Improving the calibration of building simulation with interpolated weather datasets

Pablo Eguía Oller (Phd)[a], José María Alonso Rodríguez (MSc)[a*], Ángeles Saavedra González (Phd)[b], Elena Arce Fariña (Phd)[c], Enrique Granada Álvarez (Phd)[a]

[a]E.T.S. Ingenieros Industriales, University of Vigo, Campus Lagoas-Marcosende, 36200, Vigo, Spain

[b]ETS Ingeniería de Minas, University of Vigo, Lagoas-Marcosende s/n, 36200, Vigo, Spain

[c]Defense University Center, Marin, Spain

*Corresponding author: José María Alonso Rodríguez

University of Vigo. School of Industrial Engineering. GTE Research Group.

Rúa Maxwell, s/n. Campus Universitario Lagoas-Marcosende. 36310 Vigo, Pontevedra, Spain

Email: josemaria.alonso@uvigo.es

How to cite: P. Eguía Oller, J.M. Alonso Rodríguez, Á. Saavedra González, E. Arce Fariña, E. Granada Álvarez, Improving the calibration of building simulation with interpolated weather datasets, Renewable Energy 122 (2018) 608–618. https://doi.org/10.1016/j.renene.2018.01.100.

Improving the calibration of building simulation with interpolated weather datasets.

Abstract

The building sector offers huge potential for energy savings, which helps to achieve environmental objectives and social benefits. A good approach to determine both the energy consumption of new buildings and the energetic refurbishment of existing buildings is through thermal simulation.

This paper studies how building energy simulation calibration can be improved using interpolated weather data to determine on-site meteorological parameters at the building location.

The lack of precise meteorological data in the exact location of buildings means that data from nearby stations is generally used, not knowing how far the error spreads in the results of heating demands and loads. The novelty of this paper lies in the analysis of error propagation to the results of demands and loads of thermal simulation, as well as in the method used to reduce these errors by TPS interpolation.

As an interesting conclusion, the average (CV(RMSE)) obtained in the simulation of the studied building, placed successively in each one of the 70 meteorological station locations, decreases from 74% when using the nearest neighborhood to each site to 26% using the TPS interpolation technique. The error in the building simulations is almost three times lower using the studied method.

Keywords: Building simulation; Weather data; Interpolation; Thin plate spline (TPS)

Nomenclature

AEMET	Spanish State Meteorological Agency
ASHRAE	American Society of Heating, Refrigerating and Air-Conditioning Engineers
CTE	Spanish Technical Building Code
CV(RMSE)	Coefficient of Variation of the Root-Mean-Square Error
EPW	EnergyPlus Weather Format
GIS	Geographic Information System
INE	National Statistical Institute of Spain
MAE	Mean Absolute Error
MeteoGalicia	Galician Meteorology Agency
NN	Nearest Neighbourhood
RMSE	Root mean square error
StID	weather station ID
TPS	Thin Plate Splines
TRNSYS	Transient System Simulation Tool

Introduction

Building energy consumption accounts for between 20% and 40% of global energy consumption in developed areas [1]. The housing sector represents approximately 40% of EU energy consumption [2]; more specifically, in Spain, the building sector energy consumption represents 33% of global energy consumption [3]. According to the Eurostat analysis, approximately 50% of overall energy consumption for a typical household in Spain corresponds to space heating or cooling [4]. Thus, in view of its importance, the building sector offers huge potential for energy savings.

Lowering building energy costs while increasing in-house comfort will help to achieve global environmental objectives and contribute to social well-being.

When designing an efficient building or analyzing an existing building, the energy demand and consumption must be evaluated to recommend measures for reducing energy consumption. A good method to evaluate building energy demands is the use of thermal simulations [5], [6]. In addition, most current building regulations require energy demand estimates that can only be obtained by thermal simulation [7].

There is a gap between simulated and real building energy performance [8]. This difference between the measured and simulated energy consumption of real buildings can be adjusted through model calibration [9]. ASHRAE's Guideline 14-2002 for Measurement of Energy and Demand Savings includes calibrated simulations as an evaluation method [10]; however, in this field, there is a deficiency in standards for building calibration [11].

Weather is a fundamental parameter when calibrating a simulation [12]. For example, M. Royapoor et al. concluded that, to have a calibrated model, on-site weather parameters should be used [13]. Another example is the study of H. Yoo et al., who analyzed different weather parameters and their influence on heating and cooling loads [14]. G. Mustafaraj et al. improved the accuracy of model calibration, showing the need for the creation of an hourly year weather data file [15].

Six weather variables were interpolated at 70 meteorological station locations to generate the annual meteorological archives and simulate a representative building in Galicia using the Transient System Simulation Tool TRNSYS.

When simulations are performed at locations without weather recorded data, two techniques of providing data are primarily used: interpolation and synthetic data generation [16], [17]. The purpose of using interpolation functions is to obtain intermediate values consistent with those obtained at nearby locations [18]. Global techniques are characterized by using all the measured data to make predictions. In other words, a single function is applied to the entire study area. This approach usually results in a smooth predicted surface, which may be reasonable as long as the studied surface is known to have a global trend. Local interpolation techniques are usually based on the same methodology but are applied on a region of the total sample set. From this point of view, global techniques can be considered as a simplification of local techniques because the latter result in interpolated surfaces that are more flexible and better adapted to the local characteristics of the sampled data. Deterministic interpolation techniques calculate a continuous surface by using the geometric characteristics of the measurements. We can divide these methods into two groups: global and local techniques. Local techniques, such as nearest neighborhood (NN) and thin plate splines (TPS), calculate the interpolated values from the measured points included in the neighborhoods, which are smaller spatial areas within the global study area. In previous research [19], [20], we have evaluated different types of interpolations and determined that TPS is a good interpolation method when using these data in thermal simulations for single-family residential buildings; in this article, to generalize the methodology, we will develop thermal simulations for a building with eight above-ground stories over an area six-times larger and compare the results using the nearest meteorological station with the values obtained from interpolation by the TPS method. NN is currently the typical method used to establish the climatic conditions of the thermal simulations.

The chosen method of comparison is through maps of interpolated values represented by GIS software (QGIS [21]) because, when dealing with 70 meteorological stations, tables of values are of limited use.

The average CV(RMSE) error obtained in the simulations of the studied building, placed successively at each one of the 70 meteorological station locations, decreases from 0.74 using the closest station (NN) at each site to 0.26 using the TPS interpolation technique. The error in the building simulations is almost three-times lower using the TPS method.

Material and Methods

The experiment developed in this paper consists of simulating a typical building in the 70 locations with weather stations available, using data from the meteorological station in situ, data from the nearest meteorological station, and data obtained through TPS interpolation.

We assume that the simulation results with the station data at the location of the building are correct and compare the errors committed with the other two methods: the nearest weather station and the values obtained in the building location by interpolation with the TPS method.

A representative construction was chosen given a common geometric building structure in Galicia (northwest Spain) based on the National Statistical Institute of Spain (INE) database.

Figure 1 shows a 3D model of the building created using the TRNSYS 3D-Plugin for Trimble SketchUp. The selected building has eight above-ground stories with a floor area of 1689.4 m^2 and a total volume of 5053.2 m^3. Each story was characterized as a thermal zone in the model. The main facade is oriented towards the east. The impacts of the shadows of adjacent buildings were also taken into account (see TRNSYS 3D model in Figure 1).

