

JGR Atmospheres

RESEARCH ARTICLE

10.1029/2019JD031286

Key Points:

- A 244-member ensemble is used to show internal model variability within the Weather Research and Forecasting model
- Over 8 °C hourly differences between members were found (median 1.2 °C) simply by changing the model start times
- This overlooked feature has implications for how users should run one of the most widely used regional climate models

Supporting Information:

- Supporting Information S1
- Data Set S1

Correspondence to:

R. Bassett,
r.bassett@lancaster.ac.uk

Citation:

Bassett, R., Young, P. J., Blair, G. S., Samreen, F., & Simm, W. (2020). A large ensemble approach to quantifying internal model variability within the WRF numerical model. *Journal of Geophysical Research: Atmospheres*, 125, e2019JD031286. <https://doi.org/10.1029/2019JD031286>

Received 4 JUL 2019

Accepted 6 MAR 2020

Accepted article online 9 MAR 2020

A Large Ensemble Approach to Quantifying Internal Model Variability Within the WRF Numerical Model

R. Bassett¹ , P. J. Young^{1,2} , G. S. Blair^{3,2} , F. Samreen³ , and W. Simm³ 

¹Lancaster Environment Centre, Lancaster University, Lancaster, UK, ²Data Science Institute, Lancaster University, Lancaster, UK, ³School of Computing and Communications, Lancaster University, Lancaster, UK

Abstract The Weather Research and Forecasting (WRF) community model is widely used to explore cross-scale atmospheric features. Although WRF uncertainty studies exist, these usually involve ensembles where different physics options are selected (e.g., the boundary layer scheme) or adjusting individual parameters. Uncertainty from perturbing initial conditions, which generates internal model variability (IMV), has rarely been considered. Moreover, many off-line WRF research studies generate conclusions based on a single model run without addressing any form of uncertainty. To demonstrate the importance of IMV, or noise, we present a 4-month case study of summer 2018 over London, UK, using a 244-member initial condition ensemble. Simply by changing the model start time, a median 2-m temperature range or IMV of 1.2 °C was found (occasionally exceeding 8 °C). During our analysis, episodes of high and low IMV were found for all variables explored, explained by a relationship with the boundary condition data. Periods of slower wind speed input contained increased IMV, and vice versa, which we hypothesis is related to how strongly the boundary conditions influence the nested region. We also show the importance of IMV effects for the uncertainty of derived variables like the urban heat island, whose median variation in magnitude is 1 °C. Finally, a realistic ensemble size to capture the majority of WRF IMV is also estimated, essential considering the high computational overheads (244 members equaled 140,000 CPU hours). We envisage that highlighting considerable IMV in this repeatable manner will help advance best practices for the WRF and wider regional climate modeling community.

Plain Language Summary The Weather Research and Forecasting model is typically used to dynamically downscale atmospheric conditions for limited geographical regions (e.g., southern United Kingdom), using one or more nested grids of increasing resolution, and with externally provided boundary conditions (e.g., from reanalysis data). Research applications of the model (i.e., separate to forecasting) are wide ranging from hurricanes to urban heat islands. However, despite the model being constrained at its boundaries, subtle modifications to initial conditions (e.g., soil moisture) may propagate and lead to different results. We generate an ensemble of 244 different initial conditions simply by changing the model start time and show large differences in modeled temperature and other meteorological fields across a fixed analysis period. This effect, rarely considered by the Weather Research and Forecasting modeling community (perhaps due to large computational requirements), should be reflected in all future research.

1. Introduction

Regional climate models (RCMs), through dynamic downscaling over limited areas, allow small-scale features to be resolved at high resolutions without the large computational overheads of global climate models (GCMs). They are typically employed to explore subgrid-scale processes, such as urban heat islands (UHIs) or heavy precipitation events. RCM boundary conditions, taken from GCMs or reanalysis (observation-based) products, should largely constrain internal model variability (IMV) or noises; the overall RCM climatology will be comparable to the boundary conditions. However, initial states, code approximations to mathematical formula, and nonlinear physical processes may drive RCM IMV (Christensen et al., 2001). Despite this, IMV has yet to be fully investigated by the RCM community (Giorgi, 2019; Laprise et al., 2012). Moreover, RCM studies often draw conclusions without exploring model uncertainty (e.g., El Afandi et al., 2013; Sertel et al., 2011; Tse et al., 2018; Zhang et al., 2011; Zhou et al., 2016). Considering that IMV is reflected in RCM output (Giorgi & Bi, 2000), results from such single-run studies should at best be seen as

©2020. The Authors.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

contingent. Furthermore, where differences between single model runs and observations are interpreted as error, these may in fact be due to imperfect initial conditions (Laprise et al., 2012). Here, we quantify IMV in the Weather Research and Forecasting (WRF) model (Skamarock et al., 2008), perhaps the most widely used mesoscale numerical weather prediction model, with a global community of over 36,000 users and code contributors. WRF, as a modeling environment, has been exploited for a wide range of applications, ranging from air quality to hydrology (Powers et al., 2017). WRF is also well connected to UHI research, the nocturnal warming caused by the built environment, through coupling with several urban models (Chen et al., 2011).

RCMs like WRF cover limited areas (northwest Europe in this study) and require an initial state (e.g., land use and soil moisture) followed by frequent updates about the wider, global circulation patterns at their boundaries. To some extent, these regularly updated boundary conditions will constrain IMV as the given RCM should be dynamically consistent with the driving GCM. As such, analysis of WRF variability has typically been limited to modifying small fragments of the model. For example, studies may change: urban models (Cui & de Foy, 2012; Holt & Pullen, 2007); land-use products (Feng Chen et al., 2014; De Meij & Vinuesa, 2014; J. Wang et al., 2016); urban parameters, such as building heights or albedo (Lin et al., 2008; Miao et al., 2009; Morini et al., 2018; Touchaei & Wang, 2015); physics schemes (García-Díez et al., 2013; Kleczek et al., 2014); synoptic conditions (X. X. Li et al., 2013; X. X. Li & Norford, 2016); or initial condition data sources (Gallus & Bresch, 2006; Wu et al., 2005). Several studies combine some of the above, that is, using multiphysics or parameter approaches (Crétat et al., 2012; Evans et al., 2012; Imran et al., 2018; Jankov et al., 2007; Jerez et al., 2013; Mooney et al., 2017; Sharma et al., 2017; Stegehuis et al., 2015; Wan & Zhong, 2014; Z. H. Wang et al., 2011; Zhong & Yang, 2015). Yet these studies really fail to distinguish model sensitivity from varying its parameters versus IMV.

Fortunately, some initial condition ensembles, commonly found in weather forecasting and climate modeling (Kay et al., 2015), have started to address IMV within RCMs (Bonekamp et al., 2018; Crétat et al., 2011; Fathalli et al., 2018; Giorgi & Bi, 2000; Laprise et al., 2012; Laux et al., 2017; Lucas-Picher et al., 2008). However, this use of ensembles has not fully caught up to the RCM community (Laprise et al., 2012). This may be due to a lack of IMV awareness, something we aim to address in this study, or the availability of computational requirements to run these ensembles. Computational requirements may be addressed through access to new technology platforms, such as cloud computing (Simm et al., 2018; Siuta et al., 2016). Conversely, increased computing resources may conflict between increasing resolution and resolving subgrid-scale processes and ensembles (Berner et al., 2017).

This study aims to improve upon existing IMV knowledge by significantly increasing the number of ensemble members, allowing us to address: (i) the degree of IMV for various atmospheric fields, (ii) the number of simulations required to capture IMV, and (iii) what conditions leads to higher IMV. This is achieved through modifying initial conditions (i.e., simply changing start times) while keeping model configuration (i.e., parameters and physics) and lateral boundary conditions constant. Furthermore, we test if IMV has a knock-on effect to the magnitude of the UHI, a common application of WRF modeling. While our attention is focused on the WRF model because of its large user base, our approach is applicable to any RCM. Furthermore, the resulting IMV quantifications may impact optional coupled model components, for example, WRF-Chem (Grell et al., 2005).

