Spatial Statistics ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Spatial Statistics

journal homepage: www.elsevier.com/locate/spasta

Copula-based interpolation methods for air temperature data using collocated covariates

Fakhereh Alidoost *, Alfred Stein, Zhongbo Su

ITC, University of Twente, Enschede, 7500AE, The Netherlands

ARTICLE INFO

Article history:
Received 31 October 2017
Accepted 8 August 2018
Available online xxxx

Keywords:
Copula
Interpolation
Covariate
Data scarce
Air temperature

ABSTRACT

This paper introduces two copula-based interpolation methods to produce air temperature maps in a data-scarce area: a spatial copula interpolator including covariates, and a mixed copula interpolator. The methods allow a construction of the conditional distribution of air temperature given the collocated covariates. Our study compared the new methods with the spatial copula interpolator, the ordinary kriging predictor and the cokriging predictor. Daily mean air temperature was used from weather stations and ERA_Interim reanalysis weather data at 174 locations in the Qazvin Plain, Iran. Spatial copula interpolator including covariates resulted in more precise predictions as shown by leave-two-out cross-validation. Visual inspection of air temperature maps demonstrated that the new methods well represented spatial variability of air temperature at a 1 km spatial resolution. The results showed an improved performance of the new methods to describe both spatial variability and co-variability between variables. The methods are potentially useful for other sparsely and irregularly distributed weather data.

1. Introduction

A copula is a multivariate joint distribution that describes the dependence structure between variables (Nelsen, 2006). The joint distribution is estimated using a distribution family that can be different from the family of the marginal distributions of the involved variables. An appealing property of a copula in describing spatial dependences is that its parameter is estimated by means

* Corresponding author.
E-mail address: f.alidoost@utwente.nl (F. Alidoost).

https://doi.org/10.1016/j.spasta.2018.08.003
2211-6753/

of a correlogram that describes dependences based upon the correlation between marginals (Oden, 1984; Gräler and Pebesma, 2011). The purpose of copula-based interpolation methods is to predict the marginal probability at an unvisited location given the marginal probabilities of the nearest neighbours.

Recently, several studies have assessed the performance of copula-based interpolation methods and compared with kriging methods (Bárdossy and Li, 2008; Haslauer et al., 2016; Heißerer et al., 2016; Durocher et al., 2016). Gräler and Pebesma (2011) investigated the application of a multivariate copula that models spatio-temporal random fields using vine structures in interpolation of daily mean PM_{10} measurements. Gräler (2014) demonstrated the potential of spatio-temporal copula interpolation with a single covariate.

This study focuses on mean air temperature. The use of these data in hydrological models, e.g., crop growth simulations for assessing crop water requirement at local scales has been the key to support irrigation management. Application of hydrological models at unvisited locations remains a challenge because weather stations are usually sparse and located at irregular positions. A solution to this problem is to use gridded air temperature data from a weather forecast system. The coarse spatial resolution of those data, however, is a source of uncertainty because of the spatial variability at local scales (Aalto et al., 2013). Hence, interpolation has to take place to predict air temperature at unvisited locations.

Kriging is a well-established interpolation method (Cressie, 1993). Since the twentieth century, a variety of methods has been developed. To predict air temperature values at a 1 km spatial resolution, typical examples are spatio-temporal regression-kriging with incorporation of remote sensing images (Hengl et al., 2012), generalized additive models (Aalto et al., 2013), and residual kriging and regression kriging methods using auxiliary maps (Wu and Li, 2013; Kilibarda et al., 2014). Parmentier et al. (2015) compared universal kriging, generalized additive models and geographically weighted regression. Kilibarda et al. (2014) showed the effect of daily land surface temperature on both minimum and maximum air temperature variability.

With the aim to improve prediction of air temperature in a data-scarce area, we present two interpolation methods based upon copulas: a spatial copula interpolator including covariates, and a mixed copula interpolator. The first method considers two types of dependences: spatial dependences of air temperature at a single location and its nearest neighbours, and non-spatial dependences between air temperature and its collocated covariates at that location. The second extends the first method by including the non-spatial dependences of air temperature and its collocated covariates at the nearest neighbours. The two methods are compared with the spatial copula interpolator, the ordinary kriging and cokriging predictors.

The structure of this paper is as follows: the copula-based interpolation methods are presented in Section 2, the study area and data are introduced in Section 3, the results are described in Section 4, and the discussion and conclusion are in Section 5.

2. Method

2.1. Spatial copula interpolator

Let $f(X|X = x_1, \ldots, X = x_n)$ be the conditional density distribution of the variable X at an unvisited location conditioned on its n nearest neighbours. The conditional expectation is the optimal predictor to derive the value of the variable X at an unvisited location, denoted by $\hat{x}_0$. It can be shown that it minimizes the Bayes risk (Cressie, 1993). The conditional expectation can be either linear or nonlinear in X and it can be written using the conditional copula density function $c\,(U|U = u_1, \ldots, U = u_n)$ as:

$$\hat{x}_0 = E[X|X = x_1, \ldots, X = x_n] = \int_x x \cdot f(X|X = x_1, \ldots, X = x_n)dx = \int_0^1 F^{-1}\,(u) \cdot c(U|U = u_1, \ldots, U = u_n)du, \quad (1)$$

where F is the marginal cumulative distribution function, i.e., $u = F\,(x)$ (Bárdossy and Li, 2008). The predictor has two main parts: a marginal distribution $F\,(.)$ and a multivariate copula $c(.|.)$. The last

Please cite this article in press as: Alidoost F., et al., Copula-based interpolation methods for air temperature data using collocated covariates. Spatial Statistics (2018), https://doi.org/10.1016/j.spasta.2018.08.003.

equality in (1) can be proven by:

$$f(X|X = x_1, \ldots, X = x_n) = \frac{f(x_0, \ldots, x_n)}{f(x_1, \ldots, x_n)} = \frac{c(u_0, \ldots, u_n) \cdot f(x_0) \cdot \ldots \cdot f(x_n)}{c(u_1, \ldots, u_n) \cdot f(x_1) \cdot \ldots \cdot f(x_n)}$$
$$= c(U|U = u_1, \ldots, U = u_n) \cdot$$
$$f(x_0) = c(U|U = u_1, \ldots, U = u_n) \cdot \frac{du}{dx}, \tag{2}$$

where $c(u_0, \ldots, u_n)$ and $c(u_1, \ldots, u_n)$ are the copula density functions. The choice of a Gaussian distribution for f in (1) leads to a linear predictor that is the equivalent to the simple kriging predictor (Cressie, 1993). Such a predictor is able to capture extremes if it is based upon local nearest neighbours rather than a large set of neighbouring observations.

