Comparative study on long term climate data sources over South Korea
Dong-Ik Kim and Dawei Han

ABSTRACT

Long term climate data are vitally important in reliably assessing water resources and water related hazards, but in-situ observations are generally sparse in space and limited in time. Although there are several global datasets available as substitutes, there is a lack of comparative studies about their suitability in different parts of the world. In this study, to find out the reliable century-long climate dataset in South Korea, we first evaluate multi-decadal reanalyses (ERA-20 cm, ERA-20c, ERA-40 and NOAA 20th century reanalysis (20CR)) and gridded observations (CRUv3.23 and GPCCv7) for monthly mean precipitation and temperature. In the temporal and statistical comparisons, CRUv3.23 and GPCCv7 for precipitation and ERA-40 for temperature perform the best, and ERA-20c and 20CR also indicate meaningful agreements. For ERA-20 cm, it has only a statistical agreement, but the mean has the difficulty in representing its ensemble. This paper also shows that the applicability of each dataset may vary by region and all products should be locally adjusted before being applied in climate impact assessments. These findings not only help to fill in the knowledge gaps about these datasets in South Korea but also provide a useful guideline for the applicability of the global datasets in different parts of the world.

Key words | climate data sources, interannual variability, non-parametric trend test, reanalysis, skill score

INTRODUCTION

To adapt and mitigate climate change, it is essential to analyse the reliable long term climate dataset. Although the gauged local data are generally considered as the best values, they are usually sparse and limited in the time range (Simmons et al. 2004; Becker et al. 2015). For this reason, the availability of a highly accessible and reliable gridded dataset has been developed since the 1980s, and some research groups, such as the Climate Research Unit (CRU) and the Global Precipitation Climatology Centre (GPCC), have constructed the monthly precipitation or temperature dataset by applying their own interpolation methods based on observations worldwide (Chen et al. 2002; Becker et al. 2015; Harris et al. 2014). They have had an important role in trend analysis in areas lacking local observations and global climate change analysis (Nicholson et al. 2003; Fekete et al. 2004; Dinku et al. 2008; Zhang & Zhou 2011; Nikulin et al. 2012). The other surrogates for local observations are reanalysis products derived using modern data assimilation techniques, which have been increasingly applied in climate impact studies. Representative, the European Centre for Medium-Range Weather Forecasts (ECMWF) and the National Oceanic and Atmospheric Administration (NOAA) have produced these kinds of products. Initially, most reanalysis datasets were only able to cover from the mid-twentieth century to present (Compo et al. 2011), e.g. the first National Centers for Environmental Prediction/National Center for Atmospheric Research (NCEP/NCAR) reanalysis: 1948–present (Kalnay

Dong-Ik Kim (corresponding author)
Dawei Han
Water and Environment Research Group, Department of Civil Engineering, University of Bristol, Bristol, UK
E-mail: dikim0127@gmail.com

doi: 10.2166/wcc.2018.052
et al. 1996); ECMWF 45-year reanalysis (ERA-40): 1975–2002 (Uppala et al. 2005); Japan Meteorological Agencies Reanalysis (JRA-25): 1979–present (Onogi et al. 2007); ECMWF reanalysis interim (ERA-interim): 1989–present (Dee & Uppala 2009; Dee et al. 2011b). However, a few recent reanalyses, such as NOAA 20th century reanalysis v2c (20CR), ECMWF 20th century atmospheric model ensemble (ERA-20 cm), and ECMWF 20th century assimilating surface observations only (ERA-20c), extended the data period up to the whole of the 20th century (Compo et al. 2011; Hersbach et al. 2015; Poli et al. 2016). Moreover, those reanalyses are able to perform daily or sub-daily scales as well as monthly scales, while the interpolated datasets only provide monthly values. Nevertheless, because these datasets are not directly taken from observations, they have additional uncertainties (Dee et al. 2011a; Hersbach et al. 2015). Hence, it is essential to evaluate their qualities in order to use these products in a climate change study.

To examine the quality of these data sources, many global-, continental-, or local-scale studies have been carried out. For instance, Simmons et al. (2004) compared ERA-40 and NCEP/NCAR reanalysis to CRU data for air temperature (CRUTEM2v) at 5 × 5° resolution on global and continental scales and concluded that there was very similar variability between CRUTEM2v and ERA-40, especially in the Northern Hemisphere from 1979 onwards. In a global comparison between interpolated observations and reanalysis data with 3.75 × 2.5°, Donat et al. (2014) showed that ERA-40 and ERA-interim had a better agreement than NCEP/NCAR reanalysis and JRA-25 for extreme temperature, and the reanalysis products for extreme precipitation performed with a low agreement but still correlated significantly. In the case of a national-scale evaluation, the performance over Iran was carried out by Raziei et al. (2011) by comparing GPCC Full Data Reanalysis Product Version 3 (GPCCv3) and NCEP/NCAR precipitation dataset, which showed that GPCCv3 could complement the observations but NCEP/NCAR reanalysis had significant discrepancies before the 1970s. A recent study over China by Gao et al. (2016) statistically evaluated ERA-20 cm, the latest ECMWF 20th-century reanalysis dataset. After comparing each ensemble at 0.5 × 0.5° grids for precipitation and temperature, it was concluded that generally all ensemble simulations were able to represent the real conditions on a comparable level.

It is important that comparative studies should cover a wide range of locations around the world and gaps should be filled in for the sites lacking such studies so that a clear pattern can be understood. In South Korea, the long term climate trend analysis on precipitation and temperature has generally been based on observed values (Chung & Yoon 2000; Chung et al. 2004; Chang & Kwon 2007; Bae et al. 2008; Jung et al. 2011) and the time range of these studies were limited in the late 20th century. There were a few trials to apply the interpolated datasets or reanalysis products on the climate trend research over Korea, but these datasets were applied to estimate the features of the comparable region like East-Asia, not South Korea itself (Ho et al. 2003; Jeong et al. 2015; Choi et al. 2016). In other words, the previous studies used the observation data to evaluate the climate trend in Korea, and compared the result with the Asian outputs based on the modelled datasets. In South Korea, the number of stations operated over 50 years is less than 15, although hundreds of gauging stations have been installed. For this reason, the time period for climate impact assessment has been limited up to the mid-20th century in South Korea. Thus, if the researchers would like to extend the study period, it is essential to attempt to determine a reliable long term dataset with high resolution, which should be explored. However, as aforementioned, there has been a lack of evaluation for the reliability and applicability of century-long reanalyses, as well as observation-based global climate data over South Korea.

Given this background, this study selected several century-long precipitation datasets (ERA-20 cm, ERA-20c, 20CR, CRU TS v.3.23 (CRUv3.23) and GPCC Full Data Reanalysis Product Version 7 (GPCCv7)) and temperature datasets (ERA-20 cm, ERA-20c, 20CR and CRUv3.23), covering the whole of the 20th century. ERA-40 was also considered as a benchmark for a half century reanalysis. By estimating the temporal variability, trend and statistical agreement for monthly values of each dataset in South Korea, this study focuses on the applicability, uncertainty and limitation of those multi-decadal datasets in the country-scale climate change study. For evaluation, we have assessed correlation coefficient $r$, the significance of trend by the Mann–Kendall test, and the skill score based on the
probability density functions (PDFs). The specification of the datasets and methodology applied in this study are introduced in the next section and the main results for precipitation and temperature follow. Finally, the discussion and conclusions are presented.

