Machine learning and geographic information systems for large-scale wind energy potential estimation in rural areas

Dan Assouline\textsuperscript{1}, Nahid Mohajeri\textsuperscript{2}, Dasaraden Mauree\textsuperscript{1}, Jean-Louis Scartezzini\textsuperscript{1}

\textsuperscript{1}Solar Energy and Building Physics Laboratory, Ecole Polytechnique Federale de Lausanne, Station 18, 1015 Lausanne, CH
\textsuperscript{2}Sustainable Urban Development Programme, Department for Continuing Education, University of Oxford, Oxford, UK
E-mail: dan.assouline@epfl.ch

Abstract.
Clean, safe, affordable and available in the long-term, wind is one of the most promising sources of renewable energy. Its optimized and profitable use, however, requires an estimation of the potential in locations of interest, given its very volatile behavior in various settings. In the present study, we propose a methodology using a combination of Machine Learning (Random Forests), Geographic Information Systems and wind parametric models to estimate the large-scale theoretical wind speed potential in rural areas over the entire Switzerland. The monthly wind speed over rural areas is estimated based on wind speed measurements and several meteorological, topographic, and wind-specific features available across the country. Wind speed values and their associated uncertainty are computed at the scale of 200 x 200 [m$^2$] pixels covering the territory, at a typical height for rural commercial wind turbine installation, that is, $z=100m$. The developed methodology, is, however, applicable to any large region, given the availability of data of interest. The results show that in the case of Switzerland, wind turbines could approximately represent an non-negligible installed power capacity of, for each pixel and for each turbine installation, on average 80 kW in Swiss rural areas, and up to 1600 kW in most suitable pixels.

1. Introduction
Given its clean nature and the efficiency of modern wind turbines, wind energy has become very popular over the last decade and have seen its installed capacity growing all around the world [1]. As a result, many studies have been conducted in order to exploit the electricity generated by wind turbines in large regions and countries. As regards the wind energy potential, wind speed is the most critical variable to be estimated. The theoretical wind energy potential is therefore expressed by the wind speed at locations and heights of possible turbine installation.

In the literature, wind potential mapping studies have been conducted using multiple methodologies. A common strategy consists of performing statistical analysis, notably parameter estimation for the Weibull Probability Distribution Function, commonly considered to describe the distribution of wind speed [2, 3]. Several other studies use a multi-criteria decision methodology together with GIS data processing to compute a wind energy potential. [4, 5, 6].
Machine Learning (ML) methods have been explored for multiple environmental variables, including wind speed. The focus of ML methods in this domain, however, has been on forecasting rather than mapping tasks [7]. Regarding mapping applications, Neural Networks (NNs) became popular in the 2000’s for energy-related estimations [8], and several mapping studies were proposed using NNs trained with wind data. Other ML methods were recently used for wind speed spatial estimation, including ensemble learning and kernel methods [9, 10, 11, 12].

Concerning the (rural) wind potential in Switzerland specifically, the main studies have been mentioned in the previous paragraph [9, 11]. The data itself, however, is often not publicly available. The main source of available wind maps Switzerland is offered by the Swiss Wind Atlas, of the Swiss Federal Office of Energy (SFOE) (www.uvek-gis.admin.ch/BFE/storymaps/EE_Windatlas). These maps are computed using Weibull distributions fitted with parameters estimated based on wind measurements.

To perform an accurate mapping of monthly wind speed, we here propose a hybrid methodology that combines Machine Learning (Random Forests) and GIS. It is then applied for the theoretical potential estimation of wind energy in Swiss rural areas. Given the availability of data and the typical resolution of the data sources, we consider a pixel grid of size (200 × 200) [m²] covering the entire Swiss territory. The estimations of wind speed are therefore performed within each of these pixels.

The steps of the methodology are as follows: (i) collect data and extract significant features impacting the behavior of wind speed in Switzerland (meteorological, topographic, and other wind-related variables), (ii) train monthly Random Forests (RFs) with wind monitored data at 10m as output and the features previously extracted as inputs, (iii) use trained RFs to estimate monthly wind speed maps at z=10m in rural areas, (iv) vertically extrapolate the wind speed to a height of 100m, using a traditional log wind profile, (v) compute the potential wind power generated by turbines over the territory.

