On the Spin-Up Period in WRF Simulations Over Europe: Trade-Offs Between Length and Seasonality

Sonia Jerez1, Jose María López-Romero1, Marco Turco1, Raquel Lorente-Plazas1,2, Juan José Gómez-Navarro1, Pedro Jiménez-Guerrero1, and Juan Pedro Montávez2

1Regional Atmospheric Gómez-Navarro Group, Department of Physics, University of Murcia, Murcia, Spain, 2Department of Meteorology, Meteored, Spain

Abstract Regional climate models (RCMs) are usually initialized and driven through the boundaries of their limited area domain by data provided by global models (GCMs). The mismatch between the low-resolution GCM initial conditions and RCM’s high resolution introduces physical inconsistencies between the various components of the RCM. These inconsistencies can be resolved by running the RCM during a period that is considered unreliable: the spin-up period. There is no deterministic definition of the length that the spin-up period should have. Here we try to provide general guidelines that can be used to the advantage of the community. We base our analysis on Weather Research and Forecasting (WRF) simulations over a Euro-Cordex compliant domain and find that for 2-m temperature and precipitation, rather short spin-up periods (1 week) can be sufficient. Furthermore, longer periods (6 months) are advisable, and start dates in non-winter months should be pursued, as this ensures a more realistic representation of the snow cover. Thus, the issue is not only about the spin-up length. As the soil subsystem evolves slowly and requires longer periods to reach equilibrium than the longest considered here (1 year), seasonality plays an important role in minimizing the impact of the unreliability of the soil initialization. Fortunately, except for goals where the deep soil-atmosphere feedback are critical, the lack of equilibrium between them can be ignored, as it seems to have little effect on the simulation of the atmospheric variables most frequently used in RCM studies.

Plain Language Summary High-resolution climate simulations performed with the so-called regional climate models (RCMs) are usually initialized from coarser databases that provide an inconsistent picture at the higher resolution of the RCM. To overcome it and obtain reliable RCM simulations, RCMs should run over a time span to reach physical equilibrium. During this “trash” period (called spin-up period), the RCM outputs should be discarded for later analysis, for example, for assessing climate change impacts. In order to guarantee the reliability of RCM simulations and minimize their high computational cost, determining an optimal length of the spin-up period is critical. For a European domain with ~50-km resolution and for a particular, but widely used, RCM, the WRF model, here we find that half a year of spin-up would be enough for atmospheric variables (temperature and precipitation), always that the RCM initialization occurs in summer months; otherwise, the snow cover initial state is misrepresented, making it inadvisable longer spin-up periods implying winter start dates. However, soil variables (especially water content) cannot be trusted even after up to one full year of spin-up. Fortunately, the negligible impact of this issue for most RCM applications involving only atmospheric variables emerges also from our analysis.

1. Introduction

The primary tool for providing trustworthy information for climate projections is the use of both global and regional climate models (GCMs and RCMs, respectively). However, high-resolution global climate simulations are computationally very expensive. Conversely, RCMs have demonstrated their skill in downscaling global information and better reproducing mesoscale and local features over a limited region at affordable computational costs (e.g., Gómez-Navarro et al., 2011; Jacob et al., 2014; Jerez et al., 2013; Rummukainen, 2010; Schewe et al., 2019). GCMs and RCMs numerically solve a set of differential equations governing the climate system over a spatial domain, after required discretization and parameterization procedures. For that they need to be provided with both initial and boundary conditions (IC and BC, respectively). In
the case of GCMs, periodic BC make the problem exclusively dependent on IC, while RCMs require continuous updates of the conditions at the domain borders, aside from the external forcing (Jerez et al., 2018). Hence, RCMs are usually nested into GCMs (to perform dynamical downscaling), which drive the regional simulations through their domain boundaries. This makes RCM outputs less dependent on the IC and more dependent on the BC as compared to global simulations.

Based on the belief that RCMs actually forget IC after some execution time, with the aim of optimizing computational resources and therefore of minimizing delivery times, a common practice (e.g., Gómez-Navarro et al., 2011, Jerez et al., 2013, 2018) consists in partitioning the period to be simulated into sub-periods (chunks) that are then run in continuous executions of the regional model. Each chunk should include a spin-up period guaranteeing that the model outputs at the beginning of the target period are as they would have been from a continuous non-partitioned run. This is required to avoid inhomogeneities when sub-periods are concatenated, but also in continuous runs to avoid masking effects in the model outputs due to IC issues. This spin-up period is thus the execution time required so that the RCM has fully reached its physical equilibrium state, within the path indicated by the BC, forgetting about the IC (Anthes et al., 1989; Denis et al., 2002; Giorgi & Mearns, 1999; Yang et al., 1995). During this period, the model outputs may be tainted and must be discarded (Cosgrove et al., 2003; Seck et al., 2015).

If the required spin-up period has length $S$ (in time units), the target period has length $P$ (in time units) and the number of partitions or chunks is $N$, the set of model executions should simulate a total of $N(S/P/N+S)=P+NS$ time units. Without the partitioning approach (i.e., $N=1$), the length of the simulated period reaches its minimum: $P+S$ time units. Nonetheless, depending on the computational architecture and code parallelization constraints, the partitioning approach could minimize the total time needed for the pursued climate simulation to be ready, especially for small values of $S$ (Jerez et al., 2009). Let us call $T_s$ the execution time needed to simulate one unit of time with the model in its serial version, that is, using one single processor per job, and $T_p$ the time needed to simulate one unit of time with the model in its parallel version, that is, using $X$ processors per job (with $X \geq 1$). $T_p$ is always lower than $T_s$, but not as much as $T_s/X$ because of the time employed in the transfer of information between processors (especially in the case of distributed memory parallelization, while the overhead shall be very little in shared modes), so $T_s/X < T_p < T_s$. Let us assume that one has $X$ processors that can be employed either in serial or parallel mode, and let us compare the cases $N=1$ with the model code running in parallel in the $X$ processors to simulate a period of $P+S$ time units, and $N=X$ with the model code in $X$ serial runs, one per processor simulating a period of $P/N+S$ time units. The execution time in the first case is given by $(S+P)T_p$; in the second case by $(S+P/N)T_s$. The length of $S$ is thus critical in order to decide which strategy is the best.

