ARTICLE

Check for updates

High resolution global spatiotemporal assessment of rooftop solar photovoltaics potential for renewable electricity generation

Siddharth Joshi [1,2,3 ✉], Shivika Mittal[4], Paul Holloway [2,5], Priyadarshi Ramprasad Shukla[6], Brian Ó Gallachóir [1,2,3] & James Glynn [1,2,3,7]

Rooftop solar photovoltaics currently account for 40% of the global solar photovoltaics installed capacity and one-fourth of the total renewable capacity additions in 2018. Yet, only limited information is available on its global potential and associated costs at a high spatiotemporal resolution. Here, we present a high-resolution global assessment of rooftop solar photovoltaics potential using big data, machine learning and geospatial analysis. We analyse 130 million km^2 of global land surface area to demarcate 0.2 million km^2 of rooftop area, which together represent 27 PWh yr^{-1} of electricity generation potential for costs between 40–280 \$ MWh^{-1}. Out of this, 10 PWh yr^{-1} can be realised below 100 \$ MWh^{-1}. The global potential is predominantly spread between Asia (47%), North America (20%) and Europe (13%). The cost of attaining the potential is lowest in India (66 \$ MWh^{-1}) and China (68 \$ MWh^{-1}), with USA (238 \$ MWh^{-1}) and UK (251 \$ MWh^{-1}) representing some of the costliest countries.

[1] SFI MaREI Centre for Energy Climate and Marine, Cork, Ireland. [2] Environmental Research Institute, University College Cork, Cork, Ireland. [3] School of Engineering, University College Cork, Cork, Ireland. [4] Grantham Institute–Climate Change and the Environment, Imperial College London, London, United Kingdom. [5] Department of Geography, University College Cork, Cork, Ireland. [6] Global Centre for Environment and Energy, Ahmedabad University, Ahmedabad, India. [7] Center on Global Energy Policy, Columbia University, New York, USA. ✉email: Siddharth.joshi@ucc.ie

From powering the National Aeronautics and Space Administration (NASA's) Vanguard satellites in 1958 to lighting homes in sub-Saharan Africa, solar photovoltaics (PV) technology has come a long way. Rooftop Solar photovoltaics (RTSPV) technology as a subset of the solar photovoltaic electricity generation portfolio can be deployed as a decentralized system either by individual homeowners or by large industrial and commercial complexes. Over the past decade, reduction in the deployment cost coupled with policy-driven initiatives has led to a rapid uptake of RTSPV globally. Between 2006 and 2018, the installed capacity of the RTSPV has grown from 2.5 GW to 213 GW- an 85-fold increase globally[1]. With an additional capacity installation of 41 GW, RTSPV currently accounts for 40% of the global cumulative installed capacity of the solar PV and nearly one-fourth of the total renewable capacity additions in 2018—surpassing the combined new installed capacities of both coal and nuclear. At the same time, RTSPV technology has demonstrated a steep decline in its deployment costs which ranged between 63 and 265 $ MWh^{-1} in the year 2019—a reduction of between 42 and 79% over 2010 values[2].

Globally, nearly 800 million people were without electricity in 2018, the majority of who are living in rural areas[3]. Here, the role of decentralized rooftop PV in advancing the ethos of the Sustainable Development Goal (SDG) 7 becomes very important. The fast installation time and low levelised cost of RTSPV can aid in mitigating the problem of energy access by making citizens or communities a prosumer. The prosumer can generate and consume electricity as per their requirements without depending exclusively on a centralized grid infrastructure. As the fastest deployable energy generation technology with the highest year-on-year growth rate[4], solar PV technology is projected to supply 25–49% of the global electricity needs by 2050 while providing employment for up to 15 million people between 2018 and 2050[5]. Out of this, RTSPV deployment will contribute up to 40% of the total solar PV-derived electricity generation by 2050.

Increased deployment of RTSPV can support displacing fossil fuels out of the current energy generation mix as can be observed in the successful implementation of rooftop photovoltaics in Germany. As the demand for electricity as an energy source increases in the future, RSTPV based generation sources will form a large part of the future renewable-based generation portfolio. This shift in the current generation mix coupled with the future low carbon generation capacity expansion can aid in reducing the energy-derived greenhouse gas emissions and also aid in advancing the SDG13 goal of combating climate change with co-benefits for the SDG3. RTSPV technology can thus lead to consumer-driven breakthroughs in tackling climate change, reducing local air pollution, accelerating development, and providing affordable energy access to areas lacking electrification.

To better understand the role that an RTSPV system can fulfill in the future, a global harmonized geo-mapped assessment of its technical potential and the costs associated with attaining the technical potential is pertinent especially when such assessments at a global level are lacking. RTSPV systems are deployed as a decentralized system contrary to the utility-scale solar PV systems, which increases the complexity of its assessment as the smallest unit of deployment becomes a rooftop as opposed to a large plot of a green or brownfield site. Along with the complexities associated with accurately determining the rooftop area, assessment of seasonal variations of its potential is also important to understand the supply dynamics of variable renewable energy (VRE) technologies like RTSPV. This highlights the need for a high-resolution spatiotemporal assessment that accurately represents the geographical variability of the built environment along with impacts of seasonal changes in solar insolation.

Current research primarily focuses on utility-scale solar PV resource assessment at a global scale. A similar assessment has not been done for decentralized RTSPV at a scale greater than regional/national levels[6–9]. As a result, energy system models and research informing climate change policy have not fully considered the role of solar PV in meeting the climate change mitigation goals[10]. Assessment of RTSPV potential requires an underlying dataset of building footprints, solar insolation mapping, and technology-specific information like panel size, conversion efficiency, and system losses. The current literature is adequate in providing global information on the latter two categories, with the largest inaccuracies[11] attributed to the demarcation and calculation of building footprints which require large data and costly information processing hardware to extract buildings from satellite imagery[12].

Two major approaches are currently used to determine built-up area, or more specifically the extent and area occupied by building rooftops, Supplementary Table 7. The first approach addresses the problem from a "Bottom–up"[13–23] perspective and is the most common approach currently implemented to calculate rooftop area at scale. Such approaches establish the relationship between building footprint data (cadastral, crowd-sourced, satellite-derived) and socio-economic metrics (Gross Domestic Product (GDP), Population) for a small sample set and then estimate the extent of building footprints across a wider scale. In the study by Jacobson et al.[24], the authors have used the bottom-up methodology to calculate available rooftop areas for 179 countries by establishing relationships between population, GDP, and floorspace area per capita based on sample data from the USA and several European countries. For a full set of global countries, Gernaat et al.[25] have used the relationship between population density, household expenditure, and rooftop area to calculate available rooftop area per country. They have calibrated their relationship equations on World Bank's data and realized an $R^2 = 0.66$. These methods are quick to implement and are relatively accurate when predicting building footprint areas in the neighborhood regions, but inaccuracies arise when the analysis is upscaled towards country/regional levels. This reduction in accuracy[26] can be attributed to the inaccuracies in the coarse geospatial mapping of socioeconomic data[27] and the heterogeneity of built-up landscapes.

The second approach addresses the problem from a "Top–Down"[28–39] perspective, utilizing aerial imagery to determine built-up area along with building footprints contained in it. "Top–Down" methods include earth observation, drone-mounted Light Detection and Ranging (LiDAR), and machine learning (ML) classification algorithms to detect buildings. On a country-wide scale, Gagnon et al.[7] have used LiDAR datasets for 128 US sample cities to calculate available rooftop areas for the continental USA using statistical inferencing to extrapolate beyond their sample site. They have generated statistical measures to generalize rooftop orientation, slope, and availability based on high-resolution LiDAR imagery. Collecting, processing, and analyzing the aerial imagery is a costly and computationally intensive task requiring datacentre scale infrastructure. These shortcomings have led to limitations in the scale of studies using top-down methods, where only a few commercial[40–42] ventures have been able to provide country-wide building footprints, and to date, a global-scale analysis has yet to be implemented. One solution to mitigate the processing and data bottleneck is to use predefined land cover classification[43–49] to demarcate global land areas which are artificially built up and then downscale the built-up area to actual building footprint[49] using simplifying assumptions. This hybrid approach was used in a study by Bodis et al.[6], where they have established a relationship between cadastral data
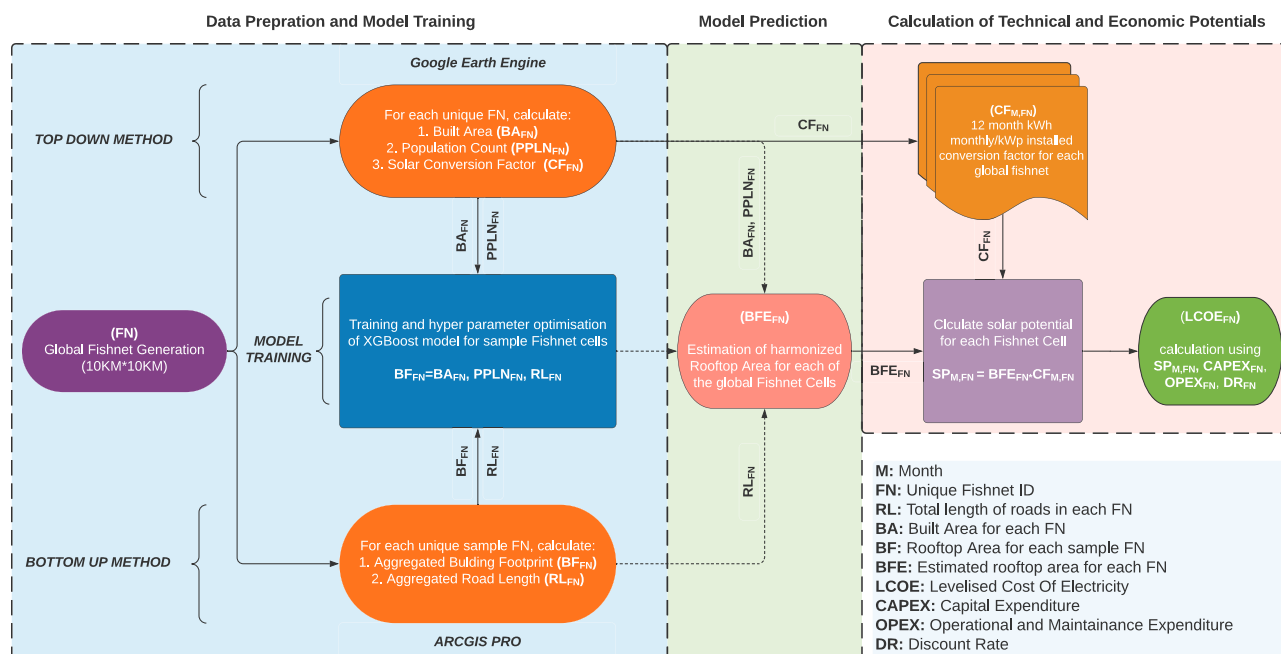
Fig. 1 Framework for assessment of RTSPV's global technical potential and costs. The framework developed in this study starts with data preparation and mapping of various geospatial metrics to both Top-down and Bottom-up pathways. Further, the machine learning model is trained and used to estimate BFE_{FN} values from BA_{FN}, $PPLN_{FN}$, and RL_{FN} values. Next, the BFE_{FN} values are used to calculate the technical potential (SP) with the aid of the conversion factors (CF). Finally, the calculated potential dataset is used to map the Levelised Cost of Electricity (LCOE) values using IRENA's renewable power generation cost data. The detailed framework is described in the "Results" and "Methods" sections. Model parameters and regional mappings are provided in Supplementary Tables 6, 8–10 and Supplementary Fig. 3a–d.

of Netherlands, CORINE landcover, and European Settlement Map to infer the rooftop area in the other 27 EU countries.

