Central-Eastern China persistent heat waves: Evaluation of the AMIP models.

Citation for published version:
Freychet, N, Tett, S, Hegerl, G & Wang, J 2018, 'Central-Eastern China persistent heat waves: Evaluation of the AMIP models.', *Journal of Climate*. https://doi.org/10.1175/JCLI-D-17-0480.1

Digital Object Identifier (DOI):
10.1175/JCLI-D-17-0480.1

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Journal of Climate

Central-Eastern China persistent heat waves: Evaluation of the AMIP models

N. Freychet*

School of Geosciences, University of Edinburgh, Crew Building, The King's Buildings, Edinburgh EH9 3FF, UK

S. F. B. Tett, G. C. Hegerl

School of Geosciences, University of Edinburgh, Crew Building, The King's Buildings, Edinburgh EH9 3FF, UK

J. Wang

Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China

**Corresponding author address:* Nicolas Freychet, Crew Building, The King's Buildings, Alexander Crum Brown Road, Edinburgh EH9 3FF, UK

E-mail: nicolas.freychet@ed.ac.uk

ABSTRACT

Large scale and persistent heat waves affecting Central-Eastern China are investigated in 40 different simulations of sea surface temperature driven global atmospheric models. The different models are compared with results from reanalysis and ground station datasets. It is found that the dynamics of heat wave events is well reproduced by the models. However, they tend to produce too persistent heat wave events (lasting more than 20 days) and several hypothesis were tested to explain this bias. The daily variability of the temperatures or the seasonal signal did not explain the persistence. However, interannual variability of the temperatures in the models, and especially the sharp transition in the mid-90s, has a large impact on the duration of heat waves. A filtering method was applied to select the models closest to the observations in terms of events persistence. The selected models do not show significant difference with the other models for the long term trends. Thus, the bias on the duration of the events do not impact the reliability of the model positive trends, mainly controlled by the changes in mean temperatures.

1. Introduction

Large scale and persistent heat waves (HW) over East China have a large environmental and socio-economic impact (e.g. Luber and McGeehin 2008; Wang et al. 2015) and have been the focus of many studies (see for example Perkins (2015) and Lu and Chen (2016) for a review). During the past few decades, the frequency of these events have been found to increase (Wei and Chen 2011; Wang and Fu 2013; Ren et al. 2005, 2016; Zhou and Wang 2016). But this trend is not always consistent and can vary in some regions (Yan et al. 2011b; Ding and Qian 2011; Dong and Huang 2015). Freychet et al. (2017) showed that, for large scale heat waves, this trend is mainly due to increase in the mean temperature. This study also showed that HW are related to strong mid-troposphere positive anomaly and to an enhanced heat and moisture transport in the lower troposphere. On the other hand, Luo and Lau (2017) indicated dry conditions associated with HW over Southern China. Other works have also pointed out the role of the reduction in the snow cover over the western Tibetan Plateau (e.g. Wu et al. 2012; Sun et al. 2014) and of the Eurasian teleconnection pattern (Wang et al. 2016a). Thus, different processes are involved in the formation and magnitude of the HW events.

Adaptation to such events for the next few decades is important and was investigated by the Working group II of the IPCC5 Fifth Assessment (IPCC 2014, Kripalani et al. (2007)). Many studies, relying on global climate model projections such as the CMIP5 (Coupled Model Intercomparison Project Phase 5) ensemble, indicate an increase in HW events for the future decades in terms of frequency, intensity and duration (e.g. Guo et al. 2017). As many different models are used for such ensemble experiments, the confidence on these projections can be questioned, especially for extreme or rare events (Freychet et al. 2015, 2016). The main objective of this study is to conduct an evaluation of the AMIP models for persistent and large scale heat waves over Central-

Eastern China (CEC) and use these evaluated models to estimate the changing risk of such events. The region is chosen to be close to Lin et al. (2015) definition. It is heavily populated and extreme temperature events can impact a large population. Urbanisation is also important and can locally impact the temperatures. However, this aspect is not included in the current global climate models and should not change the results of this study. It must also be noted that the results presented in this study are specific to the definition of the region. Other area could lead to different findings depending on the dynamics (e.g. Wang et al. 2016b). Even if using realistic SST forcing, AMIP simulations are not reanalyses, thus it is not expected that they can reproduce the same heat waves at the same dates. In this study only statistical approaches are considered, different from a case analysis such as Luo and Lau (2017) for instance.

Our focus is on the atmospheric component of the climate models and the evaluation is based on two different reference datasets, defined in Section 2. Another ensemble of 15 members of the Met Office HadGEM3-GA6-N216 model (Walters et al. 2017) is also used to examine the intra-variability of the models. The study investigates if the AMIP ensemble is consistent in terms of dynamics (Section 3) and if the models can reproduce HW signals in the observational datasets (Section 4). A major question is to verify that the models are consitent in terms of risk change. This point is addressed in section 5, before concluding in Section 6.

2. Data and heat waves definition

a. Data

1) Reanalysis and observations

Maximum and minimum temperatures (Tmax and Tmin) and atmospheric circulation variables from ERA Interim reanalysis (ERAI, Dee et al. (2011)) are used as a reference for this study. Daily data are extracted at 0.75 degree resolution, and the 1979-2010 period is used. Homogenized ground station observations of temperature (OBS, Li and Yan (2009)) are also used. OBS are first regridded on the ERAI grid (shown in Fig.1a-b for Tmax) by averaging, for each grid point, the corresponding available data from OBS. If no OBS data is available for a grid point, then it is masked.

A significant bias exists between ERAI and OBS (not show). ERAI is too cold, especially over the central and Southern China. Part of this bias may be related to the urban effect that can impact locally the ground station temperatures (Yan et al. 2011a). Another part of this bias may be due to elevation effect, that is directly recorded in OBS but could be missing in ERAI due to the resolution. A modification is applied to OBS so that it is more consistent with ERAI. To do so, a linear temperature gradient coefficient (C_Z = 0.6K/100m) is combined with the difference between the elevation of each station (Z_{OBS}) and the elevation of the co-located ERAI grid point (Z_{ERAI}) to obtain an adjustement term (dT) equivalent to: $dT = C_Z \times (Z_{OBS} - Z_{ERAI})$. This term is then applied to the temperatures at the station. The station observations are then regridded on the ERAI grid. Also note that the choice of a fixed coefficient C_Z is arbitrary and can vary significantly according to the land type (Li et al. 2013). Thus the adjustment method employed here should not be considered as perfect.

After adjusting the elevation effect, the differences between ERAI and OBS are reduced (Fig.1c) compared to the raw data differences (not shown). This indicates that part of the differences between reanalysis and observation are due to the fact they represent temperatures at different elevations, stations being more often located in the valley while reanalysis grid point correspond to the mean elevation of the region. Other processes impacting temperatures at a very local scale such as aerosols or urban effect (e.g. Gong and Wang 2002; Heisler and Brazel 2010; Yan et al. 2011a) could explain the remaining differences. Results for Tmin show lower biases compared to

Tmax, and the elevation correction also reduces the differences between ERAI and OBS (Fig.1f). Hereafter, OBS will refer to the regridded ground station observation, after elevation correction. Moreover, the term “observations”will be used to include both ERAI and OBS, when comparing the results with the models.

2) Model data

Daily data from 1979 to 2008 from an ensemble of 40 members of the AMIP multi-model ensemble (AMIP) is investigated. As some models have several members, the total of independent models is 21 (Table **??**). AMIP models correspond to the same CMIP5 models but are forced by prescribed sea surface temperature (SST) during the historical period, removing uncertainties due to ocean models. The study does not investigate individual performance of each model. However, for each diagnostic performed, the list of the five models with the lowest and the highest scores is given in Table **??**. The user may refer to this table to see individual model performances.

Another ensemble of 15 members from the Met Office HadGEM3-GA6-N216 atmospheric model is used (N216). It also follows an AMIP-like experiment, i.e. forced by prescribed SST during the historical period, and data are extracted for the same period. The N216 ensemble is mainly used to estimate the internal variability and uncertainties. It runs from 1960 to 2013, but the same period (1979-2008) as the AMIP is used for analysis.

b. Heat waves definition and computation of the composites

1) Heat waves definition

There are many ways to define HW events and trends can be different depending on the index definition (You et al. 2016). Here we focus on large scale and persistent events, and the definition of HW used in the study follows that of Freychet et al. (2017). Daily Tmax and Tmin are both averaged over the Central-Eastern China (CEC) region (105E-125E, 30N-40N), and the 90^{th} percentile is computed for each temperature, using the extended summer period (May-September) of each year. A warm day is defined as when both Tmax and Tmin are above their respective 90% values on the same calendar day. A HW event is defined when at least 5 consecutive days are warm days. Note that this methodology is applied independently to each dataset (ERAI, OBS, and each model member) to define their own 90^{th} percentile removing mean temperature bias.

