Advances in Science and Technology Research Journal

Advances in Science and Technology Research Journal, 2027, 21(1), 128–145
https://doi.org/10.12913/22998624/231208
ISSN 2299-8624, License CC-BY 4.0

Received: 2026.07.22
Accepted: 2026.11.15
Published: 2026.12.01

Machine learning-based algorithm for estimating energy efficiency of photovoltaic panels in public transport: Case study of Lublin

Marcin Matczuk[1*], Łukasz Grabowski[2], Grzegorz Kozieł[1], Marta Dziuba-Kozieł[1], Dominik Strzałka[3], Piotr Andrzej Jakliński[2], Kamil Żyła[1], Jacek Kęsik[1]

[1] Department of Computer Science, Lublin University of Technology, Nadbystrzycka 36B, 20-618 Lublin, Poland
[2] Department of Thermodynamics, Fluid Mechanics and Aviation Propulsion, Lublin University of Technology, Nadbystrzycka 36, 20-618 Lublin, Poland
[3] Department of Complex Systems, Rzeszów University of Technology, Al. Powstańców Warszawy 12, 35-029 Rzeszów, Poland
* Corresponding author's e-mail: m.matczuk@pollub.pl

ABSTRACT

The study investigates the energy efficiency of photovoltaic (PV) panels installed on public transport buses. The focus is on assessing urban routes for their solar energy harvesting potential and the impact of generated solar energy on fuel consumption. The analysis is based on real-world data collected from the public transport system in the city of Lublin, Poland. Due to the prevalence of GPS signal loss, particularly in dense urban environments, a machine learning-based algorithm was developed to enhance bus trajectory reconstruction. The proposed methodology comprises four main stages: (1) filtering and processing GPS data to reconstruct bus routes, (2) augmenting the dataset by computing additional metrics and factors from the original dataset, (3) dividing bus routes into hexagonal spatial units and calculating average metric values for each region, and (4) identifying areas with higher shading factors and solar energy potential for each spatial unit. The reconstruction stage uses anomaly detection techniques to mitigate the effects of incomplete or noisy GPS signals. The results indicate that several areas in Lublin are suboptimal for installing PV panels, that energy generated by roof-mounted solar panels positively influences the vehicle's energy balance, and that the issue raised in the article could help improve urban planning. Furthermore, the proposed approach facilitates the identification of spatial patterns affecting PV performance and supports data-driven decision-making for planning energy-efficient transport systems.

Keywords: energy efficiency, photovoltaic panels, machine learning, isolation forest, algorithm, public transport, spatial distribution, urban planning.

INTRODUCTION

The continuous and inevitable development of technology is a double-edged sword, offering benefits but also presenting challenges [1,2]. Life is improving in quality, humans are exceeding their limits, and more and more processes are being automated. However, these phenomena also have negative side effects that frequently impact the environment [3].

Nowadays, it is no longer possible for humanity to focus solely on profit. It is crucial to thoroughly analyse the environmental impact of technological developments [4]. This issue pertains, inter alia, to the development of the automotive and transport sectors [5]. Currently, the majority of people own a

car, which significantly facilitates communication and travel. However, vehicles also contribute to air pollution and traffic congestion, causing people to spend increasing amounts of time in traffic.

Fortunately, advancements in renewable energy technologies have provided solutions that help mitigate the aforementioned issues. For example, photovoltaic panels (PV) can be installed on the roofs of vehicles, particularly public transport vehicles, as they operate on fixed routes and under relatively stable conditions. Consequently, additional energy can be generated, making the time spent in traffic congestion more productive.

This paper investigates the energy efficiency of photovoltaic panels installed on public transport buses, with the objective of assessing bus routes in terms of their potential for maximising photovoltaic energy harvesting. The scope of the study also includes the detection of areas exhibiting a high shading factor under clear-sky conditions. These areas are important for urban planning and the placement of PV panels [6,7].

The research is based on data collected from the public transport system in Lublin, Poland. Similar telemetry-based analyses of propulsion system operating conditions have previously been applied in aviation research [8]. The methodology consists of four main steps: (1) filtering and processing GPS data to reconstruct bus routes, (2) data augmentation through the computation of additional metrics and factors based on the original dataset, (3) division of bus routes into hexagonal spatial units and calculation of average metric values for each region, and (4) identification of areas characterized by the highest shading factor.

The first step in the analysis process is to reconstruct the bus route using GPS data. However, this data contains many imprecise samples due to the frequent loss of GPS signal in urban canyons. To overcome this problem, a machine learning–based authorial algorithm was devised. The filtering procedure consists of multiple stages, one of which uses Isolation Forest (IF), an unsupervised learning model designed to detect anomalies in datasets [9]. IF is a popular and efficient algorithm because it does not require a reference set, such as those used in Bayesian filters [10], and operates directly on the data.

The present study proposes a novel interdisciplinary and geographically specific framework that integrates GPS trajectory reconstruction with an assessment of the energy performance of vehicle-mounted photovoltaic (PV) systems.

Previous assessments of photovoltaic potential along bus routes have used, inter alia, simulations of shadows cast on roads and GIS-based three-dimensional urban models [11], fixed-point fisheye imagery, mobile sensors and street-view panoramas [12]. In contrast, the proposed framework uses real-world data and combines operational routes with segment-level measurements of PV power output and fuel consumption. Moreover, to the best of the authors' knowledge, no comparable integrated analysis has previously been conducted for the study region.

Another methodological contribution is a GPS trajectory reconstruction procedure designed for severely corrupted data with clusters of invalid observations. In such cases, the standard Isolation Forest algorithm may perform poorly because it assumes that anomalies are relatively few and distinguishable from normal observations [13]. Before IF is applied, several simple preprocessing algorithms are used to remove or reduce the most problematic error patterns. The novel contribution is therefore a lightweight, modular processing pipeline adapted to the error patterns observed in operational bus data.

The findings indicate that several areas in Lublin are suboptimal for installing PV panels, that energy generated by roof-mounted solar panels positively influences the vehicle's energy balance, and that the issue raised in the article could help improve urban planning. This work provides localised insights necessary for optimising sustainable transit infrastructure in temperate urban environments.

RELATED WORKS

Over the past decade, the declining cost of photovoltaic (PV) technology has driven its widespread adoption across various sectors, including public transportation. Existing literature in this domain primarily focuses on urban infrastructure, such as assessing the effectiveness of PV panels at public transport stops [14] or analysing their integration with building energy utilisation [15]. Concurrently, optimisation of the underlying technology remains a prominent research avenue. This paragraph mentions a few of the initiatives.

In the town of Gillingham, United Kingdom, a study was conducted on the optimisation of the energy efficiency in electric buses with curved surface solar panels [16]. According to simulations, curved PV panels generated approximately 18% more energy than flat panels under similar conditions.

Other technical interventions, such as incorporating 1 mm and 2 mm copper layers into heat sinks, have been shown through simulations to improve PV panel efficiency by 0.1% to 0.2%, depending on environmental conditions [17].

In paper [18] the possible integration of vehicle-integrated photovoltaic systems in urban public transportation was considered. It can reduce the reliance on the electrical grid, but the irradiance variability and partial shading pose significant challenges to system stability and efficiency.

Globally, the deployment of photovoltaic systems in public transportation is heavily influenced by regional meteorological conditions. For example, a feasibility study conducted in Barranquilla, Colombia, investigated the use of public transport bus-mounted PV panels designed to offset the high energy demands of air conditioning systems in an environment where daily temperatures regularly exceed 30 °C [19].

For smaller-scale transport systems, for instance, a university campus transport system, paper [20] shows that such a solution is a viable option, but only for a given location with considerable amounts of solar radiation year-round.

The authors searched the major bibliographic databases, including Google Scholar, Scopus, and Web of Science, for similar solutions that address the stated objectives. Some of the identified studies were related, as they focused on the utilisation of photovoltaic systems in public transportation – for example, the study conducted in Barranquilla, Colombia [19]. Other works provided valuable insights into the domain of the issue, such as a study conducted in the town of Gillingham, United Kingdom [16]. Nevertheless, no works similar to the authors' were found. Thus, the issue under investigation and the case of the city of Lublin are unique.

MATERIALS AND METHODS

Procedure overview

To achieve the research objective of estimating the energy efficiency of photovoltaic panels in public transport, the study was conducted in four main steps. The process begins with defining new features, metrics, and indicators derived from existing telemetry data. Detailed descriptions of these indicators are provided in the section titled "Energy efficiency assessment indicators".

The procedure then proceeds to the GPS data filtering stage, as the raw GPS data contains numerous imprecise and invalid samples. In the *GPS data filtering* section, a multistage algorithm developed by the authors was applied to mitigate this issue.

The subsequent section, *Spatial mapping*, describes a method for reconstructing a bus route using a hexagonally partitioned spatial map. During the GPS data filtering stage, some samples are discarded; however, these samples contain only invalid GPS data, while all remaining features are valid. Therefore, the *Spatial mapping* section also presents a method for mapping all data samples onto the spatial map.