Beyond strategies, the overall question concerns when RCM stores adjust from the IC to an equilibrium state, if they do. However, the determination of the optimum spin-up period is very complex since it depends on many factors. First, the climate system is composed of an extensive number of components with different response time scales. For instance, longer spin-up periods are required for surface hydrological variables, such as soil moisture (Christensen, 1999), than for fine-scale atmospheric components, which respond faster to dynamical and thermo-dynamical forcings (Cholette et al., 2015; De Elía et al., 2002; Doblas-Reyes et al., 2013; Skamarock, 2004; Zhong et al., 2007). Second, each RCM has a different internal variability due to different internal model physics and dynamics (Tebaldi & Knutti, 2007), which makes the problem model dependent. The third factor is the domain size, because the BC influence decays as we move far away from the boundaries (Leduc & Laprise, 2009), which gives rise to the so-called spatial spin-up (Matte et al., 2017). Fourth, IC are obtained by spatial interpolation of data provided by GCMs that usually have different soil physics and structure to RCMs. The greater the inconsistencies between IC and RCM physics (Turco et al., 2013), the longer the required spin-up periods (Jacob & Podzun, 1997). Indeed, some authors have explored alternative strategies based on spinning up the land surface model (and any other more slowly evolving components, e.g., the ocean model) for a longer period than the atmospheric model (Angevine et al., 2014; Koster et al., 2010) or initializing the RCM runs from own RCM outputs of previous realizations (Christensen, 1999). Fifth, the synoptic/meteorological conditions can also affect the spin-up requirement (Yang et al., 2011; Zhong et al., 2007). Finally, the spin-up will be constrained by extreme conditions, for example, the presence of ice and snow cover, extremely dry conditions, or extremely hot or cold temperatures (Day et al., 2014; Seck et al., 2015).
Several studies have focused on establishing optimum spin-up periods for land surface models (Chen et al., 1997; Cosgrove et al., 2003; Yang et al., 1995) and mesoscale atmospheric models (Angevine et al., 2014; Christensen, 1999; Zhong et al., 2007), while less attention has been paid to hydrological models (Ajami et al., 2014; Seck et al., 2015). The proposals range from days (Gomez et al., 2015; Zhong et al., 2007) to decades (Christensen, 1999; Cosgrove et al., 2003), depending on the target of study. In particular, Cosgrove et al. (2003) showed how soil moisture spins up slowly, while evaporation and deep soil temperature spin-up comparatively more quickly. Any spin-up behavior that was identified for soil temperature and evaporation disappeared within the first 6 months, while for soil moisture it took over 10 years to disappear. These findings are in overall agreement with those by Christensen (1999). Seck et al. (2015) compared the spin-up for dry and wet IC, indicating longer system memory when dry initial states were considered. On the other hand, Zhong et al. (2007) and Gómez-Navarro et al. (2015) found that a 4- to 8-day spin-up time was enough for atmospheric variables in the context of the abnormal climate event during the summer of 1998 in China and recent wind storms in Europe, respectively. Therefore, there is not a simple straightforward answer to how long the optimum spin-up period should be.

Indeed, a variety of different approaches are currently being implemented over Europe under the umbrella of Euro-Cordex (Jacob et al., 2014, https://euro-cordex.net). Vautard et al. (2013) used 21 runs of several state-of-the-art RCMs to assess the ability of this ensemble of models to accurately simulate heat waves at the regional scale. Among them, 10 model runs initialized soil variables using reanalysis data and used no spin-up period, while, at the opposite extreme, two simulations used a spin-up period of over a year. Katragkou et al. (2015) used a spin-up period of 1 year to initialize six hindcast WRF (Weather Research and Forecasting model; Skamarock et al., 2008) simulations with different configurations affecting the microphysics, convection, and radiation schemes. According to the authors, the 1-year spin-up time allowed for adjustment of the soil moisture and temperature. The same spin-up period (1 year) was used in Katragkou et al. (2017). Last, Prein et al. (2016) used simulations from eight different models of the Euro-Cordex ensemble (or model versions in the case of WRF) to check whether there was a modification in simulated precipitation when increasing the horizontal resolution. Four of the runs used no spin-up period (soil was initialized from reanalysis data), including the two WRF experiments; two used a 1-year spin-up period; and two were initialized 10 years in advance. No objective criterion is given for the selection of the spin-up time in any of the aforementioned references.

Hence, the objective of this contribution is to assess the sensitivity to the spin-up period of WRF simulations over an extensive domain covering Europe, for both land-surface and atmospheric variables, in order to suggest recommendations of optimum spin-up times for future regional climate simulations.

2. Data and Methods

2.1. RCM Configuration

The regional climate simulations were carried out with the state-of-the-art widely used WRF model in its 3.6.1 version (available at https://www.mmm.ucar.edu/weather-research-and-forecasting-model, last accessed on 19 July 2019). IC and BC were taken from the Coupled Model Intercomparison Project Phase 5 (CMIP5) experiment r1i1p1 MPI-ESM-LR historical run (Giorgetta et al., 2012; doi:10.1594/WDC/RCM_CMIP5_historical-LR; available at https://cera-www.dkrz.de, last accessed on 19 July 2019), avoiding the use of reanalysis data that may reduce spin-up requirements. BC were updated every 6 hr for the RCM runs without nudging, including the low BC for sea surface temperature. The spatial configuration of the RCM consists of a Euro-Cordex compliant domain covering Europe with a spatial resolution of 0.44° in both latitude and longitude. In the vertical dimension, 29 unevenly spaced eta levels were specified, with more levels near the surface than upward, and the model top was set to 50 hPa. The physics configuration of the WRF model consisted of the Lin microphysics scheme (Lin et al., 1983), the RRTM radiative scheme (Iacono et al., 2008), the Grell 3-D ensemble cumulus scheme (Grell, 1993; Grell & Dévényi, 2002), the University of Yonsei boundary layer scheme (Hong et al., 2006), and the Noah land surface model (Chen & Dudhia, 2001; Tewari et al., 2004). The latter consists of a model with four layers (10, 30, 60, and 100 cm wide, from top to bottom), with gravitational moisture flow and constant temperature at the deepest boundary, which includes frozen-ground physics, patchy snow cover, time-varying snow density and snow
roughness length, heat flux treatment under snow pack, and modified soil thermal conductivity according to the water content.

2.2. Experiments and Data

A set of ten 1-year long simulations was carried out, each simulation with a different spin-up period. The simulated year was arbitrarily selected as 1982 (the target period). The longest spin-up period considered here covers 2 years before the start date of the target period. The corresponding simulation, thus starting on 1 January 1980 and spanning a total of 3 years, is taken here as a reference. The other simulations have spin-up periods (in ascending order) of 0 hr, 12 hr, 1 day, 1 week, 1 month, 3 months, 6 months, 9 months, and 1 year, thus starting at some point in the year before the target one (i.e., 1981 in this case).

Two additional experiments were carried out using RCM data to initialize the soil variables. For that, first, the period from 1 January 2008 to 31 January 2010 was simulated (chosen so that it did not comprise the target year). Then, the 31-day running means of 2009 from this simulation were used to initialize the soil variables in the simulations of the year 1982. This way, two simulations were performed: with 1 day and 3 months of spin-up, thus using the 31-day running means centered on 31 December 2009 and 1 October 2009 from the aforementioned run, respectively, as IC for the soil variables.

