Received: 12 November 2020 | Revised: 5 March 2021 | Accepted: 3 April 2021 | Published on: 4 May 2021

DOI: 10.1002/joc.7130

RESEARCH ARTICLE

International Journal of Climatology RMetS

A wavelet transform method to determine monsoon onset and retreat from precipitation time-series

Jorge L. García-Franco[1] | **Scott Osprey**[1,2] | **Lesley J. Gray**[1,2]

[1]Atmospheric, Oceanic, and Planetary Physics, Department of Physics, University of Oxford, Oxford, UK

[2]National Centre for Atmospheric Science, Leeds, UK

Correspondence
Jorge L. García-Franco, Clarendon Laboratory, Parks Road, Oxford, United Kingdom OX1 3PU.
Email: jorge.garcia-franco@physics.ox.ac.uk

Funding information
Natural Environment Research Council, Grant/Award Number: NE/P006779/1; NERC North Atlantic Climate System Integrated Study (ACSIS), Grant/Award Number: NE/N018028

Abstract

A new method to determine monsoon onset and retreat timings using wavelet transform methodology applied to precipitation time-series at the pentad scale is described. The principal advantage of this method is its portability, since it can be easily adapted for any region and dataset. The application of the method is illustrated for the North American Monsoon and the Indian Monsoon using four different precipitation datasets and climate model output. The method is shown to be robust across all the datasets and both monsoon regions. The mean onset and retreat dates agree well with previous methods. Spatial distributions of the precipitation and circulation anomalies identified around the onset and retreat dates are also consistent with previous work and illustrate that this method may be used at the grid-box scale, not just over large area-averaged regions. The method is also used to characterize the strength and timing of the Midsummer drought (MSD) in southern Mexico and Central America. A two peak structure is found to be a robust structure in only in 33% of the years, with other years showing only one peak or no signs of a bimodal distribution. The two-peak structure analysed at the grid-box scale is shown to be a significant signal in several regions of Central America and southern Mexico. The methodology is also applied to climate model output from the Met Office Hadley Centre UKESM1 and HadGEM3 CMIP6 experiments. The modelled onset and retreat dates agree well with observations in the North American Monsoon but not in the Indian Monsoon. The start and end of the modelled MSD in southern Mexico and Central America is delayed by one pentad and has a stronger bimodal signal than observed.

KEYWORDS

atmosphere, climate, geographic/climatic zone, geophysical sphere, monsson, observational data analysis, physical phenomenon, rainfall, scale, tools and methods

Abbreviations: ABC, a black cat; DEF, does not ever fret; GHI, goes home immediately.

Jorge L. García-Franco and Scott Osprey contributed equally to this study.

1 | INTRODUCTION

Monsoon regions are characterized by a strong contrast between the rainy and the dry seasons, with at least 70% of the total annual rainfall observed during summer in a monsoon (Zhou *et al*., 2016; Wang *et al*., 2017). The start and end of the wet season, defined by onset and retreat (or demise) dates, are physically associated with sudden changes in the moisture transport driven by the changes to the circulation. Around the onset and retreat dates, changes in the predominant wind direction or strength also cause abrupt changes in meteorological conditions such as rainfall, humidity and cloud cover (Zhou *et al*., 2016; Gadgil, 2018). A current view is that monsoons can be defined based solely on the contribution of summer rainfall to the annual total highlighting the importance of rainfall to several societal sectors (Zhang and Wang, 2008; Wang *et al*., 2017).

The timing and strength of the rainy season are key aspects of the seasonal cycle of monsoon regions such that the length of the rainy season greatly influences sectors such as agriculture (Sultan *et al*., 2005; Gadgil and Rupa Kumar, 2006; Jain *et al*., 2015; Harvey *et al*., 2018) and water management (Turner and Annamalai, 2012; Bussmann *et al*., 2016). The objective determination of onset and retreat dates is key for climate and weather research aiming to understand the short and long-term variability, trends and predictability of the regional monsoons (e.g., Kitoh and Uchiyama, 2006; Cook and Buckley, 2009; Htway and Matsumoto, 2011; Lucas-Picher *et al*., 2011; Nieto-Ferreira and Rickenbach, 2011).

For this reason, a plethora of methods have been used to diagnose the onset and retreat dates from a range of variables and datasets. Bombardi *et al*. (2020) provides a recent review of methods for objectively defining monsoon onset and retreat dates and highlighted the technical differences and purposes of each. Methods can be divided into those that aim to evaluate monsoon onset and retreat on a regional scale (e.g., Webster and Yang, 1992; Fasullo and Webster, 2003; Garcia and Kayano, 2013) or at a local or grid-box scale (e.g., Liebmann and Marengo, 2001; Cook and Buckley, 2009).

Threshold methods are the most commonly used local-scale methods that typically diagnose onset and retreat from a precipitation time-series (Bombardi *et al*., 2020). These methods evaluate the accumulated (Liebmann and Marengo, 2001) or daily/pentad-mean rainfall rates (Geil *et al*., 2013) and determine the onset and retreat dates when the time-series exceeds or falls below a pre-defined value (threshold) for a given amount of time (persistence). The persistence parameter is used to reduce the method's sensitivity to noise in the precipitation time-series. The threshold parameter can be a statistical measure of the seasonal cycle such as the total annual mean rainfall (Arias *et al*., 2012) but the threshold and persistence parameters can also be subjectively determined. The parameters for threshold methods vary distinctly from study to study for several reasons. First, the parameters vary from one monsoon region to another because the wet and dry seasons in each region have different timings, strengths and dynamical features (Wang *et al*., 2017). For instance, the North American Monsoon occurs over a much shorter time than the South American Monsoon and is smaller geographically (Vera *et al*., 2006). Within a given monsoon region, for example, in the South American Monsoon, several methods are used for different purposes depending on the temporal and spatial scales of interest (see, for example, Liebmann and Marengo, 2001; Marengo *et al*., 2001; Nieto-Ferreira and Rickenbach, 2011; Carvalho *et al*., 2011; Garcia and Kayano, 2013). The differences between the characteristics of the datasets, for example, in horizontal resolution, relate to differences in the seasonal cycle of rainfall which means that the implementation of threshold methods in different datasets also requires normalization or statistical treatment of the threshold and persistence parameters.

In other words, each threshold method is tailored to a monsoon region using a specific dataset and a specific variable for a given purpose. This characteristic of the threshold methods poses various shortcomings. Firstly, practical shortcomings of the threshold methods, particularly rigid thresholds, include false hits (Moron and Robertson, 2014) or some years not meeting the threshold and persistence criteria (Arias *et al*., 2012) requiring further relaxation of the parameters. Second, the conclusions of most analyses of threshold methods arise from the output of one single-method and one single dataset. These studies would benefit from the input of a second independent method to confirm the results. However, due to the differences between regional monsoons most techniques are not suitable to be used in other regions.

The Coupled Model Intercomparison Project (CMIP) assessments of monsoon onset and retreat typically use precipitation threshold methods due to the lack of high temporal or vertical resolution output from all models to estimate vertically integrated quantities need for some methods (e.g., Geil *et al*., 2013; Zou and Zhou, 2015; Ha *et al*., 2020). Threshold methods have multiple shortcomings for CMIP assessments as the persistence and threshold parameters are tuned for observations with a specific seasonal cycle but models have a range of biases in the seasonality, magnitude and spatial distribution of rainfall (Pascale *et al*., 2019; García-Franco *et al*., 2020). The use of pre-defined threshold values may also not be suitable to compare different model experiments with changes in forcing where the climatological mean rainfall or the

seasonal cycle may change within the model run. These shortcomings are relevant because a proper diagnosis of the seasonal cycle in CMIP assessments is key to understand and diagnose current and future changes to monsoon seasonality as a result of greenhouse forcing (Zhou *et al*., 2016; Wang *et al*., 2017).

The observational or model-based analysis of the monsoons as a global phenomena also require a systematic approach to determine onset and retreat dates that is not tailored to any specific region (Zeng and Lu, 2004). For example, projects such as the Global Monsoons Model Inter-comparison Project (GMMIP) which is part of CMIP6 may benefit from broader and more systematic methods to be used across models and monsoon regions (Zhou *et al*., 2016) and particularly methods that can be implemented with standard output from modelling centres. These systematic or large-scale methods may also be relevant in the evaluation of trends in the timings of the rainy season on hemispheric or global scales in observations and models (Zeng and Lu, 2004; Zhang and Wang, 2008).

The objective diagnosis of shorter time-scale rainfall variability, such as bimodal regimes and active and break phases of a monsoon, also requires methods that can separate relatively drier and wetter periods within the rainy season. For instance, the characteristics of the seasonal cycle of precipitation in northwestern Central America and southern Mexico fit the definition of a monsoon climate (Wang *et al*., 2017), characterized by the majority of precipitation observed during local summer, and for this reason this region is sometimes considered to be part of the American monsoon (Vera *et al*., 2006; Pascale *et al*., 2019). However, this region shows a unique climatological bimodal signal (Karnauskas *et al*., 2013). After monsoon onset, rainfall decreases considerably around the midsummer; this decrease is followed by a secondary increase in precipitation in the late summer (Mosiño and García, 1966; Magaña *et al*., 1999).

Several studies have documented the characteristics of this bimodal signal, most commonly referred to as the Midsummer drought (MSD; e.g., Mosiño and García, 1966; Karnauskas *et al*., 2013; Perdigón-Morales *et al*., 2018) and have linked the cause of the bimodal distribution of rainfall to regional changes in the circulation and the moisture transport (Magaña *et al*., 1999; Herrera *et al*., 2015). However, the objective determination of the strength, spatial distribution and robustness of bimodal signals is not straightforward. For example, the global method used in Bombardi *et al*. (2020) fails to diagnose the region of southern Mexico, Central America and the Caribbean as a bi-modal regime.

The majority of existing methods to diagnose bimodal signals in this region use geometric or statistical measures of the monthly mean rainfall that measure the difference between the months of maximum rainfall and the drier months. However, this approach fails to capture the shorter-scale changes that have been shown to occur in both observations and model data, as the MSD does not start or end exactly on given calendar months (Magaña *et al*., 1999; García-Franco *et al*., 2020). Zhao and Zhang (2020) review and compare several methods to detect and measure the MSD, finding that using monthly-mean data and prior assumptions of the dates of the first and second peaks can lead to errors. Only a handful of methods exist that can determine the characteristics of the MSD in Central America on sub-monthly timescales. Anderson *et al*. (2019) analysed the pentad-mean time-series from the Climate Hazards Infrared Precipitation with Stations (CHIRPS) dataset. After a double temporal smoothing of the time-series, Anderson *et al*. (2019) determined the MSD timings through a threshold method with parameters tailored to the CHIRPS dataset. In turn, Zhao *et al*. (2020) used daily mean time-series and determined the two-peak structure through linear-regression analysis.

In short, research on monsoon regions requires objective methods to determine the timings of onset, retreat and short-scale changes within the rainy season such as the MSD and active and break phases. Multiple methods exist, each with various parameters fit for different purposes, but these methods present shortcomings for studies that compare results from multiple datasets or investigate model experiments where climatological rainfall and the seasonal cycle change. Similarly, studies that investigate the impact of decadal modes of variability (e.g., Arias *et al*., 2012), greenhouse warming (e.g., Geil *et al*., 2013) or general trends (e.g., Sa*hana et* al., 2015) rely solely on the output of one single method whereas the use of two or more methods may help to test the sensitivity of their results to the chosen method and parameters. Both the objective determination of monsoon onset and retreat and the timings of bimodal regimes require a method that can analyse temporal changes to precipitation on several scales and that can be used on any gridded dataset.

The goal of this study is then to present an objective approach that is more portable across datasets, regions and robust for various purposes. This article introduces a wavelet transform method to determine monsoon onset and retreat dates using pentad-mean precipitation time-series. Wavelet algorithms have been extensively used in atmospheric research for multiple purposes, such as the detection of the boundary layer height (e.g., Brooks, 2003), as well as to analyse time-frequency features of a signal (e.g., Whitcher *et al*., 2000; Dimdore-Miles *et al*., 2021). In fact, Allen and Mapes (2017) used a wavelet analysis to determine monsoon onset and retreat using daily OLR data. The method is constructed such that the determination of monsoon onset and withdrawal dates is less

sensitive to the characteristics of the time-series, that is, the characteristics of the seasonality of each monsoon or of a given observational dataset. Furthermore, the method is expanded to characterize bimodal regimes which is illustrated for the MSD of southern Mexico, Central America and the Caribbean.

The remainder of this article is organized as follows: Section 2 describes the methods and datasets. Section 3 shows the results of applying the method to the Indian and North American Monsoons and the MSD. Section 4 summarizes the method and discusses the results.

2 | DATA AND METHODS

2.1 | Data

2.1.1 | Observations and reanalysis data

Table 1 summarizes the datasets and variables used in this study. We use three gridded precipitation datasets, the Tropical Rainfall Measurement Mission (TRMM) v7 3B42, Climate Hazards Infrared Precipitation with Stations (CHIRPS) and the Climate Prediction Center Merged Analysis of Precipitation (CMAP). These three precipitation datasets are merged products, TRMM and CMAP mainly use microwave satellite measurements complimented by several other sensors and calibrated with rain-gauge data whereas CHIRPS uses several products from TRMM, as well as high-resolution station data. These datasets also differ in their end-product horizontal resolutions.

The precipitation output from the latest reanalysis from the European Centre for Medium-Range Forecasting: ERA-5, which has been shown to exhibit a relatively good representation of the temporal characteristics of rainfall in monsoon regions (e.g., García-Franco *et al*., 2020). Other variables from ERA5 used to diagnose changes to the circulation associated with monsoon onset were geopotential height at 500 hPa and wind speed ($\vec{u}$).

2.1.2 | Model data

Daily precipitation data from the CMIP6 archive are used, retrieved from: https://esgf-index1.ceda.ac.uk/projects/cmip6-ceda/, to illustrate the method using standard climate model output. In particular, we use results from the Met Office Hadley Centre models (MOHC): HadGEM3 GC3.1 and UKESM1 and we have chosen the pre-industrial control and historical experiments (Menary *et al*., 2018; Sellar *et al*., 2019; Andrews *et al*., 2020) which are amongst the most commonly used experiments in CMIP6. The daily precipitation data were converted to pentad-scales.

Table S1 provides a summary of the experiments used and the acronyms of each experiment for each model (Menary *et al*., 2018; Ridley *et al*., 2018, 2019; Sellar *et al*., 2019). HadGEM3 GC3.1 was run at two horizontal resolutions, N96 and N216 for the pre-industrial control experiment (Table S1). UKESM1 differs from HadGEM3 by representing Earth System processes, such as interactive ocean biogeochemistry and atmospheric chemical interactions (Sellar *et al*., 2019). The pre-industrial Control experiments use constant forcing estimated for 1850 whereas the historical experiments aim to represent time-varying greenhouse emissions, volcanic eruption and solar signals in the historical period (1850–2014) Eyring *et al*. (2016). These simulations were recently evaluated in the North American Monsoon and the MSD region, showing a good representation of the seasonal cycle in these regions (García-Franco *et al*., 2020).

