BioOne COMPLETE

GIS-based Tests for Quality Control of Meteorological Data and Spatial Interpolation of Climate Data

Authors: Chiu, Ching-An, Lin, Po-Hsiung, and Lu, King-Cherng

Source: Mountain Research and Development, 29(4) : 339-349

Published By: International Mountain Society

URL: https://doi.org/10.1659/mrd.00030

BioOne Complete (complete.BioOne.org) is a full-text database of 200 subscribed and open-access titles in the biological, ecological, and environmental sciences published by nonprofit societies, associations, museums, institutions, and presses.

Your use of this PDF, the BioOne Complete website, and all posted and associated content indicates your acceptance of BioOne's Terms of Use, available at www.bioone.org/terms-of-use.

Usage of BioOne Complete content is strictly limited to personal, educational, and non - commercial use. Commercial inquiries or rights and permissions requests should be directed to the individual publisher as copyright holder.

BioOne sees sustainable scholarly publishing as an inherently collaborative enterprise connecting authors, nonprofit publishers, academic institutions, research libraries, and research funders in the common goal of maximizing access to critical research.

Downloaded From: https://complete.bioone.org/journals/Mountain-Research-and-Development on 28 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

Mountain Research and Development (MRD)
An international, peer-reviewed open access journal
published by the International Mountain Society (IMS)
www.mrd-journal.org

MountainResearch
Systems knowledge

GIS-based Tests for Quality Control of Meteorological Data and Spatial Interpolation of Climate Data

A Case Study in Mountainous Taiwan

Ching-An Chiu[1], Po-Hsiung Lin[2], and King-Cherng Lu[3]*
** Corresponding author: cachiu@nchu.edu.tw*
[1] Experimental Forest, National Chung Hsing University, 250 Kuo Kuang Road, Taichung 40227, Taiwan, Republic of China
[2] Department of Atmospheric Sciences, National Taiwan University, Number 1, Section 4, Roosevelt Road, Taipei 10617, Taiwan, Republic of China
[3] Department of Forestry, National Chung Hsing University, 250 Kuo Kuang Road, Taichung 40227, Taiwan, Republic of China

Open access article: please credit the authors and the full source.

Constructing climate layers is more difficult and important in mountainous areas as a result of sparse meteorological stations and complex topography. This requires a 2-stage process: quality control of meteorological data and spatial interpolation of climate data. For this article, unscreened metadata and observed data were collected from all stations in Taiwan for the period 1961–2002. A quality-control procedure based on a geographic information system (GIS) allowed us to reject 13.5% of stations because of missing or erroneous metadata and filter out 8.3% of the observed data because of extreme errors or unreasonable temporal sequence and spatial patterns. After applying the quality-control procedure, the monthly mean temperature and total monthly precipitation were calculated as spatial interpolation sampling points. We evaluated the performance of 6 kriging-based spatial interpolation methods with regard to their errors by cross-validation. For interpolating the monthly mean temperature, the strong relation between temperature and elevation led us to favor modified residual kriging. For interpolating the total monthly precipitation, log-transformed kriging was chosen for practical reasons (steadier and simpler). We compared our product layers with pre-existing climate layers. The overall spatial patterns of these layers were similar, except for certain extremes in the mountains. Consequently, the GIS-based approaches presented here could help in rapid construction of adequate climate layers for regions with unconfirmed data.

Keywords: *Quality control (QC); spatial interpolation (SI); geographic information system (GIS); kriging; climate data; Taiwan.*

Peer-reviewed: *August 2009* ***Accepted:*** *September 2009*

Introduction

Climate is one of the most important environmental factors in terrestrial ecosystems (eg Tuhkanen 1980; Walter 1985). Many environmental and ecological models require spatially continuous climate data, usually in the form of interpolated grids (eg Franklin 1995; Guisan and Zimmermann 2000). To construct a spatial climate layer, two critical issues must be considered: quality control (QC) of meteorological data and spatial interpolation (SI) of climate data.

The QC of meteorological data has to be undertaken prior to any data analysis to eliminate erroneous values (Štěpánek et al 2009); it is an essential prerequisite for the SI of climate data. It is well known that some datasets recorded in meteorological stations may contain errors as a result of typing, transmitting, coding, and even missing data (eg Meek and Hatfield 1994). One must pay attention to the correctness of station metadata (coordinates, elevation, etc) as well as to the quality of meteorological observed data (air temperature, precipitation, etc). The incorrectness of station metadata and observed data is a significant problem in Taiwan as well as elsewhere (eg Peterson et al 1998).

It is necessary to explore which SI method is the most appropriate in each situation as no single one is optimal for all regions and data (Price et al 2000). Many SI methods such as distance-weighting algorithms, geostatistics, splines, and multivariate analysis have been rapidly developed because of the advancement of computer software and hardware (eg Boer et al 2001; Marquínez et al 2003). The geostatistical method known generically as kriging is a common approach because it takes into account spatial relations between the experimental data (ie sample points) and quantifies the uncertainty of the model (Martinez-Cob 1996; Moral 2009).

In the last few decades, numerous attempts have been made to construct spatial climate layers (Stahl et al 2006).

 http://dx.doi.org/10.1659/mrd.00030 © 2009 by the authors

Downloaded From: https://complete.bioone.org/journals/Mountain-Research-and-Development on 28 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

The topic is more difficult and challenging in mountainous areas because of sparse stations and complex topography (eg Benavides et al 2007; Guan et al 2009). The climate layers produced are relevant in mountain environments to assess impacts of climate change on the distribution and diversity of species (eg Thuiller 2008; Ashiq et al 2009).

Recently, the QC and SI of meteorological and climate elements have often been linked to a geographic information system (GIS) (eg Ashiq et al 2009; Guler et al 2009; Štěpánek et al 2009). This article discusses the construction of a high-resolution climate grid for mountainous Taiwan, with 2 specific goals. The first was to test the QC procedure for meteorological data, and the second was to test the SI technique through GIS-based approaches working with a digital elevation model (DEM). The performance of this combined approach was then evaluated and discussed using Taiwanese meteorological data.

Material and methods

Study area and data collection

Taiwan is a mountainous island at the fringe of East Asia's continental shelf, covering about 36,000 km^2 and with a complex topography ranging from 0–3952 m, with only 31.3% of its area below 100 m. The Taiwanese climate is controlled mainly by orographic relief and by an alternation between the summer southwest monsoon and the winter northeast monsoon (Su 1984a). Previous studies (eg Su 1984b, 1985) have been based on individual analysis of climate stations and empirical inference.