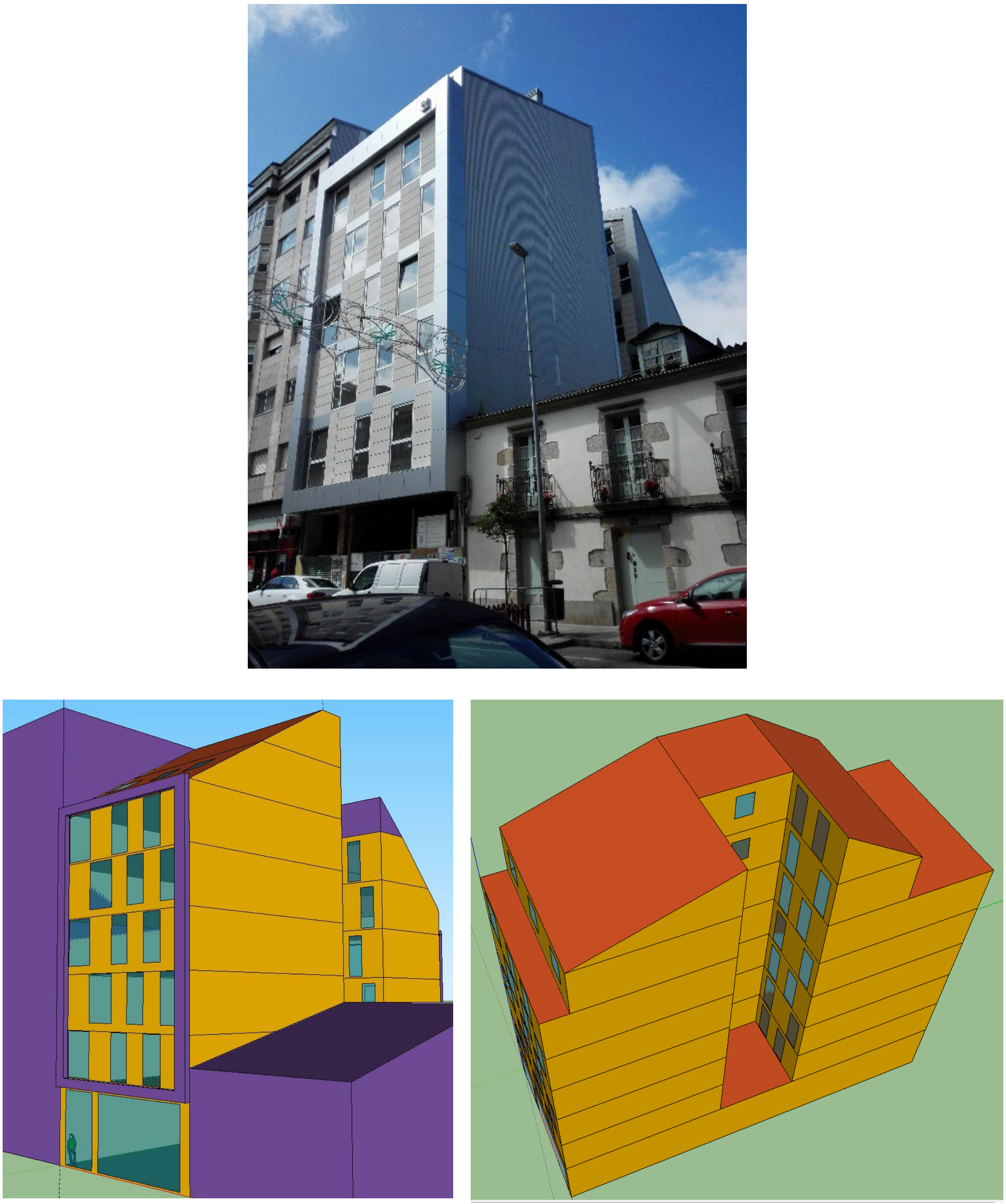


Figure 1. Photograph and TRNSYS 3D model of the building.

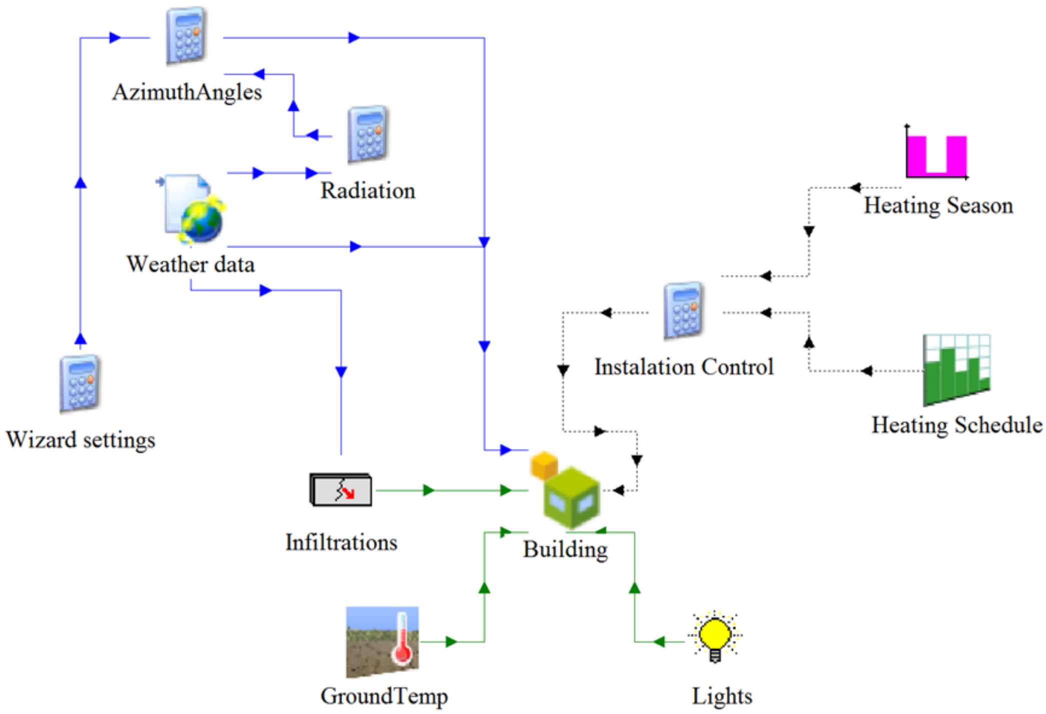


Figure 2. Simulation studio panel from TRNSYS.

Figure 2 illustrates the model in the TRNSYS simulation environment. Sensible heating demands were evaluated assuming unlimited thermal resources for the building. The set point temperatures, internal gains (e.g., lighting) and building air infiltration values correspond to the Spanish Technical Building Code (CTE) specifications. The Spanish Technical Building Code, promulgated by Royal Decree 314/2006, 17 March 2006, is the regulatory framework governing the basic quality requirements for buildings in Spain, including facilities. Based on the occupancy profiles of the Spanish Institute for Energy Diversification and Saving (IDAE) for this type of dwelling, the temperature controls were fixed to 20°C from 3 pm to 11 pm. The heating season regime was set from the October's last Sunday to April 1st. Based on the Spanish procedures to calculate building energy performance, the climate zones for the analyzed area do not require calculation of the cooling; thus, only heating demand was assessed. For more information about the simulation TRNSYS model, refer to [20].

Table 1 and Table 2 show the TRNSYS wall and window type manager definition entries, respectively.

Wall category	Wall type	Thickness. [mm]	Transmittance. [W/m^2 K]
External	Roof	400	0.500
	Facade	201	0.748
Adjacent	Ceiling	384	0.940
	Wall	90	0.508
Boundary	Basement wall	200	3.892
	Ground	400	2.314

Table 1. Building envelope properties

	Values	Characteristics
Glazing	4/16/4	Pilkington Optitherm.
	0.586	G-value.
	1.06	Transmittance [W/m^2 K].
Frame	0.3%	Frame window area.
	3.03	Transmittance [W/m^2 K].
	0.6	Solar absorptance.
	0.9	Emissivity.