As the main objective of this study is to focus on the most threatening events for society, HW highlights the warmest events in an absolute way. As the temperatures are warmer during mid July, it is expected that most of the HW events will be identified during this period too. Thus, HW events can be seen as a phenomenon that amplifies the seasonal transition and increases the temperature during the warmest period. It also implies that HW events are related to the seasonal transition. This point will be further discussed in the Section 4.

2) COMPOSITES

To study the atmospheric circulation during the HW, a composite method is applied to an atmospheric variable labelled X. When a day d is identified as part of a HW event, the corresponding variable X_d at time d is extracted, and the climatology (as a 5-days running mean) of X at the same calendar day (X_{d-clim}) is removed. To remove any long term trend and variability and to focus on anomalies due to the HW, the difference between the annual mean around the time d (X_{d-ann}) and the mean 1979-2008 climatology of X (X_{clim}) is also removed. Thus, only the anomaly (X_d) due to the HW remains.

$$X_d = X_d - X_{d-clim} - (X_{d-ann} - X_{clim}) \quad (1)$$

The composite of X corresponds to the averaging of the anomalies from all the HW days during the studied period (see Appendix for a schematic view).

3. Heat Wave Dynamics

It is first important to verify if the model can reproduce the observed dynamics of events. For that, a composite analysis is used, as described in Section 2.

The dynamical processes correlated with persistent HW events have been described in details in Freychet et al. (2017). Here we verify that the models can reproduce the composite ERAI signals. The ensemble mean of the AMIP models can reproduce the observed dynamical patterns (Fig.2). A mid-troposphere high pressure (Z500) along with a subsidence anomaly (W500) and northward shift of the subtropical jet (U200) leads to an increase in surface solar radiation (SSR) and favour higher Tmax. The specific humidity (S.Hum.) is also higher than usual during these events and is important to reduce the night time cooling and keep Tmin higher. Finally, the low level circulation (SLP) pattern corresponds to the development of a meridional cell anomaly with an upward motion over the the North-East of the CEC region. This anomaly has been hypothesised to lead to return wind from the North and to increase the heat convergence over CEC during the HW (Freychet et al. 2017).

The individual member performances are tested (Fig.3a,b). Most of the models are close to the reference (ERAI) in terms of correlation (between 0.7 and 0.9). The scatter of the N216 members, especially for the SLP, indicates a high intra-model variability. Poor results may be due to a too strong control of the seasonal transition in some members instead of an anomaly of the circulation (i.e. HW events may be triggered by an overall large increase in temperature during the peak of the summer). The ensemble mean is overall consistent with ERAI in terms of patterns (correlation) but tend to have a weaker signal due to the ensemble averaging. The dynamical signal is tested furthermore with a lag-composite analysis (Freychet et al. 2017), from 10 days before to 10 days after the HW events. The anomalies are averaged over the [105E-125E, 30N-40N] region for Z500, and over the [115E-140E, 40N-50N] region for the SLP. The evolutions of these anomalies are compared with ERAI results and displayed in Fig.3c,d. The ensemble mean is able to reproduce the signal with a good correlation (0.8 for Z500 and 0.9 for SLP), but individual results are more scattered. Interestingly, the ensemble is more consistent for the SLP, indicating that the low-level dynamical response in the models is a robust result. Other variables are also tested (surface solar radiation, 500hPa vertical wind and 850hPa specific humidity, not shown). Results are overall similar to findings in Fig.3: the ensemble mean is consistent with ERAI, but individual models can have weaker performances.

Overall the AMIP ensemble is able to reproduce the main spatial and temporal evolutions of the dynamical patterns of HW events, even if some individual members are less consistent.

4. Representation of heat waves in the AMIP models

This section investigates if models can reproduce HW events compared to observations, in terms of number and duration during the historical period (Section a and b). Following this, the possible reasons for the model differences are explored. Finally, Section d discusses the interannual variability of the events.

a. Estimation of heat waves events in models

The difference between the reanalysis and observations shows that the estimated number of observed heat waves has considerable dataset uncertainty (Fig.4a). For example, Fig.4 of Luo and Lau (2017) shows another example of different heat wave number and intensity estimates (over South China) based on different datasets (reanalysis or weather station). High variability is also seen between the different models or even between different simulations of a same model. Indeed, when looking at the different members of the N216 ensemble, the number of days may vary from

30 to 60, and the standard deviation of N216 ensemble is about the same order as the observed uncertainty. Thus, the statistics on these events are very sensitive to the sampling processes and both modelled or observed events must be considered within a margin of error. The fact that these events are rare and the period is limited suggests that part of the difference may be simply due to the variability. Considering, the actual number of heat waves per year (Fig.4b), results are more consistent between observations but the AMIP models still tend to produce too many events and have a large scatter.

To verify that the differences between models and observations are not an artefact due to an incorrect seasonal signal, the seasonal climatology is corrected in each model and OBS, using the seasonal climatology of ERAI. To do so, the 31-day smoothed climatology is removed from the simulated temperatures (or OBS), and the 31-day smoothed climatology from ERAI is added. Then heat waves are computed using the corrected data (Fig.4c and d). The total number of heat wave days in the AMIP ensemble is not improved by such methodology. Interestingly, the number of events in OBS is enhanced, increasing the uncertainties in the observations. The seasonal signal may influence the production of heat waves, and with the same seasonal climatology the reanalysis or ground stations have a different estimation of the number of events. Consequently, the uncertainties on the true estimate is larger and models are more consistent with observations. As correcting the seasonal climatology does not improve the results, the actual temperatures are used from hereon.

b. Event Persistence

To investigate in more detail the reasons for the overestimation of the number of heat wave days, the persistence of the warm events is displayed in Fig.5 (a warm event being a combination of both Tmin and Tmax above their respective threshold during a same day). As defined before, an event is defined as a heat wave if it lasts at least 5 days, but in Fig.5 shorter events (1 to 4 days) are also plotted to obtain a full spectrum of the warm events persistence. For each models, results are displayed as a percentage relative to the total number of warm days in this same model (or observation). For instance, if a model has 12 warm events lasting for 2 days, and in total it has 300 days of warm events (all length grouped), then it would have 8% of events with a persistence of 2 days. The mean persistence of the events of more than 5 days is also displayed for OBS, ERAI each AMIP and N216 members.

ERAI is, overall, consistent with the station data (Fig.5), though there are more short events in ERAI (4 days), and less long lasting heat waves than in the gridded station observations. The maximum heatwave length in ERAI is 9 days, while in OBS it can reach 12 days, and the percentage of long lasting events is larger in OBS than in ERAI. However, this differences are relatively small compared to the differences with the models, and could be due to local effects (e.g. urban effect) not resolved in ERAI. Many AMIP members produce very persistent events that can last for 20 days or more. The mean duration is found to be 6 days in ERAI and 8 days in OBS, and ranges from 5 to 11 days in the models. Thus the mean duration may be considered as realistic in some models, but specific longer events could be problematic and some models are outside the range of the observational uncertainties. Possible reasons for such behaviour are explored below.

c. Hypothesis for the over-persistent heat waves

Three main hypothesis are investigated in this section: the variability of the models, the variability of the temperatures in the models, and the influence of the seasonal signal.

1) Internal variability and observation error

Even if long persistent HW events (more than 10 day events) are observed in many simulations, considerable internal variability exists in the models, illustrated in Fig.6 for the N216 simulations. Some members can simulate a reasonable ratio of long persistent events whereas other simulations produce mostly long lasting events. These differences are also observed in the AMIP ensemble (not shown). Thus, the persistence of the events cannot be attributed to a systematic bias of a model, but may be linked to the internal variability of the model.

A crude estimation of the realistic range of the maximum persistence is made, based on the observations mean (μ_{obs}=10.5 days) and differences (σ_{obs}=3 days) and on the N216 standard deviation (σ_{N216}=5.1 days). Considering that the uncertainties are simply independent and cumulative, the maximum realistic persistence could be considered as:

$$\mu_{obs} + \sqrt{(\sigma_{obs}^2 + \frac{\sigma_{N216}^2}{N})} \quad (2)$$

with N the number of members (for example 15 for the N216 ensemble mean). The result would be 14 days for the N216 ensemble mean and 16.5 days for a single member. It means that an event persistence of 16.5 days in a single member can be considered as reasonable, given the range of the intra-model variability and observation uncertainties. This explains part of the differences between the models and the observation, but not the most persistent events. It is still important to understand if a specific factor controls the variability of the persistence, or could be attributed to chaos. Thus, other factors are investigated below.