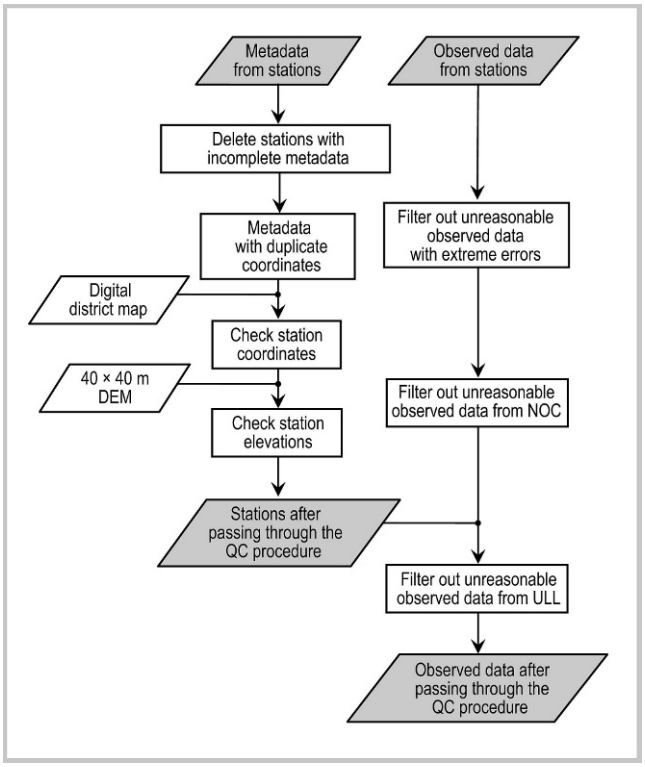
This article is based on 2 raw datasets describing all the meteorological stations in Taiwan: (1) the metadata, including each station's code, name, coordinates, elevation, and address and (2) the observed data, including daily mean temperature (Td) and total precipitation (Pd). Raw metadata and observed data were collected from a total of 1728 stations: 33 permanent specialized stations (SPs), 362 unmanned remote stations (REs) managed by the Central Weather Bureau, and 1333 unspecialized cooperator stations (COs). These data were archived in a Central Weather Bureau database over a 42-year period (1961–2002); the total number of Td and Pd records was thus 13,051,457. Generally speaking, only SPs were considered to provide very high-quality data. We also made use of a digital township map and a 40-m resolution DEM. The coordinates of all layers are transformed by ArcGIS to the 2° zone Transverse Mercator projection with TWD67 datum generally used in Taiwan.

Methods

Quality control (QC) of meteorological data

First we looked for stations with doubtful metadata and unreasonable observed data. A summary of the QC procedure for meteorological data is shown in Figure 1 and is described in detail here:

FIGURE 1 Flow chart of the quality control (QC) procedures for metadata and observed data from meteorological stations. The criteria and explanations of "continuous no-observed-change with time limits" (NOC) and "upper/lower limits" (ULL) are given in the text.



Stations with doubtful metadata

1. Incomplete metadata: They could not be used in the subsequent climate SI procedure.
2. Different station codes with identical coordinates: One of them was transferred to another organization at some point in the past. Both their metadata and observed data were merged.
3. Checking the correctness of station coordinates: If the coordinate point lay more than 3 km (value based on experience) outside the township of its address in the metadata, the coordinate was considered to be erroneous. Consequently, such stations were deleted.
4. Checking the correctness of station elevations: If a station's coordinate was confirmed, its elevation in the metadata was compared with its altitude in the DEM. If the difference was greater than 200 m (value based on experience), the elevation was considered to be erroneous. Consequently, such stations were deleted.

Unreasonable observed data

1. Extreme errors: (a) All Td records below −15°C or above 36°C were filtered out because all the historical Td records from reliable SPs range from −12–33°C.

Downloaded From: https://complete.bioone.org/journals/Mountain-Research-and-Development on 28 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

(b) Any Pd records below 0 mm (not including 0 mm) or above 2000 mm were filtered out because the historical Pd records from SPs indicated a maximum value of 1135 mm.

2. Unreasonable temporal sequence of observed data: Any record that reported an identical observed value over 3 consecutive days was filtered out, except for sequences in which Pd was equal to 0 mm. This criterion is denoted as "continuous no-observed-change with time limits" (NOC) by Meek and Hatfield (1994).
3. Unreasonable spatial pattern of observed data: (a) Td values of each station were compared with the average Td of the 5 vertically nearest stations within 1000 m of vertical and 70 km of horizontal distance. If the difference was more than 7°C, the Td record was considered unreasonable and was deleted. (b) Pd values of each station were compared with the average Pd of the 5 horizontally nearest stations within a vertical range of 300 m. If the difference was more than 300 mm, the Pd record was considered unreasonable and was deleted. This criterion is referred to as the "upper/lower limits" (ULL) on the spatial variation of meteorological data.

Spatial interpolation of climate data

We summarize 4 issues from the climate spatial interpolation (SI) procedure that we followed.

Selecting climate data

In the present work, we analyzed only 2 climate parameters: monthly mean temperature (Tm) and total monthly precipitation (Pm). Tm and Pm were calculated for all stations that passed the QC procedure. In the SI procedure for climate data, Tm and Pm were the dependent variables; for each station's elevation, X and Y coordinates (abbreviated as E, X, and Y, respectively; units in m) were the independent variables.

SI methods

Geostatistical method or kriging has several advantages and is widely used in climate mapping (Benavides et al 2007; Moral 2009). This article examines 6 variants of the ordinary kriging technique: ordinary kriging (OK), detrended kriging (detOK), anisotropic kriging (aniOK), cokriging (COK), modified residual kriging (resOK), and log-transformed kriging (logOK). The expediency of these methods is dependent on the characteristics of environment and data (eg Price et al 2000). When the correlation between environment and data is strong, for example between elevation and temperature, resOK could be a better method (Stahl 2006). Readers interested in a comprehensive description of these methods can refer to the literature (eg Goovaerts 1997). The 6 kriging variants in this paper were implemented using the ArcGIS 8.1 Spatial Analyst extension and Geostatistical Analyst extension (Johnston et al 2001) and SPSS 11.0 statistical software.

Assessing prediction error

The most common method for assessing prediction errors of different SI methods is leave-one-out cross-validation (eg Benavides et al 2007), using a single observed data element from the original dataset as the validating data and the remaining observed data as the training data, until each observed data element has been used once as the validating data. The deviations were summarized here by root-mean-square error (RMSE); other prediction error indices were considered. For a detailed presentation on predictive assessments, interested readers should again refer to textbooks (eg Goovaerts 1997).

Results

QC of meteorological data

QC of metadata

The first QC step, checking on incomplete metadata, found 71 COs that lacked coordinates or elevation. The second QC step, finding different station codes with identical coordinates, found 176 COs. The metadata and observed data for these duplicate stations were merged. In the third QC step, a few erroneous COs were found when checking the correctness of station coordinates. When the elevation of metadata was checked in the fourth QC step, we found many doubtful COs and several doubtful REs. Ignoring the merged duplicate stations, we distinguished 233 COs (13.5% of all) with doubtful metadata in the QC procedure. By contrast, all SPs and REs administered by the professional Central Weather Bureau passed the metadata QC procedure.

