading and turbulence, deposition	

Vehicle-induced turbulence	Observation-based parameterization used to modify near-surface coefficients of thermal diffusivity	Makar *et al.* (2020)

1260

1261
1262

Event	Event Observed	
Forecast	Yes	No
Yes	A	B
No	C	D

1263

1264
1265
1266
1267

Chemical	Region	Simulation	FO2	MB	MGE	NMGE	R	RMSE	COE	IOA
PM2.5	Western Canada	No Feedback	0.412451	4.8050.777	6.6884.335	1.3220.878	**0.259278**	10.1637.642	1.4760.669	0.192165
		Feedback	**0.414453**	4.5780.236	**6.5314.215**	**1.2910.829**	0.238219	**9.8036.976**	**1.4180.534**	**0.173233**
	Western USA	No Feedback	0.556536	1.953639	5.3493.793	0.82360.5	**0.254358**	8.5716.061	0.538166	0.231417
		Feedback	0.556524	**1.805786**	**5.2873.773**	**0.81360.2**	0.252361	**8.4435.978**	**0.520162**	**0.240419**
O₃	Western Canada	No Feedback	0.741773	5.988-3.683	11.0897.76	0.495347	0.527634	15.4459.996	0.223138	0.388569
		Feedback	**0.745778**	**5.891-3.553**	**10.9697.693**	**0.490344**	0.527635	**15.2689.854**	**0.210145**	**0.395573**
	Western USA	No Feedback	0.865879	**1.731-3.584**	10.7029.667	0.285257	0.693763	14.27912.534	0.249319	0.62566
		Feedback	**0.866881**	1.770-3.456	**10.6639.607**	**0.284256**	**0.694763**	**14.22512.458**	**0.252322**	**0.626661**
NO₂	Western Canada	No Feedback	0.437528	**0.997417**	**2.75729.**	0.594	0.564519	3.965353	0.154053	0.577527
		Feedback	**0.42952**	1.0370-.533	2.758274	0.595591	**0.565513**	**3.936314**	0.154055	0.577527

39

| | Western USA | No Feedback | **0.493428** | **0.346669** | 2.341278 | 0.572759 | **0.653585** | 3.674917 | 0.177161 | 0.588419 |
| | | Feedback | 0.483427 | 0.427578 | **2.332224** | **0.570746** | 0.651581 | **3.657858** | **0.180441** | **0.590429** |

Table 3: Summary performance metrics for ozone, nitrogen dioxide, and PM2.5. 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. NMGE: Normalized Mean Gross Error. R: Correlation Coefficient. RMSE: Root Mean Square Error. COE: Coefficient of Error. IOA: Index of Agreement.

Figures:

41

Figure 1: GEM-MACH domains: (a) ~~Operational~~ **GEM** meteorology **10km resolution forecast domain. (b)** GEM-MACH 10km resolution~~Experimental 2.5 km grid cell size~~ **forecast domain.** (c) GEM-MACH inner 2.5-km grid resolution forecast domain for comparison to observations. ~~used here.~~ **Red lines indicate locations of illustrative South to North and West to East cross-sections appearing in subsequent analysis in the text.**

Time

(a) GEM archived 10km forecasts, North American domain, created using meteorological analyses and original microphysics, Day 1 to Day 3.

12 UT, Day 1   6 UT, Day 2   12 UT, Day 2   12 UT, Day 3

Meteorological Initial and boundary conditions from archived North American 10km P3 simulations. Chemical initial and boundary conditions from ECMWF Reanalysis.

(b) GEM-MACH 30 hour, 10km no-feedback forecasts, Western North American domain, using P3 microphysics.

6 UT, Day 2   12 UT, Day 2   6 UT, Day 3   12 UT, Day 3

Meteorological Initial and boundary conditions from archived North American 10km P3 simulations

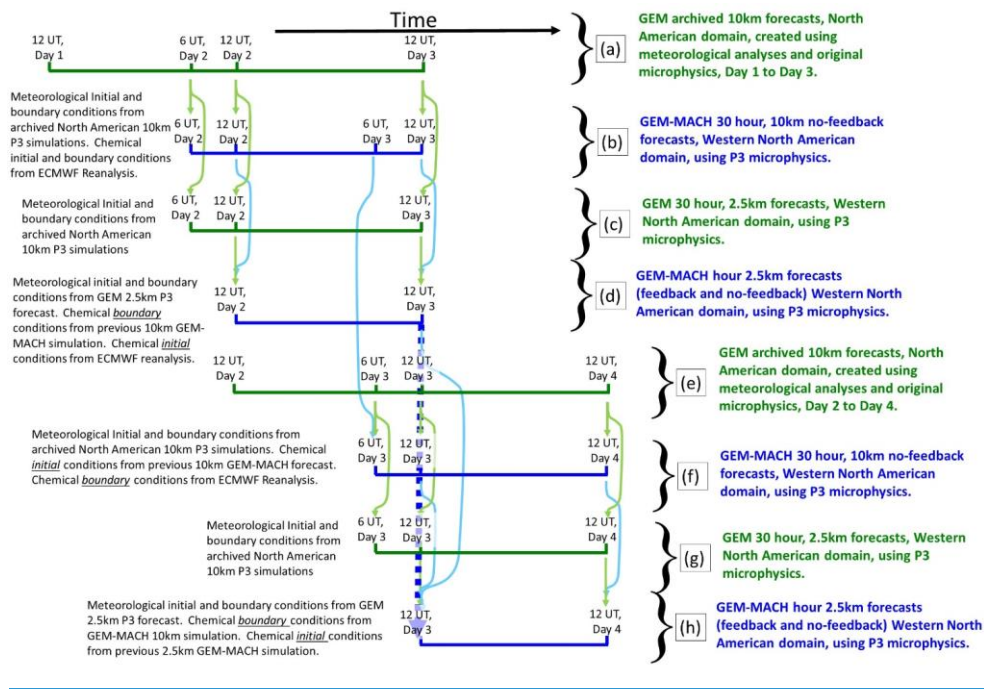(c) GEM 30 hour, 2.5km forecasts, Western North American domain, using P3 microphysics.

6 UT, Day 2   12 UT, Day 2   12 UT, Day 3

Meteorological initial and boundary conditions from GEM 2.5km P3 forecast. Chemical boundary conditions from previous 10km GEM-MACH simulation. Chemical initial conditions from ECMWF reanalysis.

(d) GEM-MACH hour 2.5km forecasts (feedback and no-feedback) Western North American domain, using P3 microphysics.

12 UT, Day 2   12 UT, Day 3

(e) GEM archived 10km forecasts, North American domain, created using meteorological analyses and original microphysics, Day 2 to Day 4.

12 UT, Day 2   6 UT, Day 3   12 UT, Day 3   12 UT, Day 4

Meteorological Initial and boundary conditions from archived North American 10km P3 simulations. Chemical initial conditions from previous 10km GEM-MACH forecast. Chemical boundary conditions from ECMWF Reanalysis.

(f) GEM-MACH 30 hour, 10km no-feedback forecasts, Western North American domain, using P3 microphysics.

6 UT, Day 3   12 UT, Day 3   12 UT, Day 4

Meteorological Initial and boundary conditions from archived North American 10km P3 simulations

(g) GEM 30 hour, 2.5km forecasts, Western North American domain, using P3 microphysics.

6 UT, Day 3   12 UT, Day 3   12 UT, Day 4

Meteorological initial and boundary conditions from GEM 2.5km P3 forecast. Chemical boundary conditions from GEM-MACH 10km simulation. Chemical initial conditions from previous 2.5km GEM-MACH simulation.

(h) GEM-MACH hour 2.5km forecasts (feedback and no-feedback) Western North American domain, using P3 microphysics.

12 UT, Day 3   12 UT, Day 4

Time

(a) GEM archived 10km forecasts, North American domain, created using meteorological analyses and original microphysics, Day 1 to Day 3.

(b) GEM-MACH 30 hour, 10km no-feedback forecasts, Western North American domain, using P3 microphysics.

(c) GEM 30 hour, 2.5km forecasts, Western North American domain, using P3 microphysics.

(d) GEM-MACH hour 2.5km forecasts (feedback and no-feedback) Western North American domain, using P3 microphysics.

(e) GEM archived 10km forecasts, North American domain, created using meteorological analyses and original microphysics, Day 2 to Day 4.

(f) GEM-MACH 30 hour, 10km no-feedback forecasts, Western North American domain, using P3 microphysics.

(g) GEM 30 hour, 2.5km forecasts, Western North American domain, using P3 microphysics.

(h) GEM-MACH hour 2.5km forecasts (feedback and no-feedback) Western North American domain, using P3 microphysics.

Meteorological Initial and boundary conditions from archived North American 10km P3 simulations. Chemical initial and boundary conditions from ECMWF Reanalysis.

Meteorological Initial and boundary conditions from archived North American 10km P3 simulations

Meteorological initial and boundary conditions from GEM 2.5km P3 forecast. Chemical *boundary* conditions from previous 10km GEM-MACH simulation. Chemical *initial* conditions from ECMWF reanalysis.

Meteorological Initial and boundary conditions from archived North American 10km P3 simulations. Chemical *initial* conditions from previous 10km GEM-MACH forecast. Chemical *boundary* conditions from ECMWF Reanalysis.

Meteorological Initial and boundary conditions from archived North American 10km P3 simulations

Meteorological initial and boundary conditions from GEM 2.5km P3 forecast. Chemical *boundary* conditions from GEM-MACH 10km simulation. Chemical *initial* conditions from previous 2.5km GEM-MACH simulation.

Figure 2: Example time sequencing of model simulations used to generate the 2.5-km GEM-MACH simulations carried out here. Green lines and print indicate GEM (weather forecast only) simulations), blue lines and print indicate 2.5-km GEM-MACH simulations. Arrows indicate data flow (light green: meteorological information; light blue: chemical information). Steps (a) through (f) illustrate the sequence of forecasts used to generate two consecutive days of 2.5km GEM-MACH simulations. Note that on-line coupling occurs only at the 2.5km GEM-MACH forecast level, in this sequencing.

Operational Configuration (CFFEPSv2.03)

- Satellite Retrievals (hotspots)
- Meteorology: 72 hour forecast sfc, 850, 700, 500, 250 mb temperatures
- Offline Fire behaviour model
- Speciated, size resolved, vertically distributed emissions (output file)
- GEM-MACH

Experimental Configuration (CFFEPSv4.0)

- Satellite Retrievals (hotspots)
- GEM-MACH
  - GEM Dynamics
  - GEM Physics
    - GEM-MACH Chemistry
      - Online Fire behavior model, providing speciated, size-resolved, vertically distributed emissions
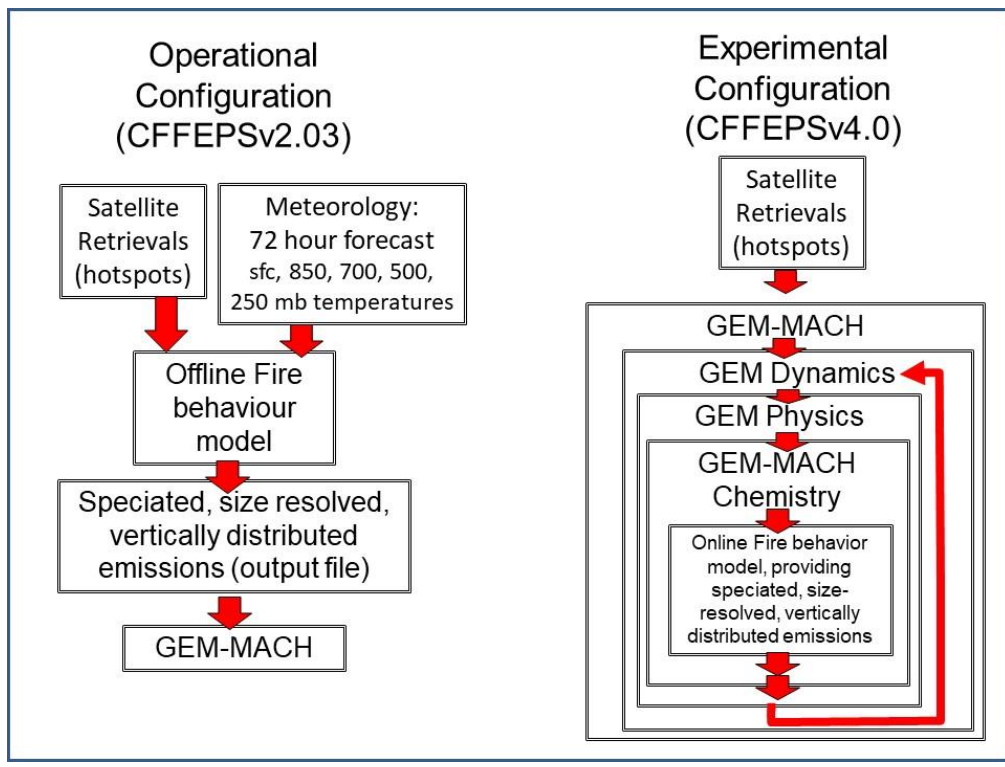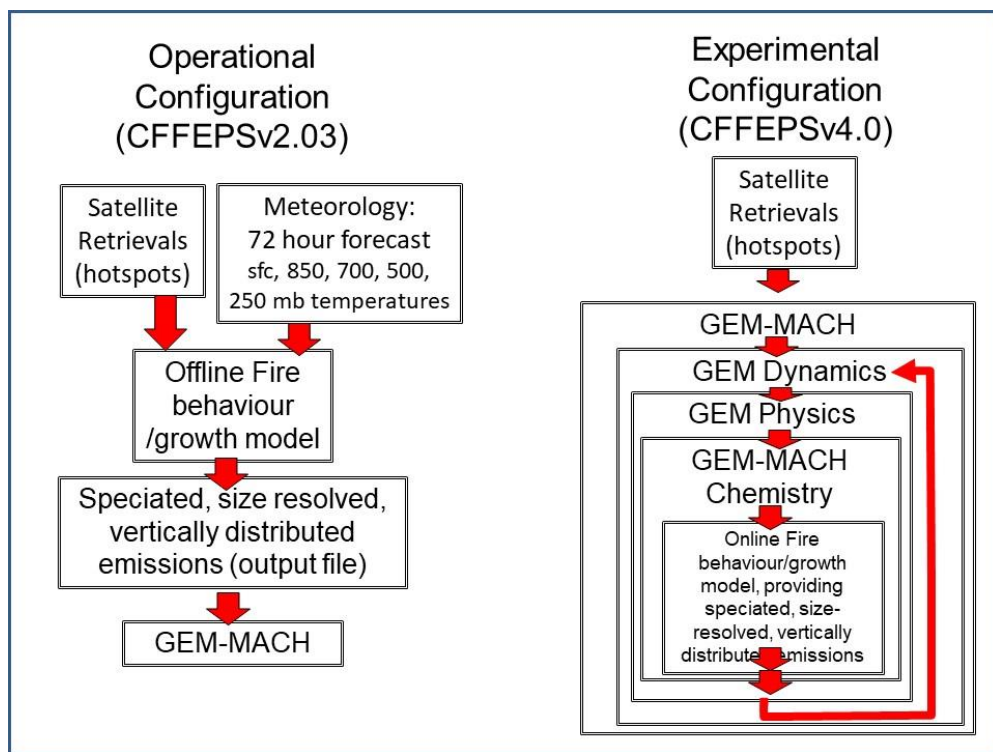
Figure 3: Process comparison between original (CFFEPSv2.03, left) and on-line (CFFEPSv4.0, right) forest fire emissions and vertical plume distribution algorithms.
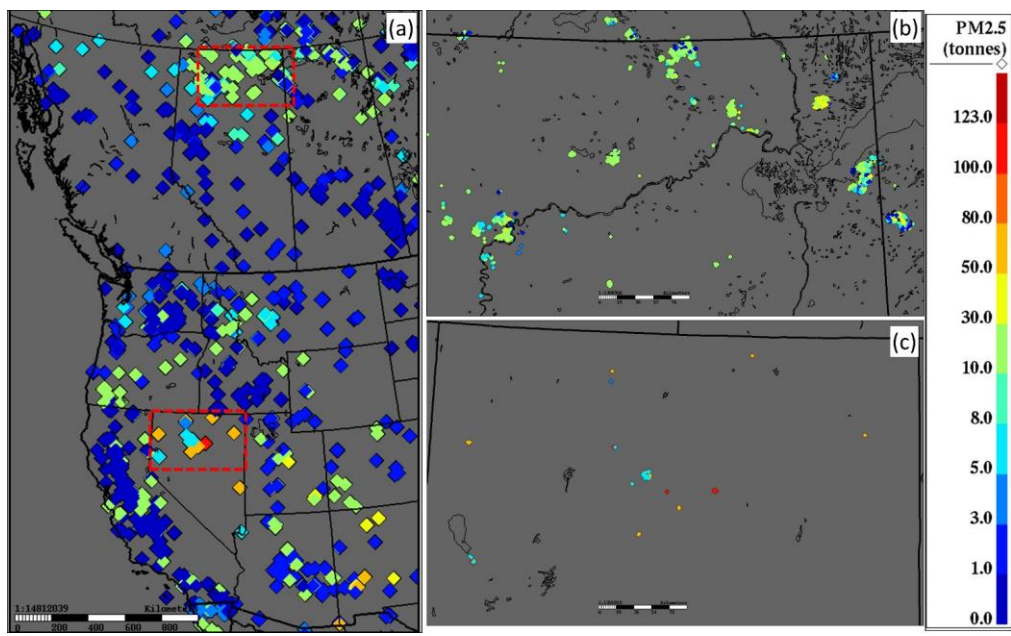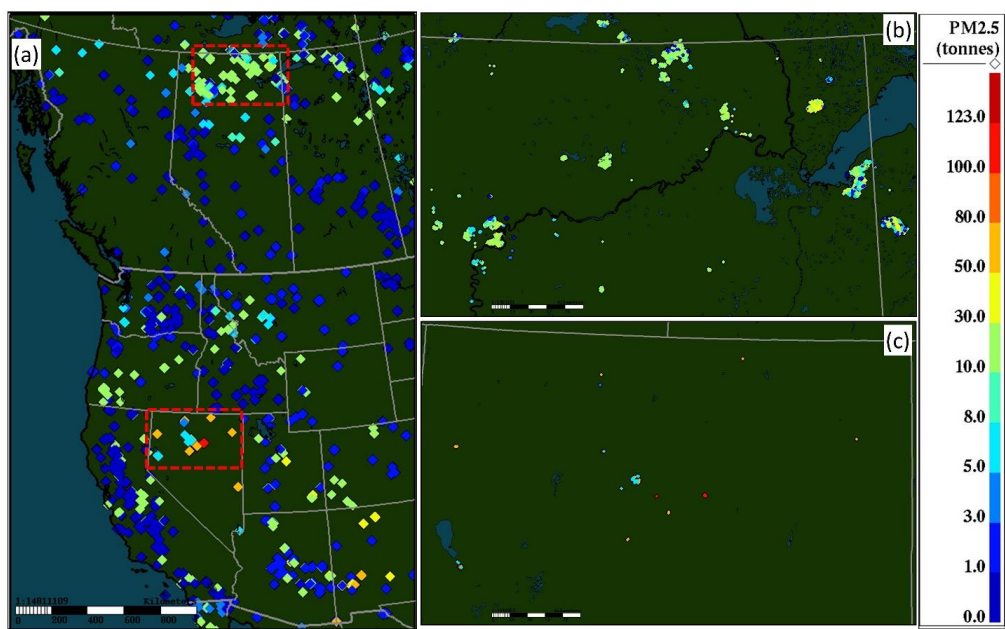
Figure 4: Hotspot locations during the study period, colour-coded by daily total tonnes PM2.5 emitted. (a) Entire model 2.5-km domain, with northern Alberta and northern Nevada sub-regions as red dashed boxes; (b) northern Alberta zoom, with smaller symbols for individual hotspots showing the large fire regions; (c) northern Nevada zoom, to the same scale as (b), showing isolated hotspots with high emissions.

