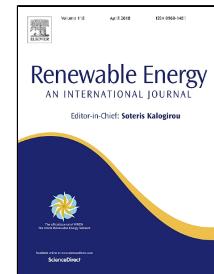


Accepted Manuscript

A city-scale roof shape classification using machine learning for solar energy applications



Nahid Mohajeri, Dan Assouline, Berenice Guiboud, Andreas Bill, Agust Gudmundsson, Jean-Louis Scartezzini

PII: S0960-1481(17)31300-9

DOI: 10.1016/j.renene.2017.12.096

Reference: RENE 9593

To appear in: *Renewable Energy*

Received Date: 31 August 2017

Revised Date: 27 November 2017

Accepted Date: 26 December 2017

Please cite this article as: Nahid Mohajeri, Dan Assouline, Berenice Guiboud, Andreas Bill, Agust Gudmundsson, Jean-Louis Scartezzini, A city-scale roof shape classification using machine learning for solar energy applications, *Renewable Energy* (2017), doi: 10.1016/j.renene.2017.12.096

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

A city-scale roof shape classification using machine learning for solar energy applications

Nahid Mohajeri^{a,b*}, Dan Assouline^a, Berenice Guiboud^a, Andreas Bill^a, Agust Gudmundsson^c, Jean-Louis Scartezzini^a

^aSolar Energy and Building Physics Laboratory (LESO-PB), Ecole Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland, ^bSustainable Urban Development Programme, Department for Continuing Education, University of Oxford, Rewley House, 1 Wellington Square, Oxford OX1 2JA, United Kingdom. *Corresponding author; e-mails: nahid.mohajeri@epfl.ch; nahid.mohajeri@conted.ox.ac.uk

^c Department of Earth Sciences, Royal Holloway University of London, Egham TW20 0EX, United Kingdom

1 Abstract

2 Solar energy deployment through PV installations in urban areas depends strongly on the
 3 shape, size, and orientation of available roofs. Here we use a machine learning approach,
 4 Support Vector Machine (SVM) classification, to classify 10,085 building roofs in relation to
 5 their received solar energy in the city of Geneva in Switzerland. The SVM correctly identifies
 6 six types of roof shapes in 66% of cases, that is, flat & shed, gable, hip, gambrel & mansard,
 7 cross/corner gable & hip, and complex roofs. We classify the roofs based on their useful area
 8 for PV installations and potential for receiving solar energy. For most roof shapes, the ratio
 9 between useful roof area and building footprint area is close to one, suggesting that footprint
 10 is a good measure of useful PV roof area. The main exception is the gable where this ratio is
 11 1.18. The flat and shed roofs have the second highest useful roof area for PV (complex roof
 12 being the highest) and the highest PV potential (in GWh). By contrast, hip roof has the lowest
 13 PV potential. Solar roof-shape classification provides basic information for designing new
 14 buildings, retrofitting interventions on the building roofs, and efficient solar integration on the
 15 roofs of buildings.

17 18 Keywords:

19 Machine learning; Roof shape classification; PV potential; Support Vector Machine

20 21 22 1. Introduction

23
 24 Photovoltaics (PV) are among the most promising emerging technologies for deployment of
 25 solar energy in urban areas. Solar PV panels can be installed on the rooftops of individual

26 buildings. They have proven to be an efficient and viable resource of sustainable renewable
27 energy for the urban areas [1, 2, 3, 4, 5, 6]. However, to reach their full potential for the urban
28 scale many socio-technological and economical decisions need to be made. Improvements are
29 also needed as regards urban characteristics such as design, geometry, and positioning of
30 buildings [7]. While roofs of buildings in urban areas provide promising and increasingly
31 important locations for PV installations, an efficient method for finding and evaluating
32 suitable roofs on existing buildings for the optimal location of PVs remains a challenge.

33 There have been several studies focusing on mapping solar energy potential and
34 feasibility of solar PV installations on building rooftops at the neighborhood and the urban
35 scales using different methods. For example, Wiginton et al. [1] use a sampling technique as
36 well as the GIS-based Feature Analyst (FA) tool to estimate the rooftop PV potential. Several
37 other studies have used aerial images and ArcGIS LiDAR data to determine roof geometries
38 and to estimate their PV potential at a regional scale [3, 4, 5, 6, 8, 9]. Assouline et al. [10]
39 introduced a method that combines Geographic Information Systems (GIS) for data
40 processing and a machine learning approach to estimate the building rooftop solar PV
41 potential for cities in Switzerland. While there are many recent studies on modelling roof
42 geometries for solar application based on simplification of roof shapes [11, 12] and geo-
43 spatial techniques [1, 13, 14], few studies explore the relation between roof shapes of new
44 buildings and their solar potential with a view of providing support for design development
45 [15, 16, 17, 18]. Recently, machine learning algorithms (e.g. Support Vector Machines,
46 Random Forests, artificial neural networks) have been extensively used for different
47 applications. These include forecasting solar radiation [19] as well as geospatial
48 environmental data modelling [20, 21, 22]. However, using machine learning algorithms so as
49 to classify urban characteristics for solar applications, including impacts of different roof
50 shapes on annual solar PV electricity production, has remained a challenge.

51 In the present study, we develop a novel approach using Support Vector Machine (SVM),
52 a kernel-based machine learning technique, to classify roof shapes based on their solar PV
53 potential. We focus on the city of Geneva in Switzerland. The city latitude is 46.2° N, the
54 longitude is 6.15° E, and its average elevation is 374 m. Geneva, with a population of about
55 200 thousand in 2015 (www.bfs.admin.ch), is the second largest city in Switzerland and the
56 largest one in the French-speaking part of the country. There are 11,806 buildings in Geneva.
57 However, all the factors that we analyse were available for only 10,085 buildings, which is
58 thus the number used in this study.

59 The main aims of this study are to: (1) explore the variability of 10,085 building roofs in
60 the city of Geneva; (2) use a machine learning methodology to classify the building roof

61 shapes in the city in relation to their solar potential; (3) rank the roof shapes based on their
62 received solar energy and potential PV electricity production; and (4) classify the building
63 types (e.g. residential, commercial, and industry) based on the roof shape and the solar PV
64 potential.

65

66 2. Methods and data

67

68 To classify the building roof shapes in the city of Geneva based on their solar energy
69 potential, we use a machine learning algorithm, that is, Support Vector Machine (SVM).
70 SVM is one of the most efficient machine learning algorithms for classification tasks. It was
71 first introduced by Cortes and Vapnik in 1995 [23]. Section 2.1 gives a short description of
72 the main principles of the method. In Section 2.2, we explain briefly the data and the
73 estimation method for yearly mean solar radiation on the rooftops. We use yearly mean solar
74 radiation data, which is freely available for building rooftops of the city of Geneva from
75 (<http://ge.ch/sitg/>).

76

77 2.1 Support Vector Machine for classification

78

79 Support Vector Machines (SVMs) are primarily an example of a linear classifier. They can,
80 however, be extended to a non-linear classifier with kernel functions [20, 23, 24, 25, 26, 27].
81 SVM is a machine learning technique based on the concept of decision planes (hyperplanes)
82 which define the decision boundaries of the classifier. It was originally used to solve a binary
83 classification problem. The principle of the algorithm for a binary classification, however, can
84 be easily extended to a multi-class classification problem. We consider a set of training data
85 with two classes of points C_1 and C_2 . The goal of a classification task is to design a function
86 f that can assign any new point x to either C_1 or C_2 , using the training data. One way to
87 design such a function is to find the boundary between the classes, separating the two
88 different classes with a linear decision surface. This is a line in 2D, a plane in 3D, or a
89 hyperplane in higher dimensions.

90 The basic idea of SVM is to find the optimal separating hyperplane between the two
91 classes (C_1 and C_2) by maximizing the margin between the closest points of the classes
92 (Fig.1). The points lying on the boundaries are called support vectors and the middle of the
93 margin is defined by the optimal separating hyperplane (Fig. 1). To start from the simplest
94 case, suppose that this separation is possible without misclassifications. Then the data set is

95 called linearly separable. Since real data are usually not linearly separable, this assumption is
 96 relaxed to obtain a linear classifier that allows for classification errors and is known as a soft
 97 margin classifier.

98 Consider the following basic set of linear functions:

99

100 $f(x) = \mathbf{w} \cdot \mathbf{x} + b$ (1)

101

102 where \mathbf{x} is a vector in R^d (input space of dimension d), b is a scalar, and \mathbf{w} is an unknown
 103 vector in R^d , that is, a weight vector or normal (perpendicular) to the hyperplane to be
 104 optimized. In a classification task, the sign of the function $f(x)$ is the output of the classifier.

105 We consider the data to be linearly separable. Thus, it is a hyperplane that separates class C_1
 106 from class C_2 (Fig. 1a). The hyperplane can be defined by $\mathbf{w} \cdot \mathbf{x} + b = 0$. For the hyperplane
 107 to be as far from the support vectors as possible, we need to maximize the margin. A simple
 108 computation leads to a margin of $\frac{2}{\|\mathbf{w}\|}$, where $\|\mathbf{w}\|$ is the Euclidean norm of \mathbf{w} , a measure of

109 its length. As mentioned above, SVMs are originally a kind of linear classifier. Consider the
 110 following decision function, defined by \mathbf{w} and b [21, 26, 27]:

111

112 $y = \begin{cases} +1, & \text{if } \mathbf{w} \cdot \mathbf{x} - b \geq 1 \\ -1, & \text{if } \mathbf{w} \cdot \mathbf{x} - b \leq -1 \end{cases}$ (2)

113

114 Given a training data $\{(\mathbf{x}_i, y_i)\}$ ($i = 1, 2, \dots, L$), L is the number of training points, where
 115 $\mathbf{x}_i \in R^d$ is the input vector of dimension d (the number of features) and $y_i \in R$ is the desired
 116 output value (called the label) for point i, we impose the above decision function, Eq. (2), as a
 117 set of constraints on the training data:

118

119 $\mathbf{w} \cdot \mathbf{x}_i + b \geq +1, \text{ if } y_i = +1$ (3)

120 $\mathbf{w} \cdot \mathbf{x}_i + b \leq -1, \text{ if } y_i = -1$ (4)

121

122 The two set of equations can be combined into:

123

124 $y_i (\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1$ (5)

125

126 Equation (5) is the constraint of our optimization problem. As mentioned above, the objective
 127 is to maximize the margin $\frac{2}{\|\mathbf{w}\|}$, representing the distance between the two dashed lines in Fig.
 128 1a, subject to the constraint in Eq.(5), which is equivalent to find:

129

130 $\min \mathbf{w}$ such that $y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1$

131

132 Minimizing \mathbf{w} is equivalent to minimizing $\frac{1}{2}\|\mathbf{w}\|^2$. Using this term makes it possible to
 133 perform Quadratic Programming (QP) optimization later on. We therefore need to [21, 23, 24,
 134 25, 26, 27]:

135

$$\begin{aligned} & \text{Minimize} && \frac{1}{2}\|\mathbf{w}\|^2 \\ & \text{subject to} && y_i(\mathbf{w} \cdot \mathbf{x}_i + b) - 1 \geq 0 \end{aligned} \quad (6)$$

137

138 We use the Lagrangian formulation to solve this optimization problem. This is a general
 139 analytical method to solve constrained optimization problems. Introducing the Lagrange
 140 multipliers α_i for the constraints, one has to minimize the Lagrange functional (Lagrangian)
 141 L_P with respect to \mathbf{w} and b and maximize the Lagrangian with respect to α_i [21, 26]:

142

$$L_P = \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^L \alpha_i y_i (\mathbf{w} \cdot \mathbf{x}_i + b) + \sum_{i=1}^L \alpha_i \quad (7)$$

144

145 subject to the following constraints:

146

$$147 \alpha_i \geq 0, \quad i = 1, 2, \dots, L \quad (8)$$

148

149 This can be done by differentiating L_P with respect to \mathbf{w} and b and setting the derivatives to
 150 zero [21, 26]:

151

$$152 \frac{\partial L_P}{\partial \mathbf{w}} = 0 \Rightarrow \mathbf{w} = \sum_{i=1}^L \alpha_i y_i \mathbf{x}_i \quad (9)$$

153

$$154 \frac{\partial L_P}{\partial b} = 0 \Rightarrow \sum_{i=1}^L \alpha_i y_i = 0 \quad (10)$$

155

156 Substituting Eqs. (9) and (10) into Eq. (7) gives a new formula which depends on α and we
 157 need to maximize, namely [21, 26]:

158

$$159 L_D = \sum_{i=1}^L \alpha_i - \frac{1}{2} \sum_{i=1}^L \sum_{j=1}^L \alpha_i \alpha_j y_i y_j \mathbf{x}_i \cdot \mathbf{x}_j \quad (11)$$

160

161 subject to the following constraints:

162

$$163 \sum_{i=1}^L \alpha_i y_i = 0, \quad \alpha_i \geq 0, \quad (12)$$

164

165 Here, L_D refers to as the *Dual* form of the *Primal* L_P . The Dual form requires only the dot
 166 product of each input vector \mathbf{x}_i to be calculated, which is important for the *Kernel Trick*
 167 described in the next part. Thus, the decision function becomes:

$$168 f(\mathbf{x}) = \sum_{i=1}^L y_i \alpha_i \mathbf{x}_i \cdot \mathbf{x} + b \quad (13)$$

169 This is a convex quadratic problem with respect to linear constraints which can be solved with
 170 Quadratic Programming (QP). Any well-known method can be applied to solve it [28].

171 In order to extend the SVM methodology to handle data that is not fully linearly
 172 separable, we relax the constraints for Eq. (5) slightly so as to allow for misclassified points.
 173 This is done by introducing positive slack variables $\xi_i \geq 0$ to the constraints (5) [23, 26]:

174

$$175 y_i (\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 - \xi_i \quad (14)$$

176

177 As few ξ_i as possible should be non-zero. To minimize the overall error brought by the slack
 178 variables, the optimization problem becomes:

179

$$180 \text{Minimise} \quad \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^L \xi_i \\ 181 \text{Subject to} \quad y_i (\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 - \xi_i \quad (15)$$

182

183 where the parameter C controls the trade-off between the slack variable penalty and the size
 184 of the margin. More specifically, the constant C determines the trade-off between the

185 generalization of the classifier and the amount of outliers tolerated (if C is too big the data
 186 will be overfitted and the capacity to adapt to a new data will be low. However, if C is too
 187 small it might adapt well but will not have enough memory of the training data to classify
 188 well). For this problem, the same approach of Lagrangian multipliers is used as for linearly
 189 separable data. However, one has to introduce the Lagrange multipliers for the constraints
 190 $\xi_i \geq 0$ as well in order to arrive at the dual formulation of the problem. The resulting
 191 optimization is also a QP-problem and can be solved by standard QP solvers. The trade-off
 192 constant C becomes an upper bound for the weights in the dual formulation, resulting in the
 193 following constraints: $0 \leq \alpha_i \leq C, i = 1, 2, \dots, L$.