Table 2. Window properties

Studied area and data sources

The analysis was conducted for the Galician Meteorology Agency (MeteoGalicia) net, using seventy different meteorological stations, in the region of Galicia (Northwest Spain). Galicia has an area of 29,574.4 km^2 (11,418.7 sq mi) with a population of 2.720.668 people (January 2016) (source: INE, Spanish Statistical Office). Galicia has a maritime influenced climate. This region has climatic characteristics analogous to those prevalent throughout Western Europe: cool summers and mild winters, with the possibility of rainfall along the whole year. In general terms, according the Köppen-Geiger climate classification, the Galician climate is considered to be Csb. However, there are different microclimates due to its complicated relief, that is, the climate of Galicia is characterized by a progression between a predominance of pure oceanic climate with climatic zones that can be considered as subtropical. Rainfall ranges between 594 mm and 3143 mm, based on MeteoGalicia records. The spatial distribution of temperatures presents a variation between the coast and the interior, related to the presence of the Atlantic Ocean. According to the CTE, the studied area is comprised of 6 climatic zones (i.e., C1, C2, C3, D1, D2 and E1), which are in relation with the climatic severity. The C climate zone has a less severe climate than those of D or E, and the D climatic zone is less severe than that of E. Climatic zones are indicated in the HE1 Section of the CTE Energy Savings Document. In our case study, the calculation table is reduced to the 4 province capitals in Galicia: Coruña, Lugo, Orense, and Pontevedra. Table 3 allows the CZ of any location in Galicia to be obtained from its altitude.

Province Capital	C.Z.	Elevation (e) [m]	C3	C2	C1	D2	D1	E1
Coruña	C1	0			e<200		e≥200	
Lugo	D1	412					e<500	e≥500
Orense	D2	327	e<150	e<300		e<800		e≥800
Pontevedra	C1	77			e<350		e≥350	

Table 2. Climatic zone (CZ) calculation correspondence for locations in the four provinces of Galicia

StID	LONGITUDE	LATITUDE	ELEVATION	Climatic zone
10045	-8.26221896	43.2413695	94	C1
10046	-7.8943978	43.3430821	651	D1
10047	-7.08301941	43.540711	51	D1
10048	-7.3446339	42.1195584	1026	E1
10049	-9.02859829	42.5551672	30	C1
10050	-8.25226804	43.491464	37	C1
10052	-8.77630574	42.7455814	661	D1
10053	-7.5444682	42.9948754	400	D1
10055	-7.78307165	43.2265962	684	E1
10056	-7.50178093	42.4731965	645	E1
10058	-7.39877783	41.9743289	546	D2
10060	-8.67944536	42.0786	484	D1
10061	-8.13729095	42.6143856	500	D1
10062	-6.92289994	42.8207736	1364	E1
10063	-8.42852879	42.2253186	371	D1
10064	-8.66421685	42.4092445	57	C1
10067	-8.70430461	42.4592563	424	D1
10085	-8.80470645	42.5800761	3	C1
10086	-8.40021838	42.3152979	705	D1
10087	-8.86830023	42.9686119	369	D1
10088	-7.44625149	43.4545921	595	E1
10089	-7.98265101	42.9073811	477	D1
10091	-8.86619845	41.9944708	473	D1
10092	-8.05246448	43.7042593	254	D1
10096	-8.69099082	43.095195	540	D1
10097	-7.78878175	43.5917779	576	D1
10099	-7.47627801	42.6476099	432	D1
10102	-7.19238705	42.5946818	777	E1
10109	-8.23823651	42.41459	553	D2
10110	-7.7083803	41.9462219	807	E1
10112	-7.96814974	42.1736211	623	D2
10115	-7.00863716	42.3553978	1229	E1
10116	-7.08961183	42.1630374	851	E1
10117	-7.29715492	42.3938725	1026	E1
10118	-7.37481102	43.6448512	421	D1
10119	-7.97190638	41.9036087	1059	E1
10122	-7.93307372	42.5797597	991	D1
10124	-8.55943646	42.875961	255	D1
10126	-8.93618636	42.3820883	121	C1
10129	-8.79879168	42.4030844	34	C1
10130	-6.78345593	42.3751641	1620	E1
10132	-7.04707469	42.7071022	1310	E1
10136	-7.04785524	43.1771242	789	E1
10137	-6.91530332	42.9559943	910	E1
10138	-6.89255989	42.2078109	1762	E1
10141	-8.1328198	43.5625001	278	D1
10144	-8.18961314	42.926242	362	D1
10154	-8.60187382	42.3206964	260	C1
10155	-7.877746	42.35299	139	C3
10161	-8.68605504	42.1699246	460	D1
10162	-7.63087786	43.6308822	59	D1
10800	-9.17831255	43.1244507	5	C1
14001	-8.72758817	42.2416662	7	C1
14003	-8.53118151	43.3472312	5	C1
19003	-8.43471917	42.7847757	225	D1
19010	-8.83688392	42.8014087	157	C1
19012	-8.85457036	42.6746637	59	C1
19014	-7.34011628	43.456918	115	D1
19018	-7.19645044	43.0359061	917	E1
19020	-7.38156359	42.7966782	416	D1
19021	-7.71832325	42.6144356	391	D1
19031	-7.83647734	42.4295336	403	D2
19033	-7.43197512	42.3875651	469	D2
19040	-7.18801093	41.8981999	778	D2
19041	-7.04890488	42.0174566	1025	E1
19045	-8.52493187	42.1804032	41	C1
19047	-8.79153818	41.9382123	52	C1
19050	-8.3186965	42.7791102	211	C1
19056	-8.71611497	42.537132	52	C1
50500	-8.52121113	42.88662974	305	D1

Table 3. Weather station network characteristics

Based on the ASHRAE handbook [10], six meteorological variables were analyzed: wind speed and wind direction, global radiation, pressure, temperature and relative humidity. Although the ASHRAE recommends the use of these six variables, it has been shown through different methodologies that weather files may be obtained from monthly values of maximum and average temperatures, average wind speed and daily insolation [22]. Obviously, the use of simplified weather data should lead to poor simulation results [23]. In this case, the wind direction and the pressure were used because an empirical method described in the ASHRAE handbook was applied to estimate infiltrations [20]. The Galician MeteoGalicia agency network comprises of 155 weather stations with a density of 0.005 station/km^2; however, because of a lack of measured variables or data gaps, only 70 (Figure 3) stations were considered. The weather station network characteristics are given in Table 4. The weather station ID (StID) is the same as that used by MeteoGalicia.

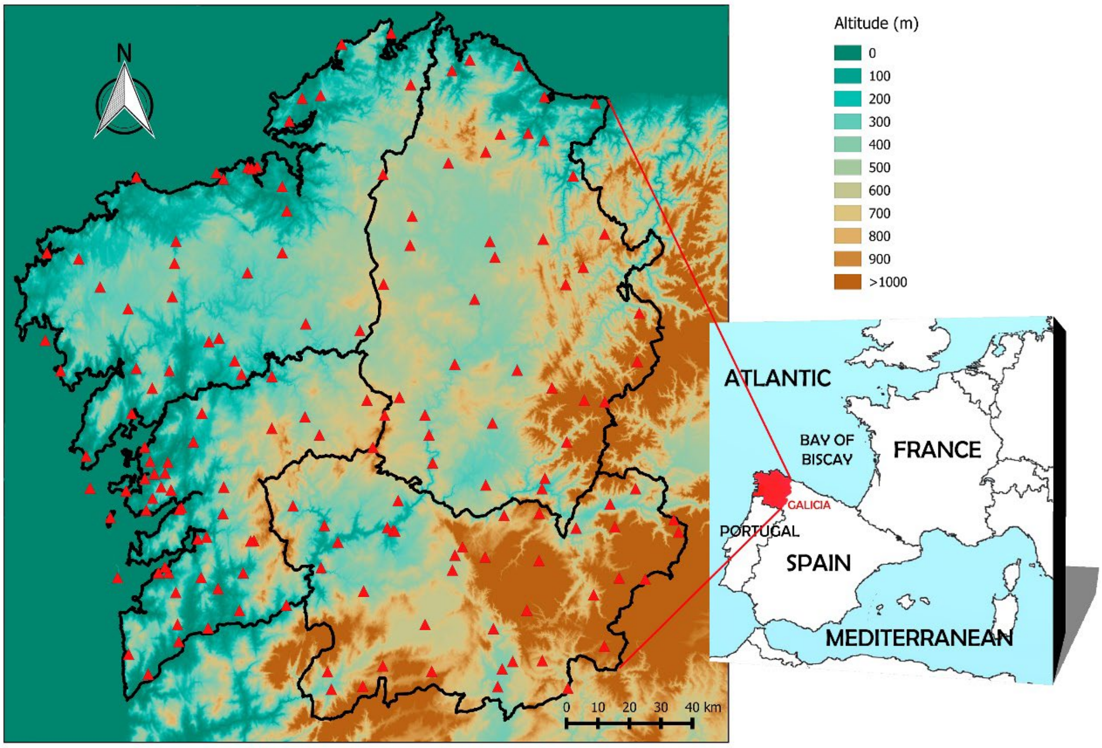


Figure 3. Weather station network placement.