As mentioned above, land-surface variables spin up slower than atmospheric ones. Here we inspect both kinds: near-surface (2-m height) air temperature (T2; units: °C), precipitation (PR; units: mm/day), soil layer moisture content (SMOIS; units: % of saturation), and soil temperature at the bottom of the layer (TSLB; units: °C); the latter two for the four soil layers considered by default in the land surface model (layer 1 at the top and layer 4 at the bottom), accordingly labeled SMOIS1-to-4 and TSLB1-to-4. The RCM outputs of these variables were recorded every 3 hr and then time averaged up to the daily time scale.

2.3. Methodology

First, we assessed the behavior of the spatially averaged (over all the land grid points of the domain) 5-day running mean daily series. The running mean was performed to filter stochastic lags in the raw daily series due to high-frequency meteorological-like fluctuations derived from model internal variability, which has nothing to do with climate behaviors. Boxplots of the differences in these series, considering only the target period (year 1982), between each experiment and the reference run were used to visualize the widths of their distributions. In order to provide a context for the amplitude of these differences, the standard deviation ($\sigma$) of the 5-day running mean reference series in the target year was used, computed after removing the low-frequency signals (i.e., the annual cycle). For that, a high-pass filter was applied to suppress all variability on timescales greater or equal to 1 year, using the Climate Data Operator (CDO, https://code.mpimet.mpg.de/projects/cdo) highpass function with a threshold for the minimum frequency per year that passes the filter larger than 1 (we used 2), as recommended in the CDO documentation, since there could be dominant frequencies around the 1-year peak in the fast Fourier transformed series (if there is one) as well due to the issue that the time series are not of infinite length. The raw and high-pass filtered 5-day running mean reference series are displayed in Figure S1 in the supporting information.

Second, we computed the root-mean-square error (RMSE) between the 1982 5-day running mean daily series retrieved from each experiment and the reference series at the grid point level. In order to elucidate statistically significant changes (reductions or increases) in the RMSE values with longer spin-up periods, we applied a resampling method as in Milelli et al. (2010). We rejected the null-hypothesis that there is no difference in the RMSE between two simulations with different spin-up periods with a 95% of confidence level according to the method.

Finally, the similarity between the cumulative probability density functions (CDFs) constructed from the daily series of each simulation was also analyzed at the grid point level (for this analysis, we do not need filter lags due to model internal variability since the CDFs do not account for chronology). For that, the nonparametric Kolmogorov-Smirnov test (Lilliefors, 1967) was used. In order to reject the null-hypothesis that two CDFs were drawn from the same sample distribution, we imposed $p<0.05$, thus a confidence of 95% according to the method. Although this analysis accounts for the similarity among all quantities in the distributions, specific impacts on their far-end right side were additionally evaluated by direct comparison of the 90th percentile values of the daily T2 and PR series (T2p90 and PRp90, respectively) between the various experiments.
and the reference run. The statistical significance of differences in T2p90 and PRp90 were again assessed by 1000 bootstrap replications as detailed in Milelli et al. (2010), imposing 95% confidence.

3. Results

3.1. Atmospheric Variables

To give insight into the general behavior of the experiments, Figure 1 shows the time series of differences between the 5-day running mean spatially averaged series (over land) of T2 and PR retrieved from each simulation (with spin-up periods varying from 0 hr to 1 year; see legend) and the reference run (with 2 years of spin-up). Differences during both the overlapping spin-up period (gray shaded side of the graph) and the target period are shown. To give a context, the horizontal black solid lines above and below the zero-line represent one standard deviation, above and below zero respectively, of the high-pass-filtered 5-day running mean reference series in the target period (see Figure S1). To the right, boxplots show the statistical distribution of the differences during the target period, characterized by the median (50th percentile; horizontal lines), the 25th–75th interquartile range (boxes) and the 5th–95th interpercentile range (whiskers extending from the box limits).

Figure 1. Difference between the spatially averaged (over all land grid points) 5-day running mean time series of (a) T2 and (b) PR reproduced by each experiment (with spin-up periods ranging from 0 hr to 1 year; see legend) and by the reference run (with 2 years of spin-up). Differences during both the overlapping spin-up period (gray shaded side of the graph) and the target period are shown. To give a context, the horizontal black solid lines above and below the zero-line represent one standard deviation, above and below zero respectively, of the high-pass-filtered 5-day running mean reference series in the target period (see Figure S1). To the right, boxplots show the statistical distribution of the differences during the target period, characterized by the median (50th percentile; horizontal lines), the 25th–75th interquartile range (boxes) and the 5th–95th interpercentile range (whiskers extending from the box limits).

The T2 differences series (Figure 1a) oscillate closely around zero, in the sense that they remain well below one standard deviation of the filtered reference series (in particular, for the experiments with 1 week or longer spin-up periods) but for a short period at the beginning of the summer season, when a large portion of the experiments present notable deviations from the zero-line. Remarkably, the confinement of the T2 differences around zero does not increase monotonously with larger spin-up periods, as could have been expected. It can be appreciated in the boxplot to the right of Figure 1a that both the 25th–75th and 5–95th interpercentile ranges sometimes diminish and others increase when moving toward experiments with larger spin-up periods. This feature becomes even more apparent in Figure 2, where the RMSE of the 5-day running mean T2 series from each experiment (when faced to the series of the reference run) is compared to the RMSE of each predecessor experiment. RMSE is significantly smaller in the experiment with 12 hr of spin-up than in the experiment with 0 hr of spin-up over a few grid points (Figure 2d), but it is larger in the
Figure 2. Panel a shows the root-mean-square error (RMSE) between the 5-day running mean T2 series from the experiment with 0 hr of spin-up and those from the reference run. Panels d to l show differences between the RMSE of each experiment (with spin-up periods ranging from 12 hr to 1 year) and the RMSE of the previous experiment (i.e., 12 hr vs. 0 hr in panel d, 1 day vs. 12 hr in panel e, and 1 week vs. 1 day in panel f, etc.) in order to display how the RMSE decreases or increases with consecutive larger spin-up periods. These differences are colored only if statistically significant (with $p<0.05$). The numbers in panels indicate the percentage of land grid points in which the RMSE for the corresponding experiment is minimum among all experiments (it is a rounded integer number). These land grid points are highlighted by squares in the corresponding maps. Panel b shows the RMSE of the experiment with the highest number of land grid points in which the RMSE is minimum (the “best” experiment). Panel c shows differences in % between the RMSE of the best experiment and the RMSE of the experiment with 0 hours of spin-up shown in panel a (colored only if statistically significant).
experiment with 1 day of spin-up than in the experiment with 12 hr of spin-up over a wider area in central Europe (Figure 2e). Over the same area, it decreases again in the experiment with 1 week of spin-up as compared to the experiment with 1 day of spin-up (Figure 2f); but then increases again in the experiment with 1 month of spin-up (Figure 2g). A significant reduction in the RMSE of the T2 series is achieved afterward with the experiment with 6 months of spin-up (Figure 2j). In fact, this experiment achieves the largest number of land grid points (52%) with minimum RMSE values (thus becoming the “best” member), even though experiments with longer spin-up periods are included in this analysis. This best member actually reduces the RMSE as compared with the experiment with 0 hr of spin-up (Figures 2a–2c), by 2°C as we move northeastward, which is by up to 50%.

