ACADEMIA JOURNAL OF BIOLOGY 2021, 43(4): 101–117
DOI: 10.15625/2615-9023/16062

CONSTRUCTION OF FINE RESOLUTION BIOCLIMATIC VARIABLES FOR CENTRAL HIGHLANDS AND SOUTHERN CENTRAL COAST OF VIETNAM

Pham Bach Viet[1], Hoang Minh Duc[2], Tran Van Bang[2], Luu Hong Truong[2], Nguyen Dang Quang[3], Diep Dinh Phong[2,*]

[1]Vietnam National Space Center, VAST, Vietnam
[2]Southern Institute of Ecology, VAST, Vietnam
[3]Center for Hydro-Meteorological Technology Application, Vietnam

Received 12 May 2021, accepted 16 December 2021

ABSTRACT

In this study, 19 surface bioclimatic variables of high spatial resolution 0.00226º (~ 250 m) are generated in a Geographic Information System by the combination of (1) the raster dataset of monthly temperature and precipitation obtained from the global WorldClim database at 0.00833º spatial resolution for the period of 1960–2000; and (2) the climate data (temperature and precipitation) of the Central Highlands and Southern Central Coast collected from the 31 temperature and 97 precipitation recording sites for the period of 1991–2015. The statistical downscaling method is applied, using multiple linear regression analysis, in which elevation, geographic coordinates, and distance from the coast are treated as independent variables, to estimate the distribution of temperature; and the B-Spline interpolation method combined with multiple linear regression analysis is employed on precipitation over the study area. The outcomes of the two main analyses are computed to create 19 high spatial resolution bioclimatic variables. While using only local climate data on analyzing the regression models results in high fluctuation of estimated temperature, the combination of the two datasets is more informative. The spatial distribution of our interpolated precipitation is similar to the WorldClim data but has a smaller difference in the mean annual precipitation. The results, which shows higher spatial resolution and are closer to the observed data than those from the WorldClim, could be better applied for predicting species distribution in the region.

Keywords: Bioclimatic variables, climate data, downscaling, multiple linear regression, precipitation, temperature.

Citation: Pham Bach Viet, Hoang Minh Duc, Tran Van Bang, Luu Hong Truong, Nguyen Dang Quang, Diep Dinh Phong, 2021. Construction of fine resolution bioclimatic variables for Central Highlands and Southern Central Coast of Vietnam. *Academia Journal of Biology,* 43(4): 101–117. https://doi.org/10.15625/2615-9023/16062

**Corresponding author email*: diepdinhphong@gmail.com

INTRODUCTION

The distribution of a species is influenced by several biological and environmental factors (Franklin, 2010), of which bioclimatic variables (i.e. variables are derived from climatic data and biologically meaningful) have been widely used for species distribution models (SMDs) in the fields of ecology research and biodiversity conservation for both vegetation and animals (Bett et al., 2012; Liu et al., 2013; Porfirio et al., 2014; Raghavan et al., 2016; Chandra, 2016). Bioclimatic variables were first introduced in 1986 in Australia, conceptually developed by Henry Nix (John, 1991; Booth et al., 2014). Since then, several significant steps have improved the concept and methods by the WorldClim team (Booth, 2018). The latest version 2 of bioclimatic variables and version 2.1 of climatic variables, as well the previous versions, are surface models constructed from surface climatic models interpolated from monthly temperature and precipitation values of climatic data gathered at global meteorological stations and other data sources (Fick & Hijmans, 2017). The WorldClim database represents the most common source of climate data for this kind of modelling, covering all of the global land areas, excluding Antarctica (Wagner et al., 2017; WorldClim[1]), with the spatial resolution between 30 arc-second (~ 0.86 km^2) to 10 minutes (~ 344 km^2) (WorldClim[2]).

Climatic surface data and bioclimatic variables were also studied and developed by other groups such as CHELSA (Karger, 2017), PRISM (Daly et al., 2008; O’Donnell, 2012), CliMond (Kriticos, 2011), and TerraClimate (Abatzoglou et al., 2018) where spatial resolution is similar or coarser. These datasets are either at the global scale or continental scale and were constructed by various approaches, some of which were partly inherited from the WorldClim, while the others incorporated new variables into the development of surface climatic data.

Applications of bioclimatic variables have been implemented on geographic regions at various scales ranging from sub-regional to continental and global scale, which are mostly between regional and continental scales, both taking more than 70% (Porfirio et al., 2014). Numerous research projects have made use of these bioclimatic variables, all 19 variables or some subsets, most of these have focused on Europe with only a small fraction (~3%) covering Asia (Porfirio et al., 2014; Booth et al., 2014).

Presently, global data of the finest spatial resolution of the mentioned bioclimatic variables is 30 arc-second, which is broadly useful for SDMs at various spatial scales. Applications at a small scale, even at sub-country size or small areas, utilize this resolution. Recent examples of studies in Asia include research on primate distribution for Vietnam (Nguyen et al., 2019), the discovery of a new crocodile-lizard population in the border region of Vietnam and China (van Schingen et al., 2016), primates in Indochina (Bett et al., 2012), and biodiversity and forest health at the continental scale of Tropical Asia (Deb, 2016; Deb et al., 2017).

The WorldClim dataset in 30 arc-second resolution is widely used for SDMs because of its availability and global coverage (Marchi et al., 2019), and it is better at capturing the environmental variables than coarser resolutions where there is a limited number of observation stations, particularly in mountainous and remote areas (Fick & Hijmans, 2017). Although the resolution of 30 arc-second is useful for modeling the distribution of widespread species, it seems to cause SDMs to omit or underestimate the distribution of rare and restricted range species. A study on habitat loss of species in the Swiss Alps under climate change impacts at a local-scale (25 m × 25 m grid cells) predicted that 100% of species persisted in their habitats under the local-scale models while all their suitable habitats were lost under the European-scale model (10’ × 10’) (Marchi et al., 2019). Organisms, especially restricted-range and locally endemic species, experience or highly adapt to the local environment in a way that is quite different from a large-scale environment (Collen et al., 2014; Peterson et al., 1998). For ecological applications, the representation of the temporal and spatial variability of temperature and precipitation is important to infer ecological niches and species distributions. Consequently, errors in the climatic dataset at this small spatial scale can accumulate in such studies, which calls for an improvement of climatic information available for such analyses (Karger et al., 2017). In order to make SMDs more detailed for localization of species distribution, a higher spatial resolution of surface climatic data must be performed relying on downscaling methods and adopted local climatic data, which results in a fine resolution of surface bioclimatic dataset.

This paper describes the construction of 19 fine spatial resolution bioclimatic variables by the integration of surface climatic dataset obtained from the WorldClim and local climatic data with geographic data for the Central Highlands (CH) and Southern Central Coast (SC) of Vietnam (Fig. 1). Statistical downscaling method and geostatistical spatial interpolation are applied to address the above problem and to yield an expected spatial resolution of equivalent 250 m for surface climate variables. These results are of significant use in constructing 19 bioclimatic variables applied for SDMs.

