

A multi-decadal assessment of the performance of gauge- and model-based rainfall products over Saudi Arabia: climatology, anomalies and trends

Ahmed M. El Kenawy^{a,b,*} and Matthew F. McCabe^a

^a *Water Desalination and Reuse Center, Division of Biological and Environmental Sciences and Engineering, King Abdullah University of Science and Technology, Jeddah, Saudi Arabia*

^b *Department of Geography, Mansoura University, Egypt*

ABSTRACT: Many arid and semi-arid regions have sparse precipitation observing networks, which limits the capacity for detailed hydrological modelling, water resources management and flood forecasting efforts. The objective of this work is to evaluate the utility of relatively high-spatial resolution rainfall products to reproduce observed multi-decadal rainfall characteristics such as climatologies, anomalies and trends over Saudi Arabia. Our study compares the statistical characteristics of rainfall from 53 observatories over the reference period 1965–2005, with rainfall data from six widely used gauge-based products, including APHRODITE, GPCC, PRINCETON, UDEL, CRU and PREC/L. In addition, the performance of three global climate models (GCMs), including CCSM4, EC-EARTH and MRI-I-CGCM3, integrated as part of the Fifth Coupled Model Intercomparison Project (CMIP5), was also evaluated. Results indicate that the gauge-based products were generally skillful in reproducing rainfall characteristics in Saudi Arabia. In most cases, the gauge-based products were also able to capture the annual cycle, anomalies and climatologies of observed data, although significant inter-product variability was observed, depending on the assessment metric being used. In comparison, the GCM-based products generally exhibited poor performance, with larger biases and very weak correlations, particularly during the summertime. Importantly, all products generally failed to reproduce the observed long-term seasonal and annual trends in the region, particularly during the dry seasons (summer and autumn). Overall, this work suggests that selected gauge-based products with daily (APHRODITE and PRINCETON) and monthly (GPCC and CRU) resolutions show superior performance relative to other products, implying that they may be the most appropriate data source from which multi-decadal variations of rainfall can be investigated at the regional scale over Saudi Arabia. Discriminating these skillful products is important not only for reducing uncertainty in climate, hydrological and environmental assessments but also for advancing model developments in the region.

KEY WORDS rainfall; gauge-based products; global climate models; trends; cross-tabulation; Saudi Arabia

Received 5 December 2014; Revised 20 March 2015; Accepted 20 April 2015

1. Introduction

Rainfall is a key component of the hydrological cycle, influencing runoff generation, groundwater recharge, soil moisture and evaporation, as well as impacting numerous components of the biosphere and human activities (Lorenz and Kunstmann, 2012). Given its role in modulating the formation and persistence of droughts and floods, as well as influencing a range of land–atmosphere interactions and processes, accurate characterization and representation of the spatial and temporal distribution of rainfall, together with its changing patterns, are of considerable interest.

Changes in the global climate over recent decades have been the subject of much investigation. Numerous studies

have highlighted changes in rainfall at various spatial scales, ranging from the global (Trenberth, 2011) to the hemispheric (Polson *et al.*, 2014), continental (Cortesi *et al.*, 2012) and regional scales (Black *et al.*, 2010). In the Middle East, Zhang *et al.* (2005) assessed observed changes in precipitation extremes from 1950 to 2003, concluding that the rates of change were spatially inconsistent over the region. Recently, Donat *et al.* (2014) studied the spatial and temporal variability of precipitation in the Arab region, suggesting insignificant changes since the mid 20th century. Nonetheless, for Saudi Arabia and the Middle East region in general, assessing the responses of rainfall to ongoing climate change has not been well characterized. This is due, in part, to the lack of a complete, long-term, spatially distributed and reliable climate data set. Issues such as the unrepresentative gauge placement, discontinuous temporal records, ageing infrastructure and the inherent costs of manipulating an observation network all complicate the task of developing a consistent data record.

In spite of these limitations, some attempts to analyse the spatial and temporal variability of rainfall and their driving

* Correspondence to: A. M. El Kenawy, Water Desalination and Reuse Center, Division of Biological and Environmental Sciences and Engineering, King Abdullah University of Science and Technology, Thuwal/Jeddah 23955-6900, Kingdom of Saudi Arabia. E-mail: kenawy@mans.edu.eg

forces have been made in Saudi Arabia. Most of the previous rainfall studies throughout the country either focus on short-term (< 30-year) analysis (e.g. AlSarmi and Washington, 2011; Almazroui *et al.*, 2014), use a limited number of spatially sparse observatories (e.g. Almazroui *et al.*, 2012; Donat *et al.*, 2014), focus on specific regions rather than the whole country (e.g. Abdullah and Almazroui, 1998; Rehman, 2010) or consider a single spatially interpolated data set (Almazroui *et al.*, 2012). For example, Almazroui *et al.* (2014) employed monthly rainfall data at 27 synoptic stations from 1985 to 2010 to classify Saudi Arabia into homogenous pluviometric sub-regions using a multivariate statistical approach. Their results indicated that the spatial variability of rainfall in the region is significantly influenced by physiography. Almazroui *et al.* (2012) also studied annual rainfall trends in the Arabian Peninsula for the period 1978–2009, finding evidence of decline ($35.1 \text{ mm decade}^{-1}$) during the wet season. Most recently, El Kenawy *et al.* (2014) investigated the connection between synoptic weather types and the intensity of rainfall over Saudi Arabia. They found a generally negative trend in patterns favouring above-normal rainfall in the region (e.g. cyclones).

Even considering these important contributions, there remain uncertainties associated with the assessment of rainfall variations in Saudi Arabia. Unfortunately, the majority of past and current rain gauge data cannot be used for reliable diagnosis of long-term rainfall variability, due to their relatively short duration, issues with temporal sampling and consistency and/or limited spatial coverage. In addition, rainfall is highly variable in space and time over Saudi Arabia as a consequence of the complex interactions between different dynamic processes, including mid-latitudes and tropical configurations (e.g. the Red Sea Trough, the Mediterranean and monsoonal influences). Detecting or assessing statistically significant behaviour in rainfall demands that data be accurately represented over space and time and with sufficient length to be able to infer inter-decadal scale variations.

With advances in computational performance and data assimilation techniques, a number of regional and global gridded meteorological products have been released to the community. Many studies have incorporated data from these products to initialize a wide variety of hydrological and environmental applications and to study implications of climate variability across different regions worldwide. For example, Raziei *et al.* (2011) assessed the spatial and temporal variability of drought from 1951 to 2005 over Iran using GPCC (V4), finding good agreement with both observations and NCEP/NCAR reanalysis. Similarly, Prakash *et al.* (2014) compared six gridded rainfall data sets against observations over India during the monsoon season (June–September, JJAS). In Saudi Arabia, few efforts have been undertaken to assess uncertainties and implications associated with gridded products in the region comprehensively. In any data-scarce country, evaluating gauge- and global climate model (GCM)-based products against available observations is useful from a number of perspectives. First, it provides an option to use spatially

and temporally continuous datasets, allowing for a detailed climatological assessment in the region. Second, good agreement between *in situ* data and some of these products allows a broader perspective to study local and larger scale dynamics associated with wet events in the region. Third, defining skillful products is important not only for reducing uncertainty in climate and hydrological assessments but also for advancing model developments and improving on model physics (Deng *et al.*, 2015). Finally, a realistic rainfall data set can provide an input to initialize a wide variety of hydrological and environmental applications (e.g. hydrological modelling, streamflow forecasts, agricultural and water resources management) and can thus help decision makers to adopt appropriate policies for impact mitigation and adaptation.

The overriding aim of this study is to assess the quality and efficiency of six gauge-based and three model-based gridded products to represent the observed rainfall characteristics over Saudi Arabia. In particular, this study quantitatively assesses the differences in rainfall climatologies, anomalies and trends between observations and the different products for a common period 1965–2005. The products used in this work have a relatively high-spatial resolution of approximately 1° or less and include APHRODITE v1101 (Yatagai *et al.*, 2012), GPCC v6 (Schneider *et al.*, 2011), PRINCETON (Sheffield *et al.*, 2006), UDEL v3.01 (<http://www.esrl.noaa.gov/psd/>), CRU v3.22 (Harris *et al.*, 2014) and PREC/L (Chen *et al.*, 2002). We also use rainfall data from three GCMs, including CCSM4 v4 (Gent *et al.*, 2011), EC-Earth (Hazeleger *et al.*, 2010) and MRI-CGCM3 (Yukimoto *et al.*, 2012). While retrievals from the satellite era (e.g. TRMM, PERSIANN, CSMaP and MPE) provide additional multi-spatial and multi-temporal data sources, they were not considered here as the products are not available for the entirety of the study period, limiting an evaluation of climatological features. Similarly, global reanalysis products were excluded in our analysis due to either their coarse spatial resolution (e.g. ERA40, NCEP/NCAR v1, GPCP, TARCAT v2 and 20CR v2) or incompatible temporal coverage (e.g. ERA-Interim, MERRA and JRA-25). Overall, research on characterizing the uncertainties associated with the gauge- and GCM-based products is still lacking in Saudi Arabia, and this study presents the first attempt to provide an examination of the agreement between observed data and this number of products in the region.

2. Study area

Saudi Arabia lies in southwestern Asia between the latitudes of $15^\circ 22' \text{ N}$ and $32^\circ 09' \text{ N}$ and longitudes of $34^\circ 50' \text{ E}$ and $55^\circ 50' \text{ E}$ (see Figure 1). The study domain is located in one of the driest regions of the world, reflected by low rainfall totals averaging $100\text{--}120 \text{ mm year}^{-1}$. Rainfall climatology is characterized by short, intense, infrequent and very local wet events, which generally occur from October to May, with June–September representing

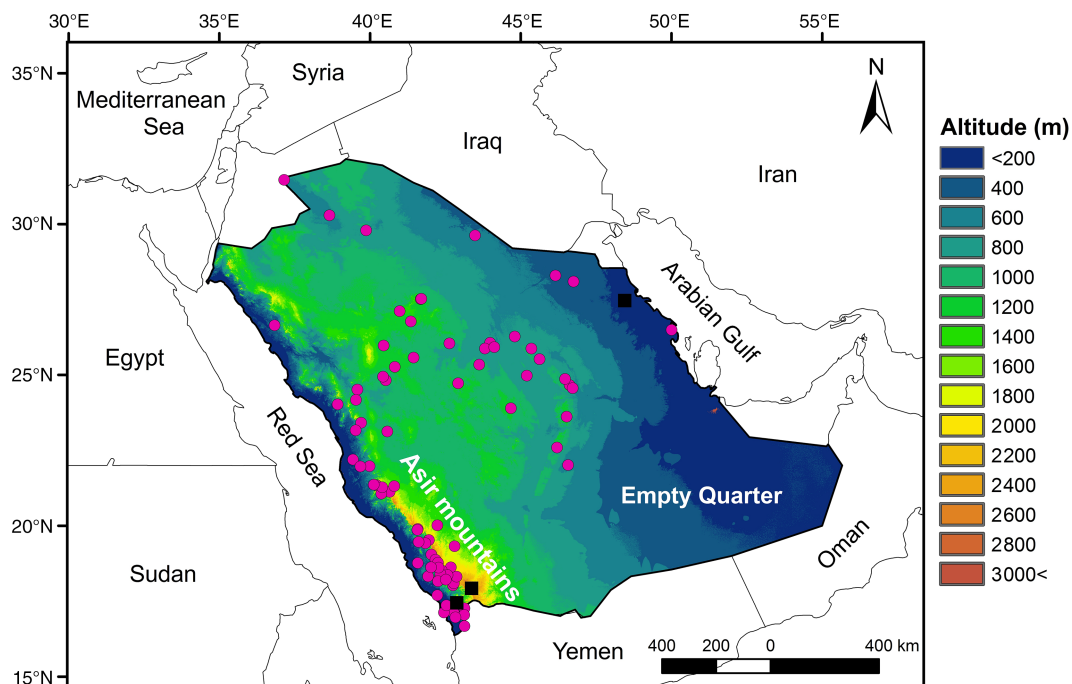


Figure 1. Location of the study area and the spatial distribution of the rain gauges. The black squares denote the three locations given in Figure 2.

the driest months. Rainfall is described by its high variability in space and time. The heterogeneous spatial distribution of rainfall presents a greater rainfall frequency (and total) over the southwest and the west, relatively low rainfall in the central and northern regions and almost no rainfall in the southeast (the Rub al Khali, or ‘empty quarter’ region). Figure 2 compares monthly variations of rainfall totals (mm) at three different sites across the region. Rainfall maxima vary markedly among the observatories, with the highest amounts recorded at Al Nuayra station (eastern province) during winter, at Horoub station (Jazan province) in summer and early autumn and at Al Harja (Assir-Abha province) in spring. These spatial and temporal variations can be explained by the interactions between mid-latitude and subtropical configurations including, among others, the Red Sea Trough (Goldreich, 2003) as well as Mediterranean and monsoonal influences. However, these spatial variations also give indications on a considerable low spatial dependency among rainfall observatories, even the nearby sites (e.g. Horoub and Al Harja observatories).