For PR, difference series in Figure 1b have amplitudes that reach easily the magnitude of the standard deviation of the filtered reference series. The 5th–95th interpercentile ranges of these differences lie within the 1–σ interval only for the experiments with 6 and 9 months of spin-up. Contrary to the case of the T2 series, the confinement of the differences in PR around zero increases gradually with larger lengths of the spin-up period, with the sole exception of the experiment with 1 year of spin-up. However, the maps showing differences in RMSE at the grid point level between consecutive experiments (Figures 3d–3l) display noisy patterns, with both positive and negative isolated patches, suggesting the prevailing role of model internal variability. Nevertheless, the experiment with 6 months of spin-up stands out again as the best member, with nearly 30% of the land grid points holding the lowest RMSE in this experiment (Figure 3j), which is twice as much as the next best member. Compared to the experiment with no spin-up, these lower RMSE values imply an error reduction of around 30–60%, mainly over the westernmost and easternmost parts of the domain (Figure 3c). The distances between the rest of the experiments are not as evident as in the case of T2 and not as structured in terms of the added value of long spin-up periods. For instance, the second and third best members, with 9 months and 1 week of spin-up lengths, respectively, hold 15% and 14% of grid points with minimum RMSE, respectively, and the fourth and fifth best members, with 1 year and 4 months of spin-up lengths, respectively, hold 10% and 9% of grid points with minimum RMSE, respectively.

### 3.2. Soil Variables

The soil temperature series in the upper layer (Figures 4a and 5a–5c) exhibit a nonsurprising similar behavior to that already discussed for T2 (T2 is interpolated between the soil surface and the lowest atmospheric model level), although some distinct features stand out. First, the medians of these differences are systematically negative except for the runs initialized in summer months, that is, those with 3, 4, and 6 months of spin-up (boxplots in Figure 4a). Second, while the RMSE maps in Figures 5a and 5b display generally softer intensities as compared to the RMSE maps for T2, for both the experiment with no spin-up and the best member (again, the experiment with 6 months of spin-up, which performs the best over more than a half of the grid points), the RMSE in the experiment with no spin-up is larger over some mountainous regions (e.g., The Alps) and north-easternmost areas as compared to the map in Figure 2a.

The medians of the differences between TSLB series in deep soil layers are also negative except for the runs initialized in summer months (boxplots in Figures 4b and S2a and S2b). For the deepest layer (Figure 4b), these differences have magnitudes larger than one standard deviation of the filtered reference series in those runs with spin-up periods shorter than 1 month. This occurs in spite of the fact that the series evolve relatively fast to correct the striking initial biases, mostly positive (thus following the Newton’s exponential law of cooling), likely because of the very inertial nature of deep soil variables. RMSE in TSLB4 exhibits, however, very low values (~0.5°C) in the best experiment, the one with 6 months of spin-up, in which 30% of the grid points present the lowest RMSE values (Figures 5d–5f), showing a strong reduction in the signals depicted by the experiment with 0 hr of spin-up (see a discussion about these signals in the supporting information, around Figure S11). Notably, as we go deeper in the soil layers, the percentage of land grid points presenting the lowest RMSE in the experiment with 6 months of spin-up decreases in favor of the experiments with longer spin-up periods, in particular the one with 9 months of spin-up (see Figures S3–S6). This indicates that longer spin-up periods are required to overcome the deviations in the IC as we go deeper, likely due to the lag in the response to the atmospheric forcings in deep soil layers because the layers above smooth and filter the downward traveling signals, and also because deep soil layers are wider and so more inertial.
Regarding moisture content, soil layers are initially too wet (Figures 4c and 4d) when compared to the results from the reference simulation (so the curves first exhibit a decay). This bias is most overcome, in terms of spatial averages, in the runs with long spin-up periods in the case of upper layers (Figures 4c and S2c) and for those initialized in summer months in the case of bottom layers (Figures 4d and S2d), suggesting

**Figure 3.** As Figure 2 for PR.

Regarding moisture content, soil layers are initially too wet (Figures 4c and 4d) when compared to the results from the reference simulation (so the curves first exhibit a decay). This bias is most overcome, in terms of spatial averages, in the runs with long spin-up periods in the case of upper layers (Figures 4c and S2c) and for those initialized in summer months in the case of bottom layers (Figures 4d and S2d), suggesting
a prominent role of seasonality in the initialization of the simulations. In terms of percentage of land grid points with minimum RMSE in the SMOIS variable (Figures 5i and 5l and S7–S10), small differences exist between the experiments with 6, 9, and 12 months of spin-up. Interestingly, although these percentages increase in the experiments with longer spin-up periods as we focus on deeper layers, these longest best experiments tend, however, to present worse RMSE values than the experiment with 0 hr of spin-up over Figure 4. As Figure 1 for TSLB in layers (a) 1 and (b) 4 and SMOIS in layers (c) 1 and (d) 4.
some areas (remarkably, over the Scandinavian region and a patched latitudinal strip in central Europe; see a discussion about these signals in the supporting information, around Figure S12), a feature absent for other variables.

Figure 5. As the first row in Figure 2 for TSLB in layers 1 (a-c) and 4 (d-f) and SMOIS in layers 1 (g-i) and 4 (j-l).
3.3. A synthesis From a Different Perspective: Probability Distributions

Figure 6 shows, for each variable, the percentage of land grid points where the CDFs constructed from the daily data of the experiments with spin-up periods ranging from 0 hr to 1 year are statistically different from those corresponding to the reference run according to the K-S test described in section 2.3. In agreement with the results of sections 3.1 and 3.2, this analysis further highlights two facts. First, there is an
increasing number of land grid points where CDFs are discordant as we move from atmospheric to soil variables, and as we move into deeper soil layers. Also, the percentage of land grid points with discordant CDFs is greater for PR than for T2, and, among the soil variables, greater for SMOIS than for TSLB. Second, the percentage of land grid points with discordant CDFs diminishes with longer spin-up periods (some nuances), at least up to a point. For T2, such a decline is hardly appreciable, as it is for PR from the fourth experiment (with 1 week of spin-up) onward. Conversely, it is most evident for TSLB. For instance, in the case of TSLB3, 70% of land grid points present discordant CDFs in the experiment with no spin-up, which drops to 40% in the experiment with 6 months of spin-up. For all these variables (T2, PR, and TSLB from layers 1 to 4), the minimum (thus, the best result) is achieved in the experiment with 6 months of spin-up, in agreement with previous analyses. SMOIS CDFs do not pass the K-S test for most of the domain whatever the spin-up period is, but the best results are generally achieved after 9 months of spin-up (although with very small differences with the results obtained after 6 months of spin-up). The experiment with 1 year of spin-up does not provide the best results in any case.

