Les données éclairent les toits suisses

Lever le voile sur le potentiel photovoltaïque suisse à l'ère numérique | Si les approximations des modèles et le manque de données à haute résolution empêchaient encore récemment d'estimer de manière réaliste le potentiel photovoltaïque en toiture à grande échelle, le big data et l'apprentissage automatique offrent désormais une approche avec une résolution temporelle et spatiale sans précédent.

ROBERTO CASTELLO

Sur la route longue et sinueuse menant à la décarbonisation du secteur de l'énergie, le déploiement du photovoltaïque en toiture (rooftop photovoltaics, RPV) a fait l'objet d'une attention croissante au cours des dernières années. Pour pouvoir formuler des politiques efficaces en vue de son intégration dans l'environnement construit, une évaluation quantitative du potentiel électrique du RPV est nécessaire. Or, cette dernière nécessite des données précises et à haute résolution afin de saisir la variation spatiale et temporelle de l'électricité produite.

Si l'estimation du potentiel photovoltaïque à l'échelle d'une ville est très utile pour l'élaboration de politiques et la planification de stratégies, elle présente néanmoins certains inconvénients: elle n'est, par exemple, pas adaptée pour estimer le retour sur investissement des panneaux solaires pour les ménages privés. Ceci rend, par conséquent, plus difficile l'intégration de ces derniers dans un système électrique décentralisé, avec stockage d'énergie et connexion au réseau électrique.

Très peu de méthodologies ont permis pour l'instant d'estimer le potentiel du RPV à grande échelle; et, jusqu'à récemment, aucune n'a réussi à le faire avec une résolution temporelle horaire et une résolution spatiale à l'échelle du toit, tout en tenant compte de la propagation systématique des incertitudes introduites par le processus de modélisation. L'absence d'études à l'échelle nationale à de telles résolutions s'explique par les difficultés de calcul liées au traitement des ensembles de données requis et à la gestion des données manquantes, ainsi que par l'absence d'un ensemble complet de données relatives aux zones étudiées. Pour relever ces défis, des approches [1-3]

Figure: Charlie Wilde/pixabay

basées sur l'apprentissage automatique (machine learning, ML) et les données massives (big data) ont permis de réaliser des progrès significatifs dans cette estimation en Suisse, en améliorant la précision du modèle et en fournissant une quantification de l'incertitude.

Le pouvoir aux données

Chaque fois que des variables quantitatives sont mesurées, des données sont produites. Comme il est souvent impossible de sonder une population tout entière, on utilise les données d'un échantillon pour décrire au mieux la distribution inconnue de la population. Un petit échantillon peut cependant être biaisé par rapport à la distribution réelle : donc plus l'échantillon est inclusif, plus cette description sera exacte et précise.

La grande quantité de données spatiales disponibles dans les bureaux fédéraux suisses constitue la source de données environnementales et urbaines nécessaires à la conception d'un modèle précis pour l'estimation du potentiel RPV. Néanmoins, dans le cadre de l'ambitieux projet de prévision du potentiel RPV à haute résolution spatio-temporelle pour l'ensemble de la Suisse, les chercheurs sont inévitablement confrontés à l'absence de certaines données ou à l'indisponibilité de mesures dans certaines zones géographiques. Or, un modèle de régression peut apprendre la tendance d'une certaine variable par rapport à d'autres, et ses paramètres peuvent être utilisés pour déduire la valeur d'une variable lorsque les mesures manquent.

Les techniques d'apprentissage automatique sont particulièrement efficaces pour traduire les dépendances linéaires et non linéaires d'une variable par rapport à d'autres dans un modèle compact et « prêt à l'emploi ». Les algorithmes de régression supervisés apprennent à partir d'un ensemble de données qui contient des paires constituées d'entrées (features) et de leurs valeurs de sortie connues (targets). Cette partie, appelée entraînement (training) des algorithmes, permet de découvrir les interdépendances et de les encoder dans un modèle. Le modèle est ensuite appliqué pour prédire des valeurs de sortie inconnues pour un nouvel ensemble d'entrées.

