Regional Decadal Climate Predictions Using an Ensemble of WRF Parameterizations Driven by the MIROC5 GCM

EHUD STROBACH

Department of Solar Energy and Environmental Physics, Blaustein Institutes for Desert Research, Ben-Gurion University of the Negev, Sde Boker, Israel, and Earth System Science Interdisciplinary Center, College of Computer, Mathematical, and Natural Sciences, University of Maryland, College Park, and Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, Maryland

GOLAN BEL

Department of Solar Energy and Environmental Physics, Blaustein Institutes for Desert Research, Ben-Gurion University of the Negev, Sde Boker, and Department of Physics, Ben-Gurion University of the Negev, Beer Sheva, Israel, and Center for Nonlinear Studies, Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico

(Manuscript received 16 February 2018, in final form 24 December 2018)

ABSTRACT

Regional climate models (RCMs) are expected to provide better representations of the climate dynamics because of their higher spatial resolutions. Here, we generated an ensemble of decadal (2006–36) RCM predictions for the area of Israel, which spans a considerable climatic gradient and comprises complex terrain. We used the WRF Model forced by the MIROC5 global climate model (GCM). The ensemble was generated by choosing different combinations of radiation, microphysics, surface layer, and planetary boundary layer parameterizations. The simulation results were compared with meteorological station data for the first simulated decade. For the minimum surface temperature, all the RCM configurations performed better than the driving GCM, while for the maximum surface temperature, only three out of eight configurations improved the predictions. The RCM configurations had higher errors in predicting the precipitation, but four configurations had comparable errors to the GCM. For the next two decades, the ensemble average predicts an increase of $0.5^{+0.8}_{-0.4}$ and $0.4^{+0.8}_{-0.1}$ C decade$^{-1}$ for the average daily minimum and maximum surface temperatures, respectively. No significant change is predicted in the precipitation. We found that all the parameterizations affect the predictions of the surface temperatures and precipitation [e.g., the CAM radiation scheme simulates colder temperatures than the RRTM for GCMs (RRTMG)] but the PBL and surface layer scheme has the largest effect on the errors. Spectral nudging was found to have a considerable effect on the deviations of the precipitation predicted by the WRF configurations from the predictions of the GCM and a much smaller effect on the surface temperature predictions.

1. Introduction

Global decadal climate simulations are designed to provide information on the dynamics of the climate system over time scales of one to three decades (Smith et al. 2007; Keenlyside et al. 2008; Pohlmann et al. 2009). On these time scales, it is not only the response of the climate system to changes in the atmospheric composition and changes in the land surface that are of interest but also the natural variability of the climate system (Meehl et al. 2009, 2014). Global decadal prediction experiments are part of phase 5 of the Coupled Model Intercomparison Project (CMIP5; Taylor et al. 2012) and have shown an improved prediction skill in several regions around the globe compared with the predictions of the CMIP5’s long-term climate projection experiments (Doblas-Reyes et al. 2013; Müller et al. 2012; Pohlmann et al. 2013; Müller et al. 2014). A demand for decadal climate predictions also exists (Cane 2010).

To provide meaningful information, climate model predictions must be accompanied by uncertainties. A common method to estimate these uncertainties involves using an ensemble of climate predictions (Strobach and Bel 2015, 2016). Ideally, an ensemble
should account for all the sources of uncertainties. In decadal climate predictions, uncertainties can be divided into two main sources: internal variability and model variability (Hawkins and Sutton 2009; Meehl et al. 2009). Internal variability is due to the uncertainty in the initial conditions and initialization methods, and model variability is due to the differences between different climate models. It has been shown that the relative contribution of model variability is larger than that of internal variability in global decadal climate predictions of surface temperature (Stroobach and Bel 2017a). Scenario variability due to projected changes in the atmospheric composition has been shown to be of less importance in decadal climate predictions (Hawkins and Sutton 2009; Meehl et al. 2009).

Global simulations are limited by their spatial resolution; therefore, high-resolution regional climate simulations that are driven by lower-resolution global simulations are commonly used (Giorgi and Mearns 1999). In regional climate predictions, there is an additional source of variability due to the differences between simulations with different driving global models. These driving models provide the boundary conditions and sometimes also the large-scale-spatial dynamics (Tebaldi and Knutti 2007). Because of the large computational resources needed to generate an ideal ensemble, there are coordinated efforts to produce both global (Taylor et al. 2012) and regional (Giorgi et al. 2009; Giorgi and Gutowski 2015) ensembles of climate predictions. Nevertheless, one should bear in mind that accounting for all sources of model uncertainties still does not guarantee sampling of the entire probability space.

In this study, we focused on the area of Israel. This area is interesting for regional climate predictions because of its relatively small size ($\approx 22,000$ km$^2$), complex terrain (ranging from 400 m below sea level to 2000 m above sea level), and considerable climatic gradients (annual average precipitation amounts vary from less than 50 mm yr$^{-1}$ in the south to more than 1000 mm yr$^{-1}$ in the north) (Goldreich 2003). The region lies between the subtropical and the temperate zones, and therefore, a change of tens of kilometers in the expansion of the Hadley cell may affect the climate of this area considerably by expanding the desert region northward several tens of kilometers within a few decades (Johansson and Fu 2009). The area is also subject to water shortages, and changes in precipitation, even small ones, can have a large impact on the amount of water available for agriculture.

Long-term climate predictions for the area of Israel, based on global climate models, suggest that by the mid-twenty-first century, the temperature is expected to increase by $\approx 1.0^\circ$–$2.25^\circ$C (Evans 2009; Lelieveld et al. 2016), and the annual precipitation is expected to decrease by $5$–$25$ mm yr$^{-1}$. Regional climate models for this area predict a decrease in precipitation during the twenty-first century of 5%–25% relative to the 1961–90 period (Black 2009; Hemming et al. 2010; Chenoweth et al. 2011). One of the studies found that in the southern part of Israel, the precipitation in the period from 2021 to 2050 is actually expected to increase by 5%–25% relative to the 1961–90 period (Chenoweth et al. 2011). A study based on statistical downscaling showed that the precipitation in the area is expected to decrease by $\approx 10$ mm yr$^{-1}$ in the 2020–30 period and by $\approx 20$ mm yr$^{-1}$ in the 2040–50 period, relative to 2000–09.

The regional climate model (RCM) used in this study is the Weather Research and Forecasting (WRF) Model, version 3.6.1 (Skamarock et al. 2008), which is able to use different parameterization schemes for different components of the model. We used this capability to generate an eight-member ensemble; therefore, our ensemble represents only a subspace of the parameterization variability and not the other sources mentioned earlier. This kind of variability does not fall within the defined categories of internal, model, and boundary condition variability mentioned above. Here, we define it as intramodel variability.