The maps at the bottom of Figures 6d to 6m provide, for the best case (i.e., for the experiment with the highest number of land grid points which passed the K-S test), a spatial disaggregation of the aforementioned information. Since T2 and PR are not problematic variables, the corresponding maps display mostly white. For TSLB, the area with discordant CDFs grows southward from the top right corner of the domain (northeastern Europe) as we go into deeper soil layers. Most of the domain displays black for SMOIS, more and more as we go deeper and deeper.

Given the potential impact of inaccurate soils on, specifically, the far-end right side of the distribution of the atmospheric variables, the 90th percentile values of the daily series of T2 and PR (T2p90 and PRp90, respectively) from the various experiments and the reference run were directly compared (see Figure S13). This evaluation revealed overall non-significant differences in PRp90 between any of the simulations and the reference one, and non-significant differences in T2p90 between the simulation with 6 months of spin-up and the reference one. All the other experiments depict significant differences in T2p90 somewhere (positive or negative; the former mostly northward, the latter mostly southward), likely related to the deficit and excess of the soil moisture content of the first soil layer during the summertime (when the highest temperatures are reached) as compared to the reference state (see Figure S14), as this leads, respectively, to an enhanced/reduced exchange of sensible heat at the land-atmosphere interface and thus to higher/lower extreme temperatures (Seneviratne et al., 2010). These results are, therefore, in overall agreement with the findings presented so far.

### 3.4. Soil Initialization and Seasonality

Soil initialization plays a dominant role in determining the spin-up requirement. On the one hand, soil variables, especially soil moisture and, especially, at deep soil layers, do not converge to the reference state even after 1 year of spin-up (e.g., Figure 6). On the other hand, as pointed out above, the simulation with 6 months of spin-up overall outperforms others with longer spin-up periods. We analyzed this in more detail to find the reason for this counterintuitive finding. It seems that, throughout Europe, soil variables are more stable and better modeled during the summer season, at least at the surface, and thus easier to be correctly initialized using coarse GCM outputs. Figure 7 illustrates this argument. It shows time series of the percentage of land area covered by snow during the target year (1982) and the preceding one (1981) for all the runs, including the reference run with 2 years of spin-up. For the latter, during the year preceding the target period (along which all the other experiments start), the temporal standard deviation of the variable (snow coverage) in a 30-day moving window is displayed by shadows around the corresponding time series. Snow coverage is underestimated in all runs initialized in non-summer months. The runs initialized before summer (thus with more than 6 months of spin-up) converge to zero in summer (the “correct” value according to the reference run) and, from there on, they follow the reference series, as do the runs initialized in summer (see also panels b to i in Figure 7). The runs initialized after summer (0 hr to 1 month of spin-up), especially those with the shortest spin-up periods (1 week or less), do not fully converge to the reference series until the summer of the following year (when all series drop to zero again), with visible errors thus still persisting during the first half of the target year. But the risk of initializing in non-summer months is not only about the risk of biased soil variables initial conditions, but also about the temporal variability of the soil variables.
Figure 7. (a) Time series of the percentage of land area covered by snow during the target year (1982) and the preceding year (1981) for all the runs, including the reference run with 2 years of spin-up (see legend). For this reference run, during the year preceding the target period (along which all the other experiments start), the temporal standard deviation of the variable (snow coverage) in a 30-day moving window is displayed by shadows around the corresponding time series. Panels b to i show the snow coverage per grid cell (in % of the grid cell area) at different dates (1st January 1981, the start date for the experiment with 1 year of spin-up, in the first column; 1st July 1981, the start date for the experiment with 6 months of spin-up, in the second column; and 1st January 1982, the start date of the target period, in the third column) in different experiments (the reference run with 2 years of spin-up in the first row; the experiment with 1 year of spin-up in the second row, and the experiment with 6 months of spin-up in the third row).
being initialized, which are most stable in summer (see the shadowed range in Figure 7a), when thus initial conditions are less dependent on the exact initial date of the run.

One way to correct the aforementioned inconsistencies between the GCM-provided initial conditions and the RCM-developed ones could consist in taking the land IC from a previously spun-up RCM simulation. The asterisks in Figure 6a represent the results from the additional experiments using RCM data to initialize the soil variables, although these mismatch the initial dates of the runs, and thus, inconsistencies are still to be expected. With this approach (described in detail in section 2.2), the simulation with 1 day of spin-up behaves similarly to that with 3 months of spin-up initialized with GCM data, and the simulation with 3 months of spin-up performs the best, thus even (slightly) better than the simulation with 6 months of spin-up initialized with GCM data. This means that this approach reduces to about one half the spin-up length required to obtain the best results (always within the range of spin-up lengths studied here).

However, in this new best case, CDFs of soil moisture still remain significantly different to those from the reference run in more than 50% of land grid points. Hence, it is still difficult to assume that the soil is in full equilibrium.

In order to further disentangle the contribution of seasonality (spin-up initialization date) and length of the spin-up to the results presented in sections 3.1–3.3, we conducted an additional analysis shown in Figures 6b and 6c. Here, the similarity of the CDFs obtained from the different experiments and the reference run is again compared in the same way as in Figure 6a, with the difference that these CDFs were constructed for 1-year long periods that do not go from 1st January to 31st December of the target year (1982), but from 1st October 1981 to 30th September 1982 (Figure 6b), and from 1st July 1981 to 31st August 1982 (Figure 6c). Hence, for instance, the experiment that had 6 months of spin-up for a target period starting on 1st January provides 3 months of spin-up for a target period starting on 1st October and does not include any spin-up for a target period starting on 1st July. Accordingly, the labels on the x axis in panels b and c of Figures 6 refer to the actual spin-up times in each case. This new analysis shows that the optimum experimental setup (giving rise to the lowest number of land grid points with discordant CDFs) for a target period starting on 1st October consists of 3 months of spin-up, getting worse when longer spin-up periods are considered. The same applies for a target period starting on 1st July: the optimum is achieved after 3 months of spin-up. Therefore, the simulation that had been initialized in July is (again) the best choice for a target period starting on 1st October; but for a target period starting on 1st July, the best choice is the simulation that had been initialized in April. All cases together (target periods starting on 1st January, 1st October, and 1st July) highlight the importance of initializing the simulation in summer or non-winter months, which, depending on the start date of the target period, can reduce the optimum spin-up from 6 months (for usual target periods starting on 1st January) to three. The use of spin-up periods implying winter initialization dates should be avoided whenever possible.