2. Data and methods

2.1. Machine Learning / Random Forests

Machine Learning (ML) models learn patterns from data (examples) and are able to perform predictions/estimations by reproducing similar patterns. If we denote $Y$ an output variable (the target) and $X_1, ..., X_d$ input variables of interest (features), the principle of a supervised regression task in ML is as follows: given some training data $(x_i, y_i) (i = 1, ..., N)$ representing couples of (input,output) for $N$ points (training samples), a regression task aims at learning a function $\psi : \mathbb{R}^d \rightarrow \mathbb{R}$ using the training data so that predictions $\psi(x)$ are the closest possible to the corresponding output $y$. Each $x_i \in \mathbb{R}^d$ is an input vector and $y_i \in \mathbb{R}$ is the corresponding output value (label) for point $i$. To measure the performance of the model, the labeled data (known examples) is split into a training set (75% of the data, used to train the model) and a test set (25% of the data, used to test the model). We use the Root Mean Square Error (RMSE) and the normalized RMSE (NRMSE) as error metrics within the test set.

Random Forests (RFs) [13] is a popular method which is part of the ensemble learning family of ML. The main concept of the algorithm is to train many decision trees based on resampled versions of the training data (a low performance model) and aggregate their prediction in order to obtain one model with good performance, as the aggregation reduces the variance of the prediction. RFs has numerous practical advantages, including its speed, its robustness with regards to hyperparameters, and several embedded measures offering information about the a trained model. These latter measures include the Variable Importances, offering a metric to estimate the impact of each feature on the model, and the Out-Of-Bag score (OOB), giving information about how performant the regressor is compared with a simple average estimator. The OOB score ranges from -1 to 1, 0 showing the same performance as an average estimate. For more details about the Random Forests please see [13]. A useful extension of RF, called
Table 1. Datasets used in the study, along with their characteristics and reference.

| Data          | Data type | Region | Inputs                        | Period     | Time resolution | Spatial resolution | Error     | Source                  |
|---------------|-----------|--------|-------------------------------|------------|-----------------|--------------------|-----------|-------------------------|
| Sunshine duration | Time series | Switzerland | Meteo stations               | 1981-2010 | Monthly means   | 66 stations        | Negligible | MeteoSwiss [15, 16]     |
| Precipitation  | Time series | Switzerland | Meteo stations               | 1981-2010 | Monthly means   | 417 stations       | 4-40%     | MeteoSwiss [15, 16]     |
| Temperature    | Time series | Switzerland | Meteo stations               | 1981-2010 | Monthly means   | 91 stations        | Negligible | MeteoSwiss [15, 16]     |
| Cloud Cover    | Time series | Switzerland | Human observations          | 1981-2010 | Monthly means   | 23 stations        | NA        | MeteoSwiss [15, 16]     |
| Air Pressure   | Time series | Switzerland | Meteo stations               | 2000-2017 | Daily means     | 257 stations       | NA        | MeteoSwiss [15, 16]     |
| Wind speed     | Time series | Switzerland | Meteo stations               | >2000      | Monthly means   | 197 stations       | NA        | MeteoSwiss [15, 16]     |
| Digital Height Model (DHM) | Raster | Switzerland | Elevation maps + aerial images | NA | NA | 25m × 25m | ±2m | Swisstopo [17]      |
| CORINE land cover data (Switzerland) | Vector polygons | Switzerland | + various vector and raster layers | 2012 | NA | NA | ±3-8m | WSL [18]        |
| Building footprint data (TLM3D) | Vector polygons | Switzerland | LiDAR + 3D buildings | 2018 | NA | NA | ±30-50cm | Swisstopo [19]     |

Quantile Regression Forests (QRFs) [14], allows to estimate the uncertainty attached to an RF prediction by creating a Prediction Interval (PI) around a prediction with a certain confidence level (e.g. 95% confidence), so that, schematically: pred. − Plow ≤ real value ≤ pred. + Pup.