QC of observed data

For the first step, the QC rule for screening out extreme errors allowed us to detect instrument malfunction codes (eg −9999.5) as well as any Td and Pd that lay far outside the historical range observed by SPs, such as 50.1°C at station R2F340 on 1985/01/30. The continuous NOC records found in the NOC step, such as 888.0 mm at station F2N480 on 26–28 March 1986, were filtered out. Station C0T870 recorded 1702.6 mm on 31 May 1995. This case failed in theULL step. Most of the records filtered out by the ULL step contained more than 400 mm of precipitation. A total of 1,084,252 records (8.3% of all observed data) were rejected in the QC procedure. Most of the records belonged to COs and some belonged to REs, but none belonged to SPs. Filtering out these doubtful data allowed us to raise the accuracy of Tm and Pm as well as the interpolation of these parameters.

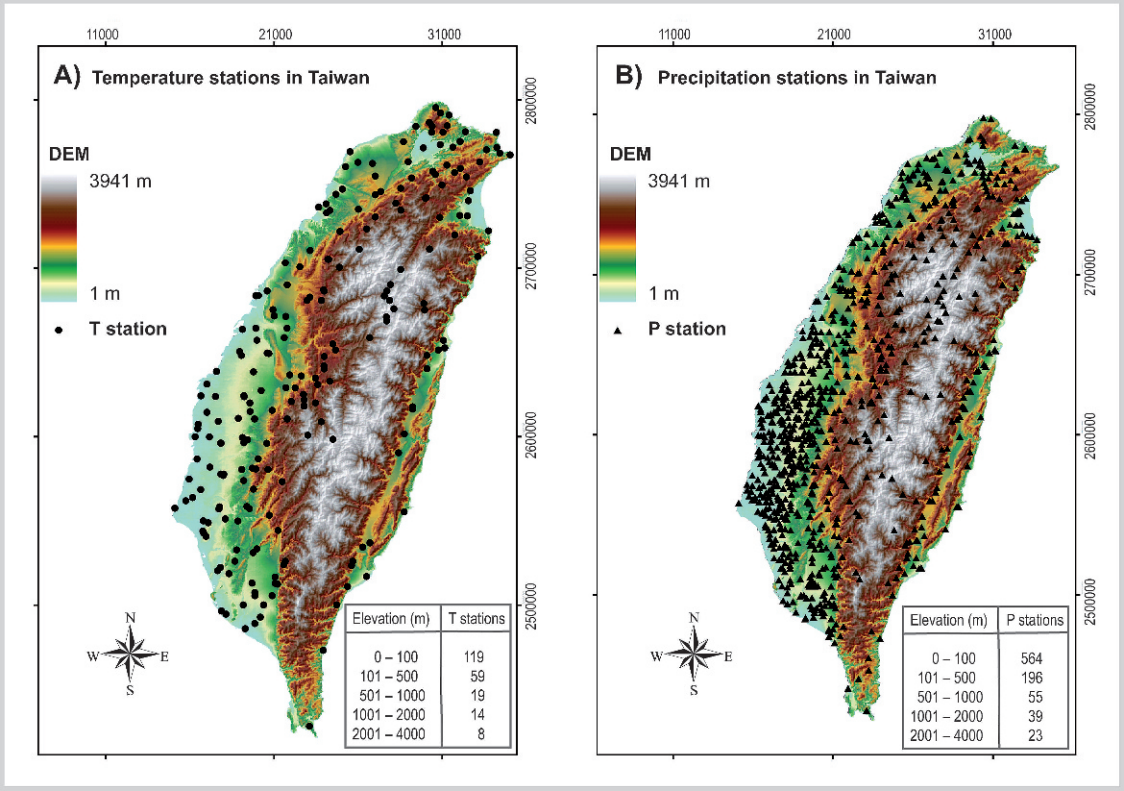
SI of climate data

Selection of climate variables

After assuring the quality of raw meteorological data, Tm and Pm were calculated using long-term daily observed data, from stations with a minimum length of at least 7 years and 12 years, respectively. The 219 selected long-

Downloaded From: https://complete.bioone.org/journals/Mountain-Research-and-Development on 28 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

FIGURE 2 Locations of selected and long-term stations from which data were used in the interpolations: (A) temperature (219 stations); (B) precipitation (877 stations). The marginal numbers are X and Y coordinates (in m) in 2-degree Transverse Mercator projection.



term temperature stations and 877 precipitation stations, shown in Figure 2, were strongly biased toward the lowlands (see statistics at the bottom right corner of Figure 2). Only 41 temperature stations (18.7% of all) and 117 precipitation stations (13.3% of all) were located above 500 m in mountainous area. The summary statistics of Tm and Pm are presented in Tables 1 and 2, respectively.

Interpolating temperature layers

Here we applied 6 different kriging methods to interpolate Tm layers. Each method, except for OK, used elevation as an additional independent variable because of the noticeable relation between temperature and elevation (Rolland 2003; see Table 1). The temperature lapse rate was 4.28–6.14°C per 1000 meters for different months and regions in Taiwan. Figure 3A tracked the RMSE of each method for 12 months. The minimum prediction error was obtained by resOK.

The adjusted coefficients of determination (adj. R^2) for the relation of Tm (abbreviated T1–T12 for January to December) with predictors E, X, and Y ranged from 0.013 (T8 with Y) to 0.963 (T6 with E, X, and Y). We selected the variables by stepwise backward elimination to determine

TABLE 1 Summary statistics of monthly mean temperature (Tm, such as T1 for January, units in °C). Min: minimum; Max: maximum; SD: standard deviation; *r*: correlation coefficient between Tm and elevation.

	Min	Max	Mean	SD	Skewness	*r*
T1	–1.29	20.07	15.17	3.18	–2.22	–0.90
T4	3.35	25.92	21.41	3.37	–2.20	–0.91
T7	7.68	30.46	26.65	3.50	–2.71	–0.98
T10	6.06	27.31	22.96	3.36	–2.29	–0.94

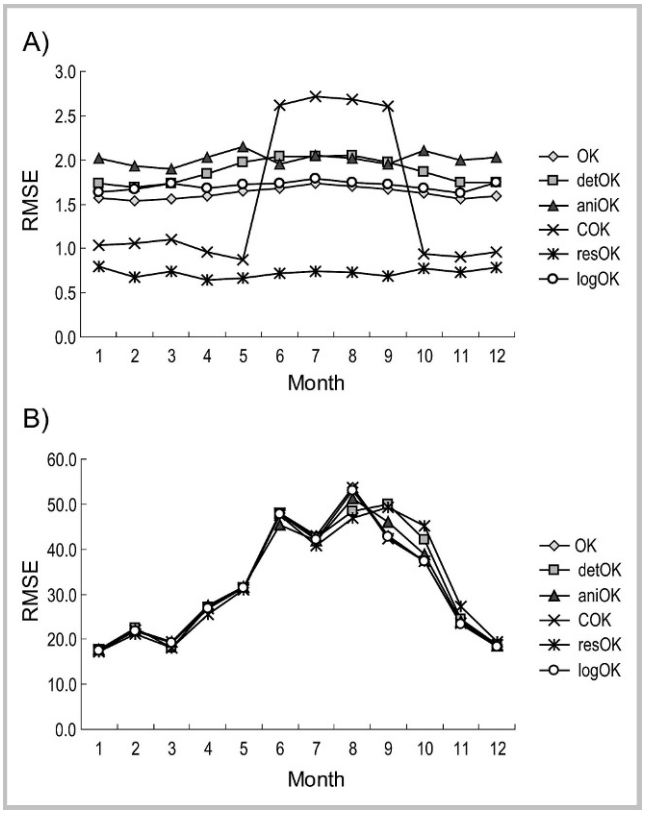
Downloaded From: https://complete.bioone.org/journals/Mountain-Research-and-Development on 28 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

TABLE 2 Summary statistics of total monthly precipitation (Pm, such as P1 for January, units in mm). Log-P: log-transformed values of Pm; Min: minimum; Max: maximum; SD: standard deviation; *r*: correlation coefficient between Pm and elevation.

