ductances 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 paper, 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 Fig. 1 for the NA domain and Fig. 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 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 $\mathrm{Tg\,yr^{-1}}$ over the NA domain and 67.4 $\mathrm{Tg\,yr^{-1}}$ over the EUR domain. Individual model estimates range from 59.5 (GEM-MACH (Zhang)) to 91.4 $\mathrm{Tg\,yr^{-1}}$ (WRF-Chem (UPM)) over the NA domain and from 47.2 $\mathrm{Tg\,yr^{-1}}$ (LOTOS/EUROS) to 82.9 $\mathrm{Tg\,yr^{-1}}$ (WRF-Chem (UPM)) over the EUR domain. For context, past global modeling studies reported global annual O_3 deposition totals generally ranging between 800 and 1200 $\mathrm{Tg\,yr^{-1}}$ (e.g., Hardacre et al., 2015; Stevenson et al., 2006; Wild, 2007; Ganzeveld et al., 2009).

Over the NA domain all models estimate that O_3 deposition is highest over the eastern US 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 US 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 US, in addition to the spread over the southeastern US. Over the EUR domain, the largest relative spread in deposition to land occurs over northwestern Scandinavia, Ireland, Great Britain, and Türkiye, 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.

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 slightly higher dry deposition fluxes than GEM-MACH (Base)). Consistent with the discussion in Sect. 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. 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 numbers are for 2010.

	Annual total O_3 dry deposition flux (Tg)	Annual mean O_3 dry deposition velocity ($\mathrm{cm\,s^{-1}}$)
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

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 US and the generally lower V_d over the southwestern US compared to WRF-Chem (UPM). The role of differences