$$|MB|_{feedback} - |MB|_{no-feedback}$$
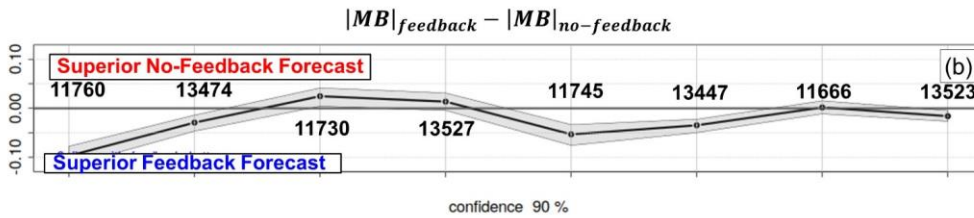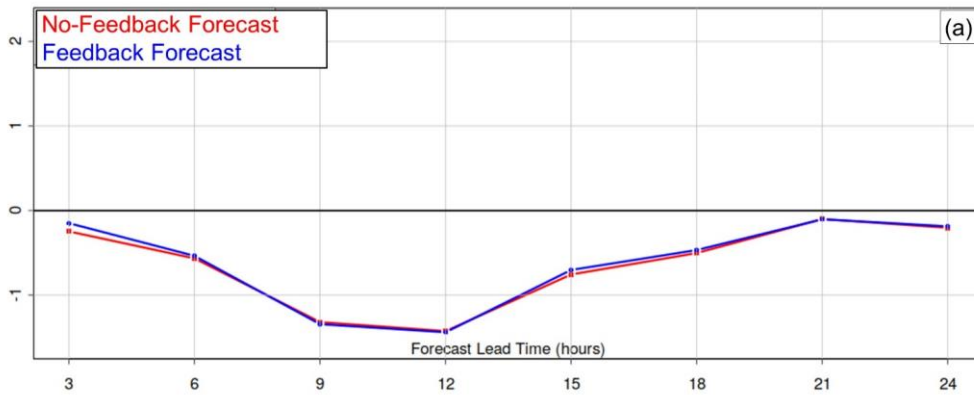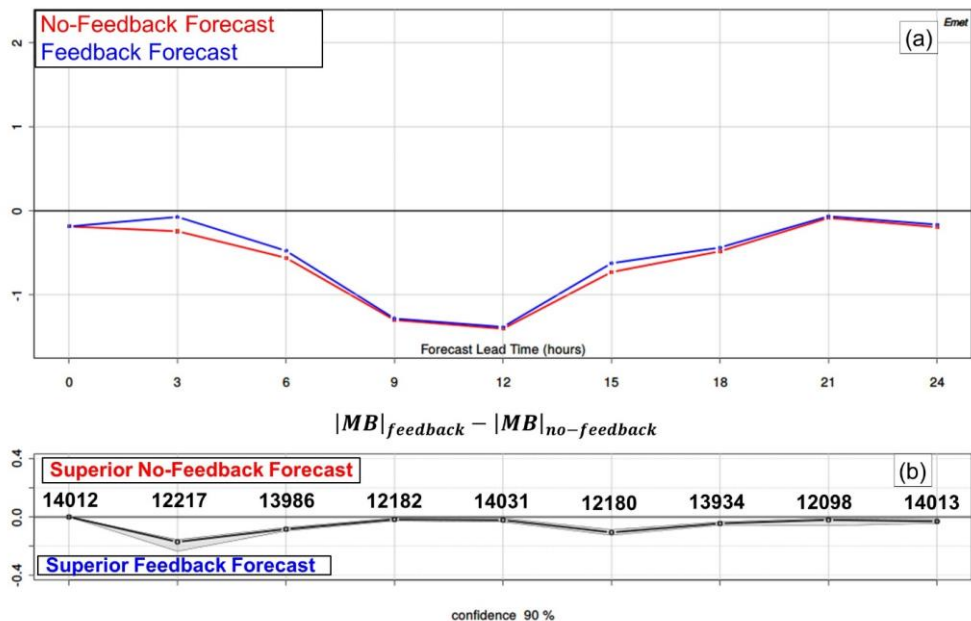
Figure 5: Mean bias in surface temperature (°C) at forecast hours starting at 0 UT. (a) Red line: no-feedback forecast values; blue line: feedback forecast values. (b) Difference in absolute value of mean bias between the two forecasts $\left(|MB|_{feedback} - |MB|_{no-feedback} \right)$, with the region *below* 90% confidence level shown shaded grey. Mean values above above/below the '0' line, and outside of the shaded region thus indicate differences in the mean between the two forecasts which differ at or above the 90% confidence level. Values of the difference which appear below/above the zero line and outside of the grey area thus indicate superior domain average performance for the feedback/no-feedback forecasts at each of the 3-hourly intervals, respectively. Numbers appearing above the metric differences are the number of observations contributing to the calculated metrics.
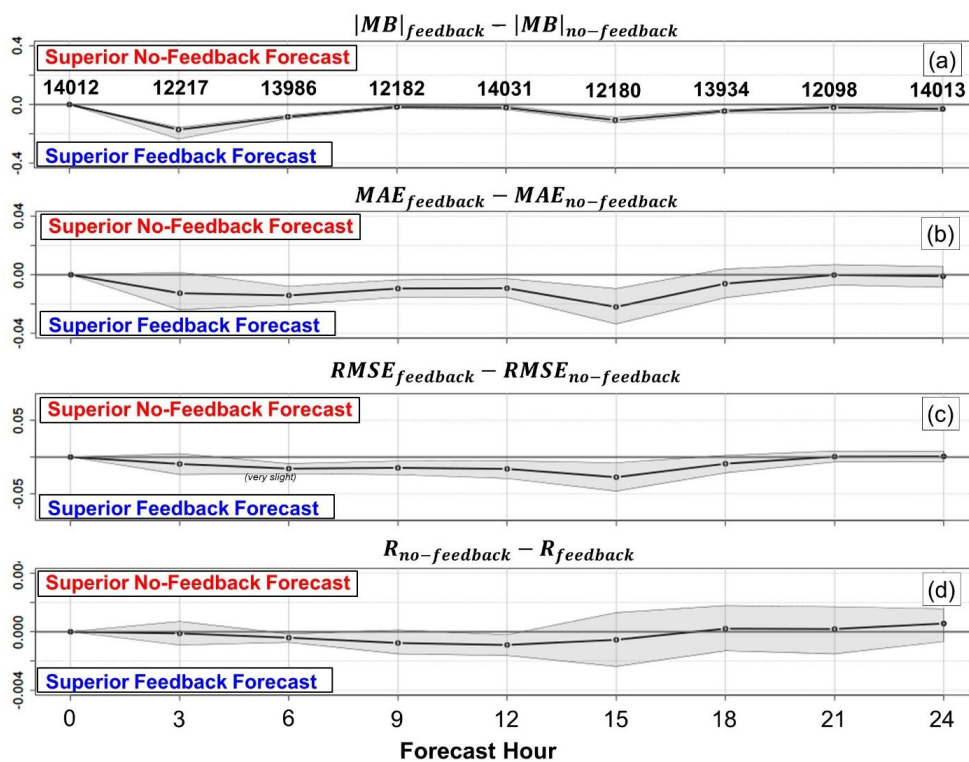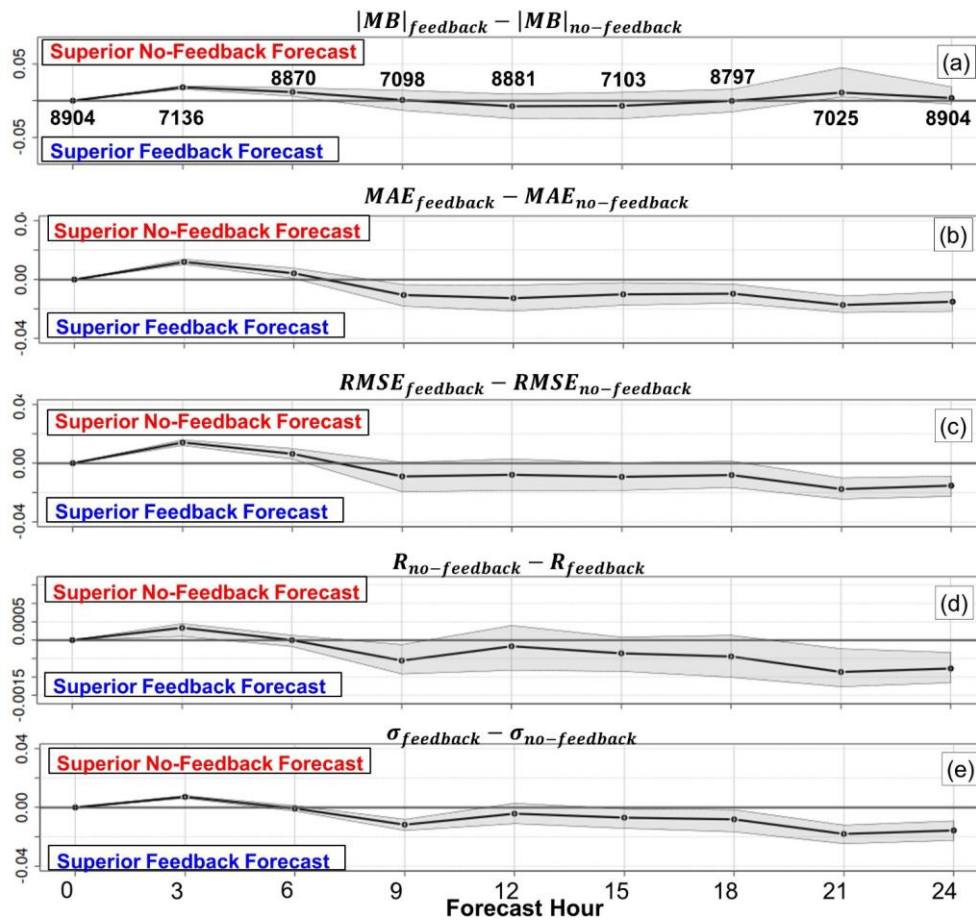
$|MB|_{feedback} - |MB|_{no-feedback}$

Superior No-Feedback Forecast

(a)

11760   13474   11745   13447   11666   13523

11730   13527

2.nirex_with_   Superior Feedback Forecast

$MAE_{feedback} - MAE_{no-feedback}$

Superior No-Feedback Forecast

(b)

Superior Feedback Forecast

$RMSE_{feedback} - RMSE_{no-feedback}$

Superior No-Feedback Forecast

(c)

(very slight)

Superior Feedback Forecast

$R_{no-feedback} - R_{feedback}$

Superior No-Feedback Forecast

(d)

