A Diagnostic Intercomparison of Modeled Ozone Dry Deposition Over North America and Europe Using AQMEII4 Regional-Scale Simulations

Christian Hogrefe[1], Stefano Galmarini[2], Paul A. Makar[3], Ioannis Kioutsioukis[4], Olivia E. Clifton[5,6], Ummugulsum Alyuz[7], Jesse O. Bash[1], Roberto Bellasio[8], Roberto Bianconi[8], Tim Butler[9], Philip Cheung[3], Alma Hodzic[10], Richard Kranenburg[11], Aurelia Lupascu[12], Kester Momoh[7], Juan Luis Perez-Camanyo[13], Jonathan E. Pleim[1], Young-Hee Ryu[14], Roberto San Jose[13], Martijn Schaap[11], Donna B. Schwede[1,*], and Ranjeet Sokhi[7]

[1]Center for Environmental Measurement and Modeling, US Environmental Protection Agency, 109 T.W. Alexander Dr., P.O. Box 12055, RTP, NC 27711, USA
[2]JRC, European Commission, Ispra, Italy
[3]Environment and Climate Change Canada, Toronto, Canada
[4]Department of Physics, University of Patras, Patras, Greece
[5]Center for Climate Systems Research, Columbia University, New York, NY, USA
[6]NASA Goddard Institute for Space Studies, New York, NY, USA
[7]Centre for Climate Change Research (C3R), U. Hertfordshire, UK
[8]Enviroware srl, Concorezzo, MB, Italy
[9]Research Institute for Sustainability – Helmholtz Centre Potsdam, Germany
[10]NCAR, Boulder, CO, USA
[11]TNO, Utrecht, the Netherlands
[12]ECMWF, Bonn, Germany
[13]Technical University of Madrid (UPM), Madrid, Spain
[14]Yonsei University, Seoul, South Korea
[*]retired

Correspondence to: Christian Hogrefe (hogrefe.christian@epa.gov)

Abstract. This study analyzes ozone (O3) dry deposition fluxes and velocities (Vd) from twelve regional-scale simulations that were performed over North America and Europe in Phase 4 of the Air Quality Model Evaluation International Initiative (AQMEII4). AQMEII4 collected grid-aggregated and land use (LU)-specific O3 Vd and effective conductances and fluxes for the four major dry deposition pathways. Consistent with recent findings in the AQMEII4 point model intercomparison study, analysis of the grid-aggregated fields shows that grid models with similar Vd can exhibit significant differences in the absolute and relative contributions of the different depositional pathways. Analysis of LU-specific Vd and effective conductances reveals a general increase in model spread compared to grid-aggregated values. This indicates that an analysis of only grid-aggregated deposition diagnostics can mask process-specific differences that exist between schemes. An analysis of AQMEII4 LU distributions across models revealed substantial differences in the spatial patterns and abundance of certain LU categories over both domains, especially for non-forest partially vegetated categories such as agricultural areas, shrubland, and grassland.

We demonstrate that these differences can significantly contribute to or even drive differences in LU-specific dry deposition fluxes which can increase the variability in model predictions of ozone. Two recommendations for future deposition-focused modeling studies emerging from the AQMEII4 analyses presented here are to 1) routinely generate diagnostic outputs to advance a process-based understanding of modeled deposition and support impact analyses, and 2) recognize the importance of documenting and analyzing the representation of LU across models and work towards harmonizing this aspect when using air-quality grid models and model ensembles for deposition analyses.

1 Introduction

Dry deposition to the surface is one of the largest sinks in the budget of tropospheric ozone (O_3) (Wild, 2007) and can be a key contributor to O_3 variability (Clifton et al., 2020a; Lin et al., 2020; Baublitz et al., 2020). As summarized in Hardacre et al. (2015) based on previous global modeling studies (Stevenson et al., 2006; Wild, 2007; Ganzeveld et al., 2009; Young et al., 2013), the magnitude of this sink has been estimated to be on the order of 1000 Tg/yr, with about 2/3 of the dry deposition flux occurring over land and the remaining 1/3 occurring over oceans. Most global and regional chemistry transport models (CTMs) employ resistance frameworks to represent O_3 dry deposition, such as the frameworks introduced in Wesely (1989) and Zhang et al. (2003). In these frameworks, dry deposition fluxes are represented as the product of the O_3 mixing ratio and a deposition velocity (V_d) which in turn is represented as a combination of process-specific resistances acting in parallel or in series to impede the transfer of mass to the surface. However, the implementation of these resistance frameworks can differ significantly across models, especially with respect to representing the surface resistance and its split between the stomatal and non-stomatal components (Hardacre et al., 2015; Wu et al., 2018; Wong et al., 2019; Clifton et al., 2020b and 2023; Galmarini et al., 2021).

The ability of CTMs to provide spatially continuous fields of O_3 mixing ratios, V_d, and dry deposition fluxes has led to their use in numerous studies quantifying ecosystem damages due to O_3 dry deposition. In some studies, ecosystem impacts were calculated using O_3 mixing ratios simulated by global CTMs in conjunction with crop distribution datasets and literature-based concentration-response relationships without explicitly taking into account modeled stomatal dry deposition fluxes (Wang and Mauzerall, 2004; Van Dingenen, 2009; Avery et al., 2011). In a study focused on assessing O_3 dry deposition in global models, Hardacre et al. (2015) analyzed fields of monthly grid-aggregated O_3 dry deposition fluxes from 15 CTMs that had participated in the first Task Force on Hemispheric Transport of Air Pollution (TF HTAP) model intercomparison project (Fiore et al., 2009). Because the native land use (LU) databases used by each model were not available, the study overlaid the modeled fluxes on two independent global LU databases to estimate O_3 fluxes to specific LU types. A similar approach was taken by Schwede et al. (2018) to estimate global nitrogen deposition to forests using deposition fluxes archived from the second TF HTAP model intercomparison (Tan et al., 2018) and combining them with a global forest cover dataset independent of the models' internal LU databases. While this approach makes direct use of modeled dry deposition fluxes, potential

inconsistencies between the LU datasets used in the CTMs and those used in the post-processing attribution of modeled fluxes to specific ecosystems can introduce uncertainties. Moreover, utilizing archived grid-aggregated deposition fluxes does not account for potential subgrid variations in LU that could have a strong impact on ecosystem specific deposition. To explore the impacts of such subgrid effects, Schwede et al. (2018) also presented results from the EMEP MSC-W model (Simpson et al., 2012) which employs a mosaic approach to calculate and optionally output V_d and deposition fluxes for each LU present in a grid cell. Results showed that the total global estimate of nitrogen deposition to forest was 12% higher using forest-specific EMEP fluxes compared to grid-aggregated fluxes and that locally fluxes estimated using the two approaches could differ by more than a factor of 2 depending on the LU heterogeneity in a given grid cell (Schwede et al., 2018). A companion paper in this issue (Makar et al., 2025) showed large land-use dependent variations in deposition fluxes for both sulfur and nitrogen species (particularly for bidirectional fluxes of ammonia gas) and noted that LU databases used for post-processing critical load analysis may also differ from those used in regional models, with harmonization of these databases recommended for future modeling studies. ~~(Makar et al., 2024).~~

As discussed above, CTM-based approaches to assess ecosystem impacts of air pollution can be affected by uncertainties in input datasets (emissions, meteorology, and LU) as well as the representation of a variety of chemical and physical processes. Hardacre et al. (2015) noted that advancing a process-level understanding of how uncertainties and errors in modeled dry deposition contribute to overall CTM error and variability would require the generation of detailed diagnostic model outputs at an hourly resolution, including stomatal and non-stomatal conductances and fluxes at both the grid-aggregated level and for individual LU categories. While a few global modeling studies using individual CTMs have incorporated such diagnostic fields and demonstrated their utility (Paulot et al., 2018; Clifton et al., 2020a) and prior to AQMEII-4 this practice was recommended by a recent review (Clifton et al. 2020b), to date it had not been adopted in model intercomparison studies at either the global or regional scale. Partially motivated by this research gap, the fourth phase of the Air Quality Model Evaluation International Initiative (AQMEII4) was designed around complementary grid and single point modeling activities for a diagnostic intercomparison and evaluation of simulated deposition with a specific focus on dry deposition of gaseous species (Galmarini et al., 2021; Clifton et al., 2023).

In this study conducted within the framework of AQMEII4, we apply the diagnostic approaches described in the AQMEII4 technical note (Galmarini et al., 2021) and the manuscript presenting the evaluation of single point models at eight long-term O_3 flux measurement sites (Clifton et al., 2023) to the analysis of O_3 dry deposition simulated by the AQMEII4 regional-scale models for analysis domains over North America (NA) and Europe (EUR). The extent of these analysis domains was defined in Galmarini et al. (2021) and is shown in Figures 1 and 2. ~~EU).~~ Our analysis also leverages the common AQMEII4 LU types defined in Galmarini et al. (2021). As discussed in that study, assigning model native LU types to a common set of 16 AQMEII4 LU types allows consistent LU-specific comparisons (note that some of the most common LU types, such as evergreen needleleaf forest, are present in all native model LU databases). First, we quantify temporal and spatial variations in estimated O_3 deposition by analyzing different deposition pathways and comparing grid-aggregated vs. AQMEII4 LU-

specific diagnostics. We also demonstrate that the LU-specific diagnostics can be used to connect grid model results to the evaluation of single point models against long-term O_3 flux measurement sites described in Clifton et al. (2023). Next, we assess the impacts of differences in LU between the AQMEII4 grid models on simulated fluxes to different ecosystems. Finally, we discuss the implications of our analyses to O_3 deposition impact studies, particularly impact studies based on datasets from multi-model intercomparison activities, as well as to future model development aimed at improving dry deposition process representation.

2 Datasets

Table 1 lists the participating regional-scale models, the institutions performing the simulations, the continent each simulation was performed for, their dry deposition schemes, ~~and~~ the LU classification scheme used in their dry deposition calculations, and the thickness of the first model layer. Detailed descriptions of the model configurations were provided in Makar et al. (2024) while their dry deposition algorithms were documented in Galmarini et al. (2021), Clifton et al. (2023), and Hogrefe et al. (2023). The 16 AQMEII4 LU types to which native LU types were assigned are described in Galmarini et al. (2021) and discussed in detail in section 3.3. An operational and probabilistic evaluation of simulated O_3 mixing ratios against ground level network observations over NA and EUR~~EU~~ is presented in Kioutsioukis et al. (2025).

~~Modeling System~~	~~Domain~~	~~Modeling Group~~	~~Dry Deposition Scheme~~	~~LU for Dry Deposition Scheme~~
~~WRF/CMAQ (M3Dry)~~	~~NA~~	~~U.S. EPA~~	~~M3Dry~~	~~MODIS~~
~~WRF/CMAQ (STAGE)~~	~~NA~~	~~U.S. EPA~~	~~STAGE~~	~~AQMEII4~~
~~GEM-MACH (Base)~~	~~NA~~	~~Environment and Climate Change Canada~~	~~Wesely~~	~~Makar et al., (2018)~~
~~GEM-MACH (Zhang)~~	~~NA~~	~~Environment and Climate Change Canada~~	~~Zhang~~	~~Zhang et al. (2003)~~
~~GEM-MACH (Ops)~~	~~NA~~	~~Environment and Climate Change Canada~~	~~Wesely~~	~~Makar et al., (2018)~~
~~WRF-Chem (RIFS)~~	~~NA~~	~~Research Center for Sustainability (RIFS)~~	~~Wesely~~	~~USGS24~~
~~WRF-Chem (UPM)~~	~~NA~~	~~Technical University of Madrid (UPM)~~	~~Wesely~~	~~USGS24~~
~~WRF-Chem (NCAR)~~	~~NA~~	~~National Center for Atmospheric Research / Yonsei University~~	~~Wesely~~	~~USGS24~~
~~WRF-Chem (RIFS)~~	~~EU~~	~~Research Center for Sustainability (RIFS)~~	~~Wesely~~	~~USGS24~~
~~WRF-Chem (UPM)~~	~~EU~~	~~Technical University of Madrid (UPM)~~	~~Wesely~~	~~USGS24~~
~~LOTOS/EUROS~~	~~EU~~	~~TNO~~	~~DEPAC~~	~~DEPAC~~
~~WRF/CMAQ (STAGE)~~	~~EU~~	~~University of Hertfordshire~~	~~STAGE~~	~~AQMEII4~~

~~**Table 1:** Model simulations analyzed in this study. See Makar et al. (2024) for additional details on the configuration of each simulation.~~

The analysis in this manuscript utilizes dry deposition fluxes and the O_3 dry deposition diagnostic variables described in detail in Galmarini et al. (2021) and Clifton et al. (2023), namely O_3 dry deposition velocities (V_d) and effective conductances (Paulot et al., 2018; Clifton et al., 2020a; Galmarini et al., 2021; Clifton et al., 2023) for the stomatal, cuticular, soil, and lower canopy surface deposition pathways reported at both the grid-aggregated scale and for each of the 16 AQMEII4 LU categories. Following the general framework introduced in Wesely et al. (1989), all models in this study except WRF/CMAQ STAGE calculate Vd as the inverse of the sum of the aerodynamic resistance r_a, the quasi-laminar sublayer resistance r_b, and the surface resistance r_c which consists of most or all of the parallel deposition pathways listed above but is implemented differently across schemes as documented in Galmarini et al. (2021). The aerodynamic resistance r_a is independent of deposition pathway in all models, and r_b is independent of deposition pathway in all models except WRF/CMAQ STAGE which computes different r_b values for stomatal and cuticular vs. soil pathways and therefore includes this term in its pathway-specific r_c calculations. As stated in Galmarini et al. (2021), "an effective conductance is the contribution of a given depositional pathway to the deposition velocity, expressed in the same units as the deposition velocity. The sum of the effective conductances for all deposition pathways is the deposition velocity." Importantly, Galmarini et al. (2021) noted that the equations used to compute the effective conductances are specific to a given dry deposition scheme and provided these equations for the different AQMEII4 modeling systems in Appendix B.) ~~for the stomatal, cuticular, soil, and lower canopy pathways reported at both the grid-aggregated scale and for each of the 16 AQMEII4 LU categories.~~ No effective conductances were available for LOTOS/EUROS but reported grid-aggregated O3 V_d, deposition fluxes, and LU distributions are included in the analysis.

Modeling System	Domain	Modeling Group	Dry Deposition Scheme	LU for Dry Deposition Scheme	Thickness of First Model Layer
WRF/CMAQ (M3Dry)	NA	U.S. EPA	M3Dry	MODIS product MCD12Q1v6 for 2017 (Friedl and Sulla-Menashe, 2019)	Approximately 19 m
WRF/CMAQ (STAGE)	NA	U.S. EPA	STAGE	AQMEII4 (defined in Galmarini et al., 2021) mapped from MODIS product MCD12Q1v6 for 2017 (Friedl and Sulla-Menashe, 2019) as shown in Table S1	Approximately 19 m
GEM-MACH (Base)	NA	Environment and Climate Change Canada	Wesely	Makar et al. (2018) mapped from Zhang et al. (2003) as shown in Table S2	Approximately 42 m
GEM-MACH (Zhang)	NA	Environment and Climate Change Canada	Zhang	Zhang et al. (2003)	Approximately 42 m
GEM-MACH (Ops)	NA	Environment and Climate Change Canada	Wesely	Makar et al. (2018) mapped from Zhang et al. (2003) as shown in Table S2	Approximately 42 m
WRF-Chem (RIFS)	NA	Research Center for Sustainability (RIFS)	Wesely	USGS24 classification using 1-km Advanced Very High Resolution Radiometer (AVHRR) data from Loveland et al, 2000	Approximately 25 m
WRF-Chem (UPM)	NA	Technical University of Madrid (UPM)	Wesely	USGS24 classification using 1-km AVHRR data from Loveland et al, 2000	Approximately 25 m
WRF-Chem (NCAR)	NA	National Center for Atmospheric Research / Yonsei University	Wesely	USGS24 classification mapped from MODIS product MCD12Q1v5.1 climatology (Broxton et al., 2014) as shown in Table S4	Approximately 52 m
WRF-Chem (RIFS)	EUR	Research Center for Sustainability (RIFS)	Wesely	USGS24 classification mapped from CORINE Land Cover (EEA 2020) as shown in Table S3	Approximately 25 m

WRF-Chem (UPM)	EUR	Technical University of Madrid (UPM)	Wesely	USGS24 classification using 1-km AVHRR data from Loveland et al, 2000	Approximately 25 m
LOTOS/EUROS	EUR	TNO	DEPAC	DEPAC LU as described in Manders-Groot et al. (2023), mapped from Global Landcover 2000 dataset (https://forobs.jrc.ec.europa.eu/glc2000, last accessed May 12, 2025)	Approximately 20 m
WRF/CMAQ (STAGE)	EUR	University of Hertfordshire	STAGE	AQMEII4 (defined in Galmarini et al., 2021) mapped from MODIS (Friedl et al., 2002) as shown in Table S1	Approximately 20 m

Table 1: Model simulations analyzed in this study. See Makar et al. (2024) for additional details on the configuration of each simulation.

WRF/CMAQ, GEM-MACH, and LOTOS-EUROS calculate and report V_d and effective conductances at the midpoint of the first model layer while values in WRF-Chem are calculated as average over the first layer following McRae et al. (1982). As shown in Table 1, the height of the first model layer varies across models and may therefore affect the comparison of V_d across models. However, only the computation of r_a has a dependency on layer height. When comparing the magnitude of the annual domain-average r_a, r_b, and r_c terms for O_3 across participating models, we find that r_c is typically at least a factor of 10 higher than r_a, suggesting that the effect of different layer heights on our comparison of V_d is relatively minor and that differences in the representation and computation of the parallel surface resistance pathways constituting r_c are the main driver of differences in V_d. In addition, the independence of the different resistances included in the computation r_c (Galmarini et al., 2021, Clifton et al., 2023) on layer height also implies that our analysis of effective conductances to quantify the split between different surface deposition pathways is likely not affected by model-to-model differences in layer heights.

For WRF-Chem (NCAR), LU-specific dry deposition calculations for its native "mixed shrubland / grassland" category were inadvertently mapped to and then reported for the AQMEII4 "savanna" category (along with its native "savanna" category) rather than the intended AQMEII4 "shrubland" category. This inadvertent mapping was corrected for the comparison of LU categories across models but could not be corrected for the analysis of LU-specific V_d and effective conductances. For all models except GEM-MACH (Base) and GEM-MACH (Ops), the spatial fields of LU categories in both the land-surface model (LSM) and dry deposition code were time independent and considered the "snow and ice" category to represent permanent snow and ice. For the GEM-MACH (Base) and GEM-MACH (Ops) dry deposition calculations, the LU categories used in the dry deposition calculations and subsequently reported to AQMEII4 were modified from their time-independent LSM values to represent the time-varying effects of transient snow cover through the "snow and ice" category, and proportionally reduced the fractions of all other LU categories in the grid cell if snow cover was present. To enable a comparison of these time-

dependent LU fractions to the fractions reported by other models, time-independent fractions for GEM-MACH (Base) and GEM-MACH (Ops) were derived by scanning the hourly reported fields and determining the LU fractions in each grid cell when the coverage of the "snow and ice" category reached its annual minimum (zero for the vast majority of grid cells) for that grid cell. This approach allows the use of LU fractions as stratification criterion for all models in our analysis while retaining the effects of transient snow cover (represented through separate snow cover fields rather than time-dependent LU fractions in other dry deposition schemes) in the reported dry deposition diagnostics.

As described in Galmarini et al. (2021), each simulation reported gridded fields of V_d and effective conductances as 288 individual values representing monthly median diurnal cycles (24 hours times 12 months) at each grid cell for each simulated year. This approach was chosen because reporting and storing these diagnostic fields for all hours in the year would have been logistically impossible, and the selected approach still allows an analysis of their diurnal and seasonal variations. However, it should be noted that this approach of using monthly median values for individual diagnostics can introduce a temporal mismatch between reported V_d and effective conductance diagnostics, which in turn may cause the sum of the effective conductances on a monthly median basis to differ from monthly median V_d, even though on an hour-by-hour basis their sum equals V_d (Paulot et al., 2018; Clifton et al., 2020a; Galmarini et al., 2021; Clifton et al., 2023).~~)~~ Despite this potential for deviations, the sum of the reported effective conductances was found to be within 3.5% of V_d on an annual domain mean basis for all simulations. The largest deviation (3.5%) occurred for WRF/CMAQ (M3DRY) for which grid-aggregated V_d was calculated within the model while effective conductances were estimated through post-processing (Hogrefe et al., 2023). For the remainder of the discussions in this manuscript, we therefore treat the sum of the monthly median effective conductances as equivalent to V_d.

3 Results

3.1 Grid-Aggregated Results

Spatial patterns of these annual deposition fluxes for each model as well as the multi-model mean and normalized standard deviation are displayed in Figure 1 for the NA domain and Figure 2 for the EUR domain. Table 2 lists the annual total O_3 dry deposition fluxes estimated by the participating models over both the NA and EUR~~EU~~ domains. These totals are calculated over all non-water grid cells that were common to all simulations. The multi-model mean value is 74.3 Tg/yr over the NA domain and 67.4 Tg/yr over the EUR~~EU~~ domain. Individual model estimates range from 59.5 Tg/yr (GEM-MACH (Zhang)) to 91.4 Tg/yr (WRF-Chem (UPM)) over the NA domain and from 47.2 Tg/yr (LOTOS/EUROS) to 82.9 Tg/yr (WRF-Chem (UPM)) over the EU domain. For context, past global modeling studies reported *global* annual O_3 deposition totals generally ranging between 800 Tg/yr and 1200 Tg/yr (e.g. Hardacre et al., 2015; Stevenson et al., 2006; Wild 2007; Ganzeveld et al., 2009).

Spatial patterns of these annual deposition fluxes for each model as well as the multi-model mean and normalized standard deviation are displayed in Figure 1 for the NA domain and Figure 2 for the EU domain. Over the NA domain all models estimate that O_3 deposition is highest over the Eastern U.S. as well as along the West Coast, consistent with our expectations that reflect a combination of higher O_3 mixing ratios due to higher O_3 precursor emissions in these regions and higher V_d over areas with higher vegetation density. The pronounced differences in deposition fluxes over the densely vegetated Southeastern U.S. between WRF-Chem (UPM) and GEM-MACH (Zhang) suggest that differences in the methodology used to represent deposition to vegetation are a key driver of model spread. The map of the multi-model normalized standard deviation over NA also shows significant spread in deposition to land over the generally more arid regions in the Western U.S., in addition to the spread over the Southeastern U.S. Over the EUREU domain, the largest relative spread in deposition to land occurs over Northwestern Scandinavia, Ireland, Great Britain, and Turkey, as well as the Alps and Northern Africa. Over both domains, there is a large relative spread in O_3 deposition to water, with generally higher values for WRF-Chem than all other models, but the absolute magnitude of the flux is much smaller than the flux over land.

Table 2. Annual total O3 dry deposition fluxes and annual mean O3 dry deposition velocities estimated by the participating models over both the NA and EU domains. These totals are calculated over all non-water grid cells that were common to all simulations. The NA numbers are for 2016 while the EU number are for 2010.

