Key Points:
- Thermodynamical versus dynamical sources of biases can be identified using atmospheric nudging of horizontal winds in a climate model
- Atmospheric nudging improves simulated temperature and precipitation in CESM; however, more than half of the initial bias remains
- For temperature and precipitation extremes the larger part of the biases remains after correcting for the dynamics

Supporting Information:
- Supporting Information S1

Correspondence to:
K. Wehrli,
kathrin.wehrli@env.ethz.ch

Assessing the Dynamic Versus Thermodynamic Origin of Climate Model Biases

Kathrin Wehrli[1] [ID], Benoit P. Guillod[1,2] [ID], Mathias Hauser[1] [ID], Matthieu Leclair[1], and Sonia I. Seneviratne[1] [ID]

[1]Institute for Atmospheric and Climate Science, Department of Environmental Systems Science, ETH Zurich, Zurich, Switzerland, [2]Institute for Environmental Decisions, Department of Environmental Systems Science, ETH Zurich, Zurich, Switzerland

Abstract Global climate models present systematic biases, among others, a tendency to overestimate hot and dry summers in midlatitude regions. Here we investigate the origin of such biases in the Community Earth System Model. To disentangle the contribution of dynamics and thermodynamics, we perform simulations that include nudging of horizontal wind and compare them to simulations with a free atmosphere. Prescribing the observed large-scale circulation improves the modeled weather patterns as well as many related fields. However, the larger part of the temperature and precipitation biases of the free atmosphere configuration remains after nudging, in particular, for extremes. Our results suggest that thermodynamical processes, including land-atmosphere coupling and atmospheric parameterizations, drive the errors present in Community Earth System Model. Our result may apply to other climate models and highlight the importance of distinguishing thermodynamic and dynamic sources of biases in present-day global climate models.

Plain Language Summary Global climate models have become indispensable tools to simulate past, present, and future climate. However, present-day models display systematic errors, which are found, for example, in the simulation of surface temperature and precipitation. In this study, we demonstrate that the origin of climate model biases can be more closely isolated, by distinguishing between those due to dynamic processes, that is, related to atmospheric circulation, including wind and pressure systems, and those due to thermodynamic processes, that is, energy exchanges and phase changes, in particular, related to land-atmosphere interactions and convective processes. To separate these two contributions, we prescribe observed winds into simulations with a global climate model, thereby forcing it toward reproducing the correct atmospheric circulation. Strikingly, the largest part of the temperature and precipitation biases are still found when the winds are corrected. This highlights that thermodynamic processes play a dominant role for biases in the representation of today's climate in the employed model. Because the model has similar biases as other climate models, it is likely that these conclusions would also apply there. Hence, in order to make more precise climate projections, it is important that the representation of thermodynamic processes is further improved by the climate modeling community.

1. Introduction

Although large progress has been made during the development of global climate models (GCMs), they still include large biases. Studies analyzing simulations from the Coupled Model Intercomparison Project Phase 5 (CMIP5; K. E. Taylor et al., 2012) revealed systematic errors in simulated global sea surface temperature (SST; Wang et al., 2014), extratropical cyclones (e.g., Zappa et al., 2013), and also processes at the land surface, such as evapotranspiration (Mueller & Seneviratne, 2014) and the seasonal cycle of albedo (Li et al., 2016). It is noteworthy that the main biases in surface temperature and precipitation are fairly consistent between many of the CMIP5 models (Intergovernmental Panel on Climate Change (IPCC), 2013; Mueller & Seneviratne, 2014). It is not trivial to track down where these systematic model biases originate from. First, the manifold nonlinear interactions and feedback processes between the components of the climate system render an identification of biases' origin difficult. Second, the scarce or incomplete observational record for some variables can hinder model evaluation.

A crucial variable driving surface temperature is incoming solar radiation. However, other processes influence the local temperature, such as atmospheric circulation, cloud cover, or the interactions with land and ocean. The simulation of these processes, that is, land surface processes, convection, clouds, and precipitation, relies heavily on parameterizations, introducing additional sources of error (Chen et al., 2010; Gilmore et al., 2004; IPCC, 2013; Neale et al., 2008). Hence, GCM biases can originate from the parameterized processes, as well as the simulation of dynamic and thermodynamic processes. The dynamic component generally refers to the circulation-induced influences, which range from geostrophic wind, weather regimes, and overall large-scale circulation patterns to local wind direction and speed. The thermodynamic component includes phase transitions, differences in lapse rate and land-sea contrast as well as effects from the partitioning of radiative and turbulent fluxes. In practice, it is not simple to separate the contribution of these components, due to their complex interactions with each other (Otto et al., 2016). For example, for precipitation formation or for the buildup of a drought event, both the thermodynamics and dynamics are involved (Schaller et al., 2016; Vautard et al., 2016).

A way to study the thermodynamic contribution in isolation is to remove circulation effects by prescribing a circulation state using the so-called *nudging* approach (Jeuken et al., 1996). Nudging can be used to force winds toward observations, which produces a realistic representation of the observed weather patterns in the model. Nudging has been used in various studies on climate model development and validation (e.g., Feichter & Lohmann, 1999; Jeuken et al., 1996; Kooperman et al., 2012; Lin et al., 2016; Ma et al., 2013; Telford et al., 2008). These studies have shown that nudging reduces model variability and uncertainty without requiring long simulations or large ensembles. One drawback is that nudging is added as a forcing term to the model equations and thus can potentially alter the model climate significantly. For example, Kooperman et al. (2012) show that there are nonnegligible changes to the hydrological cycle and K. Zhang et al. (2014) note differences in the radiation budget. This implies that nudging can potentially generate new biases and produce unrealistic fields. Thus, attention has to be paid to the detailed implementation of nudging and the experiment design (Lin et al., 2016; K. Zhang et al., 2014).

Here we investigate the role of dynamics and thermodynamics for biases in the Community Earth System Model (CESM; Hurrell et al., 2013). CESM is a widely used GCM that is well established in the climate community and has contributed to high-impact studies (Deser et al., 2012; Fischer et al., 2013; Fischer & Knutti, 2014; Sanderson et al., 2017). To separate the dynamic and thermodynamic contribution to the surface climate biases, we apply nudging of the horizontal winds. Hence, conceptually, the model bias is examined conditionally on the observed large-scale circulation to make statements about which contributions drive model bias and where future research might be most beneficial to reduce biases in CESM and potentially other GCMs.