Superior Feedback Forecast

3   6   9   12   15   18   21   24

Forecast Hour

Figure 6: Summary meteorological performance comparison for surface temperature (C). (a) mean bias, (b) mean absolute error, (c) root mean square error and (d) Pearson correlation coefficient. 90% confidence level shown in grey. Numbers appearing above the absolute mean bias differences are the number of stations contributing to the calculated metrics.
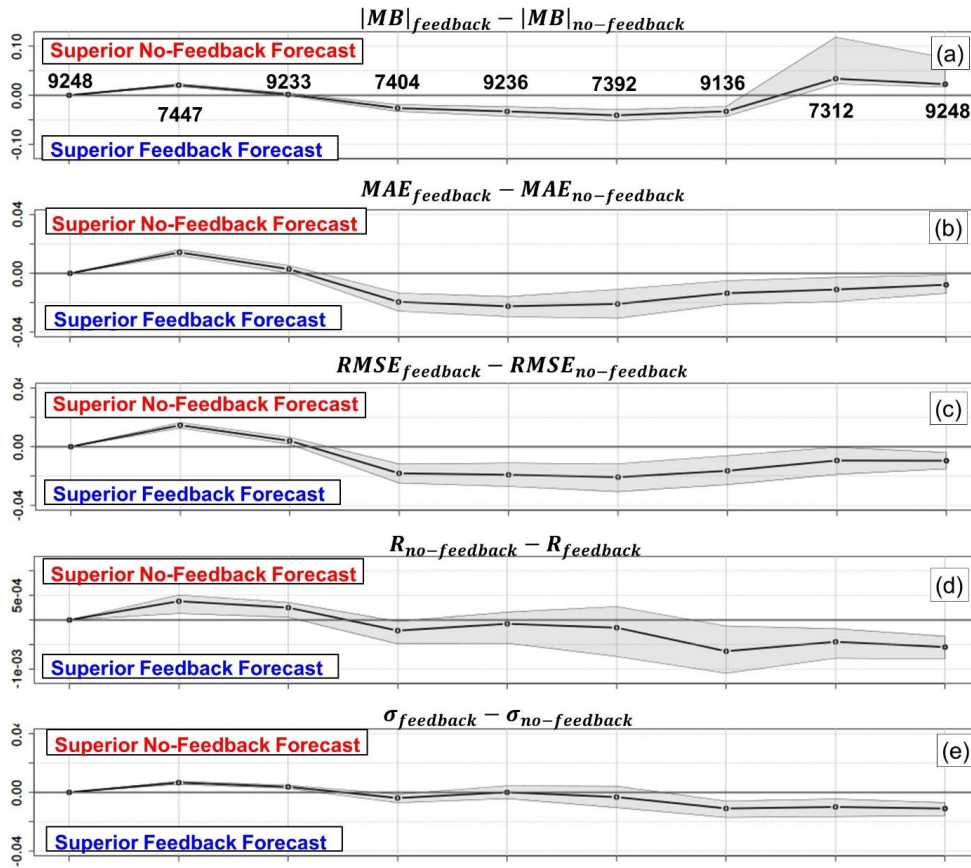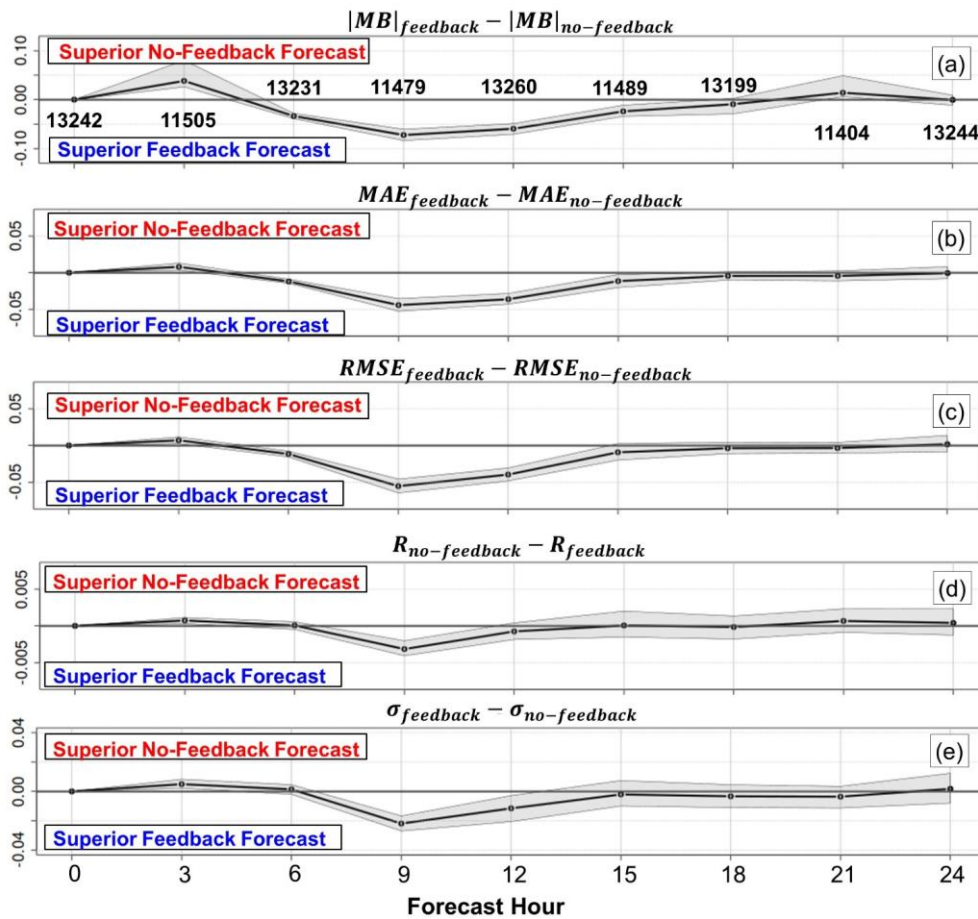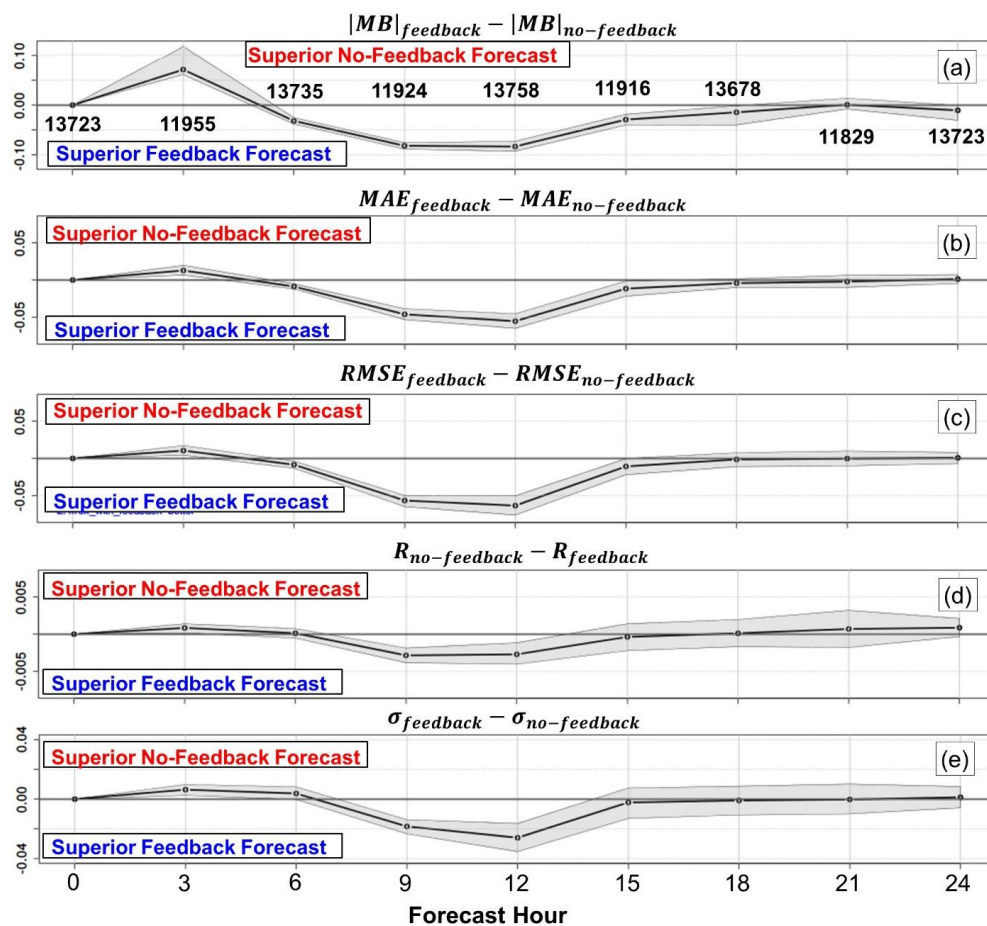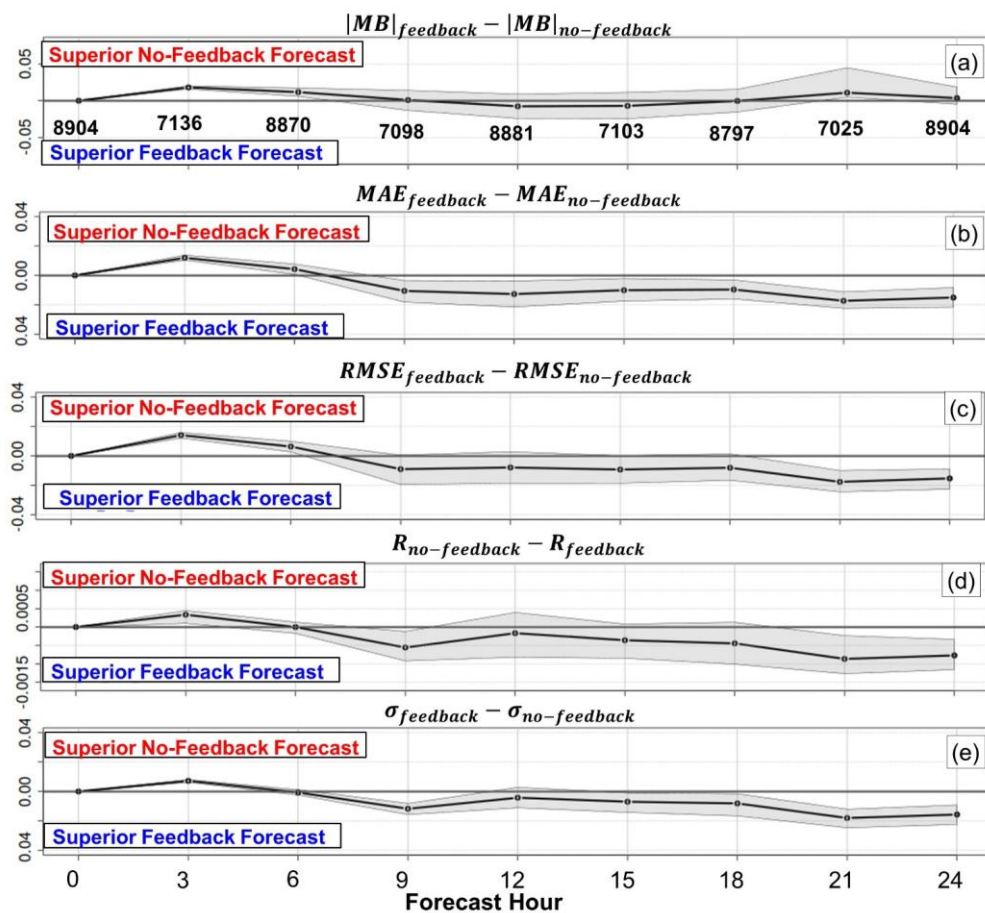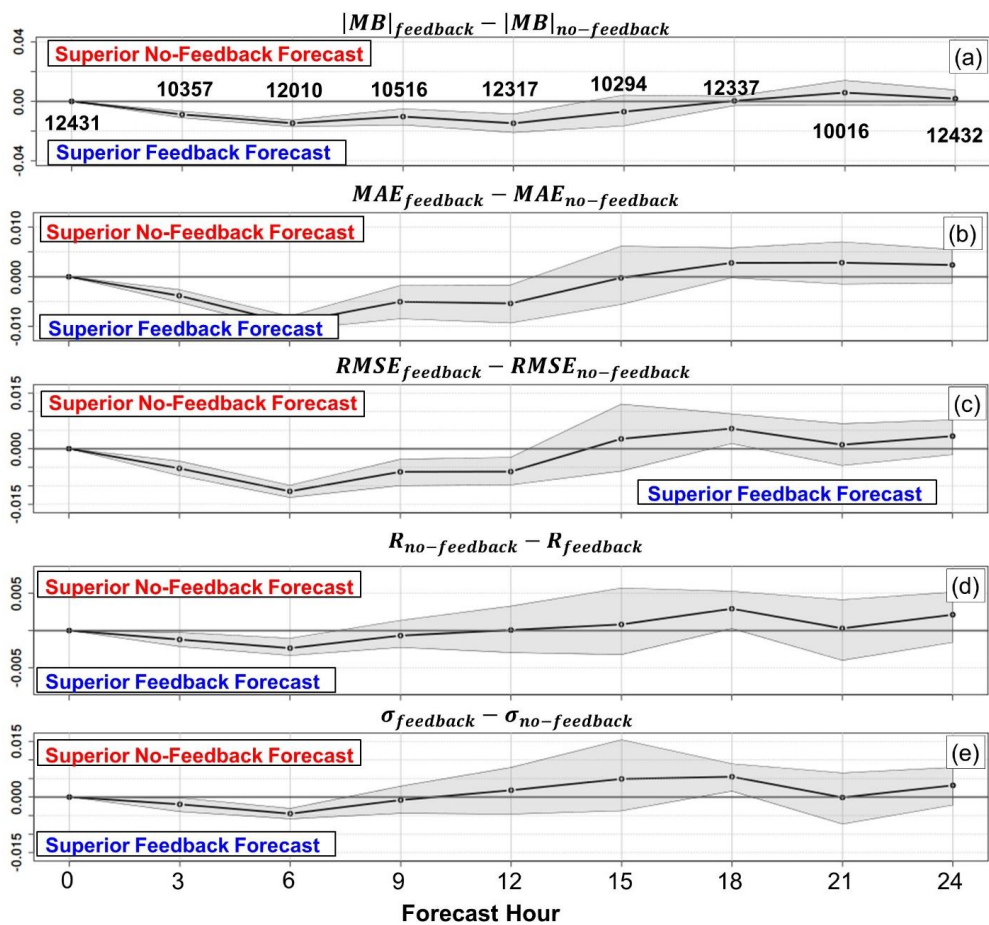
$$|MB|_{feedback} - |MB|_{no-feedback}$$

Superior No-Feedback Forecast
8870   7098   8881   7103   8797
8904   7136   7025   8904
Superior Feedback Forecast
(a)

$$MAE_{feedback} - MAE_{no-feedback}$$

Superior No-Feedback Forecast
Superior Feedback Forecast
(b)

$$RMSE_{feedback} - RMSE_{no-feedback}$$

Superior No-Feedback Forecast
Superior Feedback Forecast
(c)

$$R_{no-feedback} - R_{feedback}$$

Superior No-Feedback Forecast
Superior Feedback Forecast
(d)

$$\sigma_{feedback} - \sigma_{no-feedback}$$

Superior No-Feedback Forecast
Superior Feedback Forecast
(e)

Forecast Hour

$$|MB|_{feedback} - |MB|_{no-feedback}$$

(a)

Superior No-Feedback Forecast

9248   9233   7404   9236   7392   9136

7447   7312   9248

Superior Feedback Forecast

$$MAE_{feedback} - MAE_{no-feedback}$$

(b)

Superior No-Feedback Forecast

Superior Feedback Forecast

$$RMSE_{feedback} - RMSE_{no-feedback}$$

(c)

Superior No-Feedback Forecast

Superior Feedback Forecast

$$R_{no-feedback} - R_{feedback}$$

(d)

Superior No-Feedback Forecast

Superior Feedback Forecast

$$\sigma_{feedback} - \sigma_{no-feedback}$$

(e)