	Annual total O3 dry deposition flux (Tg)	Annual mean O3 dry deposition velocity (cm/s)
North America (2016)		
WRF/CMAQ (M3Dry)	66.1	0.28
WRF/CMAQ (STAGE)	68.3	0.3
GEM-MACH (Base)	73.9	0.37
GEM-MACH (Zhang)	59.5	0.27
GEM-MACH (Ops)	80.4	0.42
WRF-Chem (RIFS)	79.0	0.34
WRF-Chem (UPM)	91.5	0.38
WRF-Chem (NCAR)	75.5	0.28
Europe (2010)		
WRF-Chem (RIFS)	77.5	0.32
WRF-Chem (UPM)	82.9	0.37
LOTOS/EUROS	47.2	0.21
WRF/CMAQ (STAGE)	61.9	0.25

To assess the role of variations in simulated V_d on the dry deposition fluxes discussed above, Table 2 includes the annual mean O_3 V_d over both domains alongside the annual total O_3. The relative differences between the annual mean V_d generally match those between the annual total deposition fluxes with only small exceptions (e.g., WRF-Chem (RIFS) has slightly lower V_d but a slightly higher dry deposition fluxes than GEM-MACH (Base)). Consistent with the discussion in Section 2, the model-

to-model differences in V_d shown in Table 2 indeed do not appear to be caused by the different first layer thicknesses shown in Table 1. Notably, the GEM-MACH (Ops) simulation has the highest V_d while the GEM-MACH (Zhang) simulation has the lowest V_d despite both simulations having one of the thickest first layer heights. This is consistent with the notion that the surface resistance r_c (independent of first layer thickness) rather than the aerodynamic resistance r_a (dependent on first layer thickness) is generally the limiting factor controlling ozone dry deposition.

Table 2. Annual total O_3 dry deposition fluxes and annual mean O_3 dry deposition velocities estimated by the participating models over both the NA and EUR domains.~~Examining~~ These totals are calculated over all non-water grid cells that were common to all simulations. The NA numbers are for 2016 while the EUR number are for 2010.

Figures 3-4 show the spatial patterns of annual mean V_d for each model and the multi-model mean and normalized standard deviation. While these maps visually confirm the results from Table 2 that GEM-MACH (Ops) has the highest mean V_d and GEM-MACH (Zhang) the lowest mean V_d over the NA domain and WRF-Chem (UPM) has the highest and LOTOS-EUROS the lowest mean V_d over the EUR domain, they also show important spatial differences. For example, both GEM-MACH (Base) and WRF-Chem (UPM) have very similar annual mean V_d when averaged over the entire domain (Table 2), but this agreement in the means masks the generally higher V_d in GEM-MACH (Base) over the Southeastern U.S. and the generally lower V_d over the Southwestern U.S. compared to WRF-Chem (UPM).The role of differences in surface deposition pathways and LU distributions in causing model-to-model differences in V_d magnitude and patterns is discussed in subsequent Sections.

Comparing the model-to-model differences in Figures 3-4 to those in the corresponding dry deposition flux maps in Figures 1-2 shows a high level of similarity. ~~in Figures 3-4 and comparing them to the patterns of the corresponding dry deposition flux maps in Figures 1-2 shows a high level of agreement.~~ For example, the areas over the Southeastern U.S. where GEM-MACH (Base) and GEM-MACH (Ops) have larger fluxes than other models coincide with areas where these two simulations also have the highest O_3 V_d. Moreover, the areas with the highest normalized flux standard deviation discussed above for both the NA and EUR domains also show the highest normalized standard deviation for V_d. Despite these generally similarities, it is important to note that models may have above-average O_3 fluxes while also having below-average O_3 V_d or that a given model may have areas with similar V_d but noticeably different dry deposition fluxes, due to the influence of factors other than V_d (e.g., the representation of regional transport and chemistry, the parameterizations for sub-grid-scale turbulence, and the location and density of precursor emission sources) on simulated O_3 mixing ratios. Maps of annual mean O_3 for each model as well as the multi-model mean and standard deviation are shown in Figures S1 – S2 and we refer to Kioutsioukis et al., (2025) for both an operational evaluation of simulated O_3 fields and analyses partitioning variability in these fields to variability in O_3 V_d and other variables such as wind speed and the height of the planetary boundary layer. Makar et al. (2025) showed that WRF-Chem (NCAR) significantly underestimated observed precipitation and Kioutsioukis et al. (2025) hypothesized that a corresponding underestimation of clouds and overestimation of radiation was the main driver for the large positive ozone

bias reported for that model. The above-average O_3 mixing ratio for WRF-Chem (NCAR) shown in Figure S1 is consistent with this hypothesis and may provide at least a partial explanation for the above-average O_3 flux but below-average O_3 V_d shown for this model in Figures 1 and 3. As a second example of confounding factors when comparing spatial patterns of O_3 V_d and O_3 fluxes, the GEM-MACH (Ops) panel in Figure S1 shows generally higher O_3 mixing ratios over the Southeastern U.S. vs. parts of the Canadian boreal forest region, consistent with corresponding spatial differences in O_3 fluxes for this model despite similar O_3 V_d in these two regions. The influence of factors other than V_d in shaping spatial O_3 variability is analyzed in more detail in Kioutsioukis et al., (2025).~~EU domains also show the highest normalized standard deviation for V_d. In some instances, areas may have similar V_d but noticeably different modeled dry deposition fluxes (e.g. the Southeastern U.S. vs. parts of the Canadian boreal forest region for model GEM-MACH (Ops)), suggesting that factors other than V_d, such as the location and density of precursor emission sources as well as other parameterizations for regional transport and chemical lifetime also impact modeled dry deposition flux patterns. GEM-MACH (Ops) and GEM-MACH (Base) share the same gas-phase deposition code, though GEM-MACH (Base) more explicitly simulates transport through the forest canopy (Makar et al., 2017), and incorporates a parameterization for the influence of vehicle motion on vertical diffusive transport of mobile emissions (Makar et al., 2021), both of which have been shown to have a substantial impact on ozone concentrations near the surface.~~ However, despite these potentially confounding influences, the general agreement of both the model rankings for the domain-wide dry deposition fluxes vs. V_d (Table 2~~, Figures S1 vs. S2~~) and their spatial patterns (Figure 1 vs. Figure 3 and Figure 2 vs. Figure 4) confirms the influence of O_3 V_d towards deposition fluxes estimated in these regional-scale simulations. This in turn establishes that a diagnostic understanding of simulated O_3 V_d can aid in interpreting model-to-model differences in dry deposition fluxes.

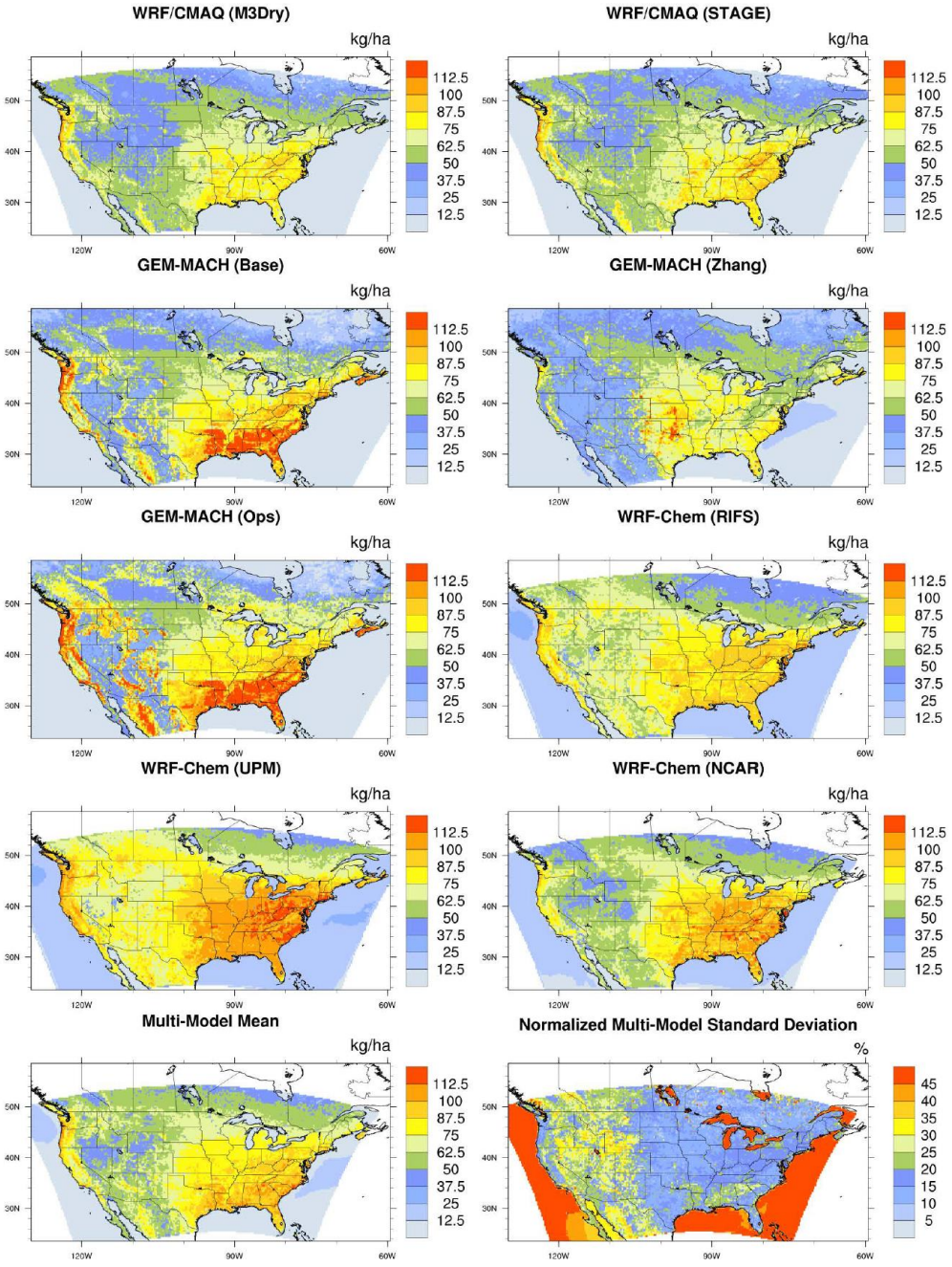


Figure 1. 2016 annual total O3 grid-scale dry deposition fluxes for each model, the multi-model mean, and the normalized multi-model standard deviation over the NA domain. Note that the plots for individual models are not clipped to the domain common to all simulations and show the maximum spatial extent submitted for each model. The multi-model mean and normalized standard deviations are calculated and shown over the common domain.

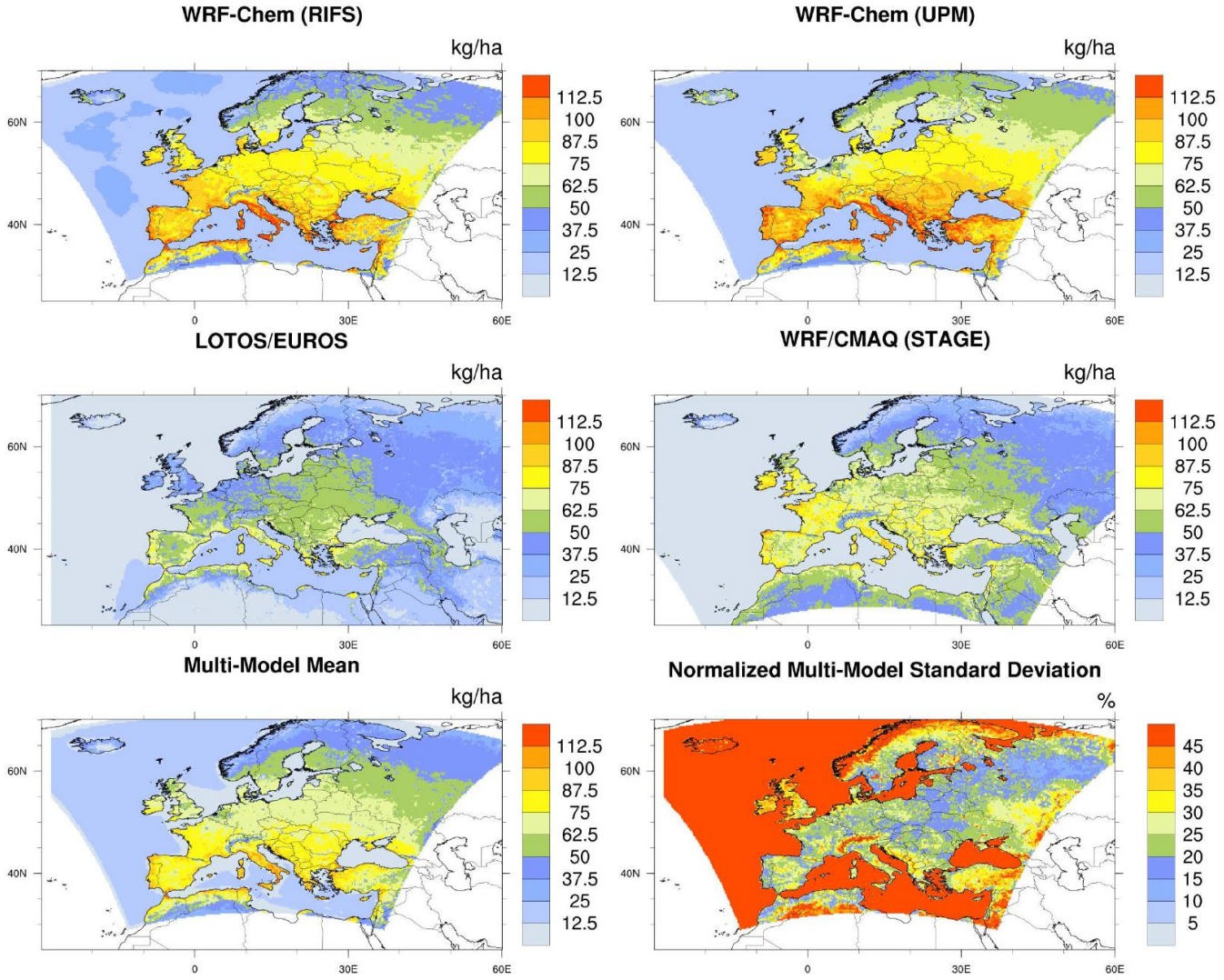


Figure 2. 2010 annual total O3 grid-scale dry deposition fluxes for each model, the multi-model mean, and the normalized multi-model standard deviation over the EUR~~E~~U domain. Note that the plots for individual models are not clipped to the domain common to all simulations and show the maximum spatial extent submitted for each model. The multi-model mean and normalized standard deviations are calculated and shown over the common domain.

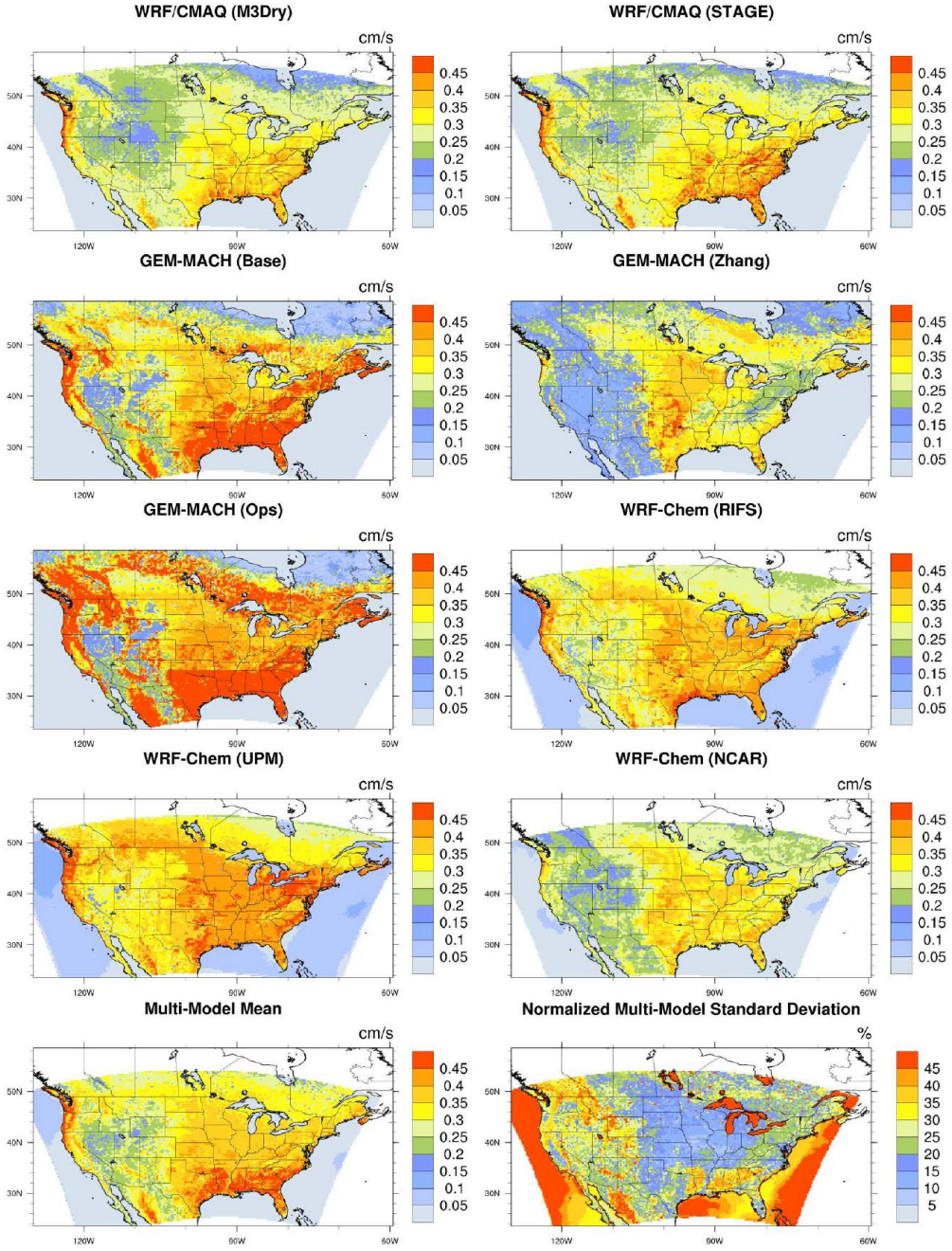


Figure 3. 2016 annual mean O3 grid-scale dry deposition velocities for each model, the multi-model mean, and the normalized multi-model standard deviation over the NA domain. Note that the plots for individual models are not clipped to the domain common to all simulations and show the maximum spatial extent submitted for each model. The multi-model mean and normalized standard deviations are calculated and shown over the common domain.

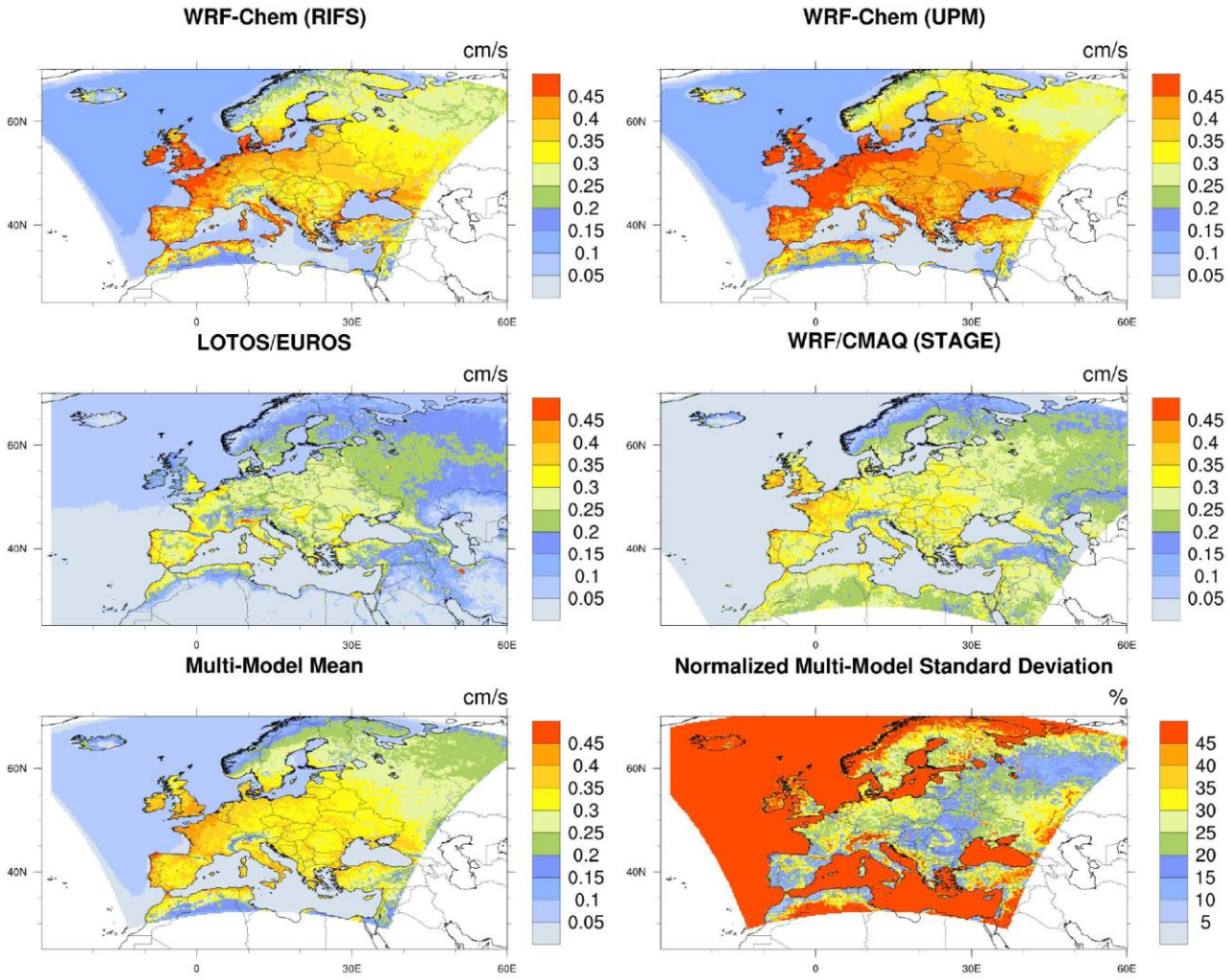


Figure 4. 2010 annual mean O3 grid-scale dry deposition velocities for each model, the multi-model mean, and the normalized multi-model standard deviation over the EUR~~EU~~ domain. Note that the plots for individual models are not clipped to the domain common to all simulations and show the maximum spatial extent submitted for each model. The multi-model mean and normalized standard deviations are calculated and shown over the common domain.