The final stage focuses on calculating energy-efficiency metrics for individual segments of the bus route. For each hexagonal segment, the mean values of the metrics presented in the section *Data enhancement* were calculated. Furthermore, the analysis identifies regions characterised by the highest shading factors, i.e., areas where the performance of photovoltaic panels is adversely affected by shading. The study's findings are reported and discussed in the *Results and Discussion* chapter.

The presented study is based on two assumptions: only the period when the engine was operational is considered, and samples recorded in the early morning and late evening are excluded. This is related to the low solar radiation intensity during these periods.

Furthermore, the examined buses were equipped with PV panels mounted in a fixed orientation parallel to the roof, with no provision to adjust their tilt angle. Consequently, the panel orientation could not be actively optimised during operation. The study, therefore, evaluates the performance of the existing vehicle-mounted configuration and does not consider a theoretically optimal PV orientation.

In the research, the term "route" refers to the physical road sections traversed by a bus — both directions. The results presented in the *Results and Discussion* section were calculated using the mean values of the introduced metrics obtained from data collected during the analysed period, from late morning to early evening. Thus, the analysis covers multiple bus runs on the same route within the same day.

Data description

The present article is the result of an extensive analysis of real-world data collected from public

transport vehicles in Lublin, Poland. Among the tested objects were buses that operated on various routes, covering both suburban and city centre areas. The data was collected over the course of a year and stored in its original, unprocessed form as raw telemetry data. The sampling frequency was set at two samples per second (2Hz), and samples were only collected while the bus was operational.

Table 1 presents all monitored parameters and their descriptions as they appear in the original raw data. However, this article focuses on a subset relevant to the study's subject. The parameters essential to the study include GPS data, the solar panel's output voltage and current, the battery and alternator currents, and fuel consumption. In the following section on bus route reconstruction, speed is also used to improve the results of the Isolation Forest algorithm.

Energy efficiency assessment indicators

To analyse the energy efficiency of photovoltaic (PV) panels installed on heavy-duty vehicles, such as public transport buses, it was necessary to derive a set of primary indices and performance coefficients from the raw telemetry data. The first fundamental metric defined was the aggregate solar power output (P_{solar}). This quantity represents the total instantaneous power generated by the roof-mounted array and is calculated as follows:

$$P_{solar} = \sum_{k} (U_k^o \cdot I_k^o) \tag{1}$$

Equation 1 is relatively straightforward. It is merely a sum of products of current (I) and voltage (U). The upper index, denoted by the letter 'o', is used to denote the output value from the kth panel, where index k indicates the specific panel in the sequence. The P_{solar} metric was calculated for each sample in the dataset.

It should be emphasised that P_{solar} does not represent the theoretical maximum power output of the PV modules. Instead, it represents the instantaneous aggregate electrical power actually delivered by all PV panels under the prevailing operating and environmental conditions, based on the measured output voltage and current of each panel.

While P_{solar} quantifies the power supply, it must be contextualised against the vehicle's total power demand to determine the effective efficiency gain. Consequently, an additional quantity of total electrical power demand (P_{elec}) was calculated (2).

$$P_{elec} = (I_{alt_1} + I_{alt_2}) \cdot U \tag{2}$$

The total electrical power demand generated by the vehicle's conventional charging system is derived from the alternator outputs. Here, I_{alt_1} and I_{alt_2}, denote the instantaneous current values from the bus's primary alternators, and U represents the instantaneous internal bus system voltage. Among the buses from which the data were collected were vehicles with two alternators and those with a single alternator. In the latter case, the variables I_{alt_1} and I_{alt_2} were replaced by I_{alt}.

Given the discrete power data, the corresponding energy consumption can be readily calculated as $E = \Delta t \cdot \sum_i P_i$, since the constant sampling frequency (2 Hz) results in equal time intervals between consecutive measurements. The duration for which each bus was operational was recorded and stored in the data.

The next step in the analysis was to determine the relative contribution of the PV system to the total electrical demand of the vehicle. Given the aggregate solar power P_{solar} and the alternator power P_{elec}, the solar displacement ratio (SDR) was defined as the proportion of the total electrical load satisfied by solar energy:

$$SDR = \frac{P_{solar}}{P_{solar} + P_{elec}} \tag{3}$$

The metric established in Equation 3 serves as a primary coefficient for the subsequent spatial and statistical analyses. By normalising the solar contribution against the total instantaneous electrical load, the SDR enables an objective comparison of PV performance across diverse operational conditions, independent of the absolute magnitude of the vehicle's electrical load.

The subject of the research was Lublin's public buses. A considerable proportion of these buses' operation is spent in the idle stage (engine running at idle speed), particularly when traversing high-traffic areas. The bus is known to idle at intersections, in daily congestion, and at bus stops. However, when a bus is idling in a sunlit location with no shade, the presence of photovoltaic panels allows this time to be utilised productively. Although the vehicle remains stationary, the photovoltaic panels continue to generate electrical power. It is therefore essential to include the idling stage in the analysis. To accomplish the aforementioned objective, another quantity was introduced.

Table 1. List of data channels in the telemetry dataset

Name	Description
I_AK	Battery current
I_ALT	Main alternator current
I_ALT_1	Additional alternator current
I_ALT_2	Additional alternator current
U	Bus voltage
E	Energy
Przebieg	Odometer
Predkosc	Speed
Czas pracy	Engine runtime
TPS	Throttle position
T_silnika	Engine temperature
T_zew	Outdoor temperature
n	Engine RPM
Poziom_ON	Fuel level
l/h	Fuel consumption (litres per hour)
km/l	Fuel consumption (kilometres per litre)
Total_zu_ON	Total fuel consumption
LATITUDE	GPS latitude
LONGTITUDE	GPS longitude
U_we_{i}	Input voltage of the i-th solar panel
U_wy_{i}	Output voltage of the i-th solar panel
I_we_{i}	Input current of the i-th solar panel
I_wy_{i}	Output current of the i-th solar panel

$$EII = \frac{Fuel\ consumption}{Speed + \epsilon} \tag{4}$$

Equation 4 is the defining formula for the energy inefficiency index (EII). The symbol $\in$ represents a small constant (e.g., 1.0) used to avoid division by zero. When the vehicle was idling at a crossroads or crawling in a traffic jam, its speed was low while fuel consumption was high. Consequently, the EII value increased during these periods. The fuel consumption and speed values were derived from raw telemetry data and expressed in litres per hour and kilometres per hour, respectively. Hence, the unit of measurement for EII is litres per kilometre; however, it was omitted as it was not relevant to the subsequent analysis.

The analysis of the energy efficiency of photovoltaic (PV) panels in public transport is based on fuel consumption. The available telemetry data indicate that buses may be equipped with either a single alternator (denoted as I_{alt}) or two alternators (I_{alt_1}, I_{alt_2}), depending on the vehicle configuration. These alternators supply electrical power to onboard systems, including display monitors, Wi-Fi, USB chargers, door mechanisms, lighting and air conditioning.

In the absence of a PV system, or during nighttime operation, the electrical demand is supplied entirely by engine-driven alternators, which increases the mechanical load on the engine and, consequently, fuel consumption. When a PV system is present, part of the auxiliary electrical demand may be supplied by the PV panels, thereby reducing the load on the alternator(s). This reduction is expected to translate into lower fuel consumption, although the relationship may be indirect and dependent on operating conditions.

The analysis is therefore conducted at the system level, using measured electrical and fuel consumption data, without explicitly modelling the vehicle's internal energy management strategy.

A pivotal factor in establishing the relationship between fuel consumption and electrical load was the article by Grabowski et al. [21]. Their study was conducted on Lublin's public transport buses – the same vehicles used to collect data for this study. The aforementioned paper determines the precise amount of fuel (diesel) required to generate 1 kWh of electrical energy. The average conversion factor, G = 0.364 liters/kWh, was determined by Grabowski et al. and represents the amount of fuel required to generate 1 kWh of electrical energy.

Using the G factor, the gross amount of fuel used in a hypothetical scenario involving buses without the PV system was computed (5).

$$Fuel\ gross = Fuel\ consumption + \\ +(P_{solar} \cdot G) \tag{5}$$

It is important to note that the fuel gross quantity is theoretical; however, this does not significantly affect the remainder of the analysis. Next, a new metric, denoted EII_{gross}, was computed. The symbol $\in$ in (6) represents a small constant (e.g., 1.0) used to avoid division by zero.

$$EII_{gross} = \frac{Fuel\ gross}{Speed + \epsilon} \tag{6}$$

Equation 6 differs from Equation 4 only in the numerator, and their difference defines the theoretical energy savings attributable to the presence of PV panels (7).

$$EII_{gain} = EII_{gross} - EII \tag{7}$$

It can be deduced that $EII_{gross} \geq EII$; therefore, $EII_{gain} \geq 0$ always holds.

The last quantity introduced was the solar efficiency gain (SEG) index (8). It measures the theoretical efficiency gain due to the presence of PV panels. In other words, it quantifies how much the PV system reduces engine load, resulting in a hypothetical reduction in fuel consumption.

$$SEG = \frac{EII_{gain}}{EII_{gross} + \epsilon} \cdot 100 \tag{8}$$

The SEG value in Equation 8 is expressed as a percentage and is expected to be very small. The energy required to move the bus and keep it in motion is extremely greater than the energy generated by the PV system, even on a sunny summer day.

