International Institute for Applied Systems Analysis
IIASA www.iiasa.ac.at

Address: IIASA, Schlossplatz 1, A-2361 Laxenburg, Austria
Email: repository@iiasa.ac.at **Telephone:** +43 (0)2236 807 342

YSSP Report
Young Scientists Summer Program

Role of Rooftop Solar Photovoltaics in Global Energy Transitions

Author: Siddharth Joshi
Email: siddharth.joshi@ucc.ie

Approved by

Behnam Zakeri

Laxenburg 30.09.2021

Supervisor: Behnam Zakeri, Alessio Mastrucci and Volker Krey

Program: Energy, Climate, and Environment Program

30th September 2021

This report represents the work completed by the author during the IIASA Young Scientists Summer Program (YSSP) with approval from the YSSP supervisor.

It was finished by 30th September 2021 and has not been altered or revised since.

This research was funded by IIASA and its National Member Organizations in Africa, the Americas, Asia, and Europe.

ZVR 524808900

Table of Contents

Abstract

Rooftop solar PV (RTSPV) currently accounts for 40% of the global solar PV installed capacity and one-fourth of the total renewable capacity additions in 2019. Being decentralised in nature, RTSPV technology can lead to consumer driven breakthrough in tackling climate change, reducing local air pollution, and providing affordable energy access to the areas lacking electrification. Thus, it is pertinent to assess the role of RTSPV in the future energy systems. Currently some global Integrated Assessment Models (IAMs) represent solar PV technology as an aggregation of Utility (UTSPV) and RTSPV deployment application, which makes it difficult to assess the potential of RTSPV, including related socio-economic transitions at a global scale. This can be attributed to limited studies being conducted to assess RTSPV as a separate deployment application from utility scale application.

To mitigate these gaps in the literature, here we assess the role of RTSPV technology at a global scale by designing a Machine Learning based assessment framework that is capable of estimating RTSPV potential for electricity generation at a global scale with a high spatiotemporal resolution and also provide a global building dataset that can aid in cross disciplinary research utilising geomapped global building data. The datasets generated by this framework were then used to create a new tool that will aid researchers in identifying and selecting fully mapped building footprint mapping from Open Street Map (OSM) dataset for use in their research. Further, we utilise this framework to estimate future changes (medium time horizon) in potentials using spatially explicit built-up extent and population datasets derived from Shared Socioeconomic Pathways (SSP). Finally, with the aid of 12-region MESSAGEix-GLOBIOM model and updated country-wise technology supply cost curves (derived from estimated potentials and costs), we assess the role of RTSPV in achieving low carbon energy and climate mitigation scenarios.

From the assessment, we find that for different growths envisioned under the SSPs, global rooftop area will increase between 3-26% by 2050 compared to its 2020 values. We also find that a rooftop area growth of 7-200% by 2050 over its 2020 values can be observed for different continents for SSP narratives. This highlights the importance of policy formulation with consideration to variation in growth rates between regions. The current global building stock has enough rooftop area to meet the current yearly aggregated global electricity demand. When converting the available global rooftop area to technical potential of RTSPV technology, a growth in technical potential of between 25-50% over 2020 values can be observed for different worlds envisioned by the SSP narratives. The largest growth in technical potential between years 2020-2050 occurs in regions that have high quality solar resource endowment. In terms of technology attributes and building archetypes, net rooftop availability is the largest constraint in deployment of RTSPV technology. Our analysis show that a threshold of global net 50% rooftop availability is present beyond which only marginal uptake of RTSPV occurs. This threshold reduces to 30% when emission bounds are introduced into the global energy systems. Finally, the spatially explicit assessment showed that there is a mismatch between quality of solar resource and the hotspots of built-up areas which highlights the need for informed sub-national policy formulation dealing with RTSPV technology.

Acknowledgments

I would first acknowledge my supervisors and staff at IIASA whose efforts have made the virtual YSSP project a success. I would like to thank my main supervisors Dr. Behnam Zakeri, Dr. Alessio Mastrucci and Dr. Volker Krey who have guided me throughout this project and provided valuable feedback on the research. My special thanks and appreciation go out to Dr. Behnam Zakeri and a very congratulations to him for becoming father to LEO.

From my home institute of MaREI, University College Cork, I would like to convey my gratitude to Prof. Brian Ó Gallachóir, Dr. James Glynn and Dr. Paul Holloway who have provided excellent financial, emotional, and mental support along with research feedback during the YSSP tenure. I would also like to express my gratitude to Dr. Shivika Mittal and Distinguished Prof. P.R. Shukla for their guidance on the scenario analysis and data harmonization.

Lastly, I would like to thank my parents Meenakshi and Ramesh, my brother Sauhard, and Shatabdi for family support during these trying yet exciting months of YSSP.

About the authors

Siddharth Joshi is a final year PhD student in the Research Centre for Energy, Climate and Marine (MaREI) at University College Cork, Ireland. His research focuses on the use of high-performance computing and machine learning frameworks for global high-resolution spatiotemporal supply and demand modelling in energy systems. Before his PhD he worked as Head of R&D at a telemedicine start-up, business consultant and instrumentation engineer in oil and natural gas sector. He holds a MEngSc in Sustainable Energy from University College Cork and BE in Electronics, Instrumentation and Control from Thapar University, India. His research interests include the application of big-data, nature inspired machine learning architectures and Geographic Information Systems in tackling the real-world problems associated with the supply and demand estimation in the energy system models. (Contact: Siddharth.joshi@ucc.ie, Twitter: @Siddhar28031625)

1. Introduction

Low carbon and cost-effective electricity generation has the potential to contribute significantly in tackling the problem of climate change, energy access and sustainable development. Current electricity generation from renewable energy sources is primarily derived from wind, hydro and solar. Amongst these, solar photovoltaics (PV) are the fastest growing technology, of which rooftop solar PV (RTSPV) currently occupies 40% of the total installed solar PV capacity[1]. To better understand the role that RTSPV will play in the future energy systems, we need to assess the global electricity generation potential of it and analyse the dynamics of RTSPV in a global energy system using scenario analysis and modelling.

A global assessment of RTSPV technology is a complex task as the smallest unit of assessment is a rooftop. This complexity is compounded by the fact that building stock archetypes change between geographies and are dependent on the socio-economic and cultural factors prevalent in the region of interest (ROI). In the past, bottom-up[2-7] modelling approaches were used to assess the roof area at sub-national and national scales. Here, the studies focussed on extrapolation of relationships between socioeconomic drivers and rooftop area from a small sample region to a larger ROI. Although these methods are good for rapid estimation of rooftop area, they often report lower accuracies[8] than the highly spatially resolved methods that utilise large scale surveying of building stocks. On the other hand the highly spatially resolved techniques like LiDAR[9,10] based rooftop mapping which use drone mounted laser to map landscape in 3D, and Machine Learning (ML) based Object detection[11,12] have shown promising results for ROI covering continental scale.

The Lidar based rooftop mapping is currently the most accurate method of determining the rooftop area along with capturing the rooftop attributes at scale. But these methods require significant investment in aerial imaging and computational costs as a result of which the most common implementation of LiDAR based rooftop mapping is limited to a city scale analysis. ML based models form the next class of methods that can aid in detection of building rooftops at scale, but these methods have shown limited suitability for a global scale study as the trained ML models require heavy investment in training data that should have enough diversity to cover a global ROI[13]. Additionally, server scale computational environment is required to train and generate inference from these trained ML model which requires significant cost and time investment. As a result of this, the largest ROI tackled by a ML based approach covers the continent of Africa[14]. It can be inferred from the progress made by ML based algorithm that a global scale study is feasible, but due to the complexities of computation time and generation of data for model training it may be sometime till we reach the milestone of open source global geomapped building dataset.

A third stream of methods that can aid in rapid assessment of rooftop area at ROIs spanning continental scales is to use a hybrid approach. This approach utilises spatial relationship between the landcover mapping (derived from remote sensed imagery), socioeconomic metrices and actual on ground building stock attributes to infer rooftop area for out of sample regions. Two[9,15] studies that have demonstrated this hybrid approach utilise either statistical inferencing or ML based modelling to generate these relationships.

Generation of a harmonised dataset that documents the global rooftop area is of extreme importance to not only energy system modellers but also to the national and international research institutions as this spatially explicit dataset can aid in energy access, disaster management,

sustainable growth studies and in tackling the global problem of climate change mitigation. Of more importance is that a global harmonised spatially explicit dataset be generated that documents the future spatial growth in the rooftop area to aid in cross domain scenario analysis and policy formulation. Here, with an aim to assess the role of RTSPV in global energy transitions we extend the methodology that the author of this study have developed in a previous study published in *Nature Communications*[16] where they have assessed the global RTSPV potential for renewable electricity generation in 2015, by incorporating the Shared Socioeconomic Pathways (SSP) narratives to assess the global rooftop area at a 1/8 degree spatial resolution and at a monthly temporal resolution for years spanning 2020-2050. The SSP narratives[17] examine how global society, demographics and economics might change over the next century by quantifying the narratives into numerical metrics that can be interpreted by mathematical models.

The global rooftop area estimates for SSP narratives along with updated global Utility scale potential assessment and costs were then used to run technology uptake and climate change mitigation scenarios using IIASA's global 12 region MESSAGEix-GLOBIOM model. This way we analysed the role of rooftop solar PV in addressing the twin challenges of sustainable development and climate change with co-benefits in advancing the sustainable development goals (SDG) 7 and SDG 13 in global energy systems using an updated technology specific potential and costs along with scenario analysis. The generated datasets can also aid in advanced spatially explicit energy access, energy planning and updated energy demand analysis by cross disciplinary teams housed at IIASA. Additionally, spatially explicit datasets will aid in not only informing global energy policy questions but also aid in analysing local energy policy questions using national MESSAGEix-GLOBIOM models.

The novelty of this study lies in the first high resolution spatiotemporal global assessment of RTSPV technology for the current year and into the medium-term time horizon (2020-2050). Further, the tools developed in this study have utility not only for the energy system modellers but also for the research working in cross disciplinary research dealing with urban growth, energy access and climate change adaptation. The study also documents the first research into the effects of rooftop availability, panel efficiencies and emission bounds on RTSPV technology. Finally, this study demonstrates how a high-resolution spatiotemporal assessment of technology can be ingested into global energy models to inform energy policies ranging from city level to a global level.

The study is divided into three parts – Methods, MESSAGEix framework implementation and Results. In methods, we briefly describe the methodological basis of the study and how the generated datasets can be accessed and used. In MESSAGEix framework implementation we discuss how the datasets generated in this study can be used in a global energy system model. Finally, we document our findings in the results section and conclude with discussion of the results in a broader context.

2. Methods

To assess the role of RTSPV technology in global energy transitions, we divide the study into work packages - Data collection, Machine Learning framework development, and implementation of new potentials and costs in 12 region global MESSAGEix-GLOBIOM Model. The primary role of the ML framework is to train on the collected data and estimate rooftop area present in each tile of a custom global grid. The three work packages were executed sequentially with check and balances being implemented at the end of each work package to form the methodological basis of the study. The following paragraphs will discuss in detail how these work packages were implemented.

2.1 Data collection

We started the task of data collection by defining a global fishnet (FN) grid at a spatial resolution of 1/8 degree. This grid cell has an approximate resolution of 14 km^2 at equator and the size of the grid cell is dynamic based on the latitude it lies in. This spatial resolution of the grid was chosen to match the FN resolution of SSP derived population and built-up extent datasets which we discuss later in this report. A 14km^2 FN also provides us with a large enough extent to capture city limits at scale and small enough extent to not cover the entire conurbations within itself.

Next, we chose 2020 as our base year with 2030, 2040, and 2050 as our medium-term time horizon projection years. Primary datasets collected during this study can be categorised into either a *Vector Dataset* (building footprint polygons and geo-mapped roads) or *Raster Datasets* (base year population, base year built-up extent, future SSP derived griddled population, future SSP derived griddled built-up extent and future country wise SSP derived GDP). The attributes of the different base year datasets and their sources are documented in Table 1 with visual depiction in Figure 1.

From the collected data from big data sources, we observed that full country coverage of base year building polygon data is available for USA, UK, Australia, and Canada. Full continental coverage is available for Africa[18] sans the North Africa covering countries above Sahara Desert. For the rest of the world, building polygon data is available in Open Street Maps[19], but the coverage is sporadic with good spatial coverage available for European continent. This mismatch between completeness of OSM derived building footprints encouraged us to create our own OSM Gap Detection Tool to capture selected data that has full completeness based on our FN grid. The base year population count data covers the entire global landmass hence no further filtering or sampling of the dataset was required.

The base year global built-up extent dataset has global coverage for the year 2019. We assumed that the coverage at the end of 2019 will be equivalent to the year 2020 built-up extents. The built-up layer captures the extent of human made modifications on the earth. Using a suite of remote sensing techniques, these structures can be isolated from the natural landscape and the area occupied by these structures can be converted into a raster grid where each grid cell can represent either the built-up area contained within it or the percentage of area that is built-up. Naturally, built-up extent will capture roads, tennis courts, carparks, airport runways etc. that do not form part of the building footprint and can sometime cover 2-3 times more area than a building footprint in a built-up raster cell. To account for this, we created a ML model to downscale the built-up extent to estimated rooftop area which we will discuss in next sections.

