Modified temperature-based global solar radiation models for estimation in regions with scarce experimental data

Jesús-Ignacio Prieto [a,*], David García [b]

[a] Department of Physics, University of Oviedo, c/Federico García Lorca 18, E-33007 Oviedo, Spain
[b] Department of Energy, University of Oviedo, c/Wifredo Ricart s/n, 33204 Gijón, Spain

ARTICLE INFO

ABSTRACT

Reliable data on the availability of solar energy is needed, as solar energy is an essential resource for sustainable development worldwide. However, ground-based radiometric stations are scarce, at least in large areas far from population and in developing countries, so there are difficulties in validating methods for estimating solar radiation. Indirect models mitigate the problem by providing radiation data from other meteorological variables, which can be measured with low-cost equipment and calibrated with data from secondary station networks. However, models' accuracy decreases if estimations are required far away from the calibration stations. It is hoped that modified models that include the influence of geographical and topographical variables can attenuate this drawback in data-scarce regions. This paper evaluates the accuracy and generality of 14 existing models of monthly global solar radiation based on temperature, which is a routinely measured variable. At first, models are locally calibrated at 105 stations in three large areas in Spain. Then, from the local coefficients of eight stations selected in each area, general equations are derived for the coefficients of each model as function of the ratio between elevation and distance to the sea. The predictions of these modified models, i.e., using coefficients derived from general equations, are compared both for the eight base stations and the remaining ones used for validation. In the comparisons, not only errors averaged in groups of stations are considered, but also local results. Several models perform well in some areas, but a simple homogeneous model is the only one whose indicators are good in all areas and hardly vary when using general coefficients derived from the data measured at all available stations.

1. Introduction

Solar energy is one of the main alternatives to sustainably meet the world's growing energy demand and thus contribute to the development of the actions urgently called for by the Intergovernmental Panel on Climate Change to reduce the severe environmental impact [1]. Consequently, there is currently strong interest in both solar thermal and photovoltaic systems for agricultural [2], industrial and, above all, residential applications [3,4], as cities in developed countries are thought to be the main consumers of energy worldwide [5]. However, reliable data on solar energy availability are not always known for project development, especially in remote areas and developing countries.

Due to the scarcity of weather stations with solar radiation records [6,7], models for indirect estimation of global solar radiation (GSR) have been developed for decades, using abundantly measured meteorological variables or, more recently, satellite images. Alternative methodologies using artificial intelligence techniques have also been developed, such as those based on Adaptive Neuro-Fuzzy Inference Systems (ANFIS) and Artificial Neural Network (ANN) methods [8]. Traditional GSR models have occasionally been preferred to soft-computing techniques, because of their computational simplicity [9], and have been suggested as complementary tools to more complex physical models [10].

The predictions of models are usually accurate for locations where they can be calibrated, but so far no model has outperformed all others everywhere [11]. Some authors questioned the need to develop new models [12], while others argued that most existing models have yet to be analysed over a wide geographical and climatic range [13]. Depending on the input variables, the models can be classified into sunshine-based, temperature-based, cloud-based, and other climatic parameters-based models [14]. Among the hundreds of existing GSR models [12], a small number of hybrid models include the influence of geographical variables. Many of these models are the result of replacing

Nomenclature

a_i	Empirical parameter
E'	Centred pattern RMSE (kWh/m^2) = $\sqrt{\sum_{i=1}^{n}\left[(s_i-\overline{s_i})^2-(o_i-\overline{o_i})^2\right]\Big/n}$
E'_n	Normalised centred pattern RMSE $= E'/\sigma_o$
H	Global solar irradiation on horizontal surface (kWh/m^2)
H_0	Extraterrestrial global solar irradiation on horizontal surface (kWh/m^2)
KGCC	Köppen-Geiger climate classification
L	Distance to the sea (m)
MBE	Mean bias error (kWh/m^2) = $\sum_{i=1}^{n}(s_i-o_i)/n$
NMBE	Normalised mean bias error $= \left(\sum_{i=1}^{n}(s_i-o_i)/n\right)\big/\overline{o_i}$
NRMSE	Normalised root mean square error = $\left(\sqrt{\sum_{i=1}^{n}(s_i-o_i)^2/n}\right)\big/\overline{o_i}$
NSE	Nash-Sutcliffe model efficiency = $1-\sum_{i=1}^{n}(s_i-o_i)^2\big/\sum_{i=1}^{n}(o_i-\overline{o_i})^2$
o_i	Observed value
R^2	Coefficient of determination = $\left[\sum_{i=1}^{n}(s_i-\overline{s_i})(o_i-\overline{o_i})\Big/\sqrt{\sum_{i=1}^{n}(s_i-\overline{s_i})^2\sum_{i=1}^{n}(o_i-\overline{o_i})^2}\right]^2$
RMSE	Root mean square error (kWh/m^2) = $\sqrt{\sum_{i=1}^{n}(s_i-o_i)^2/n}$
RMBE	Relative mean bias error $= \sum_{i=1}^{n}((s_i-o_i)/o_i)/n$
RRMSE	Relative root mean square error $= \sqrt{\sum_{i=1}^{n}((s_i-o_i)/o_i)^2/n}$
s_i	Simulated value
T_m	Mean air temperature (K)
T_{max}	Maximum air temperature (K)
T_{min}	Minimum air temperature (K)
z	Elevation above sea level (m)
ΔT	Temperature difference (K) = $T_{max}-T_{min}$
φ	Latitude (rad)
λ	Longitude (rad)
σ_o	Standard deviation of experimental solar irradiation (kWh/m^2) = $\sqrt{\left(\sum_{i=1}^{n}(o_i-\overline{o_i})^2\right)\Big/n}$
σ_s	Standard deviation of simulated solar irradiation (kWh/m^2) = $\sqrt{\left(\sum_{i=1}^{n}(s_i-\overline{s_i})^2\right)\Big/n}$
σ_{sn}	Normalised standard deviation $= \sigma_s/\sigma_o$

the coefficients of previous models with functional relationships involving geographical or meteorological variables, which increases the complexity of the original model. For example, hybrid models derived from the Ångström-Prescott model have up to 6 variables and 10 coefficients more than the original model [15,16]. Potential benefits of complexity have to be assessed according to objectives. If the only objective were to improve accuracy by using locally calibrated coefficients, a model with a larger number of variables and adjustment coefficients could be advantageous. However, the calibrated coefficients often vary from site to site, and increasing the number of variables and coefficients reduces the degrees of freedom of a model. Complexity can then limit the generality of a model, i.e., its applicability to large geographical areas. It is therefore understandable that the most comprehensive review to date has recommended seeking the optimal compromise between accuracy and complexity [12].

Daily sunshine fraction has been mostly used in hybrid models involving geographical variables. In fact, out of a total of 94 recently compiled hybrid models [17], latitude, longitude or altitude were included as influencing variables in 16 cases, 11 of which were modifications of classical sunshine-based models, with two regression coefficients in one case and 4 to 12 coefficients in the remaining cases. One of this models, having relative sunshine, altitude and the month index as input variables, explained roughly 90% of the variability in the data from 59 stations in 19 European countries [18]. Precipitation and latitude were the only variables used in another model, with three coefficients [19], while temperature, relative humidity, precipitation, and wind speed were model inputs in other two cases [15,20], with 9 and 11 coefficients, respectively. In the remaining two models [21,22], temperature was the only meteorological variable used.

This paper aims to assess the trade-off between accuracy and generality of models. The target is limited to models of monthly average GSR on horizontal surface, which can be used to estimate irradiance

using semi-empirical models [23]. As air temperature measurement requires low-cost equipment and is widespread [7], temperature-based models have been chosen as the subject of the study. Only one of the above-mentioned temperature-based models is dimensionally homogeneous, which has the advantage of reducing the number of influential variables that may be implicit in the coefficients [17]. An equation was previously proposed to express the only coefficient of this model as a function of the z/L ratio, which provided acceptable results for a large coastal region of northern Spain [17,22]. In this article such a model is compared for the first time with other 13 temperature-based models of low or moderate functional complexity. The coefficients of the 14 models are locally calibrated using data from 105 weather stations, located in three areas covering a large part of peninsular Spain, and then analysed for the first time as a function of the z/L ratio. The analysis leads to modified models that are actually new hybrid models. Given the geographical and climatic variety of the area studied, it is expected that specific results may be of interest to other countries. Since the ratio between coastal and inland stations could influence the model quality, comparisons are made not only based on averages calculated over a set of stations but also by looking at the accuracy of local predictions.

The procedure consists of the following stages, carried out sequentially in each of the three areas:

1) Obtaining, for each model, site-calibrated coefficients at each station, from monthly experimental data averaged over the corresponding time period.

2) Comparison of model performance with site-calibrated coefficients, using appropriate statistical quality indicators.

3) Obtaining new general hybrid models for each geographical area, i.e., obtaining equations that allow the local coefficients to be expressed as a function of geographical variables, using data from eight stations representative of the climatic variety of each region.

4) Comparison between the performance of the new general hybrid

models, using scatter plots and appropriate statistical quality indicators, both at the eight base stations and at the remaining stations used for validation purposes in each area.

As the optimisation of the selection of representative stations in each area is beyond the scope of the article, an approximate evaluation of the representativeness of the eight base stations in each area is finally carried out. For this purpose, the performance of the modified models using general equations derived from data at eight stations is compared with that obtained using general equations derived from data at all stations in each area.

2. Materials and methodology

2.1. Selection of temperature-based GSR models

The models described below have been selected from dozens of existing GSR models that allow indirect estimation of the H/H_0 ratio from air temperature measurements [9,13], with a maximum of three parameters, seeking a representative variety of the variables used and the types of functional relationship.

2.1.1. One-parameter models

Model No.1: Hargreaves and Samani [24] suggested that the clearness index could be estimated from the amplitude of the air temperature oscillation by means of the following equation:

$$H/H_0 = a_1 \Delta T^{0.5} \tag{1}$$

The coefficient a_1 was initially set at 0.17 for arid and semiarid climates, but later values of 0.16 and 0.19 were recommended respectively for inland and near the coast [25]. The site dependence of a_1 is to some extent due to the fact that equation (1) is not homogeneous [22], as it is impossible to construct a dimensionless group using only the variables H, H_0 and ΔT. This parameter should therefore be interpreted as containing implicit variables, despite which this classical model has achieved satisfactory results in many locations.

Model No.2: Meza and Varas [26] *proposed* the following variant of the classical Bristow and Campbell approach [27] to analyse monthly averages of GSR in Chile:

$$H/H_0 = 0.75\left(1 - e^{-a_1 \Delta T^2}\right) \tag{2}$$

Site-calibrated values of a_1 falling in the range from 0.00150 to 0.01944 were obtained for 20 locations in Chile. This dispersion may also be related to some extent to the inhomogeneity of the model. [17].

Model No.3: The model proposed by Weiss et al. [28], also non-homogeneous, has the same functional form and number of parameters as the previous model, but employs the same variables as the Goodin et al. model [29], i.e.:

$$H/H_0 = 0.75\left(1 - e^{-a_1 \Delta T^2/H_0}\right) \tag{3}$$

Model No.4: The model proposed by Annandale et al. [21] is a modification of model No.1 that has outperformed other temperature-based models in regions of complex orography such as India [30], due to the incorporation of an altitude correction factor, i.e.:

$$H/H_0 = a_1\left(1 + 2.7 \bullet 10^{-5}z\right)\Delta T^{0.5} \tag{4}$$

The originally recommended value for the parameter is $a_1 = 0.16$, but a simple dimensional analysis allows to detect that both a_1 and the factor $2.7 \bullet 10^{-5}$ contain implicit variables.

Model No.5: Prieto et al. [22] derived this model from the dimensional analysis of model No.1:

$$H/H_0 = a_1(\Delta T/T_{\min})^{0.5} \tag{5}$$

with a_1 being dependent on z/L, in agreement with the models of Hargreaves and Samani, Allen [31] and Annandale. It has been validated

in the coastal provinces of northern Spain [17].

2.1.2. Two-parameter models

Model No.6: Hargreaves et al. [32] proposed this other evolution of model No.1, which obtained good results under different climatic conditions [33,34] due to the use of two parameters, i.e.:

$$H/H_0 = a_1 + a_2 \Delta T^{0.5} \tag{6}$$

Model No.7: Chen et al. [35] developed this model from daily GSR data at 48 stations in China, obtaining the best results among temperature-based models, such as model No.6. It consists of the following logarithmic equation, which is not homogeneous, so there are implicit variables in at least one of the parameters:

$$H/H_0 = a_1 + a_2 \ln \Delta T \tag{7}$$

Model No.8: Pandey and Katiyar [36] proposed to express the clearness index as polynomial functions of the ratio $T_{\max}/T_{\min}$ instead of the difference ΔT. Although no mention is made of respect, the procedure yields homogeneous equations. The linear version corresponds to the following equation:

$$H/H_0 = a_1 + a_2 T_{\max}/T_{\min} \tag{8}$$

Model No.9: Adaramola [37] proposed this model for various locations in Nigeria, in the absence of data required by sunshine-based models:

$$H/H_0 = a_1 + a_2 T_{\mathrm{m}} \tag{9}$$

Model No.10: Chen and Li [38] developed this model for monthly GSR estimates at 13 stations in Yangtze River Basin in China. It can be considered as the linear, also non-homogeneous, version of model No.6:

$$H/H_0 = a_1 + a_2 \Delta T \tag{10}$$

2.1.3. Three-parameter models

Model No.11: Li et al. [39] developed this model for daily GSR estimates and obtained somewhat better results than model No.6 with data measured in Chongqing, China. The same variables as in model No.8 were used in a non-homogeneous equation with three parameters, i.e.:

$$H/H_0 = a_1 + a_2 T_{\max} + a_3 T_{\min} \tag{11}$$

Model No.12: This model is the second degree polynomial version of equation (8), proposed by the same authors [36]:

$$H/H_0 = a_1 + a_2 T_{\max}/T_{\min} + a_3 \left(T_{\max}/T_{\min}\right)^2 \tag{12}$$

Model No.13: This model is the second-degree polynomial version of equation (9), i.e.:

$$H/H_0 = a_1 + a_2 T_{\mathrm{m}} + a_3 T_{\mathrm{m}}^2 \tag{13}$$

It was proposed by Ohunakin et al. [40] from experimental measurements in Nigeria, assuming $T_{\mathrm{m}} \approx (T_{\max} + T_{\min})/2$, due to data availability.

Model No.14: This model, proposed by Hassan et al. [41], obtained the most accurate GSR estimates among 20 temperature-based GSR models evaluated in Egypt at different locations, especially at coastal sites, using T_{m} as a variable. It is based on the following non-homogeneous equation:

$$H/H_0 = a_1 + a_2 H_0 T_m^{a_3} \tag{14}$$

2.2. Meteorological stations.

Tables A1-A3 in the Appendix show the geographical and climatic characteristics of the 105 meteorological stations used for the model analysis. Stations 1–71 are in Andalusia, a region in southern Spain with an area of approximately 87,600 km2, corresponding to 17.3% of the Spanish territory. Two of its eight provinces are inland, while the
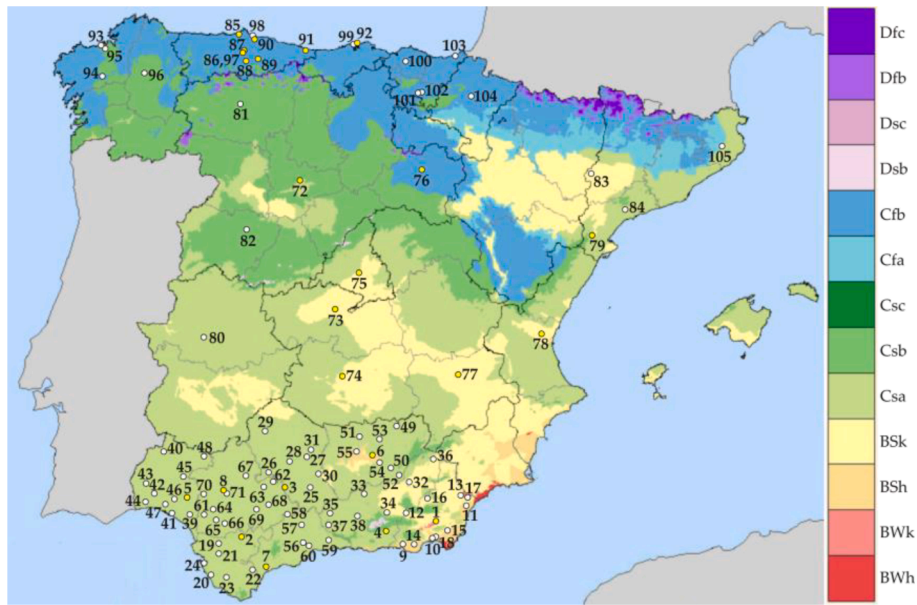
Fig. 1. Köppen-Geiger climate classification of meteorological stations [42].

remaining provinces have abundant coastline with either the Atlantic Ocean or the Mediterranean Sea. Stations No.72–84 are distributed over a wide area, from relatively arid provinces of the central plateau of Spain to the Mediterranean coast. Finally, stations No.85–105 are in coastal provinces of northern Spain, covering a strip of about 1000 km in length but little variation in latitude. As can be seen from the tables and Fig. 1, the geographic, orographic, and climatic diversity is wide.

2.3. Experimental data.

Monthly mean values of maximum and minimum air temperatures and horizontal solar irradiation at both ground and extraterrestrial levels for each weather station during the respective time period are provided as Appendix A in Supplementary Data.

2.4. Statistical indicators

Various methods have been proposed to assess the fit between experimental observations and model estimates but none of them are free of limitations [43]. Since non-expert stakeholders can easily interpret percentage results [44], in this article the performance of the models is assessed using dimensionless statistical indicators, namely the relative root mean square error RRMSE, the relative mean bias error RMBE, and the coefficient of determination R^2. To facilitate comparisons with results from other authors, the Nash-Sutcliffe coefficient of model efficiency, NSE, and the normalised values of the root mean square error and mean bias error, NRMSE and NMBE, have also been calculated, using the mean values of the measurements as references for the latter two.

