A Data-Driven Approach to Monitor Sustainable Development Transition in Italian Regions Through SDG 11 Indicators

Poli, Giuliano; Cuntò, Stefano; Muccio, Eugenio

DOI
10.1007/978-3-031-65285-1_22

Publication date
2024

Document Version
Final published version

Published in
Computational Science and Its Applications – ICCSA 2024 Workshops

A Data-Driven Approach to Monitor Sustainable Development Transition in Italian Regions Through SDG 11 Indicators

Giuliano Poli[1], Stefano Cuntò[1], and Eugenio Muccio[1,2](✉)

[1] Department of Architecture, University of Naples Federico II, Via Toledo 402, Naples, Italy
{giuliano.poli,stefano.cunto,eugenio.muccio}@unina.it

[2] Department of Management in the Built Environment, Faculty of Architecture and the Built Environment, Delft University of Technology, Julianalaan 134, 2628 BL Delft, the Netherlands

Abstract. Monitoring SDG 11 targets is crucial for making informed decisions and supporting multidimensional transitions in European cities. Among all the goals, SDG 11 emerges as a cornerstone for cities, offering a comprehensive framework to tackle their multifaceted challenges. Composite indicators and indices, as suited evaluation tools to monitor city progress or decline, allow sustainability problems to be included in local agendas by aggregating multi-dimensional variables at different time spans through data-driven approaches. The primary concerns about using indicators as evaluation tools to compare performances are inherent to inconsistencies related to different assessment frameworks and methods, data downscaling from global to local levels, choice of aggregation rules to obtain synthetic results, and data gaps. This contribution, in particular, focuses on data gaps by elaborating on a testing case, while critically discussing related issues. The research was addressed to identify normative, assessment, and methodological gaps in monitoring progress towards SDG 11 at global, European, and Italian levels. Application of Machine Learning algorithms to predict null values within an SDG 11 regional dataset was implemented to compare three Italian regions according to 18 common indicators. The contribution is part of the Research Project of National Relevance "GLOSSA - GLOcal knowledge System for Sustainability Assessment of urban projects", coordinated by Polytechnic of Turin (Italy), and it supports its first-step knowledge phase aimed at identifying gaps in SDG 11 indicators downscaling and monitoring.

Keywords: SDG 11 monitoring · Data-driven approach · Machine Learning · Indicators · Sustainable Development Strategies

1 Introduction

Among the 17 Sustainable Development Goals (SDGs), SDG 11 emerges as a cornerstone for cities, offering a comprehensive framework to tackle their multifaceted challenges [1]. Fostering inclusivity, resilience, and sustainability, its targets should unfold the city's potential to be more vibrant in social-cultural terms, resilient, and economically equitable [2]. Monitoring SDG 11 progress is, thus, crucial for making informed decisions and supporting multidimensional transitions in European cities [3].

O. Gervasi et al. (Eds.): ICCSA 2024 Workshops, LNCS 14820, pp. 337–355, 2024.
https://doi.org/10.1007/978-3-031-65285-1_22

Indicators, in their single or composite form, constitute suited evaluation tools to monitor city progress or decline since they allow sustainability problems to be included in local agendas by aggregating multi-dimensional variables at different timespans [4, 5]. Furthermore, these metrics can be incorporated into diverse models to elucidate current patterns or to simulate and forecast potential outcomes under varying conditions, envisioning future iterations of urban landscapes [6].

Data-driven approaches provide a robust framework for measuring, analysing, and interpreting various phenomena through indicators [7]. By harnessing the power of data, these approaches offer insights into trends, patterns, and correlations that enable informed decision-making across diverse domains. Moreover, they facilitate the analysis of indicators through advanced techniques such as statistical modeling, machine learning, and data visualisation, in order to uncover hidden relationships within the data, predict future trends, and derive actionable insights.

Among the various approaches, Exploratory Data Analysis (EDA) examines data from as many perspectives as possible, constantly on the lookout for identifying recurring data patterns. The purpose is discovering evidence about the data, and it is neither motivated by a need to confirm the existence of a specific impact nor is it backed by a statistical model that includes a mathematical formulation for such an effect [8]. Furthermore, EDA plays a propaedeutic role in preparing data for Machine Learning algorithms by providing fast insights into data characteristics, identifying data validation strategies, guiding feature selection, and using data visualisation to better represent the shape of the dataset [9].

The primary concerns about using indicators as evaluation tools to compare performances are inherent to: inconsistencies related to different assessment frameworks and methods, data downscaling from global to local levels, choice of aggregation rules to obtain synthetic results, and data gaps [10]. This contribution focuses on data gaps by elaborating on a testing case, while critically discussing related issues.

One of the main gaps in monitoring and forecasting indicators concerns data absence since it prevents downscaling at local levels. On the counterpart, the data gap can unlock indicators research in selecting consistent and context-aware proxy variables by integrating the bottom-up with the top-down dimension [11, 12]. Bridging this gap is crucial in the context of SDG 11, whose monitoring and evaluation require comprehensive data at both the macro and micro levels to understand urban dynamics, assess progress, and inform policy interventions. By addressing data gaps and adopting a data-driven approach, stakeholders can better track urban development indicators, identify disparities, and tailor strategies to promote equitable and sustainable urbanisation [13].