Superior No-Feedback Forecast

Superior Feedback Forecast

Figure 7: Summary meteorological performance comparison for surface pressure (hPa). (a) mean bias, (b) mean absolute error, (c) root mean square error, (d) Pearson correlation coefficient, and (e) standard deviation. 90% confidence level shown in grey. Numbers appearing above the absolute mean bias differences are the number of stations contributing to the calculated metrics.
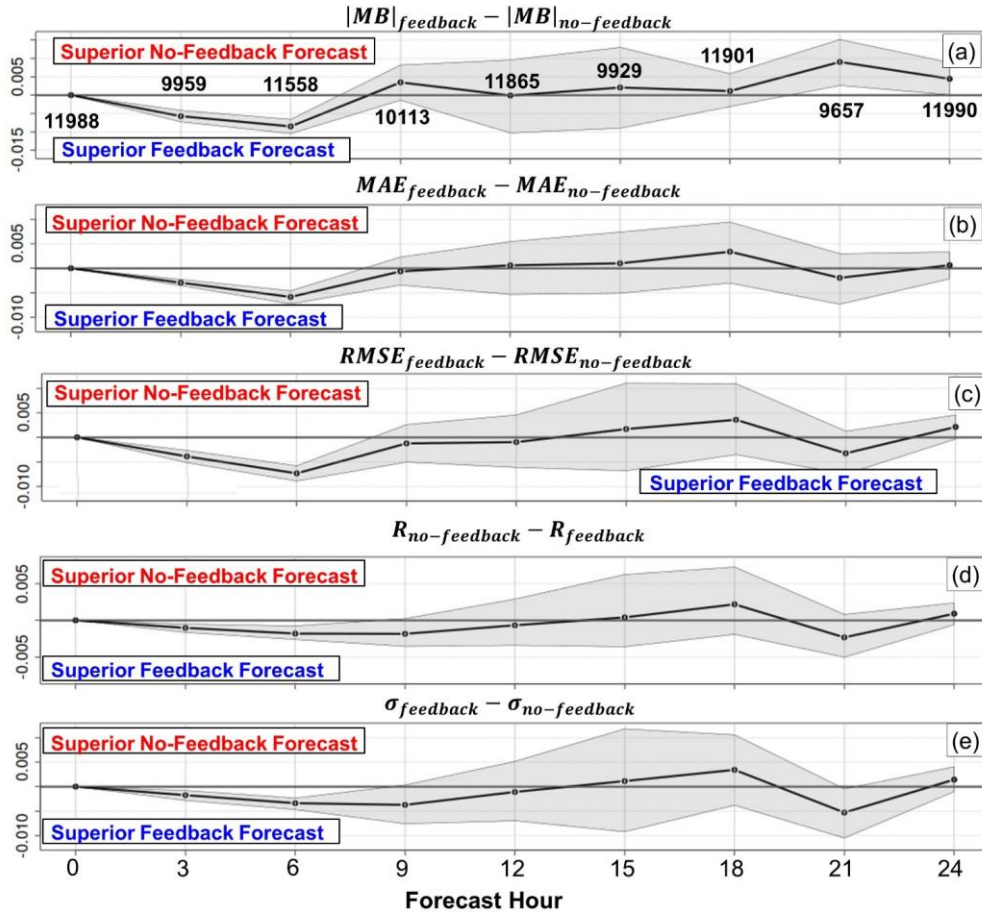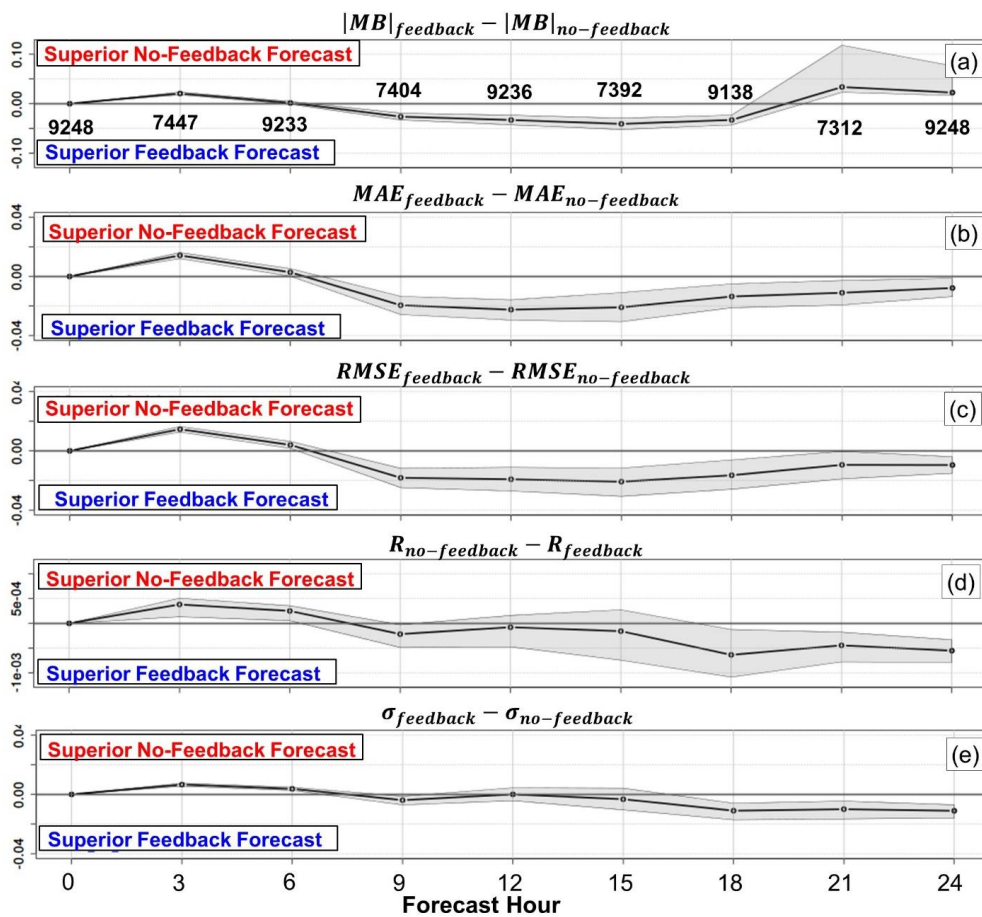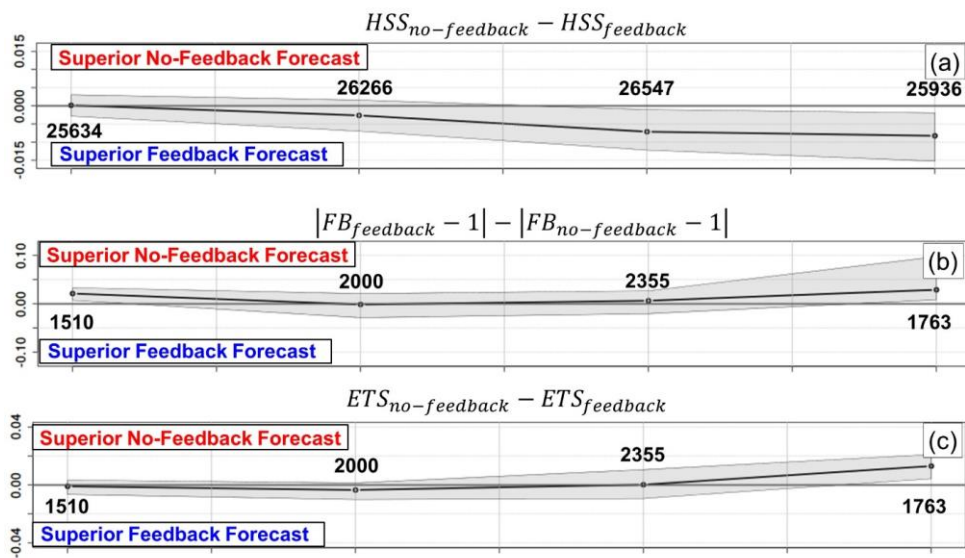
$$|MB|_{feedback} - |MB|_{no-feedback}$$

(a)

Superior No-Feedback Forecast

13231   11479   13260   11489   13199

13242   11505

11404   13244

Superior Feedback Forecast

$$MAE_{feedback} - MAE_{no-feedback}$$

(b)

Superior No-Feedback Forecast

Superior Feedback Forecast

$$RMSE_{feedback} - RMSE_{no-feedback}$$

(c)

Superior No-Feedback Forecast

Superior Feedback Forecast

$$R_{no-feedback} - R_{feedback}$$

(d)

Superior No-Feedback Forecast

Superior Feedback Forecast

$$\sigma_{feedback} - \sigma_{no-feedback}$$

(e)

Superior No-Feedback Forecast

Superior Feedback Forecast

Forecast Hour

Figure 8: Summary meteorological performance comparison for dewpoint temperature (C). (a) mean bias, (b) mean absolute error, (c) root mean square error, (d) Pearson correlation coefficient, and (e) standard deviation. 90% confidence level shown in grey. Numbers appearing above the absolute mean bias differences are the number of stations contributing to the calculated metrics.
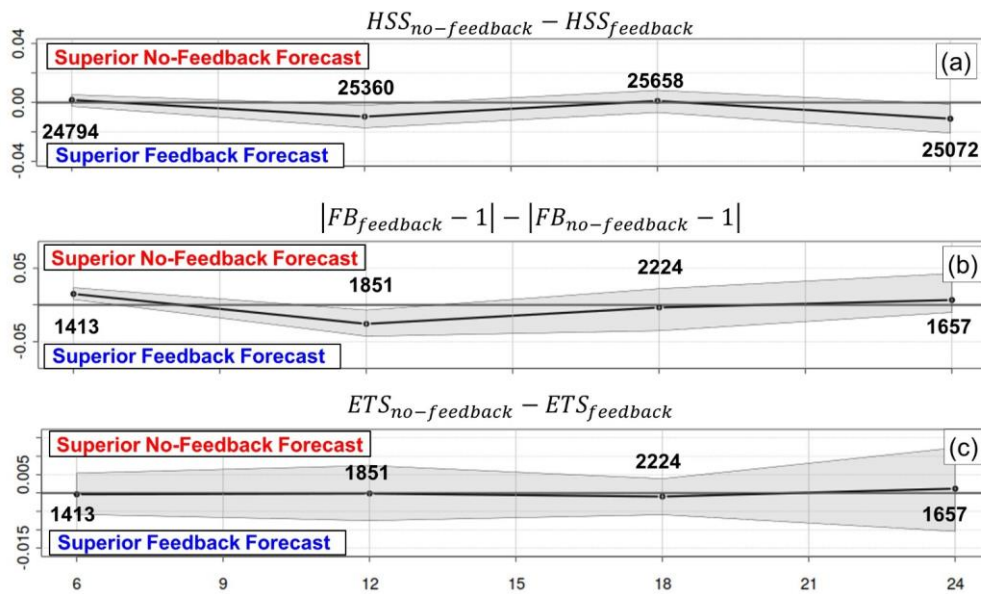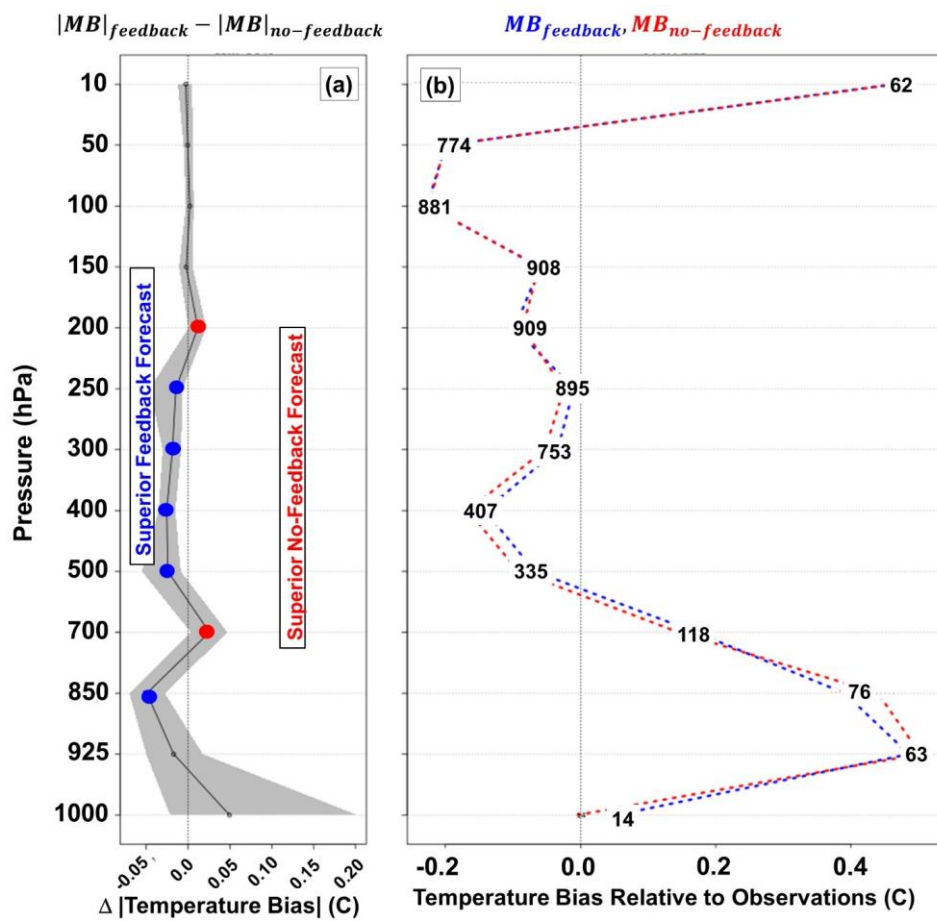
$|MB|_{feedback} - |MB|_{no-feedback}$

Superior No-Feedback Forecast (a)

8904   7136   8870   7098   8881   7103   8797   7025   8904

Superior Feedback Forecast

$MAE_{feedback} - MAE_{no-feedback}$

Superior No-Feedback Forecast (b)

Superior Feedback Forecast

$RMSE_{feedback} - RMSE_{no-feedback}$

Superior No-Feedback Forecast (c)

Superior Feedback Forecast

$R_{no-feedback} - R_{feedback}$

Superior No-Feedback Forecast (d)

Superior Feedback Forecast

$\sigma_{feedback} - \sigma_{no-feedback}$

Superior No-Feedback Forecast (e)

Superior Feedback Forecast

Forecast Hour

59

$|MB|_{feedback} - |MB|_{no-feedback}$

(a)

Superior No-Feedback Forecast

10357   12010   10516   12317   10294   12337

12431

10016   12432

Superior Feedback Forecast

$MAE_{feedback} - MAE_{no-feedback}$

(b)

Superior No-Feedback Forecast

Superior Feedback Forecast

$RMSE_{feedback} - RMSE_{no-feedback}$

(c)

Superior No-Feedback Forecast

Superior Feedback Forecast

$R_{no-feedback} - R_{feedback}$

(d)

Superior No-Feedback Forecast

Superior Feedback Forecast

$\sigma_{feedback} - \sigma_{no-feedback}$

(e)

Superior No-Feedback Forecast

Superior Feedback Forecast

Forecast Hour

: Summary meteorological performance comparison for sea-level pressure (hPa). (a) mean bias, (b) mean absolute error, (c) root mean square error, (d) Pearson correlation coefficient, and (e) standard deviation. 90% confidence level shown in grey. Numbers appearing above the absolute mean bias differences are the number of stations contributing to the calculated metrics.