2. Model and Experiments

2.1. Community Earth System Model

Simulations are carried out using version 1.2 of the CESM (Hurrell et al., 2013). SSTs and sea ice fractions are prescribed to transient monthly values from a merged product combining the Hadley Centre sea ice and SST data set version 1 (HadISST1) and the weekly optimum interpolation (OI) SST analysis version 2 by the National Oceanic and Atmospheric Administration (NOAA), as described in Hurrell et al. (2008). CESM couples the Community Atmosphere Model version 5.3 (CAM5; Neale et al., 2012) to predict the Earth's atmosphere and version 4.0 of the Community Land Model (CLM4; Lawrence et al., 2011; Oleson et al., 2010) for the land surface. CAM5 and CLM4 are run on a horizontal resolution of $0.9° \times 1.25°$ with 30 vertical layers in the atmosphere and 15 soil layers. Greenhouse gases are prescribed from transient observed concentrations. We extend CAM5 with a nudging module as described in section 2.2.

An analysis of the model performance reveals that CESM shares systematic biases with the majority of CMIP5 models (see Figures S1a and S1b in the supporting information). The CMIP5 multimodel mean has, for example, a hot and dry bias during Northern Hemisphere summer for Eastern Europe and the Great Plains and a cold and wet bias over the Iberian Peninsula and West Coast of the United States. We find similar biases examining the CMIP5 configuration of CESM model alone. The setup of CESM used as control in this study (see section 2.3 for experiment description) has larger error magnitudes than the CMIP5 multimodel mean due to the effect of compensating biases when averaging over models (Figure S1c).

2.2. Nudging

Nudging (also called Newtonian relaxation) introduces an extra term into the governing equations of the model to nudge the dynamical model variables toward given meteorological fields. We use a similar approach

as used for CAM5 in Kooperman et al. (2012) and K. Zhang et al. (2014). Both studies show that the physical tendencies in the model are not predominated by nudging and that the model climatology is not characteristically different from the free-running model, especially when only winds are nudged. Here nudging is applied to the horizontal wind. Six-hourly fields of zonal and meridional wind are retrieved from the ERA-Interim reanalysis (Dee et al., 2011), which serves as a proxy for the observed circulation. The reanalysis results are regridded to the model resolution and interpolated linearly in time to provide the input fields for each model time step (30 min). The input fields are prescribed by adding a relaxation term to the prognostic equations of horizontal wind:

$$\frac{\partial U}{\partial t} = \dots - \frac{K(z)}{\tau}\Big(U(x, t) - U_{\text{target}}(x, t)\Big) \tag{1}$$

where U stands for the horizontal wind components simulated by the model and U_{target} refers to the respective value from ERA-Interim. Following Kooperman et al. (2012), the nudging time scale τ is 6 hr. We use a height-dependent nudging profile, denoted by $K(z)$, which allows us to prescribe the large-scale circulation to observations (mostly above 700 hPa, see the right panels in Figure S2). By contrast, the boundary layer is not prescribed, allowing surface turbulent fluxes to evolve freely and the atmosphere to interact with the land and equilibrate. Technically, nudging is applied at the end of the CAM model sequence, just before coupling of the atmosphere model with land and ocean (Kooperman et al., 2012).

2.3. Modeling Experiments

CESM is run using prescribed ocean fields. The simulations are initialized in 1979 and run until 2012. Ensemble members are generated by initial perturbation of the temperature fields (Kay et al., 2015). A five-member control ensemble (*CTL*) is run without nudging. The second ensemble of five members is identical to *CTL* except that it has the nudging module switched on during the whole simulation period (*NDG*). As the first three simulation years are regarded as spin-up time, the years analyzed are 1982–2012 (31 years).

2.4. Data Sets Used for Analysis

The performance of CESM with free and nudged atmosphere is assessed by comparing the model output to gridded observational data sets. For comparison of the temperature fields we use monthly data with 0.5° resolution from a climate data set provided by the Climate Research Unit (CRU), called CRU TS (Harris et al., 2014). Precipitation is compared to the Full Data Reanalysis product provided by the Global Precipitation Climatology Centre (GPCC), called GPCC-FD, which has a horizontal resolution of 1° (Becker et al., 2013; Schneider et al., 2014). Surface radiation is compared to two satellite products: the Surface Radiation Budget (SRB) Release-3.0 data set (Stackhouse et al., 2013) and the Clouds and the Earth's Radiant Energy System, Version 3a (CERES) product (Kato et al., 2013). Surface turbulent fluxes are evaluated against data from the Global Soil Wetness Project 2 (GSWP-2; Dirmeyer et al., 2006) while for land evaporation we use the Global Land Evaporation Amsterdam Model, Version 3 (GLEAM) data set (Martens et al., 2017), which gives qualitatively the same results as we get for latent heat flux from GSWP-2. For extreme temperatures we use ERA-Interim as a reference. For the extreme indices related to drought and extreme precipitation we use daily precipitation data from the Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA-2; Gelaro et al., 2017). The gridded data products are remapped to the resolution of CESM using second-order conservative remapping. The analysis uses the ensemble median and excludes ocean points and also Antarctica, unless the latter is covered in the observational data sets. Precipitation values are set to missing where the observed seasonal mean of the daily mean precipitation is smaller than 0.1 mm to avoid displaying large relative errors for very small absolute errors.

3. Results

Nudging forces the speed and direction of horizontal winds toward observations and strongly reduces the variability between the ensemble members of the nudged simulation (not shown). Evaluating the resulting wind fields at the grid point level and the geopotential against ERA-Interim confirms that the large-scale circulation in *NDG* closely follows the prescribed winds (see Figures S2 and S3). The fact that the geopotential and the wind field are close to ERA-Interim suggests that the nudged model is in geostrophic balance. The remainder of this section presents results for the surface climatology of *CTL* and *NDG*.