Table 1 | Base year layers used in this study along with their attributes and sources

Layer	Type	Region	Attribute	Format	Size
FN	Fishnet Grid	Global	~ 3 million polygons	Vector Polygon	N.A.
BF20[20]	2020 Building Footprint	USA, Canada, UK, Australia, Africa	~ 800 million buildings	Vector Polygon	~100 GB
BF20[19]	2020 Building Footprint	Rest of the world - OSM	~ 250 million buildings	Vector Polygon	~200 GB
PPLN20[21]	2020 Population Count	Global	100m Resolution	Raster	~1GB
BU20[22]	2020 Built-up Area	Global	100m Resolution	Raster	~3GB
RL20[19]	Road Length	Global	~ 34 million km	Vector Polylines	~100 GB

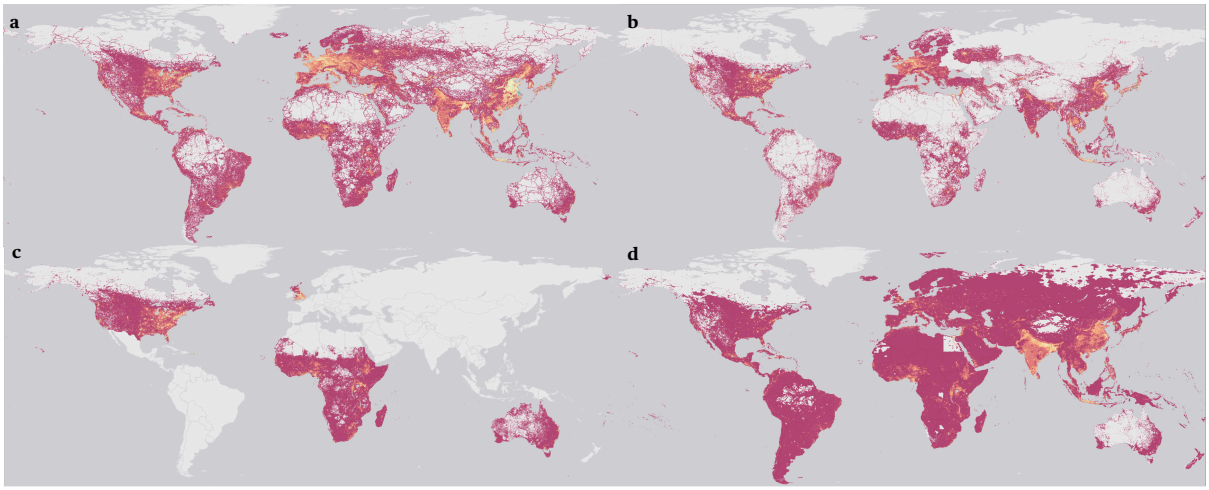

Figure 1 | Spatial spread of the base input datasets. Global geo-mapped roads extracted from Open Street Maps, **a**. Global human made built-up areas extracted from Copernicus Land Monitoring Program GLC V3.1 2020, **b**. Building footprint polygons derived from big data sources, **c**. Global geo-mapped population count for 2020 at 100m resolution derived from World POP project, **d**. The brighter yellow colour represents relatively high values of respective metrics in the datasets, with gradation to red colour representing low values of respective metrics in the datasets. Presence of light grey colour represents absence of data in the respective datasets with dark grey representing ocean.

The next step in our study after collection of base datasets for the year 2020 is to collect SSP derived datasets for the years 2020, 2030, 2040 and 2050. In total, we collected SSP derived data for gridled population[23], built-up extent[24], and GDP[25] per country data for the years 2020-2050. The gridled population count dataset and built-up extent dataset are available as raster dataset at 1/8-degree resolution, with GDP per country dataset being mapped to respective country boundaries using administrative boundary dataset from GADM project V3.6 (https://gadm.org/data.html). The attributes of the different SSP derived datasets are documented in Table 2 with visual depiction of change in metrics between year 2020 and 2050 depicted in Figure 2.

Table 2 | SSP derived layers used in this study along with their attributes and sources

Layer	Type	Region	Attribute	Format	Size
$PPLN_{X,Y}$*	Future SSP derived griddled population count	Global	1/8 degree	Raster	N.A.
$BU_{X,Y}$*	Future SSP derived griddled built-up extent	Global	1/8 degree	Raster	N.A.
$GDP_{X,Y}$*	Future SSP derived country-wise GDP	Global	Country wise	Vector Polygon	N.A.

* where "X" is the SSP narrative number, "Y" is the year for which the respective metric is provided.

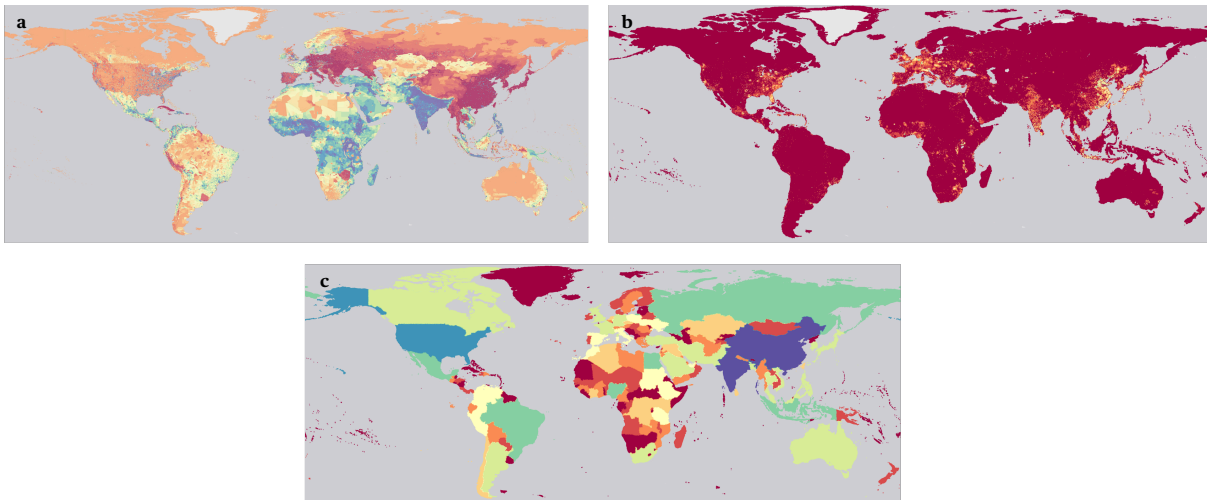

Figure 2 | Relative change in metric for SSP2 derived datasets between 2020 and 2050. **a,** Global change in geo-mapped population. Red coloured areas have relatively lowest change in population between 2020 and 2050, with blue-coloured areas representing relatively highest change in population. **b**, Global change in geo-mapped built-up areas. Red coloured areas have relatively lowest change in Built-up area between 2020 and 2050, with yellow-coloured areas representing relatively highest change in Built-up area. **c**, country-wise change in GDP. Red coloured areas have relatively lowest growth in GDP between 2020 and 2050, with blue-coloured areas representing relatively growth in GDP.

2.2 Base year calibration and spatial harmonisation

After collection and verification of base year datasets, SSP derived datasets we conducted a harmonisation of base years across the datasets. As discussed previously, 2020 was chosen to be our base year and in line with that except for the base year building polygon and base year Built-up extent, all other layers have 2020 datapoints. To account for this, we assumed that the 2019 Built-up extent of our BU20 layer will represent the 2020 datapoints. Additionally, the BF20 layer polygon will be assumed to represent building footprints for 2020 year. These assumptions add-in a component of uncertainly in the harmonisation, but at a global scale these assumptions will have very slight effect on the final output of the study due to the design of our ML framework which we discuss in next sections.

After aligning the datasets to a common base year, we aligned the datasets on a common spatial resolution. For this we first started mapping the base year datasets to the FN grid. We overlayed the FN grid on top of the BF20, PPLN20, BU20 and RL20 datasets and used a cookie cutter approach to cut and aggregate the datasets within each unique FN grid cells. Next, BU20 layer boundary inside each FN was chosen as the region of interest and any datapoint outside this BU20 boundary but inside the FN boundary was not considered. This provided us with the first stage of spatial harmonisation where only datapoints inside the BU20 layer extents were considered. To achieve this, we used area outside the BU20 layer as a masking layer to select datapoints that are not masked, Figure 3.

The base year *vector* datasets representing non-masked BF20 and RL20 datasets were processed on the ARCGIS PRO V2.8 platform, where we used the inbuilt multicore enhancements to process the cutting and aggregation of vector dataset at scale. After the cutting step, each building polygon and road polyline feature inside each unique FN grid cell was aggregated to represent a single value per FN grid cell. It should be noted that a polygon falling on the FN grid cell boundary was intersected at the boundary and only the area of the polygon inside of the respective FN was attributed to that FN.

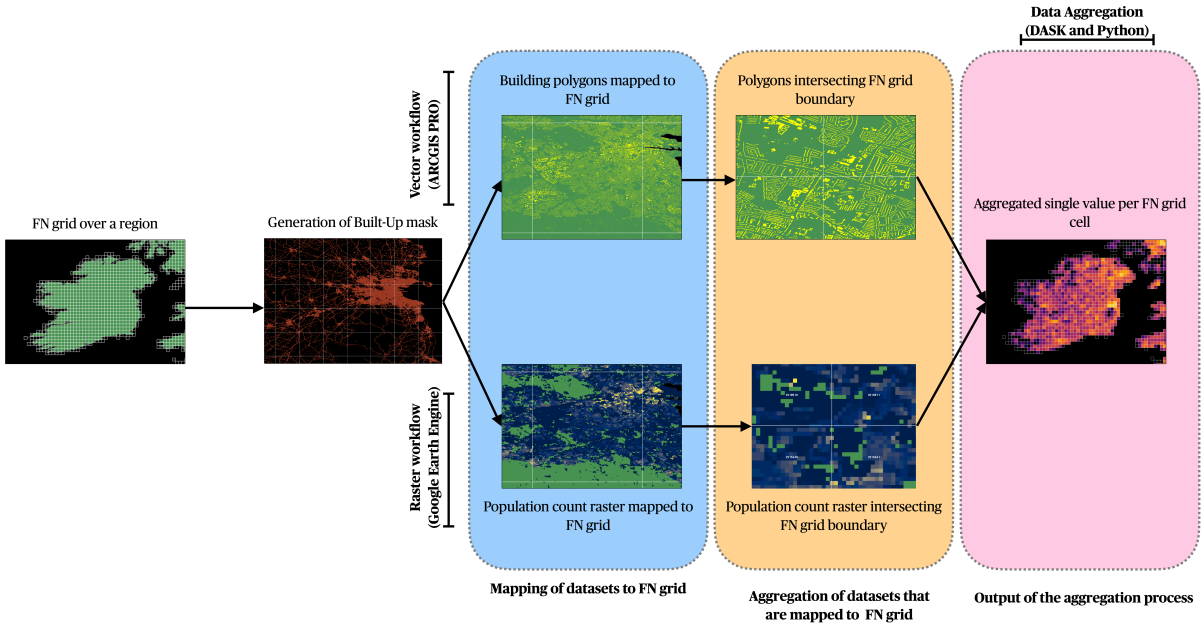

Figure 3 | Process flow of data aggregation for FN grid
The process starts with creation of a FN grid of 1/8 degree resolution over global land mass. Next, BU20 layer is used as a masking layer to delineate areas where built-up structures are present in year 2020. The masking layer along with the FN grid are then used to map vector and raster datasets to FN grid that underlie the masking layer. Finally, the vector and raster dataset values are aggregated for each fishnet to generate single value per FN grid cell. Here the vector datasets intersecting FN boundary are split at the boundary and aggregated to the respective FN grid cells while the raster datasets are aggregated by the means of weighted sum.

The base year *raster* datasets representing non-masked PPLN20 and BU20 datasets were processed on the Google Earth Engine[26] platform. Both the datasets were clipped at the boundary of the overlapping FN and the pixels completely inside the FN were aggregated as is, with pixels falling on the boundary being aggregated using weighted summation where the value attribution of the pixel in consideration is calculated based on the area of the pixel inside the FN. It should be noted that while PPLN20 dataset represents simple population count at 100m resolution, the BU20 layer pixel

represents the percentage of built-up area inside each 100m pixel. Hence, the aggregation of BU20 pixel was done by multiplying the pixel area by pixel value to represent true built-up area represented by each 100m resolution pixel.

The SSP derived population $P_{X,Y}$ and $B_{X,Y}$ for Y equal to 2020 were spatially harmonised to FN grid by mapping the values from spatially harmonised PPLN20 and BU20 datasets derived in the previous steps. This aids in first providing a common base year value for estimation of future aggregated rooftop areas per FN grid cell and second removes any mismatch of datapoints and data values between the base datasets and SSP derived datasets. The mismatch between the datapoints occur due to $P_{X,2020}$ and $B_{X,2020}$ using exogenous methodologies and frameworks to estimate the values in their respective datasets. As an example the $B_{X,2020}$ dataset points depicting presence of built-up area is derived from a model that uses GHSL[27] layer from JRC for the year 2015 thereby not incorporating some newly developed areas in east China, Figure 4. Additionally, mismatch between data values can occur when for a FN grid cell $B_{X,2020}$ layer either under or over represents the value depicted by BU20 dataset. As a result of these mismatches, for a BU20 layer's global aggregated built-up area of 1.46 million km^2, $B_{X,2020}$ layer only represents 0.98 million km^2 of global aggregated built-up area. This highlights the importance of harmonising the datasets both at a common temporal and spatial slice.