194 To move from a linear classifier to a non-linear classifier (Figs. 1b, 1c), a set of
 195 mathematical functions, kernels, are defined. The performance of SVM depends heavily on
 196 the choice of the kernel function, that is, $k(\mathbf{x}_i, \mathbf{x}_j)$. Using the kernel function, we implicitly
 197 map the input data to a so-called feature space, where f can be defined as a linear function.
 198 We compute $k(\mathbf{x}_i, \mathbf{x}_j) = \langle \varphi(\mathbf{x}_i), \varphi(\mathbf{x}_j) \rangle$, where $k(\mathbf{x}_i, \mathbf{x}_j)$ is the kernel function and φ is a
 199 mapping from the original input space to the feature space [29, 30, 31]. The feature space is
 200 constructed via the projection, that is, kernel trick. Only $k(\mathbf{x}_i, \mathbf{x}_j)$ needs to be computed and
 201 not the mapping φ . After replacing the inner product with the kernel in Eq. (13), the solution
 202 for f is as follows:

$$203 \quad f(\mathbf{x}) = \sum_{i=1}^L \alpha_i k(\mathbf{x}_i, \mathbf{x}) + b \quad (16)$$

204 There are number of kernels that can be used in Support Vector Machines models. These
 205 include, for example, Linear, Polynomial, Radial Basis Function (RBF), and Sigmoid. The
 206 RBF is by far the most popular choice of kernel types used in Support Vector Machines, due
 207 to its high performance. The non-linear radial basis function kernel is defined by Eq. (17)
 208 [21]:

$$209 \quad k(\mathbf{x}_i, \mathbf{x}_j) = e^{-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}} \quad (17)$$

210
 211 where $\|\mathbf{x}_i - \mathbf{x}_j\|^2$ is the squared Euclidean distance between two feature vectors namely, \mathbf{x}_i
 212 and \mathbf{x}_j , σ is a parameter of the Gaussian RBF which defines the width of the ‘bell’ curve. In
 213 order to choose the optimal parameter configuration, we perform Cross Validation (CV) on
 214 the training data using grid-search. The grid search considers a range of values for each

215 parameter configuration and the best parameter configuration is the one that offers the lowest
216 cross validation error. K -fold cross validation is used. In K -fold CV, the training data points
217 are randomly partitioned into k equal-sized parts. $k-1$ parts are used for training and the
218 remaining part for validation, considering all different permutations. We estimate the cross-
219 validation error by averaging the errors that are obtained from each permutation.

220

221 *2.2 Solar PV electricity production on rooftops*

222

223 Yearly mean solar radiation for the rooftops of more than 10,000 buildings (city scale) has
224 been estimated, using fine resolution LiDAR point data, by SITG (Le système d'information
225 du territoire à Genève) [32, 33] and which is freely accessible from <http://ge.ch/sitg/>. We used
226 solar radiation analysis tools in Geographic Information System (GIS), in combination with
227 MATLAB, to model solar potential on building roofs at building and neighbourhood scales
228 to verify the city-scale solar model provided by SITG. Figure 2 shows a visualisation of the
229 solar potential availability on the rooftops in three different scales, namely building scale (Fig.
230 2a), neighbourhood scale (Fig. 2b), and city scale (Fig. 2c). The solar PV electricity potentials
231 for rooftops have been estimated as follows: (i) LiDAR point data are used to make a Digital
232 Surface Model (DSM) with precision of 0.5 m x 0.5 m which includes buildings with their
233 different roof shapes as well as details on the landscape. (ii) DSM is then used to extract the
234 roof area, slope, aspect, and orientation of roof surfaces which will be used in the calculation
235 of solar potential. (iii) The average 20 years meteorological data (e.g. solar radiation, cloud
236 cover, temperature) from 1980 to 2000 is obtained from Meteonorm
237 (<http://www.meteonorm.com/>) in order to estimate the solar radiations components (direct-
238 beam, sky-diffuse, and ground reflected) on the tilted surfaces. (iv) The calculations of solar
239 radiation components are performed in MATLAB using Hay anisotropic model for sky-
240 diffuse radiation [34, 35]. (v) Shading co-efficient on the direct and diffuse radiation
241 components is taken into account [32, 33]. (vi) Useful roof areas for PV installation are
242 defined using the following thresholds: (a) the superstructures from roof surfaces are
243 removed; (b) only surface having mean solar radiation ≥ 1000 [kWh/m²] are considered [7,
244 36]; (c) a 1-m-buffer is generated around each remaining roof surface [10]; (d) only roof areas
245 larger than 5m² are considered, the minimum useful areas obtained after applying thresholds
246 a, b, c; (vii) The energy output, that is, the maximum potential of solar PV electricity
247 production on the rooftops is estimated through multiplying the annual mean solar radiation
248 (GWh/year) and the useful roof area for PV installation, taking into account the average

249 efficiency (20%) and the 79% performance ratio for mono-crystalline PV panels
250 (<http://ge.ch/sitg/>).

251

252 **3. Roof characteristics and solar potential**

253

254 We analysed the roofs in terms of slopes and facing directions or aspects for all available
255 buildings of the city of Geneva (Fig. 2c). We also analysed the mean annual solar radiation
256 received by the roofs as a function of roof slope and aspect (Fig. 3). Considering first the
257 slope, the results show that the most common roof slopes are in the sectors of 10-20° and 30-
258 40°, with the sector of 20-30° being a close second (Fig. 3a). Generalising, we show that
259 slopes between 10° and 40° are by far the most common in the city of Geneva. Conversely,
260 slopes in the range 70-80° are rare, and so are slopes in the ranges 60-70° and 0-10°. The
261 latter represents flat roofs, which are thus comparatively rare in Geneva.

262 The effect of roof slope on received annual solar radiation is shown in Fig. 3b. The results
263 show that the flat to gently sloping roofs, those sloping 0-40°, receive the greatest solar
264 radiation. In fact, there is a clear inverse relationship between roof slope and received solar
265 radiation: the greater the slope the less the received radiation. Thus, the steepest sloping roofs,
266 with slopes of 70-80°, receive the least amount of solar radiation.

267 Considering next the roof aspect, the results show that south-facing roofs are the most
268 common, followed by roofs facing either southeast or southwest (Fig. 3c). As expected, the
269 roofs facing south receive the greatest solar radiation (Fig. 3d). These are followed by roofs
270 facing west and east, with the roofs facing, northwest, northeast and, in particular, north
271 receiving the least radiation, as expected.

272 These results can now be combined so that the monthly and yearly received solar
273 radiations are shown as functions of both roof slope and aspect or facing direction (Fig. 4). In
274 Fig. 4a and Fig. 4b we show monthly and annual mean solar radiation (kWh/m²) for a typical
275 neighbourhood in Geneva (see Fig. 2b for the location of neighbourhood). In Fig. 4c we only
276 show the annual mean solar radiation (kWh/m²) as a function of slope and aspect of the roof
277 surfaces for all the studied buildings in the city of Geneva. The illustrations demonstrate very
278 well the following principal result: Figures 4a shows that in winter only south-facing and
279 inclined roofs (with slopes up to 40 degrees) get high values. On approaching summer the
280 gently sloping roofs become more favourable and the potential more equally distributed over
281 all aspects. In June and July, roofs with slopes up to 30 to 40 degrees reach almost the same
282 solar radiation irrespective of orientation (aspect).. Fig. 4b and Fig. 4c show that the most
283 favourable surfaces have a slope of up to 40 degrees and an aspect between southeast, south

284 and southwest. Remarkably, for north-facing roofs with gentle slopes the potential is not that
285 much different from gently sloping east and west oriented roofs. The steeper the slope, the
286 lower the values for annual radiation although for southeast and southwest aspects, even roofs
287 with slopes up to 70 degrees get radiation values of 800kWh/m² and more.

288 Several studies for Geneva [37, 38], and generally for Switzerland [39], show that a tilt
289 angle between 30° and 40° is commonly considered as optimal roof slope for installing solar
290 panels. For example, for the city of Zürich, located at latitude 47°N, an optimum angle of 30°
291 for the tilt of the roof is reported by [39]. Similarly, for the city of Geneva, located at latitude
292 46°N, the optimal tilt angle for installation of PV on rooftops is regarded as 30 degrees [37,
293 38]. The knowledge of the optimum tilt angle is very important in order to obtain the highest
294 possible annual or seasonal solar energy yield. There have been several studies analysing the
295 correlation between the optimal angle for a fixed Building Integrated Photovoltaic (BIPV)
296 system and the latitude of the system's site. For example, Chen et al. [40] suggests using only
297 the latitude angle for the tilted panel. Several other studies, however, [41, 42] show that the
298 annual and seasonal optimum tilt angle depends on not only the latitude of the site, but also
299 the weather conditions (e.g. seasonal weather patterns such as winter clouds) as well as
300 surrounding obstacles, providing shading in urban areas. Several studies indicate the annual
301 optimum tilt of a PV panel is never greater than the site latitude, but can be up to 10 degrees
302 less than the latitude. More specifically, a tilt angle equal to the site latitude is optimal for a
303 for an ideal, totally clear skies throughout the year. For different seasons of the year, however,
304 the optimum tilt varies by up to 15 from that of the site latitude (more in the winter, less in the
305 summer).

306

307 **4. Roof-shape classification**

308

309 The total number of buildings chosen and classified for this study of the city of Geneva is
310 10,085. Exploration of the buildings shows that, for Geneva, the roofs fall into 13 shapes (Fig.
311 5), namely flat, shed, gabled, hipped, gambrel, mansard flat, mansard hipped, cross-hipped,
312 cross-gabled, corner-hipped, corner-gabled, pyramidal, and complex – the last one being a
313 hybrid or mixture of several of the above roof shapes. Some of these 13 roof shapes, however,
314 are not easy to classify properly and use in practice. For the purpose of the present analysis,
315 we have thus divided the above 13 roof shapes into 6 main roof-shape classes, which are flat
316 and shed, gabled, hipped, gambrel and mansard, cross/corner hipped and gabled, and complex
317 (Table 1). We then apply the Support Vector Machines algorithm (SVMs) for classifying the
318 roof shapes, using the following steps (a to d).

319 **(a) Feature selection.** We select the following 5 feature types for the roof shapes (Table 2):
320 (a) Number of roofs facing given directions, or roof aspects (bin width 22.5° , from 0 to 360°).
321 (b) Number of roofs with given surface slopes (bin width 10° , from 0 to 90°). (c) Number of
322 roofs with given roof shapes. (d) Percentage of total roof area within a certain slope range (bin
323 width). (e) Percentage of total roof area within a certain aspect range (bin width). Altogether,
324 there were 35 features or parameters that were selected for. These features are listed in Table
325 2.

326 **(b) Scaling of the dataset.** Each feature is normalised by subtracting its mean and dividing
327 the result by the standard deviation. The features in the labelled data are also scaled.

328 **(c) Data labelling.** A labelled data (data where the classes are known) is used to train and test
329 the classifier using SVM. The classifier is then applied to the rest of the data. Since no
330 labelled data exist for shapes of the roofs of Geneva, we manually label the data. In this step,
331 a total of 717 buildings (some 6% of total data) were manually labelled using Google Earth
332 and, in particular, a high-quality aerial map from Swisstopo (<https://map.geo.admin.ch>).
333 Examples of labelling roof shapes using high resolution aerial photographs of roofs for
334 different classes are shown in Figure 6. The number of labelled buildings for each roof type is
335 as follows. Flat: 127, gabled: 127, hipped: 124, gambrel and mansard: 118, cross and corner
336 hipped and gabled: 115, complex roofs: 106; Table 3). SVM was then used to classify the roof
337 shapes for the rest of the buildings (9368).

338 **(d) SVM testing and training.**

- 339 • For the purpose of testing and training, the labelled data are divided into two unequal
340 parts: 75% for training and cross-validation and 25% for testing the classifier (Table
341 3).
- 342 • Radial Basis Function (RBF) kernel is used to perform the SVM classification within
343 Python, using the scikit-learn library [43] and it offers the greatest accuracy.
- 344 • Cross-validation is used to assess how the training dataset can be generalised to an
345 independent dataset. K-fold cross-validation (the original sample is randomly
346 partitioned into K equally-sized subsamples) is used in this study. We choose K= 6
347 which is regarded as a common value.
- 348 • SVM classifier from the above step is used to predict the roof shapes for the rest of the
349 buildings of Geneva.
- 350 • Classification accuracy metric is used to evaluate the performance of the classifier.
351 The accuracy metric is defined as Eq. (18):

352
$$\text{accuracy} = \frac{\psi}{\Omega} * 100\% \quad (18)$$

353 where ψ is the number of correctly classified roof shapes and Ω is the total number
354 of roof shapes.

355

356 5. Results

357

358 Our results indicate that the SVM classifier is able to identify the 6 types of roof shapes, that
359 is, flat, gable, hip, gambrel & mansard, cross/corner gable & hip, and complex roofs, on
360 average, in 66% of cases (Table 4). The number of testing data and the accuracy of the
361 classification for each of the 6 roof-shape classes is given in Table 4. The aim is to maximize
362 the diagonal numbers of the table (marked in bold) as it means that buildings belonging to the
363 class 0 (flat & shed), for example, are correctly classified as part of class 0. Also, 2 out of 32
364 flat & shed roof shapes are misclassified as complex. Thus, the obtained accuracy for the flat
365 roof-shape is 94%. The results in Table 4 indicate that, on average, flat and complex roofs are
366 classified more accurately than the other types. The worst accuracy belongs to the hip &
367 pyramidal roofs in which mostly are misclassified as gable roof due to their similar
368 characteristics.

369 After the validation, the classification is applied to the rest of data. Distribution of the
370 various classes of roof shapes in Geneva as a whole is shown in Fig. 7. While the flat & shed
371 are very common, the complex roof shapes (red) are very extensive. To get more detailed
372 information about the roofs, it is not enough just to look at entire roofs, whose spatial
373 distribution and numbers are shown in Fig. 7, but also at the individual parts or surfaces of
374 which the roofs are composed. Thus, we distinguish here between the roofs themselves, the
375 number of roofs being equal to the number of buildings (a total of 10,085 analysed), and the
376 surfaces that constitute the roofs. More specifically, the number of roof surfaces (in
377 parentheses) that belong to each roof shape is as follows (Fig. 8): flat and shed (11,521), gable
378 (4411), hip (3696), gambrel an mansard (13,529), cross/corner (7252), and complex
379 (26,1002). The total number of analysed roof surfaces is thus 66,811.