2. Methods

2.1. WRF Configuration

WRF Version 3.9.1.1 was configured using three nested domains, each containing 100×100 grid cells and increasing in resolution from 27 to 9 to 3 km (Figure 1a). The model was centered over London, a city historically well observed in terms of urban climate (e.g., Jones & Lister, 2009). All model physics and dynamics options were kept as default, except setting the model time step to 120 s and turning urban physics on. For this, the single-layer urban canopy model (Kusaka et al., 2001) was used to account for radiation processes within an urban canyon (e.g., trapping) and friction caused by increased drag from buildings, before passing heat and momentum fluxes back to WRF. To complement single-layer urban canopy model, urban parameters specific to London were used (Loridan et al., 2013). Two-way nesting that allows a domain to feed information back to its parent domain is kept on. While the use of a triple two-way nested domain will

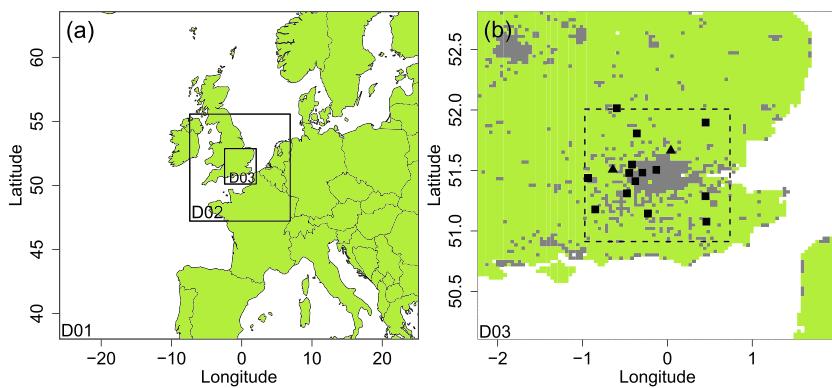


Figure 1. (a) WRF domain configuration: D01 (outer) 27-km, D02 (middle) 9-km, and D03 (inner) 3-km resolution. (b) D03 land use: urban (gray) and mixed rural categories (green). The dashed box around London represents the analysis zone in which the spatial means for both the observations and ensemble members were calculated. The black squares show the location of the observation records for temperature and precipitation, and black triangles for two stations that contained temperature only, further detailed in supporting information Table S1.

significantly increase the degrees of freedom in the model, this is a commonly used setup in WRF studies. WRF default land use (Moderate Resolution Imaging Spectroradiometer satellite classification) was kept for simplicity (Figure 1b). As the default land use only contains a single urban category, we recommend using location-specific land use for more detailed investigations. For initial and lateral boundary conditions, NCEP (National Centers for Environmental Prediction, 2000) Final Operational Model Global Tropospheric Analyses data at 1° horizontal and 6-hourly temporal resolution were used (<https://rda.ucar.edu/datasets/ds083.2/>). Sea surface temperatures (SSTs) were updated during the simulation with data also provided by NCEP. The SST update is not “on” by default in WRF, and instead, the model uses a fixed value for the whole simulation. While SSTs may not be too influential for a short timescale or continental setups, due to the coastal location of our domain, it has a significant impact on our simulations (tests not shown). Switching on the SST update also updated other time-dependent variables during the simulation, including vegetation fraction, leaf area index, and albedo. Although NCEP is commonly used in WRF studies (see Loridan et al., 2013), a variety of alternatives are available (see http://www2.mmm.ucar.edu/wrf/users/download/free_data.html). To account for model spin-up, we disregarded the first 24 hr of each simulation.

2.2. Observations for Evaluation

For comparisons between our simulations and observations, we confined the analysis to a subset of the inner domain (see the dashed box in Figure 1b). Within this box, hourly observations were taken from 16 Met Office surface stations (Met Office, 2012). The details of each station are listed in Table S1 in the supporting information. These locations were chosen based on consistent data records, spatial spread, and small variations in altitude. Furthermore, as the focus of this study is not fine tuning WRF to a specific place, we took the mean air temperature and precipitation of all stations to represent a spatial mean within the area defined by the dashed box in Figure 1b. This simplified spatial mean was also calculated using the same dashed box for each ensemble member. This box represents 40 grid cells or 120 km in each direction. A time series for observed variables are provided in Figure S1.

2.3. Ensemble Framework

We created a 244-member ensemble with run lengths ranging from 2 to 4 months. WRF ensemble members were started at 6-hr intervals covering a 2-month period from 1 May 2018 to 1 June 2018 (i.e., 244 members from 61 days × 4 starts per day), illustrated in Figure 2. Although this method differs from more traditional ensemble perturbations, like Giorgi and Bi (2000), WRF simulation lengths typically range from days to months, and start times are usually decided on an ad hoc basis. For each ensemble member, all WRF preprocessing steps were repeated, to ensure each member had a unique set of initial conditions. Bash scripts were used to automatically run the preprocessing, update the WRF namelist for each member and submit these to a high-performance computer (HPC). During the simulations, each member contained identical boundary

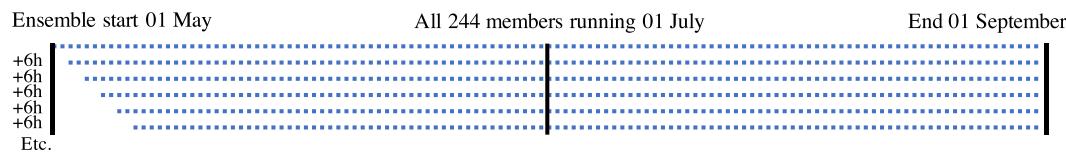


Figure 2. Ensemble schematic. From 1 May, each member was started at 6-hr intervals until 1 July. From this point, all 244 members were left to run for a further 2-month period, ending 1 September 2018.

conditions ending 1 September 2018, hence creating a 2-month period when all members were running. The summer 2018 period was chosen because of its notable (near heat wave) warmth that became one of the joint-hottest summers on record in the United Kingdom (<https://www.metoffice.gov.uk/about-us/press-office/news/weather-and-climate/2018/end-of-summer-stats>). This case is likely to contain higher IMV than an equivalent winter case due to the increased frequency of stable weather patterns (Giorgi & Bi, 2000).

The ensemble configuration required over 140,000 CPU hours and had the potential to generate almost 30-TB output data. We therefore only saved a small fraction of the output variables: 2-m air temperature, 10-m wind (u and v components), surface pressure, precipitation, boundary layer height, sensible and latent heat fluxes, and 2-m specific humidity. All ensemble members were run on a HPC at Lancaster University, using 32 CPUs for each simulation, and running 10 simulations concurrently. The longest (4 months) member took approximately 24 hr to run and shortest (2 months) 12 hr. Excluding the preprocessing steps and queue times, the 244-member ensemble took 37 days to complete.

2.4. Bit-for-Bit Reproducibility

Different hardware options were available on Lancaster University's HPC, which could either be specified at job submission or left unspecified in order to reduce queuing times. In unspecified mode, the scheduler would automatically assign our WRF jobs to the first available hardware type. Due to the number of simulations, we ideally needed to run these in unspecified mode to reduce queue times. However, different hardware types and compilers have been known to produce different results (R. Li et al., 2016), largely due to floating point rounding errors. To test this, we conducted a 2-month test simulation specifying (i) two different hardware types with each run using 32 CPUs and (ii) a different number of CPUs, 32 versus 40, on each of these two hardware types. All four test simulations used the same initial conditions and boundary conditions, as well as Intel compilers with default flags as set by the WRF model installation scripts. Different compilers (e.g., ifort vs. gfortan) may also cause variability (R. Li et al., 2016); however, we do not test this here, using only the Intel compiler for our ensembles. Our tests showed no difference in simulations using different hardware types using the same number of CPUs. We did find significant variability in results between 32 and 40 CPUs, and the same amount of variability was found between the different number of CPU cases on both hardware types. This likely relates to how WRF splits domains into patches for each CPU. For example, 32 CPUs may contain 8×4 patches and 40 CPUs 10×4 patches. Therefore, our ensembles were permitted to run on unspecified hardware types but only using 32 CPUs, in the knowledge that the variability between the simulations would just be due to IMV.

3. Results and Discussion

The ensemble mean temperature, wind speed, pressure, and precipitation time series are provided in Figure S1. The ensemble mean temperature bias compared to the observations is shown in Figure 3a. The mean bias during the 4-month period is 0.04°C (s.d. 1.87°C). Figure 3b shows the root-mean-square error (RMSE) between the simulated and observed 2-m temperature for each ensemble member over a 2-month period from 1 July to 1 September (i.e., when all 244 members were running), where the mean RMSE is 1.16°C . Our results show that, in terms of model performance against observations, the difference between RMSE only slightly improves as members are started nearer the 2-month analysis period. Using the R package segmented (<https://cran.r-project.org/web/packages/segmented/segmented.pdf>), we detected a breakpoint (see Figure 3b), occurring at member 138 (near the beginning of June); however, mean RMSE before and after this change represents less than a 3% reduction.