The advantage of such an ensemble is that it allows us to study the effects of different parameterizations and identify their impacts on the different variables. The differences between different parameterizations have attracted much attention. It has been shown that non-local planetary boundary layer (PBL) schemes reduce the bias in the forecast of surface temperature and air moisture (Hu et al. 2010) and that surface temperature, wind, and relative humidity are sensitive to the PBL scheme (Banks et al. 2016). It has also been shown that surface temperature (minimum and maximum daily temperatures) and precipitation are sensitive to the radiation, PBL, cumulus physics, and microphysics schemes (Evans et al. 2012; Zittis et al. 2014). These effects suggest that it is indeed useful and informative to study an ensemble of different WRF parameterizations.

The main objective of this paper is to investigate the near-term dynamically downscaled climate predictions from a decadal climate simulation—that is, to use a driving model that was initialized with interpolated observed conditions and to simulate continuously the regional climate over three decades. A secondary objective is to test the reliability of a relatively small ensemble that only accounts for different parameterizations. We also investigate the physical mechanisms that control the differences between ensemble members. To capture the effects of the complex topography, we used a
2. Methods

a. WRF setup

We used the WRF Model to simulate the climate of a region in the eastern Mediterranean during the period from 2006 to 2036. The WRF regional model was driven by the MIROC5 decadal experiment (Mochizuki et al. 2012; Tatebe et al. 2012; Watanabe et al. 2010) (the results of the GCM provided the boundary conditions for the RCM). The MIROC5 was chosen because of the availability of all the variables required for the WRF, including the SST, and the fact that previous works showed that it has a skill in simulating the climate of the region considered here with relatively small biases (Kim et al. 2012; Chikamoto et al. 2013; Flato et al. 2013). Moreover, it was found to be suitable for dynamical downscaling in the region simulated in this work (McSweeney et al. 2015). The atmospheric grid of the MIROC5 is a T85 spectral truncation, which is equivalent to \(1.4°\) resolution, and it has 40 vertical levels (Mochizuki et al. 2012; Tatebe et al. 2012; Watanabe et al. 2010). The temporal resolution of the driving data was 6 h for the atmospheric variables (3D temperature, wind, and humidity; surface pressure and skin temperature) and the SST, and one month interpolated to 6 h for the soil variables (four levels of soil temperature and moisture at 0–5, 5–25, 25–100, and 100–200 cm).

The WRF simulations were continuously run (one initialization). The choice of continuous simulations of the whole period reduces the effects of the initialization method and allows the model to better converge to its own climatology (the dynamic boundary conditions still affect the dynamics). The large-scale atmospheric dynamics of the driving model was imposed on the experimental domain using spectral nudging (Waldron et al. 1996; von Storch et al. 2000). The spectral nudging was applied only to the coarser grid (of the three nested grids), and it was limited to wavenumbers three and below (in both lateral directions); in addition, it was not applied in the PBL and in the lowest 10 vertical levels of the model. This spectral nudging setup is in agreement with previous works that showed that wavelengths of 2000 km and longer contain most errors in limited domain simulations and should be nudged in order to correctly simulate the synoptic scales (Vukicevic and Errico 1990; Miguez-Macho et al. 2004) (here, we nudged wavelengths of 1800 km and longer). The nudged variables include the 3D temperature, wind, specific humidity, and geopotential fields. The relaxation time was set to \(\approx 50\) min (nudging coefficient of 0.0003 s\(^{-1}\)), which is the WRF default value and was shown to improve the predictions of precipitation and surface temperature (it was shown that at least in some regions, the results are not sensitive to the precise value of the nudging coefficient) (Miguez-Macho et al. 2004; Glisan et al. 2013; Spero et al. 2014). The nudging was limited to levels above the PBL and above the 10th vertical level of the model. These settings allowed the model to evolve to its own state in the lowest part of the atmosphere, which is directly influenced by the high-resolution surface.

The configuration of the WRF Model included three nested grids (domains) of 54 km \(\times\) 54 km, 18 km \(\times\) 18 km, and 6 km \(\times\) 6 km spatial resolutions and 40 vertical levels (similar to the vertical resolution of the driving data). The three domains were centered on Israel, and each of them included \(\approx 100 \times 100\) grid cells (100 \(\times\) 100 for the outer domain, 94 \(\times\) 97 for the middle domain, and 94 \(\times\) 118 for the inner domain). Figure 1 depicts a map of the domains. The model output for several surface variables (temperature, precipitation, and wind components) was sampled every 10 min to capture properties such as minimum and maximum daily temperatures appropriately.

We tested eight different parameterization configurations of the model, consisting of all the possible combinations of two microphysics, two radiation, and two PBL/surface layer schemes. The different parameterizations are detailed below.

1) Microphysics

WRF single-moment 6-class (WSM6) (Hong and Lim 2006): Based on Lin et al. (1983) with changes made to the ice-phase according to Hong et al. (2004) and a new method for representing the mixed-phase particle fall speeds for snow and graupel particles according to Dudhia et al. (2008).

Goddard Cumulus Ensemble (GCE) (Tao and Simpson 1993): Also based on Lin et al. (1983) with additional processes from Rutledge and Hobbs (1984) and several other modifications.

The GCE scheme was shown to produce higher cloudiness (Zittis et al. 2014).

2) Longwave and shortwave radiation

NCAR Community Atmosphere Model (CAM) (Collins et al. 2004): A spectral band scheme (19 bands for shortwave radiation and 8 bands for longwave radiation). It uses the \(\delta\)-Eddington approximation to calculate radiative scattering and absorption, and it accounts for the absorption of \(\text{CO}_2\), \(\text{CH}_4\), \(\text{N}_2\text{O}\), \(\text{O}_2\), \(\text{O}_3\), \(\text{H}_2\text{O}\), CFC-11, and CFC-12.
The Rapid Radiative Transfer Model for GCMs (RRTMG) (Iacono et al. 2008): A spectral band scheme (14 bands for shortwave radiation and 16 bands for longwave radiation). It uses the correlated $k$-distribution method for the calculation of radiative transfer, and it accounts for the absorption by the atmospheric gases listed for the CAM model and also for the absorption by CFC-22, N$_2$, and CCl$_4$.

Both the CAM and the RRTMG account for sub-grid clouds by considering the cloud cover fraction.

3) Planetary boundary layer/surface layer

Mellor–Yamada–Janjic’ (MYJ) (Janjic’ 1994) PBL with an Eta similarity theory (Eta) surface layer: A one-dimensional prognostic turbulent kinetic energy scheme with a local vertical mixing PBL scheme. The surface layer scheme is based on the similarity theory with a viscous sublayer and standard similarity functions. This scheme uses the 1.5 order (i.e., first-order closure for some variables and second-order closure for others) for the closure of the turbulence moments. Because of the local mixing applied in this scheme, the simulated PBL tends to be cooler and moister relative to observations (Cohen et al. 2015).

