

PAPER • OPEN ACCESS

## Using Machine Learning to estimate the technical potential of shallow ground-source heat pumps with thermal interference

To cite this article: Alina Walch *et al* 2021 *J. Phys.: Conf. Ser.* **2042** 012010

View the [article online](#) for updates and enhancements.

### You may also like

- [Low-carbon Technology Application in Wuhan Based on Ground Source Heat Pump System](#)  
Hongmin Li and Zuxu Zou
- [Integration of a heat exchanger on the supply air with the ground-source heat pump in a passive house – case study](#)  
B Radomski, L Drojetzki and T Mróz
- [Collocated geothermal resources of the South of Western Siberia \(Russia\) and prospects of their use](#)  
A A Shevchenko, O V Skrobot, E G Bord et al.



**IOP | ebooks™**

Bringing together innovative digital publishing with leading authors from the global scientific community.

Start exploring the collection—download the first chapter of every title for free.

# Using Machine Learning to estimate the technical potential of shallow ground-source heat pumps with thermal interference

Alina Walch<sup>1</sup>, Roberto Castello<sup>1</sup>, Nahid Mohajeri<sup>2</sup>, Agust Gudmundsson<sup>3</sup>, Jean-Louis Scartezzini<sup>1</sup>

<sup>1</sup> Solar Energy and Building Physics Laboratory, EPFL, Lausanne, Switzerland

<sup>2</sup> Institute for Environmental Design and Engineering, UCL, London, UK

<sup>3</sup> Department of Earth Sciences, Royal Holloway University of London, Egham, UK

**Abstract.** The increasing use of ground-source heat pumps (GSHPs) for heating and cooling of buildings raises questions regarding the technical potential of GSHPs and their impact on the temperature in the shallow subsurface. In this paper, we develop a method using Machine Learning to estimate the technical potential of shallow GSHPs, which enables such an estimation for Switzerland with limited data and computational resources. A training dataset is constructed based on meteorological and geological data across Switzerland. We analyse correlations and the importance of each of the input data for estimating the GSHP potential and compare different input feature sets and Machine Learning models. The Random Forest algorithm, trained on the full dataset, provides the best performance to estimate the GSHP potential. The resulting model yields an  $R^2$  score of 0.95 for the annual energy potential, 0.86 for the heat extraction rate, and 0.82 for the potential number of boreholes per GSHP system.

## 1. Introduction

The extraction of shallow geothermal energy using ground-source heat pumps (GSHPs) of depths up to 400 m is growing worldwide. Switzerland is the country with the world's highest density of shallow geothermal installations, most of which are GSHPs with vertical borehole heat exchangers (BHEs) [1]. A dense installation of BHEs may, however, lead to an excessive long-term cooling of the subsurface due to thermal interactions between boreholes. Thus, thermal interaction must be considered when quantifying the technical potential of BHEs, that is, the long-term maximum energy which can be extracted annually using GSHPs [2,3]. While analytical methods exist to quantify this potential, they are computationally intensive and require the availability of high-resolution data of the thermal properties of the shallow ground. These limitations may hinder the use of analytical models to quantify the technical potential of shallow GSHPs at national scale. By contrast, Machine Learning (ML) can handle missing data in analytical models and, once trained, is computationally highly efficient [4].

Here we use ML in combination with geospatial tools for data pre-processing to estimate the technical potential of BHEs. More specifically, we optimise the set of features and the ML design to (i) use available data from the entire Switzerland and (ii) estimate the technical GSHP potential with high accuracy. For this, we first design the input dataset for the ML model, using an existing regional-scale study of shallow GSHP potential based on analytical modelling [2] as input. Second, we analyse the feature set and compute the importance of all features in the dataset for predicting the technical potential and compare the predictions to the analytical model [2]. Third, we train four ML algorithms (K-Nearest Neighbours, Support Vector Machine, Random Forest, Artificial Neural Network) on different feature sets and compare their performance for estimating the technical geothermal potential.