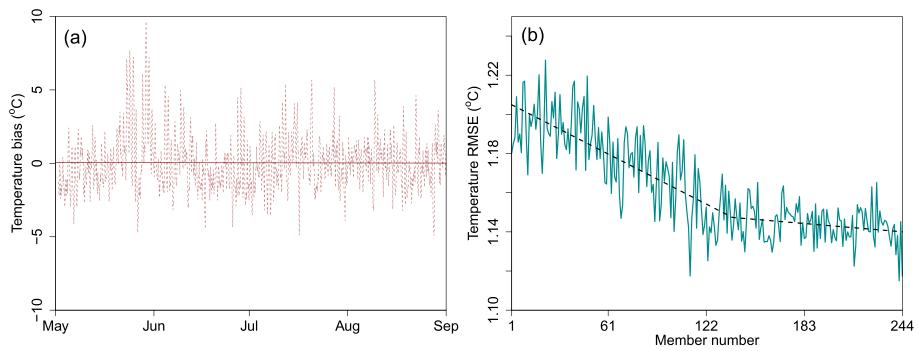


Figure 3. (a) Ensemble mean bias against the mean temperature observations. (b) Temperature RMSE evaluation for each individual member (x axis). RMSE is calculated between the modeled spatial mean in the analysis box shown in Figure 1b and the mean of the observed 2-m air temperatures.

3.1. IMV and the Impact on the Derived UHI

The spread or bias relative to the ensemble mean 2-m temperature (i.e., the IMV) is shown in Figure 4. Note the spread is calculated from the spatial mean of the analysis box centered over London (see extent in Figure 1b). The magnitude of the IMV increases as more members start, and by the start of July all members are included. From Figure 4 it is clear that notable differences exist between ensemble members. To explore this, at each hour in the 4-month analysis period, the range between the ensemble members are calculated using the spatial means from the analysis box. We consider the range between these values to represent WRF IMV. The IMV time series for temperature, surface pressure, wind speed, and precipitation are presented in Figures 5a–5d. Across the 4-month period, the median temperature range is $1.20\text{ }^{\circ}\text{C}$ (interquartile range, IQR, $0.67, 2.21\text{ }^{\circ}\text{C}$) and maximum range over $8\text{ }^{\circ}\text{C}$. This temperature range is significantly higher than previous IMV studies (Laux et al., 2017), perhaps explained by the significantly larger ensemble size used in our study. It is also comparable to the spread created using large multiphysics ensembles, for example, a 3-month spatial mean up to $3\text{ }^{\circ}\text{C}$ (Jerez et al., 2013) or maximum range up to $10\text{ }^{\circ}\text{C}$ (Stegehuis et al., 2015). Moreover, (i) a spatial mean is used to calculate variability; therefore, at the individual grid cell level variability will be higher; (ii) at the start of the 4-month period, less members are running; therefore, some of the variability will be missed in the calculation (i.e., an underestimation). Variability for the other variables we outputted from WRF (not plotted) are listed in Table 1. Note, these variables only represent a small subset of all possible WRF output as we were limited by storage.

To further emphasize IMV, cumulative precipitation covering a 2-month period, starting from 1 July 2018 when all 244 members are running, is plotted in Figure 6. When all simulations are complete (1 September 2018), the total precipitation at the end of this 2-month period ranges between 84.17 and 106.30 mm . The actual observed precipitation, taken as the spatial mean of 14 weather stations (see the black line in Figure 6), had a cumulative total of 84.00 mm , placing this at bottom of the ensemble range. For comparison, the ensemble mean cumulative precipitation is 94.41 mm (s.d. 4.86 mm). While this is not far from

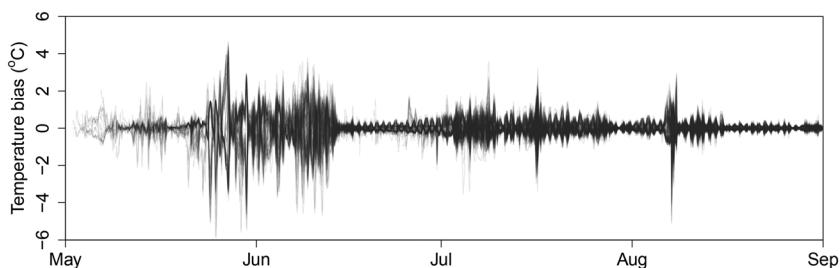


Figure 4. Domain 3 spatial mean 2-m air temperature (calculated within the analysis box) bias from the ensemble mean for each member between 1 May and 1 September 2018.

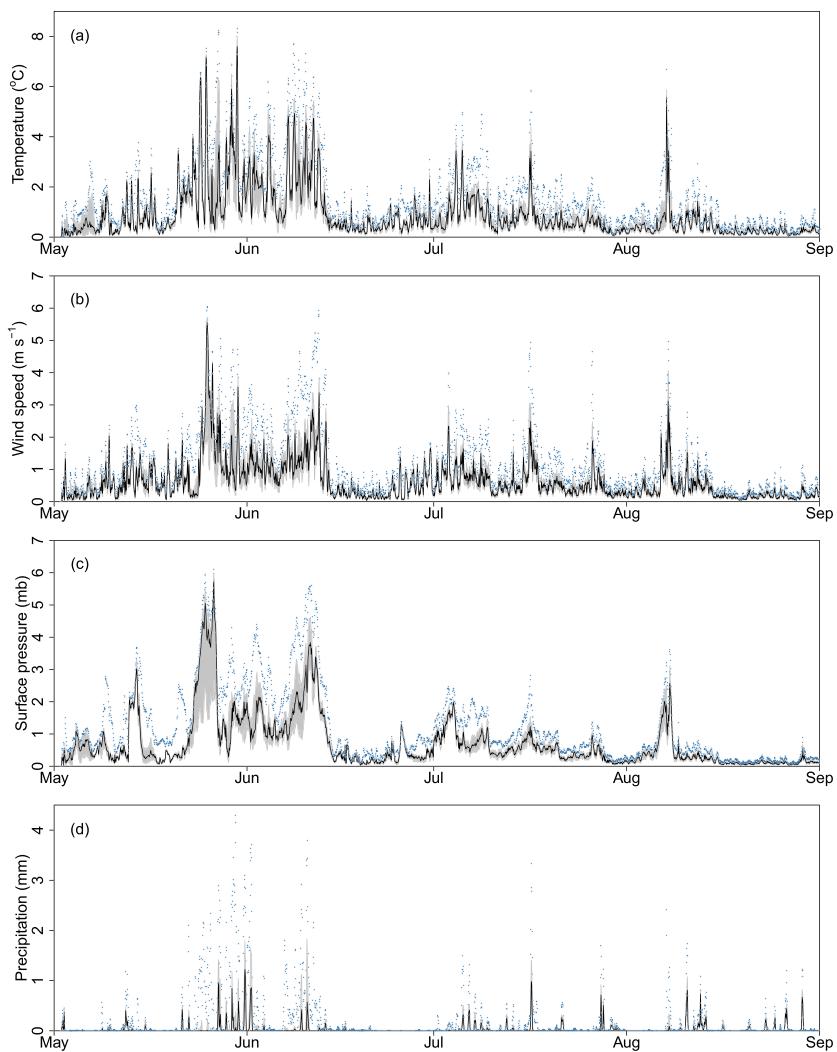


Figure 5. Hourly median (black line), IQR (gray shading), and maximum (blue crosses) range for the spatial mean of (a) 2-m air temperature, (b) 10-m wind speed, (c) surface pressure, and (d) precipitation.

Table 1

Measures of IMV for Selected Surface WRF Output Variables, Calculated Across the Hourly Spatial Mean Ensemble Range for the 4-Month Period

	Mean	Median (IQR)	Maximum
2-m air temperature (°C)	1.69	1.20 (0.67, 2.21)	8.32
Boundary layer height (m)	221.06	146.06 (66.89, 293.67)	1715.81
Latent heat flux (W/m²)	28.77	13.11 (3.50, 41.29)	231.82
Surface pressure (mb)	1.24	0.77 (0.41, 1.77)	6.12
Sensible heat flux (W/m²)	29.61	15.37 (5.68, 42.72)	197.27
2-m specific humidity (g/kg)	1.17	0.84 (0.42, 1.70)	5.59
Derived UHI (°C)	1.06	0.84 (0.46, 1.49)	6.79
Night (°C)	1.14	1.00 (0.55, 1.64)	3.40
Day (°C)	1.02	0.74 (0.44, 1.36)	6.79
10-m wind speed (m/s)	1.23	0.89 (0.47, 1.65)	6.03

Note. UHI is derived as the mean of the urban grid cells less the rural grid cells within the analysis box.

the observations, for precipitation the ensemble spread in Figure 6 shows a significant overestimation in July, followed by the model missing rainfall events in August, that effectively allows (cumulatively) the observations to catch up to the model.