The current state of art[50–55] methods utilize ML-based object detection algorithms to map individual buildings for Region of Interest (ROI) at the city/country level. However, none of the state of art methods have been applied to a global ROI assessment partly due to the amount of data processing required in ML-based methods and partly due to the limitations of the ML algorithm in detecting similar objects that deviate from training sample sets.

Here, we directly address the gap in the current literature by developing a hybrid framework that integrates "Top-down" and "Bottom-up" methods using a ML model to provide a high-resolution global assessment of rooftop solar PV technical potential at a monthly temporal resolution (Fig. 1). We divide the entire global landmass into multiple $10\,km^2$ assessment units. In each assessment unit, we utilize a high-resolution population dataset (at 100 m spatial resolution), road length, and built-up area boundaries to estimate the rooftop area. The estimation algorithm built upon the ML model is designed to learn the relationship between population, road length and built-up area boundaries, and actual rooftop area from a large global sample set covering countries in various stages of socioeconomic development and built-up topography. This way we utilize a comprehensive set of disparate built environment archetypes involving variable population densities in different landscape configurations to overcome the limitations imposed by the "Top-down" and "Bottom-up" methods used in previous studies. Along with the geo-mapping of the technical potential, we also map the associated Levelised Cost of Energy (LCOE) and regional supply cost curves for a global ROI covering >195 countries, spanning 130 million km^2 land area, containing buildings ranging from detached rural nucleated settlements to global conurbation dotted by multi-storied skyscrapers across varied geographies.

In our study, we define the "technical potential" of RTSPV as the maximum electricity generation that can be derived from a given rooftop area, where the rooftop area is kept consistent with the 2015 year's built-up extent and does not include additional building stock created after that. The results presented in this study have considered a 100% rooftop availability at a 10% panel efficiency. To account for the change in the potential due to different panel efficiencies and rooftop availability, we have documented global and regional potentials for a set of rooftop scaling factors and panel efficiencies (Supplementary Tables 4 and 5). Additional boundary conditions of our study and the limitations in the interpretation of the main results are documented in the "Methods" section.

Results

Model design and validation. We started by dividing the global landmass into a Fishnet Grid (FN) containing a total of 3,521,120 unique squares of $10\,km^2$ size, where each FN has a unique id and is attributed to a unique country. Aggregation of the built area (BA_{FN}) within each FN was undertaken using built area subclassifications (100 m resolution) provided by the global landcover layer[46] (LC) of the Copernicus land monitoring program. Being derived from a native 10 m resolution satellite imagery, the LC layer provided us with a significantly improved representation of built-up area over the current state of art methods that utilize coarser-resolution landcover classification. The use of pre-built landcover classifications also aided in reducing the information processing overheads related to manually classifying petabytes of satellite imagery.

It is important to highlight that the BA_{FN} layer aggregations contain buildings along with roads, green area boundaries, footpaths, parking lots, etc. These additional built-up structures are not relevant to the current analysis and can occupy significant BA_{FN} areas in a low-density built environment. Methodological uncertainties can also arise due to the heterogeneity of impervious
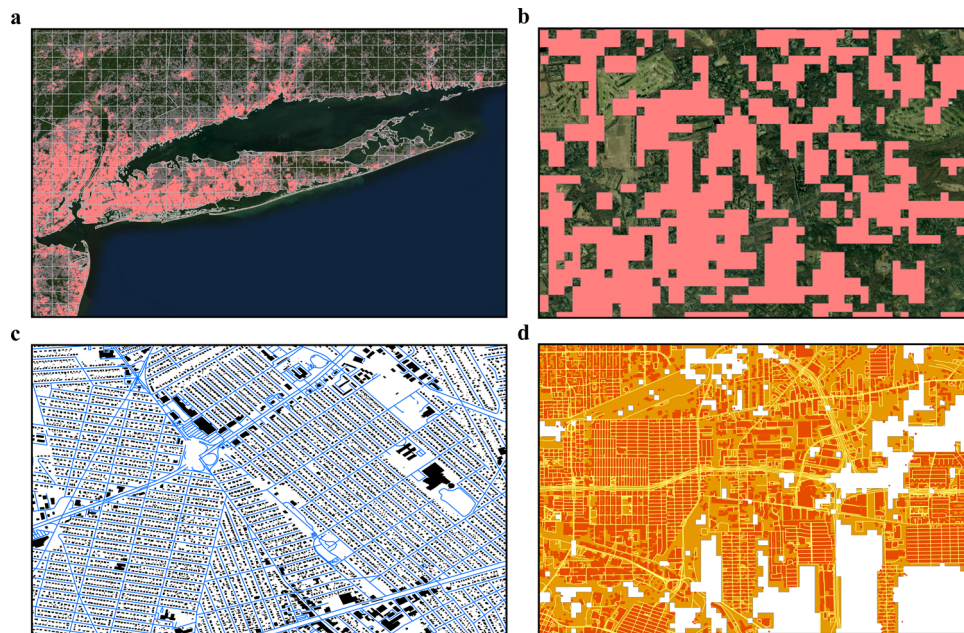
Fig. 2 Visual representation of a sample FN involved in Bottom-up methods. a The image is depicting an area over New York, USA. The gridlines are the geographical boundaries of Fishnet cells (FN), with each cell size of 10 km². The red area inside each FN is the built area (BA$_{FN}$). **b** The green areas in the image are the non-built-up geography in the year 2015, with red areas representing BA$_{FN}$ as 100 m² blocks. Each block has a value of 0–100% based on the percent of built-up area in a block. **c** The black polygons are the building footprints (BF$_{FN}$) derived from big data sources, inside a sample FN. The blue lines are the roads (RL$_{FN}$) inside the sample FN, with white areas representing empty spaces that are not used in our analysis. **d** The image represents the four main categories of objects present inside each sample FN. The orange areas are the BA$_{FN}$ blocks, with red areas representing the actual buildings, yellow lines representing roads, and white areas are the areas not considered in our study. Data Credits: Copernicus GLC Landcover, Microsoft Building Footprints, and Open Street Maps.

landscapes and due to misclassification caused by the similar spectral signatures of other built structures on the ground, but at a global ROI, these errors form a very small percentage of the ground truth. This underscores the need to downscale the BA$_{FN}$ layer globally to accurately represent a harmonized rooftop area.

To downscale the BA$_{FN}$ layer, we started by generating a subset of global FN containing actual building footprints[40,42] (BF$_{FN}$), road length[56] (RL$_{FN}$), and population count[57] (PPLN$_{FN}$). For calculating building footprints (derived from Microsoft AI, Ecopia AI, and Open Street Map-OSM datasets), we aggregated individual building footprints based on FN cells overlapping the individual buildings to generate a single building footprint area value per FN (Fig. 2). In total, 37,115 km² of building footprint area and 11.4 million km² of the total land area were covered by our sample FNs, accounting for >300 million individual buildings ranging from small outdoor sheds to mega factories. The building samples also provided us with a diverse set of building types in different geographies spanning a wide spectrum of socio-economic stages of development. These global sample sets (Fig. 3a, Supplementary Table 1) are a marked improvement over the previous literature, where a narrow sampling strategy often at a city/country level is undertaken. Using heterogeneous global building samples enables the overall analysis to be more resistant to generalization error which is introduced due to overreliance on a small set of similar built-up landscapes.

OSM provides a highly accurate open-source global roads infrastructure dataset[58], which we used for mapping road length (RL$_{FN}$). Our sample dataset showed a Pearson Correlation of 0.95 (Fig. 4b) between the building footprint sample and the road length sample. It can be observed that in the real world, the development of roads is often associated with increasing population size and with increasing building stock. Road length has not been used in any previous studies and will increase the

accuracy of downscaling over previous methods. In total, we processed >16 million km of roads for our sample FNs (Fig. 3b).

For our population count aggregation, we utilized WorldPop's high-resolution griddled population count. The high-resolution population dataset coupled with a high-resolution BA$_{FN}$ layer enabled us to accurately map the population count for each of our sample FN. Previous studies at country/regional scale have used coarse resolution (≥1 km resolution) population count which can lead to inaccuracies in downscaling due to over counting of the population in a given FN.

Aggregation and processing of big data like geospatial building and road datasets require a very high computation time and significant cost in hardware. To overcome this challenge, we used Google Earth Engine's (GEE) cloud computing platform to perform planetary-scale analysis, where the datasets were split into smaller raster files based on each FN's geographical extent and finally aggregated to represent a single value (see "Methods" section). It should be noted that the GEE platform is lacking in the processing of vector data like building and road polygons, for which we utilized multi-core processing enabled by ArcGIS PRO desktop software. This combination of GEE for raster data (PPLN$_{FN}$, BA$_{FN}$) and ArcGIS PRO for vector data enabled us to process our datasets at a fraction of the time and cost. To our knowledge, this is the first attempt at merging two fundamentally different service platforms to assess global potentials, which could be useful for other planetary-scale resource assessments. Links to the input datasets and their validation reports are documented in Supplementary Table 2.

Next, we linked the aggregated BA$_{FN}$, PPLN$_{FN}$ datasets from Top-down methods and processed BF$_{FN}$, RL$_{FN}$ datasets from the Bottom-up methods using a ML model that takes BA$_{FN}$, PPLN$_{FN}$, RL$_{FN}$ datasets as independent variables and BF$_{FN}$ as the dependent variables. The ML model trains from the relationship
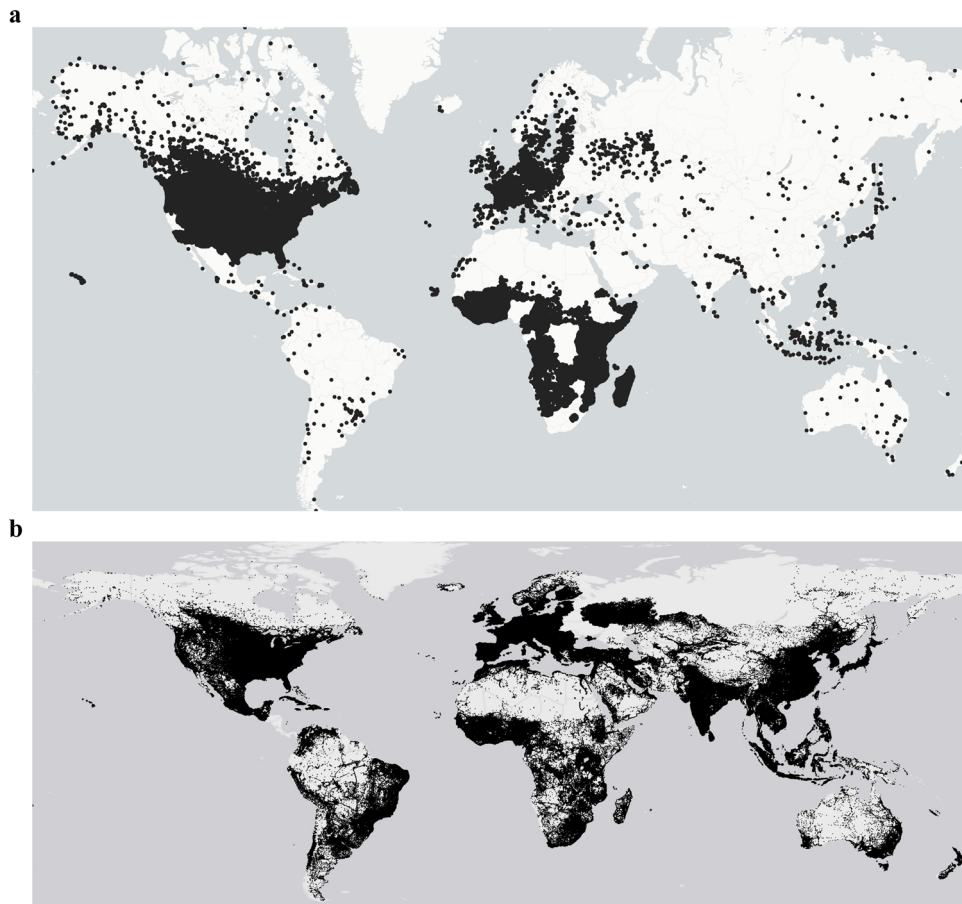
a

b

Fig. 3 Geographical distribution of global roads and sample Fishnet cells. a Global distribution of sample FN was generated in our study. Microsoft AI-based building polygons are used for the USA, Canada, and Tanzania, Ecopia AI-derived building footprint areas are used for the African countries. OSM-derived building polygons representing a total of 4000 sample FNs are used for the rest of the world. **b** Geographic location of all the global roads used in our analysis. OSM-derived global roads have near-global coverage, with each pixel on the map being for each road feature inside the country of reference. Data Credits: Fig. 3b, © OpenStreetMap contributors.

between the independent and dependent variables for our sample FNs and estimates the aggregated building rooftop area for the rest of the non-sample FNs, effectively downscaling the BA_{FN} layer to an estimated area occupied by building rooftops (BFE_{FN}). For data gaps in $PPLN_{FN}$, RL_{FN} layer, we used an iterative imputing method to interpolate missing values in the dataset using multiple regression runs.