## 2. Methods and data

### 2.1. Dataset description

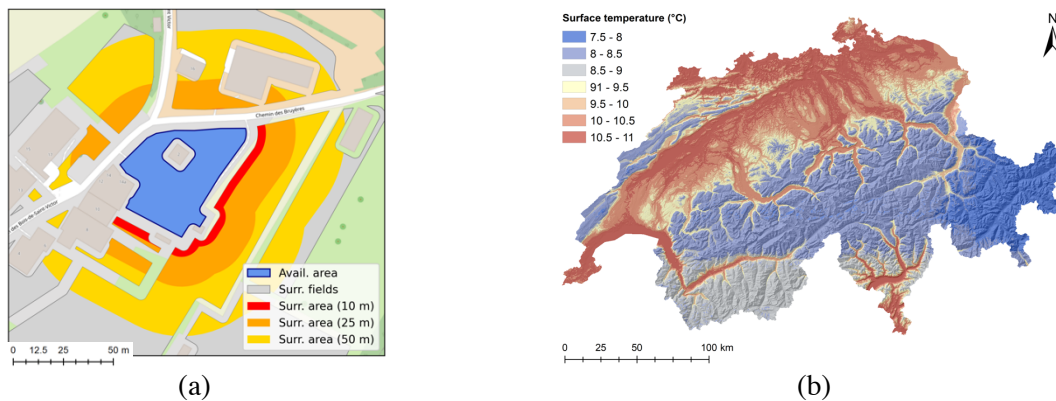
The dataset used to estimate the technical potential of GSHP consists of several groups of input data (features) and outputs (targets), as shown in Table 1. To design a well-performing ML model which can be applied anywhere in Switzerland and does not require computationally intensive geo-spatial processing steps, each of the feature groups is pre-processed separately. The borehole arrangements (the number of boreholes  $N_{BHE}$  and depth  $H$ ) are sampled with replacement from a regional-scale case study in the Swiss Cantons of Vaud and Geneva [2]. These two Cantons contain a large variety of borehole field geometries and cover urban as well as rural areas, so that the data is considered sufficiently representative for Switzerland. As a virtual installation of BHEs, performed in [2], may not be feasible at country scale, the ML features consist only of the available areas for GSHP systems (borehole fields), which can be efficiently computed using geospatial tools [2]. These include the area of a field itself ( $A_{field}$ ) and the area surrounding a field within 10 ( $S_{10}$ ), 25 ( $S_{25}$ ) or 50 m ( $S_{50}$ ), as shown in Figure 1a. Surrounding areas are taken into account as the presence of nearby boreholes may significantly impact the performance of a GSHP system, especially in small fields [2].

The geological properties, which include the thermal conductivity  $\lambda$  (at depths of 50, 100, 150, and 200 m [2]) and thermal diffusivity  $\alpha$  of the ground material, should reflect the thermal characteristics found across Switzerland. Using the values tabulated in the Swiss geothermal norm SIA 384/6 [5], we randomly sampled values for  $\lambda$  in the range of 1 – 5.3 W/mK. To obtain realistic variations of  $\lambda$  with BHE depth, we set the mean  $\lambda$  to the sampled values and keep the variation in  $\lambda$  of the original borehole fields. As the sensitivity of  $\alpha$  in the modelling is small, we leave the values for  $\alpha$  unchanged. The features determined by the meteorological conditions, namely surface temperature  $T_0$ , operating time  $t_{op}$  and maximum seasonal load  $w_{HDD}$  of GSHP systems, are all related to altitude. To obtain realistic combinations of these, we hence randomly sample a set of locations across Switzerland and use the  $T_0$  (see Figure 1b),  $t_{op}$  and  $w_{HDD}$  attributed to each location. Here we consider only altitudes below 2000 m, as few buildings in Switzerland are located above 2000 m.

The different pre-processed datasets are combined to create a synthetic dataset of 100,000 GSHP systems, to which we apply the analytical model proposed by Walch et al. [2]. The model simulates the large-scale heat extraction from shallow GSHPs of up to 200 m depth, accounting for potential thermal interaction between neighbouring boreholes. The heat extraction is simulated across a range of possible borehole arrangements, from which an optimized arrangement is derived such as to maximize the heat extraction while sustaining a minimum feasible operating power [2]. In this work, we estimate all variables describing the optimized GSHP systems, namely the annual heat extraction ( $Q_{pot}$ ), the heat extraction rate ( $q_{max}$ ) and the borehole arrangement ( $N_{BHE}, H$ ).