MATERIALS AND METHODS

Study area

The study area includes the Central Highlands and Southern Central Coast of Vietnam (10º34'N–16º12'N and 107º12'E–109º27'E), which covers nearly 100,000 km^2 distributed in Da Nang City and 12 other provinces (Fig. 1). The landscape of CH-SC comprises two typical regions known as (i) the narrow coastal plains mixed with low mountains adjacent to the west with CH, and the eastern portion of plains close to the East Sea, (ii) low mountains with plateaus of Dak Lak, Plei Ku and Lam Vien, collectively referred as the Tay Nguyen Plateau (CH) (Regalado et al., 2005). Topographically, low mountains located in this study area are referred to the southern portion of the Truong Son (Annamite) Range, which extends through Laos, Vietnam and a small area in Cambodia; the highest elevation in CH is at the top of Ngoc Linh at 2,598 m. In the study area, South Truong Son Range runs parallel to the coast, northwest-southeast in the north and then the north-south direction in the south. The eastern slope of the range rises steeply from the plain; the western slope is more gentle, forming large plateaus. Climatologically, Vietnam has a monsoon system of southwesterly in April-September and northeasterly in October-March (Phan et al., 2014). However, the rainfall characteristics over the study region are complicated, switching from summer (April, May) to winter (November, December) in different places. The rainy season in CH and north-central SC is starting in April-May, and withdrawing in October-November, meanwhile, in the southern tip of SC (Ninh Thuan, Binh Thuan provinces) the rainy season arrives late and stays in a short period time, usually begins in September and ends in December (Nguyen et al., 2013, 2014).

The total rainfall in the rainy season contributes more than 80% of the annual rainfall. The annual rainfall ranges from 1,500 mm to 2,400 mm and the inter-annual variations of rainfall are mainly influenced by seasonal winds. During the rainy season, the monthly rainfall exceeds 200 mm and reaches its peak in August and September in most of the study region. The average annual temperature is rather low in comparison with other regions, ranging from 20 ºC to 25 ºC and the highest temperatures are in April and May from 27 ºC to 32 ºC (Nguyen et al., 2013; CGIAR, 2016). In Central Vietnam, rainfall distributions and rainy seasons are significantly controlled by the combination of cold surges and tropical cyclones, and roughly two thirds of the heavy rainfall events are caused by tropical cyclones. Central Vietnam has its maximum monthly rainfall in autumn (September to November), and at some sites in the region, the annual rainfall can reach 3,600 mm (Hue station) to 4,000 mm (Ba To station) (Nguyen & Nguyen, 2004).

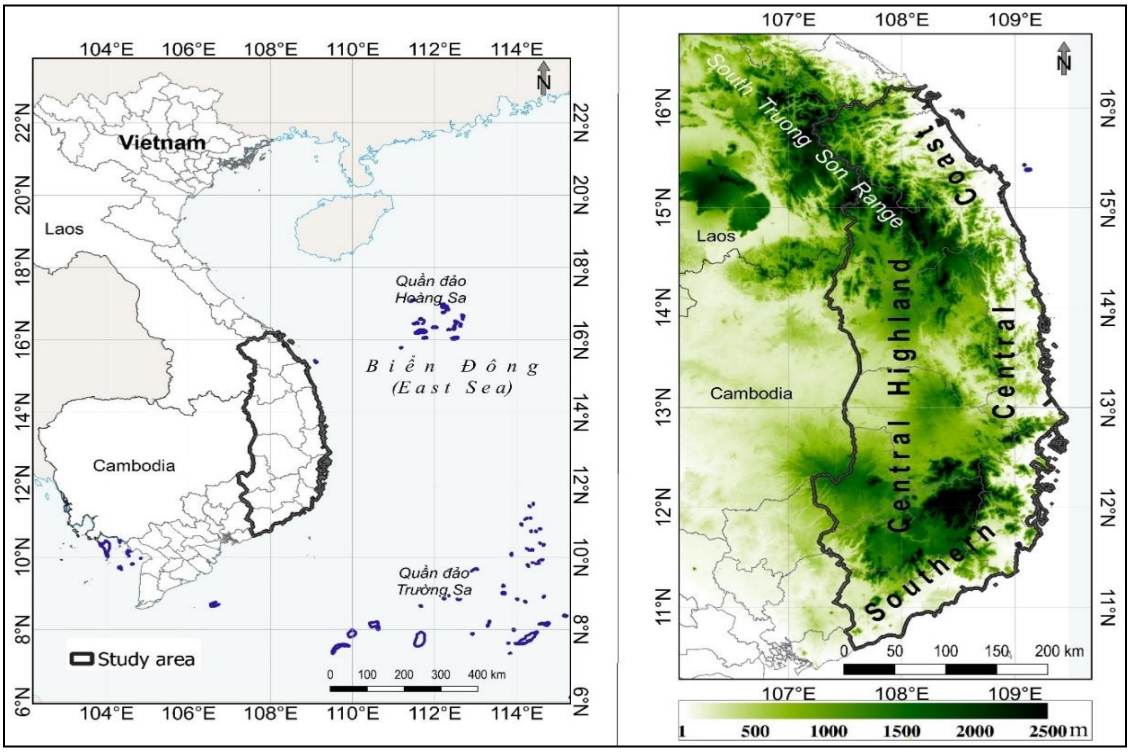

Figure 1. Location of the study area in the Indochina region. CH-SC in thick solid lines (left map) and its surroundings with detailed terrain (right map). Colour ramp bar indicates elevation in meters

Data sources

Data were assembled from the following sources: (1) Surface climatic data of WorldClim database, (2) Climate data (temperature and precipitation) recorded by local meteorological stations, (3) Digital Elevation Model (DEM) 1 arc-second, (4) Distance from the coast (DC), and (5) Geographic coordinates: longitude (LON) and latitude (LAT).

Surface climate data of the WorldClim (in GeoTiff format) are the mean monthly value for the period 1970–2000, including variables of monthly mean, maximum, minimum temperature, and precipitation, one for each month of the year. With spatial resolution 30 arc-second (or 0.00833^{o}), 0.86 km^2 at the equator, they are commonly considered at the spatial resolution of 1 km or 1 km^2. These datasets are downloadable on the WorldClim webpage (WorldClim[1]) (Fick & Hijmans, 2017). Presently, they are the finest-resolution global data and represents the Vietnam climate pattern more properly than do the others mentioned.

Precipitation and temperature data were provided by the Center for Hydro-Meteorological Technology Application (Vietnam Meteorological and Hydrological Administration) based on records from 97 sites in the national hydro-meteorological station network (Fig. 2). Among these, 31 hydro-meteorological stations observe both temperature and rainfall and 66 rain gauge sites record rainfall only. This dataset includes average, maximum and minimum temperatures in degrees Celsius and precipitations in millimetres for each month, from 1991 to 2015 (25-year period) and contains geographic coordinates and altitudes of the stations above the sea level. The spatial distribution of these sites is uneven as illustrated in Figure 2.

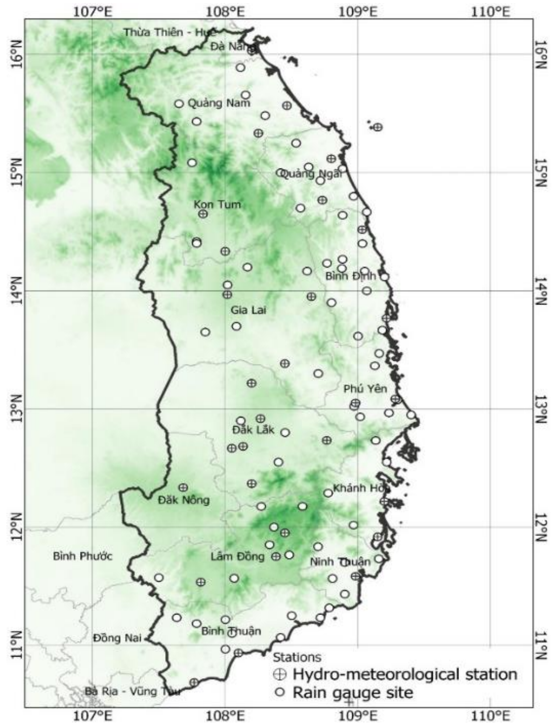

Figure 2. Geographic locations of data-recording sites used in the study

DEM of Shuttle Radar Topology Mission (SRTM) 1 Arc-second Global version 3 is available for downloading at https://earthexplorer.usgs.gov. In this version, missing data or gaps were filled with other digital elevation data from Japan and Germany, as a result, it has less error than older versions. The most recent version was released by the USGS in January 2015 (https://earthdata.nasa.gov/). Notably, high resolution of elevation data is one of the important auxiliary data, which enable spatially downscaling information of interest in a valid fashion (Hengl, 2009).