In our tests, the ML model built upon XGboost algorithm provided better accuracy (R^2 value) and better overall error reduction (considerably lower mean absolute error) in the prediction of sample data as compared to multivariate regression (Supplementary Table 3). To reduce the overfitting of the ML model, a 10-fold cross-validation strategy was used for tuning hyper-parameters to generate the best model (see "Methods" section). The ML model performed well (Fig. 4d, e) with $\pm 4\,km^2$ error in predicting total building footprint per $100\,km^2$ FN, with the majority of errors lying between $\pm 0.25\,km^2$. Also, for a total of $37,000\,km^2$ of absolute building footprint area across our samples, a total of $4\,km^2$ of absolute error was recorded. On a regional level, slight left skewness/systematic underestimation of rooftop area was observed for the Asian Region (Supplementary Fig. 1). At the end of the downscaling step, a dataset containing 3,521,120 FN cells is generated, with each cell containing aggregated building rooftop area (BFE_{FN}).

Rooftop area to solar potential conversion was undertaken by using World Bank's conversion factor[59] (CF_{FN}) for locations between 60°N and 45°S, covering over 99% of the world's

population. The CF_{FN} layer is provided as a griddled raster dataset representing kWh produced by each kWp installed capacity per day. The locations outside the latitudes were assigned a constant CF value of 3.5 kWh/kWp/day (peak). Next, CF_{FN} values for each of the twelve months were used to generate monthly solar potential per FN ($SP_{FN,M}$). Further, we aggregated monthly technical potentials to yearly technical potentials for each of the countries to generate a global RTSPV potential map, as shown in Fig. 5. To represent the costs associated with the technical potential, we used the LCOE metric (see "Methods" section). The capital expenditure (CAPEX) values range between 840 and 3874 $\$\,kW^{-1}$ of installed capacity for 17 countries in total, with the rest of the world being allocated averages for the continent they are in. The operation and maintenance cost (OPEX) and discount rates (DR) were based on whether the country is OECD or Non-OECD. For each FN, LCOE costs were calculated based on CAPEX, OPEX, SP, and DR with a project lifetime of 25 years. Our technology cost data are based on IRENA[2] residential solar PV costs for 2019.

Global technical potential and cost assessment. Our assessment shows a total global technical potential of 27 PWh yr^{-1}, of which Asia (13 PWh yr^{-1}), North America (5.5 PWh yr^{-1}) and Europe (3.6 PWh yr^{-1}) represent a majority of the potential followed by Africa (2.9 PWh yr^{-1}) and South America (1.7 PWh yr^{-1}). The hotspots for the potential are concentrated in and around the
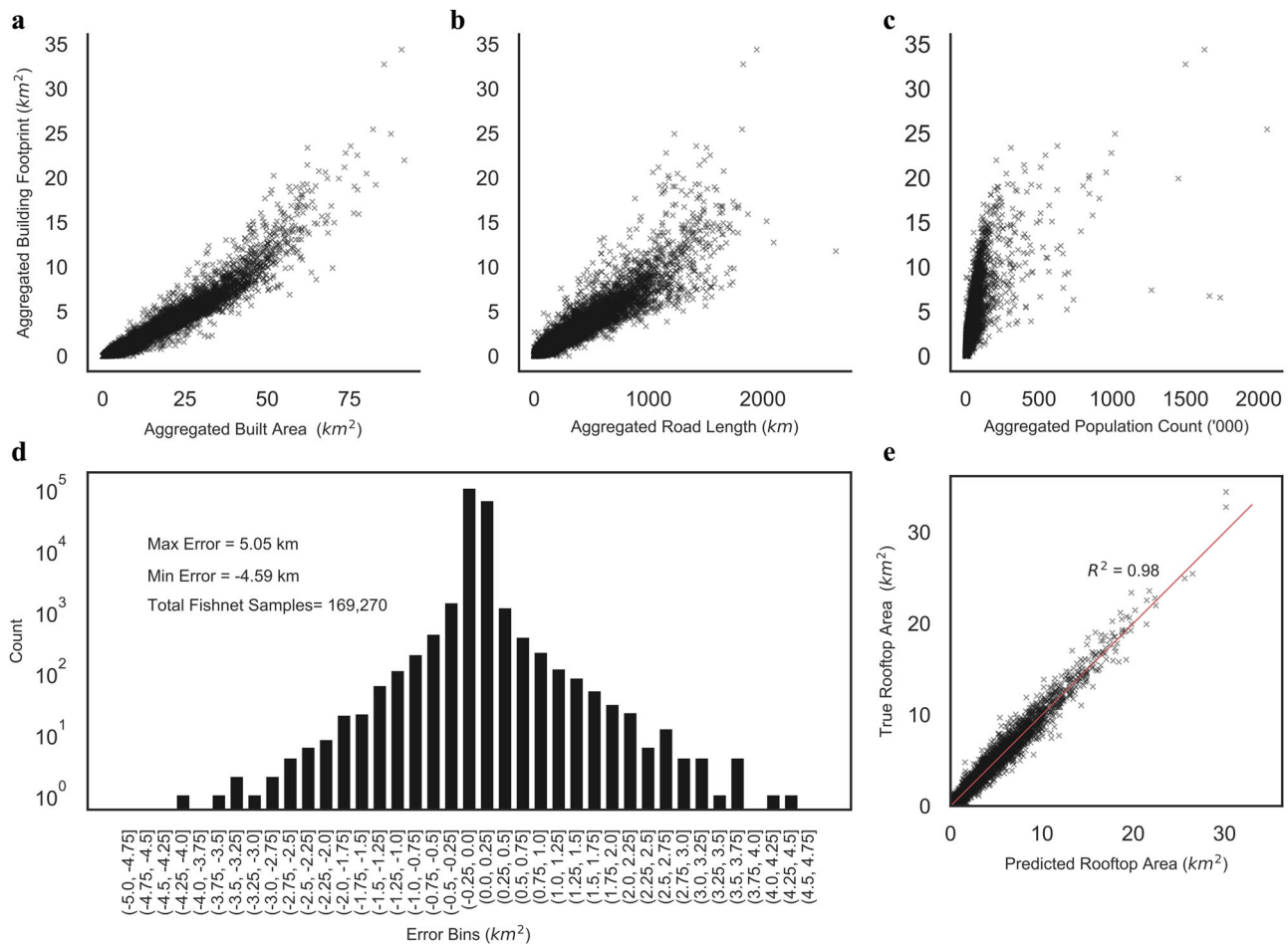
Fig. 4 Relationship between inputs, distribution of errors, and prediction accuracy of the downscaling model. The distribution of BA_{FN} (Built Area) (**a**), RL_{FN} (Road Length) (**b**), and $PPLN_{FN}$ (Population) (**c**) in relation to the BF_{FN} (Building Footprint). The distributions have a Pearson Correlation of 0.97, 0.95, 0.72, respectively. The distribution of errors in prediction (**d**) has a bell shape when using the trained model to predict the sample dataset. Each bar is the count of FN that is in each error bin. The majority of errors lie in the -0.25 to $+0.25\ km^2$ per FN range, with peak error capping at $\pm 5\ km^2$ per FN. The downscaling model built upon the machine learning model has shown adequate accuracy in prediction (**e**) with an $R^2 = 0.98$. Each cross represents the predicted value of each sample FN. The density of points is higher for low values as the majority of sample FN will have smaller built-up areas contained in them, with very few covering major cities due to the size of the FN used in this study.

densely populated nucleated settlements globally (Fig. 5a). Nearly 20% (5 PWh yr^{-1}) of the global potential is located within the areas with a high population density (>1500 people/km²), with 55% (15 PWh yr^{-1}) of the potential being dispersed within the low-density areas (<500 people/km²). Amongst the countries, China (4.3 PWh yr^{-1}), the USA (4.2 PWh yr^{-1}), and India (1.7 PWh yr^{-1}) have the highest yearly potential (Table 1). A ±1% deviation can be observed in the yearly global potential due to the aggregation methodology of the CF factor (Supplementary Fig. 2).

Although the African region has a good solar insolation endowment, the RTSPV potential is assessed as being the third lowest due to low building stock. Amongst the African Region, the largest potential is concentrated in the West African region followed by the North African Region. The combined West and North African regions have more potential than India, high-lighting the importance that low-cost RSTPV can play in future energy systems. Future population growth and a corresponding increase in the building stock may increase the overall RTSPV potential for Africa. Both North American and European regions have similar assessed rooftop areas (~30,000 km²), yet North America has nearly 1.5 times the potential of Europe due to higher solar insolation during the year especially during the winter months.

Along with spatial variability due to on-ground building distributions, seasonal variability of the RTSPV potential is also observed due to the variation in intra annual solar insolation. The seasonal variability of the monthly global potential is between 1.84 and 2.61 PWh, with December and January representing the months with the lowest global potential (Fig. 5a). Globally the highest seasonal variability of the potentials is observed above the 45° north latitude covering Europe, Russia, the USA, and Canada. Within regions, the highest intra annual variability of the potentials (Fig. 6) is observed in the West European region (EUW) with monthly potentials between 94 and 255 TWh. There is a variability of ±40% around the average monthly potential of 183 TWh in EUW, with the highest monthly potential being observed in the summer and the lowest monthly potentials in the winter. The lowest regional intra annual variability of potentials is observed in the West African region (AFW) with monthly potentials between 97 and 119 PWh. There is a variability of ±1% around the average monthly potential of 109 TWh in the AFW region, with the maximum monthly potential observed in December and January.