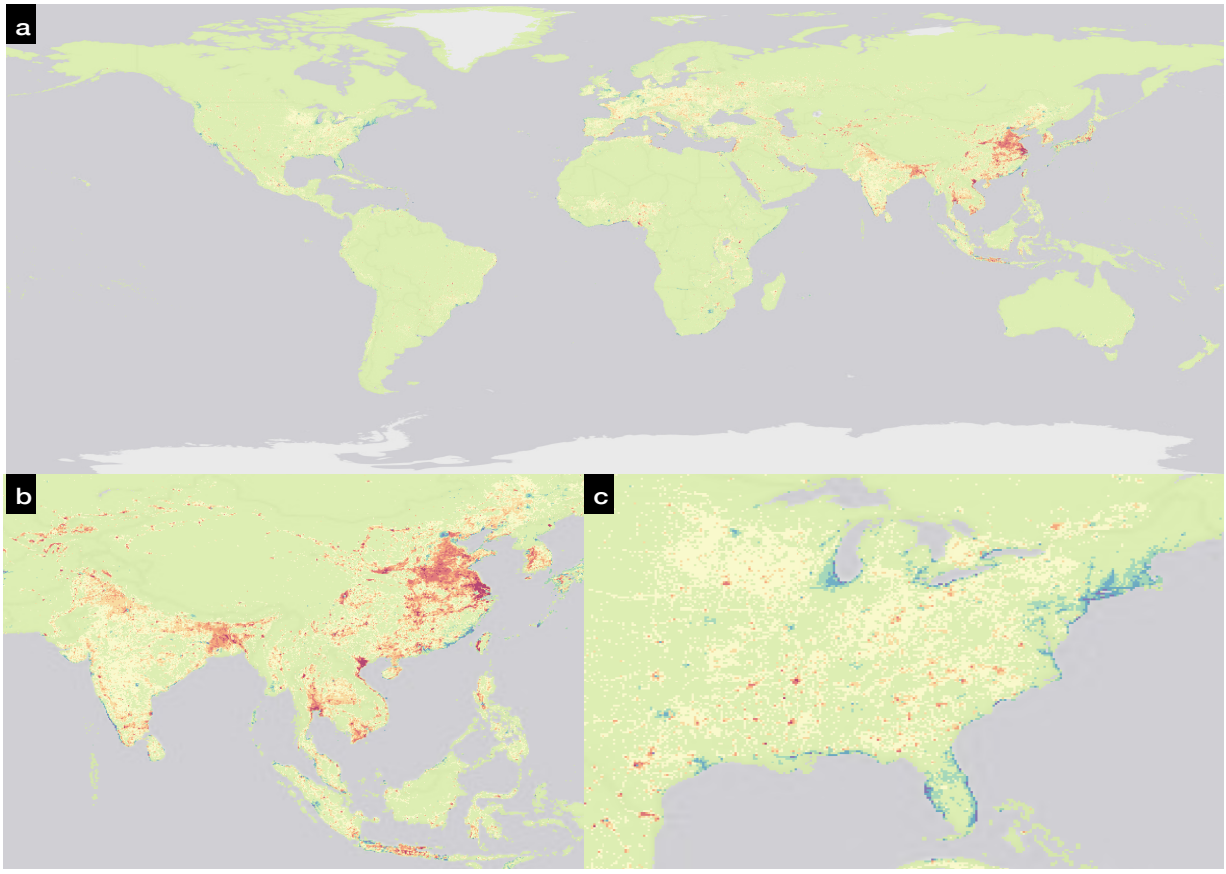

Figure 4 | Discrepancies between BU20 layer and SSP derived $BU_{X,2020}$ layer
a, Global FN grid cell depicting the discrepancies between BU20 and $BU_{X,2020}$ layer. Red and orange coloured region FN grid cells have BU20 values more than $BU_{X,2020}$ dataset values while blue FN grid cells have $BU_{X,2020}$ values more than BU20 values. In general, the blue coloured FN grid cells signify overrepresentation of built-up area in $BU_{X,2020}$ layer and red coloured regions signify underrepresentation of built-up area. **b**, zoomed in region of Asia where red coloured FN grid cell are observed in East China with blue coloured grid cells being observed in coastal regions. c, zoomed in region of east coast of USA where blue colour FN grid cells are observed in coastal regions.

After harmonising the $P_{X,2020}$ and $B_{X,2020}$ datasets for each of the SSP scenarios, the future datapoint and data values per FN grid cell of the respective datasets were recalculated using the following

$$P_{X,Y} = (P_{X,Y} - P_{X,2020}) + PPLN20 \quad (1)$$

$$B_{X,Y} = (B_{X,Y} - B_{X,2020}) + BU20 \quad (2)$$

where, for each unique FN grid cell, X is the SSP scenario, Y is the year for which datapoint and value is calculated, PPLN20 is the base year population count and BU20 is the base year Built-up area. This effectively captures the absolute growth in the metrices per FN grid cell over the harmonized base datasets. For GDP value per FN grid cell, we devised population weighted down mapping of country level GDP value using the following

$$\mathrm{GDP}_{X,Y} = \frac{\mathrm{GDP}_{C,X,Y}}{P_{C,X,Y}} * P_{X,Y} \quad (3)$$

where, for each unique FN grid cell, X is the SSP scenario, Y is the year for which datapoint and value is calculated, C is the country for which aggregated metrics are calculated at country level. This GDP downscaling methodology although being imperfect, adds in a new feature layer representing GDP weighted population count per FN grid cell for training our ML model discussed in next section. Finally, we create the population density layers for both base year datasets and SSP derived datasets using the following

$$\mathrm{PD20} = \frac{\mathrm{PPLN20}}{FN_{Area}} \quad (4)$$

$$\mathrm{PPLND}_{X,Y} = \frac{\mathrm{P}_{X,Y}}{FN_{Area}} \quad (5)$$

where, for each unique FN grid cell, X is the SSP scenario, Y is the year for which the datapoint and data value is calculated and FN_{Area} is the geodesic area occupied by the FN grid cell.

2.3 Machine Learning model

We designed a ML based framework based on XGBoost[28] ML model to estimate aggregated rooftop area per FN grid cell. The ML framework accomplishes the task of *first* extracting FN grid cell from BF20_OSM layer derived from OSM global building footprint dataset that have complete building footprint polygon mapping and *second* to estimate the aggregated rooftop area per sample FN grid cells. The flow of data and steps involved in development of ML framework is shown in Figure 5.

We start the development of the ML framework by extracting sample FN grid cells from the base year datasets. The FN grid cells that have complete coverage for PD20, BU20, RL20 and BF20 datasets are selected as sample FN grid cells and the extracted sample layers are named here as PD_S20, BU_S20, RL_S20 and BF_S20 respectively. The PD_S20, BU_S20, RL_S20 sample FN grid cells then act as independent variables with BF_S20 acting as dependent variable for the M1 model. The M1 model is then trained by using 10-fold cross validation strategy and 1000 hyper tuning iterations. The 10-fold cross validation strategy enables the use of complete input dataset for training purposes and

aids in reducing the problem of overfitting in conjunction with 1000 rounds of hyper tuning iterations. The trained M1 model then accepts PD20, BU20, RL20 layers as drivers to estimate the aggregated gross rooftop area for all the global FN grid cells, $BF_{FN}20$ layer.

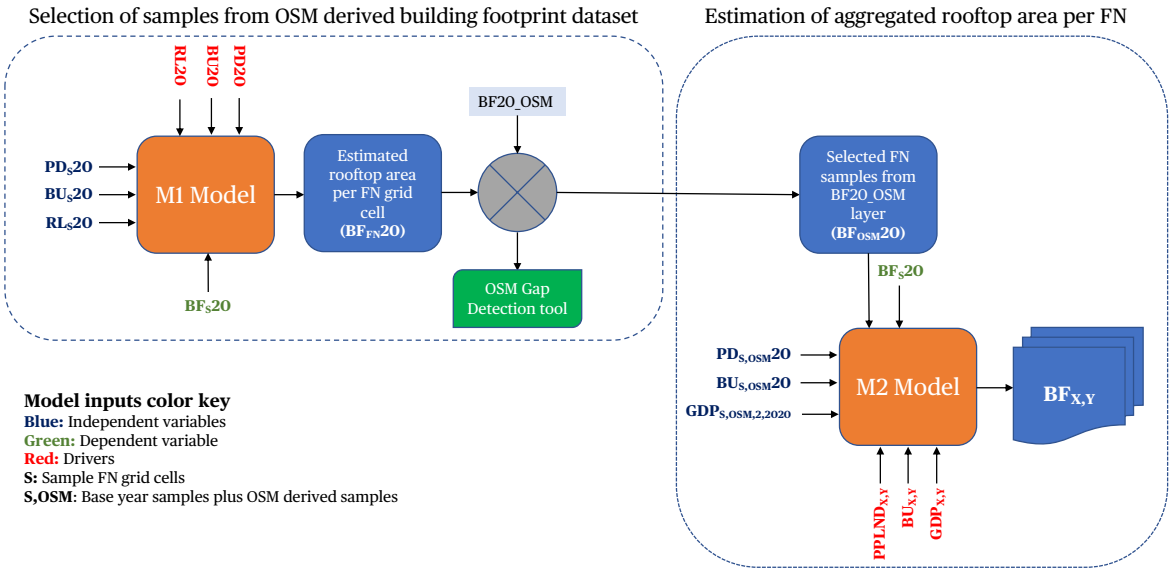

Figure 5 | Overview of ML framework

The ML framework is divided into two stacked XGBoost models. The first model "M1" aids in selection of samples from global OSM building footprint dataset and outputs result for OSM Gap Detection tool. The second model "M2" combines the samples from first model with BF_S20 samples and uses the SSP derived drivers to estimate the aggregated rooftop area per FN grid cell.

At this stage, we have a global estimate of rooftop area for the year 2020 which we then use to extract samples from BF20_OSM layer. For this, we compare at FN level the values of $BF_{FN}20$ and BF20_OSM layer. For the FN grid cells where the ratio between BF20_OSM and $BF_{FN}20$ is in between 1.1 and 0.9 i.e., where BF20_OSM values show 90- 110% of $BF_{FN}20$ values, those FN grid cells are selected for their completeness of building footprint mapping and extracted as $BF_{OSM}20$ sample layer. This comparison between M1 model predicted values and OSM derived values also lends itself to development of a OSM Gap detection tool which we discuss in further in next section.

After tuning, training, and inferencing of $BF_{OSM}20$ layer from M1 model, we shift our focus to the M2 Model which will enable the estimation of global gross aggregated rooftop area per FN grid cell for SSP narratives. For this, we combine the BF_S20 samples from base year dataset with $BF_{OSM}20$ samples. We also resample PD20, BU20 and $GDP_{X,Y}$ layers to collect samples based on FN grid cells covering our combined building footprint samples to generate $PD_{S,OSM}20$, $BU_{S,OSM}20$ and $GDP_{S,OSM,2,2020}$ layers. The $GDP_{S,OSM,2,2020}$ layer here represent population based downscaled GDP per sample FN grid cell for samples covering base year and OSM derived Building footprint FN grid cells for SSP2 narrative and 2020 year. The $PD_{S,OSM}20$, $BU_{S,OSM}20$, $GDP_{S,OSM,2,2020}$ sample FN grid cells then act as independent variables with BF_S20 and $BF_{OSM}20$ acting as dependent variable for the M2 model. The final sample FN grid cells used in our study is shown in Figure 6.

Figure 6 | Global distribution of sample FN grid cells
For FN grid cells covering USA, Canada, Africa, UK, and Australia BF20 layer was used. For the rest of the world, OSM derived FN grid cell was used after selecting them from inferencing of M1 model.

The M2 model is trained by using 10-fold cross validation strategy and 1000 hyper tuning iterations. At the conclusion of this step, we have our final M2 model which then accepts $PPLND_{X,Y}$, $BU_{X,Y}$ and $GDP_{X,Y}$ layers as drivers to estimate a global $BF_{X,Y}$ layer for five SSP narratives and years ranging from 2020-2050. The final $BF_{X,Y}$ layer is stored as individual geopackage files having 1/8 degree FN grid cell resolution with value representing the aggregated gross rooftop area inside the FN grid cell for further analysis, Figure 7.

Although, the trained M1 model in conjunction with SSP derived drivers can aid in generation of final $BF_{X,Y}$ layer, but we could did not implement this as RL20 layer data is only available for base year of 2020 and multivariate regression would be required to estimate its value beyond 2020 which would add an extra layer of uncertainty in our results. Additionally, selection of $BU_{S,OSM}20$ and merger of this layer with BF20 layer provided us with additional global data points to retrain a new model M2 which would be more compliant with global trends rather than just the countries/region covered by BF20 dataset. Extended details of the attributes of the input datasets are documented in Table 3.

