INTERNATIONAL JOURNAL OF CLIMATOLOGY
Int. J. Climatol. **38**: 1216–1233 (2018)
Published online 12 September 2017 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/joc.5239

How well does the European Centre for Medium-Range Weather Forecasting Interim Reanalysis represent the surface air temperature in Cuban weather stations?

Albeht Rodríguez-Vega,[a*] Juan C. Antuña-Marrero,[a] Michel D. S. Mesquita,[b] Alan Robock,[c] Thomas Toniazzo[b] and Odd H. Otterå[b]

[a] *Atmospheric Optics Group, Meteorological Institute of Cuba, Camagüey, Cuba*
[b] *Uni Research Climate, Bjerknes Centre for Climate Research, Bergen, Norway*
[c] *Department of Environmental Sciences, Rutgers University, New Brunswick, NJ, USA*

ABSTRACT: In this research, we compare 2-m air temperature from the ERA-Interim reanalysis of the European Centre for Medium-Range Weather Forecasting with 2-m air temperature weather station observations in Cuba, with the goal of evaluating the behaviour and uncertainties of the ERA-Interim data set with respect to station-based observations. Three interpolation methods are used to determine 2-m temperatures from the ERA-Interim data set at the station locations. The differences were analysed utilizing root mean square error (RMSE), mean absolute error (MAE) and bias. The comparison was conducted for daily, monthly and annual time scales, and for the rainy (May–October) and less rainy (November–April) seasons. The best interpolation method is the mean of four grid points method. We find a warm bias in the ERA-Interim reanalysis for most Cuban stations. The smallest differences are at 1800 UTC and the largest differences are at 1200 UTC. All differences are greater than 0.3 K, although many of the stations show differences in the range of 1.5–2.0 K. In some stations the differences are greater than 5.0 K. At the daily scale more than 50% of the stations show significant differences at the 95% confidence level. The differences are caused by the altitude difference between the stations and the nearest grid point of ERA-Interim, the land-sea mask of ERA-Interim and the station location respect to this mask, and by local processes, such as a local breeze. At the monthly scale there are fewer stations with significant differences than for the other time scales. The ERA-Interim reanalysis better represents the surface 2-m temperature for coastal stations than for inland stations. Years with moderate and strong El Niño or La Niña show significant differences between ERA-Interim and observations. The amplitude between the maximum bias and the minimum bias is greater in those years.

KEY WORDS ERA-Interim reanalysis; surface air temperature; Cuba; data assimilation; interpolation methods; ERA-Interim land-sea mask

Received 26 November 2016; Revised 27 June 2017; Accepted 17 July 2017

1. Introduction

Long temperature records are necessary to understand the influence of climate on surface biophysical processes (Minder *et al.,* 2010; Mooney *et al.,* 2011). The most used surface temperature data sets for studies of climate variability and climate change and their impact on human activity and the environment are those of meteorological stations that are located near to the study area. However, geographical distance or inhomogeneities in terrain type or elevation may make temperature data from the nearest available station unrepresentative of the climate processes taking place in the area of study (Rolland, 2003).

Point data from stations are integrated with other physical information on the state of the atmosphere via numerical models of the atmosphere in reanalysis datasets, which have global coverage on a grid of representative points with a given spatial resolution (Kalnay *et al.,* 1996; Dee *et al.,* 2011; Mesquita *et al.,* 2015). In recent years, the use of reanalysis has become common practice for the study of atmospheric and ocean processes. However, atmospheric observations used in data assimilation are not perfect; they contain several kinds of errors, including instrumental errors and errors of human origin (Kistler *et al.,* 2001; Kalnay, 2003; Dee *et al.,* 2011). These atmospheric observations may contain errors of representativeness, in which some observations that are presumed correct may have associated local atmospheric phenomena. These phenomena cannot be resolved by the model because they are not representative of the average behaviour of the variable in the grid area required by the model for data analysis (Kistler *et al.,* 2001; Kalnay, 2003). Some local processes, such as deep convection and sea breeze, are not well represented in numerical simulation models. So, to reproduce these local processes it is necessary to use convective parameterization and others

* Correspondence to: A. Rodríguez-Vega, Atmospheric Optics Group, Meteorological Institute of Cuba, Ave. Finlay km 7 1/2, CP 70300, Camagüey, Cuba. E-mail: albert@cmw.insmet.cu; albertrodriguezvega@gmail.com

(i.e. radiation fluxes, microphysical processes) to take into account dynamical processes happening at smaller scales than the model resolution (Mayor and Mesquita, 2015).

Therefore, evaluation of reanalysis output with *in situ* observations is a necessary procedure in the field of climatology. This procedure allows the determination of the representativeness of the data with regard to the actual measured values. Also, it allows for the evaluation of the ability of climate data to represent the weather conditions in the station area, determining possible biases. Second, it allows the establishment of statistical relationships to complete missing data and eliminate outliers from the observations.

Previous studies reported a comparison of 2-m temperature from ERA-Interim with a European daily high-resolution gridded data set of near-surface temperature (minimum, mean and maximum), the high-resolution (8 km) SAFRAN atmospheric reanalysis, with weather station observations (Szczypta *et al.,* 2011; Gao *et al.,* 2012a; Chung *et al.,* 2013; Linsay *et al.,* 2013; Zou *et al.,* 2013). Here we conduct a similar study for Cuba, comparing 2-m temperature from 68 weather stations (T2m_Obs) with 2-m temperatures from the ERA-Interim reanalysis (T2m_ERA) at several temporal scales. In particular, the following questions will be addressed: (1) Can ERA-Interim accurately represent the 2-m temperature for the Cuban stations? and (2) Which is the best interpolation method to obtain the 2-m temperature from the Cuban stations?

The next section describes the selection process for the T2m_Obs stations used in this research and the methods for interpolation the four time-coincident T2m_ERA grid points surrounding the individual stations. In addition, it describes the statistics used for the comparison of the T2m_ERA with the T2m_Obs series. The comparison was conducted for daily, monthly and annual time scales and for the rainy (May–October) and less rainy (November–April) seasons. The following section shows the results of the comparison of the three interpolation methods.

2. Data and methods

2.1. Cuban weather stations

There are 68 Meteorological Institute of Cuba (INSMET) weather stations. All are synoptic stations and some also conduct specialized observations, such as actinometric and agrometeorological observations. Table 1 lists all the stations and how complete their records are for the four observation times for the period 1991–2012, and the station locations. Geographically, the stations have a relatively homogeneous distribution throughout the island (Figure 1), giving good coverage of the regional meteorology.

Station observations are stored in a digitized meteorological database (Báez *et al.,* 2009). The data underwent a process of quality control and standardization from the metadata available following the methodology suggested in the Guide to Climatological Practices of the World Meteorological Organization (World Meteorological Organization, 2011). No gap filling has been applied. Some outliers, not related to weather conditions, were removed (Báez *et al.,* 2009; Ramón Pérez, 2015; personal communication).

2.2. Selection of observed stations

For each of the stations from the INSMET climatological database, we created a data set of T2m_Obs daily observations at 0000, 0600, 1200 and 1800 UTC for the period 1979–2012 to match T2m_ERA. The WMO Technical Regulations state that climatological normals are calculated for consecutive periods of 30 years (1 January 1901 to 31 December 1930, 1 January 1931 to 31 December 1960, and so on), but also recommended the calculation of an average (also known as a provisional normal) at any time for stations that do not have 30 years of data (Angel *et al.,* 1993; Arguez and Vose, 2011). These averages are calculated for a period of ten or more years, from 1 January of the first year in a decade (e.g. from 1 January 1991 to 31 December 2004). The optimal time for normal temperatures may be less than 30 years (Arguez and Vose, 2011; World Meteorological Organization, 2011), depending on the type of investigation that is to be performed. When it is used for comparison between the data sets of the same variable, but from different sources, 30 years of data are not needed (Angel *et al.,* 1993; Arguez and Vose, 2011; World Meteorological Organization, 2011). This is the case for the comparison between T2m_ERA and T2m_Obs.

According to the Guide to Climatological Practices of the World Meteorological Organization (World Meteorological Organization, 2011) the normals or period average should be calculated only when values are available for at least 80% of the years of record, with no more than three consecutive missing years. Also in cases when there is an extended period of missing data but reasonably complete data after that time, it is recommended to calculate a period average using only data from the years following the break in the record. We applied these principles and some stations were detected with data gaps not fulfilling the former conditions for the period 1979–1990. As a result, stations with more than 20 years of data and less than 15% missing data in the period 1991–2012 at the four observation times were selected. Figure 2 shows the temporal distribution of T2m_Obs from the 68 Cuban stations for the period 1979–2012. The 0600 UTC time shows many data gaps between 1979 and 1991, because many stations had an observation regimen from 1200 to 0000 UTC (0700 to 1900 local time), and 0600 UTC corresponds to 0100 local time.

As a result of the selection process, we use T2m_Obs from 61 stations in this research, which represents 90% of all available stations in the country. The period chosen for the study was 1991–2012, because of the large amount of missing data between 1979 and 1990 (Figure 2). In addition, the period chosen meets WMO requirements for climatological normals. Figure 1 shows the geographical distribution of the 61 selected stations, with over 20 years

Table 1. List of the Cuban weather stations. The last five columns show the percent of completeness of the data sets for the period 1991–2012 at four observation times and the station location. The rows marked gray are the stations rejected because of data coverage.

WMO Code	Station	Latitude (°N)	Longitude (°W)	0000 UTC	0600 UTC	1200 UTC	1800 UTC	Location
78308	La Piedra	22.11	79.98	31	31	31	31	Land
78309	Amistad Cuba-Francia	21.84	82.85	92	90	99	99	Land
78310	Cabo de San Antonio	21.87	84.95	94	94	94	94	Coastal
78312	Santa Lucia	22.66	83.96	99	99	99	99	Coastal
78313	Isabel Rubio	22.15	84.11	100	100	100	100	Land
78314	San Juan y Martínez	22.28	83.83	100	100	100	100	Land
78315	Pinar del Río	22.40	83.65	99	99	99	99	Land
78316	La Palma	22.77	83.56	100	99	100	100	Land
78317	Paso Real de San Diego	22.56	83.31	86	86	86	86	Land
78318	Bahía Honda	22.93	83.22	99	99	99	99	Land
78319	Valle de Caujerí	20.15	74.84	89	83	89	89	Land
78320	Güira de Melena	22.78	82.51	95	95	95	95	Land
78321	La Fé	21.73	82.76	100	100	100	100	Land
78322	Batabanó	22.74	82.29	99	98	100	100	Coastal
78323	Güines	22.85	82.04	90	90	90	90	Land
78324	Punta del Este	21.56	82.56	100	98	100	100	Coastal
78325	Casablanca	23.14	82.34	100	100	100	100	Coastal
78326	Santo Domingo	22.59	80.23	100	100	100	100	Land
78327	Unión de Reyes	22.77	81.54	100	100	100	100	Land
78328	Varadero	23.16	81.23	72	70	72	72	Coastal
78329	Indio Hatüey	22.82	81.02	98	96	98	97	Land
78330	Jovellanos	22.80	81.14	100	100	100	99	Land
78331	Jagüey Grande	22.53	81.14	79	65	79	79	Land
78332	Colón	22.68	80.93	99	99	99	99	Land
78333	Playa Girón	22.07	81.03	99	99	99	99	Coastal
78334	Palenque de Yateras	20.37	74.96	88	88	88	88	Land
78335	Aguada de Pasajeros	22.13	80.83	94	94	94	94	Land
78337	Trinidad	21.78	79.99	99	99	99	99	Coastal
78338	Sagua la Grande	22.81	80.09	100	100	100	100	Land
78339	Cayo Coco	22.54	78.37	95	95	95	95	Coastal
78340	Bainoa	23.00	81.94	91	91	91	91	Land
78341	El Jíbaro	21.74	79.23	99	99	99	99	Land
78342	Topes de Collantes	21.92	80.02	100	100	100	100	Land
78343	El Yabú	22.46	79.99	100	100	100	100	Land
78344	Cienfuegos	22.19	80.44	94	94	94	94	Coastal
78345	Júcaro	21.52	78.33	99	76	100	100	Land
78346	Venezuela	21.76	78.80	99	100	99	99	Land
78347	Camilo Cienfuegos	22.16	78.77	100	100	100	100	Land
78348	Caibarién	22.50	79.47	100	100	100	100	Coastal
78349	Sancti Spiritus	21.97	79.45	99	99	99	99	Land
78350	Florida	21.52	78.25	100	100	100	100	Land
78351	Santa Cruz del Sur	20.74	78.00	99	99	99	99	Coastal
78352	Esmeralda	21.84	78.12	99	99	99	99	Land
78353	Nuevitas	21.56	77.25	100	100	100	100	Coastal
78354	Palo Seco	21.14	77.32	100	100	100	100	Land
78355	Camagüey	21.42	77.85	100	100	100	100	Land
78356	Jamal	20.30	74.45	87	80	87	87	Coastal
78357	Las Tunas	20.95	76.94	100	100	100	100	Land
78358	Puerto Padre	21.21	76.61	100	100	100	100	Coastal
78359	Manzanillo	20.31	77.16	100	100	100	100	Coastal
78360	Cabo Cruz	19.84	77.72	100	100	100	100	Coastal
78361	Jucarito	20.69	76.90	98	97	100	100	Land
78362	La Jíquima	20.93	76.54	99	99	99	99	Land
78363	Contramaestre	20.28	76.27	100	100	100	100	Land
78364	Santiago de Cuba	20.04	75.82	99	99	99	99	Land
78365	Cabo Lucrecia	21.07	75.62	100	100	100	100	Coastal
78366	La Gran Piedra	20.01	75.63	86	86	86	86	Land
78368	Guantánamo	20.13	75.23	95	95	95	95	Land
78369	Punta de Maisí	20.24	74.15	100	100	100	100	Coastal
78370	Guaro	20.67	75.78	100	99	100	100	Coastal
78371	Pinares de Mayarí	20.49	75.79	99	99	99	99	Land
78372	Pedagógico	20.88	76.22	91	91	91	91	Land
78373	Santiago de las Vegas	22.98	82.39	83	83	83	83	Land
78374	Tapaste	23.00	82.14	95	95	95	95	Land
78375	Melena del Sur	22.78	82.11	94	93	94	94	Land
78376	Bauta	22.98	82.54	90	89	90	90	Land
78377	Veguitas	20.33	76.89	100	98	100	100	Land
78378	Velasco	21.06	76.35	98	97	98	98	Land