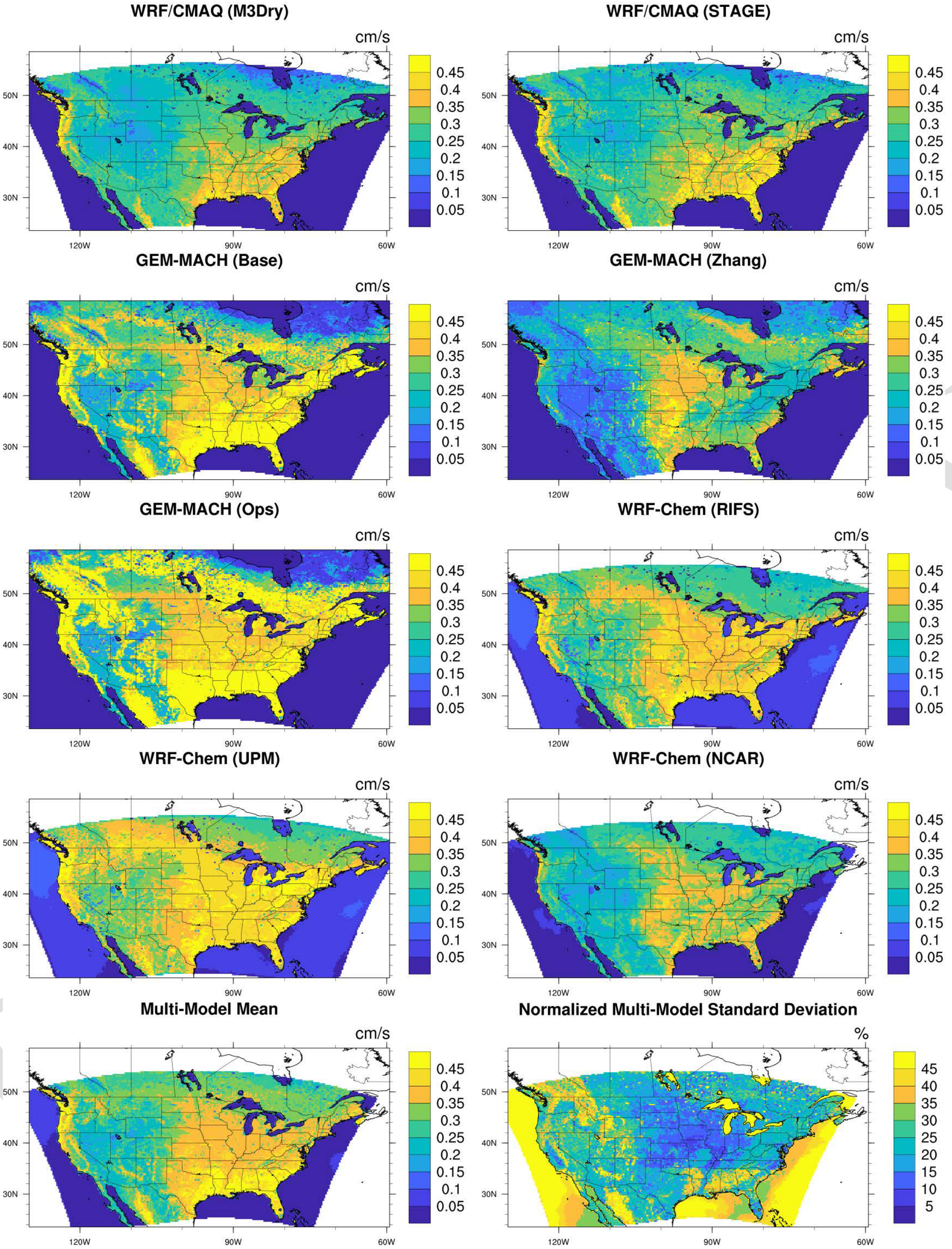

Figure 3. 2016 annual mean O_3 grid-scale dry deposition velocities for each model, the multi-model mean, and the normalized multi-model standard deviation over the NA domain. Note that the plots for individual models are not clipped to the domain common to all simulations and show the maximum spatial extent submitted for each model. The multi-model mean and normalized standard deviations are calculated and shown over the common domain.

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 Figs. 3–4 to those in the corresponding dry deposition flux maps in Figs. 1–2 shows a high level of similarity. For example, the areas over the southeastern US 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

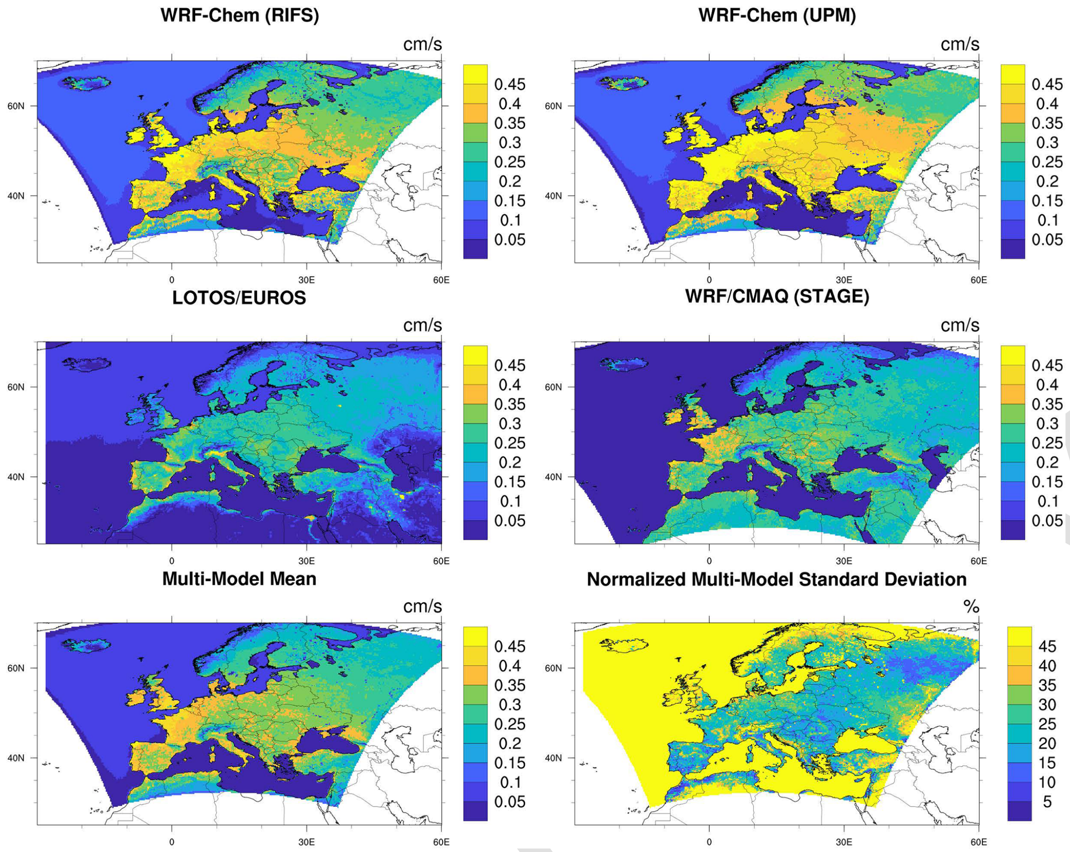

Figure 4. 2010 annual mean O_3 grid-scale dry deposition velocities for each model, the multi-model mean, and the normalized multi-model standard deviation over the EUR domain. Note that the plots for individual models are not clipped to the domain common to all simulations and show the maximum spatial extent submitted for each model. The multi-model mean and normalized standard deviations are calculated and shown over the common domain.

standard deviation for V_d. Despite these general 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 a 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 subgrid-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 Figs. S1–S2 in the Supplement, 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 Fig. 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 Figs. 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 Fig. S1 shows generally higher O_3 mixing ratios over the southeastern US 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). 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) and their spatial patterns (Fig. 1 vs. Fig. 3 and Fig. 2 vs. Fig. 4) confirms the influence of O_3 V_d on 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.

