



Empirical modelling and spatiotemporal analysis of average daytime air temperature for outdoor work in Bulgaria

Krastina Malcheva*, Tania Marinova

*National Institute of Meteorology and Hydrology,
Tsarigradsko shose 66, 1784 Sofia, Bulgaria*

Abstract: A methodology has been developed to estimate the average daytime air temperature by the forecasted daily maximum and minimum temperatures in the context of the Ministry of Labour and Social Policy of Bulgaria regulations for outdoor work in unfavorable temperature conditions. In accordance with the assignment, the average daytime temperature is defined as the average of the five air temperature measurements in synoptic stations between 06 UTC and 18 UTC. An empirical model using the multiple linear regression method, with daily maximum and minimum temperature as explanatory variables, has been created by data for the period 2001–2016 from 28 synoptic stations located in non-mountainous regions of the country. The accuracy of the model in predicting average daytime temperatures below 10 °C and above 30 °C has been evaluated by data for 2017. The study also analyzed the spatial variability of the average daytime temperature and regression model coefficients.

Keywords: daytime air temperature, exploratory data analysis, multiple linear regression method, unfavorable temperature conditions

1. INTRODUCTION

Working outdoors in hot or cold weather can have different adverse effects on human health, decrease work productivity and increase accidents, injuries and fatalities. Identifying heat- and cold-related risks is the first step in proper outdoor working process management (Mäkinen&Hassi, 2009). Changing climate and increasing frequency of extreme temperature events require adaptive governance to ensure healthy conditions at work. Bulgarian legislation provides many regulations in the field of occupational safety and health that are harmonized with European Union laws. Ordinance No. 11 of 21 December 2005 on determining the terms and conditions for providing free

* krastina.malcheva@meteo.bg

food and/or supplements to it (DV, No. 1, 2006) requires the employers to provide suitable food and drinks when working outdoors in adverse thermal conditions, but for years the Guidelines for the implementation of the Ordinance No. 11 had missed methodology for estimating daytime air temperature. Therefore, it was assigned to the National Institute of Meteorology and Hydrology (NIMH) to propose an approach and straightforward methodology for calculating expected average daytime temperatures for a few days ahead using the forecast of daily minimum and maximum temperatures by cities, available online at <https://meteo.bg/bg/forecast/gradove>. The daytime hours were defined to coincide with day working hours – from 06:00 to 22:00 Eastern European Time (EET). The adverse thermal conditions for outdoor work are determined explicitly (Guidelines for the implementation of Ordinance No. 11, 2006): “working at average daytime temperatures below +10 °C and above +30 °C for more than half of the maximum working hours established by the Labour Code”.

2. DATA AND METHODS

Since the climate regionalization of the country differs significantly from the administrative one, we considered it inappropriate to derive separate formulas for the cities listed on the website because doing so would result in a complicated methodology and significant errors if employers tried to utilize these formulas for other settlements, even in the same administrative district. The specific assignment was framed as the development of a simple-to-use empirical formula that allows the calculation of average daytime air temperature for the next few days with acceptable accuracy for the entire non-mountainous part of the country (below 800 m a.s.l.) and throughout all seasons using only the forecasted daily maximum and minimum temperatures. Therefore, the well-known time series forecasting methods could not be applied. Most of the meteorological information in the NIMH archive consists of data from conventional measurements at a time scale determined by the observation program – every 3 hours for synoptic stations and 3 times a day for climatic stations. The closest to the actual average air temperature for a specific time interval, such as from 6 to 22 EET, can be obtained only using data from synoptic stations.

Bearing in mind these considerations, the following approach was adopted to solve the task:

1. Synoptic stations representative of the climate in the non-mountainous parts of the country, with at least 85% available data in recent decades, were selected.
2. Average daytime temperature (ADT), defined as the average of the five air temperature measurements in synoptic stations from 06 to 18 UTC (08 to 20 EET), was calculated for each day and station.
3. The multiple linear regression (MLR) method, considering daily maximum and minimum temperature as explanatory variables, was applied to the entire dataset to derive the empirical formula for ADT calculation.

4. The suitability of the obtained formula to predict ADT in the critical intervals below 10 °C and above 30 °C was estimated by a test dataset.

Data from selected climatic and automatic meteorological stations were processed and included in the study to provide an adequate climatology and spatiotemporal analysis of ADT. The selection of meteorological stations was based on their climatological representativeness and the quality control criteria for air temperature data and derived time series (WMO, 2017). The inspection of observational data revealed that in recent decades, almost all climatic stations from the meteorological network of NIMH met these criteria. For automatic stations, the acceptable period is since 2012. We chose 2001–2017 as a long enough recent period to obtain reliable statistics and good MLR performance. The regression model was built on the dataset comprising daily data from all selected synoptic stations for 2001–2016 and has been tested on the respective dataset for 2017.

Figure 1 shows the location of stations used in the study: 28 synoptic (labeled as S1 to S28), 66 climatic, and 35 automatic stations. Table 1 presents the distribution of meteorological stations in the four climate subareas, according to the Sabev&Stanev (1959) climate regionalization.

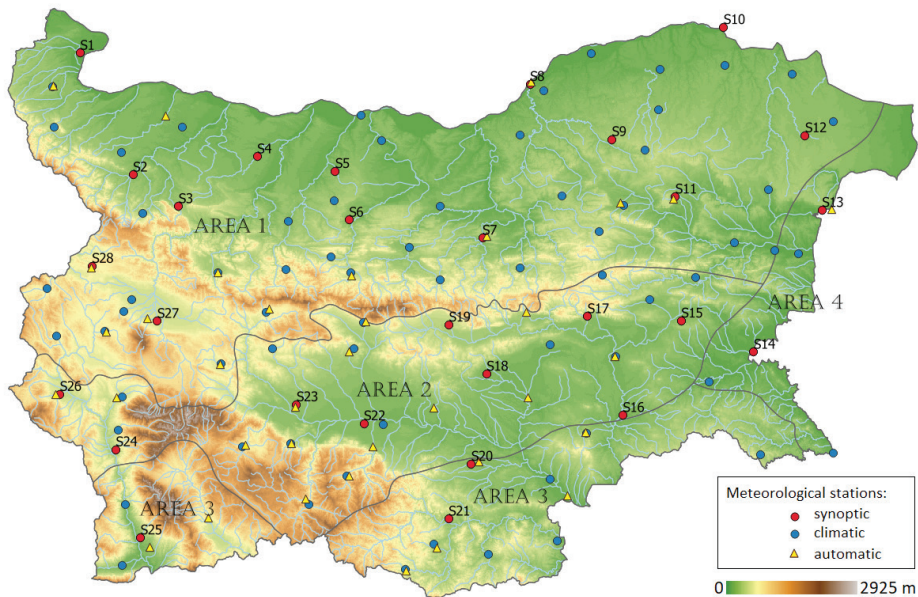


Fig. 1. Hypsometric map of Bulgaria with locations of meteorological stations used in the study: synoptic (red circles), climatic (blue circles), and automatic (yellow triangles); the boundaries of the main climate subareas, according to the Sabev&Stanev (1959) climate regionalization, are shown in dark grey (subareas are labeled as AREA1 to AREA4).

