

ESTIMATION OF DAILY MEAN TEMPERATURES: AN ACCURATE METHOD FOR THE DOURO VALLEY

ESTIMAÇÃO DAS TEMPERATURAS MÉDIAS DIÁRIAS: UMA METODOLOGIA PRECISA PARA A REGIÃO DO DOURO

António C. Real^{1,2,*}, José Borges², Carlos B. Oliveira²

¹ ISVOUGA Institute, Rua António de Castro Corte Real, 4520-181 Sta. M. Feira, Portugal.

² INESC TEC, Faculty of Engineering, University of Porto, Rua Dr. Roberto Frias, 4200-465 Porto, Portugal.

*corresponding author: Tel. +351256377550, e-mail: a.sousa@doc.isvouga.pt

(Received 06.07.2017. Accepted 08.06.2018)

SUMMARY

Air temperature data from many locations worldwide are only available as series of daily minima and maxima temperatures. Historically, several different approaches have been used to estimate the actual daily mean temperature, as only in the last two or three decades automatic thermometers are able to compute its actual value. The most common approach is to estimate it by averaging the daily minima and maxima. When only daily minima and maxima are available, an alternative approach, proposed by Dall'Amico and Hornsteiner in 2006, uses the two daily extremes together with next day minima temperature and a coefficient related to the local daily astronomical sunset time. Additionally, the method uses two optimizable coefficients related to the region's temperature profile. In order to use this approach it is necessary to optimize the region's unknown parameters. For this optimization, it is necessary a dataset containing the maxima, minima, and the actual daily mean temperatures for at least one year. In this research, for the period 2007-2014, we used three datasets of minima, maxima and actual mean temperatures obtained at three automatic meteorological stations located in the Douro Valley to optimize the two unknown parameters in the Dall'Amico and Hornsteiner approach. Moreover, we compared the actual mean daily temperatures available from the three datasets with the correspondent values estimated by using *i*) the usual approach of averaging the daily maxima and minima temperatures and *ii*) the Dall'Amico and Hornsteiner approach. Results show that the former approach overestimates, on average, the daily mean temperatures by 0.5°C. The Dall'Amico and Hornsteiner approach showed to be a better approximation of mean temperatures for the three meteorological stations used in this research, being unbiased relative to the actual mean values of daily temperatures. In conclusion, this research confirms that the Dall'Amico and Hornsteiner is a better approach to estimate the mean daily temperatures and provides the optimized parameters for three sites located at each of the three sub-regions of the Douro Valley (*Baixo Corgo*, *Cima Corgo* and *Douro Superior*).

RESUMO

Os dados relativos à temperatura do ar, em muitos locais do mundo, estão apenas disponíveis na forma de séries diárias de temperaturas mínimas e máximas. Historicamente, várias abordagens diferentes têm sido usadas para estimar a temperatura média diária, dado que só nas últimas duas ou três décadas termômetros automáticos são capazes de calcular seu valor real. A abordagem mais comum, é estimá-la utilizando a média aritmética das temperaturas mínimas e máximas diárias. Quando apenas os mínimos e máximos diários estão disponíveis, uma abordagem alternativa, proposta em 2006 por Dall'Amico e Hornsteiner, usa os dois extremos diários juntamente com a temperatura mínima do dia seguinte e um coeficiente relacionado com a hora astronómica do pôr do sol, no local. Além disso, o método usa dois coeficientes otimizáveis, relacionados com o perfil de temperatura da região. Para usar essa abordagem é necessário otimizar os parâmetros desconhecidos da região. Para esta otimização, é necessário um conjunto de dados contendo os máximos, mínimos e as temperaturas médias diárias reais de, pelo menos, um ano. Neste trabalho, para otimizar os dois parâmetros utilizados na abordagem proposta por Dall'Amico e Hornsteiner, no período de 2007 a 2014, foram utilizados três conjuntos de dados de mínimos, máximos e temperaturas médias reais, obtidas em três estações meteorológicas automáticas localizadas na Revião do Douro. Além disso, comparamos as temperaturas médias diárias reais, disponíveis em cada um dos três conjuntos de dados obtidos nas três estações meteorológicas, com os valores correspondentes, estimados usando *i*) a abordagem usual da média aritmética das temperaturas máximas e mínimas diárias e *ii*) a abordagem de Dall'Amico e Hornsteiner. Os resultados mostram que a primeira abordagem superestima, em média, as temperaturas médias diárias em 0,5°C. A abordagem de por Dall'Amico e Hornsteiner mostrou ser uma melhor aproximação das temperaturas médias nas três estações meteorológicas utilizadas neste trabalho, sendo não-enviesadas em relação aos valores médios reais das temperaturas diárias. Em conclusão, este trabalho confirma que a abordagem de Dall'Amico e Hornsteiner é melhor para estimar temperaturas médias diárias. São fornecidos os parâmetros otimizados para os três locais, situados em cada uma das três sub-regiões do Douro Vale (*Baixo Corgo*, *Cima Corgo* e *Douro Superior*).

Key words: daily mean temperature, monthly mean temperature, minimal temperature, maximal temperature, Douro Valley.

Palavras-chave: temperatura média diária, temperatura média mensal, temperatura mínima, temperatura máxima, região do Douro.