380 In Fig. 8 the number of roof surfaces is plotted against roof slope (Fig. 8a) and against
381 roof aspect or facing direction (Fig. 8b). There are various points that can be inferred from
382 Fig. 8a. The first is that surfaces belonging to complex roofs are the most common for all the
383 slopes, except the steepest one. This is partly a reflection of the comparatively high number of
384 complex roof surfaces in general (26,102 or 39% of the total). Partly, however, this fact
385 reflects the geometry of the complex roofs which tend to have roof surfaces of a greater

variety of slopes than the other shapes (Figs. 5 and 6). The complex roofs generally belong to buildings with complex shape and large surface roof area. These include museums, railway stations, and churches. A second point is that the flat & shed roof class includes surfaces that slope as much as 80°. This is primarily because the flat roof class includes also the shed roofs, many of which have steep slopes. The third point is that there are some surfaces sloping in excess of 80°. These are included here (and in Fig. 4) for the sake of completeness, but were omitted for the results presented in Fig. 3.

As for the direction of facing or aspect of roof surfaces, the results agree with those obtained for the entire roofs (Figs. 3 and 4), namely that most of roof surfaces face in a southerly direction (Fig. 8b). More specifically, the aspect peak is roughly at the azimuth of 160-200°, that is, from south-southeast through south to south-southwest.

In order to analyse the solar PV potential of buildings, we need to calculate the available or useful roof areas of the buildings for PV installation. This we did by removing the superstructures from the roofs, as well as 1m² from the margin of roof surfaces, and considering the 28m² threshold (only areas larger than 28m² are considered) for roofs due to the size of a typical PV panel (Fig. 9) [10]. The resulting available area is always smaller than the total roof area (Fig. 10). In addition, footprint areas are widely available so that the available roof areas can also be compared with the corresponding footprint areas.

Because of the wide availability of footprint areas, we also calculated the ratio of useful roof area for each type of roof shape to that of the corresponding building footprint area. The results show that for most roof shapes this ratio is close to one (Fig. 11). The only exceptions, and those are not great, are the hip (ratio of 1.10) and gable (ratio 1.18). The results thus suggest that the footprint area can be used as a measure of the useful roof area for the installation of PVs.

Once the useful area has been calculated, the potential PV potential for the various roof shape groups could be estimated. Here we did so by calculating the received mean solar radiation (kWh/m²) and the PV potential (MWh/year; solar electricity production per year) for each of the 6 main roof shape groups (1 kWh = 3.6×10^6 J; 1 MWh = 3.6×10^9 J). The results (Fig. 12a) show the flat & shed roofs having the highest received mean solar radiation of 809 kWh/m². This high value is followed by somewhat lower values for cross/corner (806 kWh/m²) and hip (800 kWh/m²). Somewhat lower than these are the received radiation by gambrel and mansard (789 kWh/m²), followed by the considerably lower values for the complex (761 kWh/m²) and the gable (744 kWh/m²) roofs.

419 Since received mean solar radiation is second lowest for complex roofs, it may come as a
420 surprise that they have by far the greatest average PV potential in Geneva (Fig. 12b). The
421 reason is that complex roofs have the greatest number of roof surfaces (Fig. 8), the largest
422 cumulative roof area, as well as available or useful roof area for PVs (Fig. 10). Following the
423 complex roofs in terms of solar PV potential are the flat & shed roofs (Fig. 12b). This is less
424 surprising since these have the highest annual mean solar radiation per unit area (Fig. 12a).
425 The other roof shapes, namely gable, hip, gambrel & mansard, and cross/corner have similar
426 PV potential, the gable being the lowest.

427

428 **6. Discussion and conclusions**

429

430 The present study shows that the machine-learning technique can be very useful in classifying
431 roof shapes for city scale, as demonstrated here for the city of Geneva. By implication, the
432 same method can be extended to include larger cities and, eventually, entire countries [44], as
433 well as other European countries. The results of the study are particularly useful for designers,
434 investors, owners, and stockholders in providing quantitative information on the effects of
435 roof shape on the PV solar potential at the design stage. The results are also important from
436 planning and policy-making perspectives in that they provide helpful input for choosing the
437 right type of roof when retrofitting existing buildings or designing new buildings. The
438 classification of roofs based on their useful area for PV installations and potential for
439 receiving solar energy should encourage policy makers and investors to select roof design and
440 construction that (i) maximizes the possibility of mounting solar panels on roofs, (ii) provides
441 greater capacity for solar energy per building and (iii) lowers installation cost due to
442 simplification of the installation.

443 A further development of the concepts and methods presented here is not only to look at
444 the roof shapes and their potential for PV installations, but also to consider what the buildings
445 are used for. While we do not propose to go deeply into these aspects here, as an indication of
446 this further development of the present methods, we have explored the main uses of the
447 buildings in relation to their roof shapes and the associated PV potential (Fig. 13).

448 The buildings in Geneva are grouped according to the following building types
449 (<http://ge.ch/sitg/>): mixed, activity, residential, collective, and others. (1) Mixed buildings are
450 those that function partly as residential and partly as commercial. These include, for example,
451 buildings where family flats are on the first and/or second floor but shops on the ground floor.
452 (2) Under the term activity we include commercial stores, offices, hotels, restaurants, and

453 buildings or spaces for industry. (3) Residential covers the buildings where people live - their
454 homes. (4) Collective denotes those buildings used for educational and cultural purposes as
455 well as for health and sport. (5) Others are buildings that do not fit into one of the previous
456 four categories.

457 The results (Fig. 13a) show that residential building types are the highest percentage of
458 buildings for flat & shed, gable, and hip, whereas mixed buildings are the highest percentage
459 for gambrel & mansard, cross/corner hip & gable, and complex buildings. For those roof-
460 shape groups where residential is highest, mixed is the second highest, while residential is the
461 second highest percentage for those shape groups where mixed is the highest percentage.
462 Clearly, residential and mixed together dominate and jointly range from 54% for complex
463 shapes to 75% for hip.

464 In terms of contribution to the total solar PV potential in the shape classes, the picture is
465 somewhat different (Fig. 13b). While residential is the highest percentage (44%) for gable and
466 hip, and mixed for gambrel & mansard and cross/corner (45%), collective is the highest
467 percentage for buildings with complex roof shape (38%), and mixed is highest for flat and
468 shed (27%). For flat & shed, however, three of the building activity types make a very similar
469 contribution to the solar PV potential, namely mixed, activity, and residential. Clearly, the
470 contribution of the building types to the solar PV is somewhat complex and further studies on
471 this topic may be of interest.

472 There are other possibilities for expanding and developing the work presented here, with
473 a view of providing input and help for urban design, aiming at PV installations on the roofs,
474 in the future. For example, the roof-shape classification could be refined, adding more roof
475 shapes into account. In Fig. 5 we list 13 roof shapes, which are then reduced to the 6 shape
476 classes used in the present paper. While we do not think that it is realistic to try to use all the
477 13 shapes, a more refined classification, with more than 6 main groups, would allow us to
478 explore finer aspects of the variation in roof shapes in cities, as well as their solar PV
479 potential. Another aspect is to try to use a more adapted machine learning algorithm (e.g.
480 Random Forests) and expand the classification of roof shapes to the national scale. It is also
481 important to improve the training of the machine learning procedure so as to get higher
482 percentage of correct identification of the roof shapes. Presently, the SVM correctly identifies
483 the 6 main roof shapes in 66% of the cases. While this is certainly acceptable, with further
484 refinement and training on new and different data the success rate may increase in this
485 respect.

486 In conclusion, in this paper we show that machine-learning approach through the Support
487 Vector Machine classification is a promising method for classifying roof shapes. We use 6
488 main roof shape classes and rank them based on their useful areas for PV installations. We
489 also show that for most of the roof shapes the ratio between the useful roof area and the
490 building footprint is close to one. Since footprint areas are widely available, this result is very
491 important since it suggests that footprint areas can be used as approximate substitutes to
492 assess useful roof areas for PVs. We conclude that the solar roof shape classification provides
493 basic information that should be of help in designing new buildings, retrofitting innervations
494 on the building roofs, as well as for efficient solar integration on roofs.

495

496 **Acknowledgments**

497 This research has been financed partly by CTI (Commission for Technology and Innovation)
498 within SCCER FEEB&D (CTI.2014.0119) and partly by Swiss National Science Foundation
499 under Mobility Fellowship P300P2 174514.

500

501 **References**

- 502 [1] L. Wiginton, H. Nguyen, and J. Pearce. Quantifying rooftop solar photovoltaic potential
503 for regional renewable energy policy. *Comput. Environ. Urban.* 34 (2010) 345–357.
- 504 [2] S. Nowak, M. Gutschner, D. Ruoss, P. Togweiler, T. Schoen. Potential of building
505 integrated photovoltaics. Report IEA-PVPS T7-4 (2002).
- 506 [3] J. Goodling, R. Crook, A.S. Tomlin. Modelling of roof geometries from low-resolution
507 LiDAR data for city-scale solar energy applications using a neighboring buildings method.
508 *Appl. Energy* 148 (2015) 93–104.
- 509 [4] F. Tarsha-Kurdi, T. Landes, P. Grussenmeyer P, M. Koehl, Model-driven and data-driven
510 approaches using LiDAR data: analysis and comparison, Munich (2007) 87–92.
- 511 [5] M.C. Brito, N. Gomes, T. Santos, J.A. Tenedorio. Photovoltaic potential in a Lisbon
512 suburb using LiDAR data. *Sol. Energy* 86 (2012) 283–288.
- 513 [6] C. Carneiro, E. Morello, and G. Desthieux. Assessment of solar irradiance on the urban
514 fabric for the production of renewable energy using lidar data and image processing
515 techniques. In *Advances in GIScience* Springer (2009) 83–112.
- 516 [7] N. Mohajeri, A. Gudmundsson, G. Upadhyay, D. Assouline, J. Kämpf, J.L. Scartezzini,
517 Effects of urban compactness on solar energy potential. *Renew. Energy* 93 (2016) 469-482.
- 518 [8] C. Carneiro, E. Morello, T. Voegtle, and F. Golay. Digital urban morphometrics:
519 automatic extraction and assessment of morphological properties of buildings. *T. GIS.* 14
520 (2010) 497–531.
- 521 [9] N. Lukac, S. Seme, D. Zlaus, G. Stumberger, B. Zalik. Buildings roofs photovoltaic
522 potential assessment based on (Light Detection And Ranging) data. *Energy* 66 (2014) 598-
523 609.

- 524 [10] D. Assouline, N. Mohajeri1, J.-L. Scartezzini. Quantifying rooftop photovoltaic solar
525 energy potential: A machine learning approach. Sol. Energy 141 (2017), 278–296.
- 526 [11] J. Gooding, R. Crook, A. S. Tomlin. Modelling of roof geometries from low-resolution
527 LiDAR data for city-scale solar energy applications using a neighboring buildings method,
528 Appl. Energy 148 (2015) 93–104.
- 529 [12] M. Kada, Scale-dependent simplification of 3D building models based on cell
530 decomposition and primitive instancing. In: Proceedings of the International Conference on
531 Spatial Information Theory. COSIT'07. Melbourne, Australia (2007) 222–237.
- 532 [13] J. Khan, M. H. Arsalan, Estimation of rooftop solar photovoltaic potential using geo-
533 spatial techniques: A perspective from planned neighborhood of Karachi – Pakistan, Renew.
534 Energy 90 (2016) 188e203.
- 535 [14] S. Izquierdo, M. Rodrigues, N. Fueyo, A method for estimating the geographical
536 distribution of the available roof surface area for large-scale photovoltaic energy-potential
537 evaluations. Sol. Energy 82 (2008), 929–939.
- 538 [15] I. Shapiro, The Receptivity of Roofs to Solar Panels. In Proceedings of the 2nd World
539 Sustain. Forum, 1–30 November 2012; Sciforum Electronic Conference Series, 2 (2012)
- 540 [16] E. Nault, G. Peronato, G. Rey, M. Andersen, Review and critical analysis of early-design
541 phase evaluation metrics for the solar potential of neighborhood designs, Build Environ. 92
542 (2015) 679e691.
- 543 [17] D. Li, G. Liu, S. Liao, Solar potential in urban residential buildings, Sol. Energy 111
544 (2015) 225e235.
- 545 [18] K. Konis, A. Games, K. Kensek, Passive performance and building form: An
546 optimization framework for early-stage design support. Sol. Energy 125 (2016) 161-179.
- 547 [19] C. Voyant, G. Notto, S. Kalogirou, M. L. Nivet, C. Paoli, F. Motte, A. Fouilloy, Machine
548 learning methods for solar radiation forecasting: A review, Renew. Energy 105 (2017)
549 569e582.
- 550 [20] M. Kanevski, Statistical learning in geoscience modelling: Novel algorithms and
551 challenging case studies. Computers & Geosciences 85 (2015) 1–2.
- 552 [21] M. Kanevski, A. Pozdnukhov, V. Timonin, Machine Learning for Spatial Environmental
553 Data: Theory, Applications, and Software. EPFL Press, Spain (2009).
- 554 [22] M. Kanevski, M. Maignan, Analysis and Modelling of Spatial Environmental
555 Data. EPFL Press, Lausanne (2004).
- 556 [23] C. Cortes, V. Vapnik, Support-vector networks. Machine Learning 20 (1995) 273–297.
- 557 [24] A. Hur-Ben, J. Watson. 2010. A user's guide to Support Vector Machines. Methods Mol.
558 Biol. 609 (2010) 223-39. doi: 10.1007/978-1-60327-241-4_13.
- 559 [25] T. Hastie, R., Tibshirani, J. Friedman. The Elements of Statistical Learning: Data
560 Mining, Inference, and Prediction. Springer Series in Statistics, USA (2013).
- 561 [26] T. Fletcher T, Support Vector Machines Explained. University College London (2009).
- 562 [27] C. Burges. A tutorial on support vector machines for pattern recognition. Data Min.
563 Knowl. Disc. 2 (1998) 121–167.