3.1. Surface Temperature

The seasonal mean surface temperature fields are compared to CRU TS, but qualitatively similar results are obtained using ERA-Interim as reference (not shown). In the first two columns of Figure 1 we show the
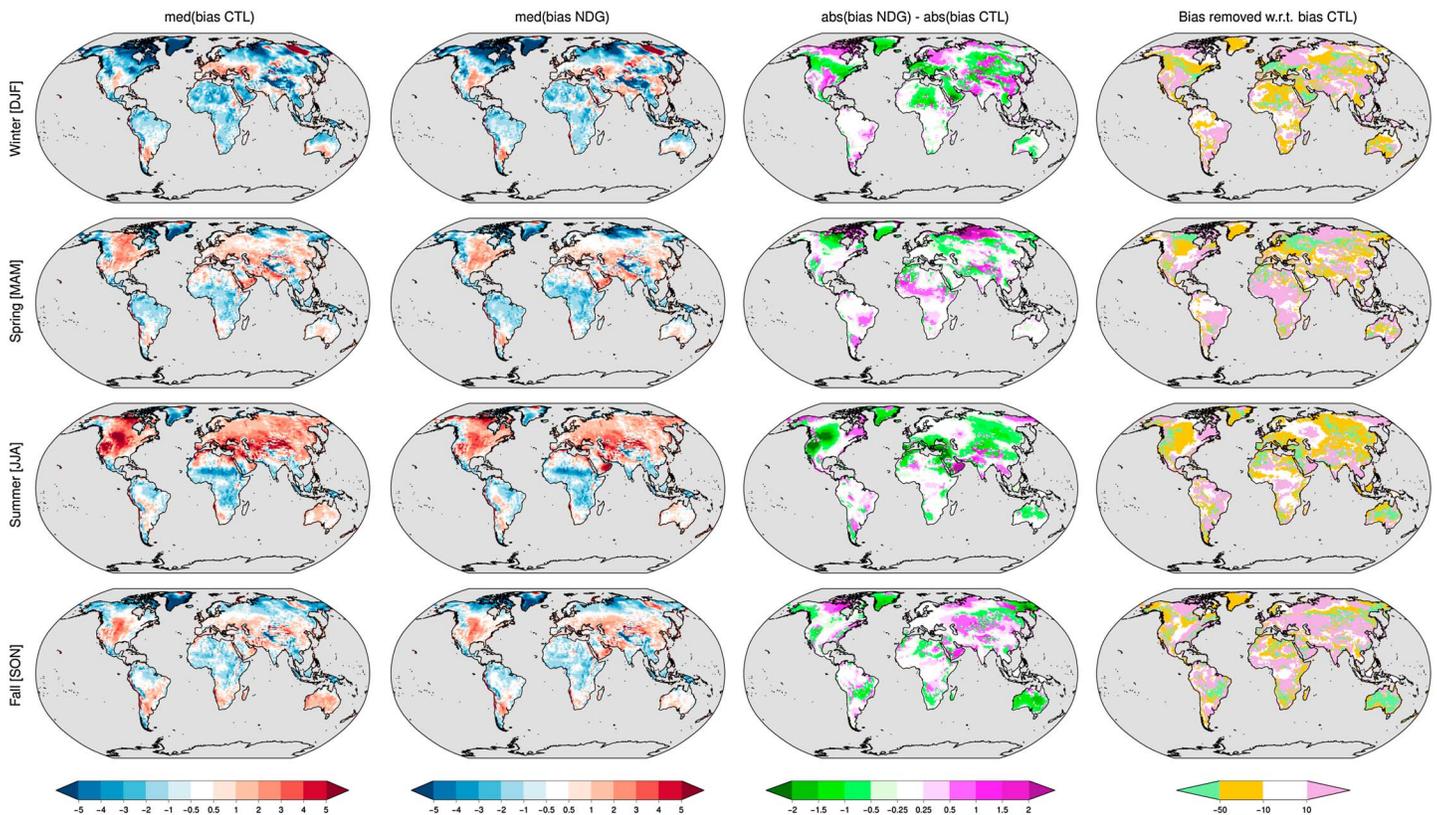
Figure 1. Model biases in 2-m temperature [°C] compared to CRU TS for 1982–2012. From left to right: bias of median from the control simulation (free atmosphere); bias of median from the nudged simulation; comparison of absolute biases in nudged versus control simulation (taking the median bias of both experiments; negative values shown in green indicate smaller biases in *NDG*); Bias removed in the nudged simulation with respect to bias in the control simulation [%].

median bias of the control and the nudged simulations. In many regions the error magnitudes decrease with the introduction of nudging. However, most of the biases remain and the patterns are nearly unchanged. Both ensembles simulate too hot Northern Hemisphere summers (June-July-August, JJA). The third and fourth columns show the absolute and relative difference of the bias between *NDG* and *CTL*. In summer, nudging generally reduces error magnitudes, indicated by predominantly green colors in the third column. But still, in many regions the reduction is less than half of the bias (last column in Figure 1). In autumn (September-October-November, SON) the picture is more mixed and in spring (March-April-May, MAM) a large bias is introduced by nudging in the Arctic. Hence, nudging does not remove most of the GCM biases. Since circulation, for example, frequency of a specific weather regime, can be excluded as contributing factor to the bias in the nudged simulations, there must be non-dynamical mechanisms responsible for the remaining errors.