Given that a bus traverses the same route repeatedly, the amount of data collected per route is large. In the following chapters, each bus route is divided into spatial segments, and the mean values of the metrics established in this section are calculated for each segment. Although the introduced metrics are based on theoretical assumptions, the large volume of available data mitigates the impact of these assumptions. It should be emphasised that in the research, the term "route" refers to the physical road sections traversed by a bus. Thus, the analysis covers all bus runs on the same route on the same day.

GPS data filtering

The aim of this section is to present a novel, author-developed algorithm based on machine learning for filtering high-noise and imprecise GPS data. This issue is evident in our dataset. The presence of urban canyons and poor road conditions contributes to GPS signal loss or sudden jumps to random locations. It is worth noting that the GPS data in our dataset consists of two values: latitude and longitude.

The filtering process begins by overlaying a bounding box over Lublin, as the recorded GPS signal may indicate a location in another country. The bounding box is defined by two opposite points: (51.13977986048797, 22.450585637634955) and (51.292449565797476, 22.68458320457082). Any GPS sample that lies outside the bounding box is rejected. Figure 1 shows a bus route plotted based on unfiltered GPS data.

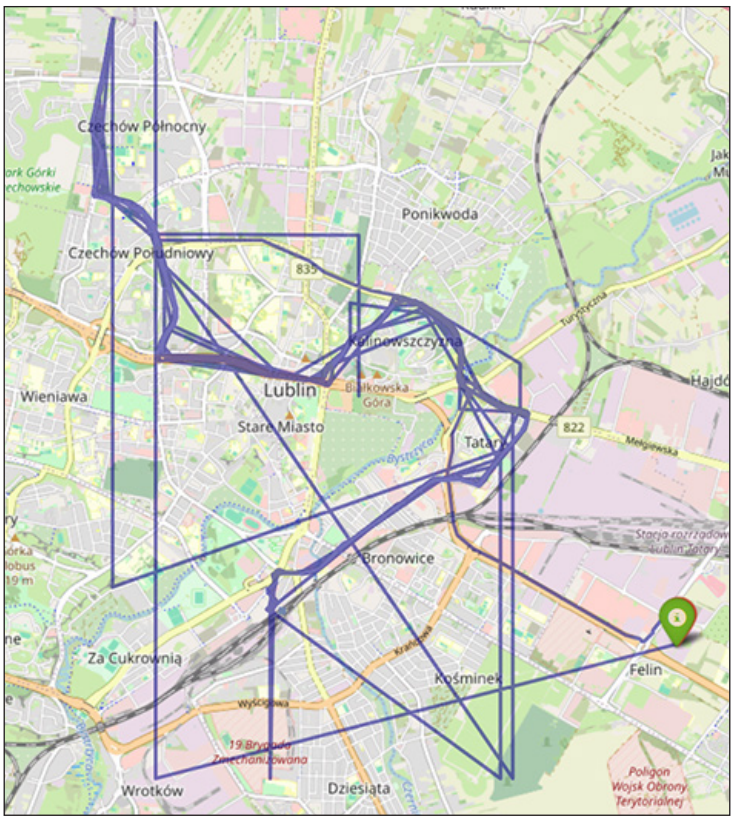


Figure 1. Lublin's public bus route plotted using unfiltered data
Only a bounding box constraint was applied.
The figure in its original resolution is provided in the supplementary materials.

As shown in Figure 1, the route is relatively clearly delineated. The main issue is the presence of noise, i.e., imprecise GPS locations. However, due to the high sampling frequency and the repetitive nature of the route, the visible outlier points form clusters rather than single points. Furthermore, the telemetry data are spatially and temporally dependent, as the samples were collected concurrently with the movement of the bus (for clarity, also during idle periods).

Removing samples containing invalid GPS data is not a viable solution, as it would result in the loss of other valuable variables. In summary, the objective is to reconstruct the route, create a spatial map, and ultimately relate imprecise points to it.

Following the bounding box step, the Last Known Good Point (LKGP) algorithm is applied. The fundamental principle underlying LKGP is the calculation of implied speed based on the haversine distance [22]. If the implied speed exceeds a predefined maximum (e.g., 65 km/h), the subsequent point is deemed invalid in terms of localisation. Listing 1 presents the pseudocode of the LKGP algorithm.

It is evident that the LKGP algorithm removes samples. However, at this stage, the objective is to reconstruct the bus route. Later, a methodology is employed to redistribute the discarded samples along the route.

According to Listing 1, computing the implied speed requires the time and distance values. The latter is straightforward to obtain — the haversine distance can be used — but the former is more problematic. In the dataset, there is no record of the time at which a sample was collected. Therefore, the best approach is to rely on indices, as the data is sequential.

```
Input:
points: GPS points in the form of (latitude, longitude)
max_speed: predefined maximum speed
rate: sampling rate–seconds per sample
Output: an array storing samples with valid GPS locations
Data:
lkgp_idx = 0: index of the last known good point
valid_idxs = [0]: array of valid indices
1  Function lkgp_filter (points, max_speed, rate):
2  for i = 1 to size(points):
3     lat1, lon1 = points[lkgp_idx]
4     lat2, lon2 = points[i]
5     d = haversine_distance(lat1, lon1, lat2, lon2)
6     time = (i – lkgp_idx) * rate / 3600
7     implied_speed = (d / time) if time > 0 else 0
8     if implied_speed <= max_speed then
9        append index i to the valid_idxs array
10       lkgp_idx = i
11 return points[valid_idxs]
```

Listing 1. Last Known Good Point algorithm

Given that the sampling frequency is known, the difference between the index of the last known good point and that of the currently processed sample is multiplied by the reciprocal of the frequency. The resulting value is then divided by 3600 to express the time in hours rather than seconds. If the GPS signal suddenly deviates from the bus's actual location, the computed distance travelled increases, while the time interval between consecutive samples remains small, resulting in an abnormally high speed. Figure 2, panel (a), depicts the result of applying the aforementioned LKGP algorithm. The value of max_speed (Listing 1) was set to 65 km/h, as certain sections in Lublin permit speeds of up to 70 km/h. It is evident that most of the noise has been eliminated, but further improvements are still possible.

The penultimate step in the process of filtering GPS data is the application of a standard median filter [23]. The methodology is as follows: the median point is selected from x consecutive GPS points, and the remaining points are discarded. Given the sequential nature of the data, a median filter is a reasonable choice. The value of x = 15was chosen empirically. The final results are presented in Figure 2, panel (b), and in Figure 3.

The overall visible effect of the median filter is modest. Figure 3 presents magnified views of two different regions. It is worth noting that the current objective is to reconstruct the route and convert it into spatial regions.

The Isolation Forest (IF) algorithm constitutes the final component of the proposed filtering method. As previously noted, outliers in unfiltered GPS data tend to cluster; consequently, IF does not yield satisfactory results when applied directly to raw data. Therefore, the preceding steps are essential.

Prior to applying IF, a data augmentation step is performed [24,25]. The augmented dataset is derived from the three aforementioned steps. Two additional features are defined: *jump_dist*, representing the Euclidean distance between consecutive GPS points, and *bearing*, defined as the angle between vectors with a common origin formed by consecutive GPS points (9).

$$bearing = arctan2(\Delta latitude, \Delta longitude) \quad (8)$$

The interpretation of jump_dist is straightforward, and the Euclidean distance is used instead of the haversine distance, since the two differ little over small regions [26]. Moreover, the former is computationally less expensive. The bearing represents the angle of the movement vector.

The introduction of the bearing and jump_dist features enhances the ability of Isolation Forest to detect anomalous GPS points – those that appear suddenly in an unexpected direction or at an unusual distance from the previous point, likely due to GPS noise or measurement errors. In summary, the following set of features constitutes the input for the Isolation Forest algorithm:

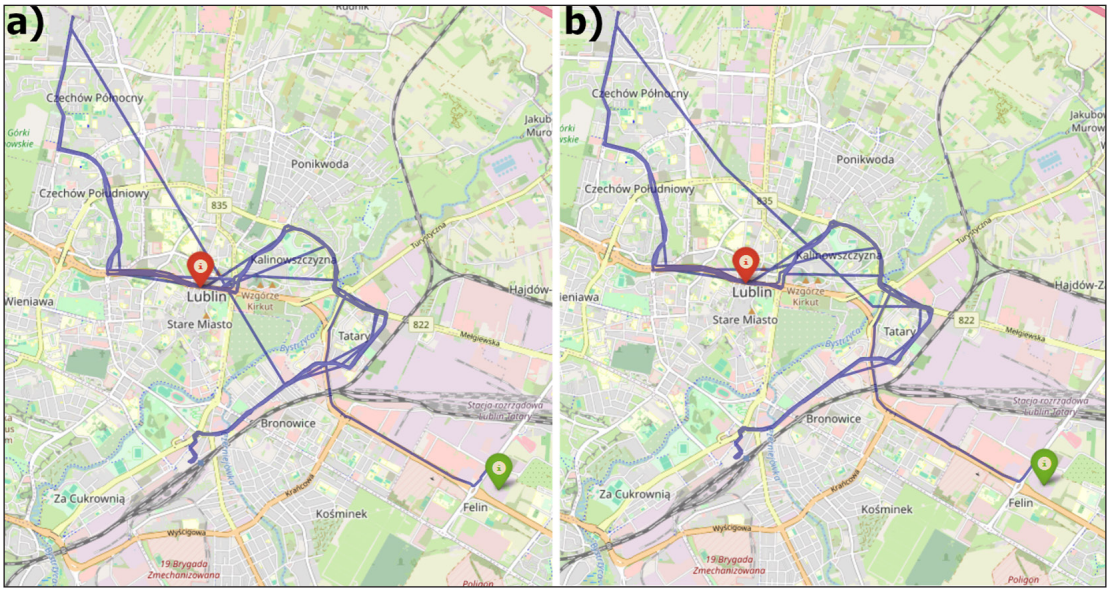


Figure 2. Lublin's public bus route plotted using partially filtered data; in panel (a), the bounding box constraint and the LKGP algorithm were applied, in panel (b), the bounding box constraint, the LKGP algorithm, and the median filter were applied. The figure in its original resolution is provided in the supplementary materials

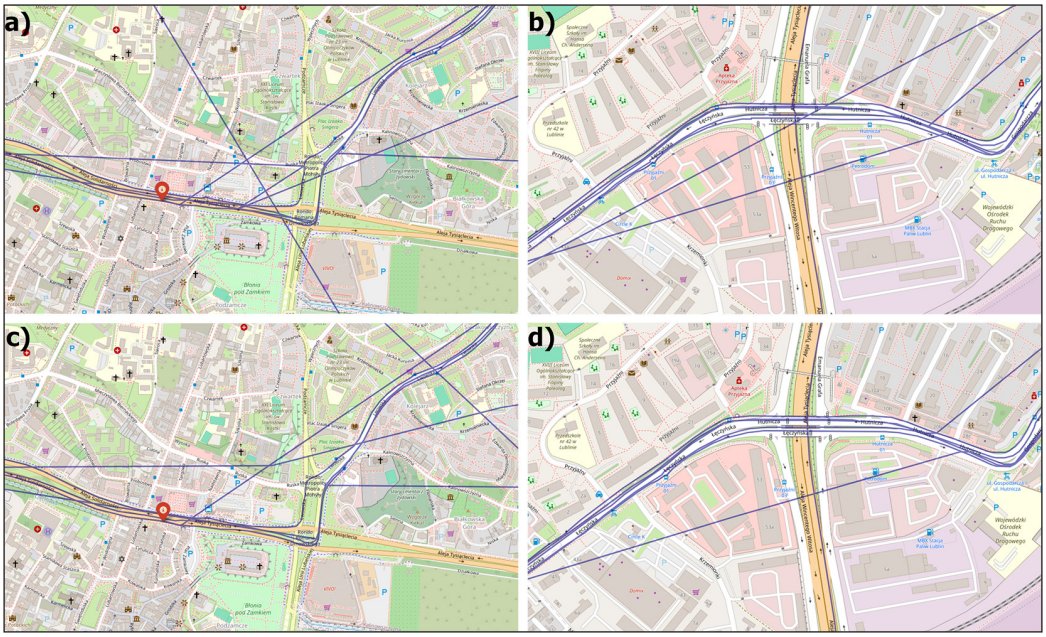


Figure 3. Effect of the median filter
Panels (a) and (b) show the pre-filter state, while panels (c) and (d) show the post-filter state. The figure in its original resolution is provided in the supplementary materials

latitude, longitude, speed, jup_dist, and bearing. It is important to note that not all parameters listed in Table 1 are used by IF.

Each sample designated as an anomaly by IF is discarded. The final result of the filtering process is illustrated in Figure 4. Some invalid points remain in the data, but they are negligible. At this stage, it is possible to divide the route into spatial regions with high accuracy.

Spatial mapping

The estimation of photovoltaic (PV) energy efficiency in public transport is based on previously computed metrics and spatial segmentation. This means that the bus route is divided into spatial segments (hexagons), and each sample is assigned to a specific segment. The mean values of the metrics are then computed. Figure 5 illustrates the expected outcome. To create a map such as the one shown in Figure 5, the Python library *Folium* and Uber's *H3* hexagonal hierarchical geospatial indexing system were used.

H3 allows each geographic coordinate to be assigned to a predefined-size hexagonal segment. Thus, a hexagon ID is assigned to each sample in the GPS-filtered dataset, thereby creating a new column called h3_id. Next, a set of unique hexagon IDs is created; this set serves as the reference set of valid hexagons in the following steps.

One of the adjustable parameters in *H3* is the resolution, which determines the size of the hexagonal cells. A resolution of 11 was selected as a compromise between spatial granularity and map readability. Since the spatial maps represent discrete data, the metric values near the cell boundaries may be problematic. One possible solution would be to use a higher resolution, resulting in smaller hexagonal cells.

The subsequent stage involves incorporating all available data. That is to say, the objective is to snap all samples from the raw telemetry dataset to the map. Therefore, the following steps are performed. First, the metrics P_{solar} (1), SDR (3), EII (4), fuel_gross (5), EII_{gross} (6), and SEG (8) are calculated for each sample. Second, each sample is assigned a hexagon ID. Third, a Boolean flag (true/false) is assigned to each sample, indicating whether the sample lies on the route, thereby creating a new column named in_route. This is determined by checking whether the hexagon ID belongs to the reference set of valid hexagons. The two aforementioned steps are illustrated in Figure 6 as a flowchart.

At this point, the majority of samples have been snapped to the map. Nonetheless, there are still data points with invalid GPS localisation but valid other features in the processed dataset. The goal is to include them in the analysis as well. Therefore, another stage of the mapping process is performed.

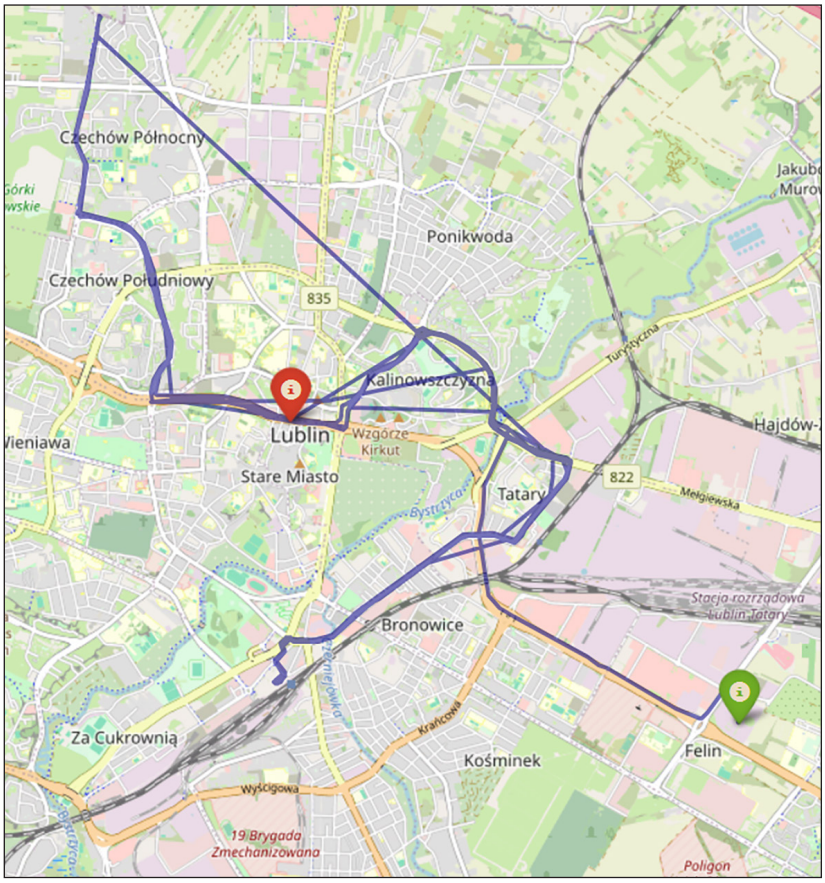


Figure 4. Lublin's public bus route plotted using filtered data
The bounding box constraint, the LKGP algorithm, the median filter, and the Isolation Forest algorithm were applied. The figure in its original resolution is provided in the supplementary materials.

