Renewable Energy 95 (2016) 413–421

Contents lists available at ScienceDirect

Renewable Energy

journal homepage: www.elsevier.com/locate/renene

Developing multiple regression models from the manufacturer's ground-source heat pump catalogue data

F. Simon [a,*], J. Ordoñez [a], T.A. Reddy [b], A. Girard [c], T. Muneer [d]

[a] School of Civil Engineering, University of Granada, Av. Severo Ochoa s/n, 18071, Granada, Spain
[b] School of Sustainable Engineering and the Built Environment, Arizona State University, PO Box 871605, Tempe, AZ, 85287-1605, USA
[c] Faculty of Engineering and Sciences, Adolfo Ibáñez University, Av. Padre Hurtado 750, Altos del Sporting, Viña del Mar, Chile
[d] School of Engineering and Built Environment, Edinburgh Napier University, 10 Colinton Road, Edinburgh, EH10 5DT, Scotland, UK

ARTICLE INFO

Article history:
Received 11 August 2015
Received in revised form
15 April 2016
Accepted 19 April 2016

Keywords:
GSHP (ground-source heat pump)
Performance prediction
Manufacturer data
Multiple regression (MR)

ABSTRACT

The performance of ground-source heat pumps (GSHP), often expressed as Power drawn and/or the *COP*, depends on several operating parameters. Manufacturers usually publish such data in tables for certain discrete values of the operating fluid temperatures and flow rates conditions. In actual applications, such as in dynamic simulations of heat pump system integrated to buildings, there is a need to determine equipment performance under operating conditions other than those listed. This paper describes a simplified methodology for predicting the performance of GSHPs using multiple regression (MR) models as applicable to manufacturer data. We find that fitting second-order MR models with eight statistically significant *x*-variables from 36 observations appropriately selected in the manufacturer catalogue can predict the system global behavior with good accuracy. For the three studied GSHPs, the external prediction error of the MR models identified following the methodology are 0.2%, 0.9% and 1% for heating capacity (*HC*) predictions and 2.6%, 4.9% and 3.2% for *COP* predictions. No correlation is found between residuals and the response, thus validating the models. The operational approach appears to be a reliable tool to be integrated in dynamic simulation codes, as the method is applicable to any GSHP catalogue data.

1. Introduction

In order to reduce the fossil fuel consumption, ground-source heat pumps (GSHP) are becoming a widely employed technology for heating and cooling buildings and for domestic hot water production [1–5]. GSHP attract worldwide interest due to their environmental protection, energy efficiency and from being a renewable energy form [6–9].

In dynamic simulation applications, such as in the modeling of heat pump systems integrated to buildings, it is required to evaluate the heat pump performance under particular operating conditions. However, it becomes difficult to estimate the correct performance value at operating conditions which do not exactly correspond to those listed in the manufacturer's data tables [10,11].

In the present study we introduce a methodology based on multiple regression (MR) modeling which can be used to predict GSHP performance with good accuracy at particular design conditions. The parameters in the models are the operating secondary fluids (source and load) inlet temperatures and flow rates, the heat pump heating capacity (*HC*) and the coefficient of performance (*COP*). The analysis presented is based on three manufacturer performance tables of three commercially available GSHPs in heating mode.

The operational method for the identification of MR models can be integrated in dynamic simulation tools such as EnergyPlus and TRNSYS in order to predict the performance of GSHP at particular operating conditions. The same approach can be employed for the development of MR models for the performance simulation of any GSHP system in heating or cooling mode using the corresponding capacity table from the manufacturer catalogue. Our hypothesis is that the proposed method can help in the selection of the most appropriate GSHP evaluating with precision its performance, and thus increase the potential of GSHP implementation in buildings.

The paper is organized as follows. Next section introduces the methodology employed in the study based on step forward multiple regression modeling. The regression analysis, including a

* Corresponding author.
E-mail address: fsimon@correo.ugr.es (F. Simon).

http://dx.doi.org/10.1016/j.renene.2016.04.045
0960-1481/

Nomenclature

Acronyms

COP	coefficient of performance
CV	coefficient of variation
GSHP	ground-source heat pump
HC	heating capacity
HE	heat extracted
MR	multiple regression
P	compressor power input
RMSE	root mean square of error

Symbols

β	regression coefficient
ε	residual
μ	mean
t	temperature (°C)
v	flow rate (L/s)
x	independent variable
y	dependent variable

Subscripts

i	inlet condition
j	number of response
l	load
n	number of predictor variables
o	outlet condition
s	source

description of the manufacturer specification tables, a general overview of regression models, the statistical evaluation of the identified MR models, and the proposed approach validation is presented in Section 3. Section 4 presents a discussion about the findings and Section 5 is some concluding remarks.

2. Methodology

The purpose of the study is to introduce a general methodology based on MR modeling able to determine the heat capacity and the coefficient of performance of GSHPs in heating mode from different working fluid temperatures and flow rates. Essentially, an optimal set of operating parameters from the manufacturer data tables would be to find relationships between the response variables (*HC* and *COP*) versus four operating parameters: the source flow rate (v_s), the load flow rate (v_l), inlet load temperature at the heat pump condenser (t_{il}) and inlet source temperature at the heat pump evaporator (t_{is}). The compressor power input (*P*) and heat extracted (*HE*) can then be deduced from *HC* and *COP* values. In the study, t is temperature, v is flow rate, subscripts i, l and s are short for the heat pump inlet conditions, load and source, respectively. Fig. 1 shows the schematic diagram of the GSHP system for space heating.

The method consists in selecting a sample of observations from the specification table, which can be used for the development of MR models. The performance data of the remaining observations in the manufacturer table are then compared to the predictions calculated from the identified models. The statistical analysis further evaluates the models' robustness and prediction accuracy, determining the models' goodness-of-fit and the coefficients of variation (CV) of the prediction residual errors.

An important aspect of the selection of the observation sample is that it needs to provide a fair representation of the entire input space. In this concern, the observation sample is selected in a way that there are an equal number of low-range, mid-range and high-range values for each of the variables, and the combination selection is appropriately spread out so that the whole input space is represented. Each combination of particular operating conditions corresponds to an observed value of *HC* and *COP* from the manufacturer catalogue. Since there are four independent variables (v_s, v_l, t_{il} and t_{is}), a complete factorial design as used in statistical experiments would require 81 observations. For the purpose of the study, we propose three incomplete factorial designs using Latin squares, i.e. three samples containing 12, 24 and 36 observations respectively, in order to evaluate the influence of the sample size on model accuracy.