The accuracy of the models is considered high when the percentage errors are less than 10% [45], while, depending on the degree of
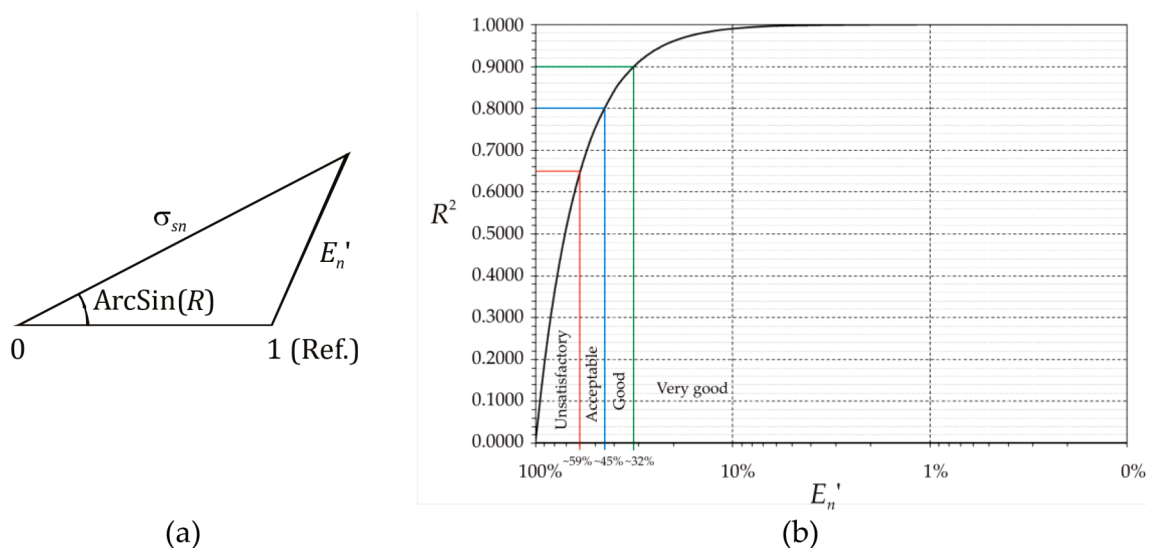

(a)

(b)

Fig. 2. (a) Geometric basis of the Taylor diagram; (b) Variation of the coefficient of determination as a function of the centred pattern RMS difference for $\sigma_{sn} = R$.

agreement between estimates and measurements, R^2 values range from 0 to 1 and NSE values range from $-\infty$ to 1. Negative NSE values suggest that the mean of the measurements is a better predictor than the model estimates themselves [46]. Some authors warn against the risk of identifying the meaning of R^2 and NSE [47], while others try to avoid confusion by classifying the former as an indicator of dispersion and the latter as an indicator of overall performance [44]. For example, while $R^2 = 0.8$ indicates that the model explains 80% of the variance in the observed data, the value $NSE = 0.8$ has a totally different meaning, i.e., that the model mean squared error represents 20% of the observed variance. The cause of possible confusion may be that R^2 can be interpreted as a maximum potential value for NSE, as the following equation reduces to $NSE = R^2$ for the optimal case of $\overline{s_i} = \overline{o_i}$ and $\sigma_{sn} = R$ [46]:

$$NSE = 2\sigma_{sn}R - \sigma_{sn}^2 - \left(\frac{\overline{s_i} - \overline{o_i}}{\sigma_o}\right)^2 \tag{15}$$

Although not yet widely used in GSR model analysis [48], Taylor diagrams have been recommended for visualising results [49]. In this article, this type of chart is used as an alternative to the usual scatter-plots, because it is a graphical tool that provides a concise statistical summary of how well patterns match each other in terms of their correlation, their root-mean-square difference, and the ratio of their variances. The Taylor diagram is based on the definition of the centred pattern RMS difference [49]:

$$E' = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left[(s_i - \overline{s_i})^2 - (o_i - \overline{o_i})^2\right]} \tag{16}$$

which is related to the key statistical indicators by means of the following equations:

$$RMSE = (\overline{s_i} - \overline{o_i})^2 + E'^2 \tag{17}$$

$$(E')^2 = \sigma_o^2 + \sigma_s^2 - 2\sigma_o\sigma_s R \tag{18}$$

The following relationship is derived from normalising equation (18) by σ_o, and is the basis for the graphical representation of the degree of closeness between a model and the reference data set, using dimensionless variables (Fig. 2a):

$$E'_n = \sqrt{1 + \sigma_{sn}^2 - 2\sigma_{sn}R} \tag{19}$$

For the value $\sigma_{sn} = R$, which corresponds to the maximum value of NSE, this equation reduces to the following expression, which can be used as a basis for proposing rating intervals based on E'_n, with criteria similar to those based on NSE [43] (Fig. 2b):

$$R^2 = 1 - \left(E'_n\right)^2 \tag{20}$$

However, it is recommended to perform assessments based on equation (19), as dispensing with σ_{sn} would reduce advantages with respect to other procedures. For example, $E'_n = 20\%$ is obtained for $\sigma_{sn} = R \approx 0.9800$, but if $\sigma_{sn} = 0.8$ it is necessary to reach $R = 0.9999$ to obtain the same deviation.

3. Results and discussion

3.1. Model performance with site-calibrated coefficients

The regression coefficients obtained for each model and station when fitting the experimental monthly GSR data, as well as the corresponding annual mean relative errors, are provided as Appendix B in

Supplementary Data.

In general, all models perform well at all stations in Andalusia (Table B1). The highest errors are obtained by models No.2, 3 and 14, with RRMSE values in the range of 10–20% at 7, 55 and 2 stations, respectively.

The results are not very different for the stations in the central zone (Table B2), with most RRMSE values below 10%. Models No.1, 4, 5 and 9 reach RRMSE values slightly higher than 10% at one station, while model No.2 has values in the range 10–20% at 6 stations and model No.14 reaches values in the range 15–50% at 3 stations. Model No.3 still shows the worst performance, with errors of 10–20% at 10 stations and higher values at the remaining 3 stations.

Similar trends are observed at the northern stations (Table B3), although with slightly higher errors for all models. Models No.1 and 4 reach RRMSE values a few tenths above 10% at station No.92, like model No.5 at stations No.92 and 94. Model No.2 obtains errors in the range 10–20% at 8 stations, while the values of model No.3 are in the range 10–50% at all stations. For each of the stations, it is observed that the lowest RRMSE values are generally obtained using models with the highest number of coefficients. However, at some stations model No.14 shows less favourable results than models No.11–13, or other models with a smaller number of coefficients.

On the other hand, the results obtained at the two stations with data for different periods allow a brief evaluation of the influence of the period on the comparisons. Thus, in station No.86 the site-calibrated coefficients hardly differ from those obtained in station No.97 for models No.1, 2, 4 and 5, as well as in stations No.101 and 102 for models No.1, 2, 4–7 and 10. Therefore, pending more complete studies, it can be deduced that the period is more influential for model No.3 and the models with more parameters.

Table 1 shows the averages of statistical indicators obtained by each model for the set of stations in each geographical area. Normalised values can be converted to ordinary variables taken into account that the average of the 852 monthly GSR data in Andalusia is $\overline{H} = 5.016\,\text{kWh}/(\text{m}^2 \bullet \text{day})$, with a standard deviation $\sigma_o = 1.969\,\text{kWh}/(\text{m}^2 \bullet \text{day})$ caused by the natural variability of irradiation, while the average of the 156 monthly data in central Spain is $\overline{H} = 4.311\,\text{kWh}/(\text{m}^2 \bullet \text{day})$, with standard deviation $\sigma_o = 1.914\,\text{kWh}/(\text{m}^2 \bullet \text{day})$, and the average of the 252 monthly data in northern Spain is $\overline{H} = 3.350\,\text{kWh}/(\text{m}^2 \bullet \text{day})$, with standard deviation $\sigma_o = 1.509\,\text{kWh}/(\text{m}^2 \bullet \text{day})$. When the signs of RMBE and NMBE are not coincident, bias analyses based on the sum of relative values are considered more advisable. As can be seen, all models provide high quality estimates, except model No.3 in all geographical areas and model No.14 in the central area. Model No.11 achieves the best results and errors increase with latitude except for model No.14.

Fig. 3(a), 3(b) and 3(c) show the Taylor diagrams obtained for the performance of the models at the stations of each geographical area. In short, the charts are a clear and meaningful tool for corroborating observations previously derived from tabulated results.

3.2. Model performance using general equations derived from data at eight stations

For each model, an analysis of the geographical dispersion of site-calibrated coefficients has been carried out. In each geographical area, the calibrated coefficients of eight stations have been used to obtain trend lines by the least squares method and the coefficients of the remaining stations have been used for validation purposes.

The analyses are based on the z/L ratio because longitude is not an influential variable, variations in latitude are small in each area, and the

Table 1
Summary of model performance using site-calibrated coefficients.

| Station\Model		No.1	No.2	No.3	No.4	No.5	No.6	No.7	No.8	No.9	No.10	No.11	No.12	No.13	No.14
Stations No.1–71	RRMSE (%)	3.55	6.55	13.28	3.54	4.04	2.43	2.45	2.86	2.86	2.46	1.58	2.64	2.75	4.39
	RMBE (%)	0.20	−1.95	0.21	0.12	0.31	0.06	0.06	0.08	0.07	0.06	0.03	0.07	0.07	0.19
	NRMSE (%)	3.36	4.15	13.19	3.37	3.89	2.41	2.43	2.83	2.70	2.41	1.54	2.55	2.64	4.57
	NMBE (%)	−0.37	−1.05	−4.24	−0.45	−0.56	−0.15	−0.13	−0.23	−0.34	−0.16	−0.08	−0.18	−0.37	0.48
	NSE	0.9927	0.9888	0.8870	0.9926	0.9902	0.9962	0.9962	0.9948	0.9953	0.9962	0.9985	0.9958	0.9955	0.9864
	R^2	0.9935	0.9913	0.9464	0.9936	0.9924	0.9963	0.9962	0.9950	0.9960	0.9963	0.9985	0.9959	0.9962	0.9873
	σ_{sn}	0.9687	1.0383	0.7545	0.9679	0.9509	0.9904	0.9899	0.9858	0.9718	0.9905	0.9950	0.9723	3.2729	1.0210
	$E_n^{'}$(%)	8.52	10.24	31.83	8.51	9.81	6.12	6.18	7.19	6.83	6.14	3.92	6.49	6.65	11.58
Stations No.72–84	RRMSE (%)	5.21	10.33	19.22	5.20	5.70	4.14	4.16	4.11	5.37	4.17	3.43	3.74	5.02	18.14
	RMBE (%)	0.39	−3.60	−1.94	0.26	0.51	0.17	0.16	0.16	0.29	0.17	0.11	0.14	0.25	1.12
	NRMSE (%)	4.97	7.21	20.94	5.00	5.48	4.21	4.26	4.20	4.80	4.20	3.32	3.95	4.45	20.42
	NMBE (%)	−0.61	−1.70	−8.27	−0.74	−0.86	−0.20	−0.20	−0.18	−0.46	−0.20	−0.06	−0.12	−0.43	3.18
	NSE	0.9874	0.9736	0.7776	0.9873	0.9847	0.9910	0.9908	0.9910	0.9883	0.9910	0.9944	0.9921	0.9900	0.7884
	R^2	0.9885	0.9803	0.8337	0.9886	0.9873	0.9911	0.9909	0.9911	0.9886	0.9911	0.9944	0.9921	0.9903	0.8626
	σ_{sn}	0.9643	1.0623	0.7667	0.9629	0.9468	0.9888	0.9869	0.9906	0.9808	0.9903	0.9977	0.9800	0.9555	1.1915
	$E_n^{'}$(%)	11.12	15.78	43.33	11.13	12.20	9.48	9.59	9.45	10.75	9.46	7.47	8.88	9.98	45.44
Stations No.85–105	RRMSE (%)	6.37	10.43	29.66	6.36	6.70	5.09	5.07	5.55	6.32	5.11	4.01	5.25	6.19	6.19
	RMBE (%)	0.53	−2.35	−10.46	0.46	0.56	0.26	0.26	0.31	0.39	0.26	0.17	0.27	0.37	−0.53
	NRMSE (%)	6.00	8.94	36.53	6.02	6.39	5.41	5.38	5.94	5.57	5.45	4.03	5.41	5.36	6.86
	NMBE (%)	−0.95	−1.94	−20.51	−1.02	−1.15	−0.46	−0.45	−0.65	−0.27	−0.47	−0.14	−0.56	−0.25	0.33
	NSE	0.9823	0.9606	0.3420	0.9821	0.9799	0.9856	0.9857	0.9826	0.9847	0.9854	0.9920	0.9856	0.9859	0.9768
	R^2	0.9852	0.9632	0.5764	0.9853	0.9842	0.9862	0.9864	0.9839	0.9847	0.9860	0.9920	0.9866	0.9859	0.9826
	σ_{sn}	0.9422	1.0086	0.5952	0.9414	0.9313	0.9699	0.9699	0.9597	0.9950	0.9697	0.9906	0.9639	0.9952	1.0672
	$E_n^{'}$(%)	13.15	19.38	67.12	13.17	13.95	11.97	11.91	13.11	12.36	12.05	8.94	11.95	11.88	15.22
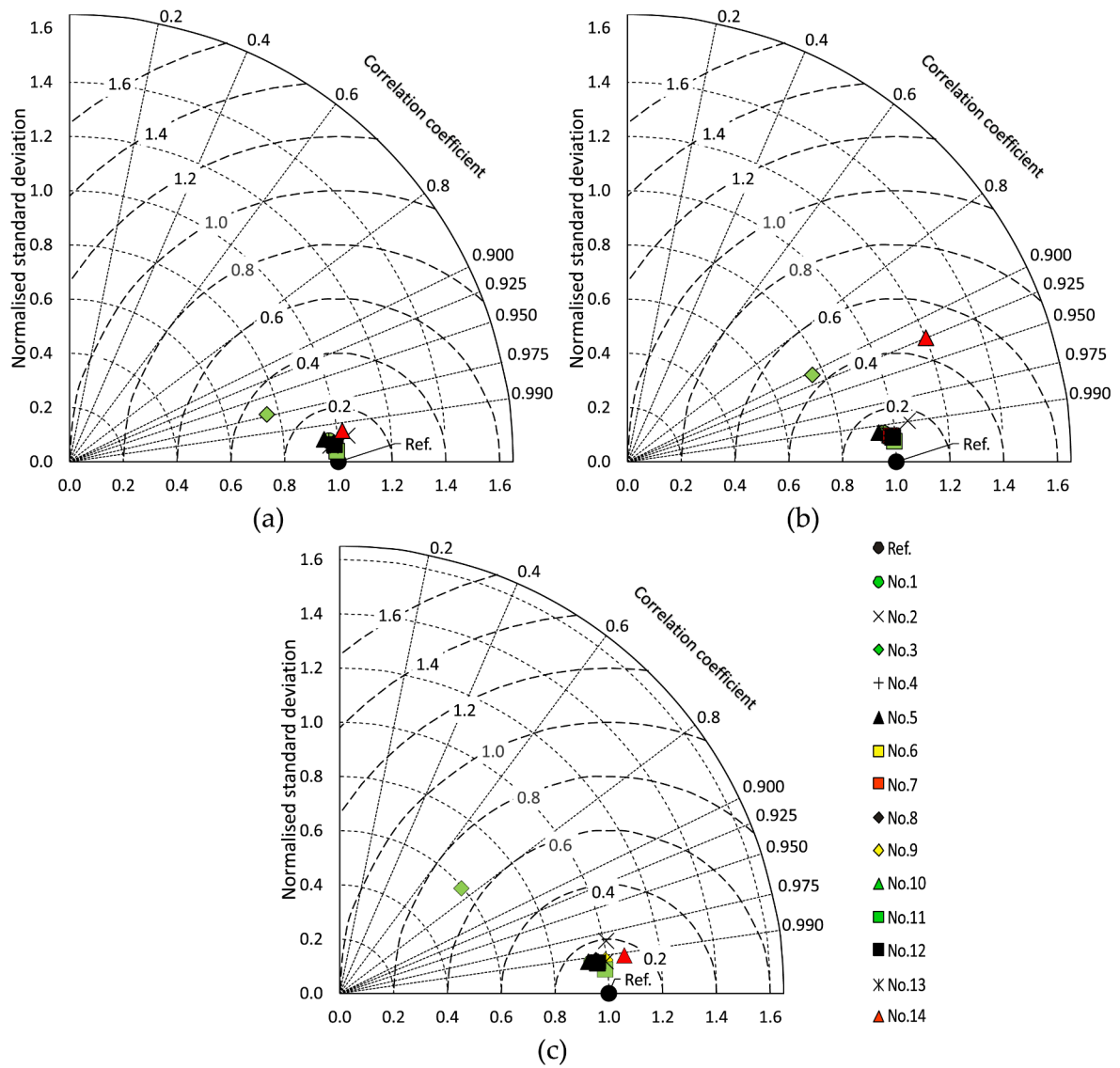
Fig. 3. Comparison of model performance using site-calibrated coefficients: (a) Stations No.1–71; (b) Stations No.72–84; (c) Stations No.85–105.

use of z and L as separate independent variables would lead to non-homogeneous equations. In the search for models of maximum simplicity, polynomial trend lines with a degree greater than two have not been considered.

3.2.1. Andalusia

Although an optimal selection of representative stations for the region is beyond the scope of the article, care has been taken to cover the variety of climatic and geographic characteristics. In this sense, few stations in the region are outside the range $0.5 < z/L < 50$, which roughly corresponds to the selected stations No.1–8. With respect to climate, six stations were selected with Csa climate, which is predomi-

nant in the region (56 stations), another one among the eight stations with BSh climate, and the remaining one is the only station in the region with BWk climate. The six stations with BSk climate are not represented, so their data can be used for validation purposes.

Equations (21) to (48) are the result of the analysis for each model at stations No.1–8. The trend lines derived from the regression analyses are shown in Figs. 4 to 6, where the site-calibrated coefficients for stations No.1–71 are also plotted to visualise the degree of dispersion.