Le modèle le plus simple est la régression linéaire, mais les techniques de ML offrent la possibilité d'en créer des plus complexes, et aussi plus exacts. Essayons par exemple de prédire le rayonnement

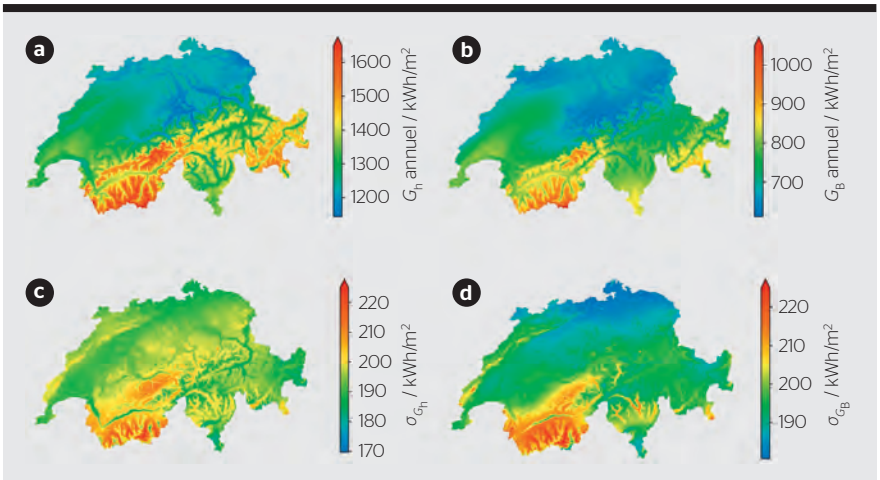

Figure 1 Distribution spatiale du rayonnement solaire horizontal annuel prévu (a) et de sa composante directe (b), ainsi que leurs incertitudes totales respectives (c) et (d).

Figure 2 Installation virtuelle de panneaux photovoltaïques après retrait des superstructures du toit : la meilleure configuration de panneaux orientés verticalement et horizontalement est sélectionnée.

solaire incident sur chacun des pixels de 200 x 200 m² couvrant la Suisse, ce qui représente l'un des défis de l'étude décrite dans [1]. Il est toujours possible de trouver dans la littérature scientifique des formules physiques à appliquer pour calculer le rayonnement solaire sur une certaine zone géographique. Ce serait toutefois une façon lente et laborieuse d'atteindre cet objectif. Un autre moyen consiste à s'appuyer sur les quelques mesures de l'irradiation solaire déjà existantes et à essayer de modéliser la manière dont celle-ci dépend d'un certain nombre de features.

Les mesures satellitaires et la télédétection (remote sensing, RS) fournissent une couverture à une résolution limitée : l'objectif est de prédire les valeurs pour des pixels à plus haute résolution spatiale, là où aucune mesure n'est disponible. Le principe consiste à laisser la machine apprendre quelles sont les corrélations entre les features (la position géographique du pixel, son altitude, etc.) et les targets (dans ce cas, le rayonnement solaire incident sur le pixel). Une fois la relation apprise (par le biais du training), la machine est capable de donner pour chaque nouveau pixel (et donc ses features donnés en entrée) sa meilleure estimation du rayonnement solaire incident, ainsi que son incertitude statistique et celle due au modèle.

Combien d'énergie le soleil procure-t-il ?