Table 3 | Data attributes for input building footprint datasets for M2 Model

Sample Areas	Input Rooftop Area (km^2)	# Individual Polygons	FN Grid Cell Covered
Australia	2,418	~10 million	
UK	3,450	~33 million	
USA	29,930	~ 144 million	~100,000
Canada	2,500	~19 million	
AFRICA	17,166	~300 million	
$BF_{OSM}20$	21,000	~140 million	~2,500

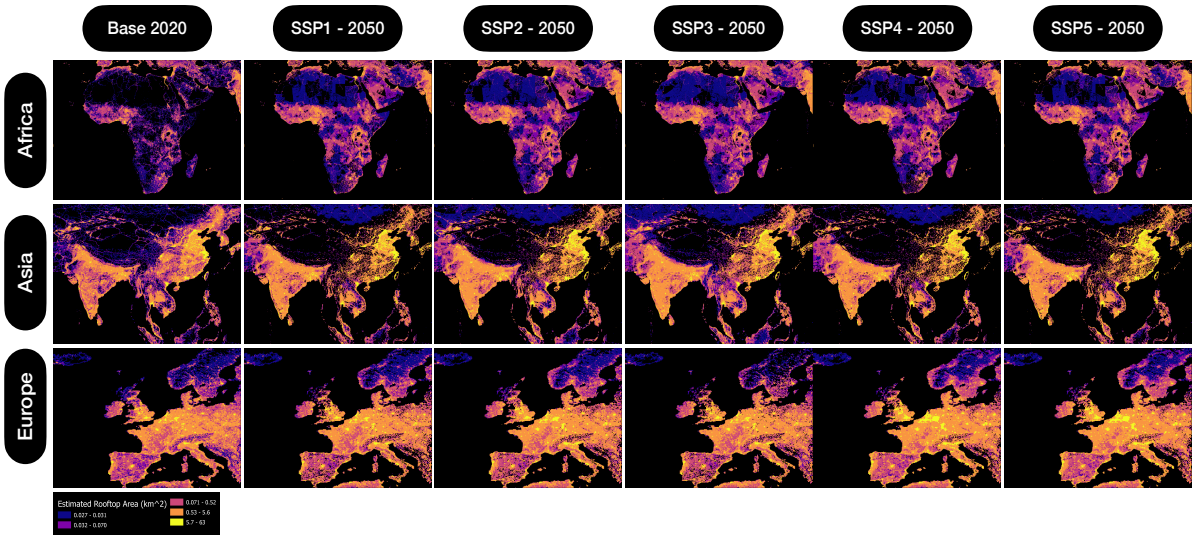

Figure 7 | Visual depiction of $BF_{X,Y}$ layer for selected global regions.
The image panels depict the pixelwise output of $BF_{X,Y}$ layer classified by a graduated colour ramp. Each pixel in the panel represents the aggregated gross rooftop area per FN grid cell. Growth in rooftop area per FN grid cell can be observed for East China, West Africa, and Central European areas.

2.4 Open Street Map gap detection tool

Open Street Map derived data is being used in many studies as a source of ground truth mapping and for calibration of big data models. Additionally raw OSM data in the form of building polygons, road mapping is being used extensively in resource accessibility studies and vulnerability mapping[29]. Primary reason for uptake of OSM data can be attributed to its free accessibility and presence of more than a million active users who are updating the digital planet files on an hourly basis. Although the quantity of data that is present inside the OSM database is vast, studies using them often must do significant pre-processing to extract data that is suitable for their use case. Additionally, users of the OSM dataset struggle with lack of validation studies done on OSM datasets.

For data attributes dealing with global roads, one study[30] highlights that's the OSM global road dataset is 80% complete. Similar studies for global building footprint datasets are currently limited to either country level studies[31] or regional studies (https://github.com/hotosm/osm-analytics). As a by-product of the output of our M1 model, we have created a tool that can overlay our predicted rooftop area mapped to the FN grid on top of the raw building polygon dataset and hence estimate the completeness of the base OSM dataset. We have created a layer that is derived from M1 models output that represents percentage completeness of the OSM building polygon dataset by providing values between 0-6. The base dataset for OSM comparison was procured in August 2021.

A value of 0 represents that either OSM data is missing, or data cannot exist at that FN grid cell. A value of 1 represents that OSM dataset coverage is 100% in that FN grid cell. Any value between 0.9-1.5 can be considered as representing 100% completeness of the OSM dataset as our M1 model does have under or over prediction characteristics in some regions based on driver metrics. A value greater than 1.5 should be representative of regions in OSM that may not have population presence but have OSM building polygon tags e.g., greenhouses, industrial complexes around major shipping ports etc. Since our M1 model relies on population as an important metric, in FN grid cells having completeness value greater than 1.5, our model gives lower value than OSM dataset value. Another reason for this can be attributed to wrong tag being assigned to building polygons or misclassification of non-building built-up structure as building polygon inside OSM dataset. An example of

completeness value dataset is shown for Europea in Figure 8a, with example cases of completeness value greater than 1 shown in Figure 8b, Figure 8c and Figure 8d.

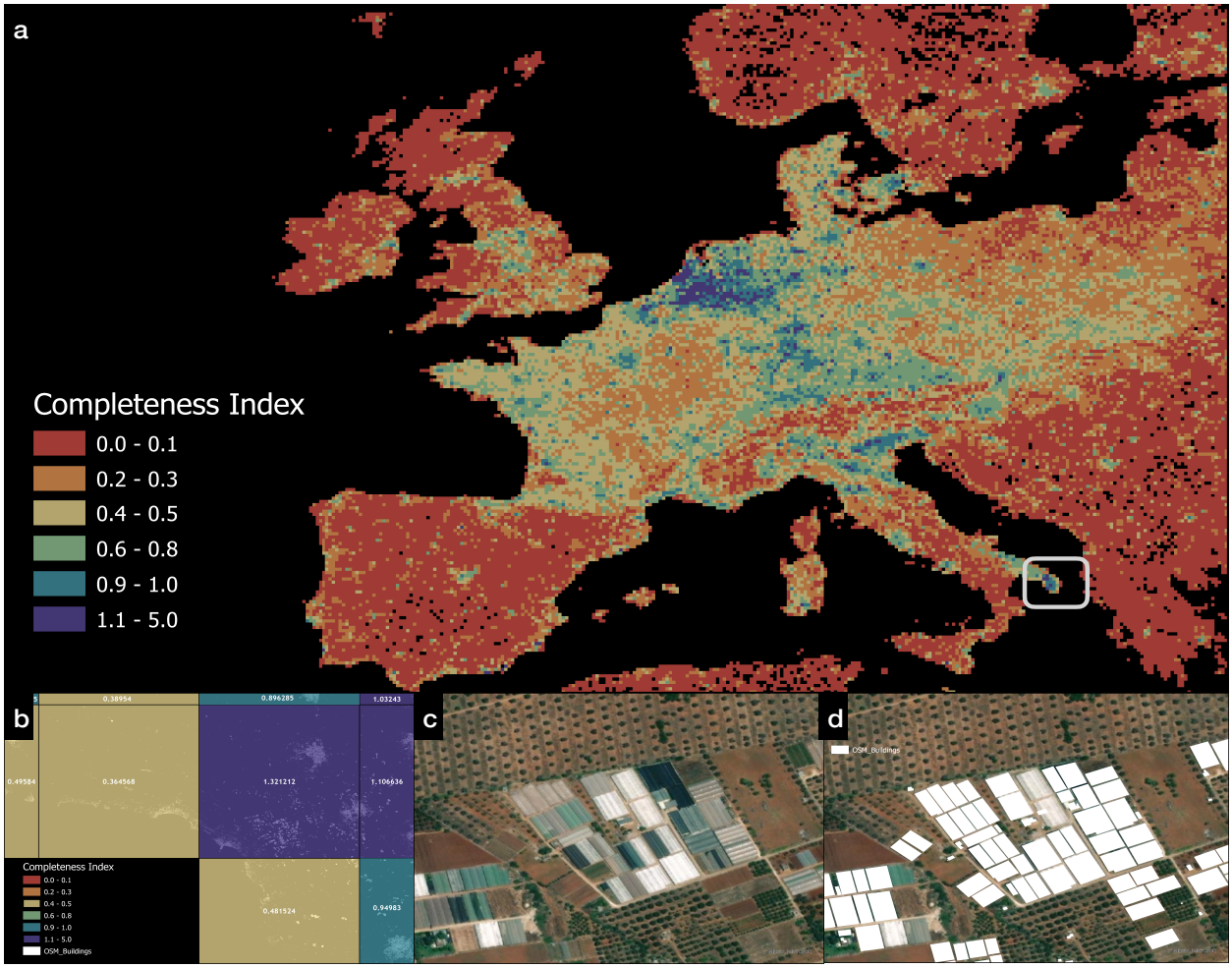

Figure 8 | Output visualisation of OSM gap detection tool.
Output of the gap detection tool dataset with each individual FN grid cells classified by the completeness index, **a**. **b**, zoomed in view of the ROI bound by white bounding box in figure 8a. The values inside the FN grid cell represent the completeness index value for that FN grid cell. **c**, image displaying a sample of area marked inaccurately in OSM dataset inside the FN grid cell with completeness index of 1.32. **d**, overlay of OSM polygon on incorrectly identified buildings in figure 8c where greenhouse installations have been marked as buildings leading to the FN representing value greater than 1.

2.5 Technical potential assessment

For converting the estimated rooftop area to electricity generation potential per month, we utilised the PVOUT factor from World Bank's PV potential dataset[32] for locations between 60^0N and 45^0S, covering over 99% of the world's population. This dataset is provided as a griddled raster dataset representing kWh produced by each kWp installed capacity per month. The locations outside the latitudes were assigned a constant PVOUT value of 3.5 kWh/kWp/day (peak).

We first start by calculating the kWh/day of solar PV derived electricity from a 1kWp panel for each of the global FN grid cell by taking the mean of the raster cell values from PVOUT dataset. The PVOUT dataset is presented at a 1km raster cell resolution. Next, we assume that for each FN grid cell the estimated rooftop area is the total available rooftop area for rooftop solar panel installation.

Additionally, we assume that the rooftops are flat, the orientation is east-west, slope of the building is non-existent, the panels fully cover the available rooftop area and are placed at the optimal angle for the given latitude. These assumptions represent the ideal case for installation of rooftop solar PV technology and represent the highest possible potential that can be generated from a given rooftop area. We have included details in Usage Notes section for tailoring the datasets generated using these assumptions for different use cases.

Next, we assume that all the installed panels have a 10% efficiency and require 10m^2 area for 1kWp installation capacity. This assumption aids in direct conversion of estimated rooftop area into installed capacity. The installed capacity is then converted into monthly potentials using the FN aggregated PVOUT dataset. Additionally, the PVOUT layer was converted into annual capacity factors for the FN grid cell they lie in. Conversion from estimated rooftop area to technical potential is achieved using following equations

$$\mathrm{IC}_{X,Y} = \frac{BF_{X,Y}}{10} \tag{6}$$

$$\mathrm{POT}_{X,Y,M} = \mathrm{IC}_{X,Y} * PVOUT_{FN,M} * D_M \tag{7}$$

Where,
$\mathrm{IC}_{X,Y}$ is the installed capacity in kWp per FN for SSP narrative X and year Y
$\mathrm{BF}_{X,Y}$ is the aggregated gross rooftop area per FN for SSP narrative X and year Y
$\mathrm{POT}_{X,Y,M}$ is the monthly potential in kWh for M month per FN for SSP narrative X and year Y
$\mathrm{PVOUT}_{FN,M}$ is the aggregated PVOUT value for M month per FN for SSP narrative X and year Y
D_M is the days in the respective month M

3. Data records

The datasets generated in this study can be categorised into

1) OSM gap detection dataset.
2) High spatial resolution global estimated rooftop area datasets for SSP narratives for years 2020-2050.
3) Monthly temporal resolution global RTSPV technical potential assessment datasets for SSP narratives for years 2020-2050.

The details of the datasets are documented in Table 4. The OSM gap detection tool dataset is provided as a geopackage file. In this file, individual FN grid cells are represented by a polygon enclosed by the FN boundary with percentage completeness values being stored for each polygon. The dataset is provided in an EPSG:4326 format and is valid for base OSM building dataset accessed on 1st August 2021. We plan to update this dataset on a biannual basis as the processing time for the entire globe is approximately 4 days on a server cluster comprising of 24 cores.

The global estimated rooftop area per FN grid cell dataset for each SSP narratives is provided as a csv file with one column containing FN_ID representing the FN grid cell ID, one column containing the respective metrics and finally one column documenting average yearly capacity factors for the FN grid cell. Similar structure is used for the monthly global technical potential dataset for SSP narratives. To enable conversion of csv files to a format compatible with GIS environment, we have provided a FN_Poly dataset as geopackage file along with a csv file with WKT formatted FN polygon geometries. The FN_Poly file in conjunction with other datasets can be used to convert csv-based datasets into GIS compatible layers for spatial analysis. For this we have provided an ipython notebook file for python-based dataset conversions.

Table 4 | Dataset Attributes for output datasets

Name	Usage	Type	Platform	# FN polygons
OSM_GDT	OSM Gap Detection Tool	Geopackage	Github	3.5 mil
SSP_X_RA_Y	Global rooftop area for SSP narrative X and year Y	CSV	figshare	3.5 mil
SSP_X_POT_Y_M	Global technical potential for SSP narrative X, year Y and month M	CSV	figshare	3.5 mil
FN_Poly	Fishnet grid cell polygon boundaries	Geopackage and CSV with WKT geometry	figshare	3.5 mil
Python script	Combining Datasets with FN_poly to visualise in GIS environment	.ipynb file	figshare	3.5 mil

4. Technical validation

The datasets presented in this study have undergone end to end technical validation for the base year of 2020. The validation is performed for M1 and M2 model inputs, performance of M1 and M2 models, estimation validity of outputs of M1 and M2 models and finally verification of estimations generated by the M1 model. For datasets covering years 2030-2050, we unfortunately could not provide a true verification of validity of the results as the drivers used to estimate these values represent values in future which are difficult put an accuracy value on.