It becomes evident that the performance of the models is discriminated if the coefficients of the models are expressed by general equations as a function of the z/L ratio. In the group of one-parameter models, the R^2 values are relatively low and similar. For model No.5, despite having
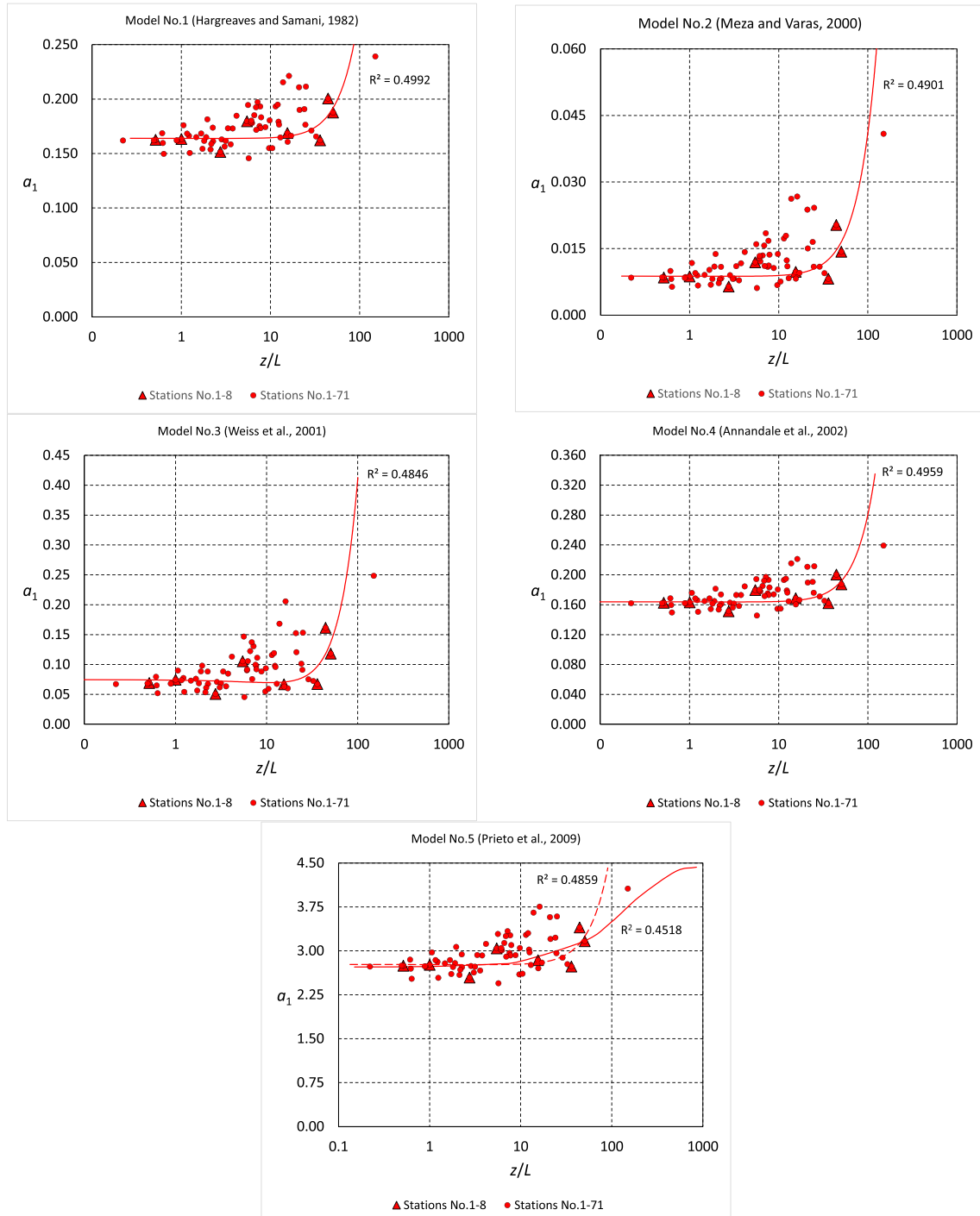

Fig. 4. Dispersion of regression parameters for one-parameter models in Andalusia.
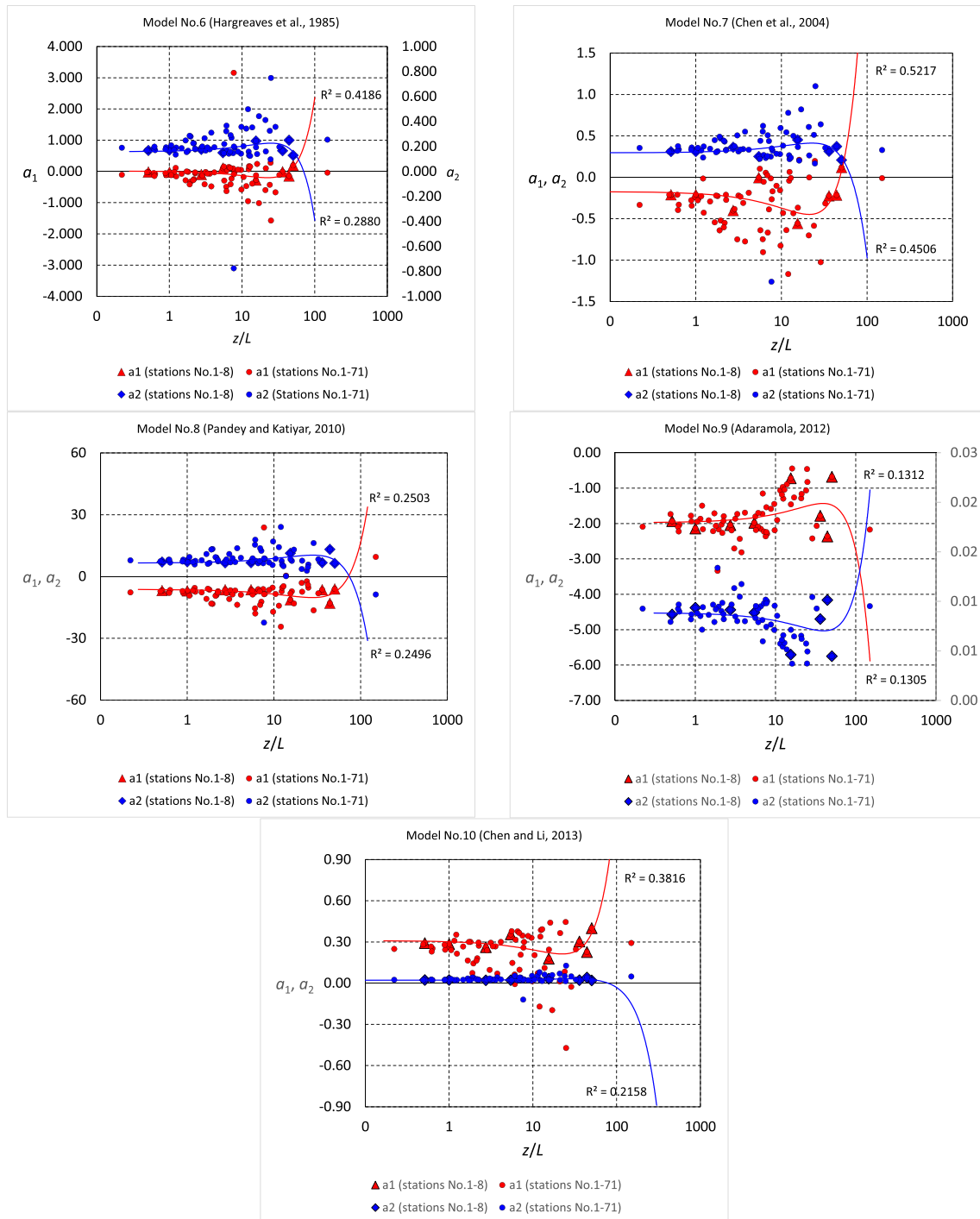
Fig. 5. Dispersion of regression parameters for two-parameter models in Andalusia.
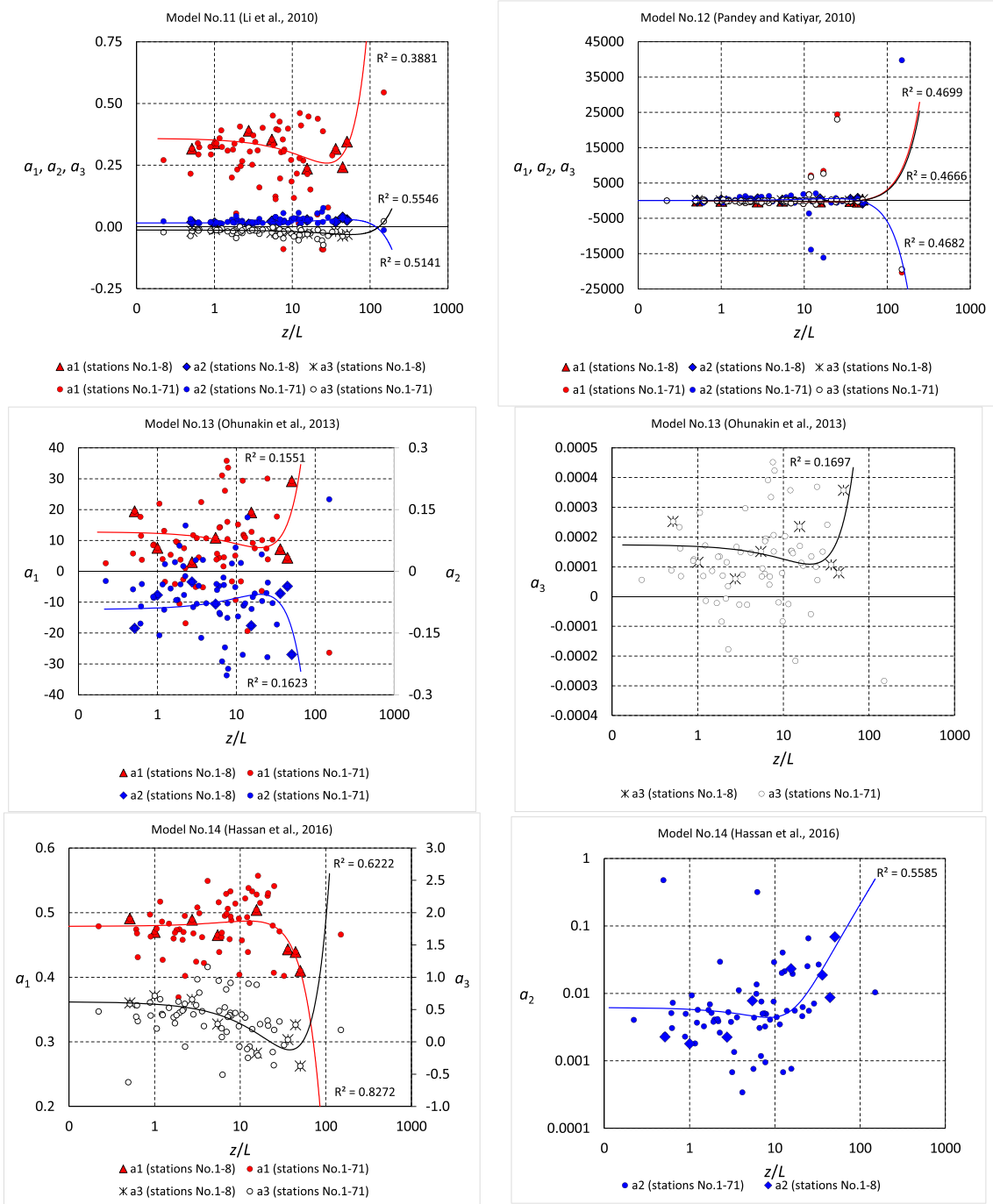
Fig. 6. Dispersion of regression parameters for three-parameter models in Andalusia.

a slightly lower R^2 value, equation (26) is preferred to equation (25), represented by a dashed line, as the functional form of the former avoids physically meaningless parabolic branches that do not agree with observations about the influence of the distance to the sea [6]. The dispersion is greater in the groups of models with two or three parameters, being models No.11, 12 and 14 the only ones that meet or exceed the R^2 values obtained with one-parameter models.

Model No.1:

$$a_1 = 0.164013 - 0.000079z/L + 1.26 \bullet 10^{-5}(z/L)^2 \tag{21}$$

Model No.2:

$$a_1 = 0.008786 - 0.000019z/L + 3.42 \bullet 10^{-6}(z/L)^2 \tag{22}$$

Model No.3:

$$a_1 = 0.074666 - 0.000921z/L + 4.27 \bullet 10^{-5}(z/L)^2 \tag{23}$$

Model No.4:

$$a_1 = 0.163878 - 0.000073z/L + 1.24 \bullet 10^{-5}(z/L)^2 \tag{24}$$

Model No.5:

$$a_1 = 0.163878 - 0.000073z/L + 1.24 \bullet 10^{-5}(z/L)^2 \tag{25}$$

$$a_1 = 4.439 - 1.715e^{-0.0060z/L} \tag{26}$$

Model No.6:

$$a_1 = 0.014949 - 0.019053z/L + 0.000423(z/L)^2 \tag{27}$$

$$a_2 = 0.157909 + 0.005445z/L - 0.000109(z/L)^2 \tag{28}$$

Model No.7:

$$a_1 = -0.171993 - 0.026114z/L + 0.000617(z/L)^2 \tag{29}$$

$$a_2 = 0.295337 + 0.010375z/L - 0.000229(z/L)^2 \tag{30}$$

Model No.8:

$$a_1 = -6.229355 - 0.285993z/L + 0.005130(z/L)^2 \tag{31}$$

$$a_2 = 6.513271 + 0.273878z/L - 0.004865(z/L)^2 \tag{32}$$

Model No.9:

$$a_1 = -1.974561 + 0.027861z/L - 0.000358(z/L)^2 \tag{33}$$

$$a_2 = 0.008850 - 0.000092z/L + 0.000001(z/L)^2 \tag{34}$$

Model No.10:

$$a_1 = 0.309877 - 0.008672z/L + 0.000193(z/L)^2 \tag{35}$$

$$a_2 = 0.021009 + 0.000719z/L - 0.000012(z/L)^2 \tag{36}$$

Model No.11:

$$a_1 = 0.359028 - 0.007241z/L + 0.000130(z/L)^2 \tag{37}$$

$$a_2 = 0.015814 + 0.000577z/L - 0.000006(z/L)^2 \tag{38}$$

$$a_3 = -0.012665 - 0.000651z/L + 0.000006(z/L)^2 \tag{39}$$

Model No.12:

$$a_1 = -16.571563 - 25.516618z/L + 0.566704(z/L)^2 \tag{40}$$

$$a_2 = 24.279863 + 48.696416z/L - 1.081515(z/L)^2 \tag{41}$$

$$a_3 = -7.535744 - 23.233044z/L + 0.516037(z/L)^2 \tag{42}$$

Model No.13:

$$a_1 = 12.797385 - 0.514723z/L + 0.012958(z/L)^2 \tag{43}$$

$$a_2 = -0.092683 + 0.003654z/L - 0.000091(z/L)^2 \tag{44}$$

$$a_3 = 0.000174 - 0.000006z/L + 1.59 \bullet 10^{-7}(z/L)^2 \tag{45}$$

Model No.14:

$$a_1 = 478721 + 0.001332z/L - 0.000054(z/L)^2 \tag{46}$$

$$a_2 = 0.006199 - 0.000420z/L + 0.000025(z/L)^2 \tag{47}$$
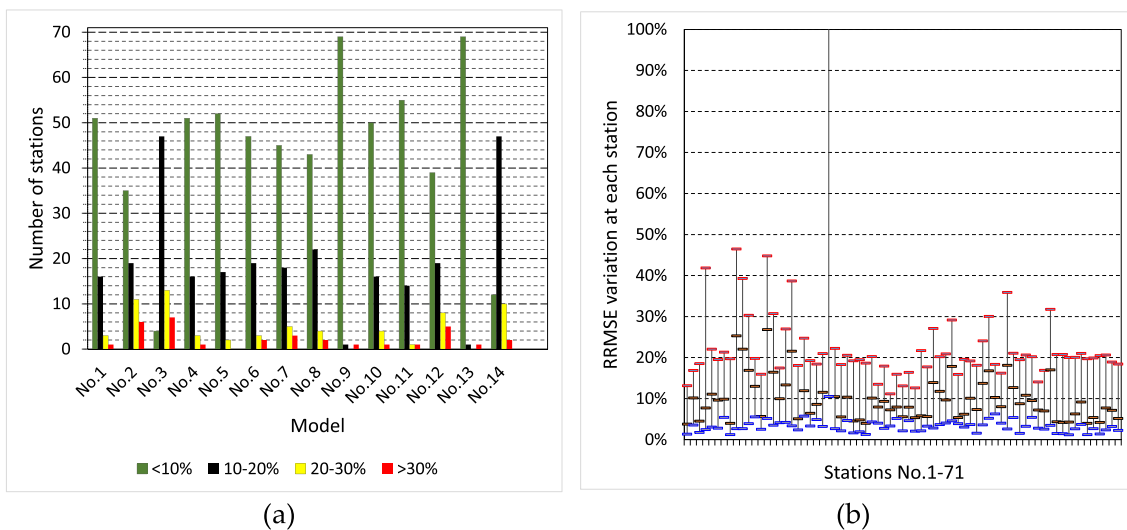

Fig. 7. Andalusia: (a) Number of stations in RRMSE ranges; (b) RRMSE variations at each station.

Table 2
Summary of model performance using general equations derived from data at eight stations in Andalusia.