2) Temperature variability

The daily variability of the temperatures is an important aspect that can explain over-persistent warm events. Indeed, if a model has a systematic too low daily variability of the temperature (thus with temperatures more stable from one day to another), it may lead to more stable temperatures and thus longer events. This hypothesis is investigated in Fig.7 (a,b). The variability is computed by removing the 3-days running mean and taking the standard deviation of the anomaly (for Tmin and Tmax separately). No clear relationship can be found between the variability of Tmin and the HW persistence. But many models producing long HW events (red circles) tend to correspond to weaker variability of Tmax (with an overall correlation of -0.61). Thus, a too weak daily variability of the maximum temperature in the models could lead to more systematic long HW events. However, this signal is not observed for N216, and the models with a similar (or lower) observations variability have too many long heat waves. Thus the biases cannot be explained by variability alone, though it has an impact on the duration of the events in the models.

3) EFFECT OF THE SEASONAL CYCLE

The amplitude of the summer range (i.e. the difference between the coldest and warmest period of the summer based on the 5-day smoothed climatology) could also impact the HW persistence. Too large a summer range would lead to systematically too persistent heat waves, as the warmest period would be above the threshold used to detect HW. This hypothesis is tested in Fig.7 (c,d). The summer ranges for Tmax and Tmin correspond to the difference between their highest and their lowest magnitudes respectively (based on the daily climatology smoothed by a 5 days running mean). Again, no clear relationship is found between this signal and the persistence of HW, either in terms of inter- (AMIP ensemble) or intra-model (N216 ensemble) variability. However, it is noticeable that all the members (AMIP and N216) have a larger seasonal range for Tmax, compared to ERAI.

As the simulated summer range is generally larger than observations, persistence is analysed after correcting the seasonal climatology as explained in Section 4. It is clear that even after correcting the seasonal climatology, differences in the persistence (Fig.8a.) are still noticeable for both AMIP and N216.

A last case is to consider heat waves events in terms of anomalies, by removing the seasonal climatology from the temperatures before computing HW events. This correspond to the methodology described before to correct the climatology, except that the ERAI climatology is not added after removing the model climatology. In this case the events are independent from the seasonal signal. As expected, the events tend to be shorter (Fig.8b), because they are not amplified by the seasonal transition. There is a better agreement between EARI and OBS, but the models still tend to produce to many long lasting events.

Errors in the seasonal cycle cannot on their own explain the persistence of simulated events. However, the influence of the seasonal signal in the models is larger than in OBS or ERAI. For the models, the persistence of high temperatures may be partly due to an anomalous high seasonal range rather than by circulation anomalies, or a combination of both. There are also large uncertainties associated with both intra-model variability and differences between observations. These results also indicate that statistics on HW events are highly dependent on the choice of the index (absolute or anomalies), in accordance with You et al. (2016).

Next it will be investigated if the models can still reproduce the historical trends of the events despite their bias.

d. Evolution and trend of the heat waves

ERAI and OBS have a good agreement in terms of inter-annual evolution of HW events (Fig.9). They both have a clear decadal oscillation and an overall positive trend. Models tend to reproduce the positive trend, but the decadal oscillation is less clear (though it is still visible), especially for the N216 ensemble. A major transition occurs between the mid-90s and 2000, with a peak just after 2000. In the observations this transition is also visible, but in the models it is particularly sharp.

Fig.10 shows the same evolution but for long HW events only (more than 10 days). In the observations, the two peaks (corresponding to the few long events in OBS) are concurrent with the higher phases of the decadal signal. This indicates that the persistence of the events can be influenced by the decadal variability of temperatures. In the models, the signal is mostly controlled by the mid-90s transition, with most of the long HW occurring after this transition. This is also visible for the signal without running mean where the interannual variability is larger (Fig.10b). Two periods are clearly visible in the models (before 1995 and after 2000), with a transition between the two and a peak just after 2000. The influence of the interannual variations of the temperatures is tested furthermore. The HW events and their persistence are computed after removing the yearly summer mean temperatures from the signals without the seasonal climatology (as described in Section 3). Doing so, the persistence is reduced (not shown), though the impact is not as large as the seasonal signal. This indicates that the interannual variability of the temperatures can also influence the length of the HW events.

Finally, it is noticeable that both models and observations indicate a steady increase in the number of HW (days or events per year), even if models reproduce less clearly the observed decadal oscillations. This is not surprising given the ensemble averaging that tend to reduce the variability. When computing HW events after removing both the interannual summer means, the signal is more constant (Fig.10c) in the models. This clearly indicates that the trend in the models is mainly controlled by the trends in the mean temperatures, which is consistent with Freychet et al. (2017). An interesting difference between observations and models is the clear decadal oscillation still visible in ERAI and OBS but not in the models (though their signal tends to oscillate too). It may indicate again some missing chaos in the model ensemble due to averaging.

5. Risk and confidence in the models trends

Fig.11a,b compares the AMIP ensemble probability distribution for the number of HW days or events between the first and the last decade (1980-1990 and 1998-2008 respectively) of the investigated period. Both distributions shift to a higher number of days or events between the two periods, indicating an increased risk of HW days, but the ensemble spread is also large. Similar results are found for the N216 ensemble (Fig.11c,d). Interestingly, the most recent period (2009-2013) does not show a significant difference. Thus the major increase in the heat waves events occurred during the mid-90s transition. It may be due to a change in aerosols emission and transport during these years (and a high sensitivity of the models to these changes), but this hypothesis could be investigated in future work.

In previous sections, it has been shown that the signals in the AMIP simulations is often biased, especially in terms of the length of the events they produce. Thus, the reliability of the trend of heat waves in the AMIP (and the long-term forecasts) can be questioned. An approach to improve the confidence of the ensemble (and its projection) is to filter the best models based on their consistency with observations and reanalysis results. In the following, a filtering method is applied, based on the statistics of heat wave events (number of events or days). Two sources of error are considered: the observational error (estimated from the difference between ERAI and OBS) and the internal variability of the models (estimated with the N216 ensemble spread). This gives a margin of uncertainties within which the differences between a model and the observations can be considered as reasonable.

As the biases are observed on the number and the duration of HW events, two variables are considered to evaluate the models performance: the total number of heat wave days per years ($HW_{d/y}$) and the ratio of days included in long heat waves (more than 10 days) compared to the total number of heat wave days (HWL_{rat}). The reference values and associated uncertainties are computed using both OBS and ERAI, using the following formula:

$$\mu_{obs} = \frac{OBS + ERAI}{2} \tag{3}$$

$$\sigma_{obs} = |OBS - ERAI| \tag{4}$$

with μ_{obs} the mean, σ_{obs} the error and || symbols denoting the absolute value. In a similar way, the mean of a model (μ_{mod}) is computed by averaging, if necessary, the results from each of its members. The model error is estimated from the N216 ensemble (σ_{mod}), as it has the largest number of members, and corresponds to the standard deviation of the N216 ensemble. This error is particularly important as it is common to have only one member for a model, and thus a large uncertainty comes from the sampling process. As shown before, different members may have very different results, and thus the model cannot be evaluated correctly with a single member. A model results termed good when its difference with the observation is lower than the total error, i.e.:

$$Criteria : |\mu_{mod} - \mu_{obs}| \leqslant \sqrt{\sigma_{obs}^2 + \frac{\sigma_{mod}^2}{N}} \tag{5}$$

where N is the number of ensemble members. When several members are available for one model, only the ensemble mean is evaluated (and all members are considered retained or excluded based on the result on the ensemble mean). The criteria is verified for both variables ($HW_{d/y}$ and HWL_{rat}) and a model is termed good if it meets both criterion. The linear trends of the models is displayed in Fig.12. Even if the selection criteria is sharp many models are considered as good. However, the ensemble of good models does not show a significant difference compared to the ensemble mean of other models. Both groups indicate a positive trend, either in terms of events (about 0.25 events per decade) or days (about 2 HW days per decade). These results are consistent with ERAI and OBS, when considering the margin of error (ensemble scatter and difference between ERAI and OBS), especially in terms of HW events. The weaker trend in the observation may be related to a stronger decadal variability while the models have a more steady increase (Figure 9). Selecting only the best models does not significantly affect the results in this case, thus the results from the overall ensemble (in terms of trends) can be considered as reliable.

6. Concluding remarks

The representation of persistent large-scale heat waves over Central-Eastern China have been investigated in 40 AMIP members and compared with the results from ground stations and ERA Interim reanalysis. An ensemble of 15 members of the HadGEM3-A-N216 model was used to estimate the intra-model variability.