**Table 1.** Overview of input features (groups 1-3) and targets used to model GSHP performance.

Data group	ID	Description	Range/Values	Source
Borehole arrangements	$A_{field}$	Available area of building parcel	Fields in VD/GE	Walch et al. [2]
	$S_{10,25,50}$	Surrounding avail. area in 10, 25, 50 m	Fields in VD/GE	Walch et al. [2]
Geological properties	$\lambda_{50-200}$	Th. conductivity, at depths 50 – 200 m	1 – 5.3 W/mK	SIA [5]
	$\alpha$	Thermal diffusivity (average)	0.6 – 1.4 $\mu\text{m}^2/\text{s}$	Walch et al. [2]
Meteorological conditions	$T_0$	Surface temperature, 1m below ground	7.5 – 11 °C	Assouline et al. [4]
	$t_{op}$	Nominal operating time for heating	1800 – 2800 h	SIA [5]
	$w_{HDD}$	Maximum heating load (see [2])	0.11 – 0.22	Derived from [6]
Targets	$Q_{pot}$	Energy potential (kWh/a)	ML output	
	$q_{max}$	Heat extraction rate (W/m)	ML output	
	$N_{BHE}$	Number of installed boreholes	ML output	
	$H$	Borehole depth (m)	Computed as $H = \frac{Q_{pot}}{q_{max} t_{op} N_{BHE}}$ [2]	

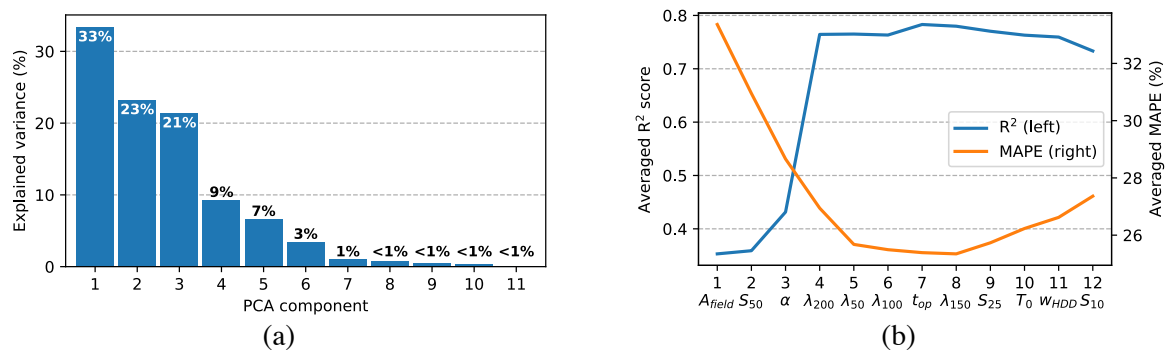


**Figure 1.** Feature pre-processing steps, namely (a) computation of available area for BHE installations (fields) and surrounding areas for each field within 10, 25 and 50 m distance from the installation, and (b) national mapping surface temperature [4] as one of the meteorological conditions (see Table 1).

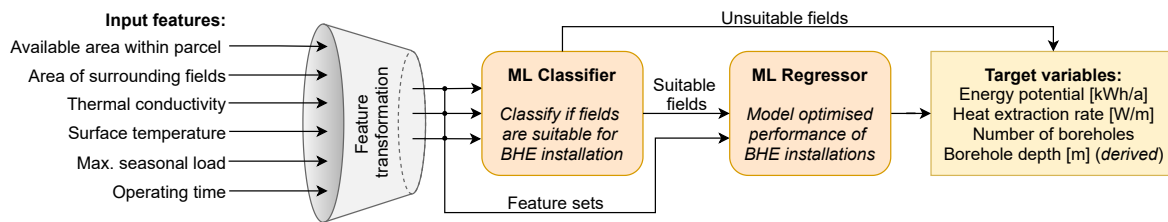
2.2. Feature analysis and transformation

Feature analysis is performed to understand correlations between the input features and the target. Strong correlations between features, such as those between the  $\lambda_{50-200}$ , between the features related to meteorological conditions and between those related to the borehole arrangement, may negatively affect the performance of some ML algorithms. We thus apply principal component analysis (PCA) to obtain a feature set of independent components, without losing significant variance in the features. As Figure 2a shows, the first three PCA components alone account for 77% of the variance within the feature set, while the last 4 components account for less than 1% of variance. From the PCA we hence derive four feature sets for the ML model design, using the first 3 (PCA\_3), 5 (PCA\_5), 7 (PCA\_7) and all PCA (PCA) components, explaining 77%, 93%, 97% and 100% of variance, respectively.