The approach involves using multiple linear regression of the first and second order to estimate the heat pump *HC* and *COP* values in heating mode under specified conditions of the four operating parameters (t_{is}, t_{il}, v_s and v_l). As seen in Figs. 2–5 below, although some of the relationships are close to linear, others are distinctly non-linear, and thus the higher order terms are introduced to allow linear regression on non-linear relationships.

3. Regressions analysis of manufacturer's performance data

3.1. Description of manufacturer's data tables

The principle of GSHP technology is to make use of the low-grade geothermal energy of the earth at a relatively lower depth through ground heat exchangers. A GSHP system will extract/discharge thermal energy for all of its applications [12–15]. In order to make it easier for engineers when it comes to GSHP selection and sizing for a particular building, GSHP manufacturers offer performance data tables from their catalogue. In heating mode, such specification tables give data of compressor power input (*P*), heat extracted (*HE*), heat capacity (*HC*) and *COP* as a function of the inlet and outlet load temperatures (t_{il} and t_{ol}) of the distribution circuit fluid entering and leaving the heat pump condenser, the inlet and outlet source temperatures (t_{is} and t_{ls}) of the earth connection fluid entering and leaving the heat pump evaporator, and the load and source flow rates (v_l and v_s), in the distribution and ground loops respectively. The specification tables of the cooling mode are arranged identically, giving observations of cooling capacity (*CC*), heat rejected (*HR*) and cooling energy efficiency (*EER*) as a function of temperature and flow rate parameters. Unlike the capacity tables in heating mode, the distribution and the ground circuits are respectively the source and the load side in cooling mode. Table 1 provides the technical features of the three commercially available GSHPs considered in the study (called HP1, HP2 and HP3).

The working operations in heating mode of the studied GSHPs range between the interval limits shown in Table 2. However there are some conditions for which operation is not recommended by the manufacturer. These extreme conditions for the three studied equipments are as following:

- HP1: When t_{is} varies between −1.1 and 10 °C, v_s cannot vary between 0.95 and 1.45 L/s;
  When t_{is} varies between 21.1 and 32.2 °C, t_{il} cannot vary between 26.7 and 48.9 °C.

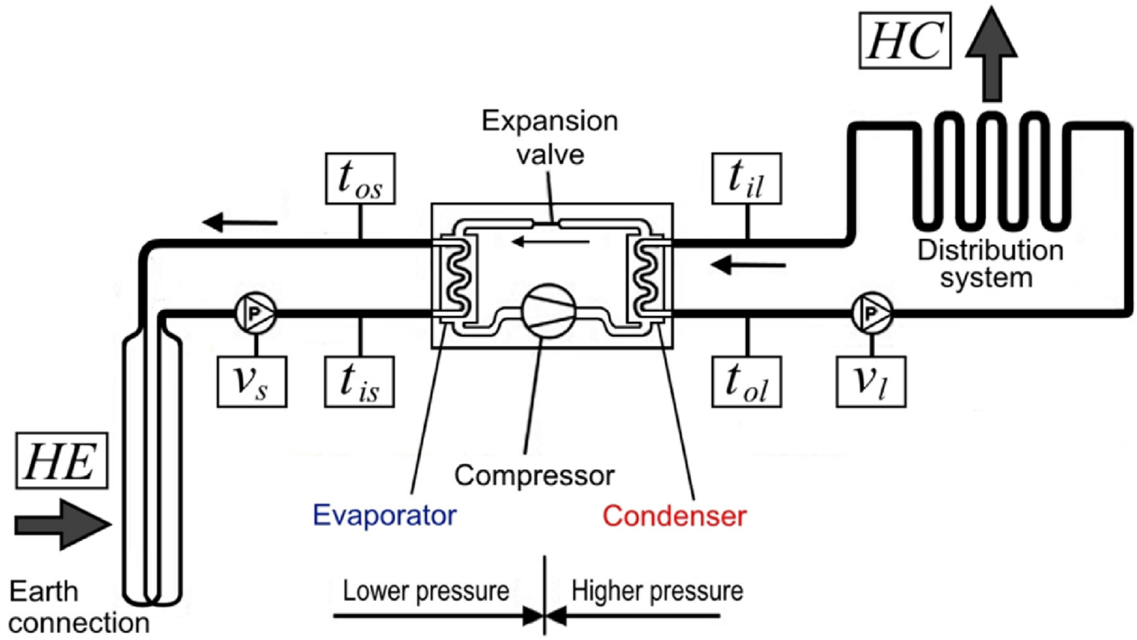

Fig. 1. Schematic diagram of GSHP in heating mode.

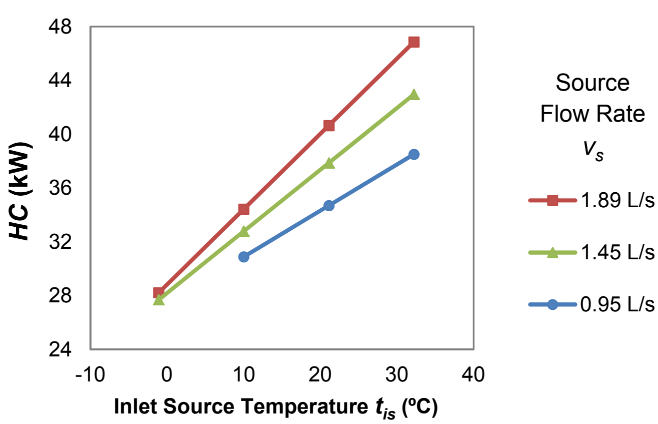

Fig. 2. Heat capacity against inlet source temperature for t_{il} = 15.6 °C and v_l = 1.45 L/s.