It was found that models tend to overestimate the number of heat wave days during the historical period, mostly because the events are too persistent. In the observations and reanalysis, the length of the events reaches a maximum of 12 or 9 days respectively, while in the models it can be more than 20 days.

Possible reasons to explain this bias were investigated: the magnitude of the summer range between the coldest and warmest temperatures, the climatology and the daily variability of the temperatures. None of these possible factors showed a significant relationship with the persistence of the heat waves, though it seems that the models are particularly sensitive to the seasonal signal. When investigating the decadal variability of the signals, it was found that most of the long heat waves occurs during the warmest periods. Thus, a possible explanation is that the heat wave signal in the models is more impacted by interannual to long term variability of the temperatures, while in the observations it is more sensitive to short term variations. It was also noticed that the large internal variability of the models could explain part of the long heat waves.

The circulation signal during heatwave events was verified with a composite analysis. The AMIP ensemble mean was consistent with reanalysis though individual members were less consistent. It was also verified that the composites of short heat waves (5-10 days) were consistent with the composites of all events, i.e. that the too persistent heat waves were not related to an incorrect dynamics. Finally, models were selected based on their heat wave length agreement with observations taking account of internal variability and observational error. These filtered models had similar trends in the number of heat waves and heat wave days as the other members of the ensemble. Thus the biases on the persistence of HW events do not affect significantly the trends, the later being mainly controlled by the interannual variability of the temperature. Thus, if a model can reproduce the mean change in the temperatures, it is expected that it can also reproduce the trends of the heat waves. Other dynamical factors (such as the jet streams, the Circumglobal teleconnections, the Western North Pacific High or the South Asia High) have been shown to influence the summer temperatures in China (e.g. Wang et al. 2013). We haven't investigated these processes in this study, thus they should be considered in the future as possible factors impacting heat waves in the models and eventually leading to biases in the persistence of the events.

Based on this study, the AMIP models were found reliable in terms of dynamics for the heat waves over Central-East China. Despite their tendencies to produce too persistent events, most of the AMIP members are able to reproduce the positive trends observed in both ground stations and reanalysis, and all results indicate an increase in the risk of such events during the past decades (from 4 events during the first decade to 8 events during the last decade). However, the long term trends in the models should be considered carefully due to some missing signals in the models (the decadal oscillation observed in ERAI and OBS). The mid-90s transition, especially clear in the models, should also be investigated in future work, as it raises the question of possible large scale impact of aerosols emissions. Finally, it is also noticeable that some uncertainties come from the difference between observation and reanalysis. Larger datasets, such an ensemble of reanalysis, could be used to improve the estimation of these uncertainties.

Using directly the raw temperature threshold is justified as it impacts human health. However the methodology used to define heat waves may lead to uncertain results. Indeed, the signal may result from a mix between the natural warming due to the seasonal transition and a warming due to weather type circulation anomaly. Thus the persistence of the event could be attributed to one or the other. Moreover, the use of a fixed threshold to identify the duration of an event can lead to sensitive statistics (as an event could be cut in too with one day in the middle just below the threshold for instance). Thus a final advise is that statistics on heat waves should always be carefully associated with a margin of error due to the methodology and definition, the data used and the sampling.

Acknowledgments. This work and all contributors were supported by the UK-China Research and Innovation Partnership Fund through the Met Office Climate Science for Service Partnership (CSSP) China as part of the Newton Fund. The author(s) wish to acknowledge use of the Ferret program for analysis and graphics in this paper. Ferret is a product of NOAA's Pacific Marine Environmental Laboratory.

APPENDIX

a. Computation of composites

The composite of X at a day d (X_d) of a specific year (*ann*) is given by equation 1. The corresponding daily climatology (X_{d-clim}) of the variable is first removed (Fig.A1). The difference between the annual mean of the year *ann* and the climatology (annual mean, X_{clim}) is also removed from the composite. This method removes any long term trend effect (for instance, an elevation of the geopotential height due to a global temperature warming) and only highlight the differences due to short terms anomalies.

References

Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quaterly J. of the Roy. Met. Soc.*, **137(656)**, 553–597.

Ding, T., and W.-H. Qian, 2011: Geographical patterns and temporal variations of regional dry and wet heatwave events in China during 1960-2008. *Adv. in Atm. Sc.*, **28**, 322–337.

Dong, D., and G. Huang, 2015: Relationship between altitude and variation characteristics of the maximum, minimum temperature and diurnal temperature range in China. *Chinese J. of Atm. Sc.*, **39**, 1011–1024.

Freychet, N., H.-H. Hsu, C. Chou, and C.-H. Wu, 2015: Asian summer monsoon in CMIP5 projections: A link between the change in extreme precipitation and monsoon dynamics. *J. Climate*, **28**, 1477–1493.

Freychet, N., H. H. Hsu, and C.-H. Wu, 2016: Extreme precipitation events over East Asia: Evaluating the cmip5 model. *Atmospheric Hazards - Case Studies in Modeling, Communication, and Societal Impacts*, J. S. M. Coleman, Ed., InTech, Rijeka, chap. 05, doi:10.5772/62996, URL http://dx.doi.org/10.5772/62996.

Freychet, N., S. F. B. Tett, J. Wang, and G. C. Hegerl, 2017: Summer heat waves over Eastern China: dynamical processes and trend attribution. *Env. Res. Lett.*, **12**, 024 015.

Gong, D. Y., and S. W. Wang, 2002: Uncertainties in the global warming studies [in chinese]. *Earth Sci. Front.*, **9(2)**, 371–376.

Guo, X., J. Huang, Y. Luo, Z. Zhao, and Y. Xu, 2017: Projection of heat waves over China for eight different global warming targets using 12 CMIP5 models. *Theor. and App. Clim.*, **128(3-4)**, 507–522.

Heisler, G. M., and A. J. Brazel, 2010: *The urban physical environment: Temperature and urban heat islands*. Am. Soc. of Agron., Crop Sci. Soc. of Am., Soil Sci. Soc. of Am.

Kripalani, R. H., J.-H. Oh, and H. S. Chaudhari, 2007: Response of the East Asian summer monsoon to doubled atmospheric CO2: Coupled climate model simulations and projections under IPCC AR4. *Theor. Appl. Climatol.*, **87**, 1–28.

Li, X., L. Wang, K. Yang, B. Xue, and L. Sun, 2013: Near-surface air temperature lapse rates in the mainland China during 19622011. *J. Geophys. Res. Atmos.*, **118**, 7505–7515, doi:10.1002/jgrd.50553.

Li, Z., and Z.-W. Yan, 2009: Homogenized daily mean/maximum/minimum temperatures series for China from 1960-2008. *Atm. and Oc. Sc. Let.*, **2(4)**, 237–243.

Lin, C., and Coauthors, 2015: Impact of wind stilling on solar radiation variability in China. *Nature Sc. Reports*, **5(15135)**, doi:10.1038/srep15135.

Lu, R.-Y., and R.-D. Chen, 2016: A review of recent studies on extreme heat in China. *Atm. and Ocean. Sc. Lett.*, **9(2)**, 114–121.

Luber, G., and M. McGeehin, 2008: Climate change and extreme heat events. *Am. J. of Prev. Med.*, **35**, 429–435.

Luo, M., and N.-G. Lau, 2017: Heat waves in southern china: Synoptic behavior, long-term change, and urbanisation effects. *J. of Clim.*, **30(2)**, 703–720.

Perkins, S. E., 2015: A review on the scientific understanding of heatwavestheir measurement, driving mechanisms, and changes at the global scale. *Atmos. Res.*, **164**, 242–67.

Ren, G.-Y., M.-Z. Xu, Z.-Y. Chu, A.-Y. Zhang, J. Guo, H.-Z. Bai, and X.-F. Liu, 2005: Changes of surface air temperature in China during 1951-2004. *J. Geophys. Res.*, **10(4)**, 717–727.

Ren, Y.-Y., D. Parker, G.-Y. Ren, and R. Dunn, 2016: Tempo-spatial characteristics of sub-daily temperature trends in mainland China. *Clim. Dyn.*, **46**, 2737–2748.

Sun, Y., X. Zhang, F. W. Zwiers, L. Song, H. Wan, T. Hu, H. Yin, and G. Ren, 2014: Rapid increase in the risk of extreme summer heat in eastern china. *Nature Clim. Change*, **4(12)**, 1082–1085.

Walters, D., and Coauthors, 2017: The Met Office Unified Model Global Atmosphere 6.0/6.1 and JULES Global Land 6.0/6.1 configurations. *Geosci. Model Dev.*, **10**, 1487–1520, doi:10.5194/gmd-10-1487-2017.