To estimate the joint density function $c(u_0, \ldots, u_n)$ with $m = n + 1$ variables, it is decomposed into $m \cdot (m-1)/2$ bivariate copulas (Gräler, 2014) based on a canonical vine or C-vine structure (Aas et al., 2009). The joint density is the product of all bivariate copula densities within this structure (Aas et al., 2009). For $m = 3$, $c(u_0, u_1, u_2)$ is decomposed as

$$c(u_0, u_1, u_2) = c(u_0, u_1) \times c(u_0, u_2) \times c(C(u_1|u_0), C(u_2|u_0)), \tag{3}$$

where $C(.|.)$ is the conditional copula. Several copula families have been identified that are able to construct a multivariate joint distribution (Nelsen, 2006). These families describe different tail dependence structures (Manner, 2007). We use five families in our study: Student's t (Demarta and McNeil, 2005), Gaussian, Clayton, Gumbel and Frank (Joe, 1993), being sufficiently flexible families to capture the dependence structures. The Student's t family has two parameters: one for the correlation and one for the degrees of freedom whereas the other families have one parameter. Therefore, the number of parameters in the C-vine structure equals $m.(m-1)/2$. The parameter of bivariate copulas is related to Kendall's τ (Nelsen, 2006) and maximum likelihood for estimation. Based upon their results, the most suitable copula is selected according to Akaike's Information Criteria (AIC) (Akaike, 1974).

The first tree in the vine structure consists of spatial bivariate copulas, e.g., $c(u_0, u_1) \times c(u_0, u_2)$, taking the influence of the neighbours into account. The parameter of the spatial bivariate copula is obtained from the correlogram obtained with binned data pairs (Gräler, 2014). Pairs with distances larger than the distance in last spatial bin are considered independent and are described by the Product copula family (Nelsen, 2006). A polynomial of degree two fitted to Kendall's τ values estimates the correlation function. The remaining trees in the vine structure consist of non-spatial bivariate copulas, e.g., $c(C(u_1|u_0), C(u_2|u_0))$.

2.2. The spatial copula interpolator including covariates

To introduce the spatial copula interpolator including covariates, we consider one variable X and two covariates, e.g., Y and Z. The aim is to predict $\hat{x}_0$ with a finite sample of X. Samples of Y and Z are available at all locations. The conditional copula density function in (1) is then written as $c(U|U = u_1, \ldots, U = u_n, V = v_0, W = w_0)$, where $v_0 = F_Y(y_0)$, $w_0 = F_Z(z_0)$, $_0$denotes an unvisited location, F_Y and F_Z are marginal distribution functions of the covariates. The mean predictor in (1) equals:

$$\hat{x}_0 = \int_0^1 F^{-1}(u) \cdot c(U|U = u_1, \ldots, U = u_n, V = v_0, W = w_0)\, du. \tag{4}$$

By conditioning on V and W, the collocated covariates at an unvisited location, i.e., v_0 and w_0 are incorporated to the predictor. The conditional distribution can be extended to higher dimensions by including more than two covariates in a straightforward way.

In this study, we will use the empirical marginal probability u_i at location i is defined using the following rank-order-transformation $u_i = \frac{\text{rank}(x_i)}{N+1}$, where N denotes the total number of observations. A similar transformation is also applied to y_i and z_i. The empirical marginal distribution avoids using the theoretical marginal distributions that might affect the estimation of copula parameter. By

Please cite this article in press as: Alidoost F., et al., Copula-based interpolation methods for air temperature data using collocated covariates. Spatial Statistics (2018), https://doi.org/10.1016/j.spasta.2018.08.003.

means of kernel density estimation (Silverman, 1986), a continuous approximation of the marginal distribution F is obtained under the assumption of stationary. With a Kolmogorov–Smirnov test, the null hypothesis is tested that the values of X are drawn from the empirical distribution F (Conover, 1971). Note that the empirical probabilities are limited to observations and therefore, the interpolation methods are unable to predict extreme values outside the range of the observations.

2.3. Mixed copula interpolator

Next, we introduce the second method, the mixed copula interpolator. The conditional distribution of X conditioned on Y and Z at location i is equal to $C\left(U|V=v_i, W=w_i\right)$, where C is a conditional copula, $v_i = F_Y\left(y_i\right)$ and $w_i = F_Z\left(z_i\right)$. The conditional probability p_i equals:

$$p_i = C\left(u_i|V=v_i, W=w_i\right). \tag{5}$$

The conditional copula C is estimated in a similar way for the spatial copula interpolator including covariates. The conditional probability p_i is used as the probability of nearest neighbour i for copula in (4) and the final form of the predictor equals:

$$\hat{x}_0 = \int_0^1 F^{-1}\left(u\right) \cdot c\left(U|U=p_1, \ldots, U=p_n, V=v_0, W=w_0\right) du. \tag{6}$$

Hence, the collocated covariates at the nearest neighbour, i.e., y_i and z_i are incorporated into the predictor. The conditional distribution can be extended to higher dimensions for including more than two covariates.

2.4. Comparison and evaluation of the interpolation methods

We compare the spatial copula interpolator including covariates (4) and the mixed copula interpolator (6), with the spatial copula interpolator (1), the ordinary kriging predictor (Cressie, 1993) and the cokriging predictor (Stein and Corsten, 1991). We treat available observations from n weather stations as benchmarks for leave-k-out cross-validation to quantify the performance of the interpolation methods. To this end, k stations are removed from the n weather stations and predictions $\hat{x}_i, i = 1, \ldots, k$ are obtained using observations from the reminder of the stations. Each interpolator is then applied on $m = \frac{n!}{k!(n-k)!}$ replications of dependence structures. The mean absolute error (MAE) and error score (ES) are determined as:

$$MAE = \frac{1}{m}\sum_{j=1}^{m}\left(\frac{1}{k}\sum_{i=1}^{k}|x_{ij} - \hat{x}_{ij}|\right), \tag{7}$$

$$ES = rank\left(MAE\right), \tag{8}$$

for each method (Durai and Bhradwaj, 2014). The smallest ES indicates the best interpolator. The overall prediction quality depends upon a good estimation of the copula and the marginal distributions as well as the number of the observations.

The coverage of 90%, 95% and 99% prediction intervals from the conditional distributions $F\left(X|.\right)$ are investigated at each weather station. The number of observed values that fall in the intervals provides insight into the performance of the copula-based methods. This should be interpreted with care, because the type and number of covariates can be different in the copula-based methods. In addition, the spatial variation of mean and standard deviation of the conditional distributions are compared at each weather station.