Local radiative processes as well as turbulent fluxes can be the origin of temperature biases in GCMs. For surface net radiation the comparison against satellite products (e.g., SRB, see Figure S4) shows a positive error for JJA in central Europe and western North America, where also temperatures are too high. Even though the magnitude of this radiation error is sensitive to the choice of the observational product (compare with CERES, Figure S5), there are no substantial differences between the experiments. The patterns and magnitude of the temperature biases found in the simulations are not explained by radiation alone. Misrepresentation of local processes, such as land-atmosphere feedback, might also contribute to these biases. The land surface interacts with the atmosphere in manifold nonlinear ways, mostly via evapotranspiration (ET), which can modulate surface temperature, cloud formation, and precipitation (e.g., Guillod et al., 2015; Koster et al., 2014; Seneviratne et al., 2010; C. M. Taylor et al., 2011). If the model underestimates soil moisture in a region, it will experience more limitation of ET than observed and the partitioning of the surface fluxes will be biased toward sensible heat flux, inducing a warming at the surface (Seneviratne et al., 2010). This overestimation of land-atmosphere coupling is found in GCMs, which can influence their temperature bias (Mueller & Seneviratne, 2014; Sippel
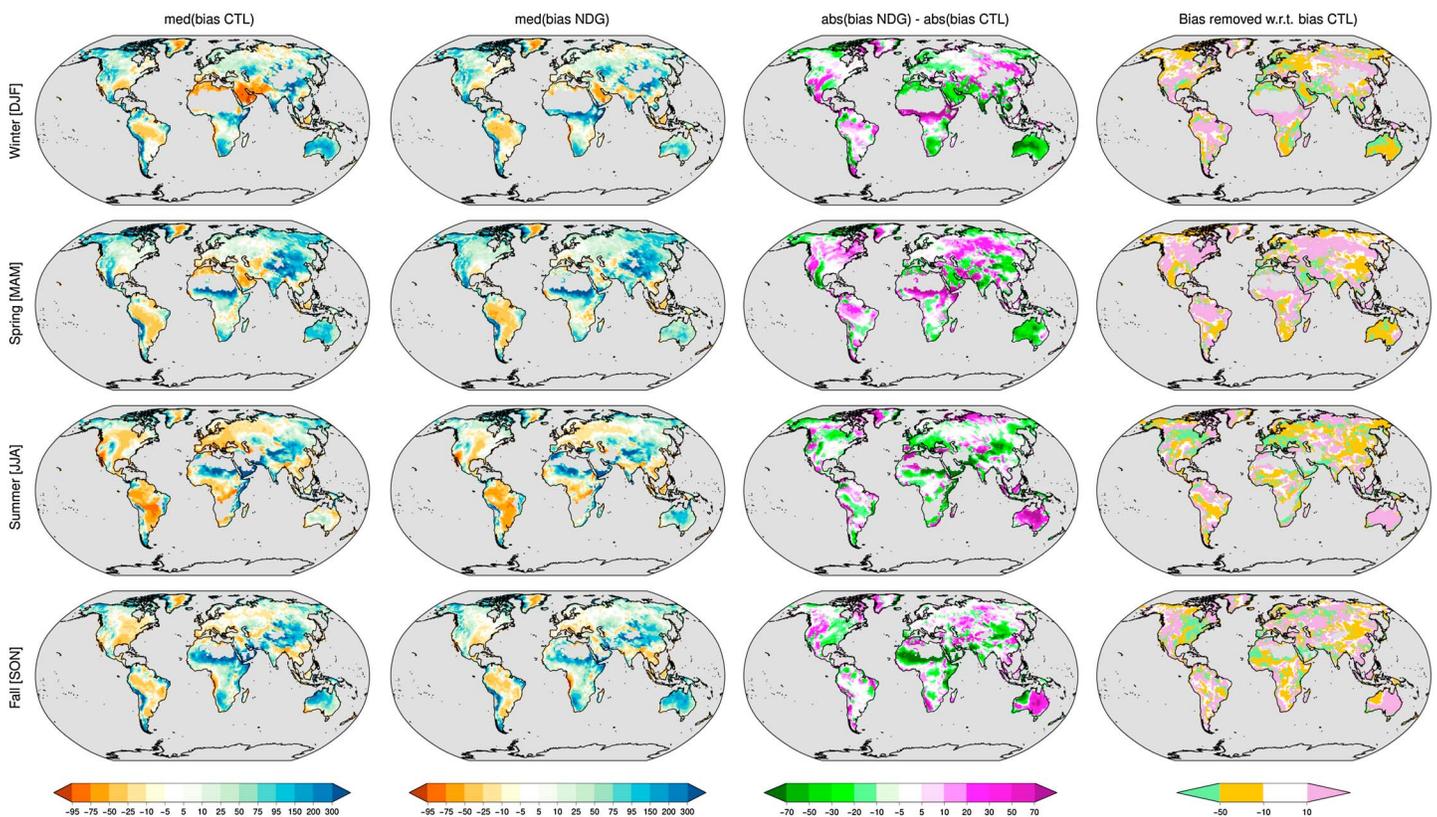
Figure 2. Same as Figure 1 but showing the bias of mean daily precipitation [%] in the model compared to GPCC-FD.

et al., 2017). Using GSWP-2 as a reference, we find an overestimation of sensible heat flux over the European and North American continent for all seasons but strongest during MAM for both experiments (except for the Great Plains and Rocky Mountains, where it is correct or slightly underestimated, see Figure S6). ET in Europe is similar in *CTL* and *NDG*: It is too small in MAM and JJA, especially in Scandinavia and the Mediterranean, but it is better captured in winter (December-January-February) and autumn (SON) (Figure S7). The same applies to the Northern American continent but ET is overestimated over the Rocky Mountains and Great Plains.

A tendency toward too much sensible and too little latent heat flux together with a radiation surplus for JJA in Europe and Northern America may explain the temperature bias of the models. Low values of ET (Figure S7) also suggest that these regions might be too dry in the model. On the other hand, where ET is positively biased, such as in the Great Plains and Rocky Mountains for JJA, we can assume that the simulation is wet enough to keep the temperature close to the observed. The negative temperature bias seen in MAM in the Arctic (Figure 1), does not seem to be related to an issue with the partitioning of turbulent fluxes. Instead, it might originate from a radiation bias, probably related to clouds, or misrepresentation of snow/ice albedo feedback processes.

3.2. Precipitation

Results show a decrease of the area-weighted global mean precipitation from 3.04 mm/day to 2.94 mm/day by nudging, which corresponds to −3.5%. There is a substantial spread in estimates of global mean precipitation from different gauge-based, satellite or reanalysis data sets (Sun et al., 2017). We have computed the global values by regridding the data sets to the model resolution for the ERA-Interim reanalysis: 2.90 mm/day, the Multi-Source Weighted-Ensemble Precipitation (MSWEP) satellite product: 2.54 mm/day and the Global Soil Wetness Project 3 (GSWP 3) multimodel analysis: 2.80 mm/day (for years 1982–2010). Despite the spread within the global estimates, all agree that nudging slightly reduces the bias. Note that the nudged simulations are closer to ERA-Interim, which we nudge toward, than to other reference data sets.