3. Data and methods

3.1. Rainfall observations and products

3.1.1. Observational data (OBS)

As a reference for comparisons, we use daily records of rainfall totals (mm) from a network of meteorological stations. The data were provided by the Saudi Ministry of Water and Electricity (MOWE) for the period 1960–2012. Observatories were obtained from a network of over 350 rain gauges and constrained based on their consistency and

time stability. As data gaps can have a detrimental influence on long-term assessments, the analysis was restricted to 53 stations with largely complete records. More specifically, stations with less than 5% of values missing during the period 1965–2005 were selected. Figure 3 shows the temporal evolution of the number of rain gauges in Saudi Arabia and the percentage of missing values for the selected stations with respect to the entire network. Given that metadata that might indicate changes in instruments, calibration techniques or relocations of rain gauges is unavailable, the data were screened for the presence of erroneous records originating from data entry (e.g. negative values, non-existent dates, daily amount > 200 mm). Here, no attempts were made to test the homogeneity of the time series because the current network of rain gauges shows low spatial and temporal dependencies, even among nearby observatories, which constrains any homogenization protocol. The spatial distribution of the 53 rain gauges used in this work is illustrated in Figure 1. As shown, the station elevations range between 6 and 2648 m above sea level, with 20 rain gauges located above 1000 m, capturing the different climate conditions in both high and low altitudes and also representing differences in the maritime and continental weather regimes across the region.

3.1.2. Gauge-based gridded data

As indicated in Table 1, we evaluated six major gauge-based products for Saudi Arabia. These products cannot be considered completely independent, as they use overlapping rain gauge observations from the region. Nonetheless, they employ varying numbers of rain gauges as well as different interpolation models to re-scale the data to a regular grid. For example, APHRODITE uses

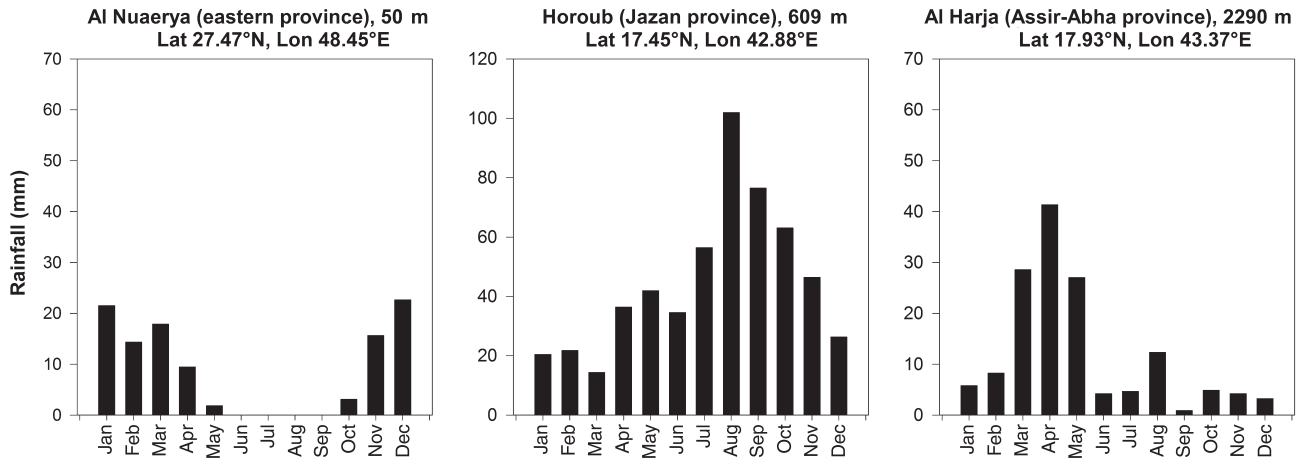


Figure 2. Monthly variations of rainfall totals (mm) at three different stations in the study domain. Long-term averages were calculated for the base period 1965–2005. The altitude (in metres above mean sea level), the latitude and the longitude of each station are provided.

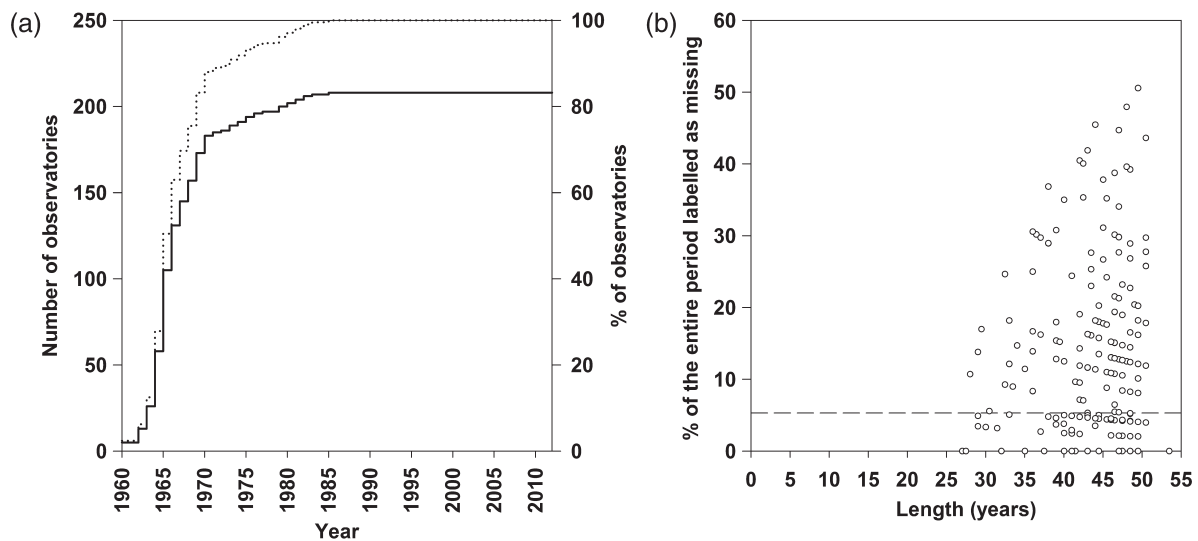


Figure 3. (a) The absolute and relative cumulative distribution of the number of rain gauges in Saudi Arabia from 1960 to 2012, (b) the percentage of missing values in relation to the total count of records for each observatory, the dashed line shows the threshold of missing values used in this study.

the weighted mean interpolation method as a means of data integration, whereas UDEL employs the climatologically aided interpolation (CAI) approach (Willmott and Matsuura, 1995). The changing number of observations and the variety of interpolation techniques employed can result in significant differences between the gauge-based gridded data, which may not just limit the utility of some of them but might also complicate the identification of climate signal uncertainty. The six gauge-based products are briefly described below:

- APHRODITE (v1101) was developed within the Asian Precipitation-Highly Resolved Observational Data Integration Towards the Evaluation of Water Resources (APHRODITE) project, representing the first comprehensive long-term gridded data undertaken in Asia (Yatagai *et al.*, 2012). In this comparison, we examine the daily gridded precipitation data (V1101) for the Middle East, which were provided at a spatial resolution

of 0.25° for the period 1951–2007. The data set was compiled through interpolating a collection of gauge observations in the region, in addition to measurements available through the World Meteorological Organization (WMO) Global Telecommunication System (GTS). Generally, The APHRODITE project collected data from 2.3 to 5.5 times the number of rain gauges of GTS (Yatagai *et al.*, 2012).

- CRU (v3.22) is a monthly global rainfall data set provided by the Climatic Research Unit (CRU) at the University of East Anglia (UK) and produced at a 0.5° grid resolution for the period 1901–2010 (Harris *et al.*, 2014). This product uses 5159 meteorological land stations worldwide.
- GPCC (v6) is a global product developed by the Global Precipitation Climatology Centre (GPCC) (Schneider *et al.*, 2011). Similar to APHRODITE, the time series were developed from multiple sources of observations by optimally merging monthly land

surface precipitation from 67 200 rain gauges based on the GTS and historical records (with at least ten complete years of data). Here, we use the latest version of GPCC (V6), which covers the period from 1901 to 2010 with a 0.5° spatial resolution. Further details on the development of this product can be found in a series of GPCC project reports available online at <http://gpcc.dwd.de>.

- PREC/L, NOAA's Precipitation Reconstruction data set over Land (PREC/L), was developed on a monthly scale for the period 1948 to present (Chen *et al.*, 2002). Precipitation is available at three spatial resolutions, with the 0.5° resolution version used in this work. This global product was compiled by interpolating measurements from 17 000 rain gauges over land collected from both the Global Historical Climate Network (GHCN) version 2 and the Climate Anomaly Monitoring System (CAMS).
- PRINCETON is a global meteorological forcing data set developed by the terrestrial hydrology research group at Princeton University, which is used as driving data for land surface models and other terrestrial modelling systems (Sheffield *et al.*, 2006). The product provides near-surface meteorological data, using a combination of both reanalysis and observational data. In this analysis, we use the 0.25° resolution version, which is available on a daily basis for the period 1948–2010.
- UDEL (v3.01) is a product developed by the University of Delaware (USA) for the period from 1901 to 2010. The result consists of monthly gridded high-resolution land data for precipitation, covering the globe at a spatial resolution of 0.5°. UDEL was created from a large number of rain gauges, GHCN version 2 and the archive of Willmott and Matsuura (1995). Further details and information on this data set can be found at <http://www.esrl.noaa.gov/psd/>.

3.1.3. GCM data

In addition to evaluating gauge-based products, we also examined model simulations for which daily rainfall fields are available at a reasonable spatial scale. In particular, we integrated three GCM outputs for the period 1965–2005 that included CCSM4, EC-Earth and MRI-I-CGCM3. These models formed part of the Fifth Coupled Model Intercomparison Project (CMIP5) for the Intergovernmental Panel on Climate Change (IPCC) Fifth Assessment Report (AR5). From the numerous CMIP5/AR5 models (Taylor *et al.*, 2012), we only selected those with the smallest grid spacing (approximately 1° or less, see Table 1). The available simulations use state-of-the-art parameterizations for short-wave and long-wave radiation, clouds, cumulus convection, boundary layer and land surface processes. Table 2 summarizes the main characteristics of the three models and their physics. A comprehensive review of the CMIP5 models used herein is described in Taylor *et al.* (2012).

Table 1. Overview of the characteristics of the nine different rainfall products used in this study.

Product	Grid size (Lat × Lon)	Time resolution	Time period	Type	Institute	References
Gauge-based APHRODITE	0.25° × 0.25°	Daily	1951–2007	Gauge-based (weight mean interpolation method)	Japanese Meteorological Research Institute	Yatagai <i>et al.</i> (2012)
CRU TS 3.2	0.5° × 0.5°	Monthly	1901–2013	Gauge-based	Climate Research Unit	Harris <i>et al.</i> (2014)
GPCC-v6	0.5° × 0.5°	Monthly	1901–2010	Merged (WMO GTS, SYNOP and monthly climate reports)	Global Precipitation Climatology Centre	Schneider <i>et al.</i> (2011)
PREC/L	0.5° × 0.5°	Monthly	1948–present	Gauge-based (GHCN v2, CAMS)	NOAA/OAR/ESRL PSD, Boulder, Colorado, USA	Chen <i>et al.</i> (2002)
PRINCETON	0.25° × 0.25°	Daily	1948–2010	Merged (reanalysis and observational)	Princeton University	Sheffield <i>et al.</i> (2006)
UDEL	0.5° × 0.5°	Monthly	1901–2010	Merged [climatologically aided interpolation (CAD)]	NOAA/OAR/ESRL PSD, Boulder, Colorado, USA	–
GCMs CCSM4	1.25° × 1°	Daily	1850–2005	GCM atmospheric component model	National Center for Atmospheric Research, USA	Gent <i>et al.</i> (2011)
EC-Earth	1.125° × 1.12°	Daily	1850–2005	GCM atmospheric component model	EC-Earth Consortium	Hazeleger <i>et al.</i> (2010)
MRI-I-CGCM3	1.1° × 1.1°	Daily	1850–2005	GCM atmospheric component model	Meteorological Research Institute, Japan	Yukimoto <i>et al.</i> (2012)

Table 2. Overview of the physics parameterization and configurations of the GCMs used in this study.

Model	AGCM	OGCM	Initialization		Perturbation		Aerosol	
			Atmosphere/ land	Ocean	Atmospheric	Ocean	Concentration (C)/emission (E)	Direct (D)/ Indirect (I1,I2)
CCSM4	1.3°L26	1°L60	NO	Ocean assimilation (POPDART)	Single atm from AMIP run	Ensemble assimilation	E	D
EC-Earth	1.1°L62	1°L42	ERA40/Interim	Ocean assimilation (ORAS4/ NEMOVAR S4)	Start dates with singular vectors	Ensemble ocean assim. (NEMOVAR)	C	D
MRI-I-CGCM3	1.1°L48	1°L51	NO	SST	Start dates and ensemble assimilation		E	D, I1, I2

In contrast to gauge-based products, GCMs are largely independent from observations, as they seek to explicitly resolve the physical processes that govern mass, momentum and energy transfer in the coupled Earth system. Accordingly, GCMs are not expected to reproduce the day-to-day variations of the climate system, especially when compared directly to observations or other observation-driven modelling approaches. However, in order to evaluate the capacity of GCMs to simulate future climate states, they should display some ability to reproduce the predominant climatological features in retrospective simulations. For this reason, the current study focuses on assessing the GCMs capacity to reproduce key characteristics of the climate, such as anomalies, trends and climatology, features that are critical in understanding the larger scale and longer term hydroclimatology of a region.