Synoptic and climatic stations cover the country's territory relatively evenly, but not the automatic ones – they are located mainly along the river valleys in Southern Bulgaria, where the Mediterranean climate affects the temperature regime more strongly. Therefore, one can hypothesize that climatology, summarized for the whole territory, will return biased results. But, on the other hand, some climatic peculiarities commented on by Sabev&Stanev (1959) should be considered: 1) the climate south of the Balkan Mountains is distinguished by a much greater diversity; 2) the temperature differences between northern and southern lowlands almost disappear during the summer months; 3) the climate of the country features with clear enough continentality even in southern regions. To preserve the climatological representativeness of the study, since the hourly data from automatic weather stations (AWS) are needed to obtain the diurnal course of unfavorable thermal conditions during different months, we selected AWS as close as possible to the chosen conventional stations.

Table 1. Distribution of the used in the survey meteorological stations by climate subarea; the number of synoptic, climatic and automatic stations is labeled with N_{syn} , N_{clim} and N_{AWS}

Climate subarea	Label	Synoptic stations	N_{syn}	N_{clim}	N_{AWS}
Moderate-continental	AREA1	S1÷S12, S27, S28	14	39	13
Transitional-continental	AREA2	S15÷S19, S22÷S24, S26	9	15	14
South-Bulgarian	AREA3	S20, S21, S25	3	8	7
Black-Sea	AREA4	S13, S14	2	4	1

Multiple Linear Regression is a method for statistical modeling of linear relationships represented by well-known equations (Montgomery&Runger, 2003):

$$y_i = \beta_0 + \sum_{j=1}^k \beta_j x_{ij} + \epsilon_i = \mathbf{x}_i^T \boldsymbol{\beta} + \epsilon_i, \quad i = 1, 2, \dots, n \quad (1)$$

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad (2)$$

where \mathbf{y} is the dependent variable vector of size $(n,1)$; \mathbf{X} is the matrix of k explanatory variables of size (n, p) ; $\boldsymbol{\beta}$ is the vector of regression coefficients $(p,1)$; $\boldsymbol{\epsilon}$ is the vector of random errors $(n,1)$; $p = k + 1$ is the number of parameters to be estimated.

Although there is no universal solution, selecting appropriate metrics for assessing the performance and accuracy of regression models is essential for a number of applied tasks. We chose three of the most widely used metrics for the present study:

1) Adjusted R-squared (R_{adj}^2) is a goodness-of-fit measure accounting for the number of explanatory variables in the regression model.

$$R_{adj}^2 = 1 - \frac{SSE/(n - p)}{SST/(n - 1)}, \quad (3)$$

where: $SSE = \sum_{i=1}^n (\hat{y}_i - y_i)^2$ is the sum of squared errors (residuals); $SST = \sum_{i=1}^n (y_i - \bar{y})^2$ is the total sum of squares; $(n - p)$ denoted the degrees of freedom of the model; \bar{y} and \hat{y}_i are the mean and the i -th estimated values of y .

2) Standard error of regression (SE) is a measure showing the average deviation of errors in the regression model:

$$SE = \sqrt{\frac{SSE}{n - p}}, \quad (4)$$

3) The mean absolute percentage error (MAPE) is a commonly used performance metric defined as the mean of absolute relative errors (Myttenaere et al., 2016):

$$MAPE = \frac{100}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right|, \quad (5)$$

The MLR analysis is based on several key assumptions (Montgomery&Runger, 2003):

- 1) Linear relationship between the dependent and independent variables, which can be visually estimated by scatter plots.
- 2) Acceptable multicollinearity between the independent variables, usually evaluated using the variance inflation factor (VIF).
- 3) Normally distributed residuals, typically assessed by a Q-Q plot.
- 4) Consistent variance of residuals across all predicted values (homoscedasticity) – can be assessed visually by plotting residuals vs. predicted values
- 5) Residuals should not be correlated with each other, which can be visually estimated by autocorrelation plots.

Exploratory data analysis (EDA) allows the preliminary quality evaluation of large datasets. Outliers can distort the regression model significantly. Therefore, they must be analyzed and carefully handled (by correcting or removing) before building the model. Some frequently used techniques for outlier detection are based on standard deviation (z-score) or threshold values (Smith-Miles, 2011). Applying EDA to each of the used datasets is described in the relevant sections of the paper.

All computations and scripts were made using freely available software R 3.6.2 (R Core Team, 2019) and RStudio 1.2 (RStudio Team, 2019). The maps were made using Geostatistical Analyst tools in ArcGIS Pro 2.4 (Environmental Systems Research Institute, 2019) and the digital surface model AW3D30 of the Japan Aerospace Exploration Agency (JAXA EORC homepage, 2019).

3. RESULTS

3.1. Modeling average daytime temperature by data from synoptic stations

As the process of encoding and digitizing daily synoptic data is associated with a larger amount of technical errors due to the significant volume of information and many fields to be filled in, daily data from climate observations in synoptic stations were used as control samples. The values of ADT, calculated as the average of measurements at synoptic hours from 08 to 20 EET (*tds*), were compared with the corresponding values, calculated as the average of measurements at climate observation hours 07, 14 and 21 local time (*tdc*). In the same manner, readings of maximum temperature at 20 EET (*tx20*) and minimum temperature at 08 EET (*tn08*) were compared with the measured daily maximum (*tmax*) and daily minimum (*tmin*) temperature. The frequency distribution and main statistics of computed differences are presented in Figure 2.

Outliers are often due to technical errors. However, high kurtosis can indicate heavy-tailed distributions and requires caution when using tools and methods that assume a normal distribution. According to Chebyshev’s inequality, at least 88.89% of the values lie within three standard deviations of the mean (Amidan et al., 2005; Alsmeyer, 2011). The appropriate thresholds for filtering discordant observations were determined as ± 3 °C for the differences *tn08-tmin* and *tx20-tmax* and $-2/+5$ °C for *tds-tdc*.