Wang, A., and J. Fu, 2013: Changes in daily climate extremes of observed temperature and precipitation in China. *Atm. and Oc. Sc. Lett.*, **6**, 312–319.

Wang, S., X. Yuan, and Y. Li, 2016a: Does a strong el nio imply a higher predictability of extreme drought? *Nat. Scient. Rep.*, **7**, 40 741.

Wang, W., W. Zhou, X. Li, X. Wang, and D. Wang, 2016b: Synoptic-scale characteristics and atmospheric controls of summer heat waves in China. *Clim. Dyn.*, **46(9-10)**, 2923–2941.

Wang, W., W. Zhou, Y. Li, X. Wang, and D. Wang, 2015: Statistical modeling and cmip5 simulations of hot spell changes in china. *Clim. Dyn.*, **44**, 28592872.

Wang, W. W., W. Zhou, X. Wang, S. K. Fong, and K. C. Leung, 2013: Summer high temperature extremes in southeast China associated with the East Asian jet stream and circumglobal teleconnection. *J. Geophys. Res.*, **118**, 8306–8319.

Wei, K., and W. Chen, 2011: An abrupt increase in the summer high temperature extreme days across china in the mid-1990s. *Adv. Atmo. Sc.*, **28**, 1023–1029.

Wu, Z., Z. Jiang, J. Li, S. Zhong, and L. Wang, 2012: Possible association of the western Tibetan Plateau snow cover with the decadal to interdecadal variations of northern China heatwave frequency. *Clim. Dyn.*, **39**, 2393–2402.

Yan, X., Y. Hou, and B. Chen, 2011a: Observed surface warming induced by urbanization in East China. *J. of Geop. R.: Atm.*, **116**, 2156–2202, doi:10.1029/2010JD015452.

Yan, Z., J. Xia, C. Qian, and W. Zhou, 2011b: Changes in seasonal cycle and extremes in China during the period 1960-2008. *Adv. in Atm. Sc.*, **28(2)**, 269–283.

You, Q., Z. Jiang, L. Kong, Z. Wu, Y. Bao, S. Kang, and N. Pepin, 2016: A comparison of heat wave climatologies and trends in china based on multiple definitions. *Clim. Dyn.*, **48**, 39753989.

Zhou, C.-L., and K.-C. Wang, 2016: Coldest temperature extreme monotonically increased and hottest extreme oscillated over Northern hemisphere land during last 114 years. *Nature Scient. Report*, **6:24721**.

LIST OF FIGURES

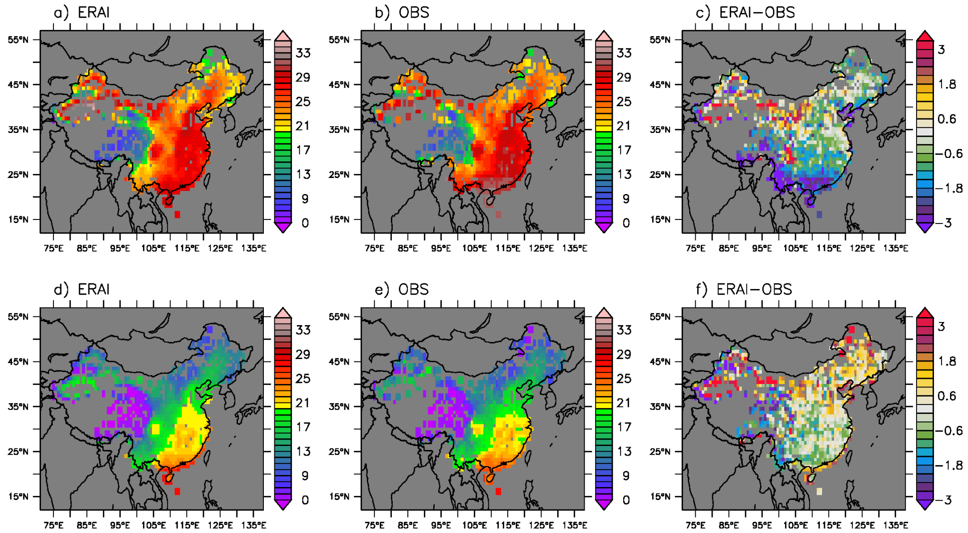


FIG. 1. Summer mean Tmax and Tmin (°C) for ERAI, OBS (corrected by the difference of elevation with ERAI at each point) projected on ERAI grid and difference between the two datasets (ERAI-OBS). All datasets have been masked where no ground station data were available. (a-c) show Tmax and (d-f) show Tmin.

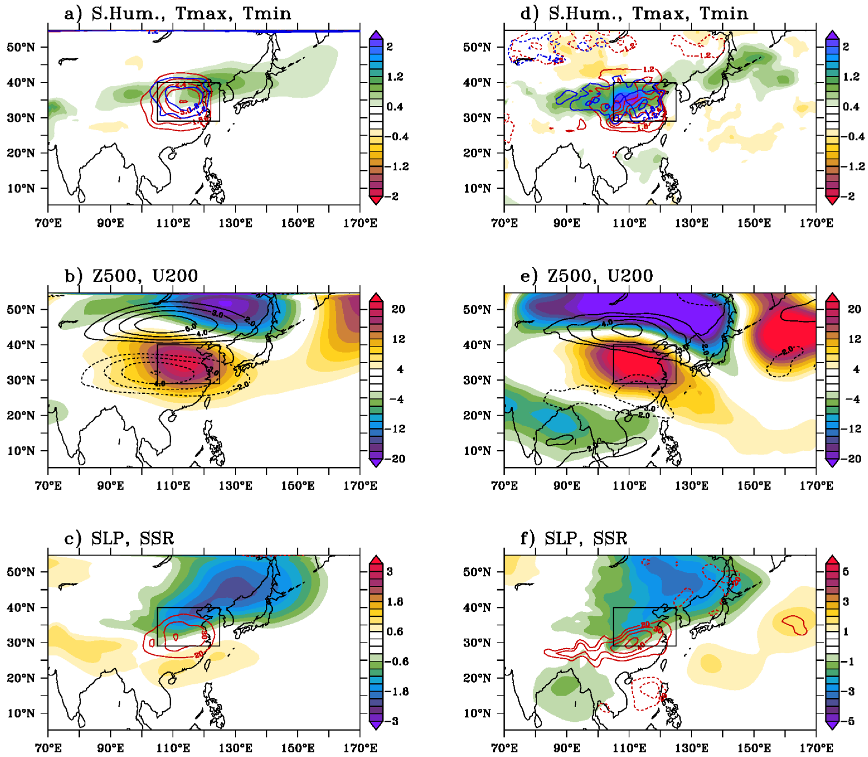


FIG. 2. Composite of the dynamics during the HW events from the AMIP ensemble mean (left) and ERAI (right). The variables displayed are: (a,d) specific humidity (S.Hum., shading, g.kg^{-1}), maximum temperature (Tmax, red contours, °C), minimum temperature (Tmin, blue contours, °C), (b,d) 500 hPa geopotential height (Z500, shading, m), 200 hPa zonal wind (U200, black contours, m.s^{-1}), (c,f) sea level pressure (SLP, shading, hPa) and surface shortwaves radiation (SSR, red contours, W.m^{-2}). The black box indicates the Central-Eastern China region.