- 564 [28] R. Vanderbei. LOQO: An Interior Point Code For Quadratic Programming, Technical
565 Report (Revised) SOR 94-15, Princeton University (1998).
- 566 [29] V. Vapnik, Statistical Learning Theory, John Wiley & Sons (1998).
- 567 [30] B. Scholkopf, A. Smola, Learning with Kernels: Support Vector Machines,
568 Regularization, Optimization, and Beyond. MIT Press (2002).
- 569 [31] A. Smola, B. Scholkopf, A tutorial on support vector regression, Statistics and
570 Computing 14 (2004) 199-222.
- 571 [32] G. Desthieux, C. Carneiro, E. Morello, Cadastre Solaire Du Canton De Geneve. Raport
572 final. Etude réalisée par hepia, EPFL et Politenico di Milano pour le compte du Service de
573 l'énergie-ScanE et des Services industriels genevois (SIG) (2011).
- 574 [33] G. Desthieux, P. Gallinelli, R. Camponovo, Cadastre Solaire Du Canton De Geneve –
575 Phase 2. Analyse du potentiel de production énergétique par les panneaux solaires thermiques
576 et PV. Rapport final. Etude pour le compte de l'Office cantonal de l'énergie (OCEN) et des
577 Services industriels genevois (SIG) (2014).
- 578 [34] D. H. W. Li, T. N. T. Lam, Determining the optimum tilt angle and orientation for solar
579 energy collection based on measured solar radiance data. International Journal of Photoenergy
580 (2007) 85402, doi:10.1155/2007/85402.
- 581 [35] D. H. W. Li, G. H. W. Cheung, Study of models for predicting the diffuse irradiance on
582 inclined surfaces, Appl. Energ. 81 (2005) 170–186.
- 583 [36] D. Robinson, J.L. Scartezzini, M. Montavon, R. Compagnon, SOLURBAN Project:
584 Solar Utilisation Potential of Urban Sites, Swiss Federal Office for Energy, Bundesamt für
585 Energie BFE, Bern (2005).
- 586 [37] NET Nowak Energie & Technologie SA. Le Potentiel Solaire dans le Canton de Genève .
587 Technical report, Waldweg 8, Switzerland, St. Ursen (2004).
- 588 [38] G. Desthieux, P. Gallinelli, R. Camponovo. Analyse du potentiel de production
589 énergétique par les panneaux solaires thermique et PV. Final report. Etude pour le compte de
590 l'Office cantonal de l'énergie (OCEN) et des Services industriels genevois (SIG), Geneva
591 (2014).
- 592 [39] International Energy Agency (IEA) . Potential for Building Integrated Photovoltaics.
593 Achievable levels of electricity from photovoltaic roofs and façades: methodology, case
594 studies, rules of thumb and determination of the potential of building integrated photovoltaics
595 for selected countries. Summary of Report IEA - PVPS T7-4.
- 596 [40] C.L. Cheng, C. S. Sanchez Jimenez, M. Lee. Research of BIPV optimal tilted angle, use
597 of latitude concept for south orientated plans. Renew. Energy 34 (2009) 1644-1650.
- 598 [41] A. Gharakhani Siraki, P. Pillay. Study of optimum tilt angles for solar panels in different
599 latitudes for urban applications. Solar Energy 86 (2012) 1920–1928.
- 600 [42] M. Lave, J. Kleissl. Optimum fixed orientations and benefits of tracking for capturing
601 solar radiation in the continental United States. Renewable Energy 36 (2011) 1145-1152.
- 602 [43] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel,
603 P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, J., A. Passos, D. Cournapeau, M.,
604 Brucher, M. Perrot, E. Duchesnay, Scikit-learn: Machine learning in Python. J. Mach. Learn.
605 Res. 12 (2011) 2825–2830.

606 [44] D. Assouline, N. Mohajeri1, J.-L. Scartezzini, Building rooftop classification using
607 Random Forests for large-scale PV deployment. Earth Resources and Environmental Remote
608 Sensing/GIS Applications VIII, edited by Ulrich Michel, Karsten Schulz Proc. of SPIE Vol.
609 10428 (2017), 1042806, doi: 10.1117/12.2277692.

610

611

612 **Figure captions**

613 **Fig. 1.** A schematic graph showing the main concept of SVM for classification. (a)
614 Hyperplane through two arbitrary linearly separable classes. (b) Two non-linearly separable
615 classes in 2D. The data cannot be discriminated with a linear function. (c) Nonlinear mapping
616 of 2D data into 3D using kernels in which the maximum margin hyper-plane can be
617 constructed.

618

619 **Fig. 2.** Yearly mean solar radiation of building roofs in the city of Geneva at three different
620 scales, namely (a) a building scale, (b) a neighbourhood scale, and (c) a city scale.

621

622 **Fig. 3.** Number of building roof surfaces for different ranges of (a) slope and (c) aspect.
623 Annual mean solar radiation (kWh/m^2) of roof surfaces for different ranges of (b) slope and
624 (d) aspect.

625

626 **Fig. 4.** (a) Monthly and (b) annual mean solar radiation (kWh/m^2) as a function of slope and
627 aspect of roof surfaces for a typical neighbourhood, shown the location in Fig. 2b, and (b)
628 annual mean solar radiation (kWh/m^2) as a function of slope and aspect of roof surfaces for
629 buildings in the entire city of Geneva (Fig. 2c).

630

631 **Fig. 5.** Schematic presentation of the most common roof shapes (perspective and plan view)
632 of the buildings in the city of Geneva.

633

634 **Fig. 6.** Examples of labelling roof shapes using high resolution aerial photographs of roofs for
635 different classes. For each shape class, the aerial photograph is on the left, the labelled roof
636 shape on the right. (a) Flat roof (class 0). (b) Gable (class 1). (c) Hip (class 2). (d) Mansard
637 (class 3). (e) Cross/corner Gable and Hip (class 4). (f) Complex (class 5).

638

639 **Fig. 7.** Results of roof-shape classification for building data in the city of Geneva. The colours
640 show distribution of different roof-shape class for total 10,085 buildings in the city of Geneva.

641

642 **Fig. 8.** Frequency distribution of roof surfaces versus (a) roof slope and (b) aspect for
643 different roof-shape groups. The number (in parentheses) of roof surfaces are as follows: flat
644 and shed (11,521), gable (4411), hip (3696), gambrel and mansard (13,529), cross/corner
645 (7252), and complex (26,102). The total number of analysed roof surfaces is thus 66,811.

646

647 **Fig. 9.** Schematic presentation of different steps taken to estimate available roof area for a
648 building. (a) A building with detailed roof surfaces including superstructure (e.g., a chimney,
649 dormers, and a staircase). (b) Removal of the superstructures from roof surfaces. (c) Creation

650 of a 1m² buffer around each remaining roof surface. (d) The final available roof useful area
651 for PV installation after removing the areas less than 28m².

652

653 **Fig. 10.** Average roof area (m²), average building footprint area (m²), and average available
654 roof area (m²) for PV installation for different roof-shape classes.

655

656 **Fig. 11.** (a) Ratio of available or useful roof areas for PV installations to footprint areas of
657 buildings in different roof-shape groups. (b) Ratio of useful roof areas to the roof areas
658 considering the slope of each roof surface.

659

660 **Fig. 12.** (a) Received annual mean solar radiation in kWh/m² and (b) average PV potential for
661 different roof-shape groups. In figure b, the darker colours for each roof-shape group show
662 the estimated average PV potential based on the total solar radiation received by the roof
663 surfaces. By contrast, the lighter colours show the estimated average PV potential based only
664 on the solar radiation in excess of 800 kWh/m² received by the roof surfaces.

665

666 **Fig. 13.** Roof-shape classes versus building types (building activity) and solar PV potential.
667 The building types are: (1) Mixed, which are partly used as residential and partly as
668 commercial. (2) Activity, where the use includes functions as commercial stores, offices,
669 hotels, restaurant, and industrial functions. (3) Residential, where the buildings function only
670 as family homes. (4) Collective, where the use is for educational, cultural, and health and
671 sport purposes. (5) Others, where the buildings are primary for parking or function as
672 transport infrastructures. (a) Total number of buildings of the above types given as percentage
673 of all the buildings for a given roof shape. (b) Total PV potential for the building types given
674 as percentage of all the buildings for a given roof shape.

675

676

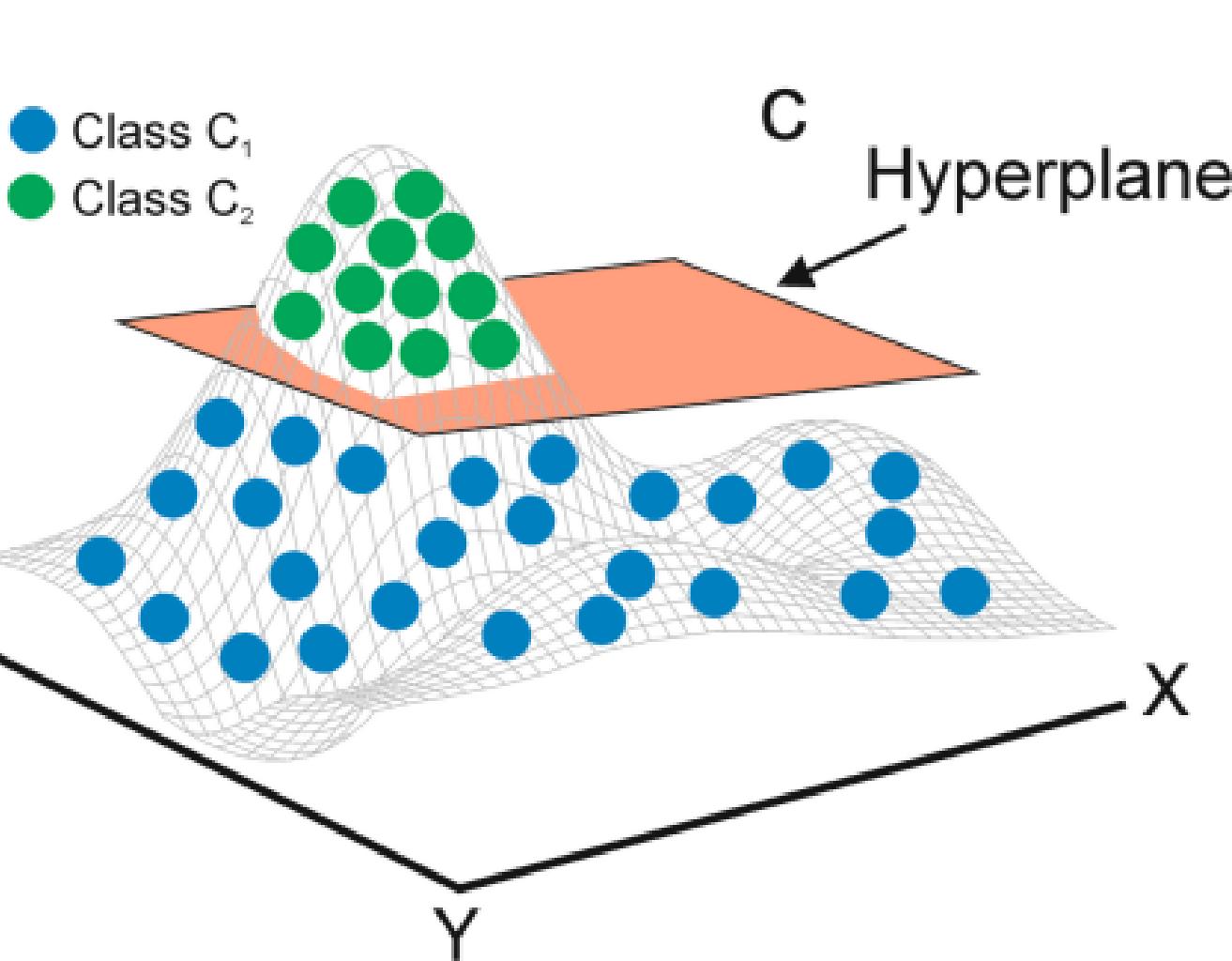
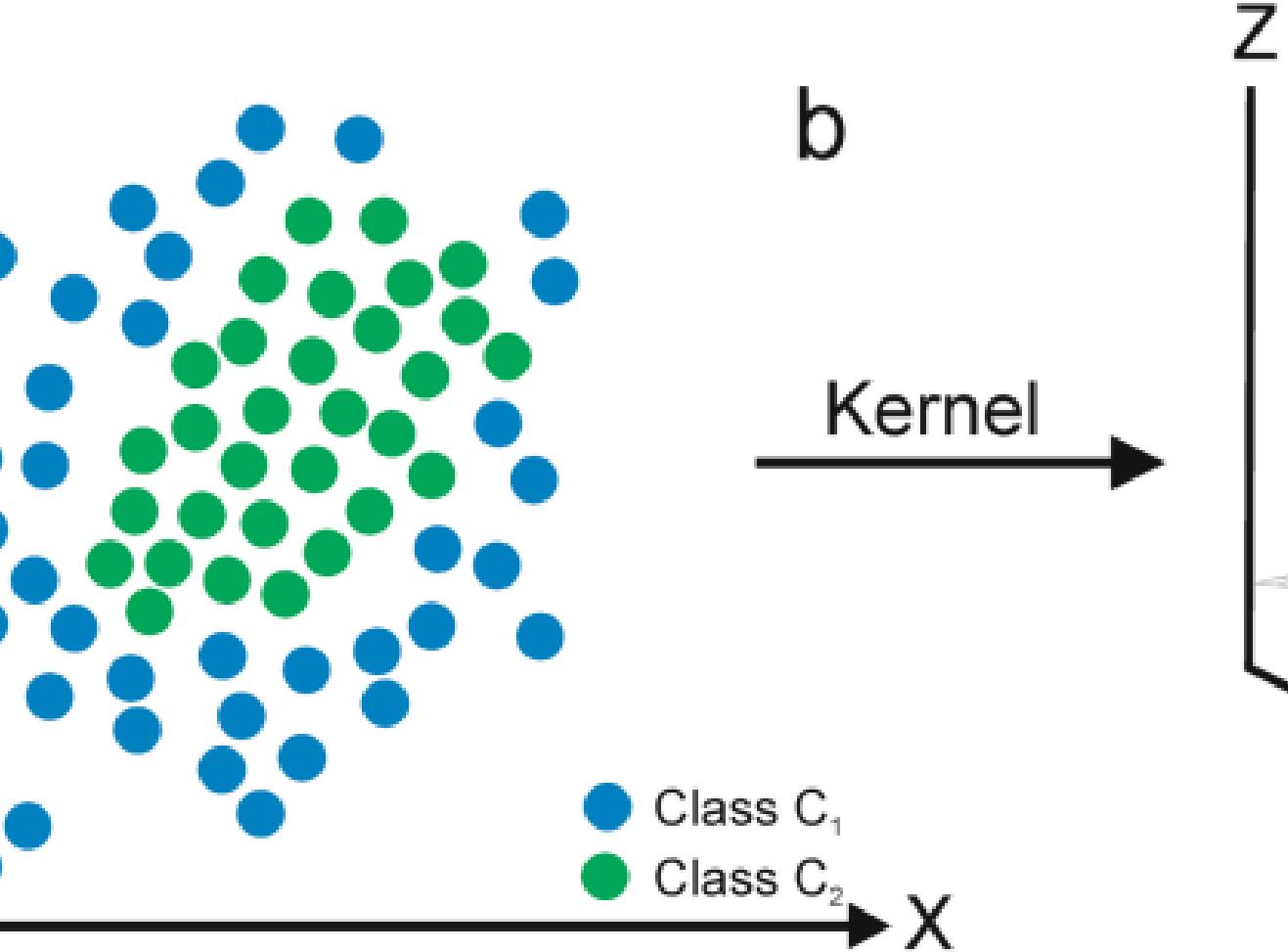
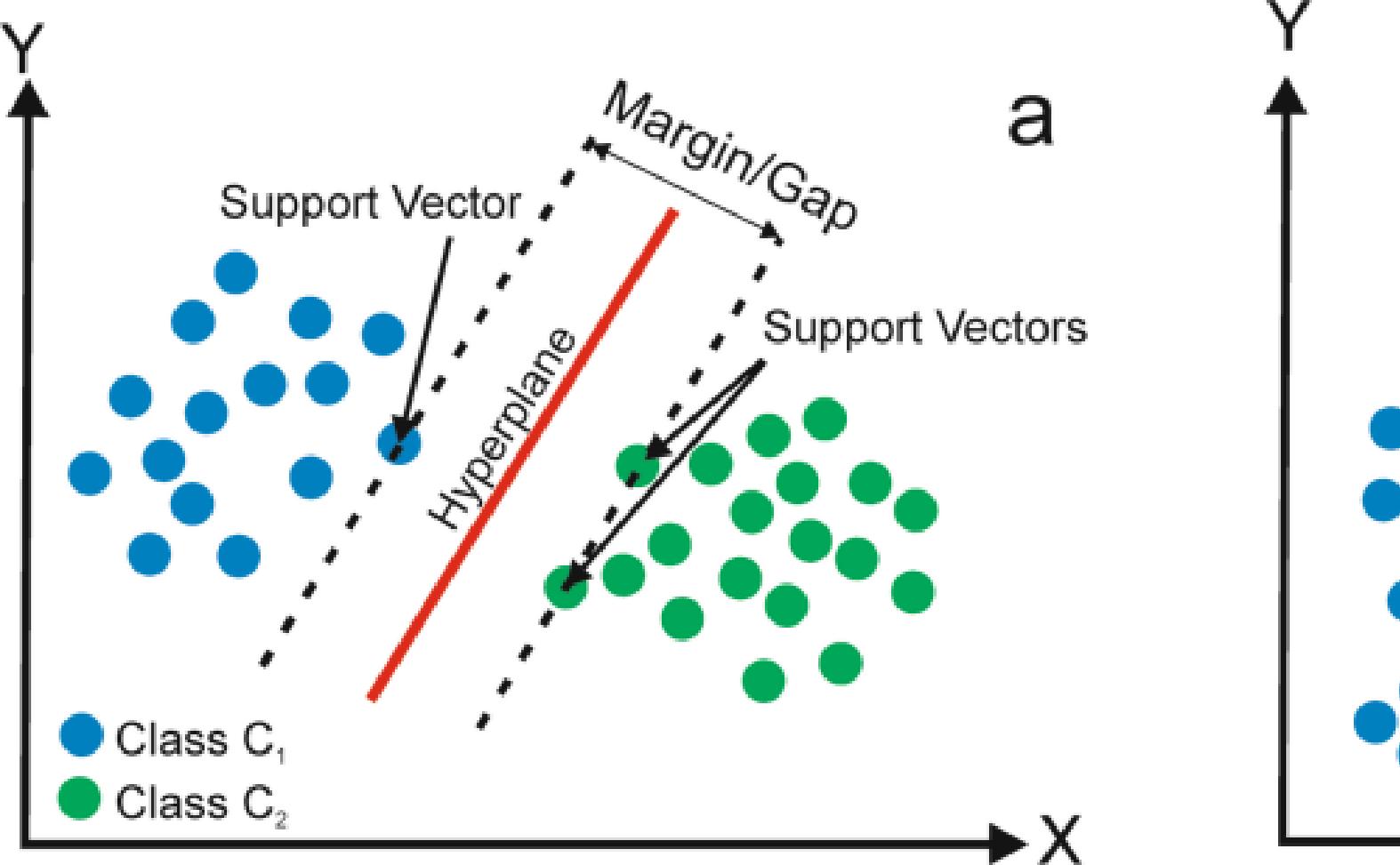
677