DC, LON and LAT were processed and extracted from DEM. The coastline was vectorized from DEM and utilized as a baseline to calculate distance from all locations to the line. Small bays were generalized as these were considered small water bodies. Distance from the coast was identified by the straight line, from a cell to the coastline, which is relatively perpendicular to the coastline and measured in kilometres. Geographic coordinates were extracted from DEM for each cell by latitude and longitude to produce two separated raster data, whose unit is decimal degree as other datasets.

Data preparation

The quality of the raster data is highly dependent on the amount of input data, which varies in space and time (Maeda et al., 2020). There is less error and consequently higher interpolation accuracy in regions with high station densities than those with low spatial data density. Particularly, in places with very few weather stations and strong gradients or topographically complex terrain, climate interpolation is difficult, leading to a poor representation of climatic variability (Fick & Hijmans, 2017; Karger et al., 2017). It becomes a constraint to make SDMs more detailed for localization of species distribution (Fick & Hijmans, 2017). This sense describes the study area of CH-SC alike, where meteorological stations are sparse and spatially unevenly distributed. To improve this situation, long-term climate data of precipitation and temperature of weather stations in CH - SC were obtained to integrate with the surface climatic dataset of WorldClim. These datasets were interpolated to make fine spatial resolution surface climatic data as the basis for high resolution of bio-climatic variables.

WorldClim data: The values of the climate data were drawn from the 30 arc-second surface dataset (or 0.00833°) at the interval of 0.0416° (roughly 4.5 km distance) to generate a dataset of point vectors. These datasets of point vectors, which were featured in values of temperature (mean, minimum, maximum) and precipitation for 12 months, were used to interpolate.

Station data: A dataset of monthly temperature and precipitation for each year for the period 1991–2015 was checked to remove any outliers. Missing data at rain-gauge stations are categorized as (i) no data for all months of the year, as observed at some stations up to 4 or 5 years; (ii) no data for months, usually in the dry season, for all years of the period; and (iii) no data in some months of the dry season for a few years. Despite missing data at those stations for years and months, the remaining available data were kept for calculation for those stations to limit error caused by low spatial data density (Fick & Hijmans, 2017). Those stations are in remote mountainous areas where station density is low and thus those temporal data were yet included in calculation later. There are no issues with the temperature data (Fig. 2). In all cases of missing data, they were treated as no data and were not counted in averaging the monthly mean. The two datasets were averaged out at the monthly long-term mean for the period up to 25 years.

The 31 stations with temperature data included for analysis range from 2.0 m to 1,500 m in altitude (while the topography of the study area includes elevations ranging from 0 m to above 2,500 m), with 68% stations situated below 500 m, 21% stations from 500 m to 1,000 m and 3% (one station) above 1,000 m (i.e. Da Lat Station: 1,508 m). The average density is 0.31 stations per 1,000 km^2 or one station for more than 3200 km^2. The distance between two adjacent stations ranges from 9.4 km to 113 km. The 97 sites with precipitation data have a better density, 0.97 stations per 1,000 km^2, or one station for roughly 1,030 km^2. There are 71% of the stations situated below 500 m in altitude, 24% from 500 m to 1,000 m and 5% (equal to 5 stations) above 1,000 m. Some stations are quite close to one another, even less than 2 km. In general, the spatial distribution of precipitation stations is like the situation of temperature stations.

The dataset of 0.0416^o interval was integrated with the station data to produce a new one, which is named hereafter WorldClim-Station data. These two datasets were synchronized together with variables of mean monthly temperature (average, minimum, maximum) and monthly precipitation. Such a combination would contribute more local specific features of temperature and precipitation to interpolated models than using WorldClim data solely. These datasets were used for spatial interpolation and regression calculation for the next procedures.

For surface modelling, the DEM of 1 arc-second resolution was resized to 0.00226^o (~250 m) spatial resolution by the nearest neighbour resampling technique, which is preferable when the result is used for further model analysis, because the original values were preserved, the terrain of the study area was intact (Prajapati et al., 2012; Le Coz et al., 2009) and, in particular, values of elevation of individual mountaintops were not shifted (Sharma et al., 2016). Distance from the coast, longitude, and latitude values were extracted from this 0.00226^o DEM. This step created four surface gridded datasets with a spatial resolution of 0.00226^o (~250 m), of which the elevation dataset (hereafter called ALT) was applied as the baseline data for afterwards spatial interpolation.

Like Hutchinson (1995), we expanded the spatial extent of all datasets used for the study area, especially the western margin of CH-SC (see the right image of Figure 1), to improve the accuracy of spatial interpolation. This expansion included WorldClim, ALT, DC, LON, and LAT datasets. This also holds for the WorldClim-Station point dataset to ensure a high statistical confidence level.

Data analysis for downscaling climatic variables

The statistical approach was used to generate interpolated climate surfaces, which varies with temperature and precipitation variables as described below.

Temperature

The correlation between altitude and temperature (Hutchinson, 1991; Andrews, 2010), which varies from place to place and in time depending on actual atmospheric conditions, was estimated using a linear regression analysis of temperature versus altitude (Loomis et al., 2017; Shen et al., 2016). This analysis has been widely used in developing empirical models (Lanzante, 1996;

Ninyerola et al., 2000). Like the techniques applied for climate variables of the WorldClim (Hijmans et al., 2005), datasets on longitude and latitude, and distance from the coast were used as independent variables. It is noted that WorldClim used the commercial ANUSPLIN software package to generate surface climatic data with the thin-plate spline technique (Hutchinson, 1995; Hijmans et al., 2005; Fick & Hijmans, 2017).

In our study, multiple linear regression with independent variables of altitude, distance from the coast, and geographical coordinates, and a dependent variable of monthly temperature, which had been obtained from the previous step, was used to model the relationship between the temperature and the other four independent variables by fitting a linear equation to sample data to estimate temperature according to the explanatory variables. These models then were applied to project back on the elevational surface of the altitude data (ALT) to create monthly surface temperature variables (average, minimum, maximum). The general equation is:

$$t = a_0 + b_1*(\mathrm{ALT}) + b_2*(\mathrm{LON}) + b_3*(\mathrm{LAT}) + b_4*(\mathrm{DC}) \quad (1)$$

Where: t is estimated temperature corresponding to the temperature at a given altitude, distance from the coast (DC) and x, y location (LON, LAT); and a_0 and $b_{(n)}$ are the regression coefficients estimated from a statistical analysis of datasets. All the calculations were done with α less than 0.05 regarding the relation between temperature, dependent variable, with independent variables to get statistical significance. Multiple linear regression with four independent variables might occur multicollinearity, which reduces the precision of the estimated coefficient values and weakens the statistical model. In this setting, Variance Inflation Factors (VIF) values were used to identify the correlation between independent variables and the strength of that correlation. If VIFs values were greater than 5, they were considered with care or eliminated from the regression, and the calculation process was repeated without that. VIFs between 1 and 5 were acceptable (Hair et al., 2014) and other values were regarded, such as the correlation coefficient, standard error, confidence of interval and p-value.