A 95% prediction interval width (PIW) at an unvisited location is obtained as $PIW_0 = F^{-1}\left(C^{-1}\left(0.975|.\right)\right) - F^{-1}\left(C^{-1}\left(0.025|.\right)\right)$, describing the uncertainty of the predictions (Li, 2010). The kriging methods result in the prediction error variance s_0^2 (Cressie, 1993; Kutner et al., 1996). A 95% PIW at an unvisited location under the assumption of a Gaussian joint distribution is obtained as $PIW_0 =$

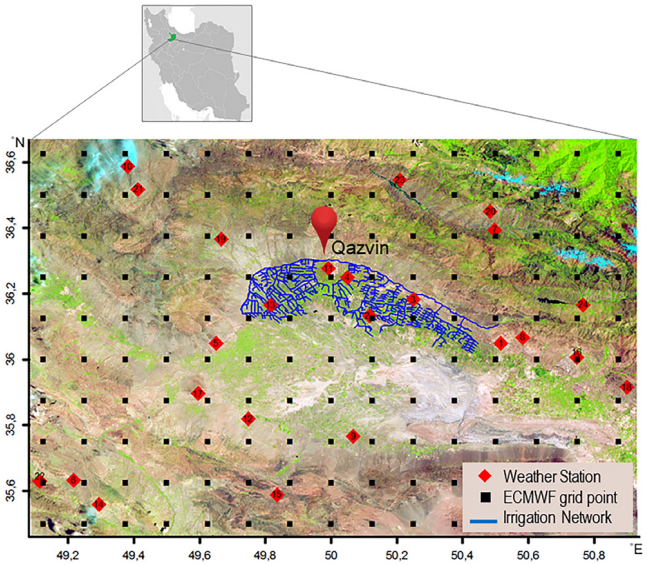

Fig. 1. Study Area located in Qazvin Plain, Iran. The area covers an irrigation network, 24 weather stations, and a sample subset of 10 × 15 grid cells of ECMWF dataset at 0.125° lat/lon distances.

$(\hat{x}_0 + 1.96 \cdot s_0) - (\hat{x}_0 - 1.96 \cdot s_0)$. The methods were implemented in R using the packages gstat (Pebesma, 2004), copula (Kojadinovic and Yan, 2010), spcopula (Gräler and Pebesma, 2011), and VineCopula (Brechmann and Schepsmeier, 2013). We contributed to spcopula and VineCopula packages in R to interpolate the random field spatially including more than one covariate.

3. Application

The Qazvin irrigation network located in the Qazvin plain, Iran (Fig. 1) has served as a pilot study for a project aiming at the development of a planning and monitoring system to support irrigation management (Sharifi, 2013). One objective of this project is to produce daily mean air temperature maps using measurements from weather stations. The study area is enlarged between 35.44° and 36.68° latitudes (N) and 49.09° and 50.92° longitudes (E) (Fig. 1) to have sufficient weather stations. The extended area includes 24 weather stations (Table 1), agricultural fields, urban areas and natural vegetation. The European Centre for Medium-Range Weather Forecasts (ECMWF) provides reanalysis weather data using the ERA-Interim data assimilation system (Persson, 2013). The air temperature is available at three hourly intervals at 10 × 15 grid cells at a 0.125° lat/lon resolution (Fig. 1). The daily mean air temperature is determined by averaging the minimum and maximum temperatures for each of the 150 grid cells and 24 weather stations, in total 174 locations, at each day of June in 2014. This month is chosen considering its importance in the crop calendar of the irrigation network: it is the end of winter crops and the beginning of summer crops especially maize.

We applied the interpolation methods to mean air temperature on June 6th and 22nd 2014 denoted by d_6 and d_{22}, respectively. These two days were selected as these were Landsat 8 overpass days and thus provided three covariates for the 19 of the 24 weather stations: land surface temperature (LST), leaf area index (LAI) and elevation (Fig. 3). Five weather stations were outside the coverage of Landsat 8 images (Zanter, 2016). For LST, we followed the method introduced by Jiménez-Muñoz et al. (2014) and for LAI, that from Allen et al. (1998). Landsat 8 images have a 30 m spatial resolution and 16 days temporal resolution. Surface elevation was obtained from the SRTM dataset (Jarvis et al., 2008) at a 90m spatial resolution. Investigating the correlations (Table 2), we ignored LST as a covariate at d_{22}.

Please cite this article in press as: Alidoost F., et al., Copula-based interpolation methods for air temperature data using collocated covariates. Spatial Statistics (2018), https://doi.org/10.1016/j.spasta.2018.08.003.

Table 1
Twenty-four weather stations measure air temperature in the study area. Daily minimum and maximum air temperature are available in June 2014.

Station ID	Station name	Latitude	Longitude	Elevation (m)	Type	Air temperature measurements
1	Abeyk	36.05	50.52	1278	Climatology type1	6 hourly
2	Magsal	36.13	50.12	1205	Climatology type1	6 hourly
3	Nirougah	36.18	50.25	1299	Climatology type1	6 hourly
4	Qazvin	36.25	50.05	1280	Synoptic	3 hourly
5	Takestan	36.05	49.65	1326	Synoptic	3 hourly
6	Avaj	35.63	49.22	1888	Climatology type1	6 hourly
7	Baghkelaye	36.39	50.50	1256	Climatology type2	min and max
8	Baghkosar	36.07	50.58	1541	Climatology type2	min and max
9	Bouinzahra	35.77	50.07	1213	Synoptic	3 hourly
10	Bourmanak	36.59	49.38	578	Climatology type2	min and max
11	Camp	36.28	49.99	1311	Climatology type2	min and max
12	Danesfahan	35.82	49.75	1303	Climatology type2	min and max
13	Dolatabad	36.17	49.82	1249	Climatology type2	min and max
14	Estalaj	35.56	49.29	2340	Climatology type2	min and max
15	Hajiarab	35.59	49.84	1707	Climatology type2	min and max
16	Hashtgerd	36.01	50.75	1601	Synoptic	3 hourly
17	Jahanabad	35.90	49.60	1372	Climatology type2	min and max
18	Karaj	35.92	50.90	1657	Synoptic	3 hourly
19	Kouhin	36.37	49.67	1498	Climatology type2	min and max
20	Moalem	36.45	50.48	1569	Synoptic	3 hourly
21	Niarak	36.52	49.41	1184	Climatology type2	min and max
22	Qouzlo	35.63	49.11	2061	Climatology type2	min and max
23	Razmiankia	36.55	50.21	1010	Climatology type2	min and max
24	Taleghan	36.17	50.77	1827	Synoptic	3 hourly

Table 2
Correlations between mean air temperature and its covariates on d_6 and d_{22}. The temperature values are the combination of bias corrected values and measurements from weather stations. The covariates are elevation, land surface temperature (LST) and leaf area index (LAI).