Looking at the seasonal and spatial distribution of the bias in mean daily precipitation in Figure 2 shows that the improvements are not uniform. Summer precipitation (JJA) in Europe is generally underestimated, sup-
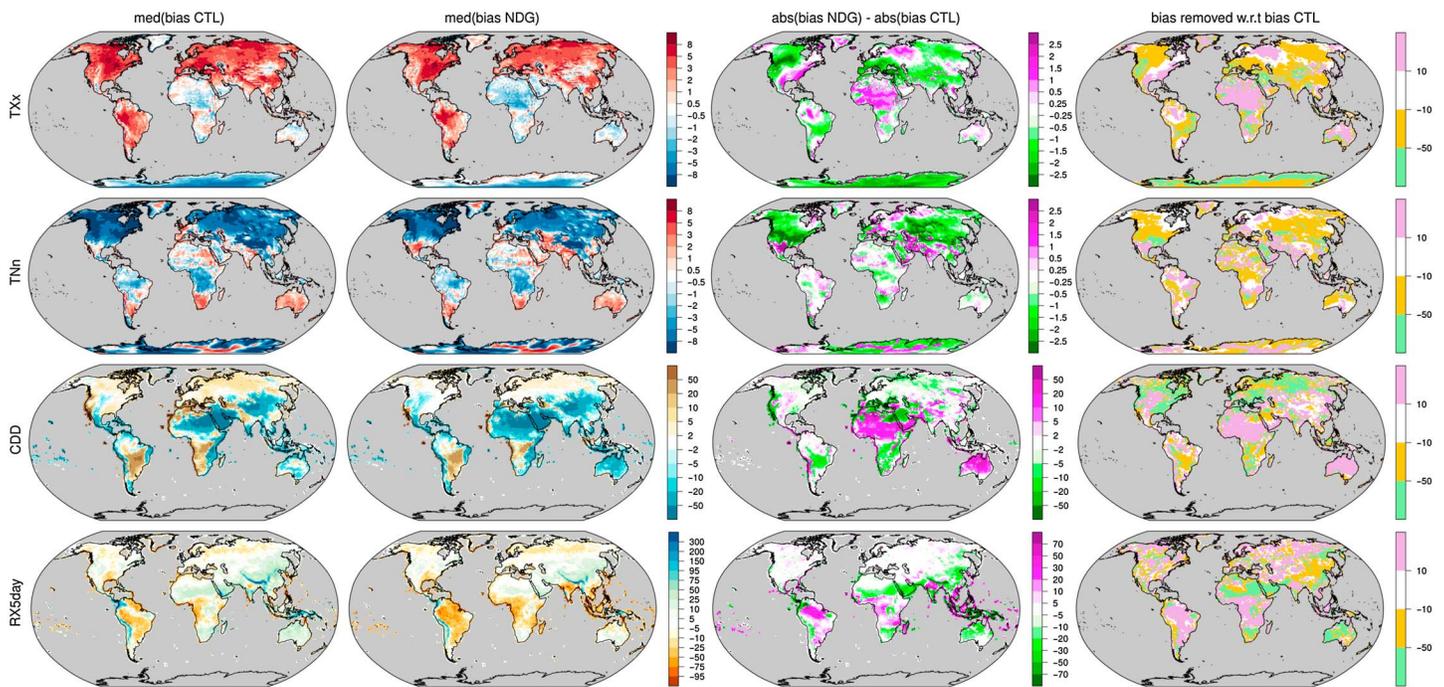
Figure 3. Bias in hot and cold extremes, drought, and precipitation indices computed as 31-year average (taken over 1982-2012) of the yearly indices. From left to right: as in Figure 1. Shown from top to bottom are the following: Yearly maximum of the daily maximum temperature (*TXx*) compared to ERA-Interim; Yearly minimum of the daily minimum temperature (*TNn*) compared to ERA-Interim; Number of consecutive dry days in a year (*CDD*) compared to MERRA-2; Yearly maximum 5-day accumulated precipitation (*RX5day*) compared to MERRA-2. The ocean is masked for *TXx* and *TNn* for better readability.

porting our interpretation of the temperature results (Figures 1 and S7). Most of this bias is smaller in *NDG* than in *CTL* for Central Europe but not for Northern Europe (where the bias is of similar magnitude) and for the Iberian Peninsula (where the bias is larger in *NDG* due to overcompensation toward a too wet state). For many regions the changes are small compared to the total bias, and in some cases *NDG* performs worse than *CTL*, as illustrated in the third and fourth columns of Figure 2. Bias reduction in *NDG* is limited to only a few regions (shown with green color in the fourth column). Other mechanisms must be the main contributors to the precipitation bias in other regions.

3.3. Global Mean Error

To quantify the impact of nudging on model performance at the global scale, we compute the spatial root-mean-square error (RMSE) for global land areas (excluding Antarctica). Generally, nudging leads to only a small reduction in model biases. The results for surface air temperature show that the global bias is reduced by 3% to 11% in *NDG* for all seasons, except MAM (see Table S1). Similarly, for mean daily precipitation *NDG* performs better than *CTL* for all seasons except MAM.

We also compute the RMSE for the tropics (30°S–30°N) and northern midlatitudes (35°N–66°N, see Tables S2 and S3). Tropical regions generally have a smaller RMSE of surface temperature than the global average and a larger RMSE of precipitation. The opposite is found for the northern midlatitudes. Of the two experiments, *NDG* generally performs better than *CTL* for both precipitation and temperature, except for MAM in the tropics. A comparison with the biases shown in Figure 2 confirms that nudging rather degrades the performance of the model for MAM, especially in the tropics. Constant solar heating and convergence of air masses in the tropics generate convective systems. Although there is an indirect influence on tropical precipitation through the large-scale winds generating pressure systems, there is no direct constraint on the small scale convective cells in our nudging approach. Substantial biases originating from the simulation of convection might be reflected in the global performance of the model.