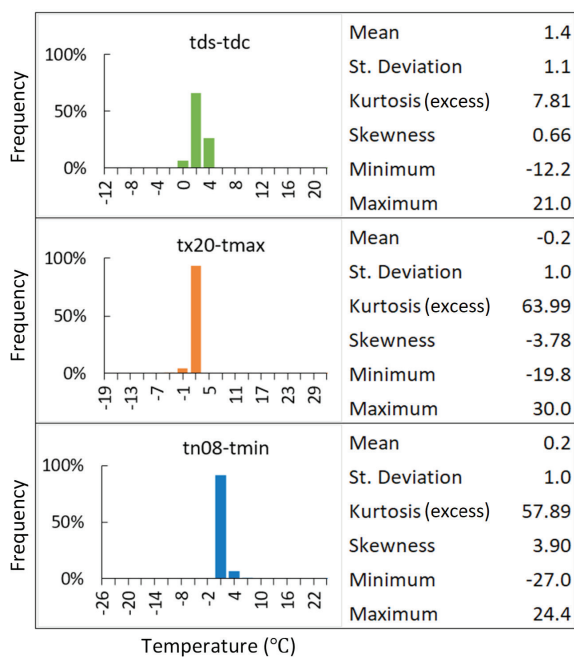


Fig. 2. Main statistics of differences between daytime, maximum and minimum air temperatures calculated by synoptic and climate observation data (2001–2016).

In fact, 99.7% of *tds-tdc* values and about 97.5% of *tn08-tmin* and *tx20-tmax* values fall within the respective ranges of $-2\text{ }^{\circ}\text{C}$ to $5\text{ }^{\circ}\text{C}$ and $-3\text{ }^{\circ}\text{C}$ to $3\text{ }^{\circ}\text{C}$. The thorough analysis showed that, excluding the technical errors, the larger differences correspond to larger deviations from the long-term average weather conditions associated with rapid weather changes. Therefore, all dates with missing data or differences beyond the specified ranges were excluded from the dataset.

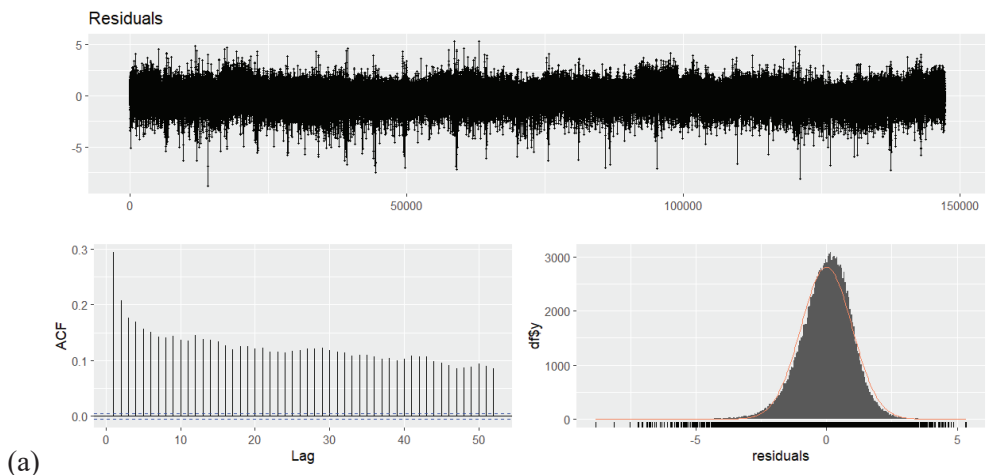
The empirical formula has been derived by the filtered dataset using the MLR model considering the average daytime temperature (*tds*) as a response (dependent) variable and daily maximum (*tmax*) and minimum temperature (*tmin*) as explanatory variables:

$$td = 0.71tmax + 0.31tmin - 1.2, \quad (6)$$

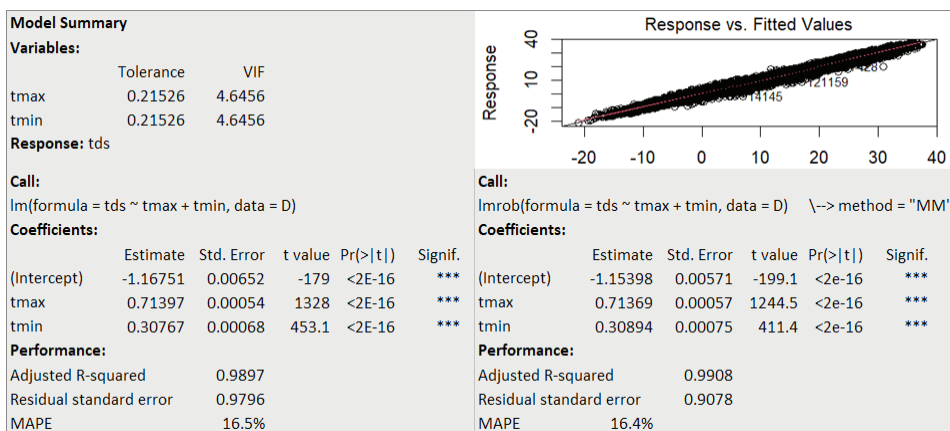
where *td* is the estimated ADT value.

The MLR summary statistics are shown in Figure 3b (left). The adjusted R-squared is 0.9897, the standard error of the regression (SE) is $1\text{ }^{\circ}\text{C}$ (to be precise, $\sim 0.98\text{ }^{\circ}\text{C}$), and the mean absolute percentage error (MAPE) is 16.5% computed for non-zero values of *tds* (zeroes are approximately 0.2% of all *tds* values). The calculated VIF value is below 5, indicating an acceptable multicollinearity. However, the Q-Q plot of residuals vs. normal distribution and additional plots and tests reveal heavy tails, influential points and autocorrelation (Figure 3a). All of that can lead to unreliable standard errors and confidence intervals, misleading conclusions about the significance of relationships between variables and inefficient predictions.

The robust regression procedure ‘lmrob’ implementing the MM-estimator proposed by Yohai (1987) from robustbase R-package (Rousseeuw et al., 2015) was used in order to reduce the influence of the discordant observations and to mitigate some of the LSE inconsistencies (Croux et al., 2003). To avoid autocorrelation, we applied this function repeatedly on samples of randomly ordered but evenly picked (with a varying step) data. The medians of coefficients of samples-based regression, namely 0.71371 and 0.30905 for the slopes and -1.14617 for the intercept, are very close to values obtained by applying the ‘lmrob’ function over the entire dataset (Figure 3b, right). The median of calculated regression standard errors is $\sim 0.9\text{ }^{\circ}\text{C}$. The average bias after coefficients rounding, as in Eq. 6, is $-0.12\text{ }^{\circ}\text{C}$. The prediction accuracy of the model (denoted hereafter as *Model 1*) is $\pm 1.2\text{ }^{\circ}\text{C}$ (80% prediction interval) and $\pm 1.8\text{ }^{\circ}\text{C}$ (95% prediction interval).