Station\Model		No.1	No.2	No.3	No.4	No.5	No.6	No.7	No.8	No.9	No.10	No.11	No.12	No.13	No.14
Stations No.1–8	RRMSE (%)	6.57	11.74	14.41	6.58	7.15	7.10	7.31	8.35	4.17	7.14	6.21	8.90	3.92	21.58
	RMBE (%)	0.42	−2.80	1.16	0.44	0.59	0.53	−0.01	1.18	−0.10	1.04	1.45	−0.51	−0.09	−2.03
	NRMSE (%)	6.76	9.34	13.51	6.76	7.19	7.45	7.14	9.27	3.53	8.00	7.11	8.54	3.45	27.32
	NMBE (%)	0.34	−1.41	−2.45	0.32	0.23	0.59	0.09	1.26	−0.53	1.09	1.48	−0.25	−0.54	−2.50
	NSE	0.9698	0.9423	0.8793	0.9698	0.9658	0.9633	0.9663	0.9431	0.9917	0.9577	0.9665	0.9517	0.9921	0.5063
	R^2	0.9707	0.9548	0.8954	0.9707	0.9661	0.9656	0.9678	0.9496	0.9926	0.9617	0.9705	0.9552	0.9929	0.6410
	σ_{sn}	1.0156	1.0829	0.8360	1.0134	1.0005	1.0289	1.0227	1.0481	0.9706	1.0378	1.0350	1.0360	0.9723	1.1620
	$E_n'(\%)$	17.37	23.74	34.17	17.36	18.49	19.10	18.36	23.63	8.98	20.39	17.89	21.96	8.77	69.97
Stations No.1–71	RRMSE (%)	14.10	18.13	20.42	14.01	9.81	46.66	60.33	49.19	7.45	36.31	25.50	117.79	8.77	###
	RMBE (%)	−2.97	−9.94	−6.01	−2.97	−3.58	0.03	0.85	0.29	−1.13	−0.48	0.20	5.37	−1.20	###
	NRMSE (%)	14.23	16.37	22.35	14.14	10.38	47.43	61.36	50.41	7.06	36.91	25.22	121.80	8.03	###
	NMBE (%)	−3.02	−7.77	−9.42	−3.05	−3.94	0.30	1.29	0.49	−1.52	−0.38	0.13	6.27	−1.62	###
	NSE	0.9597	0.8945	0.8145	0.9596	0.9641	0.9450	0.9354	0.9348	0.9972	0.9521	0.9697	0.8932	0.9971	0.9200
	R^2	0.8782	0.8767	0.7365	0.8795	0.9410	0.3962	0.2874	0.3668	0.9693	0.5152	0.6973	0.1033	0.9598	0.0053
	σ_{sn}	0.9983	1.0439	0.8016	0.9954	0.9398	1.5553	1.8518	1.6139	0.9725	1.3501	1.1672	3.2729	0.9776	###
	$E_n'(\%)$	35.43	36.73	51.64	35.19	24.47	120.87	156.32	128.45	17.56	94.06	64.27	309.97	20.04	###

$$a_3 = 0.622002 - 0.039309z/L + 0.000518(z/L)^2 \qquad (48)$$

The annual mean relative errors obtained from general equations (21) to (48) for each model at each station in Andalusia are included as Table C1 of Appendix C in Supplementary Data. It is observed that the lowest errors are obtained with models No.9 and 13, and the highest with models No.3 and 14, while models No.1, 4, 5, 10 and 11 lead to results that seem acceptable and similar. On the other hand, it is observed that all models except model No.5 lead to very high errors for the highest value of z/L, i.e., at station No.24. Model No.5 obtains a value of RRMSE = 10.49% at this station, while it reaches the maximum value of 23.36% at station No.14. The remaining models have errors greater than 30% at some station. With respect to the six BSk climate stations, no differences with results obtained in other zones can be observed.

Fig. 7(a) shows for each model the number of stations obtaining RRMSE values in four percentage ranges, while Fig. 7(b) shows the RRMSE variations obtained with the 14 models at each station.

The averages of statistical indicators calculated for stations No.1–8 and for all 71 stations, listed in Table 2 and depicted in Fig. 8(a) and 8 (b), facilitate a more precise analysis. For the eight stations on which equations (21) to (48) are based, the performance of all models except No.3 and 14 could be qualified as very good using the criteria in Fig. 2 (b). With the same equations and criteria, the estimates of models No.9, 13 and 5 would be very good for all 71 stations, while those of models No.1, 2 and 4 would be good and the performance of the remaining models worsens significantly.

3.2.2. Central Spain

In this case, the dispersion analysis of the model coefficients is based on data from stations No.72–79, i.e., six inland stations and two on the Mediterranean coast, while three stations to the west and two to the east are used for validation purposes. A large area in the interval $1 < z/L < 12$ and a wide climatic variety are thus covered.

Equations (49) to (76) are the result of the analysis carried out for each model at stations No.72–79, following the same procedure as in the previous case. The trend lines derived from the regression analyses are shown in Figs. 9 to 11, where the site-calibrated coefficients for stations used for validation purposes are also plotted to visualise the degree of dispersion.

Except in the case of model No.13, the coefficients of determination obtained at stations No.72–79 are lower than those obtained at stations No.1–8 in Andalusia whatever the number of parameters in the model. For model No.5 a near-zero value of R^2 is obtained by fitting the locally calibrated coefficients to a functional relationship formally similar to equation (26), so it is not surprising that the result is $a_1 = 4.386 - 1.714e^{0.0000} = 2.672$, which is close to the mean of the coefficients, i.e., $a_1 = 2.646$. As the figures show, models No.1 and 4 are not far from analogous performance in the z/L range studied.

Model No.1:

$$a_1 = 0.164955 - 0.002044z/L + 8.58 \bullet 10^{-5}(z/L)^2 \qquad (49)$$

Model No.2:

$$a_1 = 0.007436 + 0.001135z/L - 1.77 \bullet 10^{-4}(z/L)^2 \qquad (50)$$

Model No.3:

$$a_1 = 0.064900 + 0.003023z/L - 7.10 \bullet 10^{-4}(z/L)^2 \qquad (51)$$

Model No.4:

$$a_1 = 0.164349 - 0.001815z/L + 5.66 \bullet 10^{-5}(z/L)^2 \qquad (52)$$

Model No.5:

$$a_1 = 2.760237 - 0.022105z/L + 2.760237(z/L)^2 \qquad (53)$$

$$a_1 = 2.672 \qquad (54)$$

Model No.6:

$$a_1 = -0.724725 + 0.351466z/L - 0.031761(z/L)^2 \qquad (55)$$

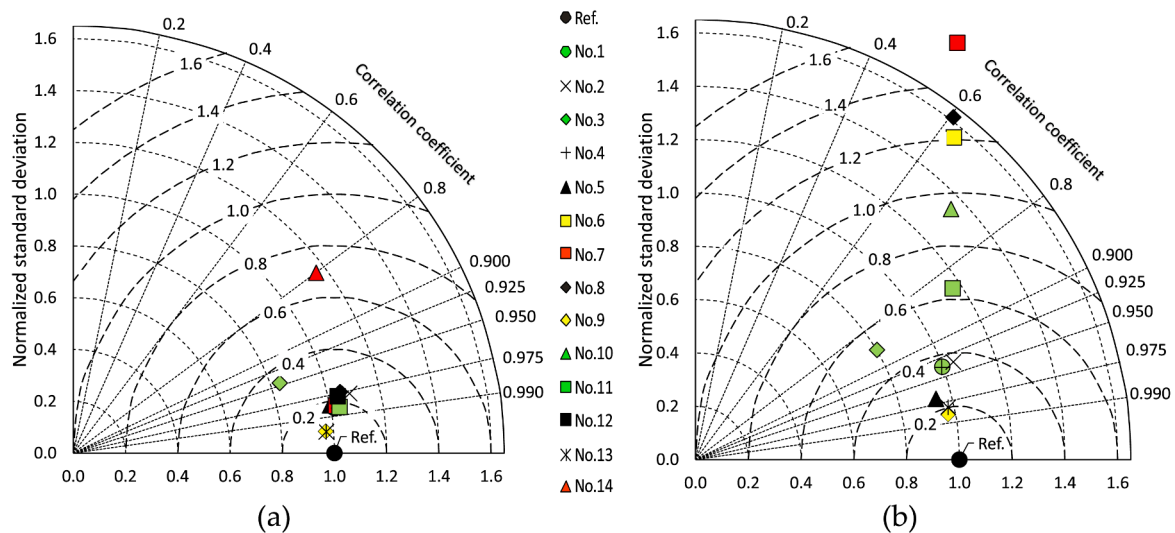$$a_2 = 0.382516 - 0.109364z/L + 0.009957(z/L)^2 \qquad (56)$$

Fig. 8. Comparison of model performance using general equations derived from data at eight stations in Andalusia: (a) Stations No.1–8; (b) stations No.1–71.
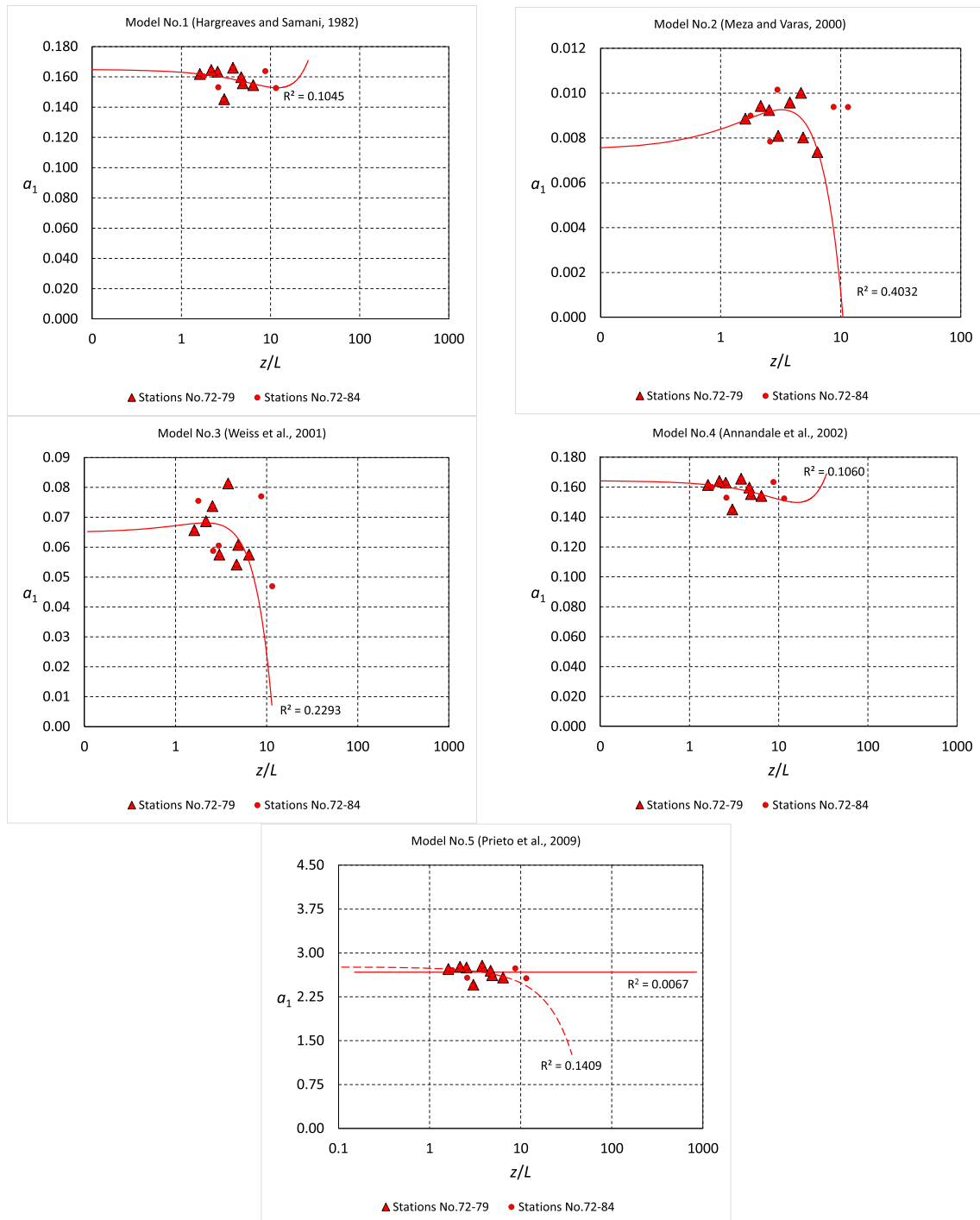
Fig. 9. Dispersion of regression parameters for one-parameter models in central Spain.
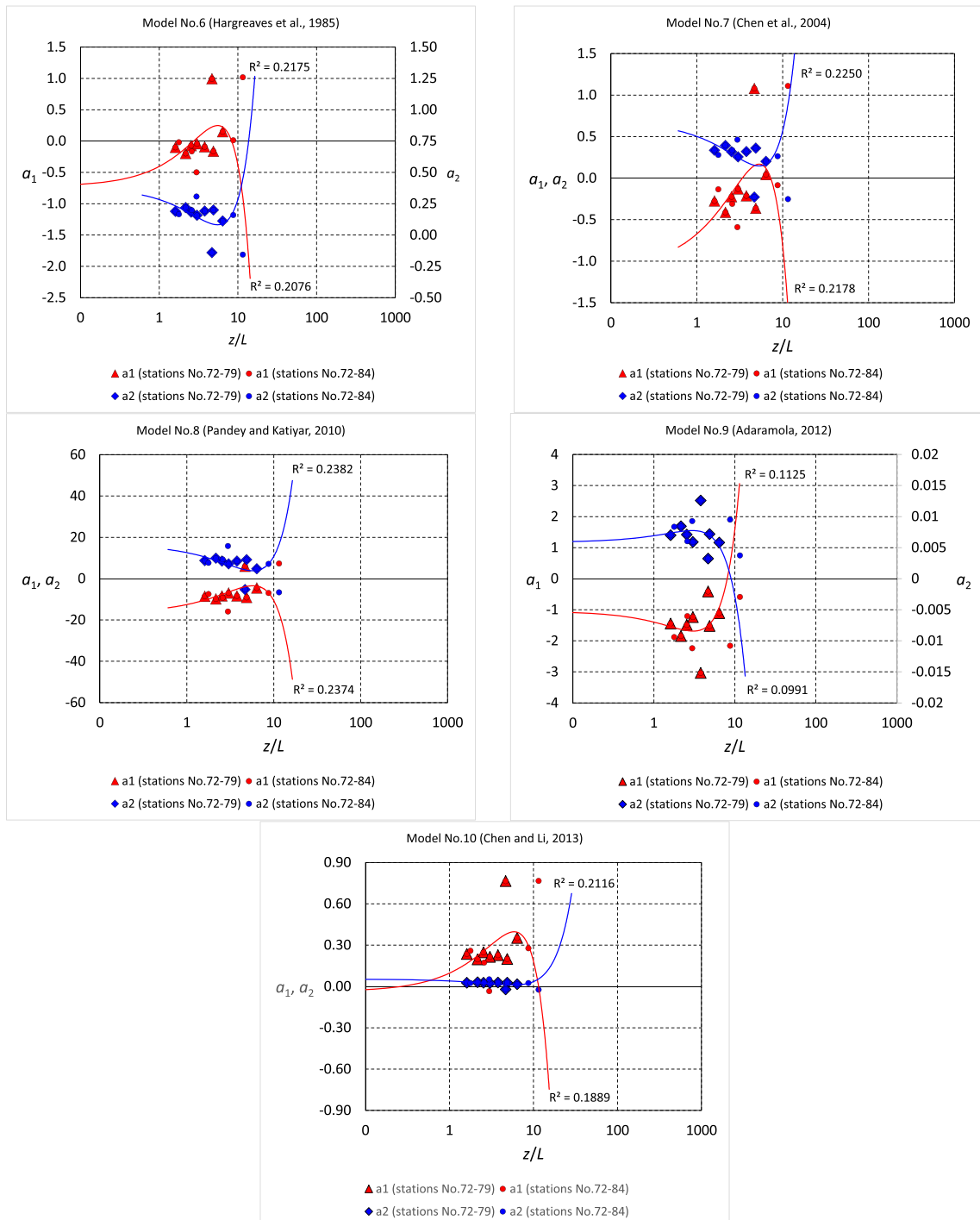
Fig. 10. Dispersion of regression parameters for two-parameter models in central Spain.
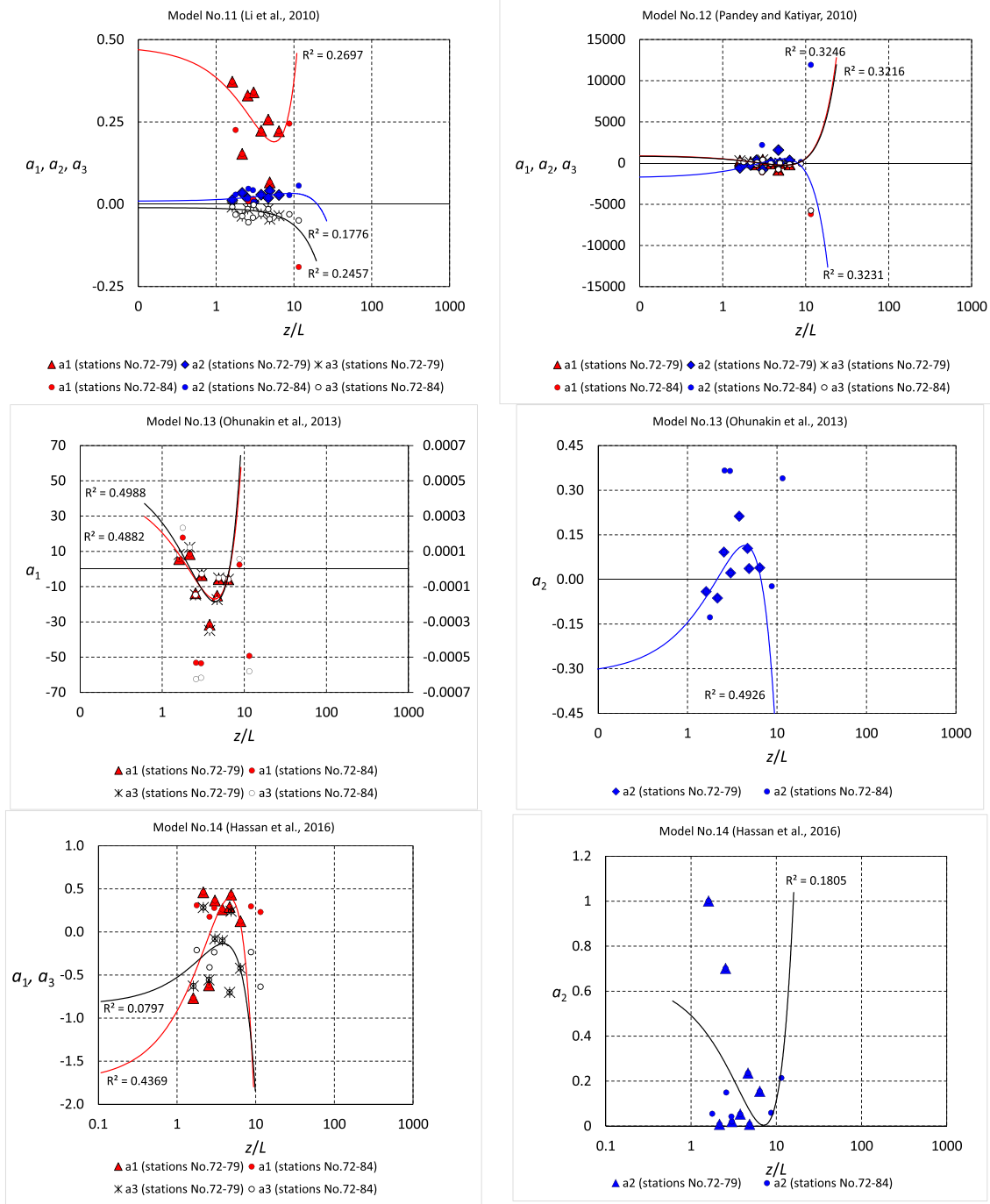
Fig. 11. Dispersion of regression parameters for three-parameter models in central Spain.