INTRODUCTION

Temperatures gridded datasets start in 1850 because there are too few observations available from before this date (Brohan *et al.*, 2006). However, local meteorological datasets exist since the 16th century as, for example, the 1756-1998 daily temperature and pressure dataset for Stockholm (Moberg *et al.*, 2002). Until a few decades ago, temperatures were recorded from several manual daily readings. For the Stockholm dataset, during 1756–1760, the temperature readings were made around sunrise and 13:00; during 1761-1783 the temperature readings were made around 6:30-8:00, 13:00 and 22:00-23:00; and in 1784 observations at 06:00, 14:00 and 21:00 were introduced for all variables. At the present time, effective practices for carrying out meteorological observations and measurements of temperature according to internationally agreed standards may be found in (World Meteorological Organization, 2014). Presently, there is still not a widespread consensus among meteorological organizations relative to the definition of mean daily temperature. For example the World Meteorological Organization defines the mean daily temperature as “*Mean of the temperatures observed at 24 equidistant times in the course of a continuous interval of 24 hours (normally the mean solar day, from midnight to midnight, according to the zonal time or the mean solar time of the station); or a combination of temperatures observed at less numerous times, so arranged as to depart as little as possible from the mean defined above*” (World Meteorological Organization, 2016). On the other hand, the Government of Canada defines the mean daily temperature as “*The average of the maximum and minimum temperature at a location for a specified time interval*” (Government of Canada, 2016).

Daily mean temperatures are used as input variables in a number of climatological, physical, ecological, agricultural, biological, and technical processes and applications. Historically, several different approaches have been used to estimate the actual daily mean temperature as it was only in the last two or three decades that automatic thermometers that operate continuously are able to compute its

mathematical value:
$$\bar{T} = \frac{1}{24} \int_{00:00:00}^{23:59:59} T dT$$
. Several approaches estimate the daily mean temperature by averaging a collection of several daily temperature readings taken at different times of the day (Sakellariou and Kambezidis, 2016; Ma and Guttorp, 2013; Weiss and Hays, 2005).

In many countries and in many places, the daily mean temperature still is estimated using the common approach of averaging the maxima and minima temperatures of a day. In this research, this approach will be called the TAvG method. It is long known that the use of this method produces poor estimates of the actual mean daily temperature Kaemtz *et al.* (1845). In the Douro Valley (see Figure 1), meteorological and climatological studies about its present and future climate, as well as agricultural studies on the region’s crops and agricultural products (grapes, olives, almonds, figs, oranges and cherries), use estimations of mean daily temperatures. Most of these studies have used the TAvG method to estimate the unknown real mean temperature (Santos and Malheiro, 2011; Jones, 2012; Jones and Alves, 2012; Jones 2013; Fraga *et al.*, 2014; C. Real *et al.*, 2014; C. Real *et al.*, 2016; Fraga and Santos, 2017).

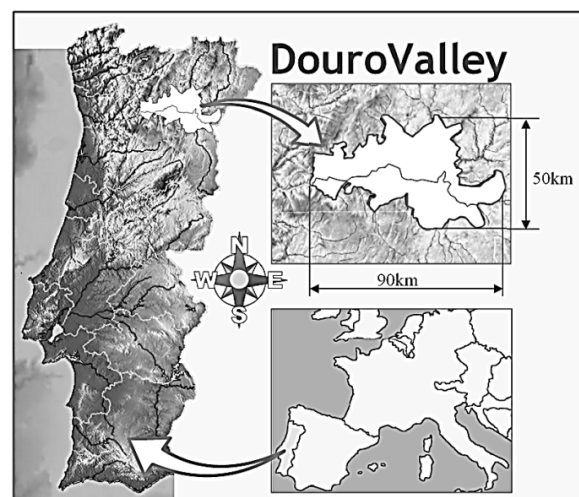


Figure 1. Portugal and the location of the Douro Valley (relief maps: www.maps-for-free.com).

Portugal e a localização da Região do Douro (mapas de relevo: www.maps-for-free.com).

An alternative estimation approach, proposed by Dall’Amico and Hornsteiner (2006) that we will call the Dall’Amico method, uses the two daily extreme temperatures together with the next day minima temperature and a coefficient related to the local daily astronomical sunset time (US Naval Observatory, 2017). Additionally, the method uses two optimizable coefficients, related to the temperatures profiles of the region. For the optimization of the two unknown parameters the authors suggest that a dataset containing the maxima, minima, and the actual daily

mean temperatures for at least one year should be used to fit the coefficients to the available data.

This research used three datasets of maxima, minima and mean daily temperatures collected at three meteorological stations located at each of the three sub-regions of the Douro Valley (*Baixo Corgo*, *Cima Corgo* and *Douro Superior* – see Figure 2). As these datasets were obtained in automatic meteorological stations that are able to measure temperatures continuously, the mean temperatures in the datasets are the actual values. We compared the actual mean daily temperatures to the estimates obtained by using the TAVg method and by using the alternative method proposed by Dall’Amico and Hornsteiner (2006). The quality of the estimates obtained by the two methods was assessed.

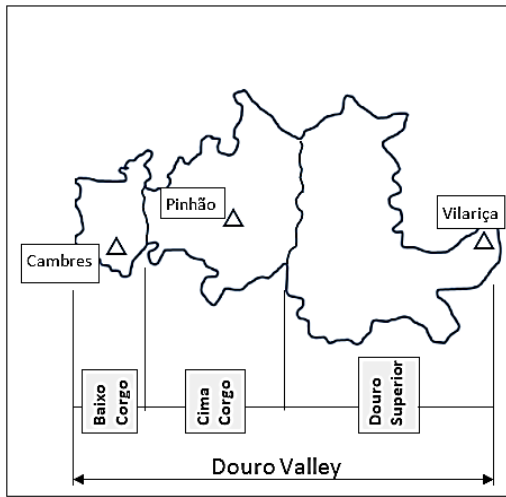


Figure 2. Location of the local meteorological stations.