Une première étape consiste à estimer le **potentiel « physique »**, déterminé

Figures : [1]

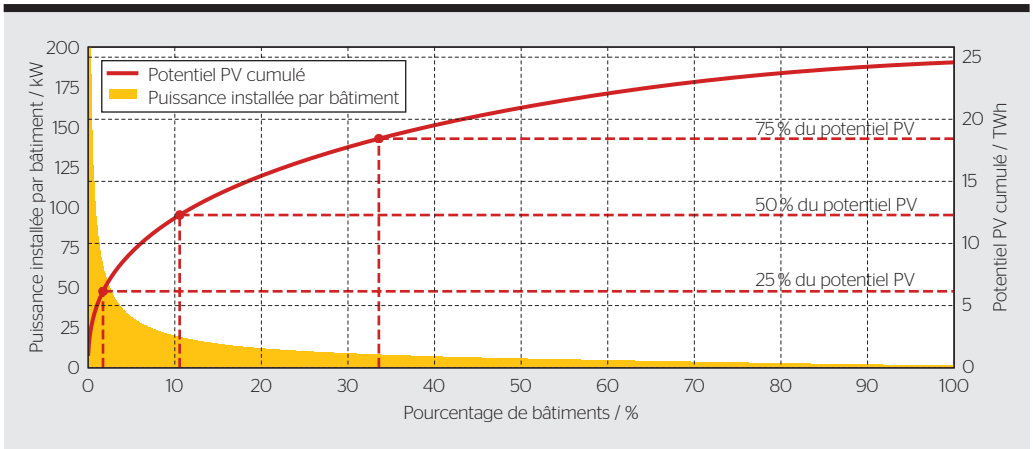

Figure 3 Puissance installée par bâtiment et potentiel technique RPV cumulé en fonction du pourcentage de bâtiments considérés comme appropriés.

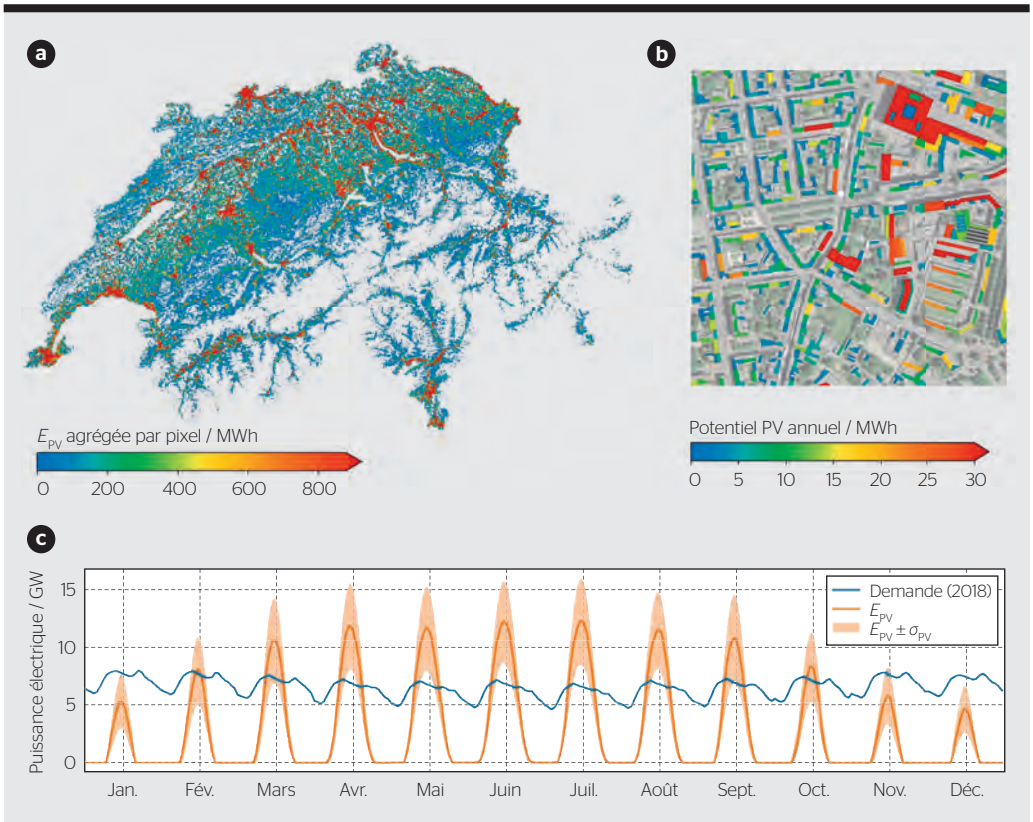

Figure 4 Distribution spatiale du potentiel technique annuel, agrégé en pixels de 500 x 500 m² à des fins de visualisation (a), potentiel technique annuel pour les toitures appropriées d'un pixel de 500 x 500 m² choisi au hasard dans la ville de Genève (b), et profils mensuels moyens horaires du potentiel technique additionné pour toutes les toitures considérées comme appropriées, leurs incertitudes, et demande suisse d'électricité en 2018 (c).