Model No.7:

$$a_1 = -1.105429 + 0.478799z/L - 0.044989(z/L)^2 \qquad (57)$$

$$a_2 = 0.686702 - 0.203369z/L + 0.019159(z/L)^2 \qquad (58)$$

Model No.8:

$$a_1 = -16.766892 + 4.616474z/L - 0.400258(z/L)^2 \qquad (59)$$

$$a_2 = 16.651543 - 4.451209z/L + 0.386091(z/L)^2 \qquad (60)$$

Model No.9:

$$a_1 = -1.049686 - 0.417381z/L + 0.068256(z/L)^2 \qquad (61)$$

$$a_2 = 0.005867 + 0.001271z/L - 0.000215(z/L)^2 \qquad (62)$$

Model No.10:

$$a_1 = -0.038519 + 0.148164z/L - 0.012605(z/L)^2 \qquad (63)$$

$$a_2 = 0.053217 - 0.014714z/L + 0.001289(z/L)^2 \qquad (64)$$

Model No.11:

$$a_1 = 0.479686 - 104432z/L + 0.009400(z/L)^2 \qquad (65)$$

$$a_2 = 0.009127 + 0.005210z/L - 0.000284(z/L)^2 \qquad (66)$$

Fig. 12. Central Spain: (a) Number of stations in RRMSE ranges; (b) RRMSE variations at each station.

$$a_3 = -0.010229 - 0.002188z/L - 0.000317(z/L)^2 \qquad (67)$$
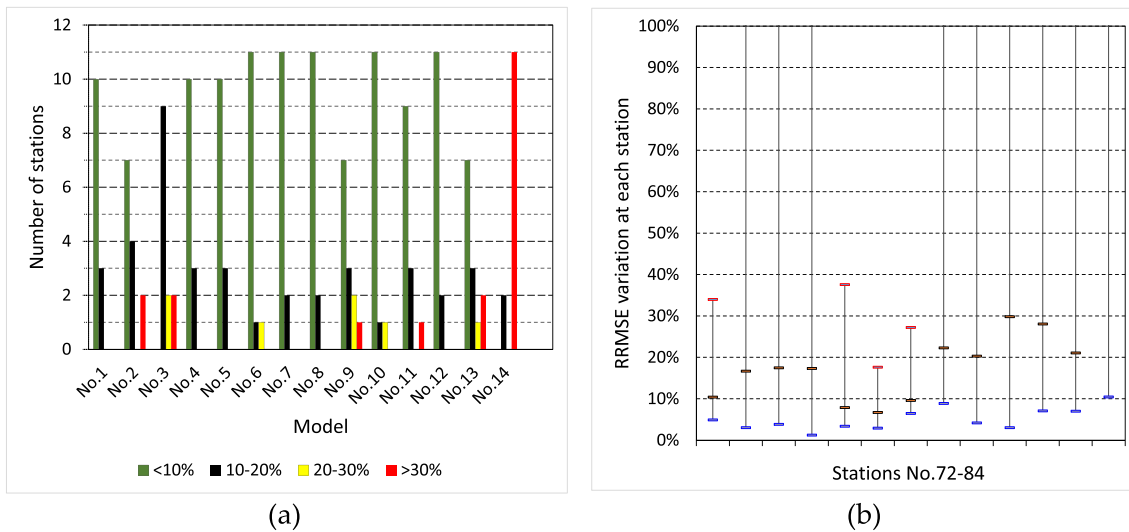
Model No.12:

$$a_1 = 919.852486 - 438.474010z/L + 40.404874(z/L)^2 \qquad (68)$$

$$a_2 = -1782.197657 + 846.999992z/L - 78.061084(z/L)^2 \qquad (69)$$

$$a_3 = 863.702912 - 409.039734z/L + 37.703265(z/L)^2 \qquad (70)$$

Model No.13:

$$a_1 = 46.092197 - 28.956920z/L + 3.321779(z/L)^2 \qquad (71)$$

$$a_2 = -0.320486 + 0.198886z/L - 0.022739(z/L)^2 \qquad (72)$$

$$a_3 = 0.000565 - 0.000342z/L + 0.000039(z/L)^2 \qquad (73)$$

Model No.14:

$$a_1 = -1.733084 + 0.911482z/L - 0.097708(z/L)^2 \qquad (74)$$

$$a_2 = 1.683487 - 0.706755z/L + 0.074787(z/L)^2 \qquad (75)$$

$$a_3 = -0.846567 + 0.366024z/L - 0.047162(z/L)^2 \qquad (76)$$

The annual mean relative errors obtained for each model at each station in central Spain are provided as Table C2 of Appendix C in Supplementary Data. It is observed that models No.1, 4 and 5 predict the lowest RRMSE values, while the highest values correspond to models No.3 and 14, which obtains errors greater than 10% at all stations. Most models predict high RRMSE values for the highest value of z/L, i.e., at station No.84, where models No.1, 4 and 5 reach values slightly higher than 10%. Models No.6–8, 10 and 12 lead to similar results to the best ones except for station No.84.

Fig. 12(a) shows for each model the number of stations obtaining RRMSE values in four percentage ranges, while Fig. 12(b) shows the RRMSE variations obtained with the 14 models at each station.

The averages of statistical indicators calculated for stations No.72–79 and for the 13 stations as a whole, listed in Table 3 and

depicted in Fig. 13(a) and 13(b), facilitate a more precise analysis. For the eight stations on which equations (49) to (76) are based, the performance of all models except No.3 and 14 could be qualified as very good using the criteria in Fig. 2(b). With the same equations and criteria, the estimates of models No.1, 4–8, 10 and 12 would also be very good for the set of 13 stations, while the performance of the remaining models worsens significantly.

3.2.3. Northern Spain

In this case, the dispersion analysis of the model coefficients is based on data from the only seven stations in the Principality of Asturias with GSR records during 2003–2016, and from the nearest station in neighbouring provinces with data during the same period. For validation purposes, another 13 data series are used, corresponding to stations located along a wide strip of the northern coast of Spain, with little climatic diversity but a range of z/L that is wider than in previous cases, i.e., $4 < z/L < 400$, approximately.

Equations (77) to (104) are the result of the analysis carried out for each model at the eight stations used as the basis of the study, following the same procedure as in the previous cases. The trend lines derived from the regression analyses are shown in Figs. 14 to 16, where the site-calibrated coefficients for stations used for validation purposes are also plotted to visualise the degree of dispersion. As the figures show, except for models No. 13 and 14, the coefficients of determination obtained in stations No.85–92 are higher than in the previous regions, especially for the one-parameter models. In the set of 21 stations, the degree of dispersion increases for all models except for model No.5, using equation (82).

Model No.1:

$$a_1 = 0.139833 + 0.000181z/L + 2.11 \bullet 10^{-8}(z/L)^2 \qquad (77)$$

Model No.2:

$$a_1 = 0.014543 - 0.000171z/L + 6.91 \bullet 10^{-7}(z/L)^2 \qquad (78)$$

Model No.3:

$$a_1 = 0.067672 - 0.000259z/L + 1.71 \bullet 10^{-6}(z/L)^2 \qquad (79)$$

Table 3
Summary of model performance using general equations derived from data at eight stations in central Spain.

Station\Model		No.1	No.2	No.3	No.4	No.5	No.6	No.7	No.8	No.9	No.10	No.11	No.12	No.13	No.14
Stations No.72–79	RRMSE (%)	5.76	10.07	16.31	5.79	6.20	6.11	6.11	5.95	8.49	6.15	6.26	7.16	8.36	124.15
	RMBE (%)	0.43	−3.28	−0.70	0.47	0.57	−0.56	−0.53	−0.81	1.24	−0.59	−1.38	−1.30	1.10	84.08
	NRMSE (%)	6.11	8.28	17.09	6.13	6.79	7.15	7.21	6.99	9.02	7.12	7.04	9.31	8.78	159.75
	NMBE (%)	−0.45	−1.34	−6.55	−0.50	−0.71	−1.71	−1.65	−1.91	0.36	−1.76	−1.83	−2.61	0.20	96.29
	NSE	0.9801	0.9633	0.8441	0.9799	0.9754	0.9727	0.9722	0.9739	0.9565	0.9729	0.9735	0.9537	0.9588	−12.6343
	R^2	0.9811	0.9716	0.9190	0.9811	0.9779	0.9764	0.9759	0.9780	0.9566	0.9767	0.9755	0.9594	0.9588	0.4861
	σ_{sn}	0.9610	1.0710	0.7305	0.9574	0.9414	0.9418	0.9412	0.9426	0.9810	0.9422	0.9720	0.9342	0.9759	3.5550
	E'_n(%)	14.08	18.89	36.47	14.12	15.60	16.04	16.22	15.54	20.83	15.94	15.72	20.66	20.29	294.63
Stations No.72–84	RRMSE (%)	7.27	47.87	30.17	7.34	7.52	8.74	8.42	8.05	25.97	9.32	20.12	8.89	25.76	298.12
	RMBE (%)	1.24	−18.85	−8.99	1.15	1.67	1.92	1.41	1.10	11.34	2.41	−1.36	1.43	11.17	3.64
	NRMSE (%)	6.44	46.03	31.96	6.55	6.56	7.29	7.22	6.84	21.10	7.59	22.60	8.61	20.49	309.47
	NMBE (%)	−0.07	−15.50	−14.34	−0.23	−0.04	0.46	0.13	−0.39	8.49	0.78	−3.83	−0.21	8.16	23.08
	NSE	0.9790	−0.0751	0.4816	0.9782	0.9782	0.9730	0.9736	0.9762	0.7741	0.9708	0.7409	0.9624	0.7869	−47.5952
	R^2	0.9797	0.4893	0.5813	0.9791	0.9805	0.9738	0.9742	0.9776	0.8268	0.9716	0.7606	0.9631	0.8329	0.0814
	σ_{sn}	0.9579	1.4139	0.8843	0.9543	0.9401	0.9610	0.9626	0.9524	1.0641	0.9604	1.0120	0.9555	1.0515	7.3618
	E'_n(%)	14.61	101.05	65.84	14.89	14.82	16.39	16.26	15.40	44.40	17.03	50.89	19.39	43.17	714.10
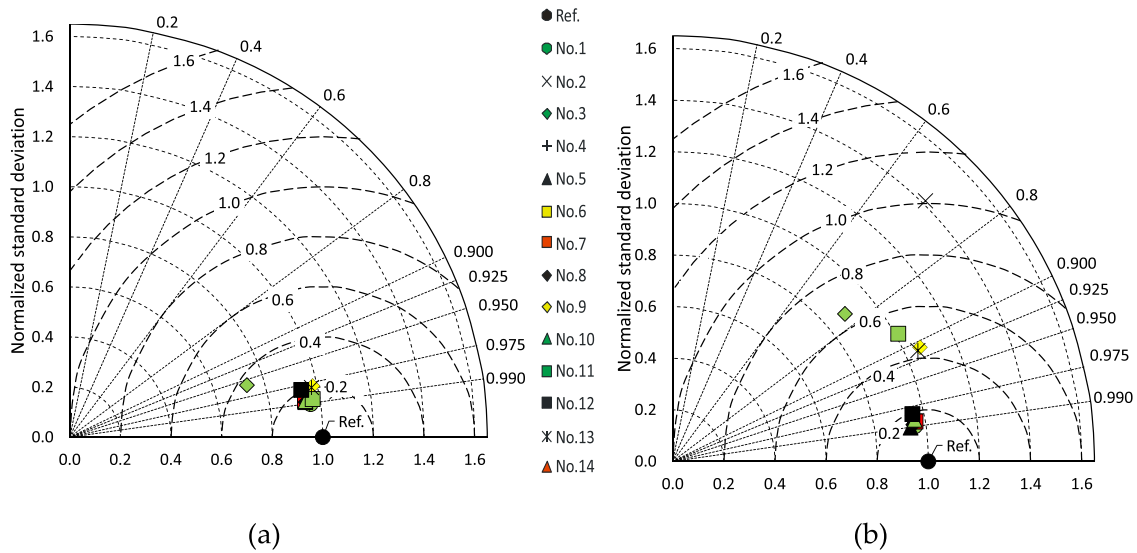
Fig. 13. Comparison of model performance using general equations derived from data at eight stations in central Spain: (a) Stations No.72–79; (b) Stations No.72–84.

Model No.4:

$$a_1 = 0.139742 + 0.000183z/L + 1.68 \bullet 10^{-8}(z/L)^2 \tag{80}$$

Model No.5:

$$a_1 = 2.351488 + 0.003207z/L + 6.15 \bullet 10^{-9}(z/L)^2 \tag{81}$$

$$a_1 = 3.485 - 1.350e^{-0.0191z/L} \tag{82}$$

Model No.6:

$$a_1 = 0.106637 - 0.023557z/L + 0.000065(z/L)^2 \tag{83}$$

$$a_2 = 0.072081 + 0.011131z/L - 0.000030(z/L)^2 \tag{84}$$

Model No.7:

$$a_1 = 0.027675 - 0.018355z/L + 0.000052(z/L)^2 \tag{85}$$

$$a_2 = 0.137861 + 0.012377z/L - 0.000034(z/L)^2 \tag{86}$$

Model No.8:

$$a_1 = -2.522739 - 0.611678z/L + 0.001679(z/L)^2 \tag{87}$$

$$a_2 = 2.752747 + 0.602542z/L - 0.001653(z/L)^2 \tag{88}$$

Model No.9:

$$a_1 = 0.387247 - 0.035580z/L + 0.000083(z/L)^2 \tag{89}$$

$$a_2 = 0.001388 - 0.003579a_1 \tag{90}$$

Model No.10:

$$a_1 = 0.244437 - 0.011180z/L + 0.000031(z/L)^2 \tag{91}$$

$$a_2 = 0.007666 + 0.002467z/L - 0.000007(z/L)^2 \tag{92}$$

Model No.11:

$$a_1 = 0.156048 - 0.006212z/L + 0.000017(z/L)^2 \tag{93}$$

$$a_2 = 0.025634 + 0.001109z/L - 0.000003(z/L)^2 \tag{94}$$

$$a_3 = -0.025711 - 0.001009z/L + 0.000003(z/L)^2 \tag{95}$$

Model No.12:

$$a_1 = 3540.407793 - 211.316172z/L + 0.502365(z/L)^2 \tag{96}$$

$$a_2 = -6909.900957 + 412.544347z/L - 0.980072(z/L)^2 \tag{97}$$

$$a_3 = 3371.847460 - 201.339882z/L + 0.477983(z/L)^2 \tag{98}$$

Model No.13:

$$a_1 = 4.826063 - 0.835812z/L + 0.002077(z/L)^2 \tag{99}$$

$$a_2 = -0.030319 + 0.005664z/L - 0.000014(z/L)^2 \tag{100}$$

$$a_3 = 0.000052 - 0.000010z/L + 2.39 \bullet 10^{-8}(z/L)^2 \tag{101}$$

Model No.14:

$$a_1 = 0.360945 - 0.000221z/L + 0.000001(z/L)^2 \tag{102}$$

$$a_2 = 0.007763e^{-2.064736a_3} \tag{103}$$

$$a_3 = -2.020620 - 0.038092z/L + 0.000125(z/L)^2 \tag{104}$$

The annual mean relative errors obtained for each model at each station in northern Spain are provided as Table C3 of Appendix C in Supplementary Data. It is observed that all models predict higher RRMSE values than in the previous regions, which are approximately less than 20% at all stations except for models No.2, 3 and 12. The error
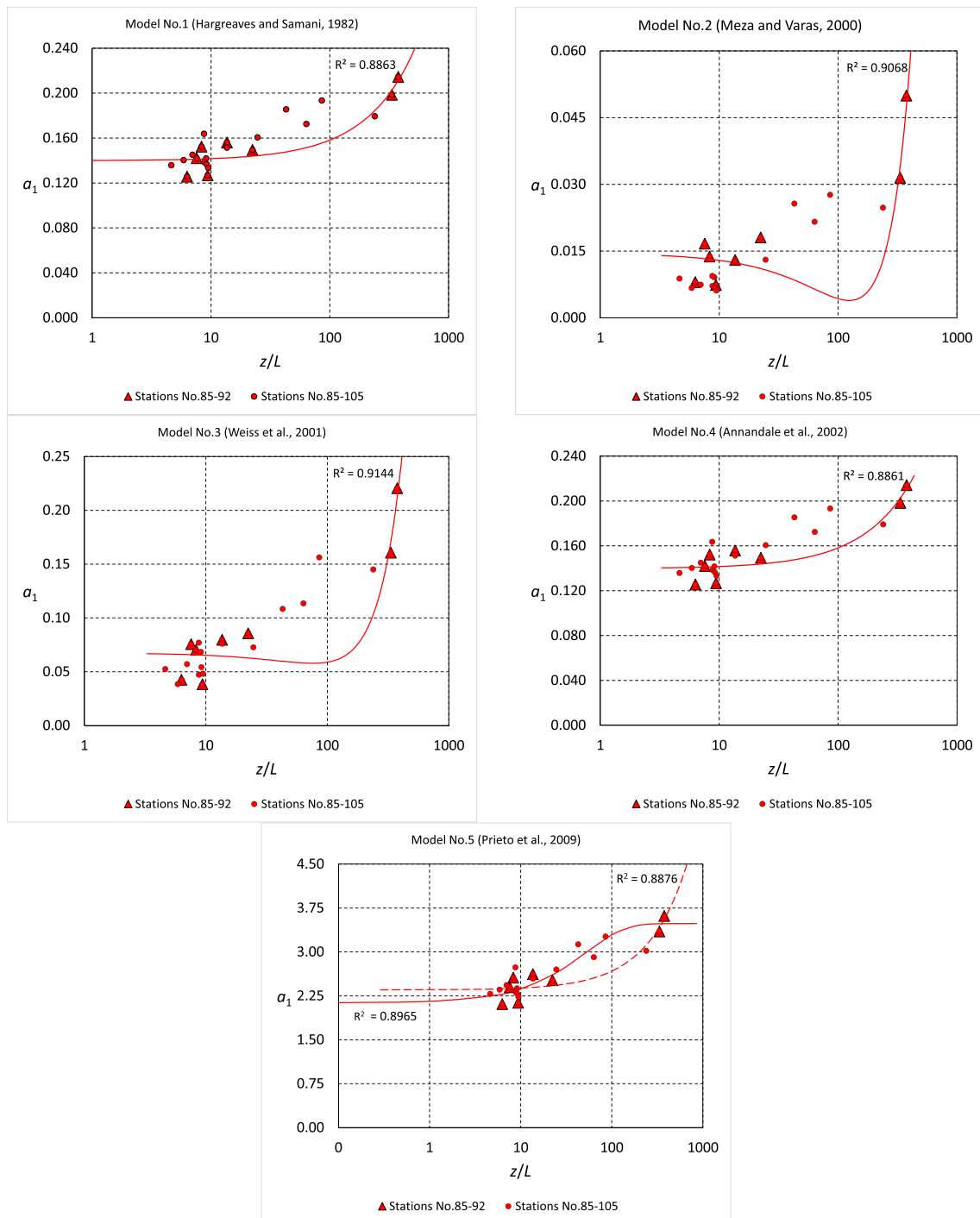
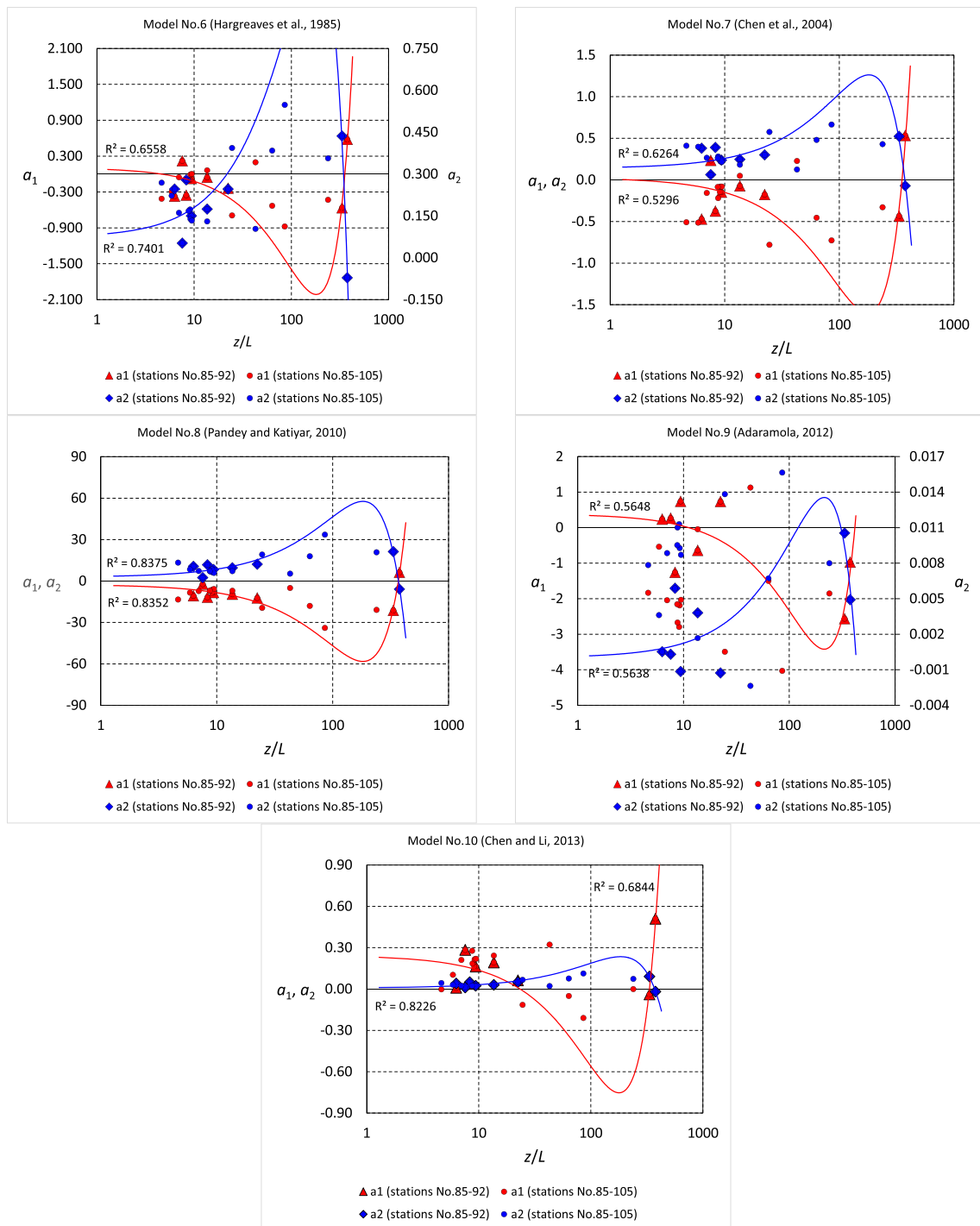**Fig. 14.** Dispersion of regression parameters for one-parameter models in northern Spain.

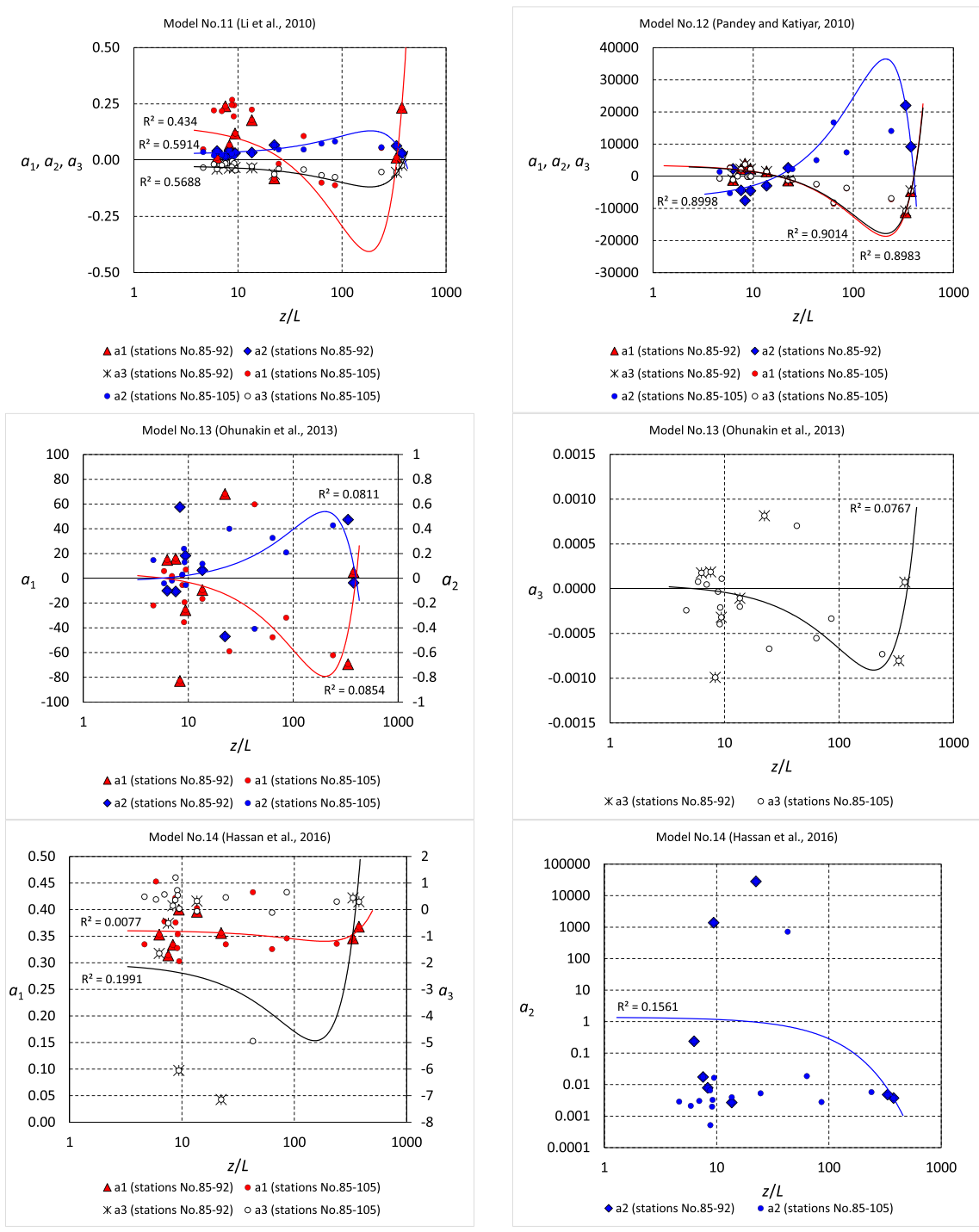**Fig. 15.** Dispersion of regression parameters for two-parameter models in northern Spain.

Fig. 16. Dispersion of regression parameters for three-parameter models in northern Spain.
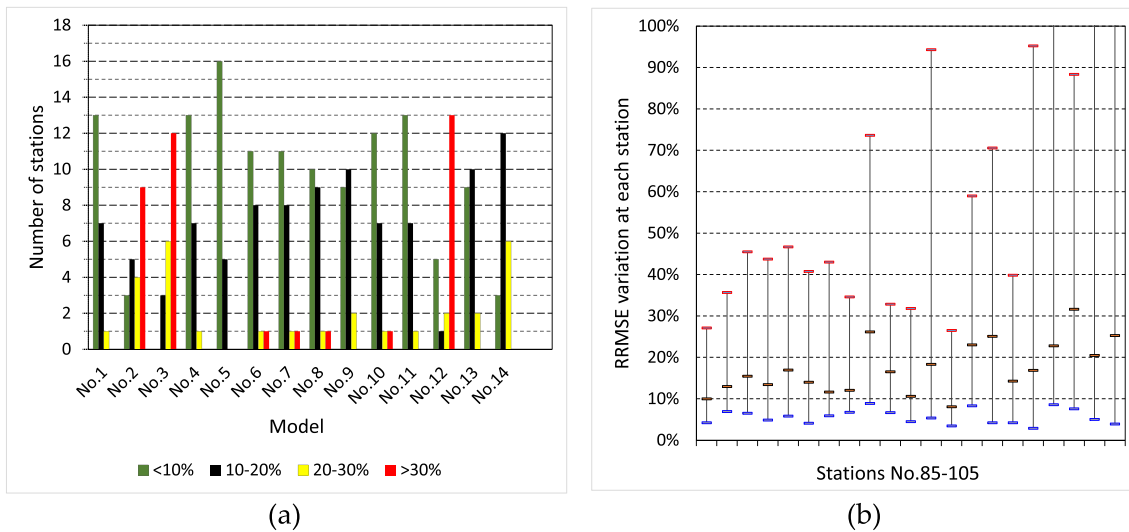
Fig. 17. Northern Spain: (a) Number of stations in RRMSE ranges; (b) RRMSE variations at each station.

variability is not associated with extreme values of the z/L ratio, which seems to indicate that stations No.85–92 are representative of the model performance for the set of 21 stations.

Fig. 17(a) shows for each model the number of stations obtaining RRMSE values in four percentage ranges, while Fig. 17(b) shows the RRMSE variations obtained with the 14 models at each station. It is observed that all models obtain values above 20% at some station except model No.5, that the mean RRMSE value of the 14 models is between 10 and 20% at each of the stations No.85–92, and that this error is higher at six other stations.

The averages of statistical indicators calculated for stations No.85–92 and for the set of 21 stations, listed in Table 4 and depicted in Fig. 18(a) and 18(b), facilitate a more precise analysis. For the eight stations on which equations (77) to (104) are based, the performance of all models except No.2, 3, 12 and 14 could be qualified as very good using the criteria in Fig. 2(b). With the same equations and criteria, the estimates of models No.1, 4, 5 and 11 would also be very good for the set of 21 stations. The best performance is achieved by model No.5, while models No.2, 3 and 12 are not suitable for this area.

3.3. Model performance using general equations derived from data at all stations.

In order to roughly assess the representativeness of the eight stations selected as the basis for each regional analysis, trend lines of the model coefficients as a function of z/L have been obtained using calibrated data at all stations in each area. The results derived from the same procedure used in the previous sections, which are summarised in Table 5 and Fig. 19, broadly corroborate those obtained from the inter-model comparisons based on eight stations in each area.

As can be seen, models No. 9 and 13 obtain the best results in Andalusia, followed by models No. 1, 4 and 5, with small differences between the latter. In the central area, models No.1, 4–8, 10 and 11 obtain the best results, again with differences in $E_n^{'}$ of the order of 1%. Finally, in the northern zone, model No.5 obtains the best results, with differences in $E_n^{'}$ of the order of 3% with respect to models No.1 and 4.

On the other hand, among the models with good performance, model No.5 is the only one with values of $E_n^{'}$ that differ little from those given in Tables 2-4.

Equations (105) to (116) correspond to the best performing models in each area, while Tables 6 and 7 facilitate the comparison between results obtained by the best models using respectively equations (21) to (104) and equations (105) to (116). The differences between the averages indicated in both tables can be considered as margins of improvement that a model could reach depending on the climatic representativeness of the stations selected to obtain the general equations of the model coefficients. With the data used, there is no doubt that the averages of model No.9 outperform other results in Andalusia. However, it should also be noted that these averages are favoured because, among the 71 stations in Andalusia, the z/L ratio is high only at station No.24, where this model obtains RRMSE = 50.32% and is therefore not acceptable from the perspective of local performance. As a low average error calculated in a large set of stations may mask important errors at some individual stations, the distribution of stations in different z/L ranges is of paramount importance in the averaging of local results, which has not been noticed so far.

• Model No.1:

Stations No.1–71:

$$a_1 = 0.166715 + 0.000801(z/L) - 2.25 \bullet 10^{-6}(z/L)^2 \qquad (105)$$

Stations No.72–84:

$$a_1 = 0.158368 - 0.000258(z/L) - 3.56 \bullet 10^{-6}(z/L)^2 \qquad (106)$$

Stations No.84–105:

$$a_1 = 0.139656 + 0.000477(z/L) - 8.42 \bullet 10^{-7}(z/L)^2 \qquad (107)$$

• Model No.5:

Stations No.1–71:

Table 4

Summary of model performance using general equations derived from data at eight stations in northern Spain.

| Station\Model		No.1	No.2	No.3	No.4	No.5	No.6	No.7	No.8	No.9	No.10	No.11	No.12	No.13	No.14
Stations No.85–92	RRMSE (%)	9.94	24.13	36.94	9.95	9.70	9.51	10.02	9.50	10.64	9.24	10.43	26.61	10.56	13.99
	RMBE (%)	0.96	2.33	−13.78	0.97	0.87	−2.04	−2.85	−2.02	1.18	−1.33	0.27	15.34	1.24	−3.84
	NRMSE (%)	10.63	25.53	44.91	10.62	10.49	10.81	11.35	10.77	11.38	10.53	11.59	26.96	11.29	18.80
	NMBE (%)	−0.37	1.43	−24.96	−0.38	−0.66	−2.88	−3.68	−3.08	0.37	−2.16	0.06	13.13	0.41	−4.47
	NSE	0.9372	0.6375	−0.1221	0.9373	0.9388	0.9350	0.9283	0.9355	0.9279	0.9383	0.9252	0.5958	0.9291	0.8034
	R^2	0.9374	0.7198	0.2322	0.9375	0.9396	0.9397	0.9360	0.9413	0.9292	0.9409	0.9290	0.7517	0.9304	0.8254
	σ_{sn}	0.9582	1.1333	0.5699	0.9581	0.9454	0.9602	0.9565	0.9467	0.9994	0.9647	1.0254	1.1118	0.9987	1.0126
	E'_n (%)	25.05	60.11	88.07	25.02	24.70	24.57	25.32	24.34	26.84	24.31	27.35	55.52	26.61	43.06
Stations No.85–105	RRMSE (%)	10.52	34.11	35.99	10.55	9.20	14.98	15.06	14.75	13.10	15.23	10.99	68.97	12.93	18.28
	RMBE (%)	−1.82	−1.94	−11.79	−1.79	0.91	2.21	1.30	2.28	−4.19	3.17	1.46	44.05	−4.87	−11.66
	NRMSE (%)	12.33	39.26	43.20	12.28	9.11	16.94	17.02	16.14	16.64	17.30	12.97	93.65	16.65	24.79
	NMBE (%)	−3.39	−1.47	−21.11	−3.58	−0.92	1.67	0.59	1.44	−6.58	2.83	1.97	50.79	−7.22	−14.89
	NSE	0.9250	0.2400	0.0800	0.9257	0.9590	0.8586	0.8571	0.8716	0.8635	0.8525	0.9171	−3.3238	0.8633	0.6970
	R^2	0.9316	0.5695	0.3459	0.9316	0.9612	0.8764	0.8707	0.8838	0.8886	0.8782	0.9322	0.5948	0.8945	0.8221
	σ_{sn}	0.9354	1.3277	0.8033	0.9347	0.9382	1.0642	1.0488	1.0459	0.8815	1.0844	1.0804	2.3981	0.8719	0.7810
	E'_n (%)	26.32	87.12	83.69	26.33	20.13	37.42	37.78	35.69	33.93	37.89	28.46	174.70	33.31	44.01
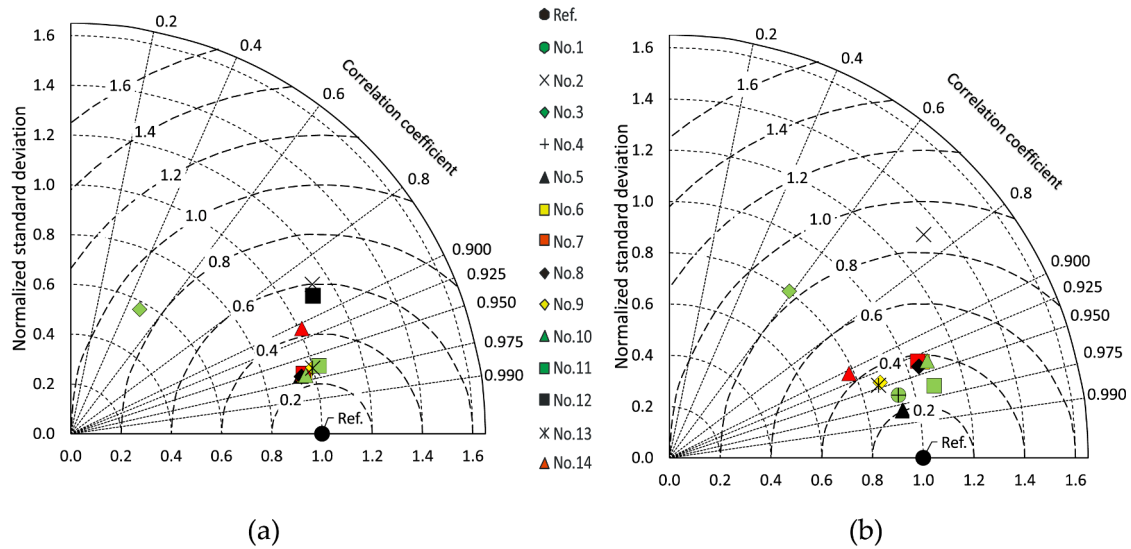
Fig. 18. Comparison of model performance using general equations derived from data at eight stations in northern Spain: (a) Stations No.85–92; (b) Stations No.85–105.