Localização das estações meteorológicas.

We are aware that the use of simple methods as the one proposed by Dall’Amico and Hornsteiner are prone to lead to poorer results when used in sites where a significant spatial variability exists. For these sites, local landscape may determine highly variable patterns for the evolution of temperatures along each day. This is the case of mountain regions such as the Douro Valley, interrupted longitudinally by the Douro river valley, where mesoscale weather processes, the existence of dynamically and thermally wind systems, sunshine exposition, clouds at site level and other factors, influence the daily evolution of the weather at the site. Nevertheless, after having the optimized coefficients for the Dall’Amico and Hornsteiner expression its use is intended to estimate the unknown real value of the mean temperature, using only the daily extreme temperatures.

MATERIAL AND METHODS

Data

Three datasets of maxima, minima and mean daily temperatures for the period 2007-2014, collected at *Cambres* (91m elevation), *Pinhão* (107 m elevation) and *Vilarica* (171 m elevation) stations, located at each of the three sub-regions of the Douro Valley (*Baixo Corgo*, *Cima Corgo* and *Douro Superior*), have been used (Figure 2). Considering a constant lapse rate of 5.1 °C/km (elevation) (Stone and Carlson 1979), the difference in temperatures from station to station that may be inputable to the difference in the elevation is less than 0.4 °C. A methodology developed by Feng *et al.* (2004) was used to identify data that was suspect of being erroneous. In order to homogenize the cleaned dataset the RHtestsV3 software package (Wang, 2011) was used.

Table I shows the maximum, minimum and mean temperatures at *Cambres*, *Pinhão* and *Vilarica* for each quarter in 2007-2014. *Cambres*, *Pinhão* and *Vilarica* stations are located at each of the three sub regions of the Douro Valley (*Baixo Corgo*, *Cima Corgo* and *Douro Superior*). *Vilarica* is the station that is located further inland and *Cambres* is the station that is closer to the Atlantic Ocean coast. Analysis of the minimum temperatures in each station at the four quarters show that during the coldest months, minimum temperatures decrease from the coast towards inland and maximum temperatures are similar in the three stations. During the warmest months, minimum temperatures are similar in the three stations and maximum temperatures increase from the coast towards inland.

Methods

Estimates of mean temperatures, obtained by using two methods of estimation (TAVg and Dall’Amico), were compared with the correspondent real mean temperatures.

TAVg method

The TAVg method is the most common and simple method for estimating the mean daily temperature. The mean daily temperature is estimated by averaging minima and maxima daily temperatures, see equation 1.

$$\hat{T}_{mean}^{day i} = \frac{(T_{max}^{day i} - T_{min}^{day i})}{2} \quad (1)$$

Table I

Temperatures at Cambres, Pinhão and Vilarica (2007-2014)

Temperaturas em Cambres, Pinhão e Vilarica (2007-2014)

Station	months	2007			2008			2009			2010		
		T _{mean}	T _{max}	T _{min}	T _{mean}	T _{max}	T _{min}	T _{mean}	T _{max}	T _{min}	T _{mean}	T _{max}	T _{min}
Cambres elevation: 91m	1-3	10.2	15.3	5.9	10.7	16.1	6.5	10.4	16.7	5.2	9.4	14.0	5.8
	4-6	17.6	24.8	11.3	17.5	24.1	11.6	17.9	25.0	11.6	18.3	25.7	11.8
	7-9	22.1	30.7	14.3	21.9	29.5	14.6	23.3	31.7	15.8	24.3	33.1	16.7
	10-12	10.3	18.2	5.0	10.6	16.7	5.9	13.1	17.9	9.5	11.0	16.4	7.0
Pinhão elevation: 107m	1-3	10.0	14.6	5.9	10.2	15.4	6.3	10.0	16.2	5.1	9.2	13.9	5.7
	4-6	17.8	24.2	12.0	17.8	23.8	12.2	18.4	25.2	12.1	18.3	25.1	12.1
	7-9	23.0	30.7	15.9	22.5	29.8	15.7	24.0	31.9	16.7	25.0	33.4	17.6
	10-12	10.3	16.9	5.8	10.5	15.9	6.4	13.2	17.8	9.6	11.0	16.1	7.3
Vilarica elevation: 171m	1-3	9.5	14.8	4.3	9.6	15.5	4.8	9.6	16.4	3.7	8.8	13.7	4.6
	4-6	17.7	24.8	10.8	17.8	24.8	11.2	19.0	26.0	11.7	18.0	25.5	10.9
	7-9	23.5	32.0	15.0	23.3	31.7	15.1	25.0	33.6	16.5	25.6	34.7	16.7
	10-12	9.3	17.1	3.3	9.6	15.7	4.6	12.9	17.9	8.6	10.3	16.1	5.5