	Elevation	LST	LAI
Mean air temperature on d_6	−0.25	0.24	−0.23
Mean air temperature on d_{22}	−0.26	−0.02	−0.23

The covariates are at different spatial resolutions (Fig. 3). Throughout we maintained a 1 km resolution that represents spatial variation of air temperature.

We defined bias as a systematic overestimation and underestimation of reanalysis weather data with respect to measurements (Persson, 2013; Mao et al., 2015). The average bias for all stations equals 3.9 °C and 3.4 °C at d_6 and d_{22}, respectively. We applied a bias correction method to obtain bias corrected values (Alidoost and Stein, 2016). A two-sample Kolmogorov–Smirnov test was performed of the null hypothesis that measurements and bias corrected values are drawn from the same distribution. The p values were equal to 0.22 and 0.65 at d_6 and d_{22}, respectively, did not reject the null hypothesis. Based upon these results, we used a combination of measurements and the bias corrected values as observations for fitting purposes in the interpolation methods (Fig. 2).

4. Results

4.1. Distribution of the observations

The empirical marginal distribution is shown in Fig. 4. The p values obtained using the Kolmogorov–Smirnov test were equal to 0.79 and 0.48 at d_6 and d_{22}, respectively. The number of observations in the tails of the distributions was low, i.e., there were two extremes in the upper tail at d_6 and one in the lower tail at d_{22} (Fig. 4). For copula-based interpolators, in contrast to kriging

Please cite this article in press as: Alidoost F., et al., Copula-based interpolation methods for air temperature data using collocated covariates. Spatial Statistics (2018), https://doi.org/10.1016/j.spasta.2018.08.003.

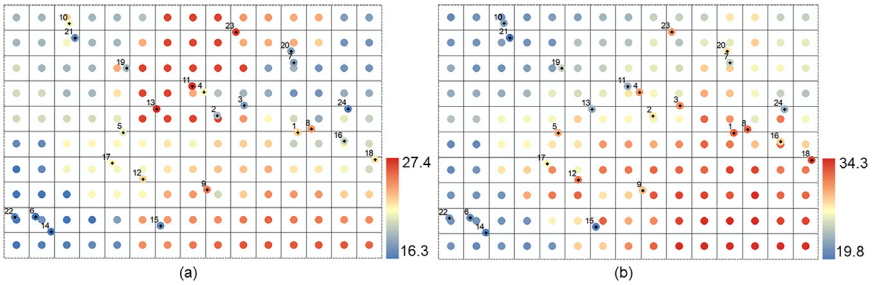

Fig. 2. Spatial variation of mean air temperature at 174 locations from the weather stations and the bias corrected reanalysis weather data on d_6 (a) and d_{22} (b).

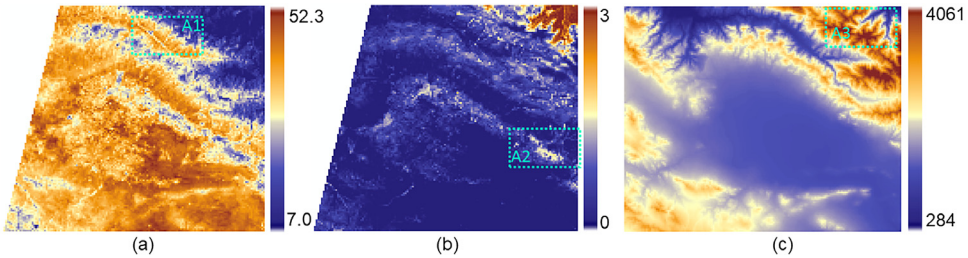

Fig. 3. Three covariates for air temperature. (a) LST and (b) LAI are obtained using Landsat 8 bands at a 30 m spatial resolution. (c) Surface elevation is obtained from the SRTM dataset at a 90 m spatial resolution. The areas A1, A2 and A3 are selected to investigate the co-variability of the air temperature.

Table 3
The number of observed values that fall in the 90%, 95% and 99% prediction intervals of the conditional cumulative probabilities $F(X|.)$ for 19 weather stations on d_6. The observed values are the measurements from weather stations. The covariates are elevation, land surface temperature (LST) and leaf area index (LAI).

Prediction interval	Spatial copula interpolator using covariates	Mixed copula interpolator	Spatial copula interpolator
90	15	14	15
95	17	17	19
99	19	18	19

predictors, it is a challenge to estimate a skewed marginal distribution with two extreme values out of 174 observations. They are not able to predict the extremes in leave-k-out cross validation for $k \geq 2$. Hence, the marginal distribution function has to be well estimated.

Fig. 5 shows the fit to Kendall's τ values in the correlogram for six and five spatial bins at d_6 and d_{22}, respectively. Apparently, the correlogram changes over the range of $[-0.2, 0.7]$ describing the positive and negative dependences. The Student t and Clayton copulas are selected according to the lowest AIC values at each bin at d_6 whereas Student t and Gumbel copulas are selected at d_{22} (Fig. 5).

The multivariate distributions were estimated using the C-vine structures and the conditional cumulative probabilities $F(X|.)$ for 19 weather stations are shown in Fig. 6. The number of observed values that fall within the 90%, 95% and 99% prediction intervals for spatial copula interpolator using covariates were equal to 15, 17 and 19 whereas for mixed copula interpolator were equal to 14, 17 and 18, respectively (Table 3). Hence, it showed a good performance of the methods in the fitting of the distributions.

For the ordinary kriging, the variogram is obtained for the same number of spatial bins as the correlogram, followed by fitting a Gaussian variance function to the variogram of the mean air

Please cite this article in press as: Alidoost F., et al., Copula-based interpolation methods for air temperature data using collocated covariates. Spatial Statistics (2018), https://doi.org/10.1016/j.spasta.2018.08.003.