Figures S3 and S4 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 of 18.2 %; winter daytime flux of 14.5 %; summer nighttime flux of 22.7 %; winter nighttime flux of 22.8 %; summer daytime V_d of 18.6 %; winter daytime V_d of 17.2 %; summer nighttime V_d of 34.2 %; winter nighttime V_d of 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.

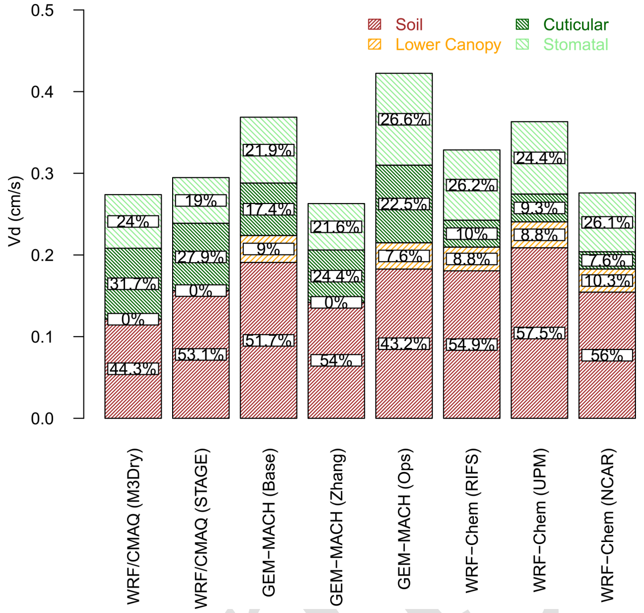

Figure 5. Annual domain-average grid-scale effective conductances and ozone deposition velocities for 2016 over the NA domain. Averages were calculated over all non-water grid cells in the portion of the analysis domain shared by all model simulations.

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 Sect. 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 a 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 for 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, Fig. 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.

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 fore-

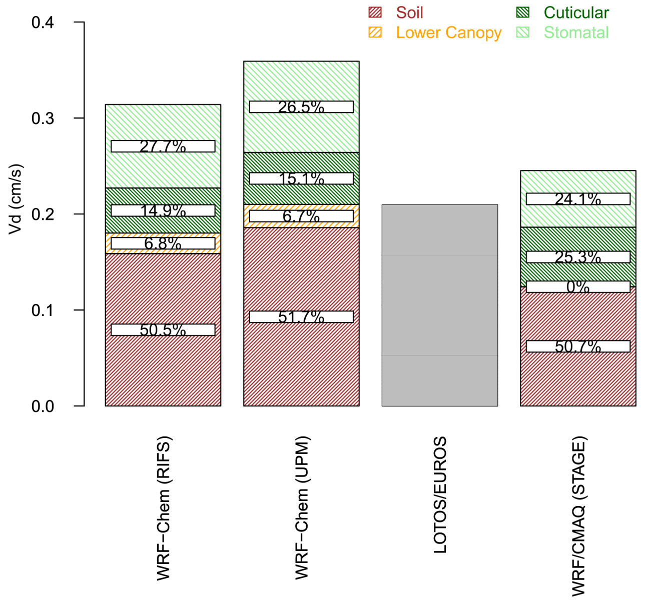

Figure 6. Annual domain-average grid-scale effective conductances and ozone deposition velocities for 2010 over the EUR domain. Averages were calculated over all non-water grid cells in the portion of the analysis domain shared by all model simulations. The bar for LOTOS/EUROS shows only the deposition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

casted meteorology, in turn influencing the deposition velocity components.

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. 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 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 US and central Europe. This is consistent with the domain-average contributions shown in Fig. 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^{-1} throughout both domains (10 % or less of the domain-wide annual means as shown in Figs. 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 (RIFS) and WRF-Chem (UPM) simulations have noticeably higher absolute soil pathway contributions in the south-central and southeastern US than the other simulations, indicating that the higher values in the domain-average soil pathway contributions for these four simulations shown in Fig. 3 originate in these regions.

Figures S5–S6 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 Figs. S7–S8. 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 US and of the soil pathway over the stomatal and cuticular pathways in the western US that is simulated by WRF/CMAQ and GEM-MACH is less pronounced in the WRF-Chem simulations. Over the EUR 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 are highest for WRF/CMAQ over the northeastern US, followed by GEM-MACH and then WRF-Chem. These features are especially pronounced during summer (Fig. S7). For the EUR 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.