Spatially, we show the 2-m temperature field for a single time slice (20 August 2018 at midnight) for several ensemble members in Figure 7a, effectively depicting London's UHI. Even though the ensemble members are started only a few hours apart, large spatial differences between the temperature fields (i.e., the UHI pattern) exist. Aside from city size and form (fixed in these simulations), UHI intensity is predominantly controlled by wind speed and cloud cover; proxies for stability (Pasquill & Smith, 1984), as captured in many observational studies (e.g., Tomlinson et al., 2012). Therefore, based on the considerable variability already found for wind speed (see Figures 5b and Table 1), it is not surprising that temperature

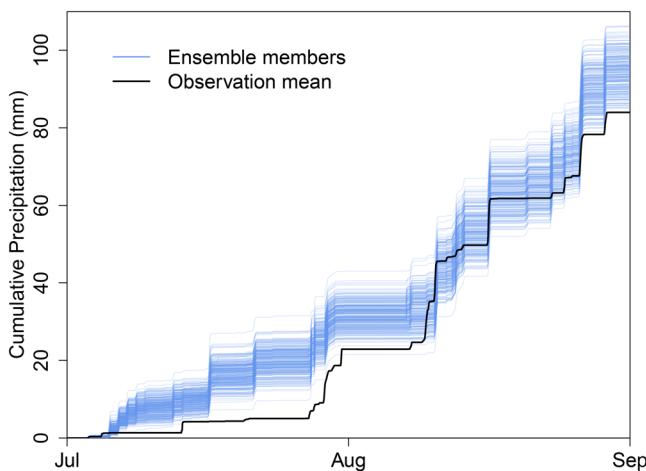


Figure 6. Spatial mean cumulative precipitation over a 2-month period (July and August) when all 244 members were running. The black line represents the mean observed precipitation across 14 weather stations.

fields depicting the UHI also exhibited large IMV between ensemble members.

We then classify rural areas as all grid cells not containing urban or water, shown as green in Figure 1b. Separate means for all urban and all rural grid cells within the analysis box are calculated for each member. Prior to the UHI calculation, temperatures were adjusted to sea level to account for altitude changes (0–217 m) within the domain using the mean environmental lapse rate ($0.0065\text{ }^{\circ}\text{C/m}$). Across the 4-month modeling period, there is considerable IMV for UHI intensity with a median of $0.84\text{ }^{\circ}\text{C}$ (IQR 0.46 , $1.49\text{ }^{\circ}\text{C}$). Split diurnally (based on daily sunrise and sunset times), larger IMV is found at night, $1.00\text{ }^{\circ}\text{C}$ median versus $0.74\text{ }^{\circ}\text{C}$ during the day. These UHI ranges created through modifying the start time are greater in magnitude to that found by improving urban parameters (e.g., $0.6\text{ }^{\circ}\text{C}$ improvement in modeled nighttime UHI was found by Toucheai & Wang, 2015). An example UHI spread for this single time slice (20 August 2018 at midnight) that contains all ensemble members is shown in Figure 7b. Considering a large number UHI studies use single-run methods, these results show considerable UHI variability that should really be accounted for in future model based UHI studies.

3.2. IMV Relationship With Boundary Conditions

It is evident in Figure 5 that IMV levels waxed and waned during the 4-month modeling period. For example, sustained high IMV is found mid-May to mid-June, with a median temperature range of $3.04\text{ }^{\circ}\text{C}$ (IQR 1.88 , $4.46\text{ }^{\circ}\text{C}$). This is also the period where the maximum temperature range over $8\text{ }^{\circ}\text{C}$ occurs and where differences between the ensemble mean and observations are highest (see Figure S1). Note that by mid-May only 60 members are running; therefore, IMV during this period is likely underestimated. Between the variables plotted in Figures 5a–5d, we note a similarity in the periods where either high or low IMV occur. We show the relationships between some of the outputted WRF variables in Figure 8, with r values up to 0.85 (Spearman's rank).

Fast-moving boundary conditions, that is, a high through flow of weather driven by large-scale circulations, will cause conditions within the WRF domain to be more strongly forced by these external inputs (Giorgi & Bi, 2000; Laprise et al., 2012). However, if conditions at the domain boundaries are not active, like persistent high pressure, WRF will have more freedom to generate its own weather: Hence, there will be a higher IMV as the model is more sensitive to initial conditions. We demonstrate this effect in our results by calculating the mean of the maximum wind speed input from each of the first 10 model levels (surface to 700 mb or $\sim 10\text{ km}$) within the Outer Domain 1 (Figure 1a), using the coarse (1°) 6-hourly boundary condition data from NCEP. This cross-domain wind speed is a proxy for the “refreshing” capability of the boundary conditions imposed onto the internally generated conditions inside of WRF.

Figure 9a shows the comparison between the temperature range for our ensemble and the input wind speed. We immediately see that lower input wind speeds correspond to higher temperature IMV across the ensemble. We further show this relationship as a scatterplot in Figure 9b. Here we find a significant relationship between input wind speed and modeled 2-m temperature ($r = -0.66$, Spearman's rank). Therefore, we find slow-moving boundary conditions are linked to higher levels of IMV in WRF; these results are in agreement with earlier studies (e.g., Giorgi & Bi, 2000). A cross-correlation function (Figure 9c) was used to test preceding input wind speed effects on IMV, finding peak correlations at zero lags. This conveys that IMV at a given time does not relate to earlier input wind conditions. We note the secondary lags likely relate to synoptic regimes, that is, the main high wind speed (mid-May to mid-June) or low wind speed (mid-June to early July) periods. Although we do not test effects of seasonality in this study, considering the relationship with input wind speed, we would expect findings similar to Giorgi and Bi (2000) where higher IMV was found in summer due to less influence from prevailing westerly winds. Overall, these findings will have direct implications for UHI modeling, where periods of calm conditions are usually chosen due to their relationship with maximum UHI intensity.

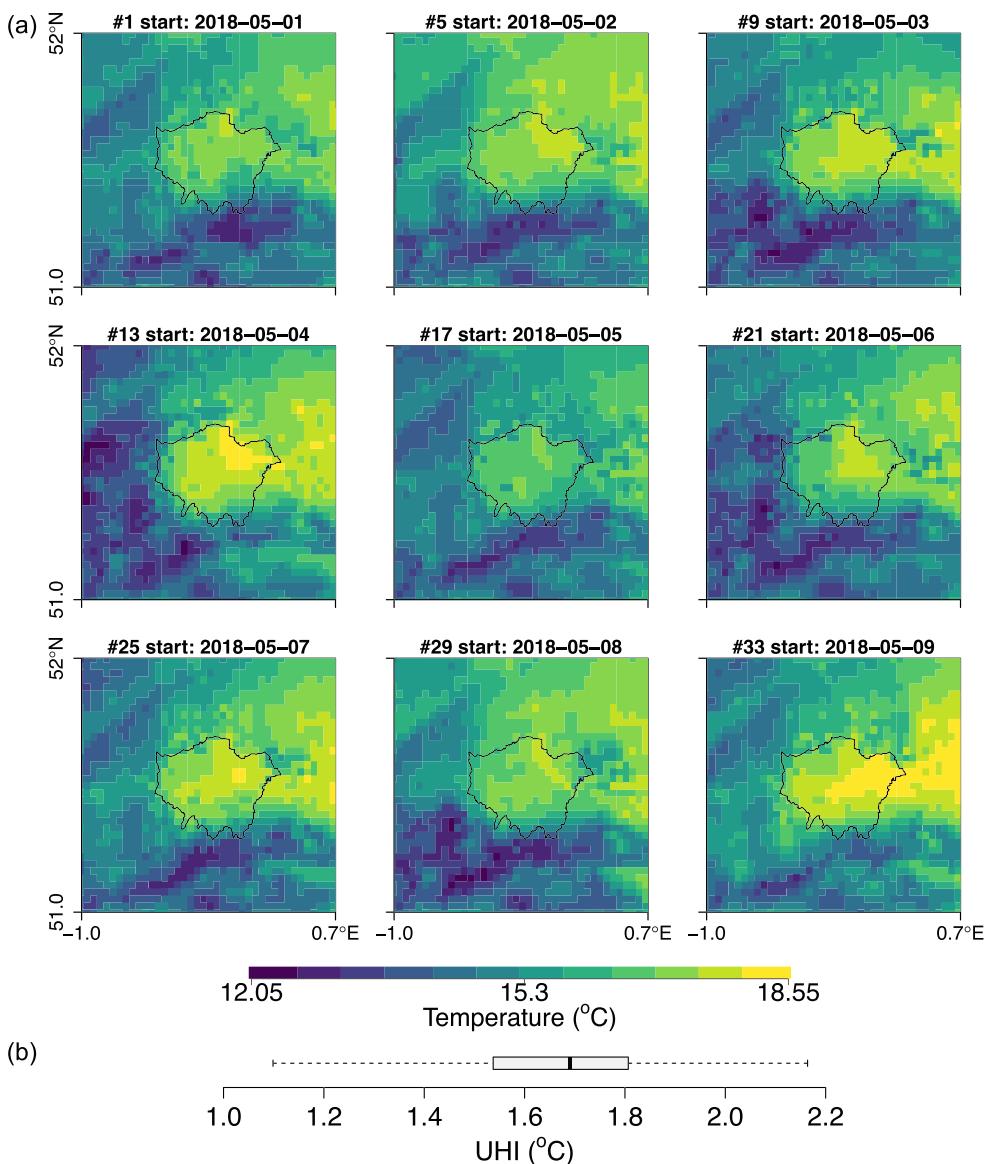


Figure 7. (a) Domain 3 analysis box 2-m temperature for ensemble members (#) 1, 5, 9, 13, 17, 21, 25, 29, and 33 at a single time slice, midnight 20 August 2018. The black outline represents the Greater London region. (b) Box plot showing the UHI spread containing all 244 members on midnight 20 August 2018, the whiskers represent the minimum and maximum.