DATA

Observed local data

To analyse the precipitation and temperature change over the mainland of South Korea, daily total precipitations and daily mean 2-m air temperatures from 13 ground gauge stations, spanning 1961–2010, were taken from the data archive of Korea Meteorological Administration (KMA) (https://data.kma.go.kr/cmmn/main.do) and merged to the monthly values. In order to compare the datasets for the common period, the stations were evenly selected excluding islands of Korea from 1961 to 2001 with no empty values, although three of them were available from 1966, 1968 and 1973, respectively. The quality of the observations is strictly controlled by KMA. Detailed information on the location and data period of the stations is provided in Figure 1 and Table 1.

Reanalysis data

ERA-20c is the first atmospheric 20th century reanalysis of the ECMWF. This dataset, covering 1900–2010, is produced by assimilating observations of surface pressure and surface marine winds only (Poli et al. 2016). Considering the data availability and resolution of other datasets, we extracted total precipitation from the 24-hour accumulated forecasts and 2-m air temperature from 6-hourly analysis data with a 0.5 × 0.5° grid from January 1901 to December 2010 via the ECMWF web server. The products in South Korea were accumulated into monthly data and the values over the sea were excluded.

In addition to ERA-20c, the ECMWF also released ERA-20 cm data with a ten-member ensemble from January 1900 to December 2010 (Hersbach et al. 2015). Compared with ERA-20c, this dataset was produced with the same Integrated Forecasting System (IFS) version Cy38r1, but it includes no data assimilation (Donat et al. 2016). Three-hourly total precipitation and temperature data with a 0.5 × 0.5° grid from January 1901 to December 2010 were extracted from the web server and calculated as inland monthly datasets. In this study, to explore the general feature of ERA-20 cm ensemble, we used the ensemble mean and the ensemble member 0 (hereafter ‘En0’) only. A more detailed assessment on all ten ensemble members will be covered in another study.

To determine the difference among the ECMWF products, data named ERA-40, the 45-year reanalysis data from September 1957 to August 2002 (Uppala et al. 2005), were extracted from the ECMWF archive in the same way as ERA-20c. We collected the 6-hourly convective precipitation data, large-scale precipitation data and 2-m air temperature data on a 0.5 × 0.5° grid from January 1961 to December 2001. The total precipitation was produced by the sum of convective and large-scale precipitation, excluding the values on the sea, and the products were aggregated into monthly data.

20CR is one of the long term reanalysis datasets provided by the NOAA. Its latest version 2c, spanning 1850–2014 with a resolution of 1.875 × 1.9°, was produced by assimilating only surface pressures and using the Ensemble Kalman Filter technique to produce 56 ensemble members (Donat et al. 2016). Because each ensemble dataset is not
available in the web server, we collected only 8-times daily ensemble means for total precipitation and 2-m air temperature from 1901 to 2010 and accumulated them on a monthly basis. As with other datasets, the data over the sea were ignored.

**Gridded observations by CRU and GPCC**

CRU TS v.3.23 (CRUv3.23) is the recently updated time-series land-only dataset from 1901 to 2014, which covers the whole world, except the Antarctic (Harris et al. 2014). This dataset was constructed by using the Climate Anomaly Method based on worldwide observations providing monthly total precipitation and monthly mean 2-m air temperature with its highest resolution (0.5° × 0.5° latitude/longitude) (Harris et al. 2014). In this paper, for comparison with the observations and reanalysis datasets, the data over South Korea from 1901 to 2010 were extracted.

GPCC produced the global land-surface precipitation data, and its recent version, GPCC Full Data Reanalysis Version 7.0 (GPCCv7), covers a 111-year analysis period from 1901 to 2013 based on the rain gauge database over 51,000 stations worldwide ( Schneider et al. 2015). In this study, the monthly total precipitation product with the highest resolution of 0.5° × 0.5° over South Korea from 1901 to 2010 was taken from this dataset.

**METHODOLOGY**

**Evaluation of interannual variability**

To explore the temporal strength of the linear relationship between the model products and the observed values, the Pearson’s linear correlation coefficients (r) mean between the products and the observations of 13 stations from 1961 to 2001 were calculated. This method has been widely used to measure the degree of collinearity between the observed and the modelled data in the multi-decadal climate variability studies, although it is oversensitive to high extreme values and insensitive to proportional gaps between two variables (Legates & McCabe 1999; Deser et al. 2004; Herrmann et al. 2005; Dickinson et al. 2006; Wu et al. 2010; Gholami et al. 2015; Wang et al. 2015). Here, we focus on the variability between the observation and the modelled datasets using r, while the absolute differences between them are simply explored through figures on seasonal/annual change.

For this analysis, the seasonal/yearly total precipitation and mean temperature variables were derived from all the datasets. Every seasonal dataset was collected, i.e. for spring (March–May), summer (June–August), autumn (September–November), and winter (December–February).

In the case of r, considering the differences in coordinate and resolution between datasets (Figure 1), we

<table>
<thead>
<tr>
<th>No.</th>
<th>Name</th>
<th>Longitude</th>
<th>Latitude</th>
<th>Observation period</th>
<th>Elevation (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Seoul</td>
<td>126-57-56 E</td>
<td>37-54-17 N</td>
<td>1961-2010</td>
<td>11.1</td>
</tr>
<tr>
<td>2</td>
<td>Incheon</td>
<td>126-37-29 E</td>
<td>37-28-39 N</td>
<td>1961-2010</td>
<td>69.6</td>
</tr>
<tr>
<td>3</td>
<td>Seosan</td>
<td>126-29-45 E</td>
<td>36-46-25 N</td>
<td>1968-2010</td>
<td>30.3</td>
</tr>
<tr>
<td>4</td>
<td>Chuncheon</td>
<td>127-44-08 E</td>
<td>37-54-09 N</td>
<td>1966-2010</td>
<td>79.1</td>
</tr>
<tr>
<td>5</td>
<td>Gangneung</td>
<td>128-53-27 E</td>
<td>37-45-05 N</td>
<td>1961-2010</td>
<td>27.4</td>
</tr>
<tr>
<td>6</td>
<td>Jeonju</td>
<td>127-09-17 E</td>
<td>35-49-17 N</td>
<td>1961-2010</td>
<td>54.8</td>
</tr>
<tr>
<td>7</td>
<td>Chupungnyeong</td>
<td>127-59-40 E</td>
<td>36-13-11 N</td>
<td>1961-2010</td>
<td>246.1</td>
</tr>
<tr>
<td>8</td>
<td>Yeongju</td>
<td>128-31-00 E</td>
<td>36-52-18 N</td>
<td>1973-2010</td>
<td>212.2</td>
</tr>
<tr>
<td>9</td>
<td>Gwangju</td>
<td>126-53-29 E</td>
<td>35-10-22 N</td>
<td>1961-2010</td>
<td>73.8</td>
</tr>
<tr>
<td>10</td>
<td>Yeosu</td>
<td>127-44-26 E</td>
<td>34-44-21 N</td>
<td>1961-2010</td>
<td>66.0</td>
</tr>
<tr>
<td>11</td>
<td>Daegu</td>
<td>128-37-08 E</td>
<td>35-53-06 N</td>
<td>1961-2010</td>
<td>65.5</td>
</tr>
<tr>
<td>12</td>
<td>Pohang</td>
<td>129-22-46 E</td>
<td>36-01-57 N</td>
<td>1961-2010</td>
<td>3.7</td>
</tr>
<tr>
<td>13</td>
<td>Busan</td>
<td>129-01-55 E</td>
<td>35-06-16 N</td>
<td>1961-2010</td>
<td>71.0</td>
</tr>
</tbody>
</table>
interpolated the data in each station point by using an inverse distance (ID) method, one of the most applied deterministic methods (Babak & Deutsch 2009). Compared with other preferred methods, kriging, the ID method is simple to calculate, more applicable to spatial estimation with small-sized observation networks and does not require prior information such as a semi-variogram model (Tomczak 1998; Lu & Wong 2008; Babak & Deutsch 2009). For this reason, this study applied the ID method as follows:

\[ w(x, y) = \sum_{i=1}^{N} \alpha_i w_i, \quad \alpha_i = \frac{(1/d_i)^p}{\sum_{i=1}^{N} (1/d_i)^p} \]  

(1)

where \(N\) is the number of grids used in calculation, \(w\) is the evaluated value from the data product in each station point, \(w_i\) is the \(i\)th data point among the selected values, \(d_i\) is the distance from the station to the \(i\)th grid, and \(p\) is the specified weighting power. In this equation, the weighting parameter, \(p\), can vary from 0 to infinite, and when the value increases, the estimated is less influenced by the further stations (Chang et al. 2006). In this analysis, all inland gridded values were used and the most common value, 2, was applied for the power \(p\) (Teegavarapu & Chandramouli 2005; Babak & Deutsch 2009). After calculating the \(r\) in 13 stations, the mean \(r\) values were compared.

### Trend test

The correlation coefficient, \(r\), does not represent the slope of the line of the best fit, although it shows the relationship between the observed and the model. Thus, to determine the significance of linear trends in each dataset, the Mann–Kendall test was applied for the reference period 1961–2001. The trends from 1901 to 2010 of several precipitation datasets (ERA-20 cm, ERA-20c, 20CR, CRU v3.23 and GPCC v7) and temperature datasets (ERA-20 cm, ERA-20c, 20CR and CRU v3.23) were also evaluated in order to assess long term patterns throughout the 20th century. The Mann–Kendall trend test created by Mann (1945) and Kendall (1955) is one of the widely used nonparametric tests for detecting the trends of environmental data, such as precipitation, temperature and streamflow (Xu et al. 2005; Bae et al. 2008; Shadmani et al. 2012; Zang & Liu 2013). Compared with parametric tests such as linear regression, which require data normality as well as independence, this method only requires the independence of data (Hamed & Rao 1998; Xu et al. 2005). In the Mann–Kendall test, the test statistic \(S\) and the standardized test statistic \(Z\) are estimated by the related equations as follows:

\[ S = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \text{sgn}(x_j - x_i) \]  

(2)

\[ \text{sgn}(x) = \begin{cases} 1 & \text{for } x > 0 \\ 0 & \text{for } x = 0 \\ -1 & \text{for } x < 0 \end{cases} \]  

(3)

\[ V(S) = \frac{n(n-1)(2n+S)-\sum_{i=1}^{m}t_i(t_i-1)(2t_i+5)}{18} \]  

(4)

\[ Z = \begin{cases} \frac{S-1}{\sqrt{V(S)}} & \text{for } S > 0 \\ 0 & \text{for } S = 0 \\ \frac{S+1}{\sqrt{V(S)}} & \text{for } S < 0 \end{cases} \]  

(5)

where \(x_1, x_2, x_3, \ldots, x_n\) are the time series of length \(n\), \(V(S)\) is the variance of \(S\), \(m\) is the number of tied groups, \(t_i\) is the number of ties for the \(i\)th value, and \(Z\) follows a standard normal distribution (Xu et al. 2008). The significance of trends is evaluated by comparing \(Z\) with the standard normal variate at the desired significance (Hamed & Rao 1998).

When \(|Z| > Z_{1-a/2}\), the null hypothesis is rejected and it means that there is a significant trend in the time series in the test. Here, \(Z_{1-a/2}\) is obtained from the standard normal distribution corresponding to the significance level \(a\). In this study, both 0.05 and 0.10, the most commonly used values, were applied for \(a\), although significance levels can vary from 0.1 to 0.001 according to the study (Hamed & Rao 1998; Xu et al. 2005; Bae et al. 2008; Jung et al. 2011; Zang & Liu 2013). The magnitude of the linear trend for this method is estimated by the Theil–Sen approach, sometimes referred to as ‘Kendall Slope Estimator’, defined by the median value of the ranked slope estimates as follows (Theil 1950; Sen 1968; Hirsch et al. 1982; Xu et al. 2008; Zang & Liu 2013):

\[ \beta = \text{Median} \left( \frac{x_j - x_i}{j-i} \right), \quad \forall i < j \]  

(6)

In this equation, the positive value of \(\beta\) represents the increasing trend over time, while the negative value means the opposite trend. The advantage of this method is that it is
less sensitive to outliers or extreme values than the least-square method (Shadmani et al. 2012; Sayemuzzaman & Jha 2014).

PDF-based evaluation method

To assess the statistical similarity between the observations and each dataset from 1961 to 2001, we estimated the skill score based on the PDF suggested by Perkins et al. (2007). This method was very simple but powerful in capturing the relative compatibility between observation and model distribution. In addition, compared with the traditional mean-based method, this performance showed more credible climate variations and it is flexible to collect data with different time periods from multiple stations (Perkins et al. 2007; Gao et al. 2016). By calculating the overlapped area between two distributions at each bin, this skill estimated how much the climate dataset distribution was similar to the observed. If a dataset matches the observed values perfectly in PDF, the skill score will be 1, which equals the sum of the probability. Otherwise, if the skill score is close to zero, it means that there is no common area between the model values and observations. In other words, the more overlapped the two curves, the closer to 1 this score is. The skill score is calculated as follows:

$$S_{score} = \sum_{i=1}^{n} \min \left( P_m, P_0 \right),$$

where $n$ is the number of bins for the calculation, $P_m$ is the frequency of values in a given bin from a comparison target, and $P_0$ is the frequency values in a given bin from observations. In this study, the monthly variables were applied and the square root of 1 mm month$^{-1}$ for precipitation and 1°C for temperature were used as the intervals of bins to effectively compare two PDFs as in earlier studies (Perkins et al. 2007; Gao et al. 2016).