We further assess the importance of the features towards the target prediction by means of recursive feature elimination (RFE). This method trains an ML model (k-nearest neighbour, KNN), first with all features, and then recursively eliminates the feature which contributes least to the model’s performance, measured using the coefficient of determination,  $R^2$ . Figure 2b shows the  $R^2$  score and mean absolute percentage error (MAPE) for the RFE, indicating also the feature eliminated at each iteration (starting from the right). The KNN algorithm is used in this analysis due to its speed and robust performance. The figure shows that the  $R^2$  is high with > 4 features and reaches its maximum for 7 features, where also the MAPE is nearly minimal. We hence create three more feature sets for model testing, using the first 4 (FTR\_4), 7 (FTR\_7) and all features (FTR). We further standardize the features and log-transform  $A_{field}$  and  $S_{10,25,50}$ , as these features span a logarithmic range.



**Figure 2.** Feature importance measures. (a) Explained total variance by the PCA components, and (b)  $R^2$  score and MAPE obtained from recursive feature elimination using KNN, showing the number of considered features and the eliminated feature at each iteration on the x-axis.



**Figure 3.** Conceptual design of the Machine Learning framework.

### 2.3. Machine Learning model design

We apply a two-step approach for the design of the ML framework, as shown in Figure 3. First, we classify the samples into *suitable* and *unsuitable* fields for BHE installations. Second, we estimate the potential of the suitable fields ( $Q_{pot}$ ,  $q_{max}$ ,  $N_{BHE}$ ) using an ML regression model. These targets have been obtained from the analytical model (Section 2.1). From the targets, we compute  $H$  using the formula in Table 1. This two-step approach allows to train the regression model only in the range of the optimized target values, which increases the performance and physical consistency of the results.

For the classification and the regression models, we consider four widely used ML algorithms, namely k-nearest neighbours (KNN), Random Forests (RF), single-layer artificial neural networks (ANN) and Support Vector Machines (SVM), which are described in [7] and implemented using python’s scikit-learn library. To tune the algorithms (to choose their hyper-parameters), we use 5-fold cross-validation (CV) based on randomized search. For this, random hyper-parameters are selected from a set of distributions, CV is performed for each combination of hyper-parameters and the highest-scoring model is retained. We use 60,000 data samples and 200 CV iterations, which results in excessive CV times for the ANN and SVM. We thus use a reduced dataset (see Table 2) for these models, which may cause a small reduction in performance, particularly if smaller samples are used.

As we use a multivariate regression approach, we further consider two ways of constructing the regressors. First, we train one regressor to predict all outputs at once (MO), which is possible for any algorithm supporting native multivariate regression (all except SVM). Second, we train separate models for each target (MM), which can lead to the violation of physics constraints among the targets but may perform better if the importance of features varies significantly between targets. Table 2 shows a comparison of the CV scores for all considered combinations of features and models. The best performance (highlighted cell) is obtained for using a Random Forest (RF), with separate models fitted for each target (MM) and trained on the entire feature set (FTR). The comparison is provided for the regression model only, as this is the most important stage in the modelling of GSHP performance.

For the classification, the same model (RF, FTR) was selected and tuned. As the classes in this classification task are unbalanced (there are many more suitable fields than unsuitable ones), we use the average f1-score as performance evaluation metric, which represents the harmonic mean of the precision (positive predictive value) and recall (true positive rate) and weighs both classes equally.

**Table 2.** Cross-validation scores ( $R^2$ ) for all ML models and feature sets (see Section 2.2 and 2.3). The highlighted field shows the best model (multi-model Random Forest, trained using all features).