4. Discussion and Conclusions

Dynamical downscaling is fundamentally a BC problem, in which the information to conduct the simulation is provided through the boundaries of a limited area domain. Nevertheless, RCMs require an IC that drives the evolution of the simulation during its first stages, in the so-called spin-up period. During this phase, the different components of the simulation are not yet completely in equilibrium, and therefore, it is generally regarded as an unreliable part of the simulation, which should be discarded. Right after initialization, the different components of the system tend to relax toward a solution in full equilibrium, albeit with different time scales. The atmospheric component, represented in variables such as T2 or PR, is generally the fastest, and spin-up periods as short as a few hours or days have proven to be sufficient for certain applications (De Elía et al., 2002, Skamarock, 2004, Zhong et al., 2007, Doblas-Reyes et al., 2013, Cholette et al., 2015, Gómez-Navarro et al., 2015). However, the evolution of the soil, represented in TSLB or SMOIS, is much slower and requires longer spin-up periods (Ajami et al., 2014; Christensen, 1999; Cosgrove et al., 2003; Seck et al., 2015). But even if simulated soil moisture perfectly matches observations, the atmospheric simulation may still be degraded by model errors (Hacker & Angevine, 2013) and internal variability (Gómez-Navarro et al., 2012). The heterogeneity in the time response of different climate variables makes the definition of a single suitable spin-up time, enough to consider the model in full equilibrium, a very difficult task.
This leads to a large variety of approaches within the RCM community (Gómez-Navarro et al., 2011, Vautard et al., 2013, Jerez et al., 2013 & 2018, Katragkou et al., 2015 & 2017, Prein et al., 2016).

To try to address this complex problem and provide general practical guidelines, we focus our analysis on variables that are of foremost interest in most applications of RCM simulations, that is, T2 and PR. Further, we put emphasis on soil variables, as they have the slowest time response of the system considered in most RCMs, thus contributing to the need for long spin-up periods. We base our analysis on a number of simulations with the WRF model driven by the MPI-ESM-LR historical run. All the simulations cover the same target period, the year 1982, arbitrarily selected from a rather recent period covered by large projects such as CORDEX. The only difference across simulations is the initialization time, so that the same year is simulated using spin-up periods that range from 1 year to 0 hr. An additional simulation with an exceptionally long 2-year spin-up time is used as reference.

Regarding the atmospheric variables, we find modest differences in the spatially averaged T2 series across simulations with 1 week or longer spin-up periods. In almost all cases, the 5–95th percentile range of the spatially averaged series of T2 differences between each simulation and the reference one during the simulated target period is within the range of the noisy fluctuations (natural variability) of the T2 series in the reference simulation. Nevertheless, and as expected, the amplitude of these differences becomes smallest after a relatively long spin-up period. In 50% of the land grid points the optimal spin-up is 6 months, which leads to a reduction of the RMSE with respect to the simulation with no spin-up of about 50%, especially in western Europe. The impact of the spin-up length on PR is much noisier, likely indicating a predominant role of model internal variability on this variable. In this case, the range span by the differences between each simulation and the reference one is similar to (so not clearly smaller than) the variability of this variable in the reference simulation. Still, these differences are generally narrower when longer spin-up periods are considered. We consistently find that 6 months of spin-up leads to the greatest number of land grid points with the strongest reduction in the RMSE, although the picture is patchier and only 30% of such grid points agree on this. For both variables, the CDFs constructed from all simulations (in particular, with spin-up periods of 1 week or more) agree with those from the reference run in more than 90% of the land grid points, with the best results always provided by the simulation with 6 months of spin-up. Thus, for atmospheric variables and within our experimental set, 6 months of spin-up provides better results than longer spin-up periods.

The picture becomes different when soil variables are considered. In the soil, longer spin-up periods have a more noticeable and positive effect, especially in the simulation of the soil moisture content. Yet equilibrium is far from being reached even after 1 year of spin-up, except for TSLB in the first layer. The CDFs of TSLB1 from the simulation with 6 months of spin-up agree with the reference CDFs in more than 90% of the land grid points. This percentage is reduced as we move deeper, dropping to 50% for TSLB4. The results for the SMOIS variable are worse. The percentage of land grid points with CDFs in good agreement with the reference ones drops to 25% for SMOIS1 and to <10% for SMOIS4 - which thus seems to be apparently uncoupled with the model performance regarding atmospheric variables - with these best results corresponding to the simulation with 9 months of spin-up.

However, seasonality, rather than period length alone, plays an important role regarding optimum spin-ups. For our target period starting on 1st January, an optimum spin-up period of 6 months implies an initialization date in summer, which appeared to be more suitable than longer spin-up periods implying initialization dates in winter. In fact, reinforcing this idea, for target periods starting on 1st October or 1st July, the optimum balance between length and seasonality is found just after 3 months of spin-up, since longer spin-up periods implied winter initialization dates. In order to understand why, we put the focus on the initialization of the snow cover, as it is an important player in the soil-atmosphere coupling and presents a strong annual cycle with a low peak in late summer and early autumn. Spin-up initial dates within this period of general absence of snow provided simulations of the snow cover in perfect agreement with the reference run. On the contrary, simulations initialized earlier were affected by an underestimation of the snow cover, which makes the simulations less reliable despite their longer spin-up times.

The inconsistencies between GCM initial conditions and RCM developed ones can be overcome in different ways. Here we performed additional experiments in which soil variables were initialized using RCM outputs from a different run covering a non-coincident, while representative, period. The results indicated a reduction to a half in the optimum length of the spin-up period, with the RCM relaxing faster from its own outputs.
(even if these mismatch the dates of the run being carried out), than from time-coincident GCM states. So if one already has an RCM simulation (representative of the period to be simulated), it may be a good idea to reuse it for initialization. One step further, the atmospheric outputs from this reused simulation could be first used as the atmospheric forcing to run offline the land-surface model alone for a decade or two, considering the relaxation times of soil variables (Angevine et al., 2014; Koster et al., 2010; Yang et al., 1995). When this offline simulation reaches the end-date of the available atmospheric forcing, it can be cycled back to the available start-date, and this procedure can be repeated as many times as necessary, until the deep soil fully spun-up, as in Christensen (1999). Then the fully spun-up land-surface model outputs could be used to initialize the land surface within the RCM. This type of cycling to spin up the most slowly evolving components of the Earth system is frequently used by global climate modelers, mostly for spinning up the ocean in this case (Séférian et al., 2016), but is not yet commonly used by regional climate modelers. The approach takes advantage from the fact that the land component requires a very small fraction of the computational resources consumed by an RCM run, and thus, it can be accomplished reasonably inexpensively. While there still could be some imbalances when the RCM is then run again in coupled mode for the target period (Angevine et al., 2014), this affordable method could overcome some of the extremely long spin-up times needed for deep soils (Yang et al., 1995). Although we have not tested it here, the results obtained suggest the need for future research in this line.