(a)



(b)

Fig. 3. MLR summary (a) plots and (b) statistics; plots in (a) are produced by the 'checkresiduals' function of the R-package 'forecast' (Hyndman&Khandakar, 2008)

Although the regression model demonstrated good performance on the entire temperature range of *tds*, special attention is needed on predicting ADT in the critical intervals defined in regulations – below 10 °C and above 30 °C. The suitability of the obtained empirical formula was assessed by data from the same 28 synoptic stations for 2017. The year was distinguished by a prolonged retention of extreme temperatures. Winter was cold, with three cold spells in January. Minimum temperatures in some parts of West Bulgaria dropped below -25 °C. Two severe heat waves were registered in late June and August when maximum temperatures exceeded 42 °C in some places.

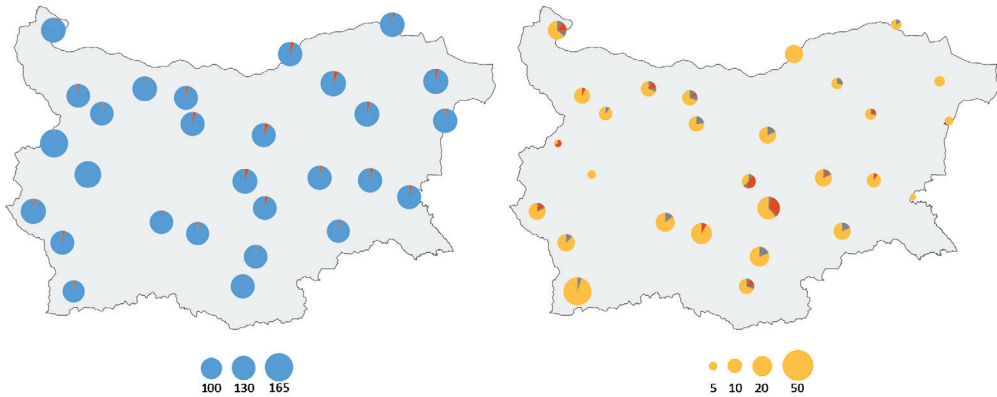


Fig. 4. Number of cases of poorly predicted ADT, with deviations larger than 1 °C (in red) and 0.5 °C (in dark grey), presented as a portion of the total number of cases of actual ADT below 10 °C (left panel) and above 30 °C (right panel) for 2017 by stations; the size of circles is proportional to the number of cases of actual ADT falling in the critical intervals.

As seen in Figure 4, the prediction is more successful in cold weather. The relative error in hot weather is higher in lowlands in the northwestern and central parts of the country and in some regions where extremely hot weather is not typical. On average, 8.2% of predicted values are less than 29 °C when actual ADT is above 30 °C, and 2.1% are higher than 11 °C when actual ADT is below 10 °C. Correspondingly, 19.2% of predicted values are less than 29.5 °C for ADT > 30 °C, and 3.6% are higher than 10.5 °C for ADT < 10 °C.

The detailed analysis of hot weather conditions showed that if $31 \geq \text{ADT} < 32$ °C, the prediction is 77% successful, and if $\text{ADT} \geq 32$ °C, it is 100% successful. Such ADT values were reached during the 2017 heat waves. As minimum temperatures above 20 °C (tropical nights) are not typical for most of the country's inland, they cause significant thermal stress and should be considered potentially dangerous for health, especially when followed by extremely high temperatures during the day. In 2017, the tropical nights exceeded 50% for $\text{ADT} \geq 32$ °C but remained around or below 30% for the lower ADT values.

The relationship between the occurrence frequency of ADT in the critical intervals and measured daily maximum and minimum air temperature in the period 2001–2016 follows clear but non-linear patterns (Figure 5). The analysis of adverse daytime temperature conditions across the country, presented in section 3.3, allowed us to suggest additional rules for the application of the formula (Eq. 6) in practice:

- If forecasted maximum temperatures are below 13 °C, it is assumed that the average daytime air temperature shall be below 10 °C;
- If forecasted minimum temperatures are above 24 °C, it is assumed that the average daytime air temperature shall be above 30 °C.

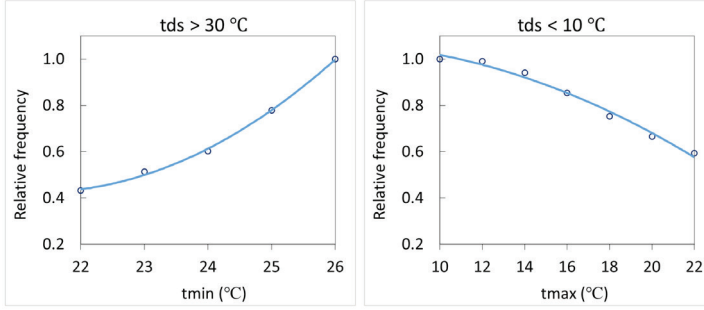


Fig. 5. Relationship between the frequency of occurrence of ADT in the critical intervals and measured daily maximum and minimum air temperature in the period 2001–2016.

3.2. Modeling average daytime temperature by data from climate observations

The diurnal course of air temperature is related to the daily course of incoming solar radiation through the surface energy balance equation. Various methods have been developed to estimate hourly air temperature by combining sinusoidal and exponential models and fitting them to the air temperature diurnal cycle, accounting for the relationship between incoming solar radiation and diurnal temperature range (DTR). Using data only for daily maximum and minimum temperature, Hungerford et al. (1989) derived an empirical formula for estimating daylight average air temperature (T_{ave}) by integrating the sine function modeling the diurnal temperature cycle over three quadrants (from $-\pi/2$ to π):

$$T_{ave} = T_{mean} + TEMCF(T_{max} - T_{mean}), \quad (7)$$

where $T_{mean} = (T_{max} + T_{min})/2$ is the daily average air temperature; T_{max} and T_{min} are daily maximum and minimum temperatures; TEMCF is a local temperature correction factor, with the lowest value at stations where the diurnal temperature curve more closely approximates the sine function. It is evident that the difference $T_{max} - T_{mean}$ equals half of DTR. This formula is successfully used for extrapolating daylight average temperatures over complex mountain regions in the USA by data from the neighbor lowland stations.

Solantie (2004) proposed a method for calculating daytime temperature sums, which is accurate enough concerning spatial and seasonal variability and can be used for various climatic applications. Following this approach, the average daytime temperature Td for a specific daytime interval can be given as:

$$Td = ta + (1 - d)aDTR, \quad (8)$$

where ta is the daily average air temperature; $DTR = tmax - tmin$; d is the ratio of daytime hours to twenty-four hours (since we calculate the average daytime temperature over a

fixed 12-hour interval, the value of d is 0.5); a is an empirical coefficient accounting for environmental features.

We calculated ta using the formula adopted by NIMH:

$$ta = (t_{7:00} + t_{14:00} + 2t_{21:00})/4, \tag{9}$$

where $t_{(7:00)}$, $t_{(14:00)}$ and $t_{(21:00)}$ denote measured air temperature at climate observation hours.