3.4. Extreme Indices

Extreme events, such as heavy precipitation, heat waves or droughts have large societal and economic impacts (Barriopedro et al., 2011; Gilbert, 2010; IPCC, 2014; Poumadère et al., 2005). As extreme events are usually linked with particular synoptic circulation patterns, such as blockings for heat waves (Quesada et al.,

2012), one might expect nudging to improve the simulation of extreme statistics. To examine whether this is indeed the case, we calculate four extreme indices following the definition by X. Zhang et al. (2011). Biases in extremes are generally larger than biases in seasonal averages, as shown in Figure 3. The first row shows the yearly maximum of the daily maximum temperature (TXx) and the second row the yearly minimum of the daily minimum temperature (TNn). The temperature of hot extremes is generally overestimated, except in Australia and Africa, and while there is improvement in the nudged approach, this is only a small error reduction compared to the total bias. Cold extremes, on the other hand, are generally simulated too cold, especially in the northern middle and high latitudes. Again, the errors are comparable between the two experiments. The number of consecutive dry days (CDD), shown in the third row, is too large in the northern high latitudes and rather too small in the low latitudes. Similarly, the maximum 5-day precipitation amount (RX5day; last row in Figure 3) is slightly overestimated in the midlatitudes and underestimated in the tropics and high latitudes. RX5day and CDD together indicated that climate extremes in North America and northern Europe are simulated too dry because of too long dry periods and also too little precipitation during intense events.

4. Discussion and Conclusions

In this study we nudge atmospheric winds to reanalysis, with the aim to disentangle the contributions of dynamics and thermodynamics to model biases in a GCM. Nudging has been used previously in model development and validation studies, which showed that the large-scale circulation is reliably adjusted to observations while the underlying model climatology is well retained (Lin et al., 2016; K. Zhang et al., 2014). Here we show that even if the dynamic component is prescribed, large errors in the surface fields persist in the CESM model. A hot and dry bias seen in Europe and Northern America for northern midlatitude summers is partly reduced by nudging, but the bias pattern remains common to both experiments.

We hypothesize that a large part of the errors that remain in the nudged simulations has a thermodynamic source and is caused by local processes. We find evidence for this hypothesis in the partitioning of the surface heat fluxes and ET. Computing the land-atmosphere coupling strength as the correlation of JJA mean temperature with latent heat flux shows that the model exhibits too strong soil moisture limitation in eastern Europe and central North America (not shown). This bias is improved but not removed in the nudged experiment, providing one possible explanation on why these regions belong to the hot spots of summer temperature error for both experiments. For extreme temperatures and droughts as well as heavy precipitation events, we find that model biases are even larger than for seasonal averages and that the effect of prescribing the atmospheric circulation is small.

In the tropics, a substantial part of the biases might originate from the convection parameterizations. Convective cells, leading to most of tropical precipitation, are not directly constrained by nudging. Additionally, nudging can introduce large biases in the model climatology, as found for temperature in the Arctic for MAM. This effect suggests that nudging interferes with model tuning, which is used to adjust the top of atmosphere radiation balance but might not match with the prescribed circulation. Compensating errors might indeed exist in the setup with free atmosphere, in which case nudging corrects one of the errors, leading to new apparent biases. In such a case, the results might appear more biased although the model simulation is in fact more physically correct.

In conclusion, conducting CESM experiments with atmospheric nudging reveals that thermodynamic processes are a dominant source of biases in the model. As this study was carried out using a single model, different results might be found with other models. However, it is also possible that the conclusions would apply to other GCMs, since the highlighted model biases (in particular, too dry and hot conditions in midlatitude summer) have been shown to be systematically found in present-day GCMs. Our results show the strengths and limitations of atmospheric nudging for climate simulations and put the role of thermodynamics for biases in climate modeling into a new light. They call for multimodel investigations assessing the robustness of the dominance of thermodynamics for temperature and precipitation biases across GCMs. Although circulation changes are an important source of uncertainty in climate predictions (Hall, 2014; Shepherd, 2014; Xie et al., 2015), our findings show a possibly dominant role of biases in thermodynamical processes for the representation of present-day climate. This highlights the urgent need

to reduce the associated uncertainty in climate models in order to achieve better constrained climate change projections.

References

Barriopedro, D., Fischer, E. M., Luterbacher, J., Trigo, R. M., & García-Herrera, R. (2011). The hot summer of 2010: Redrawing the temperature record map of Europe. *Science*, *332*(6026), 220–224. https://doi.org/10.1126/science.1201224

Becker, A., Finger, P., Meyer-Christoffer, A., Rudolf, B., Schamm, K., Schneider, U., & Ziese, M. (2013). A description of the global land-surface precipitation data products of the Global Precipitation Climatology Centre with sample applications including centennial (trend) analysis from 1901–present. *Earth System Science Data*, *5*(1), 71–99. https://doi.org/10.5194/essd-5-71-2013

Chen, H., Zhou, T., Neale, R. B., Wu, X., & Zhang, G. J. (2010). Performance of the new NCA RCAM3.5 in East Asian summer monsoon simulations: Sensitivity to modifications of the convection scheme. *Journal Climate*, *23*(13), 3657–3675. https://doi.org/10.1175/2010JCLI3022.1

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., et al. (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, *137*(656), 553–597. https://doi.org/10.1002/qj.828

Deser, C., Knutti, R., Solomon, S., & Phillips, A. S. (2012). Communication of the role of natural variability in future North American climate. *Nature Climate Change*, *2*, 775–779. https://doi.org/10.1038/nclimate1562

Dirmeyer, P. A., Gao, X., Zhao, M., Guo, Z., Oki, T., & Hanasaki, N. (2006). GSWP-2: Multimodel analysis and implications for our perception of the land surface. *Bulletin of the American Meteorological Society*, *87*(10), 1381–1398. https://doi.org/10.1175/BAMS-87-10-1381

Feichter, J., & Lohmann, U. (1999). Can a relaxation technique be used to validate clouds and sulphur species in a GCM? *Quarterly Journal of the Royal Meteorological Society*, *125*(556), 1277–1294. https://doi.org/10.1002/qj.1999.49712555609

Fischer, E. M., Beyerle, U., & Knutti, R. (2013). Robust spatially aggregated projections of climate extremes. *Nature Climate Change*, *3*, 1033–1038. https://doi.org/10.1038/nclimate2051

Fischer, E. M., & Knutti, R. (2014). Detection of spatially aggregated changes in temperature and precipitation extremes. *Geophysical Research Letters*, *41*, 547–554. https://doi.org/10.1002/2013GL058499

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., et al. (2017). The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). *Journal Climate*, *30*(14), 5419–5454. https://doi.org/10.1175/JCLI-D-16-0758.1