To analyze the cost of attaining the potentials, we generated supply cost curves for seven world regions and also at an aggregated global level (Fig. 7a). Nearly 10 PWh yr^{-1} (40%) of the global potential can
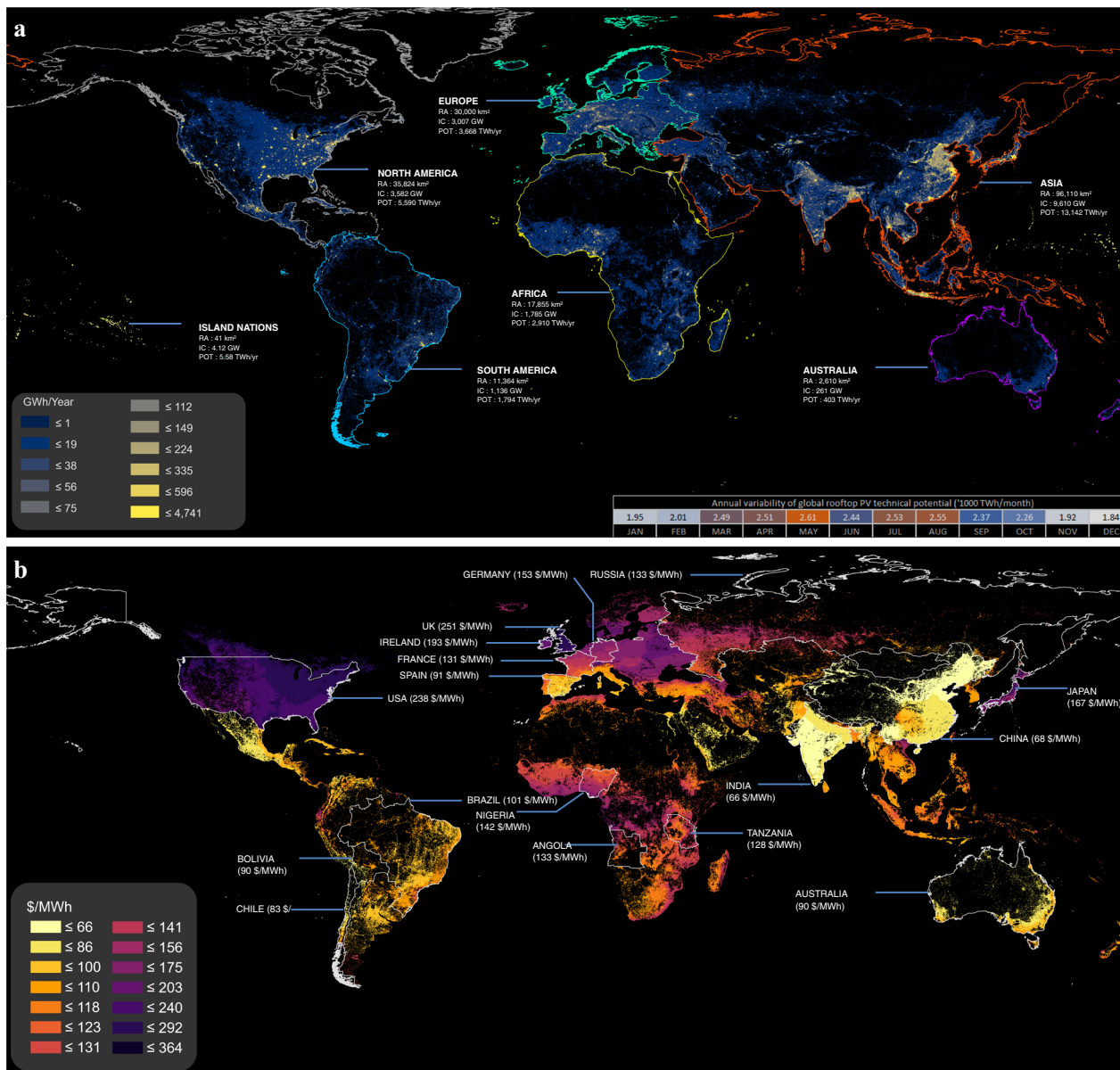
Fig. 5 Global Distribution of RTSPV technical potential and LCOE values. a The geographical distribution of global RTSPV technical potential generated in our study. Major world regions are highlighted in the image and data is provided for the total estimated rooftop area (RA), installed capacity (IC), and potential generation (POT) for each region. The global landmass is color-coded into 11 RTSPV potential bins. Seasonal variation in aggregated global potential is highlighted in the bottom right. **b** The geographical distribution of LCOE values generated in our study across the globe with major countries highlighted in each world region. Each highlighted country has a corresponding aggregated LCOE value for the whole country. The LCOE values are color-coded into 14 cost bins. Map boundary data from the Database of Global Administrative Areas (GADM: https://gadm.org).

be achieved below 100 $ MWh yr^{-1}, with the majority of the potential being realized below 200 $ MWh^{-1} (Fig. 7b). At a global level, nearly 40% of the potential can be achieved with an investment equivalent to 10% of the 2015 global GDP value, and with an investment equivalent to 30% of the 2015 GDP value, nearly 100% of the global potential can be realized (Fig. 7c). We found that realizable potential doubles with each subsequent doubling of capital investment in RTSPV until invested capital is equivalent to 20% of the global GDP value in 2015. An increase of investment from 20 to 30% of the global 2015 GDP value increases the realizable potential by only 27% indicating the areas where the cost of implementation of RSTV is very high. These areas are represented by large warehouse/industrial complexes in Alaska and Canada, where yearly solar insolation is low and CAPEX investment is high to cover the entire large rooftops with solar panels.

At a global level, spatial variability in the LCOE is also observed (Fig. 5b). In the northern hemisphere, the LCOE values gradually increase from 40 $ MWh^{-1} to 280 $ MWh^{-1} with increasing latitude. Here North-East China is an exception which has shown a decrease in LCOE values with increasing latitudes. For Asia, a majority of the potential can be realized between 40 and 100 $ MWh^{-1} making the RTSPV competitive with fossil fuel technologies. The cost of attaining the country-specific potential is lowest in India at 66 $ MWh^{-1} compared to China (68 $ MWh^{-1}).

For Europe, Africa, South America, and Island Nations a majority of total potential can be realized below 180 $ MWh^{-1}. Within Europe, Spain has the lowest LCOE cost of 90 $ MWh^{-1}, with an increasing trend in cost being observed when moving towards the higher latitudes. Within each country of the European region, further variability in LCOE is also observed with some

Table 1 32 Regions global distribution of rooftop area and solar potentials.

32 World regions[a]	Land area (1000 km^2)	BA$_{FN}$ (km^2)	BFE$_{FN}$ (km^2)	Potential (TWh yr^{-1})[b]
AFE (Africa East)	4355	6921	1246	214
AFN (Africa North)	5485	18631	3398	592
AFS (Africa South)	5537	18132	2443	413
AFW (Africa West)	13130	59604	8613	1317
ANZ (Australia and NZ)	7971	15581	2610	404
ARG	2780	9330	1749	285
ASC (Asia Central)	6622	36053	4981	748
ASE (Asia South East)	2334	62144	9568	1354
ASO (Asia other)	354	13121	2117	304
ASR (rest of Asia)	1642	8942	1561	219
BRA	8533	31962	6386	997
CAN	9904	13851	2394	327
CHN	9380	273585	35156	4375
ENE (Non-EU East)	1035	34634	4111	499
ENW (Non-EU West)	466	4199	630	75
EUE (Europe East)	1138	47700	5457	648
EUW (Europe West)	2984	134286	17467	2210
GBR	244	15798	2400	238
IDN	1899	35710	6163	878
IND	3160	78971	11731	1815
IRN	1623	12437	2004	341
JPN	372	41533	7942	1044
KOR	100	8758	1439	201
LAM (Latin America)	7231	27268	5028	805
MEA (Middle East)	1619	15742	2628	462
MEX	1956	20502	4037	720
ROW (rest of the world)	2426	1302	212	35
RUS	16832	58175	8038	941
SAU	1920	6333	918	169
TUR	780	13526	1716	265
USA	9457	160479	27585	4247
ZAF	1221	11954	2150	374
TOTAL	134491	1297163	193875	27512

[a]ISO alpha-3 codes are the region names for individual countries, a group of countries as a region has names described in parenthesis.
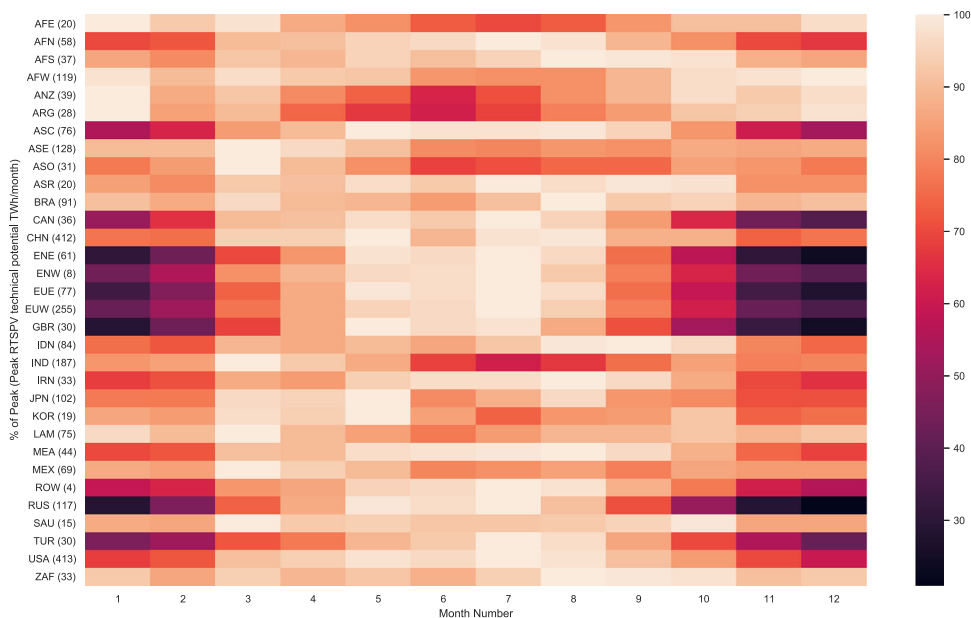[b]Considering 100% rooftop availability at 10% panel efficiency.

Fig. 6 Monthly potential variability for 32 world regions. The heatmap represents the monthly variability of regional potentials. The values in the parenthesis represent the highest monthly potential for the region. The color-coded heatmap represents % of the maximum monthly potential of the region. Europe (EUW, EUE) has shown high intra-annual variability of the potential, with the maximum potential being realized in the summer months. In the southern hemisphere, the maximum potential is realized in the winter months. Africa West (AFW) has shown the lowest variability in the monthly potential.
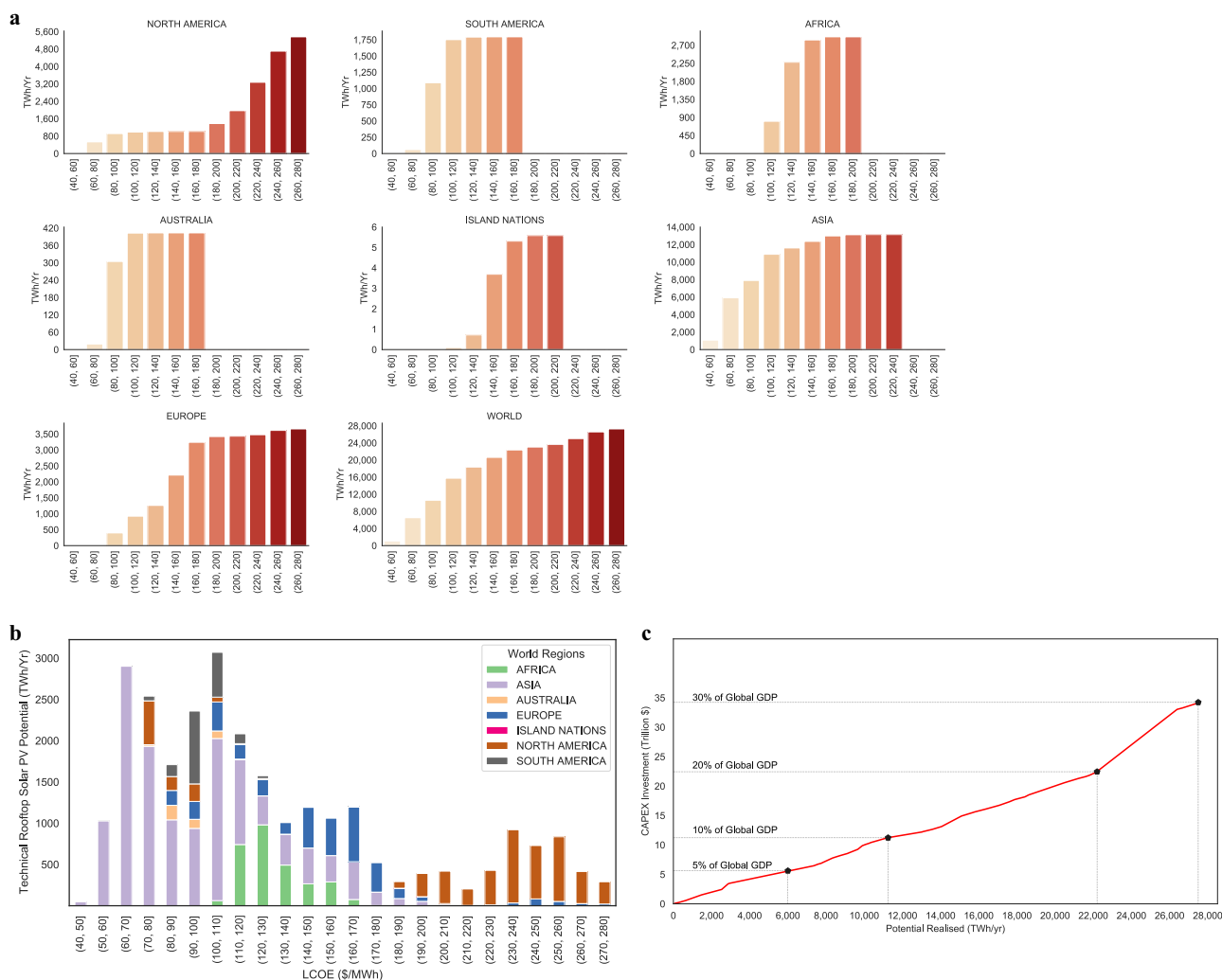
Fig. 7 Regional supply cost curves based on a 20 $/MWh bin size and distribution of LCOE across regions. Regional Supply cost curves (**a**), showing the cumulative technical potential for a given LCOE bin size. The absolute increase in the bar height represents the change in the added potential that can be realized in a given LCOE bin. Near constant bar heights for consecutive bins are representative of a very small additional potential being realized at an increasing cost. For a single bar, the height of the bar represents the cumulative potential that can be realized within a specific LCOE band. **b** Distribution of potentials across LCOE bins of 10 $/MWh size. Potential extraction is the cheapest in Asia, followed by Africa and Europe. **c** CAPEX investment required to achieve a specific total potential as a percentage of global GDP in 2015. Nearly 50% of global potential can be achieved at investment equivalent to 15% of 2015 global GDP value.