	Min	Max	Mean	SD	Skewness	*r*
P1	9.56	486.65	48.39	51.75	3.38	0.28
P4	32.67	412.67	112.84	51.40	1.46	0.51
P7	80.33	694.04	293.07	111.43	0.74	0.15
P10	7.75	968.35	104.59	148.29	2.52	0.19
Log-P1	0.98	2.69	1.54	0.32	0.90	0.41
Log-P4	1.51	2.62	2.01	0.18	0.31	0.45
Log-P7	1.90	2.84	2.44	0.17	–0.22	0.16
Log-P10	0.89	2.99	1.71	0.49	0.60	0.34

the best multilinear regression formula (Table 3), such as T1 = 57.398 – 0.00461E + 0.00001038X – 0.0000163Y (adj. R^2 of 0.938). The resOK seemed to be the best formula for every Tm, explaining a significant amount of variation ($P < 0.01$). Thus, T1–T12 were interpolated by resOK and displayed through ArcGIS, as Figure 4 shows for T1 and T7. It was clear that the mean temperature varies principally with E, but there were slight variations with X and Y, except for July and August temperatures (see Table 3). According to SI results, temperatures in January ranged from –1.2–20.2°C (Figure 4A) and in July from 6.7–29.3°C (Figure 4B). The mean annual temperature was found to range from 4.0–25.0°C.

FIGURE 3 Root-mean-square error (RMSE) of Tm (A) and Pm (B) interpolation assessed by cross-validation for six methods (OK: ordinary kriging; detOK: detrended kriging; aniOK: anisotropic kriging; COK: co-kriging; resOK: modified residual kriging; logOK: log-transformed kriging).



Interpolating precipitation layers

Table 2 reveals that the distribution of Pm values for the 877 stations were right-skewed (Kolmogorov-Smirnov normality test, $P < 0.01$). This might hinder the interpolative accuracy. For this reason, Pm values were decimal-log-transformed to more closely approximate a normal distribution (see Table 2) and were also interpolated using the logOK method. The RMSE of Pm interpolation assessed by cross-validation for 6 methods is shown in Figure 3B, but no statistical difference was found between the different interpolations ($P < 0.05$). Here we used logOK to interpolate Pm layers not only because log-transformed normalized data can raise the predictive accuracy (Phillips et al 1992; Martinez-Cob 1996; Price et al 2000) but also for practical reasons. This option is discussed in the Discussion section below.

January to December precipitation (abbreviated as P1 to P12) layers were interpolated by logOK and laid out through ArcGIS. The value of P1 ranged from 10–486 mm, for example, while the value of P7 ranged from 103–573 mm. During the summer half-year (April to September), precipitation ranged from 854–2979 mm (Figure 5A) and was concentrated on the slopes windward of the southwest monsoon. In the Taiwan region, frequent typhoon events in the summer half-year affect most precipitation (Su 1984a). During the winter half-year (October to March), precipitation was found to range from 99–3138 mm (Figure 5B) and to be concentrated in

Downloaded From: https://complete.bioone.org/journals/Mountain-Research-and-Development on 28 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

FIGURE 4 Climate grid layers interpolated using resOK for the following: (A) January temperature; (B) July temperature. Note that different color scales have been used for the 2 layers.

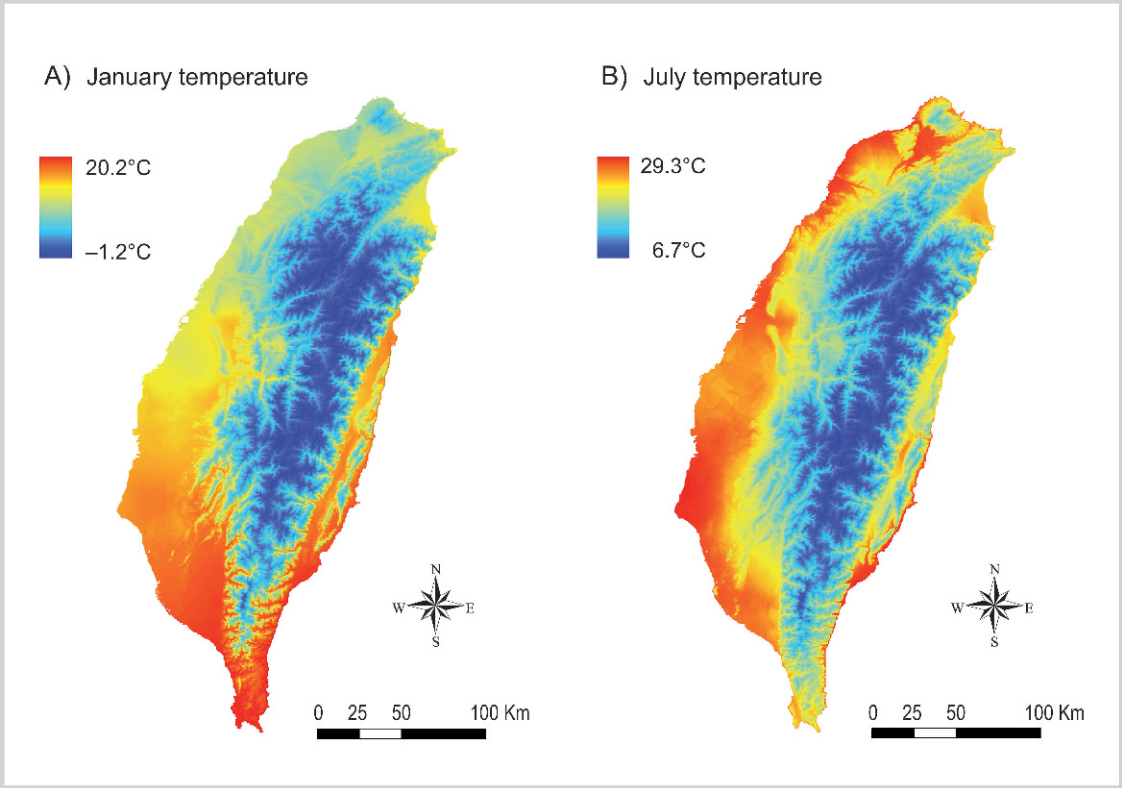
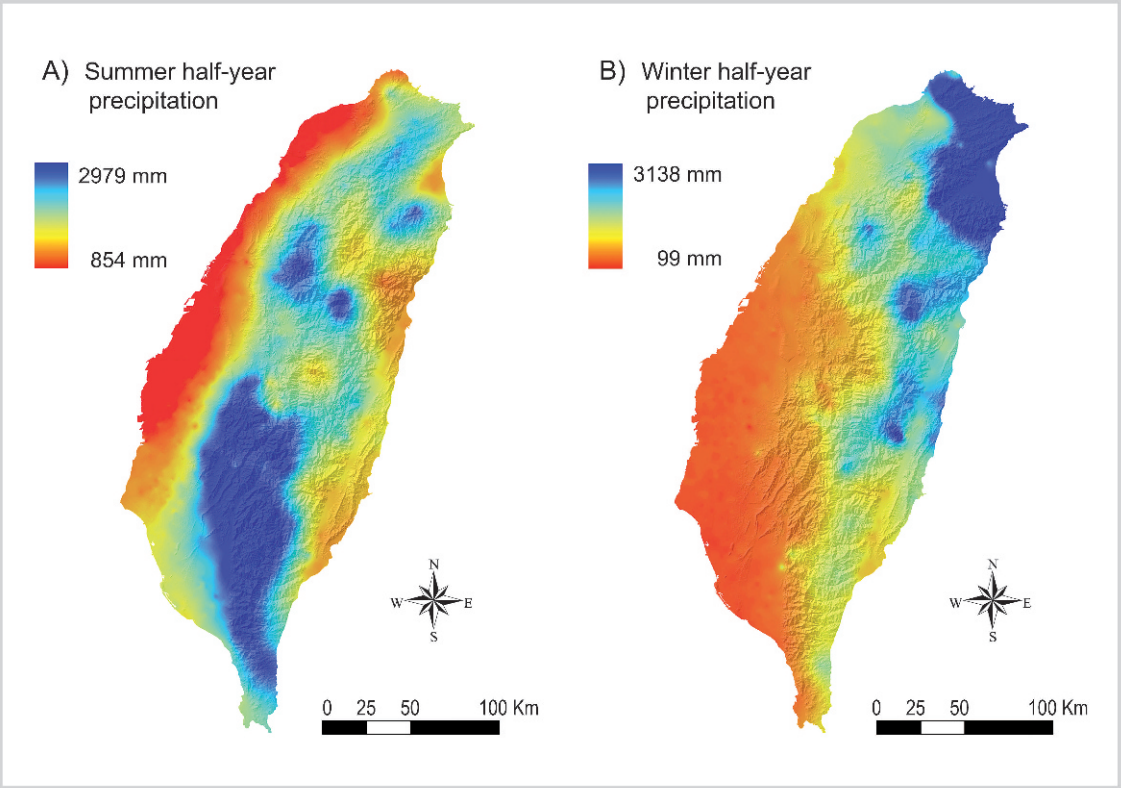