The main statistics of the difference $tds-ta$ and DTR calculated by daily data from synoptic stations (S1 to S28) during the period 2001–2016 are shown in Figure 6.

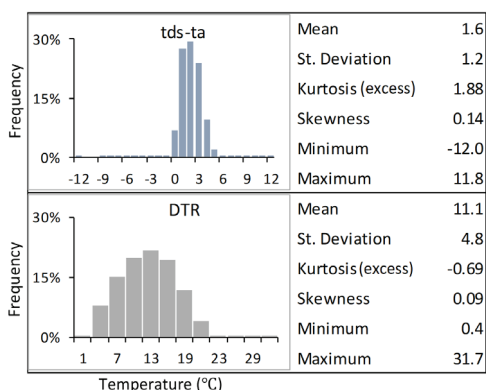


Fig. 6. The main statistics of $tds-ta$ and DTR for the period 2001–2016, calculated by daily data from synoptic stations.

The appropriate thresholds for filtering discordant observations so that at least 88.89% of the values lie within three standard deviations of the mean, according to Chebyshev’s inequality, are $-2/+5.5$ °C for $tds-ta$ and $0/+25$ °C for DTR . So, all dates with missing data or deviations beyond the specified ranges were excluded from the further analysis.

Figure 7 shows the scatter plots of $tds-ta$ vs. DTR by month. The slope of the regression line k reaches its lowest values (0.08-0.11) in winter and highest values (0.15-0.17) during the warm half-year (April–October).

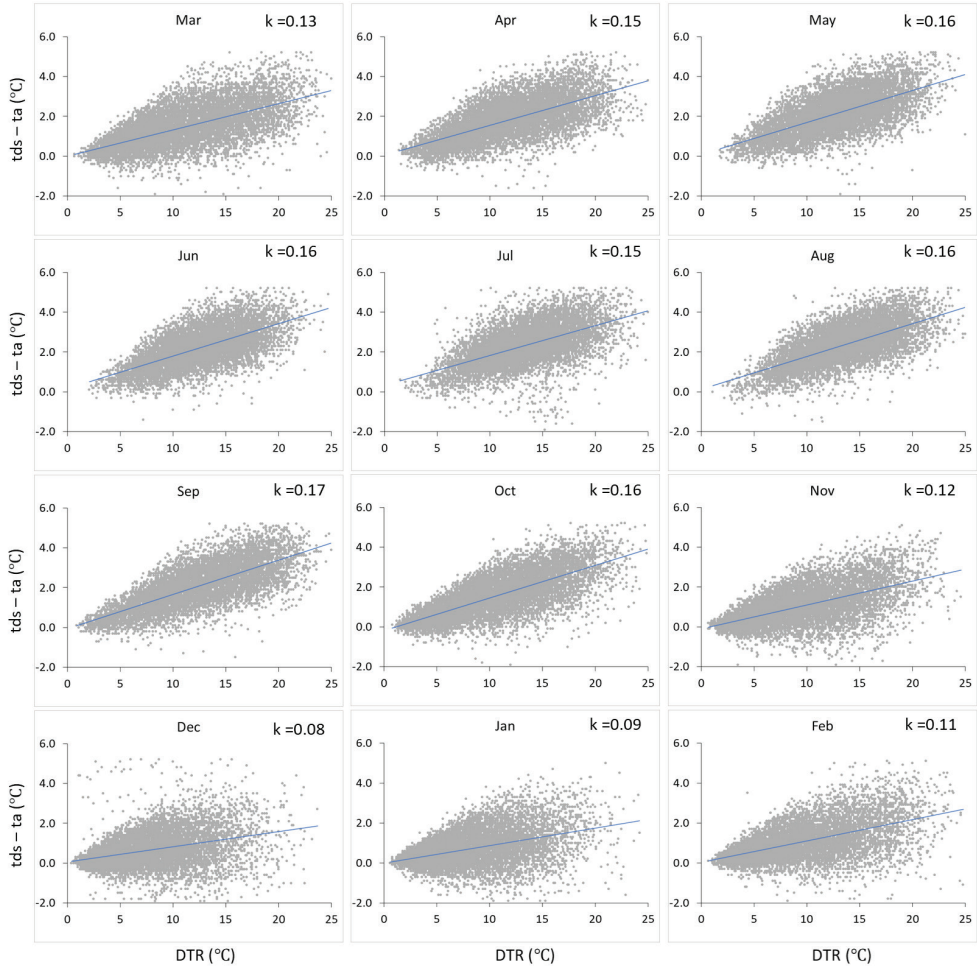


Fig. 7. Scatter plots of the difference between the average daytime and average daily air temperature ($tds-ta$) vs. diurnal temperature range (DTR) by month.

As seen in Figure 8 (left), the correlation between $tds-ta$ and DTR also varies by month and season – it is lower in winter and higher in autumn. The analysis by stations revealed that the intra-annual variability of k is larger in the mountains' foothills and the eastern part of the country (Figure 8, right).

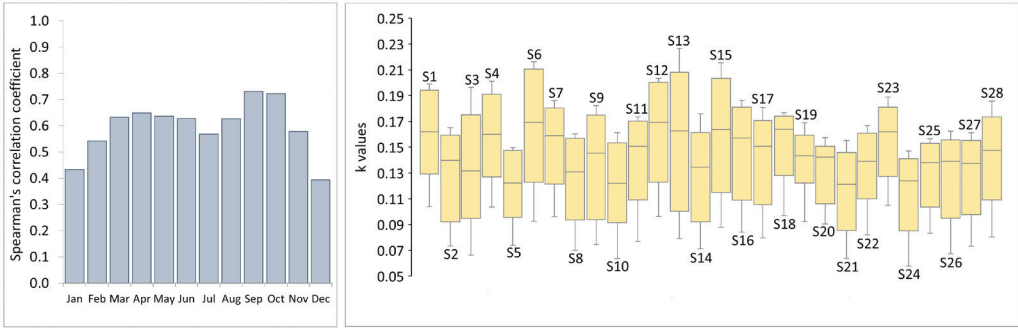


Fig. 8. Left: Spearman's correlation coefficients between *tds-ta* and *DTR* by months; Right: box plots of the intra-annual distribution of the *k* values summarized by stations.

Indeed, 99.7% of *DTR* values and 99.4% of *tds-ta* values fall within the ranges of 0.4 °C to 23 °C and -1.5 °C to 5 °C, respectively. Using filtered data, we built a robust regression model (*Model 2*) for estimating the average daytime temperature (*td*) over the entire dataset, considering the average daily temperature (*ta*) and the diurnal temperature range (*DTR*) as explanatory variables. The adjusted R-squared is 0.994, the standard error (computed as a median of samples-based regression SEs) is ~0.73 °C, and the mean absolute percentage error is 13.4%.