Figures S3~~S1~~ and S4~~S2~~ provide insight into temporal differences in modeled deposition by showing domain-wide deposition fluxes and O_3 V_d calculated for winter vs. summer and daytime vs. nighttime periods for the NA domain. The largest deposition fluxes occur during summer daytime hours while the lowest deposition fluxes occur during winter nighttime hours, a pattern also present for V_d. Photochemical production of O_3 is also higher in summer than in winter, which will also contribute to higher summertime fluxes. Fluxes during winter daytime hours are generally of the same order of magnitude as fluxes during

summer nighttime hours, both times corresponding to low photochemical production or ozone destruction through titration. These figures also suggest that the low annual total deposition value of model GEM-MACH (Zhang) and the high annual total deposition value of model WRF-Chem (UPM) shown in Table 2 were driven by low and high summer daytime flux values, respectively, and/or that other terms such as photochemical production in WRF-Chem (UPM) are enhancing summer ozone concentrations, hence increasing fluxes. While the absolute model-to-model differences are smaller during winter than summer for both fluxes and V_d, the normalized standard deviation is comparable across seasons for both variables (summer daytime flux 18.2%; winter daytime flux 14.5%; summer nighttime flux 22.7%; winter nighttime flux 22.8%; summer daytime V_d 18.6%; winter daytime V_d 17.2%; summer nighttime V_d 34.2%; winter nighttime V_d 28.2%). During daytime, fluxes and V_d exhibit similar relative spread. The larger nighttime spread in V_d compared to the fluxes suggests that nighttime chemical sinks and/or a smaller ozone reservoir in the shallower nighttime mixing layer may be more important factors controlling O_3 concentrations, and hence O_3 deposition fluxes, than the magnitude of the deposition velocity itself. However, it should also be noted that Clifton et al. (2020a) found that even small differences in wintertime V_d can have substantial impacts on the tropospheric O_3 budget given the longer chemical lifetime.

Figures 5-6 depict the contributions of the four effective conductance pathways to the annual domain-average V_d for each model over both domains. The total height of the bars for each model closely corresponds to that model's V_d shown in Table 2, within the caveats discussed in Section 2. Note that some of these simulations (WRF/CMAQ (M3DRY), WRF/CMAQ (STAGE), and GEM-MACH (Zhang)) do not incorporate a separate lower canopy pathway in their dry deposition schemes, and thus only have three deposition pathways in total (Galmarini et al., 2021; Clifton et al., 2023). These Figures show that models with similar V_d (e.g. WRF/CMAQ (M3DRY), GEM-MACH (Zhang), and WRF-Chem (NCAR) over NA) can show significant differences in the absolute and relative contributions of different pathways to V_d. Moreover, these Figures also demonstrate that effective conductances allow an attribution of model differences in V_d to specific processes. For example, GEM-MACH (Base), GEM-MACH (Ops), and WRF-Chem (RIFS) all have similar soil and lower canopy effective conductances, revealing that the differences in V_d between these simulations stem from differences in the cuticular and, to a lesser extent, stomatal pathways. Specifically, both GEM-MACH (Base) and GEM-MACH (Ops) have a higher cuticular effective conductance than WRF-Chem (RIFS). For the two CMAQ simulations over NA, Figure 5 confirms the generally larger contributions from the stomatal and cuticular pathways for WRF/CMAQ (M3DRY) and the generally larger contribution from the soil pathway for WRF/CMAQ (STAGE) that was reported in Hogrefe et al. (2023). Comparing GEM-MACH (Base) with GEM-MACH (Zhang) shows lower contributions from the soil and stomatal pathways for the latter, which also does not include a separate lower canopy pathway.

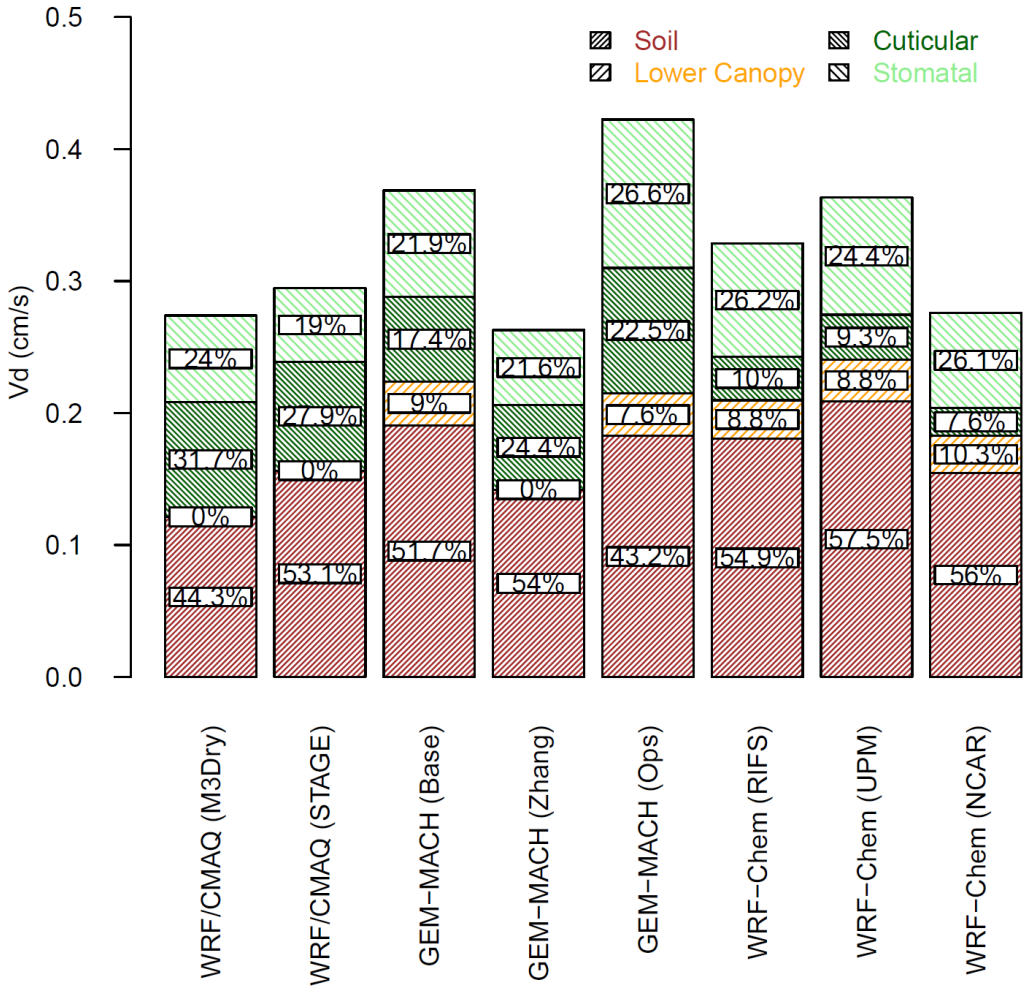


Figure 5. Annual domain-average grid-scale effective conductances and ozone deposition velocities for 2016 over the NA domain. Averages were calculated over all non-water grid cells in the portion of the analysis domain shared by all model simulations.

Comparing the two GEM-MACH simulations with the same dry deposition scheme, GEM-MACH (Base) and GEM-MACH (Ops), reveals that differences other than the dry deposition scheme that exist between these simulations (e.g. meteorology and leaf area index (LAI)) have a noticeable impact on the magnitude of the cuticular and stomatal pathways. It should be noted that both GEM-MACH (Base) and GEM-MACH (Ops) are coupled models with chemistry residing within a weather forecast model – however, GEM-MACH (Base) is "fully coupled", i.e. includes aerosol direct and indirect feedbacks on the predicted meteorology. GEM-MACH (Base) also includes several other parameterizations affecting chemical transport as noted above. The difference in the strength of the cuticle and stomatal conductances between these two models thus reflects differences in the forecasted meteorology in turn influencing the deposition velocity components.

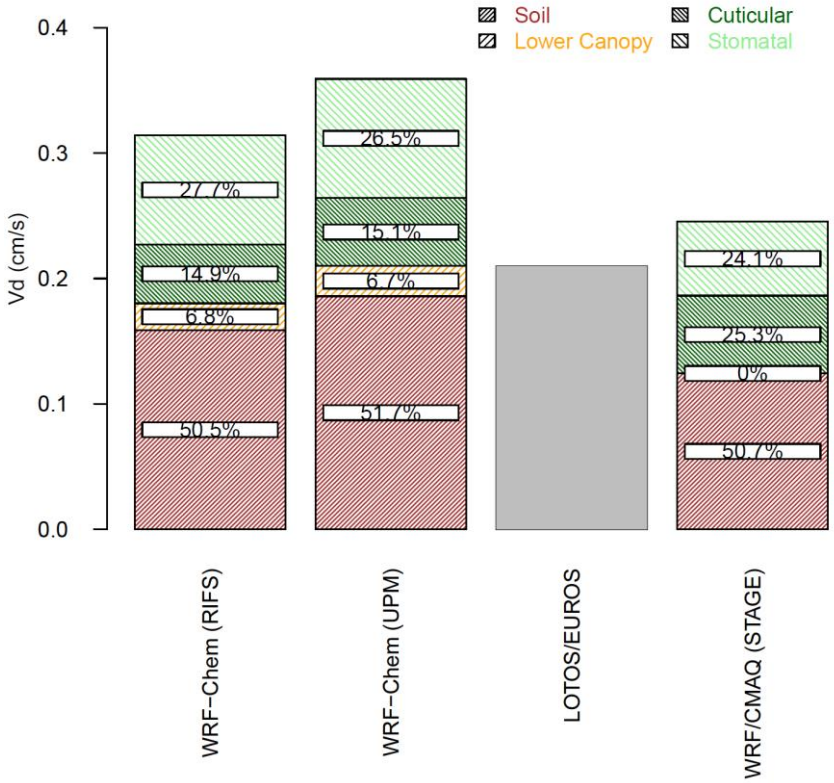


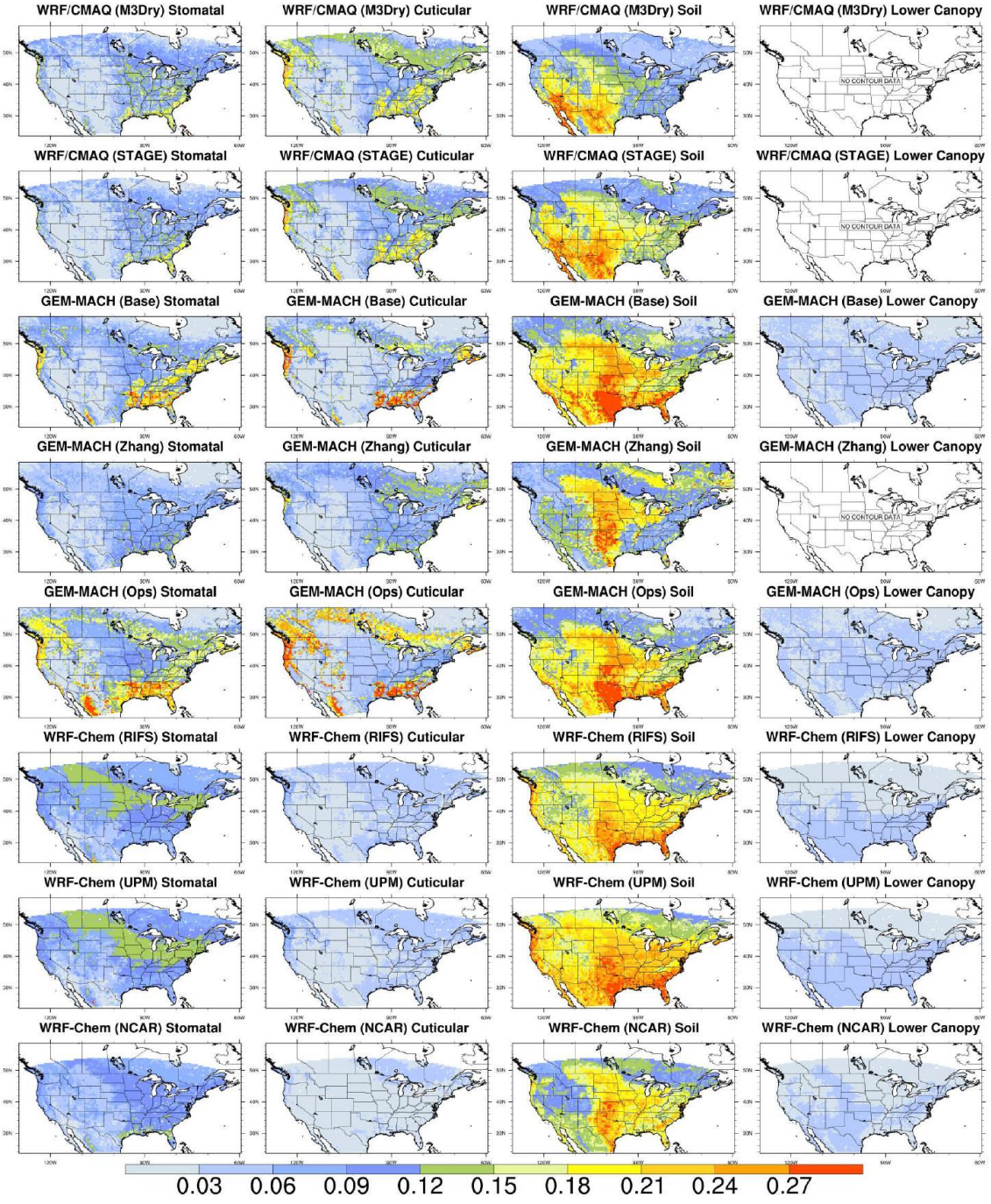
Figure 6. Annual domain-average grid-scale effective conductances and ozone deposition velocities for 2010 over the EUR~~E~~U domain. Averages were calculated over all non-water grid cells in the portion of the analysis domain shared by all model simulations. The bar for LOTOS/EUROS shows only the deposition velocity since no effective conductances were reported.

The impact of factors other than the dry deposition scheme itself on simulated V_d and pathway contributions is further illustrated by the comparison of the three WRF-Chem simulations performed over NA and the two WRF-Chem simulations performed over EUR~~E~~U. These simulations all use the WRF-Chem implementation of the Wesely scheme and as a result all share similar relative pathway contributions to V_d, but V_d itself shows a variation of roughly 25% over NA and 15% over EUR~~E~~U between these simulations. All three of these models make use of feedbacks between aerosols and meteorology – and hence other parameterization differences in addition to the gas-phase deposition code influence the resulting gas-phase deposition velocities, through changes in the predicted temperature, relative humidity, and the other meteorological terms influencing deposition velocity.

Figures 7-8 depict maps of absolute pathway contributions to annual mean V_d over both domains, where the mean is from the average of the monthly median values. The maps show the lower contributions from the cuticular pathway for WRF-Chem compared to other simulations, especially over the Eastern U.S. and Central Europe. This is consistent with the domain-average contributions shown in Figure 5. For the models that include the lower canopy pathway as a separate term from the other

pathways (all WRF-Chem simulations and two of the GEM-MACH simulations), its contribution to V_d is less than 0.06 cm/s throughout both domains (10% or less of the domain-wide annual means as shown in Figures 5-6). Both CMAQ simulations and the GEM-MACH (Zhang) simulation tend to have stronger longitudinal gradients over the NA domain for the soil pathway contribution compared to other simulations. The two GEM-MACH simulations using the Wesely dry deposition scheme (GEM-MACH (Base) and GEM-MACH (Ops)) and the WRF-Chem (CIFS) and WRF-Chem (UPM) simulations have noticeably higher absolute soil pathway contributions in the South-Central and Southeastern U.S. than the other simulations, indicating that the higher values in the domain-average soil pathway contributions for these four simulations shown in Figure 3 originate in these regions.

Figures S5 – S6~~S3 – S4~~ show the relative, rather than absolute, contribution of the four pathways to annual mean V_d over both domains while corresponding maps for summer and winter over NA are included as Figures S7 – S8~~S5 – S6~~. These figures show that the WRF-Chem simulations tend to have a weaker spatial variation than the WRF/CMAQ and GEM-MACH simulations in the split between different pathways over both domains, especially for the relative contribution of the soil pathway. For example, the dominance of the stomatal and cuticular pathways over the soil pathway over the Eastern U.S., and of the soil pathway over the stomatal and cuticular pathways in the Western U.S. that is simulated by WRF/CMAQ and GEM-MACH is less pronounced in the WRF-Chem simulations. Over the EUR~~EU~~ domain, the WRF-Chem simulations show a larger relative contribution of the soil pathway than the WRF/CMAQ simulation throughout much of Central Europe. For the NA domain, the relative contributions of the cuticular component is highest for WRF/CMAQ over the Northeastern U.S., followed by GEM-MACH and then WRF-Chem. These features are especially pronounced during summer (Figure S7~~S5~~). For the EUR~~EU~~ domain, the relative contribution of the stomatal component is highest for WRF/CMAQ over eastern Central Europe and exceeds that of the WRF-Chem simulations.



365 **Figure 7.** Absolute grid-scale ozone effective conductances (cm/s), averaged over the entire year. Results are for the NA domain during 2016. Note that these maps are not clipped to the domain common to all simulations and show the maximum spatial extent of non-water cells submitted for each model.

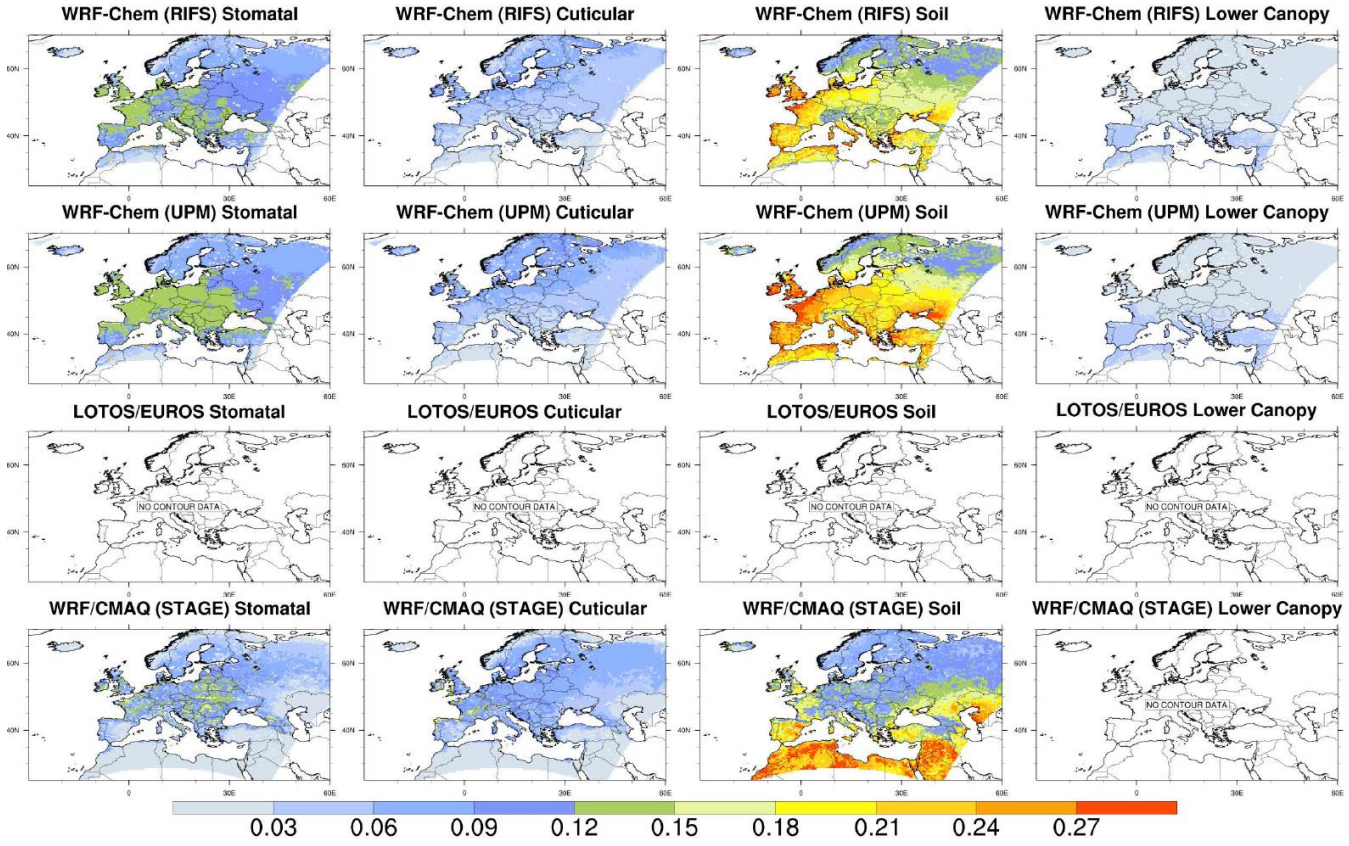


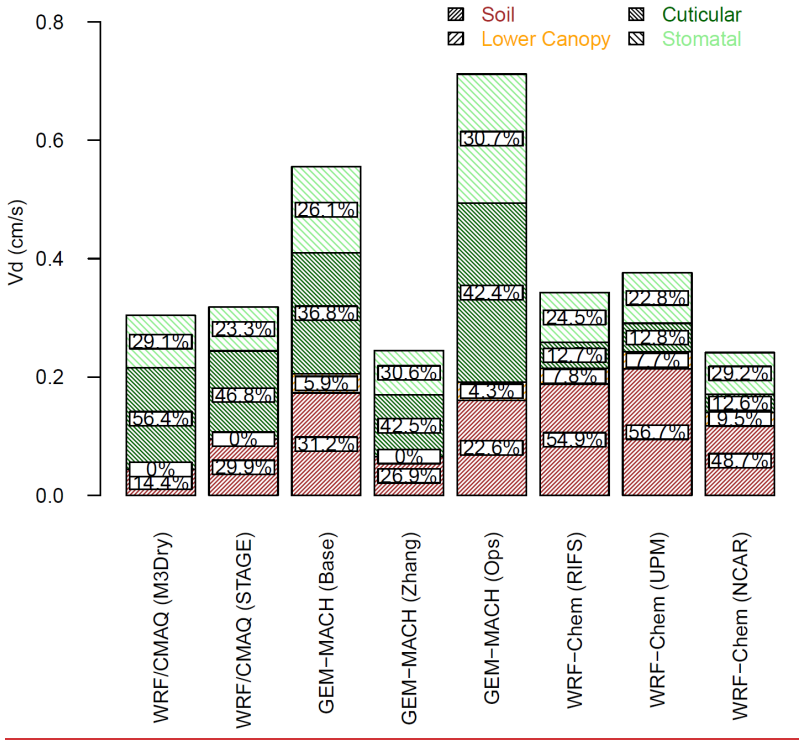
Figure 8: Absolute grid-scale ozone effective conductances (cm/s), averaged over the entire year. Results are for the NA domain during 2016. Note that these maps are not clipped to the domain common to all simulations and show the maximum spatial extent of non-water cells submitted for each model.