Feature set	KNN		RF		ANN		SVM
	MO	MM	MO	MM	MO	MM	MM
PCA_3	0.42	0.42	0.44	0.44	0.40 <sup>a</sup>	0.42 <sup>a</sup>	0.32 <sup>a</sup>
PCA_5	0.78	0.78	0.77	0.79	0.72 <sup>a</sup>	0.79 <sup>a</sup>	0.61 <sup>a</sup>
PCA_7	0.75	0.75	0.76	0.80	0.72 <sup>a</sup>	0.65 <sup>b</sup>	0.54 <sup>a</sup>
PCA	0.73	0.73	0.77	0.82	0.78 <sup>a</sup>	0.67 <sup>b</sup>	0.47 <sup>b</sup>
FTR_4	0.77	0.77	0.81	0.84	0.73 <sup>a</sup>	0.81 <sup>a</sup>	0.49 <sup>a</sup>
FTR_7	0.79	0.79	0.83	0.87	0.78 <sup>a</sup>	0.86 <sup>a</sup>	0.53 <sup>b</sup>
FTR	0.74	0.74	0.81	0.87	0.73 <sup>a</sup>	0.68 <sup>b</sup>	0.48 <sup>b</sup>

<sup>a</sup> Reduced CV iterations (20); <sup>b</sup> Reduced sample size (5,000)

### 3. Results and discussion

The results below report the performance of the best model, as outlined in Section 2, for a test set of 20,000 borehole fields, which were excluded from the cross-validation procedure.

#### 3.1. Classification model

The classification model achieves an overall accuracy of 97% in predicting the suitability of fields for BHE installation. Furthermore, the results in Table 3 show that the performance is better for the *Suitable* class than the *Unsuitable* class, which is expected given the imbalance between the classes. The comparison of precision against recall shows that out of those samples predicted as unsuitable, only 27% are truly suitable, while nearly half of the truly unsuitable fields are predicted as suitable. As these misclassifications occur only for small fields (1-2 BHE), the impact of these misclassifications on the annual total geothermal potential is negligible, increasing the potential estimate by 0.3%.

**Table 3.** Confusion matrix of the classification results, as well as precision, recall and f1-scores. The last row shows the average across both classes and the overall accuracy.

Predicted class:		<i>Unsuitable</i>	<i>Suitable</i>	Precision	Recall	f1-score
True class	<i>Unsuitable</i>	481	455	73%	51%	0.69
	<i>Suitable</i>	177	18887	98%	99%	0.98
Average (macro)		Overall accuracy: 97%		85%	75%	0.79

#### 3.2. Regression model

The performance of the regression model varies between the different target variables, showing the highest  $R^2$ -score for  $Q_{pot}$ , followed by  $q_{max}$  and  $N_{BHE}$  (see Table 4). We further compare the root-mean-squared error (RMSE), mean absolute error (MAE), MAPE and the log-weighted mean-squared error. The MAE of  $Q_{pot}$  of 11,466 kWh/a corresponds to the mean annual production of a single BHE, which is in line with the MAE of  $N_{BHE}$  of 1.5 boreholes. The MAPE is much lower (7%) for  $q_{max}$  than for the other two targets (30-35%), mainly due to deviations for small residential fields, as shown in Figure 4. While Figure 4 shows no systematic bias in the prediction, a small overestimation of  $Q_{pot}$  is observed for the smallest GSHP systems. This implies that the results should be interpreted carefully for small borehole fields. At the large scale, the results however show a very high consistency between targets and predictions, with only 0.1% of deviation for the annual sums of  $Q_{pot}$ .

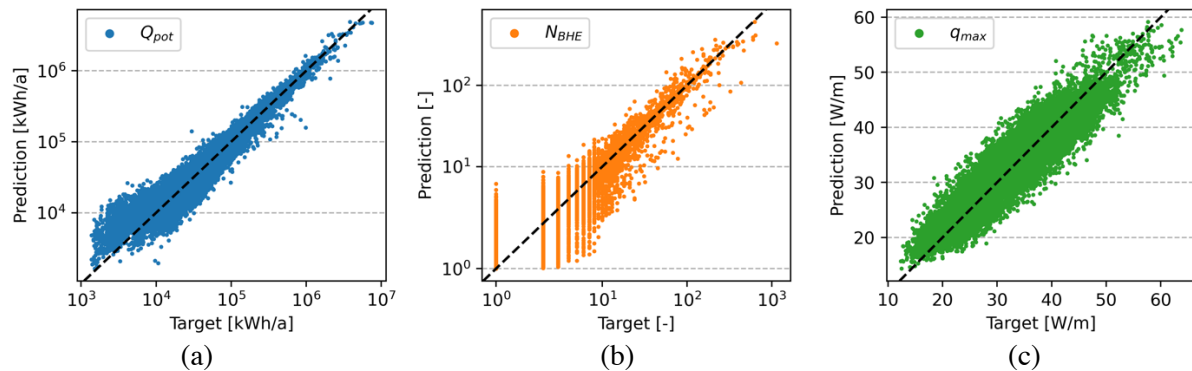
In addition to the three variables analysed previously, we compute  $H$ , the borehole depth, as the fourth variable representing the borehole arrangement. The computed  $H$  is consequently no longer in the discrete set used in the simulation (50-200m), but instead is a continuous variable. Results show that 92.7% of  $H$  are in the simulated range of 50-200m, suggesting a good physical consistency between the estimated target variables.