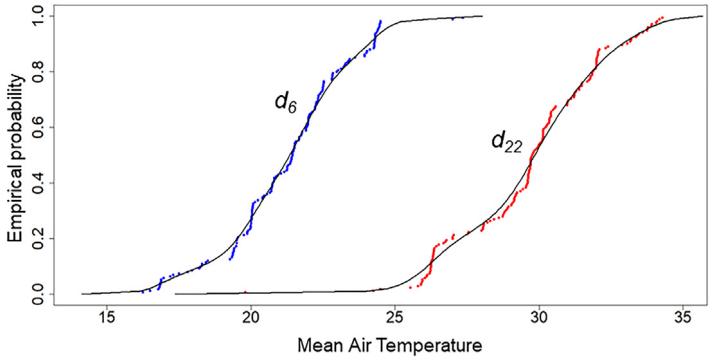

Fig. 4. Empirical marginal probabilities obtained on d_6 and d_{22}. The empirical marginal distribution function is obtained using kernel density estimation.

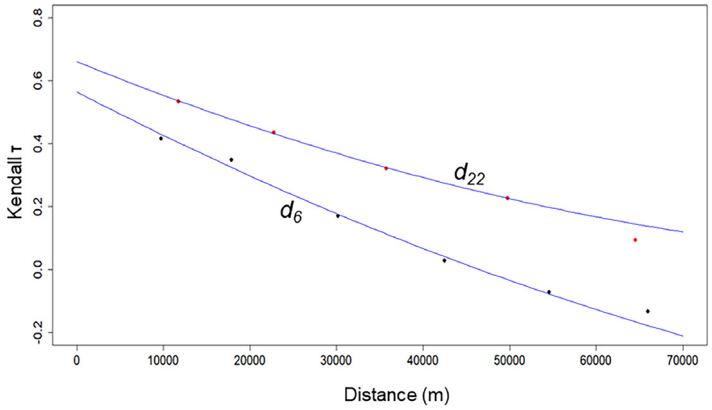

Fig. 5. Kendall's τ is obtained using observations at 174 locations on d_6 and d_{22}. A polynomial function is fitted to obtain τ at each distance. The parameters of five spatial bivariate copulas are then estimated by maximum likelihood. The best fitting copula is selected according to the lowest AIC values at each bin.

temperature (Fig. 7). We applied the cokriging in this study based upon the proportional model using the same variance and covariance functions. Gaussian covariance functions were fitted to cross variograms obtained for air temperature and its covariates.

4.2. Evaluation and comparison

For the leave-*k*-out cross-validation, we took $k = 2$ due to low number of the weather stations. The spatial copula interpolator including covariates resulted into the lowest MAE using different covariates at the two days (Table 4). The ES shows that the spatial copula interpolator including covariates improved predictions of the mean air temperature with 58% comparing with the cokriging predictor (Table 4). In addition, cross-validation showed that the use of LAI as a covariate resulted into more precise predictions.

4.3. Prediction

For making predictions, we considered $n = 8$ nearest neighbours and three covariates at d_6. The mixed copula interpolator was able to capture extremes (Fig. 8b) in contrast to the spatial copula

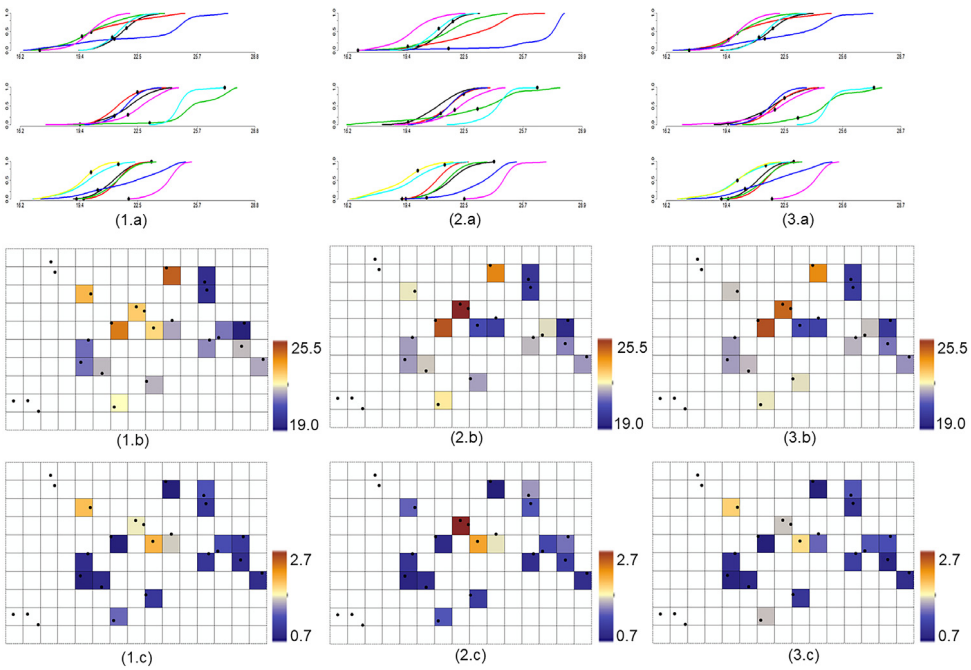

Fig. 6. (a) The conditional cumulative probabilities $F(X|.)$ for 19 weather stations and the spatial variation of (b) mean and (c) standard deviation of the conditional distributions of the predictions on d_6 obtained by: (1) the spatial copula interpolator including covariates, (2) the mixed copula interpolator, (3) the spatial copula interpolator. The observed values in the conditional cumulative distributions are denoted by black dots.

interpolator including covariates (Fig. 8a). The vegetated and non-vegetated areas based upon LAI, and highest and lowest elevated areas (Fig. 3) are interesting areas to consider to which degree the new interpolation methods take the spatial variation of covariates into account. The new methods resulted in the most heterogeneous map and, visually, more realistic spatial patterns than the kriging methods and the spatial copula interpolator, whereas the latter was more heterogeneous (Fig. 8c) than the map obtained by the kriging predictors (Figs. 8d and 7e). We further note that a low number of spatial bins leads to unrealistic spatial patterns as shown in Fig. 8.

The boxplots (Fig. 9) shows that copula-based methods well represent the mean values of the observations. The issue of failing to represent extremes using copula-based methods is related to low number of observations in the tails of the marginal distribution. The ranges of 95% PIW for copula-based methods are equal to [0.6, 12.6] °C whereas for kriging methods are equal to [4.9, 6.3] °C (Fig. 10).

The spatial variation of the mean and standard deviation of the conditional distributions (Fig. 6) shows that there is no reason to assume any lack of homogeneity. In fact, any pattern in the standard deviations would indicate such lack of homogeneity. Values, however, are relatively low as compared to the standard deviation of the observations (2.9 ° C). A few relatively high values occur in the centre of the study area. These are caused by the presence of extreme values at locations covered by the same pixels.