3.2. Analysis procedure

3.2.1. Performance evaluation statistics

Both the gauge-based and GCM products have certain limitations, which may originate from a combination of the forecast model, data integration or data sources (e.g. local station errors) (Hodges *et al.*, 2011; Lin *et al.*, 2014). Therefore, it is important to identify the quality of the products prior to using them for assessing rainfall variations in the region. Identifying the limitations of these products can establish a baseline for understanding how they are restricted and what types of improvements could be made to provide more consistent climate information that can be employed for reliable broader scale analyses. To do this, rainfall measurements from the gauge-based and GCM products were compared against observational data to find those that best represent rainfall characteristics in the region. The comparison was restricted to the 41-year period of overlap between all products between 1965 and 2005. A base level evaluation was performed by directly comparing the data from each single observatory with the data from the nearest grid box for each product. It should be noted that we avoid interpolating the observed data for a number of reasons. First, the grid-points from the individual products do not have exactly the same locations, as the products have varied spatial resolutions. Second, the patterns of rainfall can vary significantly at the site

level as a function of the chosen interpolation algorithm. Despite making best use of all available observations in Saudi Arabia, the lack of station density in some areas, the local character of wet events and the terrain characteristics all combine to complicate the results of any interpolation algorithm in the region.

The agreement between observations and the different rainfall products is investigated using a set of commonly used statistical estimators, including the bias, the mean absolute error (MAE), standard deviations of differences (σ_{diff}) and Spearman ρ correlation. For each grid point, all comparisons were made for monthly, seasonal and annual rainfall totals (mm) over the 41-year period. It should be noted that no attempts were made to evaluate the capacity of the different products to reproduce a wet/dry state or extreme events in the region, because the products were provided at different temporal scales, which constrains any analysis of daily extreme events (Table 1).

The bias was calculated as a measure of average error between the observed (O) and modelled (P) data, providing context on whether the observed values are being overestimated or underestimated. This accuracy estimator is written simply as:

$$Bias = N^{-1} \sum_{i=1}^N (P_i - O_i) \tag{1}$$

where N is the sample size, O is the observed value, P is the modal value, and i is the counter for grid boxes (i.e. the observed and model values for a particular grid box at a definite time step).

The MAE is calculated as the average of the absolute difference between observed and predicted data, and formulated as:

$$MAE = N^{-1} \sum_{i=1}^N |P_i - O_i| \tag{2}$$

The standard deviation of differences (σ_{diff}) was used to measure changes in variability. This measure uses the standard deviation of the differences between the records of two products and is computed following:

$$\sigma_{diff} = \sqrt{\left(\sigma_p^2/N_p + \sigma_o^2/N_o\right)} \tag{3}$$

We also used the Spearman's (ρ) rank correlation to assess the capability of the different products to capture the inter-annual variability of rainfall in the region. The ρ coefficient is given by:

$$\rho = \frac{6 \sum d_i^2}{N(N^2 - 1)} \quad (4)$$

where d_i is the difference between the ranks of corresponding values of rainfall, and N is the number of values in the data set.

3.2.2. Trends

The ability of the different products to characterize the spatial and temporal variations of rainfall pattern in the region was examined by analysing the long-term (1965–2005) trends for the time series of seasonal and annual rainfall totals (mm). Given that rainfall distribution in arid and semi-arid regions is highly skewed (Jackson, 1977), it is inappropriate to assume a normal distribution function of the time series data. Therefore, inter-annual trends in the seasonal and annual accumulated rainfall series were calculated by applying linear regression analysis, while the non-parametric Mann–Kendall (MK) trend test was used to test the significance of the resulting trends (Kendall, 1975). Non-parametric tests are usually more robust compared with parametric ones, among which the MK test has been used by many researchers to evaluate the significance of monotonic trends in many hydrological and meteorological applications (e.g. Zhang *et al.*, 2001; Liu *et al.*, 2013; El Kenawy *et al.*, 2014).

In order to test the ability of the gauge-based and GCM products to reproduce the sign of the observed trends, we compared and summarized the detected trends of rainfall by means of cross-tabulation and asymmetry measures (Gonzalez-Hidalgo *et al.*, 2003). The rationale behind this procedure was to statistically explore pairwise relationships between the categorical trends. Such a comparison can provide insights into whether trends are consistent among the different products in the region. Here, we used the rain gauges and grids as the random variable, while the categories of the trends were the characteristics of the random variable. More specifically, for each data set, we classified the trends into four groups according to their statistical significance ($p < 0.05$) (i.e. statistically negative, statistically positive, insignificant positive and insignificant negative). Pivot tables were constructed to represent the cross-categorized frequency data in a matrix format following the results of the trend analysis. To test the association between the trends of the observed rainfall and those of each data set, we converted the contingency table into a frequency matrix of one row and one column for each trend category using the absolute frequencies. Then, the strength (degree) of association was assessed and compared using two different estimators: the coefficient of contingency (Clark and Hosking, 1986) and phi coefficient (Agresti, 1996). The larger the value of these coefficients, the greater the degree of association. The significance of this association was assessed by means of

the non-parametric chi-square (χ^2) test with degrees of freedom of $(r - 1)^2$. A similar approach has been used to measure the spatial consistency of climate variables in previous studies (e.g. Gonzalez-Hidalgo *et al.*, 2003; López-Moreno *et al.*, 2010).

4. Results and discussion

4.1. Validation results

The performance of a set of multiple gauge-based products and GCM simulations in reproducing inter-annual characteristics of rainfall over Saudi Arabia is evaluated using the same commonly used statistical metrics described in Section 3.2 (Vicente-Serrano *et al.*, 2003; El Kenawy *et al.*, 2015). The comparison is performed on seasonal and annual time scales for a common period 1965–2005, considering the accumulated rainfall series. Table 3 summarizes the validation results for all seasons by comparing observed mean rainfall values with rainfall data from the different products for the entire study domain. The results indicate that the majority of the gauge-based products are skillful in reproducing rainfall amounts in Saudi Arabia across all seasons. Results generally show small biases at the annual scale on the order of 0.5 mm for GPCC, -0.7 mm for CRU, -1.6 mm for PRINCETON and 3.4 mm for PREC/L. UDEL presents the poorest performance among all gauge-based products, with an average annual bias of -7.5 mm. The low biases on the annual scale are also evident for GPCC, CRU, PRINCETON and PREC/L in all seasons.

As presented in Table 3, the GCM data are generally less accurate, showing larger biases throughout the year. The results indicate that the performance of GCMs is seasonally dependent, with the maximum bias found in summer and spring, whereas the least errors observed during winter and autumn. The performance of GCMs oscillates between seasonal over- and underestimation of the observed rainfall. For example, while CCSM4 tends to overestimate the observed rainfall during summer and autumn, it underestimates rain gauge data in winter and spring. In the same context, MRI-I-CGCM3 shows a negative bias in all seasons. The results also imply that GCMs are unable to detect light rainfall amounts in the dry season. One representative example is EC-Earth, which overestimates rainfall during the dry season (summer) by 121.5 mm, while it shows the best performance among all products during the wet season (winter), with a slight bias of 0.3 mm. Given that summers are typically characterized by convective rainfall in Saudi Arabia, although the magnitude is less relative to other seasons (Almazroui, 2006; El Kenawy *et al.*, 2014), this finding suggests that while the parameterization for rainfall performs quite well in winter, important land–atmosphere parameters that contribute to summer rainfall are not well represented in the GCMs. Giorgi and Marinucci (1996) attributed much of the bias observed in climate models during summertime to a lack of adequate treatment of cumulus convection, a predictable parameter that plays a key role in rainfall variations across Saudi Arabia during this

Table 3. Results of the different validation statistics.

	DJF			MAM			JJA			SON			Annual		
	Bias	MAE	Rho	Bias	MAE	Rho	Bias	MAE	Rho	Bias	MAE	Rho	Bias	MAE	Rho
APHRODITE	-4.60	11.90	0.33	-10.10	21.30	0.33	-6.80	11.70	0.55	-6.10	12.20	0.42	-6.90	14.30	0.35
CRU	0.90	13.70	0.41	-0.40	20.70	0.41	-0.40	14.00	0.47	-2.90	13.60	0.44	-0.70	15.50	0.43
GPCC	1.00	12.00	0.51	3.60	20.30	0.51	-2.60	11.40	0.48	-3.90	11.30	0.52	-0.50	13.80	0.55
PREC/L	1.80	14.30	0.42	12.10	28.10	0.38	0.40	13.70	0.57	-0.70	14.20	0.50	3.40	17.50	0.51
PRINCETON	-0.60	13.40	0.34	-0.50	21.90	0.32	-1.10	12.40	0.44	-4.30	13.10	0.37	-1.60	15.20	0.41
UDEL	-4.90	12.50	0.39	-11.40	21.90	0.34	-5.70	11.40	0.49	-7.80	12.10	0.43	-7.50	14.50	0.42
CCSM4	-8.70	11.50	0.13	-10.10	22.30	0.11	54.90	59.60	0.35	2.20	18.70	0.25	9.60	28.10	0.20
EC-Earth	0.30	15.90	0.11	11.30	31.00	0.17	121.50	123.60	0.36	32.10	43.80	0.17	41.30	53.60	0.24
MRI-I-CGCM3	-7.20	12.10	0.13	-14.20	21.30	0.07	-9.80	10.50	0.48	-6.00	14.10	0.25	-9.30	14.50	0.24

season. These results highlight the limitations of GCMs to undertake climate variability studies over the region during the summer season. In a study over the Middle East, Evans (2009) concluded that the performance of GCMs in simulating precipitation was generally inadequate, with many models not performing well. Here, it is important to note that the gauge-based products are expected to be quite close to the observed data given that they exploit ground measurements in the interpolation schemes. On the contrary, the GCMs do not directly incorporate observed rainfall data, but rather rely on physics-based descriptions and cloud resolving parameterizations that inevitably influence the performance of these models, particularly with their coarse grid spacing.

While the annual bias values are quite low for some of the products, as shown in Table 3, these biases tend to be balanced out by the mix of positive and negative values, as illustrated in Figure 4, which illustrates the spatial distribution of the mean annual bias averaged for the period from 1965 to 2005. It is noted that both the gauge-based and GCM results generally agree on a gradient of the rainfall bias from southwest to north and northeast. However, while the gauge-based products and MRI-I-CGCM3 indicate a positive gradient from south to north, CCSM4 and EC-Earth present a negative gradient. This is an interesting response, particularly as the GCMs are inconsistent not only in the sign of the gradient but also in its magnitude (CCSM4 shows a reduced bias signal relative to EC-Earth).

The gauge-based products reflect characteristics present in the observational data set, showing lower overall biases. Nonetheless, they do underestimate rainfall in the more rainy southwest regions. This finding can probably be seen in the context that the gauge-based data are obtained from interpolating scarce observations at high altitudes in the southwest parts of the region. In contrast, CCSM4 and EC-Earth model results show enhanced rainfall over the Asir Highlands in the southwest, whereas they slightly overestimate rainfall in areas coinciding with the northern and northeastern portions of the region. For example, EC-Earth suggests an annual overestimation of about 80 mm in the southwestern region, compared to roughly 0–30 mm in the central and northern areas, a response that is reflected by CCSM4, but with a much reduced amplitude. In accordance with the gauge-based data,

MRI-I-CGCM3 suggests a negative bias in rainfall totals along the southwest corner of the region and a positive bias in the central and northern portions. Overall, the spatial patterns of bias indicate in the GCMs are consistent with gauge-based products, with higher absolute errors in the south (where there is more rainfall) and lower errors in areas with low mean annual rainfall (Figure 4).

The MAE of the gauge-based products indicates annually averaged values between 13.8 mm (GPCC) and 17.5 mm (PREC/L) (Table 3). In accordance with the bias, the MAE values are much higher for GCMs, with annual MAE of 28.1 mm for CCSM4 and 53.6 mm for EC-Earth. On the other hand, MRI-I-CGCM3 shows the lowest annual MAE, with an averaged difference of only 14.5 mm from the observations, providing a value comparable to the gauge-based products. The dependencies between observations and the different products are also compared using the Spearman rank correlation, with the correlations at the seasonal and annual time scales summarized in Table 3. As presented, the gauge-based products provide higher correlations with observations with values between 0.35 (APHRODITE) and 0.55 (GPCC). GCMs generally present weaker correlations of less than 0.25 in all cases.