The stations of the MeteoGalicia network are not placed following a spatial pattern; in addition, there are data gaps in the measured meteorological variables. These two aspects imply problems for the interpolation analysis. Thus, the data gaps were filled using TPS data estimates. The data gaps were not considered when calculating the errors. As expected, the absence of data for a single weather variable (i.e., there is no data during one or more hours) means that we cannot generate an input weather data file during this period.

We used EnergyPlus software to generate the EPW meteorological data files necessary to develop the TRNSYS simulations. Each EPW data file is a text file with 8760 rows in our case (hours in a year), with data at ordered time intervals.

The simulation time step was fixed at one hour.

Introduction of interpolation techniques

The local interpolation techniques tested in this paper are nearest neighborhood (NN) and thin plate splines (TPS). NN is the easiest and least time-consuming method to approximate missing measurements. NN is based on calculating weighted averages of values from a set of nearby locations. The simplest version of this technique considers only one location in the neighborhood; thus, it selects the value of the nearest point and produces a piecewise constant surface as a result. One of the most important problems of weighted-average methods such as NN is that the interpolated value will never be higher or lower than those used to work out the interpolation.

Thin plate splines (TPS) method is a local interpolation method whose effectiveness has been proven to obtain accurate surface weather datasets. TPS assumes that the variable in a given location x, say X(x), is a sum of two theoretical functions, X(x)=s(x)+ ε(x), where s(x) denotes the signal function and ε(x) for the noise.

The signal, s(x), is estimated by a function, in our case f(x); this function minimizes the squared error for the observations and is subject to the restriction that this function f(x) has a certain level of smoothness, that is:

$$\sum_{i=1}^{n}[X(x_i) - f(x_i)]^2 + \lambda \int [f''(x)]^2 dx,$$

where n is, in this case, the number of measurements in the neighborhood, and the smoothing parameter determines the estimator degree of smoothness.

All the interpolations were conducted using the R environment [24]. Specifically, to perform the TPS analysis, the library "fields" [25] was implemented.

Error Calculation

A collection of skill scores was established to evaluate the interpolation techniques (Table 5) [26], [27]. The scores were chosen to evaluate the difference between real and estimated values, both in terms of the locations and in terms of thermal simulation results. The scores were calculated hourly, thus allowing for a better assessment of the time series.

Mean absolute error (MAE) is a measure of average error that shows the errors in the same unit as the weather variable itself [26]. MAE ranges from 0 (best case) to an unbounded value. Root mean square error (RMSE) is used to reflect the variability distribution of the errors. This skill score has a high sensitivity to large outliers [28]. The coefficient of variation of the mean square error (CV(RMSE)) is the root mean square error normalized to the observed input values range.

Skill score	Equation[a]
Mean absolute error (MAE)	$MAE = \sum_{i=1}^{n} \frac{\lvert X_i - Y_i \rvert}{N}$
Root mean square error (RMSE)	$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(X_i - Y_i)^2}{N}}$
Mean square error coefficient of variation (CV(RMSE))	$CV(RMSE) = \frac{RMSE}{\frac{\sum_{i=1}^{n} Y_i}{N}}$

[a]Explanation of the variables:
N: sample size, 8,760 for hourly counts.
$(X - Y)$ = Difference between the observed (X) and actual values (Y)

Table 4. Definition of the skill scores

Results and discussion

The results obtained by running the interpolation at the 70 locations in 2015 during the 8760 hours of a year are analyzed in this section.

Weather variable estimation

The relative skill of each climatic zone and weather variable, based on TPS interpolation results, is summarized in Table 6. As scores for each variable have different scales, the average ranks were calculated to produce an overall rating. On this basis, C1 emerges as the best climatic zone for almost all weather variables except wind direction, where D1 performs better. C3 has been excluded because only one station is located in this climatic zone. These results indicate the influence of altitude in the interpolation estimates. Although C1 is the climatic zone with the best estimation results, the differences between the climatic zones are relatively small. Nevertheless, the domain-wide statistics may hide local variations in weather variables that average out when considering all stations together. Next, the interpolation results for each location and weather variable are analyzed separately.

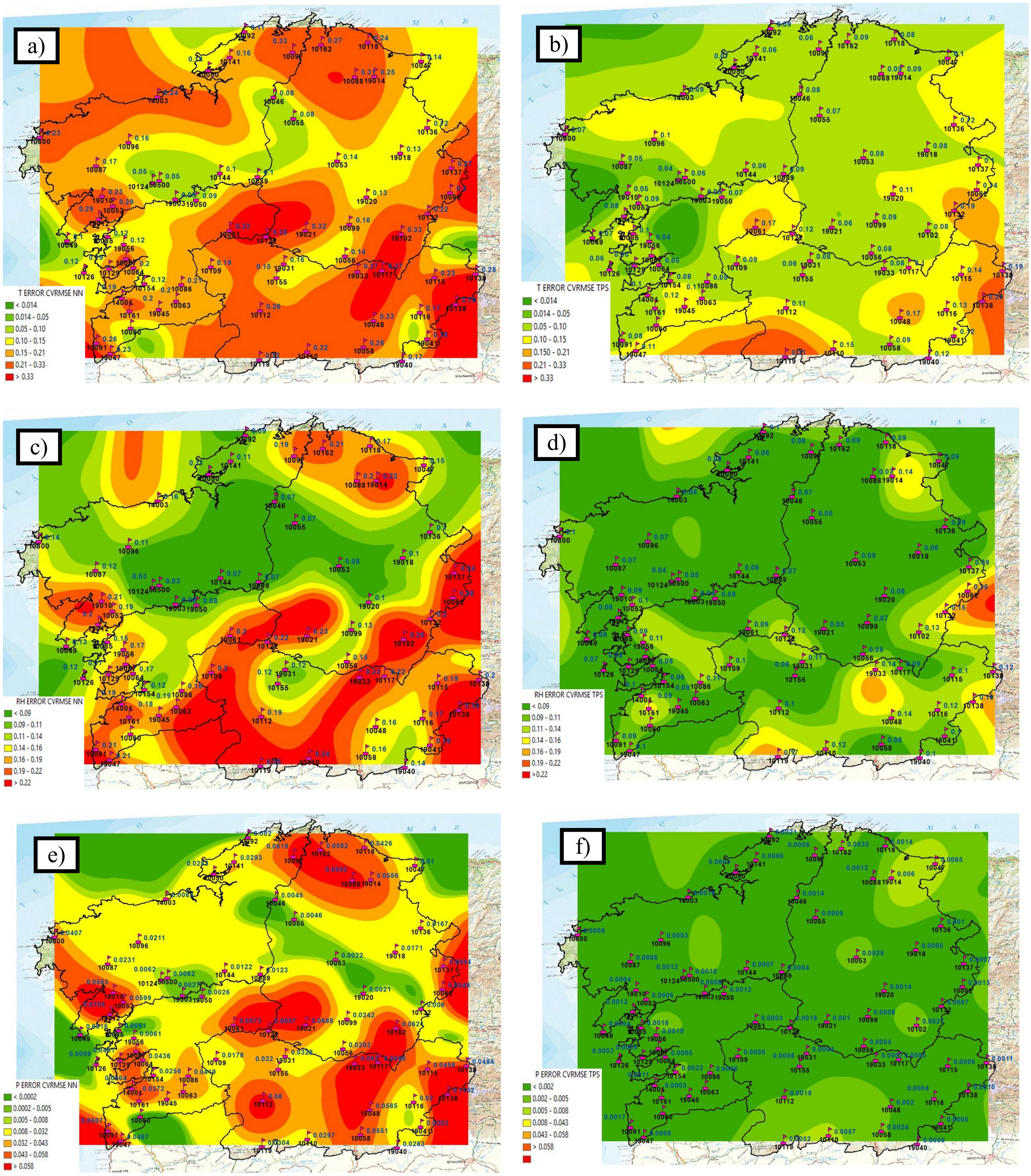
	Temperature				Relative-Humidity				Pressure				Global-Radiation				Wind-direction				Wind-speed			
Climatic zone	MAE	#	RMSE	#	MAE	#	RMSE	#	MAE	#	RMSE	#	MAE	#	RMSE	#	MAE	#	RMSE	#	MAE	#	RMSE	#
C1	0.84	3	1.16	3	5.23	3	6.74	2	82.89	5	102.71	2	25.25	3	49.70	1	69.46	2	106.71	2	1.14	2	1.84	4
C3	0.78	2	0.99	1	4.16	1	5.40	1	72.08	3	78.71	1	22.54	1	57.71	3	112.88	5	148.69	5	2.52	5	0.74	1
D1	0.75	1	1.08	2	4.87	2	6.75	3	81.13	4	148.15	5	28.42	5	57.48	2	64.50	1	99.11	1	1.73	4	1.75	3
D2	0.93	4	1.27	4	5.73	4	7.50	4	55.11	1	107.14	3	26.83	4	60.56	4	99.33	4	139.15	4	1.25	3	1.10	2
E1	1.03	5	1.43	5	6.10	5	8.44	5	71.10	2	136.40	4	23.55	2	67.15	5	75.32	3	112.89	3	1.07	1	2.25	5