Precipitation

While temperature is closely correlated with altitude, particularly in the long-term mean, precipitation is considered as a highly non-linear phenomenon (Karger et al., 2017), and although interpolation and statistical downscaling approaches may also integrate land-surface predictors such as elevation and slope aspects, acceptable outcomes still require a more or less regular distribution of meteorological stations and a proper representation of topo-climatic settings. The relationship is rather complicated, though precipitation is also significantly affected by the terrain (Bohner et al., 2018). Orographic effect with windward-leeward, rain shadow and topographic aspect was included in interpolation models, such as PRISM (Daly et al., 2008) or ANUCLIM mentioned above, a set of coefficient values is needed to generate interpolated values at required locations to develop the WorldClim dataset (Xu et al., 2010). The development of these surface models is mathematically sophisticated; for example, the ANUCLIM takes a multivariate regression of the thin-plate smoothing spline algorithm for spatial interpolation (Hijman et al., 2005).

For interpolating precipitation, we applied the same multiple linear regression analysis as in the case of temperature described above, with the minor modification that the variable of DC was not used, and a parameter of the residual was added to the regression model to optimize predicted variables with least error. The regression function is:

$$P = a_0 + b_1*(\mathrm{ALT}) + b_2*(\mathrm{LON}) + b_3*(\mathrm{LAT}) + \varepsilon \quad (2)$$

Where: P is predicted precipitation at a given elevation and its location (LON, LAT); a_0 and $b_{(n)}$ are calculated from a statistical analysis of the dataset, ε (or the residual) is the difference between the observed value (y) and the predicted value ($\hat{y}$). Each data point has one residual computed from the observed value minus the predicted value ($\varepsilon = y - \hat{y}$). The regression function of this step was used to create a new grid with regression-based values on the surface of the DEM data with B-Spline interpolation. The B-Spline interpolation technique (or basic spline) is based on a mathematical model for surface estimation that fits a minimum-curvature surface through the input points while passing through the sample points (Ly et al., 2013). Spline makes a two-dimensional minimum curvature spline interpolation on a point dataset yielding a smooth surface that passes exactly through the input points, and the general performance of this computation is similar to (but simpler than) the ANUCLIM (Yang et al., 2015). This approach was employed to take advantage of the World Bioclim dataset and locality of precipitation. Similar to the temperature regression analysis, VIFs of variables of ALT, LON and LAT were also considered.

Bioclim variables computation

The obtained monthly surface temperature and rainfall data of 0.00226^{o} were used to generate the 19 bioclimatic variables representing annual trends (mean annual temperature and annual precipitation), seasonality (annual variability in temperature and precipitation) and extreme or limiting environmental factors (temperature of the coldest and warmest month, and precipitation of the wet and dry quarters). A quarter was three consecutive months. The 19 variables were coded from Bio1 to Bio19. Bio1 to Bio11 are relevant to temperature and Bio12 to Bio19 correspond to precipitation. These variables were constructed based on the description from WorldClim (WorldClim[3]) as follows:

Bio 1 = Annual Mean Temperature.

Bio 2 = Mean Diurnal Range (Mean of monthly (max temp - min temp)).

Bio 3 = Isothermality (Bio 2/Bio 7) (×100).

Bio 4 = Temperature Seasonality (standard deviation ×100).

Bio 5 = Max Temperature of Warmest Month.

Bio 6 = Min Temperature of Coldest Month.

Bio 7 = Temperature Annual Range (Bio 5 – Bio 6).

Bio 8 = Mean Temperature of Wettest Quarter.

Bio 9 = Mean Temperature of Driest Quarter.

Bio 10 = Mean Temperature of Warmest Quarter.

Bio 11 = Mean Temperature of Coldest Quarter.

Bio 12 = Annual Precipitation.

Bio 13 = Precipitation of Wettest Month.

Bio 14 = Precipitation of Driest Month.

Bio 15 = Precipitation Seasonality (Coefficient of Variation).

Bio 16 = Precipitation of Wettest Quarter.

Bio 17 = Precipitation of Driest Quarter.

Bio 18 = Precipitation of Warmest Quarter.

Bio 19 = Precipitation of Coldest Quarter.

Evaluation of the performance of new bioclimatic variables

Two statistical tests were implemented on the surface bioclimatic variables of the WorldClim and the calculated surface bioclimatic variables. One hundred spatially random points were sampled over the study area to implement the tests. First, the paired sample *T*-test was used to determine whether the mean difference between two sets of observations was zero. If there was no difference between the mean of the WorldClim bioclimatic variables and the calculated bioclimatic variables then the null hypothesis H_0: $\mu = 0$. Second, testing the significance of the correlation coefficient was used to evaluate the relationship between the two datasets, i.e. the correlation between the WorldClim and the calculated variables. The null hypothesis (H_0) of the test was that the correlation coefficient was not significantly different from zero (the coefficient is close to zero, or no correlation between the two datasets). The two tests complemented one another to assess the quality of 19 constructed bioclimatic variables.

In addition, another test of the paired sample *t*-test was applied to compare differences between the station data with the interpolated data and the station data with the WorldClim data. The data set used to test was from 21 stations which account for about 2/3 of the total stations recording both precipitation and temperature. The test was performed by interpolating with and without these 21-station data, whose outcomes were tested with the station data. This helps to appraise how the interpolation models work and to see the similarity between the WorldClim data and the local station data.

The Open Source Geographic Information System, QGIS version 3x (QGIS, 2019) that supports functions on vector and raster analysis as well spatial interpolation was used to process and analyze all surface datasets. Moreover, statistical calculations of this study were done in worksheet MS Excel@ 365 with data analysis Add-in.

RESULTS AND DISCUSSIONS

High spatial resolution surface bioclimatic variables of CH-SC

Nineteen bioclimatic variables with a high resolution of 0.00226° (about 250 m) for the CH-SC (Fig. 3) were constructed from monthly surface temperature (average, maximum, minimum) and monthly surface precipitation based on the WorldClim data for the period of 1970−2000 and the local dataset for the period of 1991−2015.

