Geospatial Big Data analytics to model the long-term sustainable transition of residential heating worldwide

Diego Moya[1, 2, 3]
[1]*Carrera de Ingeniería Mecánica, FICM*
Universidad Técnica de Ambato
Ambato, Ecuador
da.moya@uta.edu.ec

[2]*Department of Chemical Engineering*
Imperial College London
London, UK
d.moya17@imperial.ac.uk

[3]*Institute for Applied Sustainability Research (IIASUR)*
Quito, Ecuador

Sara Giarola
Department of Chemical Engineering
Imperial College London
London, UK
s.giarola10@imperial.ac.uk

Adam Hawkes
Department of Chemical Engineering
Imperial College London
London, UK
a.hawkes@imperial.ac.uk

Abstract*— Geospatial big data analytics has received much attention in recent years for the assessment of energy data. Globally, spatial datasets relevant to the energy field are growing rapidly every year. This research has analysed large gridded datasets of outdoor temperature, end-use energy demand, end-use energy density, population and Gros Domestic Product to end with usable inputs for energy models. These measures have been recognised as a means of informing infrastructure investment decisions with a view to reaching sustainable transition of the residential sector. However, existing assessments are currently limited by a lack of data clarifying the spatio-temporal variations within end-use energy demand. This paper presents a novel Geographical Information Systems (GIS)-based methodology that uses existing GIS data to spatially and temporally assess the global energy demands in the residential sector with an emphasis on space heating. Here, we have implemented an Unsupervised Machine Learning (UML)-based approach to assess large raster datasets of 165 countries, covering 99.6% of worldwide energy users. The UML approach defines lower and upper limits (thresholds) for each raster by applying GIS-based clustering techniques. This is done by binning global high-resolution maps into re-classified raster data according to the same characteristics defined by the thresholds to estimate intranational zones with a range of attributes. The spatial attributes arise from the spatial intersection of re-classified layers. In the new zones, the energy demand is estimated, so-called energy demand zones (EDZs), capturing complexity and heterogeneity of the residential sector. EDZs are then used in energy systems modelling to assess a sustainable scenario for the long-term transition of space heating technology and it is compared with a reference scenario. This long-term heating transition is spatially resolved in zones with a range of spatial characteristics to enhance the assessment of decarbonisation pathways for technology deployment in the residential sector so that global climate targets can be more realistic met.

Keywords*— *Heat demand; spatiotemporal; integrated assessment; spatial datasets; spatial big data analytics.

I. INTRODUCTION

The Conference of the Parties, COP26 goals have established ambitious 2030 emissions reductions targets to reach net zero by the middle of the century, mobilising required finance to accelerate the phase-out of fossils [2]. Previously, the 2015 Paris Agreement and the United Nations Framework Convention on Climate Change (UNFCCC) have also established an instrument by which national governments have committed to pursue efforts in limiting the global average temperature increase to 1.5 °C [3]. The fact that the ambitious targets outlined in the Paris Agreement are difficult to meet incentivises the adoption of new energy technologies, disrupting established fossil fuel-based technologies [4]. However, successful implementation of new energy technology deployment presents information challenges [5]. In order to support national, regional and global decarbonisation policy, energy systems modelling that uses systematically processed big data approaches are emerging to address these challenges [6], [7]. Overtaking information challenges requires, among other things, to understand the spatiotemporal interactions between climate change, population spreading, along with the spatial distribution of wealth and human development, when investing and consuming energy.

A large proportion of the global energy demand and related global greenhouse gas (GHG) emissions are concentrated in the residential sector (RS). The main end-use energy in buildings comes from space heating (SH), space cooling (SC), water heating (WH), lighting, and appliances. The energy consumption in these end-use applications can vary between 30% and 70% of total energy demand depending on the economic, cultural, and geographical features of a country or region [8, 9]. Overall, SH, SC, and WH can represent up to 80% of a building's total energy consumption, accounting for

32% of total global end-use energy demand worldwide and 20% of all global anthropogenic GHG emissions [10]. With such high energy consumption and such a large share of global energy related GHG emissions, focusing on reducing the energy consumption and emissions of the RS would play a major role in decarbonising the whole energy system worldwide. However, to better represent the energy shift required in order to achieve climate targets and support energy planning, it is important to consider the spatio-temporal distribution of energy demand in modelling the future sustainable energy transition. The current lack of spatially and temporally resolved end-use energy demand assessments makes it challenging to identify the pathways to decarbonise the residential sector at a global scale.

This paper presents a novel GIS-based methodology for such an assessment to produce temporally explicit and spatially resolved energy demand zones (EDZs). A systematic, spatio-temporal data-driven approach, based on existing GIS data of end-use energy demand and end-use energy demand density from [11], Gross Domestic Product [12] and population [13], has been developed to derive the EDZs for each end-use energy demand via spatial clustering techniques. The associated end-use energy demand (SH, SC and WH) for each EDZ is estimated across 165 countries covering 99.96% of global energy users. The outputs of this analysis can be integrated into residential sector modules of Integrated Assessment Models and can equally be used for input in other energy systems models.

The remainder of this paper is organised as follows. Section two identifies key elements in the literature to consider when assessing global end-use energy demand in the residential sector. In section three, a description of the GIS-based clustering techniques and the EDZ approach is given. Results are presented in section four adopting the breakdown of regions used in the Energy Technology Perspectives of the International Energy Agency (IEA). Here, we present results for several EDZ systematically founded in this research. We provide a snapshot of results in terms of EDZ distribution and the spatio-temporal phenomena of the heating transition. Section four also discusses the implications of these results for energy and climate policy followed by a discussion of implications for future research.

II. Literature review

A. Modelling challenges

The RS is a classic case of all of the typical challenges of modelling technology deployment in energy systems models [5]. Buildings are heterogeneous; they can be categorised as residential, services or industrial, each with their own unique characteristics (age, construction, occupancy/use, services, location, and climate). Buildings are also hard to decarbonise, mainly due to their heating requirements currently being served with cheap and effective fossil-fuelled technology (e.g., gas furnaces). Alternative low carbon options such as heat pumps (when powered by renewable electricity), or district heating fuelled by bioenergy or waste heat sources, are expensive and face many barriers to implementation [14, 15]. Therefore, characterising building decarbonisation options in models is challenging. This is specifically because of a lack of highly spatially and temporally resolved approaches to assess the spatio-temporal characteristics of energy demand. GIS-based approaches that assess large amounts of data are needed to capture the local conditions of energy demand to evaluate the technical and economic characteristics of a range of technologies.

B. Geograohical Informartion Systems

Geographical Information Systems (GIS) methodologies are attracting widespread interest due to their potential to geospatially locate the demand for energy in the RS [16, 17]. Little of the work so far has been focused on understanding the temporal (intermittency) and spatial (geographical) characteristics of the end-use energy demand density in the RS. The end-use energy demand density is the total demand for a heating or cooling service for a defined area: street, neighbourhood, city, country or region [18]. In the assessment of transition pathways to decarbonise the RS, energy density is a significant factor in overall investment costs due to associated infrastructure costs (e.g., distribution pipes for DHC systems). As a result, researchers have begun to use gridded population density, combined with temperature profiles and other data, to derive spatially resolved energy demand [19, 20], but energy demand density assessment is still an unresolved task.

A few studies have focused on estimating the heating demand across Europe, USA, and China. The main approaches so far have been the use of national energy balances such as those presented by the International Energy Agency [21, 22], the national energy authorities [19, 23, 24], the United Nations database as in [25], and other research initiatives as in [26, 27]. In comparison to SH demand studies, cooling demand studies are rarely found in the scientific literature. An exception is Isaac and van Vuuren [26], who estimated the SC demand for eleven regions. SC demand estimations are also presented in more detail for the European Union at the country and regional level in [21, 28, 29]. However, to our knowledge, a systematic assessment of the spatio-temporal variations of end-use energy demand that considers demand density for the world at the country scale has not been presented in the literature prior to the publication of this study. There remains a need for a methodology that can temporally and spatially estimate the end-use energy demand considering the spatio-temporal distribution of a range of end-use energy demand density values.