$|MB|_{feedback} - |MB|_{no-feedback}$

Superior No-Feedback Forecast

9959   11558   11901
                11865   9929

11988   10113   9657   11990

Superior Feedback Forecast

(a)

$MAE_{feedback} - MAE_{no-feedback}$

Superior No-Feedback Forecast

Superior Feedback Forecast

(b)

$RMSE_{feedback} - RMSE_{no-feedback}$

Superior No-Feedback Forecast

Superior Feedback Forecast

(c)

$R_{no-feedback} - R_{feedback}$

Superior No-Feedback Forecast

Superior Feedback Forecast

(d)

$\sigma_{feedback} - \sigma_{no-feedback}$

Superior No-Feedback Forecast

Superior Feedback Forecast

(e)

Forecast Hour

Figure 109: Summary meteorological performance comparison for 10m windspeed (m s^{-1}). (a) mean bias, (b) mean absolute error, (c) root mean square error, (d) Pearson correlation coefficient, and (e) standard deviation. 90% confidence level shown in grey. Numbers appearing above the absolute mean bias differences are the number of stations contributing to the calculated metrics.

$$|MB|_{feedback} - |MB|_{no-feedback}$$

(a)

Superior No-Feedback Forecast

7404   9236   7392   9138

9248   7447   9233   7312   9248

Superior Feedback Forecast

$$MAE_{feedback} - MAE_{no-feedback}$$

(b)

Superior No-Feedback Forecast

Superior Feedback Forecast

$$RMSE_{feedback} - RMSE_{no-feedback}$$

(c)

Superior No-Feedback Forecast

Superior Feedback Forecast

$$R_{no-feedback} - R_{feedback}$$

(d)

Superior No-Feedback Forecast

Superior Feedback Forecast

$$\sigma_{feedback} - \sigma_{no-feedback}$$

(e)

Superior No-Feedback Forecast

Superior Feedback Forecast

Forecast Hour

Figure 10: Summary meteorological performance comparison for sea-level pressure (hPa). (a) mean bias, (b) mean absolute error, (c) root-mean-square error, (d) Pearson correlation coefficient, and (e) standard deviation. 90% confidence level shown in grey. Numbers appearing above the absolute mean bias differences are the number of stations contributing to the calculated metrics.

$$HSS_{no-feedback} - HSS_{feedback}$$

(a)

Superior No-Feedback Forecast

26266　26547　25936

25634

Superior Feedback Forecast

$$\left|FB_{feedback} - 1\right| - \left|FB_{no-feedback} - 1\right|$$

(b)

Superior No-Feedback Forecast

2000　2355

1510　1763

Superior Feedback Forecast

$$ETS_{no-feedback} - ETS_{feedback}$$

(c)

Superior No-Feedback Forecast

2000　2355

1510　1763

Superior Feedback Forecast

$$HSS_{no-feedback} - HSS_{feedback}$$

$$\left|FB_{feedback} - 1\right| - \left|FB_{no-feedback} - 1\right|$$

$$ETS_{no-feedback} - ETS_{feedback}$$

Figure 11: Precipitation performance evaluation (mm precipitation). (a) Heike skill score of 6-hour accumulated precipitation (No-Feedback − Feedback). (b) Frequency bias index of 6-hour accumulated precipitation (threshold of 2 mm, No-Feedback − Feedback). (c) Equitable Threat Score of 6-hour accumulated precipitation (threshold of 2 mm, No-Feedback − Feedback).

$|MB|_{feedback} - |MB|_{no-feedback}$

$MB_{feedback}, MB_{no-feedback}$

(a) Superior Feedback Forecast / Superior No-Feedback Forecast

(b)

Pressure (hPa)

Δ |Temperature Bias| (C)

Temperature Bias Relative to Observations (C)
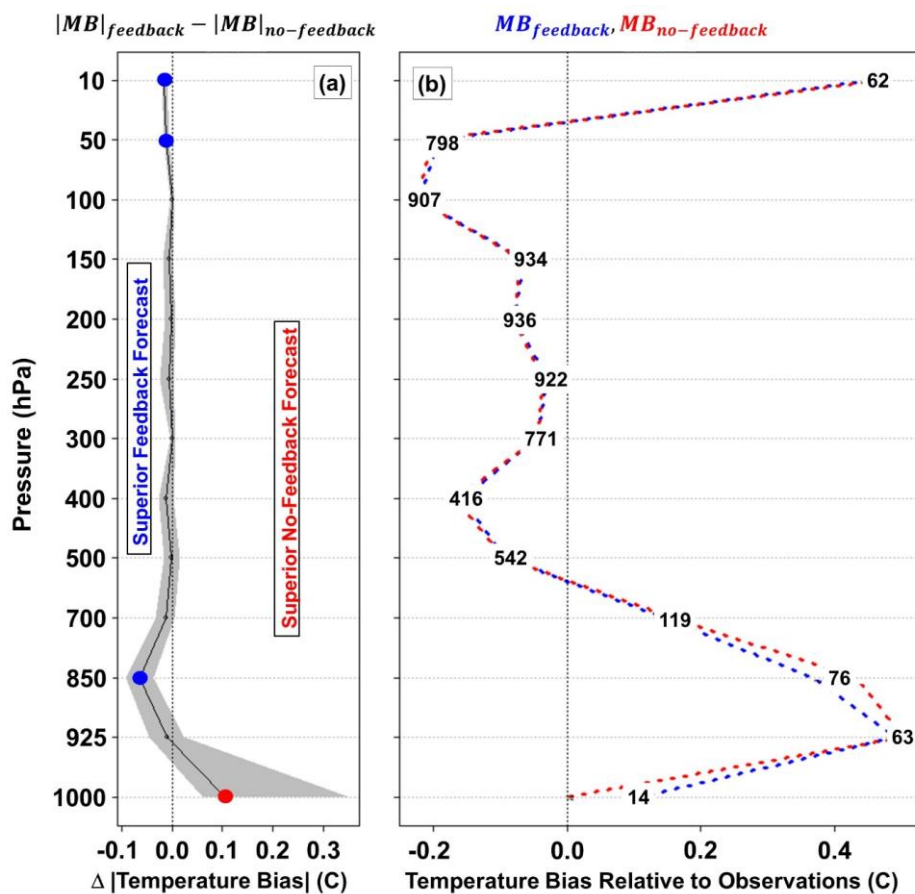
Figure 12: Forecast hour 12 (0 UT) summary upper air temperature performance comparison for air temperature (mean bias, C). (a) Difference in absolute value of mean bias in temperature, (feedback forecast – no-feedback forecast). Grey regions represent 90% confidence levels, blue symbols: pressure levels at which the feedback mean bias outperforms the no-feedback mean-bias at > 90% confidence. Red symbols: pressure levels at which the no-feedback mean bias outperforms the feedback mean bias at > 90% confidence. 90% confidence level shown in grey. (b) Mean bias in upper air temperature for feedback (blue) and no-feedback (red) (C). Numbered values on the profiles indicate the number of observed data-model pairs at each pressure level.
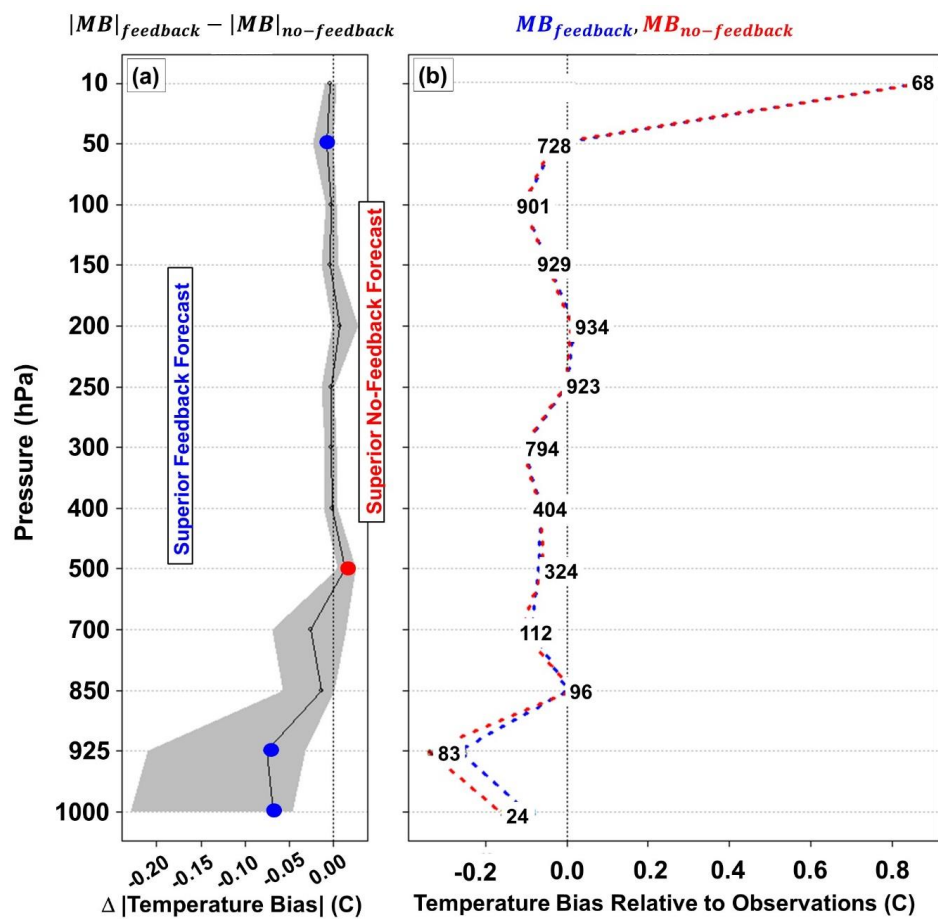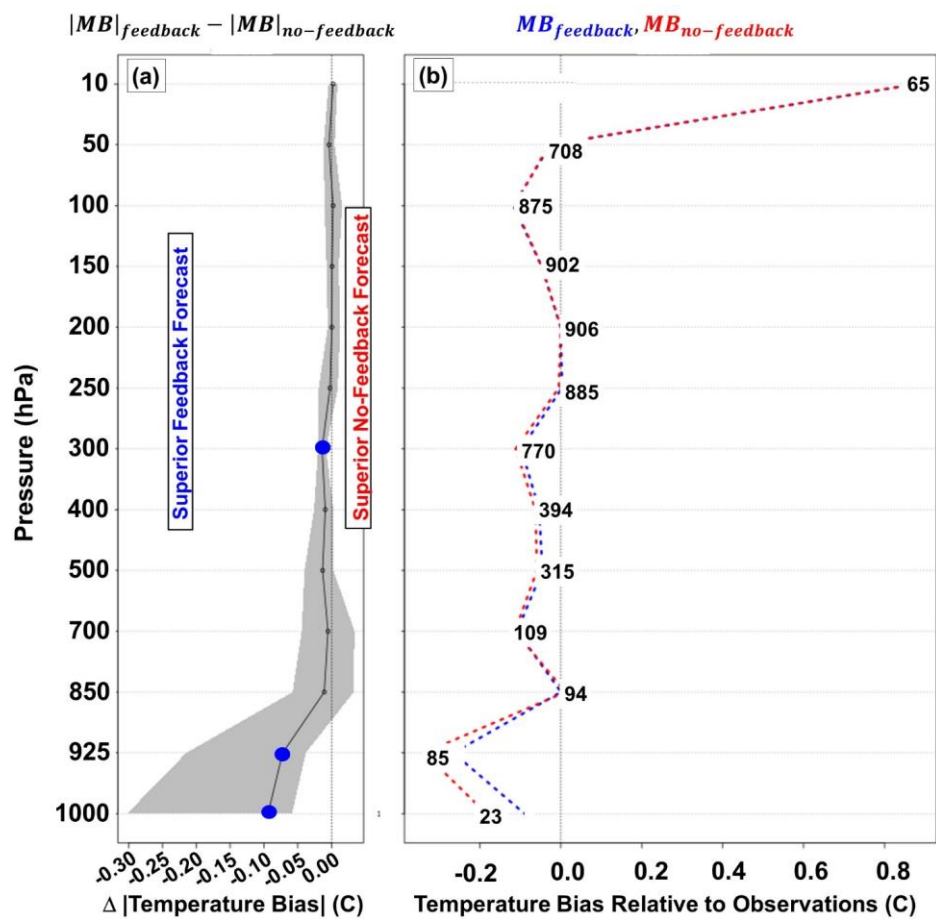
Figure 13: Forecast hour 24 (12 UT) summary upper air temperature performance comparison for air temperature (mean bias, C). (a,b) as in Figure 12.
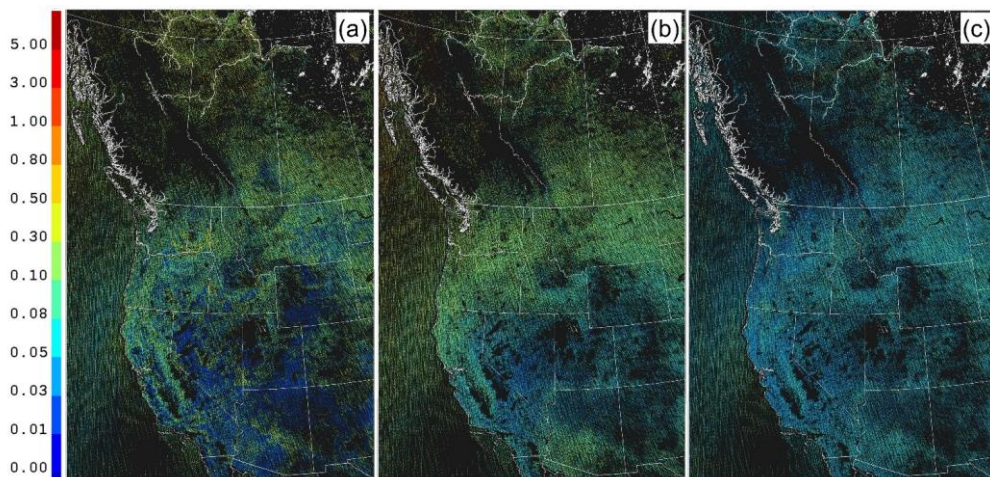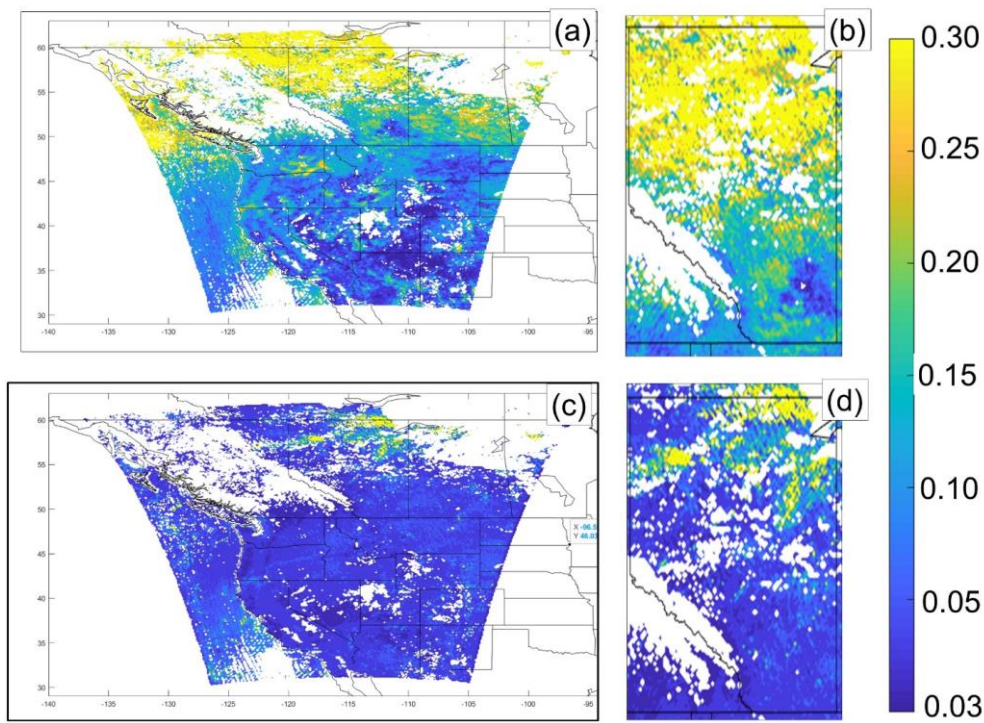
Figure 14: 550nm AOD comparison. (a) ~~All~~ MODIS observations ~~spatiotemporally averaged and~~ sampled over the model domain and forecast duration and (b) ~~zoom in on the province of Alberta.~~ (c,d) GEM-MACH 2.5km simulation, driven by 10km GEM-MACH simulations, in turn driven by ECMWF Reanalysis ~~nearest output hour AOD values corresponding to satellite sensor data availability~~ for 2.5km domain boundary conditions . (c) GEM-MACH 2.5km simulation, driven by MOZART climatological boundary conditions. ~~the same regions.~~
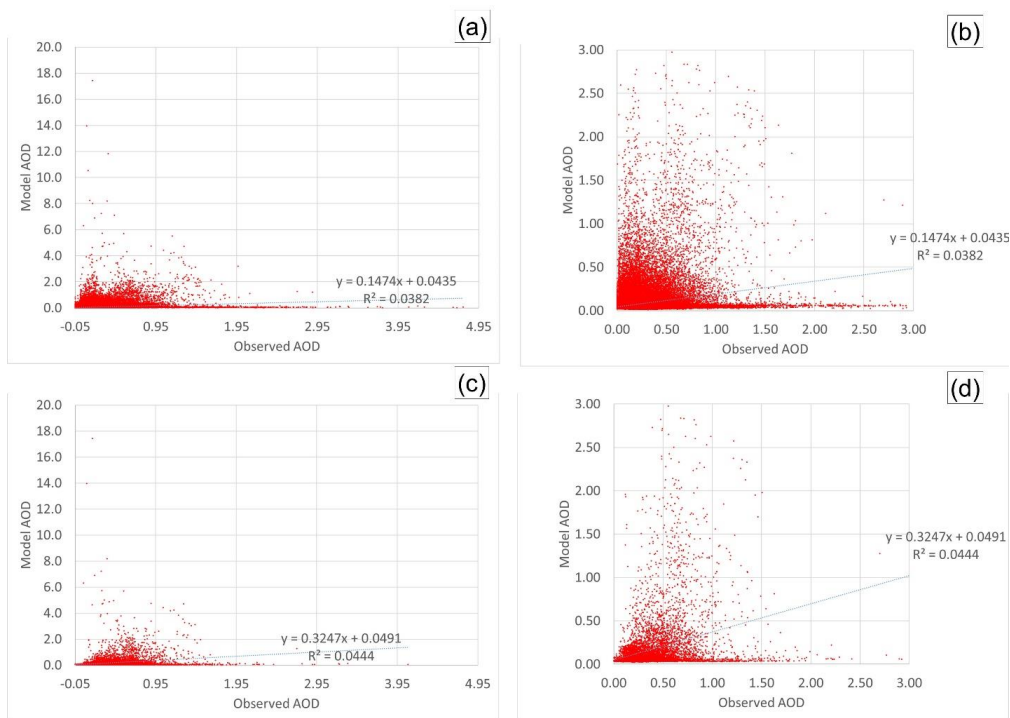
Figure 15: Scatterplots comparing model and observed AOD values. (a) All model observation pairs. (b) As in (a), rescaled to focus on AODs in the 0-3 range. (c) Model observation pairs in the region bounded by 56N, 60N, 110W and 120 W (northern Alberta). (d) As in (c), rescaled to focus on AODs in the 0-3 range.
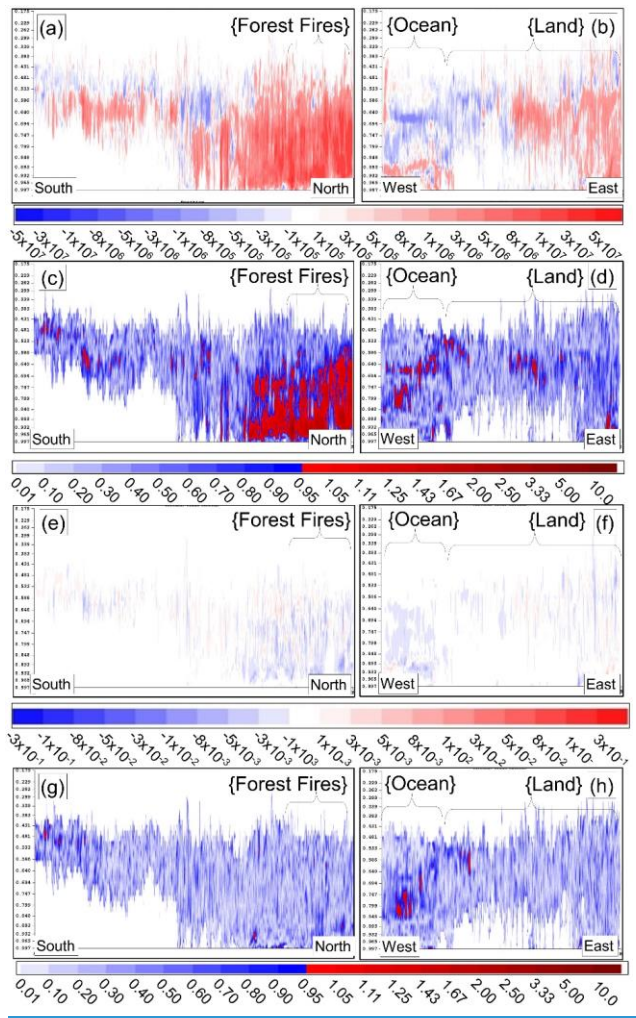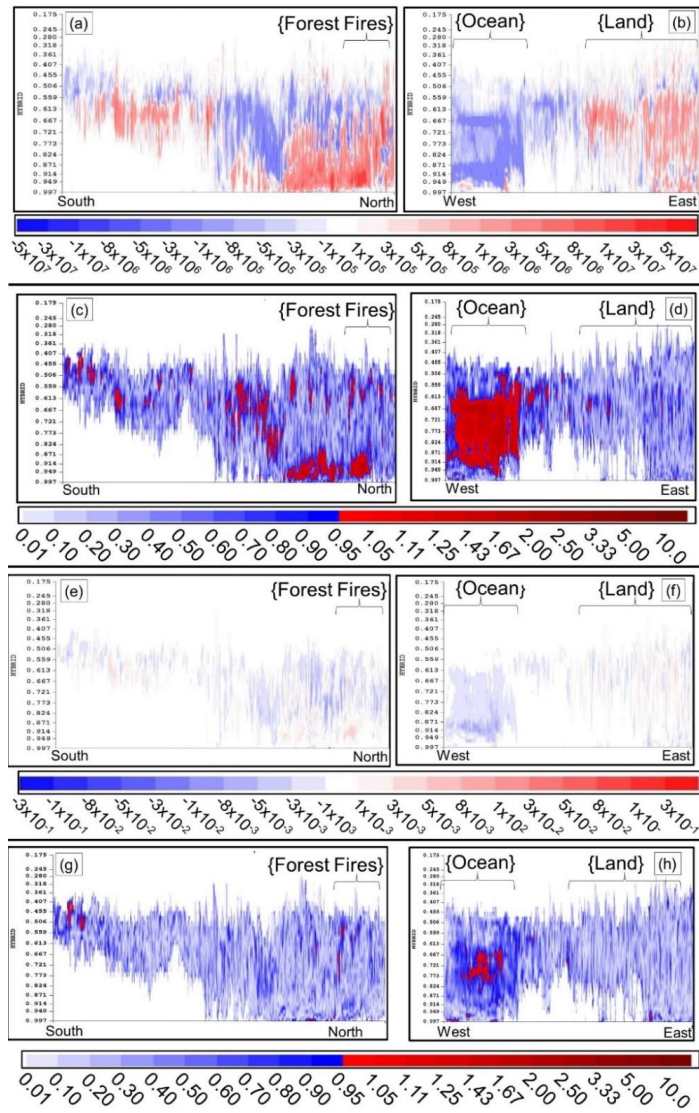
Figure 15~~16~~. (a,b) Difference in mean (Feedback – No-Feedback) cloud droplet number simulations along south to north and east to west cross-sections through the middle of the model domain. (c,d) Corresponding significance level of mean cloud droplet number