$$a_1 = 4.509 - 1.701\mathrm{e}^{-0.0085z/L} \tag{108}$$

Stations No.72–84:

$$a_1 = 2.646 \tag{109}$$

Stations No.84–105:

$$a_1 = 3.332 - 1.225\mathrm{e}^{-0.0261z/L} \tag{110}$$

- Model No.9:

Stations No.1–71:

$$a_1 = -2.044716 + 0.033104(z/L) - 0.000231(z/L)^2 \tag{111}$$

$$a_2 = 0.009101 - 0.000111(z/L) + 0.000001(z/L)^2 \tag{112}$$

Stations No.72–84:

$$a_1 = -1.559798 - 0.065362(z/L) + 0.010916(z/L)^2 \tag{113}$$

$$a_2 = 0.007196 + 0.000270(z/L) - 0.000041(z/L)^2 \tag{114}$$

Stations No.84–105:

$$a_1 = -0.924541 - 0.016093(z/L) + 0.000040(z/L)^2 \tag{115}$$

$$a_2 = 0.001460 - 0.003538a_1 \tag{116}$$

Consistent with other analyses [50], it can be observed that: a) decreasing the number of constant coefficients may reduce accuracy but increases simplicity, i.e., degrees of freedom, as is the case for models No.1–5; b) such an increase implies that a lower R-squared is needed for the significance of the model at a given confidence level; and c) generality is improved by using ratios of variables as model inputs, as is the case for model No.5. On the other hand, the ability of model No.5 to adapt to different climatic and geographical conditions is favoured by the use of functional relationships of the type of equation (26), with physically meaningful asymptotes at the extremes of the z/L definition

range. This advantage is most noticeable in the northern coastal area, where the percentage of stations with high z/L is notorious. Therefore, pending criteria to avoid parabolic branches characteristic of other models for high z/L values, model No.5 is recommended for areas in Spain with scarce radiometric data. In addition, Fig. 20 suggests that further improvements can be expected from future work by analysing the influence of latitude on the geographical variability of model coefficients.

4. Conclusions

Fourteen temperature-based GSR models have been compared using data from 105 stations located in three areas covering a large part of peninsular Spain and its climatic variety.

Using site-calibrated coefficients, accuracy generally improves for models of higher functional complexity. Model No.11 achieves the best results in all three areas, however, most models have high accuracy and none improve on the others at any location.

The variation of the local coefficients in each geographical area was analysed as a function of the z/L ratio, obtaining equations based on data from eight stations and using the remaining ones for validation purposes. In Andalusia, models No.9 and 13, followed by model No.5, provided the most accurate estimates. All models except model No.5 lead to very high errors for the highest value of z/L in this area, i.e., at station No.24, where the model No.5 obtains a RRMSE value of 10%, approximately. In the central area, models No.1, 4 and 5 achieve the best performance, with RRMSE values in the order of 10% at the station with the highest z/L value, i.e., station No.84, where the remaining models predict inaccurate values. In this area, the coefficient of model No.5 is independent of z/L, even though some stations are close to the border with Portugal and others on the Mediterranean coast. In the coastal area of northern Spain, model No.5 achieves the best results, followed at a moderate distance by models No.1 and 4, and further away from the remaining models. Thus, the z/L ratio seems to be a suitable variable to represent geographical and topographical influences on the model performance, but inter-model comparisons should take into account that

Table 5
Summary of model performance using general equations derived from data at all stations in each area.

Station\Model		No.1	No.2	No.3	No.4	No.5	No.6	No.7	No.8	No.9	No.10	No.11	No.12	No.13	No.14
Stations No.1–71	RRMSE (%)	9.10	12.93	17.55	9.13	9.54	10.69	10.78	10.38	4.51	10.86	8.75	18.32	4.44	21.49
	RMBE (%)	0.49	−1.34	0.65	0.51	0.64	0.56	−0.19	−0.21	0.12	1.33	1.10	−0.78	0.17	−20.45
	NRMSE (%)	9.17	11.60	17.16	9.18	9.51	11.15	10.79	10.91	3.97	11.82	9.50	22.52	3.99	26.69
	NMBE (%)	0.20	−0.38	−3.38	0.17	0.05	0.81	0.10	−0.16	−0.34	1.56	1.48	−1.53	−0.31	−22.03
	NSE	0.9745	0.9547	0.8997	0.9744	0.9728	0.9616	0.9623	0.9660	0.9972	0.9587	0.9739	0.9272	0.9971	0.8756
	R^2	0.9459	0.9214	0.8256	0.9458	0.9414	0.9258	0.9288	0.9259	0.9904	0.9202	0.9488	0.7285	0.9903	0.9719
	σ_{sn}	0.9948	1.0528	0.8118	0.9928	0.9786	1.0409	1.0298	1.0188	0.9706	1.0559	1.0517	1.0906	0.9716	0.6403
	E'_n (%)	23.36	29.53	42.87	23.38	24.23	28.35	27.49	27.80	10.08	29.86	23.92	57.24	10.14	38.41
Stations No.72–84	RRMSE (%)	7.00	10.87	18.96	7.05	7.29	6.94	6.88	6.74	11.18	7.04	7.22	12.90	11.27	86.15
	RMBE (%)	0.59	−3.50	−1.29	0.62	0.68	0.62	0.43	0.63	1.51	0.82	−0.62	−2.78	1.12	60.09
	NRMSE (%)	6.37	8.76	19.56	6.40	6.78	6.68	6.78	6.59	10.39	6.63	6.92	17.85	10.38	110.36
	NMBE (%)	−0.74	−1.57	−7.83	−0.79	−1.01	−0.71	−0.91	−0.57	0.00	−0.49	−0.98	−5.16	−0.33	69.10
	NSE	0.9794	0.9610	0.8059	0.9792	0.9767	0.9774	0.9767	0.9780	0.9452	0.9777	0.9757	0.8384	0.9454	−5.1798
	R^2	0.9813	0.9695	0.8755	0.9814	0.9805	0.9790	0.9789	0.9788	0.9452	0.9787	0.9761	0.8520	0.9455	0.6074
	σ_{sn}	0.9487	1.0710	0.7359	0.9455	0.9310	0.9509	0.9455	0.9599	0.9660	0.9572	0.9926	0.9119	0.9605	2.6187
	E'_n (%)	14.31	19.49	40.55	14.38	15.16	15.01	15.18	14.86	23.42	14.95	15.47	38.49	23.37	194.31
Stations No85–105	RRMSE (%)	9.94	20.38	31.99	9.98	8.91	10.15	10.98	10.04	12.41	9.72	24.76	23.56	12.42	13.44
	RMBE (%)	1.15	1.60	−7.99	1.16	0.83	−1.40	−2.41	−0.92	1.80	−0.45	17.45	11.60	1.77	−5.23
	NRMSE (%)	10.59	23.66	37.31	10.48	9.13	11.77	12.63	11.69	13.29	11.38	27.35	29.65	13.29	17.25
	NMBE (%)	−0.46	2.54	−17.72	−0.75	−1.01	−2.15	−3.21	−1.89	0.11	−1.14	17.55	11.44	0.05	−7.38
	NSE	0.9447	0.7241	0.3135	0.9458	0.9589	0.9317	0.9214	0.9326	0.9129	0.9361	0.6312	0.5666	0.9129	0.8534
	R^2	0.9449	0.8116	0.4697	0.9449	0.9614	0.9343	0.9266	0.9345	0.9129	0.9375	0.8864	0.7802	0.9130	0.8843
	σ_{sn}	0.9592	1.1913	0.7220	0.9584	0.9358	0.9838	0.9754	0.9777	0.9633	0.9955	1.2629	1.2695	0.9600	0.8762
	E'_n (%)	23.50	52.22	72.91	23.52	20.15	25.69	27.12	25.61	29.52	25.15	46.58	60.73	29.50	34.62
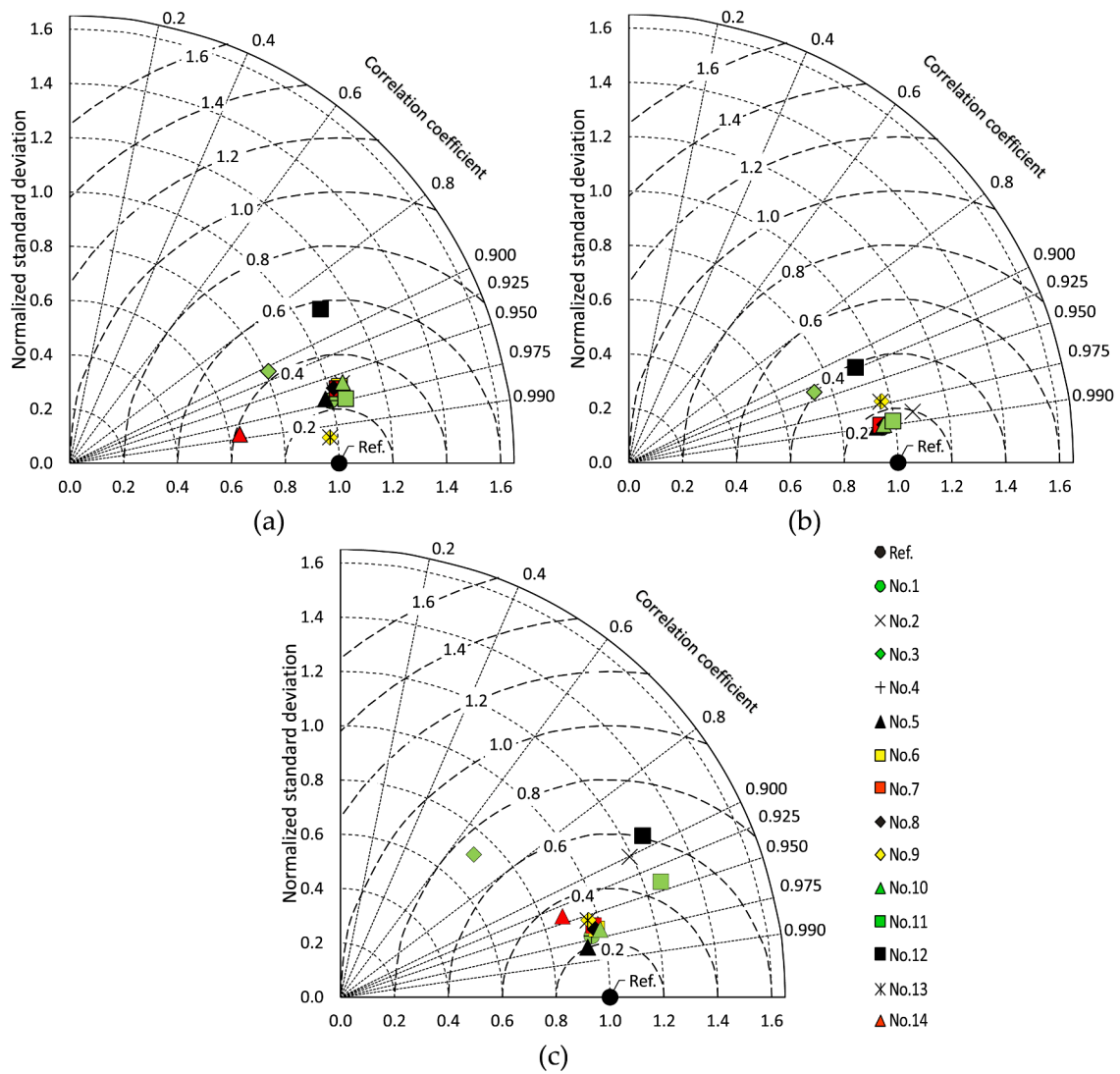
Fig. 19. Comparison of model performance using general equations derived from data at all stations in each area: (a) Stations No.1–71; (b) Stations No.72–84; (c) Stations No.85–105.
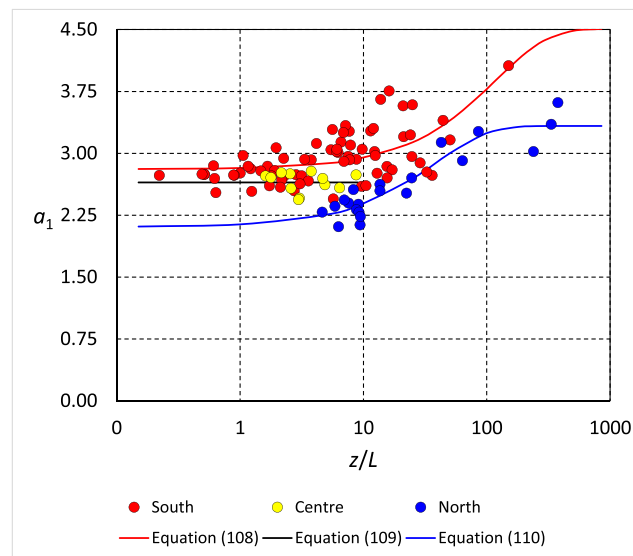
Table 6

Comparison between best performing models in each area using coefficients derived from data measured at eight stations.

Model	Stations No.1–71		Stations No.72–84		Stations No.85–105	
	RRMSE (%)	E_n' (%)	RRMSE (%)	E_n' (%)	RRMSE (%)	E_n' (%)
No.9	7.45	**17.56**	25.97	44.40	13.10	33.93
No.1	14.10	35.43	7.27	**14.61**	10.52	26.32
No.5	9.81	24.47	7.52	14.82	9.20	**20.13**

Table 7

Comparison between best performing models in each area using coefficients derived from data measured at all stations.

Model	Stations No.1–71		Stations No.72–84		Stations No.85–105	
	RRMSE (%)	E_n' (%)	RRMSE (%)	E_n' (%)	RRMSE (%)	E_n' (%)
No.9	4.51	**10.08**	11.18	23.42	12.41	29.52
No.1	9.10	23.36	7.00	**14.31**	9.94	23.50
No.5	9.54	24.23	7.29	15.16	8.91	**20.15**

Fig. 20. Comparison between experimental data and predictions of model No.5 in three areas of Spain.

low average errors in a group of stations may include unacceptable local results.

The accuracy of the models increases with coefficients derived from equations based on data from all available stations in each region, but the ranking of the models does not vary, with model No.5 being the only one that obtains good results in all areas and with hardly any variation of centred pattern RMS difference in both scenarios. The ability of this model to adapt to different locations is due to its simplicity, its dimensional homogeneity and the equation used to express its unique coefficient, with physically significant asymptotes for the extreme values of z/L instead of the parabolic branches of other models.

The results obtained are applicable to the models and experimental data used, but the climatic variety of the Spanish stations considered allows predicting that the analysis procedure may be of interest for other countries.

Funding

CRediT authorship contribution statement

Jesús-Ignacio Prieto: Conceptualization, Methodology, Writing – original draft, Supervision. **David García:** Data curation, Formal analysis, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The article contains a link to a Supplementary Data file.

Appendix A

.

Table A1

Geographical and climatic characteristic of weather stations in Andalusia.