2.2 | Methods

2.2.1 | Wavelet transform method

Wavelets are band-limited wave-like functions with specific mathematical properties that include a finite energy and zero-mean (Whitcher *et al*., 2000; Addison, 2017).

TABLE 1 Summary of the datasets used in this study

Dataset/version	Acronym	Period	Data type	Resolution	Reference
Climate Prediction Center Merged Analysis of Precipitation	CMAP	(1979–2016)	Surface station and satellite	2.5°°× 2.5°	Xie and Arkin (1997)
Climate Hazards Infrared Precipitation with Stations	CHIRPS	(1981–2018)	Surface station and satellite	0.05°°× 0.05°	Funk *et al*. (2015)
Tropical Rainfall Measurement Mission	TRMM	(1999–2018)	Surface station and satellite	0.25°°× 0.25°	Huffman *et al*. (2010)
European Centre for Medium-Range Forecasting ERA-5	ERA-5	(1979–2018)	Reanalysis	0.75°°× 0.75°	C3S (2017)

Note: For each dataset, the acronym used hereafter, the period of coverage, the field used and the horizontal resolution are shown.

The wavelet function is defined using two parameters, a dilation (a width or temporal scale) and a translation (centroid in time/space).

Wavelet transforms are the result of the inner product (convolution) of a wavelet function with a time-series or a signal (Addison, 2017). The wavelet transform can be thought of as a local comparison between the wavelet function and the observed signal for different frequencies or temporal scales. The information provided by a wavelet transform largely depends on the characteristics of the wavelet function used, thus, different wavelet functions are used for different purposes (Addison, 2017). For the purpose of finding the onset and retreat dates, the wavelet based on the Haar function is useful as this wavelet is suitable to find sudden changes in a signal (Addison, 2017). The Haar wavelet is defined as the non-continous piece-wise function:

$$\psi\left(\frac{t-b}{a}\right)=\begin{cases} 1 & b\le t\le b+\frac{a}{2} \\ -1 & b-\frac{a}{2}\le t\le b \\ 0 & \text{elsewhere,} \end{cases} \quad (1)$$

where, a is the dilation coefficient, b is the centre of the wavelet or the translation coefficient and t is the time coordinate.

The wavelet covariant transform is then the inner product of the Haar wavelet with a timeseries (Brooks, 2003), that is:

$$W_f(a,b)=\frac{1}{a}\int_{t_i}^{t_f} pr(t)\psi\left(\frac{t-b}{a}\right)dt, \quad (2)$$

where, $pr(t)$ is a time-series of precipitation, either on daily or pentad scales and $W_f(a, b)$ is the matrix of the covariant transform and t_i and t_f are the start and end time-points. No statistical treatment, normalization or anomaly, a priori, is calculated on the precipitation time-series $pr(t)$ so the units of W_f are the same as the precipitation time-series (e.g., mm day^{-1}).

Monsoon timings can be observed as sharp changes to precipitation, that is, rainfall sharply increases at onset and sharply decreases at retreat. However, measuring these changes can be difficult since precipitation time-series are typically noisy. The Haar wavelet is useful in these cases for signal detection since the wavelet transform is interpreted as gradients across different temporal scales and can smooth out the high-frequency variability using sufficiently large dilation scales. In other words, the wavelet covariant transform ($W_f(a, b)$) measures gradients on a dilation scale a for each time-step (b).

Figure 1 shows the Haar wavelet and 1 year of observed precipitation in the North American Monsoon from the CMAP dataset. Figure 1(a) illustrates how the wavelet function compares the observed signal in the interval $b<t\le b+\frac{a}{2}$ with the values of the signal in the interval $b-\frac{a}{2}\le t<b$ where b in this case is a pentad time step. The wavelet transform coefficient for dilation $a = 20$ pentads at the translation of $b = 35$, that is, pentad 35, is a measure of the precipitation difference between the sum of the observed rainfall 10 pentads after pentad 35 and the sum of the observed rainfall 10 pentads before pentad 35 as illustrated in Figure 1(b).

Figure 2 shows an example of the wavelet transform (WT) application using the observed climatological precipitation in the North American Monsoon in four different precipitation datasets. The mean climatological rainfall rates (upper panel) differ in their peak summer rainfall rates but qualitatively show similarities in the start and end dates of the rainy season. The wavelet transform coefficients ($W_f(a, b)$ in the middle panel) for a small dilation $a = 4$ are relatively noisy but show a clear maximum and minimum that correspond well with the maximum and minimum of longer dilations ($a = 14, 20$). The sum of these four coefficients at each translation or pentad b, highlight a maximum found around June 22 and a minimum found around September 21st, which agree well with previous results of mean onset and retreat dates in the North American Monsoon (e.g., Arias *et al*., 2012; Geil *et al*., 2013).

Local maxima in the wavelet transform highlight positive steps in the precipitation time-series with a coherent scale of a pentad steps. This interpretation is then extended to diagnose monsoon onset. The pentad (b^*) corresponding to the maximum of the sum of the transform over a set of scales is defined as monsoon onset (MO), that is:

$$\mathrm{MO}=b^* \Leftrightarrow \sum_{a_0}^{a_f} W_f(a,b^*)=\max\left(\sum_{a_0}^{a_f} W_f(a,b)\right). \quad (3)$$

where, a_0 and a_f are the limits of the pentad scales, that is, the dilation coefficients, b^* is the pentad of maximum $W_f(a, b)\sum W_f(a, b)$ and the monsoon onset pentad. Similarly, the monsoon retreat pentad is found at the minimum of the sum of the wavelet transforms, that is,

$$\mathrm{MR}=b^* \Leftrightarrow \sum_{a_0}^{a_f} W_f(a,b^*)=\min\left(\sum_{a_0}^{a_f} W_f(a,b)\right). \quad (4)$$

In other words, we seek to find monsoon onset and retreat using the maximum and minimum the wavelet

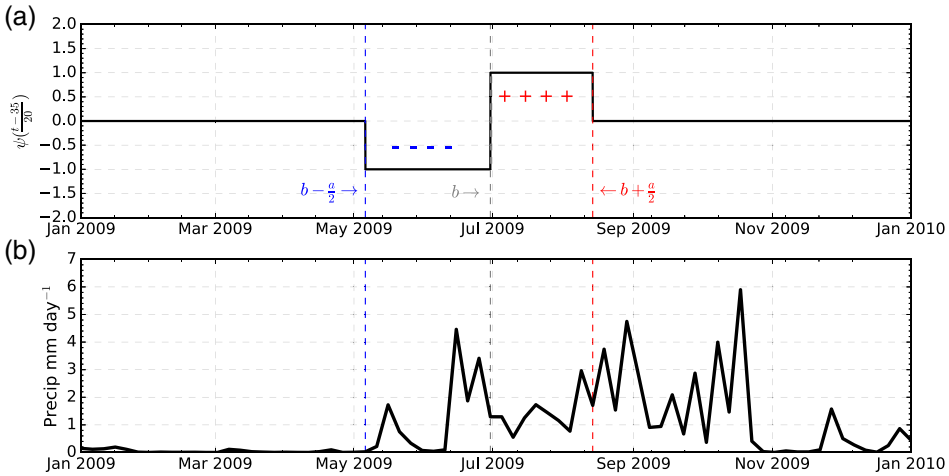

FIGURE 1 (a) Haar wavelet at a dilation $a = 20$ and translation $b = 35$, which is the 35rd pentad around June 22. The positive and negative parts of the wavelet are highlighted in red and blue, respectively. (b) CMAP precipitation in 2009 in the North American Monsoon (20–27°N, 110–103°W) [Colour figure can be viewed at wileyonlinelibrary.com]

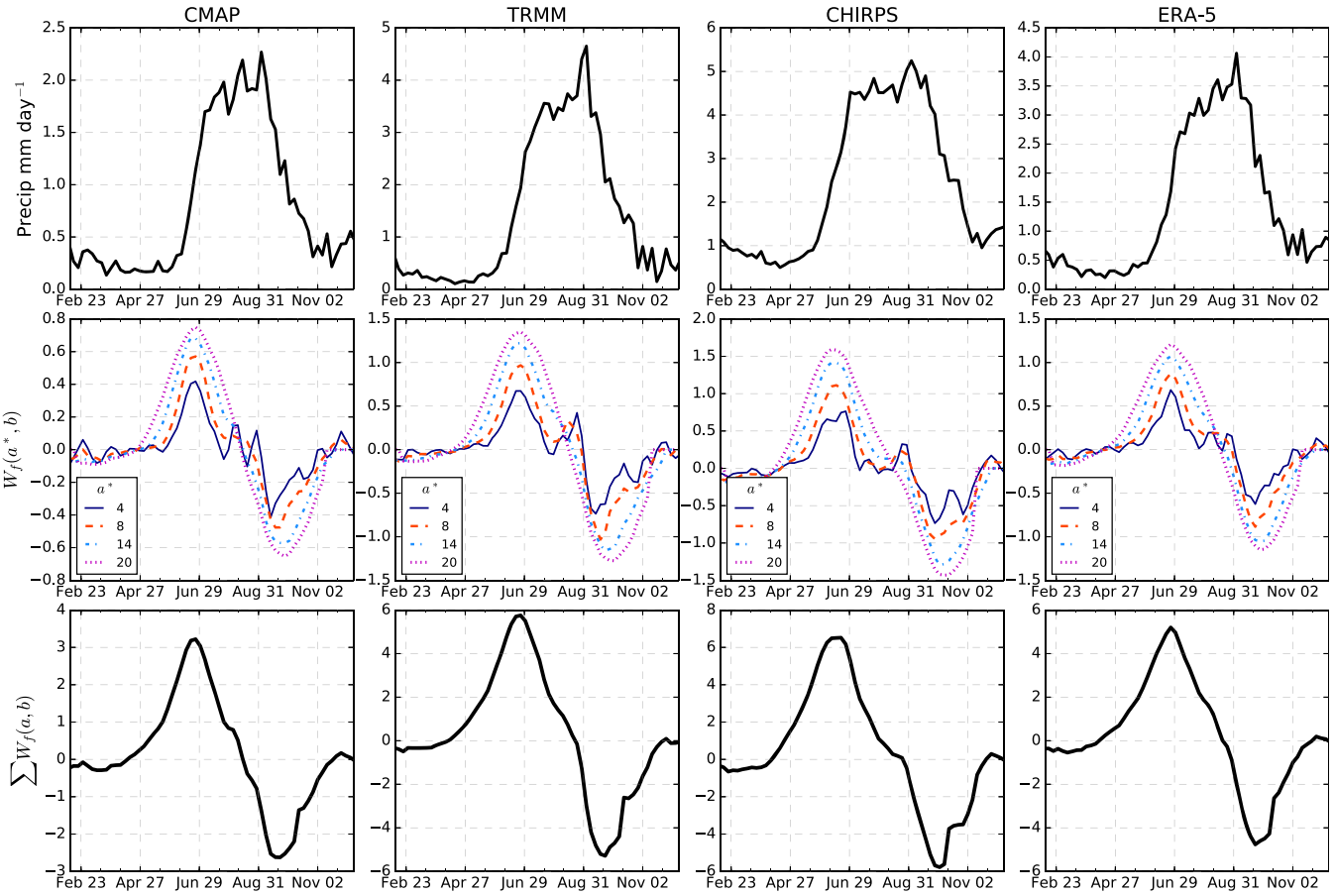

FIGURE 2 (Upper) Climatological pentad-mean precipitation in four different observational datasets in the North American Monsoon (19–35°N, 110–103°W). (Middle) The wavelet transform coefficients (mm day^{-1}) for four different dilations a. (Lower) The sum of the wavelet transform coefficients (mm day^{-1}) over dilations $a = 4, 8, 14, 20$ [Colour figure can be viewed at wileyonlinelibrary.com]

power spectrum over a range of temporal scales. Several sensitivity tests were performed with different dilation coefficients (a) in the different observational datasets, models and regions and a set or vector of dilation scales was found to be optimal to be used for all purposes. The set of dilation coefficients $\vec{a} = (28, 30, \cdots, 54)$ was found to be robust, that is, was able to capture the onset and retreat dates in all the datasets.

2.2.2 | Identification of monsoon onset and retreat

Monsoon onset is defined as the maximum sum of wavelet coefficients, capturing positive gradients within the scales of 14–27 pentads (half of the elements of vector a defined above). Monsoon Retreat has a similar definition, capturing the greatest negative gradient of precipitation over the same pentad scales.

For example, Figure 3 illustrates the method in the North American Monsoon in the TRMM dataset for 2009. Figure 3a shows the WT coefficient matrix, showing the changes in precipitation for dilations ranging from 10 to 40. A clear signal of positive coefficients is observed between pentads 28–34 and a similar negative signal observed in pentads 56–60. Figure 3b shows the time-series of the observed precipitation, which suggests that monsoon onset occurs sometime between pentads 28 and 34. Observed rainfall rapidly decreases after pentad 59 suggesting that monsoon retreat can be diagnosed around this pentad.