RESULTS

Precipitation

Interannual variability

Table 2 quantitatively explains the seasonal/annual correlation between the observation and the simulated precipitation from 1961 to 2001. In the seasonal mean comparison, the $r$ values for CRUv3.23 and GPCCv7 exceed 0.9 in every season, and ERA-20c, ERA-40 and 20CR performs moderate to high correlations (0.4 < $r$ < 0.9). Among seasonal values, spring and winter are more correlated than summer and autumn. However, the simulations for ERA-20 cm mean and En0 were located between −0.149 and 0.313, which means that there is little temporal correlation with the observation for precipitation. A similar result is described in the annual mean comparison. CRUv3.23 and GPCCv7 performed very well with the $r$ over 0.9 and ERA-20c follows with 0.621. 20CR and ERA-40 have moderate correlations with 0.498 and 0.445, respectively, but the $r$ values for ERA-20 cm mean and En0 were close to zero.

Figure 2, which illustrates the seasonal and annual precipitation change of each dataset from 1961 to 2001, supports this result. For seasonal comparison, the fluctuations of ERA-20 cm mean and En0 have little correlation with the observations in all seasons, while GPCCv7 and CRUv3.23 performed almost in similar movements with the observed values (Figure 2(a)). For ERA-20c, ERA-40 and 20CR, their movements had significant similarities to the observations, but the values themselves of each dataset were slightly different. For example, ERA-40 and 20CR had lower rainfall than the observation, especially, in summer and autumn, whereas ERA-20c was relatively close to the observation (Figure 2(a)). This means that in terms of interannual variability, ERA-20c was less biased than ERA-40 and 20CR in South Korea. The annual change showed a similar result with the seasonal trend. The annual patterns of ERA-20 cm mean and En0 were totally different from that

Table 2 | Correlation coefficient (r) for seasonal and annual total precipitation for each dataset averaged over all regions from 1961 to 2001

<table>
<thead>
<tr>
<th>Type</th>
<th>Seasonal comparison</th>
<th>Annual comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Spring</td>
<td>Summer</td>
</tr>
<tr>
<td>ERA-20 cm (Mean)</td>
<td>−0.110</td>
<td>0.070</td>
</tr>
<tr>
<td>ERA-20 cm (En0)</td>
<td>0.042</td>
<td>0.313</td>
</tr>
<tr>
<td>ERA-20c</td>
<td>0.744</td>
<td>0.407</td>
</tr>
<tr>
<td>20CR</td>
<td>0.744</td>
<td>0.407</td>
</tr>
<tr>
<td>CRUv3.23</td>
<td>0.970</td>
<td>0.938</td>
</tr>
</tbody>
</table>
Figure 2 | Total precipitation change for observation (Obs), Mean of ERA-20 cm (ERA-20 cm (Mean)), En0 of ERA-20 cm (ERA-20 cm (En0)), ERA-20c, ERA-40, 20CR, CRUv3.23 (CRU) and GPCCv7 (GPCC) averaged over the whole region from 1961 to 2001. (a) The seasonal total precipitation change (from above, spring, summer, autumn, and winter). (b) The annual total precipitation change.
of the observation, while CRUv3.23 and GPCCv7 performed very well (Figure 2(b)). For ERA-20c, ERA-40 and 20CR, they had partial similarity to the observation in the annual comparison, but only ERA-20c had the equivalent value with the observed (Figure 2(b)). In other words, ERA-40 and 20CR were clearly underestimated.

**Long term trend**

Table 3 shows the long term trends derived by the Mann–Kendall test. The standardised statistics (Z) for the reference period 1961–2001 state that there are no significant seasonal/annual trends at 90 or 95% confidence level for ERA-20 cm, ERA-20c, CRUv3.23 and GPCCv7, as well as the observation. Only ERA-40 in summer and 20CR in spring have an increasing and decreasing trend at 95% confidence level, respectively. With 90% confidence level, a further declining trend is found in the annual trend for 20CR.

The analysis from 1901 to 2010 showed more obvious trends. For ERA-20 cm, the trends of the mean and En0 were different. ERA-20 cm mean had significant increasing trends in spring, winter and annual simulations, while En0 had no significant trends. Comparing ERA-20 cm with CRUv3.23 and GPCCv7, they have no similarity in the seasonal trends and the magnitude of the slopes for ERA-20 cm were generally lower than those of CRUv3.23 and GPCCv7, except for winter. For instance, CRUv3.23 and GPCCv7 had increasing trends in summer with slopes of 1.70 and 1.79, but ERA-20 cm mean had upward trends in spring and winter with slopes of 0.20 and 0.15. In the case of ERA-20c, it performed the obvious increasing movement in every test and had a stronger increasing trend than CRUv3.23 and GPCCv7 in summer and annual tests. This showed that ERA-20c can exaggerate the long term trend for precipitation compared with the other datasets. On the other hand, 20CR performed downward trends in the summer, autumn and annual test. That is to say, the long term trend of 20CR was in contrast to the movements of other datasets.

### Table 3 | Mann–Kendall test results for precipitation trend

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Spring</th>
<th></th>
<th>Summer</th>
<th></th>
<th>Autumn</th>
<th></th>
<th>Winter</th>
<th></th>
<th>Annual</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Z</td>
<td>β</td>
<td>Z</td>
<td>β</td>
<td>Z</td>
<td>β</td>
<td>Z</td>
<td>β</td>
<td>Z</td>
<td>β</td>
</tr>
<tr>
<td>1961–2001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observation</td>
<td>−1.00</td>
<td>−1.38</td>
<td>1.13</td>
<td>2.78</td>
<td>−0.51</td>
<td>−0.74</td>
<td>0.30</td>
<td>0.14</td>
<td>−0.08</td>
<td>−0.39</td>
</tr>
<tr>
<td>ERA-20 cm (Mean)</td>
<td>1.49</td>
<td>0.42</td>
<td>−0.48</td>
<td>−0.48</td>
<td>−0.30</td>
<td>−0.19</td>
<td>−0.93</td>
<td>−0.58</td>
<td>−0.12</td>
<td>−0.27</td>
</tr>
<tr>
<td>ERA-20 cm (En0)</td>
<td>1.49</td>
<td>1.66</td>
<td>0.10</td>
<td>0.31</td>
<td>0.19</td>
<td>0.16</td>
<td>−0.35</td>
<td>−0.24</td>
<td>0.28</td>
<td>0.91</td>
</tr>
<tr>
<td>ERA-20c</td>
<td>−0.33</td>
<td>−0.43</td>
<td>0.19</td>
<td>0.45</td>
<td>1.20</td>
<td>1.06</td>
<td>0.62</td>
<td>0.26</td>
<td>0.39</td>
<td>1.43</td>
</tr>
<tr>
<td>ERA-40</td>
<td>−0.21</td>
<td>−0.16</td>
<td>2.01</td>
<td>3.44</td>
<td>1.43</td>
<td>1.02</td>
<td>0.39</td>
<td>0.16</td>
<td>1.47</td>
<td>3.95</td>
</tr>
<tr>
<td>20CR</td>
<td>−1.99a</td>
<td>−2.83</td>
<td>0.15</td>
<td>0.21</td>
<td>−0.46</td>
<td>−0.27</td>
<td>−0.86</td>
<td>−0.75</td>
<td>−1.83b</td>
<td>−5.11</td>
</tr>
<tr>
<td>CRUv3.23</td>
<td>−1.00</td>
<td>−1.43</td>
<td>1.20</td>
<td>2.36</td>
<td>−0.55</td>
<td>−0.78</td>
<td>−0.01</td>
<td>−0.02</td>
<td>−0.10</td>
<td>−0.42</td>
</tr>
<tr>
<td>GPCCv7</td>
<td>−1.02</td>
<td>−1.40</td>
<td>0.86</td>
<td>1.40</td>
<td>−0.12</td>
<td>−0.33</td>
<td>−0.06</td>
<td>−0.05</td>
<td>0.00</td>
<td>0.11</td>
</tr>
</tbody>
</table>