Yonsei University (YSU) PBL (Hong et al. 2006) with a similarity theory (MM5) surface layer: A nonlocal PBL scheme with explicit parameterization of the entrainment flux at the top of the PBL. The surface layer scheme is also based on the similarity theory with a Zilitinkevich thermal roughness length (Sportisse 2002) and standard similarity functions. The nonlocal mixing causes the PBL to be drier and warmer, which is more consistent with observations (Cohen et al. 2015). This scheme uses the first-order closure for the turbulence moments.

All the WRF configurations used the Noah land surface model (NLSM) (Chen and Dudhia 2001) and the Kain–Fritsch (KF) cumulus parameterization (Kain 2004). The NLSM is a four-layer model for the soil temperature and moisture, and it accounts for the root zone, evapotranspiration, and soil drainage and runoff. The soil texture, vegetation category, and monthly variation of the vegetation cover fraction are also considered in the model. The KF scheme is a simple cloud model with moist updrafts and downdrafts, including the effects of detrainment, entrainment, and relatively simple microphysics. The cumulus clouds were parameterized (and not explicitly resolved) in the three simulated domains including the finer-resolution domain. Table 1 summarizes the configurations of the model used in this study. All the ensemble members include the assimilation of the greenhouse gas mixing ratio [Climate-WRF (CLWRF); Fita et al. 2010; RCP4.5] similar to the scenario used in the driving data. On decadal time scales, it was already mentioned that the scenario variability is negligible (Hawkings and Sutton 2009; Meehl et al. 2009); therefore, the choice of a specific RCP is not expected to significantly affect the results (this is probably the reason why CMIP5 decadal climate predictions were not performed using different scenarios).

b. IMS data

The relatively high resolution of the WRF inner domain motivated us to compare its output with in situ observations of the Israel Meteorological Service (IMS). The IMS dataset is available online (https://ims.data.gov.il/, in Hebrew). It includes records of various meteorological observations (e.g., surface temperature, pressure, wind, etc.) from ≈300 in situ stations in Israel (the earliest records are from 1964) and precipitation data from ≈1800 stations (the earliest records are from the second half of the nineteenth century). The IMS data used in this study include monthly precipitation amounts and maximum and minimum daily temperatures. The IMS data for the first 10 years of the WRF simulations, 2006–16, were used in the comparison with the simulated climate. For the temperature, we had, on average, data from ≈57 stations per year (the actual number of stations varied from 40 to 64). For the precipitation, the
average number of IMS stations with available data was much larger than the number of stations with data for the temperature. On average, we had 237 stations with available data per year, and the actual number ranged from 215 to 252. For each of the 10 years, we only used the data of a station if they were continuous throughout the year (for the precipitation, if the data were continuous throughout the rainy season, from 1 October to 30 April). (Figure A1 in the appendix shows the location of the stations whose data were used in this study.) The comparison with in situ observations can provide an idea of the goodness of the simulations (quality of the model for simulating this specific region, etc.). It is important to note that unlike many other studies (e.g., Evans and McCabe 2013; Andrys et al. 2016), we did not use gridded interpolated measurements but rather compared each station’s data with the corresponding high-resolution WRF grid cell [for the MIROC5, the coarse-resolution driving GCM, we interpolated the coarse grid results into a grid of similar resolution to the WRF configuration using nearest destination mapping (NCAR 2011), and the comparison was done with these interpolated results].

c. Reliability

In a reliable ideal ensemble, assuming that the prediction is the average of the ensemble and that the prediction is not biased, the root-mean-square error (RMSE) is equal to the standard deviation (STD) (Palmer et al. 2006). An ensemble whose spread (quantified by the STD) is smaller than its error (quantified by the RMSE) is considered overconfident, and an ensemble with a spread larger than the error is underconfident (Leutbecher and Palmer 2008; Weigel et al. 2008; Weisheimer et al. 2009; van Oldenborgh et al. 2013). In addition to assessing the simulation’s performance using the mean absolute error (MAE), we also quantified the reliability of the small ensemble using the ratio between the STD and the RMSE. The reliability measure for an ensemble that only spans the spread of different parameterization methods and does not account for other sources of variability (such as different models, different initialization times and methods, etc.) is important for the proper assessment of the ensemble predictions. A common practice in analysis of climate predictions is to correct the bias of the simulated climate by subtracting the mean deviation of the simulated climate from observation or re-analysis data during a reference period (bias correction) (Goddard et al. 2013; Meehl et al. 2014; Hawkins and Sutton 2016; Strobach and Bel 2017a). Moreover, it is known that the choice of the reference period may affect the conclusions regarding the anomalies (Hawkins and Sutton 2016). Here, we applied three different types of bias correction. The first bias correction was done by subtracting the spatially (over all the grid cells of the high-resolution grid for which IMS data are available) and temporally (over the reference period of 2006–16) averaged deviation of each configuration from the observations (IMS data). In what follows, we denote this bias correction as “bias corrected.” The second type of bias correction (denoted hereafter as “MIROC bias corrected”) was done by removing from all the ensemble members the spatial and temporal average of the deviation of the driving MIROC5 model from the observation during the same reference period (2006–16). The last type of bias correction (denoted hereafter as “ensemble bias corrected”) was done by subtracting from all the ensemble members the spatially and temporally averaged deviation of the ensemble average from the observations during the same reference period. While the first bias correction modifies both the RMSE and the STD, the other two only modify the RMSE, while the STD remains unchanged.

3. Differences between WRF configurations and comparison with observations

The results presented below for regional climate WRF simulations are based on the inner, high-resolution (6 km × 6 km) domain. Results for the driving GCM (MIROC5) are based on interpolation of its coarse

<table>
<thead>
<tr>
<th>Member</th>
<th>Microphysics</th>
<th>Radiation</th>
<th>Surface layer</th>
<th>Planetary boundary layer</th>
<th>Land surface</th>
<th>Cumulus</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>WSM6</td>
<td>CAM</td>
<td>MM5</td>
<td>YSU</td>
<td>NLSM</td>
<td>KF</td>
</tr>
<tr>
<td>b</td>
<td>WSM6</td>
<td>RRTMG</td>
<td>MM5</td>
<td>YSU</td>
<td>NLSM</td>
<td>KF</td>
</tr>
<tr>
<td>c</td>
<td>GCE</td>
<td>CAM</td>
<td>MM5</td>
<td>YSU</td>
<td>NLSM</td>
<td>KF</td>
</tr>
<tr>
<td>d</td>
<td>GCE</td>
<td>RRTMG</td>
<td>MM5</td>
<td>YSU</td>
<td>NLSM</td>
<td>KF</td>
</tr>
<tr>
<td>e</td>
<td>WSM6</td>
<td>CAM</td>
<td>Eta</td>
<td>MYJ</td>
<td>NLSM</td>
<td>KF</td>
</tr>
<tr>
<td>f</td>
<td>WSM6</td>
<td>RRTMG</td>
<td>Eta</td>
<td>MYJ</td>
<td>NLSM</td>
<td>KF</td>
</tr>
<tr>
<td>g</td>
<td>GCE</td>
<td>CAM</td>
<td>Eta</td>
<td>MYJ</td>
<td>NLSM</td>
<td>KF</td>
</tr>
<tr>
<td>h</td>
<td>GCE</td>
<td>RRTMG</td>
<td>Eta</td>
<td>MYJ</td>
<td>NLSM</td>
<td>KF</td>
</tr>
</tbody>
</table>
grid to the WRF’s high-resolution grid (unless otherwise specified).