5. Discussion

Two dependences were characterized in the spatial interpolation of a weather variable: spatial variability and dependency with other variables, i.e., co-variability. We developed two methods based upon spatial copulas of air temperature, and non-spatial dependences between air temperature and

Please cite this article in press as: Alidoost F., et al., Copula-based interpolation methods for air temperature data using collocated covariates. Spatial Statistics (2018), https://doi.org/10.1016/j.spasta.2018.08.003.

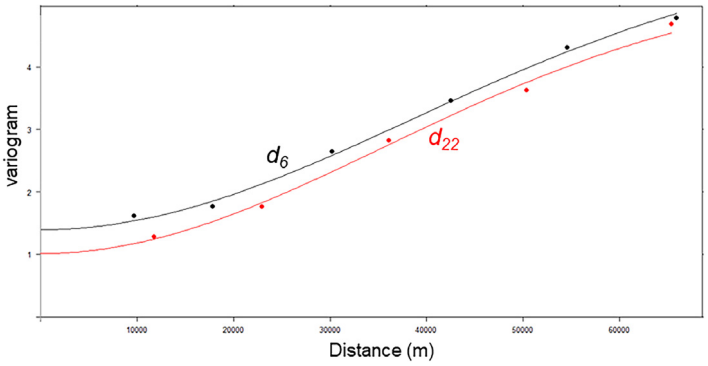

Fig. 7. The variogram obtained on d_6 and d_{22} for the same number of spatial bins as the correlogram. A Gaussian variogram model is fitted.

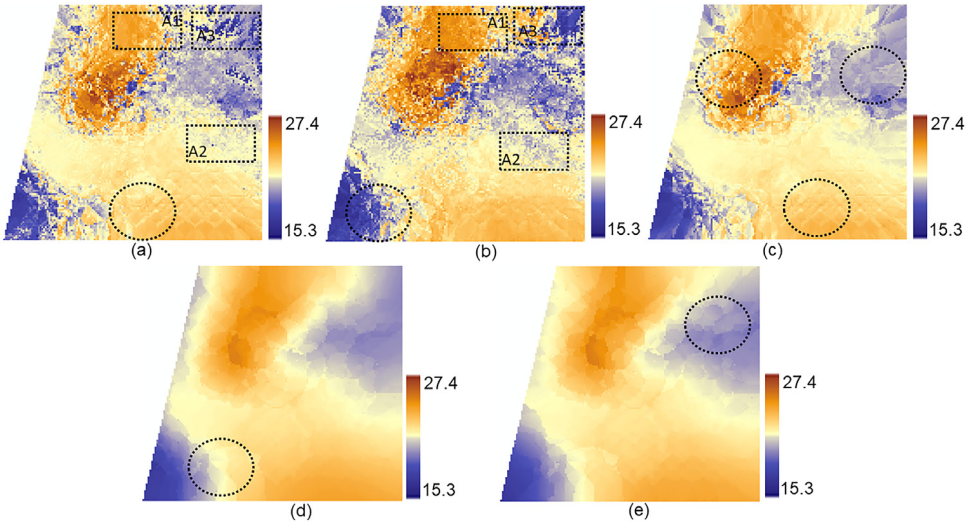

Fig. 8. Daily mean air temperature predicted at 1 km spatial resolution on d_6 based upon (a) the spatial copula interpolator including covariates, (b) the mixed copula interpolator, (c) the spatial copula interpolator, (d) the ordinary kriging predictor, (e) the cokriging predictor for a neighbourhood of eight locations. The circled areas denote squares as artefacts that represent the unrealistic spatial patterns. The areas A1, A2 and A3 as shown in Fig. 6, are examples where the co-variability becomes apparent in the results of the new methods.

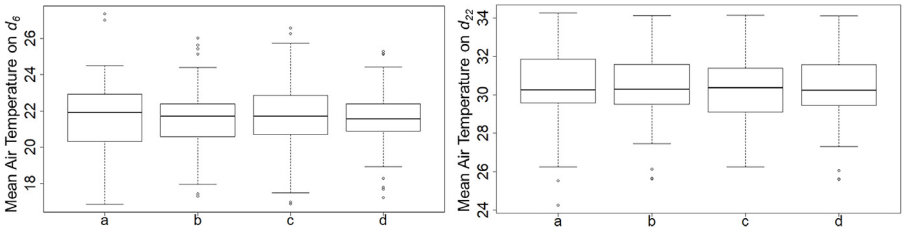

Fig. 9. Boxplots comparing the observations (a) with predicted values by: (b) the spatial copula including three covariates, (c) the mixed copula, and (d) the spatial copula interpolator. Here, observations are a combination of bias corrected values and measurements from the weather stations on d_6 and d_{22}.

Please cite this article in press as: Alidoost F., et al., Copula-based interpolation methods for air temperature data using collocated covariates. Spatial Statistics (2018), https://doi.org/10.1016/j.spasta.2018.08.003.

Table 4
Cross-validation expressed as the mean absolute error (MAE) obtained by the spatial copula interpolator using covariates, the mixed copula interpolator, the spatial copula interpolator, the ordinary kriging predictor, and the cokriging predictor. The leave-two-out cross validation is done for combinations of the covariates i.e. elevation (E), land surface temperature (LST) and Leaf area index (LAI) at two days. To compare the five methods, an error score (ES) is obtained based upon MAE for each method. The smallest ES indicates the best interpolator.

Day	Covariate	Spatial copula interpolator using covariates	Mixed copula interpolator	Spatial copula interpolator	Ordinary kriging	Cokriging
	E	1.550	1.669	1.555	1.597	1.598
	LAI	1.503	1.525	1.555	1.597	1.595
	LST	1.557	1.611	1.555	1.597	1.599
6	E, LAI	1.480	1.633	1.555	1.597	1.596
	E, LST	1.529	1.654	1.555	1.597	1.598
	LST, LAI	1.531	1.735	1.555	1.597	1.599
	E, LST, LAI	1.551	1.833	1.555	1.597	1.597
	E	1.390	1.365	1.378	1.328	1.330
22	LAI	1.291	1.308	1.378	1.328	1.328
	E, LAI	1.301	1.360	1.378	1.328	1.331
ES		**15**	41	28	30	36