TABLE 3 Formulas for linear regression of Tm (Tm are dependent variables; E, X, and Y are predictive variables). X and Y were transformed into 2-degree Transverse Mercator projection with TWD67 datum; their range in Taiwan: 149,000 < X < 351,000; 2,422,000 < Y < 2,800,000 (see Figure 2A).

Tm	Formulas for linear regression	Adj. R^2
T1	57.398 – 0.00461E + 0.00001038X – 0.0000163Y	0.938
T2	64.407 – 0.00444E + 0.000008993X – 0.0000186Y	0.944
T3	66.225 – 0.00436E + 0.00000571X – 0.0000181Y	0.941
T4	56.226 – 0.00472E – 0.0000044X – 0.0000122Y	0.956
T5	46.342 – 0.00511E – 0.00000863X – 0.00000705Y	0.961
T6	34.397 – 0.00541E – 0.0000047X – 0.00000215Y	0.963
T7	28.622 – 0.00568E	0.958
T8	28.288 – 0.00565E	0.959
T9	32.747 – 0.00535E – 0.00000767X – 0.0000015Y	0.959
T10	44.389 – 0.00498E – 0.00000567X – 0.00000694Y	0.948
T11	49.22 – 0.00476E – 0.0000015X – 0.0000103Y	0.945
T12	52.256 – 0.00469E + 0.000005803X – 0.0000133Y	0.940

Downloaded From: https://complete.bioone.org/journals/Mountain-Research-and-Development on 28 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

FIGURE 5 Climate grid layers interpolated by logOK: (A) summer half-year precipitation; (B) winter half-year precipitation. Note that different color scales have been used for the 2 layers.



the region affected by prevalent northeast monsoon rains. The overall precipitation pattern was found to be markedly affected by Taiwanese topography and the alternating monsoons.

Discussion

QC of meteorological data

In the raw database, many COs (13.5% of all) were found to have doubtful metadata. Besides, 8.3% of all observed data—the majority belonging to amateur COs—were rejected by our QC procedure. The criterion of the NOC step is based on the concept of the time series recommended by Meek and Hatfield (1994). It is true that many unreasonable continuous values exist in the observed database and were found by our NOC criterion, but the time limit of NOC may be worth discussing further. Conversely, the criterion of the ULL step is based on the concept of the spatial relation of meteorological conditions. This idea was generated mainly by the regional meteorological signal (Rhoades and Salinger 1993; Štěpánek et al 2009). The rule is subjective. Because it is possible that these filtered out data are actual values (González-Rouco et al 2001), we increased the tolerance of the ULL step (ie by allowing more distant neighbors and increasing the possible difference). The rule of ULL should be improved in future work by rethinking the mechanism and scale of precipitation (Daly 2006) and the lapse rate of temperature (Rolland 2003). No single QC method alone was found adequate; only a combination of several methods for outlier detection led to satisfying results (see also Štěpánek et al 2009). Our QC procedures coped with the doubtful stations and the extremely erroneous observed data that can provide a basic reliability of meteorological data.

SI of climate data

Selection of climate variables

According to the World Meteorological Organization, data for a 30-year period are recommended because they provide stable and reproducible monthly means (Benavides et al 2007). The most applicable length of time of observed data is practically a compromise between climate stability (ie quality) and sample size (ie quantity). Longer periods of continuous observed data can provide a steadier climatological mean state but may leave too few sample stations to accurately interpolate climate layers. Thus, the minimum length of observation has more often been set empirically and depended on the variability of the study area, including 8–30 years (eg Hevesi et al 1992; Goovaerts

Downloaded From: https://complete.bioone.org/journals/Mountain-Research-and-Development on 28 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

TABLE 4 Adjusted R^2 for Pm linear regression (Pm are dependent variables and E, X, and Y are predictive variables; see Figure 2B). The highest adjusted R^2 are written in boldface.

	E	X	Y	EX	EY	XY	EXY
P1	0.078	0.591	0.413	0.591	0.443	**0.627**	0.628
P4	0.261	0.319	0.477	0.432	**0.636**	0.500	0.637
P7	0.022	0.066	0.418	0.131	0.487	0.461	**0.505**
P10	0.035	0.591	0.086	0.597	0.105	0.661	**0.671**

2000). We set the minimum length to 7 years for Tm and 12 years for Pm. This was a subjective and empirical decision based on a compromise between the longest observed length and the adequate density of sample stations (Ninyerola et al 2000). The relation between climate stability and observed length should be explored in future.