a. Surface temperature

1) Mean temperature

The performance of climate models can be tested by their ability to reproduce past conditions with the correct spatial and temporal distributions. This fact may not be trivial in high-resolution regional experiments, in particular, when spectral nudging, which may smooth the RCM predictions, is applied. Figure 2 shows the temporal (10 years from 2006 to 2016) and ensemble (eight parameterization configurations described above) average of the mean surface temperature. Overall, the temporal and ensemble average of the simulations yielded similar patterns to those derived from extrapolated observations in the past (Goldreich 2003). The simulated pattern shows that the surface temperature is warmer in the south and decreases with the latitude and warmer in the east along the low-lying areas of the Syro-African depression and cooler along the Mediterranean coast. The lower temperatures along the high elevation regions—Mount Negev, Judea and Samaria, Mount Carmel, and Mount Hermon—are also captured. The IMS dataset that we used does not include information regarding the mean daily temperature, and therefore we could not quantify the errors in simulating this variable. There are several considerable differences between the different parameterizations.

These differences are illustrated in Fig. 3, which shows the surface temperature anomaly of the different realizations with respect to the ensemble average. The four parameterizations that include the CAM radiation scheme (Figs. 3a,c,e,g) predict colder surface temperatures than the four parameterizations that include the RRTMG radiation scheme (Figs. 3b,d,f,h) and, in a large part of the region, by more than 1°C (see also the CAM-RRTMG panel in Fig. 3, bottom row). These results are in agreement with Zittis and Hadjinicolaou (2017), who showed that, despite the large differences, each radiation scheme has its own advantages in specific regions and specific seasons. In the desert parts of the Middle East and North Africa (MENA) region, it was found that the WRF CAM radiation scheme underestimates upward and downward longwave radiation compared to an observationally based product. The overall effect of the longwave radiation underestimation was a decrease in surface temperature compared to the RRTMG. A WRF simulation using a CAM shortwave radiation scheme and an RRTMG longwave radiation scheme did not produce a large negative temperature bias (Zittis and Hadjinicolaou 2017).

The predicted surface temperature is mostly colder by \(\approx 0^\circ - 0.5^\circ\) for the CAM parameterizations with the WSM6 microphysics scheme (Figs. 3a,e) and by \(\approx 0.5^\circ - 1.0^\circ\) for the CAM parameterizations with the GCE microphysics scheme (Figs. 3c,g) relative to the RRTMG radiation scheme parameterizations (Figs. 3b,d,f,h). The differences between the four RRTMG parameterizations (Figs. 3b,d,f,h) are less pronounced, but the combination of RRTMG with
FIG. 3. Temporal (2006–16) average of the surface temperature anomaly (with respect to the ensemble average) for (a)–(h) the eight parameterizations. (bottom) The difference between the average predictions made by the two radiation, the two microphysics, and the two PBL/surface layer parameterizations.

average temperature anomaly [°C]

-1.6 -1.4 -1.2 -1 -0.8 -0.6 -0.4 -0.2 0 0.2 0.4 0.6 0.8 1 1.2 1.4 1.6
WSM6 (Figs. 3b,f) leads to higher surface temperatures relative to RRTMG with GCE microphysics (Figs. 3d,h), and the combination of RRTMG with MM5/YSU (Figs. 3b,d) leads to higher surface temperatures relative to RRTMG with Eta/MYJ (Figs. 3f,h), particularly at low elevations.

To better visualize the differences between the two alternative schemes for the three factors (radiation, microphysics, and PBL/surface layer), we included in Fig. 3 panels showing the differences between the average predictions made by the two radiation, the two microphysics, and the two PBL/surface layer parameterizations. The average for each parameterization was taken over the four configurations with the same radiation or microphysics or PBL/surface layer scheme. These panels clearly show that the CAM radiation scheme yields lower surface temperatures relative to the RRTMG parameterization (CAM-RRTMG panel); the WSM6 microphysics scheme yields surface temperatures that are \( \approx 0^\circ \text{C} \) higher than the GCE almost everywhere (WSM6-GCE panel); for the PBL/surface layer schemes, there is no one scheme that is warmer/colder throughout the simulated region (YSU/MM5-MYJ/Eta panel). However, the YSU/MM5 parameterization seems to predict higher surface temperatures (relative to the MYJ/Eta parameterization) along the coast and in the southern parts of the simulated region and colder surface temperatures in the northern parts of the region and in regions with larger elevation gradients (more complex topography).

2) MINIMUM AND MAXIMUM DAILY TEMPERATURES

To evaluate the performance of the WRF simulations, we calculated the climatology of the minimum daily temperature based on the first 10 years of the simulations (2006–16) in three different locations of meteorological stations with available IMS data [Figs. A2a(1)–a(3)]. The stations, Elat, Tel Aviv, and Haifa, are in the southern, central, and northern regions of the simulated domain, respectively (their locations are denoted in Fig. A1). In Haifa [Fig. A2a(1)], the MIROC5 and all the WRF configurations simulated higher minimum daily temperatures for most of the year with the CAM configurations (a, c, e, and g) being closer to the observations. The magnitude of the annual cycle is captured reasonably well, but the values of the simulated minimum temperature are up to \( 4^\circ \text{C} \) higher than those observed. In Tel Aviv [Fig. A2a(2)], the MIROC5 simulated lower minimum temperatures than the observed ones during the winter and higher (than the observed ones) minimum temperatures during the summer. The CAM configurations showed the opposite behavior but were overall closer to the observations, and the RRTMG configurations predicted higher (than the observed ones) minimum temperatures for most of the year. In Elat [Fig. A2a(3)], which is located in the driest region of the simulated region, the MIROC5 and the RRTMG configurations simulated higher minimum temperatures than the observations (with the WRF configurations being closer to the observations) while the CAM configurations simulated minimum temperatures that are close to the observed ones (deviations of up to \( \approx 1^\circ \text{C} \)). For the annual mean of daily minimum temperature, Fig. 4a shows that three of the four RRTMG configurations (b, d, and h) had a lower MAE than the CAM configurations (a, c, and g). However, regionally, in southern Israel and the northern part of central Israel, the CAM had a lower MAE (not shown in the figure). Figure 4a also shows that the four Eta/MYJ configurations (e–h) had a smaller MAE than the MM5/YSU configurations (a–d), demonstrating an advantage of the local PBL scheme.