Station	months	2011			2012			2013			2014			2007-2014		
		T _{mean}	T _{max}	T _{min}	T _{mean}	T _{max}	T _{min}	T _{mean}	T _{max}	T _{min}	T _{mean}	T _{max}	T _{min}	T _{mean}	T _{max}	T _{min}
Cambres elevation: 91m	1-3	10.0	15.6	5.7	9.0	16.4	3.1	9.4	14.2	5.7	10.5	15.7	6.5	10.0	15.5	5.6
	4-6	19.9	27.9	13.0	16.9	23.1	11.5	16.6	23.8	10.1	17.9	25.1	11.6	17.8	24.9	11.6
	7-9	22.8	31.2	15.5	22.1	30.2	14.8	24.0	33.1	15.9	22.0	29.6	15.7	22.8	31.1	15.4
	10-12	12.6	18.9	8.0	11.5	16.7	7.6	10.9	16.7	6.7	12.3	17.8	8.5	11.5	17.4	7.3
Pinhão elevation: 107m	1-3	9.7	14.9	5.7	8.9	15.8	3.3	9.2	13.7	5.6	10.6	15.5	6.6	9.7	15.0	5.5
	4-6	20.1	27.5	13.4	17.5	23.4	12.1	17.0	23.7	10.8	18.4	25.0	12.1	18.2	24.7	12.1
	7-9	23.5	31.3	16.5	23.3	31.1	16.2	25.1	33.4	17.5	22.9	30.3	16.8	23.7	31.5	16.6
	10-12	12.6	18.1	8.5	11.8	16.5	8.2	11.0	16.2	7.1	12.4	17.2	9.1	11.6	16.8	7.8
Vilarica elevation: 171m	1-3	9.2	15.0	4.3	8.2	16.2	1.2	8.6	13.7	4.2	10.0	15.8	4.9	9.2	15.1	4.0
	4-6	20.0	28.3	12.1	18.0	24.9	11.2	16.9	24.5	9.4	18.7	26.6	10.8	18.3	25.7	11.0
	7-9	23.8	32.6	15.3	24.2	33.1	15.3	25.9	34.9	16.8	23.1	31.7	15.1	24.3	33.0	15.7
	10-12	11.7	18.2	6.5	11.4	16.9	6.9	10.4	16.7	5.1	11.6	17.2	7.3	10.9	17.0	6.0

Dall'Amico method

Dall'Amico method (Dall'Amico and Hornsteiner 2006) is a method for estimating the mean daily temperature, $T_{mean}^{day i}$, on the basis of minima and maxima daily temperatures. The method accounts for the temperature trend by including the minimum temperature on the following day in addition to the extremes on the day in question, see equation 2.

$$\hat{T}_{mean}^{day i} = [T_{min}^{day i} + C_D(T_{max}^{day i} - T_{min}^{day i})]S^{day i} + [T_{min}^{day i+1} + C_N(T_{max}^{day i} - T_{min}^{day i+1})](1 - S^{day i}) \quad (2)$$

Where:

* $S^{day i}$ is the nondimensional astronomical sunset time on day i, $S^{day i} = t^{day i}(\text{sunset}) / 24 \text{ h}$ where $t^{day i}(\text{sunset})$ is the astronomical¹ sunset time on day i. The astronomical sunset times for any location are

¹ Dall'Amico and Hornsteiner (2006) have tested their methodology in two different sites: one located in a Euaropean flat terrain and the other located in an Alpine valley site. They concluded that the use of the actual sunset times yielded slightly worse results when compared to the astronomical sunset time.

available on the internet: e.g. http://aa.usno.navy.mil/data/docs/RS_OneYear.php

* C_D is the proportionality coefficient for the day;

* C_N is the proportionality coefficient for the night.

As suggested in Dall'Amico and Hornsteiner (2006) seasonally varying coefficients C_D and C_N in equation 2 have been optimized partitioning years into two periods: one that will be designated as cold period and the other as warm period, C_D^{Warm} , C_D^{Cold} , C_N^{Warm} , and C_N^{Cold} . The optimization consisted on the minimization of the Root Mean Square Errors (RMSE) of the estimates for the mean daily temperatures at *Cambres*, *Pinhão* and *Vilarica* stations obtained by using equation 2, relative to their actual mean daily temperatures. Parameter optimization for the Dall'Amico method was performed by using the Microsoft Excel (2016) evolutionary solver algorithm. The optimization process has also determined the boundaries of the two periods (cold and warm).

For both methods, the respective Root Mean Square Errors (RMSE) and bias have been calculated in order to assess their relative accuracy.

Assessing model for overfitting

To assess if the estimates obtained by using the Dall'Amico method were overfitting² the data, we evaluated the model by using a cross-validation methodology. To that effect, 4-fold methodology was defined. The datasets of maxima, minima and mean daily temperatures from eight years (2007-2014) were separated into four groups (folds) having each group the data from two consecutive years. Six years of data (tree folds) were used to estimate the model parameters (training the model) and the two remaining years of data (one fold) were used to validate the accuracy of the model (testing the model).

Additionally, in order to assess if the model parameters would maintain valid for a period of seven years we evaluated the quality of the estimates obtained using a single year of data to compute the model parameters (training set) and the remaining seven years of data to test the model's estimates (testing set). We have done this evaluation using each single year of data to compute the model parameters (training set). We will present results naming this methodology as "validation 1/7".

RESULTS AND DISCUSSION

The values for the boundaries of the cold and warm periods resulted from the optimization process: warm period: March 1 to September 30 and cold period: January 1 to February 28 and from October 1 to December 31. Additionally, four optimum values for the seasonally varying proportionality coefficients were obtained for the Douro Valley, which are presented in Table II.

Using each of the two pairs of C_i (C_D^{Warm} , C_N^{Warm} or C_D^{Cold} , C_N^{Cold}) in equation 2, two forms of this equation were obtained. The first, uses C_D^{Warm} , C_N^{Warm} , and is intended for the estimation of the mean daily temperature for the warm months (months 3 to 9). The second, uses C_D^{Cold} , C_N^{Cold} , and is intended for the estimation of the mean daily temperature for the cold months (months 1, 2, 10, 11, and 12).

² Overfitting refers to a model that models the training data too well. It happens when a model learns the detail and noise in the training data to the extent that it may negatively affects the performance of the model (Brownlee, 2016).

Table II

Optimized coefficients for Dall'Amico coefficients C_D and C_N in equation 1.