regions observing cheaper costs than neighboring regions in the same latitude. In the African region, the majority of the total potential can be realized between 110 and 160 $ MWh^{-1}. Within Africa, Nigeria, Gabon, and Cameroon have the highest costs (around 150 $ MWh^{-1}) to achieve their respective potentials.

North America, UK, and Japan have shown the highest cost to realize the potential. This can be attributed to the high CAPEX costs in the countries, which are expected to reduce in the future due to technological innovations and reduction in import tariffs. In North America, Canada and the northeastern states of the U.S.A around the Great Lakes have the highest LCOE costs. The cost to achieve the potential in these countries ranges from 200 to 280 $ MWh^{-1}. U.K (251 $ MWh^{-1}) has the highest country-specific costs to achieve its potential.

Globally, CAPEX required to access potential varies both with respect to the size of the GDP and with respect to the LCOE values (Fig. 8a). To realize the complete potential in their respective countries, low-income countries would need to invest capital which is multiple times (up to 3.5 times) their 2015 GDP value even at

relatively low LCOE of between 80 and 150 $ MWh^{-1} to cover for high upfront costs. For similar LCOE values, high-income countries (World Bank Income Classification) can achieve their complete potentials with a capital investment equivalent to a fraction (up to half) of their 2015 GDP value.

We categorized countries into groups based on their GDP per capita (GDPC) and on their yield factor (Fig. 8b). In this study, we defined the yield factor as the yearly potential that can be realized from 1 TW of installed capacity. Based on this categorization, we found out that emerging economics including India, Brazil, and Mexico have high yield factors (1.5–2) that would favor the deployment of RTSPV in these countries. Even though the LCOE difference between the two most populous countries (India and China) is minimal, with the greater solar endowment and with high yield factor, RTSPV technology rollout becomes more favorable in India compared to China. However, the uptake of RTSPV is still very low in these countries due to lack of credit and the inability to pay the high upfront cost of the RTSPV system. This highlights the need for global cooperation,
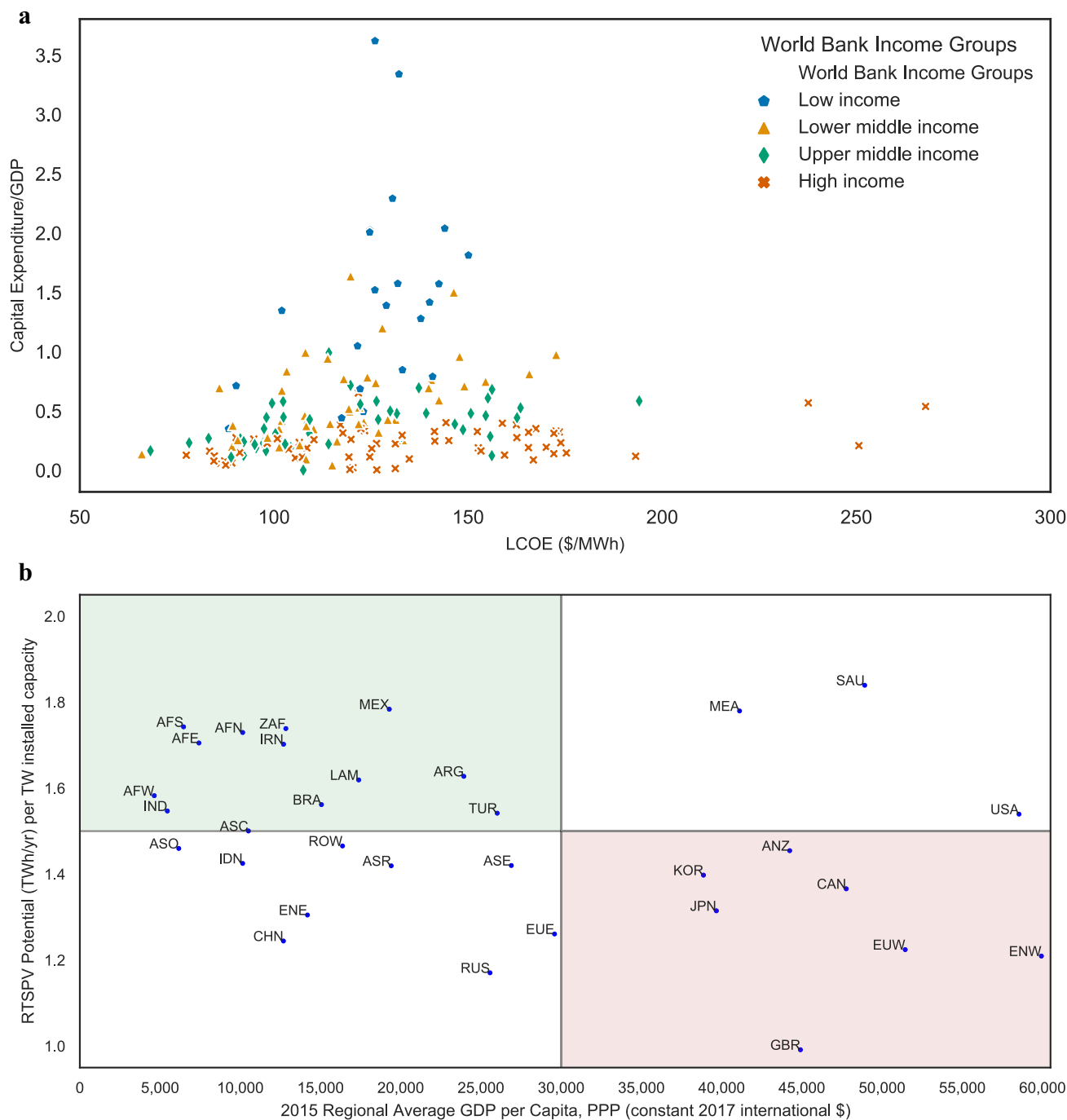
Fig. 8 Global economic and technical variation in accessing the potentials. a Disparity in achieving full potential, with low-income countries requiring capital investment greater than their respective 2015 GDP value to achieve full potential. **b** Distribution of 32 regions in relation to their relative yield factor and GDP per capita. The red region focuses on countries that currently have high rooftop PV rollout. The green region focuses on countries where RTSPV technology rollout will have maximum benefits.

technology transfer, and green financial instruments to accelerate the deployment of low carbon RTSPV technology in low-income and medium-income countries.

Discussion

The study showcases how a framework based on big data and an ML model in conjunction with cloud computing platforms can be used to undertake a planetary scale resource potential assessment. We analyzed 130 million km² of global land surface area by utilizing learnings from global samples containing 300 million

buildings with 16 million km of roads. Using Google Earth Engine and a ML model we demarcated 1.2 million km² of the built-up area containing 0.2 million km² of rooftop area. As part of the assessment, we generated (1) a global rooftop area dataset (2) a global RTSPV potential dataset at a monthly temporal resolution, (3) costs of attaining the technical potential. The datasets were further used to generate high-resolution global maps of the potential and costs. We have also advanced the current state of art by combining the top-down and bottom-up approaches at a global scale to develop a hybrid framework for resource potential assessment which can also be used in advancing the assessment of global

wind and bioenergy potentials. The assessment shows that a sizeable RTSPV potential of 27 PWh yr^{-1} exists at a global level that can be attained for costs between 40 and 280 $ MWh^{-1}. The potential is highest in Asia followed by North America and Europe. A capital investment of around 7 trillion dollars is required at current prices to achieve a global RTSPV based electricity generation of 10 PWh yr^{-1} below the LCOE of 100 $ MWh^{-1}, covering 3.72 billion people globally.

At the EU-27 regional level, our estimated rooftop area of 7596.4 km^2 is similar to the 7935 km^2 calculated in the Bodis et al. study when incorporating a rooftop scaling factor of 0.3. For the USA, our estimated rooftop area and annual potential of 8827 km^2/1.9 PWh yr^{-1} compare well with the estimates of 8130 km^2/1.4 PWh yr^{-1} presented in the Gagnon et al. study when incorporating a rooftop scaling factor of 0.32. On a city level basis, our potential of 1 TWh yr^{-1} is in alignment with the 1 TWh yr^{-1} calculated in a study by Hong et al.[36] where they have used advanced hillshade analysis to capture the effects of building induced shadows in a dense urban topography. A detailed global/regional/country and city level comparison of our results with selected research work is documented in Supplementary Table 7. From the comparison with other studies, we can conclude that the results from our framework demonstrate high veracity as they are within the margin of error of values present in the literature. In addition, the good estimation accuracy of our framework at high spatial resolution feeds into higher accuracies at aggregated lower resolutions.

Our assessment has important implications for addressing the twin challenges of sustainable development and climate change with co-benefits in advancing SDG 3 and SDG 7. First, the analysis of spatial RTSPV potential presented in this study shows that 55% of the global RSTPV potential is spread across low-density areas. This highlights an important aspect of solar transition, where the majority of its benefits in providing cost-effective and fast deployable electricity can be realized in rural areas. RTSPV can thus aid in mitigating the energy poverty being experienced in the less developed and sparsely populated areas in a country where extensive grid integration may be costly or where competition for land may exist. Nearly 20% of the global potential lies in the high-density areas where the deployment of RTSPV can aid in displacing fossil fuel-derived electricity with less polluting electricity generation thereby reducing local air pollution[60]. Second, from a perspective of energy equality and "leaving no one behind" agenda of the SDG, the most disadvantaged areas with respect to access to electricity are currently the low-income countries[3] which require the rapid and cost-effective deployment of clean electricity generation infrastructure. Our assessment shows that low-income countries may need significant capital investments as steep upfront RTSPV installation costs in the order of magnitude 2–3 times their 2015 GDP value to achieve their country-specific potentials. At current costs, governments may need to provide subsidies and seek external investments to improve the prospects of deployment of RTSPV in these areas. This highlights the vital role that the developed economies may play in enabling the deployment of RTSPV in these countries by the means of financial flows to realize the climate change co-benefits. With maturing of the technology and the emergence of an economy of scale, costs will further go down to enable a solar revolution in these areas and aid in their low carbon energy future.