C. Geospatial big data for sustainable energy

The size of spatial datasets has been growing by at least 20% every year lastly. The analytics of such amounts of geospatial data is full of challenges for researchers, importance and benefits for stakeholders [30], such as urban planners, local authorities, enterprises, researchers and energy policymakers.

SENESCYT, Award No. CZ03-35-2017 and UTA, Award No. 1895-CU-P-2017.

More recent attention has focused on the analytics of geospatial big data to enhance sustainable energy transitions [6]. The use of geospatial big data analytics can inform stakeholders on the development of sustainable energy strategies. However, stakeholders still face information challenges such as data availability, data scalability, data integration, data inconsistency, geocoding, and data privacy [31]. Although challenges are overtaking using geospatial big data analytics, understanding the spatiotemporal variation of a range of factors that influence the energy consumption in the residential sector has not been yet addressed in the literature.

Integrating GIS and big data analytics (Geospatial big data analytics) has shifted from its use on single-buildings energy analysis to its use at urban scales [7] to countries, regions and even for the whole world [32]. Geospatial big data analytics is a potential tool to integrate and compare all the possible related features when planning sustainable transitions of the residential sector [7]. Geospatial big data or spatial data mining is being used to discover hidden knowledge from large spatial datasets to inform stakeholders [33]. Geospatial big data analytics implements machine learning algorithms, statistical methods and artificial intelligence methods to solve energy problems [17]. The most common machine learning methods in energy analysis are Unsupervised Machine Learning techniques that classify each spatial data point into specific group according to similar properties. For example, Afaifia, et al. [34] implements hierarchical cluster analysis towards the development of energy transition policies for more sustainable urban areas. Deb, et al. [35] applies DBSCAN (Density-Based Spatial Clustering of Applications with Noise) algorithm to disaggregate electrical load profiles into space heating and water heating. Sachs, et al. [32] develop a method based on K-means algorithm to assess worldwide gridded energy demand density. Overall, studies focus on the analysis of spatial big data without the possibility of its use in energy systems modelling.

D. Unsupervised Machine Learning for geospatial analysis

Clustering analysis is an established Unsupervised Machine Learning (UML) technique in data mining used for classifying patterns into groups of similar characteristics, termed "clusters". Clustering techniques have been widely applied in the energy sector, especially for electricity consumption pattern recognition. Voulis, et al. [36] have provided a comprehensive review and a methodology for assessing the spatio-temporal electricity demand profiles at urban scales. Although a number of different clustering algorithms have been implemented in the energy field [37], Gianniou, et al. [38] argue that the K-means algorithm has a great potential to be adopted in energy demand studies. Gianniou, et al. [38] applied the K-means clustering algorithm to group heating customers with similar consumption patterns. The K-means approach requires the user to initially provide a number of clusters k, which makes it challenging to determine an optimal number of clusters. To address this, the Elbow Method (EM) can be applied to evaluate the evolution of the within-cluster variance as a function of the number of clusters.

The EM suggests that the elbow/knee of the curve is the most appropriate number of clusters [39]. K-means clustering techniques can also be used to segment heating and cooling demand density patterns. Unternährer, et al. [40] studied the potential of K-means for spatial clustering analysis to evaluate the use of geothermal energy in DH systems. They found that the cost-effectiveness of DH integration in urban areas is greatly affected by considering the spatial heating demand density. Unternährer, et al. [40] have also defined a range of heating demand by defining a lower and an upper bound based on clustering results. However, clustering approaches have not yet been applied in the residential sector at the global scale to further explore the tranches of end-use energy demand density where certain technologies can be introduced.

III. METHODOLOGY

Step 1. Clustering Geospatial Big Data: The first step is to classify three global high-1km2 resolution spatial datasets: GDP per capita [12], space heating demand per capita (HDpc) and heating density (HD) [32]. Here, the K-means clustering method is undertaken in two main stages. First, the optimal number of clusters (ONC) is calculated by applying the Elbow Method (EM) followed by the Hartigan-Wong algorithm for K-means (1979); this is defined by Eq. (1) and Eq. (2) [39, 41], and explained in [42]. Eq. (1) also defines the EM, which is applied to calculate the sum of squared errors (SSE) of within-cluster distances between the cluster centres and their members. When the number of clusters k is plotted against SSE, the visually determined location of the elbow of the curve indicates the appropriate number of clusters, or ONC. Once the ONC is obtained, the K-means clustering algorithm is applied to define the lower and upper bounds of each dataset. The algorithm is applied by an iterative process to minimize the intra-cluster inertia criterion defined by Eq. (2).

$$minimise\ \left(W_k = \sum_{r=1}^{k} \frac{1}{n_r} D_r\right) \qquad \text{Eq. (1)}$$

$$D_r = \sum_{i=1}^{n_r-1} \sum_{j=1}^{n_r} \left\| d_i - d_j \right\|^2 \qquad \text{Eq. (2)}$$

Where:

W_k is the average internal sum of squares,

k is the number of clusters,

n_r is the total members points in the cluster r, and

D_r is the sum of distances between points d_i and d_j, which belong to a cluster.

Each point between upper and lower bounds is representing by d_i and d_j in Eq. (2).

Step 2. Intercepting clustered Geospatial Big Data: Then, the three reclassified layers are spatially overlayed to arise a new zone that captures the three previous spatial attributes.

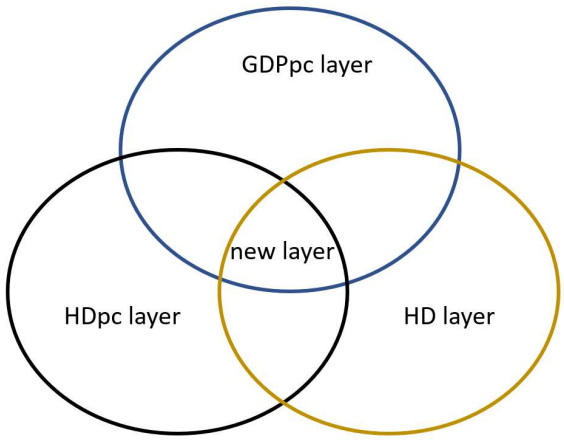

Fig. 1. Representation of the spatial overlaying of three datasets previously reclassified [1].

Step 3. Extracting energy demand in intercepted zones: In each new zone, the total demand is extracted to be used in energy modelling as explained in [32].

Step 4. Zonal-based modelling: The MUSE (ModUlar energy system Simulation Environment) model uses the demand in each zone to model a sustainable scenario of energy transition in the residential sector. MUSE models real investors behaviours [43]. To simplify the simulation, the key investment metric in MUSE will be the Levelised Cost of Electricity (LCOE). MUSE applies a technology rich approach where each technology is individually characterised regarding costs and engineering performance [43], [44].

IV. Results and Discussion

A. Optimal number of clusters

Here, we present the results of an initial exploratory analysis of the GIS data using the Elbow Method (EM) described before. As can be seen in Figure 2, this method consists of clustering global spatial data of annual (a) GDP per capita (GDPpc), (b) heating demand per capita (HDpc), and (c) heating density (HD) to obtain the optimal number of clusters (ONC) respectively. In Figure 2, we can observe that the SEE value tends to decrease toward cero as the number of clusters increases. In the case of spatial clustering of GDPpc (Fig. 2a), small SEE variation can be observed at the 5-cluster solution, while in HDpc (Fig. 2b), at the 4-cluster solution as well as in HD (Fig. 2c). The elbow indicates that cluster solutions larger than 6 for GDPpc and 4 for HDpc and HD do not exert a substantial impact on the total SSE variation. Thus, the ONC for GDPpc is selected at 5-cluster solution. Applying the same method to HDpc values, the ONC is selected at the 4-cluster solution as for HD.