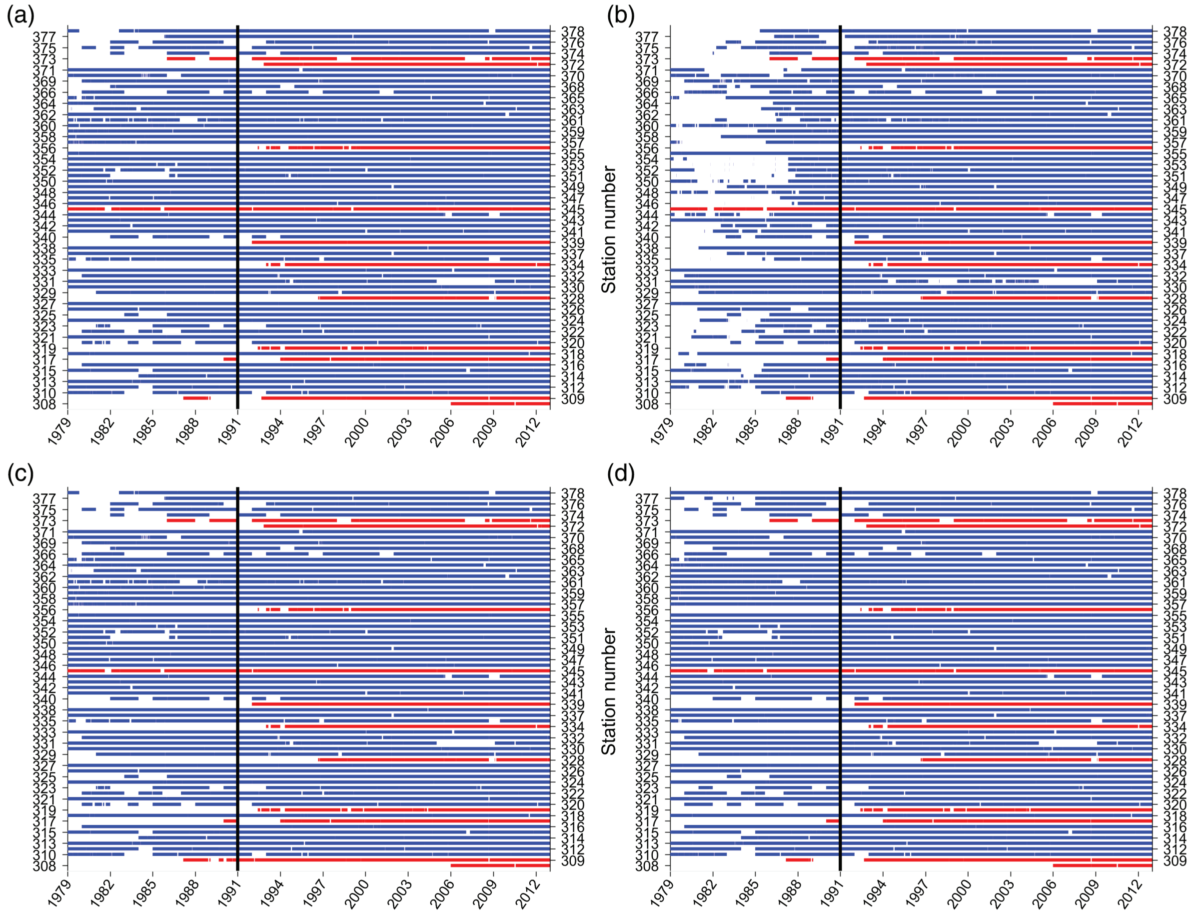

Figure 1. Temporal distribution of T2m_Obs from the 68 Cuban stations (see Table 1) in the period 1979–2012. The blue lines represent the stations with less than 15% of missing data in the period 1991–2012 and the red lines represent the stations with more than 15% of missing data. The stations are listed continuously. (a) 0000 UTC, (b) 0600 UTC, (c) 1200 UTC and (d) 1800 UTC.

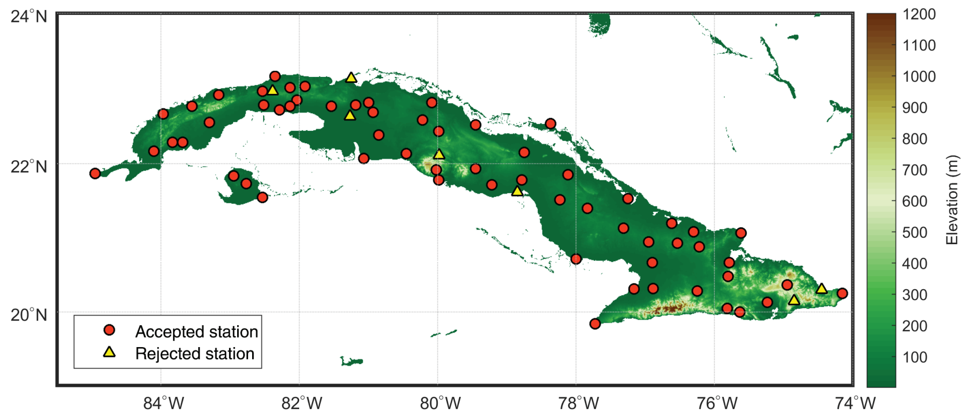

Figure 2. Geographical distribution of the available Cuban weather stations (See Table 1). Red circles are the stations used in the present study.

T2m_Obs, identified by filled circles in red. There are seven stations rejected from the selection process, identified by filled triangle in yellow, and all of them are shown in Table 1. The accepted subset of stations maintains the good spatial coverage of the parent set.

2.3. ERA-Interim reanalysis

ERA-Interim is the newest-generation reanalysis product from the European Centre for Medium-Range Weather Forecasts (ECMWF). It covers the period from 1 January 1979 onwards and continues to be extended forward in near-real time. Gridded data products include a large variety of three-hourly surface parameters and six-hourly upper-air parameters covering the troposphere and stratosphere (Dee *et al.,* 2011). The ERA-Interim project was launched to improve key aspects of ERA-40, such as the representation of the hydrological cycle and the quality of the stratospheric circulation, as well as the handling

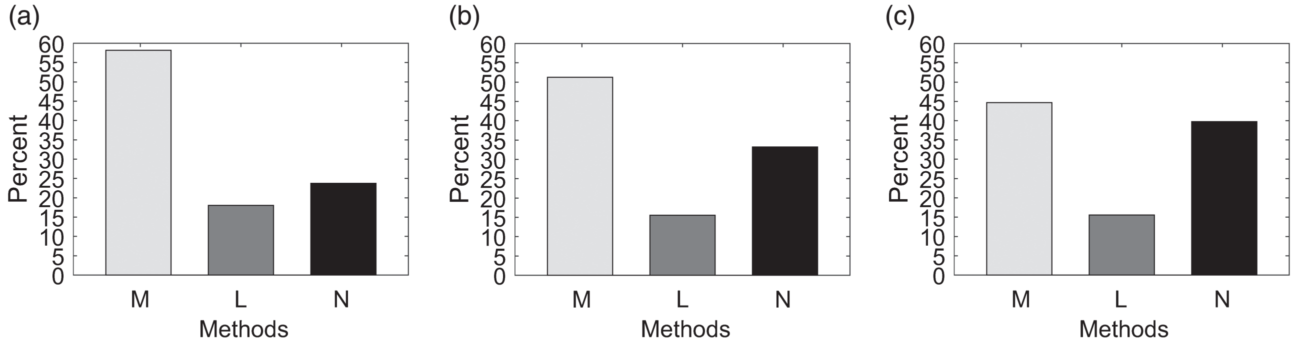

Figure 3. Percentage of cases where the RMSE shows the lowest value among all analysed methods. Individual method RMSE values were calculated from the daily T2m_Obs and T2m_ERA for all the stations at all the observation times. M is the mean of four grid points method, L is the bilinear interpolation method and N is the nearest neighbours method. (a) 0.5° resolution, (b) 0.25° resolution and (c) 0.125° resolution.

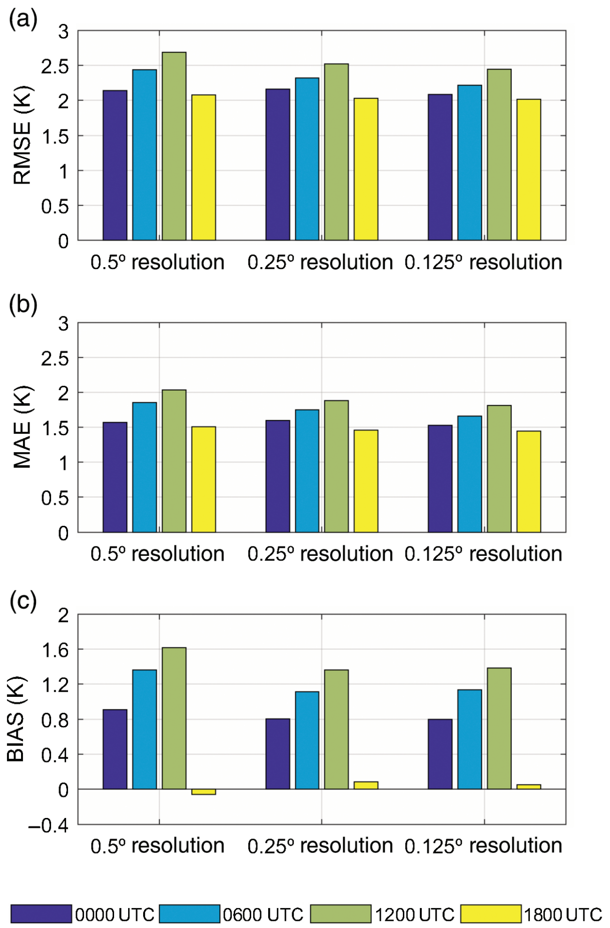

Figure 4. RMSE, MAE and BIAS values for all the stations for the three resolutions at the four observation times.

of biases and changes in the observing system (Simmons *et al.,* 2006; Uppala *et al.,* 2008; Dee and Uppala, 2009; Dee *et al.,* 2011). ECMWF provides a variety of data in uniform latitude/longitude grids (0.25°, 0.5°, 0.75°, 1°, 1.125°, 1.5°, 2°, 2.5° and 3°). The parameters (except vegetation, soil type fields and wave 2-D spectra) are interpolated from the original N128 reduced Gaussian grid using bilinear methods (Gao *et al.,* 2012a, 2012b).