3.2 LU-specific results

Model-to-model differences in domain-average grid-aggregated V_d and pathway contributions may result from not only 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 direct dependence of stomatal resistance on LAI. Another example is the effect

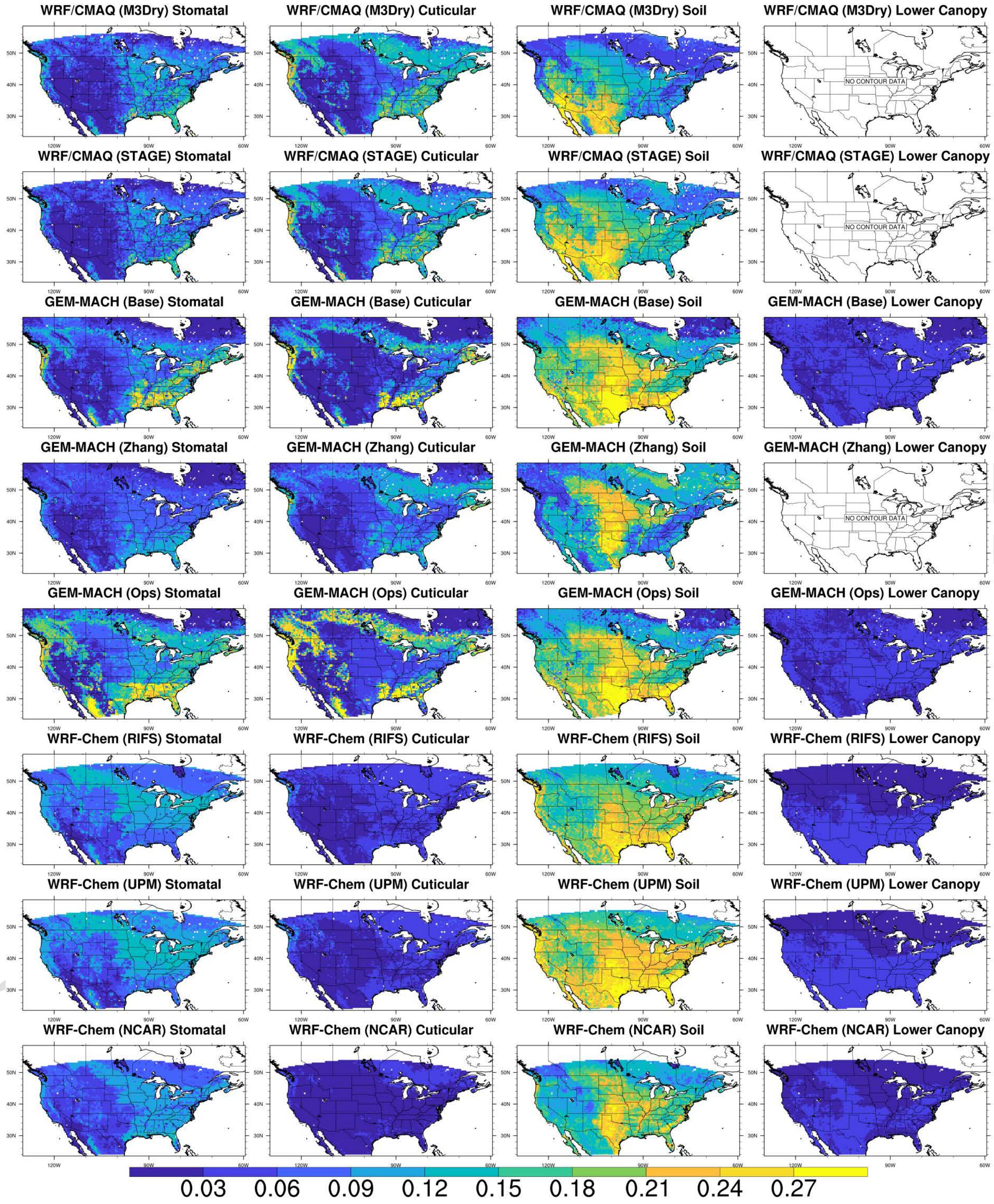

Figure 7. Absolute grid-scale ozone effective conductances (cm s^{-1}), averaged over the entire year. Results are for the NA domain during 2016. Note that these maps are not clipped to the domain common to all simulations and show the maximum spatial extent of non-water cells submitted for each model.

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 are 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 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