2.2. Data and feature extraction
A crucial step in ML training is the extraction of relevant features which are available for all training points and have an impact on the output variable (here the wind speed). The features used in the study are extracted from several Swiss datasets which are all presented in Table 1. A Digital Height Model (DHM), down-sampled to a resolution of 200 × 200 [m²], is used as the basis of the mesh pixel grid used in the study (pixel = points). For each pixel, if data is available, we simply allocate the value of a data point to the closest pixel identified by its centroid. We here briefly present the extracted features and processing leading to their extraction.

Topographic features (8) are extracted from the DHM using raster processing tools of the Spatial Analyst toolbox from ArcGIS: longitude, latitude, altitude, the terrain slope, the terrain aspect (also known as exposition or direction of the terrain), and terrain plan curvature, longitudinal curvature and transverse curvature.

Meteorological features (5) are based on meteorological maps estimated from weather measurement data (Table 1) and RF models trained in a previous study [20]: sunshine duration, air temperature, precipitation, cloud cover and air pressure.

Wind-specific features (2) are also considered: roughness length and “neighbor roughness length”. The first feature is a common wind parameter measuring how “smooth” the terrain is and defining the height of null wind speed in a log wind profile (see Eq. 1). Based on land cover (CORINE, Table 1), we use tabled values suggested by [21]. The second feature, for each pixel, is defined by the average of the roughness length of all its adjacent pixels.

3. Wind speed estimation in Switzerland
3.1. Estimation at 10 m
The computation of the wind speed at 10 m is based on the training of Random Forests (RF) models together with daily wind speed measurement data from MeteoSwiss, available for 197 stations (Table 1). The measurements are aggregated monthly, averaging the speed for each month from 2000 to 2017. The measurement values are then spatially aggregated to match the
Table 2. Errors related to the building of monthly wind speed in rural areas using RF models.

| Month | Wind speed at 10m [\(u_r(10)\)] | Testing errors | Mean Prediction Errors |
|-------|---------------------------------|----------------|------------------------|
|       | \(E_R\) \(\text{RMSE} [\text{m/s}]\) | \(E_{NR}\) \(\text{NRMSE} [%]\) | \(|\text{OOB}| [-]\) | \(\text{PE}_{s,low}\) \(\text{m/s}\) | \(\text{PE}_{s,up}\) \(\text{m/s}\) | \(\text{PE}_s\) \(\text{m/s}\) |
| Jan.  | 0.74 23.13 0.57          |               | 2.09 1.53 1.81         |               |
| Feb.  | 0.80 24.61 0.55          |               | 1.34 1.48 1.41         |               |
| Mar.  | 0.68 20.57 0.51          |               | 0.70 1.63 1.16         |               |
| Apr.  | 0.70 22.21 0.42          |               | 1.12 1.34 1.23         |               |
| May.  | 0.62 20.73 0.42          |               | 0.40 1.58 0.99         |               |
| Jun.  | 0.60 21.19 0.41          |               | 0.65 1.37 1.01         |               |
| Jul.  | 0.59 21.12 0.43          |               | 0.28 1.42 0.85         |               |
| Aug.  | 0.59 22.97 0.46          |               | 0.82 1.19 1.00         |               |
| Sep.  | 0.57 22.04 0.43          |               | 1.09 1.10 1.10         |               |
| Oct.  | 0.66 24.26 0.50          |               | 0.89 1.29 1.09         |               |
| Nov.  | 0.77 24.92 0.56          |               | 1.13 1.50 1.31         |               |
| Dec.  | 0.80 24.53 0.62          |               | 2.18 1.52 1.85         |               |

200 \(\times\) 200 [m^2] pixel resolution: if a station is alone within a pixel, its measurements represent the whole pixel; if multiple stations are located in one pixel, the monthly measurements are averaged within the pixel. We thus obtain 159 pixels with monthly wind measurements. From the 159 pixels, we select “rural pixels”, meaning those which do not include any building (defined by the TLM3D building data, presented in Table 1). We obtain 118 rural pixels with known wind speed, which will represent the data points for our training process.