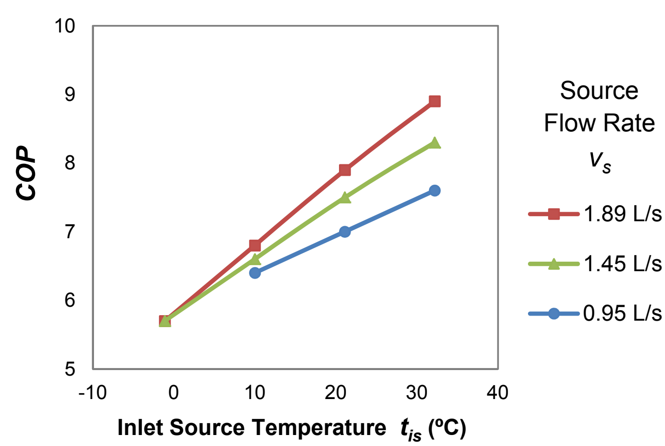

Fig. 4. *COP* against inlet source temperature for t_{il} = 15.6 °C and v_l = 1.45 L/s.

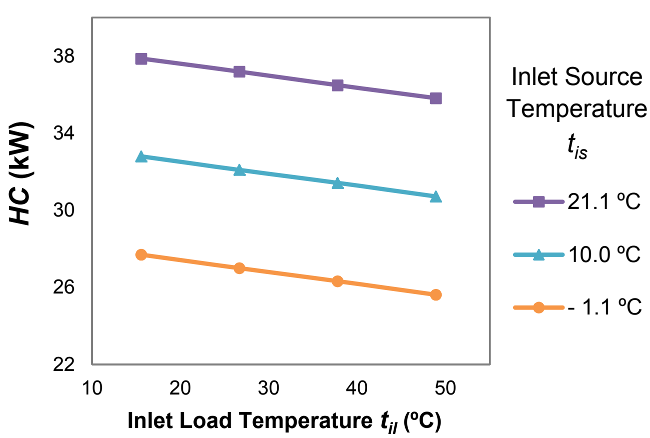

Fig. 3. Heat capacity against inlet load temperature for v_s = 1.45 L/s and v_l = 1.45 L/s.

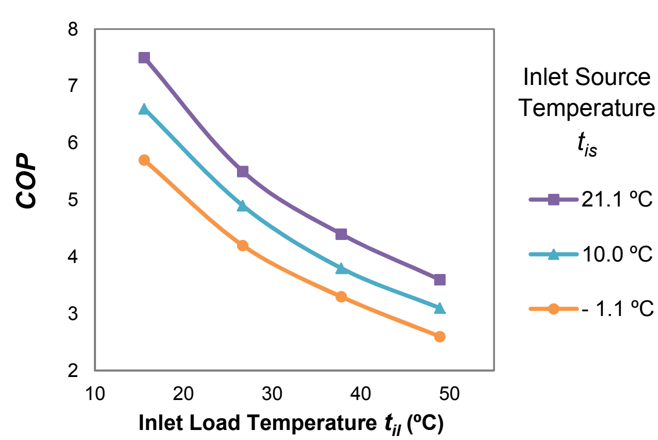

Fig. 5. *COP* against inlet load temperature for v_s = 1.45 L/s and v_l = 1.45 L/s.

Table 1
Technical features of the considered GSHP.

	HP1	HP2	HP3
Compressor (number; type)	2; scroll		1; scroll
Evaporator	Plate heat exchanger		Coaxial tube-in-tube
Condenser	Plate heat exchanger		Coaxial tube-in-tube
Refrigerant fluid	R410a	R134a	R410a
Ground loop fluid	15% propylene glycol antifreeze solution		
Distribution loop fluid	15% propylene glycol antifreeze solution		

Table 2
Parameter variation in data tables in heating mode and operating limits.

	HP1	HP2	HP3
t_{is} (°C)	−1.1, 10, 21.1, 32.2	−1.1, 10, 21.1, 32.2	−3.9, −1.1, 10, 21.1, 32.2
t_{il} (°C)	15.6, 26.7, 37.8, 48.9	15.6, 26.7, 37.8, 48.9, 60.0	15.6, 26.7, 37.8, 48.9
v_s (L/s)	0.95, 1.45, 1.89	0.95, 1.26, 1.58	0.25, 0.35, 0.44
v_l (L/s)	0.95, 1.45, 1.89	0.95, 1.26, 1.58	0.25, 0.35, 0.44

- HP2: When t_{is} varies between 21.1 and 32.2 °C, t_{il} cannot vary between 48.9 and 60.0 °C.
- HP3: When t_{is} varies between −3.9 and −1.1 °C, v_s cannot vary between 0.25 and 0.44 L/s;
  When t_{is} varies between 21.1 and 32.2 °C, t_{il} cannot vary between 26.7 and 48.9 °C.

Therefore, the total number of observations in the data tables in heating mode is 114 for HP1, 171 for HP2 and 138 observations for HP3.

Manufacturer performance data are produced on the basis of the ISO Standard 13256-2 [16] for the testing of water-to-water GSHP units. The prescribed conditions to produce the ISO data include testing at steady flow conditions for a set temperature of source and sink and the measured inputs consist of compressor energy and the circulation pump requirements to overcome the frictional resistance of the evaporator and condenser. Therefore, manufacturer's *COP*/*EER* values are generally higher than those obtained in real installations, when taking into account the full energy input for circulation pumps and system fans plus all associated controls. Kim et al. [17] compared the manufacturer's data based on the ISO standard to the actual GSHP performance, and propose a verification method allowing the identification and correction of eventual gaps.

In practice, the performance of a GSHP system also is affected by installation conditions and depends on the continual changes in the surroundings environment, i.e. variable source and sink temperature and load. Indeed, heat source and sink temperatures have a direct impact on the pressure (and thus the temperature) at which the evaporation and condensation occur in the refrigeration cycle. Any change in evaporating or condensing temperature affects the density of the refrigerant, which alters the compression ratio between the low-pressure and high-pressure sides, and thus the performance of the compressor.

The type of refrigerant fluid flowing in the cycle also influences the performance of a heat pump. Fluid density, viscosity, thermal conductivity and specific heat are all factors affecting the compression efficiency and heat transfer capacities through the evaporator and condenser. Heat pump manufacturers usually make the decision for the choice of refrigerant based on system cost and application range requirements, considering fluid boiling and condensing temperatures.