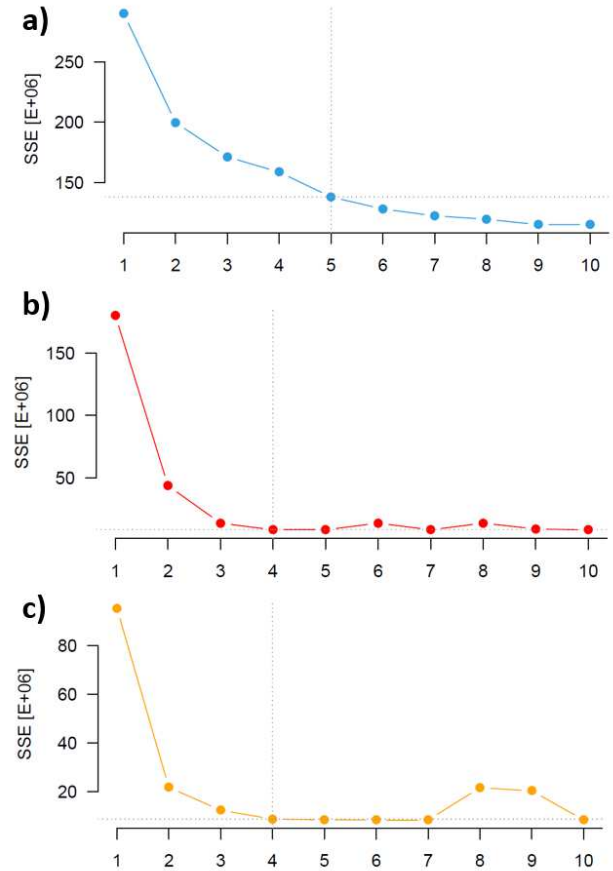

Fig. 2. Global optimal number of EDZs for (a) GDP per capita, (b) heating demand per capita, (c) heating density.

Once the ONC is selected for GDPpc, HDpc and HD, the K-means clustering algorithm segments the datasets into five, four and four groups, respectively. The calculations include the estimation of the cluster centroids along with the elements that belong to each cluster. Then, the limits of each zone are defined to be halfway between each consecutive centroid value. Table 1 shows the lower and upper limits of each reclassified zone defined here.

Table 1: Clustered results of GDPpc, HDpc and HD.

GDPpc [US$/y]			HDpc [MWh/cap]			HD [MWh/km^2]		
classes	lower	upper	classes	lower	upper	classes	lower	upper
1	min	500	1	min	0.9	1	min	1790
2	500	3785	2	0.9	3.2	2	1790	12080
3	3785	18215	3	3.2	5.3	3	12080	36927
4	18215	41667	4	5.3	max	4	36927	max
5	41667	max						

B. *Reclassified rasters*

The reclassified datasets are presented in Figure 3. The GDPpc dataset has been segmented into five groups (Figure 3a), while HDpc and HD has been segmented into four groups (Figure 3b and Figure 3c), respectively. The global atlases of the spatial distribution of the reclassified zones for GDPpc, HDpc and HD are illustrated in Figure 3. It is shown that most of the regions with high GDPpc values (classes 4 and 5) are concentrated in North America, Europe and East Asia. This would be mainly because of the higher country's economic output per person in each of those regions. A similar trend is observed for HDpc (classes 3 and 4) because of countries with both extreme seasonal weather conditions and highly populated areas. What is surprising is that other cold regions located in the Andes Cordillera (e.g., Colombia, Ecuador, Peru, Chile) or in Norther Europe (e.g., Norway, Finland) do not present high HDpc, being at classes 2 and 3 mostly. It is found that very high heating HDpc zones (class 4) are associated with small areas accounting for less than 5% of populated zones worldwide. On the other hand, lower HDpc zones (class 2) are in approximately 95% of populated zones worldwide. Figure 3c presents the global atlas of the spatial distribution of reclassified HD zones. Significant HD zones are highly concentrated in zones in East US, Europe and China. As can be identified in Figure 3c, zones located in North America, Southern Chile, Western Europe, North India and East China have the highest HD. Some places with high HD zones are among the poorest such us India, China, Chile, which are classified as lower-middle income economies. This is important to consider as income would play an important role in affording heating technologies to meet the demand.

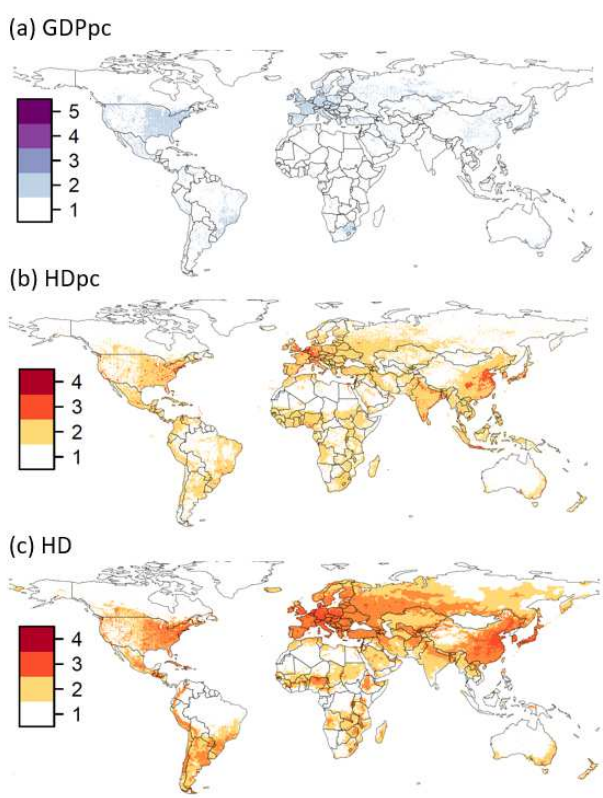

Fig. 3. Global spatial distribution of reclassified GDPpc, HDpc and

Table 2: Global Energy Demand Zones (EDZs) disaggregation. Z: zone. Refer to Table 1 for classes values.

Classes [spatially resolved and time-explicit]			Global EDZs
GDPpc	**HDpc**	**HD**	
1	1	1	EDZ1
1	3	2	EDZ2
2	3	1	EDZ3
2	3	2	EDZ4
2	3	3	EDZ5
2	3	4	EDZ6
3	4	1	EDZ7
3	4	2	EDZ8
3	4	3	EDZ9
3	4	4	EDZ10
4	4	1	EDZ11
4	3	2	EDZ12
4	4	3	EDZ13
4	2	4	EDZ14
5	2	1	EDZ15
5	3	2	EDZ16
5	3	3	EDZ17
5	2	4	EDZ18
5	3	1	EDZ19
5	3	2	EDZ20

Table 2 presents the disaggregation of the Global Energy Demand Zones (EDZs) which arise from the interception of three reclassified layers shown in Figure 3. 20 EDZs globally are identified where the total energy demand is calculated to be used in energy modelling. These EDZs captures the heterogeneity of residents when analysis household energy demand. For example, EDZ1 captures the socioeconomic conditions at GDPpc class 1 (up to 500 USD/yr) with a HDpc class 1 (up to 0.9 MWh/cap) and a HD class 1 (up to 1790 MWh/km^2). On the other hand, EDZ20 captures the socioeconomic conditions at GDPpc class 5 (41667 USD/yr or more) with a HDpc class 3 (between 3.2 and 5.3 MWh/cap) and a HD class 2 (between 1790 and 12080 MWh/km^2).