Figure 3c shows the WT coefficients as a function of pentad for several dilations (a_0). The coefficients for all scales seem to follow a very similar behaviour, increasing during spring to

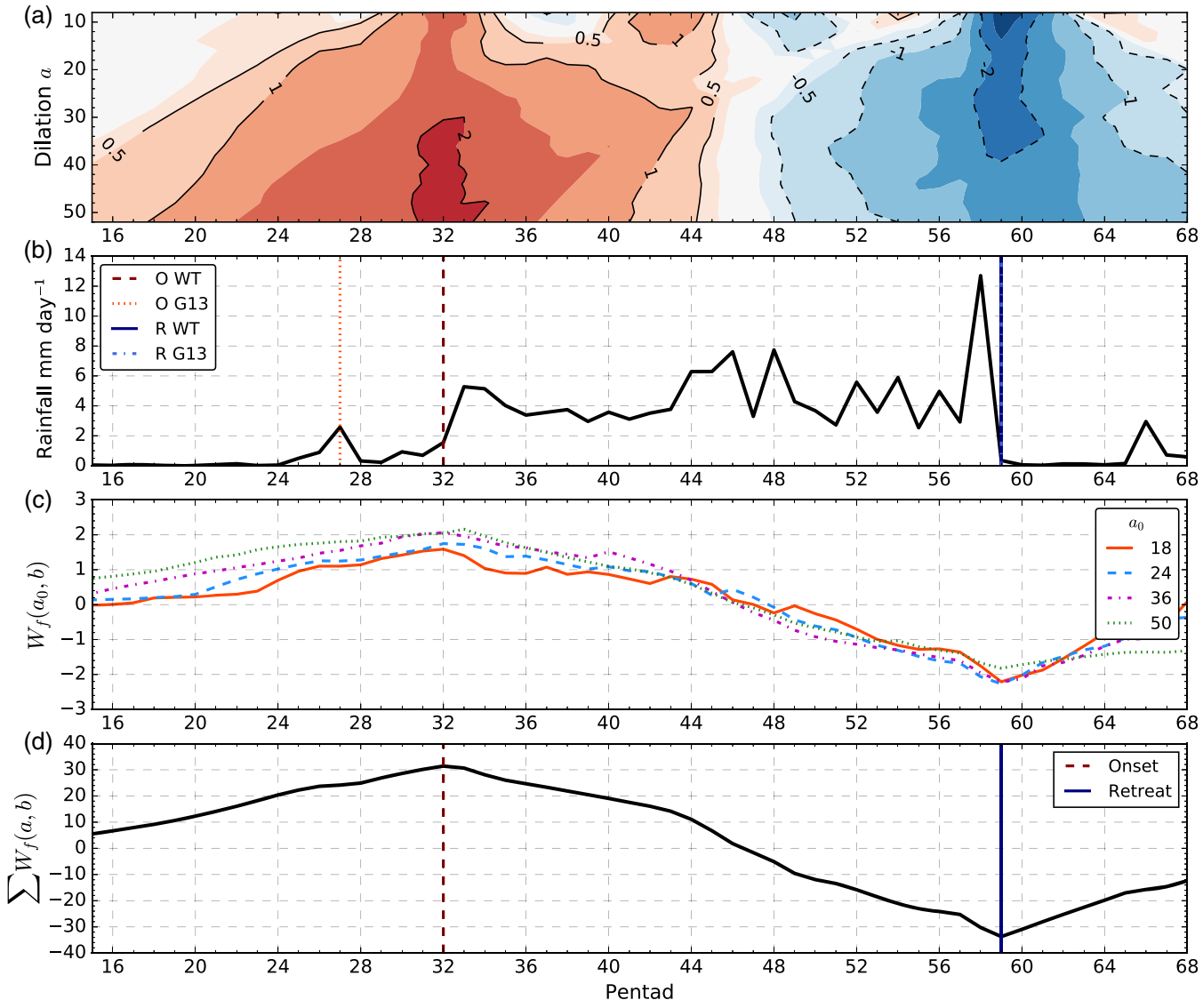

FIGURE 3 Example determination of monsoon onset and retreat dates for the north American monsoon (20–27°N, 109–103°W) in the TRMM dataset for 2009. (a) Wavelet transform coefficient matrix (mm day^{-1}) as a function of time and dilation coefficient a. The shading is from −3 to 3 mm day^{-1} with an interval of 0.5 mm day^{-1}. (b) Observed precipitation, the onset and retreat dates as determined by the wavelet transform (WT) method (dashed) and Geil *et al.* (2013)'s threshold method (solid) are shown. (c) The wavelet transform coefficients for different dilations. (d) The sum of the WT [$\sum W_f(a, b)$; mm day^{-1}]; the maximum and minimum are shown in red and blue, representing onset and retreat pentads, respectively [Colour figure can be viewed at wileyonlinelibrary.com]

reach a maximum around pentad 32 and thereafter decreasing to a minimum around pentad 59. When the sum of the wavelet transform coefficients across the dilations is computed (Figure 3d) this behaviour becomes much clearer. The maximum and minimum are found at pentads 32 and 59, respectively and these pentads define the onset and retreat times. For comparison, the results from the method of Geil *et al.* (2013) are shown in Figure 3b, indicating that this method may have found an earlier onset. Figure S1 shows another example in the same region but using model data from HadGEM3 GC3.1N216.

2.2.3 | Extension for application to the MSD signal

The wavelet method can be extended to characterize the shorter scale variations of precipitation of the MSD in Central America and the Caribbean. First, the monsoon onset and retreat dates are determined in the time-series from the area-averaged precipitation in the MSD region via the approach described in the previous section. Once the onset and retreat dates are established, an additional wavelet analysis determines the dates in which the MSD starts and ends. The onset and end of the drier period of the MSD can be found by computing the wavelet transform again but using smaller dilations and over a limited temporal range. In particular, the WT is only calculated in the 20 pentads before and after the dates defining monsoon retreat and onset, respectively. The MSD Onset (MSDO) and MSD End (MSDE) are defined as the minimum and maximum, respectively, of the sum of the wavelet transforms (Equations 3 and 4) using dilation coefficients $\vec{a}^{*} = (10, 12, \cdots, 24)$.

Figure 4 illustrates the use of the WT method to determine the dates of MSDO and MSDE for the precipitation of 2017 in ERA-5. Figure 4a depicts the wavelet covariant transform matrix, showing the W_f coefficients for each dilation a at each pentad b. The onset of rainfall is diagnosed around the time-steps of highest positive W_f coefficients—around pentads 24–32 for almost all dilations. These positive coefficients are followed by a period of negative values from pentad 32 to pentad 40, which represent the decrease in precipitation, or relative drought, in the midsummer. The MSD is followed by another period of positive coefficients from pentad 44 to pentad 52, illustrating the so-called second peak of precipitation and, finally, a period of negative coefficients associated with monsoon retreat.

The coefficients of the wavelet transform $[W_f(a_0, b)]$ for selected dilations a_0 (Figure 4c) show that the smaller dilations are more sensitive to smaller scale variations in the time series and longer dilations better highlight the long-term change of the time series. For example $a_0 = 18$ shows signs of a MSD by showing two local maxima and two local minima, whereas $a_0 = 54$ only shows a local maximum and a local minimum associated with onset and retreat.

The maximum and minimum of the sum over all dilations (Figure 4) depict the rainfall onset and retreat dates, respectively. The second wavelet transform $W_f(a^{*}, b)$ is computed over smaller dilation coefficients (a^{*}) near the onset and retreat dates as described above to highlight the MSDO and the MSDE. Figure 4e shows the sum of the wavelet transform coefficients $W_f(a^{*}, b)$ and the pentad of the MSDO, 34, and MSDE, 49, corresponding to the minimum and maximum of the sum of these wavelet transform coefficients, respectively.

The strength of the MSD can be measured through the maximum and minimum sum of the coefficients $\sum W_f(a^{*}, b)$ used to define the start and end dates of the MSD. For example, in Figure 4(e) the minimum of the $\sum W_f(a^{*}, b)$ was -20 mm day^{-1} found at pentad 35 and an opposite local maximum of $+20$ mm day^{-1} at pentad 49. These two values, hereafter *coef*1 and *coef*2, provide a quantitative measure of the strength of the MSD for this year in this dataset and will be used to measure the spatial variability of the magnitude of the MSD in the different datasets.

2.3 | Comparative methodologies

For validation purposes the wavelet transform method is compared with existing methods that determine onset and retreat in the North American and Indian monsoons. The threshold methods of Geil *et al.* (2013; hereafter G13) and Arias *et al.* (2012; hereafter A12) are compared with the results of the wavelet transform in the North American Monsoon. G13 used a threshold of 1.3 mm day^{-1} for at least 3 days for onset and 7 days for retreat for daily TRMM observations. In this study we adapt this method for TRMM to the same threshold value, but the onset pentad is the first pentad above the threshold whereas for retreat, we require rainfall to be below the threshold for two consecutive pentads. The method of A12 defines onset with two conditions. The first condition to find the onset pentad is that six out of the eight subsequent pentads must have rain-rates above the annual-mean climatological rainfall. The second condition is that at least six out of the eight previous pentads must be below the annual-mean climatological rainfall. The opposite definition is used to determine the pentad of monsoon retreat.

In the Indian Monsoon region, a commonly used method is known as the hydrologic onset and withdrawal

index (HOWI) based on moisture transport over the Arabian Sea (Fasullo and Webster, 2003; Sa*hana et* al., 2015; Chevuturi *et al*., 2019). To compute the HOWI index, first, the vertically integrated moisture transport (VIMT) is computed from daily ERA-5 data in the Arabian sea, as described by Fasullo and Webster (2003), that is:

$$\chi = \frac{1}{g}\int_{p_0}^{0} q\mathbf{V}dp, \tag{5}$$

where, χ is the VIMT, g is the gravitational acceleration, p are the pressure levels, q is the specific humidity and $\mathbf{V}$

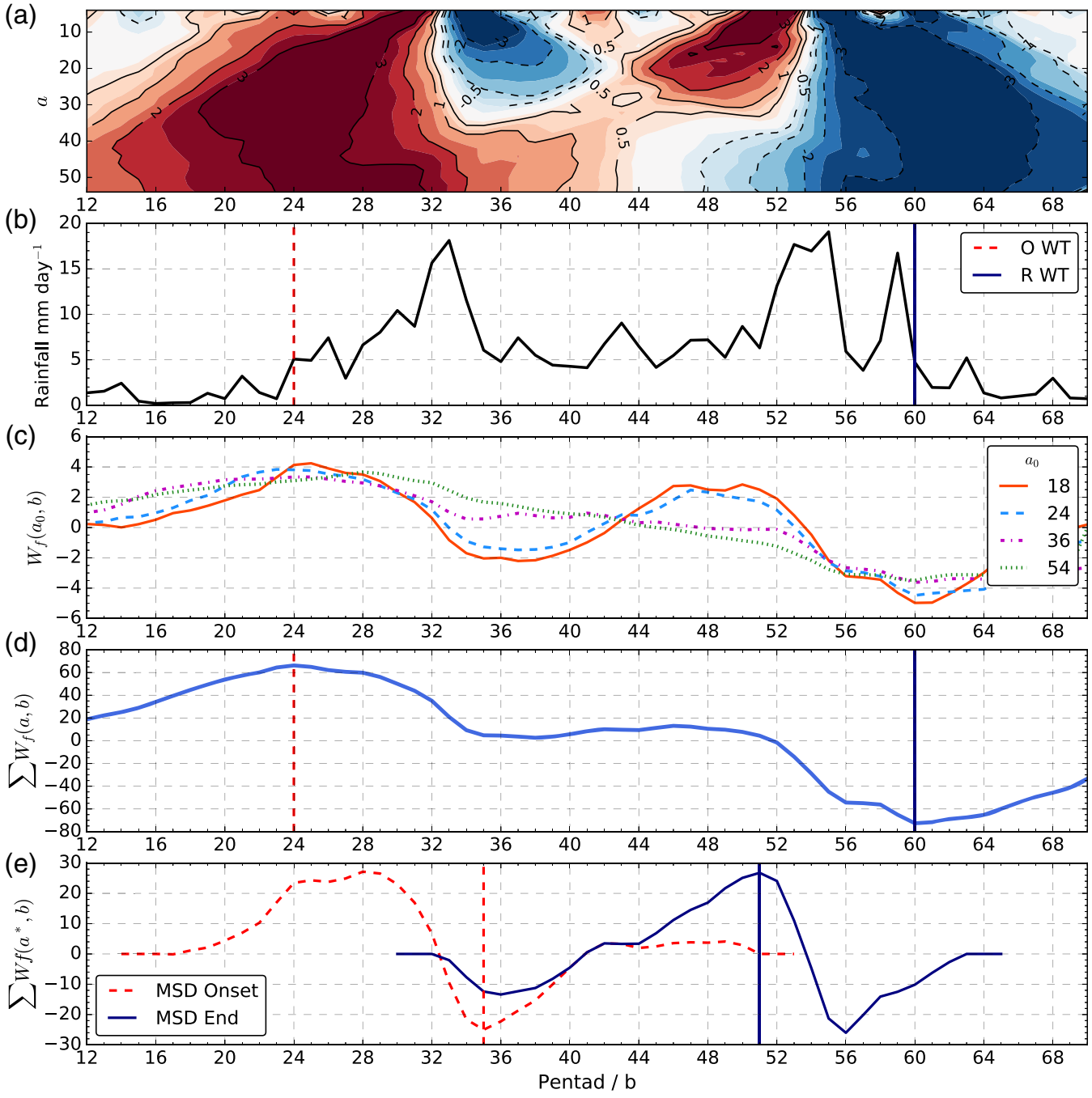

FIGURE 4 Example characterization of the MSD (11–19°N,95–85°W) using ERA-5 data for 2017. (a) Wavelet transform spectra, (b) observed precipitation with the onset and retreat pentads shown in red and blue, respectively. (c) Wavelet transform coefficients for four different dilations (mm day^{-1}). (d) The sum of the wavelet transform coefficients (mm day^{-1}) for a = 28, ..., 51. (e) The sum of $W_f(a, b)$ (mm day^{-1}) for dilation coefficients a = 12, ..., 24 showing the start (MSD onset) and end (MSD end) of the midsummer drought (MSD) [Colour figure can be viewed at wileyonlinelibrary.com]

is the wind vector. The VIMT is then normalized using the transformation:

$$\mathrm{HOWI} = 2\left(\frac{\chi - \min(\bar{\chi})}{\max(\bar{\chi}) - \min(\bar{\chi})}\right) \quad (6)$$

where, χ is the unnormalized time-series, $\bar{\chi}$ is the mean seasonal cycle of the unnormalized index and HOWI is the normalized index. The onset date is defined as the first day of each year where the HOWI index is greater than zero and the retreat date is the first day after the onset date that the HOWI index is negative (Fasullo and Webster, 2003; Sahana *et al*., 2015). The necessary daily data of moisture and wind speed on sufficient vertical levels to compute the HOWI index in the MOHC submissions to CMIP6 was not available, so the HOWI index can only be computed using ERA5 and will be compared with the WT method used on the observational gridded datasets.

3 | RESULTS

The onset and retreat dates were determined for each year in each observed and model dataset for the Indian, North American and MSD regions using the methods described in the previous section. The calculations were performed for area-averaged precipitation time-series representative of the core regions defining these monsoons. Calculations were also made at grid-box scales to illustrate the spatial distribution of the onset and retreat dates.

3.1 | The North American Monsoon

Table 2 shows the onset and retreat dates of the North American Monsoon estimated using the G13, A12 and WT methods. The table reports the results for three observational datasets, ERA-5 reanalysis and five climate model experiments. The observations generally agree that the onset date is found at pentad 33 (around June 15), according to the WT and the method by G13. However, the method of G13 generally places the retreat date one pentad after the WT method, that is, around October 7th for G13's method and October 2nd for the WT's method. The method by A12 disagrees with G13 and the WT methods on both onset and retreat mean pentads, in both cases finding an earlier onset (pentad 30) and retreat (pentad 54).

The climate models reasonably represent the mean onset and retreat dates, as only the onset dates from two experiments of UKESM1 are statistically different from two of the observational datasets. The similiarities in onset and retreat dates confirm that the seasonal cycle is very well represented by these models in the North American Monsoon. The method by A12 in the simulations also produces an earlier onset and retreat dates when compared with the other two methods by about 1.5 pentads, but this is within the uncertainty range given by the interannual variability.