Third, countries that are currently reaping their demographic dividends like India and China are better suited for rapid deployment of RTSPV. We showed that these countries have high potential with low seasonal potential variability along with the low cost of deployment. As these countries have the largest population share globally with large building stocks, they can be the first movers in decarbonizing their electricity generation infrastructure by substantially deploying a decentralized electricity generation portfolio, further aiding in climate change mitigation. Along with climate mitigation, RSTPV deployment in these countries can gain a lot from the high workforce percentage in the population in the form of cost-effective manufacturing and operational maintenance. Fourth, the high-resolution assessment can aid the local governments in identifying suitable locations for the rapid deployment of energy generation infrastructure. This way bottom-up formulation of energy policy can lead to inclusive designing of nationwide policies to provide energy justice to the citizens. Fifth, businesses and financial institutions like World Bank and International Monetary Fund can analyze in-depth the investment opportunities and risks in implementing RSTPV infrastructures leading to local job creation and sustainable development of manufacturing industries. Sixth, the information contained in the assessment along with supply cost curves is a measurable step forward and fills in a significant information gap that is present in current integrated assessment models where solar PV potentials are often represented as aggregated potentials for utility and rooftop installations. Our assessment provides insightful findings and technical potential datasets that will certainly aid in accurately modeling the future carbon-neutral scenarios to inform national energy policies[61–66]. This will no doubt aid in exploring sustainable and inclusive low-carbon future possibilities.

Our study shows pronounced variability of seasonal potentials in the higher latitudes for countries covering Europe, North America, and Australian regions. These regions have high electricity consumption per capita along with the financial capacity to introduce significant VRE in the electricity generation mix. To mitigate the variability of the potentials over the year, smart grids that optimize the generation portfolio and the introduction of regulator-driven mechanisms that balance the generation market becomes important. Further, the introduction of new market mechanisms[67] is needed to effectively integrate the prosumers and utility operators in competitive electricity generation markets. With falling prices of electricity storage technologies and smart management of interconnected grids, RSTPV technology will play a critical role in these markets by undertaking generation, storage, and system balancing roles.

In conclusion, our assessment shows that the current electricity generation potential of RTSPV exceeds the current (2018) yearly aggregated global electricity demand[68]. Our assessment also shows that a minimum of 50% of the total global rooftop area is required to meet the yearly global aggregated electricity demand. Due to the diurnal cycles of solar insolation and to balance the seasonal and daily variability of the RTSPV generation, the role of storage solutions to compliment RTSPV electricity generation is critical in realizing the maximum potential of this technology and to meet the peak daily demand. Hence, the practical realizable potential of RTSPV will depend on the future cost trajectory of storage technologies, capital expenditure related to the technology, and the overall configuration of the energy system.

Even with its limitations and shortcomings the current assessment is still the state of art and provides researchers with global analysis datasets. The underlying methods and datasets have been peer-reviewed and are of the highest quality currently available and represent a generation advancement over the datasets used in the current state of art methods. The current dataset can be improved by using a next-generation 10 m resolution landcover dataset and with an increase in the spatial resolution of the population and solar data at a global scale. Better data processing platforms can enable further work to be accomplished at a 1 km resolution providing a 100 times increase in the spatial representation of the potentials. In addition, inputs in the

form of realistic regional variation in the rooftop availability will aid in narrowing down the uncertainty in potentials and cost and should be the next logical research step in model improvement.

Methods

Top-down method. We generated a total of 3,521,120 fishnets using the ArcGIS PRO desktop application for all the global landmass except for the continent of Antarctica. The FN grid is the lowest unit of data aggregation in our method. The FN's at the boundary of two countries have a common Fishnet identifier but unique country attribution with FN being split at the boundary. Next, the fishnet polygons were uploaded to Google Earth engine platform[69] (GEE) to calculate satellite-derived Built Area (BA_{FN}), Population ($PPLN_{FN}$), and conversion factors (CF_{FN}) for each fishnet cell.

We utilized the global land cover (LC) layer from Copernicus Global Land Service v2.0 to calculate BA_{FN} values within each FN. This landcover classification layer was chosen for its robustness and near-global coverage and it is backed by exhaustive testing and validation. The LC has many categories of classifications, amongst which built-up area is one of the classifications. The built-up classification is in turn derived from European Space Agency's World Settlement Footprint 2015 layer that is derived from a 10 m resolution sentinel mission's radar and optical imagery. To calculate the total built-up area in each Fishnet cell, the LC raster file was first cut into smaller sizes based on each fishnet's geographical bounds. The individually cut LC dataset was then aggregated using the following:

$$BA_{FN} = \sum (PX_V \times PX_A) \tag{1}$$

where BA_{FN} is the Built-up area in each fishnet cell, PX_V is the pixel value (0–100) representing the percent of the built area in each pixel, PX_A is the area occupied by each pixel.

To map the number of people living in each fishnet cell, we utilized a global population raster file at 100 m resolution provided by the WorldPop project. The population raster disaggregates the recorded United Nations population counts in an administrative unit to a finer resolution using ML-based methodology. The population raster file was split into smaller entities based on each fishnet cell's geographical bounds. Then, a masking file containing areas covered by the LC layer within each fishnet cell is used to mask the population outside the bounds of the BA_{FN} area in each fishnet. The individual small population dataset was aggregated for each fishnet cell using the following:

$$PPLN_{FN} = \sum PX(NM)_V \tag{2}$$

where, $PPLN_{FN}$ is the total count of people living in each fishnet, and $PX(NM)_V$ is the pixel value of each pixel that is not masked by the overlay LC layer. The population raster file was masked to remove any population count data that is not part of the LC pixel. The masked population raster pixel can be attributed to artifacts induced due to the downscaling algorithm in the original dataset or extra built-up areas that were external to the LC layer. To maintain homogeneity in the analysis, the LC layer was taken as the base for all analysis, and any area not covered by (BA_{FN}) was assumed to be not present on the ground, even if it actually exists as ground truth.

The CF factors were calculated using World Bank's SolarGIS raster datasets. The dataset is provided as a 1 km resolution raster dataset with each pixel providing daily kWh generation for each kWp (peak) installed capacity within that pixel at a monthly resolution. The dataset has been generated using extensive simulation and validation of solar insolation, power conversion losses, effects of atmosphere, and panel aging using 20 years of documented data. For each fishnet cell, the CF factors were aggregated using the following:

$$CF_{M,FN} = \frac{1}{n} \sum_1^n PX_V \tag{3}$$

where, $CF_{M,FN}$ is the CF factor for each month for each FN, n is the number of CF pixels in the FN and PX_V is the pixel value of each pixel in the FN geographical bound. All three datasets ($PPLN_{FN}$, CF_{FN}, BA_{FN}) and FN geometries were first processed on the ArcGIS Pro desktop application to harmonize coordinate reference systems and then uploaded to the GEE platform for processing. On GEE, we split the datasets based on individual FN geometries and aggregated the three datasets based on the rules highlighted above. GEE's cloud computing architecture can process a large number of datasets in a relatively shorter time frame and outputs a tabulated dataset for further processing.

Bottom-up method. To generate ground truth building footprints, we collected building polygon shapes as vector layers from big data sources. For building footprint samples, we used AI-generated building footprints by Microsoft AI and Ecopia AI teams. These two datasets cover the entire USA, Canada, and 39 African countries. The samples from Microsoft AI and Ecopia AI (>300 million individual buildings) were split up based on the FN layer for each FN cell overlapping the sample countries, further masked to remove building footprints outside of the BA_{FN} layer. The unmasked building footprints were aggregated based on the following:

$$BF_{FN} = \sum BP(NM)_V \tag{4}$$

where BF_{FN} is the aggregated building footprint for each sample FN and $BP(NM)_V$ is the unmasked individual building footprint polygon area within the FN that is overlapping the BA_{FN} layer. The masking removed buildings constructed after the 2015 reference year. Although the BA_{FN} layer covers the entire extent of the built-up area globally, it can still miss some built-up areas due to artifacts in satellite imagery. However, this is negligible and for the purpose of our study considered as the reference layer on which to base our analysis. Overlapping building footprints were dissolved into a single polygon before splitting and polygons being intersected by an FN boundary were split up at the line of intersection. For the rest of the world, OSM-derived building footprints (accessed April 2020) were analyzed and aggregated based on Eq. (4). A total of 4000 global FN samples were selected from the OSM building dataset. The sampling strategy to generate the 4000 OSM samples was to extract buildings bound by FNs that had BF_{FN}/BA_{FN} ratio of between 0.15 and 0.11. These ratios correspond to the 75th percentile and 50th percentile of data processed from Microsoft AI and Ecopia AI datasets. In total, we were able to successfully collect samples from nearly all the global countries covering different stages of socio-economic development, cultural spread, and geographical locations.

The road length metric was derived entirely from the OSM datasets. To process the RL dataset, OSM's planetary dataset file (accessed April 2020) was used. The planetary file has roads represented in the form of lines attributed by the different types e.g., residential, highway, footpath, etc. The line feature was split up based on each global FN, masked using BA_{FN} layer, and aggregated based on the following:

$$RL_{FN} = \sum L(NM)_V \tag{5}$$

where RL_{FN} is the aggregated road length for each global FN and $L(NM)_V$ is the individual length of all the roads within each FN overlapping BA_{FN} layer. It was observed during the aggregation of the RL_{FN} dataset, that road endpoints in the OSM dataset overlap in some locations, these overlaps were dissolved into single line features prior to aggregation. Also, some roads extend beyond the BA_{FN} layer within each FN. These roads were clipped at the BA_{FN} boundary to maintain a homogenous extent for the region of interest for all the datasets used in the analysis. In total, we were successful in processing the road infrastructure for nearly all the countries of the world.

Loading, splitting, and geometry processing for the building footprints and the road lines were performed using ArcGIS PRO's multicore support. The aggregation of the datasets and mapping of the aggregated dataset with FN boundaries was performed using custom python scripts built on DASK[70] parallel compute module. Due to the sheer size of the data being processed, we found that multicore architectures and parallel computing frameworks developed in recent years can be of great use in designing and executing planetary-scale analysis with minimal cost and time requirements.

We have used cloud computation in the form of the Google Earth Engine platform. In general, parallelization of the data processing was performed in four ways, first when disaggregating global raster datasets and mapping the aggregated values to our global FNs. We did this task for 12 monthly solar rasters, 1 Landcover raster, and finally for the global population raster. Here, we ran all the raster algebra and mapping codes on Google Earth Engine where the backend data processing was split into multi parallel streams and executed on google cloud infrastructure. This led to the speeding up of the data crunching as raster algebra is highly parallel in its execution where each raster cell can be worked on independently. Second, when disaggregating global vector datasets and mapping the aggregated values to our global FNs. As of the time of writing this article, Google Earth Engine was less efficient in undertaking vector analysis. The reason for this is that instead of pixel raster surface which can be computed in a parallel fashion, vector datasets need specialized algorithms. Here, we utilized the multi-core parallel processing architecture of ARCGIS Pro where vector datasets in the form of building footprints and roads can be effectively processed by splitting the map into different regions and processing each region on a single core. This task is performed in the backend using apache spark. Third, when doing data aggregation and manipulation inside desktop python environment. Large components of the data processing ranging from calculating rooftops to LCOE calculations were performed using the DASK framework. In this framework, data frames can be broken down into small chunks and processed in parallel. This provided us with a significant reduction in processing times. Fourth, when training ML model. Custom scripts were written on python to perform hyperparameter optimization and 10-fold cross-validation in a parallel fashion. Each of the 10-folds of the cross-validation was assigned to the independent core of our 12-core machine which led to faster convergence of the hyperparameters to an optimal solution.