C. *Intercepted layers*

Figure 4 illustrates how the EDZs for China are characterised by three measures when systematically overlayed. Initially, GDP is obtained (Figure 4a), which is divided by population to obtain the GDPpc (Figure 4b). Here, GDPpc is reclassified as explain previously. We observe that China has 3 out 4 GDPpc classes. Reclassified HDpc is presented in Figure 4c, and reclassified HD in Figure 4d. Figure 4e and Figure 4f respectively show new zones when two attributes (GDPpc and HDpc) and three attributes (GDPpc, HDpc and HD) are intersected or overlayed. This example shows that not all classes might be present in each country or region. Although worldwide there are 20 EDZs, there is only 10 EDZs in China. This shows the heterogeneity of the residential sector composition that varies from country to country.

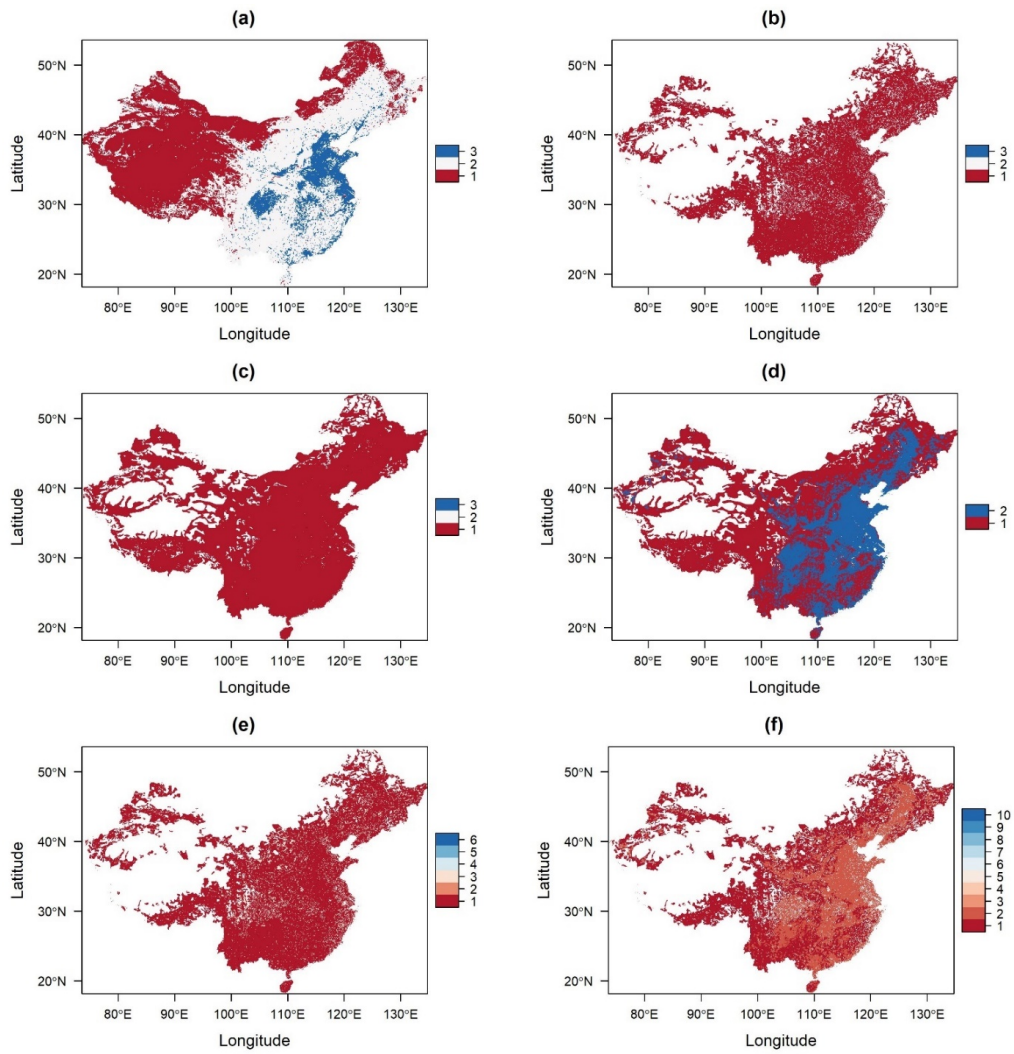

Fig. 4. Spatial characterisation of EDZs with 1, 2 and 3 spatial attributes for China. (a) Zone-based classes defined for GDP per km^2, (b) Zone-based classes defined for GDPpc per km^2, (c) Zone-based classes defined for HDpc per km^2, (d) Zone-based classes defined with HD per km^2 data, (e) Zones defined with the intersection of two attributes (GDPpc and HDpc), (f) Zones defined with the intersection of three attributes (GDPpc, HDpc and HD).

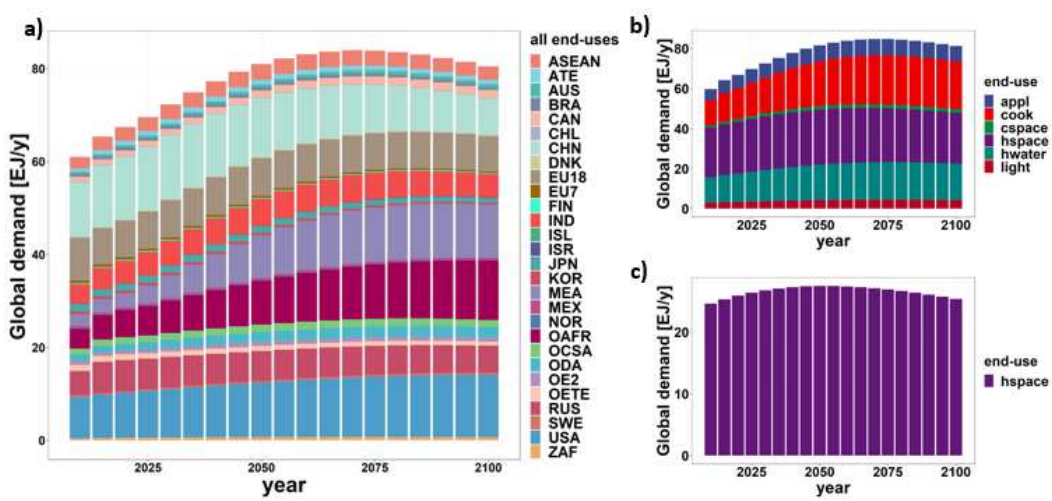

Fig. 5. Overview of the global demand of end-uses of energy by (a) region, (b) six end-uses; and (c) space heating only. Countries with the highest household demand of energy: USA, RUS, OAFR (Africa expect South Africa), MEA region (Middle East), India, EU18 and China.

D. GIS-based modeling global demand

Figure 5a summarises the global demand of end-use energy by the residential sector, disaggregated in 28 regions following the Energy Technology Perspective of the IEA [45] and considering the consumption drivers of the Shared Socioeconomic Pathways, SSP2 narrative (middle of the road) [46]. Figure 5b compares the breakdown of total end-use energy demand according to all residential energy services (i.e., appliances, cooking, space cooling, space heating, water heating and lighting). From all these energy demand services, the global demand for space heating is presented in Figure 5c. Closer inspection of the space heating demand shows the worldwide demand at 26 EJ in 2020, reaching a maximum demand of 27.5 EJ in 2050, and decreasing to 25.3 EJ by 2100. Overall, space heating moves from 41% of the total residential energy services in 2010 to 34% in 2050 and 31% in 2100. These projections are in line with the study reported by Knobloch, et al. [47] that simulates decarbonisation pathways of residential heating under the SSP2 narratives which makes both studies comparable. Knobloch, et al. [47] calculated future changes in the global heat demand, implementing the IMAGE-REMG model. While Knobloch, et al. [47] estimate the heat demand for space and water heating at 41.4 EJ in 2020, the study presented here estimated 40.2 EJ in the same year. While similarities exist between the methodologies used in both studies, the main difference is the EDZ disaggregation across regions individually presented next.