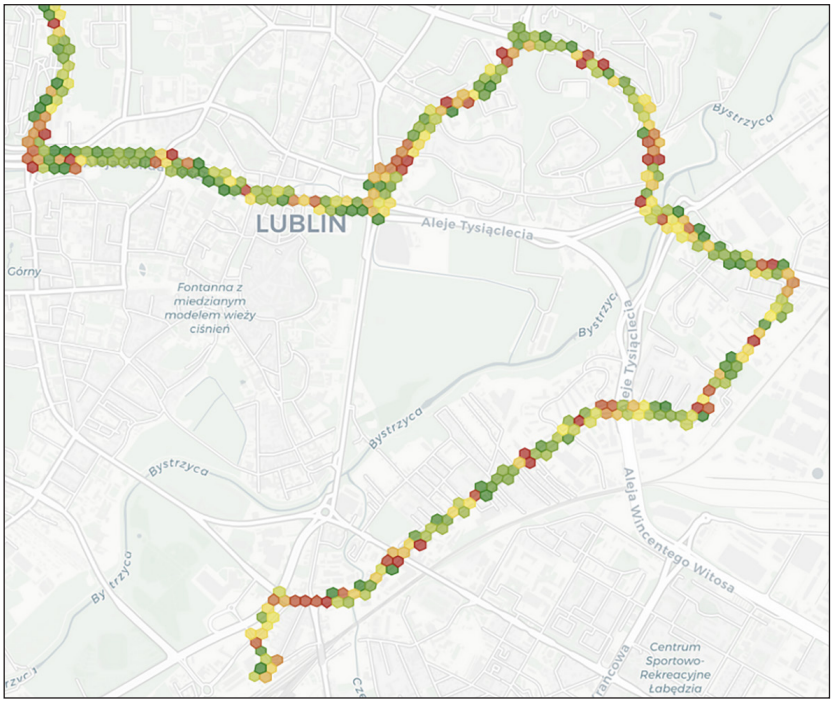


Figure 5. Example of the spatial segmentation of a bus route
The meaning of the colours is here irrelevant. The figure in its original resolution is provided in the supplementary materials.

The aim is to identify gaps in the dataset based on the Boolean flag named in_route. As has been noted on multiple occasions, the data is sequential. Hence, the process begins with the first element and proceeds sequentially through all samples. If the value of in_route is true, nothing happens; otherwise, the beginning of a gap is identified. Subsequently, the iteration continues until the first sample with the in_route flag set to true is encountered; this marks the end of the gap. The first valid point preceding and following the gap is called an anchor. The size of the gap is expressed as the number of samples between anchors (the data is sequential). Figure 7 illustrates this concept.

Once the anchors have been identified and the gap size is known, the final step is to determine the hexagon IDs and the number of hexagons between the anchors. Let N denote the size of the gap, and let K denote the number of hexagons between the anchors. Then, N samples with invalid GPS localisation are uniformly distributed over K hexagons. This is the procedure by which all raw telemetry data is included in the study.

RESULTS AND DISCUSSION

The presented study is based on two additional assumptions: only the period during which the engine was operational is considered, and samples recorded in the early morning and late evening are excluded. This is related to the low solar radiation intensity during these periods. The value of P_{solar} is used as a predictor.

For each hexagonal segment, the mean values of the metrics presented in the section *Data Enhancement* were calculated. The number of samples assigned to each hexagonal segment is determined by the spatial distribution of samples (GPS data) along the route. Furthermore, outliers – i.e., values falling below the 5th percentile and above the 95th percentile – were removed. Outliers are frequently the result of temporary measurement errors and may hinder further analysis.

Energy inefficiency index

The first element to be discussed is the spatial map of the EII metrics. Figures 8 and 9 present the results for two different bus routes in two different months. September and October were selected because solar irradiance is lower than in summer, yielding more informative results. It is well known that photovoltaic (PV) installations generate the highest amount of solar energy between June and September in Poland [27,28].

The maps indicate sections of the routes where buses consume fuel while travelling at very low speeds, typically at crossroads, bus stops, or congested areas near the city centre (e.g., *Narutowicza Street*). Such situations are classified as energy-inefficient; therefore, the EII indicator tends to be high in these regions. Although Figures 8 and 9 depict different routes and different months, the same locations are characterised by high EII values. It is important to note that Figures 8 and 9 also include the route from the bus depot to the bus line. The EII value is relative, and the figures do not imply that the red segments are extremely problematic compared to the green ones. Rather, the red segments indicate areas that are more energy-inefficient than the green segments, although the degree of this inefficiency is not explicitly defined. Moreover, the colour scale was used primarily to improve the visibility and interpretability of the results.

To summarise the preceding discussion, it can be concluded that the overall efficiency of the bus is negatively influenced by the presence of crossroads, bus stops and congested areas, i.e. periods when the bus idles. However, the gist is that photovoltaic panels can contribute positively to the energy balance.

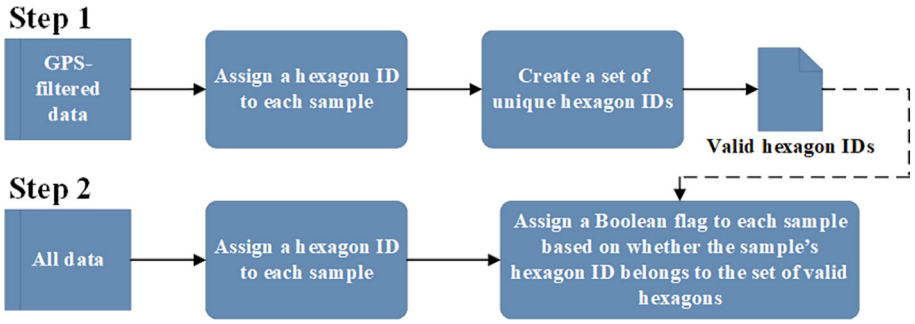


Figure 6. Diagram of the first two steps in the process of snapping data samples to the map

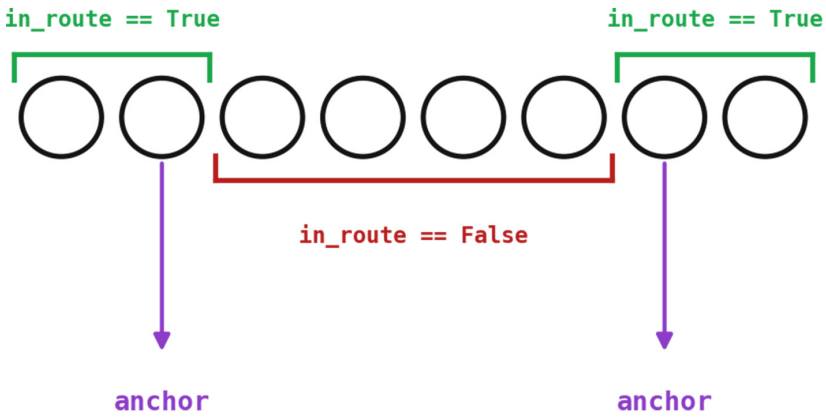


Figure 7. Schematic illustration of a GPS position gap in the dataset

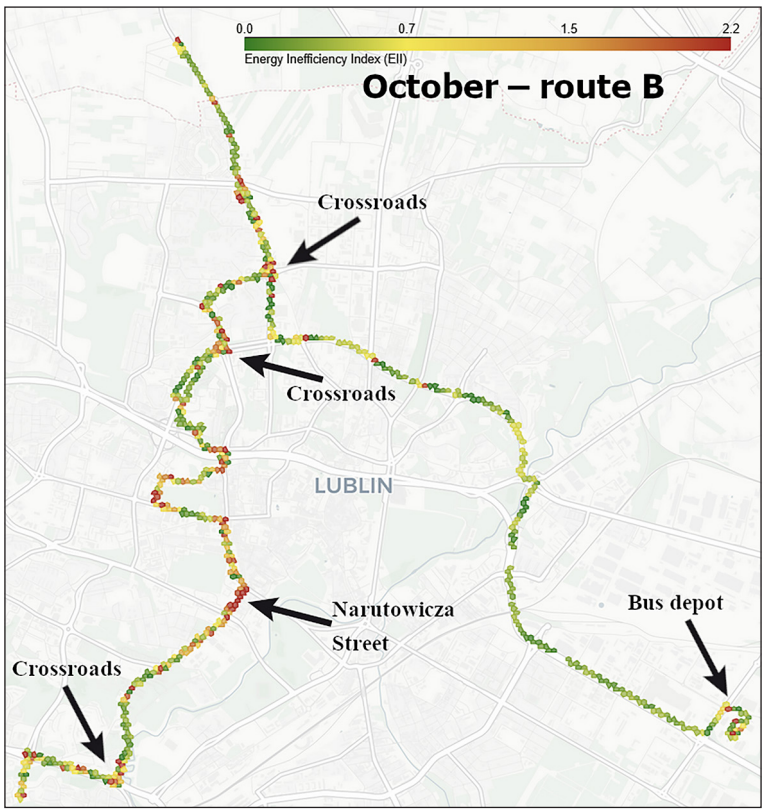


Figure 8. Spatial distribution of the mean EII values
Lower values indicate better energy efficiency – the greener the colour, the better the performance.
The illustrated month is October. The figure in its original resolution is provided in the supplementary materials.