Interpolating temperature layers

The RMSEs of all methods in Figure 3A are approximately 1.58°C. This seems to be acceptable when compared with the results presented by Boer et al (2001) and Jeffrey et al (2001). But the cross-validation process gives the average prediction error for all stations and may be heavily affected by the lopsided distribution of samples (Prudhomme and Reed 1999). The satisfying RMSE values shown in Figure 3A may only be appropriate for the lowlands (ie areas with dense stations; see Figure 2A). In the mountainous region, for example in orographic Shei-Pa National Park with only 4 stations, OK disregards the effect of topography (589–3882 m) and therefore gets a small-amplitude variability of T1 (5.85–12.35°C). According to the lapse rate of 4.28–6.14°C/km mentioned in the Results section, T1 at Mount Shei, the highest peak of Shei-Pa National Park, should be −8.29 to −1.06°C. It is easy to perceive the inadequacy of OK by the contrast between OK-interpolated 5.85°C and probable −8.29 to −1.06°C. The major reason for the impractical OK interpolation is the low density of stations, and a minor one is the disregard of altitude as an ancillary variable. This example shows that OK is not an appropriate method based on the sparse station network, which mostly affects the high mountain area.

Among the other methods, COK has the second-best performance except in June to September. This may be the result of the complex relations between temperature and orography as well as a result of the climate characteristics of Taiwan, such as the monsoon system, the prevalent cloud belt, and the *Massenerhebung* effect (Su 1984a, 1984b). Besides, Figure 3A reveals that detOK and aniOK have no special advantage as a result of the uncertainty of the detrending procedure and the anisotropy coupled with kriging in Taiwan.

Figure 3A shows the general superiority of resOK, a commonly used method in SI of temperature (eg Guler et al 2009), over other methods. Its predicted errors are significantly lower than with the other methods ($P < 0.05$). The resOK based on a regression model and less affected by station density (Marquínez et al 2003), known as trivariate regression-kriging (Boer et al 2001), integrates 2 sources of information, namely the large trend and the localized variance. The former, Tm regression formula as a function of 3 predictor variables, represents large-scale trends. The latter, OK interpolation of regression residuals, represents localized regional variances on smaller scales. That is to say, resOK is based on the principle that temperature can be described as a combination of a deterministic (trend) and a stochastic component (Engen-Skaugen et al 2007). Moreover, in mountainous areas with very few stations, the resOK method also has the best performance. If we examine the RMSE of T1 layer using only the 42 stations above 500 m (see Figure 2A), for example, the methods (and their RMSE) are ranked from best to worst as follows: resOK (0.96°C), COK (1.27°C), logOK (3.13°C), detOK (3.26°C), OK (3.35°C), aniOK (3.48°C).

Interpolating precipitation layers

Precipitation increases with elevation because of the ascent, cooling, and condensation of wet air in mountainous terrain, but the relation varies substantially with complex factors such as the orographic barrier characteristics, the distance from large water bodies, the strength and moisture content of wind, etc (eg Prudhomme and Reed 1999). In our study, E can explain only 1.2–45.4% variance of Pm. When we used all 3 predictors, 29.8–78.2% of the variance in precipitation could be explained (Table 4). This reveals the unstable relation between precipitation and station elevation and position, without a uniform gradient (eg Moral 2009). The low density of stations and the complex relation between precipitation and topography make it difficult to obtain a reliable spatial model of precipitation (Drogue et al 2002). The difficulties of modeling, especially in mountainous areas such as Taiwan, mainly derived from the interaction of weather with topography, result in a highly variable precipitation pattern (Singh and Kumar 1997). Our regression analysis also reveals that 3 geographic predictors E, X, and Y can only explain weak and unstable

Downloaded From: https://complete.bioone.org/journals/Mountain-Research-and-Development on 28 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

TABLE 5 Prediction error indices for January and July mean precipitation (P1 and P7) interpolated by 6 different methods.

	Error indices	OK	detOK	aniOK	COK	resOK	logOK
P1	Mean	–0.249	–0.038	–0.168	–0.249	–0.032	–0.619
	RMSE	17.750	17.600	17.600	17.750	17.310	17.510
	Average standard error	8.074	13.490	8.170	8.069	13.170	21.810
	Mean standardized	–0.016	–0.001	–0.010	–0.016	0.000	–0.008
	Standard RMSE	3.024	1.247	3.045	3.026	1.256	0.949
P7	Mean	0.224	–0.227	0.102	0.225	–0.067	3.413
	RMSE	42.900	42.650	41.980	42.900	40.730	42.200
	Average standard error	0.224	27.720	60.640	76.380	27.640	89.350
	Mean standardized	0.003	–0.002	0.002	0.003	–0.001	0.051
	Standard RMSE	0.557	1.824	0.678	0.557	1.666	0.463

precipitation variance. Consequently, the regression application using geographic position,
such as resOK, has not shown significant benefits of interpolating Pm in contrast with its advantage in Tm interpolation.

In this study, the RMSE of Pm interpolation assessed by cross-validation for 6 methods was 31.15 mm, which seems to be better than the performances of Marquínez et al (2003). The overall spatial patterns of interpolated layers from 6 methods are similar; the comparison of prediction error indices for interpolated P1 and P7 is summarized in Table 5. The Pm similarities of prediction errors (Figure 3B, no statistically significant difference) and spatial patterns among the 6 SI methods increase the difficulty of choosing the suitable SI method.

In fact, a perfect method for every climate variable under different environments is hard to achieve. Price et al (2000: 82) suggest that "[i]n some instances, it may be preferable to use a simple method applied to the region of interest than to use a more sophisticated approach which could be marginally more accurate, but requires considerably more time and money to implement." The more complex COK, detOK, and aniOK methods recommended by many researchers provide no particular advantage relative to the simpler OK and logOK (Figure 3B; Table 5). This outcome is mainly a result of the complex topography–precipitation interaction. Predicting precipitation's spatial pattern is also made more troublesome by the high variability and non-Gaussian character of the data (Boer et al 2001). Interpolation in large areas is more complicated because the relationship between altitude and precipitation fades in large areas. The existence of this relation is the basis of detOK and aniOK interpolation, so it is not surprising that these 2 methods do no better than the others (Phillips et al 1992).

The logOK is a simpler method. It is based on the fact that the logarithm of precipitation has a more Gaussian distribution than the precipitation itself and then leads to more stable behavior (Phillips et al 1992; Martinez-Cob 1996; Price et al 2000). Although normality is not a prerequisite for kriging, it is a desirable property (Stahl et al 2006). Kriging will only generate the best absolute estimate if the random function fits a normal distribution (Moral 2009). Rather than attempting to justify one method over another for theoretical reasons, we adopted logOK as the precipitation SI method for practical reasons.