In the meantime, for the usual procedure of executing an RCM simulation, as a general rule, we find that 6 months of spin-up is a reasonable compromise that leads to optimal performance in several of the metrics we have explored in this manuscript. But seasonality is also an important factor, and we advise starting the simulation in the warm season whenever possible to minimize inconsistencies introduced through unrealistically snow covers. In any case, it should be borne in mind that, eventually, the spin-up period required to conduct an RCM simulation depends on the application that the simulation pursues. If the interest of the simulation relies on the analysis of T2 or even PR, very short spin-up periods of a few days to 1 week can be sufficient for most applications. However, soil relaxation times are larger. Therefore, when the aim of the analysis tries to address physical mechanisms that feed back through soil processes, the spin-up period required can be much longer. Even the longest 1-year spin-up period we have considered here is clearly insufficient to reach equilibrium in the soil moisture content. Nonetheless, this lack of soil equilibrium seems to play a minor role in the simulation of the atmospheric variables employed in most climate studies, such as T2 and PR, even on the right-hand side of their distributions. In any case, this conclusion should be considered for WRF applications over Euro-Cordex compliant domains or similar, while caution is required regarding different setups.

Authors Contribution

S. J. and J. M. L. - R. contributed equally to this study by designing and running the simulations, analyzing the results, and constructing the manuscript. P. J. - G. and J. P. M. conceived the study and helped in the analysis and interpretation of the results. M. T. participated in the analysis by providing tools and suggesting appropriate approaches. All authors helped to write and revise the manuscript and contributed significantly to its readability and quality.

References

Ajami, H., McCabe, M. F., Evans, J. P., & Stisen, S. (2014). Assessing the impact of model spin-up on surface water-groundwater interactions using an integrated hydrologic model. Water Resources Research, 50, 2636–2656.
Angevine, W. M., Bazile, E., Legain, D., & Pino, D. (2014). Land surface spinup for episodic modeling. Atmospheric Chemistry and Physics, 14(15), 8165–8172.
Anthes, R. A., Kuo, Y. H., Hsie, E. Y., Low-Nam, S., & Betge, T. W. (1989). Estimation of skill and uncertainty in regional numerical models. Quarterly Journal of the Royal Meteorological Society, 115(488), 763–806.
Chen, F., & Dudhia, J. (2001). Coupling an advanced land surface–hydrology model with the Penn State–NCAR MM5 modeling system. Part I: Model implementation and sensitivity. Monthly Weather Review, 129(4), 569–585.
Chen, F., Janjić, Z., & Mitchell, K. (1997). Impact of atmospheric surface-layer parameterizations in the new land-surface scheme of the NCEP mesoscaleEta model. Boundary-Layer Meteorology, 85(3), 391–421.
Chollette, M., Laprise, R., & Thériault, J. (2015). Perspectives for very high-resolution climate simulations with nested models: Illustration of potential in simulating St. Lawrence River Valley channeling winds with the fifth-generation Canadian regional climate model. Climate, 3(2), 283–307.
Christensen, O. B. (1999). Relaxation of soil variables in a regional climate model. Tellus A, 51(5), 674–685.
Day, J. J., Tietze, S., & Hawkins, E. (2014). Pan-Arctic and regional sea ice predictability: Initialization month dependence. Journal of Climate, 27(12), 4371-4390.

De Elia, R., Laprise, R., & Denis, B. (2002). Forecasting skill limits of nested, limited-area models: A perfect-model approach. Monthly Weather Review, 130(8), 2006-2023.

Denis, B., Laprise, R., Caya, D., & Côté, J. (2002). Downscaling ability of one-way nested regional climate models: the Big-Brother Experiment. Climate Dynamics, 18(8), 627-646.

Doblas-Reyes, F. R., Andrus-Burillo, I., Chikamoto, Y., García-Serrano, J., Guemas, V., Kimoto, M., et al. (2013). Initialized near-term regional climate change prediction. Nature Communications, 4, 1715.

Giorgetta, M., Jungclaus, J., Reich, C., Legutke, S., Brovkin, V., Crueger, T., & Stevens, B. (2012). Forcing data for regional climate models based on the MPI-ESM-LR model of the Max Planck Institute for Meteorology (MPI-M): The CMIP5 historical experiment. World Data Center for Climate (WDCD) at DKRZ. World Data Center for Climate (WDCD) at DKRZ, 10.

Giorgi, F., & Mearns, L. O. (1999). Introduction to special section: Regional climate modeling revisited. Journal of Geophysical Research - Atmospheres, 104(D6), 6335-6352.

Gomez, J. J., Raible, C., & Dierer, S. (2015). Sensitivity of the WRF model to PBL parametrizations and nesting techniques: Evaluation of wind storms over complex terrain. Geoscientific model development (GMD), 8(10), 3349-3363.

Gómez-Navarro, J. J., Montávez, J. P., Jerez, S., Jiménez-Guerrero, P., Lorente-Plazas, R., González-Rouco, J. F., & Zorita, E. (2011). A regional climate simulation over the Iberian Peninsula for the last millennium. Climate of the Past, 7(2), 451.

Gómez-Navarro, J. J., Montávez, J. P., Jiménez-Guerrero, P., Jerez, S., Lorente-Plazas, R., González-Rouco, J. F., & Zorita, E. (2012). Internal and external variability in regional simulations of the Iberian Peninsula climate over the last millennium. Climate of the Past, 8(1), 25–36.

Grell, G. A. (1993). Prognostic evaluation of assumptions used by cumulus parameterizations. Monthly Weather Review, 121(3), 764-787.

Grell, G. A., & Dévényi, D. (2002). A generalized approach to parameterizing convection combining ensemble and data assimilation techniques. Geophysical Research Letters, 29(14), 36–31.

Hacker, J. P., & Angevine, W. M. (2013). Ensemble data assimilation to characterize surface-layer errors in numerical weather prediction models. Monthly Weather Review, 141(6), 1804-1821.