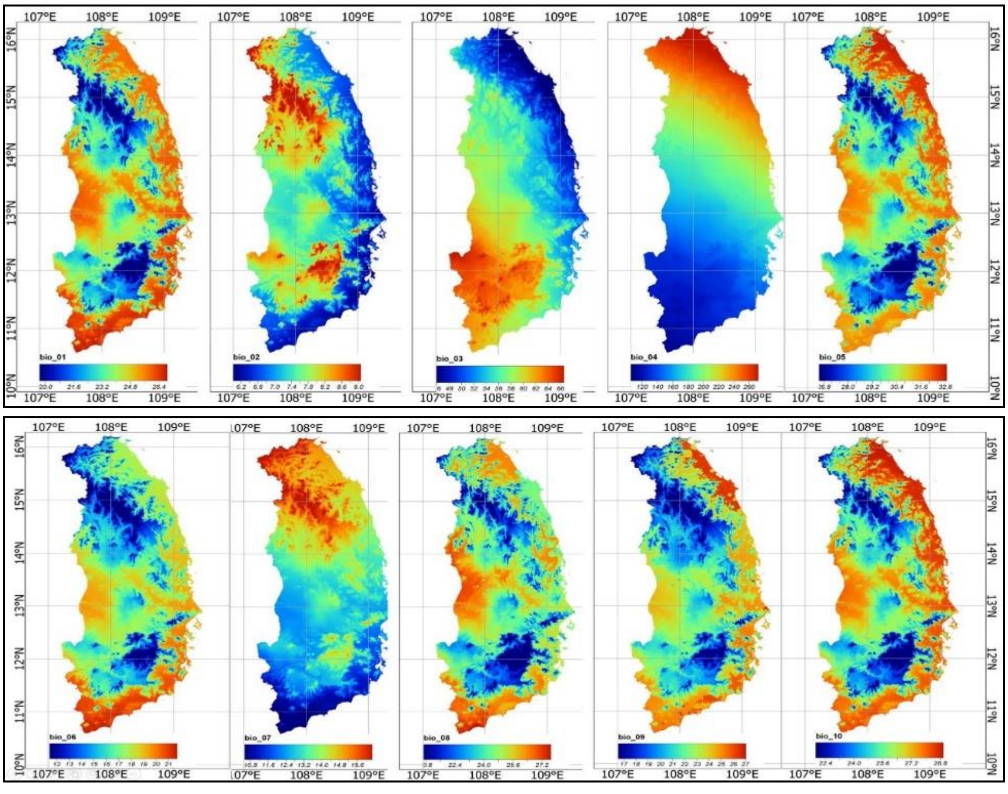

Figure 3. Nineteen bioclimatic variables, Bio 1 to bio 19 from left to right and upper to lower (in group a, b, c, d top-down). Colour bars from dark blue to dark red display values ranging from the lowest to the highest values of bioclimatic variables

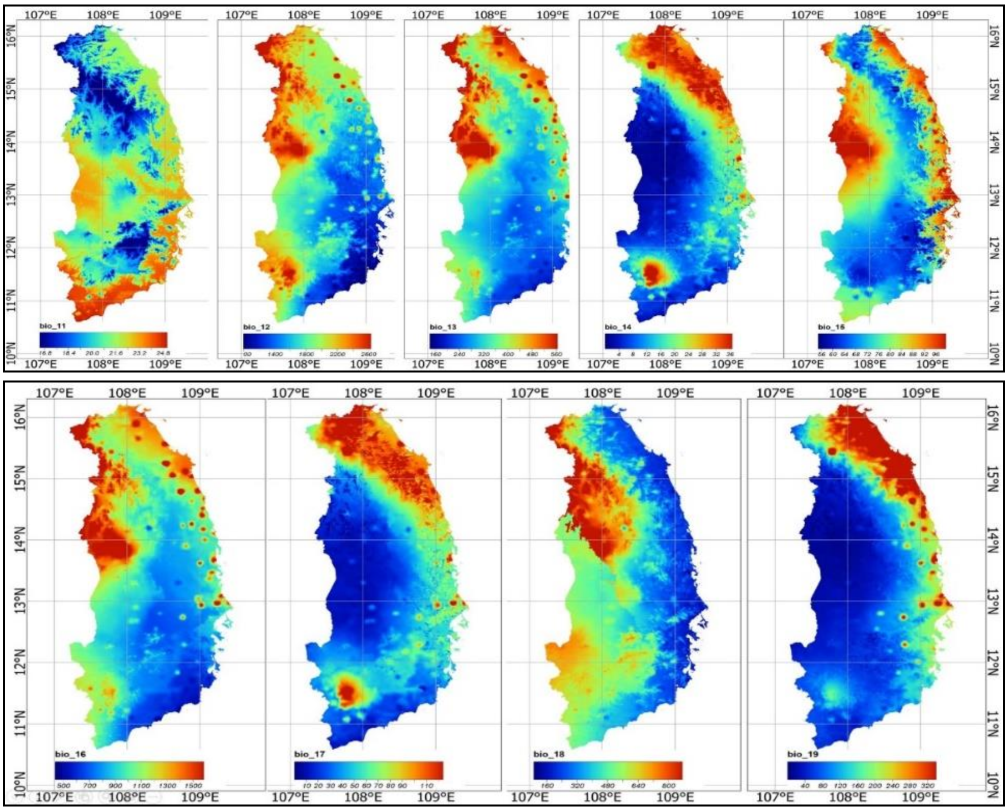

Figure 3. Nineteen bioclimatic variables, Bio 1 to bio 19 from left to right and upper to lower (in group a, b, c, d top-down). Colour bars from dark blue to dark red display values ranging from the lowest to the highest values of bioclimatic variables (next)

Temperature and precipitation variables from the regression models

Temperature

We run the multiple regression of average, minimum, and maximum monthly temperatures with two cases: The first case (only the local datasets): using local data only with 31 stations. The coefficient of correlation (R) was high (Table 1) exceptionally some months of maximum temperature were very low, resulting in all regression p-values below 0.05. p-values of independent variables highly fluctuated for all cases of months (higher or lower than 0.05 in some months). There was only the ALT with values below 0.05 for all months. Standard errors of intercept and independent variables were in the high range. VIFs values were in an acceptable range. In general, outcomes of the multiple regression analysis with the four independent variables for only temperature stations showed instability of estimated models that could affect predictive temperature values, making them lower or higher than the actual temperature for minimum, average, and maximum values in the study area. Consequently, these were not used for further analysis. This limitation is due to the uneven distribution of stations horizontally as well as in altitude and makes constraints when using only climate station data for spatial interpolation, which was mentioned above in the data section.

The second case (combined WorldClim and local datasets): The total number of points used for the analyses was 24,654 (n = 24,654). The test results indicate a significantly high correlation between the monthly average, minimum and maximum temperatures and ALT, LAT, LONG and DC with R being between 0.91 and 0.97, *p*-values approximately close to zero and VIFs between 1.29 and 1.76. With low variation and low standard errors, the lower and upper 95% confidence intervals (CIs) of the coefficient of each of four independent variables and the intercept values were close to coefficient values.

The temperature estimated from the regression model does not exceed the actual record range; for example, the estimated Min Temperature of Coldest Month in Bio 6 ranges between 4 °C and 21 °C within the altitudes of 7 m to 2,598 m matches well with those recorded at 1,508 m at Da Lat Station.

Table 1. Results of tests for multiple regression analysis for monthly average, minimum, maximum temperature values using station data only (n = 31)

Monthly temperature	R	Regression p-value, α = 0.05	p-value for intercept, ALT, LAT, LON, DC (α = 0.05)	Standard error	VIF
Average	0.96–0.99	< 0.05	ALT < 0.05. 0.05 > intercept, LAT, LON, DC > 0.05	High range	1.05–2.4
Minimum	0.88–0.94	< 0.05	ALT < 0.05. Intercept, DC > 0.05; LAT, LON mostly > 0.05	High range	1.05–2.4
Maximum	0.37–0.84	< 0.05	ALT Apr-Dec < 0.05; Intercept, LAT, LON, DC (mostly) > 0.05	High range	1.05–2.4

Precipitation

The regression analysis of precipitation (97 stations) with the three independent variables of ALT, LON and LAT and the dependent variable of the monthly precipitation has VIFs ranging from 1.07 to 1.14 indicating low collinearity is while the coefficient correlation R of monthly regression ranges from 0.53 to 0.86 with p-values of zero showing their statistical significance (Table 2).

Table 2. Coefficient of correlation R and p-values for monthly precipitation

Value	Jan.	Feb.	Mar.	Apr.	May.	Jun.	Jul.	Aug.	Sep.	Oct.	Nov.	Dec.
R	0.64	0.62	0.65	0.81	0.86	0.81	0.78	0.76	0.72	0.53	0.72	0.69
p-value	0	0	0	0	0	0	0	0	0	0	0	0

The total number of points used for the analysis was 27,683 (n = 27,683). The coefficients of regression variables (ALT, LON, LAT) have monthly differences presenting a seasonality of precipitation. The amount of precipitation of each month is different from each other, which is affected by other factors like orography, elevation, wind and location. All standard errors of coefficients were low, and the upper and lower 95% CIs of the three independent variables are close to coefficient values. The regression function with three coefficients of the predictive variables and residual analysis was applied for spatial interpolation to yield monthly surface precipitation models.