Coefficientes otimizados para C_D e C_N , na equação 1

	Coefficients C_D and C_N optimized to:			
	Cambres	Pinhão	Vilarica	the 3 stations
C_D^{Warm}	0.512	0.480	0.532	0.524
C_D^{Cold}	0.287	0.443	0.348	0.288
C_N^{Warm}	0.292	0.268	0.332	0.298
C_N^{Cold}	0.764	0.845	0.763	0.822

The bias and Root Mean Square Errors (RMSE) obtained using both methods are presented in Table III.

Table III

Bias and Root Mean Square Errors (RMSE) obtained for the estimates of the mean daily temperatures calculated using both TAVg method and Dall'Amico method, relative to the actual mean daily temperature. Coefficients optimized using temperature data from the three stations. Evaluation of bias and RMSE based on mean daily temperatures from the three stations.

Enviesamento e RMSE das estimativas das temperaturas médias diárias, calculadas utilizando os métodos de Dall'Amico & Hornsteiner e TAVg, relativamente aos valores reais. Os coeficientes otimizados para o método de Dall'Amico & Hornsteiner foram obtidos utilizando em conjunto, as séries de temperaturas médias diárias das três estações.

	TAVg Method	Dall'Amico Method
RMSE	0.86	0.65
bias	0.50	0.00

The Dall'Amico Method estimator is unbiased while the TAVg method is a biased estimator of the actual mean daily temperature, overestimating the actual value, on average, 0.5°C.

We assessed the accuracy of the mean temperatures estimates calculated using the Dall'Amico Method when using coefficients C_D and C_N optimized for a site other than the one for which the estimates are computed. We run several combinations using coefficients C_D and C_N optimized using the data from one single site and then assessing, for the two remaining sites, the accuracy of the estimates computed using the previously optimized C_D and C_N . The corresponding bias and RMSE have been computed and the results are shown in Table IV.

The results of the three combinations in Table IV show that when compared to the TAVg method, the Dall'Amico Method has comparable values for RMSE and a smaller bias. In a small region such as

the Douro Valley, despite the fact that landscape may determine highly variable patterns for the evolution of temperatures along each day, the results indicate that the use of the optimized coefficients presented in the

“the 3 stations” column in Table II will lead to estimates of the mean daily temperatures having a comparable variability but a smaller bias, relative to the TAVg method.

Table IV

Bias and RMSE computed using coefficients C_D and C_N optimized for a single site

Enviesamento e RMSE para as estimativas calculadas com os coeficientes C_D e C_N otimizados para uma só estação

		C_D, C_N optimized to Cambres						
		Pinhão	Vilarça	RMSE	bias			
Accuracy assessment using the two remaining sites	→	Dall'Amico Method		0.70	-0.20	C_D, C_N optimized to Pinhão		
		TAVg Method		0.82	0.45			
	→	Cambres	Vilarça	RMSE	bias	C_D, C_N optimized to Vilarça		
		Dall'Amico Method		1.20	0.10			
	→	TAVg Method		0.88	0.45			
		Cambres	Pinhão	RMSE	bias			
			Dall'Amico Method		0.71	0.32		
			TAVg Method		0.88	0.55		

Using the Dall'Amico Method with the C_D and C_N coefficients optimized using temperatures data from the three stations, we compared the differences between the actual mean monthly temperatures (the average of all mean daily temperatures during the correspondent month) and its estimates obtained by the two methods (TAVg and Dall'Amico) at each station (see Figure 3, Figure 4 and Figure 5).

79.51% of the mean monthly temperatures estimated using the Dall' Amico method have a smaller deviation to the actual monthly mean temperatures than the estimates obtained using the TAVg method.

In the 96-month period (Jan 2007 to Dec 2014), in the three stations (Cambres, Pinhão and Vilarça),

These results, together with the results obtained using mean daily temperatures (see Table III), show that the Dall'Amico method is a better approximation for both daily and monthly mean temperatures, having smaller bias and RMSE relative to the estimates obtained using the TAVg method.

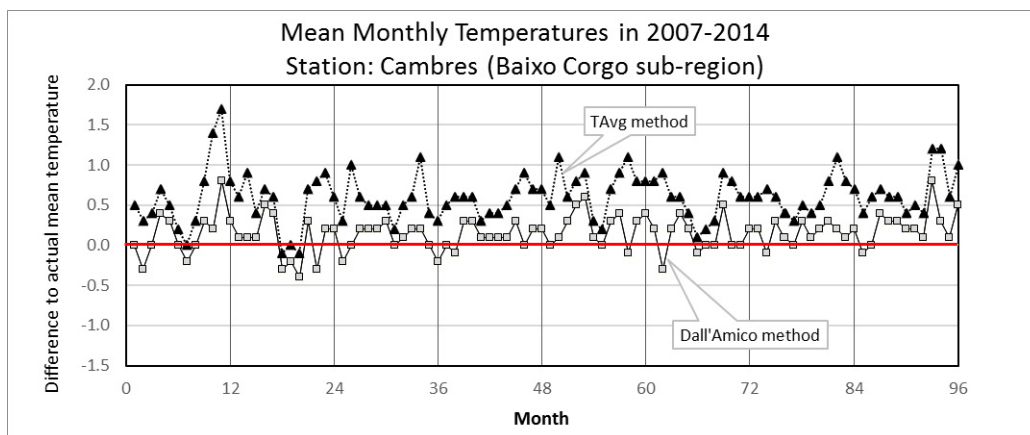


Figure 3. Differences from the estimated mean monthly temperatures to its actual value at Cambres. A constant offset of 0.5°C exists, on average, between the values of the two estimation methods. Bias Dall'Amico Method equals 0.1° . Bias TAVg method equals 0.6° .