Figure 5 compares the estimated onset and retreat dates using these three methods in the North American Monsoon in 2010 using the three observational datasets and ERA-5. The method by A12 shows an earlier onset and retreat in all the datasets whereas the WT and G13 agree in almost every dataset. The WT method is the only method that estimated the same onset date for all the four datasets

TABLE 2 Mean (*SD*) pentads of monsoon onset (O) and retreat (R) in the North American Monsoon (110°–103°W, 20°–27°N) for observational datasets, reanalysis and model output with the wavelet transform method WT, G13's and A14's method

Dataset/experiment	WT O	WT R	G13 O	G13 R	A12 O	A12 R
TRMM	33.3 (±1.8)	55.8 (±1.9)	33.0 (±1.7)	56.6 (±1.4)	30.4 (±1.7)	53.8 (±2.0)
CMAP	33.2 (±1.6)	55.0 (±2.1)	36.0 (±3.3)	55.7 (±1.8)	31.7 (±3.0)	54.5 (±3.3)
CHIRPS	32.5 (±1.5)	54.7 (±1.9)	33.6 (±1.7)	56.1 (±1.4)	30.1 (±1.7)	53.6 (±2.5)
ERA-5	33.5 (±1.8)	55.5 (±2.0)	33.6 (±1.8)	56.4 (±1.4)	30.9 (±1.7)	53.3 (±2.3)
GC3 N216-pi	33.7 (±2.0)	55.1 (±1.8)	32.6 (±2.5)	55.9 (±2.3)	32.4 (±2.5)	53.8 (±3.8)
GC3 N96-pi	33.7 (±2.2)	55.0 (±2.1)	32.7 (±2.9)	56.3 (±2.0)	31.9 (±2.6)	54.1 (±4.1)
GC3 hist	33.8 (±2.3)	55.1 (±2.1)	33.3 (±2.9)	56.0 (±2.2)	31.9 (±2.6)	53.7 (±4.0)
UKESM-pi	**34.5** (±2.1)	54.8 (±2.1)	34.1 (±2.9)	56.1 (±1.9)	33 (±2.6)	53.9 (±4)
UKESM-hist	**34.3** (±2.2)	54.3 (±2.2)	34.4 (±3.2)	55.6 (±2.1)	33.1 (±3.0)	53.2 (±4.2)

Note: Pentad 34 corresponds to the period between June 17–22 and pentad 54 to the period September 27 to October 1. The results from the WT method in the model experiments that are statistically different to both the CMAP and CHIRPS results to the 95% confidence level, according to a Welch's *t*-test, are shown in bold.

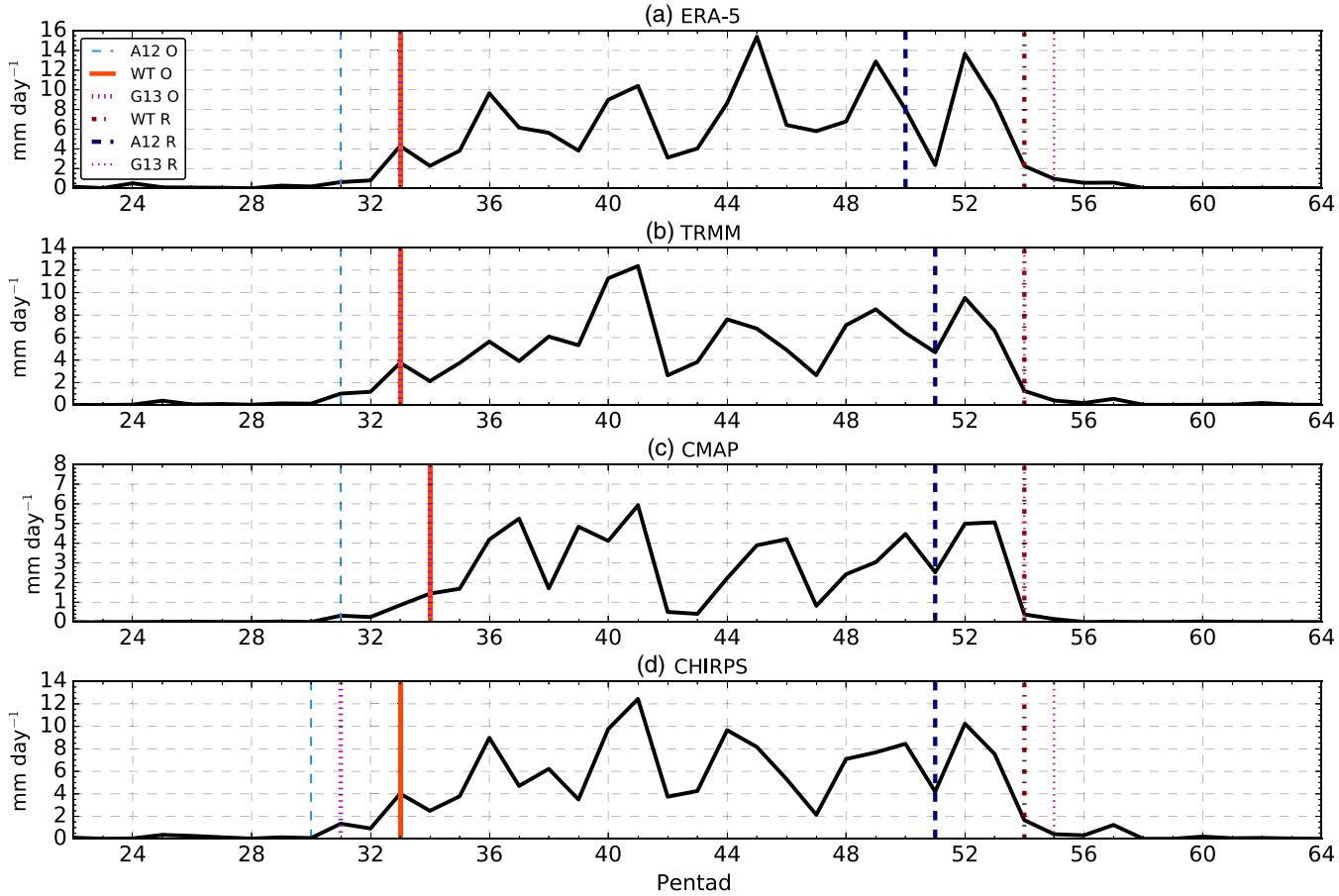

FIGURE 5 Pentad-mean precipitation for the North American Monsoon in 2010 in four precipitation datasets showing the onset and retreat pentads as diagnosed by the WT, the A12 and G13 methods. The area used to average the precipitation is illustrated in in Figure 6(b)) [Colour figure can be viewed at wileyonlinelibrary.com]

and the same retreat date in three of the four datasets, with the fourth dataset delayed by only one pentad.

The meteorological changes associated with onset and retreat in the North American Monsoon are shown in Figure 6. The composite differences of the precipitation, wind and geopotential changes 10 days prior and after onset and retreat are compared for the three methods. The impact of monsoon onset in precipitation is slightly stronger using the method by A12 than the WT or G13. The WT method shows a very similar pattern and magnitudes of the anomalies of onset and retreat when compared with the other two methods. The method by G13 produces the weakest anomalies, particularly of precipitation, but very similar patterns overall.

Figure 7 shows the spatial distribution of the mean onset and retreat dates in the North American Monsoon region for various datasets. There is high agreement between TRMM, CHIRPS and ERA5 on the spatial pattern of mean onset and retreat dates. Onset in western Mexico is around pentad 31 (around June 1st) whereas in Chihuahua and Sonora, the rainy season begins shortly after pentad 35 (June 22nd). The pattern in the medium-resolution simulation GC3 N216 piControl is consistent with observations, particularly during onset. However, the spatial pattern of the mean retreat dates in the northern regions of the monsoon show an earlier than observed retreat, possibly associated with the dry bias in this region in these models (García-Franco *et al.*, 2020).

3.2 | The Indian monsoon

Table 3 compares the mean onset and retreat dates in the Indian Monsoon computed from the HOWI index using ERA5 data and the WT used for gridded precipitation datasets. The onset and retreat dates from the HOWI index were converted from the daily to the pentad-scale to compare with the WT. The mean onset date for the HOWI index is May 27th between pentads 29 and 30, and retreat is between pentads 49 and 50, around September 3rd. The mean onset date found using the WT method

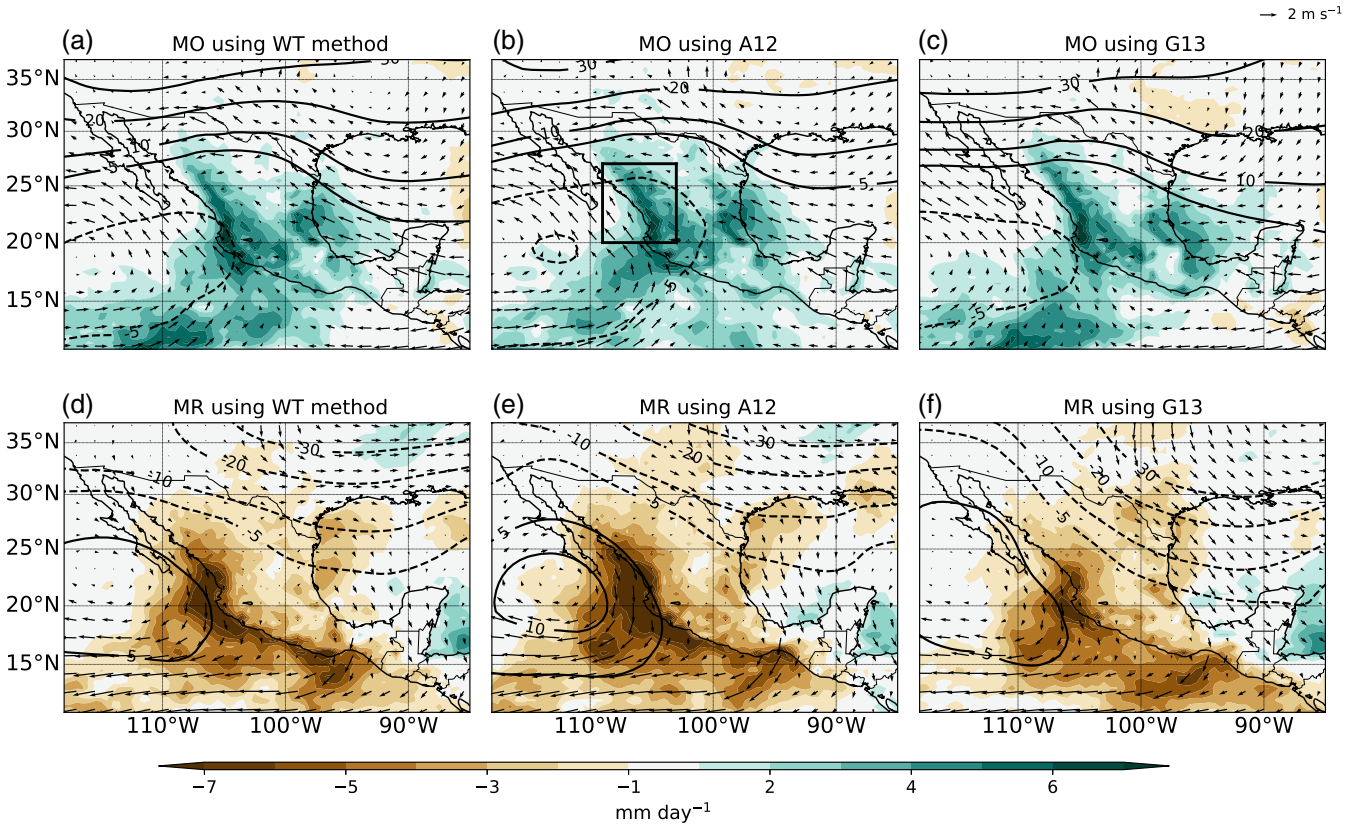

FIGURE 6 Precipitation (colour contours), low level wind at 850-hPa and geopotential height (line contours) at 500 hPa anomalies for (a–c) the difference between the 10 days after monsoon onset and 10 days prior to onset (MO) using onset dates from (a) the WT (b) Arias *et al.* (2012) and (c) Geil *et al.* (2013). Part figures (d–f) are as in (a–c) but for monsoon retreat. The data and dates are obtained from ERA-5, and the area for the average is shown in the box in (b) [Colour figure can be viewed at wileyonlinelibrary.com]

for the four observational datasets was close to pentad 32, about two pentads later than the HOWI index. The mean retreat date for the WT method was about one pentad earlier than the HOWI results. Overall, the models exhibited later than observed onset and retreat dates (4 pentads). The differences between the hydrological determination of onset and retreat dates, through HOWI, and the WT method on gridded datasets is significant, according to a Welch's *t*-test done between HOWI and all the gridded datasets. These differences may be due to the different regions where each method is defined, that is, HOWI is defined over the whole of the Arabian Sea where an earlier onset and later retreat would be expected when compared with rainfall over mainland India, where the WT method was applied.

Figure 8 shows differences between the WT (based on ERA5 precipitation) and HOWI methods, comparing precipitation and moisture fluxes at 850 hPa 10 days prior to and following monsoon onset and retreat. The HOWI index better captures the moisture transport in the Arabian Sea whereas the WT method best captures precipitation differences prior to and following monsoon onset and retreat over mainland India. The WT method is also able to capture onset and retreat dates and the associated anomalies within the climate model output. Figure S2 shows the precipitation and moisture transport anomalies around the onset and retreat of the Indian monsoon in three different climate model experiments. While the models show significant biases in the timings of the monsoon, according to a Welch's *t*-test (Table 3) the models show similar patterns of rainfall and moisture transport anomalies to the observations around both onset and retreat.