The performance (in terms of the MAE) of each of the WRF configurations was better than the performance of
the driving model MIROC5. We also tested the significance of the improvement by calculating the difference between the absolute errors of the eight ensemble members and the absolute error of the MIROC5. The statistical significance of the improvement achieved by the ensemble members was tested by introducing the null hypothesis that the temporal distribution of the difference between absolute errors of the ensemble member and the MIROC5 is symmetric around 0. We found that the improvement of the WRF ensemble members, relative to the MIROC5, is significant for all the ensemble members (i.e., the null hypothesis of a symmetrically distributed (around zero) difference between the errors of the MIROC5 and those of the WRF configurations was rejected) (von Storch and Zwiers 2002; Strobach and Bel 2016). The \( p \) values of this significance test were smaller than 0.017 for each of the eight ensemble members.

The climatology of the maximum daily temperature was also calculated [Figs. A2b(1)–b(3)] for the eight configurations, the MIROC5 and the observations in three different locations (the same locations used for the daily minimum temperature climatology). The phase of the annual cycle is captured by all the simulations in all three locations, as expected. In all three locations, the MIROC5 overestimates the maximum temperature during the summer and underestimates it during the winter (i.e., the amplitude of the simulated annual cycle during the summer and underestimates it during the winter). MIROC5 overestimates the maximum temperature by up to 3°C. The WRF configurations underestimate the maximum temperature (by up to 3°C) over most of the year. In Tel Aviv, the WRF configurations are closer to the observations and span a range that includes the observations except for the summer months for which they overestimate the maximum temperature by up to 4°C [Fig. A2b(2)]. In Elat, the WRF configurations underestimate the maximum temperature over most of the year by up to 4°C.

Figure 4b shows again that three of the four RRTMG configurations (b, d, and h) had a lower MAE than the corresponding CAM configurations (a, c, and g). Regionally, we found that the CAM configurations had a lower MAE in the south (not shown in the figure). Figure 4b also shows that the four Eta/MYJ configurations (e–h) had a smaller MAE than the corresponding MM5/YSU configurations (a–d). The mean absolute errors of only two of the WRF configurations (f and h) were significantly lower than the MAE of the driving model MIROC5 (based on the significance test described above, the null hypothesis of a symmetrically distributed difference between the ensemble member errors and the MIROC5 errors was rejected in favor of lower errors of the ensemble members with \( p \) values of 0.09 for member f and 0.06 for member h). Two ensemble members were found to be significantly worse (members a and c; the null hypothesis was rejected in favor of lower errors of the MIROC5 with a \( p \) value equal to 1), and the mean absolute errors of the other configurations were not significantly different from the MAE of the driving model (the null hypothesis could not be rejected; the \( p \) values ranged from 0.19 to 0.81).

Figure 5a shows a comparison between the annual averaged observed and predicted (by the eight ensemble members and MIROC5) daily minimum temperatures. For the observations, the average was calculated using all the stations with available measurements, and for the predictions, we used the grid cells corresponding to the station locations and calculated their average. Similarly to Fig. 3, it is shown that the four RRTMG configurations are well above the four CAM configurations, and the spread of the CAM configurations is larger. The MIROC5 predictions are well above the observations and the WRF ensemble member predictions, demonstrating that the global model had difficulty in capturing the minimum temperature.

The correlations between the observations and the predictions in all the parameterization configurations are positive but insignificant (von Storch and Zwiers 2002). The average (over the eight configurations) correlation is 0.13. Member b (WSM6-RRTMG-MM5/YSU) had the highest correlation at 0.21, and member c (WSM6-CAM-Eta/MYJ) had the lowest correlation at 0.06.

In terms of reliability, we found that the ratio between the STD and the RMSE is 1.02, suggesting that the ensemble is reliable. For the bias-corrected ensemble, the STD/RMSE \( \approx 0.09 \), suggesting that this ensemble is overconfident. For the MIROC bias-corrected ensemble (see the methods section), STD/RMSE = 0.03, and for the ensemble bias correction, STD/RMSE = 1.21. The difference between the ratios for the last two bias-correction methods reflects the fact that the ensemble bias correction reduces the error relative to the MIROC bias correction, as expected.

Figure 5b shows a comparison between the annual average observed and predicted (by the eight ensemble members and MIROC5) daily maximum temperatures. As seen in the figure, the four RRTMG configurations are above the corresponding four CAM configurations (but not all RRTMG configurations are above all CAM configurations), and the spread of the CAM configurations is larger (similar to the results for the minimum daily temperatures). The MIROC5 predictions are slightly above the WRF ensemble predictions for most of the time period but are closer to the observations.
The ratio between the ensemble STD and the RMSE of the ensemble average is 0.5, suggesting that the ensemble is overconfident. This is also reflected by the fact that the observations are outside the ensemble range in 6 out of the 10 years. For the bias-corrected ensemble, the ratio between the STD and the RMSE is 0.07, showing that this ensemble is highly overconfident. For the MIROC and the ensemble bias correction, STD/RMSE = 0.51 and 0.71, respectively.

Interestingly, for the maximum daily temperature, all the bias-correction methods and also the unmodified ensemble are overconfident.

Although neither the daily minimum nor the maximum temperatures are significantly correlated with the observations, we did find some relation between them. For the minimum daily temperature, in 7 out of 9 years, the predicted temperature increased when the observations increased and decreased when the observations decreased, relative to the previous year. The probability of this happening randomly is 2% (binomial distribution with 7 successes out of 9 trials and 50% chance of success in each trial), suggesting that it is not a coincidence. For the maximum daily temperature, the prediction had 6 successes out of 9, a probability of 9%.

Moreover, the two misses in the minimum daily temperature and two out of the three misses in the maximum daily temperature are related to the year 2010, which was reported by the IMS as an unusually hot year and the hottest observed (Israel Meteorological Service 2011). The average temperature for this year was higher by 2°–2.5°C relative to the 1981–2000 average.

Clearly, the MIROC5 and the WRF ensemble did not capture this unusual year.

b. Precipitation

Precipitation is usually more difficult to predict than surface temperature, and we began the analysis of this variable by studying the simulated spatial distribution of the precipitation and comparing it with previous studies. Figure 6 shows the seasonal accumulated precipitation during nine rainy seasons (2006/07–2014/15) averaged over the eight WRF configurations. The actual rainy season in the simulated region is from mid-October to early May (Goldreich 2003). Here, it is defined as the period from 1 October to 30 April. In the simulated region, the other months are usually dry except for some rare events. The spatial pattern of the precipitation, as predicted by the ensemble average, captures the observed north–south and west–east negative gradients (Ziv et al. 2014). The amounts also agree with the observations. In the appendix [Figs. A2c(1)–c(3)], we depict the simulated climatology of the precipitation by the eight WRF configurations, the MIROC5 and the observations for the rainy season of (2006/07–2014/15) in the same three locations considered for the temperatures (Haifa, Tel Aviv, and Elat; see the map in Fig. A1). The wet and dry seasons are well captured in the coastal locations (Haifa and Tel Aviv) and to a lesser degree in Elat, which is located at the driest part of the simulated region where rain can be considered as a rare event. The MIROC5 captures reasonably well the amounts of the precipitation in Haifa while underestimating it in Tel Aviv. In all three
locations, the WRF configurations tend to overestimate the amount of the precipitation.