The derived empirical formula has the form:

$$td = 1.03ta + 0.14DTR - 0.3 \tag{10}$$

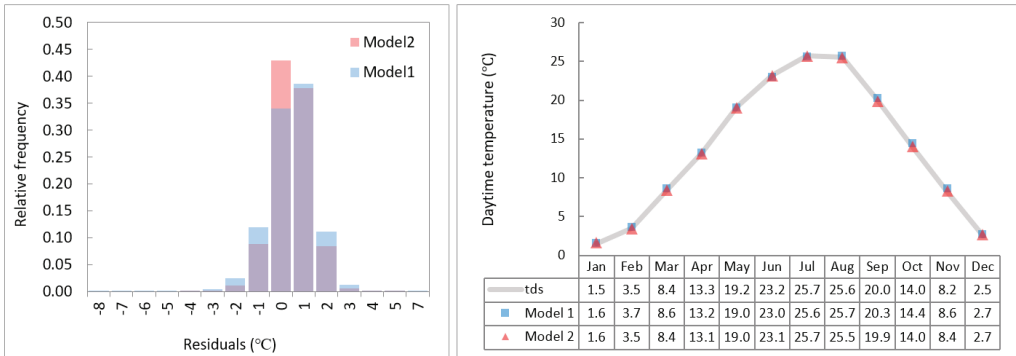


Fig. 9. Left: Distribution of residuals. Right: Annual cycle of ADT computed by observational data (*tds*) and modeled data.

Obviously, *Model 2* performs better than *Model 1*, which can also be illustrated by the frequency distributions of residuals (Figure 9, left), but both models accurately reproduced the climatological annual cycle of ADT (Fig. 9, right). Moreover, practically

the same formula, as Eq. 6, was obtained when we rebuilt *Model 1* by substituting *tds* with ADT modeled by *Model 2*.

Therefore, we could obtain a reliable climatology of the average daytime temperature and analyze the spatial distribution of regression coefficients by applying *Model 1* individually for each synoptic and climatic station using ADT modeled by *Model 2* as a dependent variable.

Figure 10 shows the spatial distribution of the ADT multiyear mean (2001–2016) for January and July, the coldest and warmest month of the year. In January, the average daytime temperature is negative in hilly regions of Western Bulgaria and some places in the Danube Plain ($-2\text{ }^{\circ}\text{C}$ to $0\text{ }^{\circ}\text{C}$); ADT is higher than $3\text{ }^{\circ}\text{C}$ in the southern lowlands (up to $5\text{ }^{\circ}\text{C}$ in the southernmost parts of the Struma Valley and Black Sea coast).

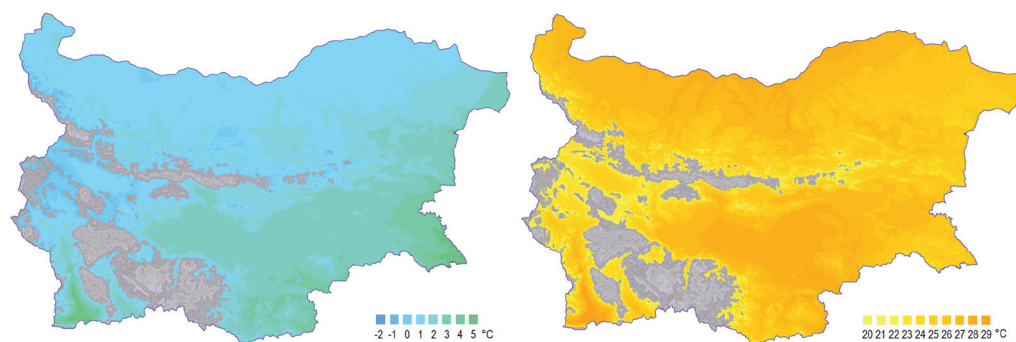


Fig. 10. Spatial distribution of the multiyear ADT mean (2001–2016) for January (left) and July (right) in non-mountainous regions of the country.

In July, ADT ranges between 20 and $29\text{ }^{\circ}\text{C}$. Temperatures are around or less than $22\text{ }^{\circ}\text{C}$ in hilly regions, $24\text{--}25\text{ }^{\circ}\text{C}$ in most of the Black Sea coastal zone and high valleys of Western Bulgaria, $26\text{--}28\text{ }^{\circ}\text{C}$ in the Thracian Lowlands and river valleys in the central part of the Danube Plain, and $27\text{--}29\text{ }^{\circ}\text{C}$ in the southern part of the Struma Valley.

The average annual number of days when $\text{ADT} < 10\text{ }^{\circ}\text{C}$ varies between 60 and 170 in the non-mountainous part of the country. Correspondingly, the average annual number of days when $\text{ADT} > 30\text{ }^{\circ}\text{C}$ is between 0 and 26.

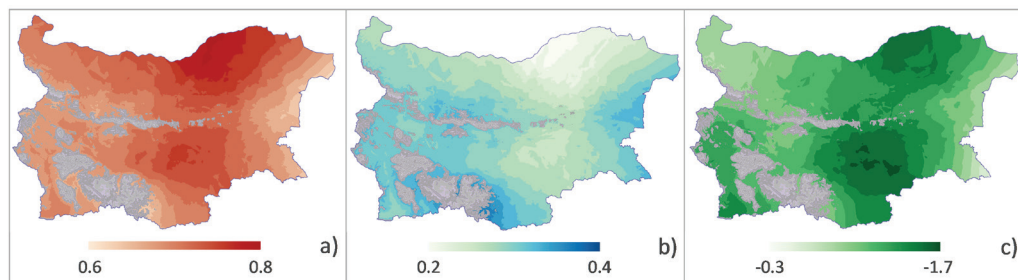


Fig. 11. Spatial distribution of the *Model 1* coefficients calculated by station data for the period 2001–2016: a) slope of *tmax*; b) slope of *tmin*; c) intercept.

The maps shown in Figure 11 (a, b) reveal a strong inverse relationship between spatial patterns of regression slopes and a relatively narrow range of values (0.6-0.8 for the slope of *tmax* and 0.2-0.4 for the slope of *tmin*). The regression intercept varies between -1.7 and -0.3 , with the lower values occurring in the areas with higher values of the *tmax* slope (and lower values of the *tmin* slope, respectively). All coefficients of Eq. 6 fall around the middle of the respective ranges, which confirms the formula's suitability.