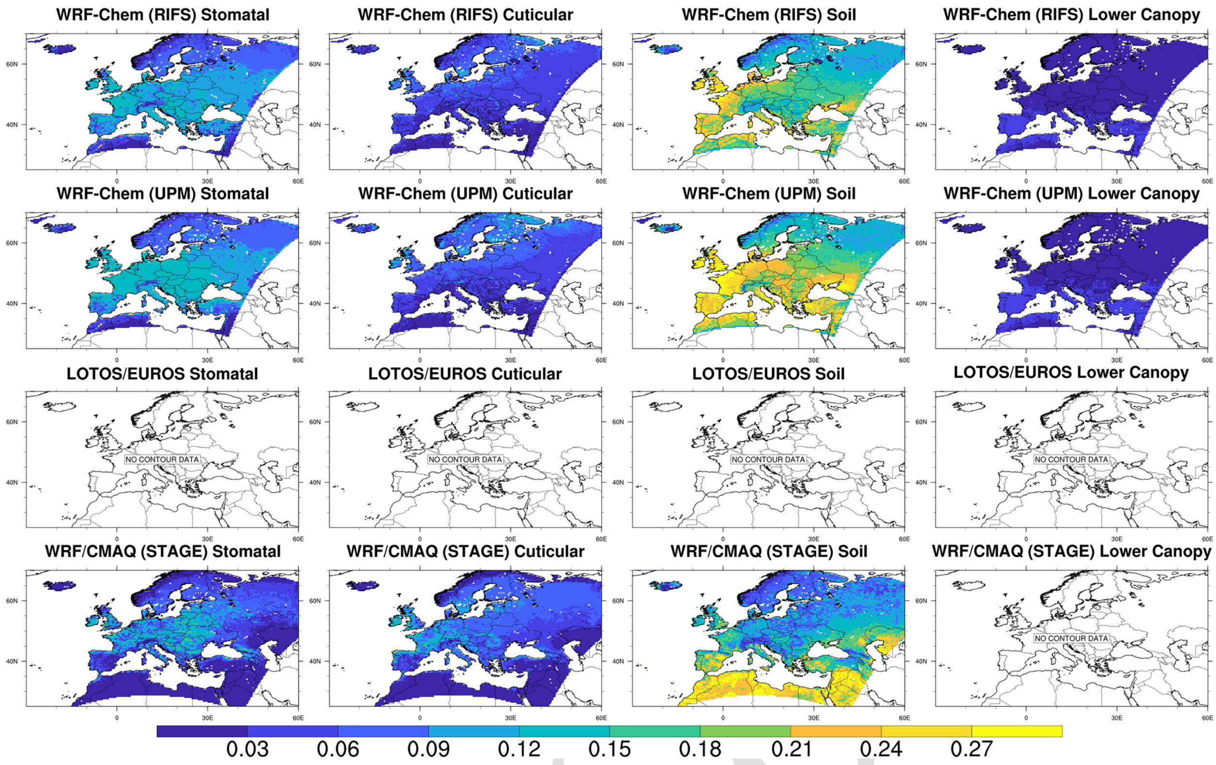
WRF-Chem (RIFS) Stomatal
WRF-Chem (RIFS) Cuticular
WRF-Chem (RIFS) Soil
WRF-Chem (RIFS) Lower Canopy
WRF-Chem (UPM) Stomatal
WRF-Chem (UPM) Cuticular
WRF-Chem (UPM) Soil
WRF-Chem (UPM) Lower Canopy
LOTOS/EUROS Stomatal
LOTOS/EUROS Cuticular
LOTOS/EUROS Soil
LOTOS/EUROS Lower Canopy
WRF/CMAQ (STAGE) Stomatal
WRF/CMAQ (STAGE) Cuticular
WRF/CMAQ (STAGE) Soil
WRF/CMAQ (STAGE) Lower Canopy

Figure 8. Absolute grid-scale ozone effective conductances ($\mathrm{cm\,s^{-1}}$), 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.

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 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 Sect. 3.3). Comparing this figure to the equivalent grid-aggregated results shown in Fig. 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 $\mathrm{cm\,s^{-1}}$ averaged over all common non-water grid cells over NA (Fig. 5), while the range is 0.24–0.71 for V_d for grid cells dominated by the evergreen needleleaf forest category (Fig. 9). Similar increases can also be seen in the ranges of absolute pathway contributions (e.g., soil effective conductance from 0.12–0.21 $\mathrm{cm\,s^{-1}}$ in Fig. 5 to 0.04–0.2 $\mathrm{cm\,s^{-1}}$ in Fig. 9) and relative pathway contributions (e.g., cuticular conductance from 7.6 % to 31.7 % in Fig. 5 to 12.6 %–56.4 % in Fig. 9). Results of individual and summed pathway contributions for a total of eight LU categories over both domains are shown in Table 3a–e. 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 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

Table 3. **(a)** Annual mean grid-scale and LU-specific sum of effective conductances (cm s^{-1}) 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 are shown in italics. Blank cells indicate that data for a given model, pathway, and/or LU were 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). **(b)** As in **(a)** but for the stomatal effective conductances (cm s^{-1}). **(c)** As in **(a)** but for the cuticular effective conductances (cm s^{-1}). **(d)** As in **(a)** but for the soil effective conductances (cm s^{-1}). **(e)** As in **(a)** but for the lower canopy effective conductances (cm s^{-1}).

	Grid scale	URB	BAR	ENF	DBF	MF	SHR	AGR	GRA
(a)									
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
(b)									
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
WRF-Chem (RIFS)	0.09	*0*	0	**0.07**	0.12	0.08	**0.04**	**0.12**	**0.06**
WRF-Chem (UPM)	**0.1**	*0*	0	**0.07**	**0.13**	**0.09**	**0.04**	**0.12**	**0.06**
LOTOS/EUROS									
WRF/CMAQ (STAGE)	*0.06*	**0.01**	0	*0.06*	*0.05*	*0.07*	*0.01*	*0.09*	*0.01*
Range EUR	0.04	0.01	0	0.01	0.08	0.02	0.03	0.03	0.05
(c)									
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	0.01	0.01	0.01	0.01	0.02	0.03	0.03	0.02	0.03

Table 3. Continued.