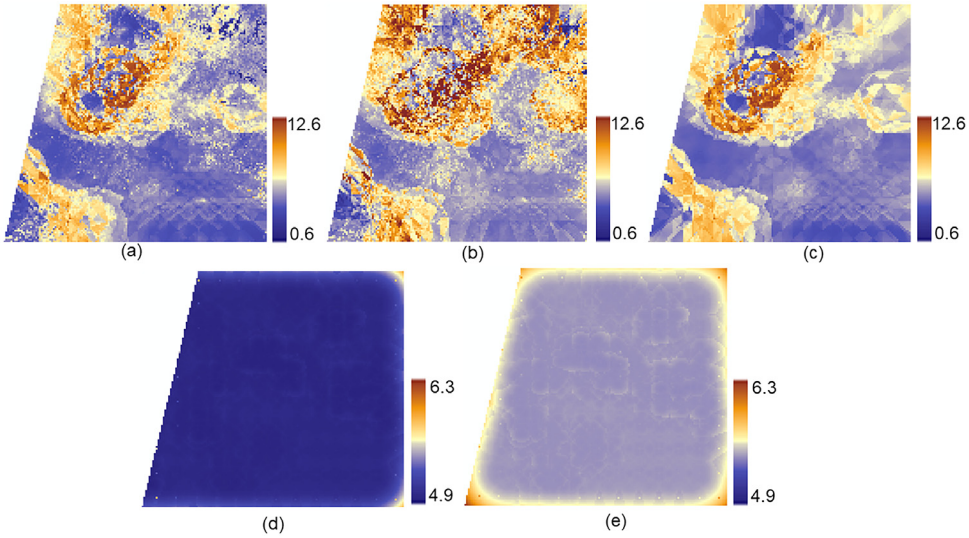

Fig. 10. 95% prediction interval widths (PIW) for each interpolation method on d_6, (a) the spatial copula including three covariates, (b) the mixed copula, (c) the spatial copula, (d) the ordinary kriging, and (e) the cokriging. The spatial copula interpolator resulted in the lowest uncertainty among copula-based methods. Ordinary kriging has smaller PIWs and is based upon assuming a Gaussian joint distribution.

its collocated covariates. The multivariate distributions are decomposed into bivariate copulas using vine structures that are generally well understood and can be estimated in a straightforward way.

The new methods provide more information about the uncertainty when interpreting the spatial variability of the PIW. We proposed to estimate the empirical marginal distribution that describes the statistical properties of daily air temperature without the knowledge of the theoretical form of the family's distribution function. The marginal distribution is, however, still assumed to be stationary. The local marginal distribution at an unvisited location (Heißerer et al., 2016) might help to improve the prediction as well as the PIW.

Please cite this article in press as: Alidoost F., et al., Copula-based interpolation methods for air temperature data using collocated covariates. Spatial Statistics (2018), https://doi.org/10.1016/j.spasta.2018.08.003.

We treated the available observations from weather stations as benchmarks during cross-validation, but we realized that the quality of measurements differs at each station. For example, stations 11 and 13 represent high extremes relative to other stations at the same day. A time series analysis of the air temperature (not shown) revealed that the quality of measurements at those stations is low. In particular, the correction for bias in the reanalysis weather data and the retrieval of covariates from remote sensing images are uncertain. A hierarchical model may be further explored to include uncertainty aspects of those observations.

We used the AIC to select the suitable copula family. The selection of families, however, depends upon the number of observations and further research is needed to develop strategies for selection optimization. Although several copula families can be found in the literature, we use five families because obtaining the inverse of the conditional copula distribution may lead to computational limitations. In addition, as all five families were symmetric, alternative families can be investigated.

For the interpolation of air temperature in a data-scarce area, we selected three covariates LST, LAI and elevation that were retrieved from remote sensing images. Our study has shown that the covariates can easily be included as additional information in estimating the joint distribution, thus allowing for a richer dependence structure. The copulas are generally able to describe both spatio-temporal and non-spatial dependences. A practical advantage of our methods is that we can analyse the joint behaviour of more than one covariate and their effects on the spatial variability of the daily air temperature locally. The availability of bias corrected values and the covariates derived from remote sensing images are, however, limitations for applying the methods on daily scales.

In order to provide a scenario that can be used to evaluate the new methods with less likely uncertain observations, we set up an experiment using the Meuse dataset (Pebesma, 2004). The leave-*k*-out cross validation showed that the average MAE values for mixed copula interpolator using Meuse variables zinc, lead, copper and cadmium were equal to 95.3, 33.7, 7.4 and 1.0, whereas for the cokriging predictor they were equal to 173.2, 55.8, 12.8 and 1.8, respectively. Further applications of the new methods in other case studies including simulation based information should provide more insight on these methods in the future.

We see several ways to further extend the current study. First, we applied the new methods in a data-scarce area, and we aimed to highlight the potential and the use of the methods for a larger dataset as well. Further comparison to other interpolation methods (Kilibarda et al., 2014) might help to assess the performance of the new methods. Second, essentially, we used the combination of the reanalysis weather data with a coarse spatial resolution and measurements from weather stations to predict mean air temperature at a higher spatial resolution. Such integration of bias correction and interpolation can be further investigated as a copula-based downscaling method. Third, in this study, we selected the number and type of covariates, the number of nearest neighbours, and the number of spatial bins in both variogram and correlogram based upon our experience. A more generally applicable sensitivity analysis might help to show the effects of these parameters on the results.

Based upon the cross-validation, the width of the prediction interval and visual inspection, we conclude that new methods allow describing both spatial variability and co-variability between weather variables and covariates using multivariate joint distributions. In addition, the use of LAI as a covariate in the interpolation of the mean air temperature reduces the uncertainties.

Acknowledgements

The authors thank the kind cooperation of the European Centre for Medium-Range Weather Forecasts (ECMWF) for providing weather forecast data, the U.S. Geological Survey for Landsat 8 images, the Consortium for Spatial Information (CGIAR-CSI) for SRTM 90 m database, and the SAJ Consulting firm in Iran for providing weather stations data.

References

Aalto, J., Pirinen, P., Heikkinen, J., Venäläinen, A., 2013. Spatial interpolation of monthly climate data for finland: comparing the performance of kriging and generalized additive models. Theor. Appl. Climatol. 13.

Aas, K., Czado, C., Frigessi, A., Bakken, H., 2009. Pair-copula constructions of multiple dependence. Insurance Math. Econom. 44, 182–198.

Akaike, H., 1974. A new look at the statistical model identification. IEEE Trans. Automat. Control 19, 716–723. http://dx.doi.org/10.1109/TAC.1974.1100705.