Downscaling model. The ML model was trained on $PPLN_{FN}$, RL_{FN}, BA_{FN} as independent variables and BF_{FN} as the dependent variable for each sample FN. The first step in model preparation was to impute the missing data in the independent variables. The imputations are necessary as some FNs have either missing population or road length data due to the global scale of the analysis. This discrepancy is expected and is present due to OSM roads not being mapped for every road on the planet and also due to the downscaling methodology used in generating the original population raster by WorldPop.

Data imputation was handled using a custom python script utilizing the Scikit-Learn[71] module's iterative impute function. Further, we generated the downscaling

model using XGBoost[72] framework. When choosing between Neural Network-based framework or Gradient boosting frameworks, we used the latter as XGBoost framework has shown superior performance, while using considerably less computation time to reach an optimum model state. Also, being run on CPU-only architecture, the XGBoost framework can generate repeatable results in each subsequent run, which is difficult to achieve on a GPU-based framework like Neural Networks due to inherent uncertainty induced by the massive parallel compute architecture of a GPU.

The base XGBoost model was customized for our task by hyper-tuning the parameters of the model using 5-fold cross-validation. Each fold of the cross-validation generated a mean square error (MSE) loss metric at the end of its run. The mean of all five MSE was chosen as the metric to reduce during the hyper-tuning process (final model parameters are present in Supplementary Table 1). The trained model was then used to estimate building footprint values for each global FN cell using $PPLN_{FN}$, RL_{FN}, BA_{FN} values. The final output of the downscaling (BFE_{FN}) was stored as a global 10 km resolution raster file where each pixel represents the estimated aggregated building footprint for each FN.

Technical potential estimation. To calculate RTSPV potential from BFE_{FN}, we made some generalizing assumptions to maintain uniformity in our calculations. We assumed that the estimated building footprint is representative of the available rooftop area in each FN i.e., 100% of the estimated rooftop is available for solar panel installation. To install 1 kWp of roof-mounted solar PV, 10 m^2 of rooftop area is required, which is in line with the thin film technology currently in use. The roof-mounted solar PV is installed at the optimum angle for each latitude and is sun-facing and shade-free to generate maximum electricity output. The building rooftops are flat in design leading to the utilization of the entire rooftop for the installation of solar panels.

Based on the assumptions i.e., 10 m^2 area for a 10% efficient panel, the technical solar potential is calculated for all the global FNs for 12 months using the following:

$$SP_{M,FN} = BFE_{FN} \times CF_{M,FN} \times \frac{Days_M}{10} \qquad (6)$$

where SP is the technical solar potential, BFE_{FN} is the estimated roof area in m^2, CF is the conversion factor, M is the month and FN is the unique fishnet cell, and $Days_M$ is the number of days in the respective month. The World Bank's solar conversion factors are available for regions between 60°N and 45°S. For regions beyond 60°N and 45°S, we assumed a constant conversion factor of 3.5 kWh/kWp/day. This has led to a slight over-assessment of the potential for countries like Sweden, Norway. However, as the density of built-up area reduces significantly beyond the 60°N and 45°S, the total global error due to this assumption remains small. The calculations for generating the solar potential are processed using custom python scripts using the DASK module to handle massive arithmetic operations. Further, the processed solar potential raster dataset is stored as a geopackage file for visualizations and economic calculation.

Cost calculation. LCOE provides an easy and robust method to compare the economic viability of a project within a specific FN. It was assumed that the capital cost of the installation (CAPEX) will be staggered to the first year of commissioning and the installed panels will have a lifetime of 25 years. The geo-mapping for CAPEX, operating expenditure (OPEX), and discount rate (DR) was sourced from IRENA's renewable energy cost report 2019. The LCOE for each FN is calculated using the following:

$$LCOE_{FN} = \frac{\sum_{t=1}^{25} \frac{CAPEX_{FN} + OPEX_{FN,t}}{(1+DR)^t}}{\sum_{t=1}^{25} \sum_{M=1}^{12} \frac{SP_{M,FN}}{(1+DR)^t}} \qquad (7)$$

where, $CAPEX_{FN}$ (2019 \$/kW) is the capital expenditure in installing the RTSPV system for the given FN, $OPEX_{FN,t}$ (2019 \$/kW) is the operational and maintenance expenditure for the given FN and for the specific year (t), t is the year number, DR is the discount rate, M is the month number, and $SP_{M,FN}$ (kWh/month) is the potential generation for the given month and FN. We have used CAPEX data for 17 different countries and allocated the average CAPEX value to the rest of the countries based on the continent they are in. OPEX and DR have values based on OECD and Non-OECD country classifications. Country-wise aggregation of LCOE, SP from each FN from their respective high-resolution intra country values has been done using the following rule

$$LCOE_{country} : \begin{cases} CAPEX : aggregation\_of\_individual\_FN\_values \\ OPEX : aggregation\_of\_individual\_FN\_values \\ SP : aggregation\_of\_individual\_FN\_values \\ DR : mean\_of\_indiviual\_FN\_values \end{cases} \qquad (8)$$

It should be noted that we did not consider the cost of additional grid expansion or storage infrastructure to attain the full technical rooftop solar PV potential. Also, the cost of decommissioning and scrap metal value of the installation was not considered at the end of the 25-year lifetime of the projects. While calculating the SP and LCOE, it was assumed that no rooftop solar PV installation exists globally, and all the additional capacities will start their commissioning from the year 2019.

Limitations. Our assessment is based on the accuracy of the global landcover layer, which with its 100 m resolution can in some locations overestimate the built-up area extents. In addition, the landcover classifies roads, parking lots, boundaries of green areas, tennis courts, and archeologically significant areas as built-up areas with misclassification varying between different regions. Our assumption that the rooftops being flat, shadow-free and sun-facing with a full rooftop available for installation adds to the methodological limitations. Next, the big data related to building footprints and global roads have inherent methodological limitations like the simplified representation of a complex rooftop with a square polygon, over-lapping roads, etc.

For technical potential calculations, we assumed that 100% of the estimated rooftop is available for installing solar panels i.e., orientation and slope of the building are not accounted for the 100% rooftop availability assumption-based results in our main analysis. These assumptions can lead to limitations in the real-world interpretation of main results as a fraction of rooftop may be available for the installation of solar panels. To account for this, we have documented regional change in potential as an uncertainty analysis for a combination of rooftop scaling factors and panel efficiencies. In the current literature, reduction of total rooftop area to available rooftop area is generally done through a rooftop scaling factor which is a proxy for loss of rooftop area due to orientation, slope, and roof superstructures like chimneys, etc. Although, some studies exist at the country level where the rooftop scaling factor is documented, on a global scale no authoritative dataset exists that can demarcate country-wise rooftop scaling factors. Further work is required to document the country-specific rooftop scaling factors which are outside the scope of the research aim of this study.

Our cost assumptions cover 17 different countries across continents with average values for the rest of the countries. These assumptions can assign increased or decreased LCOE values to a certain region like Africa and South America. Another limitation of the cost assumptions is the inability of the LCOE metric to capture intra country variation of LCOE to a high degree due to lack of high-resolution cost data. Also, cost variability due to additional grid rollout, tariff mechanisms, and global change in prices due to trade protectionism practices are beyond the scope of the current assessment. Finally, we calculated all the cost and potential metrics assuming that no installed capacity exists for the ROI, where in the present time horizon, some installed capacity does exist.

A majority of limitations can be attributed to the underlying data used in our assessment, which can be improved with subsequent advancement in the methodology of the data providers. Further research can be undertaken to reduce the methodological limitations that are currently bootstrapped by the data availability and lack of homogenous global data.

Data availability
The data that support the findings of this study are available from the corresponding author upon reasonable request. The global road map is based on OpenStreetMap (OSM), which can be freely downloaded. The planet file used in this study is downloaded on April 1, 2020. The landcover map is based on Copernicus Global Land Service: Land Cover 100 m: collection 3: epoch 2015: Globe (https://doi.org/10.5281/zenodo.2583745). Other data sources that are free to use are provided in the main text and in the "Methods" section.

Code availability
Pseudocode to undertake this analysis can be found in the supplementary material (Supplementary Note 1) and should be read in conjunction with the "Methods" section. Base XGBoost Model is available at https://xgboost.readthedocs.io. The python script for plotting and data aggregation is available from the corresponding author upon reasonable request.

References
1. IEA. Renewables 2019: Analysis and forecasts to 2024. *International Energy Agency* https://www.iea.org/reports/renewables-2019 (2019).
2. International Renewable Energy Agency, A. D. *Renewable Power Generation Costs in 2019*. https://www.irena.org/publications/2020/Jun/Renewable-Power-Costs-in-2019 (2020).
3. IEA, International Renewable Energy Agency, United Nations Statistics Division, The World Bank & World Health Organization. *Tracking SDG 7: The Energy Progress Report 2019*. The World Bank (2019).
4. IEA. *Global Energy Review 2020*. Global Energy Review 2020. https://doi.org/10.1787/a60abbf2-en (OECD, 2020).
5. IRENA. *Future of Solar Photovoltaic: Deployment, investment, technology, grid integration and socio-economic aspects (A Global Energy Transformation: paper)* (2019).

6. Bódis, K., Kougias, I., Jäger-Waldau, A., Taylor, N. & Szabó, S. A high-resolution geospatial assessment of the rooftop solar photovoltaic potential in the European Union. *Renew. Sustain. Energy Rev.* https://doi.org/10.1016/j.rser.2019.109309 (2019).

7. Gagnon, P., Margolis, R., Melius, J., Philips, C. & Elmore, R. *Rooftop Solar Photovoltaic Technical Potential in the United States: A Detailed Assessment.* NREL Technical Report https://www.nrel.gov/docs/fy16osti/65298.pdf (2016).

8. Mainzer, K. et al. A high-resolution determination of the technical potential for residential-roof-mounted photovoltaic systems in Germany. *Sol. Energy* https://doi.org/10.1016/j.solener.2014.04.015 (2014).

9. Walch, A., Castello, R., Mohajeri, N. & Scartezzini, J. L. Big data mining for the estimation of hourly rooftop photovoltaic potential and its uncertainty. *Appl. Energy* https://doi.org/10.1016/j.apenergy.2019.114404 (2020).

10. Creutzig, F. et al. The underestimated potential of solar energy to mitigate climate change. *Nat. Energy* https://doi.org/10.1038/nenergy.2017.140 (2017).

11. Freitas, S., Catita, C., Redweik, P. & Brito, M. C. Modelling solar potential in the urban environment: state-of-the-art review. *Renew. Sustain. Energy Rev.* https://doi.org/10.1016/j.rser.2014.08.060 (2015).

12. Betak, J. et al. *Global Photovoltaic Power Potential by Country (English). Energy Sector Management Assistance Program (ESMAP).* (World Bank Group, Washington, D.C.).

13. Korfiati, A. et al. Estimation of the global solar energy potential and photovoltaic cost with the use of open data. *Int. J. Sustain. Energy Plan. Manag.* 9, 17–29 (2016).