3.2 LU-Specific Results

~~Because parameter choices in dry deposition schemes are often tied to LU, the~~ model-to-model differences in domain-average grid-aggregated V_d and pathway contributions may result not only from different process representations, but also different LU spatial distributions and/or LU-dependent parameter and variable choices. It is important to note that the impacts of LU on dry deposition calculations can be both direct (e.g. the incorporation of LU-dependent LAI in some schemes' stomatal and/or cuticular resistance formulations) and indirect and can also vary across schemes depending on the specific formulations of component resistances (Clifton et al., 2023). An example of indirect impacts of LU is the effect of LU-dependent LAI on the calculation of air temperature in the LSM and its subsequent impact on stomatal resistance even in schemes without a direct dependence of stomatal resistance on LAI. Another example is the effect of LU on the calculation of ground temperature and relative humidity in the LSM, which are then used in the formulation of cuticular resistance. Thus, examining LU specific

parameters and variables (e.g., roughness length, LAI) and using them to interpret model to model differences in V_d and effective conductances is not straightforward, and untangling such direct and indirect effects across the different modeling systems is beyond the scope of the AQMEII4 grid model intercomparison activity. Instead~~choices. To investigate these effects,~~ AQMEII4 collected V_d and effective conductances for 16 standardized LU categories in addition to the grid-aggregated values analyzed above. These LU-specific fields allow us to investigate the total (both direct and indirect) impacts of LU-dependent process representations and LU distributions on modeled deposition by stratifying our analyses by LU. Put differently, in this Section we treat LU as a proxy for all LU-dependent processes and parameters, meaning that the dependence of model results on LU in this Section should be interpreted as being due to "all land-use-dependent quantities in the deposition algorithms used in AQMEII4 models.

Figure 9 shows the contributions of the four effective conductance pathways to the annual mean V_d for evergreen needleleaf forest for each model over the NA domain, averaged over only those grid cells where~~were~~ a given model had 85% coverage of evergreen needleleaf forest (note that the number of such grid cells differed across models, as discussed in greater detail in Section 3.3).



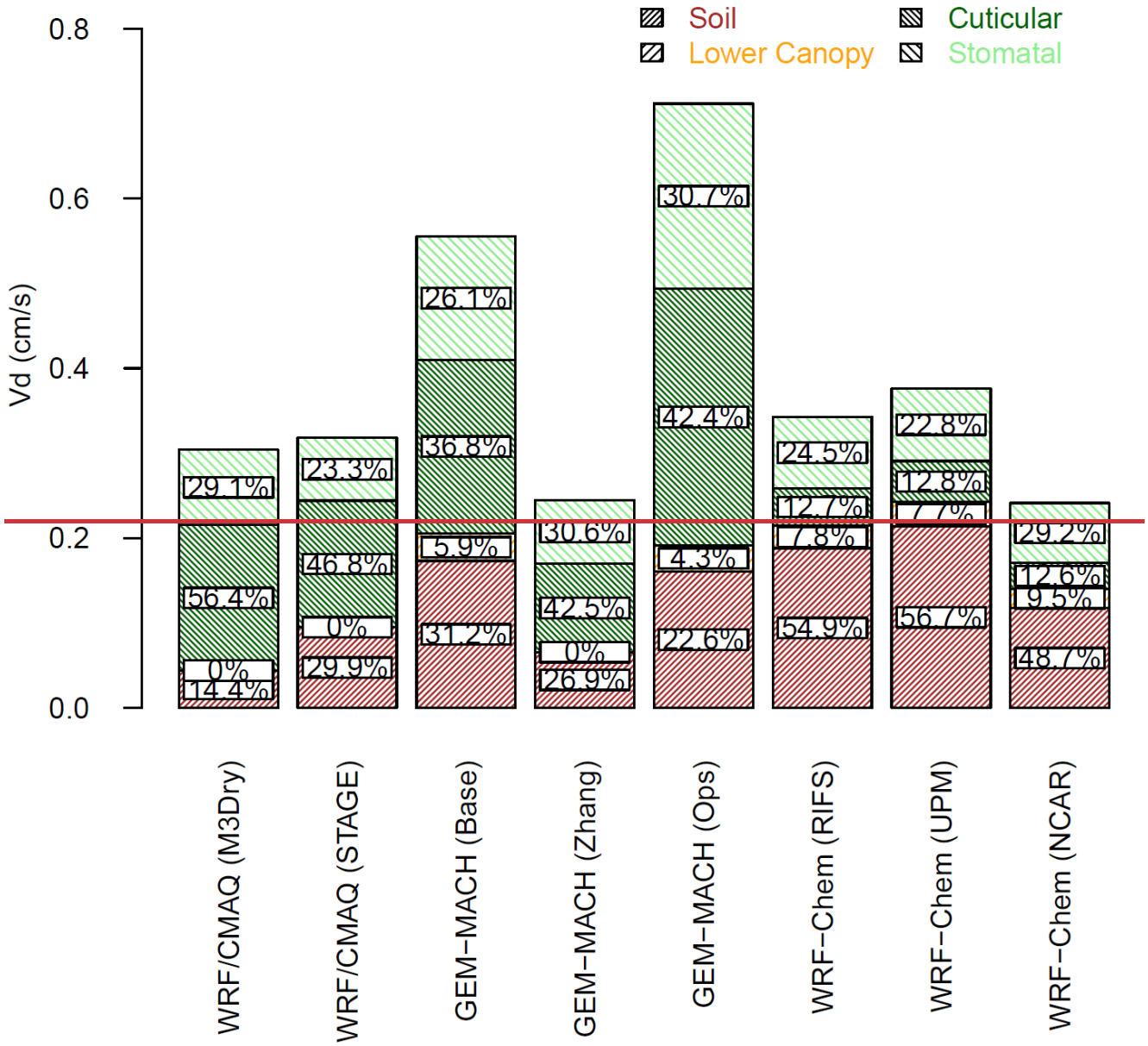


Figure 9. 2016 annual domain-average LU-specific effective conductances and ozone deposition velocities for evergreen needleleaf forest over the NA domain. Averages were calculated over all grid cells in the portion of the analysis domain shared by all model simulations for which a given model had coverage of at least 85% for the evergreen needleleaf forest LU category.

Comparing this Figure to the equivalent grid-aggregated results shown in Figure 5 when all common non-water grid cells were analyzed shows that considering instead only the model grid cells containing a specific LU type can increase model spread. For example, annual mean grid-aggregated V_d ranges from 0.26 - 0.42 cm/s averaged over all common non-water grid cells over NA (Figure 5) while the range is 0.24 - 0.71 for V_d for grid cells dominated by the evergreen needleleaf forest category (Figure 9). Similar increases can also be seen in the ranges of absolute pathway contributions (e.g. soil effective conductance from 0.12 – 0.21 cm/s in Figure 5 to 0.04 – 0.21 cm/s in Figure 9) and relative pathway contributions (e.g. cuticular conductance from 7.6% to 31.7% in Figure 5 to 12.6% – 56.4% in Figure 9). Results of individual and summed pathways contributions for a total of eight LU categories over both domains are shown in Tables 3a-3e. These results confirm the general increase in model spread of both V_d and pathway contributions when considering LU-specific diagnostics, especially for V_d and the

	Gridscale	URB	BAR	ENF	DBF	MF	SHR	AGR	GRA
WRF/CMAQ (M3Dry)	0.28	0.21	**0.25**	0.3	0.37	0.27	*0.27*	*0.28*	*0.21*
WRF/CMAQ (STAGE)	0.3	**0.27**	0.24	0.32	0.42	0.31	*0.27*	0.3	0.25
GEM-MACH (Base)	0.36	0.26	0.21	0.55	**0.48**	0.39	0.35	0.4	0.36
GEM-MACH (Zhang)	*0.26*	*0.17*	0.14	*0.24*	*0.21*	0.27	0.34	0.33	0.34
GEM-MACH (Ops)	**0.41**	0.26	0.21	**0.71**	0.45	**0.49**	**0.41**	**0.42**	**0.36**
WRF-Chem (RIFS)	0.33	0.18	0.13	0.34	0.37	0.28	0.28	0.37	0.32
WRF-Chem (UPM)	0.36	0.2	0.16	0.38	0.42	0.31	0.31	0.41	0.34
WRF-Chem (NCAR)	0.27	0.2	*0.12*	*0.24*	0.32	*0.26*		0.35	0.27
Range NA	0.15	0.1	0.13	0.47	0.27	0.23	0.14	0.14	0.15
WRF-Chem (RIFS)	0.32	*0.18*	*0.14*	0.31	0.38	0.26	**0.3**	0.36	0.3
WRF-Chem (UPM)	**0.36**	0.19	0.17	**0.33**	**0.42**	**0.29**	0.29	**0.42**	**0.31**
LOTOS/EUROS									
WRF/CMAQ (STAGE)	*0.24*	**0.26**	**0.25**	*0.22*	*0.29*	*0.24*	*0.22*	*0.29*	*0.2*
Range EUR	0.12	0.08	0.11	0.11	0.13	0.05	0.08	0.13	0.11

~~cuticular and stomatal effective conductances over forested and agricultural LU types. Such increased heterogeneity of process-level diagnostics when considering locations corresponding to specific LU categories compared to locations representing a mix of LU categories is also reported in Kioutsioukis et al. (2025). This finding suggests that LU-dependent parameter choices~~

	Gridscale	URB	BAR	ENF	DBF	MF	SHR	AGR	GRA
WRF/CMAQ (M3Dry)	0.07	*0*	0	0.09	0.13	0.09	0.01	0.09	*0.02*
WRF/CMAQ (STAGE)	*0.06*	*0*	0	*0.07*	0.1	0.08	*0*	*0.07*	*0.02*
GEM-MACH (Base)	0.08	**0.01**	0	0.15	**0.18**	0.1	0.05	*0.07*	0.04
GEM-MACH (Zhang)	*0.06*	*0*	0	*0.07*	*0.08*	*0.07*	0.05	*0.07*	**0.13**
GEM-MACH (Ops)	**0.11**	**0.01**	0	**0.22**	0.16	**0.15**	**0.1**	0.09	0.04
WRF-Chem (RIFS)	0.09	*0*	0	0.08	0.12	0.09	0.05	0.11	0.06
WRF-Chem (UPM)	0.09	*0*	0	0.09	0.13	0.1	0.04	**0.12**	0.05
WRF-Chem (NCAR)	0.07	*0*	0	*0.07*	*0.09*	0.09		0.1	0.05
Range NA	0.05	0.01	0	0.15	0.1	0.08	0.1	0.05	0.11

	~~Gridscale~~	~~URB~~	~~BAR~~	~~ENF~~	~~DBF~~	~~MF~~	~~SHR~~	~~AGR~~	~~GRA~~
~~WRF/CMAQ (M3Dry)~~ WRF-Chem (RIFS)	~~**0.09**~~	~~*0.21*~~*0*	~~***0.20***~~**0**	~~**0.07**~~	~~*0.12*~~	~~*0.08*~~	~~**0.04**~~	~~**0.12**~~	~~**0.06**~~
~~WRF/CMAQ (STAGE)~~ WRF-Chem (UPM)	~~**0.1**~~	~~***0.27***~~*0*	~~*0.20*~~**0**	~~**0.07**~~	~~*0.13*~~	~~*0.09*~~	~~**0.04**~~	~~*0.12*~~	~~**0.06**~~
~~GEM-MACH (Base)~~ LOTOS/EUROS	~~0.36~~	~~0.26~~	~~0.21~~	~~0.55~~	~~**0.48**~~	~~0.39~~	~~0.35~~	~~0.4~~	~~0.36~~
~~GEM-MACH (Zhang)~~ WRF/CMAQ (STAGE)	~~*0.06*~~	~~*0.01*~~**0.01**	~~*0.19*~~**0**	~~*0.06*~~	~~*0.25*~~	~~*0.07*~~	~~*0.01*~~	~~*0.09*~~	~~*0.01*~~
~~GEM-MACH (Ops)~~ Range ~~EURED~~	~~**0.04**~~	~~*0.01*~~	~~*0.20*~~**0**	~~**0.01**~~	~~*0.08*~~	~~*0.02*~~	~~**0.03**~~	~~**0.03**~~	~~**0.05**~~
~~WRF-Chem (RIFS)~~	~~0.33~~	~~0.18~~	~~0.13~~	~~0.34~~	~~0.37~~	~~0.28~~	~~0.28~~	~~0.37~~	~~0.32~~
~~WRF-Chem (UPM)~~	~~0.36~~	~~0.2~~	~~0.16~~	~~0.38~~	~~0.42~~	~~0.31~~	~~0.31~~	~~0.41~~	~~0.34~~
~~WRF-Chem (NCAR)~~	~~0.27~~	~~0.2~~	~~*0.12*~~	~~*0.24*~~	~~0.32~~	~~*0.26*~~		~~0.35~~	~~0.27~~
~~Range NA~~	~~0.15~~	~~0.1~~	~~0.13~~	~~0.47~~	~~0.27~~	~~0.23~~	~~0.14~~	~~0.14~~	~~0.15~~
~~WRF-Chem (RIFS)~~	~~0.32~~	~~*0.18*~~	~~*0.14*~~	~~0.31~~	~~0.38~~	~~0.26~~	~~**0.3**~~	~~0.36~~	~~0.3~~
~~WRF-Chem (UPM)~~	~~**0.36**~~	~~0.19~~	~~0.17~~	~~**0.33**~~	~~**0.42**~~	~~**0.29**~~	~~0.29~~	~~**0.42**~~	~~**0.31**~~
~~LOTOS/EUROS~~									
~~WRF/CMAQ (STAGE)~~	~~*0.24*~~	~~**0.26**~~	~~**0.25**~~	~~*0.22*~~	~~*0.29*~~	~~*0.24*~~	~~*0.22*~~	~~*0.29*~~	~~*0.2*~~
~~Range EU~~	~~0.12~~	~~0.08~~	~~0.11~~	~~0.11~~	~~0.13~~	~~0.05~~	~~0.08~~	~~0.13~~	~~0.11~~

Table 3a. Annual mean grid-scale and LU-specific sum of effective conductances (cm/s) for individual simulations as well as the range of values across all simulations. For a given column and domain, the maximum values are shown in bold and the minimum values shown in italics. Blank cells indicate that data for a given model, pathway, and/or LU was not available. The grid-scale results reflect spatial averages over all non-water grid cells in the common analysis domain. The LU-specific results reflect spatial averages over all cells for which that LU fraction exceeded 85% for a given model (note that the number of such cells can vary across models). The LU categories are abbreviated as URB (urban), BAR (barren), ENF (evergreen needleleaf forest), DBF (deciduous broadleaf forest), MF (mixed forest), SHR (shrubland), AGR (planted / cultivated), and GRA (grassland).

Table 3b. As in Table 3a but for the stomatal effective conductances (cm/s).

	Gridscale	URB	BAR	ENF	DBF	MF	SHR	AGR	GRA
WRF/CMAQ (M3Dry)	*0.12*	0.21	**0.25**	*0.04*	*0.05*	*0.04*	0.25	*0.12*	0.17
WRF/CMAQ (STAGE)	0.16	**0.27**	0.24	0.1	0.13	0.1	**0.26**	0.16	0.21
GEM-MACH (Base)	0.19	0.24	0.2	0.17	0.15	0.12	0.23	**0.26**	**0.23**
GEM-MACH (Zhang)	0.14	*0.15*	0.14	0.07	0.07	0.11	0.22	0.21	*0.16*
GEM-MACH (Ops)	0.18	0.24	0.2	0.16	0.15	0.11	0.22	0.25	**0.23**
WRF-Chem (RIFS)	0.18	0.18	0.13	0.19	0.19	0.13	*0.17*	0.2	0.2
WRF-Chem (UPM)	**0.21**	0.2	0.16	**0.21**	**0.22**	**0.14**	0.21	0.23	0.22
WRF-Chem (NCAR)	0.15	0.2	*0.12*	0.12	0.17	0.11		0.2	0.17
Range NA	0.09	0.12	0.13	0.17	0.17	0.1	0.09	0.14	0.07
WRF-Chem (RIFS)	0.16	*0.16*	*0.13*	0.15	0.18	0.11	0.19	0.18	*0.17*
WRF-Chem (UPM)	**0.19**	0.17	0.17	**0.16**	**0.21**	**0.12**	*0.18*	**0.22**	**0.18**
LOTOS/EUROS									
WRF/CMAQ (STAGE)	*0.12*	**0.24**	**0.25**	*0.08*	*0.17*	*0.09*	**0.2**	*0.14*	**0.18**
Range EUR~~EU~~	0.07	0.08	0.12	0.08	0.04	0.03	0.02	0.08	0.01

Table 3c. As in Table 3a but for the cuticular effective conductances (cm/s).

	Gridscale	URB	BAR	ENF	DBF	MF	SHR	AGR	GRA
WRF/CMAQ (M3Dry)	**0.09**	*0*	0	0.17	**0.19**	0.14	*0.01*	**0.07**	*0.02*
WRF/CMAQ (STAGE)	0.08	0	0	0.15	**0.19**	0.13	*0.01*	**0.07**	*0.02*
GEM-MACH (Base)	0.06	0	0	0.2	0.1	0.12	0.02	0.03	0.04
GEM-MACH (Zhang)	0.06	**0.02**	0	0.1	0.06	0.09	**0.07**	0.05	**0.05**
GEM-MACH (Ops)	**0.09**	*0*	0	**0.3**	0.09	**0.19**	0.05	0.04	**0.05**
WRF-Chem (RIFS)	0.03	*0*	0	0.04	0.03	0.04	0.02	0.03	0.03
WRF-Chem (UPM)	0.03	*0*	0	0.05	0.03	0.04	0.02	0.03	0.03
WRF-Chem (NCAR)	*0.02*	*0*	0	*0.03*	*0.02*	*0.03*		*0.02*	*0.02*
Range NA	0.07	0.02	0	0.27	0.17	0.16	0.06	0.05	0.03
WRF-Chem (RIFS)	*0.05*	**0.02**	**0.01**	*0.07*	*0.05*	*0.05*	0.03	*0.04*	**0.04**
WRF-Chem (UPM)	*0.05*	**0.02**	*0*	**0.08**	*0.05*	0.06	**0.04**	0.05	**0.04**
LOTOS/EUROS									
WRF/CMAQ (STAGE)	**0.06**	*0.01*	*0*	**0.08**	**0.07**	**0.08**	*0.01*	**0.06**	*0.01*
Range EUR~~EU~~	0.01	0.01	0.01	0.01	0.02	0.03	0.03	0.02	0.03

Table 3d. As in Table 3a but for the soil effective conductances (cm/s).

Table 3e. As in Table 3a but for the lower canopy effective conductances (cm/s).

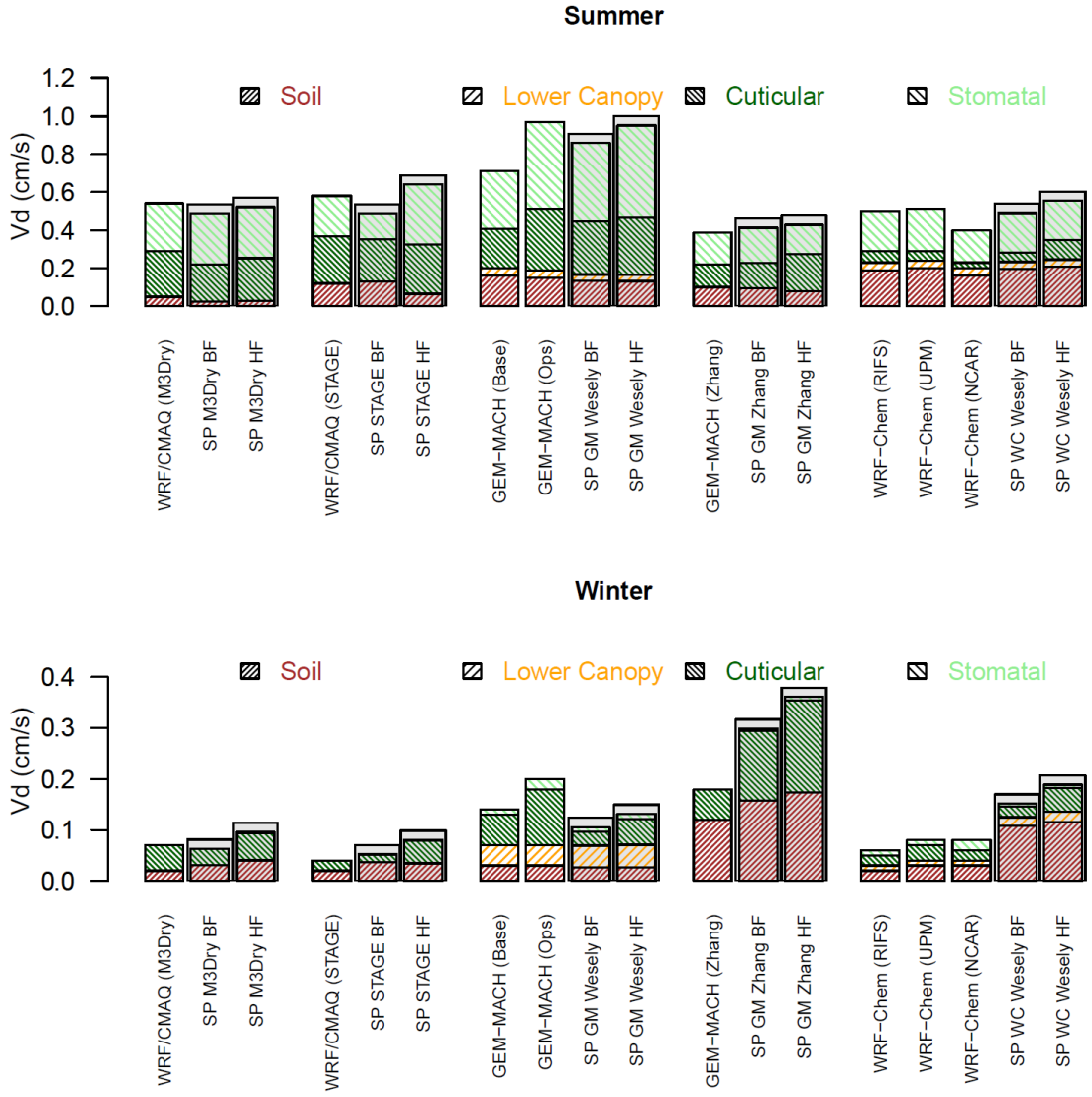
	Gridscale	URB	BAR	ENF	DBF	MF	SHR	AGR	GRA
WRF/CMAQ (M3Dry)									
WRF/CMAQ (STAGE)									
GEM-MACH (Base)	**0.03**	**0.01**	**0.01**	**0.03**	**0.05**	**0.05**	**0.05**	**0.04**	**0.05**
GEM-MACH (Zhang)									
GEM-MACH (Ops)	**0.03**	**0.01**	**0.01**	**0.03**	**0.05**	0.04	*0.04*	**0.04**	0.04
WRF-Chem (RIFS)	**0.03**	*0*	*0*	**0.03**	*0.03*	*0.02*	*0.04*	*0.03*	*0.03*
WRF-Chem (UPM)	**0.03**	*0*	*0*	**0.03**	0.04	0.03	*0.04*	*0.03*	0.04
WRF-Chem (NCAR)	**0.03**	*0*	*0*	*0.02*	0.04	0.03		*0.03*	*0.03*
Range NA	0	0.01	0.01	0.01	0.02	0.03	0.01	0.01	0.02
WRF-Chem (RIFS)	**0.02**	**0**	**0**	**0.02**	**0.03**	**0.02**	**0.04**	*0.02*	**0.03**
WRF-Chem (UPM)	**0.02**	**0**	**0**	**0.02**	**0.03**	**0.02**	*0.03*	**0.03**	**0.03**
LOTOS/EUROS									
WRF/CMAQ (STAGE)									
Range EUR~~EU~~	0	0	0	0	0	0	**0.01**	0.01	0

cuticular and stomatal effective conductances over forested and agricultural LU types. Such increased heterogeneity of process-level diagnostics when considering locations corresponding to specific LU categories compared to locations representing a mix of LU categories is also reported in Kioutsioukis et al. (2025). This finding suggests that LU-dependent parameter choices and the representation of vegetation effects on O_3 dry deposition (e.g. whether or not a given scheme accounts for soil moisture effects on stomatal conductance) are a significant source of variability in grid-aggregated V_d while also indicating that an analysis of only grid-aggregated deposition diagnostics may partially mask the effects of process-specific differences that exist between schemes. However, the attribution of LU-specific diagnostics to LU-dependent parameter choices is complicated by potential model-to-model differences in LU distribution – for example, differences in the number and location of grid cells with a given LU type in a given modeling system may cause differences in macro-scale meteorological variables like wind speed and solar radiation that affect the deposition calculations in this LU-specific analysis. Differences in LU distributions between models are analyzed in Section 3.3.