For the gauge-based products, results on the seasonal time scale generally match well with those found for the annual time series, except for the spring period (March–May, MAM), where correlation is generally lower and MAE higher than for other seasons. For the GCM-based products, the results are somewhat more variable. There is a clear increase in MAE and bias during the summer period (JJA), relative to the other seasons. Correlation coefficients are also generally lower compared to the gauge-based products, although MAE and bias values are comparable for many of the seasons. The results on the standard deviation differences (σ_{diff}), as depicted in Figure 5, are consistent with those obtained using other validation statistics (see Table 3). Almost all of the gauge-based products reflect a similar σ_{diff} at the annual scale, apart from PREC/L that has a slightly increased value and wider range relative to other products. More pronounced differences can be observed in the standard deviation of the GCMs, particularly CCSM4 and EC-Earth. Once again, MRI-I-CGCM3 shows comparable

Gauge-based products

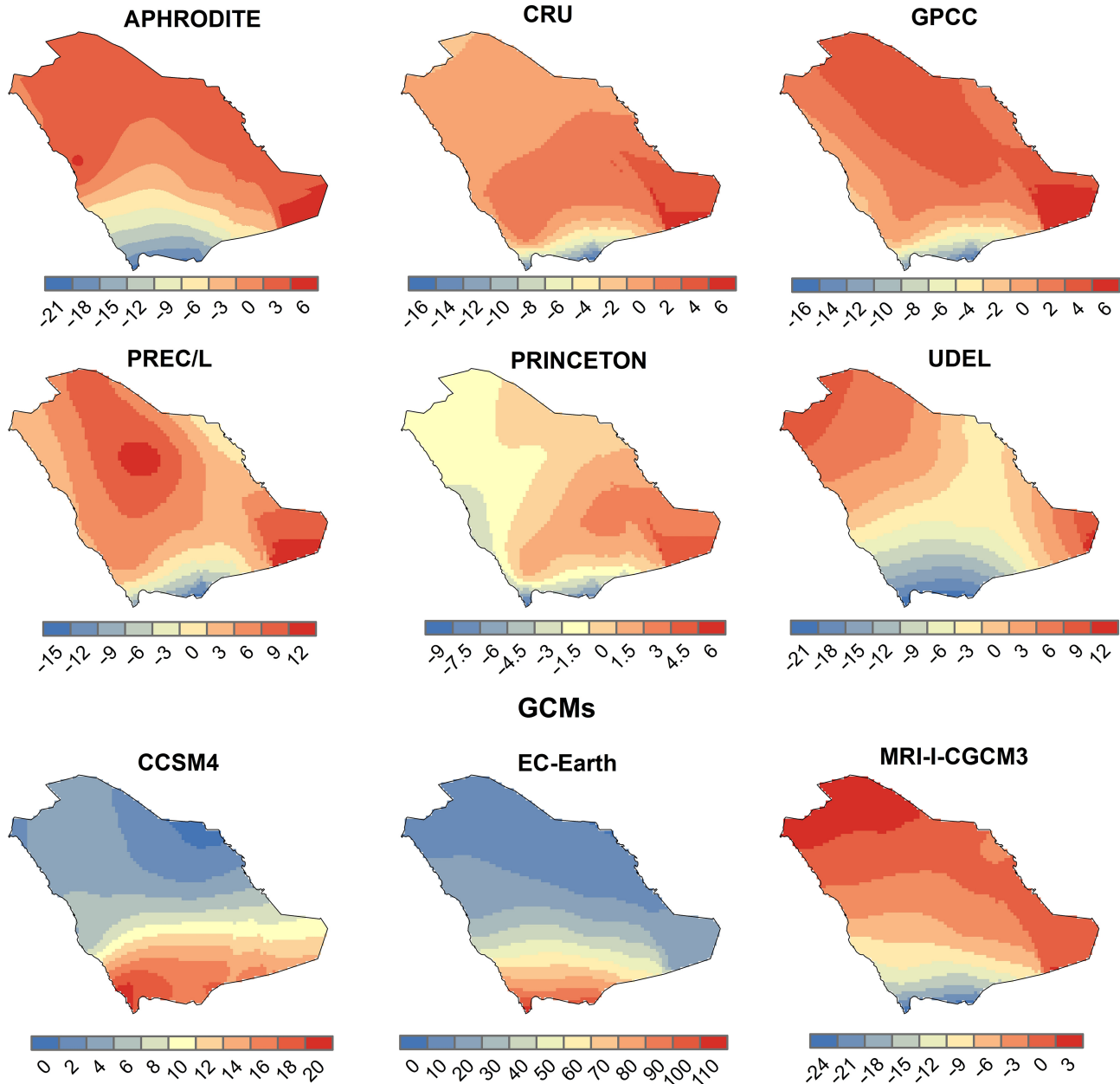


Figure 4. Spatial distributions of the mean annual bias (mm) with respect to observed data. The legend representing the bias was not homogenized for all panels to improve the final presentation of the figure given that some data sets showed high range of biases.

results with those obtained from the gauge-based outputs. In summary, although there is clear seasonal variability, the evaluation metrics generally indicate good agreement between the observed data and the gauge-based products. For the GCM-based products, there is more variability and a different seasonal response, with EC-Earth presenting the worst performance among this small sample of GCMs.

4.2. Comparative analysis of rainfall characteristics

4.2.1. Rainfall climatology

To evaluate the ability of the different products to represent the climatology of rainfall in Saudi Arabia, we

first examined the agreement of the annual rainfall cycle between observations and the different products. Then, the climatology of rainfall using the different rainfall products was compared against the available site observations using the long-term seasonal averages (1965–2005). Figure 6 represents the annual cycle of rainfall averaged for Saudi Arabia from 1965 to 2005. As illustrated, there are significant differences between the products in terms of their ability to represent the observations based annual cycle. The Pearson correlation between the GCM products and the observed seasonal cycle is generally weak, with $r = -0.01$ for CCSM4, $r = 0.01$ for EC-Earth and $r = 0.44$ for MRI-I-CGCM3. Interestingly,

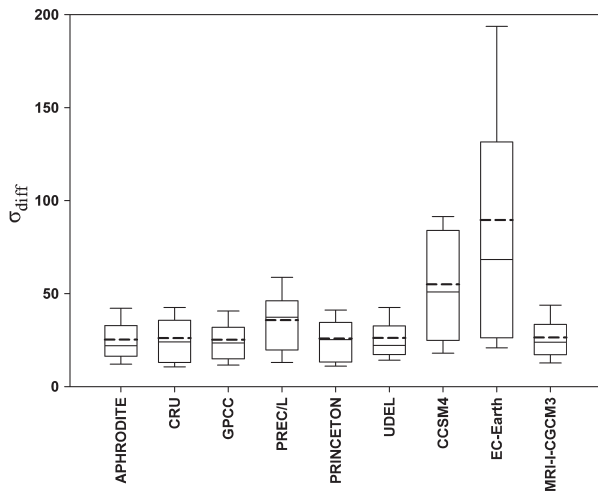


Figure 5. Standard deviation differences (σ_{diff}) between observations and each comparable product. The calculations were made for the annual accumulated rainfall (mm) from 1965 to 2005. The central solid line indicates the median, whereas the central dashed line refers to the mean. The whiskers represent the 10th and the 90th, while the 25th and the 75th are plotted as the vertical lines of the bounding boxes.

although the GCMs are driven by different model parameters (see Table 2), two of the three models show similar intra-annual variations of rainfall estimates in the region. In particular, the modelled rainfall estimates in the case of CCSM4 and EC-Earth are wetter than the observations, especially in the expected dry period of May–October. In contrast, the gauge-based products seem to reproduce the observed annual cycle quite well, particularly APHRODITE, PRINCETON, CRU and GPCC. These products present similar patterns and structures to those of the observations, with correlation coefficients of 0.90 (APHRODITE), 0.89 (PRINCETON) and 0.88 (GPCC and CRU). The overall structure of the annual cycle is very similar between these four products, particularly during the first half of the year, with peaks in March–May. The worst correlation is observed for PREC/L and UDEL (generally below 0.79). However, although the annual cycle of some gauge-based data generally fits well with observations year round, the data present conditions drier than the observations in early autumn, suggesting abrupt decreases in the mean rainfall for the months of September and October (see for example, CRU, PRINCETON and PREC/L).

The general agreement between observations and the gauge-based products in terms of the annual cycle similarity is further supported when we compare the relative contribution of rainfall amount in each season to the annual totals. Figure 7 depicts the relative contribution of seasonal rainfall to the annual rainfall totals for each data set, averaged between 1965 and 2005. As illustrated, the relative contribution of the observed December–February (DJF) mean rainfall is well represented by the majority of products. In particular, winter rainfall in CRU, GPCC, UDEL and PRINCETON represents approximately 26, 26.7, 26 and 23.9% of the annual rainfall totals, respectively, which is comparable with the observations (26%). On the other

hand, both CCSM4 (4.7%) and EC-Earth (10.8%) strongly underestimate winter rainfall, while APHRODITE slightly overestimates winter rainfall in the region, with a contribution of nearly 30.4% of the rainfall annual totals. In contrast to other GCMs, MRI-I-CGCM3 shows comparable results to observations, with winter rainfall contributing to 24.5% of the annual rainfall totals.

In the same context, while observations suggest spring as the main rainy season in Saudi Arabia, with approximately 40.7% of the annual total rainfall, the GCMs tend to underestimate the rainfall amounts during this season. While MRI-I-CGCM3 is within the observed response, CCSM4 and EC-Earth show reduced rainfall in the spring, representing no more than 18.9 and 24.1% of the total annual rainfall, respectively. In contrast, the gauge-based products give a better approximation of spring rainfall. According to APHRODITE, CRU, GPCC, PREC/L and PRINCETON, spring rainfall accounts for 39.6, 43.1, 43.7, 46.5 and 46.2% of the annual rainfall. For the summer months, CCSM4 and EC-Earth show both a warm bias and large variability compared to both observations and gauge-based products. The spread in the individual measurements during summer is high for CCSM4 and EC-Earth compared to other products, indicating that these models give rainfall estimations that are not within the observed range. The magnitudes of summer rainfall for CCSM4 and EC-Earth are almost 58.8 and 44% of the annual accumulated rainfall, compared to only 12.7% for observations. MRI-I-CGCM3 tends to predict a reduced contribution during the same season. The relative contribution of summer rainfall in the gauge-based products matches well with the observations, with the best results found for GPCC (12.9%), CRU (13.9%), PREC/L (14.1%) and PRINCETON (14%). In autumn, apart from MRI-I-CGCM3, both GCMs and gauge-based products show a reasonable representation of the observed rainfall (20.5%), with the best results found for APHRODITE (22.2%), which reproduces both the mean response and range of inter-annual variability.

Figure 8 illustrates the climatologies of annual rainfall totals computed for the multiple products as compared to observations for the period 1965–2005. It can be seen that the GCMs and gauge-based products show some consistency with the observed data, with the overall pattern of the climatologies preserves across the majority of products. Importantly, the comparison shows that most of the products reproduce the observed gradient in rainfall totals from the southwest to the northeast and to the east, reflecting the role of orography in rainfall patterns. The average annual rainfall for Saudi Arabia using the *in situ* data for the period 1965–2005 is 157.5 mm, which is comparable with CRU (148.9 mm), GPCC (151.1 mm), PRINCETON (137.6 mm) and PREC/L (196.4 mm), with a difference of only 0.05% (CRU), 0.04% (GPCC), 12.6% (PRINCETON) and 24.7% (PREC/L). Differences between the climatologies of the gauge-based products may be due to the use of different numbers of *in situ* data over space and time (see Section 3.1.2). Unfortunately, metadata on the input