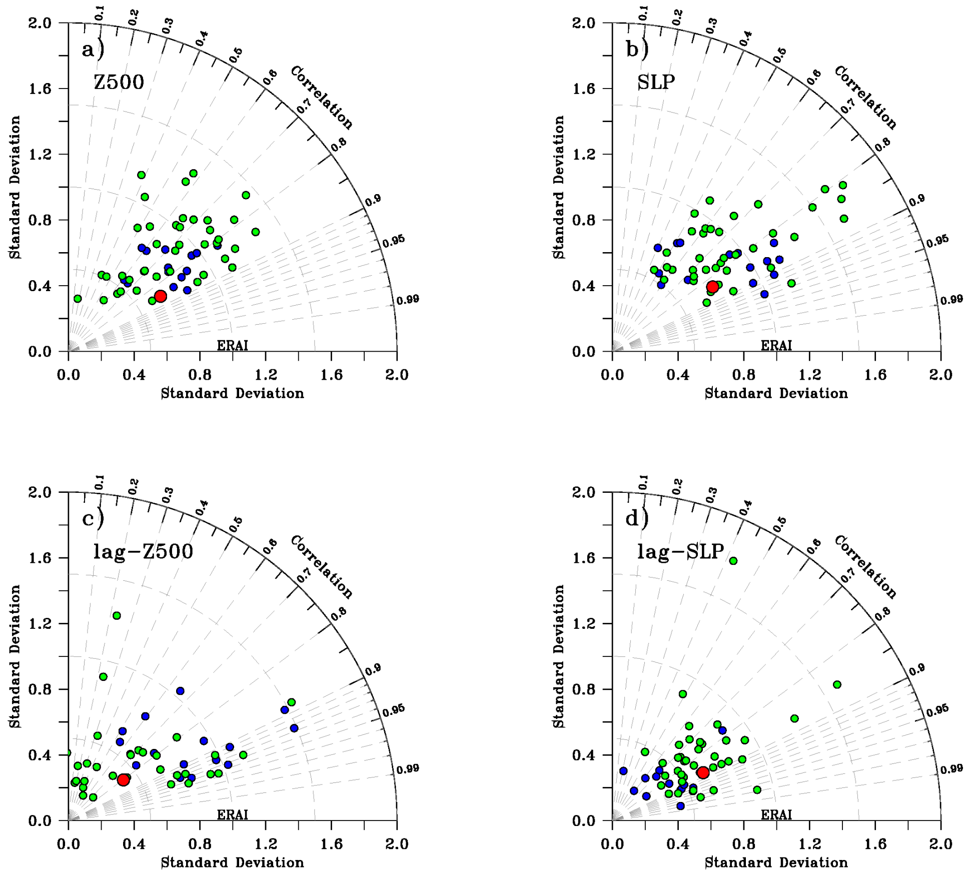


FIG. 3. Taylor diagrams for Z500 (a) and SLP (b) spatial patterns (using the 95E-155E, 20N-55N region), for each AMIP member (green circles) and N216 member (blue circles). The red circle indicates the AMIP ensemble mean, and the reference is ERAI. (c,d) are the same but for the lag-composites of Z500 and SLP (see text Section 3 for methodology).

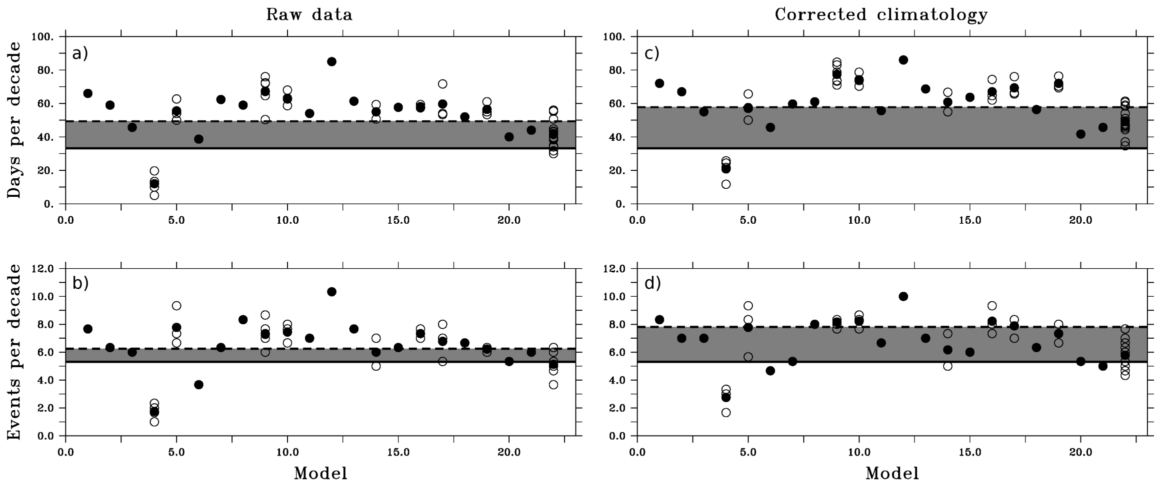


FIG. 4. HW days (a,c) and HW events (b,d) per decade, for each members (empty circles) and ensemble mean for each model (full black circle). The last model on the right of each plot is the N216 ensemble. The horizontal solid black line is ERAI and the dashed black line is OBS. The grey shading between the two indicates observational uncertainty. (a,b) are results from the raw data while (c,d) are results obtained after correcting the seasonal climatology (see text for description).

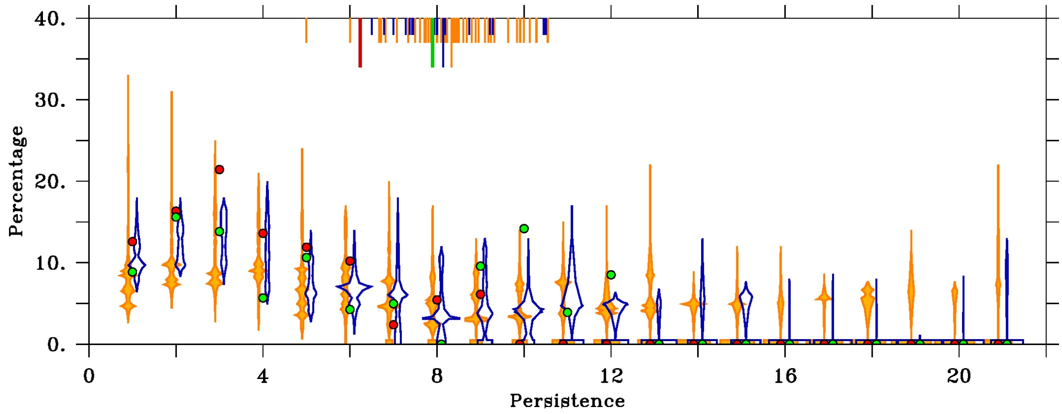


FIG. 5. Percentage of days (y axis) as a function of warm day persistence (x axis, number of days). AMIP and N216 members are represented by orange and blue density diagrams respectively. Red circles show ERAI results, and green circles are OBS. See text Section 4b for more details. The coloured tics on the top axis indicate the mean duration of HW events (more than 5 days) for ERAI (red), OBS (green), each member (short tics) and ensemble mean (long tics) of AMIP (orange) and N216 (blue).

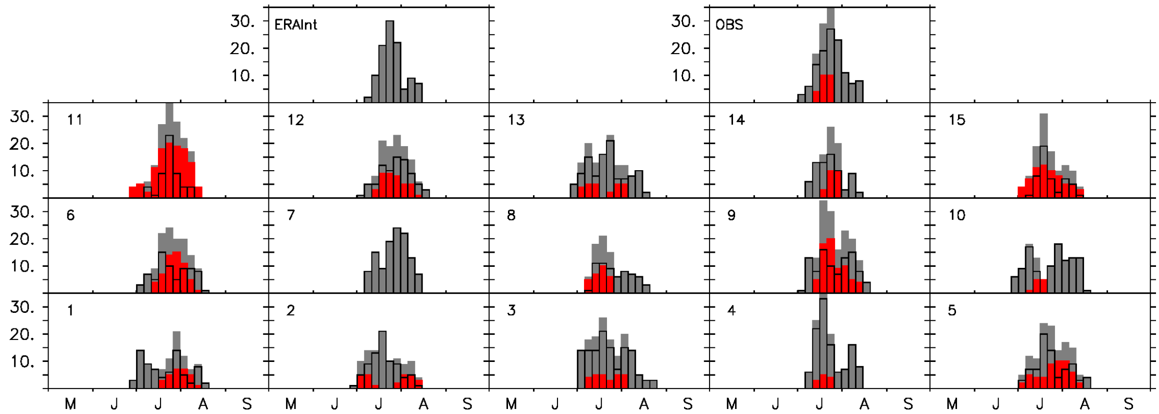


FIG. 6. Sum of all HW days (during the 1979-2008 period), grouped by climatological pentad, for each N216 members. Gray bars indicate the total number of days, red bars are the days corresponding to long lasting HW events (more than 10 days) and the black contour bars are the days corresponding to short HW events (5 to 10 days). On the top of the figures, results from ERAI and OBS are displayed on the left and the right respectively.