The availability of LU-specific dry deposition diagnostics from the AQMEII4 grid models also provides an opportunity to compare these diagnostics to the results from the point model intercomparison study by Clifton et al. (2023). At each of the eight O_3 flux measurement sites examined in Clifton et al. (2023), the point model simulations were constrained to use a common set of site-specific variables like LAI, roughness length, reference height, and soil moisture. Figures 10 and S6 – S8 show examples of this comparison. In these Figures, bars showing the single point model results from Clifton are prefixed with the label "SP" to more easily distinguish them from bars showing the grid model results and the point model labels also

use abbreviations for GEM-MACH and WRF-Chem as noted in the Figure caption. Figure 10 compares winter and summer average V_d and effective conductances simulated by the grid models for grid cells with mixed forest coverage greater than 85% against the corresponding results for single point models at the two mixed forest sites analyzed in Clifton et al. (2023), i.e. Borden Forest and Harvard Forest. The point model results are identical to those shown in Figure 5 of Clifton et al. (2023), but while 18 point simulations were included in that Figure, only the five simulations corresponding to the schemes appearing in Clifton et al (2023) that were also implemented in the AQMEII4 grid models (CMAQ M3Dry, CMAQ STAGE, GEM-MACH Wesely, GEM-MACH Zhang, and WRF-Chem Wesely) are reproduced here. Also note that while the seasonal grid model values shown in Figure 10 are derived for a single year (2016) but averaged over all grid cells for which a given model had a fractional coverage of mixed forest exceeding 85%, the point model values at Borden Forest and Harvard Forest are multi-year means at single sites. The motivation for performing this comparison despite these differences in spatio-temporal aggregation is to assess to which extent the conclusions of a single-point modeling study are consistent with results obtained from grid model deposition diagnostics and could therefore inform grid model development by providing process-level insights.

A comparison of the summertime grid model and point model results in Figure 10 leads to similar conclusions regarding the magnitudes of simulated V_d and pathway contributions across models. For example, both grid model and point model V_d range between 0.4 and 1.0 cm s^{-1} across models. The lowest V_d for the point models is simulated by GEM-MACH Zhang while the



~~highest V_d is simulated by GEM-MACH Wesely. This is consistent with the grid model results, with GEM-MACH (Zhang) showing lower V_d than all other grid models and GEM-MACH (Base) and GEM-MACH (Ops) showing higher V_d than all other grid models. Grid and point models also agree that CMAQ M3Dry has a much smaller contribution of the soil effective conductance to V_d compared to all other models and that WRF-Chem Wesely has a smaller contribution of the cuticular effective conductance to V_d compared to all other models.~~

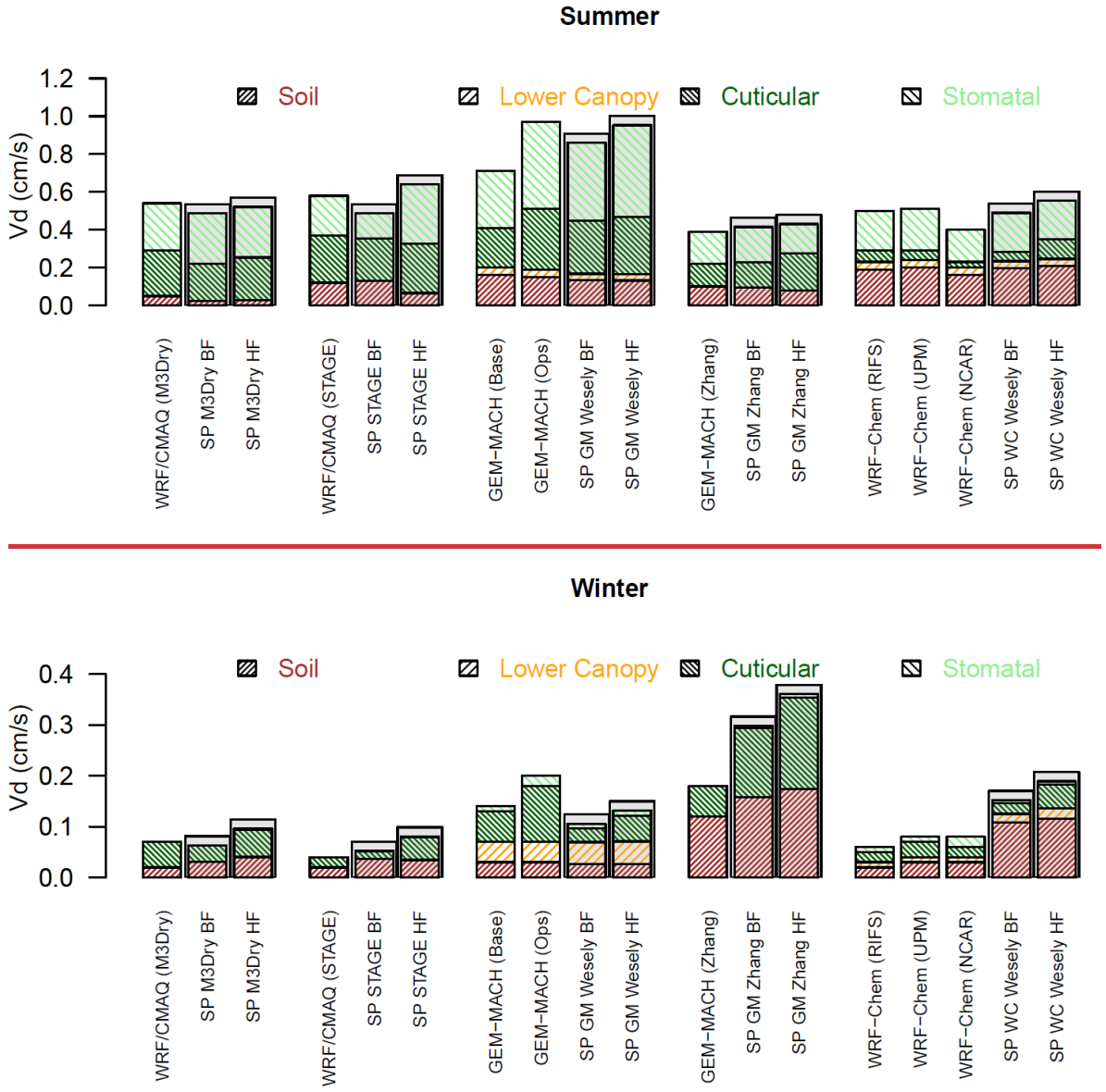


Figure 10. Summer and winter effective conductances and ozone deposition velocities calculated by the grid models for mixed forest grid cells and calculated by the corresponding subset of single point (SP) models analyzed in Clifton et al. (2023) at the Borden Forest (BF) and Harvard Forest (HF) sites. The bars for the SP models are overlayed on grey boxes to visually distinguish them from the bars representing grid models. In the x-axis labels, results for the SP GEM-MACH Wesely and Zhang simulations are shown as "SP GM Wesely" and "SP GM Zhang", respectively, while results for the SP WRF-Chem Wesely simulations are shown as "SP WC Wesely". The mixed forest grid cells selected for this analysis are those in which

a given model had at least 85% coverage for this LU category. The number of these grid cells differs across models due to underlying differences in LU (see Section 3.3).

highest V_d is simulated by GEM-MACH Wesely. This is consistent with the grid model results, with GEM-MACH (Zhang) showing lower V_d than all other grid models and GEM-MACH (Base) and GEM-MACH (Ops) showing higher V_d than all other grid models. Grid and point models also agree that CMAQ M3Dry has a much smaller contribution of the soil effective conductance to V_d compared to all other models and that WRF-Chem Wesely has a smaller contribution of the cuticular effective conductance to V_d compared to all other models.

During winter, the point and grid model results show consistency in terms of relative model rankings (GEM-MACH Zhang and the corresponding grid model simulations GEM-MACH (Base) and GEM-MACH (Ops) have the highest V_d while the CMAQ STAGE point and grid model simulations have the lowest V_d), the magnitude of V_d (ranging from about 0.1 to about 0.35 cm s^{-1}), and in terms of model-to-model variations in pathway contributions. However, the grid model results show generally lower V_d and a different ranking than the point model results, with GEM-MACH Wesely V_d roughly equal to GEM-MACH Zhang in the point model comparison rather than significantly lower when comparing the corresponding grid model simulations GEM-MACH (Base), GEM-MACH (Ops), and GEM-MACH (Zhang). Spatial variations in snow cover across the mixed forest grid cells in the grid models as well as interannual variability in snow cover at the flux measurement sites may play a role in causing these wintertime differences, although Clifton et al. (2023) found that the point models results were not very sensitive to snow cover. The results of the grid model and point model analysis agree on the non-negligible wintertime contribution of the lower canopy pathway for the models that consider it, i.e. the GEM-MACH Wesely (20-40% across the corresponding grid and point models) and WRF-Chem Wesely (10-14% across the corresponding grid and point models) point models and their corresponding grid model implementations.

Figures S9, S10~~S7, S8~~, and S11~~S9~~ show corresponding results for evergreen needleleaf forest, broadleaf deciduous forest, and grassland, respectively. For each of these LU cases, the grid model results reflect simulated LU-specific V_d and effective conductances averaged over all grid cells with coverage greater than 85% for that LU category for a given model. The point model results adapted from Figure 5 of Clifton et al. (2023) are for Hyytiälä (evergreen needleleaf forest), Ispra (deciduous broadleaf forest), and Bugacpuszta and Easter Bush (both grassland)[1]. The results for evergreen needleleaf and deciduous

[1] We note that the LU present at these observation sites may not necessarily represent the dominant land use at the broader grid scale used in this analysis, and this may account for some of the residual differences between the grid model results extracted for specific LU classes and the point model results at these sites. Ispra flux tower (45.812495N, 8.634771E) is within a forest of about 300m extent surrounded by the Joint Research Centre and the town of Ispra, and Easter Bush Field site (55.865N, 3.206W) is located 200m from the University of Edinburgh's Easter Bush Campus. The Hyytiälä site is more representative of the evergreen needleleaf forest on the grid cell scale, with the nearest changes in LU occurring ~800m from

broadleaf forest in Figures S9~~S7~~ and S10~~S8~~ are broadly consistent with those discussed above for mixed forest. In particular, agreement between grid and point modeling results is generally better during summer than winter, especially in terms of the GEM-MACH Wesely vs. GEM-MACH Zhang comparison. The non-negligible contribution of the wintertime lower canopy effective conductance simulated by GEM-MACH Wesely and, to a lesser extent, WRF-Chem Wesely discussed above for mixed forest is also visible in the deciduous broadleaf forest results, for both grid and point models, while it is less pronounced for evergreen needleleaf forest. The model-to-model comparisons for grassland show consistent behavior between the grid model and point model analyses for both winter and summer. During winter GEM-MACH Zhang and WRF-Chem Wesely show the highest V_d while GEM-MACH Wesely shows the lowest V_d. During summer, GEM-MACH Wesely V_d exceeds GEM-MACH Zhang V_d for both the grid model results and both grassland point intercomparison sites, though summertime GEM-MACH Zhang V_d for grassland is not as low relative to other models as for the different forest LU categories discussed above. Moreover, both grid model and point model results for summertime also agree that GEM-MACH Zhang has the largest relative contribution of the stomatal effective conductance to V_d for grassland.

The results presented above demonstrate that despite differences in LU in the region of the observation sites at the scale of model grid cells, the analysis of O_3 dry deposition schemes implemented in grid models can successfully be linked to the detailed point model evaluation using long-term O_3 flux measurements presented in Clifton et al. (2023) when generating LU-specific diagnostic information and limiting the analysis to specific LU categories. This in turn allows the developers of grid models to leverage process-level insights gained from point intercomparison studies (Clifton et al., 2023; Khan et al., 2024) for improving the representation of dry deposition in their modeling systems. In addition, the general agreement between the grid and point model comparisons shown in Figures 10 and S7 – S9 again suggests that the differences in first layer heights between the grid models (Table 1) are not a major factor impacting the interpretation of V_d and effective conductance differences across models since reference and displacement heights had been harmonized in the point model calculations. However, as also discussed in the first part of this section, differences in LU characterizations between models can complicate a process-level attribution of differences across models to differences in deposition schemes, especially when considering LU-specific dry deposition fluxes. Similarly, observation sites with a variety of LU within a short distance of the measurement location itself may be less useful for evaluating LU-specific aspects of deposition algorithms. ~~–~~Model differences in LU distributions and their effects on deposition fluxes are analyzed in the next section.

that site, and Bugacpuszta is also more representative of the larger region, which contains largely farmland/grassland with some trees.

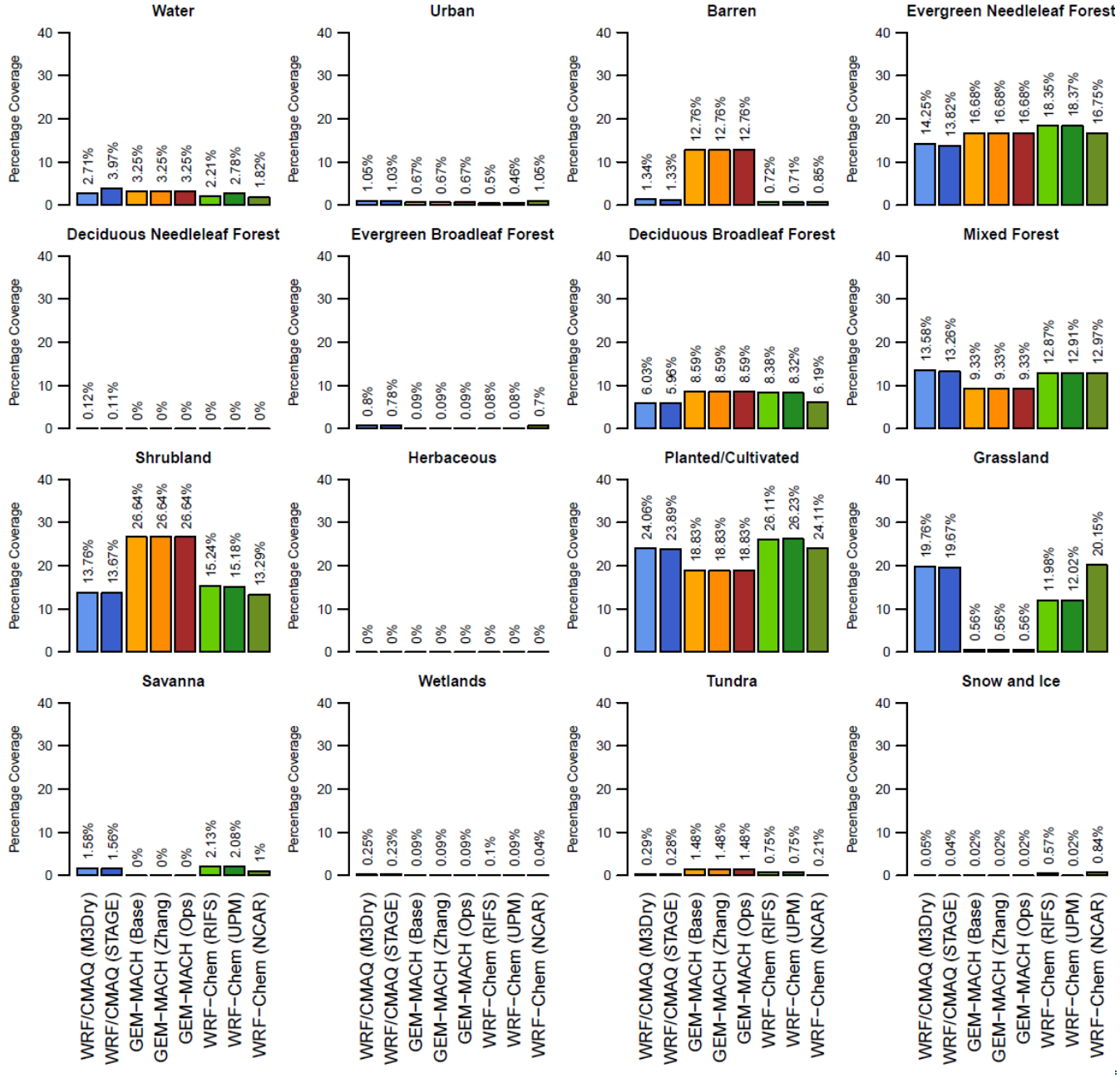
3.3 The Influence of Land Use Data on Ozone Deposition

Table 4 shows the native LU categories used in each model's dry deposition calculations and how these categories were mapped to the 16 common AQMEII4 LU categories defined in Galmarini et al. (2021) when reporting LU-specific diagnostics. For some models (WRF/CMAQ (M3DRY), GEM-MACH (Zhang), WRF-Chem (RIFS), and WRF-Chem (UPM), the native LU categories used in the dry deposition calculations are identical to those used in the land-surface model (LSM) of the driving meteorological model. For the other models, the LU classification scheme differed between the LSM in the driving meteorological model and the dry deposition calculations, and Tables S1 – S5 provide details on the internal mapping between the LSM and dry deposition calculations implemented in these models. We note that several AQMEII4 LU categories are held in common with most model native LU categories (e.g. "evergreen needleleaf trees") while others were less direct, requiring assignment into the nearest AQMEII4 LU category as documented in these Tables.

Figures 11 and 12 show bar charts of the distribution of all 16 AQMEII4 LU categories for each model across all common grid cells that aren't dominated by water for the NA and EUR~~EU~~ domains. These charts reveal that 7 of the 16 categories (barren, evergreen needleleaf forest, deciduous broadleaf forest, mixed forest, shrubland, planted/cultivated land, and grassland) account for roughly 90% of all LU over both NA and EUR~~EU~~. Over NA, the fractional coverage of some categories like evergreen needleleaf forest are similar across models, as might be expected from the commonality of this LU within the native LU categories of Table 4. On the other hand, there is considerable disagreement between models for the fractional coverage of

AQMEII4	WRF/CMAQ (M3Dry)	WRF/CMAQ (STAGE)	GEM-MACH (BASE) and GEM-MACH (Ops)	GEM-MACH (Zhang)	WRF-Chem (RIFS), WRF-Chem (UPM), and WRF-Chem (NCAR)	LOTOS/EUROS
1: Water	17: water	1: Water	13: Inland water 14: Ocean	1: Water (Ocean) 3: Inland Lake (Fresh)	16: Water Bodies	09: Inland Water; 10: Sea Water
2: Developed / Urban	13: Urban and Built-up	2: Developed / Urban	15: Urban	21: Urban	1: Urban and Built-Up Land	08: Urban
3: Barren	16: Barren or Sparsely Vegetated	3: Barren	8: Desert	24: Desert	19: Barren or Sparsely Vegetated; 23: Bare Ground Tundra	05: Bare rock
4: Evergreen needleleaf forest	1: Evergreen Needleleaf Forest	4: Evergreen needleleaf forest	1: Evergreen needle-leaf forest	4: Evergreen needleleaf trees	14: Evergreen Needleleaf Forest	02: Coniferous forest
5: Deciduous needleleaf forest	3: Deciduous Needleleaf Forest	5: Deciduous needleleaf forest	3: Deciduous needle-leaf forest	6: Deciduous needleleaf trees	12: Deciduous Needleleaf Forest	N/A
6: Evergreen broadleaf forest	2: Evergreen Broadleaf Forest	6: Evergreen broadleaf forest	2: Evergreen broadleaf forest	5: Evergreen broadleaf trees 8: Tropical broadleaf trees	13: Evergreen Broadleaf Forest	N/A
7: Deciduous broadleaf forest	4: Deciduous Broadleaf Forest	7: Deciduous broadleaf forest	4: Deciduous broadleaf forest	7: Deciduous broadleaf trees 9: Drought deciduous trees	11: Deciduous Broadleaf Forest	04: Deciduous forest
8: Mixed forest	5: Mixed Forest	8: Mixed forest	5: Mixed Forest	25: Mixed Wood Forest	15: Mixed Forest	N/A
9: Shrubland	6: Closed Shrublands 7: Open Shrublands	9: Shrubland	10: Dwarf trees, shrubs	10: Evergreen broadleaf shrubs 11: Deciduous shrubs 12: Thorn shrubs 26: Mixed shrubs	8: Shrubland 9: Mixed Shrubland/Grassland [WRF-Chem(NCAR) mapping]	N/A
10: Herbaceous	N/A	10: Herbaceous	N/A	13: Short grass and forbs	N/A	N/A
11: Planted / Cultivated	12: Croplands 14: Cropland-Natural Vegetation Mosaic	11: Planted / Cultivated	7: Crops, mixed farming	15: Crops 16: Rice 17: Sugar 18: Maize 19: Cotton 20: Irrigated crops	2: Dryland Cropland and Pasture 3: Irrigated Cropland and Pasture 4: Mixed Dryland/Irrigated Cropland and Pasture 5: Cropland / Grassland Mosaic 6: Cropland / Woodland Mosaic	01: Arable land 03: Permanent crops
12: Grassland	10: Grasslands	12: Grassland	6: Grassland	14: Long grass	7: Grassland	06: Grassland
13: Savanna	8: Woody Savanna 9: Savanna	13: Savanna	N/A	N/A	9: Mixed Shrubland / Grassland [WRF-Chem(RIFS) and WRF-Chem(UPM) mapping] 10: Savanna	11: Semi-natural vegetation
14: Wetlands	11: Permanent Wetlands	14: Wetlands	11: Wetland with plants	22: Swamp	17: Herbaceous Wetland 18: Wooded Wetland	
15: Tundra	18: Wooded Tundra 19: Mixed Tundra 20: Barren Tundra	15: Tundra	9: Tundra	22: Tundra	20: Herbaceous Tundra 21: Wooded Tundra 22: Mixed Tundra	
16: Snow and Ice	15: Snow and Ice	16: Snow and Ice	12: Ice caps and glaciers	2: Ice	24 Snow or Ice	07: Ice

Table 4. Mapping of LU classes from the categories used in the models' dry deposition code to the 16 AQMEII4 categories.