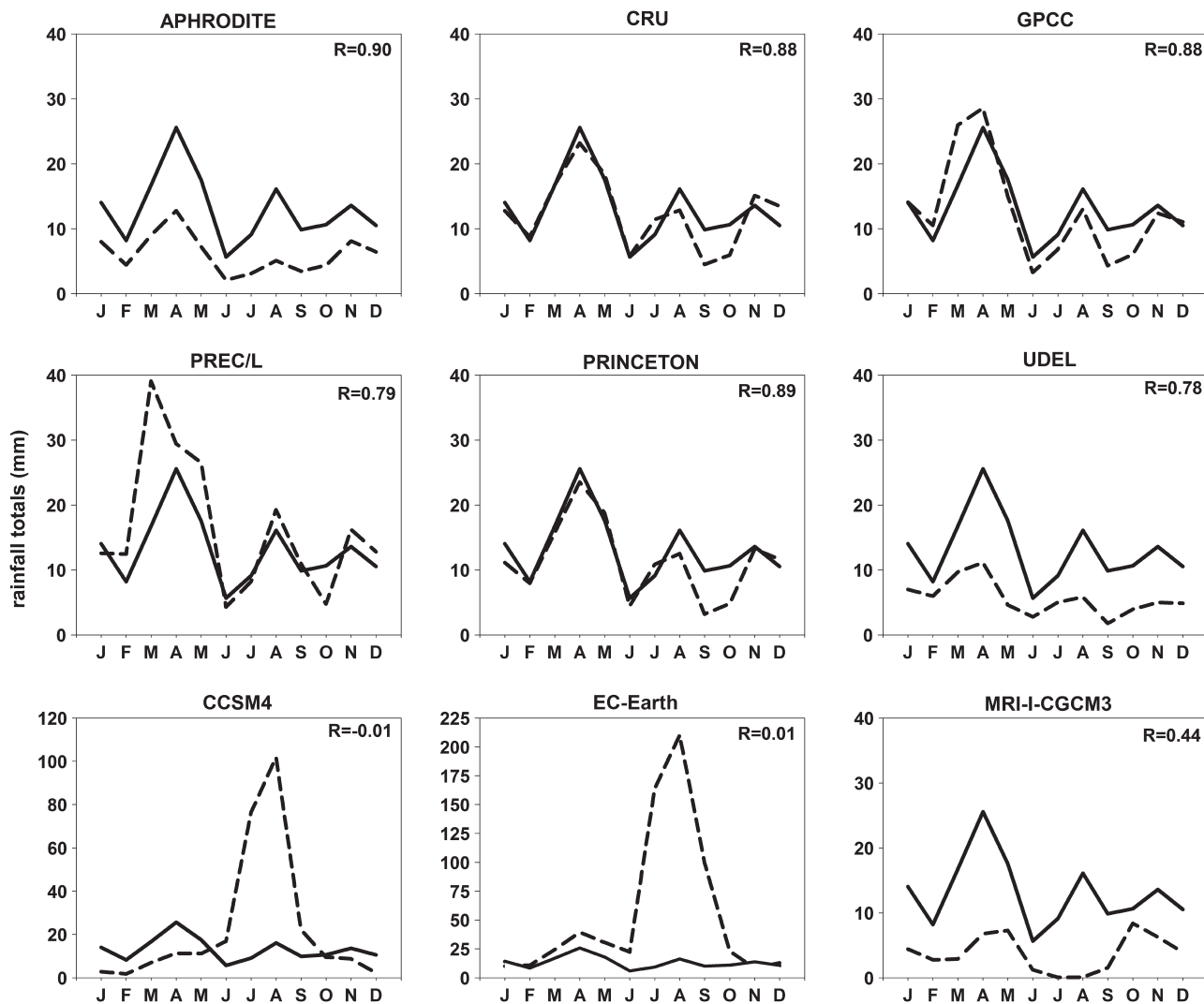


Figure 6. Annual cycle of rainfall totals (mm month^{-1}) averaged for the entire study area. For each panel, the solid line indicates the observed data, while the dashed line represents either the gauge-based product or GCM model response. Pearson correlation coefficient (R) between seasonal cycle of observations and that of each product is also provided.

stations used for calibrating each product over the region is lacking.

For the GCMs, some considerable discrepancies can be noted on the regional scale. Although the spatial pattern of rainfall totals reported by the GCMs is similar to observations and gauge-based products, CCSM4 and EC-Earth show substantial wet maxima and a sharp gradient feature in the southwestern region, particularly at higher altitudes. On the other hand, MRI-I-CGCM3 produces the lowest rainfall annual totals across the study domain, implying that the spatial gradient of rainfall totals is different to the observations. While EC-Earth reports higher rainfall totals in the southwestern areas (above 600 mm year^{-1}) that are approximately twice the observed data (less than 300 mm year^{-1}), MRI-I-CGCM3 does not show adequate skill in describing the wettest region in Saudi Arabia as it underestimates the annual total with 55 mm (approximately 18%). Given that 20 of the 53 rain gauges used in this work are located above 1000 m (refer to Figure 1), it can be seen that most of the GCM products

have major limitations in reproducing orographic rainfall over mountainous areas, while they compare relatively well over the lower altitudes across the region. Indeed, GCMs show higher rainfall than observations in the high altitudes, but almost similar rainfall for the lower altitudes. The dry and moderately dry areas, with rainfall less than 100 mm year^{-1} , are well represented by the GCMs.

Overall, it seems that the GCMs are able reproduce the spatial distribution of the observed rainfall over the region, but with larger deviations when compared to the gauge-based products. These differences might be explained by inherent scale effects, which constrain direct comparison between the point-based station estimates and the gridded outputs of GCMs. Over the study domain, real-world boundary conditions (e.g. elevation and land cover) may vary over short distances, particularly in the mountainous areas in the south. Incorporating such small scale changes and features into GCM structures and simulations is impossible. Recent efforts using regional climate models have shown the impacts of using higher resolution

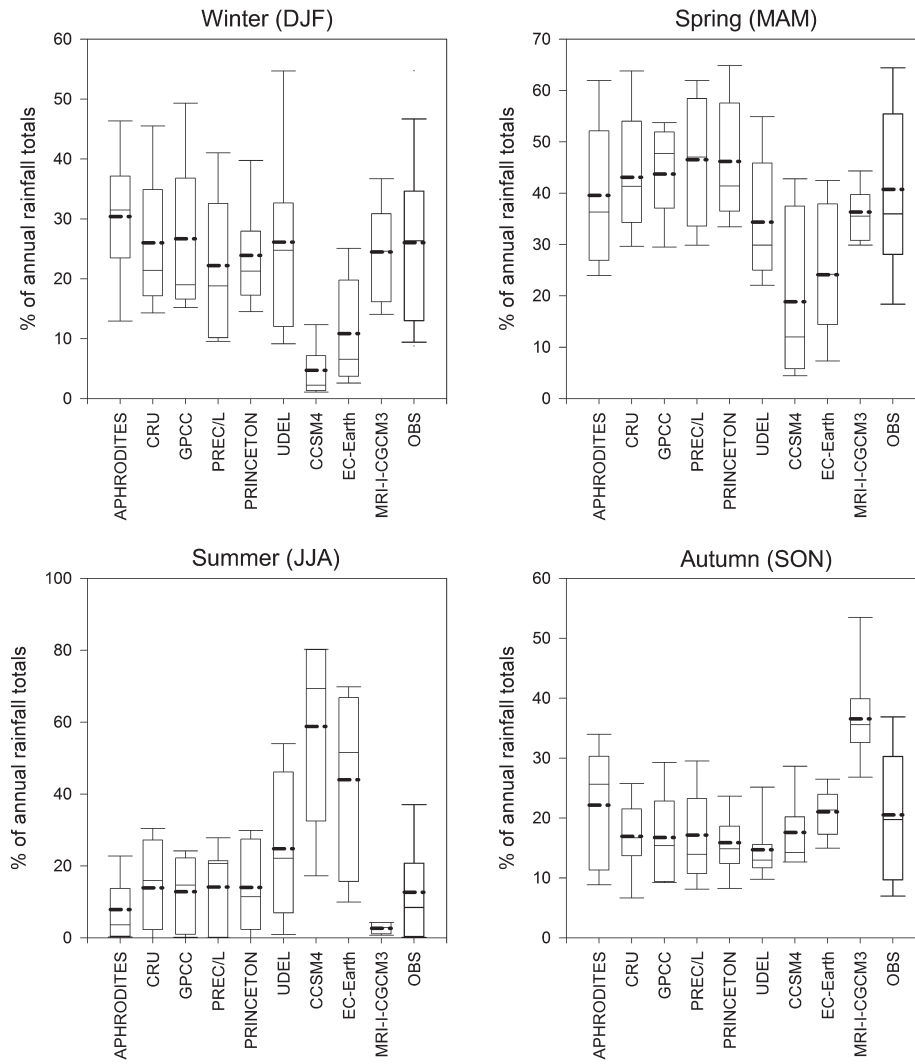


Figure 7. Relative contribution of seasonal rainfall totals to the annual totals during the period 1965–2005. The central solid line indicates the median, whereas the central dashed line refers to the mean. The whiskers represent the 10th and the 90th, while the 25th and the 75th are plotted as the vertical lines of the bounding boxes.

simulations and more realistic surface boundary conditions to better capture the sub-regional scale climate features (e.g. topography, leeward effects and land use/cover changes) that affect rainfall and other hydrometeorological responses (Evans and McCabe, 2013; Meng *et al.*, 2014a, 2014b). While increasing model resolution can clearly improve spatial representation (Deng *et al.*, 2015), the underlying issue of point rainfall *versus* grid-averaged model response remains a considerable challenge for coarser scale GCM assessments. However, this should not preclude the assessment of such approaches, but the focus should remain on reproducing seasonal patterns, climatologies and anomalies, rather than absolute values at high temporal resolutions.

It can be concluded that many of the gauge-based products, but particularly GPCC, CRU and PRINCETON, are able to reproduce realistic rainfall climatologies compared to other products, a finding that agrees well with Almazroui *et al.* (2012), who noticed that CRU was skillful in reproducing the annual rainfall climatology over

Saudi Arabia for the period from 1979 to 2009. The good performance of other gauge-based products has also been reported in a number of recent studies over nearby regions, including Iran (Raziei *et al.*, 2011), India (Prakash *et al.*, 2014) and Ethiopia (Reda *et al.*, 2014).

4.2.2. Rainfall anomaly

Figure 9 summarizes Pearson correlation coefficients calculated between the seasonal and annual anomalies of the observations and each of the tested products. For each grid point, anomalies were calculated for each season/year independently, relative to their respective 1965–2005 climatologies. An areal rainfall anomaly was then calculated using weighted averages determined by the Thiessen polygon method (Jones and Hulme, 1996). Comparing anomalies provides an overview of the general rainfall evolution and the degree of variability captured by the different products. The anomalies also assess the ability of the different products to respond over time (i.e. defining dry, wet, unchanged years) rather than any year-to-year

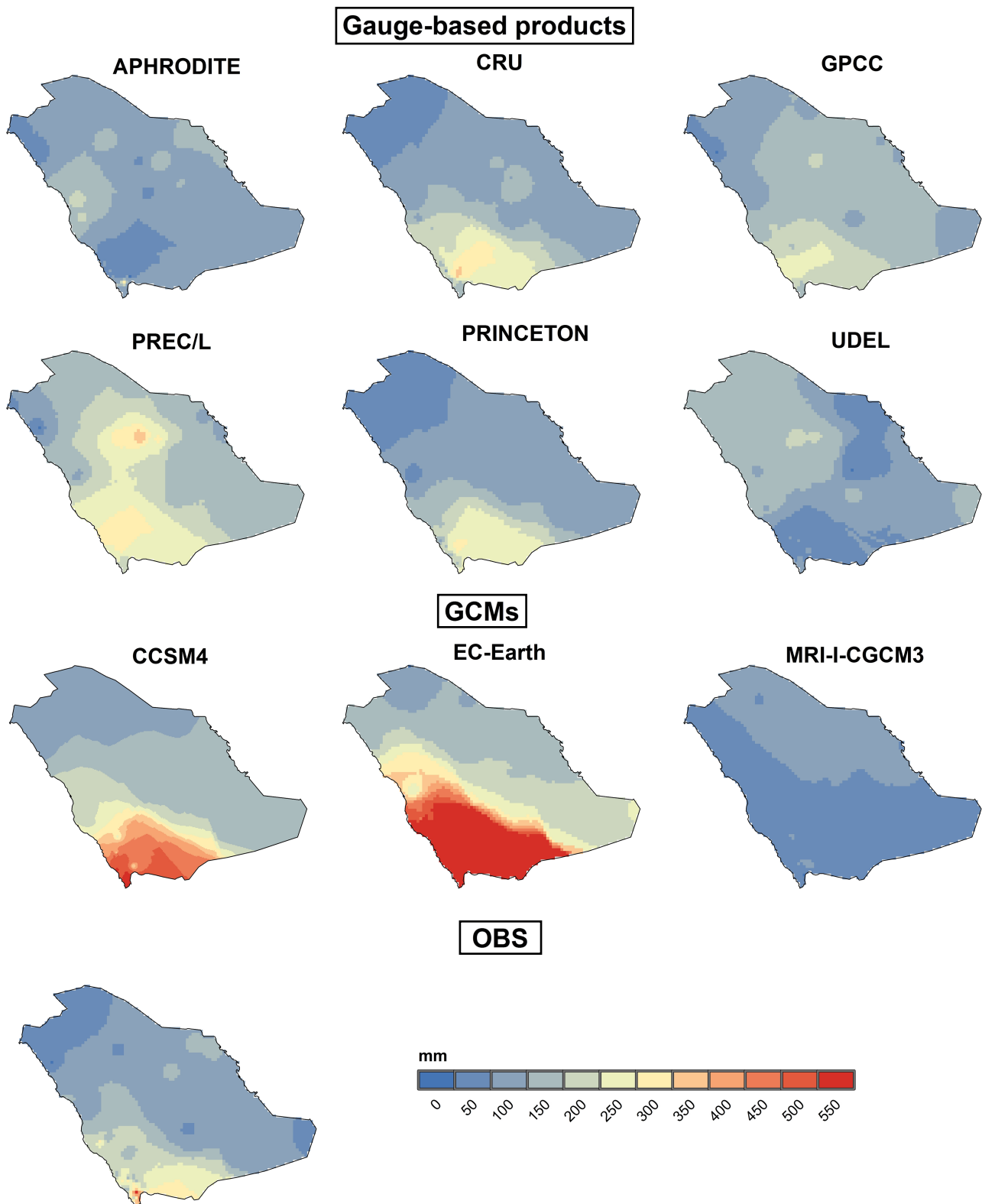


Figure 8. Spatial distributions of annual rainfall totals (mm) averaged for the period 1965–2005.