According to Kalnay (2003), the low heat capacity of the surface layer causes surface air temperatures (at 2 m) to adjust very rapidly towards equilibrium with the sea or land surface temperatures. Consequently, it is difficult to use observations of surface air temperature effectively in the reanalysis, since the model tends to forget these observations (replacing them with values adjusted by the model). As a result, at ECMWF, T2m observations are only 'active' in the surface analysis and not in the atmospheric analysis and are not used as input data (D. Tan, 2015; personal communication). The 2 m temperature in the ERA-Interim model is based on a parameterization of the surface turbulent layer that depends on quantities at the lowest model level (about 10 m above the surface) and at the surface (ECMWF, 2009; Gao, 2013).

For selecting the T2m ERA-Interim (T2m_ERA) values, a work domain around Cuba was defined, between latitudes 17° and 24°N and longitudes 70° and 90°W. The series consisted of six-hourly T2m_ERA analysis data (0000, 0600, 1200 and 1800 UTC), from 1979 to 2012, for three spatial resolutions in latitude and longitude (0.5°, 0.25° and 0.125°).

2.4. Interpolation methods for T2m_ERA

The four T2m_ERA values from the four grid points enclosing each station were selected to produce series of T2m_ERA matching in space and time those of T2m_Obs. This was done by one of three different methods: average, bilinear interpolation and nearest neighbour.

2.4.1. Processing the T2m time series

Some authors describe the climate of the Caribbean as dry-winter tropical (Rudloff, 1981; Magana *et al.,* 1999; Giannini *et al.,* 2000; Curtis, 2002; Mapes *et al.,* 2005), with two seasonal rainfall pattern (dry and wet seasons). However, in Cuba the cumulative precipitation from November to April is greater than 350 mm and the Cuban meteorologists defined this period as the less rainy season in contrast to the period from May to October, the rainy season (Lecha *et al.,* 1994). Monthly and annual means for each observation time were calculated, as were means for the rainy and less rainy seasons, for T2m_Obs and T2m_ERA using the World Meteorological Organization (2011) methodology. First the monthly means were calculated, and from monthly means the annual means and

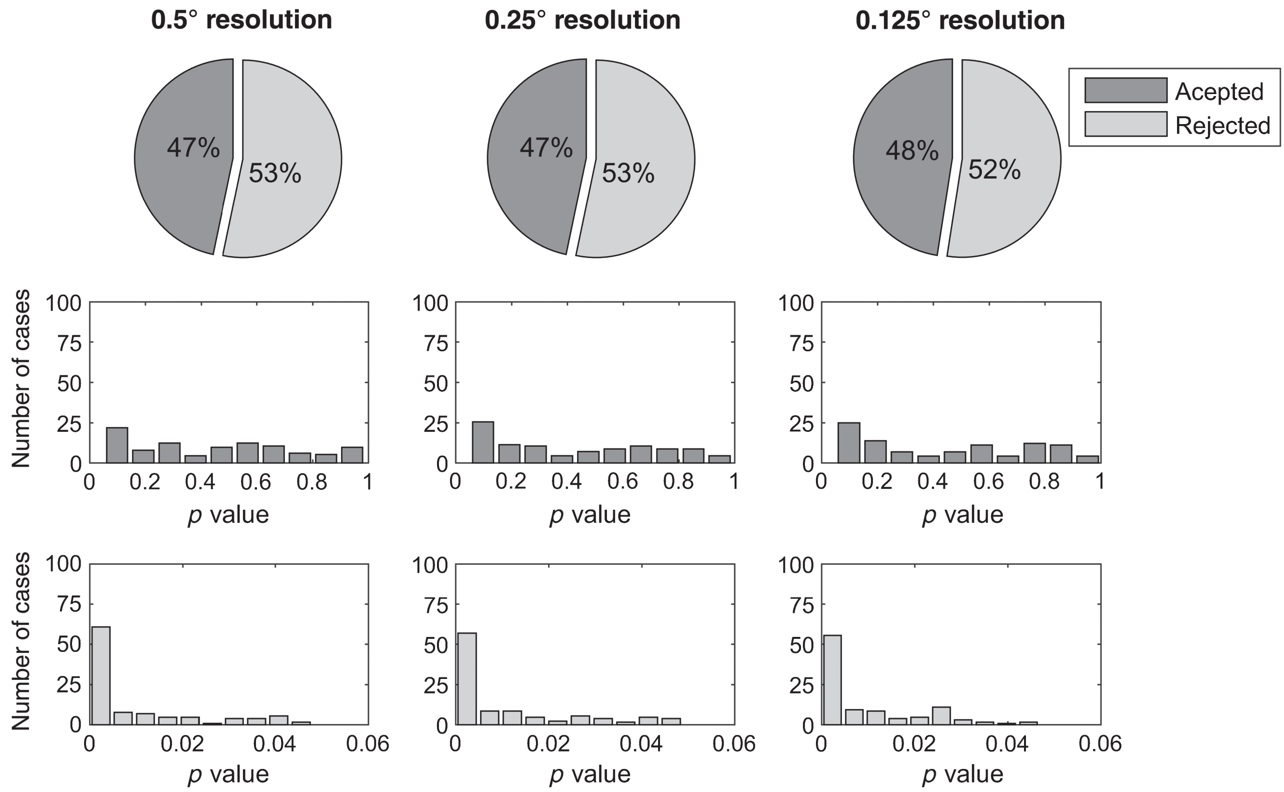

Figure 5. Graphical representation of the statistical significance for the differences between T2m_ERA and T2m_Observations for all times together. The middle panel shows p values for the accepted cases and the bottom panel shows p -values for the rejected cases.

Table 2. Hypothesis test results for accepted and rejected stations. Maximum and minimum values for the t statistic and maximum, minimum and mean for p values are listed.

Resolution	t		p		
	Max	Min	Mean	Max	Min
Accepted stations					
0.5°	1.9744	−1.9020	0.4545	0.9888	0.0512
0.25°	1.9858	−1.9786	0.4246	0.9931	0.0509
0.125°	1.9994	−1.9337	0.4408	0.9996	0.0505
Rejected stations					
0.5°	26.9556	−10.2858	0.0092	0.0479	0.0000
0.25°	27.3416	−9.8175	0.0102	0.0487	0.0000
0.125°	27.9989	−10.2620	0.0090	0.0468	0.0000

the means for the rainy and less rainy seasons were calculated. The standard deviations were also calculated for the T2m_Obs and T2m_ERA for daily, monthly and annual time scales, and for the rainy and less rainy seasons. The coefficient of variation for both T2m observations and reanalysis was calculated to determine the possible variability of T2m within the island associated with local phenomena.

2.4.2. *Selection of the interpolation method for the comparison*

Evaluation of the best interpolation method to obtain the T2m from ERA-Interim was conducted using the RMSE. Only daily data sets were used to select the best method. For the comparison of the T2m_ERA and T2m_Obs data sets were used the monthly and annual time scales, as well as the rainy and less rainy season.

2.4.3. *Statistical evaluation methods*

To assess the degree of coincidence of T2m_ERA and T2m_Obs, we used the MAE, RMSE and the bias (BIAS). MAE is an absolute measure of the deviation of the predicted (i.e. ERA-Interim) value from the observed mean at each validation station, ignoring its sign and thereby providing an indicator of the overall performance of the interpolator (Jolliffe and Stephenson, 2003; Serbin and Kucharik, 2008; Gao *et al.,* 2012a). The RMSE jointly evaluates both the magnitude of the bias (indicated by the mean error), and the dispersion (indicated by the standard deviation of the errors; Corbelle *et al.,* 2006). The BIAS meanwhile provides information on the tendency of the

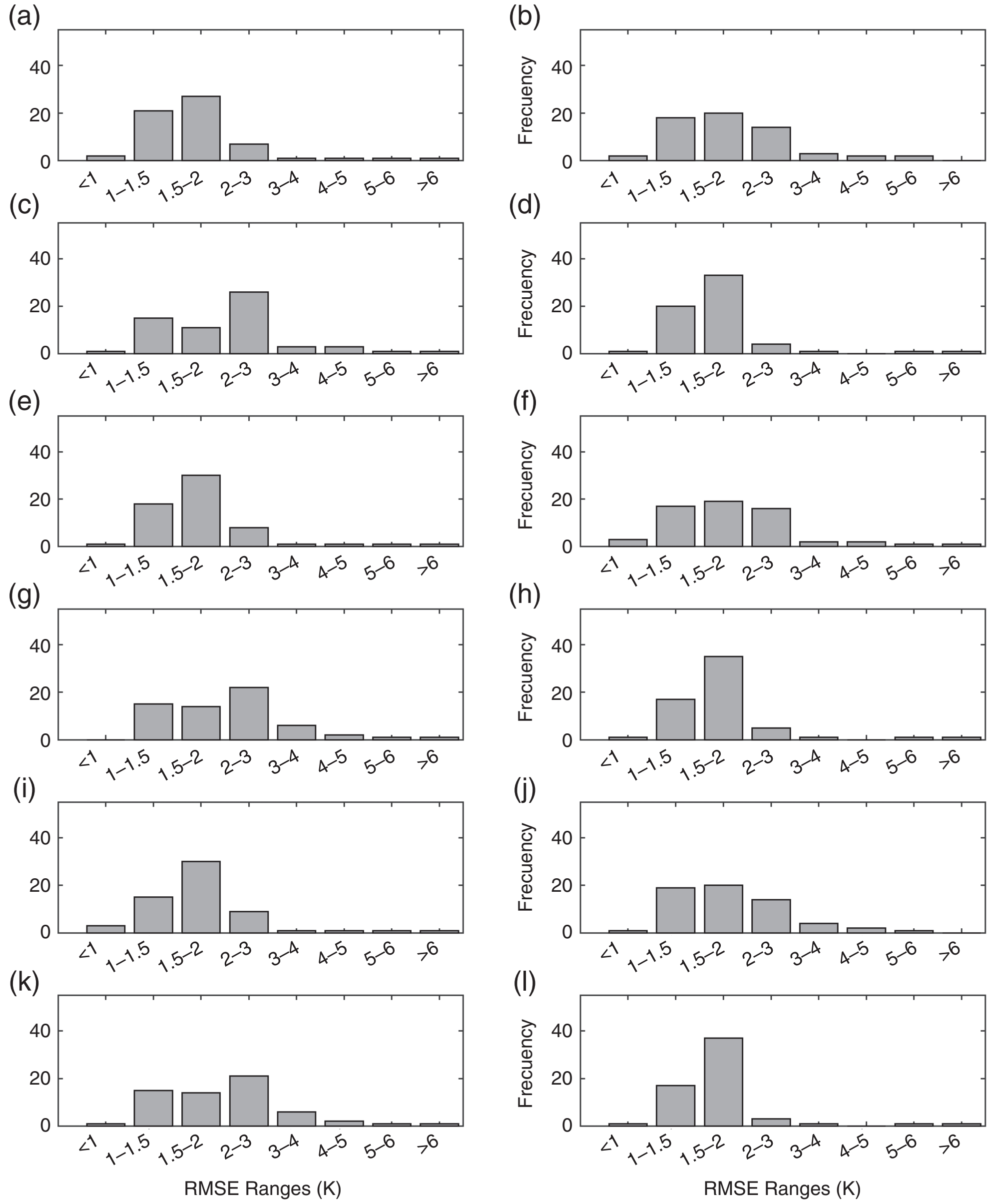

Figure 6. Frequency distribution (units of %) for the root mean square error (T2m_ERA *vs* T2m_Observations). The bars represent the distribution for all stations at each observation time and for three ERA-Interim resolutions. (a) 0000 UTC – 0.5° resolution, (a) 0600 UTC – 0.5° resolution, (c) 1200 UTC – 0.5° resolution, (d) 1800 UTC – 0.5° resolution, (e) 0000 UTC – 0.25° resolution, (f) 0600 UTC – 0.25° resolution, (g) 1200 UTC – 0.25° resolution, (h) 1800 UTC – 0.25° resolution, (i) 0000 UTC – 0.125° resolution, (j) 0600 UTC – 0.125° resolution, (k) 1200 UTC – 0.125° resolution, (l) 1800 UTC – 0.125° resolution.

model to overestimate or underestimate the variable predicted as compared to the observed. This statistic quantifies the systematic error of the model used (Jolliffe and Stephenson, 2003; Szczypta *et al.,* 2011). In the comparison, we only used the best interpolation method determined by the RMSE for the daily time scale, calculated for all months together. The analysis was developed for all stations and for coastal and land stations separately.