Comparison and evaluation

Table 3 displays the results of the tests for mean difference and the correlation coefficient between 19 variables of WorldClim and those from the combined WorldClim-Station data. There was a small mean difference between the paired variables of the two datasets except bioclimatic variable 12 (total annual rainfall). In terms of temperature, the dissimilarity of variable pairs of Bio 3, Bio 8 and Bio 9 are relevant to the monthly maximum and minimum temperatures and the mean temperature of the wettest and driest quarters. The mean difference of temperature-related bioclimatic variables is low, being less than 1 °C, for Bio

1 to Bio 11 but Bio 3 (Isothermality, calculated from Bio 2 and Bio 7) and Bio 4 (Temperature seasonality, standard deviation of the monthly mean temperatures). The mean difference of precipitation-related bioclimatic variables is also low but Bio 12 (annual precipitation). Differences of monthly rainfall bring on a discrepancy of total annual rainfall (Bio 12), where the yearly mean rainfall of local data is higher than 300 mm compared to the WorldClim data, and the rainfall of the warmest quarter (Bio18) of the local data is nearly 60 mm, which is higher than that of WorldClim. Differences in Bio 13, 14, 15, 17 and 19 are found for the wettest, driest months and coldest quarters. These are the periods with the highest, lowest rainfall and lowest temperature of the year, respectively.

Table 3. Paired sample *t*-test for difference and the correlation coefficient significant test for 19 Bioclimatic variables extracted from the WorldClim data and interpolated data. Bold numbers indicate statistical significance

Bio	Paired sample *t*-test for difference (Mean WorldClim data vs. mean interpolated data), n = 100, df = 99, α = 0.05, two tails, *t-critical value* = 1.984. Hypothesis H_0: $\mu = 0$					*t*-test for correlation significance, n = 100, df = 99, α = 0.05, two tails. Hypothesis H_0: $\rho = 0$		
	Mean WorldClim data	Mean interpolated data	Mean	*t*-value	*p*-value	*R*	*t*-value	*p*-value
1	23.773	23.786	-0.013	-0.168	0.867	0.956	32.251	0
2	7.607	7.474	0.132	1.687	0.095	0.735	10.723	0
3	57.989	56.294	1.694	3.123	0.002	0.907	21.308	0
4	182.391	181.116	1.276	0.558	0.578	0.947	29.120	0
5	30.108	30.057	0.051	0.492	0.624	0.924	23.870	0
6	16.796	16.706	0.090	0.948	0.345	0.942	27.670	0
7	13.312	13.351	-0.039	-0.342	0.733	0.806	13.468	0
8	24.025	24.252	-0.227	-2.925	0.004	0.929	24.851	0
9	22.187	21.924	0.263	2.837	0.006	0.948	29.606	0
10	26.061	25.948	0.112	1.169	0.245	0.953	31.115	0
11	20.895	20.850	0.045	0.632	0.529	0.953	31.139	0
12	1520.421	1834.798	-314.378	-35.117	1.1E-57	0.982	50.769	0
13	356.877	357.886	-1.009	-0.540	0.591	0.984	55.389	0
14	16.176	16.140	0.036	0.188	0.851	0.988	62.191	0
15	77.411	77.064	0.347	1.594	0.114	0.981	50.195	0
16	1018.241	1023.362	-5.121	-0.836	0.405	0.977	44.907	0
17	56.344	57.190	-0.846	-1.345	0.182	0.986	59.060	0
18	418.189	474.186	-55.997	-4.431	2.4E-05	0.815	13.899	0
19	130.366	133.696	-3.330	-0.947	0.346	0.954	31.599	0

These dissimilarities are minor and keep the spatial distribution pattern of the climate of the study area as the known climate type of this region. This is because the combined local climatic and WorldClim data for interpolation resulted in slight changes. Though there are minor differences, the general pattern of climatic variables, represented by bioclimatic variables, of the two datasets are similar as reflected in the coefficient of correlation being above 0.8 (R = 0.73 in Bio 2 only) and significant tests with high *t*-values were and low *p*-values (with a significant level of 95% and α = 0.05).

Repeating the paired sample *t*-test with synoptic data of the 21 stations is presented in Table 4. The results show that there was a slight difference in temperature among the first three datasets, but it was not statistically significant. This indicates that the multiple linear regression used for the temperature interpolation is a proper statistical model for generating the surface temperature variables. Comparing the precipitation of the three datasets shows a statistically significant difference for 100 random samples; this is just like that found in the previous test. However, this examination also reveals a smaller difference between the interpolated precipitation data and the station data than that between the WorldClim data and the station data. This implies that the interpolation has been improved and is close to the observed data.

Table 4. Paired sample T-test for annual precipitation (Bio 12) and mean annual temperature (Bio 1) with and without the 21-station data. n = 21, df = 20, two tail, α = 0.05, t-critical = 2.0859. Hypothesis H_0: μ = 0

Station	Interpolation with World bioclimatic data only		Interpolation with combined datasets		World bioclimatic data		Station data	
	Annual Prep.	Mean	Annual Prep.	Mean	Annual Prep.	Mean	Annual Prep.	Mean
An Khe	1508	24.074	1597	24.075	1535	24.41	1598	22.94
Ba To	1694	25.741	3667	25.742	1784	25.98	3551	25.45
Ban Me Thuat	1651	24.302	1866	24.302	1685	24.32	1862	23.89
Bao Loc	2470	22.664	2923	22.663	2575	21.11	2923	21.76
Cam Ranh	1052	26.232	1394	26.234	1091	26.77	1393	27.22
Da Lat	1750	18.917	1837	18.914	1787	18.59	1839	18.05
Da Nang	1993	25.716	2248	25.718	2052	26.46	2247	25.97
Dak To	2103	23.236	1807	23.236	2105	23.59	1809	22.6
EaKmat	1616	23.889	1933	23.889	1623	23.16	1934	23.53
Hoai Nhon	1645	25.872	2212	25.874	1676	26.5	2210	26.22
Kon Tum	1881	23.683	1907	23.683	1905	24.67	1905	23.95
Lien Khuong	1604	21.691	1603	21.690	1667	19.72	1602	21.36
Mdrak	1302	24.115	2233	24.115	1311	24.15	2230	24
Nha Trang	1083	26.208	1490	26.210	1195	26.81	1490	26.94
Phan Rang	1009	26.405	927	26.406	1007	27.11	927	27.2
Pleiku	2390	22.496	2155	22.496	2453	22.73	2157	22.06
Quang Ngai	2025	25.763	2529	25.765	2090	26.33	2530	26.09
Qui Nhon	1497	25.919	1869	25.921	1472	26.7	1867	27.27
Son Hoa	1613	26.034	1865	26.035	1366	26.7	1866	26.12
Tam Ky	1961	25.793	2911	25.795	2007	26.25	2912	25.91
Tuy Hoa	1869	26.070	2072	26.071	1457	27.04	2076	26.89
Mean	1701	24.515	2050	24.516	1707	24.719	2044	24.544
Mean difference	-343	-0.028	+6	-0.028	-337	+0.175		
t-calculated	-3.349	-0.195	1.009	-0.191	-3.326	1.204		
p-calculated	*0.003*	**0.846**	**0.325**	**0.850**	*0.003*	**0.242**		

Notes: Italic numbers: significant difference; **Bold numbers**: Insignificant difference.