In Figs. 2 and 3 the heating capacity data of HP1 are reported into charts, as a function of the inlet source and load temperatures for the three different source and load flow rates. In Figs. 4 and 5, the corresponding *COP* data given by the manufacturer for HP1 is shown.[1] At higher inlet source temperature and/or source flow rate, Fig. 2 shows that higher thermal output can be reached, because higher thermal energy is extracted. Higher inlet source temperature and/or source flow rate raises the evaporating pressure and temperature, and as the compression ratio remains the same, the condensing pressure and temperature also increases, resulting in higher thermal energy output. As the heating capacity is linear with inlet load temperature, Fig. 3 shows the degradation of compressor efficiency at higher temperatures, because it is getting near the heat pump operational limits (condensing temperature about 55–60 °C). Fig. 4 shows that the heat pump *COP* increases when the inlet temperature and/or the source flow rate increases. The interpretation of Fig. 4 is similar to the one of Fig. 2. However, the slopes of the curves in Fig. 4 seem to decrease at higher temperature, showing the degradation of the compressor efficiency at higher refrigerant temperatures (higher compressor inputs required at higher temperature in order to reach the same compression ratio). The *COP* value always increases when the temperature difference between the heat source and the heat sink decreases. This is also shown in Fig. 5 where the heat pump efficiency lowers when the inlet load temperature increases and/or the inlet source temperature decreases.

3.2. Model identification

A linear MR model may be represented by the following Equation (1). Details of such method are given in standard textbooks on regressions such as in Draper et al. [18], Neter et al. [19], James et al. [20] or Johnson et al. [21].

$$y_j = \beta_0 + \beta_1 x_{j1} + \beta_2 x_{j2} + \beta_n x_{jn} + \varepsilon_j \quad (1)$$

where y_j is the j response to be predicted using the (predictor) variables, x_{j1} to x_{jn} given as input. n is the number of predictor variables and β the regression coefficients. ε is the j residual or error between the predicted response and the observation.

Regression models assume that a straight-line relationship exists between each independent variable (x_{j1}, x_{j2}, … and x_{jn}), and the dependent variable (y). The β_n values are coefficients corresponding to each x-value, and β_0 is a constant value. The least squares

[1] Charts for the two other studied GSHPs (not reported here) indicate similar trends.

principle was used to determine the regression coefficients. The method determines the coefficients that produce the minimum sum of squared residual values, i.e. the best fitted regression line to the manufacturer's data.

The output shows the results of fitting a linear MR model to describe the relationship between a dependent variable *HC* or *COP* and 14 independent variables, which are as follows: t_{is}, v_s, t_{il}, v_l, t_{is}^2, v_s^2, t_{il}^2, v_l^2, $t_{is}\cdot v_s$, $t_{is}\cdot t_{il}$, $t_{is}\cdot v_l$, $v_s\cdot t_{il}$, $v_s\cdot v_l$, and $t_{il}\cdot v_l$. The models have been run with a stepwise forward multiple regression and so only the statistically significant model terms have been retained, i.e. those which indicate a significance level superior than 95% (p-value inferior to 0.05). This method allows a more robust model to be identified than including all variables as is done in standard regressions.

3.3. Statistical evaluation for HP1

The observations used for model fitting are data which can include some error, due to measurement inaccuracies, instrument calibration failure, test conditions such as observations taken before the operation has reached steady-state, etc. Although the intent of regression modeling is to capture only the structural behavior of the system, depending on the veracity and quantity of observations, a certain amount of these errors may be captured in the model. This is called model over-fitting.

The coefficient of variation (CV) or relative variability aims to evaluate the relative sizes of the model squared residuals and outcome values from the regression model. The lower is the CV; the smaller are the residuals relative to the predicted value. For the purpose of the study, we introduce two types of CV value in the statistical analysis, first the CV found during model fitting (also called internal prediction error or internal CV), applying to the observations used for model identification, and secondly the CV when applying the identified model to the remaining of the observations of the manufacturer table (called external prediction error or external CV). The comparison between internal and external CVs allows detecting the cases of model over-fitting. The internal CV is a measure of model goodness-of-fit over the observation sample, while the external CV is a measure of how well the model is likely to predict future system behavior. The internal CV is expected to be smaller than the external CV, because the set of data points used for model identification is an incomplete sample of the full dataset, and so they are likely to contain fewer errors than in the remaining of observations in the table.

As suggested in statistics books [18,21], in order to reduce the probability of model over-fitting, the number of data points used to fit a MR model has to be at least three times the number of independent *x*-variables in the model, otherwise one variable starts fitting noise. Therefore, the maximum number of *x*-parameters which are used to fit a model from a set of 12, 24 and 36 observations is respectively four, eight and twelve.

The test results on the statistical evaluation of MR models for the studied HP1 are listed in Table 3 for HC_1 and COP_1 predictions. Subscript *1* is the notation referring to HP1. These results reveal the R-square values adjusted for degrees of freedom of the identified models, the internal CV, the external CV, as well as an indication about the pattern formed by external residuals when plotted against the predicted responses for the remaining of the data set, where C, SE and N stand for clear, some kind of evidence and no residual specific pattern respectively. The simple R-square statistic is not reported because its value is generally biased. The magnitude of the bias in simple R-square values depends on how many observations are available to fit the model and how many variables are relative to the sample size. It is more appropriate to use the adjusted R-square value (coefficient of determination) to compare models with different numbers of independent *x*-variables, as it takes into account the size of the data set and the number of predictor variables [20,21]. For the models with the highest number of *x*-variables, the value of adjusted R-square is seen to approach unity. Results suggest that models with a lower number of *x*-variables are less robust, although they still indicate excellent goodness-of-fit within the observation sample. For instance, the evaluation of the models for HC_1 and COP_1 predictions as fitted from the 24 observations with 8 significant *x*-variables shows that the adjusted R-Squared statistic explains 99.998% and 99.88% of the variability in HC_1 and in COP_1 respectively, while the evaluation of the models identified with only three variables indicates 97.85% and 97.18% respectively.

The input parameters for the modeled results were varied accordingly within the different independent *x*-variable combinations for the observations used for fitting the models, as well as for the remaining observations in the manufacturer table, and for each case the predicted values of *HC* and *COP* were found. The coefficient of variation (CV) of the residual errors was calculated using the following Equation (2):

$$CV = \frac{RMSE}{\mu} \tag{2}$$

where *RMSE* is the root mean square of error, also called the standard error of the estimate and μ is the mean value of the observed data.