2.4.4. *Significance tests for mean comparison*

To evaluate the differences between monthly and annual T2m_ERA and T2m_Obs, as well as the rainy and less rainy season values, we used the Student's *t*-test, with null hypothesis H_0 ($\mu_1 = \mu_2$) and alternative hypothesis H_1 is ($\mu_1 \neq \mu_2$) and a significance level $\alpha = 0.05$, μ_1 and μ_2 are means of populations of T2m_ERA and T2m_Obs. This test is two-tailed, since we have no expectation a priori regarding the direction of the mean difference. The critical value of the test is obtained from the Student's *t*-table for $(n-1)$ degrees of freedom and a significance level (α).

Because T2m_ERA and T2m_Obs series are highly autocorrelated, this persistence reduces the number of statistically independent data points (Renard *et al.,* 2008; Santer *et al.,* 2008; Brands *et al.,* 2011; Daniel *et al.,* 2012), and thus the number of independent data points (the effective sample size n_e) is much lower than the sample size n_t. To account for this, we calculated the effective sample size using the expression from Santer

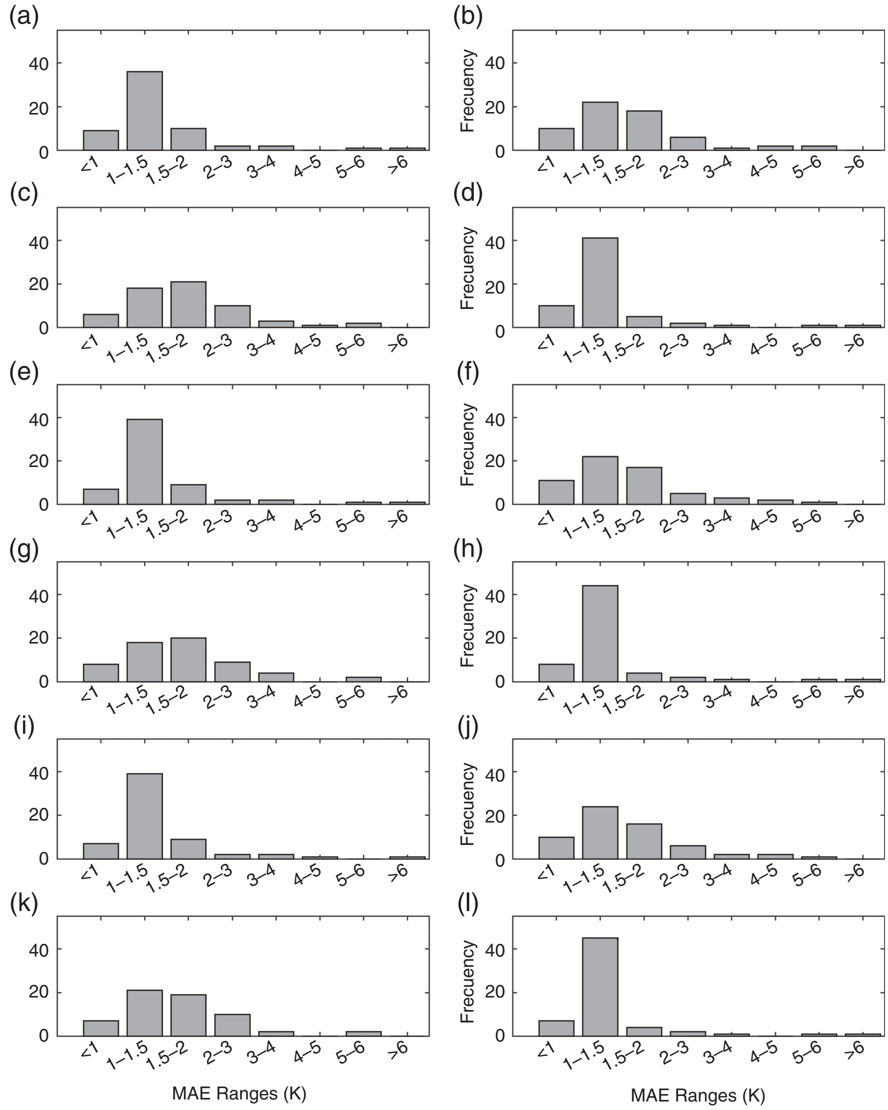

Figure 7. Frequency distribution (units of %) for the mean absolute error (T2m_ERA *vs* T2m_Observations). The bars represent the distribution for all stations at each observation time and for three ERA-Interim selected resolutions. (a) 0000 UTC – 0.5° resolution, (a) 0600 UTC – 0.5° resolution, (c)1200 UTC – 0.5° resolution, (d) 1800 UTC – 0.5° resolution, (e) 0000 UTC – 0.25° resolution, (f) 0600 UTC – 0.25° resolution, (g) 1200 UTC – 0.25° resolution, (h) 1800 UTC – 0.25° resolution, (i) 0000 UTC – 0.125° resolution, (j) 0600 UTC – 0.125° resolution, (k) 1200 UTC – 0.125° resolution, (l) 1800 UTC – 0.125° resolution.

et al. (2000, 2008) based on r_1, the lag −1 autocorrelation coefficient,

$$n_e = n_t \frac{1 - r_1}{1 + r_1}$$

In the hypothesis test n_t is replaced by n_e and the degrees of freedom are calculated as

$$df = \frac{\left(\frac{S_1^2}{n_{e1}} + \frac{S_2^2}{n_{e2}} \right)}{\frac{\left(\frac{S_1^2}{n_{e1}} \right)^2}{n_{e1}-1} + \frac{\left(\frac{S_2^2}{n_{e1}} \right)^2}{n_{e2}-1}}$$

where n_{e1} and n_{e2} are the effective sample size for T2m_ERA and T2m_Obs series, $S_1$and $S_2$are the standard deviations of the two samples. The significance of all tests reported is evaluated using these quantities.

2.4.5. *Other elements from ERA-Interim and weather stations*

We analysed the influence of the land–sea mask and topography using the land–sea mask field from ERA Interim (Zagar *et al.,* 2011). Coastal and land stations were defined by comparison of its positions with respect to the borders of the mask. The height differences between stations and ERA-Interim grid points were used to classify the stations and describe the effects of the height difference on the calculated errors.

3. Results and discussion

3.1. Selection of the interpolation method

Figure 3 shows the RMSE values among all the three analysed methods for the three resolutions used in the

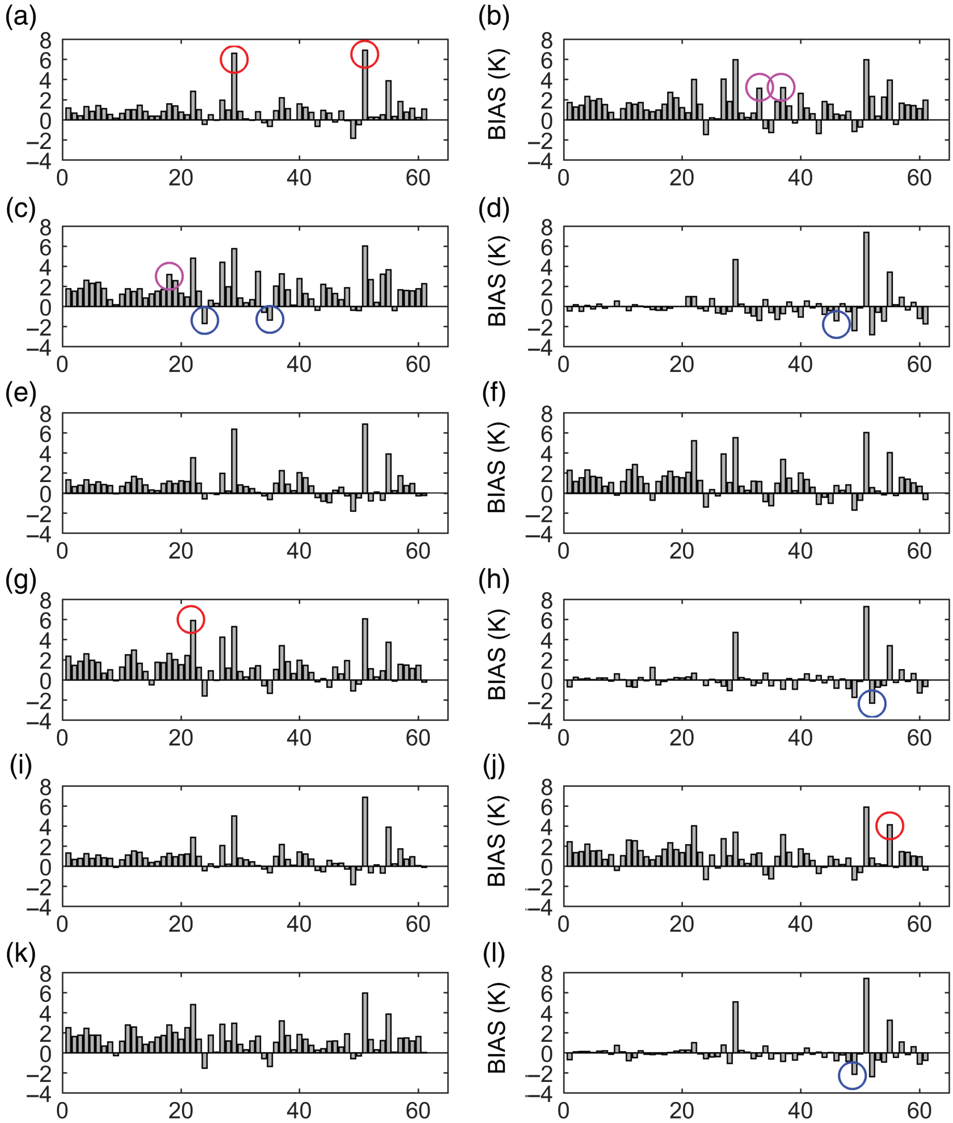

Figure 8. BIAS values (T2m_ERA *vs* T2m_Observations) for individual stations at the four observation times for the three ERA-Interim resolutions. Red circles show the four stations with positive BIAS >5.0 K (78334, 78342, 78366 and 78371). Magenta circles show the three stations with positive BIAS >3.0 K (78329, 78340 and 78351) and blue circles show the five stations with differences >1.0 K, but with negative BIAS (78337, 78349, 78361, 78364 and 78368). (a) 0000 UTC – 0.5° resolution, (a) 0600 UTC – 0.5° resolution, (c)1200 UTC – 0.5° resolution, (d) 1800 UTC – 0.5° resolution, (e) 0000 UTC – 0.25° resolution, (f) 0600 UTC – 0.25° resolution, (g) 1200 UTC – 0.25° resolution, (h) 1800 UTC – 0.25° resolution, (i) 0000 UTC – 0.125° resolution, (j) 0600 UTC – 0.125° resolution, (k) 1200 UTC – 0.125° resolution, (l) 1800 UTC – 0.125° resolution.

comparison, calculated from the daily T2m_Obs and T2m_ERA for all the stations at all the observation times. At the 0.5° resolution, the mean of four grid points method showed the most significant results, with more than 58% of cases in which the RMSE showed the lowest values among all methods. The nearest neighbour method had about 26% and the bilinear method about 18% (Figure 3(a)). There was similar behaviour at the 0.25° and 0.125° resolutions. The mean of four grid points method had 52 and 45% for 0.25° and 0.125° resolutions. However, the nearest neighbour method showed an increase for these two resolutions, with respect to the 0.5° resolution, but the values obtained from this method are about 33 and 40% for 0.25° and 0.125° resolutions. Because the mean of four grid points method had the best overall behaviour at the three resolutions, we used it for all the following statistical analysis.

3.2. General statistical analysis

We first compared the difference between T2m_ERA and T2m_Obs using the RMSE, MAE and BIAS statistics for all stations for each *observation time* at the three resolutions (Figure 4). Diurnally, the lowest values for each of these quantities were at 1800 UTC and the highest values at 1200 UTC, similar for all three resolutions. The RMSE values are greater than 1.9 K for all *observation times*, and at 1200 UTC the differences are greater than 2.4 K for all resolutions. For MAE, the differences are greater than 1.3 K for all *observation times* for the three resolutions and greater than 1.8 K for 0.5° at 1200 UTC. In general, the RMSE and MAE values decrease as the spatial resolution is increased. Like the other statistics, the BIAS shows the lowest values at 1800 UTC and the

Table 3. Number of cases of mean absolute error (MAE) and root mean square error (RMSE) for individual stations at the four observation times for the three ERA-Interim selected resolutions in the ranges greater than 3.0 K and greater than 5.0 K. Stations grouped in three categories: all stations, coastal stations and land stations. The comparison was made with monthly data sets.