The input validation of the base year datasets and SSP derived drivers are presented in Table 5 as link to the validation reports generated by either the data providers or the peer reviewed publication which form the basis of the data. Due to the scale of the dataset, assumptions and the limitation of methods used, the big datasets used in this study are expected to have errors at a higher resolution when verifying at a per building level, but at an aggregated country/ regional spatial resolution these datasets have shown acceptable performance.

Table 5 | Validation studies and articles for input datasets

Dataset	Format	Validation study link
Building Footprints	Vector	Heris, M.P., Foks, N.L., Bagstad, K.J. *et al.* A rasterized building footprint dataset for the United States. *Sci Data* **7,** 207 (2020). **https://doi.org/10.1038/s41597-020-0542-3**
Population	Raster	Lloyd, C., Sorichetta, A. & Tatem, A. High resolution global gridded data for use in population studies. *Sci Data* **4,** 170001 (2017). **https://doi.org/10.1038/sdata.2017.1**
Road	Vector	Barrington-Leigh, C., & Millard-Ball, A. (2017). The world's user-generated road map is more than 80% complete. *PloS one*, *12*(8), e0180698. **https://doi.org/10.1371/journal.pone.0180698**
PVOUT	Raster	ESMAP. 2019. Validation Report: Global Solar Atlas 2.0 Validation Report. Washington, DC: World Bank.
Built-up area 2020	Raster	Tsendbazar, N.E., Tarko, A., Linlin, L., Herold, M., Lesiv, M., Fritz, S., Maus. V; (2020): Copernicus Global Land Service: Land Cover 100m: Version 3 Globe 2015-2019: Validation Report; Zenodo, Geneve, Switzerland, September 2020; **doi: 10.5281/zenodo.3938974**
SSP derived Built-up area 2020-2050	Raster	Gao, J., O'Neill, B.C. Mapping global urban land for the 21st century with data-driven simulations and Shared Socioeconomic Pathways. *Nat Commun* **11,** 2302 (2020). **https://doi.org/10.1038/s41467-020-15788-7**
SSP derived Population 2020-2050	Raster	Jones, B. & O'Neill, B. C. Spatially explicit global population scenarios consistent with the Shared Socioeconomic Pathways. Environ. Res. Lett. 11, 084003 (2016). **https://doi.org/10.1088/1748-9326/11/8/084003**
SSP derived GDP 2020-2050	Dataset	Crespo Cuaresma, J. (2017). Income projections for climate change research: A framework based on human capital dynamics. Global Environmental Change, 42, 226–236. **https://doi.org/10.1016/j.gloenvcha.2015.02.012**

Learning accuracy of the M1 and M2 models are determined by the significance of the correlation between dependent and independent variables used to train the model. Further, a 10-fold cross validation strategy to expose the models to various combinations of input data to reduce model overfitting was used. Finally, the distribution of model output with respect to the dependent variables and spread of the errors were evaluated to choose the best model. Figure 9 documents the results of these checks.

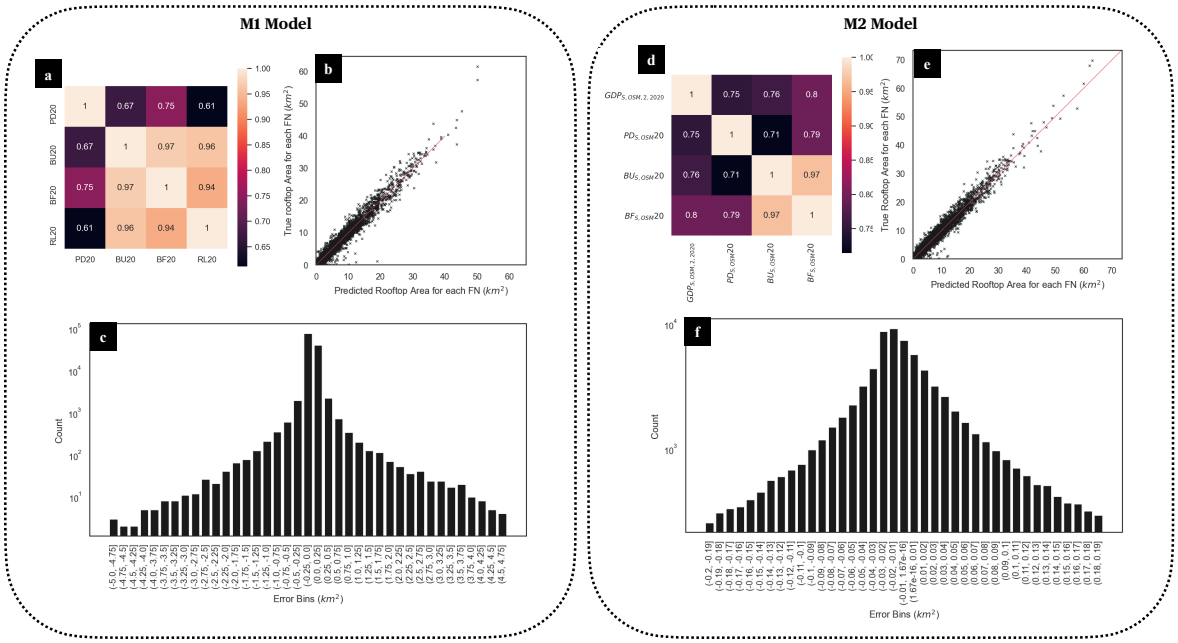

Figure 9 | Performance metrics of M1 and M2 models.
a, Correlation heatmap representing Pearson's correlation between pairs of independent and dependent variables of M1 model. High correlation can be observed for the depend variable (BF20) and independent variables (PD20, RL20, and BU20). **b**, graph representing relationship between M1 model's dependent variable and predicted values. High fidelity can be observed between dependent variable and predicted values. **c**, spread of absolute difference between dependent variable value and predicted value from M1 model at a per FN grid cell basis. Majority absolute error of ±0.25 km^2 can be observed for a 1/8 degree FN grid cell. **d**, Correlation heatmap representing Pearson's correlation between pairs of independent and dependent variables of M2 model. High correlation can be observed for the depend variable ($BF_{S,OSM}20$) and independent variables ($PD_{S,OSM}20$, $GDP_{S,OSM2,2020}20$, and $BU_{S,OSM}20$). **e**, graph representing relationship between M2 model's dependent variable and predicted values. High fidelity can be observed between dependent variable and predicted values. **f**, spread of absolute difference between dependent variable value and predicted value from M2 model at a per FN grid cell basis. Majority absolute error of ±0.05 km^2 can be observed for a 1/8 degree FN grid cell with slight left skewness in the error distribution leading to model prediction showing slight underestimation of ground truth at FN grid cell level.

The final output of the M2 model ($BF_{X,Y}$) was further evaluated for discrepancies between aggregated country-wise input base year big data derived BF20 values and aggregated country-wise M2 models estimated outputs for SSP2 narrative in year 2020 ($BF_{2,2020}$). These evaluates were conducted by aggregating the FN grid cell values for those FN grid cell that fall within the geographic boundaries of the country being evaluated. Overall, we observed high fidelity between the ground truth and estimated values at a country level. On a higher spatial resolution, we also compared the sub country level estimations for USA based on ASHRAE climatic regions. Here also high fidelity was observed between ground truth and predicted values. This way, we could validate our results to a high degree of certainty by comparing results at sub country level and country level.

Table 6 documents the results for sub country level comparison and Table 7 documents the comparison at country level.

Table 6 | Result comparison of M2 Model's output at sub country level for USA

Climatic Zone	Ground Truth (km^2)	M2 Model's output (km^2)	Absolute percentage error
3	734.2	632.5	13.8
4	76.2	108.4	42.3
6	4,102.5	3,788.5	7.7
7	700.5	706.3	0.8
9	5,530.6	5,532.2	0.0
10	1,028.1	1,073.5	4.4
11	1,426.3	1,267.3	11.1
12	5,666.1	5,818.6	2.7
13	215.9	218.4	1.1
14	950.7	832.1	12.5
15	6,632.4	6,669.0	0.6
16	68.3	77.3	13.2
17	410.8	406.8	1.0
18	1,892.0	1,900.9	0.5
19	24.9	25.9	4.2
20	12.1	16.4	35.2
21	199.6	227.1	13.8
22	1.6	1.5	9.5
24	7.9	144.5	1,736.0
Total	**29,680.5**	**29,447.1**	**0.8**

Table 7 | Result comparison of M2 Model's output at country level

Country	Ground Truth (BF20) km^2	M2 Model's Predicted value (km^2)	Note
Australia	2,418	2,527	Seen data by model
UK	3,450	3,492	Seen data by model
USA	29,930	30,025	Seen data by model
Canada	2,500	2,753	Seen data by model
Ireland	456	424	Unseen data by model

5. Usage notes

The aggregated rooftop area datasets are generated with an assumption of full one to one mapping between building footprint and rooftop area. Although it is common to have larger rooftop area than building footprint due to presence of rooftop superstructures, we have not considered this due to the scale of the analysis which looks at global region of interest rather than per building. Similarly in higher latitudes due to slope of the rooftops, total building rooftop area will be higher than the building footprint area. Hence, it is advised to use region specific conversion values when using these datasets for city level analysis. Additionally due to the nature of the ML model used for estimation of rooftop area, we recommend an error margin of ±0.1 km^2 per FN grid cell.

For the technical potential assessment, we assumed that rooftops are flat with solar panels being placed at the latitude specific optimal angle. We also assumed that the entire estimated rooftop area will be fully covered by solar panels and the panels will be devoid of shadows. This assumption culminates as our dataset representing the best-case scenario for technical potential generation. In wider literature, a rooftop availability factor of 0.3 is used to convert gross rooftop area to net rooftop area to account for unsuitable rooftops due to orientation and slope attributes of building stocks. For the users of this dataset, we recommend using region specific rooftop availability factors if known else use 0.3 as the factor for more practical results.

Our technical potential dataset is generated with an assumption of 10% panel efficiency or 10m^2 or rooftop area per 1kWp panel requirement. To re-estimate the potential for higher panel efficiency, we recommend dividing the potential dataset by 10 and multiplying with area required for the new panels of 1kWp capacity. For example, a 20% efficient panel will require 5m^2 area for a 1kWp panel thereby doubling the potentials represented in our datasets. This flexibility of after the fact analysis aids in generating datasets that meets user specifications without redoing the entire analysis from scratch.

6. Implementation of data in the MESSAGE$_{ix}$ model

Global energy system (GEM) models and Integrated Assessment models are used to inform on the policy and scenarios that are at the intersection of climate change mitigation and energy access research focus. Due to lack of global datasets and studies analysing the RTSPV technology as a separate technology from Utility scale counter parts, the GEMs and IAMs sometimes have limited insights for RTSPV focussed studies[33].

MESSAGE$_{ix}$ modelling platform[34] is a current state of art global energy system model developed by IIASA. Over the years, the solar PV technology has been represented as a single technology that represent both Utility scale and RTSPV scale technology. Naturally, updating the global energy system model of MESSAGEix-GLOBIOM[1] will be the best testing grounds for the updated RTSPV potential datasets generated in the first work package of this study. For this, we first start by calibrating the global model with the new datasets and with updated costs for RTSPV and UTSPV. Then we recalculate the UTSPV potentials based on the FN grid used in first work package. Then we undertake scenario analysis to analyse the effects of updating the potentials and costs. Finally, we discuss the scenario results and demonstrate the measurable impact of conducting the assessment at a new high spatiotemporal resolution.

The Global MESSAGEix-GLOBIOM model comprises of 12 regions covering 195 countries. The regions are created by grouping countries that are linked by common trade mechanisms and geographic affinity. The global model can project the future energy systems by optimising the supply and demand streams within the energy system along with constraints on emission and land use change. The model is technology rich and represents renewable technologies in the form of step curve of generation potential based on capacity factor bins. Additionally, investment costs, operational costs and historical technology capacity can be implemented for each capacity factor bin. We implemented the task of calibration of base global MESSAGEix-GLOBIOM model by updating 1) technology potentials and capacity factor bins, 2) investment cost of the technology 3) historical installed capacity. These tasks were performed for both RTSPV technology and UTSPV for all the 12 world regions.

The first step in model calibration is to reassess the UTSPV potentials for all the countries covered by 12 world regions. For this, we utilised a global landcover mask (L2 mask) from World Bank's solar potential by country report[35]. The L2 mask combines all the global land area which is suitable for installing UTSPV. This mask accounts for suitable distance from urban areas, terrain topology, croplands, and protected areas where UTSPV deployment is not preferred either due to international conservation treaties or due to unsuitable land use. Next, we mapped the L2 mask to our FN grid cells that we used for RTSPV assessment. This mapping provides the advantage of having both UTSPV and RTSPV technology assessments being done on a common spatial resolution. Further we assumed that the entire land area covered by L2 mask inside the FN grid cell has the potential to house solar panel arrays, the panels will have 10% efficiency or 10m^2 of area will be required to install a 1kWp panel and the panels are placed at an optimal angle suitable for the latitude where FN grid cell is present. Next, we multiplied the installed capacity per FN grid cell with the $POT_{X,Y,M}$ layer to generate monthly UTSPV technical potential layers for five SSP narratives. Here it is important to

[1] https://docs.messageix.org/projects/global/en/latest/

highlight that unlike RTSPV potentials which change year on year based on drivers, UTSPV have been kept constant across the years.