	Grid scale	URB	BAR	ENF	DBF	MF	SHR	AGR	GRA
(d)									
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	0.07	0.08	0.12	0.08	0.04	0.03	0.02	0.08	0.01
(e)									
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	0	0	0	0	0	0	**0.01**	0.01	0

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 Sect. 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 et al. (2023) 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 Fig. 5 of Clifton et al. (2023), but while 18 point simulations were included in that figure, only the 5 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 Fig. 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 dif-

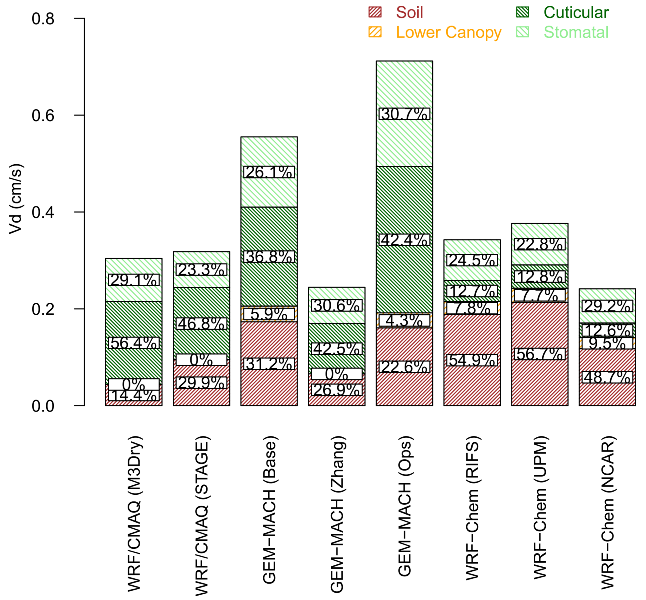

Figure 9. 2016 annual domain-average LU-specific effective conductances and ozone deposition velocities for evergreen needleleaf forest over the NA domain. Averages were calculated over all grid cells in the portion of the analysis domain shared by all model simulations for which a given model had coverage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

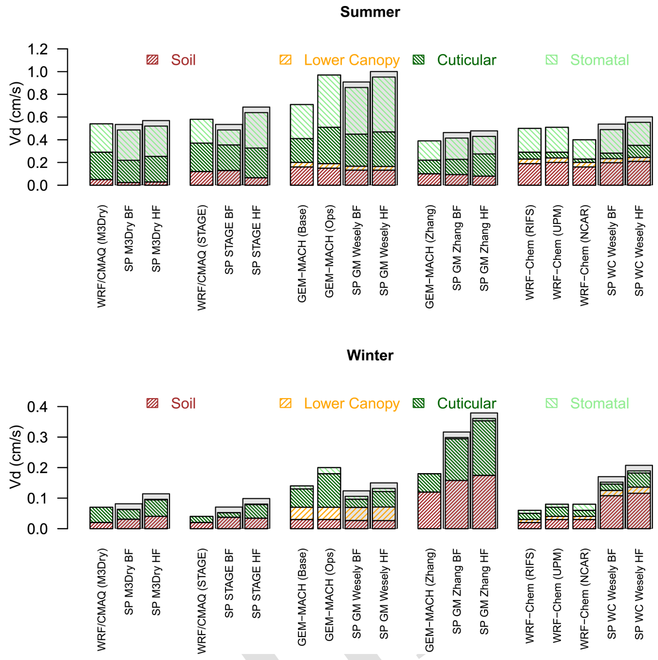

Figure 10. Summer and winter effective conductances and ozone deposition velocities calculated by the grid models for mixed-forest grid cells and calculated by the corresponding subset of single-point (SP) models analyzed in Clifton et al. (2023) at the Borden Forest (BF) and Harvard Forest (HF) sites. The bars for the SP models are overlaid 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 SP GEM-MACH (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 Sect. 3.3).

ferences 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 Fig. 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 values 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.

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 model-to-model variations in pathway contributions. However, the grid model results show a 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, and S11 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 Fig. 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 broadleaf forest in Figs. S9 and S10 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., 2025) 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 Figs. 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.

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 are not dominated by water for the NA and EUR 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

---

[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 (45.812495° N, 8.634771° E) is within a forest of about 300 m extent surrounded by the Joint Research Centre and the town of Ispra, and the Easter Bush field site (55.865° N, 3.206° W) is located 200 m 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 ∼ 800 m from that site, and Bugacpuszta is also more representative of the larger region, which contains largely farmland/grassland with some trees.