| 1901–2010        |        |       |        |       |        |       |        |       |        |       |
| ERA-20 cm (Mean) | 2.58a  | 0.20  | 0.52   | 0.11  | 1.19   | 0.12  | 1.82   | 0.15  | 2.14a  | 0.60  |
| ERA-20 cm (En0)  | 0.11   | 0.02  | 0.80   | 0.38  | −0.60  | −0.13 | 0.24   | 0.04  | 0.30   | 0.18  |
| ERA-20c          | 4.95a  | 1.21  | 3.97   | 2.04  | 4.53a  | 1.01  | 5.97a  | 0.55  | 6.09a  | 5.00  |
| 20CR             | −0.11  | −0.02 | −2.19a | −0.76 | −2.32a | −0.48 | 1.84b  | 0.25  | −1.76b | −1.13 |
| CRUv3.23         | 0.80   | 0.19  | 3.00a  | 1.70  | 1.51   | 0.46  | −0.68  | −0.07 | 2.72a  | 2.13  |
| GPCCv7           | 0.54   | 0.11  | 3.42a  | 1.79  | 1.51   | 0.46  | −1.35  | −0.16 | 2.86a  | 2.14  |

*aSignificant trend at the 0.05 significance level.

*bSignificant trend at the 0.10 significance level. β (trends for precipitation) are in mm/yr.
**Statistical comparability**

Figure 3 describes the statistical agreement between the observation and each dataset from 1961 to 2001. CRUv3.23 and GPCCv7 performed the best simulations with the skill score of approximately 0.94, and ERA-20c followed them closely with 0.93. This indicates that ERA-20c had statistical similarity with the observed at almost the same level as CRUv3.23 and GPCCv7. The scores for 20CR, ERA-40 and En0 were between 0.8 and 0.85, which shows significant agreements, whereas ERA-20 cm mean had a clearly smaller value, 0.66.

The specific discrepancies of each dataset are described in Figure 4(a), which illustrates the PDFs of the observation and each precipitation dataset over South Korea from 1961 to 2001, and Figure 4(b) which represents seasonally subdivided PDFs. It is obvious that ERA-20c as well as CRUv3.23 and GPCCv7 was one of the most fitted datasets to the observation with little discrepancies. However, the other datasets have partial gaps from the observation. For 20CR, the PDF in Figure 4(a) shows that it underestimated over 200 mm month$^{-1}$ and overestimated in the range of 25–100 mm month$^{-1}$. This result was mainly due to the underestimated values in summer, as seen in Figure 4(b). The left-biased summer rainfalls led to overestimation of moderate values and underestimation of intensive values. The PDF of ERA-40 in Figure 4(a) overall exaggerated the frequency under 50 mm month$^{-1}$ and underestimated over 200 mm month$^{-1}$. It comes from the generally underestimated distributions in all seasons, especially in summer (Figure 4(b)). In the case of ERA-20 cm mean and En0, the dry months and intensive rainfall months were underestimated but the moderate months were overestimated in Figure 4(a). It is clear that the mean of ERA-20 cm had this tendency more strongly than En0. This evaluation suggests that all datasets showed significant agreement with the observation, albeit some of them still needed a cautious approach to use in the frequency analysis.

**Temperature**

**Interannual variability**

Table 4 describes the $r$ values between the gauged temperature and the model temperature from 1961 to 2001. In seasonal comparison, CRUv3.23 and ERA-40 had the highest values over 0.9 in every season and ERA-20c followed closely with 0.830 to 0.914. 20CR had the high correlations (0.6 < $r$ < 0.9) and the values for ERA-20 mean and En0 were the lowest ones. To be more specific, ERA-20 cm mean had moderate correlations (0.4 < $r$ < 0.7) in four seasons, while En0 had low correlations, except spring. Of the four seasons, winter had the highest value except ERA-20 cm mean and En0. Theses seasonal findings were similar to the annual simulations. In annual comparison, CRUv3.23 and ERA-40 showed the most fitted correlations with the $r$ values over 0.9 and ERA-20c closely followed them with 0.879. 20CR had the 0.808 and ERA-20 cm mean (0.714) and En0 (0.523) had moderate to high correlations.

Figure 5 demonstrates the seasonal and annual mean temperature trends of each dataset over South Korea from 1961 to 2001. In Figure 5, we can see that each dataset
Figure 4 | Probability density functions (PDFs) for monthly total precipitation for observation (Obs), Mean of ERA-20 cm (ERA-20 cm (Mean)), En0 of ERA-20 cm (ERA-20 cm (En0)), ERA-20c, ERA-40, 20CR, CRUv3.23 (CRU) and GPCCv7 (GPCC) over South Korea: (a) PDFs for monthly total precipitation from 1961 to 2001; (b) PDFs for seasonally subdivided monthly total precipitation from 1961 to 2001.
performed similar movements to the observations, but the values themselves were different depending on the dataset, except ERA-40. For ERA-20 cm and ERA-20c, the seasonal/annual variations seemed to have partial correlations, but the model values were generally about 1–2 °C lower than those of observations, except the winter season. In the case of CRUv3.23, it is clear that the mean temperature for CRUv3.23 was about 2 °C lower than the observation in every comparison, although its variation trends had similarity to the observations. On the other hand, 20CR had higher values than the observation in the annual comparison, affected by the autumn and winter temperature. Only ERA-40 was very well fitted to the observed values in every comparison. This result implies that, despite the significant correlations between the observation and each dataset, the bias correction should be considered before using them.

### Long term trend

Table 5 describes the seasonal and the annual patterns of the mean temperature by the Mann–Kendall approach. In the first analysis from 1961 to 2001, the result suggested that only ERA-40 had increasing trends in spring, winter and annual simulations, as well as the observations. Seasonally, the other datasets also had an upward trend in spring, but they showed different trends in other seasons. For CRUv3.23 and ERA-20c, there were significant increasing trends in spring, autumn and winter, whereas 20CR had trends in spring and autumn. In the case of ERA-20 cm, the mean had increasing trends in spring and summer at 95% confidence level, but En0 had it only in spring at 90% confidence level. In terms of annual analysis, all datasets except En0 showed upward trends. ERA-20c, 20CR and CRUv3.23 showed significant upward trends at 95% confidence level, and ERA-40 and ERA-20 cm mean suggested them at 90% confidence level. For the slope of the annual comparison, those of CRUv3.23 and 20CR were higher than the observation’s, while ERA-20c and ERA-40 were slightly smaller than the observed. In the case of ERA-20 cm mean, the slope showed less than half of the observation’s.