Table 5. Skill scores based on the TPS interpolation results

First, we evaluate whether there is any spatial pattern in the relative skill of the interpolation methods. Figure 4 shows, for each technique (i.e., NN and TPS), the analysis of the CV(RMSE) distribution for the six studied weather variables. For ease of comparison, the color-scale used in the maps is the same. As shown in Figure 4, the NN method demonstrates poor performance in general, both for locations and variables. The interpolation techniques perform better with a greater density of points. When there is a low density of points, the interpolation fails in reproducing events in which there is a large variation of weather variables. Although there is a high quantity of stations by the sea, the interpolation methods indicate problems with these border points, especially for variables with the wind, which is affected by the sea. Altitude is another parameter that influences the interpolation results (see map, central zone). For both methods, the average skill is higher for each grid in the northern part of the map. A key reason for this is the topographic complexity of the south region, which makes interpolation more difficult.

Second, the interpolation methods perform better for temperature, relative humidity, and pressure than for global radiation, wind speed and wind direction (Table 7). Based on the CV(RMSE) values, it is clear that TPS better estimates all weather variables. However, when the density is low and the next station is close, NN is better for estimations than TPS. Because the interpolated surface is a local function of the neighboring data, TPS is appropriate for use across large heterogeneous areas [29]. Although in absolute terms, the atmospheric pressure errors are higher than the other weather variable errors based on the CV(RMSE) results in both techniques and show more homogeneous performance.

As we can see in the next figures, there are important improvements when using the TPS method versus the nearest neighborhood (NN) method. The mean CV(RMSE) error value at the 70 sites decreases from 0.21 (NN) to 0.10 (TPS) for temperature, from 0.16 (NN) to 0.09 (TPS) for relative humidity, from 0.034 (NN) to 0.001 (TPS) for pressure, from 0.44 (NN) to 0.37 (TPS) for global radiation, from 0.66 (NN) to 0.62 (TPS) for wind direction and from 1.14 (NN) to 0.63 (TPS) for wind speed.


a)
T ERROR CVRMSE NN
< 0.014
0.014 - 0.05
0.05 - 0.10
0.10 - 0.15
0.15 - 0.21
0.21 - 0.33
> 0.33
b)
T ERROR CVRMSE TPS
< 0.014
0.014 - 0.05
0.05 - 0.10
0.10 - 0.15
0.150 - 0.21
0.21 - 0.33
> 0.33
c)
RH ERROR CVRMSE NN
< 0.09
0.09 - 0.11
0.11 - 0.14
0.14 - 0.16
0.16 - 0.19
0.19 - 0.22
> 0.22
d)
RH ERROR CVRMSE TPS
< 0.09
0.09 - 0.11
0.11 - 0.14
0.14 - 0.16
0.16 - 0.19
0.19 - 0.22
>0.22
e)
P ERROR CVRMSE NN
< 0.0002
0.0002 - 0.005
0.005 - 0.008
0.008 - 0.032
0.032 - 0.043
0.043 - 0.058
> 0.058
f)
P ERROR CVRMSE TPS
< 0.002
0.002 - 0.005
0.005 - 0.008
0.008 - 0.043
0.043 - 0.058
> 0.058