Diferença entre os valores estimados para as temperaturas médias mensais em Cambres e o seu valor real. Em média, verifica-se um desfasamento de 0.5° entre os dois métodos de estimação. O enviesamento das temperaturas estimadas pelo método de Dall'Amico & Hornsteiner é de 0.1° e o enviesamento das temperaturas estimadas pelo método TAVg é de 0.6° .

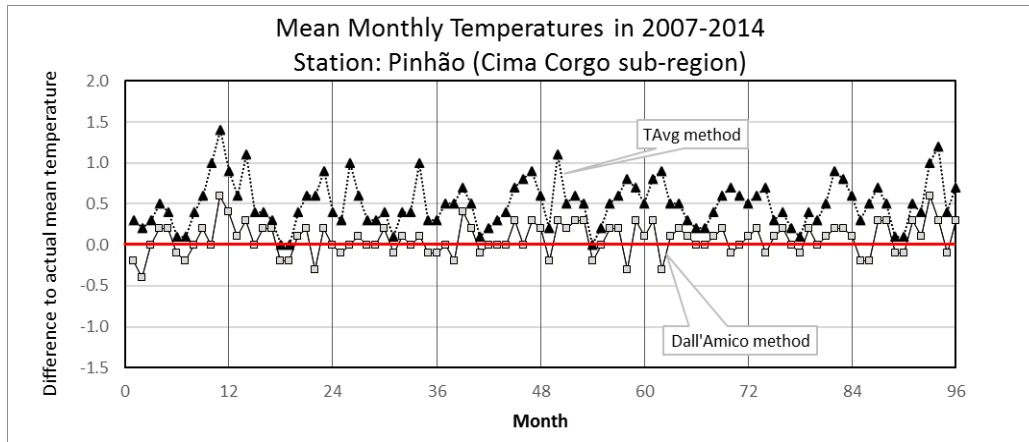


Figure 4. Differences from the estimated mean monthly temperatures to its actual value at Pinhão. A constant offset of 0.4°C exists, on average, between the values of the two estimation methods. Bias Dall'Amico Method equals 0.1. Bias TAvg method equals 0.5.

Diferença entre os valores estimados para as temperaturas médias mensais em Pinhão e o seu valor real. Em média, verifica-se um desfasamento de 0.4° entre os dois métodos de estimação. O enviesamento das temperaturas estimadas pelo método de Dall'Amico & Hornsteiner é de 0.1° e o enviesamento das temperaturas estimadas pelo método TAvg é de 0.5°.

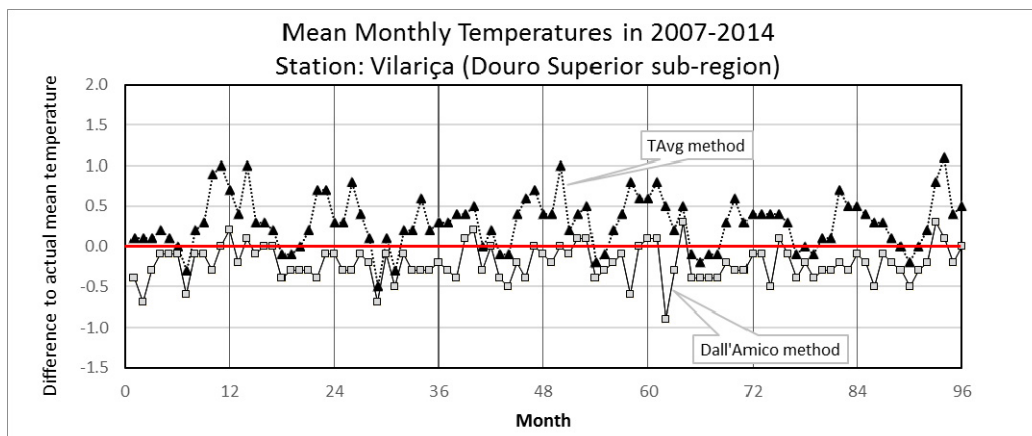


Figure 5. Differences from the estimated mean monthly temperatures to its actual value at Vilarica. A constant offset of 0.5°C exists, on average, between the values of the two estimation methods. Bias Dall'Amico Method equals -0.2. Bias TAvg method equals 0.3.

Diferença entre os valores estimados para as temperaturas médias mensais em Vilarica e o seu valor real. Em média, verifica-se um desfasamento de 0.5° entre os dois métodos de estimação. O enviesamento das temperaturas estimadas pelo método de Dall'Amico & Hornsteiner é de -0.2° e o enviesamento das temperaturas estimadas pelo método TAvg é de 0.3°.

Results of the cross-validation methodology used to assess model overfitting are presented in Table V and they show that the complete model has a RMSE and a bias comparable to the correspondent values obtained by averaging the 4-fold individual results indicating that the complete model is stable when used with new unknown data.

Results for the methodology validation 1/7 are presented in Table VI showing that even when using a single year of data to train the model, the obtained coefficients enable equation 2 to produce good

estimates of the mean daily temperatures when using new unknown data as input.

These results suggest that for a slow climate change process, the obtained coefficients will only require sparse recalibrations maintaining useable during periods of 7 or more years. Although the average of the eight resulting partial models has an RMSE and a bias both larger than correspondent values of the complete model, the estimates are still better than the ones obtained by using the TAvg method.