The internal CV values (Table 3) suggest of a good model fit for most of the HC_1 models, the COP_1 models describing a poorer fit in terms of relative closeness of the predictions to the actual values from the observation sample. Cases of model over-fitting are detected for second-order COP_1 models with three and four *x*-variables identified from the 12 observation sample, where the external CV is much larger than the internal CV. First-order COP_1 models, first-order HC_1 model with one *x*-variable identified from the 12 observation set, as well as some second-order HC_1 and COP_1 models identified with a low number of *x*-variables, are all models for which the internal CV is found larger than the external CV, thus indicating of model over-fitting. For all the remaining models developed for HP1, the internal CV values are slightly smaller than external CVs, demonstrating that the observation sample used for model identification is a fair representation of the entire data table. In these cases, the identified models have successfully captured the structural behavior of the HP1 system.

In order to validate our approach, we assume that the external CV value should be inferior to a 5% threshold limit, which is an acceptable level in statistical analysis. For the COP_1 prediction, this level of prediction accuracy is only achieved when there are a minimum of five independent variables in the model.

Another condition required to validate the method is that no specific pattern should be formed by external residuals when plotted against the predicted responses. Indeed, no specific pattern in such a plot indicates that residual errors have constant variance and that they are independent, i.e. there is no correlation with regression coefficients or the response. Although the majority of the MR models identified are found robust and relatively accurate in predicting the remaining data listed in the manufacturer's catalogue, the only tests where model residuals are found to have constant variance is for second-order models with eight significant *x*-variables identified from the 36 observation sample. A clear pattern formed by residuals was found in most of the other model tests, particularly for models with fewer parameters and identified from a lower number of observations, i.e. when the degree of freedom is low.

Table 3
Test results of MR models for HP1.

Observation sample	Model order	Number of x-variables	HC_1				COP_1			
			Adj. R^2	Int. CV	Ext. CV	Res. pattern	Adj. R^2	Int. CV	Ext. CV	Res. pattern
12 observations	First	2	0.9730	3.8%	4.9%	C	0.9206	12.5%	10.9%	C
		1	0.8793	8.1%	7.4%	C	0.8547	17.0%	19.2%	C
	Second	4	0.9987	0.8%	1.1%	C	0.9886	4.7%	16.8%	C
		3	0.9931	1.9%	2.6%	C	0.9657	8.2%	17.7%	C
		2	0.9851	2.8%	3.4%	C	0.8758	15.7%	17.3%	SE
24 observations	First	4	0.9809	2.7%	2.9%	C	0.9446	9.5%	8.1%	C
		3	0.9755	3.1%	3.8%	SE	0.9352	10.3%	9.2%	C
		2	0.9570	4.1%	4.7%	C	0.9228	11.3%	10.7%	C
	Second	8	0.99998	0.1%	0.2%	C	0.9988	1.4%	2.7%	SE
		7	0.99996	0.1%	0.2%	C	0.9984	1.6%	3.0%	SE
		6	0.9995	0.4%	0.7%	C	0.9974	2.3%	3.8%	C
		5	0.9990	0.6%	0.7%	C	0.9940	3.1%	4.1%	C
		4	0.9997	0.3%	0.4%	C	0.9869	4.6%	6.0%	C
		3	0.9785	2.9%	3.0%	C	0.9718	6.8%	7.7%	C
		2	0.9730	3.2%	3.2%	C	0.8793	14.1%	12.7%	C
36 observations	First	4	0.9798	2.8%	3.0%	C	0.9426	9.6%	7.9%	SE
		3	0.9759	3.0%	4.2%	C	0.9263	10.9%	8.4%	SE
		2	0.9583	4.0%	5.4%	C	0.9047	12.4%	9.5%	C
	Second	8	0.99996	0.1%	0.2%	N	0.9984	1.6%	2.6%	N
		7	0.9999	0.2%	0.2%	SE	0.9979	1.8%	2.8%	SE
		6	0.9998	0.3%	0.3%	C	0.9942	3.1%	3.6%	C
		5	0.9998	0.3%	0.3%	C	0.9930	3.4%	3.5%	C
		4	0.9995	0.4%	0.6%	C	0.9802	5.6%	5.4%	C
		3	0.9844	2.4%	2.1%	C	0.9705	7.5%	8.0%	C
		2	0.9765	3.0%	3.6%	C	0.9314	10.1%	9.1%	C
		1	0.9601	3.9%	4.0%	C	0.7485	20.1%	19.7%	C

3.4. Modeling approach validation

The statistical analysis shows that a validated approach is to develop second-order MR models containing eight statistically significant x-variables from 36 observations. Using this operational approach, the fitted models fulfilled the two above mentioned conditions, i.e. CV inferior to 5% and no specific pattern formed by residuals, for both HC_1 and COP_1 predictions. Following this particular approach, Tables 4 and 5 assemble the coefficients for each of the model independent variables along with their standard errors and their p-values for HC_1 and COP_1 models respectively. The MR models (shown in Tables 4 and 5) yield predictions within the range of the HP1 working operations. Temperatures are in degrees Celsius, flow rates in Liters per second, the heat capacity in kilowatt and the *COP* is dimensionless.

Each of the coefficients indicates the influence of each predictor x-variable on the GSHP working data. For instance, the coefficient of the load inlet temperature (t_{il}) is lower for the HC_1 model, indicating that the benefit of increasing t_{il} will have smaller impact on the HC_1 compared to COP_1. On the other hand, increasing the load flow rate (v_l) will have higher impact on the HC_1 compared to COP_1. The standard errors of each of the coefficients are the margins for the model output to remain within a 95% confidence interval of the predicted values. They are all seen to be minor compared to the coefficients.

Additionally, the p-values for each of the coefficients represent the probability of each of the predictor variables being insignificant for the model result. For most of the coefficients in the two models, the p-values are lower than 0.0001, i.e. it is certain that the corresponding variables are important for the predicted *HC* or *COP*. Two exceptions are the v_s^2 variable in the HC_1 model and the $v_s\, t_{il}$ variable in the COP_1 model which have p-values of 0.0007 and 0.0046 respectively. However, since these p-values remain quite low, the x-variables can be considered to be significant at the 95% level.

The mapping method for the identification of MR models was found valid for the prediction of the HP1 catalogue data. Therefore,

Table 4
Model coefficients and statistics (Model for HC_1).