3.3. Number of Ensemble Members Required to Capture IMV

Running 244 members to create a WRF ensemble is computationally demanding, and here we were limited to a small subset of the total WRF output due to storage limits. We therefore explore the minimum number of ensemble members required to capture the majority of IMV contained in all 244 members using three methods: (i) connected, (ii) sequenced, and (iii) sampled.

For the “connected” method, Figure 10a shows the median temperature range across the 2-month period of July and August as more ensemble members are included, moving back in time from the 2-month analysis window start time (1 July 2018); the ensemble starts with a single member and finishes with 244. On the left-hand side of Figure 10, by only including a single member the temperature range is unsurprisingly zero, whereas the median range is highest when all 244 members are included (0.95 $^{\circ}\text{C}$). A marked change in gradient for the mean temperature range is apparent in Figure 10a, and the R package “segmented” (<https://>

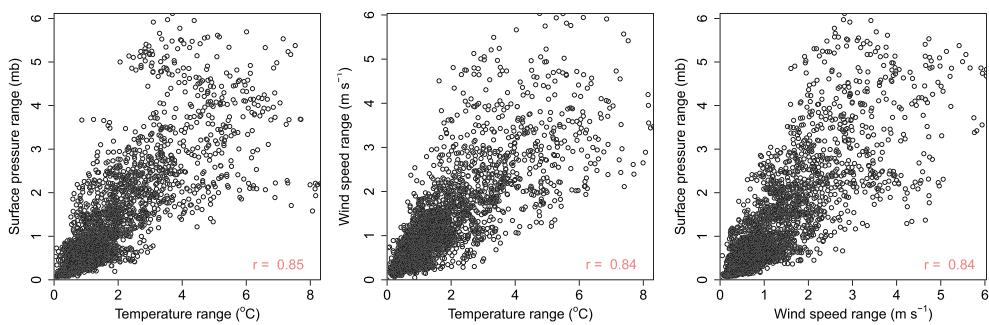


Figure 8. Relationships between ensemble spatial mean ranges of (left) surface pressure and temperature, (middle) wind speed and temperature, and (right) surface pressure and wind speed.

cran.r-project.org/web/packages/segmented/segmented.pdf) was applied to automatically detect this breakpoint. Occurring after 16 members, this accounts for approximately half (47%) of the IMV generated with all 244 simulations. From here in we refer to this “relative” IMV as rIMV (i.e., variability relative to the ensemble containing all 244 simulations), because the gradient does not plateau by 244 members suggesting that the “true” IMV is greater than captured by our simulations. Following this breakpoint, to account for 75% rIMV, we would need to include 107 members.

For the second “sequenced” method we explore using the members to greater effect, why constrict to 16 members in a row? To test this, we compared six ensemble member sequences, at 1, 2, 4, 8, 16, and 32 member intervals. Each of these new ensembles contain 244, 122, 61, 30, 15, and 8 members, respectively. Results from these sequences are plotted in Figure 10b. To further explain this sequenced method, take the purple

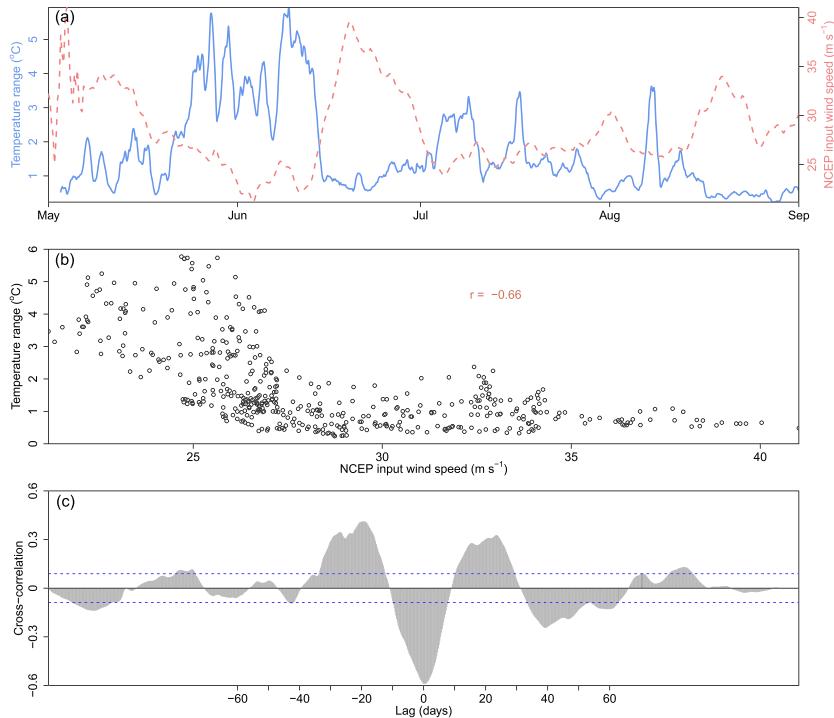


Figure 9. (a) Ensemble temperature range (IMV) shown as the blue line and the dashed red line shows the Domain 1 (outer) mean of the maximum wind speed input from each of the first 10 input model levels (surface to 700 mb or ~3 km), taken from NCEP prior to running WRF. Each line is a 24-hr running mean. (b) The relationship between the variables in (a). Note for both plots, the temperature range data were reduced to fit the 6-hourly NCEP boundary condition times. (c) Cross-correlation function between the variables in (a). The dashed blue lines are the 95% confidence bounds.

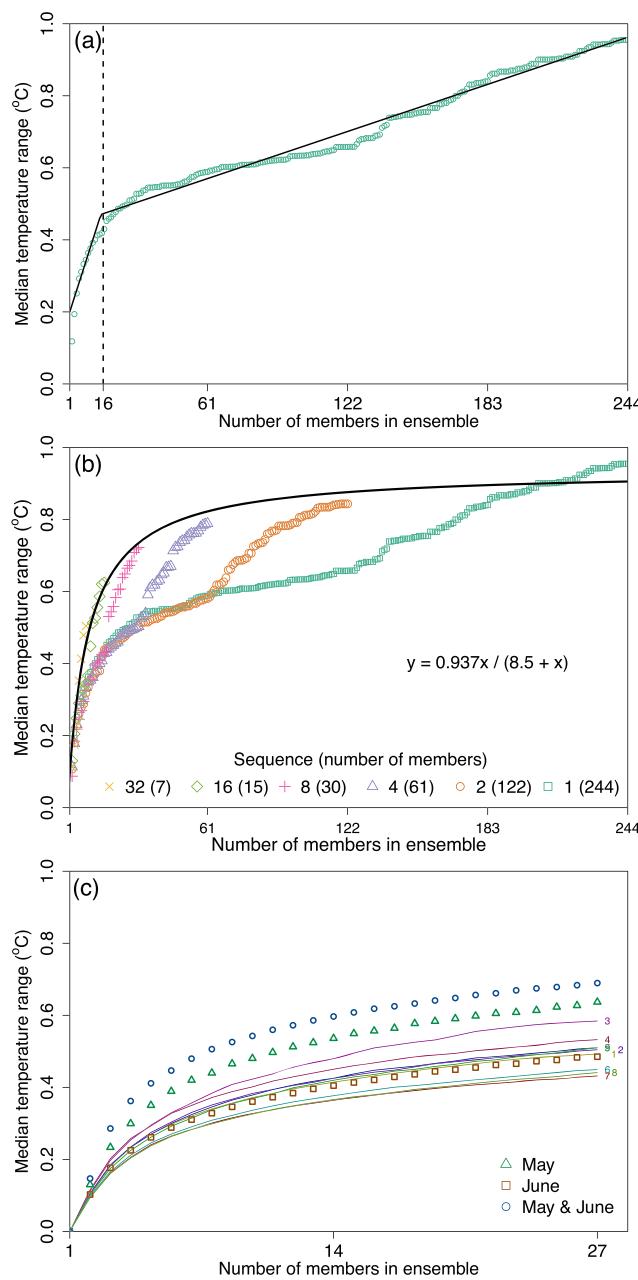


Figure 10. (a) Minimum number of ensemble members to represent the majority of WRF IMV for temperature using the “connected” approach. Breakpoint analysis was used to detect the change in gradient at 16 members, shown as the black line. (b) Minimum number of simulations required to represent the majority of WRF IMV for temperature using the “sequenced” approach. Six ensemble member sequences were created from members 1 to 244, at 1, 2, 4, 8, 16, and 32 member intervals. Each of the new ensembles contained 244, 122, 61, 30, 15, and 8 members, respectively. The black line represents the nonlinear regression for the maximum range at each of the six sequences. (c) “Sampled” method. The solid lines represent IMV from week long sampling periods (1–9) within the 2-month initialization period. The points are from member samples within monthlong periods May (triangles) and June (squares) and both months (circles). Note that for (a)–(c), these relationships are unique to our model setup, in our case, three domains centered over London.