In the case of the Indian monsoon, Figure 9 shows that the mean onset and retreat dates vary greatly spatially on the southern tip of the subcontinent. While most of northern India has a mean onset date around pentad 33, the western coast shows an earlier onset by about one or two pentads. There is high agreement in the onset date in the three observed/reanalysed datasets over mainland India and between TRMM and ERA5 over the western coast of India. Earliest onset is found on the western coast around pentad 25–27 and extending to central India by pentad 31. The GC3 N216 simulation, however, shows a later onset than observed by about two pentads in most regions. In contrast, the spatial pattern for the mean retreat date shows higher spatial

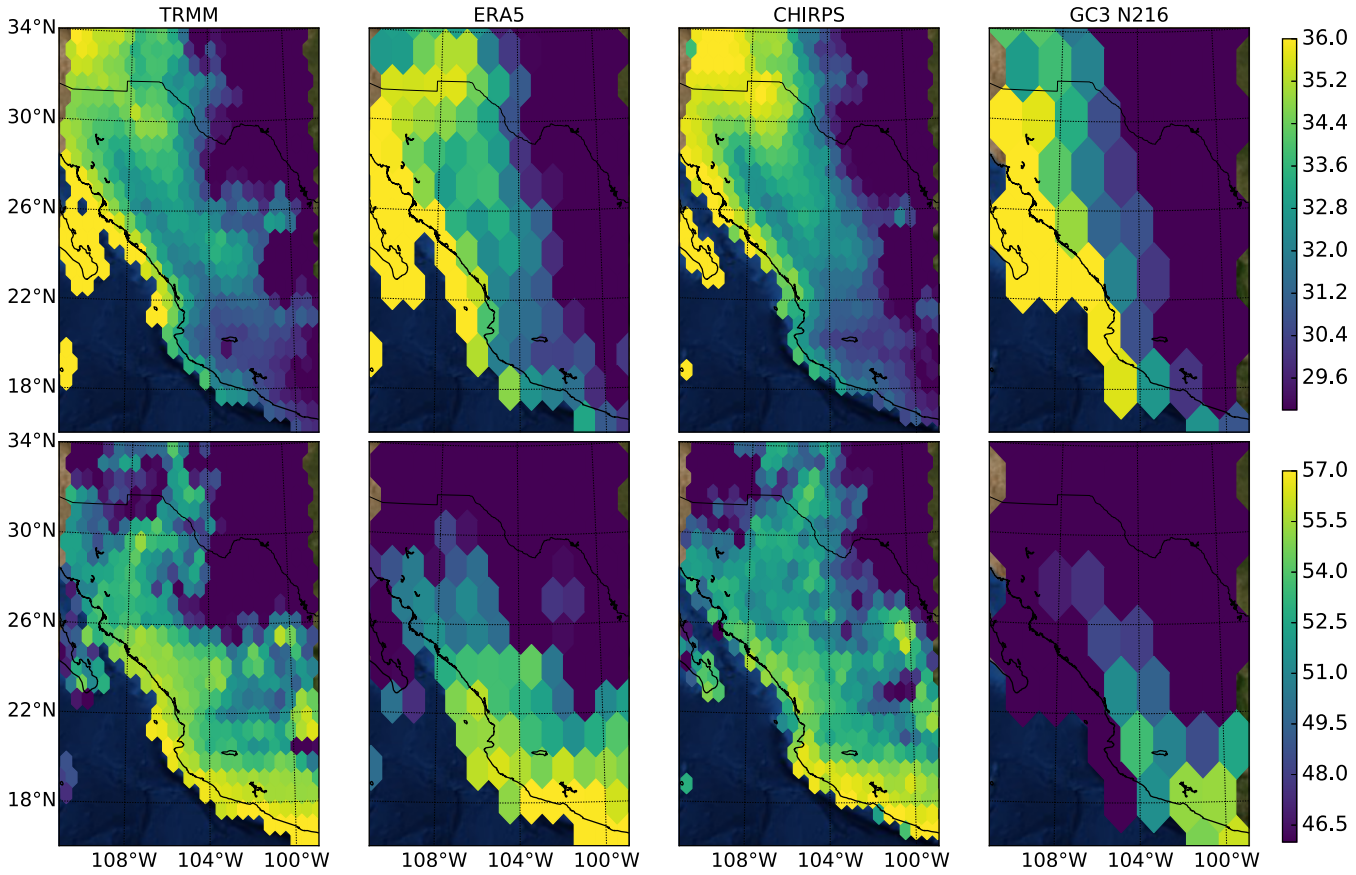

FIGURE 7 Rainfall onset (upper) and retreat (lower) mean pentads in the North American Monsoon for observations and a climate model output using the WT method [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 3 Mean (*SD*) pentads of monsoon onset (O) and retreat (R) in the Indian monsoon using the WT method on observed, reanalysed and modelled time-series as well as for the HOWI index

Dataset	O	R
TRMM	31.6 (±1.8)	53.2 (±1.9)
CMAP	31.8 (±1.6)	53.3 (±2.6)
CHIRPS	31.5 (±1.4)	53.4 (±1.9)
ERA-5	31.8 (±1.9)	52.7 (±2.6)
GC3 N216-pi	**34.4** (±1.3)	**50.5** (±1.9)
UKESM-pi	**36.1** (±3.1)	**51.9** (±3.2)
UKESM-hist	**36.0** (±3.9)	**51.8** (±3.3)
GC3 N96-pi	**35.5** (±1.8)	**51.8** (±2.3)
GC3-hist	**35.7** (±2.1)	**51.5** (±2.8)
HOWI	**29.5** (±2.3)	**49.3** (±2.4)

Note: The region over which precipitation was area-averaged for the WT method was (75°–83°E, 18°–24°N). The mean onset and retreat dates that are significantly different to the 99% confidence level to the CMAP dataset are shown in bold.

variability between the western and eastern coasts of India. CHIRPS shows the latest retreat dates over the south-central states when compared with TRMM and ERA-5.

3.3 | The midsummer drought

Results from the application of the wavelet transform to the MSD, including the mean onset and retreat dates as well as the start and end dates of the MSD period, are reported in Table 4. The mean onset date in the observations is around pentad 27 (May 14th), whereas the retreat date is around pentad 61 (October 31st). The end of the so-called first-peak period, or start of the relatively drier period (MSD), is consistently found in all the observed datasets to be around pentad 36 (around June 29th). The end of the drier period or start of the second peak is also consistently determined to be between pentads 48 and 49 in the four observational datasets. In other words, the MSD has a mean duration of 12 pentads, or around 2 months, from late June to late August. In the MOHC simulations, the MSD starts slightly later than observed by about two pentads, and ends about one pentad later than observed.

Figure 10 shows the rainfall anomalies associated with the different periods (stages) of the rainy season in southern Mexico and Central America. These include monsoon onset and retreat, and the start and end of the

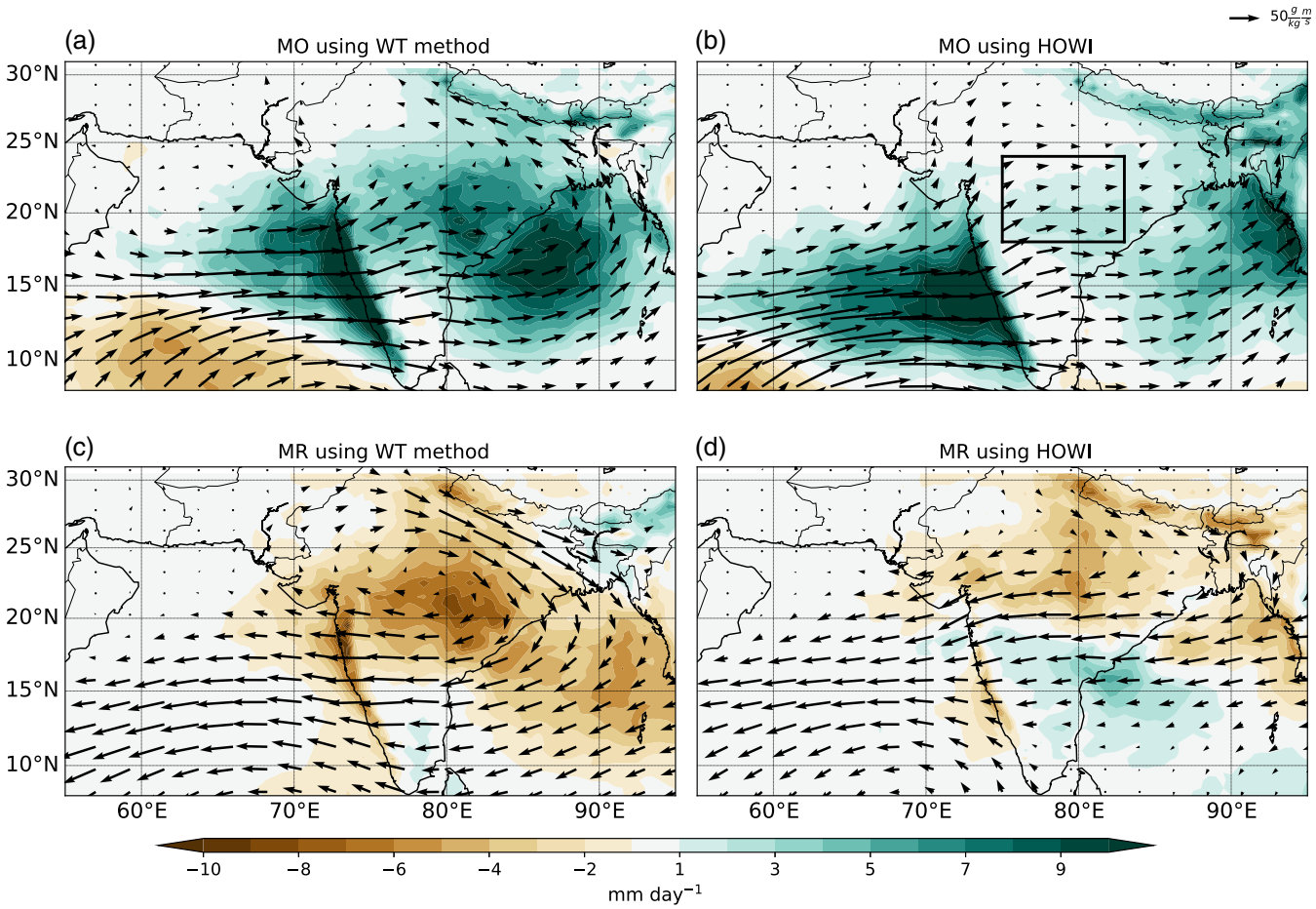

FIGURE 8 Precipitation anomalies (colour contours) and the moisture vectors showing the product of specific humidity (q) and wind ($\vec{u}$) at 850 hPa. Part figures (a, b) show the difference between the 10 days after monsoon onset and 10 days prior to onset (MO) using (a) the wavelet transform method (WT) and (b) the dates estimated using the HOWI index. Part figures (c, d) are as in (a, b) but for monsoon retreat. The data and calculations are from ERA-5 over the box in panel (b) [Colour figure can be viewed at wileyonlinelibrary.com]

MSD, the MSDO and MSDE, respectively. For each stage, we compared the anomalies computed by separating the stages using the WT method or the dates of the climatological monsoon onset, retreat, MSDO and MSDE as found in Table 4. In this way, the ability of the WT method to characterize rainfall variations is tested against a first best guess—the climatological mean dates. Overall, using the dates for MSDO and MSDE from the climatological dates results in weaker anomalies than compositing via the specific dates for each year obtained with the WT method. Even though the area-averaged signal used diagnose the different MSD stages focuses on a small region of southern Mexico and northern Central America, the anomalies associated with the onset and end of the MSD (Figure 10f,h) extend across the East Pacific warm pool, most of the western coast of Mexico and into to the Caribbean Sea and Cuba.

The analysis of individual years of observed precipitation in the selected area-averaged time-series showed that not all years showed a bimodal signal in the area-averaged precipitation (see Figure S3). In fact, a given year could be classified as having (a) a canonical two-peak structure separated by an MSD, (b) only having a first peak and an MSD but no second peak, (c) only having a second peak but no clear MSD or (d) a plateau-like monsoon season with no MSD-type variations (see Figure S3).

Due to this year-to-year variability in the characteristics of the seasonal cycle, an objective measure was defined to determine whether a signal presented a robust MSD-bimodal signature. For this purpose, the WT algorithm was applied to randomly generated precipitation time-series.

The random time-series are constructed by randomly sampling observations in the wet and dry seasons. The pentad-mean onset and retreat dates from Table 4 were used to composite the observations into dry and wet distributions. For a random time-series, the values of each pentad are randomly selected from the dry or wet distributions, depending on the pentad. In this way, the value

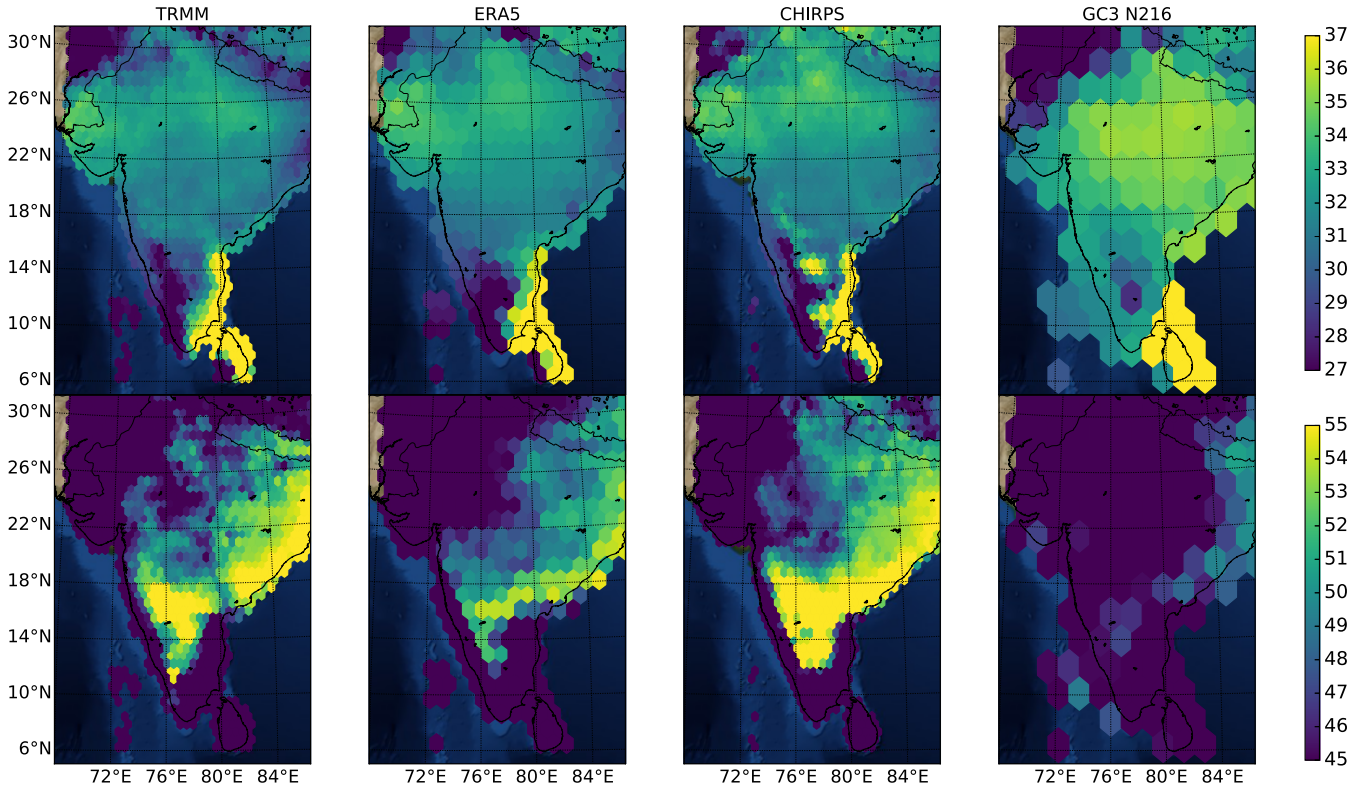

FIGURE 9 As in Figure 7 but for the Indian Monsoon [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 4 Mean pentads of rainfall onset (RO), rainfall retreat (RR), MSD onset (MSDO) and MSD end (MSDE) in the MSD region (11–19°N, 95–85°W: illustrated in Figure 10(a)) estimated through the WT method