Gilbert, N. (2010). Russia counts environmental cost of wildfires. *Nature*. https://doi.org/10.1038/news.2010.404

Gilmore, M. S., Straka, J. M., & Rasmussen, E. N. (2004). Precipitation uncertainty due to variations in precipitation particle parameters within a simple microphysics scheme. *Monthly Weather Review*, *132*(11), 2610–2627. https://doi.org/10.1175/MWR2810.1

Guillod, B. P., Orlowsky, B., Miralles, D. G., Teuling, A. J., & Seneviratne, S. I. (2015). Reconciling spatial and temporal soil moisture effects on afternoon rainfall. *Nature Communications*, *6*, 6443. https://doi.org/10.1038/ncomms7443

Hall, A. (2014). Projecting regional change. *Science*, *346*(6216), 1461–1462. https://doi.org/10.1126/science.aaa0629

Harris, I., Jones, P., Osborn, T., & Lister, D. (2014). Updated high-resolution grids of monthly climatic observations—The CRU TS3.10 dataset. *International Journal of Climatology*, *34*(3), 623–642. https://doi.org/10.1002/joc.3711

Hurrell, J. W., Holland, M. M., Gent, P. R., Ghan, S., Kay, J. E., Kushner, P. J., et al. (2013). The community Earth system model: A framework for collaborative research. *Bulletin of the American Meteorological Society*, *94*(9), 1339–1360. https://doi.org/10.1175/BAMS-D-12-00121.1

Hurrell, J. W., Hack, J. J., Shea, D., Caron, J. M., & Rosinski, J. (2008). A new sea surface temperature and sea ice boundary dataset for the community atmosphere model. *Journal Climate*, *21*(19), 5145–5153. https://doi.org/10.1175/2008JCLI2292.1

IPCC (2014). Climate change 2014: Synthesis report. In R. K. Pachauri & L. A. Meyer (Eds.), *Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (151 pp.). Geneva, Switzerland: IPCC.

IPCC (2013). Evaluation of Climate Models. In T. F. Stocker, et al. (Eds.), *Climate change 2013, the physical science basis. Working group I contribution of to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 659–740). Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press.

Jeuken, A. B. M., Siegmund, P. C., Heijboer, L. C., Feichter, J., & Bengtsson, L. (1996). On the potential of assimilating meteorological analyses in a global climate model for the purpose of model validation. *Journal of Geophysical Research*, *101*(D12), 16,939–16,950. https://doi.org/10.1029/96JD01218

Kato, S., Loeb, N. G., Rose, F. G., Doelling, D. R., Rutan, D. A., Caldwell, T. E., et al. (2013). Surface irradiances consistent with CERES-derived top-of-atmosphere shortwave and longwave irradiances. *Journal Climate*, *26*(9), 2719–2740. https://doi.org/10.1175/JCLI-D-12-00436.1

Kay, J. E., Deser, C., Phillips, A., Mai, A., Hannay, C., Strand, G., et al. (2015). The community Earth system model (CESM) large ensemble project: A community resource for studying climate change in the presence of internal climate variability. *Bulletin of the American Meteorological Society*, *96*(8), 1333–1349. https://doi.org/10.1175/BAMS-D-13-00255.1

Kooperman, G. J., Pritchard, M. S., Ghan, S. J., Wang, M., Somerville, R. C. J., & Russell, L. M. (2012). Constraining the influence of natural variability to improve estimates of global aerosol indirect effects in a nudged version of the community atmosphere model 5. *Journal of Geophysical Research*, *117*, D23204. https://doi.org/10.1029/2012JD018588

Koster, R. D., Chang, Y., & Schubert, S. D. (2014). A mechanism for land–atmosphere feedback involving planetary wave structures. *Journal Climate*, *27*(24), 9290–9301. https://doi.org/10.1175/JCLI-D-14-00315.1

Lawrence, D. M., Oleson, K. W., Flanner, M. G., Thornton, P. E., Swenson, S. C., Lawrence, P. J., et al. (2011). Parameterization improvements and functional and structural advances in version 4 of the community land model. *Journal of Advances in Modeling Earth Systems*, *3*, M03001. https://doi.org/10.1029/2011MS00045

Li, Y., Wang, T., Zeng, Z., Peng, S., Lian, X., & Piao, S. (2016). Evaluating biases in simulated land surface albedo from CMIP5 global climate models. *Journal of Geophysical Research: Atmospheres*, *121*, 6178–6190. https://doi.org/10.1002/2016JD024774

Lin, G., Wan, H., Zhang, K., Qian, Y., & Ghan, S. J. (2016). Can nudging be used to quantify model sensitivities in precipitation and cloud forcing? *Journal of Advances in Modeling Earth Systems*, *8*, 1073–1091. https://doi.org/10.1002/2016MS000659

Ma, P.-L., Rasch, P. J., Wang, H., Zhang, K., Easter, R. C., Tilmes, S., et al. (2013). The role of circulation features on black carbon transport into the Arctic in the community atmosphere model version 5 (CAM5). *Journal of Geophysical Research: Atmospheres*, *118*, 4657–4669. https://doi.org/10.1002/jgrd.50411

Martens, B., Miralles, D. G., Lievens, H., van der Schalie, R., de Jeu, R. A. M., Fernández-Prieto, D., et al. (2017). GLEAM V3: Satellite-based land evaporation and root-zone soil moisture. *Geoscientific Model Development*, *10*(5), 1903–1925. https://doi.org/10.5194/gmd-10-1903-2017

Mueller, B., & Seneviratne, S. I. (2014). Systematic land climate and evapotranspiration biases in CMIP5 simulations. *Geophysical Research Letters*, *41*, 128–134. https://doi.org/10.1002/2013GL058055

Neale, R. B., Chen, C.-C., Gettelman, A., Lauritzen, P. H., Park, S., Williamson, D. L., et al. (2012). Description of the NCAR community atmosphere model (CAM 5.0) (Tech. Rep. NCAR/TN-486+STR). Boulder, CO: National Center for Atmospheric Research.