Solar efficiency gain

Another important issue that is imperative to discuss is the SEG factor (8). It measures the theoretical efficiency gain due to the presence of PV panels. In other words, it quantifies how much the PV system reduces the engine load, which results in a hypothetical reduction in fuel consumption. This parameter constitutes the basis of the conducted research. Before proceeding further, it should be noted that the SEG value is expected to be relatively small. When comparing the energy required to keep the bus in motion with the energy generated by the photovoltaic panels, the former is several orders of magnitude greater. Therefore, even a modest energy gain provided by the photovoltaic panels can be regarded as a significant benefit. This also explains the prevalence of red and orange shades in the colour scales presented in Figures 10–12. In these figures, nearly all hexagonal segments are coloured

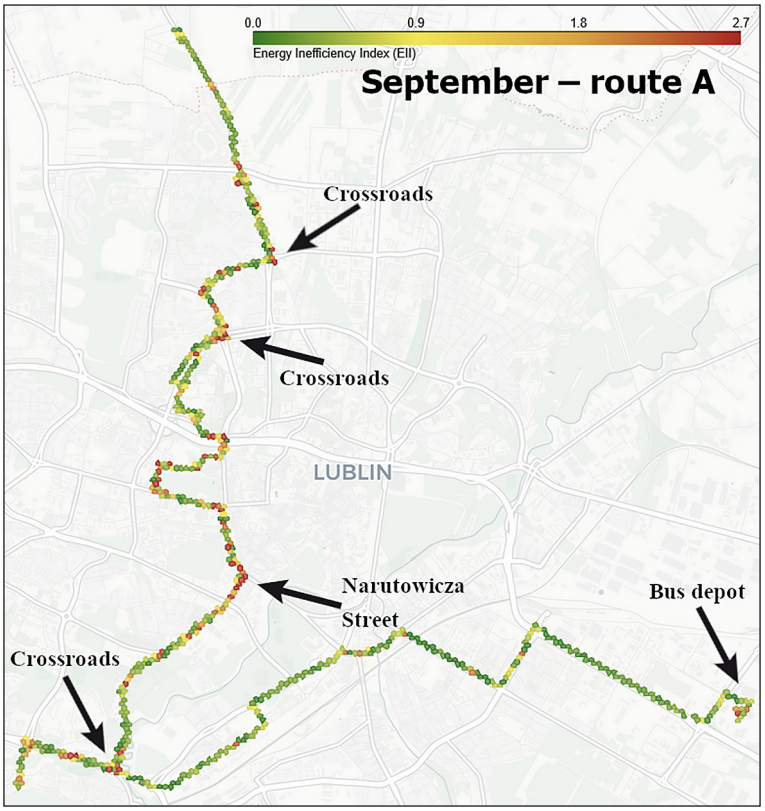


Figure 9. Spatial distribution of the mean EII values
Lower values indicate better energy efficiency – the greener the colour, the better the performance. The illustrated month is September. The figure in its original resolution is provided in the supplementary materials.

in shades of red or orange because the SEG values remain low across most of the analysed area.

Nevertheless, a discernible pattern emerges in Figures 10–12. In particular, at certain intersections and congested areas, such as Narutowicza Street, where buses move at reduced speed, orange hues predominate. This phenomenon is especially apparent in Figures 11 and 12, which present data from September, when solar irradiance is higher than in October.

This observation is particularly important, as it demonstrates that some bus manoeuvres at low speed can positively influence solar energy gain, which may become another factor for analysis, despite the characteristics of photovoltaic panels' exposure to the Sun. Nevertheless, the amount of PV energy produced still depends on the time of year, the street location relative to the Sun, shading from nearby buildings and plants, and weather conditions. All of these may serve as a starting point for identifying bus routes and courses that are most beneficial for PV energy production and solar energy gain. Such knowledge may be useful for transport companies while developing renewable energy-centred strategies.

The discussion so far leads to the conclusion that buses equipped with photovoltaic panels gain the biggest benefits during operation in inner-city areas with PV-friendly characteristics. As shown in Figure 11, the PV panels have a notably positive impact on the overall energy balance of the bus when the vehicle is idle or operating at low speed. Although the amount of energy generated is relatively small, it is renewable, green energy. Moreover, assuming that the additional energy generated by solar panels reduces engine load, thereby leading to a theoretical decrease in fuel consumption, exhaust emissions are also reduced.

The present study was not solely concerned with estimating the efficiency of photovoltaic panels. The objective of identifying locations where the solar panels were least effective was also addressed. It is well known that the urban environment adversely affects the solar performance of vehicles. Consequently, the scope of the research was expanded to include a methodology for identifying regions where the solar energy gain metrics were lowest while the energy inefficiency index was high.

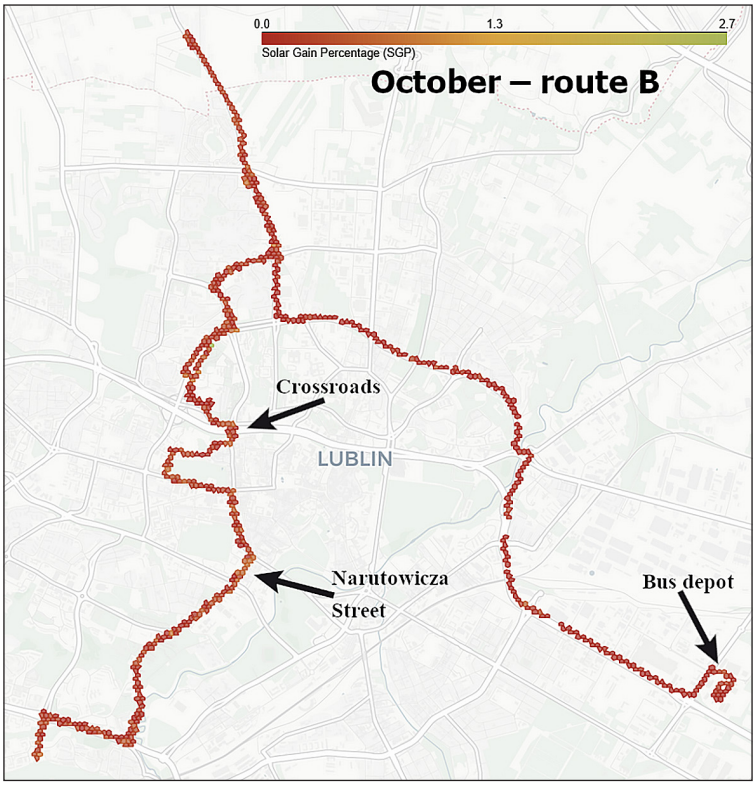


Figure 10. Spatial distribution of the mean SEG values
Higher values indicate greater solar energy gain – the greener the colour, the larger the solar energy contribution. The results shown correspond to October.
The figure in its original resolution is provided in the supplementary materials.

Shadow zones

In light of the preceding discussion, samples with bus speeds below 5 km/h were selected. That is to say, only the segments in which the bus was idle or moving at very low speed were taken into further consideration. Subsequently, the resulting dataset was divided into two distinct groups: samples where $P_{solar} > Q_3$ and samples where $P_{solar} < Q_1$, where Q_3 and Q_1 denote the third and first quartiles, respectively. The remaining samples were discarded. Thereby, two new datasets were created: sun zones ($P_{solar} > Q_3$) and shadow zones ($P_{solar} < Q_1$), and for each, the mean value of the SEG metric was calculated. The mean value in the shadow zones is important as it serves as a shadow threshold.

Next, all samples in the low-speed dataset were grouped by the hexagon index. The regions in which the number of samples was higher than a predefined threshold and the SEG metric was below the shadow threshold were selected. Hexagonal spatial maps are generated at a finite resolution; therefore, hexagons containing a small number of samples may occur. For example, some hexagons may contain 250 points, whereas others contain only 40. Thus, introducing a lower sample threshold is advantageous. The shadow threshold is adjusted individually for each route. Then, a shadow zone score was calculated for each hexagon (10).

$$shadow\ zone\ score = \frac{\#samples}{SEG + \epsilon} \tag{10}$$

The numerator in Equation 10 denotes the number of data points encompassed by a hexagon, while the denominator denotes the SEG metric value plus a small positive constant to avoid division by zero. A higher number of points indicates the bus spent more time in the segment; however, if SEG is low, this may indicate a shadowed area.

Utilising the shadow zone score, it was possible to identify locations where photovoltaic panels contributed least effectively to the energy balance. Using the *Folium* Python package, it is possible to obtain the geographic coordinates of the hexagons. Consequently, in the right column of Figure 13, real-world views of the detected segments are shown, obtained from Google Street View. The left column shows the results mapped onto the EII map.