response. The results indicate that APHRODITE and GPCC agree very well with observations in terms of the magnitude and phase (i.e. positive and negative), showing high correlation coefficients over 0.8 in the majority of the seasons. Good responses are also observed

for CRU and PRINCETON, particularly in winter and spring. As illustrated in Figure 9, the correlation-based dependencies between observation anomalies and those of the gauge-based products are stronger during autumn (September–November, SON) and spring (MAM),

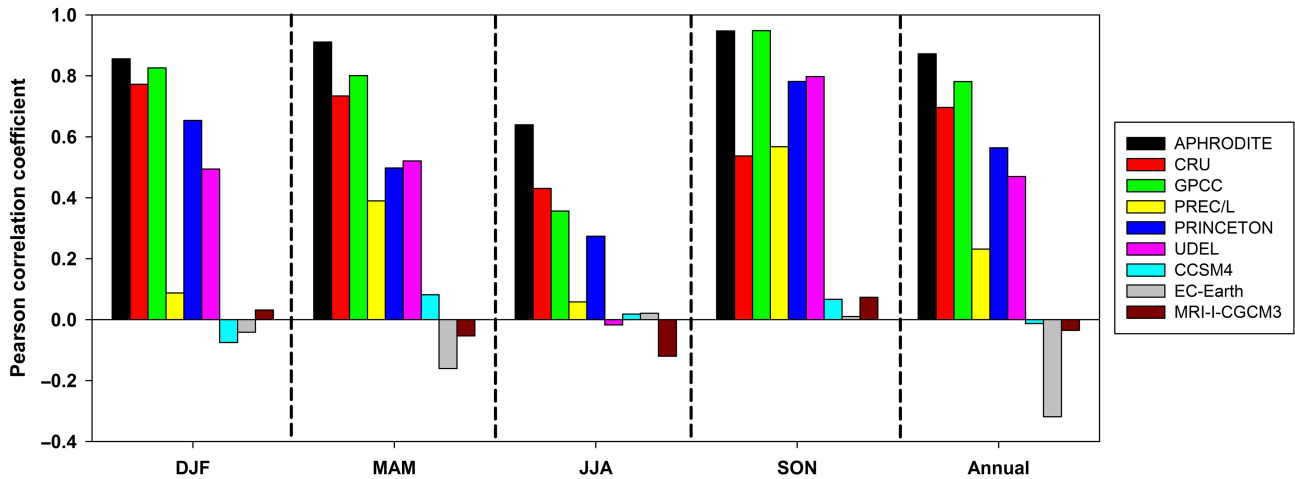


Figure 9. Anomaly correlation coefficient calculated between each product and the observed data averaged for all Saudi Arabia. Correlation coefficients above 0.3 are statistically significant at the 95% level ($p < 0.05$).

whereas they are statistically insignificant in summer (JJA). All gauge-based products show their worst performance during the summertime, particularly PREC/L and UDEL, which have correlation coefficients of 0.05 and -0.01 , respectively. The evolution of observation-based anomalies is generally well captured by GPCC, but with correlation values ranging between 0.36 in summer and 0.95 in autumn. Interestingly, while APHRODITE consistently underestimates the amount of rainfall in all seasons in the range of -10.1 (spring) to -4.6 mm (winter) (see Table 3), it provides the best approximation of observation anomaly at the annual scale. The correlation coefficients between APHRODITE and observation anomalies vary between 0.64 (summer) and 0.95 (autumn), with an overall annual correlation coefficient of 0.87 that exceeds the significance level of 0.01. The phase of APHRODITE generally matches well with that of the observed data, indicating that the APHRODITE 0.25° daily precipitation data set might be adjusted to remove the negative bias, providing a higher quality product for assessment and use in Saudi Arabia. In their study on the arid region west of China, Yang and Luo (2014) also indicated that APHRODITE data showed high agreement with rain gauges, with correlation coefficients ranging from 0.83 to 0.94. In contrast to the gauge-based products, comparisons of seasonal and annual rainfall observation-based anomaly with those derived from the GCMs show very poor agreement and even negative correlations in some cases (see for example, CCSM4 and EC-Earth in winter and EC-Earth and MRI-I-CGCM3 in spring). While CCSM4 and MRI-I-CGCM3 display the lowest correlation with observations, EC-Earth is anti-correlated at both the annual scale and across two seasons (DJF and MAM).

Figure 10 presents an illustrative example of the temporal evolution of annual observation anomalies across all rainfall products. In general, GCM data correlate insignificantly with observations and are clearly affected by spurious jumps, with individual years deviating considerably from the long-term mean rainfall. This feature

challenges their suitability for an analysis of rainfall changes in the region. It should be noted that while a robust and well performing GCM might show improved temporal agreement with seasonal and annual anomalies of observed rainfall rather than with anomalies calculated on the daily or monthly basis (where variability is higher), the correlation values for all of the investigated models shown here are statistically insignificant for the annual anomalies. However, it is perhaps more realistic to expect that GCMs only reproduce the climatological response (see Figure 7) rather than the year-to-year variation. So, while Figure 10 is instructive for the gauge-based products, it is less suited to establishing a hierarchy amongst the GCMs. While MRI-I-CGCM3 is at least able to reproduce expected ranges in annual anomalies, both CCSM4 and EC-Earth display significant departures from expectation.

The comparison of rainfall anomaly from gauge-based outputs, particularly APHRODITE, CRU and GPCC, show a more homogenous evolution and a greater agreement with *in situ* measurements. However, for some products (e.g. PREC/L and UDEL), the correlation is weaker, with coefficients of 0.23 and 0.47, respectively. The agreement between the anomalies of these products and observations is much better in the 1960s and 1970s. For example, the correlation between PRINCETON and observations in the period 1965–1980 is 0.65 compared to 0.52 from 1981 to 2005. While the reason for the difference between these periods is unclear, incorporating remotely sensed based retrievals into gauge-based products in the post-satellite era may improve their performance.

4.2.3. Long-term trends (1965–2005)

In order to compare all nine rainfall products used in this study, we assessed inter-annual and intra-annual variability of rainfall totals for the 41-year period using the ordinary least squares regression method, with their statistical significance tested using the MK tau test at the 95% confidence interval ($p < 0.05$). The agreement between the

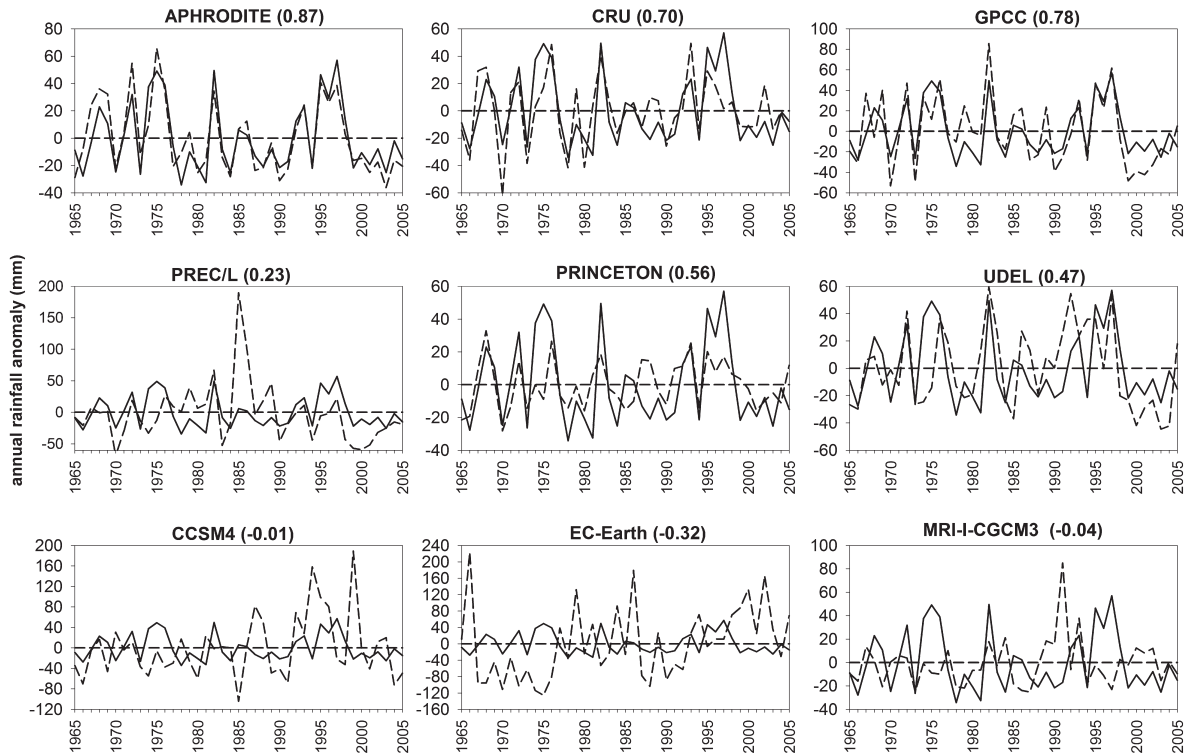


Figure 10. Anomalies of annual rainfall totals (mm) averaged for Saudi Arabia during the period 1965–2005. For each panel, the solid line indicates the observed data, while the dashed line represents either the gauge-based product or GCM model response. The dashed line refers to no differences between the annual rainfall totals and the long-term average. Pearson correlation coefficients are given in parentheses.

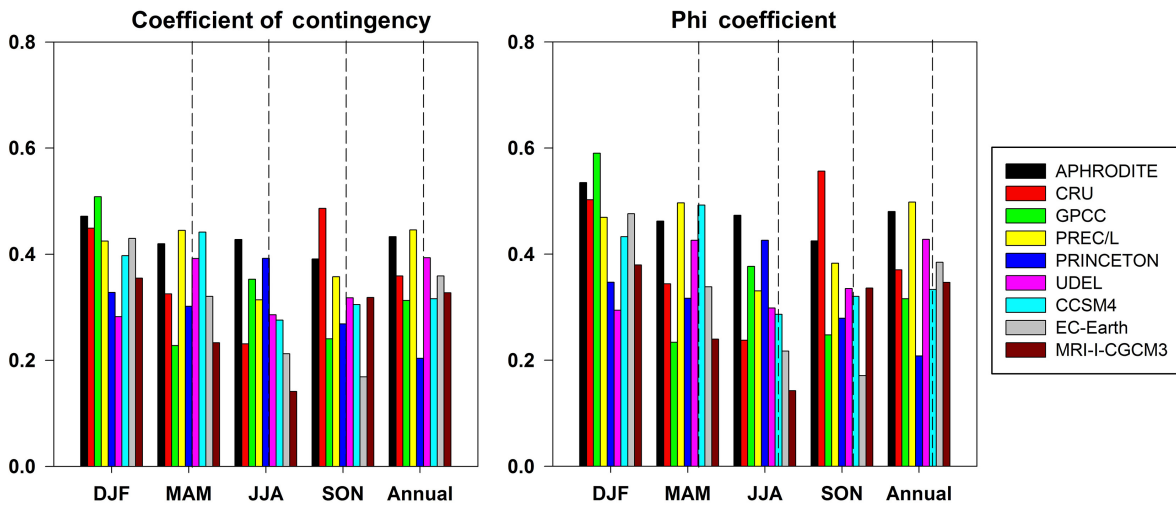


Figure 11. Agreement (asymmetry) coefficients between the categorized trends of annual and seasonal rainfall totals (i.e. statistically significant and positive, statistically significant and negative, insignificant positive and insignificant negative) of observations and those of the different products. The statistical significance was assessed using the Mann–Kendal tau statistic at the 95% level ($p < 0.05$).

categorized trends for observations against the available products was evaluated at the station level using a set of asymmetry measures (Gonzalez-Hidalgo *et al.*, 2003; López-Moreno *et al.*, 2010), with results of the goodness of fit between observed and all products shown in Figure 11. As can be seen, the majority of the products exhibit poor agreement with observations, with weaker coefficients below 0.4 in most cases. Such a response indicates that the majority of the products fail to reproduce the variations evident in the observed rainfall on the

seasonal and annual time scales and present irregular and inconsistent trends. In a few cases, some level of significant agreement is identified. In particular, APHRODITE is able to describe changes in rainfall during winter, spring and summer, but fails to do so in autumn. Similarly, GPCC and PREC/L exhibit consistent trends with observations only in winter. CRU can describe changes in rainfall in winter and autumn, but fails to do so in summer and spring.