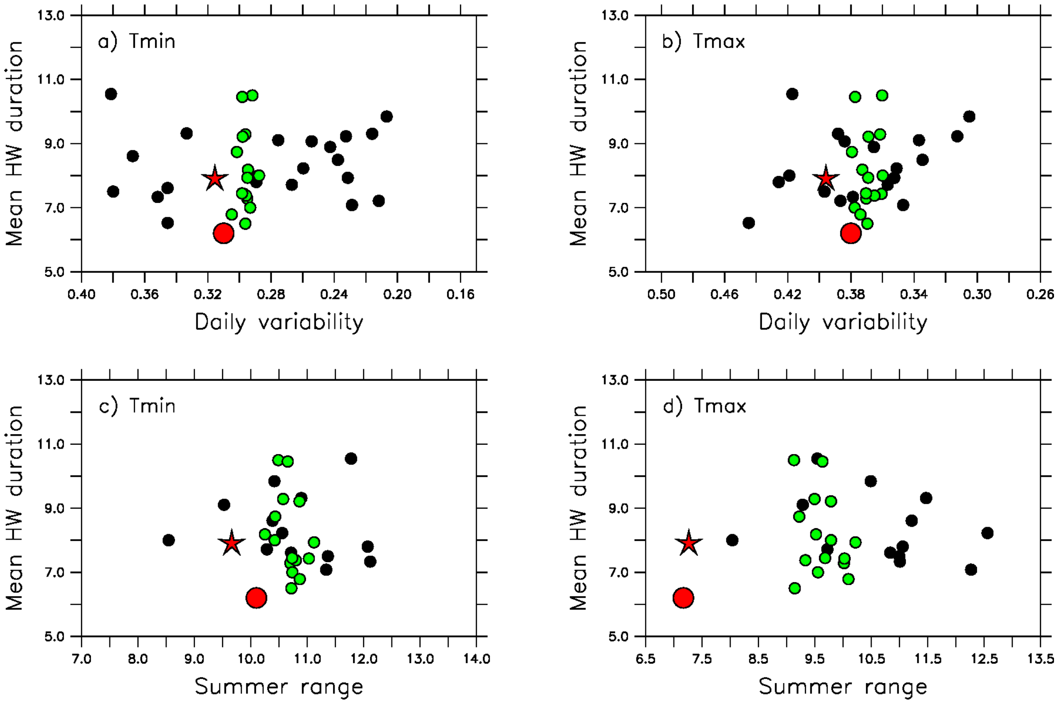


FIG. 7. Mean duration of HW events (days) for each AMIP model (ensemble mean of each model, black circles) and N216 member (grey circles), versus the daily variability (a,b) and the summer range (c,d). The red circle and star indicates results from ERAI and OBS respectively. See text Section 4c for the definition of the summer range and daily variability.

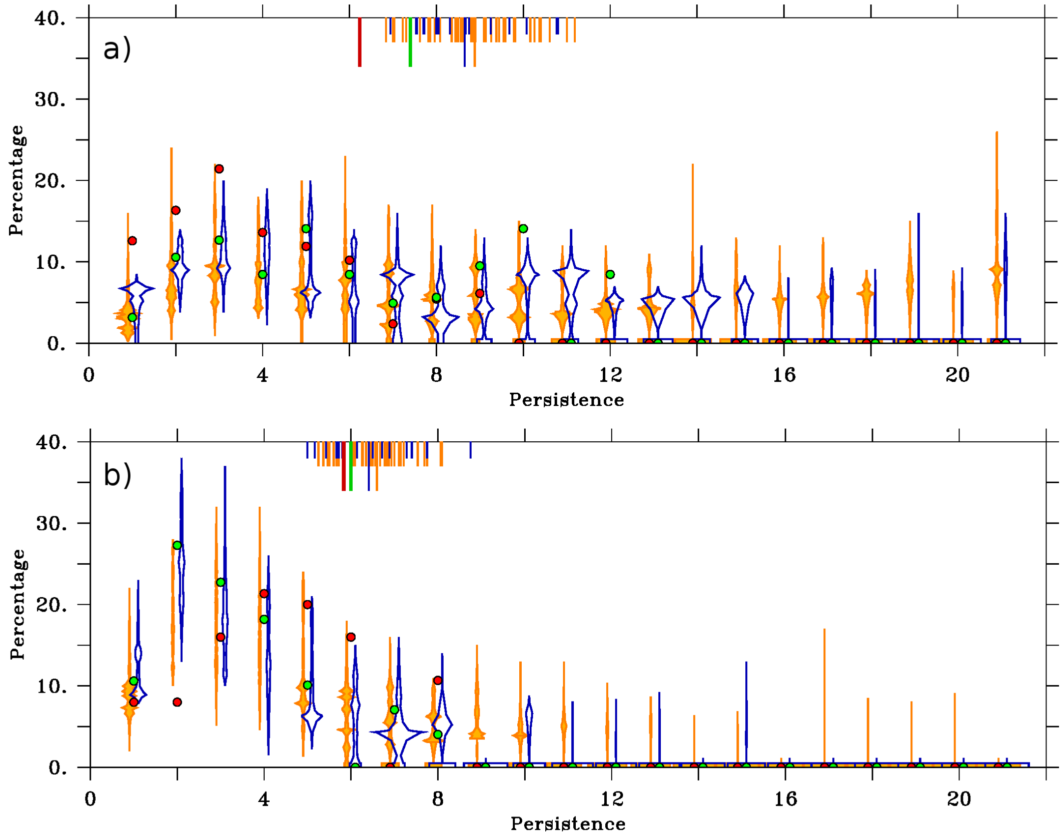


FIG. 8. As Fig.5 but based on data after correcting (a) or removing (b) the seasonal climatology.

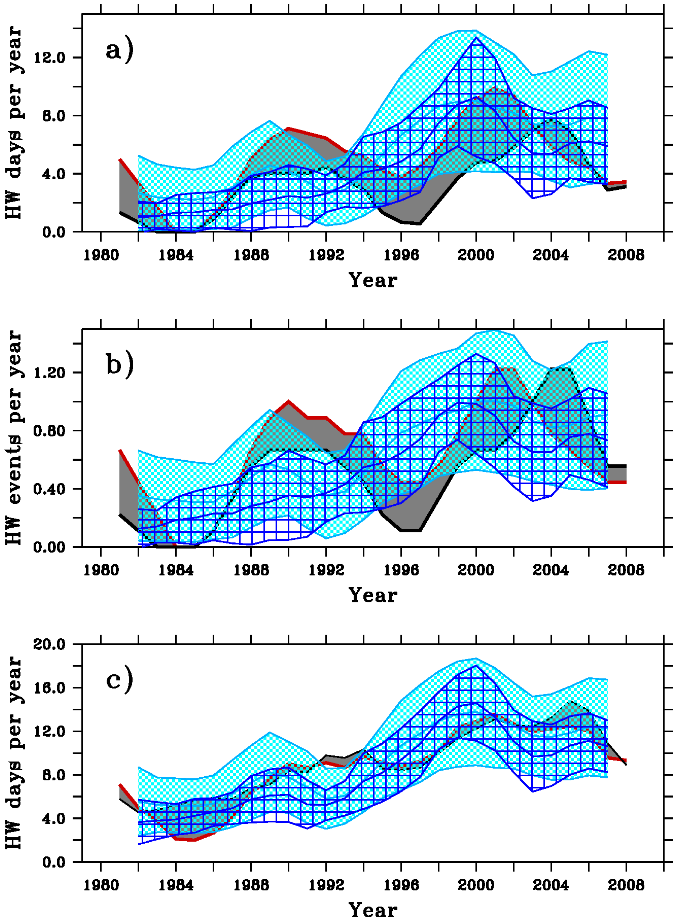
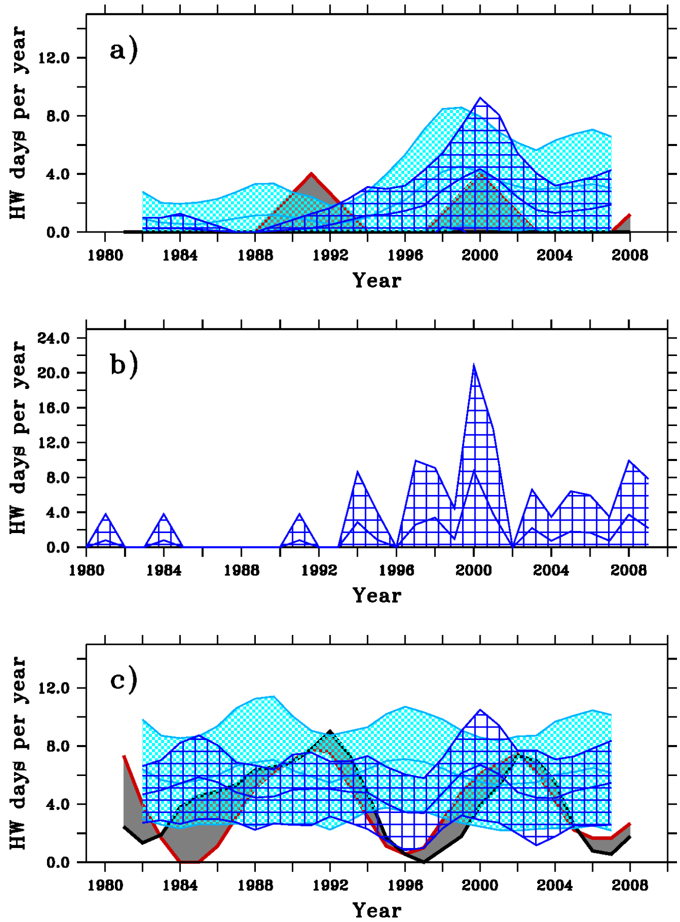


FIG. 9. Evolution of the annual number of HW days (a), HW events (b) and warm days (c), with a 5-year running mean. Solid black and red lines are ERAI and OBS respectively, and the gray shading indicates uncertainty between the two. Light blue is the AMIP ensemble mean (line in the middle) and standard deviation. Dark blue checked is the N216 ensemble mean and standard deviation.