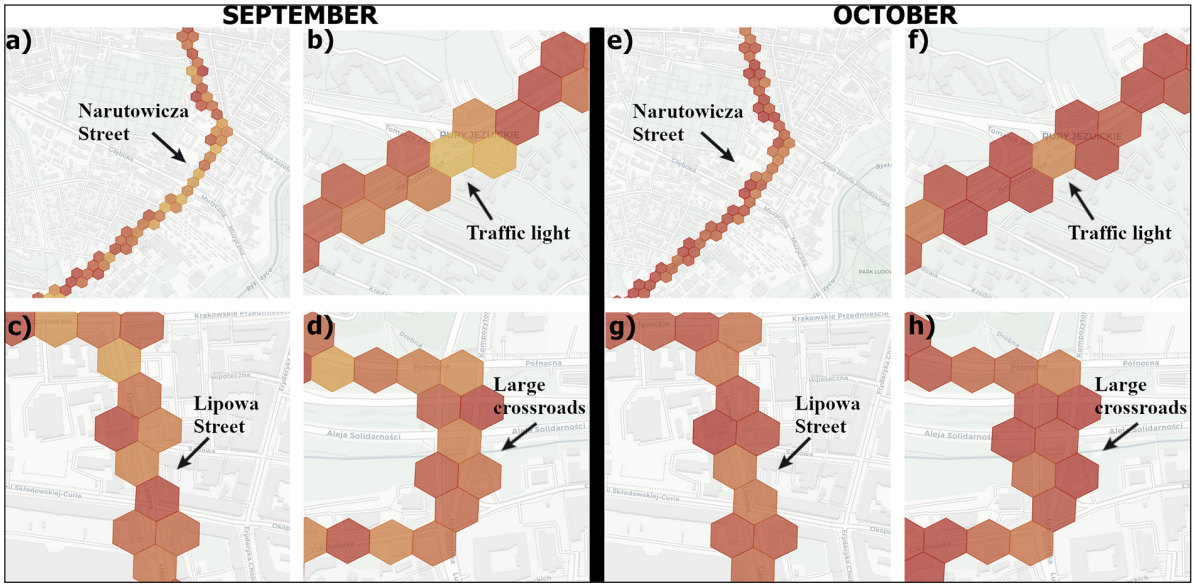


Figure 11. Spatial distribution of the mean SEG values
The greener the colour, the larger the solar energy contribution. The left column shown corresponds to September, the right column to October. The panels show: (a, e) *Narutowicza Street*, (b, f) traffic light spot, (c, g) *Lipowa Street*, and (d, h) major crossroads. The figure in its original resolution is provided in the supplementary materials.

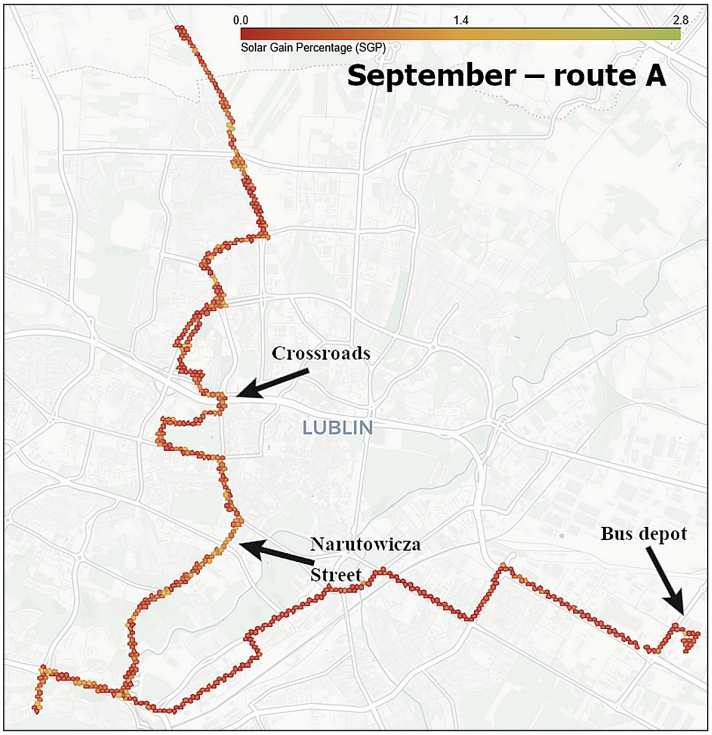


Figure 12. Spatial distribution of the mean SEG values.
Higher values indicate greater solar energy gain – the greener the colour, the larger the solar energy contribution. The results shown correspond to September.
The figure in its original resolution is provided in the supplementary materials.

The segments shown are locations that are in shadow for at least half the day. The shadows are primarily cast by trees and buildings. Panel (a) of Figure 13 depicts Gospodarcza Street and the Kresowa 01 bus stop. In this segment, the EII metric is considerably elevated due to the presence

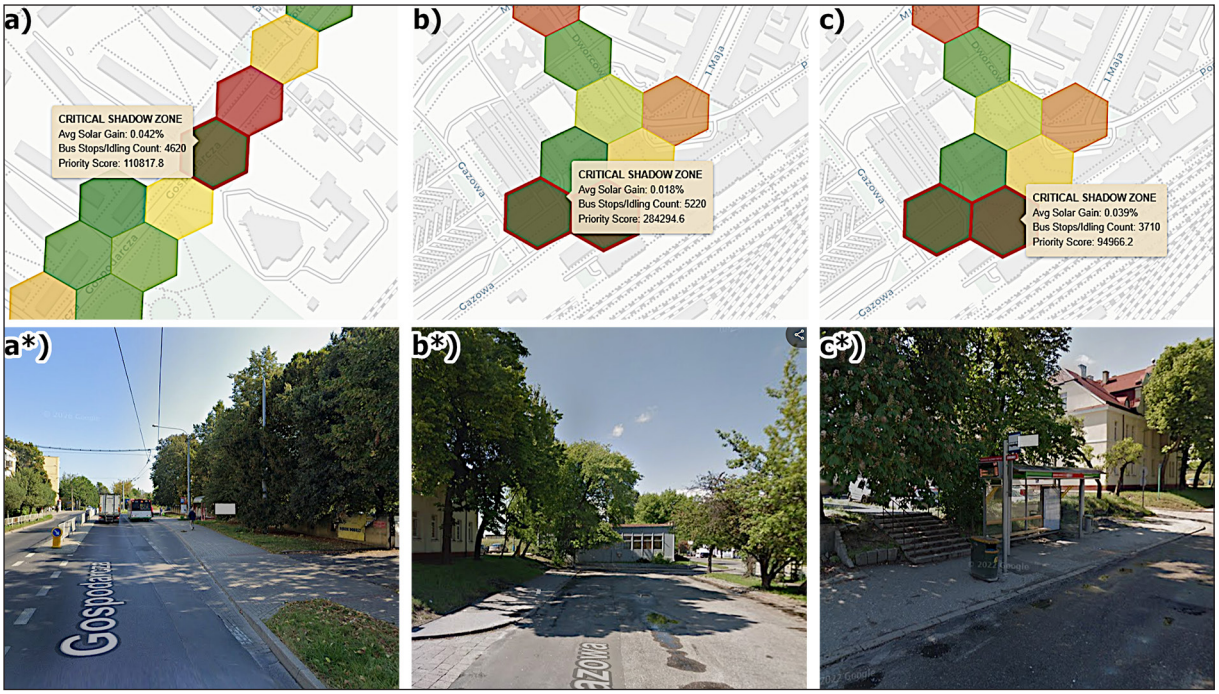


Figure 13. Results of the shadow zone analysis
The left column – panels (a), (b), and (c) – shows the detected regions on the hexagonal map, whereas the right column – panels (a*), (b*), and (c*) – shows corresponding real-world views obtained using Google Street View. The figure in its original resolution is provided in the supplementary materials.

of the bus stop. The bus idles while passengers board and alight. Given that this segment lies in a shaded zone with the high score (10), solar panels cannot effectively mitigate the elevated EII. Furthermore, panel (c) of Figure 13 depicts another bus stop that is completely in shadow. The SEG is higher here than at Gospodarcza Street because, in the segment containing the Kresowa 01 bus stop, the urban environment forms a canyon. On one side, there are tall trees, and on the other, high residential buildings. Therefore, for most of the day, the segment on Gospodarcza Street is covered in shadow.

Locations of the kind discussed so far successfully eliminate the contribution of PV panels in the energy balance. Nevertheless, they constitute a minority across all regions along the route.

CONCLUSIONS

The paper combines a pure data science approach with an important aspect of urban planning. The first part focuses on analysing real-world raw telemetry data. Based on this data, the authors establish relevant metrics and factors. These enable a deep analysis of the energy efficiency of photovoltaic panels mounted on the roofs of public buses.

Among the many analysed data channels, such as speed, revolutions per minute, voltage, and fuel consumption, there is one related to GPS. The GPS data contains many imprecise and invalid samples. Therefore, as the article title suggests, an author-developed machine learning algorithm is introduced to recalibrate the GPS data and reconstruct the bus route. The proposed method successfully processed the noisy raw telemetry data. The bus routes were accurately reconstructed and partitioned into segments.

The second part of the text is devoted to estimating the energy efficiency of photovoltaic panels across various bus routes. The issue is illustrated using georeferenced spatial maps. The proposed metrics successfully supported the achievement of the research objective by enabling the identification of regions characterised by high and low photovoltaic energy yields.

The findings may help identify routes and operating conditions that maximise photovoltaic energy yield. However, determining whether such improvements translate into economic benefits requires a separate techno-economic analysis, as investment, maintenance, additional vehicle

mass, component degradation and payback time were not included in the present study.