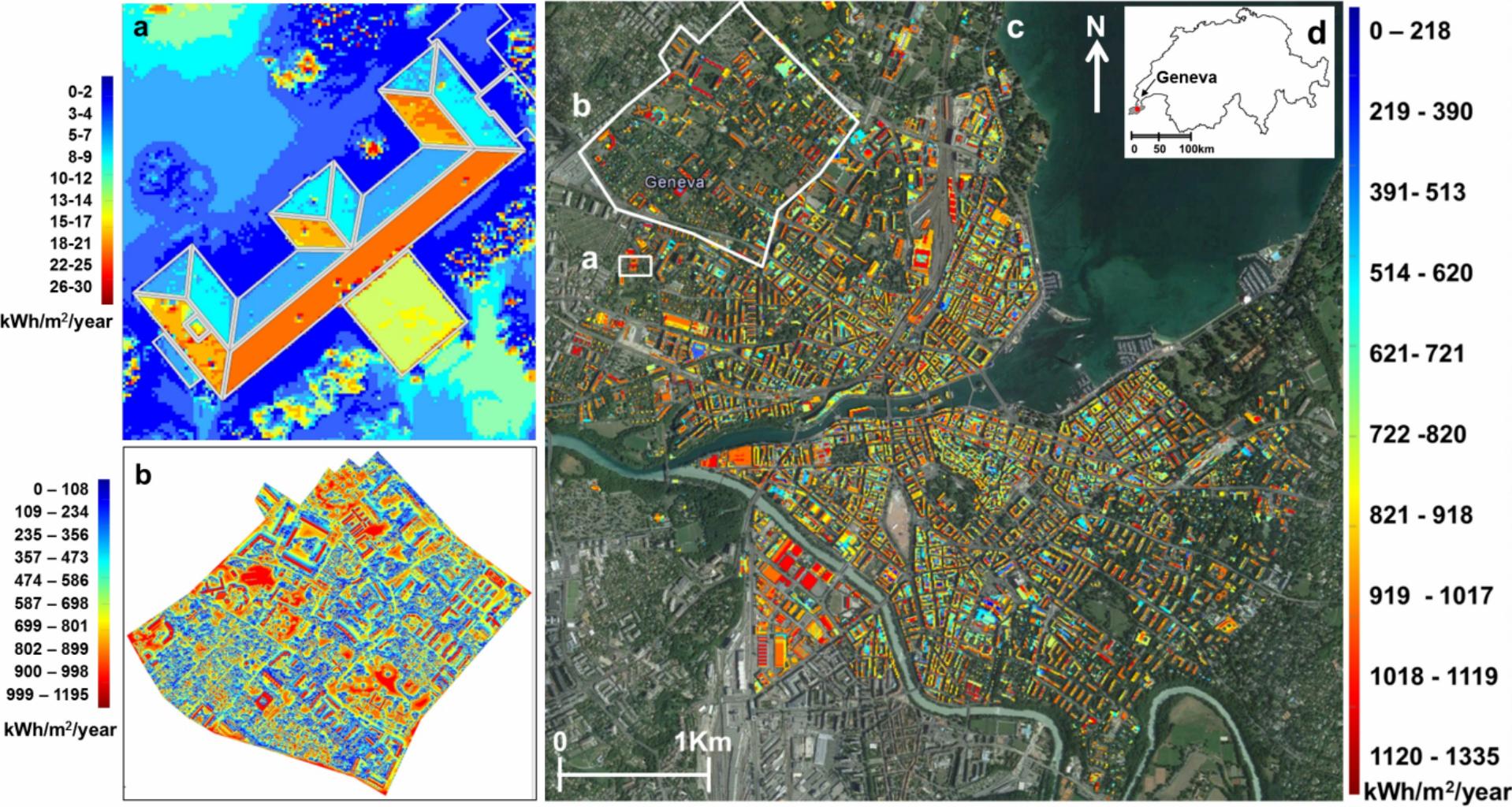
678

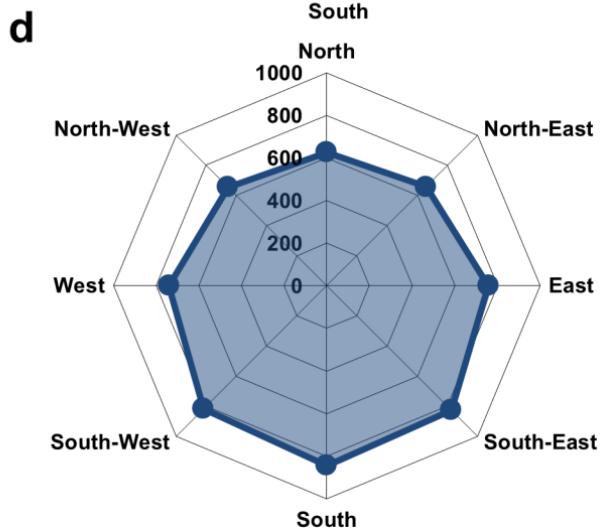
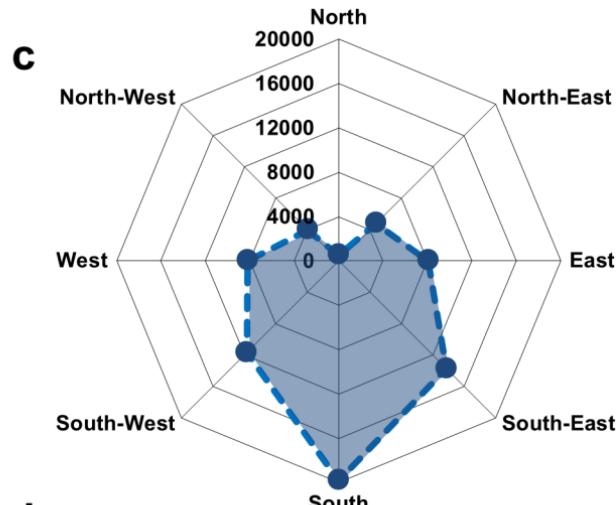
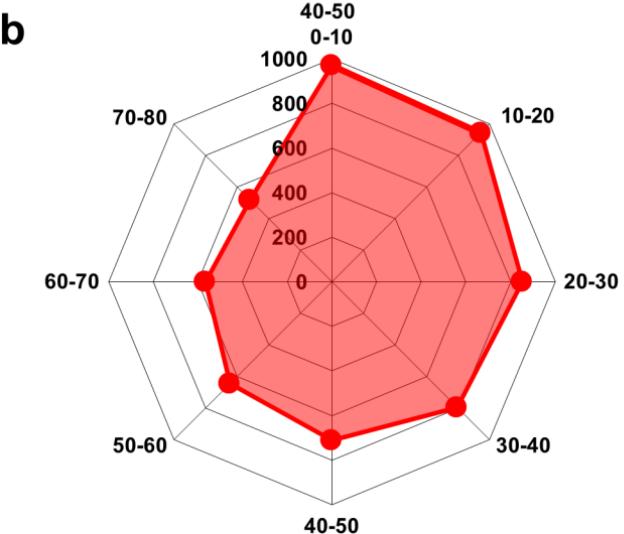
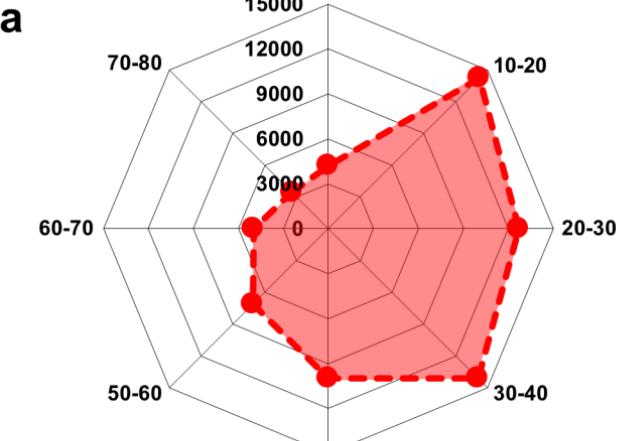
679