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

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: Waterbodies	09: Inland water 10: Seawater
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 needleleaf 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 needleleaf 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	N/A
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	N/A
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

N/A: not available.

and EUR. Over NA, the fractional coverage of some categories like evergreen needleleaf forest is 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 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 frac-

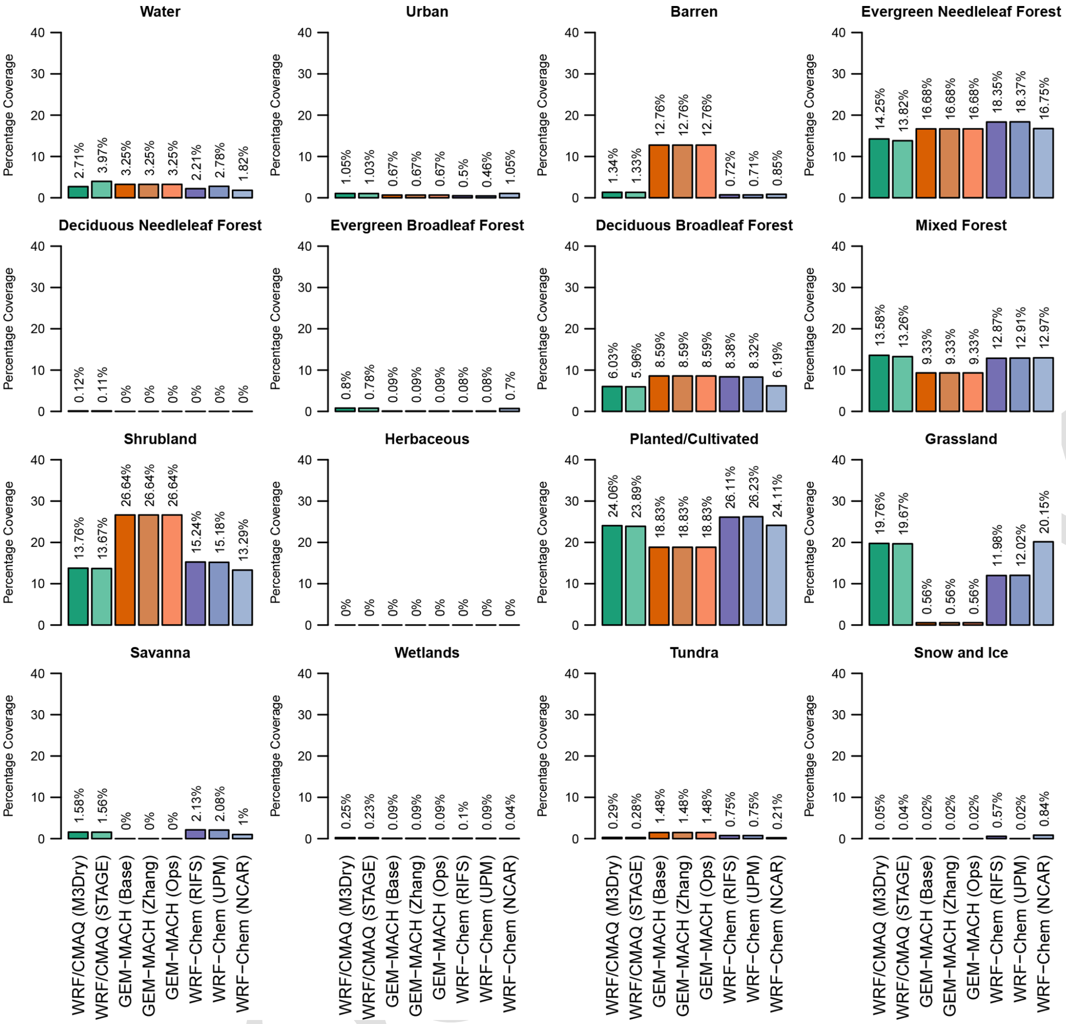

Figure 11. Fraction of the NA domain common to all models covered by each AQMEII4 LU category, excluding grid cells dominated by water by each model.

tions 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, 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, 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 access: 12 May 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 Land Cover 2000 dataset (https://forobs.jrc.ec.europa.eu/glc2000, last access: 12 May 2025)

---

[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.