The second analysis for the 20th century indicates the obvious increasing trends in all seasonal and annual simulations at 95% confidence level in Table 5. The only difference between datasets is the intensity of the slopes. As with the first analysis, the increasing magnitudes (β) of 20CR and CRUv3.23 were generally higher than those of the others. This result implies that the mean temperature in South Korea has obviously increased over the past 100 years.

### Statistical comparability

Figure 6 represents the skill score of the PDF of each dataset for monthly mean temperature from 1961 to 2001. The estimate of ERA-40 was approximately 0.90, and 20CR, En0, ERA-20c and CRUv3.23 follows with 0.74, 0.71, 0.69 and 0.69, respectively. In other words, ERA-40 reanalysis had a probability density distribution approximately equal to the observed, and 20CR, En0, ERA-20c and CRUv3.23 also had significant agreements with them. Remembering the high r values for ERA-20c and CRUv3.23 in the annual comparison (r > 0.87), this result suggests that, despite the high correlation with the observation, ERA-20c and CRUv3.23 were clearly biased and they, as well as other datasets, need bias correction for application in climate change studies over South Korea. The mean of ERA-20 cm also had meaningful agreement, but not as much as En0.

Figure 7 illustrates the PDFs of the observed, and the modelled dataset for temperature supports the skill score analysis. The performance of ERA-40 generally showed high agreement in all comparisons with the observations, but the other datasets had seasonally biased distributions.
Figure 5 | Mean temperature change for observation (Obs), Mean of ERA-20 cm (ERA-20 cm (Mean)), En0 of ERA-20 cm (En0), ERA-20c, ERA-40, 20CR, and CRUv3.23 (CRU) averaged over the whole region from 1961 to 2001. (a) The seasonal mean temperature change (from above, spring, summer, autumn, and winter). (b) The annual mean temperature change.
The seasonally subdivided PDFs help to determine the difference of each dataset by comparing the peaks of them (Figure 7(b)). For reference, the three peaks seen in spring and autumn were due to the rapid change in monthly mean temperature. For ERA-20 cm mean, En0 and ERA-20c, the distributions were located in the left of the observations in the rest of the seasons, except winter in Figure 7(b). Likewise, for CRUv3.23, the PDFs were located in the left side of the observation in every season (Figure 7(b)) and caused the generally left-biased distribution in Figure 7(a). In the case of 20CR, the PDF in Figure 7(a) seemed to perform well, except for the underestimation of the range of below 0°C and partial discrepancies, but the seasonal PDFs implied that this result has been refined in the process of combining seasonal discrepancies (Figure 7(b)). For instance, the second and third peaks of 20CR in spring represent the lower temperature than the real, but the PDF for winter shows warmer

### Table 5 | Mann–Kendall test results for temperature trend

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Spring</th>
<th></th>
<th></th>
<th></th>
<th>Summer</th>
<th></th>
<th></th>
<th></th>
<th>Autumn</th>
<th></th>
<th></th>
<th></th>
<th>Winter</th>
<th></th>
<th></th>
<th></th>
<th>Annual</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Z</td>
<td>β</td>
<td>Z</td>
<td>β</td>
<td>Z</td>
<td>β</td>
<td>Z</td>
<td>β</td>
<td>Z</td>
<td>β</td>
<td>Z</td>
<td>β</td>
<td>Z</td>
<td>β</td>
<td>Z</td>
<td>β</td>
<td>Z</td>
<td>β</td>
<td>Z</td>
<td>β</td>
</tr>
<tr>
<td>1961–2001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observation</td>
<td>2.62</td>
<td>a</td>
<td>2.53</td>
<td></td>
<td>0.57</td>
<td>0.64</td>
<td>0.84</td>
<td>0.70</td>
<td>2.30</td>
<td>a</td>
<td>3.65</td>
<td></td>
<td>3.66</td>
<td>a</td>
<td>2.09</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ERA-20 cm (Mean)</td>
<td>2.62</td>
<td>a</td>
<td>1.17</td>
<td></td>
<td>2.12</td>
<td>a</td>
<td>1.31</td>
<td></td>
<td>0.78</td>
<td>0.59</td>
<td>0.75</td>
<td>0.33</td>
<td>1.81</td>
<td>b</td>
<td>0.86</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ERA-20 cm (En0)</td>
<td>1.85</td>
<td>b</td>
<td>1.21</td>
<td></td>
<td>0.69</td>
<td>0.48</td>
<td>−0.01</td>
<td>−0.03</td>
<td>0.55</td>
<td>0.44</td>
<td>0.55</td>
<td>0.44</td>
<td>0.86</td>
<td>b</td>
<td>0.42</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ERA-20c</td>
<td>1.67</td>
<td>b</td>
<td>1.40</td>
<td></td>
<td>0.46</td>
<td>0.33</td>
<td>1.92</td>
<td>b</td>
<td>1.30</td>
<td></td>
<td>2.62</td>
<td>a</td>
<td>2.38</td>
<td></td>
<td>2.71</td>
<td>a</td>
<td>1.80</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ERA-40</td>
<td>2.55</td>
<td>a</td>
<td>2.22</td>
<td></td>
<td>−0.08</td>
<td>−0.12</td>
<td>0.46</td>
<td>0.52</td>
<td>2.26</td>
<td>b</td>
<td>2.41</td>
<td></td>
<td>1.81</td>
<td>b</td>
<td>1.47</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20CR</td>
<td>2.82</td>
<td>a</td>
<td>2.20</td>
<td></td>
<td>1.54</td>
<td>1.86</td>
<td>2.41</td>
<td>a</td>
<td>2.18</td>
<td></td>
<td>1.38</td>
<td>1.99</td>
<td>2.77</td>
<td>a</td>
<td>2.34</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CRUv3.23</td>
<td>3.36</td>
<td>a</td>
<td>3.07</td>
<td></td>
<td>1.61</td>
<td>1.38</td>
<td>2.62</td>
<td>a</td>
<td>2.16</td>
<td></td>
<td>2.86</td>
<td>a</td>
<td>3.53</td>
<td></td>
<td>3.31</td>
<td>a</td>
<td>2.91</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1901–2010</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ERA-20 cm (Mean)</td>
<td>8.69</td>
<td>a</td>
<td>1.02</td>
<td></td>
<td>7.60</td>
<td>a</td>
<td>0.91</td>
<td></td>
<td>6.61</td>
<td>a</td>
<td>0.81</td>
<td></td>
<td>6.61</td>
<td>a</td>
<td>0.79</td>
<td></td>
<td>9.38</td>
<td>a</td>
<td>0.95</td>
<td></td>
</tr>
<tr>
<td>ERA-20 cm (En0)</td>
<td>4.90</td>
<td>a</td>
<td>1.02</td>
<td></td>
<td>4.02</td>
<td>a</td>
<td>0.60</td>
<td></td>
<td>4.17</td>
<td>a</td>
<td>0.80</td>
<td></td>
<td>3.19</td>
<td>a</td>
<td>0.61</td>
<td></td>
<td>6.33</td>
<td>a</td>
<td>0.80</td>
<td></td>
</tr>
<tr>
<td>ERA-20c</td>
<td>3.25</td>
<td>a</td>
<td>0.66</td>
<td></td>
<td>2.94</td>
<td>a</td>
<td>0.49</td>
<td></td>
<td>4.06</td>
<td>a</td>
<td>0.86</td>
<td></td>
<td>6.15</td>
<td>a</td>
<td>1.55</td>
<td></td>
<td>7.02</td>
<td>b</td>
<td>1.04</td>
<td></td>
</tr>
<tr>
<td>20CR</td>
<td>6.73</td>
<td>a</td>
<td>1.71</td>
<td></td>
<td>5.14</td>
<td>a</td>
<td>1.35</td>
<td></td>
<td>6.11</td>
<td>a</td>
<td>1.43</td>
<td></td>
<td>6.36</td>
<td>a</td>
<td>2.09</td>
<td></td>
<td>8.15</td>
<td>a</td>
<td>1.80</td>
<td></td>
</tr>
<tr>
<td>CRUv3.23</td>
<td>7.93</td>
<td>a</td>
<td>1.99</td>
<td></td>
<td>3.77</td>
<td>a</td>
<td>0.85</td>
<td></td>
<td>5.96</td>
<td>a</td>
<td>1.51</td>
<td></td>
<td>5.15</td>
<td>b</td>
<td>1.62</td>
<td></td>
<td>7.85</td>
<td>a</td>
<td>1.61</td>
<td></td>
</tr>
</tbody>
</table>