Comparison with pre-existing climate layers

We compared our SI results with 2 other pre-existing climate layers in Taiwan: (1) the climate atlas (known as ATLAS, a form of isopleth map), which is manually prepared by the Central Weather Bureau, and (2) grid layers from the Agricultural Research Institute estimated by "parameter-elevation regressions using an independent slope model" (known as PRISM; Daly et al 1994). All layers were transformed to the same coordinate system and the same grid size of this study.

A comparison of our Tm layers with ATLAS shows that their general patterns are in basic agreement. The average monthly difference is only 0.02°C. The annual difference can be calculated and mapped through ArcGIS; summary statistics are presented in Table 6. Their mean, range, and standard deviation (SD) are 0.27°C, −7.53–6.82°C, and 1.52°C, respectively. Our layer is cooler than ATLAS along the mountain crests and warmer in the river valleys. A comparison of our Pm layers with ATLAS reveals that the average monthly difference is only 4.64 mm. Table 6 presents the summary statistics of annual precipitation differences. Their mean, range, and SDs are −55.68 mm, −1868.18–1616.19 mm,

Downloaded From: https://complete.bioone.org/journals/Mountain-Research-and-Development on 28 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

TABLE 6 Summary statistics of annual difference value between our own estimate and ATLAS or PRISM. Min: minimum; Max: maximum; SD: standard deviation.

Difference value		Our estimate minus ATLAS	Our estimate minus PRISM
Annual temperature (unit: °C)	Mean	0.27	–0.63
	Max	6.82	3.83
	Min	–7.53	–5.11
	SD	1.52	0.92
Annual precipitation (unit: mm)	Mean	–55.68	–152.74
	Max	1616.19	1421.77
	Min	–1868.18	–5937.40
	SD	311.27	493.54

and 311.27 mm, respectively. Further quantitative comparisons between our layers and ATLAS were hindered because of the unknown data QC procedure, data period, sample stations, and the most important mapping method (manual isopleth) used for ATLAS.

Comparing our Tm layers with PRISM showed that their general patterns are also in basic agreement. Their average monthly difference is 0.59°C, much greater than with ATLAS, however. Each monthly difference can be calculated and mapped through ArcGIS, showing that 10.78% grids have a greater than 2°C difference, with T7 as an example. Table 6 shows the difference in annual mean temperature. The mean, range, and SD are −0.63°C, −5.11–3.83°C, and 0.92°C, respectively. Our layer is cooler than PRISM almost everywhere, especially along the mountain crests of the northern and central highlands. Our layer is somewhat warmer than PRISM along the eastern slopes of the Central Mountain Range, especially its southern section. In comparing our Pm layers with PRISM, we found that the agreement is slightly less, with an average monthly difference of about 12.73 mm. Only 5.38% of the grids have a difference greater than 50 mm, with P1 as an example. Table 6 shows that the mean, range, and SD of annual precipitation are −152.74 mm, −5937.40–1421.77 mm, and 493.54 mm, respectively. Our value is lower than that of PRISM by about 2000 mm in the western mountains, and by about 4500 mm on a few peaks. The larger differences may hint that some extremes were filtered out in our QC procedure but used as sampling points in PRISM. Because the different baselines (eg the different data QC procedure, data period, sample stations) increase disagreements, we are unable to assess which method is more accurate.

Overall, the spatial patterns of temperature and precipitation in our layers are similar to those in ATLAS and PRISM. All 3 Tm layers correctly represent the relation between temperature and altitude (as well as latitude, for some regions), but our layers better illustrate the inseparability between temperature and orography. All 3 Pm layers present the expected regional and seasonal variations in precipitation. Our layers, however, seem to diminish some of the doubtful extremes that appear in ATLAS and PRISM for certain mountainous regions. In general, our results are closer to ATLAS than to PRISM. The main differences between our layers and PRISM may derive from the QC procedure, the data period, the sample stations, and the SI method.

Conclusions

In this article we propose a 2-stage process to map continuous climate layers from scattered and unchecked meteorological stations. The proposed method is based on a GIS approach to carry out a QC of meteorological data first, and then use their long-term data as sample points to proceed with the SI of climate layers. The former is a prerequisite for the latter.

In the meteorological QC procedure, many doubtful stations and unreasonable observed data were filtered out. This procedure can provide fundamental assurance of data quality and raise the accuracy of follow-up interpolation. In the climate SI procedure, we evaluated the performance of different kriging-based methods. We adopted resOK as the best temperature SI method based on cross-validation. Accurate interpolation of precipitation spatial patterns is a more complex undertaking than interpolation for mean temperature (eg Guler et al 2007; Ashiq et al 2009). We found no statistically significant difference among the 6 Pm interpolation methods. The logOK was preferred over the other methods for interpolating precipitation, not so much because of its superiority in predicting errors but for more practical reasons such as its stability and simplicity. A comparison of our SI layers with pre-existing climate layers showed that their overall spatial patterns are similar. The proposed 2-stage process is quite general and offers the possibility of mapping adequate climate layers; it could thus potentially be applied to other mountains with unchecked meteorological databases.

Downloaded From: https://complete.bioone.org/journals/Mountain-Research-and-Development on 28 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

ACKNOWLEDGMENTS

This work was supported by Shei-Pa National Park research project No. 094-301020500G-019. We would like to thank Miss Yi-Hsuan Liu (GIS Research Center, Feng Chia University) for her computer expertise. We gratefully acknowledge the advice of Professor Hsy-Yu Tzeng (Department of Forestry, National Chung Hsing University) on some ecological aspects. Miss Min-Chun Liao and Mr Hung-Chih Lin (Forestry Bureau) in particular helped us collect literature and argued that such a work would be valuable. The paper benefited from extensive and insightful reviewer comments.

REFERENCES

Ashiq MW, Zhao C, Ni J, Akhtar M. 2009. GIS-based high-resolution spatial interpolation of precipitation in mountain–plain areas of Upper Pakistan for regional climate change impact studies. *Theoretical and Applied Climatology, OnlineEarly,* 5 May 2009. http://dx.doi.org/10.1007/s00704-009-0140-y.

Benavides R, Montes F, Rubio A, Osoro K. 2007. Geostatistical modelling of air temperature in a mountainous region of Northern Spain. *Agricultural and Forest Meteorology* 146:173–188.

Boer EPJ, de Beurs KM, Hartkamp AD. 2001. Kriging and thin plate splines for mapping climate variables. *International Journal of Applied Earth Observation and Geoinformation* 3(2):146–154.

Daly C, Neilson RP, Phillips DL. 1994. A statistical-topographic model for mapping climatological precipitation over mountainous terrain. *Journal of Applied Meteorology* 33:140–158.

Daly C. 2006. Guidelines for assessing the suitability of spatial climate datasets. *International Journal of Climatology* 26:707–721.

Drogue G, Humbert J, Deraisme J, Mahr N, Freslon N. 2002. A statistical-topographic model using an omnidirectional parameterization of the relief for mapping orographic rainfall. *International Journal of Climatology* 22:599–613.

Easterling DR, Peterson TC. 1995. A new method for detecting and adjusting for undocumented discontinuities in climatological time series. *International Journal of Climatology* 15:369–377.