Dataset	RO	RR	MSDO	MSDE	coef1	coef2
TRMM	25.8 (±2.2)	61.6 (±3.1)	35.9 (±2.4)	49.0 (±4.1)	−9.5 (±4.2)	10.4 (±5.4)
CMAP	26.7 (±1.9)	60.6 (±3.3)	36.5 (±2.6)	48.0 (±4.2)	−7.1 (±4.2)	7.7 (±4.3)
CHIRPS	26.7 (±2.3)	61.4 (±3.1)	36.5 (±2.7)	48.3 (±3.5)	−4.7 (±2.7)	5.5 (±3.2)
ERA-5	26.5 (±2.2)	61.8 (±3.2)	36.1 (±2.7)	48.8 (±3.5)	−10.7 (±5.4)	11.8 (±6.6)
UKESM-pi	27.4 (±2.4)	61.9 (±3.2)	**38.2** (±2.7)	**49.1** (±2.7)	−18.2 (±8.7)	14.6 (±8.0)
GC3 N96-pi	26.9 (±2.6)	62.3 (±3.5)	**37.8** (±2.1)	**49.9** (±3.1)	−21.7 (±9.4)	16.8 (±8.0)
GC3 N216-pi	26.9 (±2.3)	62.2 (±3.5)	**38.4** (±2.1)	**50.0** (±2.7)	−23.5 (±8.0)	14.1 (±6.7)
GC3-hist	26.9 (±2.7)	**62.8** (±3.7)	**37.8** (±2.4)	**50.3** (±2.6)	−19 (±8.7)	17.1 (±8.4)
UKESM-hist	**28.5** (±2.7)	**62.8** (±3.5)	**38.7** (±2.8)	**50.1** (±2.7)	−20.3 (±10.1)	14.9 (±8.3)

Note: Pentad 35 corresponds to the period between June 22–27 and pentad 52 to the period September 13–18. The model dates shown in bold are statistically different from CMAP and CHIRPS results to the 99% confidence level according to a Welch's *t*-test.

of a given pentad of the random time-series may have been observed at a different pentad; the only constraint is that the random values come from pentads that were observed in the same season.

This approach has two advantages. First, that the approach imposes the monsoon-like feature of a sharp wet-dry contrast but secondly, the random selection in the wet season removes the possible signal of the MSD in the climatological rainfall.

The random time series are then constructed by randomly drawing values at each pentad from the wet or dry season distributions of each dataset. Then, the WT method was used on 10,000 of these random-time series. This approach rendered a distribution of coefficients (*coef*1 and *coef*2) essentially representing the variability of the WT method applied to noise.

Figure 11 shows the pentad-mean time-series from 2 years in the TRMM dataset, and two randomly

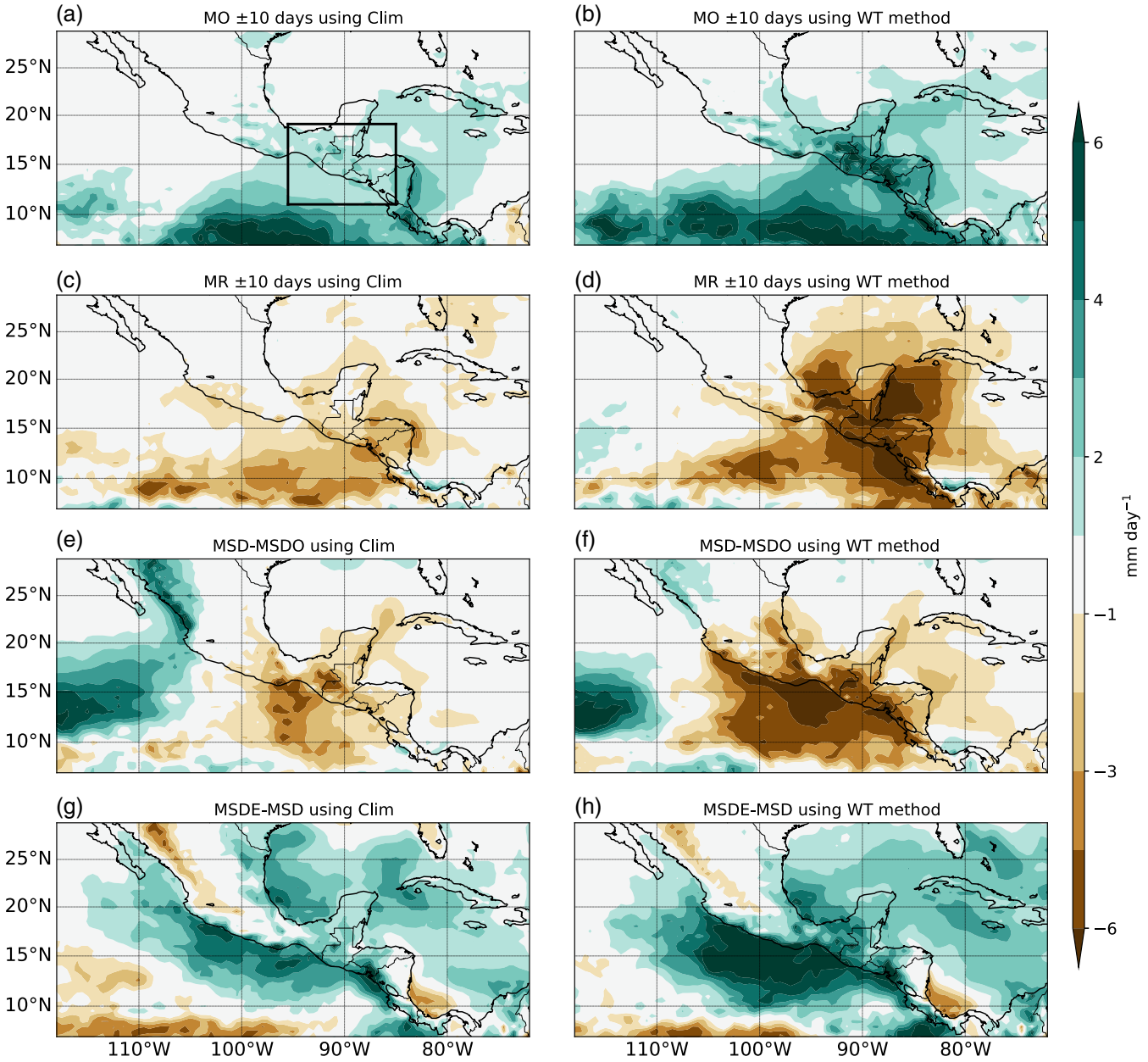

FIGURE 10 Precipitation anomalies for (a, b) the difference between the 10 days after monsoon onset and 10 days prior to onset (MO) using (a) the climatological dates of onset and (b) the dates estimated using the WT method. Part figures (c, d) are as in (a, b) but for monsoon retreat. (e, f) Difference between the MSD and the 10-day mean prior to the onset of the MSD (MSDO). Part figures (g, h) as in (e, f) but showing the difference between the end of the MSD (MSDE) and MSD. The data and calculations are from ERA-5. The black rectangle in (a) shows the MSD area used to average the precipitation throughout this study [Colour figure can be viewed at wileyonlinelibrary.com]

generated time-series. The coefficients *coef*1 and *coef*2, illustrated in Figure 11b, measure the difference in precipitation between the first peak and the MSD period and the MSD and the second peak, respectively. The first quantile of *coef*1 and the third quantile of *coef*2 provide a measure of robustness for the observed *coef*1 and *coef*2. In other words, for a year to be classified as having a robust MSD signal, the resulting *coef*1 and *coef*2 of the WT procedure must be lower and higher, respectively, than those obtained for a random time-series. The analysis of *coef*1 then determines the existence of a first-peak MSD type variability and *coef*2 determines the robustness of a possible second-peak for that year. By this procedure, a given year could fit into four categories:

- Canonical MSD: *coef*1 lower than the first quantile (25%) of random *coef*1 and *coef*2 higher than the third quantile (75%) of random *coef*2.

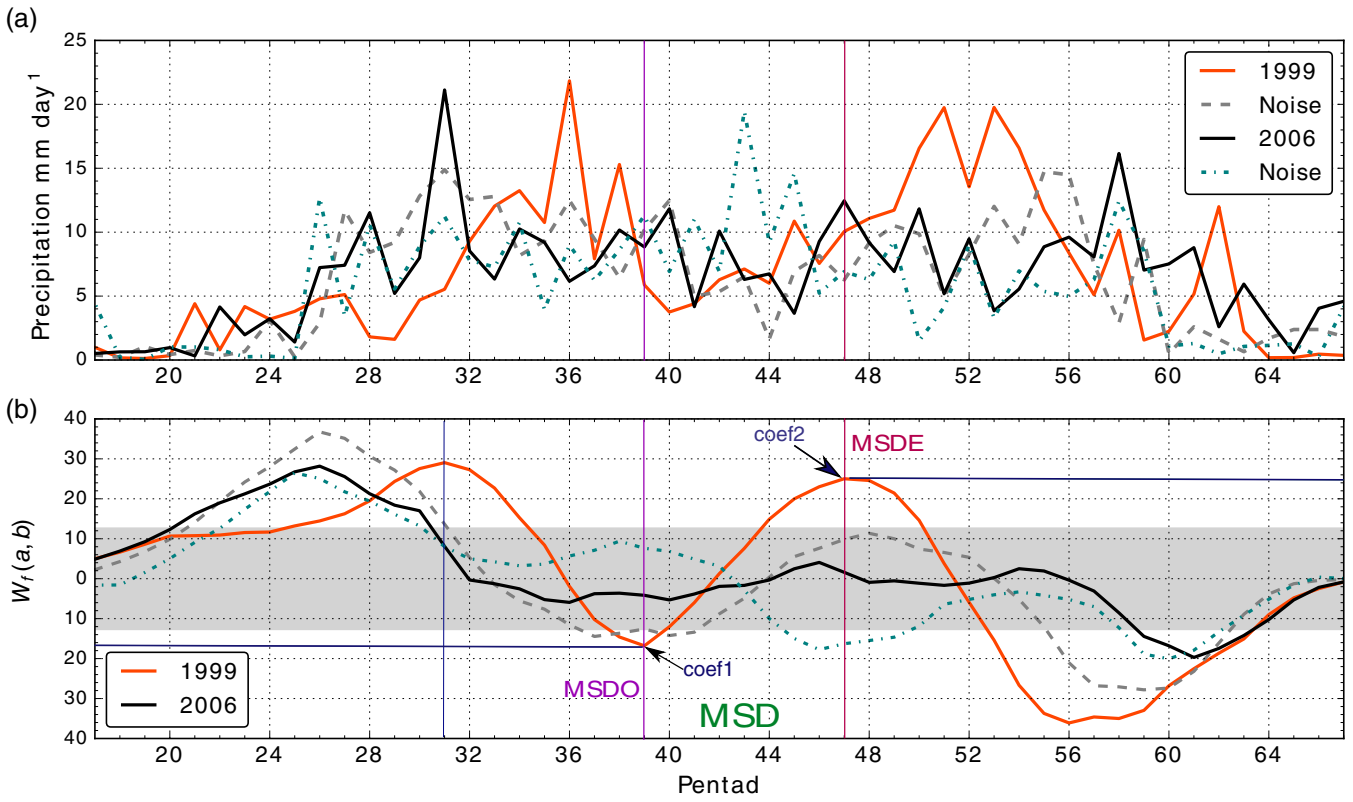

FIGURE 11 (a) Pentad-mean precipitation in 2 years of TRMM data: 1999 and 2006 and two randomly generated precipitation time-series (see text). (b) Sum of the wavelet transforms of the time series in (a). The shaded region in grey in (b) corresponds to the interval between the first quantile of *coef*1 and the third quantile of *coef*2 of 10,000 random timeseries constructed with TRMM data. The onset (MSDO) and end (MSDE) of the relatively drier period, as well as the location and values of *coef*1 and *coef*2 for 1999 are labelled in (b) [Colour figure can be viewed at wileyonlinelibrary.com]

- First peak + MSD: *coef*1 lower than the first quantile of random *coef*1 but *coef*2 lower than the third quantile of random *coef*2. In other words, the second peak is not distinguishable from noise.
- Second peak only: *coef*1 higher than the first quantile of random *coef*1 but *coef*2 higher than the third quantile of random *coef*2. In other words, the second peak is distinguishable from noise, but there is no first-peak + MSD structure.
- No MSD: *coef*1 higher than the first quantile of random *coef*1 and *coef*2 lower than the third quantile of random *coef*2. In other words, the precipitation time-series shows no robust signal of an MSD regime, with a first or second peak.

Figure S3 shows how separating years into these categories affects the pentad-mean seasonal cycle of precipitation in southern Mexico and Central America in four observational datasets. This figure also validates the above procedure as the WT method is able to robustly separate years into the different categories.

For each dataset we determine those grid-points showing a robust MSD. We use the climatological rainy and wet seasons as above, to construct the random time-series for each grid-point and estimate the random values of *coef*1 and *coef*2, repeating the procedure 10,000 times. A given grid-point is diagnosed to have a MSD when the value of *coef*2 – *coef*1 is higher than the third quantile of the PDF of the random time series. The value of *coef*2 – *coef*1 is a measure of the magnitude of the MSD since *coef*2 measures the relative strength of the second-peak compared with the MSD and therefore positive in an MSD grid-point and *coef*1 compares how dry the MSD is relative to the first-peak and thus negative if an MSD regime is observed at that grid point.