*Significant trend at the 0.05 significance level.

*Significant trend at the 0.10 significance level. β (trends for temperature) are in 10−2°C/yr.
Figure 7 | Probability density functions for monthly mean temperature for observation (Obs), Mean of ERA-20 cm (ERA-20 cm (Mean)), En0 of ERA-20 cm (ERA-20 cm (En0)), ERA-20c, ERA-40, 20CR, and CRUv3.23 (CRU) over South Korea: (a) PDFs for monthly mean temperature from 1961 to 2001; (b) PDFs for seasonally subdivided monthly mean temperature from 1961 to 2001.
temperature than the observation. This suggests that statistical usage of 20CR without considering this seasonal deviation could distort the simulation.

**SUMMARY AND DISCUSSION**

This study evaluated the multi-decadal reanalysis datasets, ERA-20 cm, ERA-20c, ERA-40 and 20CR, and two century-long gridded observation datasets, CRUv3.23 and GPCCv7, over South Korea. The authors mainly focused on temporal and statistical applicability of monthly mean values for precipitation and temperature, which are the most commonly used data in climate change studies (Gao et al. 2016).

In the temporal variability comparison for precipitation, the $r$ values for ERA-20 cm mean and En0 compared with the observation derived from the 13 gauged stations were closed to 0, while CRUv3.23 and GPCCv7 exceeded 0.9 in every seasonal/annual comparison. This result reconfirms the well-known feature of ERA-20 cm, which cannot reproduce the actual synoptic situation for precipitation (Hersbach et al. 2015). On the other hand, the other reanalyses, ERA-20c, ERA-40 and 20CR, had moderate to high correlations (0.4 $< r < 0.9$) and, of them, ERA-40 and 20CR had the seasonal gaps compared with the observations. This suggests that it is of importance to consider the local accuracy in national-scale studies using these datasets.

For the trend test on precipitation, there was no significant trend except ERA-40 for summer and 20CR in spring and annual trends for the reference period 1961–2001. However, the simulation from 1901 to 2010 shows different trends depending on the dataset. CRUv3.23 and GPCCv7 had identically increasing trends in summer and 12-month average simulations, whereas ERA-20c showed upward tendencies in all tests and 20CR had decreasing trends in summer, autumn and annual simulations. For ERA-20 cm, the mean showed increasing trends in spring, winter and annual tests, while En0 had no significant trends. It is clear that the results of the trend analysis can vary depending on the study period and regions in South Korea (Bae et al. 2008). Nevertheless, the previous long term trend studies have shown that summer precipitation observed in South Korea has generally increased (Wang et al. 2006; Chang & Kwon 2007; Choi et al. 2009; Jung et al. 2011). Chang & Kwon (2007) and Jung et al. (2011) suggested that all stations had increasing summer rainfalls since 1973. Choi et al. (2009) compared the gauged rainfalls of ten Asian countries from 1955 to 2007 and described significant increasing summer rainfall in South Korea at 95% confidence level. The longest trend analysis on Seoul, the capital of South Korea, also indicated a significant upward trend from 1778 to 2004, although the estimate for the pre-1950 period suggested no significant trend (Wang et al. 2006). Hence, the decreasing tendency of 20CR implies that despite the meaningful correlation with the observation, 20CR is able to provide distorted information in the long term trend analysis. In terms of the intensity of the annual trend, in the test from 1901 to 2010, CRUv3.23 and GPCCv7 had significant increasing annual slopes (mm/yr) of 2.13 and 2.14, respectively. These trends are different from Harris et al. (2014) which suggested 0.005 for CRU TS3.10 (CRUv3.10), the earlier version of CRUv3.23, and –0.019 for GPCC version 5, the earlier version of GPCCv7, in East Asia from 1901 to 2009. However, Choi et al. (2009), by evaluating the observations for the 1955–2007 period, showed that South Korea had a significant increasing trend (2.45) and it was much higher than China (0.33) and Japan (–1.75), the other East Asia countries. This supports the fact that the slopes of CRUv3.23 and GPCCv7 have reliability.

For statistical evaluation for precipitation, there were significant agreements between the monthly averaged observations derived from 15 gauged stations and each dataset. The skill scores for ERA-20c as well CRUv3.23 and GPCCv7, as the interpolated datasets, exceeded 0.9, and the other reanalyses were over 0.8, except ERA-20 cm mean which was 0.66. Gao et al. (2016) concluded that despite the spatial variability, all ten ensemble members of ERA-20 cm for precipitation had high skill scores, over 0.8, in China. However, the two obviously different values of ERA-20 cm mean and En0 imply that the mean has difficulty in representing individual ensemble members. In other words, an ensemble member can better describe the climate change than the mean in a certain area. Nevertheless, this evaluation suggests that all the reanalyses, including ERA-20 cm, can be applied to rainfall frequency analysis as an alternative to the observation after proper bias corrections.
For temperature, the $r$ values of ERA-20 cm mean had moderate correlations with the observations, whereas En0 performed with low to moderate correlations. An interesting point is that the most fitted dataset was ERA-40 reanalysis, not CRUv3.23 which represents the interpolated observation dataset. In CRUv3.23 as well as ERA-20c and 20CR, there were obvious gaps in the observations for temperature. The earlier comparative study for temperature from 1958 to 2001 described that CRUTEM2v, the earlier version of CRUv3.23, had significantly lower temperature than ERA-40 in the northern hemispheres from 1958 to 1967 because of the limited availability of observations (Simmons et al. 2004). However, in this study, the annual discrepancy is shown in South Korea over the whole period 1961–2001, although it has been narrow.