607 FIG. 10. (a) As Fig.9a but for the long HW events only (more than 10 days). (b) Long HW signal in N216 ensemble without the 5-years running mean smoothing. (c) As Fig.9a but for HW events computed after removing the interannual summer means from the temperatures (text Sections d).

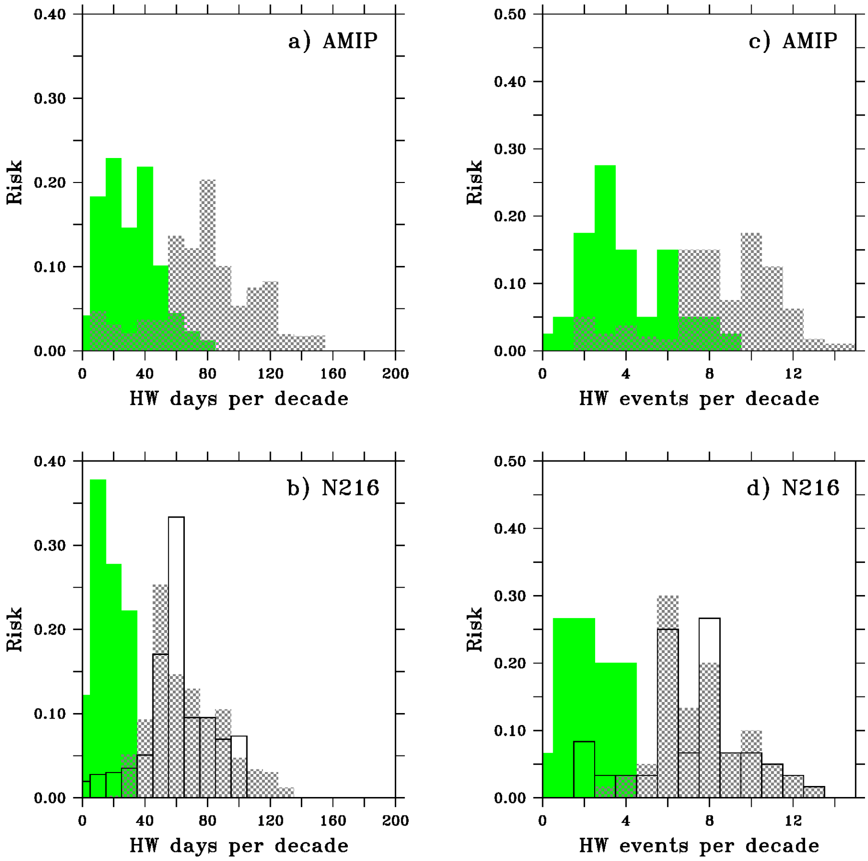


FIG. 11. Probability density function of the number of heat wave days or events during the 1980-1990 period (filled green bars) and the 1998-2008 period (grey bars), for AMIP (a,b) and N216 (c,d). 2009-2013 is also added for the N216 results (black contours).

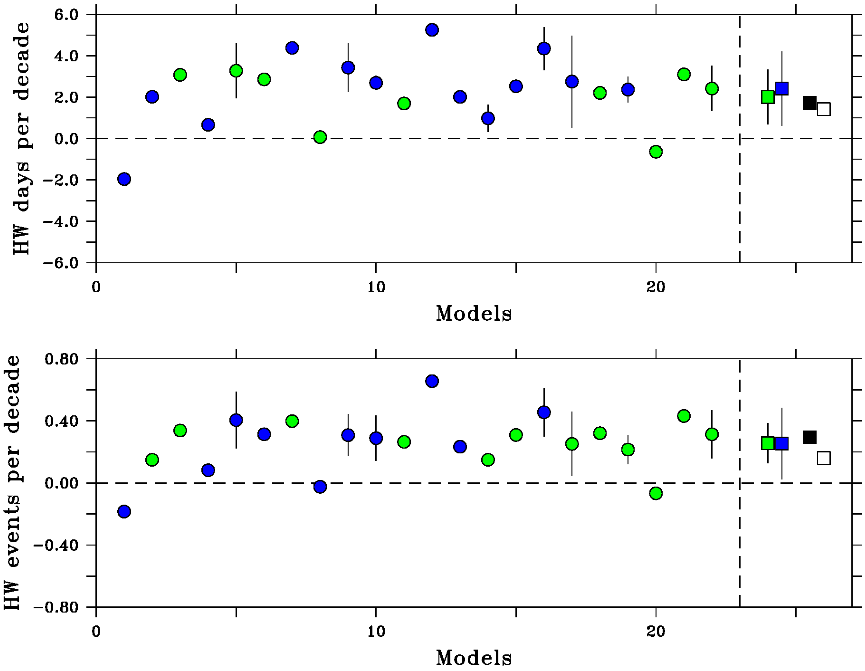
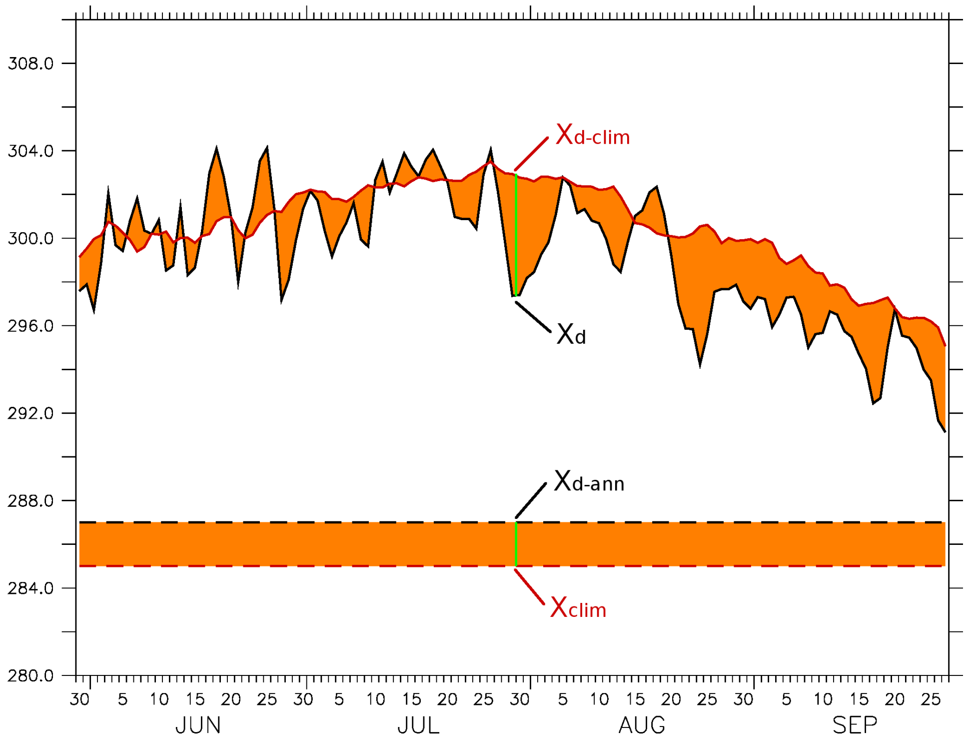


FIG. 12. Linear trends (Y axis) of the numbers of heat waves days (a) and events (b) per decade for each AMIP model (X axis) model mean (circles) and standard deviation from multi-members models (black bars). The N216 ensemble is indicated as model number 22. Green (blue) colour indicates the models considered as good (bad) by the filtering method (see text Section 5), and the ensemble means (and dispersions) of the two groups are shown by the green and blue square (and black bars). ERAI and OBS are shown with white and black squares respectively.



619 Fig. A1. Schematic representation of a composite computation (see text Section 2 and Appendix). The solid black line is the daily time serie of a variable X, the solid red line is its daily climatology and the orange shading represents the difference between the two. The dashed black line represent the annual mean of X and the dashed red line is the annual climatology (and the difference is highlighted by the orange shading).