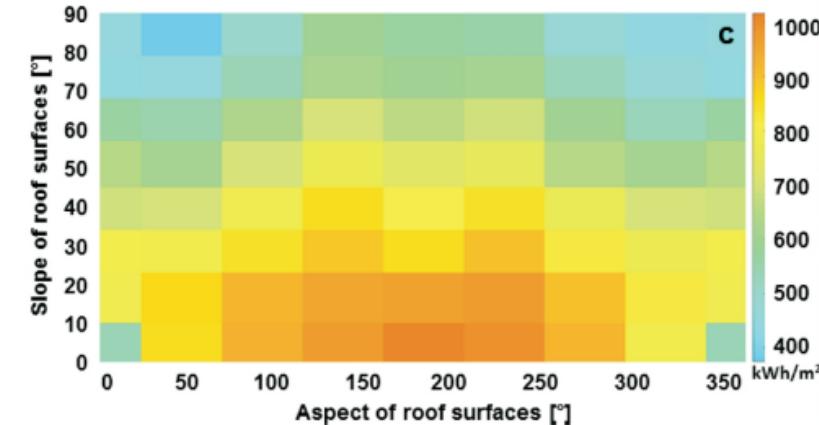
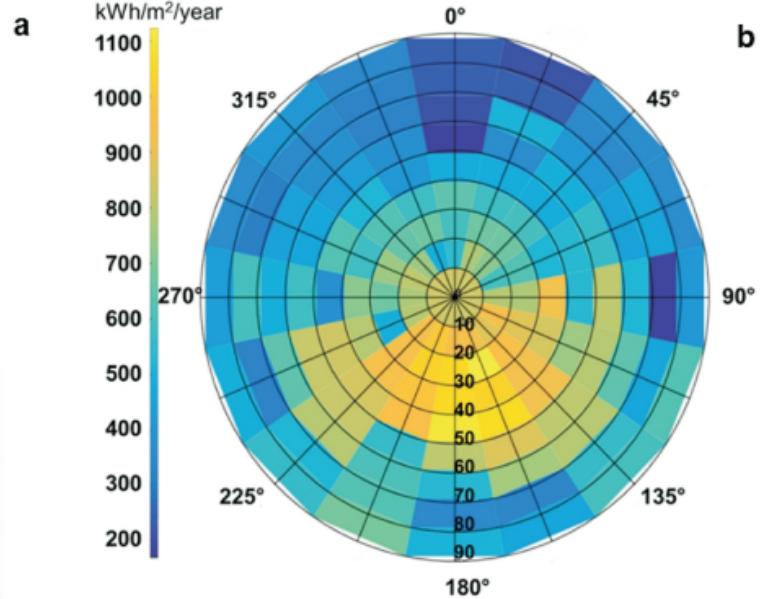
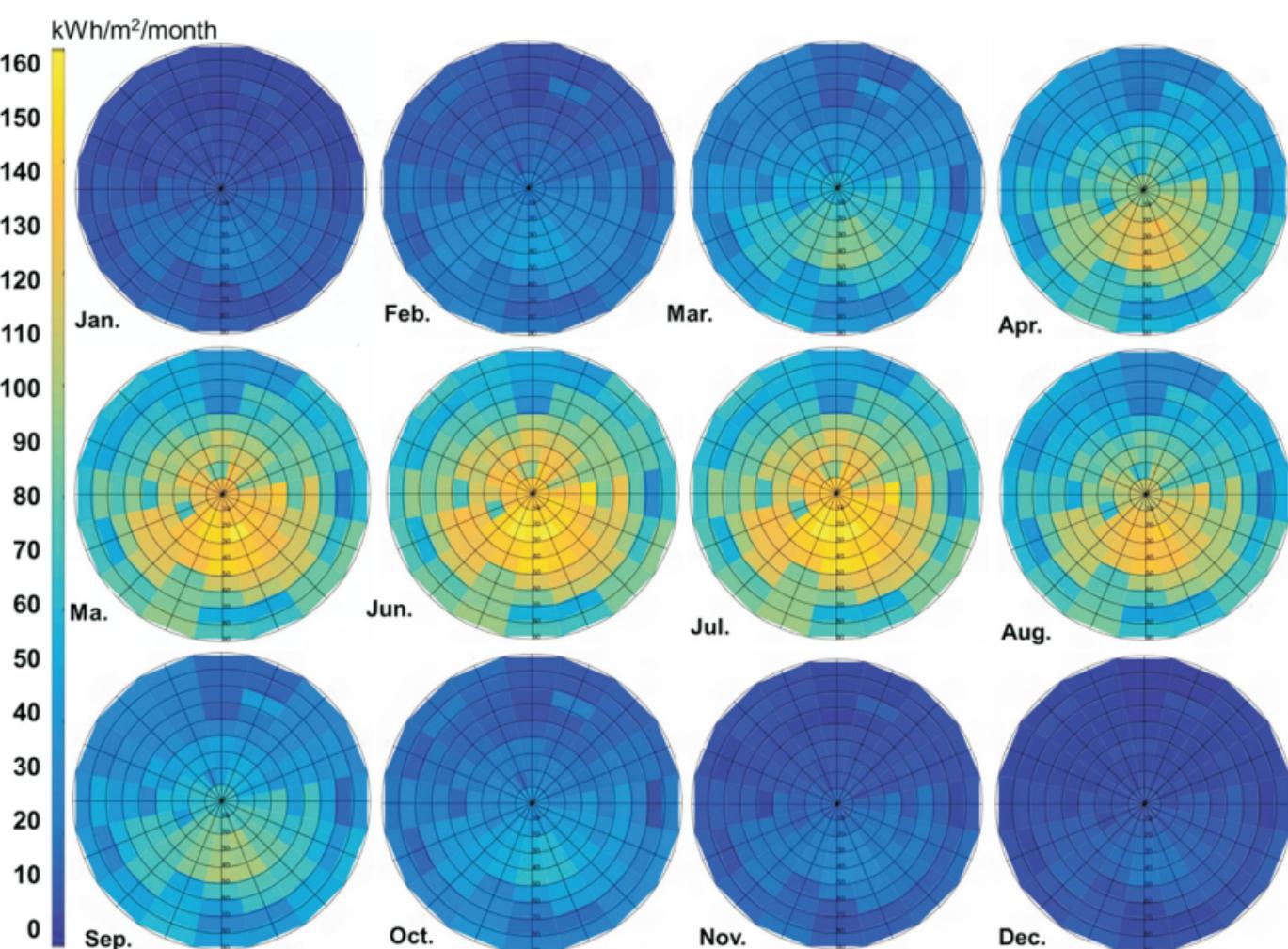
Highlights:

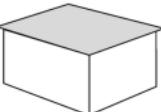
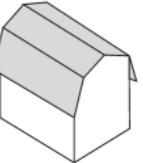
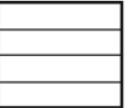
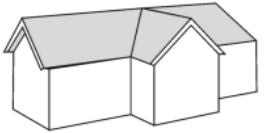
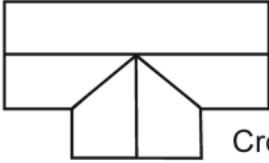
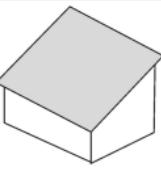
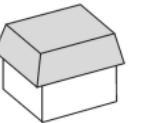
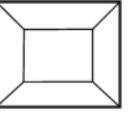
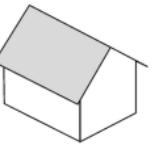
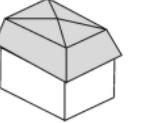
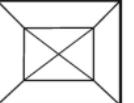
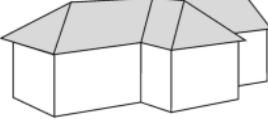
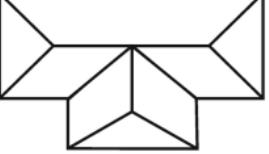
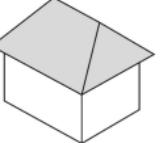
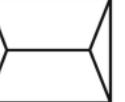
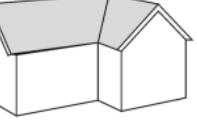
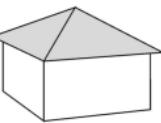
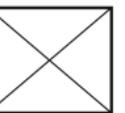
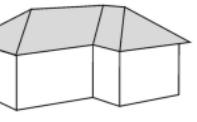
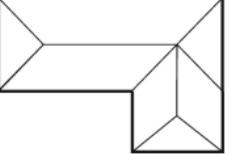
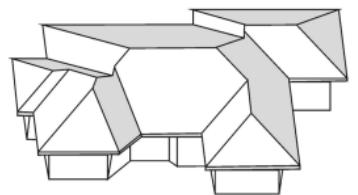
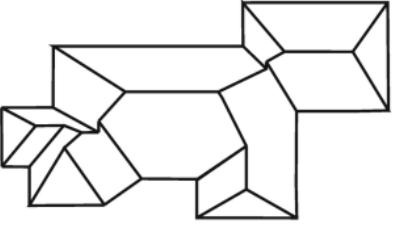
- Ranking roof shapes for receiving solar energy potential in Geneva city.
- Urban scale roof shape classification.
- Support Vector Machine (SVM) classification was used to classify roof shapes.
- Solar roof-shape classification to support early-stage design.

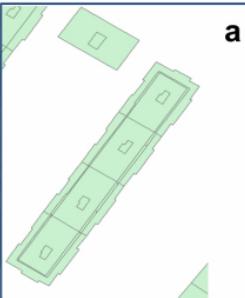




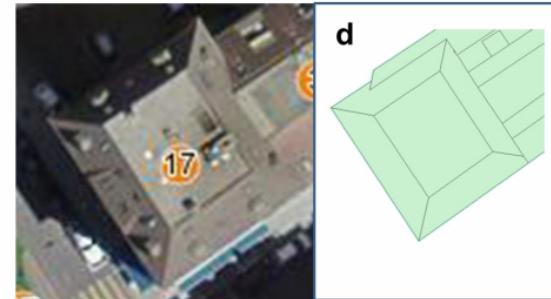
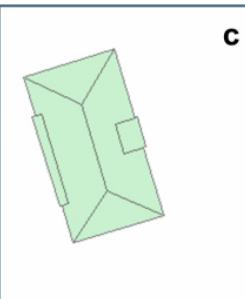
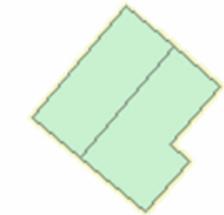




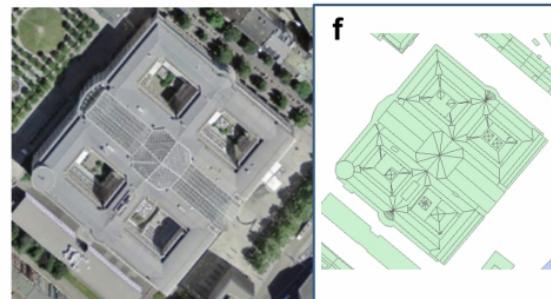
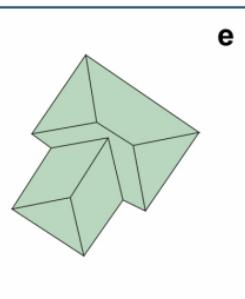
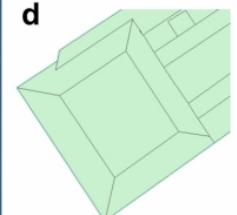
Perspective	Plan	Perspective	Plan	Perspective	Plan
					