~~Figure 11.~~ ~~Fraction of the NA domain common to all models covered by each AQMEII4 LU category, excluding grid cells dominated by water by each model.~~

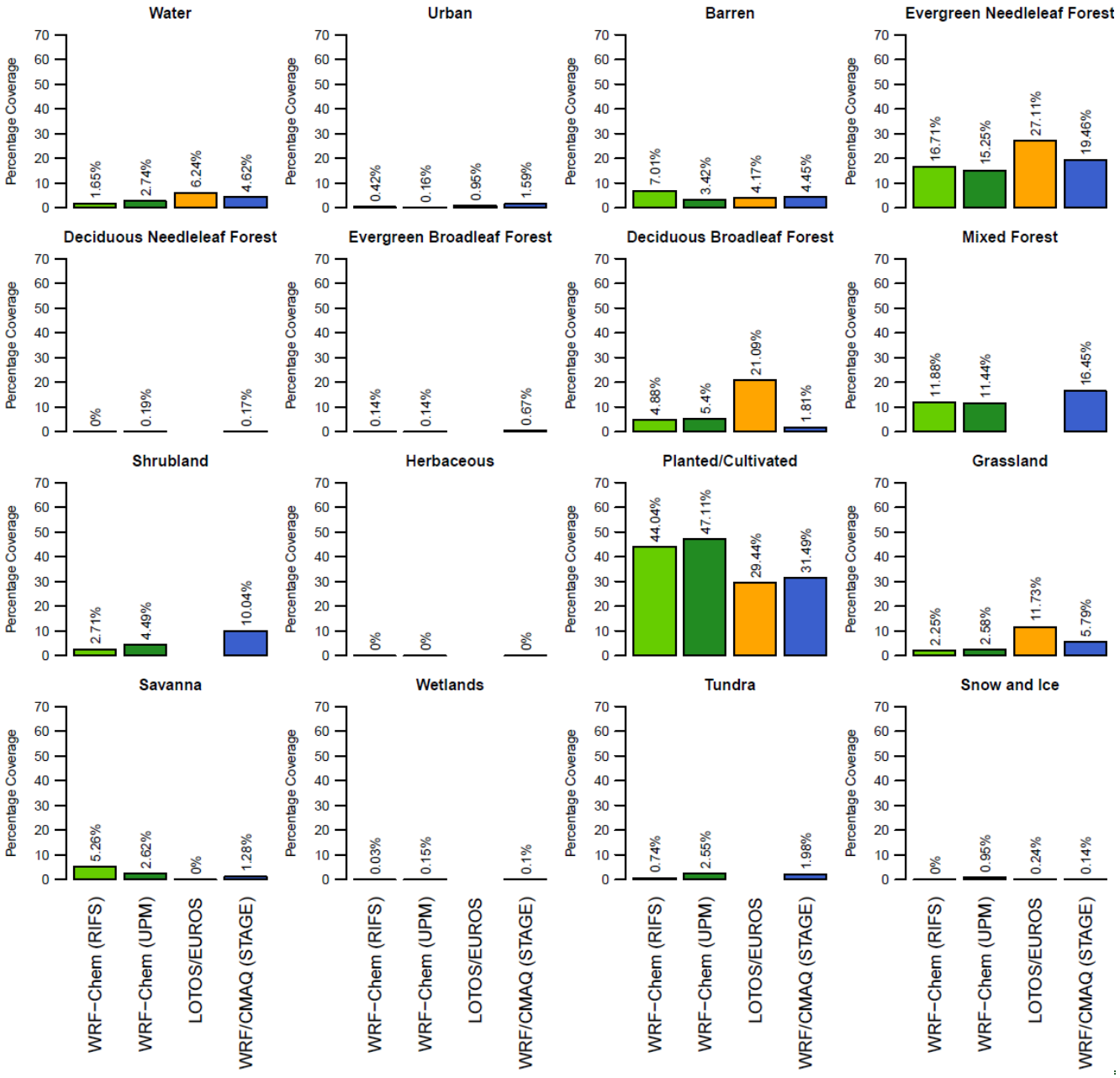


Figure 12. Fraction of the EU domain common to all models covered by each AQMEII4 LU category, excluding grid cells dominated by water by each model.

many other categories over NA, especially non-forest categories such as barren, shrubland, planted/cultivated land, and grassland. Much of this disagreement is caused by the GEM-MACH simulations having larger fractions of the barren and shrubland categories and smaller fractions for the planted/cultivated and grassland categories than either the WRF/CMAQ or

WRF-Chem simulations, pointing to ambiguities in classifying non-forest partially vegetated areas[2]. Over EUR~~EU~~, the primary driver of model variability in the LU used for deposition calculation is LOTOS/EUROS which in its DEPAC dry deposition module splits mixed forest from the driving meteorological model equally between coniferous and deciduous forest and includes shrubland in the category of semi-natural vegetation and as a result shows larger coverage than WRF/CMAQ and WRF-Chem for evergreen needleleaf forest, deciduous broadleaf forest, and grassland. In addition, while all models agree that planted/cultivated is the dominant category over EUR~~EU~~, the lowest coverage (29.44%, LOTOS/EUROS) and highest coverages (47.11%, WRF-Chem (UPM)) differ by a factor of 1.6. Here it should be noted that the WRF-Chem (RIFS) dry deposition calculations indirectly used the CORINE dataset developed for Europe (https://land.copernicus.eu/pan-european/corine-land-cover/, last accessed May 12, 2025) by mapping the CORINE categories to the USGS24 categories in WRF-Chem (RIFS). On the other hand, the WRF-Chem (UPM) dry deposition calculations relied on the global USGS24 dataset, WRF/CMAQ relied on the global MODIS dataset augmented by additional urban categories in the greater London area and mapped these to the AQMEII4 categories as shown in Table S1, and the LOTOS/EUROS dry deposition calculations relied on the Global Landcover 2000 dataset (https://forobs.jrc.ec.europa.eu/glc2000, last accessed May 12, 2025) mapped to the internal DEPAC categories (Manders-Groot et al., 2023~~LOTOS/EUROS is the only model using the official LU data of the European Union (in comparison to USGS for the WRF-CHEM simulations~~).

[2] A subsequent investigation of the GEM-MACH land use database suggested that this difference is due to an underlying dataset which included "short grass and forbs" being aggregated to "Dwarf trees and shrubs", in turn increasing the relative assignment to AQMEII4 shrubland, and decreasing the assignment to AQMEII4 grassland categories. This misassignment has been corrected in implementations of GEM-MACH subsequent to that used in AQMEII4.

AQMEII4	WRF/CMAQ (M3Dry)	WRF/CMAQ (STAGE)	GEM-MACH (BASE) and GEM-MACH (Ops)	GEM-MACH (Zhang)	WRF-Chem (RIFS), WRF-Chem (UPM), and WRF-Chem (NCAR)	LOTOS/EUROS
1: Water	17: water	1: Water	13: Inland water 14: Ocean	1: Water (Ocean) 3: Inland Lake (Fresh)	16: Water Bodies	09: Inland Water; 10: Sea Water
2: Developed / Urban	13: Urban and Built-up	2: Developed / Urban	15: Urban	21: Urban	1: Urban and Built-Up Land	08: Urban
3: Barren	16: Barren or Sparsely Vegetated	3: Barren	8: Desert	24: Desert	19: Barren or Sparsely Vegetated; 23: Bare Ground Tundra	05: Bare rock
4: Evergreen needleleaf forest	1: Evergreen Needleleaf Forest	4: Evergreen needleleaf forest	1: Evergreen needle-leaf forest	4: Evergreen needleleaf trees	14: Evergreen Needleleaf Forest	02: Coniferous forest
5: Deciduous needleleaf forest	3: Deciduous Needleleaf Forest	5: Deciduous needleleaf forest	3: Deciduous needle-leaf forest	6: Deciduous needleleaf trees	12: Deciduous Needleleaf Forest	N/A
6: Evergreen broadleaf forest	2: Evergreen Broadleaf Forest	6: Evergreen broadleaf forest	2: Evergreen broadleaf forest	5: Evergreen broadleaf trees 8: Tropical broadleaf trees	13: Evergreen Broadleaf Forest	N/A
7: Deciduous broadleaf forest	4: Deciduous Broadleaf Forest	7: Deciduous broadleaf forest	4: Deciduous broadleaf forest	7: Deciduous broadleaf trees 9: Drought deciduous trees	11: Deciduous Broadleaf Forest	04: Deciduous forest
8: Mixed forest	5: Mixed Forest	8: Mixed forest	5: Mixed Forest	25: Mixed Wood Forest	15: Mixed Forest	N/A
9: Shrubland	6: Closed Shrublands 7: Open Shrublands	9: Shrubland	10: Dwarf trees, shrubs	10: Evergreen broadleaf shrubs 11: Deciduous shrubs 12: Thorn shrubs 26: Mixed shrubs	8: Shrubland 9: Mixed Shrubland/Grassland [WRF-Chem(NCAR) mapping]	N/A
10: Herbaceous	N/A	10: Herbaceous	N/A	13: Short grass and forbs	N/A	N/A
11: Planted / Cultivated	12: Croplands 14: Cropland-Natural Vegetation Mosaic	11: Planted / Cultivated	7: Crops, mixed farming	15: Crops 16: Rice 17: Sugar 18: Maize 19: Cotton 20: Irrigated crops	2: Dryland Cropland and Pasture 3: Irrigated Cropland and Pasture 4: Mixed Dryland/Irrigated Cropland and Pasture 5: Cropland / Grassland Mosaic 6: Cropland / Woodland Mosaic	01: Arable land 03: Permanent crops
12: Grassland	10: Grasslands	12: Grassland	6: Grassland	14: Long grass	7: Grassland	06: Grassland
13: Savanna	8: Woody Savanna 9: Savanna	13: Savanna	N/A	N/A	9: Mixed Shrubland / Grassland [WRF-Chem(RIFS) and WRF-Chem(UPM) mapping] 10: Savanna	11: Semi-natural vegetation
14: Wetlands	11: Permanent Wetlands	14: Wetlands	11: Wetland with plants	22: Swamp	17: Herbaceous Wetland 18: Wooded Wetland	
15: Tundra	18: Wooded Tundra 19: Mixed Tundra 20: Barren Tundra	15: Tundra	9: Tundra	22: Tundra	20: Herbaceous Tundra 21: Wooded Tundra 22: Mixed Tundra	
16: Snow and Ice	15: Snow and Ice	16: Snow and Ice	12: Ice caps and glaciers	2: Ice	24 Snow or Ice	07: Ice

Table 4. Mapping of LU classes from the categories used in the models' dry deposition code to the 16 AQMEII4 categories.

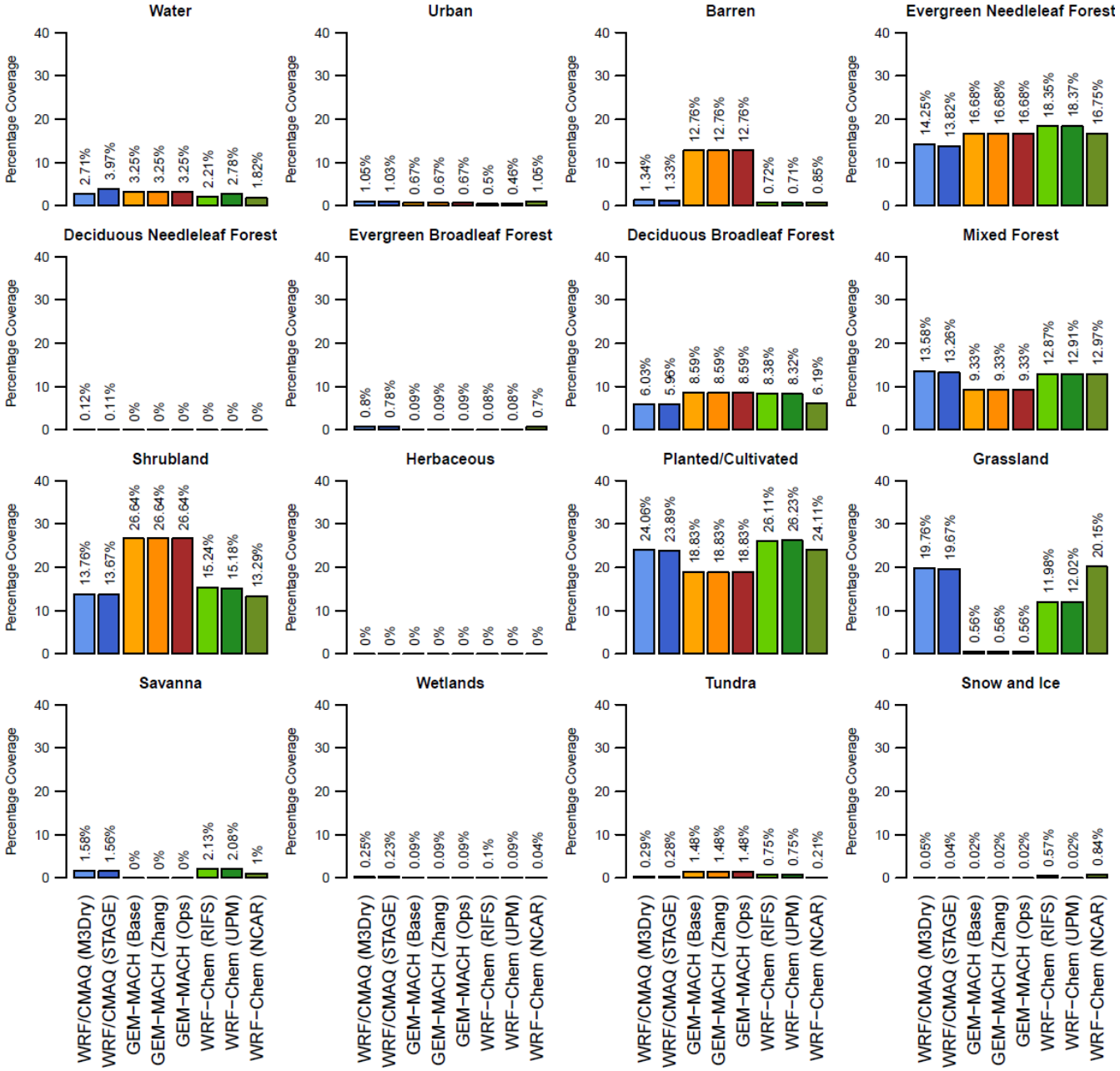


Figure 11. Fraction of the NA domain common to all models covered by each AQMEII4 LU category, excluding grid cells dominated by water by each model.

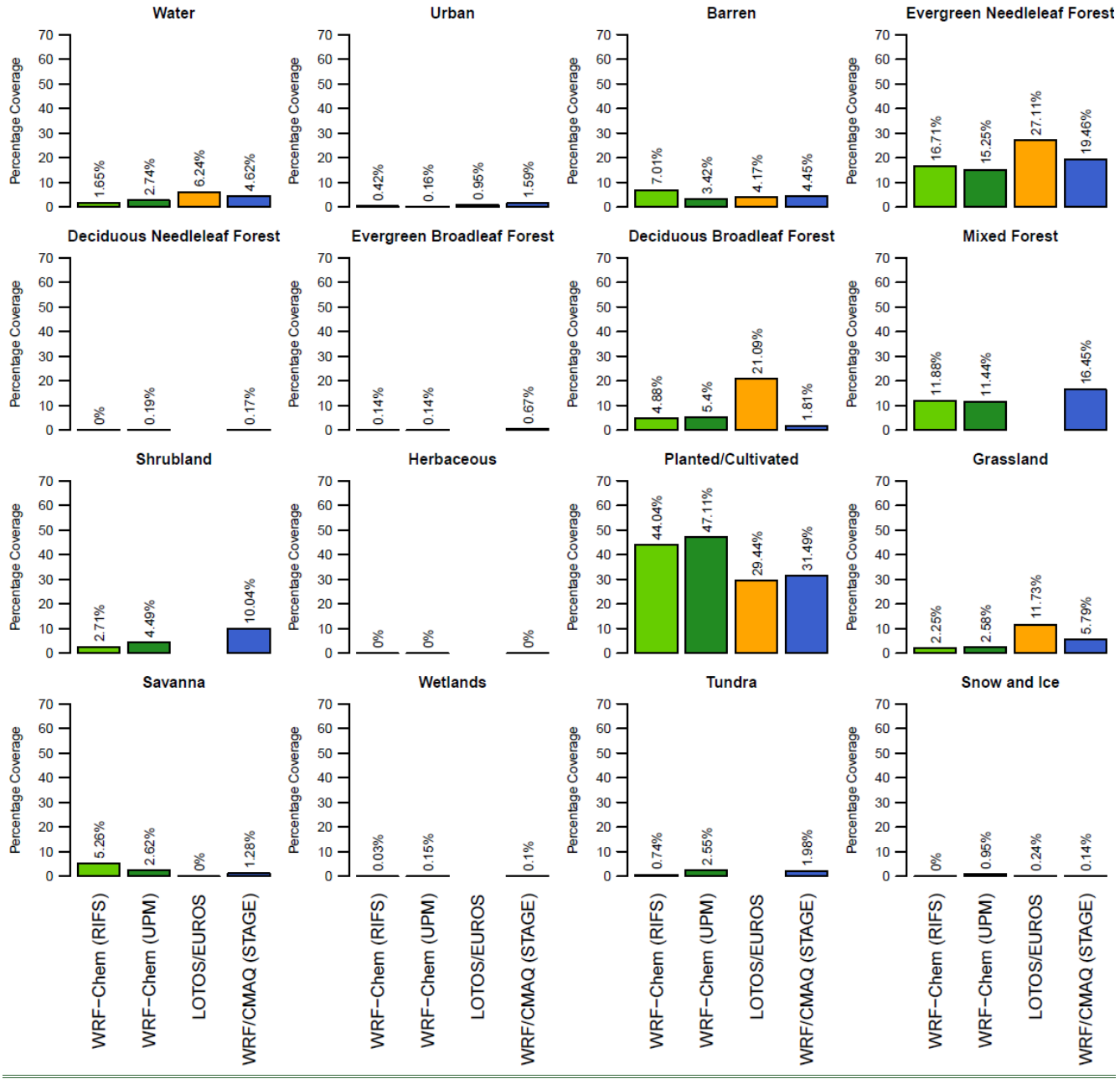


Figure 12. Fraction of the EUR domain common to all models covered by each AQMEII4 LU category, excluding grid cells dominated by water by each model.

Even for categories for which models show relatively close agreement of the total domain-wide coverage in Figures 11 and 12, spatial patterns of these coverages may still differ between models. To illustrate this, Figures S12~~S10~~ and S13~~S11~~ show maps of the fractional coverage of the evergreen needleleaf forest category for each model over NA and EUR~~EU~~. One fundamental difference between the WRF-Chem simulations and all other simulations is that the WRF-Chem simulations employed a dominant LU category approach in their LSM and dry deposition calculations (that is, only the LU with the largest LU fraction within a grid cell is used to represent that grid cell's LU for deposition calculations) while all other simulations accounted for subgrid variations in LU by employing a fractional LU category approach. Therefore, Figures S12~~S10~~ and S13~~S11~~ show evergreen needleleaf forest fractions of either 0 or 1 for the WRF-Chem simulations and fractions between 0 and 1 for all other simulations. Both figures reveal that, despite all native LU databases including evergreen needleleaf forest as an explicit LU category, the coverage for this LU can vary substantially between models, e.g. over the Southeastern U.S. in the NA domain (Figure S12~~S10~~) and Central Europe and the Iberian Peninsula in the EUR~~EU~~ domain (Figure S13~~S11~~).

To analyze the level of agreement in spatial coverage across models for all LU categories, while taking into account that the WRF-Chem simulations used a dominant LU category approach, we applied two metrics to assess model-to-model agreement for a given LU category and grid cell. The first metric simply determines whether all models agree that the LU category being assessed is the category with the highest fractional coverage (i.e. the dominant category) compared to all other categories in that grid cell, regardless of the actual fractional coverage for that dominant category for those simulations that use a fractional coverage approach. The second metric builds upon the first metric by not only determining whether the LU category being assessed is the dominant category in that grid cell, but also whether all models agree that its fractional coverage is at least 85%. By definition, all grid cells meeting the more stringent second metric also meet the first metric. Identifying grid cells meeting

	~~NA, # of common cells with "dominant coverage" for LU category for all models~~	~~NA, # of common cells with LU category coverage > 85% for all models~~	~~EU, # of common cells with "dominant coverage" for LU category for all models~~	~~EU, # of common cells with LU category coverage > 85% for all models~~
~~Water~~	~~27,925 (25.84%)~~	~~26,949 (24.94%)~~	~~81,423 (46.00%)~~	~~78,186 (44.17%)~~
~~Developed-Urban~~	~~161 (0.15%)~~	~~32 (0.03%)~~	~~63 (0.04%)~~	~~4 (0%)~~
~~Barren~~	~~112 (0.10%)~~	~~14 (0.01%)~~	~~2,203 (1.25%)~~	~~1,310 (0.74%)~~
~~Evergreen needleleaf forest~~	~~7,111 (6.58%)~~	~~1,544 (1.43%)~~	~~9,270 (5.237%)~~	~~2,531 (1.43%)~~
~~Deciduous needleleaf forest~~	~~0 (0%)~~	~~0 (0%)~~	~~0 (0%)~~	~~0 (0%)~~
~~Evergreen broadleaf forest~~	~~1 (0%)~~	~~0 (0%)~~	~~0 (0%)~~	~~0 (0%)~~
~~Deciduous broadleaf forest~~	~~2,663 (2.46%)~~	~~581 (0.54%)~~	~~362 (0.21%)~~	~~5 (0%)~~
~~Mixed forest~~	~~3,968 (3.67%)~~	~~705 (0.65%)~~	~~0 (0%)~~	~~0 (0%)~~
~~Shrubland~~	~~757 (0.70%)~~	~~43 (0.04%)~~	~~0 (0%)~~	~~0 (0%)~~
~~Herbaceous~~	~~0 (0%)~~	~~0 (0%)~~	~~0 (0%)~~	~~0 (0%)~~
~~Planted/Cultivated~~	~~10,640 (9.85%)~~	~~6,130 (5.67%)~~	~~24,627 (13.91%)~~	~~6,108 (3.45%)~~
~~Grassland~~	~~164 (0.15%)~~	~~52 (0.05%)~~	~~147 (0.08%)~~	~~0 (0%)~~

Savanna	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Wetlands	1 (0%)	0 (0%)	0 (0%)	0 (0%)
Tundra	83 (0.08%)	0 (0%)	0 (0%)	0 (0%)
Snow and Ice	1 (0%)	0 (0%)	0 (0%)	0 (0%)
Total of grid cells with common LU (incl. water cells)	53,587 (49.59%)	36,050 (33.36%)	118,095 (66.72%)	88,144 (49.80%)
Total of grid cells with diverging LU (incl. water cells)	54,471 (50.41%)	72,008 (66.64%)	58,899 (33.28%)	88,850 (50.20%)
Total of grid cells with common LU (excl. water cells)	25,662 (32.02%)	9,101 (11.22%)	36,672 (38.37%)	9,958 (10.08%)
Total of grid cells with diverging LU (excl. water cells)	54,471 (67.98%)	72,008 (88.78%)	58,899 (61.63%)	88,850 (89.92%)

Table 5. Number and percentage of grid cells within the common NA and EU analysis domains where all models agree on the LULC category for that grid cell. Two metrics are used to assess agreement: 1) all models agree that the LULC category being assessed is the category with the highest fractional coverage compared to all other categories in that grid cell ("dominant coverage"), and 2) in addition to meeting metric 1, all models also agree that a given grid cell has at least 85% coverage for the LULC category being assessed. The last four rows summarize the level of agreement across either all 16 LULC categories (including water) or all 15 non-water LULC categories. The percentages shown in the rows corresponding to individual LULC categories are calculated with respect to all grid cells in the common analysis domains (108,058 for NA and 176,994 for EU).

the first metric can be thought of as a way to assess agreement in LU categories across models if all simulations (not only WRF-Chem) had used a dominant LU category approach. The subset of grid cells identified by metric 1 for a given LU category that also satisfies the >85% criterion defined for metric 2 can be thought of as the common "dominant" cells for that LU category in which even the computations performed by the models using a fractional LU category approach were mostly impacted by the physical characteristics of that LU category, with only minor impacts from other LU categories possibly also present in the grid cell. This subset of grid cells was used in Section 3.2 when comparing V_d between the point model simulations at specific flux measurement sites and the grid model simulations.