		MAE				RMSE			
					All stations				
Difference (K)	Resolution	00	06	12	18	00	06	12	18
> 3.0	0.5°	4	5	6	3	4	7	8	3
	0.25°	4	6	6	3	4	6	10	3
	0.125°	4	5	4	3	4	7	10	3
> 5.0	0.5°	2	2	2	2	2	2	2	2
	0.25°	2	1	2	2	2	2	2	2
	0.125°	1	1	2	2	2	1	2	2
					Coastal stations				
> 3.0	0.5°	0	0	0	0	0	1	1	0
	0.25°	0	1	1	0	0	1	1	0
	0.125°	0	0	0	0	0	1	1	0
> 5.0	0.5°	0	0	0	0	0	0	0	0
	0.25°	0	0	0	0	0	0	0	0
	0.125°	0	0	0	0	0	0	0	0
					Land stations				
> 3.0	0.5°	4	5	6	3	4	6	7	3
	0.25°	4	5	5	3	4	5	9	3
	0.125°	4	5	4	3	4	6	9	3
> 5.0	0.5°	2	2	2	2	2	2	2	2
	0.25°	2	1	2	2	2	2	2	2
	0.125°	1	1	2	2	2	1	2	2

greatest values at 1200 UTC. At 0000, 0600 and 1200 UTC the differences range between +0.8 and +1.5 K. However at 1800 UTC the differences are less than 0.2 K.

Figure 5 shows the statistical significance of the differences between daily T2m_ERA and T2m_Obs for each individual station for all the four observation times together. For all resolutions more than 50% of stations had significant differences at the 95% confidence level. At 0.5° and 0.25° resolution 53% of the stations have a significant differences, with 52% at 0.125° resolutions.

The stations where the null hypothesis was accepted show *p*-values greater than 0.05. The *t* statistic shows values from −1.9786 to 1.9994. All *t* values are inside the acceptance zone (Table 2). The *p*-values for the stations where the null hypothesis was rejected show range from 0 to 0.05, with most of the values in the range from 0 to 0.0010 (Figure 6). High values of *t*-statistics are observed in the stations where the null hypothesis was rejected.

3.3. Individual station statistics

Joint analysis of the three statistics for each individual station at each observation time for the three ERA-Interim resolutions produced respective frequency distributions. For RMSE we produced eight classes or ranges, from values <1.0 to >6.0 K (Figure 6). For the 0.5° and 0.25° resolutions, the differences between T2m_ERA and T2m_Obs show a predominance of RMSE values in the range from 1.0 to 2.0 K at 0000 and 1800 UTC. The 0000 and 1800 UTC times show a similar behaviour, although at 1800 UTC, the number of RMSE values in the range from 1.5 to 2.0 K is greater than for 0000 UTC. At 0600 and 1200 UTC, there is an increase of RMSE in the range from 2.0 to 3.0 K, with respect to 0000 and 1800 UTC. This increase was greatest at 1200 UTC, where it includes around 50% of all RMSE values.

The distribution of MAE for the same eight ranges as RMSE (Figure 7) shows a predominance of MAE values in the range from 1.0 to 1.5 K with more than 70% at 0000 and 1800 UTC and 40% at 0600 UTC. In general, most MAE values are below 2.0 K, and only at 1200 UTC is there a slight increase in the range of 1.5–3.0 K, although not as significant as for RMSE. As with RMSE, there is little dependence on resolution. For both RMSE and MAE values greater than 3.0 K are observed at the same stations.

Figure 8 shows the BIAS values for individual stations at the four observation times for the three ERA-Interim resolutions. At 0600 and 1200 UTC, there is a positive BIAS for most stations for the three ERA-Interim reanalysis resolutions. In four stations the positive BIAS is >5 K. In general, the ERA-Interim reanalysis overestimates the T2m_Obs values at these two observation times. However, some stations show a negative BIAS.

In Figure 8, we indicated with circles stations with a positive BIAS greater than 3 and 5 K and stations with negative BIAS less than −1 K. Red circles are used to indicate the four stations with a positive BIAS and values greater than 5 K, magenta circles to indicate the three stations with a positive BIAS and values greater than 3 K, and blue circles to indicate the five stations with a negative BIAS and values less than −1 K.

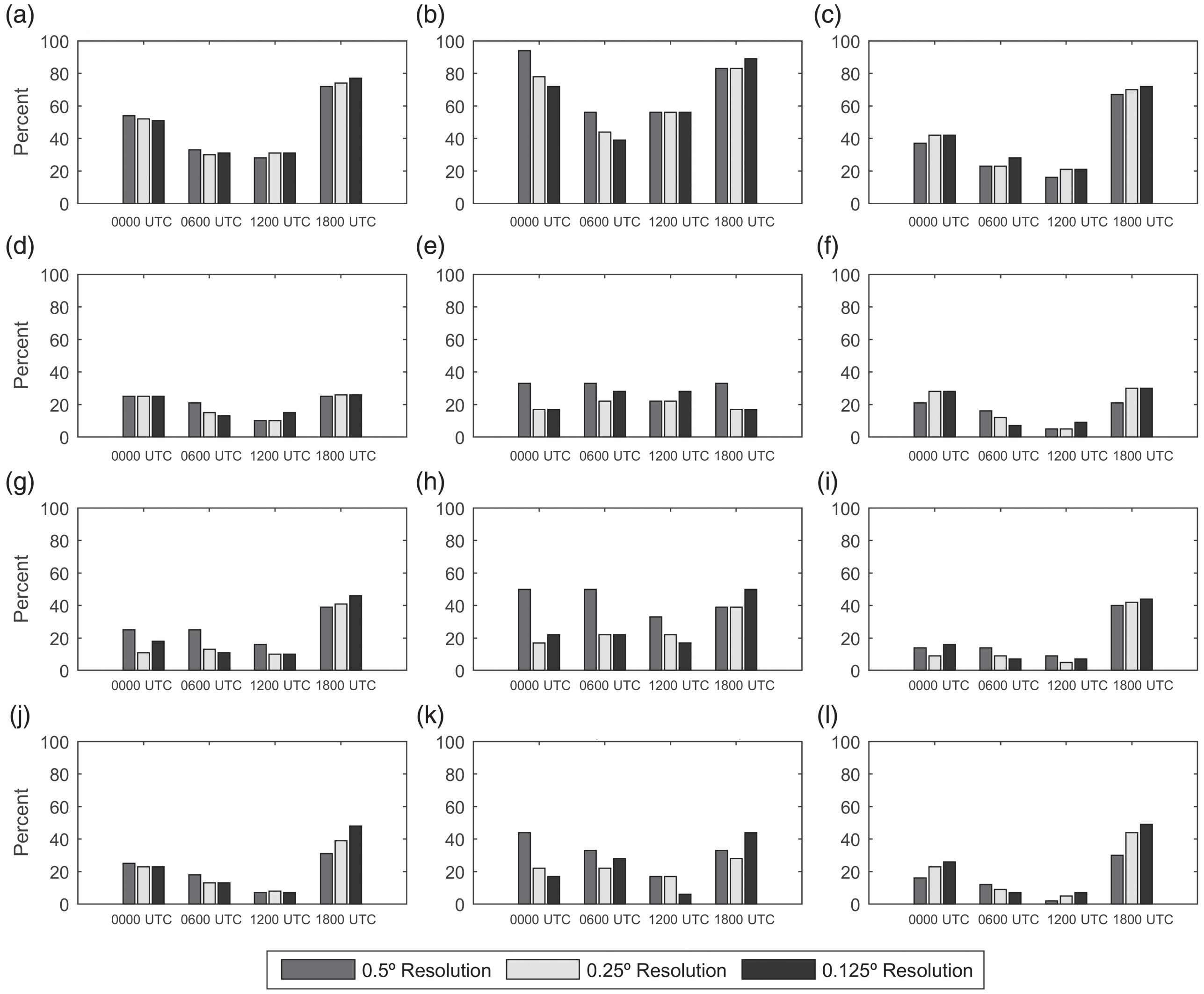

Figure 9. Percent of stations where T2m_Obs and T2m_ERA do not differ, using a significance level of 5% for different time scales. The three ERA-Interim spatial resolutions are depicted with different shadings. (a) Monthly (all stations), (b) monthly (coastal stations), (c) Monthly (land stations), (d) less raining season (all stations), (e) less raining season (coastal stations), (f) less raining season (land stations), (g) raining season (all stations), (h) raining season (coastal stations), (i) raining season (land stations), (j) annual (all stations), (k) annual (coastal stations), (l) annual (land stations).

At 0000 UTC, there is also a positive BIAS at most weather stations, but with lower values than at 0600 and 1200 UTC. However, for the stations with the higher BIAS values, they are slightly higher than at 0600 and 1200 UTC. The 1800 UTC has the best results, and most of the stations have a negative BIAS, but only in five of them is it greater than 1.0 K. At 1800 UTC the only significant positive biases are observed at the same stations with significant positive biases at the other observation times.

Table 3 lists number of cases of MAE and RMSE values for individual stations at the four observation times for the three ERA-Interim resolutions in the ranges >3.0 and >5.0 K. Most of these are land stations, pointing out that ERA-Interim reanalysis represents T2m_Obs better for coastal stations. There are no coastal stations with RMSE and MAE >5.0 K, and MAE and RMSE values >3.0 K at 0600 and 1200 UTC are found at only three coastal stations. At land stations MAE and RMSE values greater than 3.0 and 5.0 K are found at all observation times, but at 0600 and 1200 UTC there are more.

The statistical significance of the differences between T2m_ERA and T2m_Obs is shown in Figure 9 for monthly and annual time scales, and for the rainy and less rainy seasons. The most significant results are at the monthly scale, with a significance level of 5%. For the remaining three time scales, the null hypothesis is rejected for more than 60% of the stations. The best agreement is at 1800 UTC for all the three resolutions and all time scales. The results in Figure 9 agree with the RMSE and MAE values for coastal and land stations in Table 3.

At the monthly scale, for all stations together, a good agreement is observed for more than 50% of the stations for the 0000 and 1800 UTC times. At 1800 UTC there is a reasonable agreement for more than 70% of the stations. At 0600 and 1200 UTC, only about 25% of the stations show a good agreement, but even for these two observation times, the monthly scale shows a better agreement than the rest of the temporal scales. Analysing the coastal and land stations separately, the monthly scale also shows the best behaviour. In coastal stations, more than 80% of the stations have a good agreement at 0000 and 1800 UTC, and more than 50% at 0600 and 1200 UTC. For land stations a good agreement is observed at more than 60% of the stations at 1800 UTC. For the rest of the observation times, fewer than 40% show a good agreement.

These results show that the ERA-Interim and stations have a good relationship for the monthly scale, but not for the annual scale or for the rainy and less rainy seasons. This is due to the temporal and spatial inhomogeneity of the input data of the assimilation models that cause a poor estimation of the fields of variables in the long term and the best estimation at small time scales. According to Thorne and Vose (2010), the reanalyses were never primarily constructed to be long-term homogeneous (free of non-climatic influences) records but rather to provide the best possible analysis at each time step.

3.4. Reasons for agreement and differences between ERA-Interim and observations

The systematic discrepancies between ERA-Interim and station observations can have different, interdependent causes. These include the height difference between the stations and the nearest grid point of ERA-Interim (Figure 10), the geographical distance between grid points of ERA-Interim and the stations, and inconsistent surface properties, including inconsistencies between the land–sea mask of ERA-Interim and the station location (Figure 11).

In Figure 10, two zones show very large height differences >150 m. In these zones, stations have significant temperature differences and the values of t are significantly higher. The stations with large differences in RMSE and MAE (>5.0 K) and positive BIAS values are located in a mountainous zone with altitude greater than 440 m. The ERA-Interim grid points around these stations have an altitude lower than the stations. In these cases, the height differences between nearest grid point and station location are greater than 100 m, and some stations have height differences greater than 500 m (Figure 10). The t statistic at these stations has values greater than 7 and some stations show values near 28. The greater the difference in height between a station and the nearest grid point of the ERA-Interim, the more significant is the temperature differences (higher RMSE, MAE and BIAS values) and the value of t is greater. However, at the stations with temperature differences greater than 1.5 K, but with a negative BIAS the opposite happens. The grid points have higher altitude than the stations and the height difference is negative, which occurs in mountainous areas.