Performance of high spatial resolution

As illustrated, the grid data of the Bioclim 1 of 0.00833° resolution overlaid on the raster of the interpolated Bio 1 of 0.00226° displayed more details on temperature on the Ngoc Linh Mountaintop

(107.979°E, 15.067°N) (Fig. 4a). The spatial distribution of the mean annual precipitation (Bio 12) is characterized as a trend of decrease from west to east and increase from south to north of the CH-SC, which is specifically along the coastal zone (Fig. 4b). Figure 4b also displays a difference of precipitation between the interpolated data and the WorldClim data along with the cross profile of topography, and the interpolated data show less smooth change than do the WorldClim data.

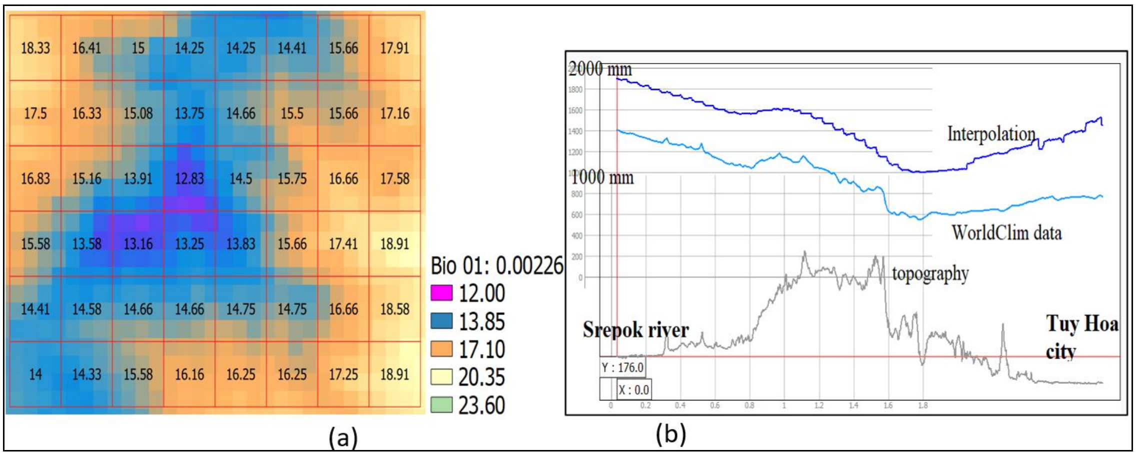

Figure 4. (a) Gridded cells of the Bioclim 1 of 0.00833° overlaid on a raster of the interpolated Bio 01 of 0.00226°. Numeric values on grids are the temperature of the 0.00833° cell (the WorldClim data), and the colour is the temperature of the 0.00226° cell (the new interpolated data). (b) Cross profile of precipitation (the interpolated data and WorldClim data) and topography, at the latitude of 13°N from the Srepok River (at the border of Vietnam and Cambodia, altitude 176 m) to Tuy Hoa City (altitude 7 m) (distance appr. 200 km)

CONCLUSIONS

Bioclimatic variables were derived by the combination of the WorldClim data (0.00833° resolution) with the local climate observed data to construct local climate factors of higher spatial resolution using multiple linear regression and spatial interpolation. Effective and low error regression models were obtained with high correlation and statistical significance (R = 0.91-0.97 and p-values close to zero). The 19 bioclimatic variables of the CH-SC region with the spatial resolution of 0.00226° (~250 m, or one cell covers an area of 1/16 km^2) have been generated and these can be utilized for other localized SDMs instead of using a coarse spatial resolution from the WorldClim database.

Acknowledgements: This research was funded by the National Program on Space Science and Technology, Vietnam Academy of Science and Technology (VAST) by the Contract Agreement VT-UD.09/18–20. We would like to thank Prof. Herbert Covert for editing the English of the manuscript.

REFERENCES

Abatzoglou J. T., Dobrowski S. Z., Parks S. A. Hegewisch K. C., 2018. Data descriptor: TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1985–2015. *Scientific Data*, 5: 170191. https://doi.org/10.1038/sdata.2017.191

Andrews D. G., 2010. An introduction to atmospheric physics (2nd ed). Cambridge University Press, pp. 237.

Bett N. N., Blair M. E., Sterling E. J., 2012. Ecological Niche Conservatism in Doucs (Genus *Pygathrix*). *International Journal of Primatology*, 33: 972–988. doi: 10.1007/s10764-012-9622-3

Bohner J., Benjamin B., 2018. GIS and climatology and meteorology. In: Cova T. J., Tsou M-H. (Eds), GIS Methods and Techniques. University of Hamburg, Hamburg, Germany Elsevier Inc.

Booth T. H., 2018. Why understanding the pioneering and continuing contributions of BIOCLIM to species distribution modelling is important? *Australasian Ecology*, 43: 852–860.

Booth T. H., Nix H. A., Busby J. R., Hutchinson M. F., 2014. BIOCLIM: the first species distribution modelling package, its early applications and relevance to most current MAXENT studies. Diversity and Distributions, 20: 1–9.

Collen B., Whitton F., Dyer E. E., Baillie J. E. M., Cumberlidge N., Darwall W. R. T., Pollock C., Richman N. I., Soulsby A. M., Böhm M., 2014. Global patterns of freshwater species diversity, threat and endemism. *Global Ecological Biogeography* 23: 40–51.

CGIAR Research Centers in Southeast Asia, 2016. The drought crisis in the Central Highlands of Vietnam. Assessment Report. https://core.ac.uk/download/pdf/132684728.pdf (last access Jan 10, 2021).

Daly C., Halbleib M., Smith J. I., Gibson W. P., Doggett M. K., Taylor G. H., Curtis J., Pasteris P.P., 2008. Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *International Journal of Climatology*. https://doi.org/10.1002/joc.1688

Deb J. C., 2016. Assessing species vulnerability to climate change in Tropical Asia: Implications for biodiversity conservation and forest management. PhD thesis. School of Earth and Environment Sciences - University of Queensland, pp. 145.

Deb J. C., Phinn S., Butt S., McAlpine C. A., 2017. The impact of climate change on the distribution of two threatened Dipterocarp trees. *Ecology and Evolution*. https://doi.org/10.1002/ece3.2846

Earthdata: https://earthdata.nasa.gov/learn/articles/nasa-shuttle-radar-topography-mission-srtm-version-3-0-global-1-arc-second-data-released-over-asia-and-australia), last access July 20, 2020.

Earthdata: NASA Shuttle Radar Topography Mission (SRTM) Version 3.0 Global 1 arc second Data Released over Asia and Australia.

Fick S. E., Hijmans R. J., 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*. Wiley Online Library. https://doi.org/10.1002/joc.5086

Franklin J., 2010. Mapping species distributions: spatial inference and prediction, Cambridge University Press.

Hair J. F. Jr., Black W. C., Babin B.J., Anderson R. E., 2014, Multivariate Data Analysis (7th ed.). Pearson New International Edition, pp 739.

Hengl T., 2009. A practice guide to geostatistical mapping. (http://spatial-analyst.net/book/).

Hijmans R. J., Cameron S. E. Parra J. L., Jones P. G., Jarvis A., 2005. Very high-resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 25: 1965–1978.

Hutchinson M. F., 1995. Interpolating mean rainfall using thin plate smoothing splines. *International Journal of GIS*, 9(4): 385–403. https://doi.org/10.1080/02693799508902045

Hutchinson M. F., 1991. The application of thin plate smoothing splines to continent-wide data assimilation. In: Jasper J.D. (Ed), BMRC Research Report No. 27, Data Assimilation Systems, Bureau of Meteorology, Melbourne: 104–113.

John R. B., 1991. BIOCLIM - A Climate analysis and prediction system. In:

Margules C. R., Austin M. P. (Eds), Natural Conservation: Cost effectively biological surveys and data analysis. CSIRO Australia.