The monthly RF models for estimating wind speed in rural areas are trained based on the 118 rural training points, using the wind speed pixel values as labels and the 15 features extracted in section 2.2 as input features. The training features are obtained by sampling the feature maps (previously derived) at the location of training pixels. The RF models are trained for each month separately, using monthly features. The testing errors (RMSE, NRMSE, and OOB score) for the training of the monthly RF models are shown in Table 2. Prediction Intervals (PIs) (with 95% confidence) are computed for each monthly model (Figure 1a shows PIs for 30 random points in February), using QRFs, offering a measure of uncertainty associated with the estimations. Also, prediction errors (average of upper and lower width of PIs) are computed: (i) over a sample of 1000 unobserved points for each month to observe the uncertainty variations with time (Table 2) and (ii) for all rural pixels in the yearly prediction to observe the uncertainty variations with space (Figure 1b).

The error table (Table 2) shows an NRMSE is around 22%, which is a reasonable error given the often very unstable behavior of wind speed through the months. Furthermore, it can be observed that the models perform better in the summer than in the winter, which is intuitively explained by generally higher and more fluctuating speed values in the winter. The spatial trend of the uncertainty showed by Fig. 1b shows, as expected, larger values (\(\pm 1.5\) to 3 m/s) in locations of high altitude, particularly mountain areas, where wind values can reach higher levels and tend to have a larger range of values. Note that high uncertainties may be the result of a lower density of measurement stations. The Variable Importances embedded in the RF algorithm are computed for each monthly RF model and averaged through the month to extract the importance of each predictor (feature) in the overall training process. Results show that as expected, the most important features are, in decreasing order: longitudinal curvature, plan curvature, air temperature, altitude, transverse curvature, and pressure.

The monthly RF models are used in order to estimate the monthly wind speed in all unknown rural areas of Switzerland. The obtained rural monthly wind maps are averaged through the year in order to obtain a rural yearly wind speed map at 10m, shown in Figure 2.
3.2. **Extrapolation at 100 m**

While wind turbines can be installed at various heights in rural areas, commercial turbines are often installed at a height of around 100m, where the wind is stronger and more stable, in order to achieve the maximum potential at the location of interest. Given the estimated wind speed in rural areas at 10m, it is possible, in the boundary layer [22], to vertically extrapolate the wind speed to another height using a log-law (in the limit of statistically neutral flow [22]):

\[
u_r(z) = \frac{u^*}{\kappa} \ln \frac{z - z_{d,r}}{z_{0,r}} \approx \frac{u^*}{\kappa} \ln \frac{z}{z_{0,r}}, \text{ for } z > z_{0,r}
\]

where \(u_r(z)\) is the rural wind speed at a height of \(z\), \(u^*\) is the rural friction velocity, \(\kappa\) is the Von Karman constant, and \(z_{0,r}\) and \(z_{d,r}\) the rural roughness length and displacement of the considered rural area (\(z_{d,r}\) is negligible outside of forest areas, which we exclude [23]).

We adopt the assumption that the friction velocity remains constant along the profile above rural areas (until 100m) [23]. Taking the ratio of Eq. 1 evaluated respectively at \(z = 100\) and \(z = 10\), we obtain, for any rural pixel \(j\):

\[
u_r^j(100) = \nu_r^j(10) \frac{\ln(100/z_{0,r}^j)}{\ln(10/z_{0,r}^j)} \quad \text{where } \nu_r^j(10) \text{ and } \nu_r^j(100)
\]

are respectively the rural roughness length and the wind speed at 10m, both previously estimated in the study. The wind speed at 100 m is computed over all rural pixels, at a monthly resolution, and is averaged over the months to obtain a yearly map. The speed values are compared with Swisstopo across Switzerland, leading to an RMSE error of 1.6 m/s, and an NRMSE of 35%.