The results of applying metrics 1 and 2 to all LU categories and models over both NA and EUREU are shown in Figures S14-S15S12-S13 and Table 5 which lists the number and percentage of grid cells within the common NA and EUREU analysis meeting each metric. The first sixteen rows of Table 5 contain results for each of the AQMEII4 LU categories. The percentages shown in these rows are calculated with respect to the total number of grid cells in the common analysis domains (108,058 for NA and 176,994 for EUREU), including water grid cells. Water is by far the category with the highest level of agreement over

both NA and EUREU with only minor differences between the two metrics, indicating that most grid cells dominated by water

	NA, # of common cells with "dominant coverage" for LU category for all models	NA, # of common cells with LU category coverage > 85% for all models	EUR, # of common cells with "dominant coverage" for LU category for all models	EUR, # of common cells with LU category coverage > 85% for all models
-Water	27,925 (25.84%)	26,949 (24.94%)	81,423 (46.00%)	78,186 (44.17%)
Developed-Urban	161 (0.15%)	32 (0.03%)	63 (0.04%)	4 (0%)
Barren	112 (0.10%)	14 (0.01%)	2,203 (1.25%)	1,310 (0.74%)
Evergreen needleleaf forest	7,111 (6.58%)	1,544 (1.43%)	9,270 (5.237%)	2,531 (1.43%)
Deciduous needleleaf forest	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Evergreen broadleaf forest	1 (0%)	0 (0%)	0 (0%)	0 (0%)
Deciduous broadleaf forest	2,663 (2.46%)	581 (0.54%)	362 (0.21%)	5 (0%)
Mixed forest	3,968 (3.67%)	705 (0.65%)	0 (0%)	0 (0%)
Shrubland	757 (0.70%)	43 (0.04%)	0 (0%)	0 (0%)
Herbaceous	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Planted/Cultivated	10,640 (9.85%)	6,130 (5.67%)	24,627 (13.91%)	6,108 (3.45%)
Grassland	164 (0.15%)	52 (0.05%)	147 (0.08%)	0 (0%)
Savanna	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Wetlands	1 (0%)	0 (0%)	0 (0%)	0 (0%)
Tundra	83 (0.08%)	0 (0%)	0 (0%)	0 (0%)
Snow and Ice	1 (0%)	0 (0%)	0 (0%)	0 (0%)
Total of grid cells with common LU (incl. water cells)	53,587 (49.59%)	36,050 (33.36%)	118,095 (66.72%)	88,144 (49.80%)
Total of grid cells with diverging LU (incl. water cells)	54,471 (50.41%)	72,008 (66.64%)	58,899 (33.28%)	88,850 (50.20%)
Total of grid cells with common LU (excl. water cells)	25,662 (32.02%)	9,101 (11.22%)	36,672 (38.37%)	9,958 (10.08%)
Total of grid cells with diverging LU (excl. water cells)	54,471 (67.98%)	72,008 (88.78%)	58,899 (61.63%)	88,850 (89.92%)

Table 5. Number and percentage of grid cells within the common NA and EUR analysis domains where all models agree on the LULC category for that grid cell. Two metrics are used to assess agreement: 1) all models agree that the LULC category being assessed is the category with the highest fractional coverage compared to all other categories in that grid cell ("dominant coverage"), and 2) in addition to meeting metric 1, all models also agree that a given grid cell has at least 85% coverage for across all models are almost fully (> 85%) or often fully covered by water, reflecting the oceans and open waters present in both analysis domains.

the LULC category being assessed. The last four rows summarize the level of agreement across either all 16 LULC categories (including water) or all 15 non-water LULC categories. The percentages shown in the rows corresponding to individual LULC categories are calculated with respect to all grid cells in the common analysis domains (108,058 for NA and 176,994 for EUR).

The LU categories with the second- and third-largest amount of agreement between models are planted/cultivated land and evergreen needleleaf forest over both domains. Over NA, the deciduous broadleaf and mixed forest LU categories also have about 600 – 4,000 [0.5% - 4%] common grid cells depending on metric and category. For the remaining LU categories, the number and percentage of grid cells matching across models is very low, especially for the more stringent metric 2. While to some extent this is expected given the low overall domain coverage of some of these categories by all models (Figures 11-12), this is also the case for shrubland and grassland which has substantial domain-wide coverage over NA (Figure 11).

The last four rows of Table 5 summarize the level of agreement across either all 16 LU categories (including water) or the 15 non-water LU categories. Figures 13a-d depict the location of common grid cells for both metrics and continents. In these figures, grid cells meeting the metric for any LU category are colored in dark red while grid cells not meeting it for any LU category (i.e. grid cells without common LU category as measured by the metric) are colored in white. These figures along with Table 5 illustrate that even for the less restrictive metric 1, 68% (62%) of non-water grid cells over NA (EUR~~EU~~) do not share a common dominant LU category across models. Over NA, many of these non-matching grid cells are located in the southern and western portions of the domain while over EUR~~EU~~, they are most prevalent in the western, northeastern, and southeastern portions of the domain. When considering the more stringent metric 2, i.e. grid cells in which the common category has at least 85% coverage for all models, this number of grid cells with diverging LU categories increases to roughly 90% of non-water grid cells over both domains. The only areas with significant contiguous clusters of such common cells are the agricultural regions in the north-central NA domain and portions of central Europe and Sweden.

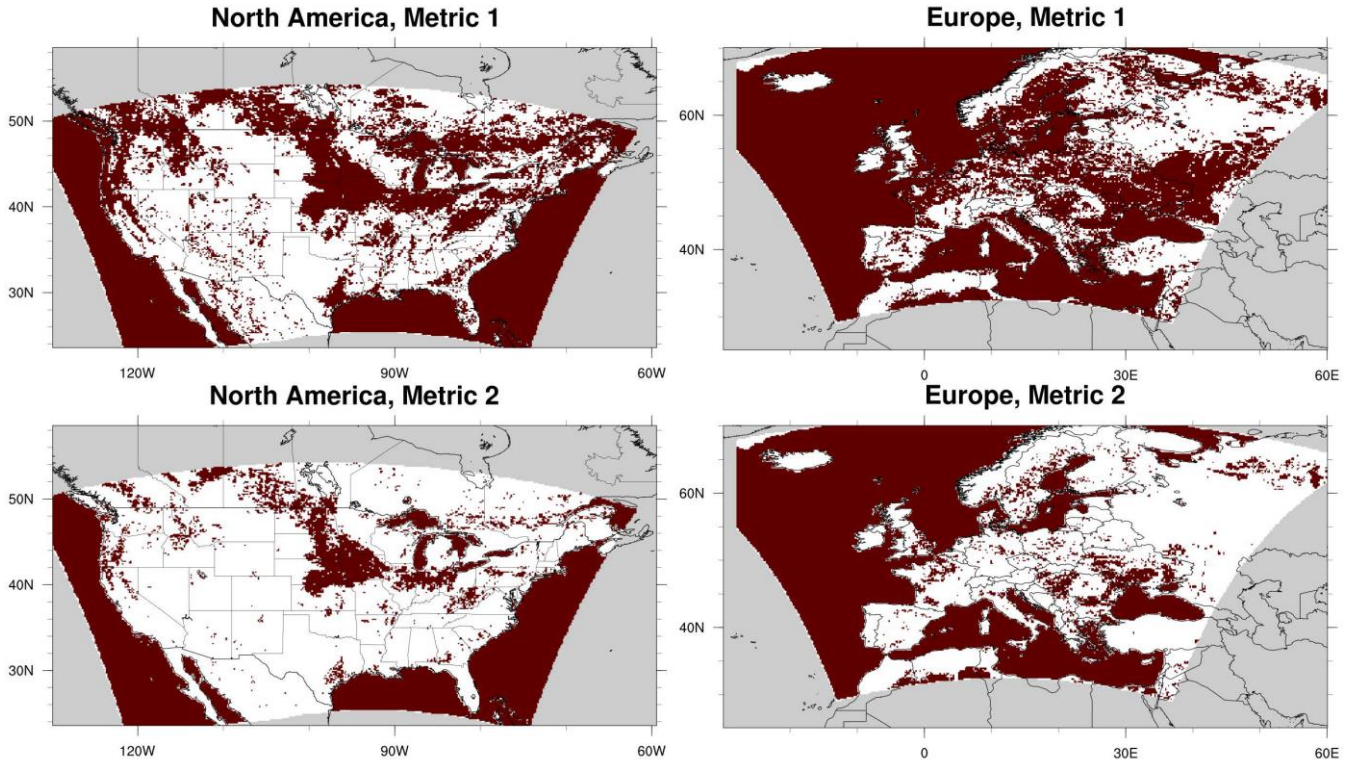


Figure 13. Red areas indicate grid cells meeting metrics 1 and 2 used to assess LU communality across models as described in the text.

This low number of grid cells with a common LU category strongly suggests that differences in LU coverage can contribute to or even drive differences in LU-specific dry deposition fluxes, in addition to any differences in process representation that exist between different models. To investigate this, Figure 14 compares LU-specific dry deposition fluxes, V_d, and LU fractions over NA for 7 selected LU categories. The LU-specific dry deposition fluxes (V_d) represent annual totals (means) over all grid cells in which the LU category being assessed has a fractional coverage of at least 85% for a given model. For some LU categories with relatively similar total coverage across the domain (e.g. evergreen needleleaf forest and deciduous broadleaf forest), differences in total O_3 dry deposition flux to that LU category closely mirror differences in LU-specific average V_d. For mixed forest, the lower fractional coverage for GEM-MACH (Base) and GEM-MACH (Ops) leads to a below-average total deposition flux to that category despite V_d being the highest. For the grassland and barren categories, differences in their dry deposition fluxes are almost entirely driven by differences in LU fractional coverage rather than differences in V_d. Corresponding results for the EUR~~EU~~ domain are shown in Figure S16~~S14~~ and confirm that differences in both V_d and LU coverage contribute to differences in LU-specific dry deposition fluxes.

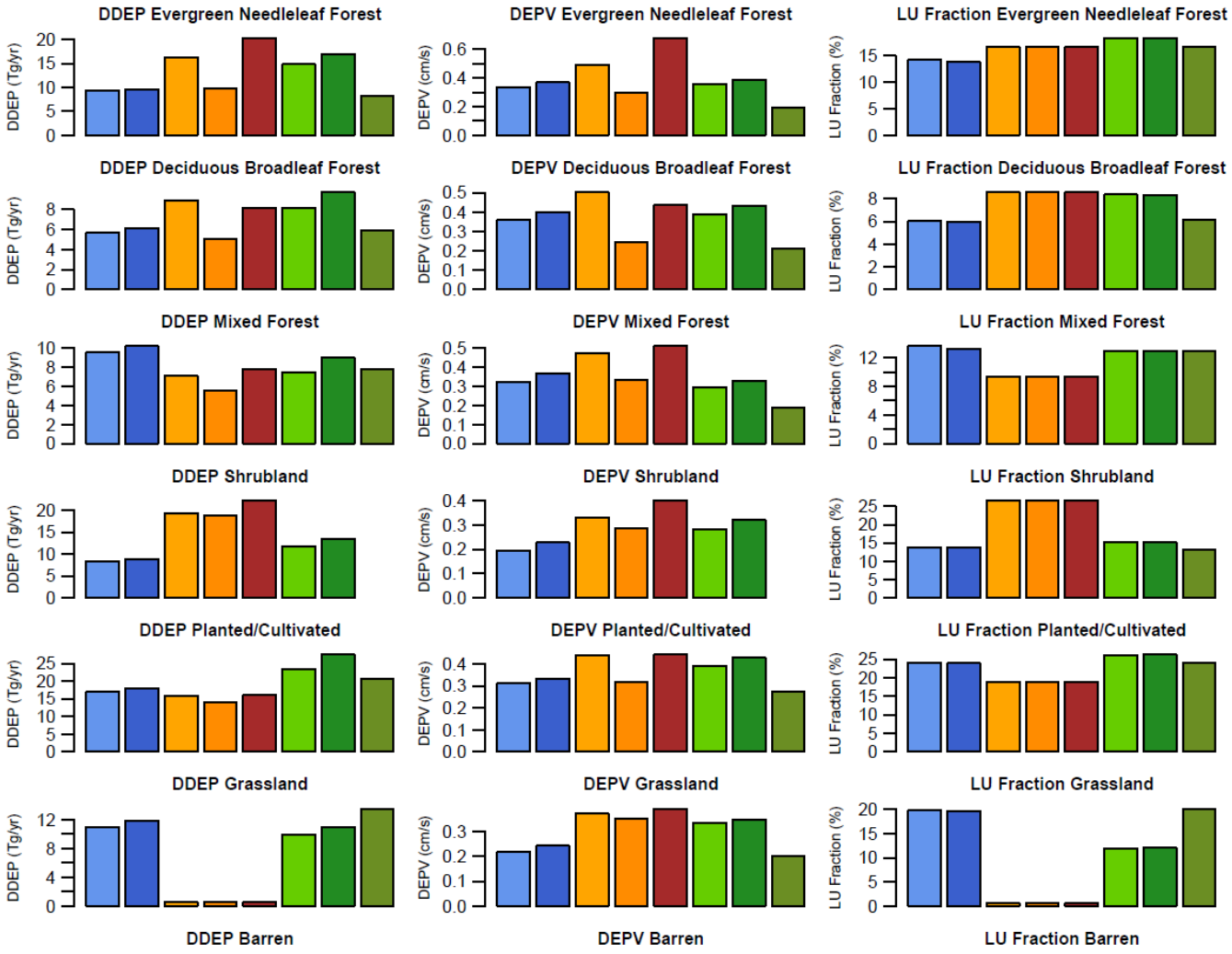


Figure 14. LU-specific annual domain-total dry deposition fluxes (Tg), LU-specific annual mean dry deposition velocity (cm/s), and percentage LU category domain coverage (excluding water grid cells) for seven selected LU categories over the NA domain. For each LU category and model, the analysis considered grid cells in the analysis domain common to all models in which a given model had at least 85% coverage for this LU category. The number of these grid cells differs across models due to underlying differences in LU (see Section 3.3).

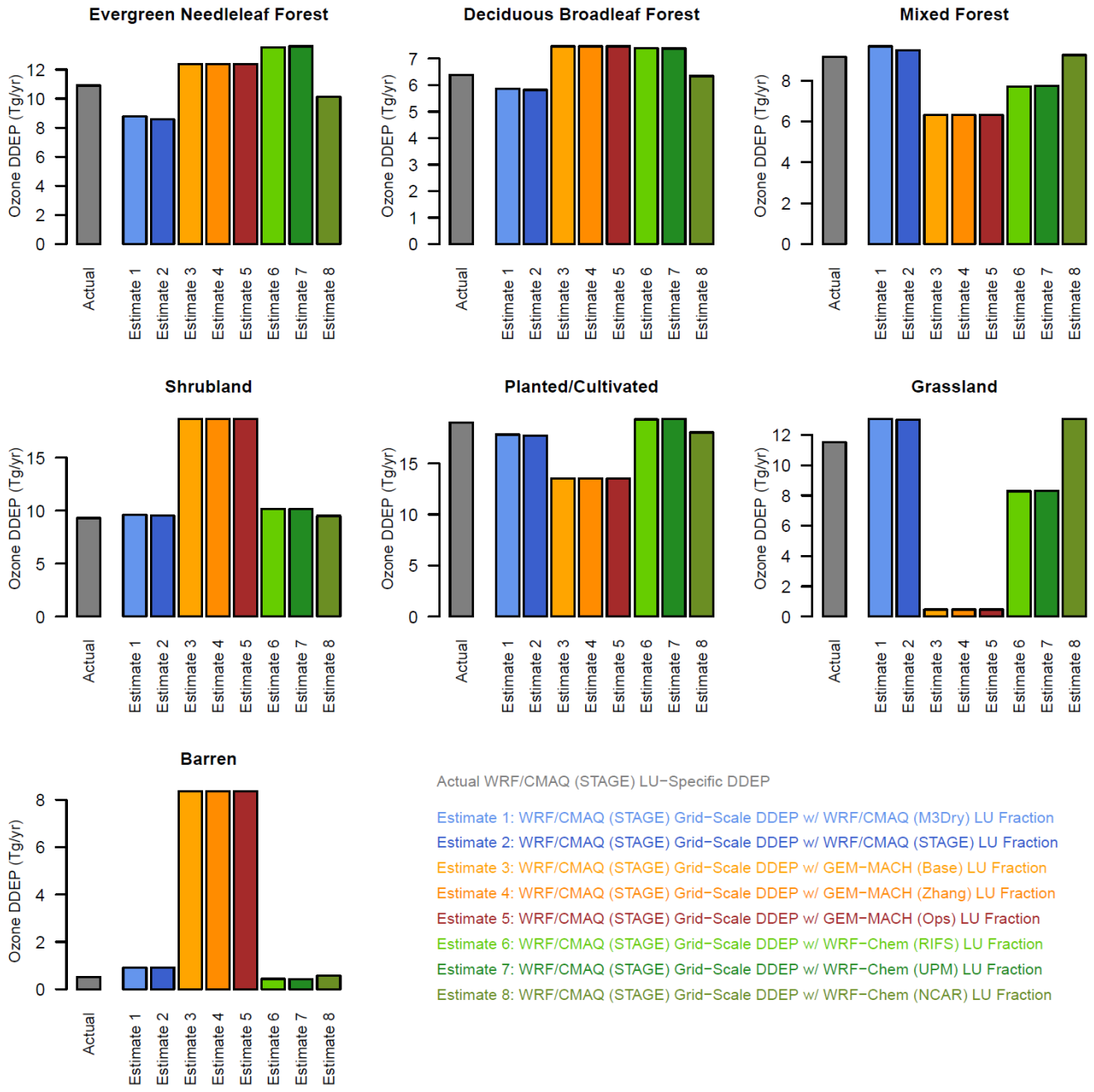


Figure 15. 2016 NA annual domain-wide total ozone dry deposition fluxes (Tg/yr) for selected LU categories. The grey bar shows LU-specific fluxes calculated by WRF/CMAQ (STAGE) while the colored bars show LU-specific fluxes estimated by combining WRF/CMAQ (STAGE) grid-scale fluxes with LU fractions from other models as described in the text.

The results shown in Figure 14 and S14 have important implications for computing estimates of deposition fluxes to specific ecosystems. In past model intercomparison studies such as those performed under the umbrella of the Task Force on Hemispheric Transport of Air Pollution (TF-HTAP, http://www.htap.org), such estimates have often been computed through post-processing by apportioning archived modeled grid-aggregated dry deposition fluxes to specific LU categories using a fixed LU database (e.g. Hardacre et al., 2015; Schwede et al., 2018). While this approach makes use of actual modeled dry deposition fluxes rather than using modeled concentrations as inputs to offline dry deposition calculations (e.g. Van Dingenen et al., 2009; Avnery et al., 2011), it may still be subject to uncertainties arising from differences between grid-aggregated vs. LU-specific dry deposition fluxes as well as differences in LU categorization – a flux associated with a model LU category for model-internal deposition may be aggregated under a different LU category in post-processing using a different post-processing database. The diagnostic information collected for AQMEII4 and analyzed in this manuscript allows us to illustrate and quantify both uncertainties.Figure 15 compares the LU-specific annual total dry deposition fluxes to seven LU categories calculated by model WRF/CMAQ (STAGE) (gray bars) to estimates derived by combining grid-aggregated dry deposition fluxes from the same model with LU fractions from all models (colored bars). Comparing the gray bars to the dark blue bars shows the impact of computing actual LU-specific deposition fluxes within the model simulation vs. estimating them by linearly scaling grid-aggregated values using the model's own LU fractions. This effect is relatively small for most LU categories but reaches about 20% for evergreen needleleaf forest and 40% for barren. Comparing the grey bars to all the other colored bars (except dark blue) shows the combined uncertainty of using a linear scaling of grid-aggregated deposition values and using different LU datasets than the one used within the model to calculate the LU-specific value. These differences are generally much larger and reflect the substantial divergence in LU categories across models, due to the variety of native-mode LU databases in use between different models. This example demonstrates the advantage of computing and collecting LU-specific diagnostic dry deposition information in model intercomparison studies so that model-to-model differences in deposition estimates can be tied to differences in model-specific process representation and LU datasets. Potential future work to harmonize LU datasets across models would allow further constraining differences to process representations only. This in turn would make it easier to leverage insights gained from point intercomparison studies (Wu et al., 2018; Clifton et al., 2023; Khan et al., submitted) for improving the representation of dry deposition in grid models.

4 Summary

This study presented a diagnostic analysis of ozone dry deposition from annual AQMEII4 photochemical grid modeling simulations performed over NA and EUR~~EU~~. Simulated annual O_3 dry deposition fluxes ranged from 59.5 Tg/yr to 91.4 Tg/yr over the NA domain and from 47.2 Tg/yr to 82.9 Tg/yr over the EUR~~EU~~ domain. Analysis of grid-aggregated effective conductances showed that models with similar V_d can exhibit significant differences in the absolute and relative contributions of different pathways to V_d. The availability of effective conductances also allowed an attribution of model differences in V_d to specific processes. For example, GEM-MACH (Base), GEM-MACH (Ops), and WRF-Chem (RIFS) all have similar soil

and lower canopy effective conductances, revealing that the differences in V_d between these simulations stem from differences in the cuticular and, to a lesser extent, stomatal pathways.