The second step in model calibration is to combine the monthly assessed potentials for RTSPV and UTSPV into a yearly value. This is done to align the potentials with the annual time resolution of current global MESSAGEix-GLOBIOM model. The yearly aggregated potential values were then further grouped by yearly average capacity factor bins for 12 world regions. This step enables the representation of two separate technologies i.e., RTSPV and UTSPV in a structure that is complaint with the current global MESSAGEix-GLOBIOM model.

The capital cost and operational expenditure attribution of installing the RTSPV and UTSPV was derived from the renewable cost of generation 2020 report of IRENA[36]. For each of the 12 regions inside the MESSAGEix-GLOBIOM Model, we chose representative country(s) from IRENA report and took average of the values. Similar exercise was conducted for historical installed RTSPV capacity per region using the historical net capacity addition data from IEA's renewables 2020 report[37]. The updated costs and potentials for 12 world regions in global MESSAGEix-GLOBIOM model are documented in Table 8. The updated potentials, costs and historical capacity were then added to the base MESSAGEix-GLOBIOM model using custom python scripts that divide potentials into capacity factor bins and ingests the data into the base global model.

Table 8 | Cost and potential of RTSPV and UTSPV, calibrated for 12 Region Global MESSAGEix-GLOBIOM Model

MESSAGEix-GLOBIOM Region	Current cost ($2005/KW)[a]	New Cost RSTPV ($2005/ KW)[b]	New Cost UTPV ($2005/ KW)[b]	Current Potential (GWa)[c]	New RTSPV Potential (GWa)[d]	New UTSPV Potential (GWa)[d]
AFR	1,238	1,166	866	28,780	469	312,691
CPA	671	909	770	22,672	105	25,921
CHN	671	563	491	22,672	1,134	76,597
EEU	839	1,446	601	35	172	8,174
FSU	983	1,104	827	7,550	378	90,990
LAM	1,178	741	857	21,399	556	219,156
MEA	1,025	1,104	770	52,620	377	140,746
NAM	1,189	2,927	897	21,384	869	123,404
PAO	1,187	1,287	1,091	23,766	251	37,844
PAS	910	1,104	827	1,969	439	25,000
SAS	606	530	470	7,136	716	45,331
WEU	839	1,446	601	4,546	685	28,731

[a] current 2020 cost deflated to year 2005 which is common for UTSPV and RTSPV
[b] new 2020 cost delated to 2005
[c] Combined potential for UTSPV and RTSPV
[d] Potential at 15% panel efficiency and 100% rooftop availability
where UTSPV is Utility Scale Solar PV, and RTSPV is Rooftop Solar PV

7. Results

7.1 Spatiotemporal assessment

The global aggregated rooftop area for year 2020 is assessed to be 0.25 million km^2 for a global aggregated Built-up area of 1.46 million km^2. Amongst the 0.25 million km^2 of assessed global rooftop area, 0.12 million km^2 is in Asia and 0.047 million km^2 is in Europe. African continent occupies 0.02 million km^2 of aggregated rooftop area in 2020 with North America occupying 0.039 million km^2 of aggregated rooftop area.

Under SSP1 narrative, the global aggregated rooftop area increases to 0.3 million km^2 in 2050 with Asia representing the highest aggregated rooftop area of 0.15 million km^2. Under SSP2 narrative, the global aggregated rooftop area increases to 0.34 million km^2 in 2050 which is an increase of 0.04 million km^2 over SSP1. In SSP2 Asia represents 0.17 million km^2 of aggregated roof area. For SSP3 narrative the global aggregated rooftop area reduces to 0.31 million km^2 in 2050, registering a reduction in global aggregated rooftop area by 0.03 million km^2 over SSP2. This can be attributed to the trends in $BU_{X,Y}$ layers. For SSP4 the global aggregated rooftop area again increases to 0.33 million km^2 in 2050 representing an increase of 0.02 million km^2 of global aggregated rooftop area over SSP3. SSP5 narrative increases the gross aggregated global rooftop area to 0.38 million km^2 in 2050 which is an increase of 0.08 million km^2 over 2020 values. From this we can observe that from a base value of 0.3 million km^2, the range of global aggregated gross rooftop area in 2050 will lie in the range of 0.31-0.38 million km^2 for worlds envisioned under SSPs which represents a percentage increase of between 3-26% compared to 2020 values.

At a global scale the highest growth rate of aggregated gross rooftop is observed for African continent across all the SSP narratives with around 2 times increase in aggregated rooftop area between years 2020-2050. The highest growth in aggregated gross rooftop area across the continents is observed for SSP5 narrative, with lowest growth being observed for SSP3 narrative. For African continent and Asian continent SSP1 narrative registers the lowest growth in rooftop area between 2020-2050. For European continent and North American continent SSP3 narrative registers the lowest growth in rooftop area between 2020-2050. In general, between SSP narratives a growth of 7-200% can be observed for different continents highlighting the importance of policy formulation with consideration to variation in growth rates between regions. Table 9 documents the percentage of global gross rooftop area across SSPs for world continents. Table 10 documents the growth rate of the gross rooftop area across the SSPs for world continents.

The assessed technical potential of RTSPV at a global yearly aggregated level for the year 2020 is observed at 16.1 PWh yr^{-1} at 30% net rooftop availability and 10% panel efficiency. In line with the growth in gross rooftop area across the SSPs, the technical potential for RTSPV also scales up. The largest technical potential in the year 2050 is observed for SSP5 narrative with a value of 24.7 PWh yr^{-1} and the lowest technical potential in the year 2050 is observed for SSP1 narrative with a value of 19.8 PWh yr^{-1}. As a result, in the year 2050 the technical potential of global RTSPV is estimated to be between 20-25 PWh yr^{-1} representing an increase of between 25-50%, Figure 10.

Table 9 | Percentage distribution of global gross rooftop area across SSPs and continents

Continents	2020	SSP1_2050	SSP2_2050	SSP3_2050	SSP4_2050	SSP5_2050
Africa	8.76	12.30	12.34	12.56	12.12	11.83
Asia	50.39	49.54	49.60	51.09	48.99	47.89
Australia	1.00	0.98	0.92	0.85	0.95	1.06
Europe	18.90	17.16	17.13	16.29	17.58	18.06
North America	15.50	14.99	14.97	13.90	15.39	16.34
Oceania	0.20	0.20	0.19	0.18	0.20	0.20
South America	5.26	4.84	4.84	5.12	4.78	4.63

Table 10 | Absolute Growth of global gross rooftop area across SSPs and continents (2020-2050)

Continents	SSP1	SSP2	SSP3	SSP4	SSP5
Africa	1.71X	1.92X	1.79X	1.80X	2.06X
Asia	1.19X	1.34X	1.27X	1.27X	1.44X
Australia	1.19X	1.26X	1.07X	1.24X	1.60X
Europe	1.10X	1.23X	1.08X	1.21X	1.45X
North America	1.17X	1.31X	1.13X	1.29X	1.60X
Oceania	1.24X	1.31X	1.16X	1.28X	1.52X
South America	1.12X	1.25X	1.22X	1.18X	1.34X

Where X is the absolute increase in 2050 values over 2020 values

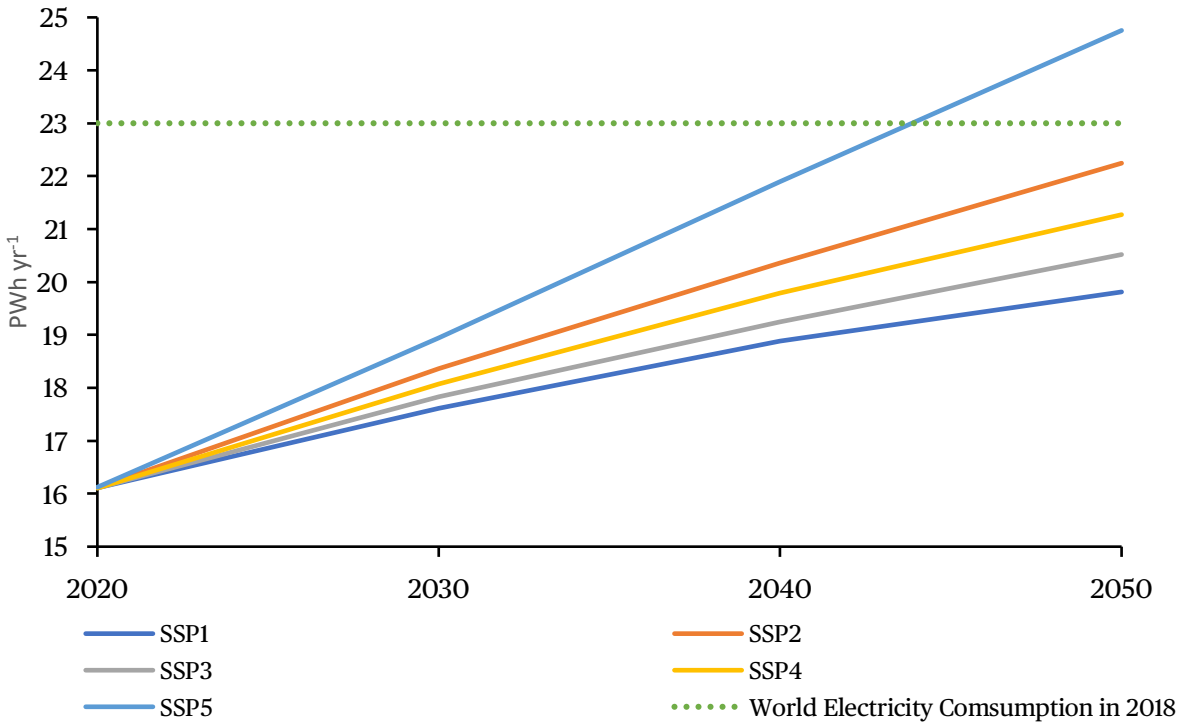

Figure 10 | Growth in yearly aggregated global RTSPV potential at 30% net rooftop area and 10% panel efficiency. The chart is displaying the growth in aggregated yearly technical potential for RTSPV across the SSP narratives. Using a conservative panel efficiency of 10% and 30% net rooftop availability, we can observe that under SSP5 scenario the assessed technical potential between years 2040-2050 will be sufficient to meet the yearly aggregated global electricity demand in 2018. Conversely, under the same assumptions 70% of the current global yearly aggregated electricity demand can be met with the current building stock.

7.2 Scenario runs

The spatiotemporal assessment of RTSPV technology at a global scale has demonstrated that the technology has the capacity to exceed the current global electricity demand. However, due to practical constraints like rooftop availability due to roof superstructures, chimney, orientation, and slope etc. the practical potential of the technology varies from region to region. The regions located near equator or in desert biomes have roof structures that can accommodate 100% solar panel deployment. Similarly, regions that occupy higher latitudes have slopes on their roofs to reduce snow accumulation and are sometime signify the cultural effects on the building archetypes. Here, 100% rooftop availability for installing solar panels is seldom possible. However, with drive towards net zero energy buildings and due to the need for retrofitting the buildings to meet the new energy efficiency standards, a considerable building stock in northern latitudes are transforming rooftops to accommodate larger number of solar panels.

The uptake of RTSPV technology in global energy systems is also affected by the interplay between costs of electricity generation by other energy technologies. Energy system models run based on optimising the supply and demand while keeping the total system costs at the lowest level and maintaining the emissions within the constraints described by the scenario. Additionally, spatial variability of the resource plays a critical role. There can be a mismatch between resource dense regions and quality of resource that can be extracted from these regions. A case in point can be China, where on a country level the potential for RTSPV is high, but on a sub national level the clustering of Built-up area is present in the east and southeast regions of China where the capacity factor of solar is the lowest. Conversely, the regions having the highest capacity factor for solar have the lowest clustering of Built-up area, Figure 11. Here, the biggest advantage of conducting the assessment at a high spatiotemporal resolution is evident where spatial variation of resource within a country can be analysed accurately.

To analyse the interplay of factors highlighted in the last paragraphs, we constructed scenarios that systematically analyse the effect of updating potentials and costs. Additionally, we ran sensitivity scenarios to analyse the effect of rooftop availability, panel efficiency and emission budget on uptake of RTSPV technology. Table 11 documents the scenario names and their narratives. Due to the time limitation of Young Scientist Summer Program 2021, we could only analyse the results for SSP2 narrative. Additional set of analysis is documented in Future Work section.