Using an equation, it is possible to estimate the wind power \(P_w\) generated by a turbine:

\[P_w = C_p \rho A_w u^3/2\]

where \(C_p\) is the coefficient of performance of the turbine, \(\rho\) the air density, \(A_w\) is the area spanned by the rotor and \(u\) the wind speed. Let us consider a turbine with a diameter of 100m, \(\rho\) of 1.2 kg/m\(^3\), and a typical \(C_p\) curve: \(C_p=0.45\) at a speed of 5-10m/s, \(C_p=0.1\) to 0.4 elsewhere (https://en.wind-turbine-models.com/turbines/130-enercon-e-101). Using the previously estimated wind speed, the power is then computed in all rural pixels to obtain the power map shown in Figure 3. It amounts to a potential installed power of, for each pixel and for each turbine installation, in a typical year, on average 80 kW in Swiss rural areas, and up to 1600 kW in the most suitable pixels.

4. **Conclusion**

This study proposes a methodology combining Machine Learning, GIS and wind models to estimate the theoretical wind speed potential over Switzerland, that is, the monthly wind speed over rural areas, at turbine height of 100m, for \(200 \times 200\) [m\(^2\)] pixel covering Switzerland. The
methodology, showing a significant potential in the country, is generalizable to any large area, given sufficient availability of data. Future work includes further processing for an accurate estimation of the potential energy generated by the turbines, and the extension to urban areas.

Acknowledgments
This research project is financially supported by the Swiss Innovation Agency Innosuisse and is part of the Swiss Competence Center for Energy Research SCCER FEEB&D. It is also partly supported by the Swiss National Science Foundation under Mobility Fellowship P300P2 174514.

References
[1] Global wind statistics 2017 accessed: 2018-11-15
[2] Cellura M, Cirrincione G, Marvuglia A and Miraoui A 2008 Renewable energy 33 1237–1250
[3] Mentis D, Hermann S, Howells M, Welsch M and Siyal S H 2015 Renewable Energy 83 110–125
[4] Aydin N Y, Kental E and Duzgun S 2010 Renewable and Sustainable Energy Reviews 14 364–373
[5] Enevoldsen P andPermien F H 2018 Journal of Renewable Energy 2018
[6] Mahdy M and Bahaj A S 2018 Renewable Energy 118 278–289
[7] Lei M, Shiyan L, Chuanwen J, Hongling L and Yan Z 2009 Renewable and Sustainable Energy Reviews 13 915–920
[8] Kalogirou S A 2001 Renewable and sustainable energy reviews 5 373–401
[9] Foresti L, Tuia D, Kanevski M and Pozdnoukhov A 2011 Stochastic Environmental Research and Risk Assessment 25 51–66
[10] Robert S, Foresti L and Kanevski M 2013 International Journal of Climatology 33 1793–1804
[11] Veronesi F, Grassi S and Raubal M 2016 Renewable and Sustainable Energy Reviews 56 836–850
[12] Grassi S, Veronesi F and Raubal M 2015 Proceedings of EWEA Conference 2015, Porte de Versailles Pavillon 1, Paris, France, 17-20 November 2015 (European Wind Energy Association (EWEA 2015))
[13] Breiman L 2001 Machine learning 45 5–32
[14] Meinshausen N 2006 Journal of Machine Learning Research 7 983–999
[15] Idaweb, link = https://gate.meteoswiss.ch/idaweb, note = Accessed: 2018-01-29
[16] Meteoswiss, link = www.meteoswiss.admin.ch, note = Accessed: 2018-01-29
[17] geodata4edu, link = http://geodata4edu.ethz.ch, note = Accessed: 2018-01-29
[18] swissbuildings3d 2.0, link = https://shop.swisstopo.admin.ch/fr/products/landscape/build3d2, note = Accessed: 2018-01-29
[19] Steinmeier C 2013 Corine land cover 2000/2006 switzerland
[20] Assouline D, Mohajeri N and Scartezzini J L 2018 Applied energy 217 189–211
[21] Silva J, Ribeiro C and Guedes R 2007 Proceedings of the European Wind Energy Conference, Milan, Italy vol 710 p 110
[22] Stull R B 2012 An introduction to boundary layer meteorology vol 13 (Springer Science & Business Media)
[23] Oke T R, Mills G, Christen A and Voogt J A 2017 Urban climates (Cambridge University Press)