Figure 11 shows the geographical distribution of the rejected stations. The colour of the circles shows the attributed causes for the rejection. Stations with positive height differences are La Gran Piedra in Santiago de Cuba, Topes de Collantes in Sancti Spiritus, Pinares de Mayari in Holguin and Palenque de Yateras in Guantanamo. Stations with negative height differences are Trinidad and Sancti Spiritus in Sancti Spiritus province, Camilo Cienfuegos in Ciego de Avila, Cabo Cruz in Granma, Guaro in Holguin, Santiago de Cuba in Santiago de Cuba province and Guantanamo in Guantanamo province.

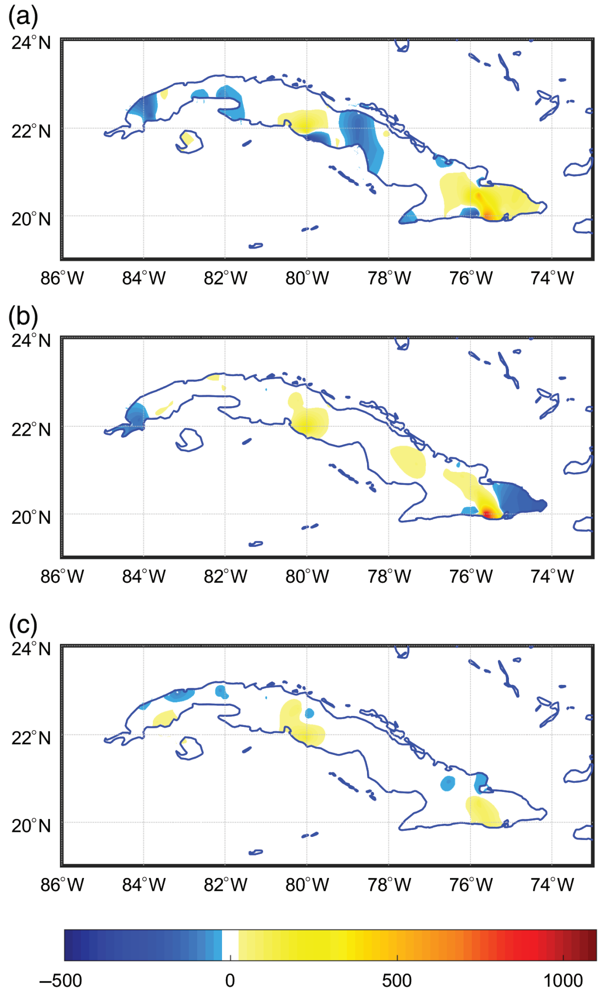

Figure 10. Elevation differences (m) between the ERA-Interim nearest grid point and the stations for the three resolutions. The white shaded zone corresponds to ±25 m. (a) 0.5° resolution, (b) 0.25° resolution, (c) 0.125° resolution.

The ERA-Interim land–sea mask in Figure 11 shows a rather gross representation of landscape details such as coasts and islands. The land mask does not completely cover the territory, and even at higher resolutions there are land areas that are defined by the mask as sea. The dynamic processes taking place in these areas are not analysed in the same way by the assimilation model, which contributes to the errors. Groen and Wolters (2011) made an analysis of the implications of the land–sea mask and topography of ERA-Interim from the dynamic point of view on wind speed forecasts. They showed how the mask setting (land–sea transition) and topography (roughness of the terrain) cause an underestimation or an overestimation of the wind speed, which can be associated with the change in patterns of air circulation at the local level, which directly influences the heat exchange

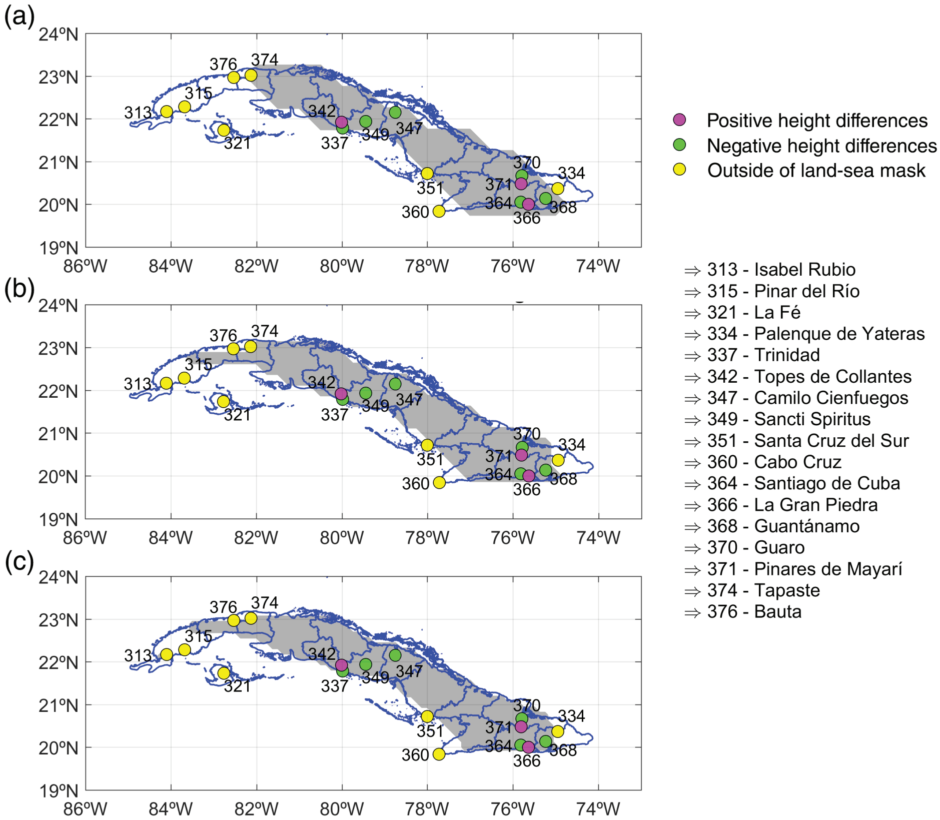

Figure 11. Locations of the rejected stations. The shaded gray area corresponds to the land mask of ERA-Interim. The colour of the circles shows the probable cause of the differences between ERA-Interim and the stations. (a) Land mask ERM-Interim 0.5°, (b) Land mask ERM-Interim 0.25° and (c) Land mask ERM-Interim 0.125°.

between the surface and atmosphere, and hence the temperature.

Eight Cuban stations with significant differences are located outside the land area in the mask of ERA-Interim, Isabel Rubio and Pinar del Rio in Pinar del Rio province, La Fe in Isla de la Juventud, Tapaste in Mayabeque, Bauta in Artemisa, Santa Cruz del Sur in Camagüey, Cabo Cruz in Granma and Palenque de Yateras in Guantanamo. The last two stations also show a large height differences with respect to the nearest grid points, but the colour of the circle was plotted in yellow to illustrate that these points are also outside the ERA-Interim land mask (Figure 11). Likewise, there are grid points that are located in the ocean and their topographical height is equal to 0 m. Some stations have an ocean point as the nearest grid point, and although Figure 11 only displays the stations with differences >3.0 K, there are many stations when the nearest grid point elevation is equal to 0 m.

There are also stations that do not show a great height difference between the stations and the nearest grid points of ERA-Interim, but have significant temperature differences, and they were also rejected in the statistical tests. These stations are located in Cuban zones where local processes produce extreme values of temperature (absolute minimum temperatures) or where the processes of convergence of local breezes take place. The Cuban zone where the lowest values of temperature are more frequent is the Habana-Matanzas plain (blue shaded areas in Figure 12).

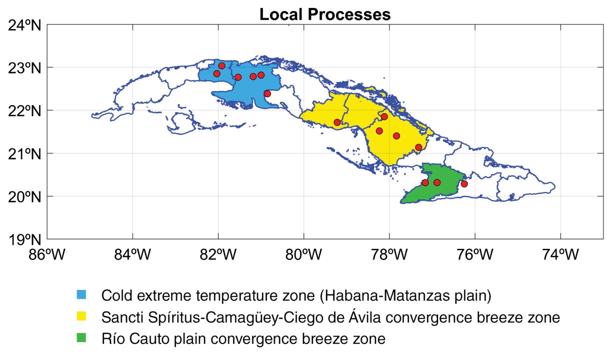

Figure 12. Locations of the rejected stations. The colour-shaded areas correspond to the probable local processes causing the differences between ERA-Interim and the stations.

The stations that are located in the Cuban zone where the lowest multiannual temperatures have been recorded are Güines (2.1 °C) and Bainoa (0.6 °C) in Mayabeque province, Indio Hatuey (1.2 °C), Unión de Reyes (1.0 °C) and Jovellanos (2.4 °C) in Matanzas province and Aguada de Pasajeros (2.6 °C) in Cienfuegos.

Figure 13 shows the coefficient of variation of T2 m within the island for the 0.5° resolution. Although there is a low variability in the station data sets, a different behaviour of the coefficient of variation along the island is observed. The highest variability is seen in the Cuban regions where the lowest values of temperature are more

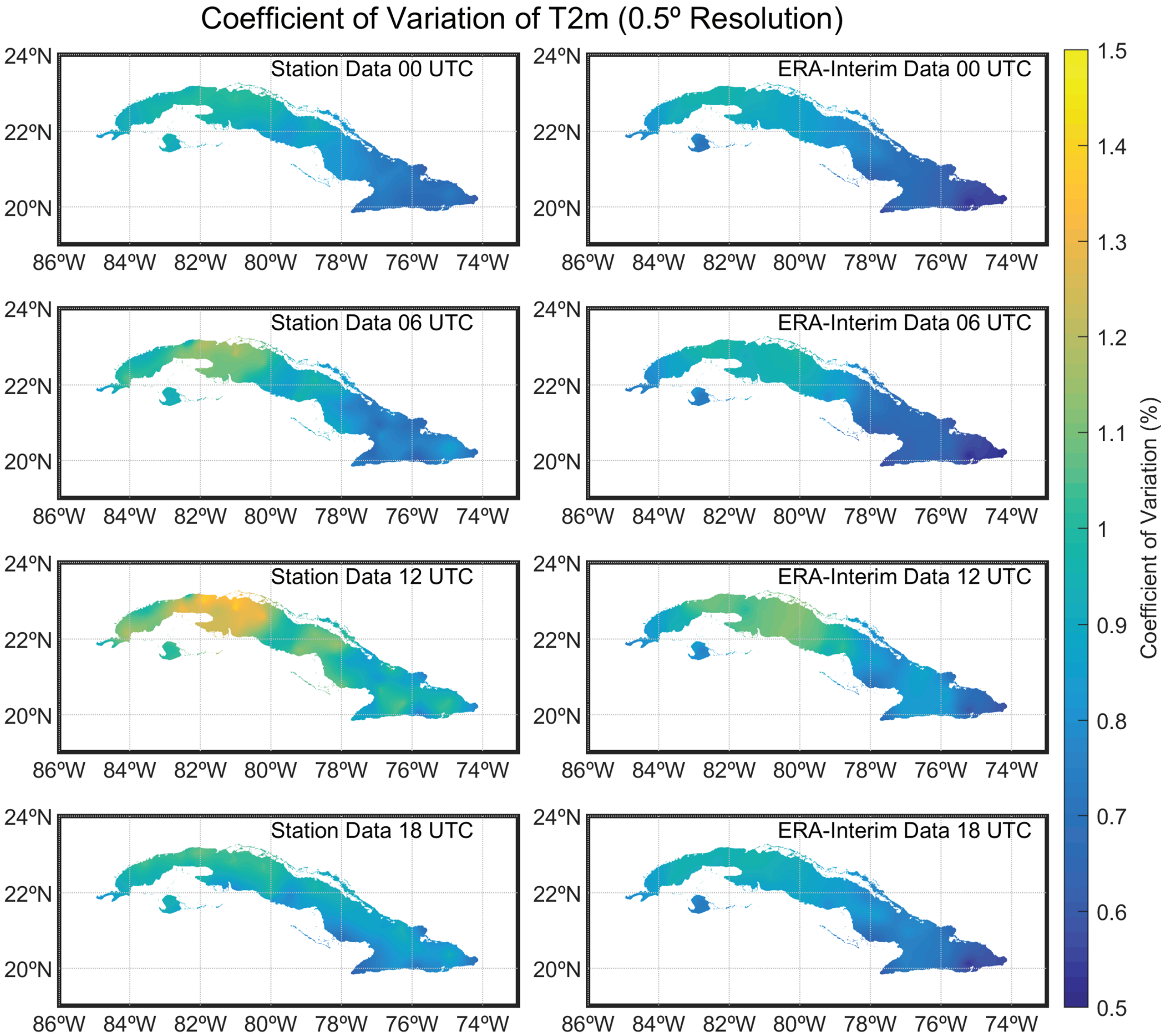

Figure 13. Coefficient of variation of T2m along the island at each observation time. The left panels show the coefficient of variation of the station dataset and the right panels show the coefficient of variation of the ERA-Interim data set.

frequent. In addition to this, the greatest variability is observed between 0600 and 1200 UTC (0001 and 0700 local time), with the highest values of the coefficient of variation at 1200 UTC. The coefficient of variation of ERA-Interim data sets shows a similar behaviour along the island. Only at 1200 UTC is a higher variability in the Cuban zone observed, when the lowest values of temperature are more frequent. However, this variability is smaller than the variability observed in station observations. The coefficient of variability at the other two resolutions (0.25° and 0.125°) shows a similar behaviour.