Analysis of LU-specific V_d and effective conductances revealed a general increase in model spread compared to analyzing grid-aggregated values, especially for V_d and the cuticular and stomatal effective conductances over forested and agricultural LU types. This finding suggests that LU-dependent parameter choices and the representation of vegetation effects on O_3 dry deposition are a significant source of variability in grid-aggregated V_d while also indicating that an analysis of only grid-aggregated deposition diagnostics can mask process-specific differences that exist between schemes. Utilizing the LU-specific diagnostics available from the grid model simulations also provided an opportunity to compare these diagnostics to the results from the single-point model intercomparison study by Clifton et al. (2023). Results showed good agreement between the behavior of different dry deposition schemes in single-point models at individual sites vs. their implementation in grid models in aggregate across either the NA or EUR~~EU~~ domain for a given LU type. For example, both grid model and point model V_d ranged between 0.4 and 1.0 cm s-1 across models for mixed forest grid model results and the corresponding point model results at Borden Forest and Harvard Forest, with the lowest V_d associated with GEM-MACH Zhang and the highest V_d associated with GEM-MACH Wesely in both comparisons. Grid and point models also agreed that WRF/CMAQ M3Dry had a much smaller contribution of the soil effective conductance to V_d compared to all other models and that WRF-Chem Wesely had a smaller contribution of the cuticular effective conductance to V_d compared to all other models. This demonstrates that the analysis of O_3 dry deposition schemes implemented in grid models can successfully be connected to detailed point model evaluation studies using long-term O_3 flux measurements when generating LU-specific diagnostic information and limiting the analysis to specific LU categories. This in turn allows the developers of grid models to leverage process-level insights gained from point intercomparison studies for improving the representation of dry deposition in their modeling systems. The importance of observation sites being representative of a single LU classification for this purpose has also been noted in this work.

This study also presented a detailed analysis of different models' LU distributions that revealed substantial differences in the spatial patterns and sometimes also the domain-wide coverage of certain LU categories over both domains, especially non-forest partially vegetated categories such as agricultural areas, shrubland, and grassland. Overall, 68% (62%) of non-water grid cells over NA (EUR~~EU~~) were found not to share a common dominant LU category across models. When considering only grid cells in which the common dominant category has at least 85% coverage for all models, this number of grid cells with diverging LU categories increased to roughly 90% of non-water grid cells over both domains. By comparing LU distributions, LU-specific V_d, and LU-specific dry deposition fluxes, we demonstrate that differences in LU coverage can contribute to or even drive differences in LU-specific dry deposition fluxes, in addition to any differences in process representation that exist between different models. For mixed forest, the lower fractional coverage for GEM-MACH (Base) and GEM-MACH (Ops) leads to a below-average total deposition flux to that category despite V_d being the highest. Differences in the grassland and barren dry deposition fluxes over the NA domain are almost entirely driven by differences in LU fractional coverage rather

than differences in V_d. On the other hand, for LU categories that have similar total domain-wide coverage across models (e.g. evergreen needleleaf forest and deciduous broadleaf forest), differences in total O_3 dry deposition fluxes to that LU category closely mirror differences in LU-specific average V_d, indicating that in this case differences in process representation drive differences in fluxes.

Finally, we contrasted the LU-specific dry deposition fluxes simulated in AQMEII4 to estimates in which grid-aggregated fluxes were linearly apportioned to specific LU categories based on LU fractions. When using a consistent set of LU information between the direct simulation of LU-specific fluxes and the apportionment of grid-aggregated fluxes to LU categories based on LU fractions, differences in domain-wide LU-specific deposition estimated by both approaches were generally within 10% but could be as high as 40% for a given LU category. These variations are caused by the non-linear effects of subgrid LU variability on V_d and deposition fluxes. When allowing for differences in LU distributions between those used in the grid model simulations generating the LU-specific results and those used to apportion grid-aggregated results to specific LU categories, domain-wide LU-specific deposition estimates differed by 50% or more for a number of categories. This illustrates that inconsistencies in LU representation between the model simulations and the post-processing apportionment approach can introduce substantial uncertainties in estimates of fluxes to specific ecosystems.

Overall, the results in this study demonstrate that the generation of grid-aggregated and LU-specific diagnostic outputs in model application and intercomparison studies can provide valuable information from (i)~~both~~ a model development point of view (~~to~~ better understand drivers of model variability and identify priorities for model improvement,~~)~~ and (ii) an impacts analysis point of view (~~to~~ distinguish between stomatal vs. non-stomatal deposition and quantify deposition to specific LU types.~~).~~ However, it should also be acknowledged that designing, implementing, collecting, and analyzing these diagnostics in AQMEII4 was resource-intensive, required a high level of coordination across research groups, and resulted in several iterations before the analyses presented in this and other AQMEII4 manuscript could be completed. Finally, the results of the LU analysis highlight the importance of documenting and analyzing the representation of LU across models. We recommend that Future work ~~should~~ be aimed at harmonizing this aspect when using CTMs for deposition analyses, in addition to the current standard practice of harmonizing emissions and boundary conditions in intercomparison studies. In addition, such future harmonization efforts should utilize current generation publicly available high resolution LU datasets to improve upon some of the currently used datasets that date back more than a decade, and would also benefit from close collaboration with the LSM community.

Data availability

The gridded annual, monthly, and monthly median diurnal AQMEII4 model fields analyzed in this manuscript can be obtained as netCDF files from zenodo at https://doi.org/10.5281/zenodo.15127614. Gridded AQMEII4 model fields at higher temporal

resolution and diagnostic outputs for additional variables collected as part of AQMEII4 constitute very large datasets (Tb), and their access requires software and documentation for their interpretation and for processing from the storage format into netCDF format. For more information and access to AQMEII4 model outputs beyond the fields analyzed in this manuscript, contact the AQMEII4 co-leads Stefano Galmarini (stefano.galmarini@ec.europa.eu) and Christian Hogrefe (hogrefe.christian@epa.gov)

~~All data used in this work are available upon request from the authors. Please contact the corresponding author to request any data related to this work.~~

Author contributions

CH led the development of this manuscript and conducted most of the analyses presented in Section 3. CH, SG, PM, IK, OEC, and DBS helped conceptualize both the AQMEII4 model intercomparison framework and the analyses presented in this study. OEC conducted the analysis of point model simulations. UA, JOB, TB, PC, AH, CH, RK, AL, PM, KM, JLPC, JEP, YHR, RSJ, and RS conducted grid model simulations. JOB, PC, PM, JEP, and RSJ conducted point model simulations. SG, Roberto Bellasio, and Roberto Bianconi developed and maintained infrastructure for data exchange and storage. All authors contributed to the editing of the manuscript.

Competing Interests

One of the co-authors (Stefano Galmarini) is an ACP editor.

Disclaimer

The views expressed in this article are those of the authors and do not necessarily represent the views or policies of the U.S. Environmental Protection Agency.

Special issue statement

This manuscript is part of the special issue “AQMEII4: a detailed assessment of atmospheric deposition processes from point models to regional-scale models”.

Acknowledgments

We gratefully acknowledge members of the AQMEII4 steering committee (Christopher Holmes, Lisa Emberson, Johannes Flemming, Sam Silva, Johannes Bieser, Jason Ducker, and Martijn Schaap) for facilitating the analysis described in this manuscript by designing and coordinating regional-scale air quality model simulations that provide diagnostic insights into modeled dry deposition.

References

Avnery, S., Mauzerall, D. L., Liu, J., & Horowitz, L. W. (2011). Global crop yield reductions due to surface ozone exposure: 1. Year 2000 crop production losses and economic damage. Atmospheric Environment, 45, 2284–2296, https://doi.org/10.1016/j.atmosenv.2010.11.045

Baublitz, C. B., Fiore, A. M., Clifton, O. E., Mao, J., Li, J., Correa, G., et al., Sensitivity of Tropospheric Ozone Over the Southeast USA to Dry Deposition. Geophysical Research Letters, 47, e2020GL087158. https://doi.org/10.1029/2020GL087158, 2020

Broxton, P. D., Zeng, X., Sulla-Menashe, D., and Troch, P. A.. A global land cover climatology using MODIS data. Journal of Applied Meteorology and Climatology, 53(6), 1593-1605. DOI: 10.1175/JAMC-D-13-0270.1, 2014

Clifton, O. E., Paulot, F., Fiore, A. M., Horowitz, L. W., Correa, G., Fares, S., Goded, I., Goldstein, A. H., Gruening, C., Hogg, A. J., Loubet, B., Mammarella, I., Munger, J. W., Neil, L., Stella, P., Uddling, J., Vesala, T., & Weng, E.: Influence of dynamic ozone dry deposition on ozone pollution. Journal of Geophysical Research: Atmospheres. https://doi.org/10.1029/2020JD032398, 2020a

Clifton, O. E., Fiore, A. M., Massman, W. J., Baublitz, C. B., Coyle, M., Emberson, L., et al.: Dry deposition of ozone over land: processes, measurement, and modeling, Rev. Geophys., 58, https://doi.org/10.1029/2019RG000670, 2020b.

Clifton, O. E., Schwede, D., Hogrefe, C., Bash, J. O., Bland, S., Cheung, P., Coyle, M., Emberson, L., Flemming, J., Fredj, E., Galmarini, S., Ganzeveld, L., Gazetas, O., Goded, I., Holmes, C. D., Horváth, L., Huijnen, V., Li, Q., Makar, P. A., Mammarella, I., Manca, G., Munger, J. W., Pérez-Camanyo, J. L., Pleim, J., Ran, L., San Jose, R., Silva, S. J., Staebler, R., Sun, S., Tai, A. P. K., Tas, E., Vesala, T., Weidinger, T., Wu, Z., and Zhang, L.: A single-point modeling approach for the intercomparison and evaluation of ozone dry deposition across chemical transport models (Activity 2 of AQMEII4), Atmos. Chem. Phys., 23, 9911–9961, https://doi.org/10.5194/acp-23-9911-2023, 2023.

EEA. (2020). CLC 2012—Copernicus Land Monitoring Service [Dataset]. Retrieved from https://land.copernicus.eu/pan-european/corine-land-cover/clc-2012

Fiore, A. M., Dentener, F. J.,Wild, O., Cuvelier, C., Schultz, M. G., Hess, P., Textor, C., Schulz, M., Doherty, R. M., Horowitz, L.W., MacKenzie, I. A., Sanderson, M. G., Shindell, D. T., Stevenson, D. S., Szopa, S., Van Dingenen, R., Zeng, G., Atherton, C., Bergmann, D., Bey, I., Carmichael, G., Collins, W. J., Duncan, B. N., Faluvegi, G., Folberth, G., Gauss, M., Gong, S., Hauglustaine, D., Holloway, T., Isaksen, I. S. A., Jacob, D. J., Jonson, J. E., Kaminski, J. W., Keating, T. J., Lupu, A., Marmer, E., Montanaro, V., Park, R. J., Pitari, G., Pringle, K. J., Pyle, J. A., Schroeder, S., Vivanco, M. G., Wind, P., Wojcik, G., Wu, S., and Zuber, A.: Multimodel estimates of intercontinental source-receptor relationships for ozone pollution, J. Geophys. Res.-Atmos., 114, D04301, doi:10.1029/2008JD010816, 2009.

Friedl, M. A., McIver, D. K., Hodges, J. C. F., Zhang, X. Y., Muchoney, D., Strahler, A. H., Woodcock, C. E., Gopal, S., Schneider, A., Cooper, A., Baccini, A., Gao, F., and Schaaf, C. Global land cover mapping from MODIS: Algorithms and early results. Remote Sensing of Environment, 83(1–2), 287–302. , 2002

Friedl, M. and Sulla-Menashe. D. MCD12Q1 MODIS/Terra+Aqua Land Cover Type Yearly L3 Global 500m SIN Grid V006. 2019, distributed by NASA EOSDIS Land Processes Distributed Active Archive Center, https://doi.org/10.5067/MODIS/MCD12Q1.006. Accessed 2025-04-16.

Galmarini, S., Makar, P., Clifton, O. E., Hogrefe, C., Bash, J. O., Bellasio, R., Bianconi, R., Bieser, J., Butler, T., Ducker, J., Flemming, J., Hodzic, A., Holmes, C. D., Kioutsioukis, I., Kranenburg, R., Lupascu, A., Perez-Camanyo, J. L., Pleim, J., Ryu, Y.-H., San Jose, R., Schwede, D., Silva, S., and Wolke, R.: Technical note: AQMEII4 Activity 1: evaluation of wet and dry deposition schemes as an integral part of regional-scale air quality models, Atmos. Chem. Phys., 21, 15663–15697, https://doi.org/10.5194/acp-21-15663-2021, 2021.

Ganzeveld, L., Helmig, D., Fairall, C. W., Hare, J., and Pozzer, A.: Atmosphere-ocean ozone exchange: A global modelling study of biogeochemical, atmospheric, and waterside turbulence dependencies, Global Biogeochem. Cy., 23, GB4021, doi:10.1029/2008GB003301, 2009.

Hardacre, C., Wild, O., and Emberson, L.: An evaluation of ozone dry deposition in global scale chemistry climate models, Atmos. Chem. Phys., 15, 6419–6436, https://doi.org/10.5194/acp-15-6419-2015, 2015.

Hogrefe, C., Bash, J. O., Pleim, J. E., Schwede, D. B., Gilliam, R. C., Foley, K. M., Appel, K. W., and Mathur, R.: An analysis of CMAQ gas-phase dry deposition over North America through grid-scale and land-use-specific diagnostics in the context of AQMEII4, Atmos. Chem. Phys., 23, 8119–8147, https://doi.org/10.5194/acp-23-8119-2023, 2023

Khan, A. M., Clifton, O. E., Bash, J. O., Bland, S., Booth, N., Cheung, P., Emberson, L., Flemming, J., Fredj, E., Galmarini, S., Ganzeveld, L., Gazetas, O., Goded, I., Hogrefe, C., Holmes, C. D., Horvath, L., Huijnen, V., Li, Q., Makar, P. A., Mammarella, I., Manca, G., Munger, J. W., Perez-Camanyo, J. L., Pleim, J., Ran, L., San Jose, R., Schwede, D., Silva, S. J., Staebler, R., Sun, S., Tai, A. P. K., Tas, E., Vesala, T., Weidinger, T., Wu, Z., Zhang, L., and Stoy, P. C.:~~Khan et al.,~~ Ozone

dry deposition through plant stomata: Multi-model comparison with flux observations and the role of water stress as part of AQMEII4 Activity 2, EGUsphere [preprint], https://doi.org/10.5194/egusphere-2024-3038 in preparation, 2024

Kioutsioukis, I., C. Hogrefe, C.,P.A. Makar, P. A.,U. Alyuz, U.,J.O. Bash, J. O.,, R. Bellasio, R.,, Bianconi, R., Buttler, T.,, Butler, O.E. Clifton, O. E.,P. Cheung, P.,A. Hodzic, A.,R. Kranenburg, R.,A. Lupascu, A.,K. Momoh, K.,J. L. Perez-CamañoCamanyo, J. L., Pleim, J.,Y. H. Ryu, Y.-H.,R. San Jose, R.,D. Schwede, D., R. Sokhi, R., and S. Galmarini, S.: Operational and Probabilistic Evaluation of AQMEII-4 regional scale ozone dry deposition. Time to harmonise our LULC masks, EGUsphere [preprint], https://doi.org/10.5194/egusphere-2025-1091, 2025.in preparation, 2025

Lin, M., Horowitz, L.W., Xie, Y. et al. Vegetation feedbacks during drought exacerbate ozone air pollution extremes in Europe. Nat. Clim. Chang. 10, 444–451, https://doi.org/10.1038/s41558-020-0743-y, 2020

Loveland, T. R., Reed, B. C., Brown, J. F., Ohlen, D. O., Zhu, Z., Yang, L., & Merchant, J. W. (2000). Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data. International Journal of Remote Sensing, 21(6-7), 1303-1330. https://doi.org/10.1080/014311600210191

Makar, P. A., Cheung, P., Hogrefe, C., Akingunola, A., Alyuz Ozdemir, U., Bash, J. O., Bell, M. D., Bellasio, R., Bianconi, R., Butler, T., Cathcart, H., Clifton, O. E., Hodzic, A., KioutsioukisKoutsioukis, I., Kranenburg, R., Lupascu, A., Lynch, J. A., Momoh, K., Perez-Camanyo, J. L., Pleim, J., Ryu, Y.-H., San Jose, R., Schwede, D., Scheuschner, T., Shephard, M. W., Sokhi, R. S., and Galmarini, S.: Critical Load Exceedances for North America and Europe using an Ensemble of Models and an Investigation of Causes offor Environmental Impact Estimate Variability: An AQMEII4 Study, Atmos. Chem. Phys., 25, 3049–3107,EGUsphere [preprint], https://doi.org/10.5194/acp-25-3049-2025, 2025egusphere-2024-2226, 2024.

Makar, P., Staebler, R., Akingunola, A., Zhang, J., McLinden, C., Kharol, S.K., Pabla, B., Cheung, P., Zheng, Q., The effects of forest canopy shading and turbulence on boundary layer ozone. Nat Commun 8, 15243, 2017. https://doi.org/10.1038/ncomms15243

Makar, P. A., Akingunola, A., Aherne, J., Cole, A. S., Aklilu, Y.-A., Zhang, J., Wong, I., Hayden, K., Li, S.-M., Kirk, J., Scott, K., Moran, M. D., Robichaud, A., Cathcart, H., Baratzedah, P., Pabla, B., Cheung, P., Zheng, Q., and Jeffries, D. S.: Estimates of exceedances of critical loads for acidifying deposition in Alberta and Saskatchewan, Atmos. Chem. Phys., 18, 9897–9927, https://doi.org/10.5194/acp-18-9897-2018, 2018.

Makar, P. A., Stroud, C., Akingunola, A., Zhang, J., Ren, S., Cheung, P., and Zheng, Q.: Vehicle-induced turbulence and atmospheric pollution, Atmos. Chem. Phys., 21, 12291–12316, 2021. https://doi.org/10.5194/acp-21-12291-2021

Manders-Groot, A.M.M. and LOTOS-EUROS team, LOTOS-EUROS v2.3.000 Reference Guide, December 2023, available from , last accessed April 30, 2025

McRae, G.J, Goodin, W.R., and Seinfeld, J.H.: Development of a second-generation mathematical model for Urban air pollution—I. Model formulation, Atmos. Env., 16, 4, 679-696, 1982, https://doi.org/10.1016/0004-6981(82)90386-9

Paulot, F., Malyshev, S., Nguyen, T., Crounse, J. D., Shevliakova, E., and Horowitz, L. W.: Representing sub-grid scale variations in nitrogen deposition associated with land use in a global Earth system model: implications for present and future nitrogen deposition fluxes over North America. Atmospheric Chemistry and Physics, 18(24), 17963-17978, 2018

Schwede, D.B., D. Simpson, J. Tan, J.S. Fu, F. Dentener, E. Du, and W. deVries, Spatial variation of modelled total, dry and wet nitrogen deposition to forests at global scale, Environmental Pollution, 243, 2018, https://doi.org/10.1016/j.envpol.2018.09.084

Silva, Sam J. and Colette L. Heald. “Investigating Dry Deposition of Ozone to Vegetation.” Journal of Geophysical Research:

Simpson, D., Benedictow, A., Berge, H., Bergstr€om, R., Emberson, L.D., Fagerli, H., Flechard, C.R., Hayman, G.D., Gauss, M., Jonson, J.E., Jenkin, M.E., Nyíri, A., Richter, C., Semeena, V.S., Tsyro, S., Tuovinen, J.P., Valdebenito, A., Wind, P., 2012. The EMEP MSC-W chemical transport model – technical description. Atmos. Chem. Phys. 12, 7825e7865.

Stevenson, D. S., Dentener, F. J., Schultz, M. G., Ellingsen, K., van Noije, T. P. C., Wild, O., Zeng, G., Amann, M., Atherton, C. S., Bell, N., Bergmann, D. J., Bey, I., Butler, T., Cofala, J., Collins, W. J., Derwent, R. G., Doherty, R. M., Drevet, J., Eskes, H. J., Fiore, A. M., Gauss, M., Hauglustaine, D. A., Horowitz, L. W., Isaksen, I. S. A., Krol, M. C., Lamarque, J. F., Lawrence, M. G.,Montanaro, V., Muller, J. F., Pitari, G., Prather, M. J., Pyle, J. A., Rast, S., Rodriguez, J. M., Sanderson, M. G., Savage, N. H., Shindell, D. T., Strahan, S. E., Sudo, K., and Szopa, S.: Multimodel ensemble simulations of present-day and near-future tropospheric ozone, J. Geophys. Res.-Atmos., 111, D08301, doi:10.1029/2005JD006338, 2006

Tan, J., Fu, J.S., Dentener, F., Sun, J., Emmons, L., Tilmes, S., Sudo, K., Flemming, J., Jonson, J.E., Gravel, S., Bian, H., Henze, D., Lund, M.T., Kucsera, T., Takemura, T., Keating, T., 2018. Multi-model study of HTAP II on sulphur and nitrogen deposition. Atmos. Chem. Phys. 18, 6847e6866.

Van Dingenen, R., F.J. Dentener, F. Raes, M.C. Krol, L. Emberson, and J. Cofala, The global impact of ozone on agricultural crop yields under current and future air quality legislation, Atmospheric Environment, 43, 3, 2009, https://doi.org/10.1016/j.atmosenv.2008.10.033.

Wang, X., and D.L. Mauzerall, Characterizing distributions of surface ozone and its impact on grain production in China, Japan and South Korea: 1990 and 2020, Atmos. Environ., 38, pp. 4383-4402, 10.1016/j.atmosenv.2004.03.067, 2004.

Wesely, M. L.: Parameterization of surface resistances to gaseous dry deposition in regional-scale numerical models, Atmos. Environ., 23, 1293–1304, 1989.

Wild, O.: Modelling the global tropospheric ozone budget: exploring the variability in current models, Atmos. Chem. Phys., 7, 2643–2660, doi:10.5194/acp-7-2643-2007, 2007.

Wong, A. Y. H., Geddes, J. A., Tai, A. P. K., and Silva, S. J.: Importance of dry deposition parameterization choice in global simulations of surface ozone, Atmos. Chem. Phys., 19, 14365–14385, https://doi.org/10.5194/acp-19-14365-2019, 2019

Wu, Z., Schwede, D. B., Vet, R., Walker, J. T., Shaw, M., Staebler, R., & Zhang, L. (2018). Evaluation and intercomparison of five North American dry deposition algorithms at a mixed forest site. Journal of Advances in Modeling Earth Systems, 10, 1571–1586. https://doi.org/10.1029/2017MS001231

Young, P. J., Archibald, A. T., Bowman, K. W., Lamarque, J.-F., Naik, V., Stevenson, D. S., Tilmes, S., Voulgarakis, A., Wild, O., Bergmann, D., Cameron-Smith, P., Cionni, I., Collins,W. J., Dalsøren, S. B., Doherty, R. M., Eyring, V., Faluvegi, G., Horowitz, L. W., Josse, B., Lee, Y. H., MacKenzie, I. A., Nagashima, T., Plummer, D. A., Righi, M., Rumbold, S. T., Skeie, R. B., Shindell, D. T., Strode, S. A., Sudo, K., Szopa, S., and Zeng, G.: Preindustrial to end 21st century projections of tropospheric ozone from the Atmospheric Chemistry and Climate Model Intercomparison Project (ACCMIP), Atmos. Chem. Phys., 13, 2063–2090, doi:10.5194/acp-13-2063-2013, 2013.

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