Parameter	Coefficient	Std. Error	P-value
Constant	25.325	0.1701	7E-41
v_s^2	0.1176	0.0309	0.0007
$t_{is}\, v_s$	0.232	0.002	4E-38
$t_{is}\, v_l$	0.0457	0.002	2E-19
$v_s\, v_l$	0.6882	0.0691	2E-10
$t_{il}\, v_l$	−0.0212	0.0021	9E-11
t_{is}	0.0543	0.004	1E-13
t_{il}	−0.0313	0.0029	3E-11
v_l	1.4856	0.1722	3E-09

Table 5
Model coefficients and statistics (Model for COP_1).

Parameter	Coefficient	Std. Error	P-value
Constant	7.8871	0.2042	4E-25
t_{il}^2	0.0025	0.0001	3E-16
$t_{is}\, v_s$	0.0432	0.0024	2E-16
$t_{is}\, t_{il}$	−0.0009	0.0001	2E-08
$t_{is}\, v_l$	0.0153	0.0026	3E-06
$v_s\, t_{il}$	−0.007	0.0022	0.0046
$t_{il}\, v_l$	−0.018	0.0032	7E-06
t_{il}	−0.222	0.0118	5E-17
v_l	1.0687	0.1215	2E-09

Table 6
Summary of statistical evaluation of MR models for HP1, HP2, and HP3.

Model	Adjusted R^2	F-significance	External RMSE	External CV
HC_1	99.996%	4E-43	0.0710	0.21%
COP_1	99.843%	3E-38	0.1290	2.62%
HC_2	99.960%	3E-45	0.2817	0.95%
COP_2	97.880%	5E-22	0.2367	4.91%
HC_3	99.942%	4E-43	0.0842	1.02%
COP_3	99.868%	3E-38	0.1564	3.21%

the same operational approach is used to develop MR models for HP2 and HP3 data. Subscripts *2* and *3* are the response *y*-variables notations as representing the studied HP2 and HP3 respectively.

The results on the statistical evaluation of MR models for HP1, HP2 and HP3 are listed in Table 6. These results indicate the R-square values adjusted for degrees of freedom, the F-significance resulting from the analysis of variance table, the external CV, as well as the external the root mean square of error (RMSE) when applying the model to the remaining of the observations. The adjusted R-square value for all the models is seen to approach unity, except for the COP_2 model, where the adjusted R-Squared statistic explains only 97.888% of the variability in COP_2. Overall, such high values indicate that the independent *x*-variables used in the models can jointly predict the variation in the outcome with a high accuracy. Moreover, all six models are statistically strong judging from the F-significances.

The external CV values suggest of a good model fit for each of the six models (i.e. CV inferior to the 5% threshold value), the *COP* models describing a slightly poorer fit in terms of relative closeness of the predictions to the actual values. The RMSE shows the standard deviation of the residuals. The standard error for all the models is relatively small compared to the ranges of *HC* and *COP*. This is illustrated in Fig. 6, where the predicted values for the HC_1 and COP_1 model are plotted against the observed. It is evident that the predicted heat capacities are very close to the observed values. The prediction is seen to be slightly poorer with the COP_1 model, confirming graphically the difference of external CV values between these two models.

The CV values in Table 6 reveal that the size of the residual relative to the predicted value is bigger for *COP* models compared to *HC* models. However, when considering the range value of *HC* compared to *COP*, the size of residuals is not so different in term of absolute value. For instance, the standard error is smaller for the COP_2 model than it is for the HC_2 model, although the range value of HC_2 is higher. This remark is illustrated in Figs. 7–9, which depict the external residuals plotted against the predicted *y*-variables. In the six charts, no specific pattern formed by the external residuals can be observed, thus verifying that residuals are independent and have constant variance.

There are only three data points which are outliers (greater than ± 2) in the model for HC_1, and two outliers in the model for COP_1. For these data points, the model prediction output is poorer. There are seven and five outliers respectively in the models for HC_2 and COP_2, while there are one and three outliers in the model for HC_3 and COP_3. However, the fit is overall excellent for each of the six models. The maximum difference of obtained predicted values as compared with the observed manufacturer data are less than 1% and 7% for COP_1 and HC_1 predictions, 3% and 16% for HC_2 and COP_2, and 3% and 12% for HC_3 and COP_3 respectively. These errors concern only a few individual points, particularly these in the lower range values.

4. Discussion

The statistical evaluation of the models identified for HP1 has demonstrated the influence of the number of significant *x*-variables and their order level on prediction accuracy. The meaning of the observation sample size on model robustness and prediction error was also investigated. It was found that the global GSHP behavior cannot be predicted accurately when fitting MR models from the 12 observations sample. When the degree of freedom is low, the models tend to capture the errors present in the observation data, thus leading to model over-fitting. The 24 observations are sufficient to achieve acceptable levels of accuracy for *HC*, even with a low number of *x*-variables, as well as for the *COP* with second-order models including at least five significant *x*-parameters. Model robustness increased and prediction error dropped for models fitted from the 36 observation set as compared to these identified from 24 observations. However, the only models which were statistically validated are second-order models identified from the 36 observations and containing eight *x*-variables, for which the residuals checking revealed no specific pattern formation.

Following the validated model identification method, the six models introduced in the study are found excellent with these of *COP* being slightly poorer. Such results indicate that the manufacturer data is certainly data generated from a controlled experiment and is not data collected from the field. The high R-squares, such as those over 99.9% are also an indication of model over-fitting. In that sense, our guess is that the observation data are actually based on a regression model developed by the heat pump manufacturers, and that the model used for *HC* data is in fact a second-order one. The proprietary regression models used by the manufacturer to generate the performance data tables were closely reproduced, particularly with the models for *HC*.