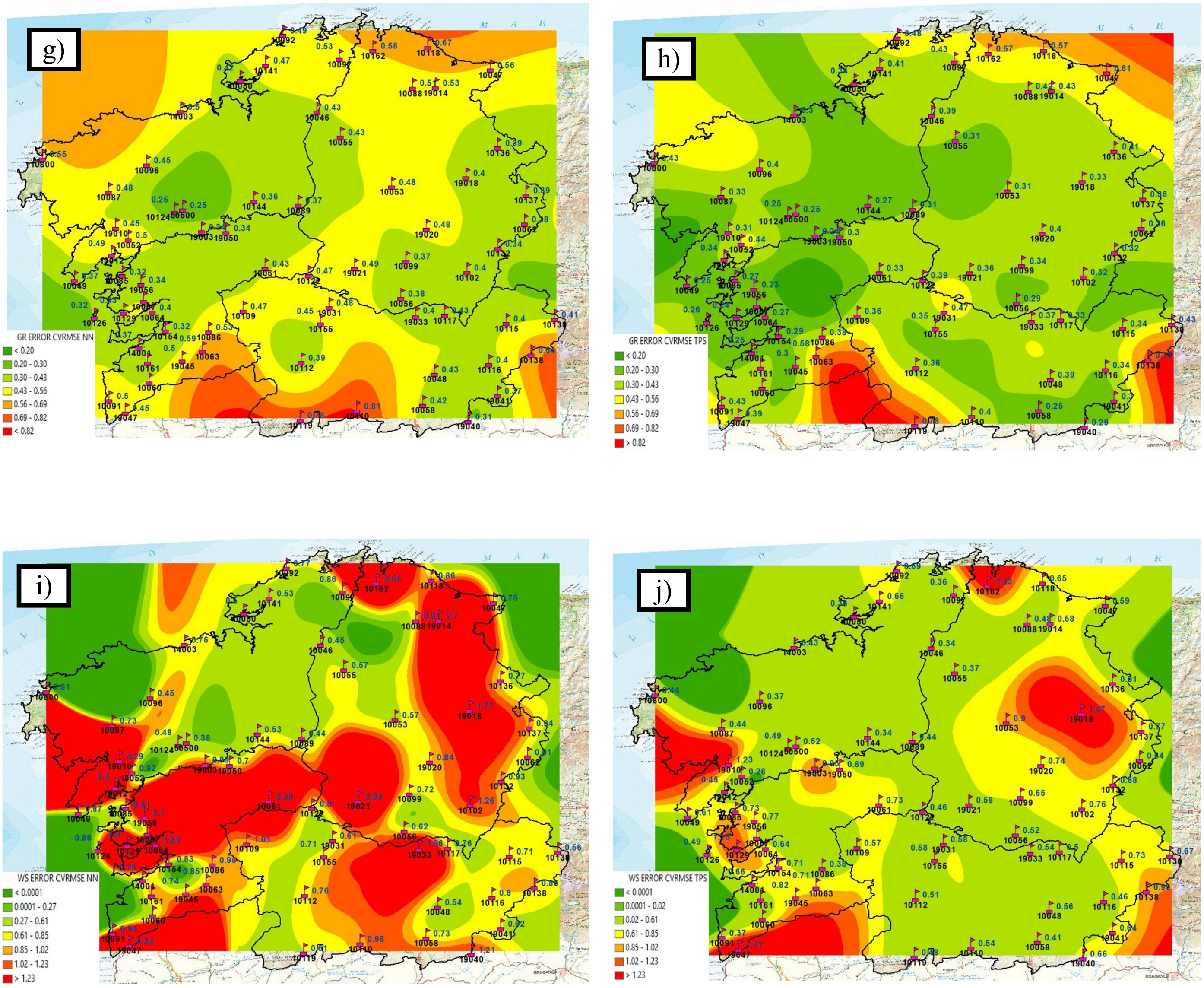


Figure 4. CV(RMSE) average values for the different weather variables at the 70 locations. a) Temperature NN, b) Temperature TPS, c) Relative humidity NN, d) Relative humidity TPS, e) Pressure NN, f) Pressure TPS, g) Global radiation NN, h) Global radiation TPS, i) Wind speed NN, and j) Wind speed TPS.

Variables	Observed-data		Estimated-values (techniques)			
			NN		TPS	
	M*	***CV****	***M***	***CV***	***M***	***CV***
Temperature (°C)	12.76	0.45	12.75	0.47	12.69	0.46
Relative humidity (%)	78.35	0.23	78.04	0.23	78.96	0.22
Atmospheric pressure (Pa)	96191.41	0.01	96162.89	0.01	96002.48	0.03
Global horizontal radiation (Wh/m^2)	161.26	1.56	159.58	1.56	161.18	1.54
Wind direction (deg)	174.47	0.56	174.00	0.56	172.94	0.58
Wind speed (m/s)	3.22	0.64	2.98	0.64	2.88	0.60
*M = Mean value **CV = $Coefficient\ of\ Variation\ = \frac{Standard\ Deviation}{Mean}$						

Table 6. Comparison between the estimated values and the observed data

Simulation results

Different simulations were performed to evaluate the influence of weather parameters on the building heating requirements. Three EnergyPlus weather datasets were generated for each location: On-site data, nearest neighborhood (NN), and TPS. On-site data datasets were generated using weather datasets from the MeteoGalicia station located at each analyzed location (a total of 70 points were analyzed). The test value was the on-site data at each location. As indicated, the heating requirements were calculated hourly. Figure 5 shows the relationship between the heating demands [kWh] and the interpolation technique over the first week of the month of February for the site of weather station 10130. This figure shows that the interpolation techniques have a tendency to underestimate heating demand. In the other locations, similar results were found. TRNSYS simulation results using on-site weather data show an average of 1230 (*SD* 79) heating hours per location, while the hours vary by 4.90% when using the NN technique and by 1.53% when applying TPS. The highest heating demand on average was in February (224 hours with heating requirements and a demand of 15 kWh).

NN is a fast and simple method, but there is a measurable lack of success. There is an annual heating demand difference, on average, for the 70 points of 32.97%, or 10630 kWh. For the TPS method, the difference is 7.12%, equivalent to 2302 kWh.

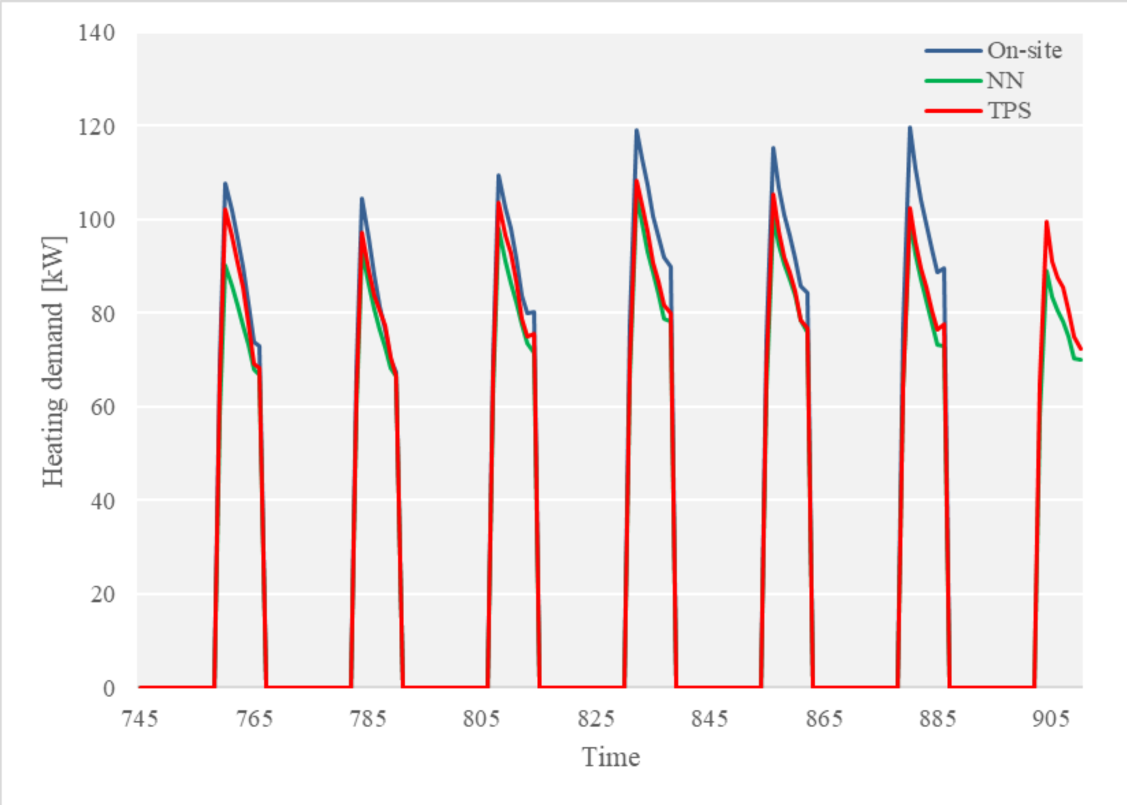


Figure 5. Heating demands during first week of February at location 10130.

The influence of latitude, longitude, and altitude was analyzed via a regression analysis. The results showed that the altitude, longitude, and latitude explain 51.7% of the MAE. A predictive model was tested for MAE, consisting of latitude, longitude and altitude variables. In testing the model, it was observed that the predictive model was significant $F (3, 66) = 23.59$, $p < 0.001$, explaining 51.7% of the variance ($R^2 = 0.517$). The coefficients of the model indicate that there is a positive relationship between altitude and MAE for TPS ($\beta = 0.666$), whereas there is a negative relationship between longitude ($\beta = -0.029$) and latitude ($\beta = -0.182$) and MAE for TPS. In other words, when the error of the simulation increases by a unit, the altitude goes up 0.666, the longitude lowers 0.029 and the latitude lowers 0.182 units. The altitude is the only variable that has a significant relationship with the error. Therefore, we proposed a more parsimonious model composed only of the altitude variable. This model was significantly explanatory, $F (1, 68) = 63.66$, $p < 0.001$, explaining 48.4% of the MAE ($R^2 = 0.484$) with a positive and significant beta coefficient ($\beta = 0.695$).

As seen in Table 8, the altitude influences the annual heating demand estimation errors. For both techniques, the E1 climatic zone contains the locations with the highest discrepancies. Nevertheless, once again, the TPS estimates are better (see MAE and RMSE values in Table 8).