The deviations of the different configurations from the ensemble average are presented in Figs. 7a–h. Some of the configurations show more (b and d) or less (e and g) precipitation over the entire region, while others (a, c, f, and h) show different (positive or negative) deviations from the ensemble mean in the northern and southern parts of the simulated region. To better illustrate the differences between the configurations, we present in the lowest row of Fig. 7 the differences between the average predictions of the two radiation, the two microphysics, and the two PBL/surface layer parameterizations. The CAM configurations predict less precipitation than the RRTMG configurations over almost the entire region. This result indicates a positive correlation between lower temperatures and less precipitation. In the northern part of the region, the difference between the two radiation schemes is larger than it is in the drier southern part. The YSU/MM5 configurations predict more precipitation than the MYJ/Eta configuration, and the difference is larger in the wetter northern part of the region. We also found that the WSM6 configurations predict more precipitation than the GCE configurations (with the difference being larger in the northern part, similarly to the differences between the PBL and radiation parameterizations), but the difference between these two microphysics schemes is less pronounced than the differences between the PBL/surface layers and radiation parameterizations.

The four Eta/MYJ configurations again have a lower MAE than the four MM5/YSU configurations (Fig. 8). The four CAM configurations have a lower MAE than the RRTMG configurations, unlike the results found for the daily minimum temperature. The MIROC5 has a significantly smaller MAE than all the WRF configurations (but it is not much smaller than the MAE of the two CAM-Eta/MYG members, e and g).

Figure 9 shows the observed annual (October–April) precipitation in the simulated region (averaged over all available stations) together with the predictions of the eight WRF configurations and the MIROC5 driving data (after interpolation to the WRF high-resolution grid). This figure shows that in terms of annual precipitation, the four Eta/MYJ configurations (e–h) performed better (with smaller MAEs) than the four MM5/YSU configurations, and similar results were found from a station-by-station comparison (Fig. 8).

Although the long-term means seem to agree with the observations, the last four configurations (e–h) have negative and insignificant correlations with the observations. The ensemble members and the driving data perform even worse in predicting an increase or decrease relative to the previous year. In terms of the ensemble spread, we found that the ratio between the STD and the RMSE is $0.57$, implying that the ensemble is overconfident. The bias-corrected ensemble is even more overconfident, with an STD/RMSE ratio of $0.19$. For the MIROC and ensemble bias-correction methods, STD/RMSE = 0.52 and 0.99, respectively. These results suggest that the ensemble reliably predicts the precipitation anomalies but is overconfident in predicting the actual precipitation values.
FIG. 7. Seasonal accumulated precipitation anomalies of the rainy seasons 2006/07–2014/15 for (a)–(h) the eight configurations. (bottom) The differences between the average predictions of the two radiation (CAM-RRTMG), the two microphysics (WSM6-GCE), and the two PBL/surface layer (MM5/YSU-Eta/MYJ) configurations.
4. Predicted decadal climate change

In this section, we examine the changes predicted by the ensemble members for the near future (the next two decades). We focus on significant changes relative to the current climate and relative to the predictions of the driving model (the deviations of the high-resolution model from the driving model are expected to be the added value of regional climate modeling).

Figure 10a shows time series of the annual average minimum daily surface temperature in Israel (averaged over all the grid cells spanning Israel) as predicted by the eight WRF configurations and the MIROC5 (after interpolating its predictions to the high-resolution grid of the WRF configurations; see the methods section for more details). The lines corresponding to the different configurations rarely cross each other, suggesting that the interannual variability of each configuration is smaller than the variability between the ensemble members (which is mostly dictated by the bias between the different configurations). Using the Mann–Kendall nonparametric test for monotonic trends and the Theil–Sen estimate of linear trends on the time series, we found that all eight configurations predict a significant (in the statistical sense) increase in the daily minimum surface temperature by $0.45^{\pm}0.56$°C decade$^{-1}$, with an ensemble average of $0.51^{\pm}$°C decade$^{-1}$. The MIROC5 predicts an increase of $0.53^{\pm}$°C decade$^{-1}$.

The simulated increase in the minimum and maximum daily surface temperatures is larger than the projected global average temperature change for RCP4.5 (Kirtman et al. 2013) (the difference between the average 2016–35 global temperature and the 1986–2005 average global temperature is predicted to be between $0.47^{\pm}$ and $1.00^{\pm}$°C) and is in agreement with the identification of the Mediterranean region as a “hot spot” (Diffenbaugh and Giorgi 2012). However, one should remember that the minimum/maximum surface temperature and the average surface temperature are different variables and may show different trends.

Figure 10c shows Israel’s annual precipitation for the 30-yr period of the experiment. All the WRF configurations show higher amounts of precipitation than the driving model and no significant monotonic trend.

The white dots represent grid cells in which a significant positive trend is predicted, and the black dots represent grid cells in which a significant negative trend is predicted (the significance was determined using the Mann–Kendall nonparametric test, and a significance level of $p = 0.1$ was used as the threshold for plotting the significance of the trends).
For all grid cells, the predicted temperature for the last simulated 15 years is larger than the simulated temperature for the first 15 years. The maximal increase (\(\pm 1.2^\circ\text{C}\)) is predicted in the northeastern part of Israel (Mount Hermon area) and the minimal increase (\(\pm 0.5^\circ\text{C}\)) is predicted in the southwestern region, the Dead Sea region, and near the Sea of Galilee. All the surface temperature variables (daily minimum, maximum, and mean) show a significant increase, but the largest increase is seen for the minimum daily temperature.

The precipitation increases in the north and the south and decreases in the center. A significant precipitation increase is seen in a small region in southern Israel. This increase is considerable since the annual precipitation in this region is smaller than 50 mm yr\(^{-1}\), and the increase is of the same order of magnitude. In some grid cells in the north, the precipitation increase was found to be larger than in the south, but it is a smaller fraction of the average annual precipitation (more than 1000 mm yr\(^{-1}\) in some locations). The zonal and meridional wind speeds show mainly insignificant negative trends with several grid cells in the south showing a significant wind decrease.

Trends in the regional climate model may stem from the driving global model or from the higher-resolution regional model. To identify the source of the simulated trends, we present in Fig. 12 the differences between the last 15 years and the first 15 years of the MIROC5 model (the same analysis shown in Fig. 11 for the WRF configurations). Israel is represented by six MIROC5 grid cells, with one of them including only one WRF grid cell (in the southwest). Figure 12 clearly demonstrates that the MIROC5 does not capture all the details of the spatial pattern seen in the predictions of the regional model. Nevertheless, the significance test shows that the average temperature is increasing—a similar conclusion to the one seen in the regional model. However, the maximum daily temperature increase predicted by the driving model is larger than the predicted increase for the daily minimum temperature. This is the opposite of the WRF ensemble predictions. In addition, none of the other variables showed a significant increase or decrease in the driving model (similarly to the regional model results).