	Flat			Gambrel	Cross gable
					
	Shed		Mansard flat		
					
	Gable		Mansard hip		Cross hip
					
	Hip		Corner gable		
					
	Pyramidal		Corner hip		Complex



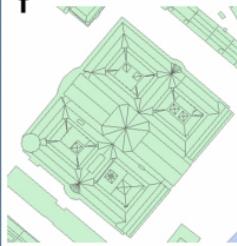
b

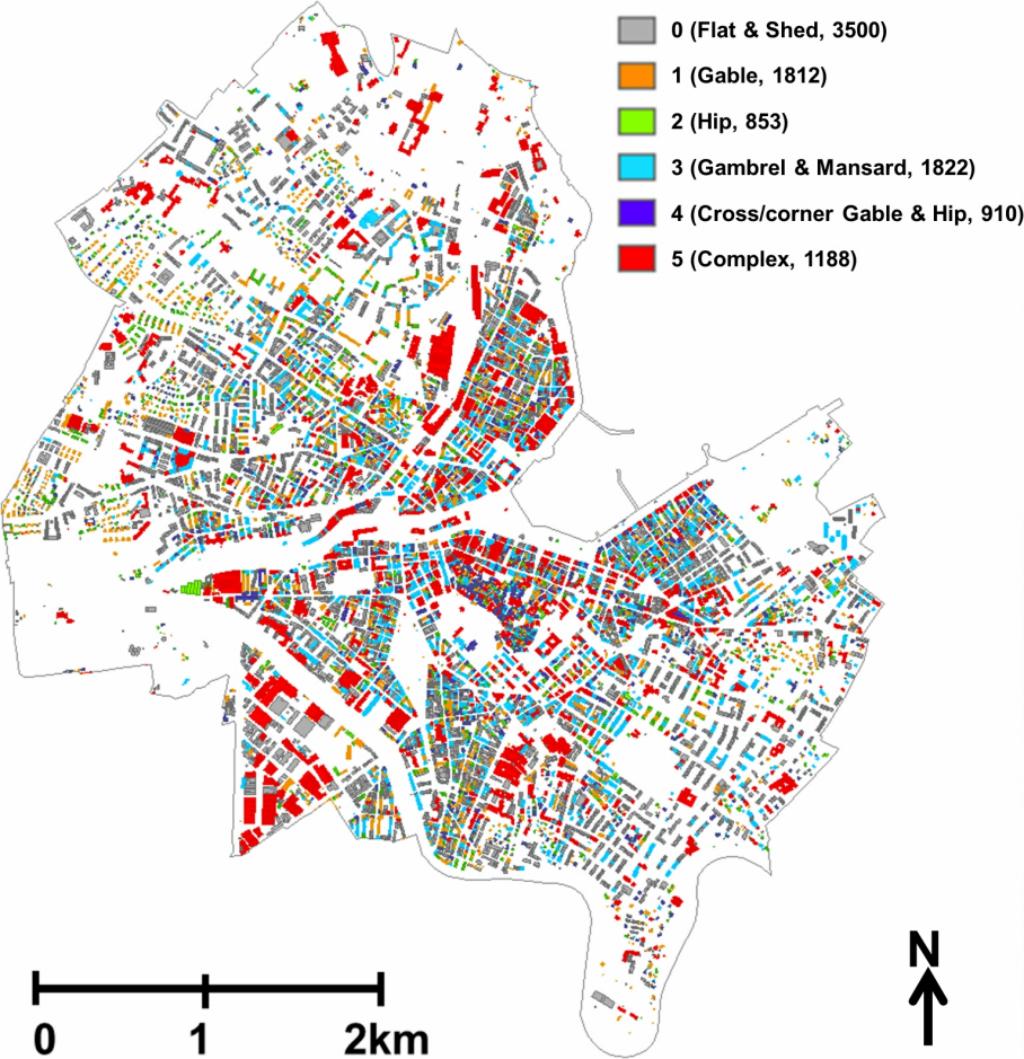


d

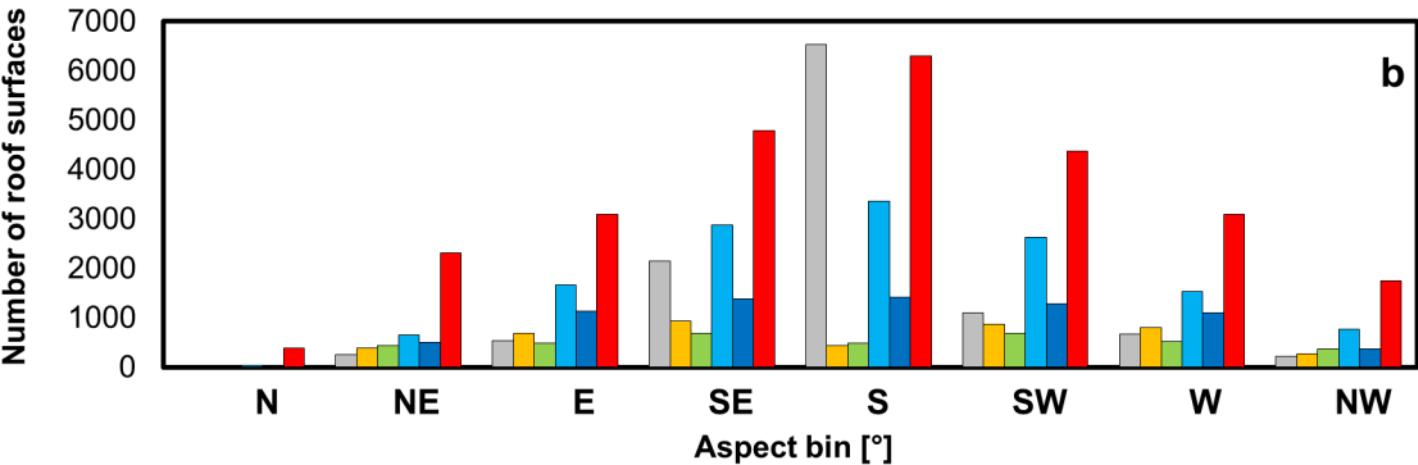
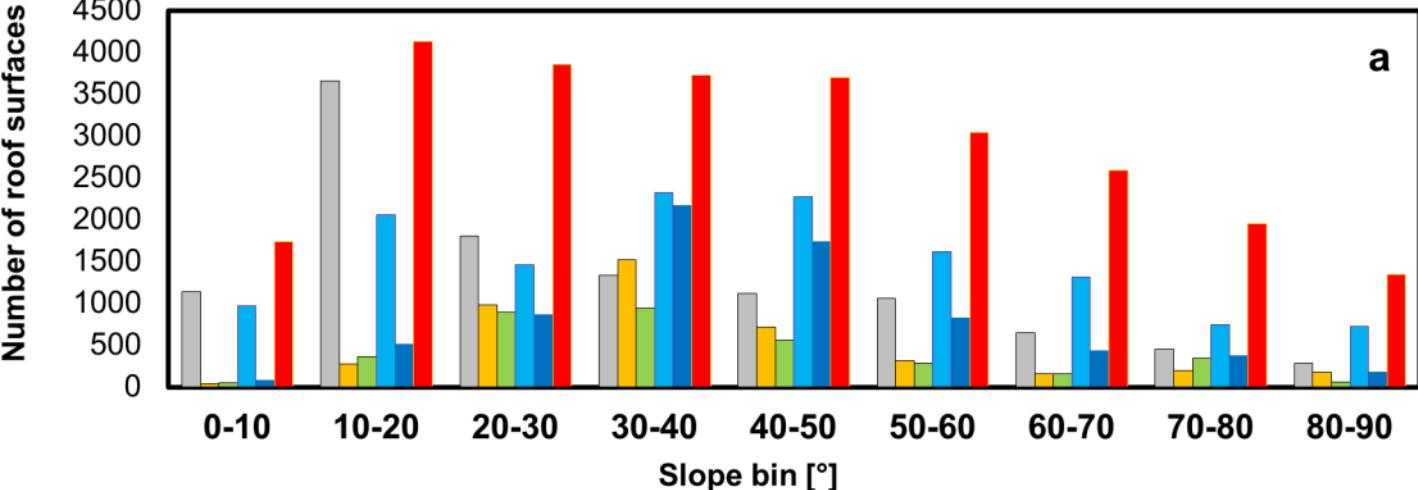


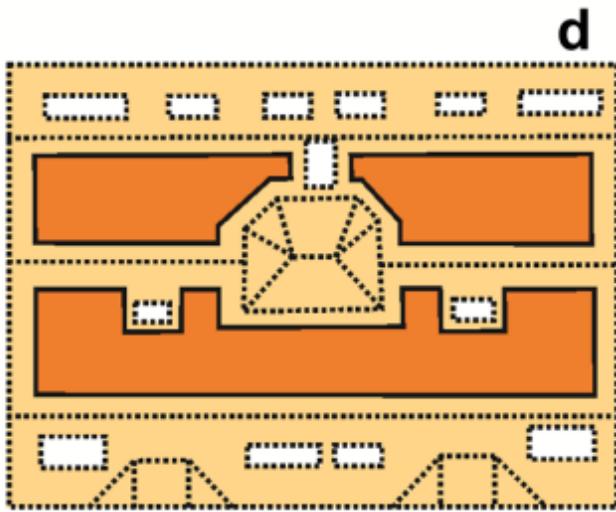
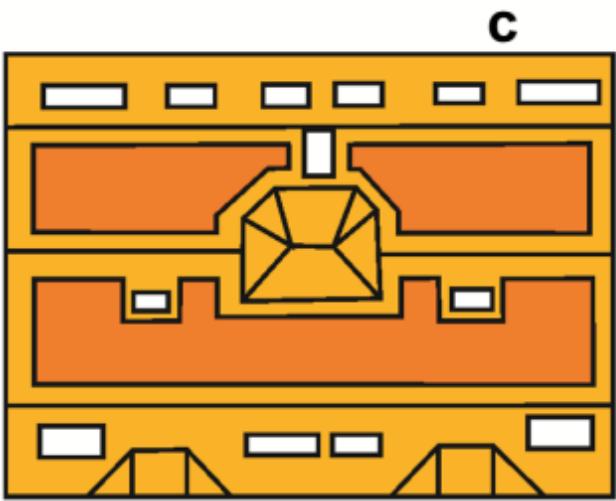
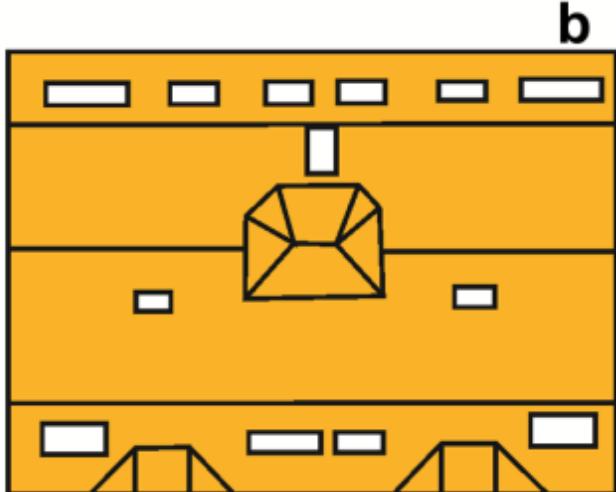
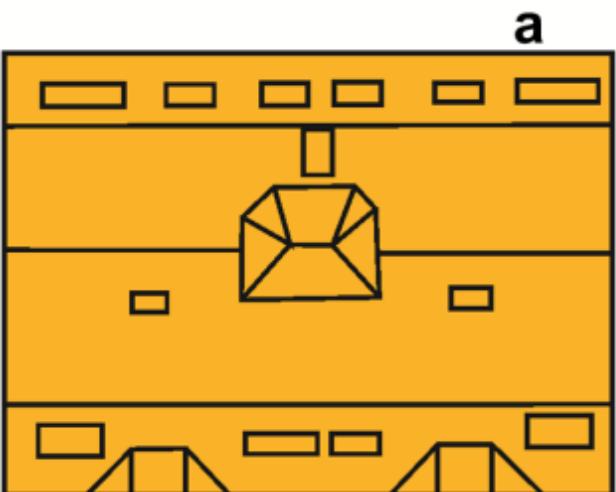
f

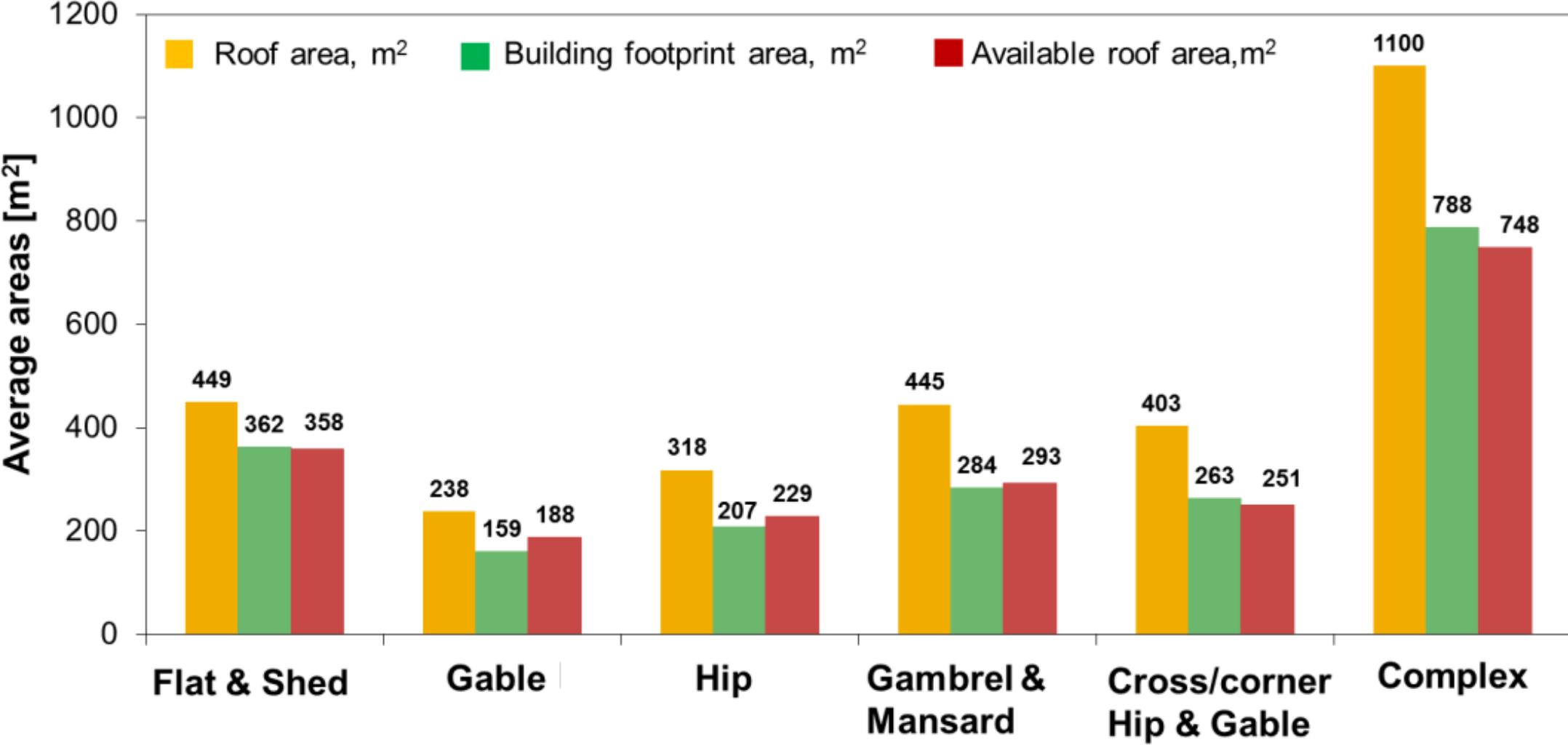


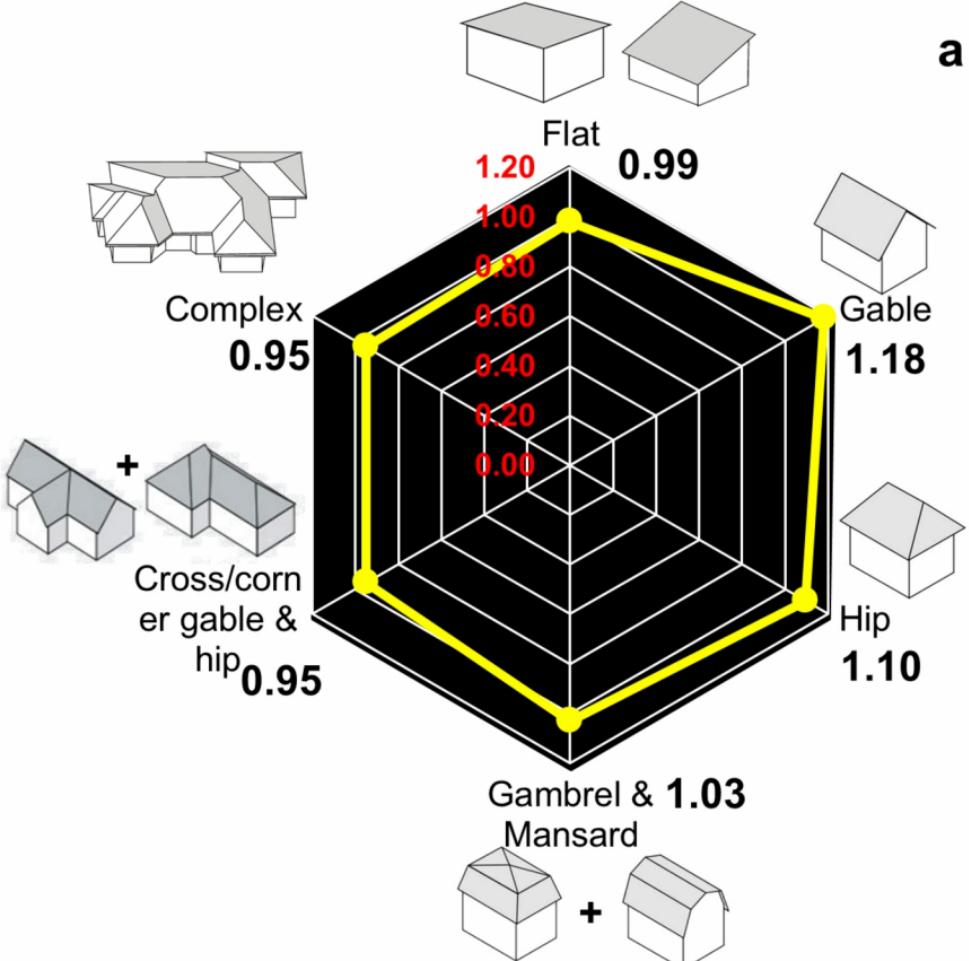
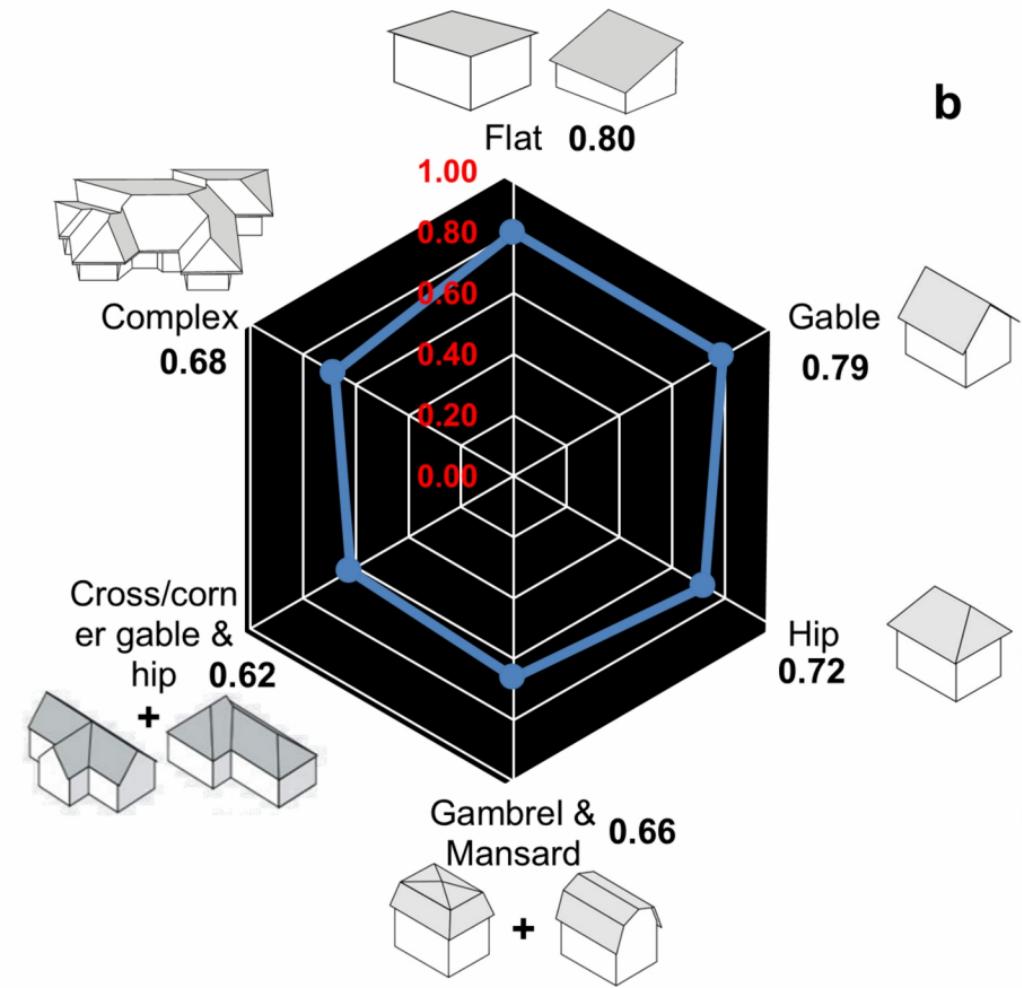