differences using the confidence ratio defined in equation (1) – red areas indicate ratio values greater than unity, i.e., significance at or above the 90% confidence level. (e,f) Difference in mean cloud droplet mass (g kg^{-1}) (g,h) Corresponding significance level of mean cloud droplet mass difference. *Note: the vertical axis in hybrid coordinates does not show all model levels for clarity; the model has much finer resolution in the lower part of the atmosphere than shown, and the portion of the vertical domain shown encompasses only the lower half of the levels in the model.*
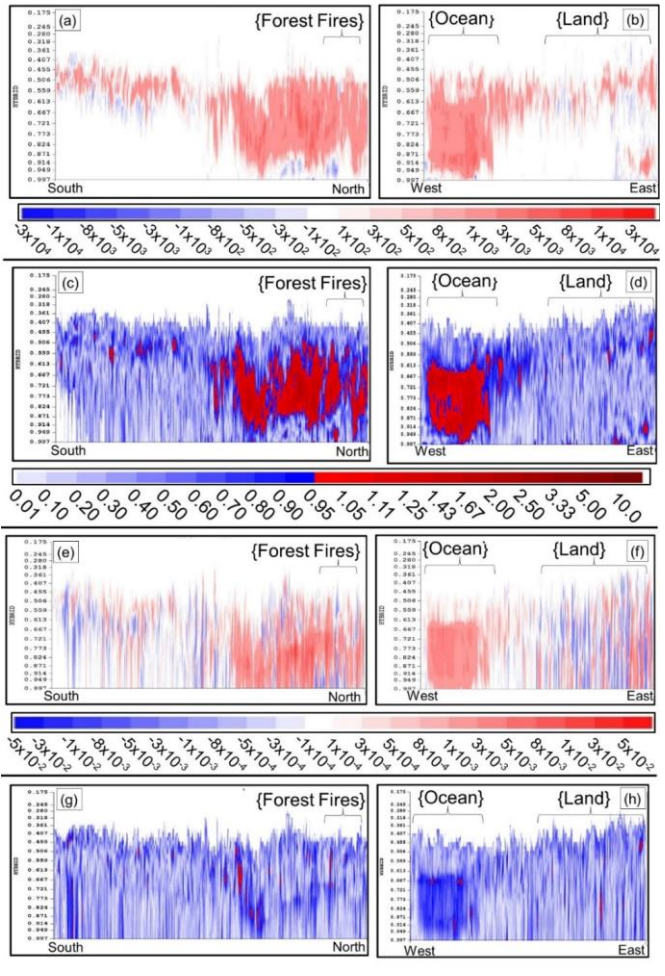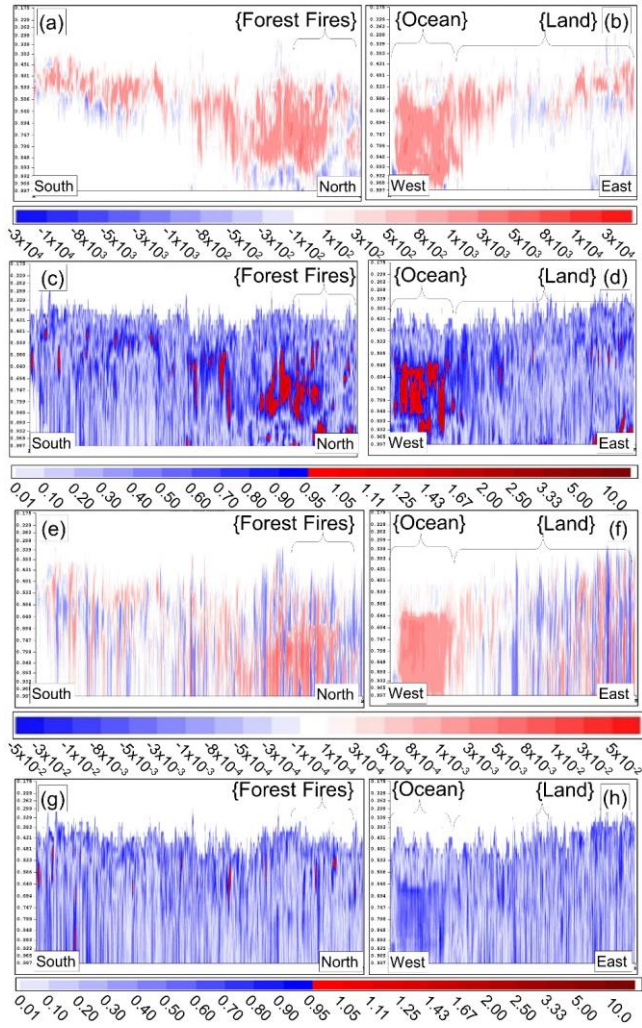
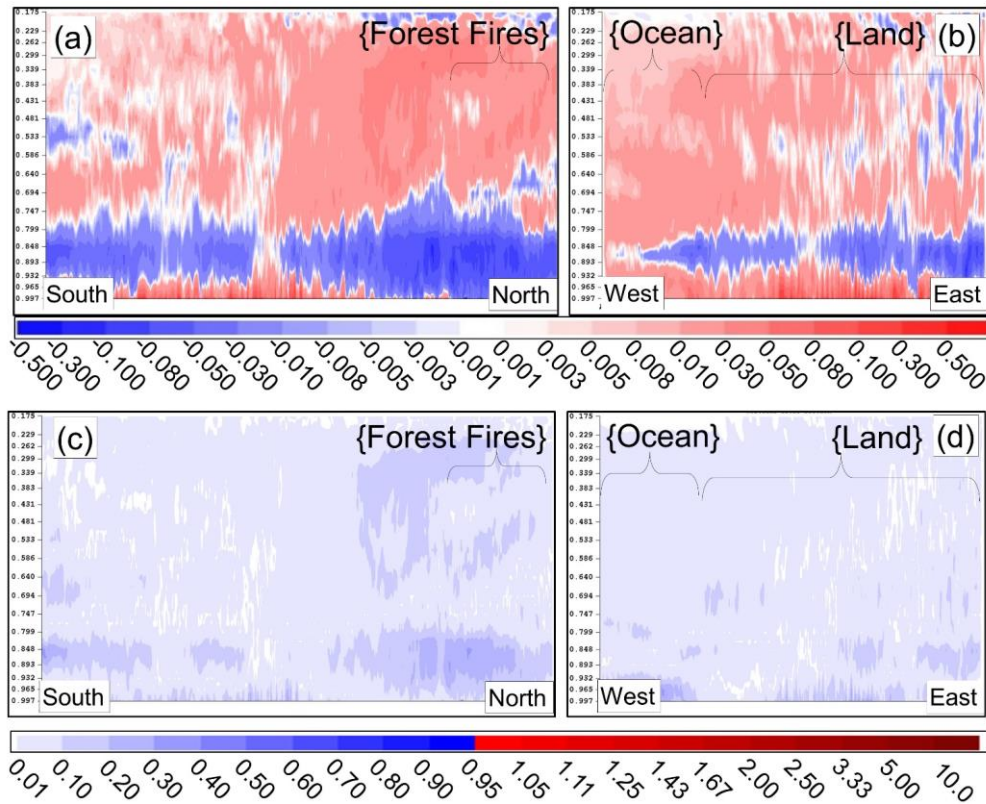**Figure 16~~17~~. (a,b) Difference in mean (Feedback – No-Feedback) rain drop number simulations along south-to-north and east-to-west cross-sections through the middle of the model domain. (c,d) Corresponding significance level of mean rain drop number differences using the confidence ratio defined in equation (1) – red areas indicate ratio values greater than unity, i.e., significance at or above the 90% confidence level. (e,f) Difference in rain cloud drop mass (g kg^{-1}) (g,h) Corresponding significance level of mean rain drop mass difference.**