Alidoost, F., Stein, A., 2016. Correction of daily ECMWF air temperature data based on copula concept. In: 'SAIL35, Eye on Foliage' International Symposium.

Allen, R.G., Pereira, L.S., Raes, D., Smith, M., 1998. FAO Irrig. Drain. Pap. 56.

Bárdossy, A., Li, J., 2008. Geostatistical interpolation using copula. Water Resour. Res. 44, 15. http://dx.doi.org/10.1029/2007WR006115.

Brechmann, E.C., Schepsmeier, U., 2013. Modeling dependence with C- and D-Vine Copula: The R Package CDVine. J. Stat. Softw. 52, 1–27. http://dx.doi.org/10.18637/jss.v052.i03.

Conover, W.J., 1971. Practical Nonparametric Statistics. John Wiley & Sons, New York, pp. 295–314.

Cressie, N., 1993. Spatial Prediction and Kriging in: Statistics for Spatial Data. John Wiley & Sons, Canada, pp. 105–110.

Demarta, S., McNeil, A.J., 2005. The t copula and related copula. Int. Stat. Rev./Rev. Int. Stat. 73, 111–129.

Durai, V.R., Bhradwaj, R., 2014. Evaluation of statistical bias correction methods for numerical weather prediction model forecasts of maximum and minimum temperatures. Nat. Hazards 73, 1229–1254.

Durocher, M., Chebana, F., Ouarda, T.B.M.J., 2016. On the prediction of extreme flood quantiles at ungauged locations with spatial copula. J. Hydrol. 533, 523–532.

Gräler, B., 2014. Developing spatio-temporal copula. In: Earth Sciences. Westfälische Wilhelms-Universität Münster, Germany, p. 139.

Gräler, B., Pebesma, E., 2011. The pair-copula construction for spatial data: a new approach to model spatial dependency. Procedia Environ. Sci. 7, 206–211.

Haslauer, C.P., Heißerer, T., Bárdossy, A., 2016. Including land use information for the spatial estimation of groundwater quality parameters –2 Interpolation methods, results, and comparison. J. Hydrol. 535, 699–709. http://dx.doi.org/10.1016/j.jhydrol.2016.01.054.

Heißerer, T., Haslauer, C.P., Bárdossy, A., 2016. Including land use information for the spatial estimation of groundwater quality parameters –1 Local estimation based on neighbourhood composition. J. Hydrol. 535, 688–698. http://dx.doi.org/10.1016/j.jhydrol.2015.12.049.

Hengl, T., Heuvelink, G.B.M., Tadić, .M.P., Pebesma, E.J., 2012. Spatio-temporal prediction of daily temperatures using time-series of MODIS LST images. Theor. Appl. Climatol. 107, 13.

Jarvis, A., Reuter, H.I., Nelson, A., Guevara, E., 2008. Hole-filled seamless SRTM data V4. International Centre for Tropical Agriculture (CIAT).

Jiménez-Muñoz, J.C., Sobrino, J.A., Skoković, D., Mattar, C., Cristóbal, J., 2014. Land surface temperature retrieval methods from landsat-8 thermal infrared sensor data. IEEE Geosci. Remote Sensing Lett. 11, 1840–1843.

Joe, H., 1993. Parametric families of multivariate distributions with given margins. J. Multivariate Anal. 46, 262–282.

Kilibarda, M., Hengl, T., Heuvelink, G.B.M., Gräler, B., Pebesma, E., Tadić, .M.P., Bajat, B., 2014. Spatio-temporal interpolation of daily temperatures for global land areas at 1 km resolution. J. Geophys. Res.: Atmos. 119, 20.

Kojadinovic, I., Yan, J., 2010. Modeling multivariate distributions with continuous margins using the copula R package. J. Stat. Softw. 34, 1–20.

Kutner, M.H., Nachtsheim, C.J., Neter, J., Li, W., 1996. Applied Linear Statistical Models, fifth ed. McGraw-Hill/Irwin, New York, p. 1415.

Li, J., 2010. Application of copula as a new geostatistical tool. (Ph.D.), Institut für isserbau der Universität Stuttgart, Universität Stuttgart, p. 161.

Manner, H., 2007. Estimation and Model Selection of Copula with an Application to Exchange Rates. Maastricht research school of Economics of TEchnology and ORganizations, Universiteit Maastricht, pp. 1–37.

Mao, G., Vogl, S., Laux, P., Wagner, S., Kunstmann, H., 2015. Stochastic bias correction of dynamically downscaled precipitation fields for germany through copula-based integration of gridded observation data. Hydrol. Earth Syst. Sci. 19, 1787–1806. http://dx.doi.org/10.5194/hess-19-1787-2015.

Nelsen, R.B., 2006. An Introduction to Copula. Springer, United States of America, p. 276.

Oden, N.L., 1984. Assessing the significance of a spatial correlogram. Geogr. Anal. 16, 1–16. http://dx.doi.org/10.1111/j.1538-4632.1984.tb00796.x.

Parmentier, B., McGill, B.J., Wilson, A.M., Regetz, J., Jetz, W., Guralnick, R., Tuanmu, M.N., Schildhauer, M., 2015. Using multi-timescale methods and satellite-derived land surface temperature for the interpolation of daily maximum air temperature in Oregon. Int. J. Climatol. 35, 3862–3878. http://dx.doi.org/10.1002/joc.4251.

Pebesma, E.J., 2004. Multivariable geostatistics in S: the gstat package. Comput. Geosci. 30, 683–691. http://dx.doi.org/10.1016/j.cageo.2004.03.012.

Persson, A., 2013. User guide to ECMWF forecast products, Livelink 4320059.

Sharifi, M., 2013. Development Of Planning and Monitoring System Supporting Irrigation Management in the Qazvin Irrigation Network. SAJ Co. Tehran, Iran.

Silverman, B.W., 1986. Density Estimation for Statistics and Data Analysis. Chapman and Hall/CRC, UK.

Stein, A., Corsten, L.C.A., 1991. Universal kriging and cokriging as a regression procedure. Biometrics 47 (2), 575–587.

Wu, T., Li, Y., 2013. Spatial interpolation of temperature in the United States using residual kriging. Appl. Geogr. 44, 112–120.

Zanter, K., 2016. Landsat 8 (L8) Data Users Handbook. Department of the Interior US Geological Survey, EROS Sioux Falls, South Dakota.

Please cite this article in press as: Alidoost F., et al., Copula-based interpolation methods for air temperature data using collocated covariates. Spatial Statistics (2018), https://doi.org/10.1016/j.spasta.2018.08.003.