3.3. Assessment of daily temperature features by hourly data from AWS

Air temperature data recorded by automatic weather stations gradually emerged as an essential source of meteorological and climatological information worldwide. Although they are free of human errors, as is in conventional observations, careful quality control is needed. For assessing temperature ranges and the percentage of erroneous values and discordant observations for each station by month, R-based exploratory data analysis has been used (package 'summarytools'; Comtois, 2019). All incorrect readings and dates with missing readings between 6 and 22 EET were excluded. The percentage of gaps is below 15% on average. The obtained temperature ranges were compared with those of conventional neighbor stations.

The temperature-frequency or bin method allows the sorting of hourly temperature data into discrete intervals (bins) based on long-term measurements (Busch, 1996). Each bin contains the average number of hours when measured temperature occurs in a particular range. Hourly temperature readings from the chosen AWS for the period 2012–2017 were separated by hours of the day and months of the year for each station. Figure 12 shows the distribution by month of the country-average number of hours per day (midnight to midnight) when air temperature falls into defined bins with a 5 °C width.

According to occupational health and safety standards, work can be considered as cold work when the ambient temperature is below 10-15 °C or when a person has cold-related symptoms at work (Mäkinen&Hassi, 2009). Many occupations in commerce, agriculture, forestry, mining, factory production, energy and communication systems maintenance, and construction involve prolonged exposure to cold for more than 3-4 hours daily in the cold half-year (November–March). Sharp changes in environmental temperature, typical for delivery and transport services in cold months, may increase the thermal stress on workers. Also, the employees in civil protection, police and the army are frequently exposed to the prolonged cold.

Working outdoors in hot environments is related to high heat stress. Morabito et al. (2006) found that hot weather conditions might represent a risk factor for work-related accidents during summer, occurring most often on days with high temperatures but not

necessarily at extreme thermal conditions. Moreover, early warming in June leads to heat discomfort and is less tolerated by workers than warming during the following months. Heat strokes at work, such as in construction activities, can occur at air temperatures of 34 °C and more when the relative humidity is less than 40% but also at temperatures of 28-30 °C and relative humidity of more than 65% (Morioka et al., 2006). The study by Ioannou et al. (2017) revealed that agriculture workers carrying out heavy manual labor spent most of the summer work shift time under conditions of increased heat (air temperature over 30 °C) and significant thermal discomfort.

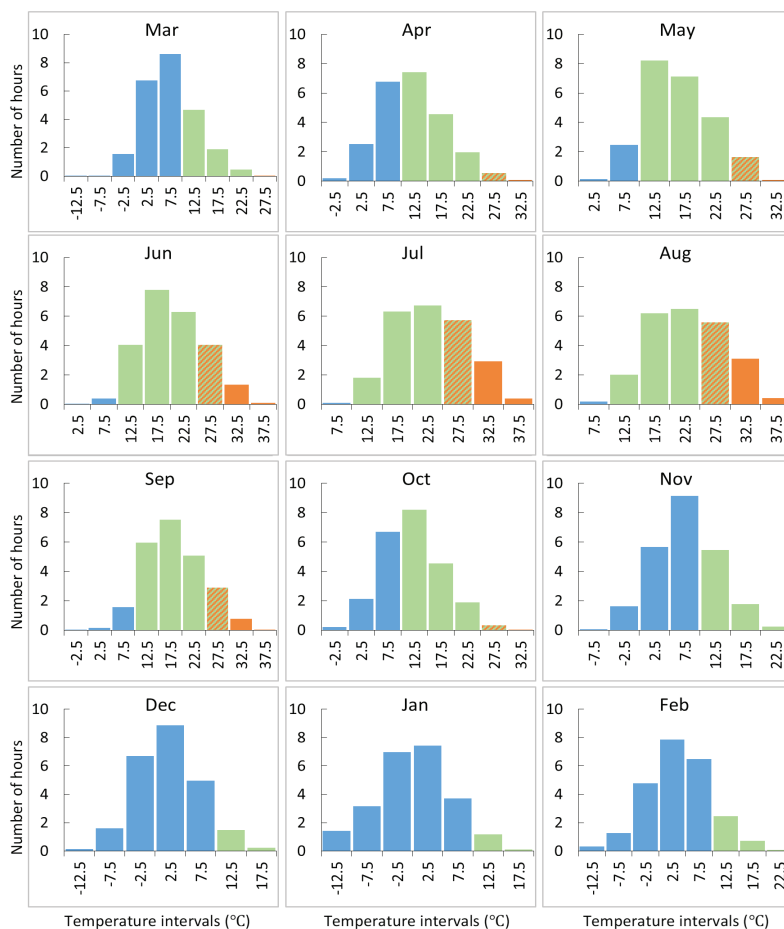


Fig. 12. Monthly distribution of the average persistence of outdoor temperature conditions in 24 hours (midnight to midnight) by hourly data from the automatic weather stations (2012–2017); the centered bin values are shown in the abscissa.

Table 2 summarizes the country-average duration of adverse temperature conditions for outdoor work in 24 hours (midnight to midnight) by month. During the cold half-year,

air temperatures remain below 10 °C most of the day (16-23 hours). Even in April and October, working outdoors is constrained in early morning and evening by temperatures below 10 °C. Extremely dangerous weather conditions for staying outdoors occur when polar air masses penetrate the country. The minimum temperatures can drop to -20 °C or less; maximum temperatures are usually well below 0 °C (ice days).

The most favorable thermal conditions for outdoor work occur during the warm half-year, except for July and August, when caution should be taken for incidents related to dangerous hot weather. Typically, air temperatures remain above 30 °C for 3.3-3.5 hours daily in July and August. However, weather conditions can be extremely hazardous to health during heat waves, particularly when the country falls under the influence of tropical air masses in slow-moving or stationary anticyclones. The maximum daily temperatures can exceed 40-42 °C, and the average daytime temperature is usually over 30 °C.

Table 2. Monthly distribution of the country-average duration (in hours) of adverse temperature conditions for outdoor work in 24 hours (midnight to midnight) in the period 2012–2017

Air temp.	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
<10 °C	22.7	20.7	16.9	9.4	2.5	0.4	0.1	0.2	1.7	8.9	16.4	22.3
>30 °C	-	-	-	-	0.1	1.4	3.3	3.5	0.8	-	-	-

4. CONCLUDING REMARKS

In the context of the Ministry of Labour and Social Policy of Bulgaria regulations for outdoor work in unfavorable temperature conditions, a methodology for estimating ADT by the forecasted daily maximum and minimum temperature has been proposed. Based on MLR analysis, an empirical formula (Eq. 6) with acceptable accuracy for the non-mountainous part of the country (below 800 m a.s.l.) has been derived. The validation results showed that for $ADT \geq 31$ °C, the prediction is 77-100% successful. On average, 8.2% of predicted values are less than 29 °C when actual ADT is above 30 °C, and 2.1% are higher than 11 °C when actual ADT is below 10 °C. The spatial distribution of the *Model 1* regression coefficients reveals a strong inverse relationship between spatial patterns of regression slopes and a relatively narrow range of values.