triangles in Figure 10b by way of an example. In total, there are 61 purple triangles. The last purple triangle represents the range in temperature or IMV from 61 members taken at 4-member intervals (or daily) between ensemble members 1 and 244. At the thirtieth purple triangle in Figure 10b, this represents the temperature range taken across the first 30 members in a 4-member sequence between 1 and 244; the new ensemble contains original members 1, 5, 9, 13, ..., 117. Using one less ensemble member (15) than the first connected-member approach (16), we increase the amount of rIMV from 47% to 65%. More dramatically, we only require 30 members to now explain 75% rIMV (whereas 107 members were needed previously). This number of ensemble members is still considerably larger than used in previous IMV studies (Bonekamp et al., 2018; Crétat et al., 2011; Fathalli et al., 2018; Laprise et al., 2012; Laux et al., 2017; Lucas-Picher et al., 2008). To explain the relationship between the maximum median temperature range captured by each of the six sequences, nonlinear regression is used, shown as the black line in Figure 10b. The resulting equation $y = 0.937 * x / (8.5 + x)$ may be used (for our WRF configuration) to estimate the likely temperature rIMV from a given number of sequenced ensemble members.

The third “sampled” method uses different arrangements of the ensemble members; that is, for 5 members out of 244 cases, members could be 3, 32, 87, 144, 201, or any other random combination, randomized using the R sample function. We also test the member sampling period, individual weeks (9 total), months and an all-member case (i.e., 2 months). We limit the number of random configurations for all tests to 702 (maximum combinations for a 2-member case over a week) and only create ensembles up to 27 members (i.e., a week). The results are shown in Figure 10c for the median temperature range for all cases. The solid lines depict members taken within individual weeks of the 2-month starting window and show a $0.15\text{ }^{\circ}\text{C}$ IMV variation between weeks when 27 members are included. This variation does not appear to relate to proximity to the analysis window. For the month cases, higher IMV occurs for our configuration when sampling ensemble members in May than June. Overall however, we find the highest IMV when sampling across all 244 members, with IMV at $0.69\text{ }^{\circ}\text{C}$, when using 27 members. For the all-member case, we also test using 30 members to make direct comparisons with the “sequenced” approach, finding 73% rIMV.

We make two deductions from these approaches: (i) the size and period chosen for sampling affect total IMV and (ii) “sequenced” and “sampled” approaches produce similar rIMV. There are caveats to recommending a minimum number of ensemble members to capture IMV. Our results are specific to this study and may not be transferable due to both the unique WRF setups in terms of location, domain sizes, resolutions, physics packages, and parameters and the fact that the gradients on the slopes in Figure 10a do not plateau after 244 members are running. Therefore, we surmise that 244 members do not capture all WRF IMV, hence *relative* IMV. This poses a challenge: Hypothetically, even if we ran enough simulations to capture 100% of IMV, we are still limited to tiny subspace of model uncertainty. Different physics packages, domain sizes, resolutions, and initial condition data will all trigger a slight variation on the IMV we found. Furthermore, for the sequencing method, the further back in time the sequence starts, the more computational power and time is required.

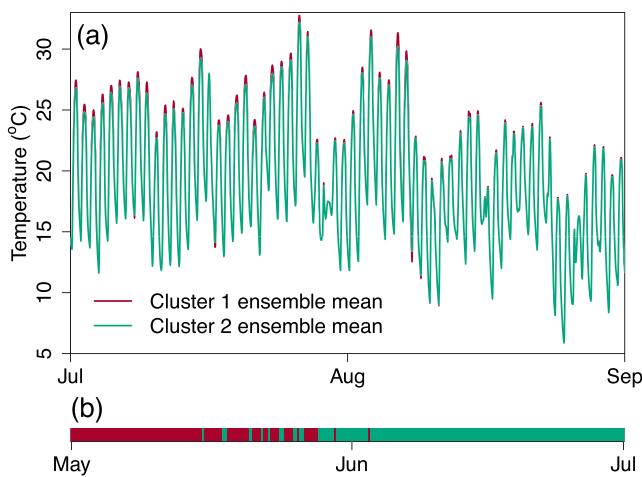


Figure 11. (a) Ensemble mean temperature for Clusters 1 ($n = 99$) and 2 ($n = 145$). See Figure S2 for a dendrogram of the clusters. (b) Member start times for each cluster.

We find the majority of Cluster 1 members to start in May and Cluster 2 in June. This difference is also reflected earlier in Figure 3b where we find reductions, albeit very small, in RMSE between observations and members started in June rather than May. Possible reasons for these effects may include fixed monthly model values for leaf area index, albedo and green fraction, or long soil moisture memory effects (Sörensson & Berbery, 2015).

4. Conclusions

A large ensemble of 244 members was used to quantify IMV within the WRF RCM for a case study covering southeast United Kingdom during June and July 2018. Using a largely default WRF model configuration, we changed model start times (initial conditions) to generate IMV, an often overlooked feature in RCMs. The simulations were completed on a local HPC at Lancaster University, taking 37 days to complete using 10 simultaneous 32 CPU jobs.

Defining the IMV as the range in a given variable between all ensemble members at each time step, we found a median IMV for the 2-m temperature of $1.20\text{ }^{\circ}\text{C}$. IMV was also found to fluctuate considerably with a maximum exceeding $8\text{ }^{\circ}\text{C}$. This IMV was found to have knock-on effects on the derived UHI for London, with median nocturnal UHI ensemble range of $1\text{ }^{\circ}\text{C}$. This is a common application of WRF, yet most similar studies use only a single model run and may only change a single parameter, such as albedo, to replicate cooling from green roofs. Such studies do not substantially address the model sensitivity in the UHI calculation, and on the basis of our results, we would therefore question their validity. Further analysis demonstrated the relationship of IMV to the boundary conditions. Coarse resolution input wind speed from the NCEP boundary condition data was statistically linked to temperature IMV within the inner domain, with higher boundary condition wind speeds reducing the IMV ($r = -0.66$). Effectively, fast-moving boundary conditions are efficient at refreshing conditions inside WRF, whereas under slow-moving boundary conditions, nonlinear (chaotic) processes within WRF are more likely to dominate.

Since running 244 ensemble members is computationally demanding, we proposed several methods to calculate a minimum number of simulations required to capture the majority of IMV (while acknowledging that 244 members will not capture all of the IMV). An analysis of IMV showed a natural breakpoint in terms of ensemble size, with large increases in IMV from increasing size up to that point and minimal increases after. For our simulations, 75% of the IMV was accounted for using 30 members in a sequence across a 2-month period. Additionally, no benefit to using randomly selected ensemble members was found, except noting the period in which ensemble members were selected could modify IMV.

Overall, our results are applicable to all WRF and indeed any RCM studies and will have knock-on implications for coupled WRF components like chemistry or hydrology. However, exact levels of IMV will be unique

Although we do not test in this study, IMV may also be introduced by perturbing initial conditions, reducing the time before the analysis period.

3.4. Ensemble Member Similarities

In the previous section, we found higher IMV when members were started in May rather than in June. To investigate this further, we used Ward's (Ward, 1963) hierarchical clustering to test if there were any similarity between the ensemble members. Ward's method effectively minimizes variance (using Euclidian distance) at each step, that is, merging the two closest members into a new cluster, before repeating this process until all members are moved into a single cluster. The analysis was conducted on the second 2 months only (July and August) when all the members were running. The results are shown as a dendrogram in Figure S2 where two main clusters are detected. The ensemble mean temperature for each of these clusters is shown in Figure 11a where differences between peak daytime temperatures are found. The differences are largest at the beginning of the 2-month period (up to $1\text{ }^{\circ}\text{C}$) and decrease toward the end. The start times of the ensemble members are provided in Figure 11b.

to individual study configurations. In addition to IMV generated through initial conditions, WRF uncertainty may also be created by changing, for example, study location, domain sizes, nest resolutions and ratios, one-way or two-way nesting, time of year, initial and boundary condition products, number of vertical levels, domain boundary relaxation zones, and use of techniques such as observation nudging. This list is by no means exhaustive as IMV could also be created by the hardware and software environment in which the model is hosted. Testing all these sources of variability will be extremely computationally demanding. However, we are excited by new cloud-computing resources that will allow for fast-finishing ensembles. To conclude, that WRF generates a large degree of IMV is not a bad problem per se so long that it is recognized and accounted for in future studies.