5. Spectral nudging

Regional climate models are driven by global models. Models that use spectral methods for the forward (in time) integration include different spatial scales. The dynamics of the large spatial scales, which are of the same order of magnitude as the scale of the simulated domain, are not expected to be well captured by the regional model. Moreover, it was shown that most of the errors in regional climate models appear for large spatial scales (>2000 km) (Vukicevic and Errico 1990; Miguez-Macho et al. 2004). Therefore, spectral nudging is often used, which means that the regional simulated dynamics of the large scales are corrected using the global simulated dynamics of these scales. We ran two sets of WRF simulations, one with spectral nudging for the entire duration of the simulation (30 years) and one with spectral nudging only during the first six years of the simulations. Figure 13 demonstrates the effect of the spectral nudging by showing the minimum daily surface temperature and precipitation time series for the WRF simulations in which the spectral nudging was stopped after the first six years.

The minimum daily surface temperature (Fig. 13a) and the maximum daily surface temperature (not shown)
Fig. 11. The difference between the last 15 years and the first 15 years of the experiment for annual minimum, maximum, and average surface temperatures; average zonal and meridional wind; and total precipitation. The white dots represent grid cells in which significant positive trends are predicted, and the black dots represent grid cells in which significant negative trends are predicted (the significance was tested using the Mann–Kendall nonparametric test).
FIG. 12. Differences between the last 15 years and the first 15 years in MIROC5 predictions for annual average surface temperature, average zonal and meridional wind, and total precipitation. The white dots represent grid cells in which significant positive trends are predicted, and the black dots represent grid cells in which significant negative trends are predicted (we used the same significance test that was used for the WRF predictions).
show qualitatively similar results to those obtained using a complete 30-yr nudging experiment. After the first six years, the annual variability is different, but the overall trend is similar, and no considerable change in the surface temperatures is observed after the nudging was stopped. On the other hand, the precipitation shows a major increase after the first six years that seems to extend up to \( \approx 2017 \). Following this period of sharp increase, the precipitation shows a gradual (insignificant) decrease. The average annual precipitation in the partially nudged simulations were larger than those in the fully nudged simulations by 54–187 mm yr\(^{-1}\) (spatial average over the simulated region) because of the sharp increase in the precipitation when the nudging was stopped. For example, the annual precipitation values in the partially nudged experiment reach above 1000 mm yr\(^{-1}\) for member b of the ensemble (WSM6/RRTMG/MM5/YSU), compared with \( \approx 800 \) mm yr\(^{-1}\) in the corresponding completely nudged experiment.

### 6. Summary and discussion

The results above presented the analysis of WRF regional climate simulations of the Middle East. We generated an ensemble of continuous 30-yr climate simulations generated by eight different WRF configurations for the region of Israel. The grid used had a 6 km \( \times \) 6 km resolution, which better captures the climatic and elevation gradients in this region. The eight configurations were generated from the permutations of two microphysics schemes, two radiation schemes and two PBL/surface layer schemes. The model was forced by the MIROC5 global climate model in its decadal prediction configuration, that is, it was initialized with interpolated observed conditions. It has been shown that initializing climate models with observed conditions can increase the forecast skill during the first decade after the initialization (Smith et al. 2007; Keenlyside et al. 2008; Pohlmann et al. 2009; Doblas-Reyes et al. 2013). However, these predictions can also be accompanied by a drift of the model to its own climatology, which degrades the forecast skill (Meehl et al. 2009, 2014).

All the WRF configurations considered here were able to reproduce the large-scale spatial distribution of the annual mean temperature, which increases in low latitudes and low elevations (Fig. 2). Specifically, the ensemble average shows that the lower temperatures over Judea and Samaria and also in the high region of Mount Negev are well captured (Fig. 2). The simulated climate also captures well the phases of the seasonal cycle (seasons). However, the values of the maximum and minimum daily temperatures and the precipitation amounts are not well captured (Fig. A2). For the minimum daily temperature, all the WRF configurations showed a smaller MAE than the driving GCM, while for the maximum daily temperature, only three of the configurations [all of which use the local PBL scheme (MYJ/Eta)] showed a smaller MAE than the MIROC5 (Fig. 4). We believe that the poorer performance of some of the WRF configurations in predicting the maximum daily temperature is related to the poorer performance of the nonlocal PBL.
scheme (YSU), which has been shown to overestimate the moisture content and underestimate the PBL height (Liu and Liang 2010; Que et al. 2016). Configuration g, which shows a somewhat poorer performance than the MIROC5 despite using the better PBL scheme (Eta), uses the GCE microphysics, which is known to overestimate the cloud cover, and the CAM radiation scheme, which is known to underestimate the temperatures; both effects contribute to the underestimation of the maximum daily temperature and the poorer performance (as shown in Fig. A2 for Haifa and Elat). The configurations using the CAM radiation scheme generated colder climates than the configurations using the RRTMG radiation scheme (Fig. 3). The results depicted in Fig. 5 clearly show that the four CAM configurations (a, c, e, and g) simulate a lower minimum surface temperature than the RRTMG configurations (b, d, f, and h). In particular, the minimum temperature panels in Figs. 5 and 10 show the grouping of the configurations according to the radiation scheme. Zittis and Hadjinicolaou (2017) showed that the longwave radiation fluxes are underestimated by the CAM radiation scheme; these underestimated fluxes reduce the loss of heat by the surface but also reduce the heating of the surface by the higher atmosphere, eventually resulting in lower surface temperatures. The effect is more significant for the minimum daily temperature because it is dictated mainly by the longwave radiation. It was also suggested that the colder temperatures simulated by the CAM configurations are due to larger cloud cover and, consequently, reduced shortwave radiation reaching the surface (Zittis and Hadjinicolaou 2017). For the maximum daily temperature, which strongly depends on the shortwave radiation, there is no grouping of the configurations according to the radiation scheme. The difference between the maximum and minimum temperatures suggests that the effect is mainly due to the longwave radiation. The temperatures predicted by the configurations using the WSM6 microphysics scheme are higher than those predicted by the GCE scheme (Fig. 3). These results are most likely related to the higher cloudiness predicted by the GCE scheme and are in agreement with the results of Zittis et al. (2014).