The present research shows that the T_{Avg} estimation method overestimates the actual mean daily air temperatures in the Douro Valley. On average, the overestimation is of 0.5°C, being larger for the warm months (0.6°C) than for the remaining months (0.3°C). The alternative estimation method, the Dall’Amico method, is also a simple method for estimating the mean daily temperature based on the

daily maxima and minima temperatures. This latter method showed to be closer to the actual values of the mean daily temperature for every month of the year, having a small bias relative to the actual values (0.0 for daily mean temperatures and 0.13 for monthly mean temperatures in the 96-month period analyzed in this research).

Table V

Results for the cross-validation methodology

Resultados da validação cruzada

		Cross-validation 4 folds				Average 4 folds	Complete model
		N training = 6570 records, N validation = 2190 records					
		Fold 1	Fold 2	Fold 3	Fold 4		
Dall’Amico Method	RMSE	0.6600	0.6500	0.7000	0.6000	0.6525	0.6500
	bias	-0.0444	0.0234	0.0148	0.0122	0.0015	0.0021
T _{Avg} Method	RMSE	0.8600	0.8800	0.8900	0.8000	0.8575	0.8600
	bias	0.4321	0.4870	0.4750	0.4536	0.4619	0.4619

Table VI

Results for the approach of using a single year of data to estimate the model parameters

Resultados obtidos quando são utilizado apenas os dados de um dos anos para estimar os parâmetros do modelo

		1 year data used for training, 7 years used for validation								Average	Complete model
		Training year									
		2007	2008	2009	2010	2011	2012	2013	2014		
Dall’Amico Method	RMSE	0.6600	0.6600	0.6600	0.6700	0.6700	0.6600	0.6600	0.6600	0.6633	0.6500
	bias	0.0030	0.0460	0.0687	-0.0412	-0.0744	0.0501	-0.0081	-0.1168	-0.0203	0.0021
T _{Avg} Method	RMSE	0.8500	0.8600	0.8600	0.8700	0.8500	0.8600	0.8600	0.8400	0.8567	0.8600
	bias	0.4600	0.4663	0.4724	0.4625	0.4504	0.4647	0.4637	0.4555	0.4615	0.4619

Several meteorological and agricultural indexes based on daily temperatures use mean daily temperatures as input. It is not always clear for which methodology of mean temperatures calculation they are calibrated to. In viticulture, for example, growing degree-days (GDD) are defined as the sum of the mean daily temperature above a threshold from January 1 to a given date: $GDD = \sum_{i=1}^{O. Date} (T_{avg} - T_{base})$, T_{avg} is the average temperature and T_{base} is a temperature used as threshold. The T_{avg} value used in the calculation of GDD is not the actual mean daily temperature but its estimation obtained by using the T_{Avg} method. For example, van Leeuwen *et al.* (2008) studied the heat requirement (GDD) for each grapevine variety to reach each phenological event defining heat requirement data for most grape varieties. Using daily average temperatures obtained by a methodology that produces estimates closer to

the unknown daily mean temperatures than the T_{Avg} estimates will lead to a definition of more precise heat requirements for the grapevine phenological events. Nowadays, however, more accurate and consistent growing heat summation indices such as growing degree hours (GDH) based on hourly temperatures, are replacing GDD for phenological prediction (Gu, 2016).

Optimized parameters for three sites located at each of the three sub-regions of the Douro Valley (*Baixo Corgo*, *Cima Corgo* and *Douro Superior*), for the period 2007-2014, are made available in this research. Interpolation of the values of the parameters will allow their use to estimate mean temperatures (with moderate confidence) in sites different from the ones used in the optimization process. As the three stations are located at sites having an average elevation of 120 meters, the use of the optimized parameters made

available in this research will make necessary to refer the daily maxima and minima to a reference elevation of 120. Rolland (2003) indicates that temperature decreases, on average, 5 °C / 1000 meters elevation increase.

The use of more precise methods for the estimation of the actual mean daily temperature, although providing estimates closer to the actual values, should be followed by a recalibration of the indices and variables dependent on mean temperatures because, as in the GDD example, the indicative values of some indices are based on estimates obtained by using the routine TAvG method.

CONCLUSIONS

In the Douro Valley, for the period of 2007-2014 and for the three meteorological stations used in this research, the common approach overestimates, on average, the daily mean temperatures by 0.5 °C (0.90 overestimates, 0.06 underestimates, 0.04 exact matches) while the Dall'Amico method has, in the same period, a bias of 0.0 °C (0.44 overestimates, 0.40 underestimates, 0.16 exact matches). Although this does not seem a large bias, it will be interesting to note that in 1980-2009, the mean annual temperature in the Douro Valley increased 1.0 °C, from an average value of 14.9 °C in the early 80s to

REFERENCES

- Brohan P., Kennedy J.J., Harris I., Tett S.F.B., Jones P.D., 2006. Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850. *Journal of Geophysical Research Atmospheres*, **111** (12): D12106.
- Brownlee J., 2016. Overfitting and Underfitting With Machine Learning Algorithms. Machine Learning Mystery. Available at: <https://machinelearningmastery.com/overfitting-and-underfitting-with-machine-learning-algorithms/> [Accessed May 11, 2018].
- C. Real A., Borges J., Sarsfield Cabral J., Jones G.V., 2014. Partitioning the grapevine growing season in the Douro Valley of Portugal: Accumulated heat better than calendar dates. *International Journal of Biometeorology*, **59** (8): 1045-59.
- C. Real A., Borges J., Cabral J.S., Jones G.V., 2016. A climatology of Vintage Port quality. *Int. J. Climatol.*, **37**, 3798-3809.
- Dall'Amico M., Hornsteiner, M., 2006. A simple method for estimating daily and monthly mean temperatures from daily minima and maxima. *Int. J. Climatol.*, **26** (13), 1929-1936.
- Feng S., Hu Q., Qian W., 2004. Quality control of daily meteorological data in China, 1951-2000: a new dataset. *Int. J. Climatol.*, **24** (7), 853-870.
- Fraga H., Malheiro A., Santos J., 2014. Climate factors driving wine production in the Portuguese Minho region. *Agric. For. Meteorol.*, **185**, 26-36.
- Fraga H., Santos J.A., 2017. Daily prediction of seasonal grapevine production in the Douro wine region based on favourable

15.9 °C at 2009 (Sousa 2015). Estimates which overestimate the actual mean temperature, on average, by 0.5 °C, correspond, if the trend in 1980-2009 maintains, to the actual mean temperatures that will happen in this region by 2032, 15 years from now.