Acknowledgments

This research was made possible through the UK EPSRC Grant EP/N027736/1: Models in the Cloud and use of the High End Computing HPC at Lancaster University (<https://www.lancaster.ac.uk/iss/hec/>). We would like to extend our gratitude to the WRF community for open access to their model, in addition to the National Centers for Environmental Prediction and Centre for Environmental Data Analysis for provision of data. Additionally, we would like to thank the wider “Ensemble” research team (<https://www.ensembleprojects.org/>) for their guidance. WRF configuration files (namelists) and model outputs for each member (inner and outer domains for the variables analyzed above) are available in R data format from the website (<https://doi.org/10.17635/lancaster/researchdata/348>).

References

- Berner, J., Achatz, U., Batté, L., Bengtsson, L., de la Cámara, A., Christensen, H. M., et al. (2017). Stochastic parameterization: Toward a new view of weather and climate models. *Bulletin of the American Meteorological Society*, 98(3), 565–588. <https://doi.org/10.1175/BAMS-D-15-00268.1>
- Bonekamp, P. N. J., Collier, E., & Immerzeel, W. W. (2018). The impact of spatial resolution, land use, and spinup time on resolving spatial precipitation patterns in the Himalayas. *Journal of Hydrometeorology*, 19(10), 1565–1581. <https://doi.org/10.1175/JHM-D-17-0212.1>
- Chen, F., Kusaka, H., Bornstein, R., Ching, J., Grimmond, C. S. B., Grossman-Clarke, S., et al. (2011). The integrated WRF/urban modelling system: Development, evaluation, and applications to urban environmental problems. *International Journal of Climatology*, 31(2), 273–288. <https://doi.org/10.1002/joc.2158>
- Chen, F., Yang, X., & Zhu, W. (2014). WRF simulations of urban heat island under hot-weather synoptic conditions: The case study of Hangzhou City, China. *Atmospheric Research*, 138, 364–377. <https://doi.org/10.1016/j.atmosres.2013.12.005>
- Christensen, O. B., Gaertner, M. A., Prego, J. A., & Polcher, J. (2001). Internal variability of regional climate models. *Climate Dynamics*, 17(11), 875–887. <https://doi.org/10.1007/s003820100154>
- Crétat, J., Macron, C., Pohl, B., & Richard, Y. (2011). Quantifying internal variability in a regional climate model: A case study for Southern Africa. *Climate Dynamics*, 37(7–8), 1335–1356. <https://doi.org/10.1007/s00382-011-1021-5>
- Crétat, J., Pohl, B., Richard, Y., & Drobinski, P. (2012). Uncertainties in simulating regional climate of Southern Africa: Sensitivity to physical parameterizations using WRF. *Climate Dynamics*, 38(3–4), 613–634. <https://doi.org/10.1007/s00382-011-1055-8>
- Cui, Y. Y., & de Foy, B. (2012). Seasonal variations of the urban heat island at the surface and the near-surface and reductions due to urban vegetation in Mexico City. *Journal of Applied Meteorology and Climatology*, 51(5), 855–868. <https://doi.org/10.1175/JAMC-D-11-0104.1>
- De Meij, A., & Vinuesa, J. F. (2014). Impact of SRTM and Corine Land Cover data on meteorological parameters using WRF. *Atmospheric Research*, 143, 351–370. <https://doi.org/10.1016/j.atmosres.2014.03.004>
- El Afandi, G., Morsy, M., & El Hussieny, F. (2013). Heavy rainfall simulation over Sinai Peninsula using the Weather Research and Forecasting model. *International Journal of Atmospheric Sciences*, 2013, 1–11. <https://doi.org/10.1155/2013/241050>
- Evans, J. P., Ekström, M., & Ji, F. (2012). Evaluating the performance of a WRF physics ensemble over South-East Australia. *Climate Dynamics*, 39(6), 1241–1258. <https://doi.org/10.1007/s00382-011-1244-5>
- Fathalli, B., Pohl, B., Castel, T., & Safi, M. J. (2018). Errors and uncertainties in regional climate simulations of rainfall variability over Tunisia: A multi-model and multi-member approach. *Climate Dynamics*, 52(1–2), 335–361. <https://doi.org/10.1007/s00382-018-4150-2>
- Gallus, W. a., & Bresch, J. F. (2006). Comparison of impacts of WRF dynamic core, physics package, and initial conditions on warm season rainfall forecasts. *Monthly Weather Review*, 134(9), 2632–2641. <https://doi.org/10.1175/MWR3198.1>
- García-Diez, M., Fernández, J., Fita, L., & Yagüe, C. (2013). Seasonal dependence of WRF model biases and sensitivity to PBL schemes over Europe. *Quarterly Journal of the Royal Meteorological Society*, 139(671), 501–514. <https://doi.org/10.1002/qj.1976>
- Giorgi, F. (2019). Thirty years of regional climate modeling: Where are we and where are we going next? *Journal of Geophysical Research: Atmospheres*, 124, 5696–5723. <https://doi.org/10.1029/2018JD030094>
- Giorgi, F., & Bi, X. (2000). A study of internal variability of a regional climate model. *Journal of Geophysical Research*, 105(D24), 29,503–29,521. <https://doi.org/10.1029/2000JD900269>
- Grell, G. A., Peckham, S. E., Schmitz, R., McKeen, S. A., Frost, G., Skamarock, W. C., & Eder, B. (2005). Fully coupled “online” chemistry within the WRF model. *Atmospheric Environment*, 39(37), 6957–6975. <https://doi.org/10.1016/j.atmosenv.2005.04.027>
- Holt, T., & Pullen, J. (2007). Urban canopy modeling of the New York City metropolitan area: A comparison and validation of single- and multilayer parameterizations. *Monthly Weather Review*, 135(5), 1906–1930. <https://doi.org/10.1175/MWR3372.1>
- Imran, H. M., Kala, J., Ng, A. W. M., & Muthukumaran, S. (2018). An evaluation of the performance of a WRF multi-physics ensemble for heatwave events over the city of Melbourne in southeast Australia. *Climate Dynamics*, 50(7–8), 2553–2586. <https://doi.org/10.1007/s00382-017-3758-y>
- Jankov, I., Gallus, W. A., Segal, M., & Koch, S. E. (2007). Influence of initial conditions on the WRF-ARW model QPF response to physical parameterization changes. *Weather and Forecasting*, 22(3), 501–519. <https://doi.org/10.1175/WAF998.1>
- Jerez, S., Montavez, J. P., Jimenez-Guerrero, P., Gomez-Navarro, J. J., Lorente-Plazas, R., & Zorita, E. (2013). A multi-physics ensemble of present-day climate regional simulations over the Iberian Peninsula. *Climate Dynamics*, 40(11–12), 3023–3046. <https://doi.org/10.1007/s00382-012-1539-1>
- Jones, P. D., & Lister, D. H. (2009). The urban heat island in Central London and urban-related warming trends in Central London since 1900. *Weather*, 64(12), 323–327. <https://doi.org/10.1002/wea.432>
- 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>
- Klecze, M. A., Steeneveld, G.-J., & Holtslag, A. A. M. (2014). Evaluation of the weather research and forecasting Mesoscale model for GABLS3: Impact of boundary-layer schemes, boundary conditions and spin-up. *Boundary-Layer Meteorology*, 152(2), 213–243. <https://doi.org/10.1007/s10546-014-9925-3>
- Kusaka, H., Kondo, H., Kikegawa, Y., & Kimura, F. (2001). A simple single-layer urban canopy model for atmospheric models: Comparison with multi-layer and slab models. *Boundary-Layer Meteorology*, 101(3), 329–358. <https://doi.org/10.1023/A:1019207923078>