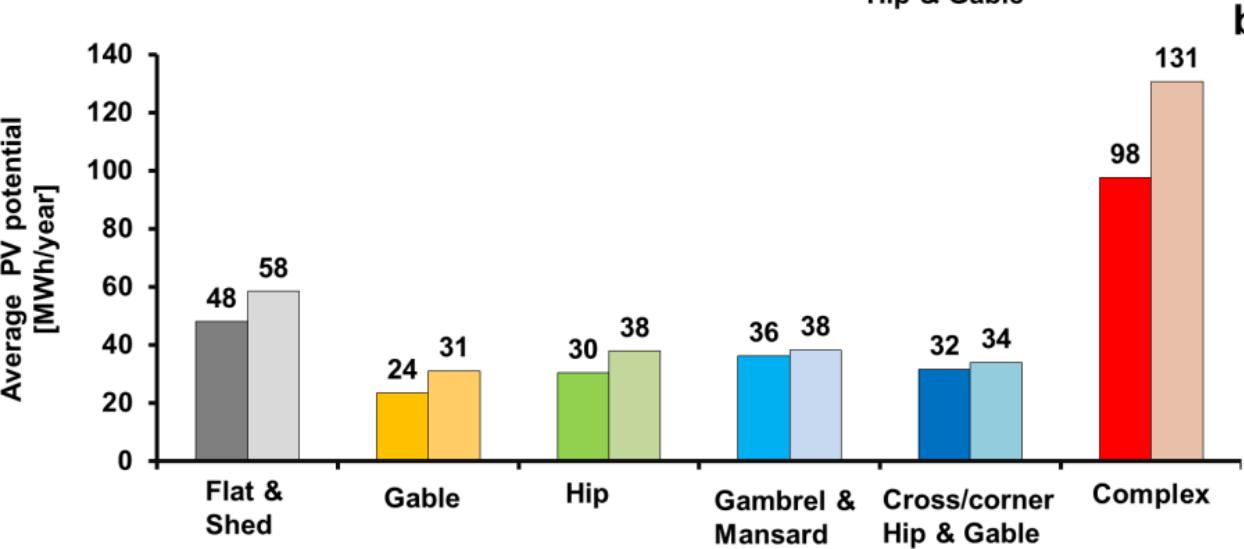
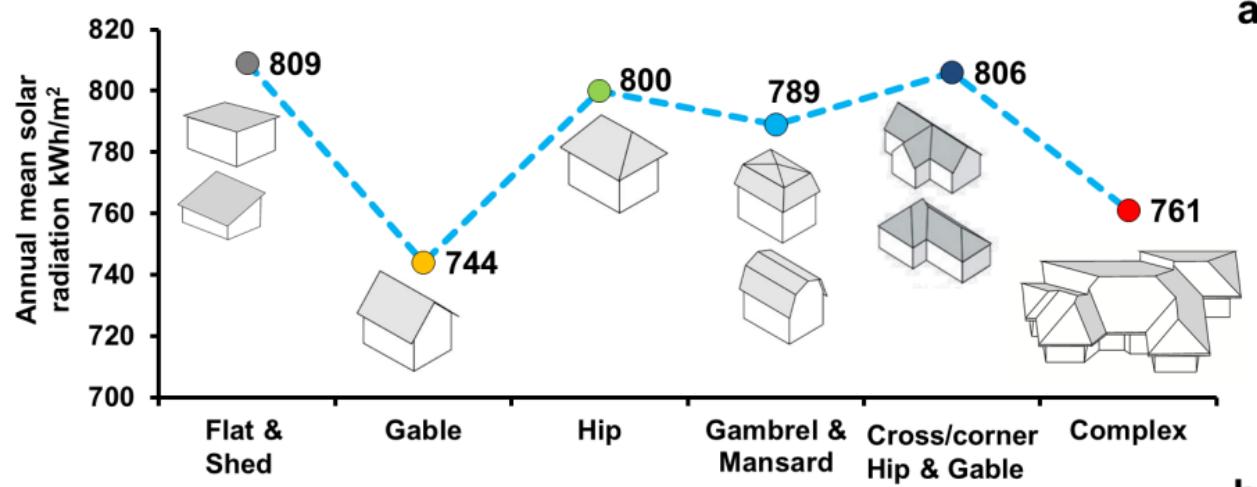
Flat & Shed Gable Hip Gambrel & Mansard Cross/corner Gable & Hip Complex







a**b**



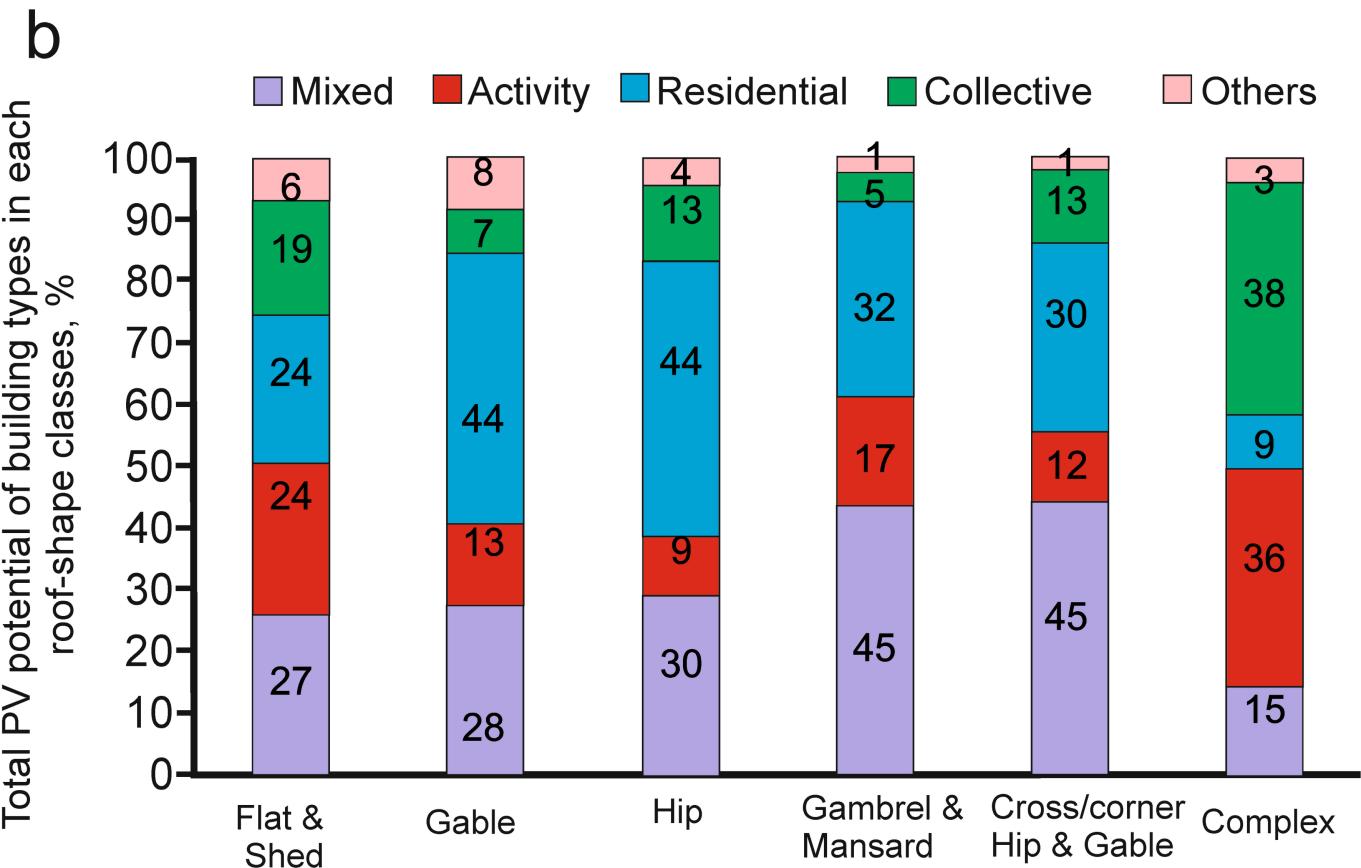
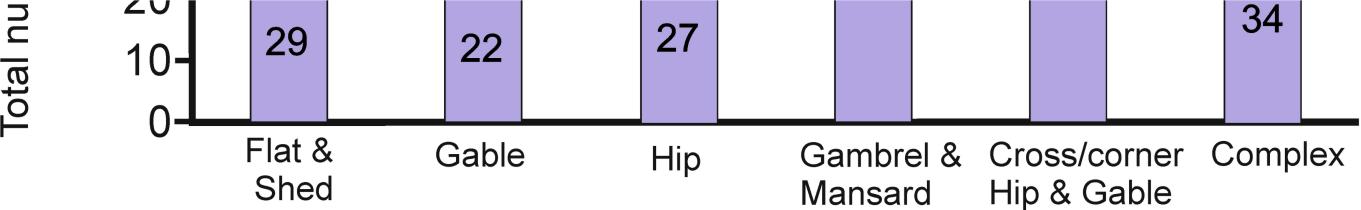


Table 1. Schematic presentation of different roof-shape classes

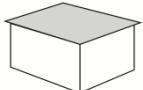
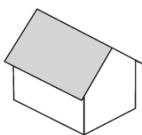
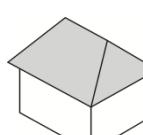
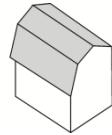
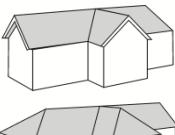
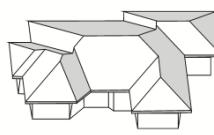
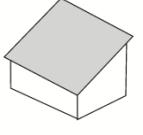
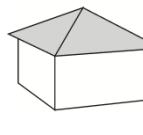
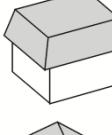
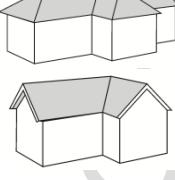
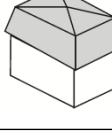
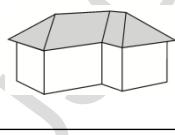
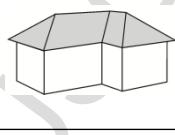
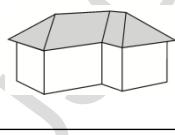
Roof classes					
Flat & Shed	Gable	Hip & Pyramidal	Gambrel & Mansard	Cross and corner Gable & Hip	Complex
0	1	2	3	4	5
					
					
					
					
					

Table 2. The description of 35 features used in the feature selection when classifying roof shapes

Number (35 features)	Feature type description
1	Number of roof surfaces for each building
9	Slope distribution (10 degree bin width) from 0 to 90 degree
9	Percentage of total roof area for each roof slope category
8	Aspect distribution (45 degree bin width) from 0 to 360
8	Percentage of total roof area for each roof aspect category

Table 3. Total number of labelled data for each roof-shape class/group, training data for each roof-shape class (75% of total labelled data), and testing data for each roof-shape class (25% of total labelled data)

Class	Flat & Shed	Gable	Hip & Pyramidal	Gambrel & Mansard	Cross and corner Gable & Hip	Complex
Total number of labelled data	127	127	124	118	115	106
Training data	95	95	93	89	86	80
Testing data	32	32	31	29	29	26

Table 4. The results of the accuracy matrix for testing data for each roof-shape class (66% overall accuracy)

Testing data	Flat & Shed	Gable	Hip & Pyramidal	Gambrel & Mansard	Cross and corner Gable & Hip	Complex	Accuracy for each class (%)
	0	1	2	3	4	5	
0 (32)	30	0	0	0	0	2	94
1 (32)	4	18	6	4	0	0	56
2 (31)	0	8	15	4	3	1	48
3 (29)	1	2	4	18	4	0	62
4 (29)	2	0	5	2	15	5	52
5 (26)	1	0	0	2	1	22	85