par le rayonnement solaire horizontal (en W/m²) mesuré à la surface terrestre. Les données météorologiques enregistrées par le RS sont des données précieuses pour calculer ce potentiel. Les satellites fournissent une couverture spatiale uniforme à haute résolution temporelle. Les méthodes basées sur le ML utilisent ces mesures à large résolution spatiale (environ 1,6 x 2,3 km²) comme targets pour apprendre les interdépendances cachées avec les features géographiques et de terrain (la pente, la courbure, etc.). Seules, les données géographiques, c'est-à-dire l'altitude, la latitude et la longitude, sont suffisantes pour concevoir un modèle de prévision précis qui estime le rayonnement solaire horizontal. Le modèle le plus performant, c'est-à-dire celui qui donne l'erreur quadratique moyenne la plus faible sur un ensemble de données de test indépendant, est choisi pour prédire les valeurs targets inconnues sur une grille plus fine de 200 x 200 m². Dans ce cas, un ensemble d'algorithmes de ML, qui prédit la variable en faisant la moyenne des résultats de ses différents estimateurs [4], donne de très bons résultats, en fournissant également l'estimation de l'incertitude du modèle.

La **figure 1** montre la distribution spatiale du rayonnement solaire horizontal (formé par les composantes directe et diffuse) annuel estimé, et son incertitude correspondante, pour l'ensemble de la Suisse. Le rayonnement solaire le plus élevé et la plus grande incertitude se trouvent à haute altitude dans le sud du pays, où les phénomènes météorologiques extrêmes sont fréquents. La majorité des bâtiments sont cependant situés à des altitudes plus basses, sur le Plateau suisse. La même carte est également compilée pour 288 pas de temps « monthly-mean-hourly, MMH » distincts : cela

signifie que le potentiel physique est estimé pour chacune des 24 heures d'une journée type pour chacun des douze mois, comme une moyenne sur tous les jours du mois. Ce niveau de résolution est essentiel pour modéliser la saisonnalité et la variation intrajournalière du potentiel physique.

La fraction des toits qui peut être utilisée

Déterminer le nombre de panneaux solaires qui peuvent être installés sur un toit implique un examen attentif de la géométrie de ce dernier, de son orientation et de la fraction quotidienne d'exposition au soleil compte tenu des bâtiments et des arbres environnants. Effectuer cette opération pour les 9,6 millions de toits qui se trouvent en Suisse constitue un objectif ambitieux.

Le **potentiel « géographique »** tient donc compte de la géométrie, des superstructures de toit existantes (cheminées, fenêtres, etc.), des effets d'ombrage et de la visibilité du ciel. Il est estimé pour chaque surface de toit en ajustant les composantes directes, diffuses et réfléchies du rayonnement solaire préalablement calculées en fonction de l'inclinaison du toit et d'autres facteurs. Le résultat est ensuite convolué avec la surface disponible pour l'installation d'un système RPV.

Le rayonnement diffus est corrigé en fonction du facteur de visibilité du ciel et le rayonnement direct en fonction de la partie non ombragée du toit à chaque heure de la journée. Ensuite, un algorithme géospatial installe virtuellement les panneaux en projetant des polygones rectangulaires sur le toit incliné. Les panneaux sont installés en alignement horizontal et vertical, et la configuration avec le nombre le plus élevé de panneaux est choisie pour chaque toit (voir l'exemple de la **figure 2**). Sur les toits plats, les panneaux sont placés en rangées orientées vers le sud et inclinées selon un angle de 30°.