All asymmetry measures (see Section 3.2.2) show that, for many products, the lowest agreement with observations

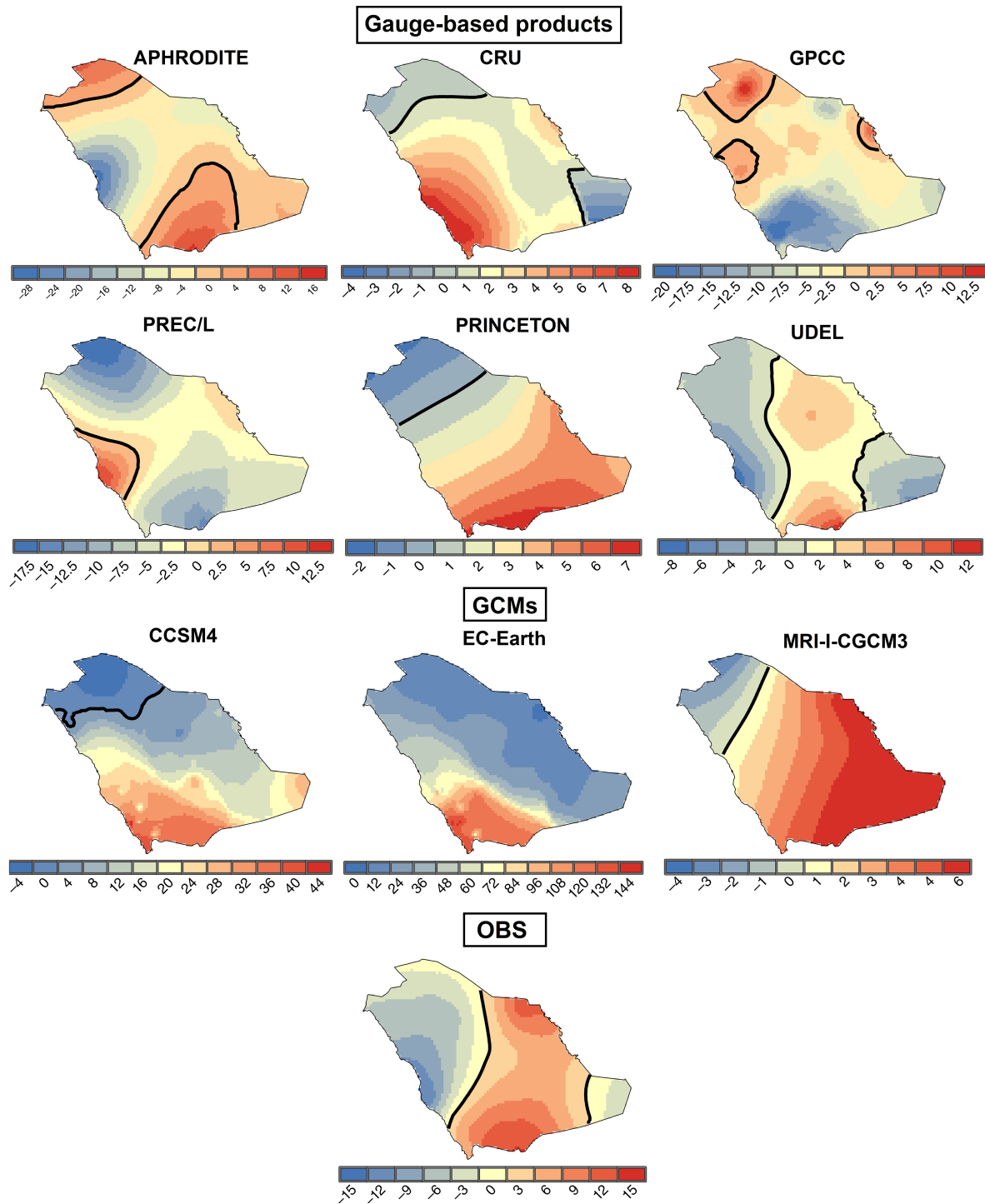


Figure 12. Spatial distribution of changes in the annual rainfall totals (mm decade⁻¹) from 1965 to 2005. The solid line indicates no changes. The legend representing amounts of change was not homogenized for all panels to improve the final presentation of the figure given that some data sets showed high range of changes.

is found during summer and autumn. In contrast, the highest agreement coefficients are generally found in winter, ranging from 0.28 (UDEL) to 0.51 (GPCC) following the coefficient of contingency, and from 0.29 (UDEL) to 0.59 (GPCC) according to the phi coefficient. Although the gauge-based products represent the climatological mean rainfall well, with low bias, higher correlation and good match with the observations anomalies, none of these

products are able to adequately reproduce the inter-annual variations of observed rainfall in all seasons (Figure 11). This result shows that the various products present different advantages, with none of them identified clearly as optimal in representing rainfall changes across all seasons.

Figure 12 illustrates the spatial distribution of the annual trends of rainfall totals over Saudi Arabia, as derived from the different products. Observations show a downward

trend in the northern portions of Saudi Arabia, whereas an upward trend is limited to the southwestern regions. The trend patterns from the nine products differ markedly, with the gauge-based products showing more comparable amounts of change with observations than GCMs, which reflect noticeably stronger changes (e.g. CCSM4 and EC-Earth). In accordance with GCMs, the gauge-based products are generally not skillful in capturing the spatial patterns of changes in annual rainfall totals. For example, GPCC exhibits contradictory trend patterns, with more rapid changes in the north than in the south. GPCC also shows a decrease in the annual rainfall totals in the range of about 10–20 mm decade⁻¹, which is clearly reversed in the observations (6–15 mm decade⁻¹). However, in a few cases (e.g. CRU, PRINCETON and UDEL), the observed wet trend in the southern zones and the observed dry trend in the northern areas are produced reasonably well.

Table 4 summarizes seasonal and annual trends in rainfall totals and their statistical significance for the different products. Table 4 indicates that the MK trend test shows an insignificant trend in the seasonal and annual rainfall, with only a few exceptions of the DJF trend for APHRODITE and the JJA and annual trends for EC-Earth. Increases in the observed rainfall are defined in winter and summer, while the spring and autumn exhibit a decrease in rainfall amounts. However, these trends are statistically insignificant at the 95% confidence level. Our results agree well with AlSarmi and Washington (2011) who found a general decreasing trend of rainfall over the Arabian Peninsula, although being statistically insignificant in most of the sites. The same finding was also confirmed for other neighbouring regions [e.g. Kwarteng *et al.* (2009) in Oman]. According to the majority of the products, it seems that the rainfall tends to decrease during the rainy seasons (winter and spring). For example, most of the products suggest a decrease in winter rainfall totals over Saudi Arabia, which is in opposition to observations that suggest a slight increase (1.53 mm decade⁻¹). As noted previously, the lack of trends in the majority of the gauge-based products may be attributed to inherent scale effects and also to the fact that these products tend to smooth rainfall values through the various interpolation schemes. However, the minimal number of significant trends in the observed data at both annual and seasonal scales confirms that the majority of the *in situ* data shows no change in their variance over time in Saudi Arabia.

5. Conclusion

Current knowledge of rainfall variations in Saudi Arabia is derived largely from *in situ* data sets, which are limited by sparsely distributed stations with often incomplete and short records. To improve the spatial and temporal resolution of current rainfall estimates in Saudi Arabia, this study assesses the performance of several gauge-based and GCM products against observations to evaluate their utility in the region. A set of skill indicators, including bias, MAE, standard deviation of differences and Spearman ρ correlation, were employed to evaluate the capacity of nine rainfall

Table 4. Seasonal and annual trends of rainfall totals calculated for the period 1965–2005 for all Saudi Arabia.

	DJF	MAM	JJA	SON	Annual
APHRODITE	-0.81	-3.13	-1.15	-1.14	-6.14
CRU	1.73	-1.36	0.08	0.49	1.25
GPCC	-0.08	-4.83	-0.30	-0.17	-4.75
PREC/L	-0.84	-2.60	0.41	-2.15	-5.82
PRINCETON	0.51	0.10	0.39	0.55	1.65
UDEL	-0.61	1.16	0.01	-0.46	-0.43
CCSM4	-1.20	-0.70	12.05	2.10	12.09
EC-Earth	-0.91	2.09	17.92	6.68	25.66
MRI-I-CGCM3	1.31	1.11	-0.12	0.74	2.86
OBS	1.53	-2.01	0.17	-0.37	-1.04

All the area-averaged time series were created following the Thiessen method. The magnitude of change is given in mm decade⁻¹. Bold numbers show statistically significant trends at the 95% confidence interval ($p < 0.05$).

products in reproducing statistical features of a network of 53 *in situ* rain gauges. As expected, the gauge-based products reproduce the statistics of observed climate features in Saudi Arabia more accurately than do the GCMs, although there is considerable within product variability. Evaluation of annual rainfall totals shows that the majority of the gauge-based products produce more or less correctly the spatial patterns of observed rainfall in the region. A comparison of the rainfall anomalies of the nine different products against *in situ* data also indicates that some gauge-based products (APHRODITE, CRU, GPCC and PRINCETON) show good agreement with observations, relative to other products, although there is also considerable seasonal variability.

The task of identifying one or several optimal products is made difficult by the variability in the different performance metrics. When comparing statistical metrics, anomalies, spatial patterns, seasonal contribution or inter-annual cycles, no single product consistently outperforms any other. As a result, the choice of rainfall product requires careful consideration against the purpose to which it will be employed. Overall, GPCC, CRU, APHRODITE and PRINCETON perform well against most metrics and could serve as baseline source for verifying and supplementing available data in the region. Certainly, an awareness of the uncertainties inherent in these products is required and also an understanding on the limits of their use. For instance, caution in the use of these data for extremes assessment is required, as their ability to reproduce extreme events has not been investigated. In addition, trend detection should be approached cautiously when using either the gauge- or GCM-based products, as they show poor skill in describing the observed inter-annual and intra-annual variability of rainfall.

The physical causes for the differences between the gauge-based and GCM products are beyond the scope of the current research. However, possible sources of differences between the gauge-based products themselves could be related to errors in source data and/or limitations in data density or interpolation schemes. For gauge-based data, there are two competing scales of interest that need to be

considered: the local scale response (which will obviously benefit from a high-density network) and the regional scale response (which benefits from a more distributed *in situ* sampling). Improving the ability of any interpolation model to predict climate in Saudi Arabia by expanding the training set to include stations from neighbouring countries like Egypt, Jordan, Yemen and the Persian Gulf makes sense, but also represents its own challenges. Most *in situ* data in the region are suited in cities and other population centres, introducing biases in the spatial distribution. There are also real costs associated with equipment installation, maintenance and archiving, which combined with historical instability in the region, goes some way to explain the paucity of long-term and consistent observations in space and time.

In comparing differences between the gauge- and GCM-based products, it is important to recognize that they are representing two quite different approaches to characterizing the climate state. Unlike gauge-based products, GCMs are largely independent from observations, employing instead physical descriptions of the governing processes that control the climate states. Therefore, it should not be expected that GCMs will be able to capture the high day-to-day variations of climate when compared to observations. This is why studies seeking to validate GCMs against *in situ* data, including the current investigation, focus on testing the ability of the models to reproduce the larger time and space scale characteristics of climate, such as anomalies, trends and climatology. Of the three GCMs examined here (albeit a small selection of the available CMIP5 products), two (CCSM4 and EC-Earth) tended to suffer from a large systematic overestimation of rainfall in Saudi Arabia, with significant uncertainties defined in the southwest. The GCM products generally failed to reproduce the observed anomalies of rainfall, both at the inter-annual and seasonal scale. Overall, using GCM products for hydroclimatological or related studies should be undertaken with caution in this region, particularly during the summertime and over complex terrain, where the model spatial resolution and forcing limit the reproduction of summer rainfall, which has a distinct regional and local character. These findings highlight the need to investigate the impact of topography on the performance of GCMs in the southwestern areas of the region. The effects of topography are complex, but including more detailed information on elevation, slope and orientation of the slope into simulations may potentially improve our understanding of the physical mechanisms that control the performance of GCMs in mountainous areas, particularly during summer. This uncertainty could also be related to the inadequate representation of some components of the land surface schemes, such as soil moisture and convective parameters. Overall, determining the reasons behind the defined differences and implementing schemes to improve these should be the focus of further investigation.

Studies of rainfall patterns and mechanisms in Saudi Arabia and similar regions are limited by the uneven distribution of observations in space and time. This study illustrated that data from selected gauge-based products can

mitigate these challenges by providing an alternative platform for synthesis of observations. Nonetheless, the issue of relying on an *in situ* network that has obvious limitations as the basis for comparing different rainfall products is an area requiring further investigation. Given the lack of a reliable *in situ* network in the region, the importance of further developing products that combine available observations, including both gauge-based and satellite data, is obviously important. However, in the pre-satellite era, evaluating product performance against available observations remains key, with the next step focusing on improving model parameterization schemes to better represent the regionally specific climate features. Also, advances in high-resolution regional climate modelling represent an obvious means of improving the representation of hydrometeorological response in the region. In this regard, satellite products have recently improved their spatial and temporal resolution, offering a promising strategy to address the limitations of gauge-based rainfall estimates. Likewise, in spite of their limited temporal coverage, a variety of reanalysis precipitation data with a reasonable spatial resolution (e.g. MERRA, NCEP, ERA-Interim and JRA25) may also provide another means to obtain precipitation data in the region. For this reason, additional validation studies are still required to test the feasibility of both reanalyses and rainfall retrievals from the more recent satellite era data. Ultimately, a combined product that optimally integrates elements from available observation and modelling platforms would be beneficial for a range of disciplines spanning agriculture, water resources planning and management and flood assessments in the region.