Figure 12 shows the regions where the climatological rainfall shows a MSD signal that is distinguishable from noise, that is, the values of *coef*2 – *coef*1 exceed the third quantile of the distribution composited with random time-series. Cuba, western Central America and most of southern and central-eastern Mexico exhibit a robust MSD signal., This map also shows the MSD amplitude; the strongest MSD signal is found on the western coast of northern Central America and northeastern Mexico. The high correspondence between the three observational datasets shows that the method is robust across datasets. These results

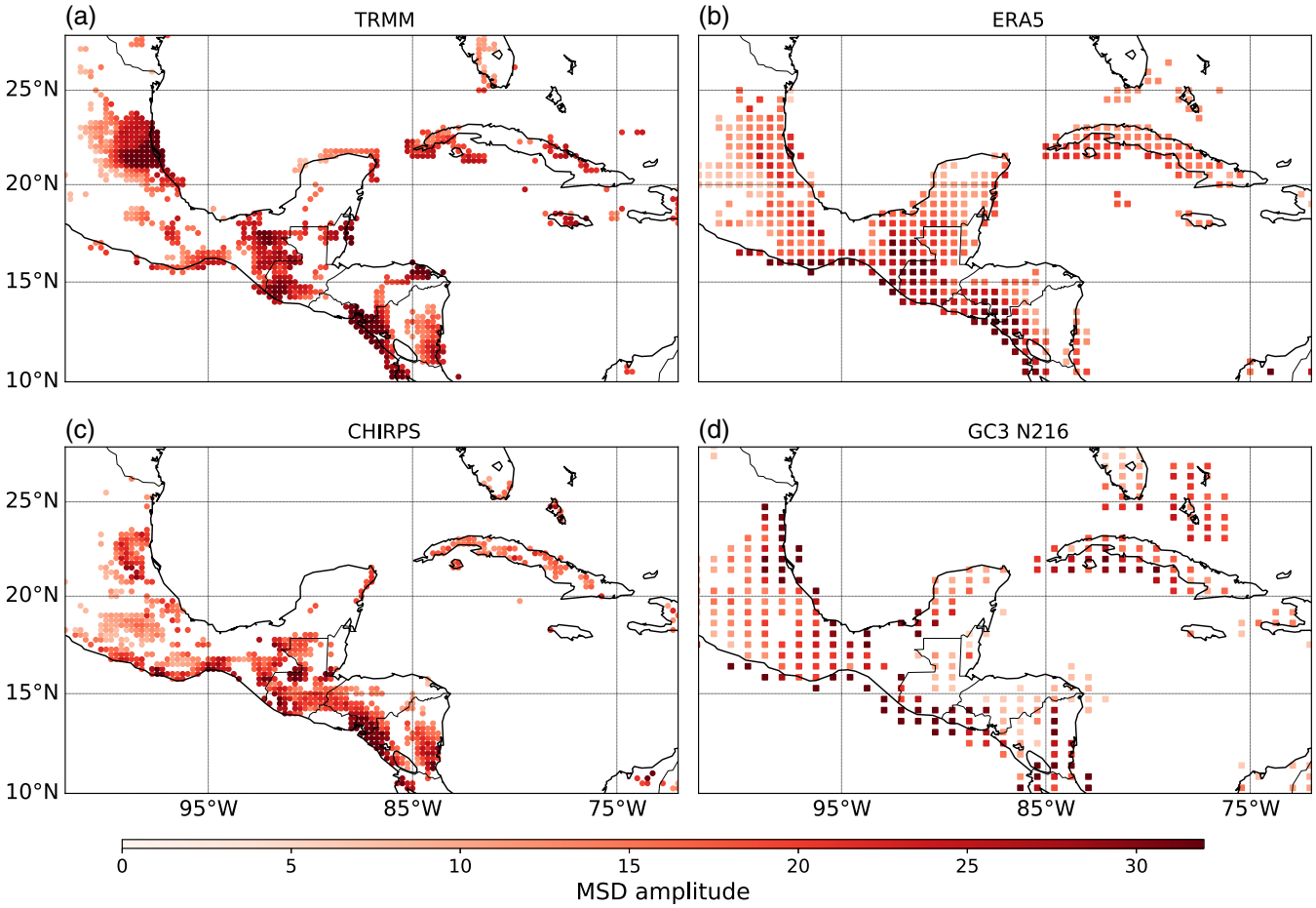

FIGURE 12 Grid points where the MSD is significantly different from noise (see Section 3.3) for (a) TRMM, (b) ERA-5, (c) CHIRPS and (d) GC3 N16-pi. The magnitude of the MSD, measured as *coef*2 − *coef*1 is shown in colour shading [Colour figure can be viewed at wileyonlinelibrary.com]

agree well with previous studies on the spatial distribution of the MSD (Magaña *et al.*, 1999; Perdigón-Morales *et al.*, 2018; Anderson *et al.*, 2019; Zhao and Zhang, 2020). In particular, the method is able to replicate the previously reported MSD signal in the Pacific Mexican coast and the stronger MSD signal in northeastern Mexico.

Figure 13 shows the spatial distribution of the mean onset and retreat pentads and the start and end of the MSD, in the grid-points where the signal is significant as in Figure 12. The earliest rainfall onset is found on the western coast of southern Mexico, Guatemala and El Salvador, as well as in Cuba, at pentad 25, whereas onset in the Yucatan peninsula is found at pentad 28 and even later, around pentad 31, in the eastern states of Mexico. In contrast, the retreat date seems spatially more homogeneous as northern Central American has a mean retreat date around pentad 59 and central Mexico around pentad 54. The MSD coherently starts over the western coast of Guatemala and Chiapas around pentad 33. In contrast, the MSD on the eastern Mexican states of Veracruz and Campeche begins after pentad 40. The earliest MSD end (Figure 13h–j) is found in central and northeastern Mexico, around pentad 42 whereas the MSD in Guatemala ends around pentad 48.

4 | SUMMARY AND DISCUSSION

A novel method is presented to diagnose monsoon onset and retreat dates using pentad-mean precipitation based on the computation of a wavelet transform over multiple temporal scales. The wavelet function used is the Haar wavelet, a wavelet typically used to find abrupt changes in signals. Onset is defined as the maximum of the sum of the coefficients of the wavelet transform computed over a range of temporal scales or dilations. These dilations were found to provide the best results in a range from 28 to 54 pentads. Monsoon retreat is similarly defined but using the minimum of this sum of wavelet transform coefficients. The use of this method is illustrated using multiple observational datasets and climate model output. The method is compared with existing

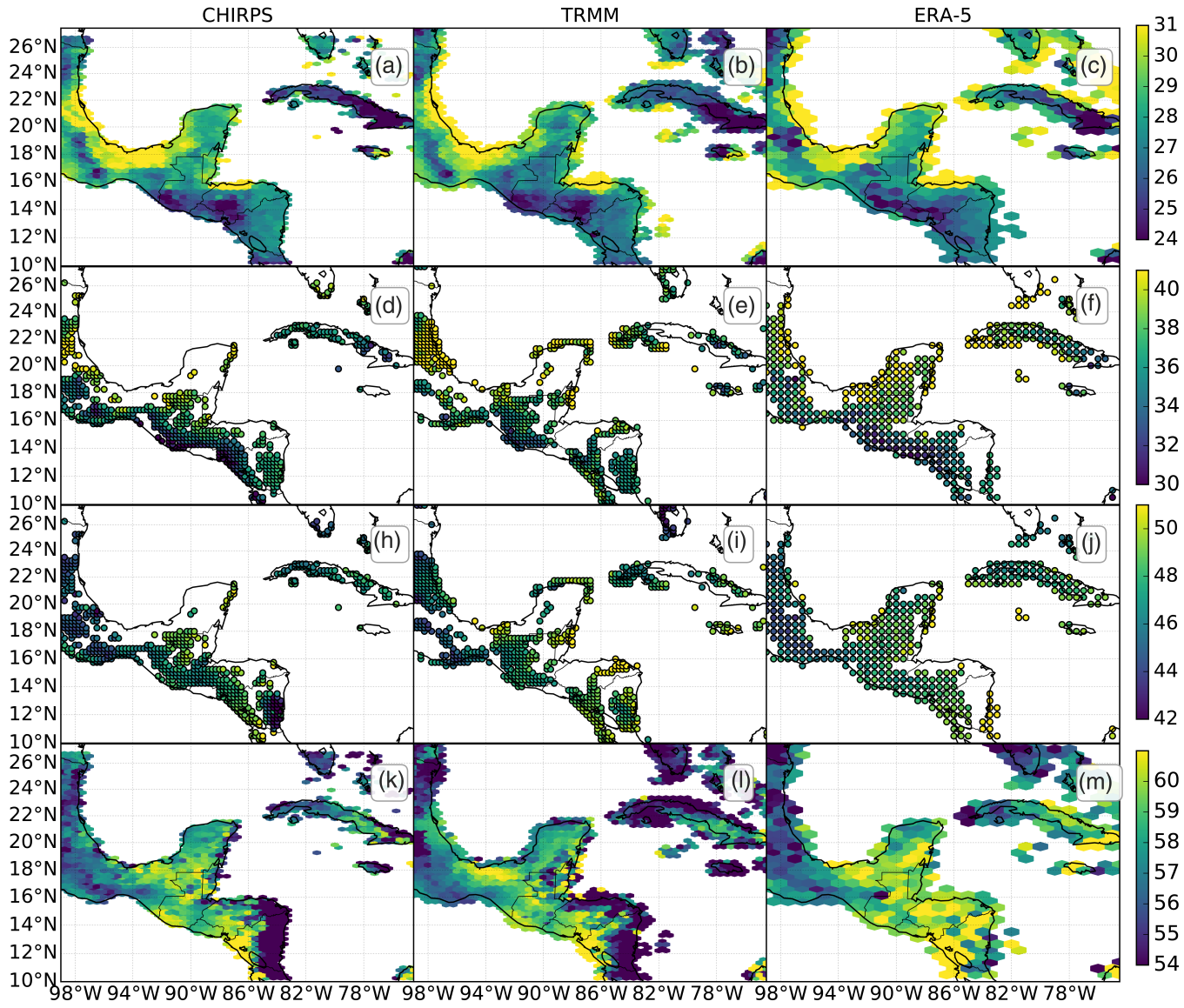

FIGURE 13 Spatial distribution of mean pentads of (a–c) onset, (k–m) retreat as well as (d–f) start and (h–j) end of the MSD in southern Mexico and northern Central America. (a, c, e, g) shows the results for the CHIRPS, (b, d, f, h) for TRMM and (c, f, j, m) for ERA-5 [Colour figure can be viewed at wileyonlinelibrary.com]

methods to find onset and retreat dates in three monsoon regions.

The method has a similar performance to existing methods that use precipitation thresholds in the North American Monsoon, as shown by the anomalies of precipitation, wind and geopotential anomalies around the onset and retreat dates. The spatial distribution of monsoon onset and retreat in this region was found to be sensibly captured by the wavelet algorithm, illustrating the earlier onset in central western Mexico and the later onset in northwestern Mexico, Arizona and New Mexico. The spatial distribution was shown to be consistent amongst the TRMM, CHIRPS and reanalysis data, suggesting that the method produces similar results in datasets with different climatologies even at the grid-box scale.

The WT method also compares well to a hydrologically defined index (HOWI) in the Indian Monsoon, although the WT better captures the precipitation variations and HOWI the moisture transport. However, the WT method is also able to capture strong differences in moisture transport around the onset and retreat dates, in both models and observations. The WT method obtains a later onset and earlier retreat than in the HOWI index, possibly associated with a lag between the moisture transport in the Arabian Sea (as diagnosed by HOWI) and the actual precipitation over mainland India (as measured by the WT method). The spatial

distribution of onset and retreat dates in the Indian Monsoon region using the WT method seem to be relatively consistent and coherent amongst the observational datasets, as the mean onset date in mainland India was found at pentad 32. Onset is earliest on the western coast of India and the onset date appears to be very homogeneous in central India.

The WT method was extended to characterize the timings and strength of the South American Monsoon bi-modal regime of precipitation and the intervening MSD period, using the same principle but additionally computing the transform over smaller dilations around the onset and retreat dates. By using randomly generated time-series, the spatial distribution of grid-points with a robust MSD signal was found in Cuba, the northwestern coast of Central America and several regions of south and north-eastern Mexico. The MSD in southern Mexico and northern Central America is found to start around pentads 35 and 36 (last week of June) and end around pentad 48 (mid-August) in most observational datasets and reanalysis. To our knowledge, this extension of the WT method provides one of very few methods to characterize the MSD on sub-monthly scales. This method may be potentially useful when diagnosing changes to the characteristics of the MSD in models or observations.

Current methods that diagnose monsoon onset and retreat using pentad or daily-mean precipitation time-series are typically rigid threshold methods. These threshold methods depend on a number of parameters that need to be tuned for a specific monsoon region and for a specific dataset. For instance, the method by G13 used a threshold value specific for the North American Monsoon and specific for the TRMM dataset but also to the limits of the area used for area-averaging the precipitation. In other words, the persistence and threshold values of most of the threshold methods require normalization, statistical treatment or additional tuning to the parameters to account for climatological differences in the datasets which introduces uncertainty. The method by A12 then uses a climatological mean value as the threshold, but in a climate model with a significantly positive bias in the dry winter season of a monsoon this method would be prone to error as the biased seasonal cycle may impose a biased calculation of monsoon onset and retreat.

The WT method is in many ways similar to the agronomical and threshold methods (e.g., Liebmann and Marengo, 2001; Moron and Robertson, 2014), as the implementation of the method uses a subjective determination of the dilation scales; these scales are comparable to the persistence and window parameters of the threshold methods. However, the WT method presented has three main advantages over most threshold methods. First, the method produces robust results for the Indian and North American monsoon of onset and retreat, and spatial distributions comparable to previous methods (Moron and Robertson, 2014) while not being subject to 'false-hits' nor years without an identification of the onset and retreat dates. In other words, the method provides robust results without requiring further treatment of years with false-hits or undetermined years.

The second advantage of the method is portability, or utility, as the method shows robust and consistent results for three observational datasets, a reanalysis and climate model experiments with varying climate forcing but without any constraint or treatment of the data beforehand and in three different regions with different seasonal cycles. In other words, this method is robust across datasets and regions. In contrast to rigid threshold techniques (e.g., Liebmann and Marengo, 2001), the identification of onset and retreat for each time-series, for example, at each grid-point, is based upon coherent temporal changes within each precipitation time-series while not using parameters determined "a priori" specifically for a region. The WT method then, can be used in any time-series, regardless of the origin of the time-series, without any further change or consideration than those established by the dilations scales determined in Section 2.2.1. The portability of the method also means that the method can be implemented as a "local-scale" method applied at the grid-box scale for high-resolution datasets such as CHIRPS as well as for regional scales using area-averaged time-series.

Third, and in contrast to typical threshold methods, the wavelet method can be applied to climate model output straightforward using the same configuration of dilation scales, a feature of the method that is illustrated by our analysis of several experiments using the Hadley Centre models. The treatment of the data does not require any normalization or statistical treatment even when used for grid-point time-series for different regions or experiments with varying forcing where the seasonal cycle or total annual rainfall may change notably within the model time-series.

While we do not suggest the WT method is better or superior than any others for all purposes, we argue that the method may be potentially useful for multiple purposes due to its portability and robustness. For instance, for climate model inter-comparison analyses which require the determination of onset and retreat in data that may have different seasonal cycles, magnitudes and resolutions. In addition to CMIP-style assessments, the method may be used in analyses of the observed trends in mean onset and retreat dates which may benefit from an independent method of validating the results from other methods or may use several datasets to support

their findings. In conclusion, the WT method is designed to be multi-purpose, to be used in varying contexts, from a local grid-scale analysis of the 'agronomical' onset or in global-scale analyses of the onset and retreat of the global monsoon while also being useful for the determination of bimodal regimes, such as those found in Central America and the Caribbean.