In terms of model-data misfit, the eight ensemble members had smaller errors than the driving data for minimum temperature, and only two ensemble members had significantly smaller maximum temperatures. However, the predictions of the four Eta/MYJ members had smaller errors than the MM5/YSU members, suggesting an advantage of the local PBL scheme over the nonlocal PBL scheme. This finding is in contrast to previous studies that found better performance of the nonlocal schemes (Hu et al. 2010; Cohen et al. 2015). The advantage of the nonlocal PBL schemes was also demonstrated for short-term forecasting (Banks et al. 2016). Figure 3 shows that the performance of the two PBL schemes varies between regions with elevation gradients and relatively flat regions; however, most of the stations used in the comparison and in determining the MAE are located in the coastal region, which is likely to cause a bias in performance assessment toward the PBL scheme that performs better in the coastal region. The RRTMG schemes had smaller errors than the CAM in three out of the four configurations (Fig. 4; the comparison is between configurations that differ only in the radiation scheme). This finding is in agreement with previous studies that found the RRTMG to perform well in desert areas (Branch et al. 2014; Zittis and Hadjinicolaou 2017). The correlations between the observations and the WRF ensemble were found to be insignificant for both the minimum and the maximum daily temperatures (Fig. 5).

The spatial distribution of the precipitation predicted by the eight configurations is in agreement with the observations (Fig. 6). In general, we found that the CAM configurations predict less precipitation than the RRTMG ones (and, as mentioned above, also lower temperatures), the MYJ/Eta configurations predict less precipitation than the YSU/Eta configurations, and the GCE configurations predict less precipitation than the WSM6 ones, despite the higher cloudiness expected by the GCE scheme. However, the difference in the precipitation between the two microphysics schemes is less pronounced than the differences between the radiation and PBL/surface layer schemes (Fig. 7). The four MYJ/Eta configurations were found to be closer to the observations than the four YSU/MM5 configurations (Fig. 9). These findings are in agreement with previous works that found the YSU/PBL scheme to overestimate the moisture content in the boundary layer and underestimate the PBL height, thereby simulating higher precipitation values (Liu and Liang 2010; Que et al. 2016). The MIROC5 showed a smaller MAE than all the WRF configurations in the prediction of the precipitation (Fig. 8).

All the WRF configurations (and the driving model) predicted a positive and significant temperature trend over Israel for the next two decades. The trend was found to be stronger than the globally projected temperature increase. However, the warming seen in the WRF simulation is smaller than or comparable to the trend of the driving data. Therefore, the main added value of the regional model is the spatial variability of the trend, which is only provided by the regional models. It is important to note that the increase in the
maximum temperature, which the RCMs predict to be smaller than the GCM prediction, is of great importance to ecosystem function and agricultural production, with related societal implications.

No significant trend was found for the precipitation in the entire domain (Fig. 10) [similar to the results of Hemming et al. (2010)]. The lack of a significant trend is in agreement with the driving data and the regional WRF simulations. Spatially, the precipitation trend was found to be positive in southern and northern Israel and negative in the central region. The trend was found to be statistically insignificant in most regions except for significant positive trends in some places in the southern part of the simulated domain. Wind trends were mostly negative and insignificant, though we found some significant negative trends in the south and in the Jordan Valley (zonal wind) in the regional WRF simulations.

Two grid cells with a significant positive trend were found in the northeast for the meridional wind (Fig. 11).

Spectral nudging improved the precipitation forecast considerably and prevented significant decoupling of the RCM from the driving GCM. The sharp increase in precipitation in the absence of nudging suggests that the prediction of this variable is dominated by large-scale processes that are better captured by the global driving model. The temperature forecast was less sensitive to the spectral nudging, suggesting the dominant role played by processes at smaller spatial scales (Fig. 13).

The ensemble generated in the research and reported in this manuscript is based on one regional model (WRF), driven by one global model (MIROC5) and different parameterization schemes. This kind of ensemble is not expected to span the full range of possible predictions because the effects of different driving
models, different parameters, different initialization methods and times, and different grids are not represented. In a global setup, using an ensemble of different climate models, it was shown that model uncertainties are the main source of uncertainties in decadal climate predictions (Strobach and Bel 2017a). The ensemble, despite spanning only eight different sets of parameterizations, was found to be reliable for the minimum daily temperature and overconfident for the maximum daily temperature. Applying bias corrections to the predictions of these variables resulted in overconfidence for both. For the precipitation, the ensemble was found to be overconfident, but applying bias correction to account for the average deviation of the ensemble average from the observations improved the reliability. Other sources of variability, such as those mentioned above, are important for generating a reliable ensemble forecast in terms of the error–spread ratio (Palmer et al. 2006; Strobach and Bel 2017b). However, this study shows that even for decadal time scales, the differences between the various parameterization schemes make a considerable contribution to the ensemble spread.

FIG. A2. [a(1)]–[a(3)] Minimum daily surface temperature climatology, [b(1)]–[b(3)] maximum daily surface temperature climatology, and [c(1)]–[c(3)] precipitation climatology for (top) Haifa, (middle) Tel Aviv, and (bottom) Elat. The colored lines represent the eight WRF configurations and the driving model, MIROC5. The black line represents the observations.
In summary, we found that in the simulated region, the main advantage achieved by the dynamical downscaling is the better spatial resolution highlighting regional differences in trends and the varying performance of the parameterizations between different regions based on their climate conditions and topography. For the minimum and maximum daily temperatures, some configurations also showed a reduced MAE, while for the precipitation, some configurations showed a similar MAE to the driving GCM, but no improvement was achieved. The lack of improvement in predicting the precipitation could be due to the fact that in the simulated region, most of the precipitation is caused by the Cyprus lows (Goldreich 2003; Alpert et al. 2004), which are large-scale synoptic systems (better captured by the GCM).

Acknowledgments. The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under Grant 293825. This research was partially supported by the Israel Ministry of Agriculture and Rural Development (Eugene Kandel Knowledge Centers) as part of the Root of the Matter: The root zone knowledge center for leveraging modern agriculture. We acknowledge the World Climate Research Programme’s Working Group on Coupled Modelling, which is responsible for the CMIP. For the CMIP, the U.S. Department of Energy’s Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals. The CMIP5 data may be accessed at https://esgf-node.llnl.gov/projects/CMIP5/.

APPENDIX

Stations Map and Representative Climatologies

Figure A1 shows the locations of stations with available observations for surface temperatures (Fig. A1a) and precipitation (Fig. A1b). These station data were used in the assessment of the simulations’ performance.

Figure A2 shows the climatology of the minimum [Figs. A2a(1)–a(3)] and maximum [Figs. A2b(1)–b(3)] surface temperatures, and the precipitation [Figs. A2c(1)–c(3)] in three different locations [Haifa; Figs. A2a(1),b(1),c(1)], Tel Aviv [Figs. A2a(2),b(2),c(2)], and Elat [Figs. A2a(3),b(3),c(3)]. The locations of these stations are denoted in the maps in Fig. A1. The climatology was calculated for the observations, the driving MIROC5 GCM, and the WRF configurations. The surface temperatures climatology is based on 10 years, 2006–16, and the precipitation climatology is based on nine rainy seasons (2006/07–2015/16).

REFERENCES