Station No.	Location	Province	φ (°)	λ (°)	z (m)	L (km)	z/L (m/km)	KGCC	Period
1	Tabernas[a]	Almería	37.09	2.30	437.5	28.24	15.49	BWk	2000–2018
2	Villamartín[a]	Cádiz	36.84	5.62	173.5	4.81	36.07	Csa	
3	Santaella[a]	Córdoba	37.52	4.89	209.5	209.05	1.00	Csa	
4	Cádiar[a]	Granada	36.92	3.18	952.5	18.88	50.45	Csa	
5	La Palma del Condado[a]	Huelva	37.37	6.54	194.5	35.68	5.45	Csa	
6	Úbeda[a]	Jaén	37.94	3.30	360.5	131.41	2.74	BSh	
7	Estepona[a]	Málaga	36.44	5.21	201.5	4.55	44.29	Csa	
8	La Rinconada[a]	Sevilla	37.46	5.92	39.5	76.87	0.51	Csa	
9	Adra[a]	Almería	36.75	2.99	44.5	3.22	13.82	BSh	
10	Almería[a]		36.84	2.40	24.5	1.17	20.94	BSh	
11	Cuevas de Almanzora[a]		37.26	1.80	22.5	2.93	7.68	BSh	
12	Fiñana[a]		37.16	2.84	973.5	45.93	21.20	BSk	
13	Huércal-Overa[a]		37.41	1.88	319.5	18.72	17.07	BSk	
14	La Mojonera[a]		36.79	2.70	144.5	8.96	16.13	BSh	
15	Níjar[a]		36.95	2.16	184.5	16.14	11.43	BSh	
16	Tíjola[a]		37.38	2.46	742.5	60.28	12.32	BSk	
17	Virgen de Fátima[a]		37.39	1.77	187.5	7.80	24.03	BSk	
18	Almería Airport-AEMET[b]		36.85	2.36	23.5	0.94	25.03	BSh	
19	Basurta-Jerez de la Frontera[a]	Cádiz	36.76	6.02	62.5	27.54	2.27	Csa	
20	Conil de la Frontera[a]		36.33	6.13	26.5	2.71	9.80	Csa	
21	Jerez de la Frontera[a]		36.64	6.01	34.5	20.67	1.67	Csa	
22	Jimena de la Frontera[a]		36.41	5.38	55.5	14.72	3.77	Csa	
23	Vejer de la Frontera[a]		36.29	5.84	26.5	13.54	1.96	Csa	
24	Cádiz-AEMET Observatory[b]		36.50	6.26	4.5	0.03	140.63	Csa	
25	Cabra-IFAPA Centre[a]	Córdoba	37.50	4.43	549.5	88.44	6.21	Csa	
26	Hornachuelos[a]		37.72	5.16	159.5	130.81	1.22	Csa	
27	El Carpio[a]		37.91	4.50	167.5	134.73	1.24	Csa	
28	Córdoba[a]		37.86	4.80	119.5	132.90	0.90	Csa	
29	Bélmez[a]		38.25	5.21	525.5	184.94	2.84	Csa	
30	Baena[a]		37.69	4.31	336.5	106.78	3.15	Csa	
31	Adamuz[a]		38.00	4.45	92.5	145.63	0.64	Csa	
32	Baza[a]	Granada	37.56	2.77	816.5	84.55	9.66	BSk	
33	Iznalloz[a]		37.42	3.55	937.5	75.07	12.49	Csa	
34	Jerez del Marquesado[a]		37.19	3.15	1214.5	49.20	24.69	Csa	
35	Loja[a]		37.17	4.14	489.5	46.70	10.48	Csa	
36	Puebla de Don Fadrique[a]		37.88	2.38	1112.5	86.08	12.92	BSk	
37	Zafarraya[a]		36.99	4.15	907.5	27.84	32.59	Csa	
38	Granada Air Base-AEMET[b]		37.14	3.63	689.5	44.27	15.57	Csa	
39	Almonte[a]	Huelva	37.15	6.48	20.5	17.67	1.16	Csa	
40	Aroche[a]		37.96	6.95	301.5	84.14	3.58	Csa	
41	El Campillo[a]		37.66	6.60	408.5	61.98	6.59	Csa	
42	Gibraleón[a]		37.41	7.06	171.5	21.78	7.87	Csa	
43	La Puebla de Guzmán[a]		37.55	7.25	290.5	38.54	7.54	Csa	
44	Lepe[a]		37.30	7.24	76.5	10.67	7.17	Csa	
45	Moguer[a]		37.15	6.79	89.5	3.11	28.77	Csa	
46	Niebla[a]		37.35	6.74	54.5	25.08	2.17	Csa	
47	Huelva-AEMET[b]		37.28	6.91	21.5	3.55	6.05	Csa	
48	Mancha Real[a]		37.92	6.27	438.5	131.56	3.33	Csa	
49	Chiclana de Segura[a]	Jaén	38.30	2.00	577.5	101.39	5.70	Csa	
50	Huesa[a]		37.75	3.06	782.5	113.87	6.87	Csa	
51	Linares[a]		38.09	3.57	284.5	149.86	1.90	Csa	
52	Pozo Alcón[a]		37.67	2.93	895.5	102.02	8.78	Csa	
53	Sabiote[a]		38.08	3.24	824.5	147.08	5.61	Csa	
54	San Jose de los Propios[a]		37.86	3.23	511.5	123.05	4.16	Csa	
55	Torreblascopedro[a]		37.99	3.69	293.5	138.08	2.13	BSh	
56	Málaga[a]	Málaga	36.76	4.54	70.5	11.56	6.10	Csa	
57	Antequera[a]		37.06	4.56	459.5	150.42	3.05	Csa	
58	Sierra Yeguas[a]		37.14	4.84	466.5	60.43	7.72	Csa	
59	Vélez-Málaga[a]		36.80	4.13	51.5	7.42	6.94	Csa	
60	Málaga CMT-AEMET[b]		36.67	4.48	23.5	1.95	12.06	Csa	
61	Aznalcazar[a]	Sevilla	37.15	6.27	6.5	29.37	0.22	Csa	
62	Écija[a]		37.59	5.08	127.5	144.11	0.88	Csa	
63	La Luisiana[a]		37.53	5.23	190.5	129.47	1.47	Csa	
64	La Puebla del Río I[a]		37.23	6.13	27.5	45.11	0.61	Csa	
65	La Puebla del Río II[a]		37.08	6.05	43.5	41.08	1.06	Csa	
66	Las Cabezas de San Juan[a]		37.02	5.88	27.5	51.63	0.53	Csa	
67	Lora del Río[a]		37.66	5.54	70.5	113.51	0.62	Csa	
68	Osuna[a]		37.26	5.13	216.5	119.98	1.80	Csa	
69	Puebla Cazalla[a]		37.22	5.35	231.5	102.73	2.25	Csa	
70	Sanlúcar La Mayor[a]		37.42	6.26	90.5	52.32	1.73	Csa	
71	Sevilla San Pablo-AEMET[b]		37.42	5.88	36.5	74.19	0.49	Csa	

[a] Environmental Climatology Information Network of the Andalusian Regional Government.

[b] Spanish State Meteorological Agency.

Table A2

Geographical and climatic characteristic of weather stations in central Spain.

Station No.	Location	Province	φ (°)	λ (°)	z (m)	L (km)	z/L (m/km)	KGCC	Period
72	Valladolid[a]	Valladolid	41.641	−4.754	738.0	195.30	3.78	Csa	2004–2017
73	Toledo[a]	Toledo	39.885	−4.045	518.0	322.07	1.61	BSk	
74	Ciudad Real[a]	Ciudad Real	38.989	−3.920	631.0	248.19	2.54	BSk	
75	Madrid[a]	Madrid	40.452	−3.724	667.0	308.79	2.16	BSk	
76	Soria[a]	Soria	41.775	−2.483	1085.0	169.56	6.40	Cfb	
77	Albacete[a]	Albacete	39.006	−1.862	679.0	139.40	4.87	BSk	
78	Valencia[a]	Valencia	39.485	−0.475	59.0	12.63	4.67	BSk	
79	Tortosa[a]	Tarragona	40.820	0.493	53.0	17.53	3.02	Csa	
80	Cáceres[a]	Cáceres	39.471	−6.339	408.0	229.50	1.78	Csa	
81	León[a]	León	42.588	−5.641	914.0	104.8	8.72	Csb	1998–2016
82	Salamanca[a]	Salamanca	40.959	−5.498	793.0	266.69	2.97	Csb	2004–2017
83	Lleida[a]	Lleida	41.626	0.598	188.0	72.68	2.59	BSk	
84	Tarragona[a]	Tarragona	41.145	1.164	74.0	6.43	11.51	Csa	

[a] Spanish State Meteorological Agency.

Table A3

Geographical and climatic characteristic of weather stations in northern Spain.

Station No.	Location	Province	φ (°)	λ (°)	z (m)	L (km)	z/L (m/km)	KGCC	Period
85	Avilés[a]	Asturias	43.584	−5.918	12.0	1.59	7.56	Cfb	2003–2016
86	Oviedo[b]		43.354	−5.873	350.0	25.72	13.61	Cfb	
87	Oviedo[a]		43.371	−5.836	189.2	22.79	8.30	Cfb	
88	Mieres[a]		43.258	−5.773	206.0	32.75	6.29	Cfb	
89	Langreo[a]		43.309	−5.706	247.0	26.32	9.38	Cfb	
90	Gijón[a]		43.531	−5.672	32.0	1.43	22.34	Cfb	
91	Niembro[c]		43.439	−4.850	136.0	0.36	376.65	Cfb	
92	Santander-CMT[b]	Cantabria	43.491	−3.801	60.0	0.18	333.33	Cfb	
93	A Coruña[b]	Coruña	43.366	−8.421	60.0	0.70	85.71	Csb	1985–2016
94	Santiago[b]		42.888	−8.411	372.0	40.90	9.10	Cfb	
95	A Coruña-Airport[b]		43.304	−8.378	100.0	4.07	24.59	Csb	2004–2016
96	Lugo[b]	Lugo	43.115	−7.456	446.0	50.82	8.78	Csb	1985–1989
97	Oviedo[b]	Asturias	43.354	−5.873	350.0	25.72	13.61	Cfb	1975–2016
98	Gijón[d]		43.545	−5.693	12.5	0.29	42.85	Cfb	1993–2005
99	Santander-Centre[b]	Cantabria	43.491	−3.819	72.0	1.13	63.52	Cfb	1989–1997
100	Bilbao[b]	Vizcaya	43.298	−2.906	44.0	9.50	4.63	Cfb	1985–2016
101	Vitoria (II)[b]	Alava	42.882	−2.735	513.0	54.20	9.46	Cfb	2000–2008
102	Vitoria(I)[b]		42.884	−2.723	508.0	55.30	9.19	Cfb	2011–2016
103	San Sebastián[b]	Guipuzcoa	43.306	−2.041	263.0	1.10	239.09	Cfb	1983–2016
104	Pamplona[b]	Navarra	42.777	−1.650	461.0	66.00	6.98	Cfb	2003–2008
105	Girona[b]	Girona	41.912	2.763	145.0	24.70	5.87	Csa	1998–2016

[a] Service of Environmental Information of the Principality of Asturias.
[b] Spanish State Meteorological Agency.
[c] Spanish Ministry of Environmental Issues.
[d] City Council of Gijón.

References

[1] IPCC. Global Warming of 1.5°C. An IPCC Special Report on the Impacts of Global Warming of 1.5°C above Pre-Industrial Levels and Related Global Greenhouse Gas Emission Pathways, in the Context of Strengthening the Global Response to the Threat of Climate Change, Sustainable Development, and Efforts to Eradicate Poverty; Masson-Delmotte V, Zhai P, Pörtner H-O, Roberts D, Skea J, Shukla PR, Pirani A, Moufouma-Okia W, Péan C, Pidcock R, et al., Eds.; IPCC: Geneva, Switzerland, 2018.

[2] VijayaVenkataRaman S, Iniyan S, Goic R. A review of solar drying technologies. Renew Sustain Energy Rev 2012;16(5):2652–70.

[3] Tiwari GN, Mishra RK, Solanki SC. Photovoltaic modules and their applications: A review on thermal modelling. Appl Energy 2011;88(7):2287–304.

[4] Parida B, Iniyan S, Goic R. A review of solar photovoltaic technologies. Renew Sustain Energy Rev 2011;15(3):1625–36.

[5] Cao X, Dai X, Liu J. Building energy-consumption status worldwide and the state-of-the-art technologies for zero-energy buildings during the past decade. Energy Build 2016;128:198–1113.

[6] Thornton PE, Running SW. An improved algorithm for estimating incident daily solar radiation from measurements of temperature, humidity, and precipitation. Agric Forest Meteorol 1999;93(4):211–28.

[7] Zang H, Xu Q, Bian H. Generation of typical solar radiation data for different climates of China. Energy 2012;38(1):236–48.

[8] Qin W, Wang L, Lin A, Zhang M, Xia X, Hu B, et al. Comparison of deterministic and data-driven models for solar radiation estimation in China. Renew Sustain Energy Rev 2018;81:579–94.

[9] Ishola KA, Fealy R, Mills G, Fealy R, Green S, Jimenez-Casteneda A, et al. Developing regional calibration coefficients for estimation of hourly global solar radiation in Ireland. Int J Sust Energy 2019;38(3):297–311.

[10] Yang D, Wang W, Gueymard ChA, Hong T, Kleissl J, Huang J, Pérez MJ, Pérez R, Bright JM, Xia X, Van der Meer D. A review of solar forecasting, its dependence on atmospheric sciences and implications for grid integration: towards carbon neutrality. Renew Sustain Energy Rev 2022;161:112348 https://doi.org/10.1016/j.rser.2022.112348.

[11] Antoñanzas F, Sanz A, Martínez-de-Pisón FJ, Perpiñán O. Evaluation and improvement of empirical models of global solar irradiation: Case study northern Spain. Renew Energy 2013;60:604–14.

[12] Chen J-L, He L, Yang H, Ma M, Chen Q, Wu S-J, et al. Empirical models for estimating monthly global solar radiation: A most comprehensive review and comparative case study in China. Renew Sustain Energy Rev 2019;108:91–111.

[13] Abraha MG, Savage MJ. Comparison of estimates of daily solar radiation from air temperature range for application in crop simulations. Agr Forest Meteorol 2008;148(3):401–16.

[14] Besharat F, Dehghan AA, Faghih AR. Empirical models for estimating global solar radiation: a review and case study. Renew Sustain Energy Rev 2013;21:798–1721.

[15] Chen R, Kang E, Ji X, Yang J, Zhang Z. Trends of the global radiation and sunshine hours in 1961–1998 and their relationships in China. Energy Convers Manag 2006; 47(18-19):2859–66.

[16] Evrendilek F, Ertekin C. Assessing solar radiation models using multiple variables over Turkey. Clim Dyn 2008;31(2-3):131–49.

[17] Prieto JI, García D. Global solar radiation models: A critical review from the point of view of homogeneity and case study. Renew Sustain Energy Rev 2022;155: 111856. https://doi.org/10.1016/j.rser.2021.111856.

[18] Paulescu M, Stefu N, Calinoiu D, Paulescu E, Pop N, Boata R, et al. Renew Sustain Energy Rev 2016;62:495–6.

[19] Reddy SJ. The estimation of global solar radiation and evaporation through precipitation– A note. Sol Energy 1987;38(2):97–104.

[20] Korachagaon I, Bapat VN. General formula for the estimation of global solar radiation on earth's surface around the globe. Renew Energy 2012;41:394–1300.

[21] Annandale JG, Jovanovic NZ, Benadé N, Allen RG. Software for missing data error analysis of Penman-Monteith reference evapotranspiration. Irrig Sci 2002;21(2): 57–67.

[22] Prieto JI, Martínez-García JC, García D. Correlation between global solar irradiation and air temperature in Asturias. Spain Sol Energy 2009;83(7):1076–85.

[23] Duffie JA, Beckman WA. Solar engineering of thermal processes. New Jersey: Wiley; 2006.

[24] Hargreaves GH, Samani ZA. Estimating potential evapotranspiration. J Irrig Drain Eng ASCE 1982;108(3):225.

[25] Hargreaves GH. Simplified coefficients for estimating monthly solar radiation in North America and Europe. Logan, Utah: Dept. Paper, Dept. Biol. and Irrig. Engrg., Utah State University; 1994.

[26] Meza F, Varas E. Estimation of mean monthly solar global radiation as a function of temperature. Agric Forest Meteorol 2000;100(2-3):231–41.

[27] Bristow KL, Campbell GS. On the relationship between incoming solar radiation and daily maximum and minimum temperature. Agr Forest Meteorol 1984;31(2): 159–66.

[28] Weiss A, Hays CJ, Hu Qi, Easterling WE. Incorporating bias error in calculating solar irradiance: implications for crop yield simulations. Agron J 2001;93(6): 1321–6.

[29] Goodin DG, Hutchinson JMS, Vanderlip RL, Knapp MC. Estimating solar irradiance for crop modeling using daily air temperature data. Agron J 1999;91(5):845–51.

[30] Bandyopadhyay A, Bhadra A, Raghuwanshi NS, Singh R. Estimation of monthly solar radiation from measured air temperature extremes. Agric Forest Meteorol 2008;148(11):1707–18.

[31] Allen RG. Evaluation of Procedures for Estimating Mean Monthly Solar Radiation from Air Temperature, Technical Report, 1995, United Nations Food and Agric. Org. (FAO), Rome, Italy.

[32] Hargreaves GL, Hargreaves GH, Riley JP. Irrigation water requirements for Senegal River Basin. J Irrig Drain Eng 1985;111(3):265–75.

[33] Supit I, van Kappel RR. A simple method to estimate global radiation. Sol Energy 1998;63(3):147–60.

[34] Li M-F, Fan Li, Liu H-B, Guo P-T, Wu W. A general model for estimation of daily global solar radiation using air temperatures and site geographic parameters in Southwest China. J Atm Solar-Terr Phys 2013;92:145–50.

[35] Chen R, Ersi K, Yang J, Lu S, Zhao W. Validation of five global radiation models with measured daily data in China. Energy Convers Manag 2004;45(11–12): 1759–69.

[36] Panday CK, Katiyar AK. Temperature base correlation for the estimation of global solar radiation on horizontal surface. Int J Energy and Environment 2010;1(4): 737–1734.

[37] Adaramola MS. Estimating global solar radiation using common meteorological data in Akure. Nigeria Renew Energy 2012;47:38–44.

[38] Chen J-L, Li G-S. Estimation of monthly average daily solar radiation from measured meteorological data in Yangtze River Basin in China. Int J Climatology 2013;33(2):487–98.

[39] Li M-F, Liu H-B, Guo P-T, Wu W. Estimation of daily solar radiation from routinely observed meteorological data in Chongqing. China Energy Convers Manag 2010;51 (12):2575–9.

[40] Ohunakin OS, Adaramola MS, Oyewola OM, Fagbenle RO. Correlations for estimating solar radiation using sunshine hours and temperature measurement in Osogbo, Osun State. Nigeria Front Energy 2013;7(2):214–22.

[41] Hassan GE, Youssef ME, Mohamed ZE, Ali MA, Hanafy AA. New Temperature-based Models for Predicting Global Solar Radiation. Appl Energy 2016;179: 437–50.

[42] Chazarra A, Flórez E, Peraza B, Tohá T, Lorenzo B, Criado E, et al. Mapas climáticos de España (1981–2010) y ETo (1996–2016) (in Spanish). Madrid: Agencia Estatal de Meteorología; 2018. https://doi.org/10.31978/014-18-004-2.

[43] Ritter A, Muñoz-Carpena R. Performance evaluation of hydrological models: Statistical significance for reducing subjectivity in goodness-of-fit assessments. J Hydrol 2013;480:33–45.

[44] Gueymard CA. A review of validation methodologies and statistical performance indicators for modeled solar radiation data: Towards a better bankability of solar projects. Renew Sustain Energy Rev 2014;39:1024–34.

[45] Lewis C. International and business forecasting methods. London: Butterworths; 1982.

[46] Gupta HV, Kling H, Yilmaz KK, Martinez GF. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. J Hydrol 2009;377(1-2):80–91.

[47] Legates DR, McCabe GJ. Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation. Water Resour Res 1999;35(1): 233–41.

[48] Despotovic M, Nedic V, Despotovic D, Cvetanovic S. Evaluation of empirical models for predicting monthly mean horizontal diffuse solar radiation. Renew Sustain Energy Rev 2016;56:246–60.

[49] Taylor KE. Summarizing multiple aspects of model performance in a single diagram. J Geophys Res 2001;106(D7):7183–17182.

[50] Mirzabe AH, Hajiahmad A, Keyhani A, Mirzaei N. Approximation of daily solar radiation: A comprehensive review on employing of regression models. Renewable Energy Focus 2022;41:143–59. https://doi.org/10.1016/j.ref.2022.02.003.