E. Global long-term transition

Figure 6 provides the EDZ-based breakdown of global space heating demand of the residential sector at 20 EDZs for 28 regions worldwide. The zones breakdown by regions illustrates the heterogeneity of residential space heating demand globally, regardless the region. Each zone, from EDZ1 to EDZ20, is spatially resolved with similar gridded characteristics anywhere on the globe. It can be seen the composition of each zone's heating demand profile for the century ahead. Interestingly, each EDZ is particularly unique by the fact that incorporates spatially resolved and time explicit attributes of each region of the world. For example, EDZ7, the zone with the highest demand, emerges from the combination of three gridded attributes, as can be seen in Table 2. This zone represents a region's economic output per person between 3785 and 18215 US$ per year within a zone where demand per capita is higher than 5.3 MWh/cap per year and less than 1790 MWh/km2 of HD. EDZ7 is spatially located in all regions of the world and evolves from a total demand of 7.5 EJ in 2010 to 9.8 EJ by 2050 and up to 8.5 EJ by 2080.

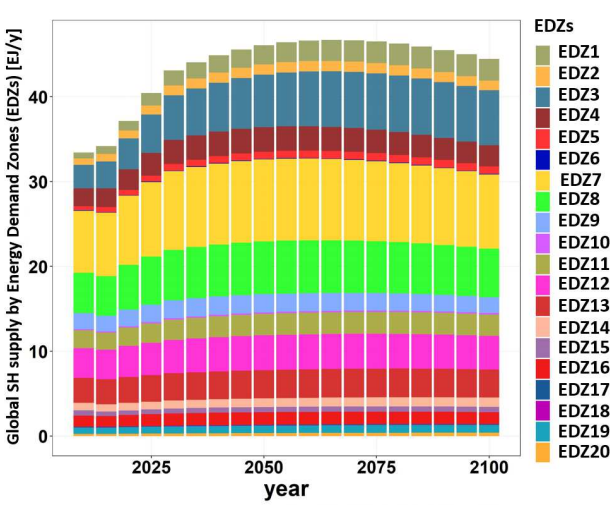

Fig. 6: Global space heating supply by EDZs.

Figure 7 shows the global technology transition for space heating supply in the residential sector, worldwide, for two scenarios: (a) Sustainable scenario with a carbon tax scheme globally; and (b) Reference scenario without a carbon tax scheme (an unsustainable scenario). The key difference between scenarios (a) and (b) is the presence of a carbon price scheme. The sustainable scenario presents a radical uptake of electricity-based heat pumps from 2030 onwards, whereas in the reference scenario, natural gas-based technologies are dominant in the same period.

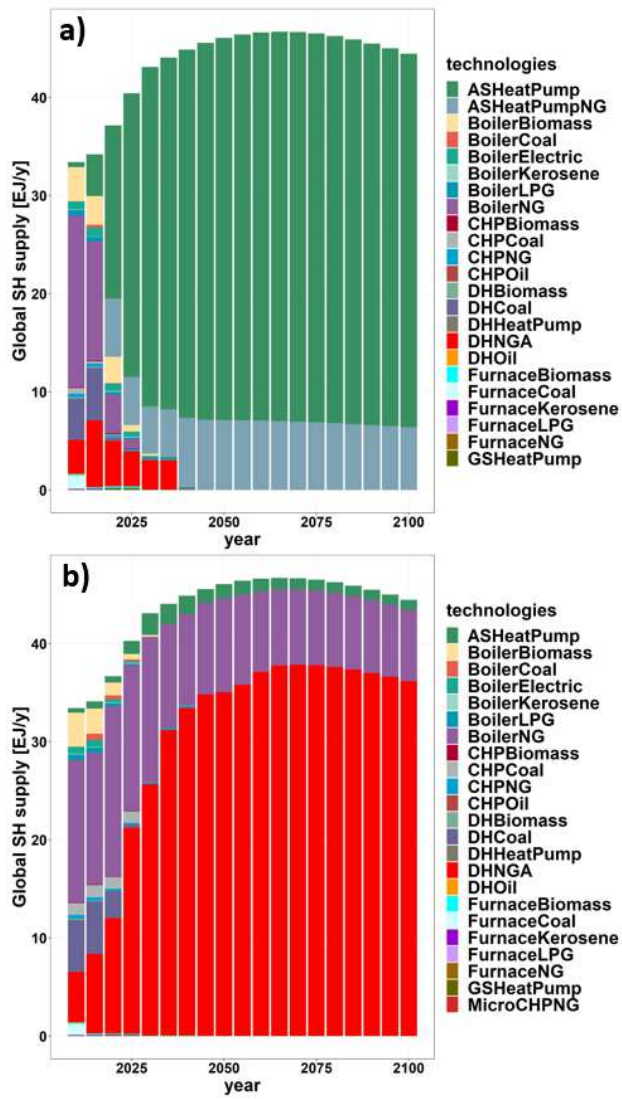

Fig. 7: Global space heating supply under two scenarios with heat density restriction. (a) Sustainable scenario with a carbon tax scheme globally; and (b) Reference scenario without a carbon tax scheme. Results are provided by technology disaggregation.

Figure 8 provides the largest set of spatially resolved and temporal explicit zones with same spatial characteristics worldwide along with the required technologies to supply them heat in a sustainable scenario that includes a carbon tax scheme globally. As observed before EDZ7 is the zone with the highest demand. In the breakdown of EDZs we can see the technology disaggregation for each zone. As expected, Air Source Heat Pumps (ASHeatPump) and ASHeatPump with natural gas back up (ASHeatPumpNG) are the dominant technologies in the transition. There are two extra key observations from the analysis: the heterogeneity and the complexity of the global space heating supply in the residential sector within the effect of gridded attributes. EDZ's heterogeneity can be seen in the different supply profiles that each zone delivers, while the zone's complexity is reflected in the technology stock that changes in time depending on budget and demand restrictions.