Sea breezes generate convergence zones in several regions inside Cuba, and they are associated with convection, rain and severe local storms in the rainy season (Pazos, 1998; Mayor and Mesquita, 2015). Different types of synoptic situations in Cuba determine the location of sea breeze convergence (Carnesoltas, 1986; Pazos, 1998). These areas are located towards the centre of the territory throughout the island, and in the Río Cauto plain (Carnesoltas, 1986; Pazos, 1998; Fernández and Díaz, 2000; Benedico, 2003; Fernández and Díaz, 2005; Benedico *et al.,* 2012). However, we will only analyse the convergence zones formed in the provinces Santi Spíritus-Ciego de Ávila-Camagüey (yellow shaded areas in Figure 12) and the one formed in the Rio Cauto plain (green shaded areas in Figure 12). Studies of mesoscale flow by Bueno *et al.* (1994), Benedico *et al.* (2012) and Pazos (1998) clearly showed these two areas of sea breeze convergence.

The first convergence zone occurs in the central zone of the island, between Sancti Spiritus and Camagüey. Six stations are located in this zone, Sur del Jibaro in Sancti Spiritus, Esmeralda, Florida, Camagüey and Palo Seco in Camagüey province and Puerto Padre in Las Tunas. Three stations are located in the other zone affected by local

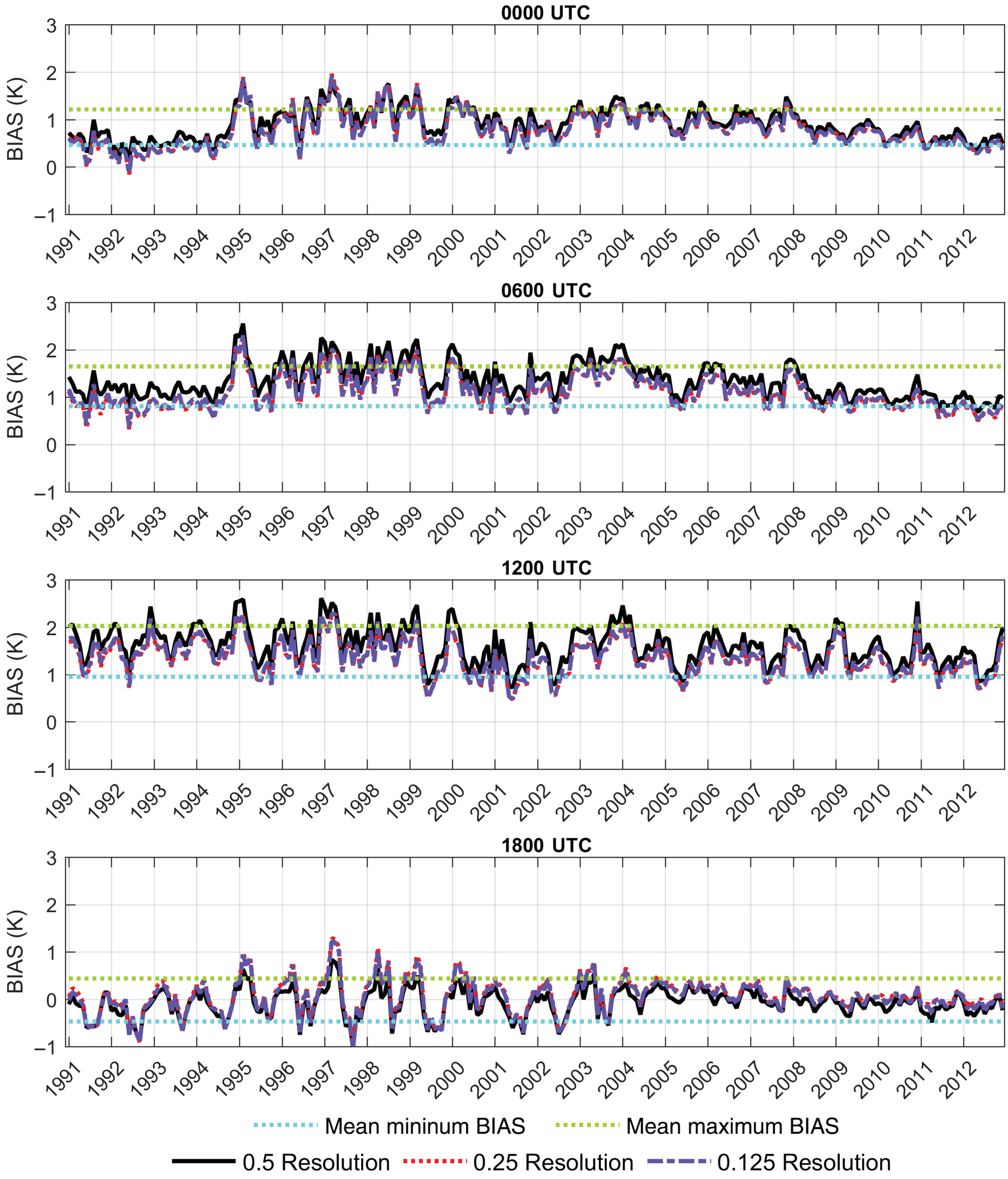

Figure 14. Monthly mean temperature differences. Dashed blue line is the mean minimum BIAS and dashed green line is the mean maximum BIAS.

breezes, in the Cauto plain, in Granma province and these stations are Manzanillo and Veguitas in Granma province and Contramaestre in Santiago de Cuba.

Another problem may be related to the limited number of stations used in the analysis, which means that episodes of extreme weather or climate are not always reflected in all their spatial or temporal dimensions (Kalnay, 2003; Dee *et al.,* 2011). However, there are other factors such as large-scale biases of ERA-Interim caused by inaccurate surface forcing, poor resolution of the boundary layer, simplified representations of moist physics and clouds and various other imperfections can affect the observed differences between the model and observations (Dee, 2005; Gao *et al.,* 2012a).

Previous studies showed that deviations between T2m_ERA and T2m_Obs are small, except for the regions where the height differences between the grid points of the model and the real world are large, such as in the central Alps (Gao *et al.,* 2012a). However, to reduce total integrated errors, corrections for the vertical temperature gradients were used. In this and other studies, stations located in mountains, valleys or near these had significant differences between the observed and modelled values (Zhao *et al.,* 2007; Durand *et al.,* 2009; Gao *et al.,* 2012a).

Some patterns such as the El Niño-Southern Oscillation (ENSO) affect the atmospheric circulation globally, and its impacts are strongly felt over Central America and the Caribbean (Jury and Gouirand, 2011; Gouirand *et al.,* 2012). The monthly differences show a seasonal behaviour with an increase of the BIAS values at the less rainy season and a decrease at the rainy season. However, the minimum or maximum values of BIAS are more significant for years with an El Niño or La Niña. Figure 14 shows that the greatest temperature differences are observed in years with moderate and strong processes, both positive and negative phases of ENSO. These years show positive differences greater than the mean maximum BIAS at less rainy season (dashed green line) and negative differences smaller than mean minimum BIAS for rainy season (dashed sky blue line). At 0600 and 1200 UTC are observed temperature differences between 0.4 and 2.6 K and at 0000 UTC are observed temperature differences smaller than 2.0 K. The most significant temperature differences at 1800 UTC show negative values and the most significant positive values are observed in 1997. In this year a very strong process of ENSO took place. Some years, such as 1995, 1998, 1999, 2000, 2004 and 2011 present significant temperature differences. These years show moderate and strong ENSO events. Also, the amplitude between the maximum BIAS on January and February, and the minimum BIAS on July and August is greater in these years.

The sea breeze circulation and local convective motions occur at small scales. The ENSO episodes may increase or decrease the magnitude of these local processes and also can change their behaviour. Although the horizontal and vertical resolution of the models have increased considerably, many important processes in the atmosphere cannot be resolved explicitly with current or future models (Kalnay, 2003), and local processes mentioned above are the type of processes that cannot be resolved by the current numerical models.

4. Conclusions

Comparison of the ERA-Interim values of 2-m air temperature and station observations in Cuba shows that the differences depend on the interpolation method used to determine the value of T2m to compare to the station. Among the three methods we investigated, the mean of the four grid points gave the most accurate results, although the nearest neighbour method gave almost as reasonable results. For this last method, the number of stations with small differences increased as the spatial scale decreased.

Different comparison statistics between the modelled and observed values show a predominance of differences in the range of 1.0–2.0 K, with over 40% of the cases. However, some stations showed differences >6.0 K. Generally, ERA-Interim overestimates the 2-m temperature for Cuba for three of the four observation times (0000, 0600 and 1200 UTC), but at 1800 UTC the values are very close for most stations. The smallest differences are at coastal stations. The monthly time scale shows a better relationship between ERA-Interim and stations than for annual, rainy and the less rainy seasons.

The large differences that appear at some stations are associated with three main elements. The first is the land–sea mask used by the ERA-Interim, as there are stations that the model places in the ocean. The second element relates to the difference in height between the stations and the four points around them. In areas with very high elevations, the four grid points around the station have a lower height than the station. In these zones, the model overestimated the temperature values. In the mountain valleys, the four grid points around the station are higher than the station, and therefore the model underestimated the temperature values in this area. This could possibly be corrected by calculating the theoretical temperature difference from the height differences. The third element is associated with local extremes of temperature caused by sea breezes in some areas of the country that are not adequately represented by ERA-Interim, which is related to the limited number of stations used in the analysis as well as to the lack of parameterizations of sea breeze process for the island of Cuba. The low variability observed in the ERA-Interim data sets may be related to the conditions described above, because the differences in the coefficient of variation observed along the island for the station data sets are not adequately represented by ERA-Interim.

The results of this study show clearly the possibility of using data from the ERA-Interim for climate type studies at monthly scales. For future research, these results could also allow us to make a statistical analysis at the daily scale, which can then serve to determine the trends of the series, including analysis of change points in these series.

INSMET in coordination with the WMO should establish agreements to increase the number of weather stations that report information to global data exchange systems. This will increase the number of stations used for data assimilation schemes for atmospheric reanalysis, provided they meet quality control criteria.

In addition, INSMET should develop further studies about local phenomena, such as the locations of the extreme value of temperature and their causes, the behaviour of temperature in complex terrain and convergence of local breezes. These types of studies will allow future modelling efforts of Cuban weather and climate to include realistic parameterizations of these kinds of phenomena. Such improvements in models simulating Cuban weather and climate will improve both weather and climate forecasting for the island.

Acknowledgements

Special thanks are given to the ECMWF public web server (http://data.ecmwf.int/data/) for providing the reanalysis data sets. This work was supported by the project entitled, ‘The Future of Climate Extremes in the Caribbean’ (XCUBE), funded by the Norwegian government, and the Bjerknes Centre for Climate Research. This project is part of the scientific cooperation between Norway and Cuba. Alan Robock is supported by the New Jersey Agricultural Experiment Station. Special thanks to the Centre for Climate Dynamics at the Bjerknes Centre for Climate Research for their contribution in this study, and also a special thanks to Dick Dee and the ERA-Interim TEAM for the assistance provided in this study

References

Angel JR, Easterling WE, Kirsch SW. 1993. Towards defining appropriate averaging periods for climate normals. *Climatol. Bull.* **27**: 29–44.

Arguez A, Vose R. 2011. The definition of the standard WMO climate normal. *Bull. Am. Meteorol. Soc.* **92**: 699–704. https://doi.org/10.1175/2010BAMS2955.1.