Mais que se passe-t-il si un toit comporte en réalité plusieurs superstructures ? La surface effective est alors d'autant plus réduite par rapport à la surface totale du toit. Une carte détaillée de ces superstructures n'est disponible que pour l'un des 26 cantons suisses (celui de Genève) : alors comment deviner de manière réaliste l'impact de ces superstructures sur les toits restants ? L'apprentissage automatique vient à nouveau à la rescousse. En utilisant la surface libre pour l'installation de panneaux sur les toits du canton de Genève (donc avec des superstructures cataloguées) comme target dans l'algorithme de ML et des features liés aux toitures et bâtiments (aspect et angle d'inclinaison, type de bâtiment, période de construction, y compris la surface avec panneaux résultante du bâtiment si les superstructures ne sont pas prises en compte), il est possible de prévoir la surface nette réelle pour l'installation de panneaux sur chaque toit de Suisse.

Les effets d'ombrage, qui rendraient inutile l'installation de systèmes photovoltaïques sur certaines parties du toit, sont aussi quantifiés à chaque heure de la journée. Un algorithme [5] permet de modéliser l'ombre pour chaque pixel d'un modèle numérique de surface (digital surface model, DSM) à une position donnée du soleil (c'est-à-dire à une certaine heure). La précision de l'estimation de la surface ombrée dépend essentiellement de la résolution du DSM. La résolution standard d'un DSM couvrant toute la Suisse est de 2 x 2 m² : seul le canton de Genève dispose d'un DSM à plus haute résolution, de 0,5 x 0,5 m². Là encore, un algorithme d'apprentissage automatique [6] est employé pour prédire le coefficient qui serait obtenu avec une carte ayant une résolution de 0,5 x 0,5 m², en utilisant comme features la valeur du coefficient calculée sur le DSM de 2 x 2 m² et d'autres caractéristiques du toit et des bâtiments. Cette procédure a considérablement amélioré la quantification de la surface ombragée des toits à l'échelle de la Suisse, un ingrédient essentiel pour calculer le potentiel en électricité finale.

Combien d'électricité pour la Suisse ?

Le **potentiel «technique»** est défini comme la production d'électricité (en kWh) pour chaque toit. Il est obtenu en multipliant le potentiel géographique (c'est-à-dire le flux d'énergie solaire ne frappant que la zone appropriée sur les toits) par le rendement du panneau solaire et le facteur de performance. Il est également tenu compte du rendement du convertisseur DC/AC (calculé pour chaque pas de temps en utilisant la température maximale au sol) et d'autres pertes telles que l'encrassement ou la dégradation des panneaux.

Pour obtenir un potentiel réel, les toits ayant une petite surface disponible (moins de 8 m²) sont exclus, car cette valeur équivaut au seuil inférieur de faisabilité économique. De même, tous les toits orientés vers le nord n'entrent pas dans le calcul, leur contribution au potentiel étant relativement faible. Ces critères ramènent la fraction appropriée de la surface de toitures à environ 57 % de la surface disponible (environ 270 km²) sur 2,3 millions de bâtiments. Le potentiel technique annuel maximum qui peut en être extrait est estimé à 24 ± 9 TWh, ce qui correspond à une production capable de couvrir 40 % de la demande en électricité enregistrée en 2018.

L'intégration dans l'environnement construit

En plus de l'estimation annuelle du potentiel technique, cette méthodologie fournit plusieurs contributions pratiques pour une intégration efficace du RPV dans les villes. La **figure 3** montre comment 25 % du potentiel estimé peut être déjà réalisé en installant des panneaux photovoltaïques sur moins de 2 % des bâtiments, ceux présentant le potentiel le plus élevé (dotés de toits plats). Les 75 % du potentiel peuvent, quant à eux, être atteints en installant des panneaux photovoltaïques sur environ un tiers des bâtiments.