Karger D. N. Conrad O., Böhner J., Kawohl T., Kreft H., Soria-Auza R. W., Zimmermann, N. E., Peter Linder H., Kessler M., 2017. Climatologies at high resolution for the earth's land surface areas. Scientific Data 4: 170122. https://doi.org/10.1038/sdata.2017.122

Kriticos D. J., Webber B. L., Leriche A., Ota N., Macadam I., Bathols J., Scott J. K., 2011. Methods in Ecology and Evolution. *British Ecological Society, Methods in Ecology and Evolution*, 3: 53–64.

Lanzante J. R., 1996. Resistant, robust and non-parametric techniques for the analysis of climate data: theory and examples, including applications to historical radiosonde station data. *International Journal of Climatology*, 16: 1197–1226.

Le Coz, M., François, D., Pierre, G., Guillaume, F., 2009. Assessment of Digital Elevation Model (DEM) aggregation methods for hydrological modeling: Lake Chad basin, Africa. *Computers & Geosciences*, 35(8): 1661–1670.

Liu C., White, M., Newell, G., Griffioen, P., 2014. Species distribution modelling for conservation planning in Victoria, Australia. *Ecological Modelling* 249: 68–74. https://doi.org/10.1016/ j.ecolmodel. 2012.07.003

Loomis S. E., James M. R., Dirk V., Carrie M., De Cort G., Jaap, S. S. D., Daniel O., Hilde, E., Street-Perrott, F. A., Meredith A. K., 2017. The tropical lapse rate steepened during the Last Glacial Maximum. *Science Advances*, 3(1). https://doi.org/10.1126/sciadv.1600815

Ly S., Charles C., Degré A., 2013. Different methods for spatial interpolation of rainfall data for operational hydrology and hydrological modeling at watershed scale. A review. *Biotechnol. Agron. Soc. Environ.*, 17(2): 392–406.

Maeda M., Yasutomi N., Yatagai A. & National Center for Atmospheric Research Staff, 2020. The Climate Data Guide: APHRODITE: Asian Precipitation - Highly-Resolved Observational Data Integration Towards Evaluation of Water Resources. (Eds). Retrieved from https://climatedataguide.ucar.edu/climate-data/aphrodite-asian-precipitation-highly-resolved-observational-data-integration-towards.

Marchi M., Sinjur I., Bozzano M., Westergren M., 2019. Evaluating WorldClim Version 1 (1961–1990) as the Baseline for Sustainable Use of Forest and Environmental Resources in a Changing Climate. *Sustainability*, 11(3043). https://doi.org/10.3390/su11113043

Ninyerola M., Pons, X., Roure J.M., 2000. A methodological approach of climatological. *Journal of Climatology*, 20: 1823–1841.

Nguyen D.N., Nguyen T. H., 2004. Climate and Climate Resources in Vietnam. Agricultural Publishing House, Hanoi, Vietnam, pp. 296 (in Vietnamese).

Nguyen D. Q., Renwick J., McGregor J. 2013. Variations of surface temperature and rainfall in Vietnam from 1971 to 2010. *International Journal of Climatology* 34(1): 249–264.

Nguyen D. Q., Renwick J., McGregor J. 2014. Variations of monsoon rainfall: A simple unified index. *Geophysical Research Letters* 41(2): 575–581.

Nguyen Tuan Anh, Le Duc Minh, Pham Viet Hung, Vu Thi Duyen, 2019. Modeling the Red-shanked Douc (*Pygathrix nemaeus*) distribution in Vietnam using Maxent. *VNU Journal of Science: Earth and Environmental Sciences*, 35(3): 61–71.

Peterson D. L., Schreiner E. G., Buckingham N. M., 1998. Gradients, vegetation and climate: spatial and temporal dynamics in the Olympic Mountains, U.S.A. *Global Ecology and Biogeography Letters*, 6(1): 7–17.

Fink A. H., Phan V. T., Pinto J., van der Linden R., Schubert D., Trinh T. L. and Ngo-Duc T., 2014. Climate change projections and selected impacts for Vietnam. In: Meon G., Pätsch M., Phuoc N. V. and Quan N. H. (eds.) EWATEC-COAST: Technologies for Environmental and Water Protection of Coastal Zones in Vietnam. EWATEC-COAST Working Group, pp. 35–56. ISBN 978-3-95404-852-6.

Porfirio L. L., Rebecca M., Harris B., Lefroy E. C., Hugh S., Gould S. F., Lee G., Bindoff N. L., Mackey B., 2014. Improving the Use of Species Distribution Models in Conservation Planning and Management under Climate Change. *PLoS One*, 9(11): e113749. https://doi.org/10.1371/journal.pone.0113749

Prajapati A., Sapan N., Sheetal M., 2012. Evaluation of Different Image Interpolation Algorithms. *International Journal of Computer Applications*, 58(12): 6–12.

QGIS Development Team, 2019. QGIS Geographic Information System. Open Source Geospatial Foundation Project. http://qgis.osgeo.org.

Raghavan R. K., Goodin D. G., Hanzlicek G. A., Zolnerowich G., Dryden M. W., Anderson G. A., Ganta R. R., 2016. Maximum entropy-based ecological niche model and bio-climatic determinants of Lone Star Tick (*Amblyomma americanum*) Niche. *Vector-Borne and Zoonotic Diseases*, 16(3): 205–211.

Regalado J. C., Nguyen Tien Hiep, Phan Ke Loc, Averyanov L., Harder D. K., 2005. New insights into the diversity of the Flora of Vietnam. *Biologiske Skrifter*, 55: 189–197.

Sharma V., Kilic A., Irmak S., 2016. Impact of scale/resolution on evapotranspiration from Landsat and MODIS images. *Water Resources Research*, 52(3): 1800–1819. doi:10.1002/2015WR017772.

Shen Y-J., Yanjun S., Jason G., Alexander B., 2016. Spatial-temporal variation of near-surface temperature lapse rates over the Tianshan Mountains, central Asia. *Advancing Earth and Space Science*, 121(23): 14006–14017. https://doi.org/10.1002/2016JD025711

van Schingen M., Ha Q. Q., Pham C. T., Le T. Q., Nguyen T. Q., Bonkowski M., Ziegler T., 2016. Discovery of a new crocodile lizard population in Vietnam: Population trends, future prognoses and identification of key habitats for conservation. *Revue Suisse de Zoologie*, 123(2): 241–251.

Xu T., Hutchinson M., 2010. Anuclim Version 6.1 User guide. The Australian National University - Fenner School of Environment and Society.

Yang X., Xie X., Liu D. L., Ji F. and Wang L., 2015. Spatial Interpolation of Daily Rainfall Data for Local Climate Impact Assessment over Greater Sydney Region. *Advances in Meteorology*, 2015: Article ID 563629. doi: 10.1155/2015/563629

Yohe L. R., Flanders J., Hoang D. M., Vu L., Phung T. B., Nguyen Q. H., Sushma Reddy S., 2014. Unveiling the impact of human influence on species distributions in Vietnam: a case study using babblers (Aves: Timaliidae). *Tropical Conservation Science*, 7(3): 586–596.

Wagner M., Trutsching W., Bathke A. C., Ruprecht U., 2018. A first approach to calculate BIOCLIM variables and climate zones for Antarctica. *Theorical Applied Climatology*, 131: 1397–1451. https://doi.org/10.1007/s00704-017-2053-5

ANU: https://fennerschool.anu.edu.au/research/products/anuclim

WorldClim[1]: https://worldclim.org

WorldClim[2]: https://worldclim.org/data/v1.4/formats.html

WorldClim[3]: https://worldclim.org/data/bioclim.html