Table 11 | Scenario Narratives

Scenario Name	Scenario Abbreviation	Narrative
Base	B	Base Global 12 Region MESSAGEix-GLOBIOM model
Old Cost New Potential	OCNP	Keeping Base cost constant, split and update the potentials for RTSPV and UTSPV
New Cost New Potential	NCNP	Update cost, split and update the potentials for RTSPV and UTSPV
Sensitivity	SR	Rooftop availability 30%, 50%, 100%
Sensitivity	SP	Panel Efficiency 15%, 20%
Sensitivity	SE	Emission bound of 800 million tonnes of carbon equivalent

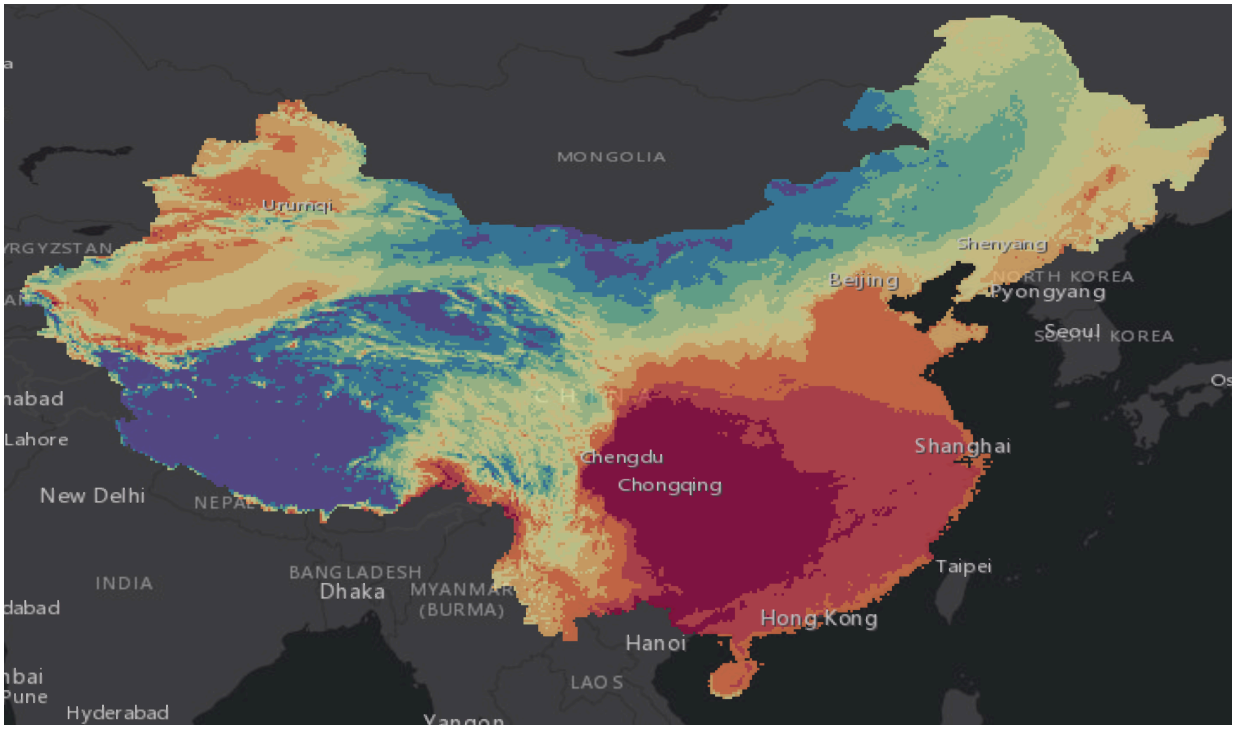

Figure 11| Visualisation of spatial distribution of Capacity Factors of RTSPV in China
The higher capacity factor regions shown here by blue colour are spatially located in the regions having low built-up area and consequently having low rooftop area. The lower capacity regions shown here by red are spatially located in the regions having high built-up area and consequently having high rooftop area. This mismatch between resource geolocation, presence of built-up area and energy demand centres exemplifies the need for a high-resolution spatiotemporal assessment of RTSPV technology.

To capture the effects of updating just the potentials while keeping the investment costs same as the base MESSAGEix-GLOBIOM model, we ran the OCNP scenario. For this, we binned our RTSPV and UTSPV potentials into 11 bins with each bin representing the maximum extractable electricity in GWa per annum binned by increasing capacity factors. We kept the panel efficiency at 15% in line with the current market trends and assumed that 100% rooftop is available for installation of RTSPV globally. Additionally, there was no bound on the emissions by the energy system and the RTSPV delivered electricity at final level with UTSPV delivering electricity at secondary level. This gradation in levels at which different solar PV technologies deliver electricity aids in accurate representation of energy systems where rooftop solar PV delivers energy directly to the prosumers as opposed to UTSPV that is connected to transmission and distribution networks. Compared to the base scenario (B), OCNP scenario results showed a significant increase in uptake of solar PV technology over the 2020-2050 time horizon. This can be gauged by the electricity extracted from solar PV technology, Figure 12. Additionally, we observed that RTSPV is the only technology being picked up by the model. This can be attributed to RTSPV being perceived as cheaper by the model compared to UTSPV due to grid losses in transmission and distribution networks and additional cost of traversing the grid infrastructure.

Next, we updated the Base scenario to capture the effects to updated costs and updated potentials (NCNP). In this scenario, we updated the investment costs for both RTSPV and UTSPV at a regional level. Assumptions and potentials from OCNP scenario were copied as is into the NCNP scenario. Here we again observed that RTSPV is the only technology being picked up by the model, but the amount of energy delivered by the technology is lower compared to OCNP scenario.

This can be attributed to higher costs of RTSPV technology represented in NCNP scenario whereas in OCNP scenario the costs are lower as we do not distinguish between RTSPV and UTSPV costs.

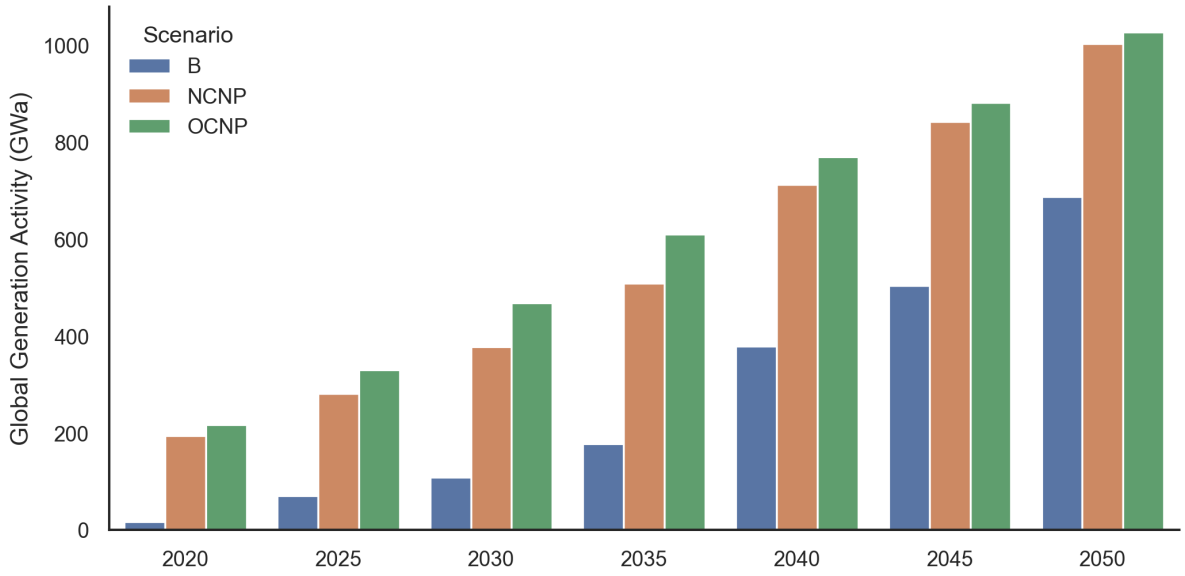

Figure 12 | Electricity generation activity of solar PV technology for B, OCNP and NCNP scenarios
Higher uptake of solar PV technology is seen when updated potentials over B scenario are used in OCNP scenario. Lower uptake of solar PV technology is seen in NCNP scenario when updating costs over OCNP scenario.

Uptake of RTSPV and its role in global energy transitions is also dependent on the rapid improvements in panel efficiencies leading to more electric eenrgy being generated from a given surface area. Another factor that will play a key role in uptake of RTSPV is the net availability of rooftop area for installation of roof mounted solar panels. This factor is sometimes the limiting constraint in the rollout of RTSPV in higher latitudes where rooftop surface can be constrained due to presence of chimneys and due to the slope and orientations of houses. Regional studies have agreed on a net rooftop availability of between 30-60%, but this percentage varies heavily between region to region and between building stock vintages. We devised a sensitivity scenario to capture the effects of rooftop availability and panel efficiencies on the uptake of RTSPV. For this, tested rooftop availability (SR) scenarios for 10%, 30%, 50% and 100% net available rooftop area and panel efficiency (SP) scenarios for 15% and 20% panel efficiency. The results of the scenarios are shown in Figure 13 where Figure 13a document the change in uptake of RTSPV and Figure 13b document the change in uptake of UTSPV for SA and SP scenarios. In all the SR and SP scenarios we kept our base model as NCNP and modified the rooftop availability and panel efficiency while keeping the UTSPV technology representation as constant. 100% rooftop availability at 15% panel efficiency sensitivity scenario is represented by NCEP scenario.

The key takeaway from the sensitivity scenarios is that a net global rooftop availability of 0.5 or 50% is the maximum amount of rooftop area that is required for maximum uptake of RSTPV technology. Any availability of greater than 50% will lead to marginal improvement in uptake of RTSPV at a global level. The increase in uptake of RTSPV when increasing the rooftop availability from 30% to 50% is less than when moving from 10% to 30%. Panel efficiencies can have small effect on the uptake of RTSPV in short time horizon of 2020-2030, but significant increase in uptake will occur in medium time horizon 2030-2050 for increased roof mounted solar panel efficiencies. For UTSPV technology, we observed limited uptake in NCNP scenarios but in SA and SP scenarios the

uptake of UTSPV increases depending on the RTSPV's rooftop availability. Significant uptake of UTSPV can occur for large constraint (10%-30%) on rooftop availability in medium time horizon. The uptake of UTSPV is slower in short time horizon but increases dramatically in successive time horizon.

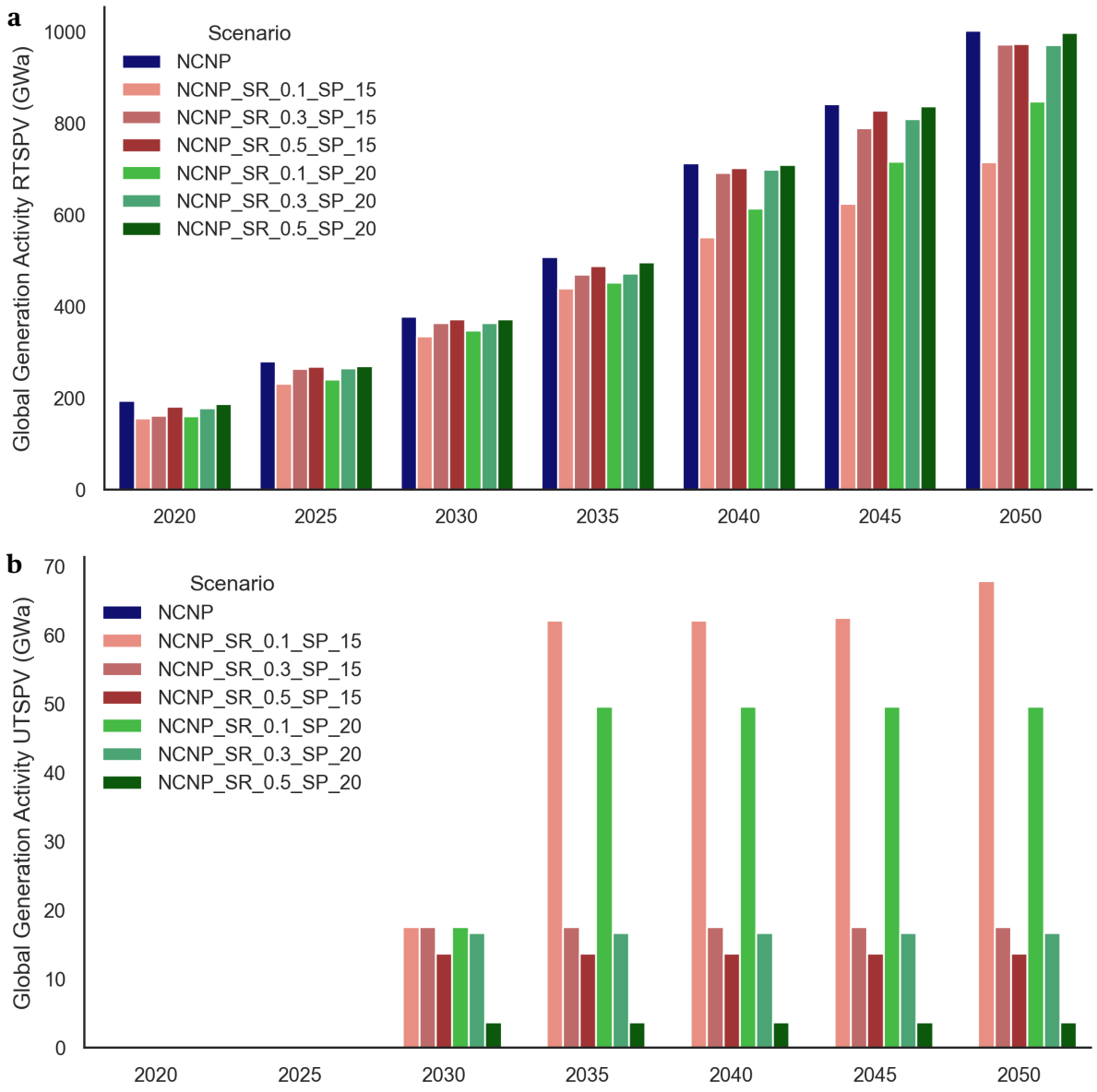

Figure 13 | Generation activity of solar PV technology for SR and SP scenarios
a, uptake of RTSPV technology for SR and SP scenarios. **b**, uptake of RTSPV technology for SR and SP scenarios. Rooftop availability are critical factors for uptake of RTSPV and UTSPV technologies and play a larger role than panel efficiencies. Large constraints on rooftop availability will lead to lower RTSPV uptake and higher UTSPV uptake. A saturation point is present at 50% rooftop availability, beyond which marginal increase in RTSPV will occur. Scenario nomenclature is NCNP_SR_X_SP_Y where X is the rooftop availability and Y is the panel efficiency and the base model is NCNP.