The operational approach consisting in the identification of MR models containing eight significant independent variables from a set of 36 observations can be employed to predict the performance of GSHPs with quite good precision. The proposed mathematical models appear to be reliable tools to be implemented into dynamic building-plant energy simulation codes or into building energy certification tools. Indeed, the proposed method can be applied to any GSHP, to determine *HC* and *COP* in heating mode or *CC* and *EER* in cooling mode using their corresponding performance data

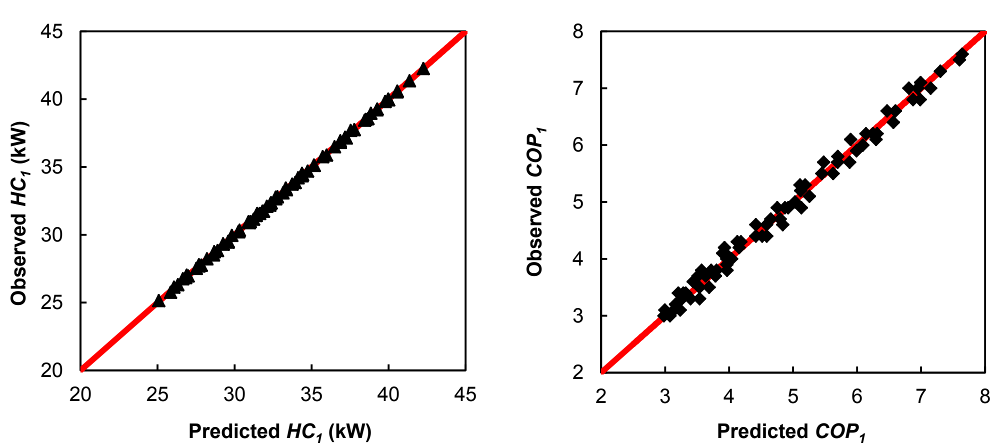

Fig. 6. Plot of HC_1 (kW) and COP_1 observed versus predicted.

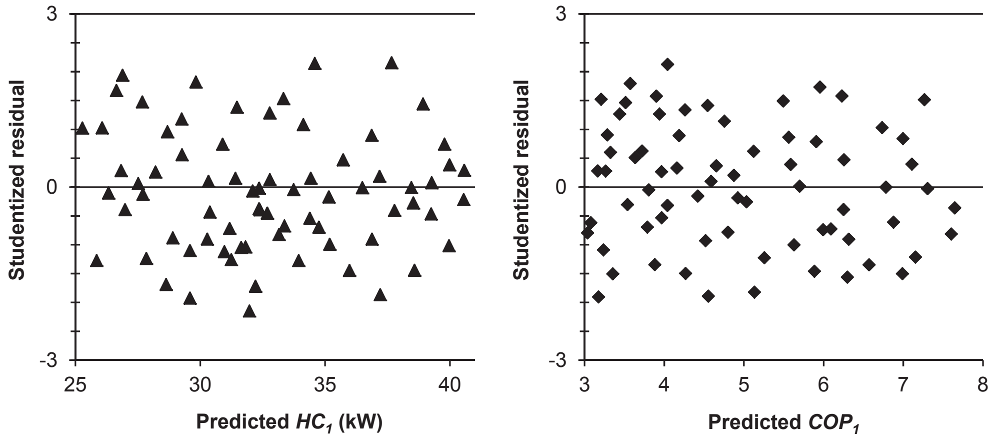

Fig. 7. Plot of residuals versus predicted HC_1 (kW) and COP_1.

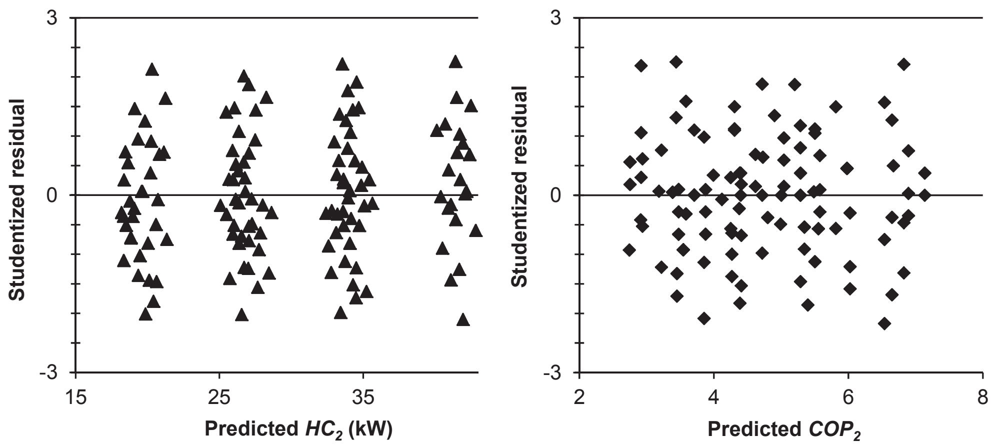

Fig. 8. Plot of residuals versus predicted HC_2 (kW) and COP_2.

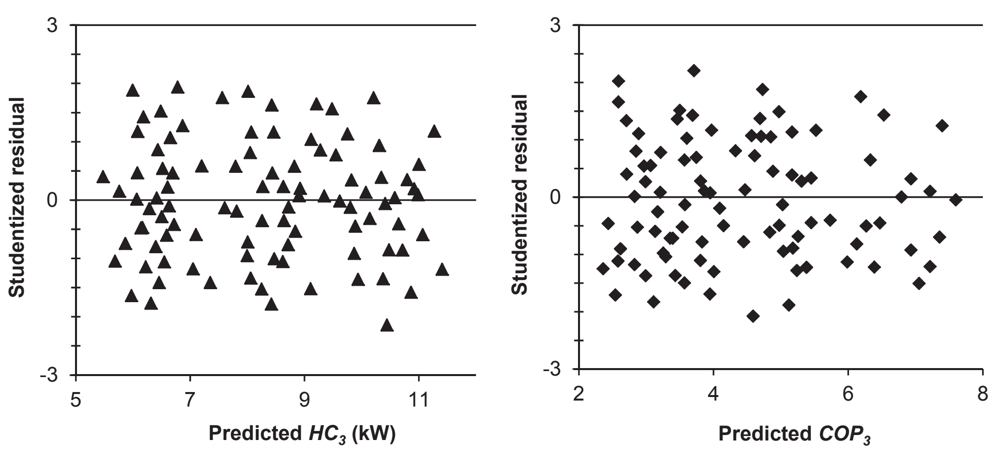

Fig. 9. Plot of residuals versus predicted HC_3 (kW) and COP_3.

tables.