- Laprise, R., Kornic, D., Rapaić, M., Šeparović, L., Leduc, M., Nikiema, O., et al. (2012). Considerations of domain size and large-scale driving for nested regional climate models: Impact on internal variability and ability at developing small-scale details. In A. Berger, F. Mesinger, & D. Sijacki (Eds.), *Climate change* (pp. 181–199). Vienna: Springer Vienna. https://doi.org/10.1007/978-3-7091-0973-1_14
- Laux, P., Nguyen, P. N. B., Cullmann, J., Van, T. P., & Kunstmann, H. (2017). How many RCM ensemble members provide confidence in the impact of land-use/land cover change? *International Journal of Climatology*, 37(4), 2080–2100. <https://doi.org/10.1002/joc.4836>
- Li, R., Liu, L., Yang, G., Zhang, C., & Wang, B. (2016). Bitwise identical compiling setup: Prospective for reproducibility and reliability of Earth system modeling. *Geoscientific Model Development*, 9(2), 731–748. <https://doi.org/10.5194/gmd-9-731-2016>
- Li, X. X., Koh, T. Y., Entekhabi, D., Roth, M., Panda, J., & Norford, L. K. (2013). A multi-resolution ensemble study of a tropical urban environment and its interactions with the background regional atmosphere. *Journal of Geophysical Research: Atmospheres*, 118, 9804–9818. <https://doi.org/10.1002/jgrd.50795>
- Li, X. X., & Norford, L. K. (2016). Evaluation of cool roof and vegetations in mitigating urban heat island in a tropical city, Singapore. *Urban Climate*, 16, 59–74. <https://doi.org/10.1016/j.uclim.2015.12.002>
- Lin, C. Y., Chen, F., Huang, J. C., Chen, W. C., Liou, Y. A., Chen, W. N., & Liu, S. C. (2008). Urban heat island effect and its impact on boundary layer development and land-sea circulation over northern Taiwan. *Atmospheric Environment*, 42(22), 5635–5649. <https://doi.org/10.1016/j.atmosenv.2008.03.015>
- Lordan, T., Lindberg, F., Jorba, O., Kotthaus, S., Grossman-Clarke, S., & Grimmond, C. S. B. (2013). High resolution simulation of the variability of surface energy balance fluxes across central London with urban zones for energy partitioning. *Boundary-Layer Meteorology*, 147(3), 493–523. <https://doi.org/10.1007/s10546-013-9797-y>
- Lucas-Picher, P., Caya, D., Biner, S., & Laprise, R. (2008). Quantification of the lateral boundary forcing of a regional climate model using an aging tracer. *Monthly Weather Review*, 136(12), 4980–4996. <https://doi.org/10.1175/2008mwr2448.1>
- Met Office (2012). Met Office Integrated Data Archive System (MIDAS) Land and Marine Surface Stations Data (1853-current). NCAS British Atmospheric Data Centre, 05/03/2018. <http://catalogue.ceda.ac.uk/uuid/220a65615218d5c9cc9e4785a3234bd0>
- Miao, S., Chen, F., LeMone, M. A., Tewari, M., Li, Q., & Wang, Y. (2009). An observational and modeling study of characteristics of urban heat island and boundary layer structures in Beijing. *Journal of Applied Meteorology and Climatology*, 48(3), 484–501. <https://doi.org/10.1175/2008JAMC1909.1>
- Mooney, P. A., Broderick, C., Bruyère, C. L., Mulligan, F. J., & Prein, A. F. (2017). Clustering of observed diurnal cycles of precipitation over the United States for evaluation of a WRF multiphysics regional climate ensemble. *Journal of Climate*, 30(22), 9267–9286. <https://doi.org/10.1175/JCLI-D-16-0851.1>
- Morini, E., Touchaei, A. G., Rossi, F., Cotana, F., & Akbari, H. (2018). Evaluation of albedo enhancement to mitigate impacts of urban heat island in Rome (Italy) using WRF meteorological model. *Urban Climate*, 24, 551–566. <https://doi.org/10.1016/j.uclim.2017.08.001>
- National Centers for Environmental Prediction/National Weather Service/NOAA/U.S. Department of Commerce (2000). NCEP FNL Operational Model Global Tropospheric Analyses, continuing from July 1999, <https://doi.org/10.5065/D6M043C6>, Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory, Boulder, Colo. (Updated daily.) Accessed 01 01 2019.
- Pasquill, F., & Smith, F. B. (1984). *Atmospheric diffusion*, *Quarterly Journal of the Royal Meteorological Society* (3rd ed.). Chichester: Ellis Horwood Limited.
- Powers, J. G., Klemp, J. B., Skamarock, W. C., Davis, C. A., Dudhia, J., Gill, D. O., et al. (2017). The weather research and forecasting model: Overview, system efforts, and future directions. *Bulletin of the American Meteorological Society*, 98(8), 1717–1737. <https://doi.org/10.1175/BAMS-D-15-00308.1>
- Sertel, E., Ormeci, C., & Robock, A. (2011). Modelling land cover change impact on the summer climate of the Marmara region, Turkey. *International Journal of Global Warming*, 3(1/2), 194. <https://doi.org/10.1504/IJGW.2011.038379>
- Sharma, A., Fernando, H. J. S., Hamlet, A. F., Hellmann, J. J., Barlage, M., & Chen, F. (2017). Urban meteorological modeling using WRF: A sensitivity study. *International Journal of Climatology*, 37(4), 1885–1900. <https://doi.org/10.1002/joc.4819>
- Simm, W. A., Samreen, F., Bassett, R., Ferrario, M. A., Blair, G., Whittle, J., & Young, P. J. (2018). SE in ES. In *Proceedings of the 40th International Conference on Software Engineering Software Engineering in Society—ICSE-SEIS'18* (pp. 61–70). New York, New York: ACM Press. <https://doi.org/10.1145/3183428.3183430>
- Siuta, D., West, G., Modzelewski, H., Schigas, R., & Stull, R. (2016). Viability of cloud computing for real-time numerical weather prediction. *Weather and Forecasting*, 31(6), 1985–1996. <https://doi.org/10.1175/WAF-D-16-0075.1>
- Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D. M., Duda, M. G., et al. (2008). A description of the advanced research WRF Version 3. Technical Report, (June), 113. <https://doi.org/10.5065/D6DZ069T>
- Sörensson, A. A., & Berbery, E. H. (2015). A note on soil moisture memory and interactions with surface climate for different vegetation types in the La Plata Basin. *Journal of Hydrometeorology*, 16(2), 716–729. <https://doi.org/10.1175/JHM-D-14-0102.1>
- Stegehuis, A. I., Vautard, R., Ciais, P., Teuling, A. J., Miralles, D. G., & Wild, M. (2015). An observation-constrained multi-physics WRF ensemble for simulating European mega heat waves. *Geoscientific Model Development*, 8(7), 2285–2298. <https://doi.org/10.5194/gmd-8-2285-2015>
- Tomlinson, C. J., Chapman, L., Thorne, J. E., & Baker, C. J. (2012). Derivation of Birmingham's summer surface urban heat island from MODIS satellite images. *International Journal of Climatology*, 32(2), 214–224. <https://doi.org/10.1002/joc.2261>
- Touchaei, A. G., & Wang, Y. (2015). Characterizing urban heat island in Montreal (Canada)—Effect of urban morphology. *Sustainable Cities and Society*, 19, 395–402. <https://doi.org/10.1016/j.scs.2015.03.005>
- Tse, J. W. P., Yeung, P. S., Fung, J. C.-H., Ren, C., Wang, R., Wong, M. M.-F., & Cai, M. (2018). Investigation of the meteorological effects of urbanization in recent decades: A case study of major cities in Pearl River Delta. *Urban Climate*, 26, 174–187. <https://doi.org/10.1016/j.uclim.2018.08.007>
- Wan, H., & Zhong, Z. (2014). Ensemble simulations to investigate the impact of large-scale urbanization on precipitation in the lower reaches of Yangtze River Valley, China. *Quarterly Journal of the Royal Meteorological Society*, 140(678), 258–266. <https://doi.org/10.1002/qj.2125>
- Wang, J., Huang, B., Fu, D., Atkinson, P. M., & Zhang, X. (2016). Response of urban heat island to future urban expansion over the Beijing-Tianjin-Hebei metropolitan area. *Applied Geography*, 70, 26–36. <https://doi.org/10.1016/j.apgeog.2016.02.010>
- Wang, Z. H., Bou-Zeid, E., Au, S. K., & Smith, J. A. (2011). Analyzing the sensitivity of WRF's single-layer urban canopy model to parameter uncertainty using advanced Monte Carlo simulation. *Journal of Applied Meteorology and Climatology*, 50(9), 1795–1814. <https://doi.org/10.1175/2011JAMC2685.1>
- Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301), 236–244. <https://doi.org/10.1080/01621459.1963.10500845>

- Wu, W., Lynch, A. H., & Rivers, A. (2005). Estimating the uncertainty in a regional climate model related to initial and lateral boundary conditions. *Journal of Climate*, 18(7), 917–933. <https://doi.org/10.1175/JCLI-3293.1>
- Zhang, N., Zhu, L., & Zhu, Y. (2011). Urban heat island and boundary layer structures under hot weather synoptic conditions: A case study of Suzhou City, China. *Advances in Atmospheric Sciences*, 28(4), 855–865. <https://doi.org/10.1007/s00376-010-0040-1>
- Zhong, S., & Yang, X.-Q. (2015). Ensemble simulations of the urban effect on a summer rainfall event in the Great Beijing Metropolitan Area. *Atmospheric Research*, 153, 318–334. <https://doi.org/10.1016/j.atmosres.2014.09.005>
- Zhou, X., Ooka, R., Chen, H., Kawamoto, Y., & Kikumoto, H. (2016). Impacts of inland water area changes on the local climate of Wuhan, China. *Indoor and Built Environment*, 25(2), 296–313. <https://doi.org/10.1177/1420326X14546774>