Hong, S. Y., Noh, Y., & Duddia, J. (2006). A new vertical diffusion package with an explicit treatment of entrainment processes. Monthly Weather Review, 134(9), 2318–2341.

Iacono, M. J., Delamere, J. S., Shephard, M. W., Clough, S. A., & Collins, W. D. (2008). Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models. Journal of Geophysical Research – Atmospheres, 113(D13).

Jacob, D., Petersen, J., Eggert, B., Aliax, C., Christensen, O. B., Bousser, L. M., et al. (2014). EURO-CORDEX: New high-resolution climate change projections for European impact research. Regional Environmental Change, 14(2), 563–578.

Jacob, D., & Podzun, R. (1997). Sensitivity studies with the regional climate model REMO. Meteorology and Atmospheric Physics, 63(1-2), 119–129.

Jerez, S., López-Romero, J. M., Turco, M., Jiménez-Guerrero, P., Vautard, R., & Montávez, J. P. (2018). Impact of evolving greenhouse gas forcing on the warming signal in regional climate model experiments. Nature Communications, 9(1), 1304. https://doi.org/10.1038/s41467-018-03527-y

Jerez, S., Montavez, J. P., & Giménez, D. (2009, May). Optimizing the execution of a parallel meteorology simulation code. In 2009 IEEE International Symposium on Parallel & Distributed Processing (pp. 1-6). IEEE.

Jerez, S., Trigo, R. M., Vicente-Serrano, S. M., Pozo-Vázquez, D., Lorente-Plazas, R., Lorenzo-Lacruz, J., Lorenzo-Lacruz, J., Santos-Alamillos, F., & Montañez, J. P. (2013). The impact of the North Atlantic Oscillation on renewable energy resources in south-western Europe. Journal of Applied Meteorology and Climatology, 52(10), 2204–2225.

Katragkou, E., García Diez, M., Vautard, R., Sobolowski, S. F., Zanis, P., Alexiadri, G., et al. (2015). Regional climate hindcast simulations within EURO-CORDEX: Evaluation of a WRF multi-physics ensemble. Geoscientific Model Development, 8(3), 603–618.

Katragkou, E., Gkotovou, I., Kartsios, S., Pavlidis, V., Tsagardis, K., Traill, M., et al. (2017). A0TH regional climate model contributions to EURO-CORDEX, In Perspectives on Atmospheric Sciences (pp. 741–746), (J. Chap: Springer.

Koster, R. D., Mahanama, S. P. P., Yamada, T. J., Balsamo, G., Berg, A. A., Boisserie, M., et al. (2010). Contribution of land surface initialization to subseasonal forecast skill: First results from a multi-model experiment. Geophysical Research Letters, 37(2).

Leduc, M., & Laprise, R. (2009). Regional climate model sensitivity to domain size. Climate Dynamics, 32(6), 833–854.

Lillicor, H. W. (1987). On the Kolmogorov-Smirnov test for normality with mean and variance unknown. Journal of the American Statistical Association, 82(38), 399–402.

Lin, Y. L., Farley, R. D., & Orville, H. D. (1983). Bulk parameterization of the snow microphysics. Monthly Weather Review, 111(1), 1005.

Matte, D., Laprise, R., Thériault, J. M., & Lucas-Picher, P. (2017). Spatial spin-up of fine scales in a regional climate model simulation driven by low-resolution boundary conditions. Climate Dynamics, 49(1-2), 563-574.

Milelli, M., Turco, M., & Oberto, E. (2010). Screen-level non-GTS data assimilation in a limited-area mesoscale model. Natural Hazards and Earth System Sciences, 10(6), 1129-1149.

Prein, A. F., Gobiet, A., Truethz, H., Keuler, K., Goergen, K., Teichmann, C., et al. (2016). Precipitation in the EURO-CORDEX 0.11° and 0.44° high-resolution, high benefits? Climate Dynamics, 46(1-2), 383–412.

Rummukainen, M. (2010). State-of-the-art with regional climate models. Wiley Interdisciplinary Reviews: Climate Change, 1(1), 82–96.

Schewe, J., Gosling, S. N., Reyer, C., Zhao, F., Ciais, P., Elliott, J., et al. (2019). State of the-art global models underestimate impacts from climate extremes. Nature Communications, 10(1), 1006.

Seck, A., Welby, C., & Maxwell, R. M. (2015). Spin-up behavior and effects of initial conditions for an integrated hydrologic model. Water Resources Research, 51(4), 2188-2210.

Séférian, R., Gehlen, M., Bopp, L., Resplandy, L., Orr, J. C., Martin, O., et al. (2016). Inconsistent strategies to spin up models in CMIP5: implications for ocean biogeochemical model performance assessment. Geoscientific Model Development, 9, 1827–1851.

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.

Skamarock, W. C. (2004). Evaluating mesoscale NWPF models using kinetic energy spectra. Monthly Weather Review, 132(12), 3019–3032.
Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D. M., Wang, W., & Powers, J. G. (2008). A description of the Advanced Research WRF version 3. NCAR Technical note-475+ STR.

Tebaldi, C., & Knutti, R. (2007). The use of the multi-model ensemble in probabilistic climate projections. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 365(1857), 2053–2075.

Tewari, M., Chen, F., Wang, W., Dudhia, J., LeMone, M. A., Mitchell, K., et al. (2004, January). Implementation and verification of the unified NOAH land surface model in the WRF model. In 20th conference on weather analysis and forecasting/16th conference on numerical weather prediction (Vol. 1115). Seattle, WA: American Meteorological Society.

Turco, M., Sanna, A., Herrera, S., Llasat, M. C., & Gutiérrez, J. M. (2013). Large biases and inconsistent climate change signals in ENSEMBLES regional projections. Climatic Change, 120(4), 859–869.

Vautard, R., Gobiet, A., Jacob, D., Belda, M., Colette, A., Déqué, M., et al. (2013). The simulation of European heat waves from an ensemble of regional climate models within the EURO-CORDEX project. Climate Dynamics, 41(9-10), 2555–2575.

Yang, Y., Uddstrom, M., & Duncan, M. (2011). Effects of short spin-up periods on soil moisture simulation and the causes over New Zealand. Journal of Geophysical Research: Atmospheres, 116(D24).

Yang, Z. L., Dickinson, R. E., Henderson-Sellers, A., & Pitman, A. J. (1995). Preliminary study of spin-up processes in land surface models with the first stage data of Project for Intercomparison of Land Surface Parameterization Schemes Phase 1 (a). Journal of Geophysical Research: Atmospheres, 100(D8), 16,553–16,578.

Zhong, Z., Yijia, H. U., Jinzhong, M. I. N., & Honglei, X. U. (2007). Numerical experiments on the spin-up time for seasonal-scale regional climate modeling. Journal of Meteorological Research, 21(4), 409–419.