Une analyse du potentiel annuel pour toute la Suisse, dans la **figure 4**, montre que la contribution du RPV serait insuffisante pendant l'hiver et la nuit, alors qu'il y aurait un surplus pendant les heures de pointe et en été. Par conséquent, une combinaison ad hoc avec d'autres sources d'énergies renouvelables (comme l'éolien et la géothermie) ainsi qu'un couplage des secteurs (sector coupling) avec des systèmes de chauffage et de refroidissement devraient être envisagés pour exploiter au mieux le solaire. L'estimation horaire du potentiel pour chaque mois de l'année est utile pour concevoir les futurs systèmes énergétiques, afin de satisfaire instantanément la demande locale en électricité.

La haute résolution spatiale (à l'échelle du bâtiment) de cette méthodologie permet aux urbanistes d'évaluer la demande en électricité qui pourrait être satisfaite par l'installation de systèmes photovoltaïques sur les toits

existants dans un certain quartier, ainsi que d'estimer plus précisément le rendement attendu pour les nouveaux toits en tenant compte de leur taille, de leur inclinaison et de leur orientation. Les décideurs politiques peuvent agréger les résultats à différentes échelles spatiales, par exemple régionale ou nationale, afin de formuler des politiques efficaces pour intégrer le PV dans l'environnement construit.

Enfin et surtout, comment prévoir le RPV pour différents scénarios urbains et climatiques? L'étude comprend également une analyse qui quantifie l'impact de chacun des paramètres utilisés dans la méthode pour extraire le potentiel technique. Le rayonnement horizontal et la fraction de toit disponible sont les paramètres les plus sensibles: leur variation de 50% peut entraîner jusqu'à 40% de changement dans le potentiel RPV final. Par conséquent, cette méthodologie peut également être utilisée afin de prévoir la production en électricité attendue à l'avenir, pour des scénarios dans lesquels les conditions climatiques et la densité urbaine devraient avoir évolué.

Une contribution importante aux politiques de décarbonisation

Tout cela peut sembler très utile, mais avec quel degré de confiance peut-on se fier à ces résultats? Ces derniers proviennent d'un modèle statistique et sont donc intrinsèquement affectés par une incertitude. Par exemple, le choix spécifique d'un ensemble d'algorithmes de ML, qui prédisent la variable «target» en faisant la moyenne des résultats de plusieurs estimateurs, permet de quantifier l'incertitude liée au modèle sous la forme d'écarts types (standard deviation) des différentes estimations. Une autre incertitude liée aux données est également prise en compte à partir des résidus entre les targets et les prédictions du modèle. Les autres incertitudes se propagent ensuite selon une procédure basée sur la variance. L'étude suggère que, si la surface de toit disponible varie d'une standard deviation par rapport à sa valeur estimée, la production d'électricité peut changer d'environ 20%.

En résumé, l'utilisation de méthodes statistiques telles que l'apprentissage automatique pour extraire des informations significatives d'un vaste ensemble de données a permis d'atteindre une résolution sans précédent dans l'estimation du potentiel photovoltaïque des toits en Suisse et, en même temps, de fournir l'incertitude correspondante. Ce travail apporte une contribution importante aux politiques de décarbonisation en Suisse, car ses résultats permettent de modéliser à grande échelle les futurs réseaux électriques. La méthode est également transférable à toute région ou à tout pays disposant d'une base de données suffisamment homogène, et peut contribuer à la transition vers des systèmes énergétiques à faible teneur en carbone.

Références

[1] Alina Walch, Roberto Castello, Nahid Mohajeri, Jean-Louis Scartezzini, «Big data mining for the estimation of hourly rooftop photovoltaic potential and its uncertainty», Applied Energy, Vol. 262, 114404, 2020. doi.org/10.1016/j.apenergy.2019.114404

[2] Dan Assouline, Nahid Mohajeri, Jean-Louis Scartezzini, «Large-scale rooftop solar photovoltaic technical potential estimation using Random Forests», Applied Energy, Vol. 217, pp. 189-211, 2018. doi.org/10.1016/j.apenergy.2018.02.118