Neale, R. B., Richter, J. H., & Jochum, M. (2008). The impact of convection on ENSO: From a delayed oscillator to a series of events. *Journal Climate*, *21*(22), 5904–5924. https://doi.org/10.1175/2008JCLI2244.1

Oleson, K. W., Lawrence, D. M., Bonan, G. B., Flanner, M. G., Kluzek, E., Lawrence, P. J., et al. (2010). Technical description of version 4.0 of the community land model (CLM) (Tech. Rep. NCAR/TN-478+STR). Boulder, CO: National Center for Atmospheric Research.

Otto, F. E. L., van Oldenborgh, G. J., Eden, J., Stott, P. A., Karoly, D. J., & Allen, M. R. (2016). The attribution question. *Nature Climate Change*, *6*(9), 813–816. https://doi.org/10.1038/nclimate3089

Poumadère, M., Mays, C., Le Mer, S., & Blong, R. (2005). The 2003 heat wave in France: Dangerous climate change here and now. *Risk Analysis*, *25*(6), 1483–1494. https://doi.org/10.1111/j.1539-6924.2005.00694.x

Quesada, B., Vautard, R., Yiou, P., Hirschi, M., & Seneviratne, S. I. (2012). Asymmetric european summer heat predictability from wet and dry southern winters and springs. *Nature Climate Change*, *2*(10), 736–741. https://doi.org/10.1038/nclimate1536

Sanderson, B. M., Xu, Y., Tebaldi, C., Wehner, M., O'Neill, B., Jahn, A., et al. (2017). Community climate simulations to assess avoided impacts in 1.5 and 2°C futures. *Earth System Dynamics*, *8*(3), 827–847. https://doi.org/10.5194/esd-8-827-2017

Schaller, N., Kay, A. L., Lamb, R., Massey, N. R., van Oldenborgh, G. J., Otto, F. E. L., et al. (2016). Human influence on climate in the 2014 Southern England winter floods and their impacts. *Nature Climate Change*, *6*(6), 627–634. https://doi.org/10.1038/nclimate2927

Schneider, U., Becker, A., Finger, P., Meyer-Christoffer, A., Ziese, M., & Rudolf, B. (2014). GPCC's new land surface precipitation climatology based on quality-controlled in situ data and its role in quantifying the global water cycle. *Theoretical and Applied Climatology*, *115*(1), 15–40. https://doi.org/10.1007/s00704-013-0860-x

Seneviratne, S. I., Corti, T., Davin, E. L., Hirschi, M., Jaeger, E. B., Lehner, I., et al. (2010). Investigating soil moisture–climate interactions in a changing climate: A review. *Earth-Science Reviews*, *99*(3–4), 125–161. https://doi.org/10.1016/j.earscirev.2010.02.004

Shepherd, T. G. (2014). Atmospheric circulation as a source of uncertainty in climate change projections. *Nature Geoscience*, *7*(10), 703–708. https://doi.org/10.1038/ngeo2253

Sippel, S., Zscheischler, J., Mahecha, M. D., Orth, R., Reichstein, M., Vogel, M., & Seneviratne, S. I. (2017). Refining multi-model projections of temperature extremes by evaluation against land–atmosphere coupling diagnostics. *Earth System Dynamics*, *8*(2), 387–403. https://doi.org/10.5194/esd-8-387-2017

Stackhouse, P., Gupta, S., Hall, F., Collatz, G., Meeson, B., Los, S., et al. (2013). ISLSCP II surface radiation budget (SRB) radiation data. https://doi.org/10.3334/ornldaac/1201

Sun, Q., Miao, C., Duan, Q., Ashouri, H., Sorooshian, S., & Hsu, K. (2017). A review of global precipitation data sets: Data sources, estimation, and intercomparisons. *Reviews of Geophysics*, *56*, 79–107. https://doi.org/10.1002/2017RG000574

Taylor, C. M., Gounou, A., Guichard, F., Harris, P. P., Ellis, R. J., Couvreux, F., & De Kauwe, M. (2011). Frequency of Sahelian storm initiation enhanced over mesoscale soil-moisture patterns. *Nature Geoscience*, *4*(7), 430–433. https://doi.org/10.1038/ngeo1173

Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, *93*(4), 485–498. https://doi.org/10.1175/BAMS-D-11-00094.1

Telford, P. J., Braesicke, P., Morgenstern, O., & Pyle, J. A. (2008). Technical note: Description and assessment of a nudged version of the new dynamics unified model. *Atmospheric Chemistry and Physics*, *8*(6), 1701–1712. https://doi.org/10.5194/acp-8-1701-2008

Vautard, R., Yiou, P., Otto, F., Stott, P., Christidis, N., van Oldenborgh, G. J., & Schaller, N. (2016). Attribution of human-induced dynamical and thermodynamical contributions in extreme weather events. *Environmental Research Letters*, *11*(11), 114009. https://doi.org/10.1088/1748-9326/11/11/114009

Wang, C., Zhang, L., Lee, S.-K., Wu, L., & Mechoso, C. R. (2014). A global perspective on CMIP5 climate model biases. *Nature Climate Change*, *4*(3), 201–205. https://doi.org/10.1038/nclimate2118

Xie, S.-P., Deser, C., Vecchi, G. A., Matthew, C., Delworth, T. L., Hall, A., et al. (2015). Towards predictive understanding of regional climate change. *Nature Climate Change*, *5*(921). https://doi.org/10.1038/nclimate2689

Zappa, G., Shaffrey, L. C., & Hodges, K. I. (2013). The ability of CMIP5 models to simulate North Atlantic extratropical cyclones. *Journal Climate*, *26*(15), 5379–5396. https://doi.org/10.1175/JCLI-D-12-00501.1

Zhang, X., Alexander, L., Hegerl, G. C., Jones, P., Tank, A. K., Peterson, T. C., et al. (2011). Indices for monitoring changes in extremes based on daily temperature and precipitation data. *Wiley Interdisciplinary Reviews: Climate Change*, *2*(6), 851–870. https://doi.org/10.1002/wcc.147

Zhang, K., Wan, H., Liu, X., Ghan, S. J., Kooperman, G. J., Ma, P.-L., et al. (2014). Technical note: On the use of nudging for aerosol–climate model intercomparison studies. *Atmospheric Chemistry and Physics*, *14*(16), 8631–8645. https://doi.org/10.5194/acp-14-8631-2014