In conclusion, this research confirms that the Dall'Amico method should be preferred to estimate the mean daily temperatures.

ACKNOWLEDGMENTS

This work was partly funded by:

Associação para o Desenvolvimento da Viticultura Duriense (ADVID);

"ERDF – European Regional Development Fund through the Operational Programme for Competitiveness and Internationalisation - COMPETE 2020 Programme within project «POCI-01-0145-FEDER-006961», and by National Funds through the FCT – Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) as part of project UID/EEA/50014/2013."

meteorological conditions. *Aust. J. Grape and Wine Research*, **23** (2), 296-304.

Government of Canada, 2016. Glossary. Available at: http://climate.weather.gc.ca/glossary_e.html [Accessed December 18, 2016].

Gu S., 2016. Growing degree hours - a simple, accurate, and precise protocol to approximate growing heat summation for grapevines. *Int. J. Biometeorology*, **60** (8), 1123-1134.

Jones G.V., 2012. A climate assessment for the Douro wine region: an examination for the past, present and future conditions for wine production, ADVID, Peso da Régua. Available at: <http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:A+Climate+Assessment+for+the+Douro+Wine+Region:+An+Examination+of+the+Past,+Present,+and+Future+Climate+Conditions+for+Wine+Production#1> [Accessed February 14, 2014].

Jones G.V., 2013. Uma Avaliação do Clima para a Região Demarcada do Douro: Uma análise das condições climáticas do passado, presente e futuro para a produção de vinho. ADVID - Associação para o Desenvolvimento da Viticultura, ed.

Jones G.V., Alves F., 2012. Impact of climate change on wine production: a global overview and regional assessment in the Douro Valley of Portugal. *Int. J. Global Warming*, **4** (3/4), 383-406.

Kaemtz, L.F., Martins, C., Lalanne, L., 1845. A complete course of meteorology. C. V. Walker, ed., London: Hippolyte Baillière.

Ma Y., Guttorp P., 2013. Estimating daily mean temperature from synoptic climate observations. *Int. J. Climatol.*, **33** (5), 1264-1269.

- Moberg A., Bergström H., Kringsman J.R., Svanered O., 2002. Daily air temperature and pressure series for Stockholm (1756--1998). *Climatic Change*, **53** (1/3): 171-212.
- Rolland C., 2003. Spatial and seasonal variations of air temperature lapse rates in Alpine regions. *Journal of Climate*, **16** (7), 1032-1046.
- Sakellariou N.K., Kambezidis H.D., 2016. Improving the estimation of the true mean monthly and true mean annual air temperatures in Greece. *Atmospheric Science Letters*, **17** (1), 13-18.
- Santos J., Malheiro A., 2011. Statistical modelling of grapevine yield in the Port Wine region under present and future climate conditions. *Int. J. Biometeorology*, **55** (2), 119-131.
- Sousa A.C.-R., 2015. Analyzing the Influence of the Douro Valley Weather on the Quality and Yield of Vintage Port. Porto. Available at: <http://repositorio-aberto.up.pt/bitstream/10216/78434/2/112947.pdf>.
- Stone P., Carlson J., 1979. Atmospheric lapse rate regimes and their parameterization. *Journal of the Atmospheric Sciences*, **36**, 415-423.
- US Naval Observatory, 2017. Sun or Moon Rise/Set Table for One Year. Astronomical Applications Department. Available at: http://aa.usno.navy.mil/data/docs/RS_OneYear.php [Accessed January 12, 2017].
- Van Leeuwen C., Garnier C., Agut C., Baculat B., Barbeau G., Besnard E., Bois B., Boursiquot J.-M., Chuine I., Dessup T., Dufourcq, T., Garcia-Cortazaz, I., Marguerit, E., Monamy, C., Koundourass S., Payan J.C., Parker A., Renouf V., Rodriguez-Lovelie B., Roby J. ., Tonietto J., Trambouze, W., 2008. Heat requirements for grapevine varieties is essential information to adapt plant material in a changing climate. In: *VIIIème Congrès International des Terroirs Viticoles - Comptes rendus - Volume 1*, 222-227.
- Wang X., 2011. RHtestsV3. Available at: http://cccma.seos.uvic.ca/cgi-bin/etccdi/main_programs/download_RHtest?a=RHtestsV3.r.
- Weiss A., Hays C.J., 2005. Calculating daily mean air temperatures by different methods. *Agric. For. Meteorol.*, **128** (1-2), 57-65.
- World Meteorological Organization, 2014. Measurement of temperature. In: *Guide to Meteorological Instruments and Methods of Observation*. Geneva, Switzerland: WMO, p. 1128.
- World Meteorological Organization, 2016. Meteoterm. Available at: <http://wmo.multicorpora.net/MultiTransWeb/Web.mvc> [Accessed December 18, 2016].