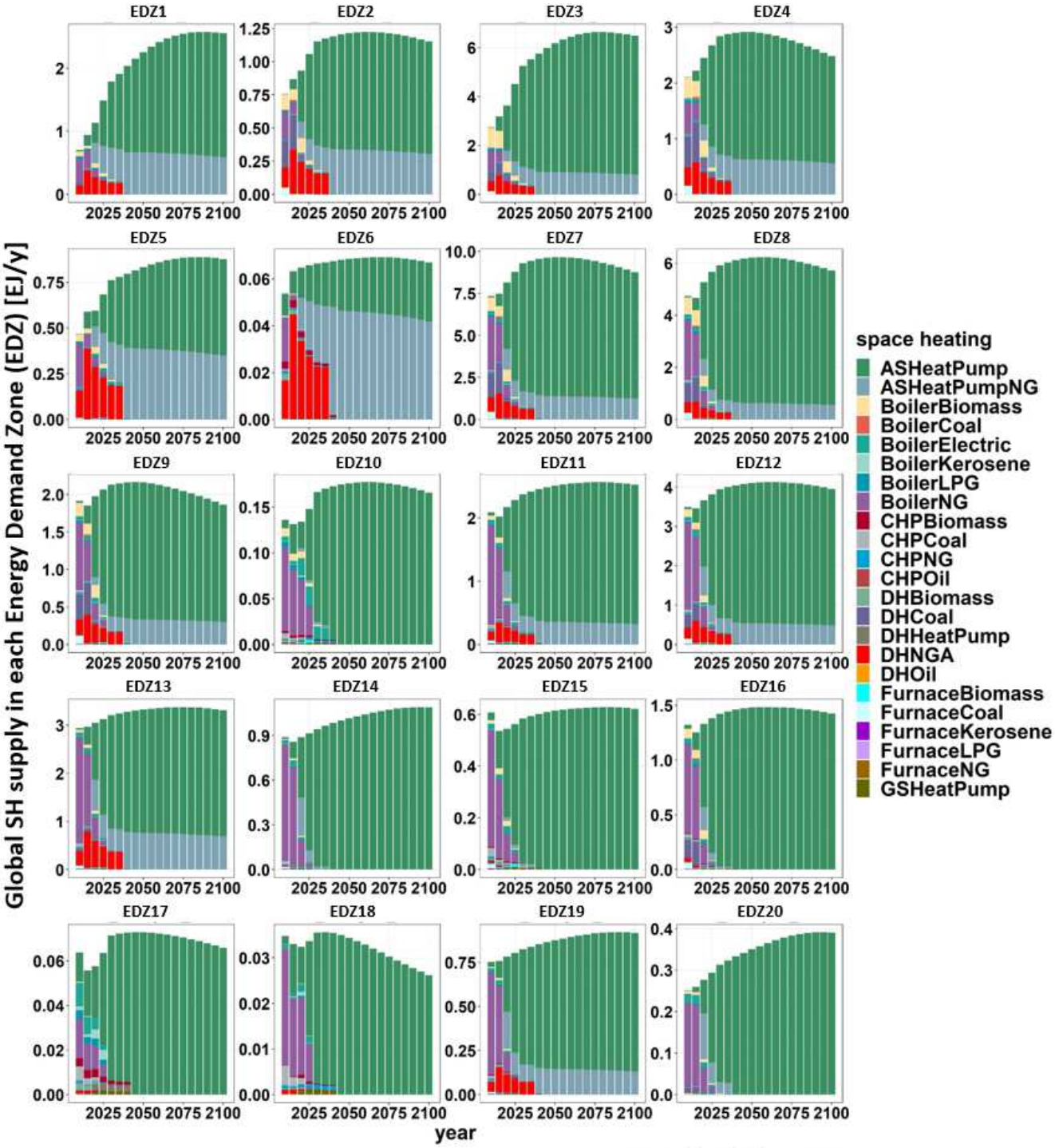

Fig. 8: Global space heating supply by EDZs in a sustainable scenario with a carbon tax scheme globally. Results are provided by technology disaggregation.

Figure 9 presents the global EDZs for a reference scenario without a carbon tax scheme: an unsustainable scenario. In this scenario, the dominant technologies are natural gas-based: boilers (BoilerNG) and district heating (DHNGA). In the case of the zone with the highest demand, DHNGA represents about 80% of the total supply from 2030 onwards. Although some EDZs uptake BoilerNG and surprisingly ASHeatPumps, the dominant technology prevails on DHNGA.

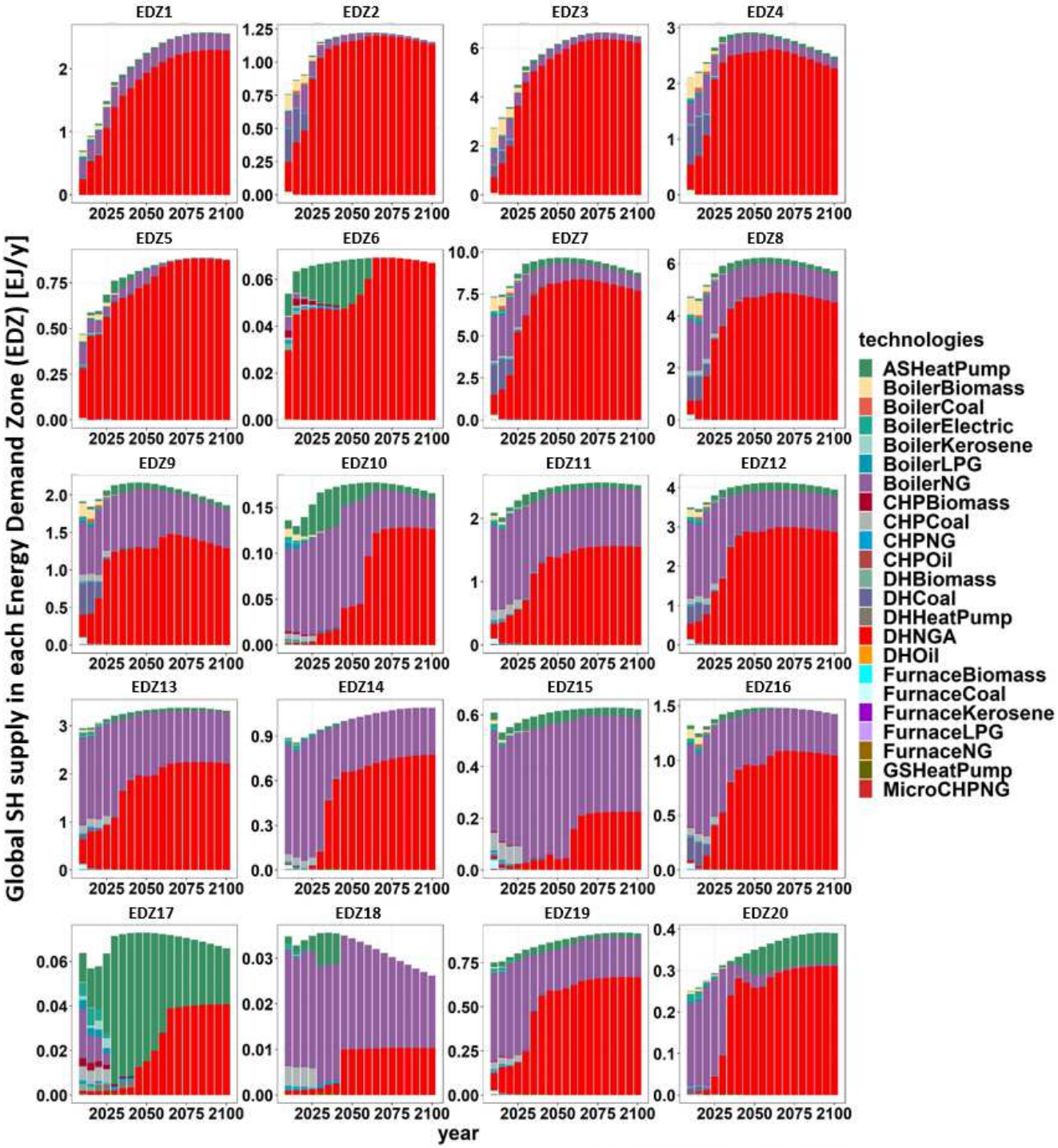

Fig. 9: Global space heating supply by EDZs in a reference scenario without a carbon tax scheme. Results are provided by technology disaggregation.

F. Global emissions

Figure 10 shows the emissions profile for two scenarios of space heating supply. The sustainable scenario with a carbon tax scheme globally (a) shows net zero emissions from the residential sector by 2045. On the other hand, the reference scenario without a carbon tax scheme (b) illustrates a dramatical increase of emission across the century ahead. These emissions profiles clearly illustrate that EDZ-based approaches can contribute with a more focus analysis, including a range of spatiotemporal variabilities.

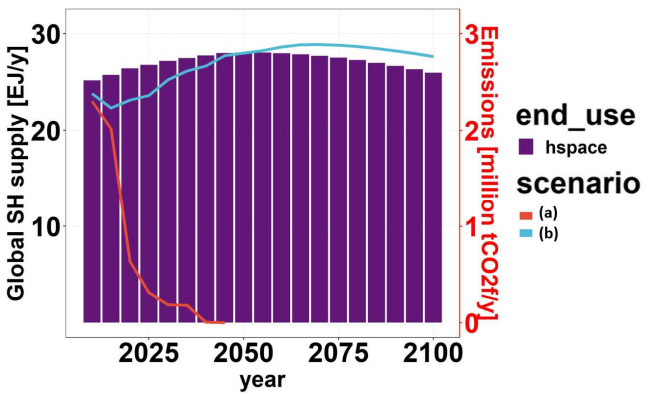

Fig. 10: Global related CO2 emissions profile for two scenarios of space heating supply. (a) Sustainable scenario with a carbon tax scheme globally; and (b) Reference scenario without a carbon tax scheme.