ACKNOWLEDGEMENTS

JGF was funded by an Oxford-Richards Scholarship in Wadham College. LG and SO acknowledge support from the National Centre for Atmospheric Science (NCAS) a research centre funded by the UK Natural Environment Research Council. The authors thank Robin Chadwick, Antje Weisheimer and Tim Woollings for their very helpful suggestions and useful discussions. Jorge L. Garciã-Franco was funded by an Oxford-Richards Scholarship. Lesley J. Grey and Scott Osprey were supported from the National Centre for Atmospheric Science (NCAS) of the Natural Environment Research Council (NERC). Lesley Grey and Scott Osprey were also supported by the NERC North Atlantic Climate System Integrated Study (ACSIS; grant no. NE/N018028) and NERC grant NE/P006779/1.

AUTHOR CONTRIBUTION

Jorge L. García-Franco conducted the analyses, Lesley J. Gray and Scott Osprey directed the research. All authors were fully involved in preparing and revising the text.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ORCID

Jorge L. García-Franco https://orcid.org/0000-0002-0396-9744
Scott Osprey https://orcid.org/0000-0002-8751-1211

REFERENCES

Addison, P.S. (2017) *The Illustrated Wavelet Transform Handbook: Introductory Theory and Applications in Science, Engineering, Medicine and Finance*. Boca Raton, FL: CRC Press.

Allen, T.L. and Mapes, B.E. (2017) The late spring caribbean rainbelt: climatology and dynamics. *International Journal of Climatology*, 37, 4981–4993.

Anderson, T.G., Anchukaitis, K.J., Pons, D. and Taylor, M. (2019) Multiscale trends and precipitation extremes in the central american midsummer drought. *Environmental Research Letters*, 14, 124016.

Andrews, M., Ridley, J., Wood, R., Andrews, T., Blockley, E., Booth, B., Burke, E., Dittus, A., Florek, P., Gray, L., Haddad, S., Hardiman, S., Hermanson, L., Hodson, D., Hogan, E., Jones, G., Knight, J., Kuhlbrodt, T., Misios, S. and Sutton, R. (2020) Historical simulations with HadGEM3-GC3.1 for CMIP6. *Journal of Advances in Modeling Earth Systems*, 12, e2019MS001995. https://doi.org/10.1029/2019MS001995.

Arias, P.A., Fu, R. and Mo, K.C. (2012) Decadal variation of rainfall seasonality in the north american monsoon region and its potential causes. *Journal of Climate*, 25, 4258–4274. https://doi.org/10.1175/JCLI-D-11-00140.1.

Bombardi, R.J., Moron, V. and Goodnight, J.S. (2020) Detection, variability, and predictability of monsoon onset and withdrawal dates: a review. *International Journal of Climatology*, 40, 641–667.

Brooks, I.M. (2003) Finding boundary layer top: application of a wavelet covariance transform to lidar backscatter profiles. *Journal of Atmospheric and Oceanic Technology*, 20, 1092–1105.

Bussmann, A., Elagib, N.A., Fayyad, M. and Ribbe, L. (2016) Sowing date determinants for sahelian rainfed agriculture in the context of agricultural policies and water management. *Land Use Policy*, 52, 316–328.

C3S (2017) ERA5: fifth generation of ECMWF atmospheric reanalyses of the global climate. Copernicus Climate Change Service Climate Data Store (CDS). Available at: https://climate.copernicus.eu/climate-reanalysis [Accessed 18th March 2019].

Carvalho, L.M., Silva, A.E., Jones, C., Liebmann, B., Dias, P.L.S. and Rocha, H.R. (2011) Moisture transport and intraseasonal variability in the South America monsoon system. *Climate Dynamics*, 36, 1865–1880.

Chevuturi, A., Turner, A.G., Woolnough, S.J., Martin, G.M. and MacLachlan, C. (2019) Indian summer monsoon onset forecast skill in the UKmet office initialized coupled seasonal forecasting system (glosea5-gc2). *Climate Dynamics*, 52, 6599–6617.

Cook, B.I. and Buckley, B.M. (2009) Objective determination of monsoon season onset, withdrawal, and length. *Journal of Geophysical Research: Atmospheres*, 114. https://doi.org/10.1029/2009JD012795

Dimdore-Miles, O., Gray, L. and Osprey, S. (2021) Origins of multi-decadal variability in sudden stratospheric warmings. *Weather and Climate Dynamics*, 2, 205–231.

Eyring, V., Bony, S., Meehl, G.A., Senior, C.A., Stevens, B., Stouffer, R.J. and Taylor, K.E. (2016) Overview of the coupled model intercomparison project phase 6 (cmip6) experimental design and organization. *Geoscientific Model Development (Online)*, 9, 1939–1958.

Fasullo, J. and Webster, P. (2003) A hydrological definition of indian monsoon onset and withdrawal. *Journal of Climate*, 16, 3200–3211.

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A. and Michaelsen, J. (2015) The climate hazards infrared precipitation with stations–a new environmental record for monitoring extremes. *Scientific Data*, 2, 150066.

Gadgil, S. (2018) The monsoon system: land–sea breeze or the itcz? *Journal of Earth System Science*, 127, 1.

Gadgil, S. and Rupa Kumar, K. (2006) *The Asian Monsoon—Agriculture and Economy*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 651–683.

Garcia, S.R. and Kayano, M.T. (2013) Some considerations on onset dates of the rainy season in western-Central Brazil with anti-symmetric outgoing longwave radiation relative to the equator. *International Journal of Climatology*, 33, 188–198.

García-Franco, J.L., Gray, L.J. and Osprey, S. (2020) The american monsoon system in HadGEM3 and UKESM1. *Weather and Climate Dynamics*, 1, 349–371.

Geil, K.L., Serra, Y.L. and Zeng, X. (2013) Assessment of cmip5 model simulations of the north american monsoon system. *Journal of Climate*, 26, 8787–8801.

Ha, K.-J., Moon, S., Timmermann, A. and Kim, D. (2020) Future changes of summer monsoon characteristics and evaporative demand over asia in CMIP6 simulations. *Geophysical Research Letters*, 47, e2020GL087492.

Harvey, C.A., Saborio-Rodríguez, M., Martinez-Rodríguez, M.R., Viguera, B., Chain-Guadarrama, A., Vignola, R. and Alpizar, F. (2018) Climate change impacts and adaptation among smallholder farmers in central america. *Agriculture & Food Security*, 7, 57.

Herrera, E., Magaña, V. and Caetano, E. (2015) Air–sea interactions and dynamical processes associated with the midsummer drought. *International Journal of Climatology*, 35, 1569–1578.

Htway, O. and Matsumoto, J. (2011) Climatological onset dates of summer monsoon over Myanmar. *International Journal of Climatology*, 31, 382–393.

Huffman, G.J., Adler, R.F., Bolvin, D.T. and Nelkin, E.J. (2010) The TRMM multi-satellite precipitation analysis (TMPA). In: Mekonnen, G. and Faisal, H. (Eds.) *Satellite rainfall applications for surface hydrology*. Dordrecht: Springer.

Jain, M., Naeem, S., Orlove, B., Modi, V. and DeFries, R.S. (2015) Understanding the causes and consequences of differential decision-making in adaptation research: adapting to a delayed monsoon onset in Gujarat, India. *Global Environmental Change*, 31, 98–109.

Karnauskas, K.B., Seager, R., Giannini, A. and Busalacchi, A.J. (2013) A simple mechanism for the climatological midsummer drought along the Pacific coast of Central America. *Atmosfera*, 26, 261–281.

Kitoh, A. and Uchiyama, T. (2006) Changes in onset and withdrawal of the east asian summer rainy season by multi-model global warming experiments. *Journal of the Meteorological Society of Japan Ser. II*, 84, 247–258.

Liebmann, B. and Marengo, J. (2001) Interannual variability of the rainy season and rainfall in the brazilian amazon basin. *Journal of Climate*, 14, 4308–4318.

Lucas-Picher, P., Christensen, J.H., Saeed, F., Kumar, P., Asharaf, S., Ahrens, B., Wiltshire, A.J., Jacob, D. and Hagemann, S. (2011) Can regional climate models represent the indian monsoon? *Journal of Hydrometeorology*, 12, 849–868.

Magaña, V., Amador, J.A. and Medina, S. (1999) The midsummer drought over Mexico and Central America. *Journal of Climate*, 12, 1577–1588.

Marengo, J.A., Liebmann, B., Kousky, V.E., Filizola, N.P. and Wainer, I.C. (2001) Onset and end of the rainy season in the brazilian amazon basin. *Journal of Climate*, 14, 833–852.

Menary, M.B., Kuhlbrodt, T., Ridley, J., Andrews, M.B., Dimdore-Miles, O.B., Deshayes, J., Eade, R., Gray, L., Ineson, S., Mignot, J., Roberts, C.D., Robson, J., Wood, R.A. and Xavier, P. (2018) Preindustrial control simulations with hadgem3-gc3. 1 for cmip6. *Journal of Advances in Modeling Earth Systems*, 10, 3049–3075.

Moron, V. and Robertson, A.W. (2014) Interannual variability of indian summer monsoon rainfall onset date at local scale. *International Journal of Climatology*, 34, 1050–1061.

Mosiño, A. and García, E. (1966) Evaluación de la sequía intraestival en la república mexicana. In Proceedings of Conference on Reg. Latinoamericana Unión Geogr. Int. pp. 500–516.

Nieto-Ferreira, R. and Rickenbach, T.M. (2011) Regionality of monsoon onset in South America: a three-stage conceptual model. *International Journal of Climatology*, 31, 1309–1321.

Pascale, S., Carvalho, L.M., Adams, D.K., Castro, C.L. and Cavalcanti, I.F. (2019) Current and future variations of the monsoons of the Americas in a warming climate. *Current Climate Change Reports*, 5, 125–144.

Perdigón-Morales, J., Romero-Centeno, R., Ordóñez, P. and Barrett, B.S. (2018) The midsummer drought in Mexico: perspectives on duration and intensity from the chirps precipitation database. *International Journal of Climatology*, 38, 2174–2186.

Ridley, J., Menary, M., Kuhlbrodt, T., Andrews, M. and Andrews, T. (2018) Mohc hadgem3-gc31-ll model output prepared for cmip6 cmip picontrol. https://doi.org/10.22033/ESGF/CMIP6.6294.

Ridley, J., Menary, M., Kuhlbrodt, T., Andrews, M. and Andrews, T. (2019) Mohc hadgem3-gc31-mm model output prepared for cmip6 cmip picontrol. https://doi.org/10.22033/ESGF/CMIP6.6297.

Sahana, A.S., Ghosh, S., Ganguly, A. and Murtugudde, R. (2015) Shift in indian summer monsoon onset during 1976/1977. *Environmental Research Letters*, 10, 054006.

Sellar, A.A., Jones, C.G., Mulcahy, J., Tang, Y., Yool, A., Wiltshire, A., O'Connor, F.M., Stringer, M., Hill, R., Palmieri, J., Woodward, S., de Mora, L., Kuhlbrodt, T., Rumbold, S., Kelley, D.I., Ellis, R., Johnson, C.E., Walton, J., Abraham, N.L., Andrews, M.B., Andrews, T., Archibald, A.T., Berthou, S., Burke, E., Blockley, E., Carslaw, K., Dalvi, M., Edwards, J., Folberth, G.A., Gedney, N., Griffiths, P.T., Harper, A.B., Hendry, M.A., Hewitt, A.J., Johnson, B., Jones, A., Jones, C.D., Keeble, J., Liddicoat, S., Morgenstern, O., Parker, R.J., Predoi, V., Robertson, E., Siahaan, A., Smith, R.S., Swaminathan, R., Woodhouse, M.T., Zeng, G. and Zerroukat, M. (2019) UKESM1: description and evaluation of the UK Earth system model. *Journal of Advances in Modeling Earth Systems*, 11, 4513–4558.

Sultan, B., Baron, C., Dingkuhn, M., Sarr, B. and Janicot, S. (2005) Agricultural impacts of large-scale variability of the west african monsoon. *Agricultural and Forest Meteorology*, 128, 93–110.

Turner, A.G. and Annamalai, H. (2012) Climate change and the south asian summer monsoon. *Nature Climate Change*, 2, 587–595.

Vera, C., Higgins, W., Amador, J., Ambrizzi, T., Garreaud, R., Gochis, D., Gutzler, D., Lettenmaier, D., Marengo, J., Mechoso, C.R., Nogues-Paegle, J., Dias, P.L.S. and Zhang, C. (2006) Toward a unified view of the American monsoon systems. *Journal of Climate*, 19, 4977–5000.

Wang, P.X., Wang, B., Cheng, H., Fasullo, J., Guo, Z., Kiefer, T. and Liu, Z. (2017) The global monsoon across time scales: mechanisms and outstanding issues. *Earth-Science Reviews*, 174, 84–121.

Webster, P.J. and Yang, S. (1992) Monsoon and enso: selectively interactive systems. *Quarterly Journal of the Royal Meteorological Society*, 118, 877–926.

Whitcher, B., Guttorp, P. and Percival, D.B. (2000) Wavelet analysis of covariance with application to atmospheric time series. *Journal of Geophysical Research: Atmospheres*, 105, 14941–14962.

Xie, P. and Arkin, P.A. (1997) Global precipitation: a 17-year monthly analysis based on gauge observations, satellite

estimates, and numerical model outputs. *Bulletin of the American Meteorological Society*, 78, 2539–2558.

Zeng, X. and Lu, E. (2004) Globally unified monsoon onset and retreat indexes. *Journal of Climate*, 17, 2241–2248.

Zhang, S. and Wang, B. (2008) Global summer monsoon rainy seasons. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, 28, 1563–1578.

Zhao, Z., Holbrook, N.J., Oliver, E.C., Ballestero, D. and Vargas-Hernandez, J.M. (2020) Characteristic atmospheric states during mid-summer droughts over central america and Mexico. *Climate Dynamics*, 55, 681–701.

Zhao, Z. and Zhang, X. (2020) Evaluation of methods to detect and quantify the bimodal precipitation over central america and Mexico. *International Journal of Climatology*, 41, E897–E911.

Zhou, T., Turner, A.G., Kinter, J.L., Wang, B., Qian, Y., Chen, X., Wu, B., Liu, B., Zou, L. and Bian, H. (2016) GMMIP (v1. 0) contribution to CMIP6: global monsoons model inter-comparison project. *Geoscientific Model Development*, 9, 3589–3604.

Zou, L. and Zhou, T. (2015) Asian summer monsoon onset in simulations and cmip5 projections using four chinese climate models. *Advances in Atmospheric Sciences*, 32, 794–806.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: García-Franco JL, Osprey S, Gray LJ. A wavelet transform method to determine monsoon onset and retreat from precipitation time-series. *Int J Climatol*. 2021;41: 5295–5317. https://doi.org/10.1002/joc.7130