75
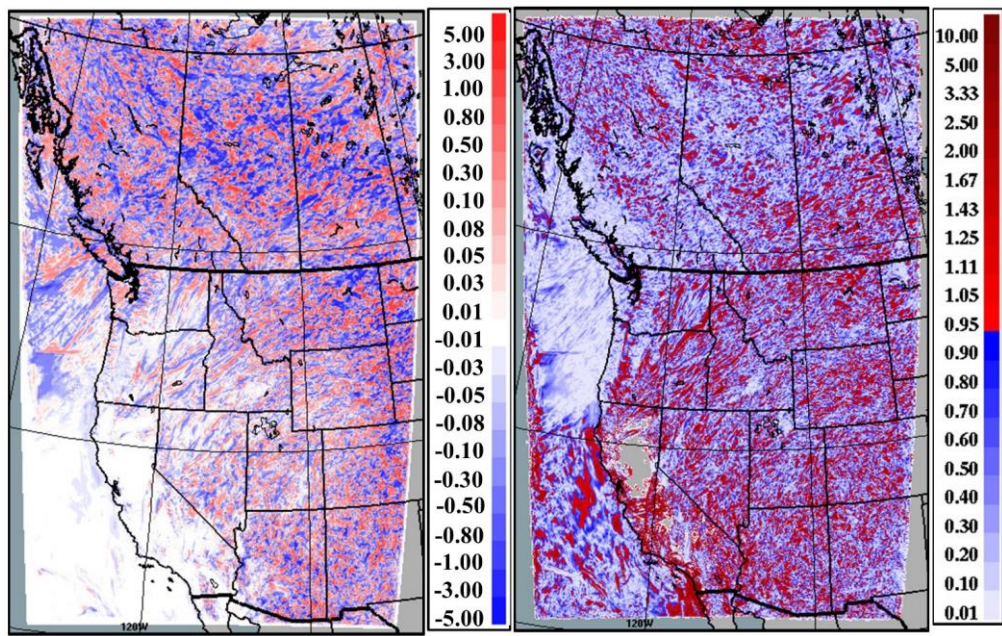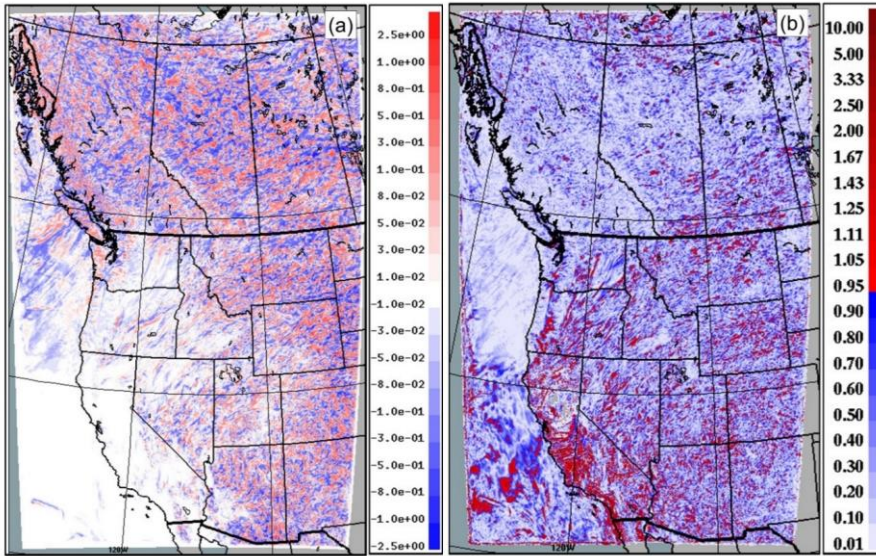
Figure 17̶18̶: (a) Average (Feedback – No Feedback) total surface precipitation during the simulation period. (b) 90% confidence ratio – values greater than 1 indicate significantly different results at the 90% confidence level.
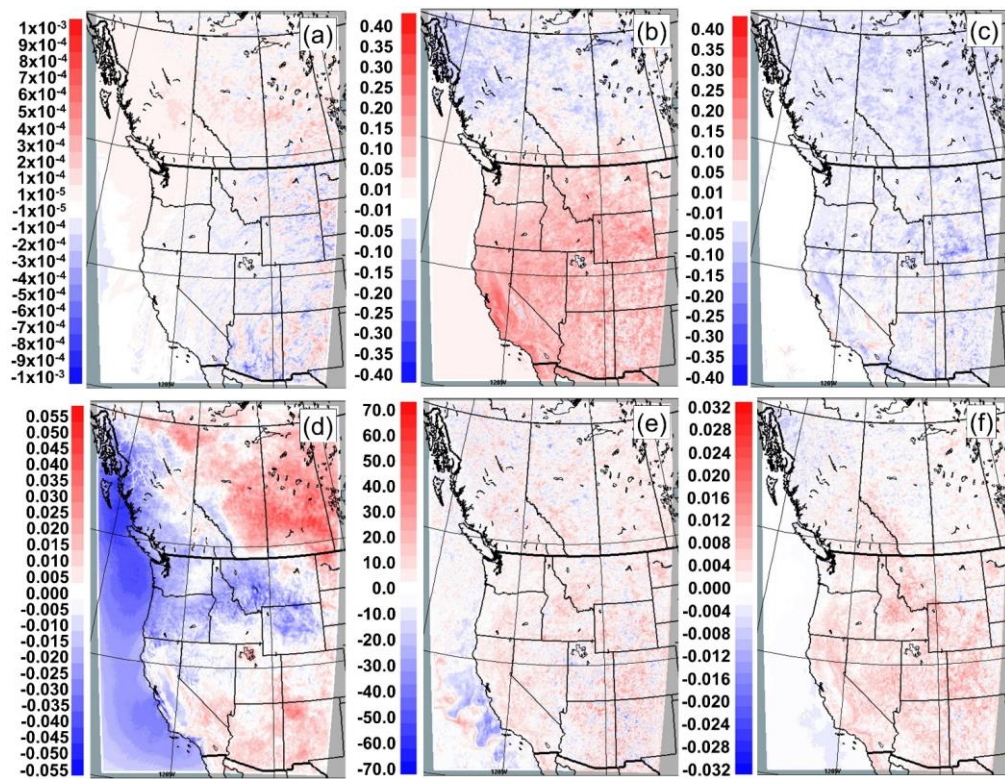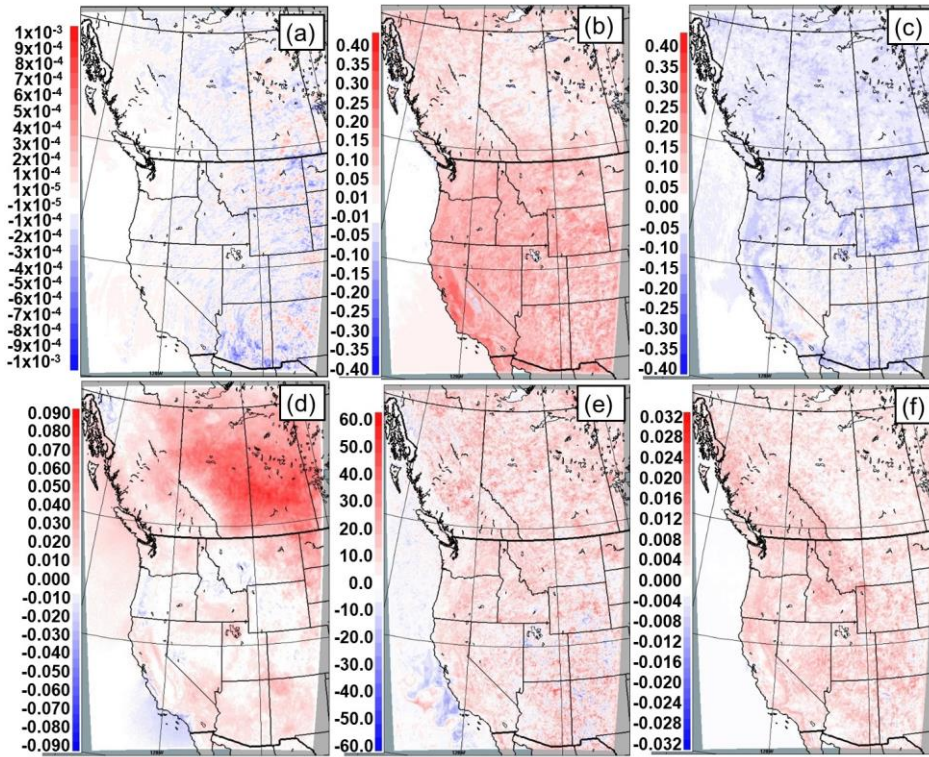
Figure 1819: Differences in average meteorological fields (feedback – no-feedback; red values indicate more positive values in the feedback simulation than in the no-feedback simulation). Panels show average difference in: (a) specific humidity (g kg^{-1}); (b) air temperature (C), (c) dewpoint temperature (C), (d) surface pressure (mb), (e) planetary boundary layer height (m), (f) friction velocity (m s^{-1}).
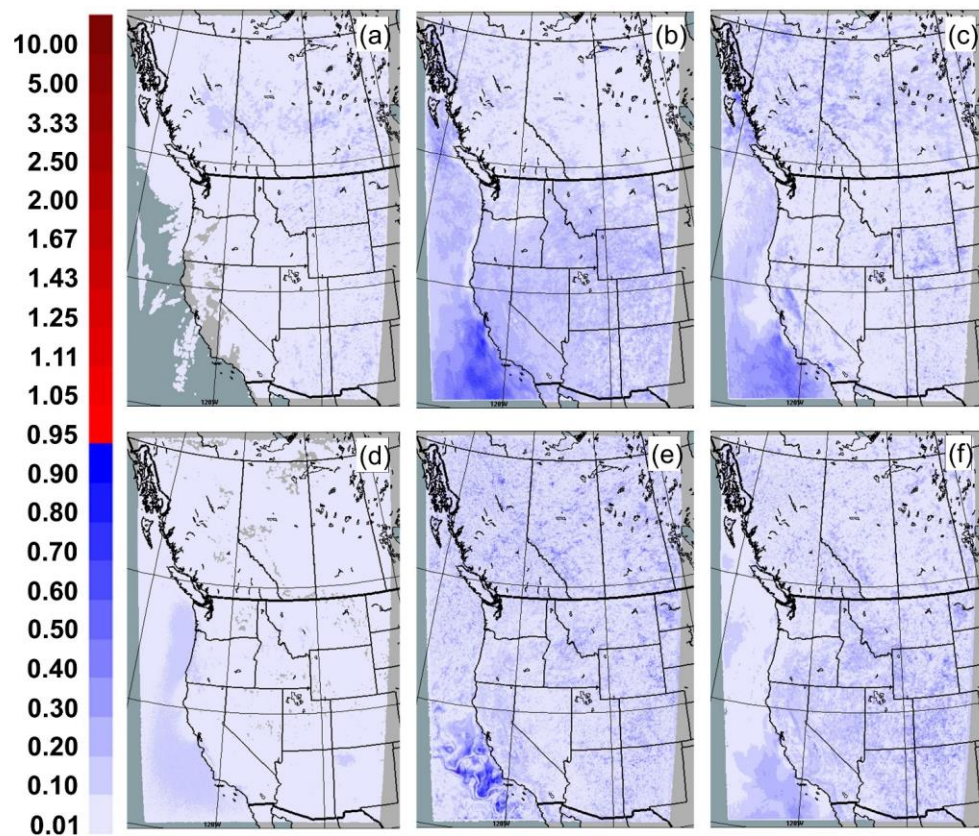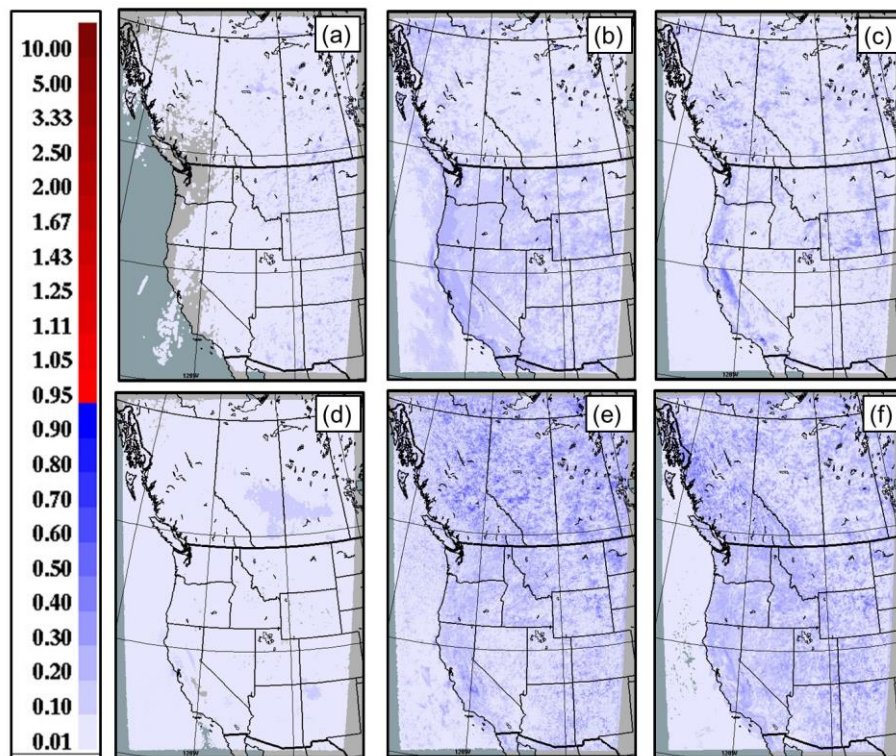
Figure 1920: 90% confidence ratios, same fields as Figure 19. Values greater than 1 indicate significantly different results at or greater than the 90% confidence level.
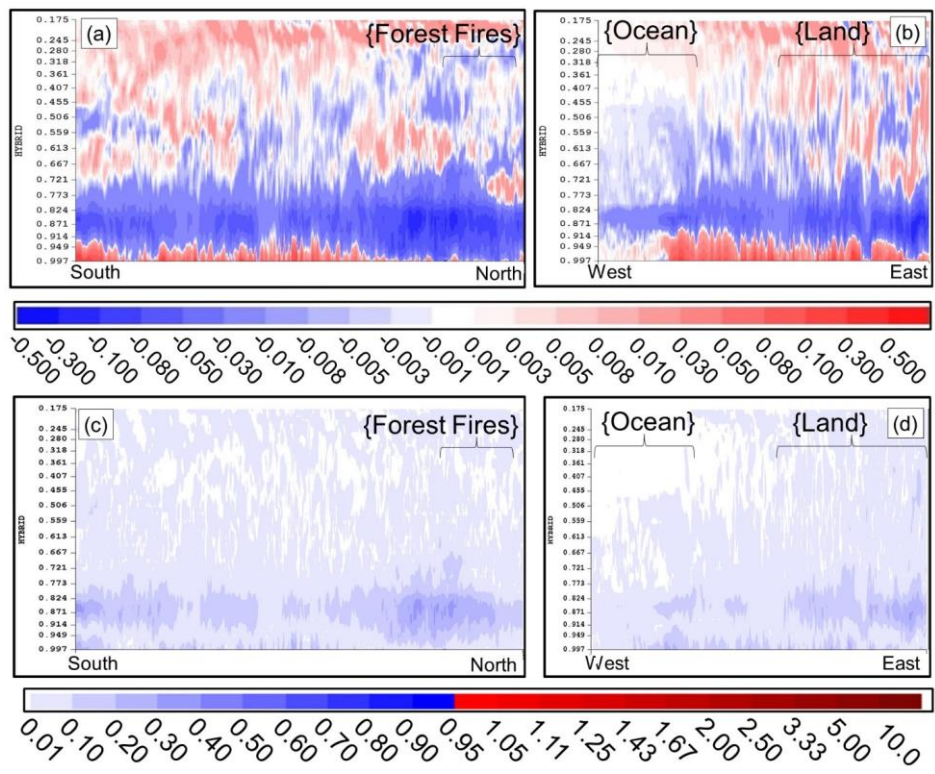
Figure 2021: (a,b) Difference in mean (Feedback – No-Feedback) temperature simulations along south-to-north and east-to-west cross-sections through the middle of the model domain. **(c,d)** Corresponding confidence ratio of mean temperature differences– red areas indicate ratio values greater than unity, i.e., significance at or above the 90% confidence level. ~~Note the reduction in temperatures between hybrid levels 0.90 to 0.70, similar to the findings of Saponaro *et al.*, (2017).~~