Conclusion

Before, the location characteristic of data has been under-utilised; this has changed now. With information technologies, spatiotemporal data for many research areas is being saved and grows rapidly every year. Geospatial big data analytics has increased awareness of the value gained by analysing big data in a geographic context. More specifically, in the energy arena, Geospatial big data analytic tools are informing stakeholders and energy policy makers the ability to discover location-based patterns and relationships from spatial data that may exist in many "unrelated" places worldwide. Geospatial big data analytic tools enable us to visualise and analyse large spatial datasets to reveal patterns and trends that would previously have remained hidden.

This research has developed an Unsupervised Machine Learning (UML)-based technique by the implementation of a Spatial Clustering Analysis to bringing together multiple data layers. The clustering method has reclassified three large data sets: GDPpc (includes GDP and population), HDpc (includes energy consumption, outdoor temperature and population) and HD (includes HDpc and population density). The reclassification process has ended with five classes for GDPpc, 4 classes for HDpc and four classes for HD. From the spatial intersection of these three reclassified datasets, a new layer has emerged with three characteristics in each zone: the called Energy Demand Zones (EDZs). Twenty EDZs have serve as input for a regional-based simulation of the global residential sector.

Two scenarios have been explored in the simulation using MUSE (a new Integrated Energy System Model developed at Imperial College London): (a) Sustainable scenario with a carbon tax scheme globally; and (b) Reference scenario without a carbon tax scheme. Results clearly show the role of GIS-based analysis of the energy transition of the residential sector. We observed different patterns on the technology uptake. Each EDZ provides a different energy supply profile which give us an idea of the key spatiotemporal features energy policy makers should focus on when planning the long-term energy transition. In the sustainable scenario, natural gas-based technologies are dominant. IN both scenarios, it can be observed the heterogeneity and the complexity of the residential sector and the importance of the effect of gridded attributes. Future research should go deeply on the gridded attributes when linking GIS-based inputs with MUSE or any other energy system model.

Acknowledgment

Mr Diego Moya has been funded by the Ecuadorian Secretariat for Higher Education, Science, Technology and Innovation (SENESCYT), Award No. CZ03-35-2017, The Technical University of Ambato (UTA), Award No. 1895-CU-P-2017 (Resolución HCU) and supported by The Science and Solutions for a Changing Planet Doctoral Training Partnership, Grantham Institute and the Sustainable Gas Institute at Imperial College London. The Institute for Applied Sustainability Research (IIASUR) supports international research on global sustainability applied to the Global South.

References

[1] M. Diego, "Global decarbonisation pathways of the residential sector applying GIS and ABM approaches," PhD, Chemincal Engineering, Imperial College London, 2021.

[2] COP26. "COP26 GOALS." https://ukcop26.org/cop26-goals/ (accessed 04 Nov, 2021).

[3] UNFCCC. "Adoption of the Paris Agreement: Proposal by the President: Draft Decision -/CP.21." https://unfccc.int/resource/docs/2015/cop21/eng/l09r01.pdf (accessed.

[4] F. G. N. Li and N. Strachan, "Take me to your leader: Using socio-technical energy transitions (STET) modelling to explore the role of actors in decarbonisation pathways," *Energy Research & Social Science,* vol. 51, pp. 67-81, 2019/05/01/ 2019, doi: https://doi.org/10.1016/j.erss.2018.12.010.

[5] S. Pfenninger, A. Hawkes, and J. Keirstead, "Energy systems modeling for twenty-first century energy challenges," *Renewable and Sustainable Energy Reviews,* vol. 33, no. Supplement C, pp. 74-86, 2014/05/01/ 2014, doi: https://doi.org/10.1016/j.rser.2014.02.003.

[6] J. Herman, H. Herman, M. J. Mathews, and J. C. Vosloo, "Using big data for insights into sustainable energy consumption in industrial and mining sectors," *Journal of Cleaner Production,* vol. 197, pp. 1352-1364, 2018/10/01/ 2018, doi: https://doi.org/10.1016/j.jclepro.2018.06.290.

[7] J. Ma and J. C. P. Cheng, "Estimation of the building energy use intensity in the urban scale by integrating GIS and big data technology," *Applied Energy,* vol. 183, pp. 182-192, 2016/12/01/ 2016, doi: https://doi.org/10.1016/j.apenergy.2016.08.079.

[8] A. Mastrucci, P. Pérez-López, E. Benetto, U. Leopold, and I. Blanc, "Global sensitivity analysis as a support for the generation of simplified building stock energy models," *Energy and Buildings,* vol. 149, pp. 368-383, 2017/08/15/ 2017, doi: https://doi.org/10.1016/j.enbuild.2017.05.022.

[9] A. Mastrucci, O. Baume, F. Stazi, and U. Leopold, "A GIS-based statistical approach to prioritize the retrofit of housing stocks at the urban scale," *Journal of Energy Challenges and Mechanics,* vol. 3, no. 4, pp. 186-190, 2016.

[10] Intergovernmental Panel on Climate Change, *Climate change 2014: mitigation of climate change*. Cambridge University Press, 2015.

[11] J. Sachs, J. Renard, D. Moya, S. Giarola, and A. Hawkes, "Global GIS-Based District Heating and Cooling Potential Assessment for the Residential Building Sector," presented at the International Energy Workshop (IEW), Gothenburg Sweden, 2018. [Online]. Available: https://iew2018.org/wp-content/uploads/2018/07/H1_Sachs.pdf.

[12] M. Kummu, M. Taka, and J. H. Guillaume, "Gridded global datasets for gross domestic product and Human Development Index over 1990–2015," *Scientific data,* vol. 5, no. 1, pp. 1-15, 2018.

[13] M. Gallery and R. Sensing, "Gridded Population of the World (GPW), v4."

[14] S. Poppi, N. Sommerfeldt, C. Bales, H. Madani, and P. Lundqvist, "Techno-economic review of solar heat pump systems for residential heating applications," *Renewable and Sustainable Energy Reviews,* vol. 81, pp. 22-32, 2018/01/01/ 2018, doi: https://doi.org/10.1016/j.rser.2017.07.041.

[15] S. Paiho and H. Saastamoinen, "How to develop district heating in Finland?," *Energy Policy,* vol. 122, pp. 668-676, 2018/11/01/ 2018, doi: https://doi.org/10.1016/j.enpol.2018.08.025.

[16] D. Groppi, L. de Santoli, F. Cumo, and D. Astiaso Garcia, "A GIS-based model to assess buildings energy consumption and usable solar energy potential in urban areas," *Sustainable Cities and Society,* vol. 40, pp. 546-558, 2018/07/01/ 2018, doi: https://doi.org/10.1016/j.scs.2018.05.005.

[17] R. Buffat, A. Froemelt, N. Heeren, M. Raubal, and S. Hellweg, "Big data GIS analysis for novel approaches in building stock modelling," *Applied Energy,* vol. 208, pp. 277-290, 2017/12/15/ 2017, doi: https://doi.org/10.1016/j.apenergy.2017.10.041.

[18] S. Werner, "District Heating and Cooling," in *Encyclopedia of Energy*, C. J. Cleveland Ed. New York: Elsevier, 2004, pp. 841-848.