Báez R, Rodríguez A, Masó N, Santana M, Pérez R, Wong A, Ramírez M, Martines CL, Casal E. 2009. Sistema de administración de datos climáticos. *Rev. Cubana Meteorol.* **15**(1): 23–37.

Benedico O. 2003. *Línea de Confluencia del Flujo a Mesoescala Sobre Ciego de Ávila: Situaciones a Escala Sinóptica Favorables para su Ocurrencia y su Relación con las Precipitaciones Convectivas*. Tesis presentada Para Optar Por el Grado Académico de Máster en Ciencias, La Habana.

Benedico O, Fernández AJ, Barcia S. 2012. Patrones de Circulación Asociados (PCA) a los Entornos de Gradiente de Presión en Superficie (EGPS) para el período lluvioso y su relación con el campo de viento y las precipitaciones sobre Sancti Spíritus, Ciego de Ávila y Camagüey. *Rev. Cubana Meteorol.* **18**(2): 138–151.

Brands S, Gutiérrez JM, Herrera S. 2011. On the use of reanalysis data for interpolation. *J. Clim.* **25**: 2517–2526.

Bueno J, Lecha L, Fernández A. 1994. Las precipitaciones diarias en la región central de Cuba y su relación con la influencia anticiclónica subtropical y el campo de viento dominante. *Rev. Ciencias Tierra Espacio* **23–24**: 73–83.

Carnesoltas CM. 1986. *La Circulación Local de Brisa en Cuba*. Tesis presentada en opción al grado de candidato a doctor en ciencias, La Habana.

Chung CE, Cha H, Vihma T, Räisänen P, Decremer D. 2013. On the possibilities to use atmospheric reanalyses to evaluate the warming structure in the Arctic. *Atmos. Chem. Phys.* **13**: 11209–11219. https://doi.org/10.5194/acp-13-11209-2013.

Corbelle E, Gil M, Armesto J, Rego T. 2006. La escala cartográfica de la imagen de satélite. Caso particular de las imágenes Ikonos y QuickBird. *Rev. Teledetección* **26**: 18–24.

Curtis S. 2002. Interannual variability of the bimodal distribution of summertime rainfall over Central America and tropical storm activity in the far-eastern Pacific. *Clim. Res.* **22**: 141–146.

Daniel JS, Portmann RW, Solomon S, Murphy DM. 2012. Identifying weekly cycles in meteorological variables: the importance of an appropriate statistical analysis. *J. Geophys. Res.* **117**: D13203. https://doi.org/10.1029/2012JD017574.

Dee DP. 2005. Bias and data assimilation. *Q. J. R. Meteorol. Soc.* **131**(613): 3323–3343. https://doi.org/10.1256/qj.05.137.

Dee D, Uppala S. 2009. Variational bias correction in ERA-Interim. *ECMWF Newsl.* **119**: 21–29.

Dee DP, Uppala SM, Simmons AJ, Berrisford P, Poli P, Kobayashi S, Andrae U, Balmaseda MA, Balsamo G, Bauer P, Bechtold P, Beljaars ACM, van de Berg L, Bidlot J, Bormann N, Delsol C, Dragani R, Fuentes M, Geer AJ, Haimberger L, Healy SB, Hersbach H, Hólm EV, Isaksena L, Kållbergc P, Köhler M, Matricardi M, McNally AP, Monge-Sanzf BM, Morcrette J-J, Parkg B-K, Peubey C, de Rosnay P, Tavolato C, Thépaut J-N, Vitart F. 2011. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc.* **137**: 553–597. https://doi.org/10.1002/Qj.828.

Durand Y, Laternser M, Giraud G, Etchevers P, Lesaffre B, Mérindol L. 2009. Reanalysis of 44 yr of climate in the French Alps (1958–2002): methodology, model validation, climatology, and trends for air temperature and precipitation. *J. Appl. Meteorol. Climatol.* **48**: 429–449.

ECMWF. 2009. *IFS Documentation-Cy40r1 Part IV: Physical Processes*. ECMWF: Reading, UK.

Fernández AJ, Díaz YA. 2000. La zona de convergencia del viento a las 1300 horas y su vínculo con la actividad de precipitaciones. *Rev. Cubana Meteorol.* **7**(1): 34–37.

Fernández AJ, Díaz YA. 2005. *Catálogo de Procesos Sinópticos del Archipiélago Cubano en el Período 1979–1993*. GeoCuba Gráfica: Cienfuegos, Cuba.

Gao L. 2013. *Validation and Statistical Interpolation of ERA-Interim Reanalysis Data for Integrated Applications*. Dissertation der Fakultät für Geowissenschaften der Ludwig–Maximilians–Universität München zur Erlangung des akademischen Grades "doctor rerum naturalium". München, Germany.

Gao L, Bernhardt M, Schulz K. 2012a. Elevation correction of ERA-Interim temperature data in complex terrain. *Hydrol. Earth Syst. Sci.* **16**: 4661–4673.

Gao L, Bernhardt M, Schulz K. 2012b. Interactive comment on "Downscaling ERA-Interim temperature data in complex terrain". *Hydrol. Earth Syst. Sci. Discuss.* **9**: C3531–C3537.

Giannini A, Kushnir Y, Cane MA. 2000. Interannual variability of Caribbean rainfall, ENSO, and the Atlantic Ocean. *J. Clim.* **13**: 297–311.

Gouirand I, Jury M, Sing B. 2012. An analysis of low and high frequency summer climate variability around the Caribbean Antilles. *J. Clim.* **25**: 3942–3952. https://doi.org/10.1175/JCLI-D-11-00269.1.

Groen G, Wolters D. 2011. Upper air climatology of Amsterdam FIR, using ERA-Interim 1989-2008. Scientific Report WR 2011-02, Part 2/2, De Bilt, The Netherlands.

Jolliffe I, Stephenson D. 2003. *Forecast Verification*. John Wiley & Sons: Chichester, UK.

Jury MR, Gouirand I. 2011. Decadal climate variability in the Eastern Caribbean. *J. Geophys. Res.* **116**: D00Q03. https://doi.org/10.1029/2010JD015107.

Kalnay E. 2003. *Atmospheric Modeling Data Assimilation and Predictability*. Cambridge University Press: New York.

Kalnay E, Kanamitsu M, Kistler R, Collins W, Deaven D, Gandin L, Iredell M, Saha S, White G, Woollen J, Zhu Y, Chelliah M, Ebisuzaki W, Higgins W, Janowiak J, Mo KC, Ropelewski C, Wang J, Leetmaa A, Reynolds R, Jenne R, Joseph D. 1996. The NCEP/NCAR 40-year reanalysis project. *Bull. Am. Meteorol. Soc.* **77**: 437–471.

Kistler R, Kalnay E, Collins W, Saha S, White G, Woollen J, Chelliah M, Ebisuzaki W, Kanamitsu M, Kousky V, van den Dool H, Jenne R, Fiorino M. 2001. The NCEP–NCAR 50-year reanalysis: monthly means CD-ROM and documentation. *Bull. Am. Meteorol. Soc.* **82**: 247–267.

Lecha LB, Paz LR, Lapinel B. 1994. *El Clima de Cuba*, 2nd edn. Editorial Academia: La Habana.

Linsay R, Wenshanan M, Schweiger A, Zhang J. 2013. Evaluation of seven different atmospheric reanalysis products in the Arctic. *J. Clim.* **27**: 2588–2606.

Magana V, Amador JA, Medina S. 1999. The midsummer drought over Mexico and Central America. *J. Clim.* **12**: 1577–1588.

Mapes BE, Liu P, Buenning N. 2005. Indian monsoon onset and America's midsummer drought: out-of-equilibrium responses to smooth seasonal forcing. *J. Clim.* **18**: 1109–1115.

Mayor YG, Mesquita MDS. 2015. Numerical simulations of the 1 May 2012 deep convection event over Cuba: sensitivity to cumulus and microphysical schemes in a high-resolution model. *Adv. Meteorol.* **2015**(8): 1–16. https://doi.org/10.1155/2015/973151.

Mesquita MDS, Erikstad KE, Sandvik H, Barrett RT, Reiertsen TK, Anker-Nilssen T, Hodges KI, Bader J. 2015. There is more to climate than the North Atlantic Oscillation: a new perspective from climate dynamics to explain the variability in population growth rates of a long-lived seabird. *Front. Ecol. Evol.* **3**: 1–14. https://doi.org/10.3389/fevo.2015.00043.

Minder JR, Mote PW, Lundquist JD. 2010. Surface temperature lapse rates over complex terrain: lessons from the Cascade Mountains. *J. Geophys. Res.* **115**: D14122. https://doi.org/10.1029/2009JD013493.

Mooney PA, Mulligan FJ, Fealy R. 2011. Comparison of ERA-40, ERA-Interim and CEP/NCAR reanalysis data with observed surface air temperatures over Ireland. *Int. J. Climatol.* **31**: 545–547.

Pazos C. 1998. *Análisis Sinóptico-Estadístico del Viento en Cuba y la Ocurrencia de Brisas Marinas y Tormentas Locales Severas en la Provincia de Ciego de Ávila*. Tesis presentada en opción al grado científico de doctor en ciencias meteorológicas, Instituto de Meteorología La Habana.

Renard B, Lang M, Bois P, Dupeyrat A, Mestre O, Niel H, Sauquet E, Prudhomme C, Parey S, Paquet E, Neppel L, Gailhard J. 2008. Regional methods for trend detection: assessing field significance and regional consistency. *Water Resour. Res.* **44**(8): W08419. https://doi.org/10.1029/2007WR006268.

Rolland C. 2003. Spatial and seasonal variations of air temperature lapse rates in Alpine regions. *J. Clim.* **16**: 1032–1046.

Rudloff W. 1981. *World-Climates, With Tables of Climatic Data and Practical Suggestions*. Wissenschaftliche Verlagsgesellschaft mbH: Stuttgart, Germany.

Santer BD, Wigley TML, Boyle JS, Gaffen DJ, Hnilo JJ, Nychka D, Parker DE, Taylor KE. 2000. Statistical significance of trends and trend differences in layer-average atmospheric temperature time series. *J. Geophys. Res.* **105**(D6): 7337–7356.

Santer BD, Thorne PW, Haimberger L, Taylor KE, Wigley TML, Lanzante JR, Solomon S, Free M, Gleckler PJ, Jones PD, Karl TR, Klein SA, Mears C, Nychka D, Schmidt GA, Sherwood SC, Wentz FJ. 2008. Consistency of modelled and observed temperature trends in the tropical troposphere. *Int. J. Climatol.* **28**: 1703–1722.

Serbin S, Kucharik C. 2008. Spatiotemporal mapping of temperature and precipitation for the development of a multidecadal climatic dataset for Wisconsin. *J. Appl. Meteorol. Clim.* **48**: 742–757.

Simmons A, Uppala S, Dee D, Kobayashi S. 2006. ERA-Interim: new ECMWF reanalysis products from 1989 onwards. *ECMWF Newsl.* **110**: 25–35.

Szczypta C, Calvet J-C, Albergel C, Balsamo G, Boussetta S, Carrer D, Lafont S, Meurey C. 2011. Verification of the new ECMWF ERA-Interim reanalysis over France. *Hydrol. Earth Syst. Sci.* **15**: 647–666.

Thorne P, Vose RS. 2010. Reanalyses suitable for characterizing long-term trends. *Bull. Am. Meteorol. Soc.* **91**: 353–361. https://doi.org/10.1175/2009BAMS2858.1.

Uppala S, Dee D, Kobayashi S. 2008. Towards a climate data assimilation system: status update of ERA-Interim. *ECMWF Newsl.* **115**: 12–18.

World Meteorological Organization. 2011. Guide to Climatological Practices WMO-No. 100, 2011 Edition, World Meteorological Organization, Geneva, Switzerland.

Žagar N, Skok G, Tribbia J. 2011. Climatology of the ITCZ derived from ERA Interim reanalyses. *J. Geophys. Res.* **116**: D15103. https://doi.org/10.1029/2011JD015695.

Zhao T, Guo W, Fu C. 2007. Calibrating and evaluating reanalysis surface temperature error by topographic correction. *J. Clim.* **21**: 1440–1446.

Zou H, Zhu J, Zhou L, Li P, Ma S. 2013. Validation and application of reanalysis temperature data over the Tibetan Plateau. *J. Meteorol. Res.* **28**(1): 139–149.