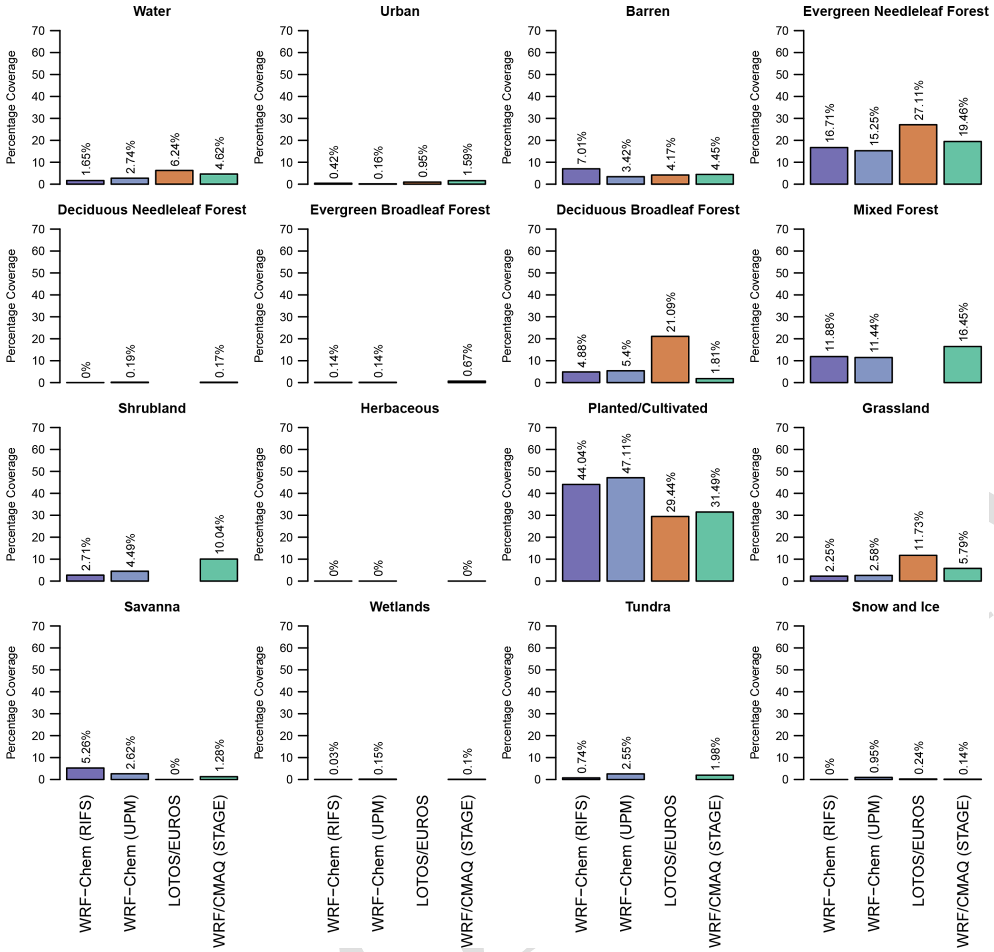

Figure 12. Fraction of the EUR domain common to all models covered by each AQMEII4 LU category, excluding grid cells dominated by water by each model.

mapped to the internal DEPAC categories (Manders-Groot and LOTOS-EUROS team, 2023).

Even for categories for which models show relatively close agreement of the total domain-wide coverage in Figs. 11 and 12, spatial patterns of the coverage may still differ between models. To illustrate this, Figs. S12 and S13 show maps of the fractional coverage of the evergreen needleleaf forest category for each model over NA and EUR. 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, Figs. S12 and S13 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 US in the NA domain (Fig. S12) and central Europe and the Iberian Peninsula in the EUR domain (Fig. S13).

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 determining not only 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

Table 5. Number and percentage of grid cells within the common NA and EUR analysis domains where all models agree on the land use–land cover (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 EUR).

	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 %)	2203 (1.25 %)	1310 (0.74 %)
Evergreen needleleaf forest	7111 (6.58 %)	1544 (1.43 %)	9270 (5.237 %)	2531 (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	2663 (2.46 %)	581 (0.54 %)	362 (0.21 %)	5 (0 %)
Mixed forest	3968 (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 %)	6130 (5.67 %)	24 627 (13.91 %)	6108 (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 (including 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 (including 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 (excluding water cells)	25 662 (32.02 %)	9101 (11.22 %)	36 672 (38.37 %)	9958 (10.08 %)
Total of grid cells with diverging LU (excluding water cells)	54 471 (67.98 %)	72 008 (88.78 %)	58 899 (61.63 %)	88 850 (89.92 %)

also meet the first metric. Identifying grid cells meeting 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 Sect. 3.2 when comparing V_d values 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 EUR are shown in Figs. S14–S15 and Table 5, which lists the number and percentage of grid cells within the common NA and EUR analysis meeting each metric. The first 16 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 EUR), including water grid cells. Water is by far the category with the highest level of agreement over both NA and EUR, with only minor differences between the two metrics, indicating that most grid cells dominated by water 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 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 600–4000 (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 (Figs. 11–12), this is also the case for shrubland and grassland, which have substantial domain-wide coverage over NA (Fig. 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 a 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) 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, 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.

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, Fig. 14 compares LU-specific dry deposition fluxes, V_d, and LU fractions over NA for seven 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 domain are shown in Fig. S16 and confirm that differences in both V_d and LU coverage contribute to differences in LU-specific dry deposition fluxes.

The results shown in Figs. 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, last access: 19 August 2025), such estimates were often 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 paper 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) (grey 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 grey 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 linear scal-