The research was addressed to identify normative, assessment, and methodological gaps in monitoring progress towards SDG 11 at global, European, and Italian levels. Application of Machine Learning algorithms to predict null values within an SDG 11 regional dataset was implemented to compare three Italian Regions according to 18 common indicators.

The contribution is part of the Research Project of National Relevance, coordinated by Polytechnic of Turin (Italy), referred to as “GLOSSA - GLOcal knowledge System for Sustainability Assessment of urban projects”, and it supports the preliminary knowledge phase aimed at identifying gaps in the SDG 11 indicators downscaling and monitoring.

2 Research Design

The research design was conceptualised to provide insights for the Glossa project which aims to define a knowledge-system to support Decision Makers in the evaluation of urban transformation projects. Whilst, in this study, the authors have selected and processed data at the regional level, the efforts were addressed to detect and solve assessment gaps linked to lacking data which generally recur in indicators downscaling.

Figure 1 shows the workflow which stems from three Research Questions (RQ), as follows:

- RQ1. How has SDG 11 been implemented and monitored through indicators at the global, European and Italian levels?
- RQ2. Which are the most recurring indicators selected by the Italian Regions within the RSDSs?
- RQ3. Which limitations emerge from the scientific literature related to SDG 11 indicators assessment methods?
- Each RQ has been stressed by using different analysis methods and evaluation tools, according to three phases:
- Phase 1. Multilevel Governance Analysis has allowed normative gaps to emerge from a critical overview of the main documents and assessment framework at global and national levels.
- Phase 2. Frequency Analysis of the most recurring indicators within the Italian Regional Sustainable Development Strategies (RSDS) has allowed a common to 18 regions set of SDG 11 indicators to be selected.
- Phase 3. Literature Review of scientific publications from 2016 to 2024 has allowed the most used SDG 11 assessment methods to emerge to identify limitations and recurring data analysis gaps. Research protocols within databases Scopus and Web of Science have been implemented in this phase.

Results of the previous 3 phases highlighted the data gap as one of the most relevant issues, requiring a specific focus on the input methods to predict missing data records for specific indicators at different timespan. Thus, a further phase follows:

- Phase 4. Machine Learning has allowed feasible solutions to data gaps to emerge, along with limitations, while EDA was addressed to show the most correlated indicators within the three regional areas of the GLOSSA project. K-Nearest Neighbours (KNN) Machine Learning algorithm has been implemented through Python 3 to predict null values within the dataset.

The last phase results consist of correlation matrices and bivariate distributions of the most correlated indicators, obtained through EDA, and prediction accuracy index graphs, obtained through KNN.

2.1 Multilevel Governance Analysis About SDG 11 Indicators (Phase 1)

A comparative analysis has allowed us to examine indicators-based frameworks at different governance levels to detect recurring issues in monitoring cities' sustainability and define a common set of indicators for Italian regions.

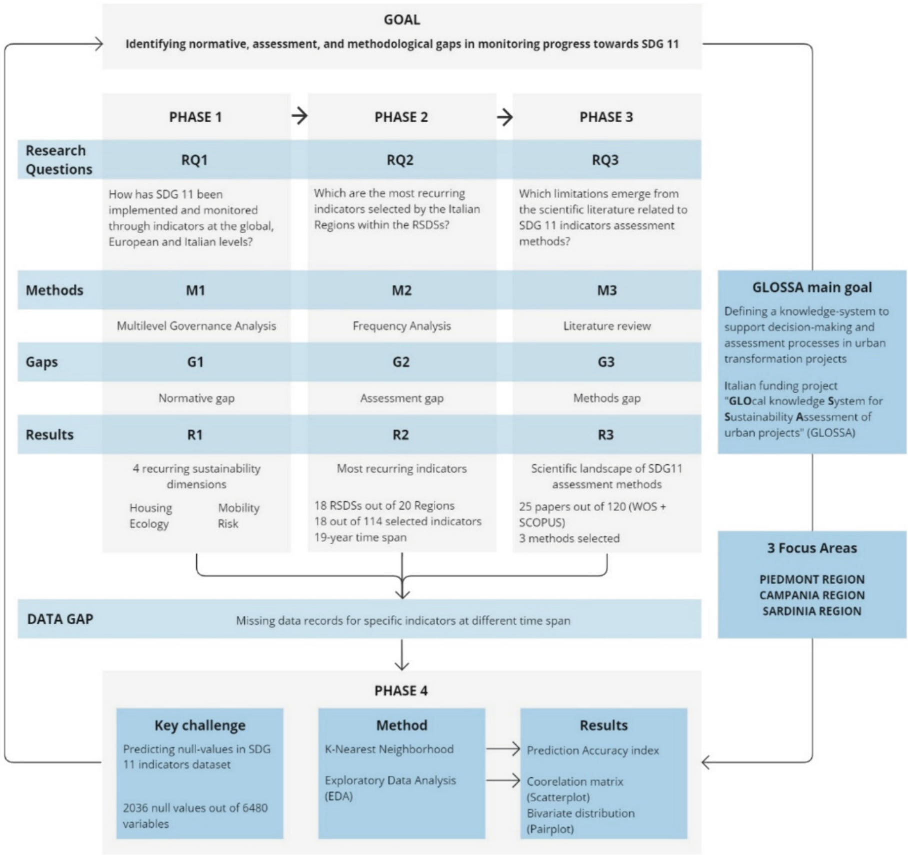

Fig. 1. Research design workflow.