Acknowledgements

Research reported in this publication was supported by the King Abdullah University of Science and Technology (KAUST), Saudi Arabia. The authors are grateful to the Saudi Arabian Ministry of Water and Electricity (MOWE) for providing the daily rainfall data. We also acknowledge the many different universities and government-based research groups for providing their data sets for analysis.

References

- Abdullah MA, Almazroui M. 1998. Climatological study of the southwestern region of Saudi Arabia. I. Rainfall analysis. *Clim. Res.* **9**: 213–223.
- Agresti A. 1996. *Introduction to Categorical Data Analysis*. John Wiley and Sons: New York, NY.
- Almazroui M. 2006. *The Relationships Between Atmospheric Circulation Patterns and Surface Climatic Elements in Saudi Arabia*. Unpublished PhD thesis, Climate Research Unit, University of East Anglia, Norwich, UK, 433 pp.
- Almazroui M, Islam MN, Athar H, Jones PD, Rahman MA. 2012. Recent climate change in the Arabian Peninsula: annual rainfall and temperature analysis of Saudi Arabia for 1978–2009. *Int. J. Climatol.* **32**: 953–966.
- Almazroui M, Dambul R, Islam MN, Jones PD. 2014. Principal components-based regionalization of the Saudi Arabian climate. *Int. J. Climatol.*, doi: 10.1002/joc.4139.
- AlSarmi S, Washington R. 2011. Recent observed climate change over the Arabian Peninsula. *J. Geophys. Res.* **116**: D11109, doi: 10.1029/2010JD015459.

- Black E, Brayshaw DJ, Rameau CMC. 2010. Past, present and future precipitation in the Middle East: insights from models and observations. *Philos. Trans. R. Soc. A* **368**: 5173–5184.
- Chen M, Xie P, Janowiak JE, Arkin PA. 2002. Global land precipitation: a 50-yr monthly analysis based on gauge observations. *J. Hydrometeorol.* **3**: 249–266.
- Clark WAV, Hosking PL. 1986. *Statistical Methods for Geographer*. John Wiley & Sons: Chichester, UK, 518 pp.
- Cortesi N, Gonzalez-Hidalgo JC, Brunetti M, Martin-Vide J. 2012. Daily precipitation concentration across Europe 1971–2010. *Nat. Hazards Earth Syst. Sci.* **12**: 2799–2810.
- Deng L, McCabe MF, Stenchikov G, Evans JP, Kucera PA. 2015. Simulation of flash-flood producing storm events in Saudi Arabia using the Weather Research and Forecasting model. *J. Hydrometeorol.* **16**: 615–630, doi: 10.1175/JHM-D-14-0126.1.
- Donat MG, Peterson TC, Brunet M, King AD, Almazroui M, Kolli RK, Bouchef D, Al-Mulla AY, Nour AY, Aly AA, Nada TAA, Semawi MM, Al Dashti HA, Salhab TG, El Fadli KI, Muftah MK, Eida SD, Badi W, Driouech F, El Rhaz K, Abubaker MJY, Ghulam AS, Erayah AS, Mansour MB, Alabdouli WO, Al Dhanhani JS, Al Shekaili MN. 2014. Changes of extreme temperature and precipitation in the Arab region: long-term trends and variability related to ENSO and NAO. *Int. J. Climatol.* **34**: 581–592.
- El Kenawy AM, McCabe MF, Stenchikov GL, Raj J. 2014. Multi-decadal classification of synoptic weather types, observed trends and links to rainfall characteristics over Saudi Arabia. *Front. Environ. Sci.* **2**: 37, doi: 10.3389/fevs.2014.00037.
- El Kenawy A, López-Moreno JI, McCabe MF, Brunzell NA, Vicente-Serrano SM. 2015. Daily temperature changes and variability in ENSEMBLES regional models predictions: evaluation and intercomparison for the Ebro Valley (NE Iberia). *Atmos. Res.* **155**: 141–157, doi: 10.1016/j.atmosres.2014.12.007.
- Evans JP. 2009. 21st century climate change in the Middle East. *Clim. Change* **92**: 417–432.
- Evans JP, McCabe MF. 2013. Effect of model resolution on a regional climate model simulation over southeast Australia. *Clim. Res.* **56**: 131–145.
- Gent PR, Danabasoglu G, Donner LJ, Holland MM, Hunke EC, Jayne SR, Lawrence DM, Neale RB, Rasch PJ, Vertenstein M, Worley PH, Yang Z, Zhang M. 2011. The Community Climate System Model Version 4. *J. Clim.* **24**: 4973–4991, doi: 10.1175/2011JCLI4083.1.
- Giorgi F, Marinucci MR. 1996. An investigation of the sensitivity of simulated precipitation to model resolution and its implications for climate studies. *Mon. Weather Rev.* **124**: 148–166.
- Goldreich Y. 2003. *The Climate of Israel. Observation, Research and Applications*. Kluwer Academic/Plenum Publishers: New York, NY.
- Gonzalez-Hidalgo JC, De Luis M, Raventos J, Sanchez JR. 2003. Daily rainfall trend in the Valencia Region of Spain. *Theor. Appl. Climatol.* **75**: 117–130.
- Harris I, Jones PJ, Osborn TJ, Lister DH. 2014. Updated high-resolution grids of monthly climatic observations – the CRU TS3.10 dataset. *Int. J. Climatol.* **34**(3): 623–642.
- Hazeleger W, Severijns C, Semmler T, Ștefănescu S, Yang S, Wang X, Wyser K, Dutra E, Baldasano JM, Bintanja R, Bougeault P, Caballero R, Ekman AM, Christensen JH, van den Hurk B, Jimenez P, Jones C, Källberg P, Koenigk T, McGrath R, Miranda P, Van Noije T, Palmer T, Parodi JA, Schmith T, Seltner F, Storelvmo T, Sterl A, Tapamo H, Vancoppenolle M, Viterbo P, Willén U. 2010. EC-Earth: a seamless Earth system prediction approach in action. *Bull. Am. Meteorol. Soc.* **91**: 1357–1363.
- Hodges KI, Lee RW, Bengtsson L. 2011. A comparison of extratropical cyclones in recent reanalyses ERA-Interim, NASA MERRA, NCEP CFSR, and JRA-25. *J. Clim.* **24**(18): 4888–4906.
- Jackson IJ. 1977. *Climate, Water and Agriculture in the Tropics*. Longman: London, 248 pp.
- Jones PD, Hulme M. 1996. Calculating regional climatic time series for temperature and precipitation: methods and illustrations. *Int. J. Climatol.* **16**: 361–376.
- Kendall MG. 1975. *Rank Correlation Methods*, 4th edn. Charles Griffin: London.
- Kwarteng AY, Dorvlo AS, Kumar GTV. 2009. Analysis of a 27-year rainfall data (1977–2003) in the Sultanate of Oman. *Int. J. Climatol.* **29**(4): 605–617.
- Lin R, Zhou T, Qian Y. 2014. Evaluation of Global Monsoon Precipitation Changes based on Five Reanalysis Datasets. *J. Clim.* **27**: 1271–1289, doi: 10.1175/JCLI-D-13-00215.1.
- Liu YY, van Dijk AIJM, McCabe MF, Evans JP, de Jeu RAM. 2013. Global vegetation biomass change (1988–2008) and attribution to environmental and human drivers. *Glob. Ecol. Biogeogr.* **22**: 692–705.
- López-Moreno JI, Vicente-Serrano SM, Angulo-Martínez M, Beguería S, Kenawy A. 2010. Trends in daily precipitation on the northeastern Iberian Peninsula, 1955–2006. *Int. J. Climatol.* **30**: 1026–1041.
- Lorenz C, Kunstmann H. 2012. The hydrological cycle in three state-of-the-art reanalyses: intercomparison and performance analysis. *J. Hydrometeorol.* **13**(5): 1397–1420.
- Meng XH, Evans JP, McCabe MF. 2014a. The impact of observed vegetation changes on land–atmosphere feedbacks during drought. *J. Hydrometeorol.* **15**: 759–776, doi: 10.1175/JHM-D-13-0130.1.
- Meng XH, Evans JP, McCabe MF. 2014b. The influence of inter-annually varying albedo on regional climate and drought. *Clim. Dyn.* **42**(3–4): 787–803, doi: 10.1007/s00382-013-1790-0.
- Polson D, Bollasina M, Hegerl GC, Wilcox LJ. 2014. Decreased monsoon precipitation in the Northern Hemisphere due to anthropogenic aerosols. *Geophys. Res. Lett.* **41**: 6023–6029, doi: 10.1002/2014GL060811.
- Prakash S, Mitra AK, Momin IM, Rajagopal EN, Basu S, Collins M, Turner AG, Achuta Rao K, Ashok K. 2014. Seasonal intercomparison of observational rainfall datasets over India during the southwest monsoon season. *Int. J. Climatol.*, doi: 10.1002/joc.4129.
- Raziei T, Bordi I, Pereira LS. 2011. An application of GPCC and NCEP/NCAR datasets for drought variability analysis in Iran. *Water Resour. Manage.* **25**: 1075–1086.
- Reda DT, Engida AN, Asfaw DH, Hamdi R. 2014. Analysis of precipitation based on ensembles of regional climate model simulations and observational databases over Ethiopia for the period 1989–2008. *Int. J. Climatol.* **35**: 948–971, doi: 10.1002/joc.4029.
- Rehman S. 2010. Temperature and rainfall variation over Dhahran, Saudi Arabia, (1970–2006). *Int. J. Climatol.* **30**: 445–449.
- Schneider U, Becker A, Finger P, Meyer-Christoffer A, Rudolf B, Ziese M. 2011. GPCC Full Data Reanalysis Version 6.0 at 0.5°: monthly land-surface precipitation from rain-gauges built on GTS-based and historic data. doi: 10.5676/DWD_GPCC/FD_M_V6_050.
- Sheffield J, Goteti G, Wood EF. 2006. Development of a 50-year high-resolution global dataset of meteorological forcings for land surface modeling. *J. Clim.* **19**(13): 3088–3111.
- Taylor KE, Stouffer RJ, Meehl GA. 2012. An overview of CMIP5 and the experiment design. *Bull. Am. Meteorol. Soc.* **93**: 485–498.
- Trenberth KT. 2011. Changes in precipitation with climate change. *Clim. Res.* **47**: 123–138.
- Vicente-Serrano SM, Saz-Sanchez MA, Cuadrat JM. 2003. Comparative analysis of interpolation methods in the middle Ebro valley (Spain): application to annual precipitation and temperature. *Clim. Res.* **24**: 161–180.
- Willmott CJ, Matsuura K. 1995. Smart interpolation of annually averaged air temperature in the United States. *J. Appl. Meteorol.* **34**: 2577–2586.
- Yang Y, Luo Y. 2014. Evaluating the performance of remote sensing precipitation products CMORPH, PERSIANN, and TMPA, in the arid region of northwest China. *Theor. Appl. Climatol.* **118**: 429–445.
- Yatagai A, Kamiguchi K, Arakawa O, Hamada A, Yasutomi N, Kito H. 2012. APHRODITE: constructing a long-term daily gridded precipitation dataset for Asia based on a dense network of rain gauges. *Bull. Am. Meteorol. Soc.* **93**(9): 1401–1415.
- Yukimoto S, Adachi Y, Hosaka M, Sakami T, Yoshimura H, Hirabara M, Tanaka TY, Shindo E, Tsujino H, Deushi M, Mizuta R, Yabu S, Obata A, Nakano H, Koshiro T, Ose T, Kito H. 2012. A new global climate model of the Meteorological Research Institute: MRI-CGCM3-Model description and basic performance. *J. Meteorol. Soc. Jpn.* **90a**: 23–64.
- Zhang X, Harvey KD, Hogg WD, Yuzyk TR. 2001. Trends in Canadian streamflow. *Water Resour. Res.* **37**(4): 987–998.
- Zhang XB, Aguilar E, Sensoy S, Melkonyan H, Tagiyeva U, Ahmed N, Kutsaladze N, Rahimzadeh F, Taghipour A, Hantosh TH, Albert P, Semawi M, Ali MK, Al-Shabibi MHS, Al-Oulan Z, Zatari T, Khelet IA, Hamoud S, Sagir R, Demircan M, Eken M, Adiguzel M, Alexander L, Peterson TC, Wallis T. 2005. Trends in Middle East climate extreme indices from 1950 to 2003. *J. Geophys Res.* **110**: D22104, doi: 10.1029/2005jd006181.