[19] Y. Zhou, J. Eom, and L. Clarke, "The effect of global climate change, population distribution, and climate mitigation on building energy use in the U.S. and China," *Climatic Change,* journal article vol. 119, no. 3, pp. 979-992, August 01 2013, doi: 10.1007/s10584-013-0772-x.

[20] M. Berger and J. Worlitschek, "A novel approach for estimating residential space heating demand," *Energy,* vol. 159, pp. 294-301, 2018/09/15/ 2018, doi: https://doi.org/10.1016/j.energy.2018.06.138.

[21] D. Connolly, "Heat Roadmap Europe: Quantitative comparison between the electricity, heating, and cooling sectors for different European countries," *Energy,* vol. 139, pp. 580-593, 2017/11/15/ 2017, doi: https://doi.org/10.1016/j.energy.2017.07.037.

[22] V. Chaturvedi, J. Eom, L. E. Clarke, and P. R. Shukla, "Long term building energy demand for India: Disaggregating end use energy services in an integrated assessment modeling framework," *Energy Policy,* vol. 64, pp. 226-242, 2014/01/01/ 2014, doi: https://doi.org/10.1016/j.enpol.2012.11.021.

[23] S. Nielsen and B. Möller, "GIS based analysis of future district heating potential in Denmark," *Energy,* vol. 57, pp. 458-468, 2013/08/01/ 2013, doi: https://doi.org/10.1016/j.energy.2013.05.041.

[24] M. Olonscheck, A. Holsten, and J. P. Kropp, "Heating and cooling energy demand and related emissions of the German residential building stock under climate change," *Energy Policy,* vol. 39, no. 9, pp. 4795-4806, 2011/09/01/ 2011, doi: https://doi.org/10.1016/j.enpol.2011.06.041.

[25] D. Ürge-Vorsatz, L. F. Cabeza, S. Serrano, C. Barreneche, and K. Petrichenko, "Heating and cooling energy trends and drivers in buildings," *Renewable and Sustainable Energy Reviews,* vol. 41, pp. 85-98, 2015.

[26] M. Isaac and D. P. van Vuuren, "Modeling global residential sector energy demand for heating and air conditioning in the context of climate change," *Energy Policy,* vol. 37, no. 2, pp. 507-521, 2009/02/01/ 2009, doi: https://doi.org/10.1016/j.enpol.2008.09.051.

[27] H. C. Gils, J. Cofala, F. Wagner, and W. Schöpp, "GIS-based assessment of the district heating potential in the USA," *Energy,* vol. 58, pp. 318-329, 2013.

[28] S. Werner, "European space cooling demands," *Energy,* vol. 110, pp. 148-156, 2016/09/01/ 2016, doi: https://doi.org/10.1016/j.energy.2015.11.028.

[29] M. Jakubcionis and J. Carlsson, "Estimation of European Union residential sector space cooling potential," *Energy Policy,* vol. 101, pp. 225-235, 2017/02/01/ 2017, doi: https://doi.org/10.1016/j.enpol.2016.11.047.

[30] J.-G. Lee and M. Kang, "Geospatial Big Data: Challenges and Opportunities," *Big Data Research,* vol. 2, no. 2, pp. 74-81, 2015/06/01/ 2015, doi: https://doi.org/10.1016/j.bdr.2015.01.003.

[31] U. Ali *et al.*, "A data-driven approach for multi-scale GIS-based building energy modeling for analysis, planning and support decision making," *Applied Energy,*

vol. 279, p. 115834, 2020/12/01/ 2020, doi: https://doi.org/10.1016/j.apenergy.2020.115834.

[32] J. Sachs, D. Moya, S. Giarola, and A. Hawkes, "Global spatially and temporally-resolved heat and cooling energy demand in the residential sector [in press]," *Applied Energy,* 2019.

[33] J. Han, J. Pei, and M. Kamber, *Data mining: concepts and techniques*. Elsevier, 2011.

[34] M. Afaifia, K. A. Djiar, N. Bich-Ngoc, and J. Teller, "An energy consumption model for the Algerian residential building's stock, based on a triangular approach: Geographic Information System (GIS), regression analysis and hierarchical cluster analysis," *Sustainable Cities and Society,* vol. 74, p. 103191, 2021/11/01/ 2021, doi: https://doi.org/10.1016/j.scs.2021.103191.

[35] C. Deb, M. Frei, J. Hofer, and A. Schlueter, "Automated load disaggregation for residences with electrical resistance heating," *Energy and Buildings,* vol. 182, pp. 61-74, 2019/01/01/ 2019, doi: https://doi.org/10.1016/j.enbuild.2018.10.011.

[36] N. Voulis, M. Warnier, and F. M. T. Brazier, "Understanding spatio-temporal electricity demand at different urban scales: A data-driven approach," *Applied Energy,* vol. 230, pp. 1157-1171, 2018/11/15/ 2018, doi: https://doi.org/10.1016/j.apenergy.2018.08.121.

[37] F. McLoughlin, A. Duffy, and M. Conlon, "A clustering approach to domestic electricity load profile characterisation using smart metering data," *Applied Energy,* vol. 141, pp. 190-199, 2015/03/01/ 2015, doi: https://doi.org/10.1016/j.apenergy.2014.12.039.

[38] P. Gianniou, X. Liu, A. Heller, P. S. Nielsen, and C. Rode, "Clustering-based analysis for residential district heating data," *Energy Conversion and Management,* vol. 165, pp. 840-850, 2018/06/01/ 2018, doi: https://doi.org/10.1016/j.enconman.2018.03.015.

[39] Y. Zhang, J. Mańdziuk, C. H. Quek, and B. W. Goh, "Curvature-based method for determining the number of clusters," *Information Sciences,* vol. 415-416, pp. 414-428, 2017/11/01/ 2017, doi: https://doi.org/10.1016/j.ins.2017.05.024.

[40] J. Unternährer, S. Moret, S. Joost, and F. Maréchal, "Spatial clustering for district heating integration in urban energy systems: Application to geothermal energy," *Applied energy,* vol. 190, pp. 749-763, 2017.

[41] R. Tibshirani, G. Walther, and T. Hastie, "Estimating the number of clusters in a data set via the gap statistic," *Journal of the Royal Statistical Society: Series B (Statistical Methodology),* vol. 63, no. 2, pp. 411-423, 2001.

[42] J. A. Hartigan and M. A. Wong, "Algorithm AS 136: A k-means clustering algorithm," *Journal of the Royal Statistical Society. Series C (Applied Statistics),* vol. 28, no. 1, pp. 100-108, 1979. [Online]. Available: https://www.labri.fr/perso/bpinaud/userfiles/downloads/hartigan_1979_kmeans.pdf.

[43] J. Sachs, S. Giarola, and A. Hawkes, "Agent-based model for energy-related investment decisions in the residential building sector," 2017: International Energy Workshop.

[44] D. Moya, S. Budinis, S. Giarola, and A. Hawkes, "Agent-based scenarios comparison for assessing fuel-switching investment in long-term energy transitions of the India's industry sector," *Applied Energy,* vol. 274, p. 115295, 2020/09/15/ 2020, doi: https://doi.org/10.1016/j.apenergy.2020.115295.

[45] PARIS REINFORCE project. "The ModUlar energy system Simulation Environment (MUSE)." The I2AM PARIS platform. https://paris-reinforce.epu.ntua.gr/detailed_model_doc/muse (accessed 19 October, 2021).

[46] B. C. O'Neill *et al.*, "The roads ahead: Narratives for shared socioeconomic pathways describing world futures in the 21st century," *Global environmental change,* vol. 42, pp. 169-180, 2017.

[47] F. Knobloch, H. Pollitt, U. Chewpreecha, V. Daioglou, and J.-F. Mercure, "Simulating the deep decarbonisation of residential heating for limiting global warming to 1.5 C," *Energy Efficiency,* vol. 12, no. 2, pp. 521-550, 2019.