Engen-Skaugen T, Haugen JE, Tveito OE. 2007. Temperature scenarios for Norway: From regional to local scale. *Climate Dynamics* 29:441–453.

Franklin J. 1995. Predictive vegetation mapping: geographic modelling of biospatial patterns in relation to environmental gradients. *Progress in Physical Geography* 19(4):474–499.

González-Rouco J, Jiménez JL, Quesada V, Valero F. 2001. Quality control and homogeneity of precipitation data in the southwest of Europe. *Journal of Climate* 14:964–978.

Goovaerts P. 1997. *Geostatistics for Natural Resources Evaluation.* New York, NY: Oxford University Press.

Goovaerts P. 2000. Geostastistical approaches for incorporating elevation into the spatial interpolation of rainfall. *Journal of Hydrology* 228:113–129.

Guan BT, Hsu HW, Wey TH, Tsao LS. 2009. Modeling monthly mean temperatures for the mountain regions of Taiwan by generalized additive models. *Agricultural and Forest Meteorology* 149(2):281–290.

Guisan A, Zimmermann NE. 2000. Predictive habitat distribution models in ecology. *Ecological Modelling* 135:147–186.

Guler M, Cemek B, Gunal H. 2009. Assessment of some spatial climatic layers through GIS and statistical analysis techniques in Samsun Turkey. *Meteorological Applications* 14:163–169.

Hevesi JA, Flint AL, Istok JD. 1992. Precipitation estimation in mountainous terrain using multivariate geostatistics, Part I: Structure analysis. *Journal of Applied Meteorology* 31:661–676.

Isaaks EH, Srivastava RM. 1989. *An Introduction to Applied Geostatistics.* New York, NY: Oxford University Press.

Jeffrey SJ, Carter JO, Moodie KB, Beswick AR. 2001. Using spatial interpolation to construct a comprehensive archive of Australian climate data. *Environmental Modelling and Software* 16(4):309–330.

Johnston K, Ver Hoef JM, Krivoruchko K, Lucas N. 2001. *Using ArcGIS Geostatistical Analyst.* Redlands, CA: ESRI.

Knotters M, Brus DJ, Oude Voshaar JH. 1995. A comparison of kriging, co-kriging and kriging combined with regression for spatial interpolation of horizon depth with censored observations. *Geoderma* 67:227–246.

Marquínez J, Lastra J, Garcia P. 2003. Estimation models for precipitation in mountainous regions: the use of GIS and multivariate analysis. *Journal of Hydrology* 270:1–11.

Martinez-Cob A. 1996. Multivariate geostatistical analysis of evapotranspiration and precipitation in mountainous terrain. *Journal of Hydrology* 174:19–35.

Meek DW, Hatfield JL. 1994. Data quality checking for single station meteorological database. *Agricultural and Forest Meteorology* 69:85–109.

Moral FJ. 2009. Comparison of different geostatistical approaches to map climate variables: Application to precipitation. *International Journal of Climatology, OnlineEarly,* 9 April 2009. http://dx.doi.org/10.1002/joc.1913.

Ninyerola M, Pons X, Roure JM. 2000. A methodological approach of climatological modelling of air temperature and precipitation through GIS techniques. *International Journal of Climatology* 20:1823–1841.

Pardo-Igúzquiza E. 1998. Comparison of geostatistical methods for estimating the areal average climatological rainfall mean using data on precipitation and topography. *International Journal of Climatology* 18:1031–1047.

Peterson TC, Easterling DR, Karl TR, Groisman P, Nicholls N, Plummer N, Torok S, Auer I, Boehm R, Gullett D, Vincent L, Heino R, Tuuomenvirta H, Mestre O, Alexandersson H, Jones P, Parker D. 1998. Homogeneity adjustments of in situ atmospheric climate data: A review. *International Journal of Climatology* 18:1493–1517.

Phillips DL, Dolph J, Marks D. 1992. A comparison of geostatistical procedures for spatial analysis of precipitation in mountainous terrain. *Agricultural and Forest Meteorology* 58:119–141.

Prentice KC. 1990. Bioclimatic distribution of vegetation for general circulation model. *Journal of Geophysical Research* 95:11811–11830.

Price DT, Mckenney DW, Nalder IA, Hutchinson HF, Kesteven JL. 2000. A comparison of two statistical methods for spatial interpolation of Canadian monthly mean climate data. *Agricultural and Forest Meteorology* 101:81–94.

Prudhomme C, Reed DW. 1999. Mapping extreme rainfall in a mountainous region using geostatistical techniques: A case study in Scotland. *International Journal of Climatology* 19:1337–1356.

Rhoades DA, Salinger MJ. 1993. Adjustment of temperature and rainfall records for site changes. *International Journal of Climatology* 13:899–913.

Rolland C. 2003. Spatial and seasonal variations of air temperature lapse rates in alpine regions. *Journal of Climate* 16:1032–1046.

Singh P, Kumar N. 1997. Effect of orography on precipitation in the eastern Himalayan region. *Journal of Hydrology* 199:183–206.

Stahl K, Moore RD, Floyer JA, Asplin MG, McKendry IG. 2006. Comparison of approaches for spatial interpolation of daily air temperature in a large region with complex topography and highly variable station density. *Agricultural and Forest Meteorology* 139:224–236.

Štěpánek P, Zahradníček P, Skalák P. 2009. Data quality control and homogenization of air temperature and precipitation series in the area of the Czech Republic in the period 1961–2007. *Advances in Science and Research* 3: 23–26.

Su HJ. 1984a. Studies on the climate and vegetation types of the natural forests in Taiwan (1) Analysis of the variation in climatic factors. *Quarterly Journal of Chinese Forestry* 17(3):1–14.

Su HJ. 1984b. Studies on the climate and vegetation types of the natural forests in Taiwan (2) Altitudinal vegetation zone in relation to temperature gradient. *Quarterly Journal of Chinese Forestry* 17(4):57–73.

Su HJ. 1985. Studies on the climate and vegetation types of the natural forests in Taiwan (3) A scheme of geographic climate regions. *Quarterly Journal of Chinese Forestry* 18(3):33–44.

Tabios GQ, Salas JD. 1985. A comparative analysis of techniques for spatial interpolation of precipitation. *Water Resources Bulletin* 21(3):365–380.

Thuiller W, Albert C, Araújo MB, Berry PM, CabezaM, Guisan A, Hickler T, Midgley GF, Paterson J, Schurr FM, Sykes MT, Zimmermann NE. 2008. Predicting global change impacts on plant species' distributions: Future challenges. *Perspectives in Plant Ecology, Evolution and Systematics* 9:137–152.

Tuhkanen S. 1980. Climatic parameters and indices in plant geography. *Acta Phytogeographica Suecica* 67:1–110.

Walter H. 1985. *Vegetation of the Earth and Ecological Systems of the Geo-biosphere.* 3rd edition. New York, NY: Springer-Verlag.

Downloaded From: https://complete.bioone.org/journals/Mountain-Research-and-Development on 28 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use