In the temperature trend test, ERA-40 showed identical tendencies to the observed which have upward trends in spring, winter and annual simulations for the 1961–2001 period, and ERA-20c, 20CR and CRUv3.23 also have similarities, except for autumn. On the other hand, ERA-20 cm mean showed significant increasing movements in spring, summer, and annual tests, whereas En0 had it only in spring. Although there are trend variations according to the study period and spatial distribution (Bae et al. 2008), the previous observed trends in South Korea for the late 20th century suggested that the winter and annual mean temperature had significant upward trends but the summer trend was weak (Chung & Yoon 2000; Jung et al. 2002; Choi et al. 2009). Hence, it could be deduced that ERA-20 cm mean, which showed strong summer and weak winter trends, has little reliability in terms of long term trend. An interesting point is that the second trend assessment from 1901 to 2010 indicates significant warming trends in all the simulations at the 0.95 confidence level, although the intensity of the slope was different depending on the dataset. This trend has been shown in recent researches. Donat et al. (2016) suggested warming trends over the world in their multidata sources analysis from 1901 to 2010, and Harris et al. (2014) showed the annual warming trend in East Asia, 0.11 °C/decade, by using CRUv3.10 from 1901 to 2008. For this reason, the increasing trends over 100 years in this paper are highly reliable, although the magnitudes of them have uncertainty.

In the case of PDFs analysis, ERA-40 performed the best with skill score of 0.90 and 20CR, ERA-20c and En0 as well as CRUv3.23 had significant agreements to the observation with values between 0.69 and 0.74. Conversely, ERA-20 cm mean had the lowest value, 0.58. This simulation indicates that these datasets have significant reliability for monthly frequency for temperature in Korea, but it is still challenging to apply them directly. In terms of ERA-20 cm, Gao et al. (2016) showed that the skill scores of all ten ensemble members averaged over all regions in China for temperature exceeded 0.9, but the skill scores for ERA-20 cm mean and En0 in this study had much lower values. This suggests that there may be a clear difference in applicability according to the region, and this gap should be explored before using the dataset on a regional scale study.

Considering the improved assimilation and ensemble technique, it is easy to hypothesise that the higher the temporal and spatial resolutions, the more accurate the reanalysis dataset should be in terms of temporal and statistical variability. However, the results in this study indicate that each dataset has its own bias and the degree of the agreement of each data can vary in space and time, as shown in previous studies (Simmons et al. 2004; Bosilovich et al. 2008; Ma et al. 2009; Bao & Zhang 2015). There may be several reasons for the data uncertainty. First of all, the inhomogeneity of input data for the simulated datasets can be one of the causes (Thorne & Vose 2010; Donat et al. 2016). In other words, the further from the present, the fewer number of stations are available and it is logical to reason the increase of uncertainty for the reanalyses as well as the interpolated observation data (Ferguson & Villarini 2012; Becker et al. 2013; Zhang et al. 2013; Harris et al. 2014). It is also known that altitude gap between the modelled data and actual terrain can be one of the reasons for significant biases in a mountainous region like South Korea (Zhao & Fu 2006; Gao et al. 2012, 2014a, 2014b). Gao et al. (2014a) showed that the biases for ERA-interim temperature data were related with the elevation difference between ERA-interim grid points and gauging stations in complex terrains, but they are able to be reduced. Regional climate events such as monsoons may explain the uncertainty of the modelled data (Shah & Mishra 2014; Gao et al. 2016). Shah & Mishra (2014) described that the reanalysis products like ERA-interim showed clear bias in the monsoon season precipitation and temperature over India. The resolution of
gridded points may also affect the uncertainty (Heikkilä et al. 2011). Heikkilä et al. (2011) compared the downscaled ERA-40 with different resolutions from 30 to 10 km with observations and concluded that 10 km resolution performed the best in complex terrains.

Likewise, there are numerous reasons for the uncertainty of the datasets and it is still challenging to reliably reproduce climate features of South Korea by directly using a single modelled data. Hence, it is necessary to evaluate the agreement between the datasets and observations and improve the quality of the products in order to apply them in regional scale analysis. Nevertheless, because there has been little attention on the global dataset in South Korea, this study can suggest the potentiality of the reanalysis data and interpolated data as an alternative data source supplementing the lack of long term observations.

CONCLUSIONS

This study first evaluated key century-long climate datasets for precipitation and temperature in South Korea. From temporal and statistical comparisons, it could be concluded that GPCCv7 and CRUv3.23 for precipitation and ERA-40 for temperature perform the best among the compared datasets for the reference period 1961–2001. ERA-40, ERA-20c and 20CR for precipitation and CRUv3.23, ERA-20c and 20CR for temperature have significant agreements with the observation, but they need to be improved for application in Korea. ERA-20 cm can be used for frequency analysis over South Korea on a monthly basis after bias correction, but are not suitable for temporal variability, including the long term trend. Moreover, ERA-20 cm mean has difficulty in representing all ten ensemble members. This paper also shows that not only reanalyses but also the interpolated datasets such as CRUv3.23, which are generally accepted as the true values in the global climate change study, are able to be biased depending on the region. It means that no long term dataset can be directly applied in climate impact analysis. The findings in this paper help to fill in the knowledge gaps about the applicability of these datasets in South Korea, and provide a useful guideline to readers from other countries on the comparative performance of the global datasets in different parts of the world.

This study has mainly explored the monthly/seasonal/annual mean change on the basis of the averaged dataset over the whole regions. This analysis is very useful for understanding the general pattern of each dataset in Korea, but it does not represent the extreme climate, which is one of the vital parameters in climate impact assessment. Spatial variations with finer resolution, such as daily or 10 km scale, should be highlighted in future studies and it is essential to correct biases of the model datasets. An advantage of the reanalysis data such as ERA-20c and ERA-20 cm by ECWMF is that they supply the daily datasets with 0.125° resolution without downscaling, while the others provide coarser data. For ERA-20 cm, it may be of importance to specifically assess the features by all ensemble members, which has simply been explored by just both of mean and En0 in this context. Hence, the bias correction for the reanalysis data with the higher spatio-temporal resolutions will be explored further in future studies, as well as the features of the ERA-20 cm ensemble.

ACKNOWLEDGEMENTS

The ERA-20c, ERA-20 cm and ERA-40 data were collected via the ECMWF’s public server (http://apps.ecmwf.int/datasets/). Support for the 20CR dataset was provided by the U.S. Department of Energy, Office of Science Biological and Environmental Research (BER), and by the NOAA Climate Program Office (www.esrl.noaa.gov/psd/). The CRU and GPCC datasets were supplied from their websites, https://crudata.uea.ac.uk/cru/data/hrg/ and http://gpcc.dwd.de, respectively. The first author is grateful for the financial support from the Government of South Korea for carrying out his PhD studies at the University of Bristol.

REFERENCES

Bao, X. & Zhang, F. 2013 Evaluation of NCEP-CFSR, NCEP-NCAR, ERA-Interim, and ERA-40 reanalysis datasets against
independent sounding observations over the Tibetan Plateau. 


First received 8 February 2017; accepted in revised form 21 January 2018. Available online 27 February 2018.