Mitigation of climate change impacts forms an important narrative for rapid deployment of low carbon energy sources and with the highest annual year on year growth, RTSPV is one of the prime candidates to spearhead the uptake of low carbon variable renewable technology. To analyse how emission bounds on energy systems (which are set because of global governmental will to aid in mitigation of climate change) will amplify the role of RTSPV in global energy transitions, we ran an emission bound scenario (B_SE_800) by introducing an 800 million tonnes of carbon equivalent

emission bound on our base model (B). Further, to analyse the effects of updated potential and costs (NCNP_SE_800) we introduced emission bounds on NCNP scenario. Additionally, the effects of rooftop availability were analysed by updating the SA scenarios with the emission bounds. The results of the SE scenarios run show significant increase in uptake of RTSPV in both short and medium time horizons, Figure 14. We also observed that compared to SA scenarios which didn't have emission bounds, the effects to rooftop availability are more pronounces where the threshold shifts to 30% rooftop availability compared to 50% in non-emission bound scenarios. This can be attributed to rapid saturation of rooftop area earlier on due to emission bounds and beyond 30% rooftop availability the increase in uptake of RTSPV is marginal ss the costs become larger.

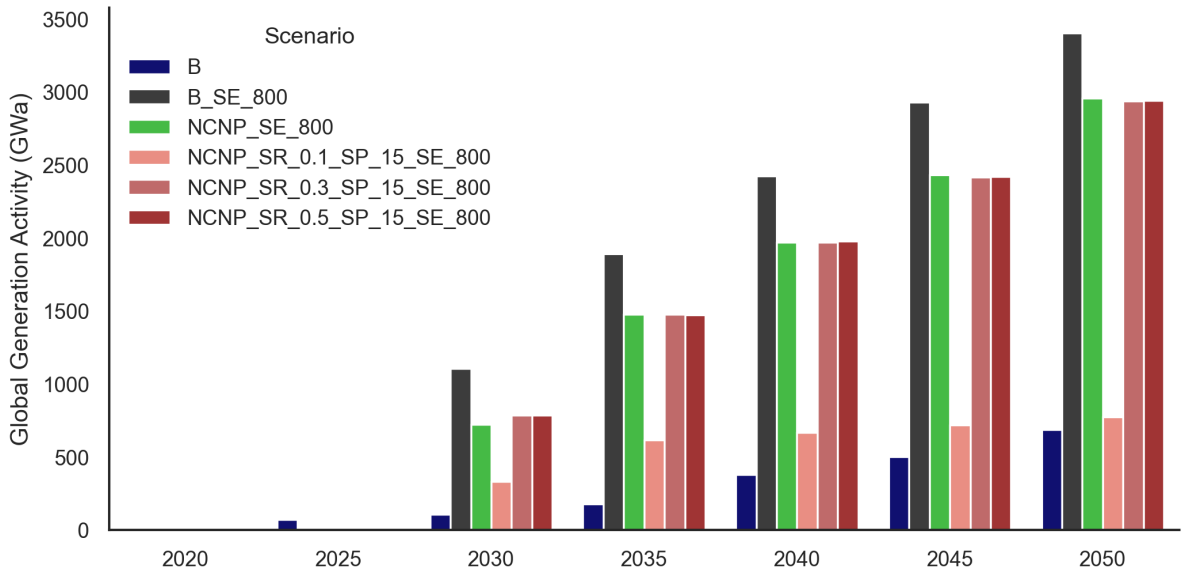

Figure 14 | Generation activity of solar PV technology for SE scenarios
Introduction of emission bounds shows significant increase in uptake of RTSPV technology in global energy system. Effect of rooftop availability is more prominent here as even smaller rooftop availability percentages show significant gains in uptake of RTSPV. Scenario nomenclature is X_SE_800 where X represent the base scenario models from previous section.

8. Discussion and Conclusion

In this study conducted during Young Scientist Summer Program of IIASA, we have attempted to highlight the role of RTSPV technologies in global energy transition. To understand how RTSPV will evolve spatially over the globe in medium term time horizon, we conducted a high-resolution spatiotemporal assessment of the technology by extending on a novel ML framework developed by the author of the report previously. As a result of this assessment, we have generated global datasets and tools which will be of importance to researchers working in the field of population mapping, climate change risk, energy systems, and energy-climate-land nexus policy formulation. We have documented and analysed how global rooftop area will evolve for SSP narratives in a spatially explicit way and used this analysis to understand the uptake of RSTPV technologies in global energy systems using the MESSAGEix modelling framework. Additionally, we have conducted a high-resolution spatial assessment of UTSPV technology that will aid in improved representation of solar PV technologies by providing potential and cost details for its sub technologies.

This study draws important conclusions that can aid in understanding the role of prosumer driven RTSPV. First, we concluded that for the SSP narratives, the global rooftop area can grow between 3-26% by 2050 compared to 2020. The largest growth will be seen in the African continent which highlights the importance of RTSPV as a decentralised energy source for countries undergoing socioeconomic transitions. It is here that the biggest benefit of clean energy and prosumer driven climate change mitigation will be observed. Second, the current global 2020 building stock is sufficient to meet the current yearly aggregated global electricity demand. Combined with increased growth of rooftop area in regions abundant with good quality solar resource, the future change in RTSPV's power generation potential can see a growth of between 25-50% which will be an important source of low carbon energy.

Third, we showcased how presence of large RTSPV potential does not automatically amount to high grade power generation by analysing the example of China where there is mismatch between high grade solar resource and hotspots of rooftop area. This conclusion highlights the importance of high-resolution sub national assessment of renewable resources especially RTSPV for designing informed energy policies. Additionally, for renewable resources that have fixed spatial attributions e.g., RTSPV whose deployment locations are fixed and cannot be spatially traversed like that of wind turbine driven energy generation, the importance of assessment of spatial distribution of the resource is evident as it can help with prioritising the locations of technology deployment.

Fourth, the result of scenario runs conducted using MESSAGEix platform have provided valuable insights on the dynamics of RTSPV deployment at a global scale. Here we analysed that rooftop availability is the largest limiting factor for the uptake of RTSPV technology. The rooftop availability has a threshold of 50% beyond which only marginal gain in uptake of RTSPV can be observed. Panel efficiencies play a lesser role in short term uptake, but in medium time horizon panel efficiencies will play a critical role. Fifth, RTSPV show important role for climate change mitigation policies. Introduction of emission bounds in energy systems have significant effect on the uptake of RTSPV where even presence of smaller rooftop availability will demonstrate significant uptake. Sixth, the OSM gap detection tool developed in this study has cross domain use cases and will aid in spatially explicit studies that utilise open source generated building polygons at a global level.

In this study we have demonstrated how a high-resolution spatiotemporal assessment of solar PV technology and its subcategories namely RTSPV and UTSPV can aid in improved representation of solar PV technology especially of RTSPV. However, additional research on transmission/distribution grid costs and losses is required to completely capture the dynamics of the role of RTSPV in future energy systems. Additionally, role of storage[38] in mitigating the intra annual variation in solar resource along with new market mechanisms[39] like utility owned rooftop[40], microgeneration utility structure and feedback to the power grid needs to be analysed. Due to the time limitation of the study period, further analysis which is underway could not be completed for this report and will form part of future research outputs. In this study we have just touched the proverbial tip of the iceberg when analysing the datasets generated here. The true advancements enabled by this study and its datasets will form the basis of future work exploring the RTSPV technology from a city level to a global ROI.

References

1. International Energy Agency (IEA). Electricity Information: Overview 2019. *Stat. IEA* (2019).
2. 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* (2020) doi:10.1016/j.apenergy.2020.115705.
3. Jacobson, M. Z. *et al.* 100% Clean and Renewable Wind, Water, and Sunlight All-Sector Energy Roadmaps for 139 Countries of the World. *Joule* (2017) doi:10.1016/j.joule.2017.07.005.
4. 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).
5. IEA. Energy Technology Perspectives 2016: Towards Sustainable Urban Energy systems. *Int. Energy Agency* (2016).
6. Hoogwijk, M. M. *On the global and regional potential of renewable energy sources*. *Assembly* (2004).
7. 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).
8. Castellanos, S., Sunter, D. A. & Kammen, D. M. Rooftop solar photovoltaic potential in cities: How scalable are assessment approaches? *Environ. Res. Lett.* (2017) doi:10.1088/1748-9326/aa7857.
9. 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* (2016) doi:https://www.nrel.gov/docs/fy16osti/65298.pdf.
10. Rottensteiner, F. & Briese, C. A new method for building extraction in urban areas from high-resolution LIDAR data. *Int. Arch. Photogramm. Remote Sens.* (2002).
11. Assouline, D., Mohajeri, N. & Scartezzini, J.-L. Quantifying rooftop photovoltaic solar energy potential: A machine learning approach. *Sol. Energy* **141**, 278–296 (2017).
12. 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).
13. 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 (2017). doi:10.1109/IGARSS.2017.8127091.
14. Sirko, W. *et al.* Continental-Scale Building Detection from High Resolution Satellite Imagery. 1–15 (2021).
15. 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.* (2019) doi:10.1016/j.rser.2019.109309.
16. Joshi, S. *et al.* High resolution global spatiotemporal assessment of rooftop solar photovoltaics potential for renewable electricity generation. *Nat. Commun.* (2021) doi:10.1038/s41467-021-25720-2.
17. Riahi, K. *et al.* The Shared Socioeconomic Pathways and their energy, land use, and greenhouse gas emissions implications: An overview. *Glob. Environ. Chang.* (2017) doi:10.1016/j.gloenvcha.2016.05.009.
18. 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. (2020). doi:doi:10.5258/SOTON/WP00678.

19. OpenStreetMap contributors. Planet dump retrieved from https://planet.osm.org . (2021).
20. 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* (2020) doi:10.1038/s41597-020-0542-3.
21. Lloyd, C. T., Sorichetta, A. & Tatem, A. J. Data Descriptor: High resolution global gridded data for use in population studies. *Sci. Data* (2017) doi:10.1038/sdata.2017.1.
22. Buchhorn, M. *et al.* Copernicus Global Land Service: Land Cover 100m: collection 3: epoch 2019: Globe. (2020) doi:10.5281/zenodo.3939050.
23. Jones, B. & O'Neill, B. C. Spatially explicit global population scenarios consistent with the Shared Socioeconomic Pathways. *Environ. Res. Lett.* **11**, (2016).
24. Gao, J. & O'Neill, B. C. Mapping global urban land for the 21st century with data-driven simulations and Shared Socioeconomic Pathways. *Nat. Commun.* **11**, 1–12 (2020).
25. Dellink, R., Chateau, J., Lanzi, E. & Magné, B. Long-term economic growth projections in the Shared Socioeconomic Pathways. *Glob. Environ. Chang.* **42**, 200–214 (2017).
26. Gorelick, N. *et al.* Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* (2017) doi:10.1016/j.rse.2017.06.031.
27. Corbane, C. *et al.* Automated global delineation of human settlements from 40 years of Landsat satellite data archives. *Big Earth Data* (2019) doi:10.1080/20964471.2019.1625528.
28. Chen, T. & Guestrin, C. XGBoost: A Scalable Tree Boosting System. (2016) doi:10.1145/2939672.2939785.
29. Herfort, B., Lautenbach, S., Porto de Albuquerque, J., Anderson, J. & Zipf, A. The evolution of humanitarian mapping within the OpenStreetMap community. *Sci. Rep.* **11**, 1–15 (2021).
30. Barrington-Leigh, C. & Millard-Ball, A. The world's user-generated road map is more than 80% complete. *PLoS One* (2017) doi:10.1371/journal.pone.0180698.
31. Orden, A., Flores, R. A., Faustino, P. & Samson, M. S. Measuring OpenStreetMap building footprint completeness using human settlement layers. *GitHub Repository* (2020).
32. [Data/information/map] obtained from 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.
33. Creutzig, F. *et al.* The underestimated potential of solar energy to mitigate climate change. *Nature Energy* (2017) doi:10.1038/nenergy.2017.140.
34. Huppmann, D. *et al.* The MESSAGEix Integrated Assessment Model and the ix modeling platform (ixmp): An open framework for integrated and cross-cutting analysis of energy, climate, the environment, and sustainable development. *Environ. Model. Softw.* **112**, 143–156 (2019).
35. Betak, Juraj; Caltik, Marek; Cebecauer, Tomas; Chrkavy, Daniel; Erdelyi, Branislav; Rosina, Konstantin; Suri, Marcel; Suriova, N. Global Photovoltaic Power Potential by Country (English). Energy Sector Management Assistance Program (ESMAP) Washington, D.C. : World Bank Group.
36. International Renewable Energy Agency, A. D. Renewable Power Generation Costs in 2019. (2020).
37. IEA. Renewables 2020. *IEA, Paris*.
38. Zakeri, B., Gissey, G. C., Dodds, P. E. & Subkhankulova, D. Centralized vs. distributed energy storage – Benefits for residential users. *Energy* **236**, 121443 (2021).
39. Parag, Y. & Sovacool, B. K. Electricity market design for the prosumer era. *Nat. Energy* **1**, 16032 (2016).
40. Barbose, G. & Satchwell, A. J. Benefits and costs of a utility-ownership business model for residential rooftop solar photovoltaics. *Nat. Energy* (2020) doi:10.1038/s41560-020-0673-y.