	MAE		**RMSE**	
Climate zone	**NN**	**TPS**	**NN**	**TPS**
C1	2.80	0.66	5.69	1.56
C3	1.98	0.76	3.98	1.63
D1	2.94	0.86	5.90	1.91
D2	3.54	0.86	7.03	1.97
E1	3.49	1.41	6.97	3.07

Table 7. MAE and RMSE mean values for the five climatic zones and the two techniques (i.e., NN and TPS)

The simulation results confirm that TPS better estimates both weather variables and heating demands (Figure 6). The NN method presents poor performance over all locations, with the exception of isolated locations (e.g., 10092). A possible explanation for this result is the position on the border of this station (see Figure 6). It is known that interpolation techniques show poor results on border points [30]. Although the NN method is fast and simple, the results do not match the reality in most locations. The kernel density results are shown in Figure 7. The figure gives the extent of the intraspecific functional variability for the heating demands results and MAE values. The median value is represented by white dashes.

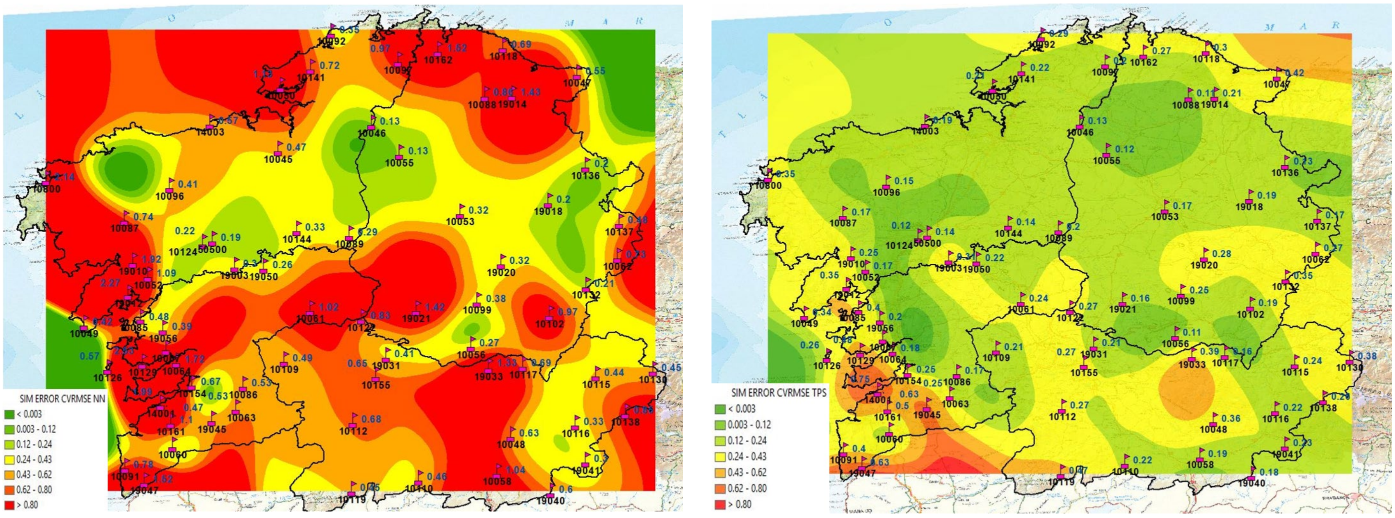


Figure 6. Maps with the calculated CV(RMSE) of thermal simulations at 70 locations for two methods over one year. (a) NN and (b) TPS.

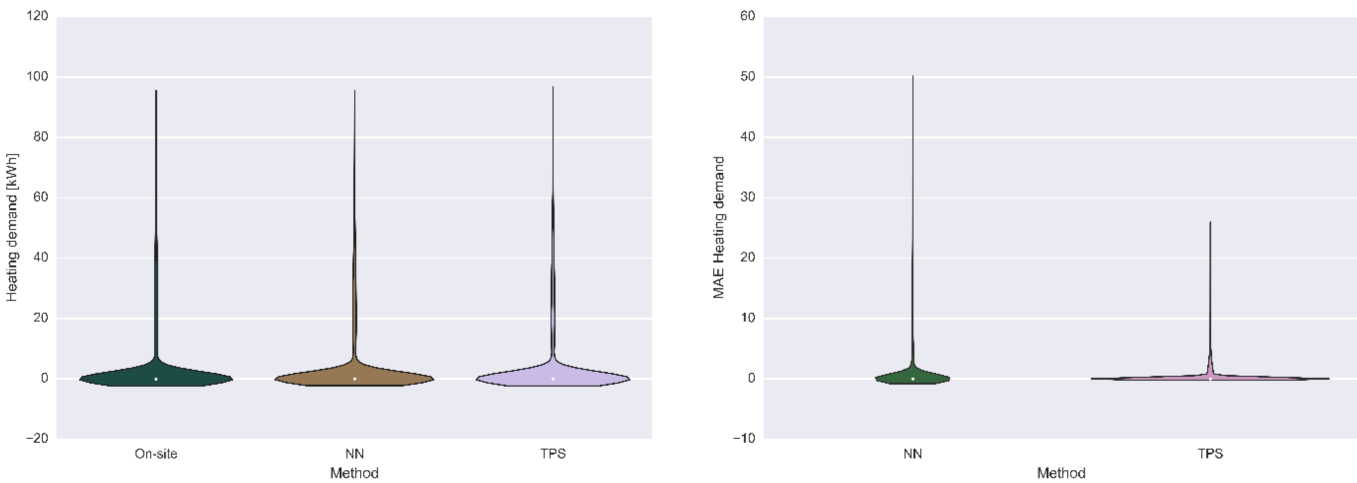


Figure 7. Violin-plots with the distribution of the heating demand results and the MAE values for the simulation results using the two techniques during 2015.

As seen in Table 9, the building location influences annual heating demands [kWh]. This quantitative difference is equivalent to a factor of 5.7 for heating load and 1.3 for heating hours. As demonstrated in a previous work, there are huge differences between the heating demands for the same building at different locations under the same climate severity [20]. If we consider the peak heating loads [kW], February is the highest month for the two techniques and for the on-site data as well. However, there are variations between the building's heating total loads calculated using the estimation techniques and on-site data. The peak heating load, in average, using on-site data was 128 kW. Again, NN has worse results (19.20% difference), whereas TPS showed better performance (16.99% difference). Furthermore, NN did not properly identify the studied location with the peak demand. In terms of the climatic zone, E1 had the highest heating demands. Once again, altitude plays an important role.

Climatic zone	**On-site**		**NN**		**TPS**	
	M*	***SD****	***M***	***SD***	***M***	***SD***
C1	1121	61	1171	102	1112	56
C3	1181	-	1268	-	1252	-
D1	1250	53	1222	83	1252	51
D2	1269	11	1253	57	1280	12
E1	1292	8	1284	32	1292	7

**M*: Mean value
***SD*: Standard deviation

Table 8. Annual heating building requirements for each climatic zone (heating load [kWh])

In summary, at the 70 locations, the TPS interpolation is the best method throughout the year (the coefficient of variation of RMSE is lower than 7%). The CV(RMSE) obtained values are within those mentioned by the ASHRAE. ASHRAE Guide 14 [10] considers that a building model is calibrated if the annual hourly CV(RMSE) obtained values are below 30%. For the calibration process, a more accurate meteorological file leads a lower deformation of the thermal model of the building to obtain the actual adjusted consumption.

In relation to climatic zone simulation results, Table 10 explains how the number of measurement points (weather stations) is important to the overall climatic zones, which results in a CV(RMSE) reduction between 42.09% and 69.79% for building simulations. Note that the error reduction (42.09%) is minor for harsh weather (E1 climatic zone); this result corresponds to locations that are colder in winter and those at a higher altitude.

CLIMATIC ZONE	NUMBER OF STATIONS	NN CVRMSE SIM MEAN	TPS CVRMSE SIM MEAN	% ERROR REDUCTION
C1	17	1.15	0.35	69.79
C3	1	0.65	0.27	58.96
D1	28	1.20	0.42	64.65
D2	6	0.76	0.24	68.38
E1	18	0.40	0.23	42.09

Table 9. Climatic zones building simulation analysis. CV(RMSE) error improvement using TPS vs. NN

Conclusions

This paper studied how building energy simulation calibration can be improved using interpolated weather data to determine on-site meteorological parameters at the building location.