Achieving maximal benefits from solar energy also requires urban planning strategies that minimise the formation of urban canyons. More open urban spaces that permit greater solar exposure are advantageous for both urban residents and the efficiency of photovoltaic systems.

The present study is also subject to several limitations. The PV panels were mounted parallel to the bus roof, with no option to optimise their tilt angle, which adversely affected their performance. Thus, the results reflect the performance of a flat, vehicle-mounted PV configuration without tilt optimisation. Moreover, as mentioned, the study evaluates the PV system's energy contribution but does not assess its economic viability, including investment and maintenance costs – the economic viability of equipping city buses with PV systems is a separate, not trivial issue.

Nevertheless, the solar energy generated by PV panels could be utilised in various ways. One potential application is to power environmental monitoring stations mounted on buses, for example, for air pollution monitoring. The operational characteristics of city buses, which continuously travel throughout urban areas, make them particularly well-suited for this purpose.

REFERENCES

1. Salmon PM, King BJ, Elstak I, McLean S, Read GJM. Tomorrow's demons: a scoping review of the risks associated with emerging technologies. Ergonomics 2025;68(9):1491–507. https://doi.org/10.1080/00140139.2024.2416554
2. Miłosz M, Żyła K, Skulimowski SP, Dakowicz AL, Szymczyk T, Badurowicz M. Virtual Reality in supporting the creation of sustainable tourism: A case study of gen z technology acceptance. Sustainability 2025;17(16). https://doi.org/10.3390/su17167173
3. Hajnrych M, Blachowski J, Worsa-Kozak M. Groundwater urban heat island (GUHI) and climate warming: Insights from two decades of monitoring in Wrocław, Poland. Water (Basel) 2025;17(21). https://doi.org/10.3390/w17213090
4. Pasik M, Bakuła K, Różycki S, Ostrowski W, Kowalska ME, Fijałkowska A, et al. Glacier geometry changes in the western shore of Admiralty Bay, King George Island over the last decades. Sensors 2021;21(4). https://doi.org/10.3390/s21041532
5. Kozłowski E, Rimkus A, Zimakowska-Laskowska M, Matijošius J, Wiśniowski P, Traczyński M, et al. Energy and environmental impacts of replacing gasoline with LPG under real driving conditions. Energies (Basel) 2025;18(20). https://doi.org/10.3390/en18205522
6. Ostrowski M, Mroczka J, Płachta K. Comparison of different types of maximum power point techniques for photovoltaic systems. Metrology and Measurement Systems 2024;31(2):401–16. https://doi.org/10.24425/mms.2024.149706
7. Mroczka J, Kaczmarek K. Modern methods of maximum power point tracking in photovoltaic systems: classification, comparison, and application under partial shading conditions. Metrology and Measurement Systems 2025;32(3):1–18. https://doi.org/10.24425/mms.2025.154674.
8. Czarnigowski J, Trendak M. Aircraft piston engine load distribution in steady state operating conditions. Combustion Engines 2023;193(2):29–35. https://doi.org/10.19206/CE-160505
9. Gałka Ł. MSTIF-IFS: Isolation forest based on minimal spanning tree and intuitionistic fuzzy sets. Pattern Recognit 2026;178:113507. https://doi.org/10.1016/j.patcog.2026.113507
10. Zientarski T, Kozar J, Pietrak K. The usability of Bayesian filters in noise reduction across interdisciplinary data. Advances in Science and Technology Research Journal 2025;19:420–35. https://doi.org/10.12913/22998624/209887
11. Oh M, Kim S-M, Park H-D. Estimation of photovoltaic potential of solar bus in an urban area: Case study in Gwanak, Seoul, Korea. Renew Energy 2020;160:1335–48. https://doi.org/10.1016/j.renene.2020.07.048
12. ur Rehman N, Hijazi M, Uzair M. Solar potential assessment of public bus routes for solar buses. Renew Energy 2020;156:193–200. https://doi.org/10.1016/j.renene.2020.04.081
13. Matczuk M, Sulowska D, Gromaszek K. On the usability of isolation forest for 3D mesh analysis and watermarking. Applied Sciences 2025;15(21). https://doi.org/10.3390/app152111364
14. Fijałkowska A, Waksmundzka K, Chmiel J. Assessment of the effectiveness of photovoltaic panels at public transport stops: 3D spatial analysis as a tool to strengthen decision making. Energies (Basel) 2022;15(3). https://doi.org/10.3390/en15031230
15. Laska M, Maciewicz M, Stefaniak Ł, Jaskóła W. Predicting building's energy utilization with the application of energy simulation software: Case study. Advances in Science and Technology Research Journal 2026;20(4):76–90. https://doi.org/10.12913/22998624/214224
16. Adefila OF, Ugwu J, Okonta ED. Optimising energy efficiency in electric buses with curved surface solar panels. Solar Energy Advances 2025;5:100109. https://doi.org/10.1016/j.seja.2025.100109

17. [Van Binh D, Quang Vu P, mạnh hải phạm. A computational study of passive cooling of photovoltaic panels using hybrid material heat sink. International Journal of Electrical and Computer Engineering (IJECE) 2025;15:4487. https://doi.org/10.11591/ijece.v15i5.pp4487-4499
18. Rodríguez-Romero S, Rabanal-Arabach J, Rojas CA, Trigo-Gonzalez M, Mondaca-Cuevas G, Arias D, et al. Analysis of advanced nonisolated topologies for vehicle-integrated photovoltaic (ViPV) systems in urban electric transport buses. IEEE J Photovolt 2026;16(1):47–53. https://doi.org/10.1109/JPHOTOV.2025.3597616
19. Pedraza-Yepes CA, Berdugo-Rolong KE, Ruiz-Muñoz DE, Higuera-Cobos OF, Hernández-Vásquez JD. Feasibility study for the implementation of photovoltaic panels in public transportation in Barranquilla. Energies (Basel) 2023;16(20). https://doi.org/10.3390/en16207161
20. Ioannidis R, Iliopoulou T, Iliopoulou C, Katikas L, Petsou A, Merakou M-E, et al. Solar-powered bus route: introducing renewable energy into a university campus transport system. Advances in Geosciences 2019;49:215–24. https://doi.org/10.5194/adgeo-49-215-2019
21. Grabowski Ł, Wendeker M, Barański G, Duk M. Study of the efficiency of on-board electricity generation in the city bus. Combustion Engines 2017;171(4):29–32. https://doi.org/10.19206/CE-2017-405
22. Azdy R, Darnis F. Use of haversine formula in finding distance between temporary shelter and waste end processing sites. J Phys Conf Ser 2020;1500:12104. https://doi.org/10.1088/1742-6596/1500/1/012104
23. Cuevas E, Rodriguez-Vazquez AN, Rivera-Aguilar BA, López-Luquín JA, Guzmán-Rosales C. Spatial Filters. Image Processing with Python: Theory, Practice, and Applications, Cham: Springer Nature Switzerland; 2026, p. 71–112. https://doi.org/10.1007/978-3-032-13285-7_5
24. Habrat M, Młynarczuk M. Impact of data space augmentation strategy on model accuracy and generalization in thin-section rock classification. Sci Rep 2026;16. https://doi.org/10.1038/s41598-026-44320-y
25. Dziuba-Koziel M, Koziel G, Markiewicz J, Kochanowicz M, Dorosz D, Miluski P, et al. Machine learning-assisted design of inverted gradient multiring optical fibres for flat-top beam generation. Opt Express 2026;34:7159–73. https://doi.org/10.1364/OE.587143
26. Roberts-Licklider K, Trafalis T. Optimizing Treatment facility locations in Oklahoma Using Haversine, Euclidean, Manhattan, and Chebyshev distance optimization. Operations Research Forum 2025;6. https://doi.org/10.1007/s43069-025-00448-7
27. Gulkowski S. Modeling and Experimental studies of the photovoltaic system performance in climate conditions of Poland. Energies (Basel) 2023;16(20). https://doi.org/10.3390/en16207017
28. Gulkowski S, Krawczak E. Long-term energy yield analysis of the rooftop PV system in climate conditions of Poland. Sustainability 2024;16(8). https://doi.org/10.3390/su16083348
29. Chwaleba K, Wach W. A convolutional neural network-driven model with adaptive feature fusion for Polish national dance music recognition. Advances in Science and Technology Research Journal 2026;20(1):354–72. https://doi.org/10.12913/22998624/210059
30. Grabowski Ł, Sochaczewski R, Biały M. Deep learning long short-term memory methods for instantaneous fuel consumption prediction: Experimental study and comparison of modeling strategies. Advances in Science and Technology Research Journal 2025;19(12):197–209. https://doi.org/10.12913/22998624/210215
31. Kłosowski G, Gola A, Świć A. Application of Fuzzy Logic Controller for Machine Load Balancing in Discrete Manufacturing System. In: Jackowski K, Burduk R, Walkowiak K, Wozniak M, Yin H, editors. Intelligent Data Engineering and Automated Learning – IDEAL 2015, Cham: Springer International Publishing; 2015, p. 256–63. https://doi.org/10.1007/978-3-319-24834-9_31