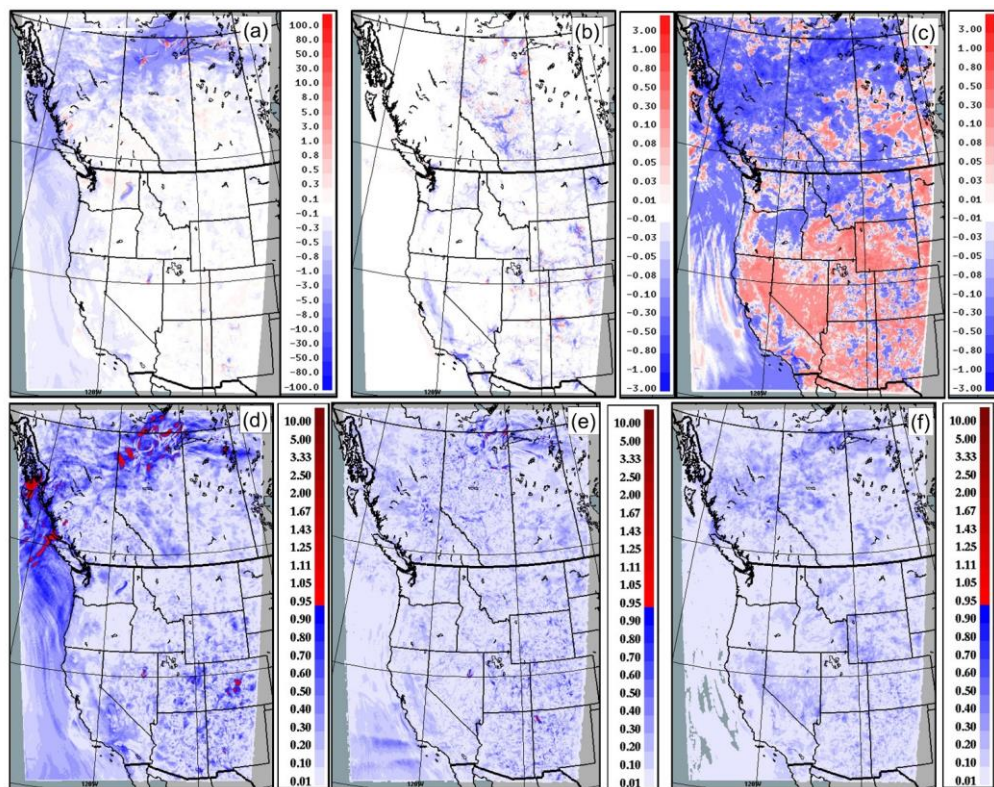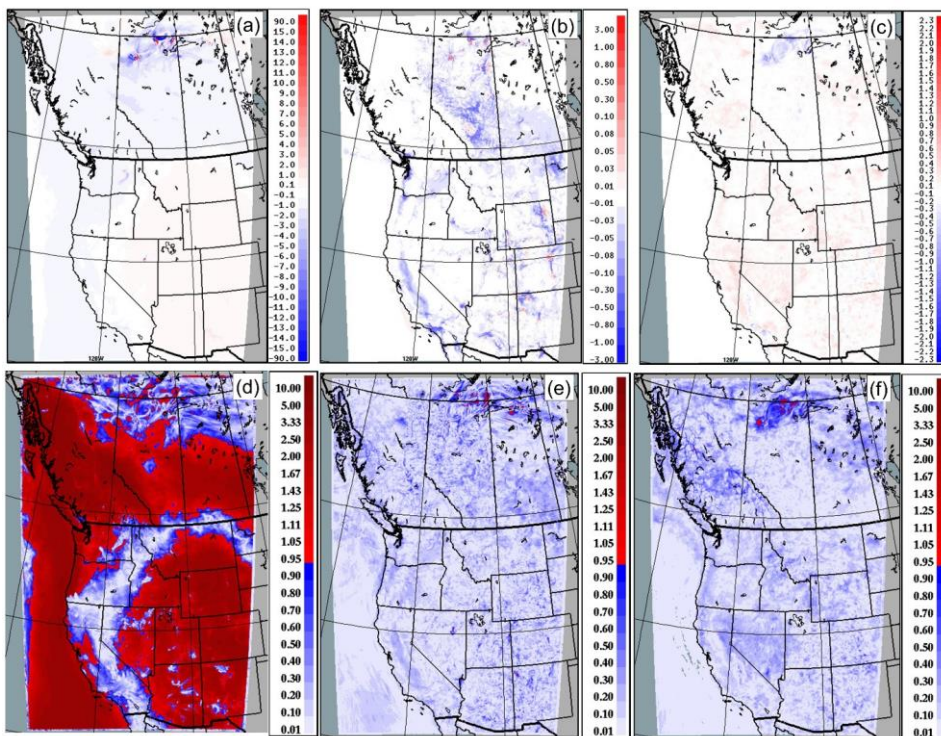
Figure 2122: (a,b,c) Difference (Feedback – No-Feedback) in surface mean PM2.5 (ug m^{-3}), NO$_2$ (ppbv) and O$_3$ (ppbv), respectively. (d,e,f) Corresponding confidence ratio of mean differences – red areas indicate ratio values greater than unity, i.e., significance at or above the 90% confidence level.
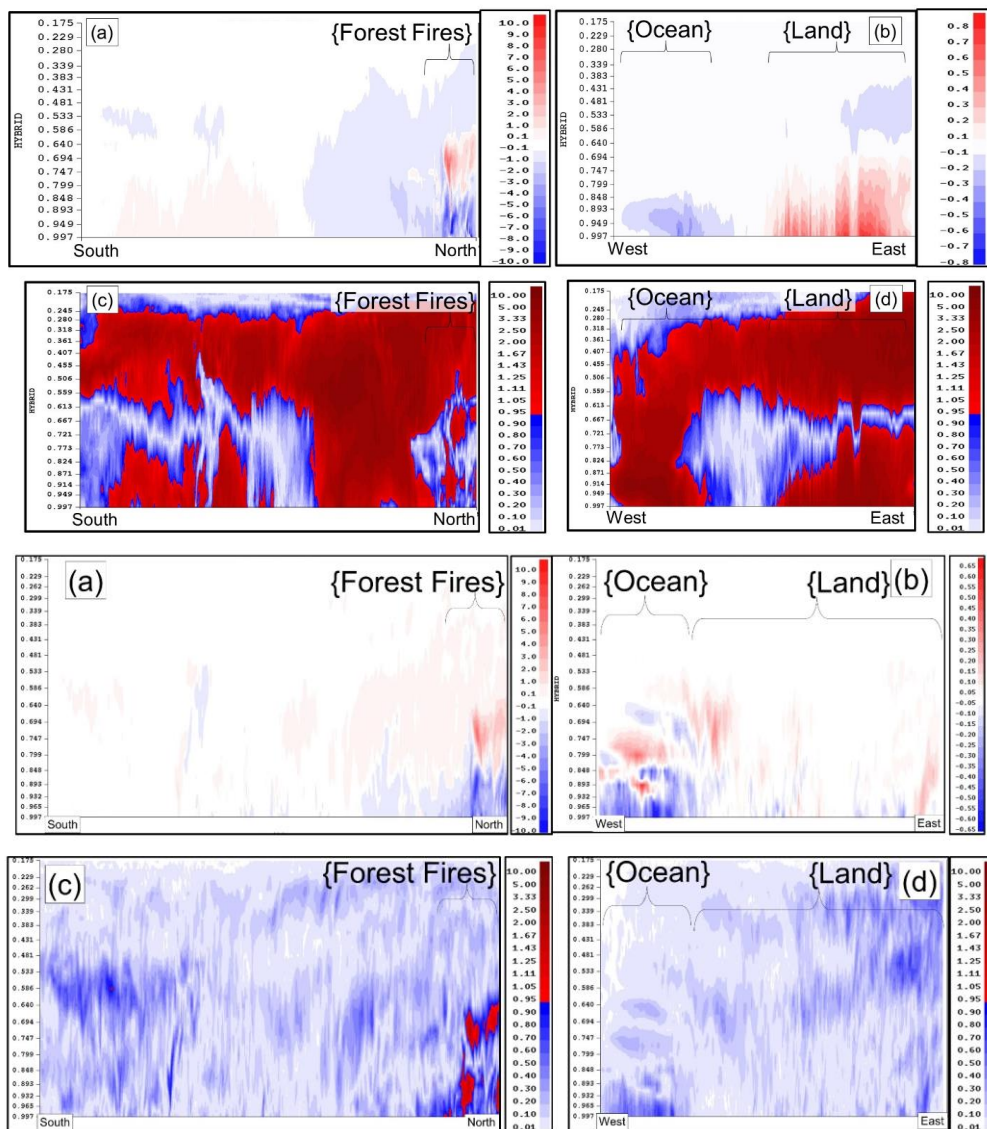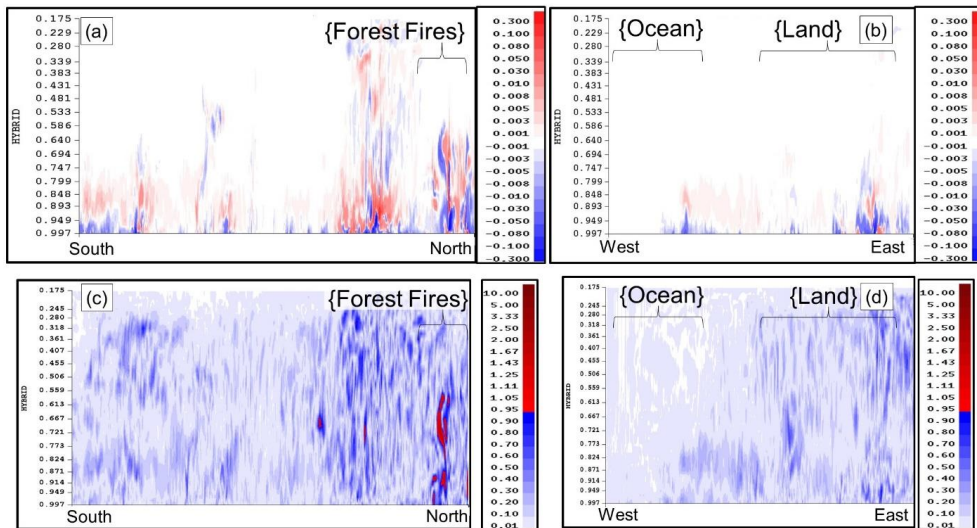
Figure 2223: (a,b) Difference (Feedback – No-Feedback) in predicted mean PM2.5 (ug m^{-3}), along domain-center South-North and West – East cross-sections. (c,d) Corresponding confidence ratio of mean differences – red areas indicate ratio values greater than unity, i.e., significance at or above the 90% confidence level. Note that colour bar scales differ between (a) and (b).
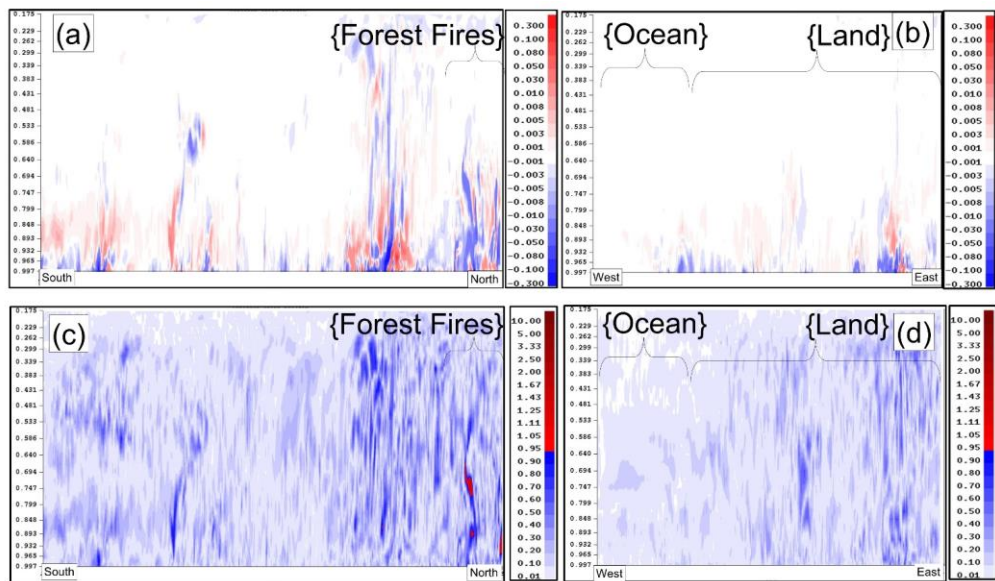
Figure 2324: (a,b) Difference (Feedback – No-Feedback) in predicted mean NO$_2$ (ppbv), along domain-center South-North and West – East cross-sections. (c,d) Corresponding confidence ratio of mean differences – red areas indicate ratio values greater than unity, i.e., significance at or above the 90% confidence level.
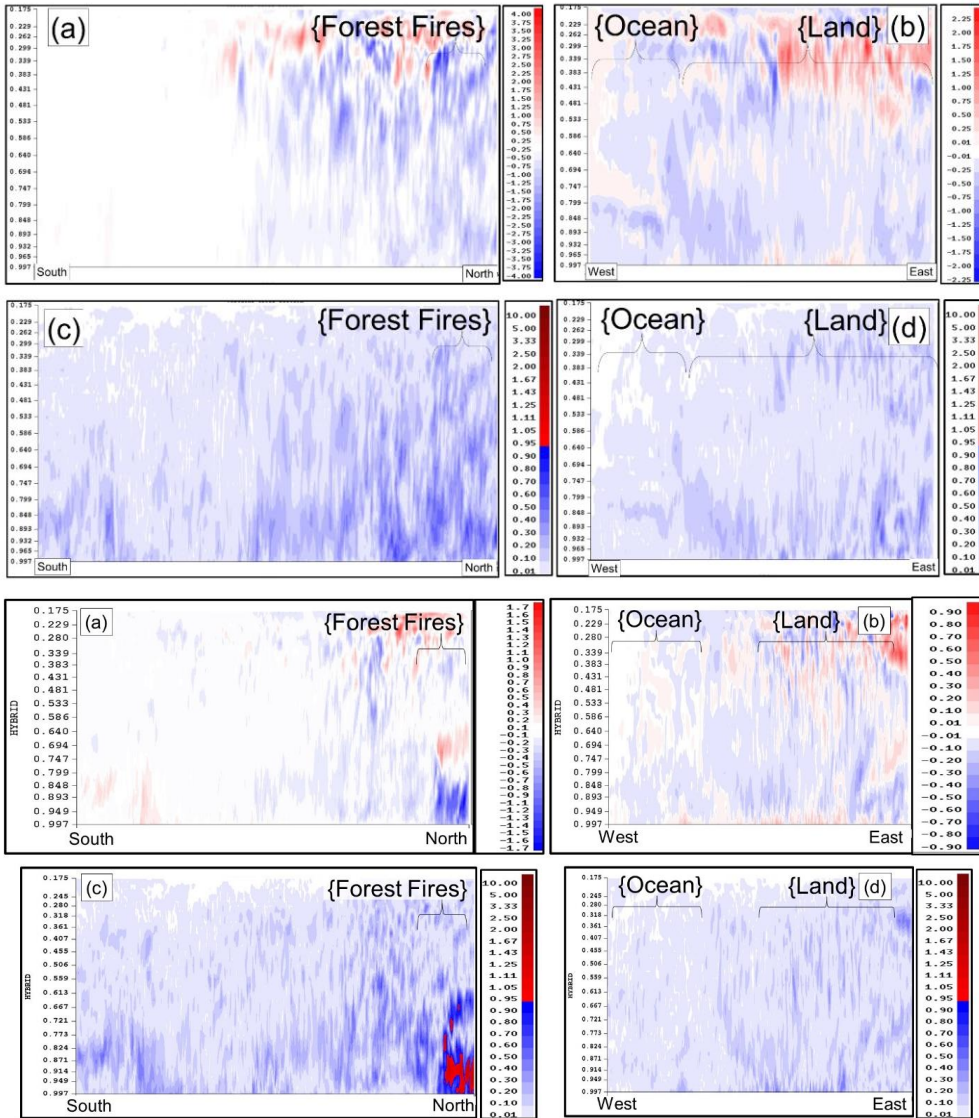
Figure 24̶25̶: (a,b) Difference (Feedback – No-Feedback) in predicted mean O$_3$ (ppbv), along domain-center South-North and West –East cross-sections. (c,d) Corresponding confidence ratio of mean differences – red areas indicate ratio values greater than unity, i.e., significance at or above the 90% confidence level. Note that colour bar scales differ between (a) and (b).