During the cold half-year, air temperatures remain below 10 °C through most of the day (16-23 hours, country averaged). In January, ADT is below 0 °C in hilly regions of Western Bulgaria and some places in the Danube Plain and higher than 3 °C in the southern lowlands (up to 5 °C in the southernmost parts of the Struma Valley and Black Sea coast). The average annual number of days when $ADT < 10$ °C varies between 60 and 170 in the non-mountainous part of the country.

The most favorable thermal conditions for outdoor work occur during the warm half-year. In the hottest months, air temperatures typically hold above 30 °C for 3.3-3.5

hours daily. In July, ADT ranges mainly from 20 to 29 °C. Temperatures are around or less than 22 °C in hilly regions, 24-25 °C in most of the Black Sea coastal zone and high valleys of Western Bulgaria, 26-28 °C in the Thracian Lowlands and river valleys in the central part of the Danube Plain, and 27-29 °C in the southern part of the Struma Valley. The average annual number of days when ADT > 30 °C is between 0 and 26. However, weather conditions during summer can be extremely hazardous to health during heat waves.

The presented analysis allowed us to suggest additional rules when estimating ADT by Eq. 6:

- If forecasted maximum temperatures are below 13 °C, it is assumed that the average daytime air temperature shall be below 10 °C;
- If forecasted minimum temperatures are above 24 °C, it is assumed that the average daytime air temperature shall be above 30 °C.

REFERENCES

- Alsmeyer, G. (2011). Chebyshev's Inequality. In: Lovric M. (eds.), *International Encyclopedia of Statistical Science*. Springer, Berlin, Heidelberg
- Amidan, B.G., Ferryman, T.A., & Cooley, S.K. (2005). Data outlier detection using the Chebyshev theorem. *IEEE Aerospace Conference, Big Sky, MT, USA, 2005*, pp. 3814–3819, <https://doi.org/10.1109/AERO.2005.1559688>
- Busch, R.D. (1996). *Methods of Energy Analysis*. In: Bruce D. Hunn, (ed.) *Fundamentals of Building Energy Dynamics*, MIT, Cambridge, MA
- Comtois, D. (2019). *Summary Tools: Tools to Quickly and Neatly Summarize Data R Package Version 0.9.4*, <https://CRAN.R-project.org/package=summarytools>
- Croux, C., Dhaene, G., & Hoorelbeke, D. (2003). Robust standard errors for robust estimators. *Discussions Paper Series (DPS) 03.16*, Center for Economic Studies, KULeuven.
- DV, No. 1 (2006). *Darzhaven Vestnik*, No. 1 of 3 January 2006 (in Bulgarian), <https://dv.parliament.bg/DVWeb/showMaterialDV.jsp?idMat=1258>
- Environmental Systems Research Institute (2019). *ArcGIS Pro Help*. Redlands, CA, <https://pro.arcgis.com/en/pro-app/help/main/welcome-to-the-arcgis-pro-app-help.htm>
- Guidelines for the implementation of Ordinance No. 11, published in DV, No. 1 of 2006 (in Bulgarian), <https://dv.parliament.bg/DVWeb/showMaterialDV>
- Hyndman, R.J., & Khandakar, Y. (2008). Automatic time series forecasting: the forecast package for R. *Journal of Statistical Software*, 27(3), 1-22. <https://doi.org/10.18637/jss.v027.i03>
- Hungerford, R.D., Running, S.W., Nemani, R.R., & Coughlan, J C. (1989). MTCLIM: A mountain microclimate extrapolation model. *USDA Forest Service. Res. Paper INT-414*, 52 p.
- Ioannou, L.G., Tsoutsoubi, L., Samoutis, G., Kajfez Bogataj, L., Kenny, G.P., Nybo, L., Kjellstrom, T., & Flouris, A.D. (2017). Time-motion analysis as a novel approach for evaluating the impact of environmental heat exposure on labor loss in agriculture workers, *Temperature*, 4:3, 330–340, <https://doi.org/10.1080/23328940.2017.1338210>
- JAXA EORC (2016). *ALOS Global Digital Surface Model "ALOS World 3D-30m" (AW3D30)*, <http://www.eorc.jaxa.jp/ALOS/en/aw3d30/index.htm>

- Krivoruchko, K., & Gribov, A. (2019). Evaluation of empirical Bayesian kriging. *Spatial Statistics* 32, 100368. <https://doi.org/10.1016/j.spasta.2019.100368>
- Mäkinen, T.M. & Hassi, J. (2009). Health Problems in Cold Work. *Industrial Health*, 47(3), 207-220, <https://doi.org/10.2486/indhealth.47.207>
- Montgomery, D.C., & Runger, G.C. (2003). *Applied statistics and probability for engineers*. Third Edition, John Wiley & Sons
- Morabito, M., Cecchi, L., Crisci, A., Modesti, P.A., & Orlandini, S. (2006). Relationship between work-related accidents and hot weather conditions in Tuscany (central Italy). *Industrial Health* 44(3), 458–464, <https://doi.org/10.2486/indhealth.44.458>
- Morioka, I., Miyai, N., & Miyashita, K. (2006). Hot Environment and Health Problems of Outdoor Workers at a Construction Site, *Industrial Health*, 44(3), pp. 474-480, <https://doi.org/10.2486/indhealth.44.474>
- Myttenaere, A.D., Golden, B., Grand, B.L., & Rossi, F. (2016). Mean Absolute Percentage Error for regression models. *Neurocomputing*, 192, 38-48
- Rousseeuw, P., et al. (2015). *robustbase: Basic Robust Statistics*. R package version 0.92-5. <http://CRAN.R-project.org/package=robustbase>
- Sabev, L., & Stanev, S. (1959). The climatic regions of Bulgaria and their climate, *Papers of IHM*, vol. V, Sofia, Nauka I izkustvo, p. 174 (in Bulgarian)
- Smith-Miles, K. (2011). Exploratory Data Analysis. In: Lovric, M. (ed.) *International Encyclopedia of Statistical Science*. Springer, Berlin, Heidelberg, https://doi.org/10.1007/978-3-642-04898-2_242
- Solantie, R. (2004). Daytime temperature sum — a new thermal variable describing growing season characteristics and explaining evapotranspiration, *Boreal Env. Res.* 9: 319–333, <https://www.borenv.net/BER/archive/pdfs/ber9/ber9-319.pdf>
- Willmott, C.J. (1982). Some comments on evaluation of model performance. *Bull. Am. Meteorol. Soc.* 63, 1309–1313.
- Yohai, V.J. (1987). High Breakdown Point and High Efficiency Robust Estimates for Regression, *The Annals of Statistics*, 15(20), 642–656, doi: 10.1214/aos/1176350366