[3] Dan Assouline, Nahid Mohajeri, Jean-Louis Scartezzini, «Quantifying rooftop photovoltaic solar energy potential: A machine learning approach», Solar Energy, Vol. 141, pp. 278-296, 2017. doi.org/10.1016/j.solener.2016.11.045

[4] Haocheng Wang, Qing He, Tianfeng Shang, Fuzhen Zhuang, Zhongzhi Shi, «Extreme Learning Machine Ensemble Classifier for Large-Scale Data» (2015), In: Cao J., Mao K., Cambria E., Man Z., Toh KA. (eds) Proceedings of ELM-2014 Vol. 1. Proceedings in Adaptation, Learning and Optimization, Vol. 3. Springer, Cham. doi.org/10.1007/978-3-319-14063-6_14

[5] Markus Neteler, M. Hamish Bowman, Martin Landa, Markus Metz, «GRASS GIS: A multi-purpose open source GIS», Environmental Modelling & Software, Vol. 31, pp. 124-130, 2012. doi.org/10.1016/j.envsoft.2011.11.014

[6] Leo Breiman, «Random Forests», Machine Learning 45, pp. 5-32, 2001. doi.org/10.1023/A:1010933404324

Auteur

Dr **Roberto Castello** est collaborateur scientifique et chef de groupe au Laboratoire d'énergie solaire et de physique du bâtiment (Leso-PB) de l'École polytechnique fédérale de Lausanne.
→ EPFL, 1015 Lausanne
→ roberto.castello@epfl.ch

L'auteur tient à remercier Alina Walch (Leso-PB) pour sa contribution principale aux travaux présentés dans cet article ainsi que Dr Dan Assouline pour ses études pionnières sur ce sujet. Ce travail de recherche a été financé par le Fonds national suisse pour la recherche scientifique (FNS) dans le cadre du programme NRP75-Big Data et par le Centre de compétence suisse pour la recherche en énergie SCCER-FEEBD (Future Energy Efficient Buildings & Districts).

IN KÜRZE

Daten beleuchten Schweizer Dächer

Ermittlung des Photovoltaik-Potenzials der Schweiz im digitalen Zeitalter

Der Einsatz von PV-Anlagen auf Dächern (Rooftop Photovoltaics, RPV) hat in den letzten Jahren zunehmend an Aufmerksamkeit gewonnen. Um eine wirksame Politik für ihre Integration in die gebaute Umwelt zu formulieren, ist eine quantitative Bewertung des elektrischen Potenzials von RPV erforderlich.

Nur wenige Methoden konnten bisher das Potenzial von RPV auf nationaler Ebene abschätzen. Bis vor Kurzem war keine Methode in der Lage, dies mit stündlicher und räumlicher Auflösung auf Dachskala zu tun, unter Berücksichtigung der systematischen Ausbreitung der durch den Modellierungsprozess eingeführten Unsicherheiten. Dies ist nun möglich: Dank Ansätzen, die auf maschinellem Lernen und grossen Datenmengen basieren, ist es dem Labor für Solarenergie und Bauphysik der EPFL gelungen, bedeutende Fortschritte bei der Verbesserung der Genauigkeit der Modelle und der Quantifizierung der Unsicherheit zu erzielen.

Ergebnis: 57% der gesamten verfügbaren Dachfläche der 2,3 Millionen Gebäude in der Schweiz sind für die Installation von Solarpanels geeignet. Das maximale jährliche technische Potenzial wird auf 24 TWh ± 9 TWh geschätzt, was einer Produktion entspricht, die 40% des Strombedarfs des Jahres 2018 abgedeckt hätte. Es hat sich auch gezeigt, dass 25% dieses technischen Potenzials bereits durch die Installation von PV-Modulen auf weniger als 2% der Gebäude mit dem höchsten Potenzial (mit Flachdächern) realisiert und 75% durch die Installation von Solarmodulen auf etwa einem Drittel der bestehenden Gebäude umgesetzt werden können. CHE