Two different interpolation techniques were evaluated over the 8760 hours of one year. Although the most important factor in terms of estimation is the topographical complexity (i.e., altitude), the number of measured points (weather station density) also has an impact on the weather variable estimation and thus on the simulation results.

The building location influences annual heating demands [kWh]. Both techniques have a heating demand estimation error 1.5 times higher in the high altitude climatic zone (E1) than in the climatic zones at the lowest altitude. In terms of power, the difference between climatic zones is a factor of 5.7 for heating load and 1.3 for heating hours. Areas with higher altitudes have a greater need for weather stations, as the error is greater.

As expected, using TPS, we obtain better results than when we use NN. Quantitatively, the result of these obtained values is the same as estimating an excess average building heating demand during a whole year of more than 700 kWh over all locations. NN showed poor performance; this method tends to overestimate the building heating requirement by 13% compared to the on-site highest heating demand value.

The average coefficient of variation of the root-mean-square error (CV(RMSE)) obtained in the simulation of the studied building, placed successively in each one of the 70 meteorological station locations, decreases from 74% when using the nearest neighborhood (NN) to each site to 26% using the Thin plate spline (TPS) interpolation technique. The error in the building simulations is almost three times lower using the studied method; the process of simulation calibration is therefore greatly benefited.

Acknowledgements

We would like to thank for the meteorological database to Spanish State Meteorological Agency (AEMET).

This investigation article was partially supported by the Spanish Government (Project: ENE2015-65999-C2-1-R).

This investigation article was partially supported by the Spanish Government (Economy and Competitiveness Spanish Ministry), through the CDTI center (Industrial Technology Development Centre), and European FEDER 2007 - 2013 Technological Fund (European Regional Development Fund) (Project: IDI-20150503).

References

[1] International Energy Agency (IEA), "Energy efficiency," *Http://Www.Iea.Org/Aboutus/Faqs/Energyefficiency/,* 2015.

[2] European Parliament, "Directive 2010/31/EU of the European Parliament and of the Council of 19 May 2010 on the energy performance of buildings," *Official Journal of the European Union,* vol. 18, *(06),* 2010.

[3] Eurostat European Commission. Ministry of industry, energy and tourism. Institute for Diversification and Saving of Energy (IDAE), "Final report. SECH-SPAHOUSEC project. analysis of energy consumption in the residential sector in spain," July 16, 2011. 2011.

[4] Eurostat European Commission. Ministry of industry, energy and tourism. Institute for Diversification and Saving of Energy (IDAE), "Consumptions of the Residential Sector in Spain. Summary of Basic Information. http://www.idae.es/uploads/documentos/documentos_Documentacion_Basica_Residencial_Unido_c93da537.pdf," 2017.

[5] A. Dutta, A. Samanta and S. Neogi, "Influence of orientation and the impact of external window shading on building thermal performance in tropical climate," *Energy Build.,* vol. 139, pp. 680-689, 2017.

[6] D. A. Vasco *et al*, "Thermal simulation of a social dwelling in Chile: Effect of the thermal zone and the temperature-dependant thermophysical properties of light envelope materials," *Appl. Therm. Eng.,* vol. 112, pp. 771-783, 2017.

[7] N. M. Mateus, A. Pinto and G. C. da Graça, "Validation of EnergyPlus thermal simulation of a double skin naturally and mechanically ventilated test cell," *Energy Build.,* vol. 75, pp. 511-522, 2014.

[8] P. De Wilde, "The gap between predicted and measured energy performance of buildings: A framework for investigation," *Autom. Constr.,* vol. 41, pp. 40-49, 2014.

[9] G. Angrisani *et al*, "Influence of climatic conditions and control logic on NO x and CO emissions of a micro-cogeneration unit serving an Italian residential building," *Appl. Therm. Eng.,* vol. 71, *(2),* pp. 858-871, 2014.

[10] ASHRAE, *Guideline 14-2002, Measurement of Energy and Demand Savings.* Atlanta, Georgia: ASHRAE, 2002.

[11] D. Coakley, P. Raftery and M. Keane, "A review of methods to match building energy simulation models to measured data," *Renewable and Sustainable Energy Reviews,* vol. 37, pp. 123-141, 2014.

[12] IPMVP New Construction Subcommittee, *International Performance Measurement & Verification Protocol: Concepts and Practices for Determining Energy Savings in New Construction, Volume III, Part I, : , 2003.* . Washington, DC.: US Department of Energy, 2006.

[13] M. Royapoor and T. Roskilly, "Building model calibration using energy and environmental data," *Energy Build.,* vol. 94, pp. 109-120, 2015.

[14] H. Yoo, K. Lee and G. J. Levermore, "Comparison of heating and cooling energy simulation using multi-years and typical weather data in South Korea," *Building Services Engineering Research and Technology,* 2014.

[15] G. Mustafaraj *et al*, "Model calibration for building energy efficiency simulation," *Appl. Energy,* vol. 130, pp. 72-85, 2014.

[16] R. Aguiar, S. Camelo and H. Gonçalves, "Assessing the value of typical meteorological years built from observed and from synthetic data for building thermal simulation," in *Proceedings of the 6th International IBPSA Conference Ob Building Simulation'99 in Kyoto,* 1999, pp. 627-634.

[17] S. Oxizidis, A. Dudek and A. Papadopoulos, "A computational method to assess the impact of urban climate on buildings using modeled climatic data," *Energy Build.,* vol. 40, *(3),* pp. 215-223, 2008.

[18] A. A. Kasam, B. D. Lee and C. J. Paredis, "Statistical methods for interpolating missing meteorological data for use in building simulation," in *Building Simulation,* 2014, pp. 455-465.

[19] P. Eguía *et al*, "Weather datasets generated using kriging techniques to calibrate building thermal simulations with TRNSYS," *Journal of Building Engineering,* 2016.

[20] P. E. Oller *et al*, "Improving transient thermal simulations of single dwellings using interpolated weather data," *Energy Build.,* vol. 135, pp. 212-224, 2017.

[21] QGIS Development Team, "QGIS geographic information system," *Open Source Geospatial Foundation Project,* 2014.

[22] F. S. Westphal and R. Lamberts, "The use of simplified weather data to estimate thermal loads of non-residential buildings," *Energy Build.,* vol. 36, *(8),* pp. 847-854, 2004.

[23] L. O. Degelman, "A statistically-based hourly weather data generator for driving energy simulation and equipment design software for buildings," in *Proceedings of Building Simulation,* 1991, pp. 592-599.

[24] R Core Team, "R: A language and environment for statistical computing. R Foundation for Statistical Computing," 2016.

[25] D. Nychka, R. Furrer and S. Sain, "fields: Tools for spatial data. R package version 7.1," *Accessed Online,* vol. 10, 2014.

[26] N. Hofstra *et al*, "Comparison of six methods for the interpolation of daily, European climate data," *Journal of Geophysical Research: Atmospheres,* vol. 113, *(D21),* 2008.

[27] C. Camera *et al*, "Evaluation of interpolation techniques for the creation of gridded daily precipitation (1× 1 km2); Cyprus, 1980–2010," *Journal of Geophysical Research: Atmospheres,* vol. 119, *(2),* pp. 693-712, 2014.

[28] C. J. Willmott and K. Matsuura, "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance," *Climate Research,* vol. 30, *(1),* pp. 79, 2005.

[29] M. F. Hutchinson, "Interpolating mean rainfall using thin plate smoothing splines," *International Journal of Geographical Information Systems,* vol. 9, *(4),* pp. 385-403, 1995.

[30] D. R. Legates and G. J. McCabe, "Evaluating the use of “goodness-of-fit” measures in hydrologic and hydroclimatic model validation," *Water Resour. Res.,* vol. 35, *(1),* pp. 233-241, 1999.