Using multiple regression calculation demonstrates that it is possible to rapidly create equation linking different variables, such a model being representative of the GSHP global behavior with acceptable accuracy and applicable over the entire solution space. However, the method consisting in developing the inverse model can be time consuming for field engineers. Moreover, there is no particular need to determine the performance under specific operating conditions, when it comes to heat-pump selection for a particular building, the equipment operating range and capacity being generally sufficient. This is why suppliers provide performance data listed in tables instead of offering MR models in their catalogue. It is more practical for engineers to utilize these tables rather than mathematical models for their professional use.

5. Conclusion

GSHP equipment manufacturers often provide performance data in the form of tables which are not convenient for higher level system modeling and simulation purposes. In this paper a simplified method for the performance prediction of GSHPs based on MR modeling from such data is presented. The mathematical models estimate the heat capacity and *COP* at particular operating secondary fluids temperature and flow rate conditions based on manufacturer data tables. The proposed operational procedure, which consists in the identification of second-order MR models containing eight statistically significant *x*-variables from a sample of 36 observations taken from the manufacturer table, was successfully validated in the statistical analysis. Predicted performance results are in good agreement with the remaining of observed data, the external prediction errors reaching 0.21%, 0.95% and 1.02% for *HC* predictions, and 2.62%, 4.91% and 3.21% for *COP* predictions for HP1, HP2 and HP3 respectively. The external residual errors as plotted against the model responses showed no correlation with the regression coefficients or the prediction response, thus validating the models. The operational approach appears to be a reliable tool to be incorporated in dynamic simulation codes developed by engineers, as the method is applicable to any GSHP catalogue data.

The study shows that global GSHP behavior can be predicted when fitting MR models from a limited number of observation data. Nevertheless, the limitation to a smaller data set for model identification may remain questionable when the full data set is available in the manufacturer catalogue. Results indicate that manufacturers do not need to provide tables with such a large amount of data points to specify the complete performance map of GSHP. On the other hand, the experimental design evaluations provide insights into how experiments in actual operating GSHP systems should be conducted.

Results from the proposed parsimonious MR models point out that manufacturer data may be generated from proprietary regression models. Since no goodness-of-fit of the latter is provided, some caution must be exercised in the use of such data and corresponding models identified.

Acknowledgements

This work was supported in Chile by the projects CONICYT/FONDAP nº15110019 (SERC-CHILE) and by the UAI Earth Research Center.

References

[1] O. Kwon, K. Bae, C. Park, Cooling characteristics of ground source heat pump with heat exchange methods, Renew. Energy 71 (2014) 651–657.
[2] A. Girard, E.J. Gago, T. Muneer, G. Caceres, Higher ground source heat pump COP in a residential building through the use of solar thermal collectors, Renew. Energy 80 (2015) 26–39.
[3] R. Gupta, R. Irving, Development and application of a domestic heat pump model for estimating CO_2 emissions reductions from domestic space heating, hot water and potential cooling demand in the future, Energy Build. 60 (2013) 60–67.
[4] H. Benli, Energetic performance analysis of a ground-source heat pump system with latent heat storage for a greenhouse heating, Energy Convers. Manag. 52 (2011) 581–589.
[5] Y. Feng, X. Chen, X.F. Xu, Current status and potentials of enhanced geothermal system in China: a review, Renew. Sustain. Energy Rev. 33 (2014) 214–223.
[6] G. Florides, K. Soteris, Ground heat exchangers — a review of systems, models and applications, Renew. Energy 32 (15) (2007) 2461–2478.
[7] X.Q. Zhai, M. Qu, X. Yu, Y. Yang, R.Z. Wang, A review for the applications and integrated approaches of GCHP systems, Renew. Sustain. Energy Rev. 15 (2011) 3133–3140.
[8] J.D. Spitler, Ground-source heat pump system research — past, present and future, HVAC&R Res. 11 (2005) 165–167.
[9] S. Karytsas, H. Theodoropoulou, Public awareness and willingness to adopt ground source heat pumps for domestic heating and cooling, Renew. Sustain. Energy Rev. 34 (2014) 49–57.
[10] N. Zhu, P. Hu, L. Xu, L. Jiang, F. Lei, Recent research and applications of ground source heat pump integrated with thermal energy storage systems: a review, Appl. Therm. Eng. 71 (1) (2014) 142–151.
[11] X. Zhang, G. Huang, Y. Jiang, T. Zhang, Ground heat exchanger design subject to uncertainties arising from thermal response test parameter estimation, Energy Build. 102 (2015) 442–452.
[12] A. Safa, A. Fung, R. Kumar, Comparative thermal performances of a ground source heat pump and a variable capacity air source heat pump systems for sustainable houses, Appl. Therm. Eng. 81 (2015) 279–287.
[13] T. Sliwa, J. Kotyza, Application of existing wells as ground heat source for heat pumps in Poland, Appl. Energy 74 (2003) 3–8.
[14] X. Wang, M. Zheng, W. Zhang, S. Zhang, T. Yang, Experimental study of a solar assisted ground-coupled heat pump system with solar seasonal thermal storage in severe cold areas, Energy Build. 42 (2010) 2104–2110.
[15] A.M. Omer, Ground-source heat pumps systems and applications, Renew. Sustain. Energy Rev. 12 (2008) 344–371.
[16] ANSI/ARI/ASHRAE/ISO 13256-2, Water-source Heat Pumps — Testing and Rating for Performance — Part 2: Water-to-water and Brine-to-water Heat Pumps, ANSI/AHRI/ASHRAE/ISO, 1998.
[17] J. Kim, J.C. Jang, E.C. Kang, K.C. Chang, E.J. Lee, Y. Kim, Verification study of a GSHP system manufacturer data based modeling, Renew. Energy 54 (2013) 55–62.
[18] N.R. Draper, H. Smith, Applied Regression Analysis, third ed., Wiley-Interscience, New York, 1998.
[19] J. Neter, M.H. Kutner, C.J. Nachtsheim, W. Wasserman, Applied Linear Statistical Models, fourth ed., Mc Graw-Hill/Irwin, Chicago, 1996.
[20] G. James, D. Witten, T. Hastie, R. Tibshirani, An Introduction to Statistical Learning with Applications in R, Springer, New York, 2013.
[21] R.A. Johnson, D.W. Wichern, Applied Multivariate Statistical Analysis, Pearson, Upper Saddle River, New Jersey, 2007, ISBN 9780135143506.