14. Naroll, R. Floor Area and Settlement Population. *Am. Antiq.* https://doi.org/10.2307/277689 (1962).

15. IEA. Energy Technology Perspectives 2016: Towards Sustainable Urban Energy Systems. *Int. Energy Agency* (2016).

16. Hoogwijk, M. M. *On the Global And Regional Potential of Renewable Energy Sources.* Assembly (2004).

17. Izquierdo, S., Rodrigues, M. & Fueyo, N. A method for estimating the geographical distribution of the available roof surface area for large-scale photovoltaic energy-potential evaluations. *Sol. Energy* 82, 929–939 (2008).

18. Ordóñez, J., Jadraque, E., Alegre, J. & Martínez, G. Analysis of the photovoltaic solar energy capacity of residential rooftops in Andalusia (Spain). *Renew. Sustain. Energy Rev.* https://doi.org/10.1016/j.rser.2010.01.001 (2010).

19. Defaix, P. R., van Sark, W., Worrell, E. & de Visser, E. Technical potential for photovoltaics on buildings in the EU-27. *Sol. Energy* 86, 2644–2653 (2012).

20. Assouline, D., Mohajeri, N. & Scartezzini, J.-L. Quantifying rooftop photovoltaic solar energy potential: a machine learning approach. *Sol. Energy* 141, 278–296 (2017).

21. Wegertseder, P., Lund, P., Mikkola, J. & García Alvarado, R. Combining solar resource mapping and energy system integration methods for realistic valuation of urban solar energy potential. *Sol. Energy* 135, 325–336 (2016).

22. Lehmann, H. & Peter, S. Assessment of roof and façade potentials for solar use in Europe. *Insitute Sustain. Solut. Innov.* https://doi.org/10.1017/CBO9781107415324.004 (2003).

23. Lee, M., Hong, T., Jeong, K. & Kim, J. A bottom-up approach for estimating the economic potential of the rooftop solar photovoltaic system considering the spatial and temporal diversity. *Appl. Energy* 232, 640–656 (2018).

24. Jacobson, M. Z. et al. 100% Clean and Renewable Wind, Water, and Sunlight All-Sector Energy Roadmaps for 139 Countries of the World. *Joule* https://doi.org/10.1016/j.joule.2017.07.005 (2017).

25. Gernaat, D. E. H. J., de Boer, H. S., Dammeier, L. C. & van Vuuren, D. P. The role of residential rooftop photovoltaic in long-term energy and climate scenarios. *Appl. Energy* https://doi.org/10.1016/j.apenergy.2020.115705 (2020).

26. Castellanos, S., Sunter, D. A. & Kammen, D. M. Rooftop solar photovoltaic potential in cities: How scalable are assessment approaches? *Environ. Res. Lett.* https://doi.org/10.1088/1748-9326/aa7857 (2017).

27. Lloyd, C. T., Sorichetta, A. & Tatem, A. J. Data Descriptor: High resolution global gridded data for use in population studies. *Sci. Data* https://doi.org/10.1038/sdata.2017.1 (2017).

28. Verma, V., Kumar, R. & Hsu, S. 3D Building Detection and Modeling from Aerial LIDAR Data. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06).* Vol. 2, 2213–2220, https://www.computer.org/csdl/proceedingsarticle/cvpr/2006/259722213/12OmNCm7Bws (2006).

29. Rottensteiner, F. & Briese, C. A new method for building extraction in urban areas from high-resolution LIDAR data. In *International Archives of Photogrammetry Remote Sensing and Spatial Information Sciences.* (Natural Resources Canada, 2002).

30. Redweik, P., Catita, C. & Brito, M. Solar energy potential on roofs and facades in an urban landscape. *Sol. Energy* 97, 332–341 (2013).

31. Zha, Y., Gao, J. & Ni, S. Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. *Int. J. Remote Sens.* 24, 583–594 (2003).

32. Buffat, R., Grassi, S. & Raubal, M. A scalable method for estimating rooftop solar irradiation potential over large regions. *Appl. Energy* 216, 389–401 (2018).

33. Ferro, A., Brunner, D. & Bruzzone, L. Automatic detection and reconstruction of building radar footprints from single VHR SAR images. *IEEE Trans. Geosci. Remote Sens.* https://doi.org/10.1109/TGRS.2012.2205156 (2013).

34. Jochem, A., Höfle, B., Rutzinger, M. & Pfeifer, N. Automatic roof plane detection and analysis in airborne lidar point clouds for solar potential assessment. *Sensors* 9, 5241–5262 (2009).

35. Lukač, N., Seme, S., Žlaus, D., Štumberger, G. & Žalik, B. Buildings roofs photovoltaic potential assessment based on LiDAR (Light Detection And Ranging) data. *Energy* 66, 598–609 (2014).

36. Hong, T., Lee, M., Koo, C., Jeong, K. & Kim, J. Development of a method for estimating the rooftop solar photovoltaic (PV) potential by analyzing the available rooftop area using Hillshade analysis. *Appl. Energy* 194, 320–332 (2017).

37. As-syakur, A. R., Adnyana, I. W. S., Arthana, I. W. & Nuarsa, I. W. Enhanced Built-Up and Bareness Index (EBBI) for Mapping Built-Up and Bare Land in an Urban Area. *Remote Sens.* 4, 2957–2970 (2012).

38. Wong, M. S. et al. Estimation of Hong Kong's solar energy potential using GIS and remote sensing technologies. *Renew. Energy* 99, 325–335 (2016).

39. Brito, M. C., Gomes, N., Santos, T. & Tenedório, J. A. Photovoltaic potential in a Lisbon suburb using LiDAR data. *Sol. Energy* 86, 283–288 (2012).

40. Heris, M. P., Foks, N. L., Bagstad, K. J., Troy, A. & Ancona, Z. H. A rasterized building footprint dataset for the United States. *Sci. Data* https://doi.org/10.1038/s41597-020-0542-3 (2020).

41. Tiecke, T. G. et al. Mapping the world population one building at a time. *arXiv* https://doi.org/10.1596/33700 (2017)

42. Leasure, D., Dooley, C., Bondarenko, M., Tatem, A. W. peanutButter: An R package to produce rapid-response gridded population estimates from building footprints, version 0.2.1. https://doi.org/10.5258/SOTON/WP00678 (2020).

43. Esch, T. et al. Breaking new ground in mapping human settlements from space–The Global Urban Footprint. *ISPRS J. Photogramm. Remote Sens.* https://doi.org/10.1016/j.isprsjprs.2017.10.012 (2017).

44. Marconcini, M. et al. Outlining where humans live, the World Settlement Footprint 2015. *Sci. Data* https://doi.org/10.1038/s41597-020-00580-5 (2020).

45. Corbane, C. et al. Automated global delineation of human settlements from 40 years of Landsat satellite data archives. *Big Earth Data* https://doi.org/10.1080/20964471.2019.1625528 (2019).

46. Buchhorn, M. et al. Copernicus Global Land Service: Land Cover 100m: epoch 2015: Globe. Dataset of the global component of the Copernicus Land Monitoring Service. https://doi.org/10.5281/zenodo.3243509 (2019).

47. Ferri, S., Siragusa, A., Sabo, F., Pafi, M. & Halkia, M. *The European Settlement Map 2017 Release. Methodology and output of the European Settlement Map.* https://doi.org/10.2760/780799 (2017).

48. Corbane, C. et al. Convolutional neural networks for global human settlements mapping from sentinel-2 satellite imagery. *arXiv* 2006.03267 (2020).

49. Bódis, K., Kougias, I., Taylor, N. & Jäger-Waldau, A. Solar photovoltaic electricity generation: a lifeline for the European coal regions in transition. *Sustainability* https://doi.org/10.3390/su11133703 (2019).

50. Maloof, M. A., Langley, P., Binford, T. O., Nevatia, R. & Sage, S. Improved rooftop detection in aerial images with machine learning. *Mach. Learn.* 53, 157–191 (2003).

51. Guo, Z. et al. Identification of village building via Google Earth images and supervised machine learning methods. *Remote Sens.* https://doi.org/10.3390/rs8040271 (2016).

52. Mohajeri, N. et al. A city-scale roof shape classification using machine learning for solar energy applications. *Renew. Energy* 121, 81–93 (2018).

53. Yang, H. L., Lunga, D. & Yuan, J. Toward country scale building detection with convolutional neural network using aerial images. In *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)* 870–873, https://ieeexplore.ieee.org/document/8127091 (2017).

54. Zhang, A., Liu, X., Gros, A. & Tiecke, T. Building detection from satellite images on a global scale. *arXiv* 1707.08952 (2017).

55. Assouline, D., Mohajeri, N. & Scartezzini, J.-L. Large-scale rooftop solar photovoltaic technical potential estimation using Random Forests. *Appl. Energy* 217, 189–211 (2018).

56. OpenStreetMapContributors. *OpenStreetMap.* https://www.openstreetmap.org (2020).

57. Stevens, F. R., Gaughan, A. E., Linard, C. & Tatem, A. J. Disaggregating census data for population mapping using Random forests with remotely-sensed and ancillary data. *PLoS ONE* https://doi.org/10.1371/journal.pone.0107042 (2015).

58. Barrington-Leigh, C. & Millard-Ball, A. The world's user-generated road map is more than 80% complete. *PLoS ONE* https://doi.org/10.1371/journal.pone.0180698 (2017).

59. [Data/information/map] obtained from 812 the "Global Solar Atlas 2.0, a free, web-based application is developed and operated by the company Solargis s.r.o. on behalf of the World Bank Group, utilizing Solargis data, with funding provided by the Energy Sector.

60. Millstein, D., Wiser, R., Bolinger, M. & Barbose, G. The climate and air-quality benefits of wind and solar power in the United States. *Nat. Energy* **2**, 17134 (2017).

61. Yalew, S. G. et al. Impacts of climate change on energy systems in global and regional scenarios. *Nat. Energy* https://doi.org/10.1038/s41560-020-0664-z (2020).

62. Barbose, G. & Satchwell, A. J. Benefits and costs of a utility-ownership business model for residential rooftop solar photovoltaics. *Nat. Energy* https://doi.org/10.1038/s41560-020-0673-y (2020).

63. Sasse, J.-P. & Trutnevyte, E. Regional impacts of electricity system transition in Central Europe until 2035. *Nat. Commun.* **11**, 4972 (2020).

64. Zhang, H. et al. Solar photovoltaic interventions have reduced rural poverty in China. *Nat. Commun.* **11**, 1969 (2020).

65. Lu, T. et al. India's potential for integrating solar and on- and offshore wind power into its energy system. *Nat. Commun.* **11**, 4750 (2020).

66. Sterl, S. et al. Smart renewable electricity portfolios in West. *Afr. Nat. Sustain.* **3**, 710–719 (2020).

67. Parag, Y. & Sovacool, B. K. Electricity market design for the prosumer era. *Nat. Energy* **1**, 16032 (2016).

68. International Energy Agency (IEA). Electricity Information: overview 2019. *Stat. IEA* https://doi.org/10.1787/e0ebb7e9-en (OECD Publishing. Paris, 2019).

69. Gorelick, N. et al. Google Earth engine: planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* https://doi.org/10.1016/j.rse.2017.06.031 (2017).

70. DASKDevelopmentTeam. *Dask: Library for Dynamic Task Scheduling.* https://docs.dask.org/en/latest/cite.html (2016).

71. Pedregosa, F. et al. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).

72. Chen, T. & Guestrin, C. *XGBoost: A Scalable Tree Boosting System.* https://doi.org/10.1145/2939672.2939785 (2016).

Acknowledgements

Author contributions

S.J. and J.G. conceived the research idea. S.J. designed and developed the framework and model along with the codes. S.J., S.M., and P.H. designed the GIS and data analysis frameworks. S.J. created the figures and drafted the manuscript. P.R.S. and B.O.G. provided valuable insights on the results. All authors discussed the results and contributed to the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-021-25720-2.

Correspondence and requests for materials should be addressed to Siddharth Joshi.

Peer review information *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.