**Table 4.** Performance comparison of the best model (Random Forest, all features) on the test set for each variable and several error scores ( $R^2$ , RMSE, MAE, MAPE, logMSE).

Target variable	R2	RMSE	MAE	MAPE	logMSE
$Q_{pot}$ [kWh/a]	0.95	47454	11466	0.31	0.13
$N_{BHE}$ [-]	0.82	9.05	1.52	0.35	0.08
$q_{max}$ [W/m]	0.86	3.02	2.37	0.07	0.01

#### 3.3. Computational efficiency

For estimating the GSHP potential in only 30s after a training time of 20 min (single core), the two-stage ML model clearly outperforms the analytical approach, which requires  $\sim 11$  hours for processing 20,000 BHE fields. The computational time of the ML approach is thus dominated by common pre-processing steps (50 min for 20,000 samples), resulting in an overall speed-up by a factor of 10.



**Figure 4.** Scatterplot of predictions against targets for all target variables, namely (a) energy potential  $Q_{pot}$  (logarithmic scale), (b) number of boreholes  $N_{BHE}$  (log. scale) and (c) heat extraction rate  $q_{max}$ .

#### 4. Conclusion

This study presents a Machine Learning method to estimate the technical potential of GSHP systems, their heat extraction rate, and the number of boreholes across a wide range of borehole field geometries, meteorological conditions, and geological properties. Several feature sets and Machine Learning models are compared, suggesting a Random Forest trained separately for each target variable on all available features as the best model, which reduces the computational time by a factor of 10. The results show a high  $R^2$  coefficient of determination for all targets, with mean absolute percentage errors of 7% for the heat extraction rate and of 30-35% for the technical potential and number of boreholes, implying that the results must be carefully interpreted at the scale of single GSHP systems.

As the deviation of the total estimated technical potential from the analytically modelled value is negligible, the proposed ML model appears well-suited for estimating the technical potential of shallow GSHPs at the national scale for Switzerland. To expand this to the national scale, a further validation of the results against real-world examples will be necessary. Further work should also address the physical consistency of the results by simulating the temperature change in the ground for the obtained solutions. Also, the estimation of uncertainties will aid the interpretability of the results. While the uncertainty related to input data must be assessed, the RF allows for the estimation of model uncertainties as standard deviations [7] or quantiles [4].

#### Acknowledgements

This research is supported by the Swiss National Science Foundation (SNSF) through the “HyEnergy” project under the National Research Program 75 (Big Data).

#### References

- [1] Link K 2019 *Statistik der geothermischen Nutzung in der Schweiz – Ausgabe 2018* (Switzerland: EnergieSchweiz)
- [2] Walch A, Mohajeri N, Gudmundsson A and Scartezzini J-L 2021 Quantifying the technical geothermal potential from shallow borehole heat exchangers at regional scale *Renewable Energy* **165** 369–80
- [3] Bayer P, Attard G, Blum P and Menberg K 2019 The geothermal potential of cities *Renewable and Sustainable Energy Reviews* **106** 17–30
- [4] Assouline D, Mohajeri N, Gudmundsson A and Scartezzini J-L 2019 A machine learning approach for mapping the very shallow theoretical geothermal potential *Geotherm Energy* **7** 19
- [5] SIA 2010 *Sondes géothermiques (SIA384/6)* (Swiss Society of Engineers and Architects)
- [6] MeteoSwiss 2017 *Daily Mean, Minimum and Maximum Temperature* (Federal Office of Meteorology and Climatology MeteoSwiss)
- [7] Walch A, Castello R, Mohajeri N and Scartezzini J-L 2019 A Fast Machine Learning Model for Large-Scale Estimation of Annual Solar Irradiation on Rooftops *ISES SWC2019 / SHC2019 Conference Proceedings* 10