With reference to the 17 SDGs of the 2030 Agenda and the corresponding 169 targets, the Statistical Commission of the United Nations established a shared framework of statistical information as a tool for monitoring and analysing sustainability. The activity programme of IAEG-SDGs includes, in addition to the implementation of indicators based on current methodologies and traditional data sources, the promotion and development of innovative elements that include non-traditional data sources, such as citizen-generated data, as complementary information sources. The current set proposed in 2023 by IAEG-SDGs consists of 231 indicators, although the total number considered is 248, as some are used for monitoring multiple targets.

At the European level, the European Commission (EC) put the SDGs at the center of its relevant governance tools such as the European Green Deal and the recovery and resilience plans under Next Generation EU.

Progress in achieving the SDGs is monitored by the European statistical office Eurostat through annual reports providing the assessment of short-term and long-term trends. The Eurostat indicator set does not directly address the 169 targets outlined in 2030

Agenda. Instead, it opted to choose suitable indicators that capture pertinent dimensions of the Sustainable Development Goals (SDGs) from the European Union's standpoint. Comprising 102 indicators distributed among the 17 SDGs, the EU indicator set allocates 6 dedicated indicators to each goal. Notably, 34 of these indicators serve the purpose of monitoring multiple SDGs simultaneously.

In the European context, member states are encouraged to develop appropriate policy and governance tools to establish national guidelines that facilitate the implementation of concrete actions at the territorial level to achieve and monitor the SDGs.

In Italy, the National Sustainable Development Strategy (NSDS) constitutes, since 2017, the reference framework for environmental and territorial planning, programming, and evaluation processes at all scales. This document reports a preliminary set of indicators, later implemented by ISTAT through the "Working Group for the definition of indicators for the National Strategy for Sustainable Development", established in 2018 with the convergence of the Ministry of Foreign Affairs and International Cooperation (MAECI), the Ministry of Economy and Finance (MEF), the Presidency of the Council, and the Higher Institute for Environmental Protection and Research (ISPRA). As of this, the first 43 indicators for the NSDS were selected, within the broader framework of ISTAT SDGs indicators. Since 2016, ISTAT has been making available numerous indicators on a semi-annual basis, many of which intersect with the Fair and Sustainable Well-being (BES) indicator system, encompassing a shared total of 62 statistical measures. In the sixth edition of the ISTAT Report on Development Objectives, an amount of 372 statistical measures are presented. Among these, 342 measures are distinct and linked to individual goals, correlating with 139 indicators outlined by the IAEG-SDGs to monitor worldwide advancement towards 2030 Agenda.

Another notable actor in SDGs monitoring is the Italian Alliance for Sustainable Development (ASviS). Since the consultations initiated by MASE in 2016 for the NSDS, ASviS has established an openly accessible web database containing 168 composite indicators aligned to UN indicators and designed to evaluate national advancement and setbacks concerning the SDGs.

The latest NSDS strategic framework comprises five areas aligned with the 2030 Agenda's five pillars. Additionally, it encompasses a sixth domain termed Sustainability Vectors, which refers to cross-cutting areas of intervention deemed essential for instigating, directing, administering, and overseeing sustainability efforts at the local level. The NSDS establishes sustainable development objectives, which are related to - yet not identical with - the SDGs. These objectives are distinguished by the interplay among various SDGs, as per the Nexus approach [14].

For each domain, National Strategic Choices (NSCs) are identified, which correspond to the missions and components of both EU and national cohesion policies. These NSCs are associated with National Sustainability Objectives (NSOs), which, in turn, are linked to "Target Values" shared by other strategies such as the PNRR or the 2030 Agenda.

The NSDS outlines a foundational set of 55 first-level indicators aligned with NSCs. These indicators are chosen based on their prevalence in key programming frameworks, including the PNRR, BES, Economic and Finance Document (DEF), Plan for Ecological Transition (PTE), and Partnership Agreement/Development Policies. Additionally, there exist 190 secondary indicators associated with NSOs. Together with the primary indicators, these secondary indicators play a crucial role in the annual monitoring and reporting of the NSDS. While the NSDS does not present a specific set of indicators explicitly aimed at assessing the achievement of individual SDGs, it indirectly indicates which indicators, aligned with the SDGs, should be chosen for the monitoring system.

Drawing from a range of indicators selected by entities such as the UN, Eurostat, ISTAT, ASviS, and those inferred from the monitoring system of NSOs within the NSDS, one can discern various recurring and pertinent themes aimed at enhancing Goal 11 across different scales, as illustrated in Table 1.

Table 1. Number and typology of indicators in global, European, and Italian normative.

Legislation/ Organization	Governance level	Total number of indicators	Number of SDG 11 indicators	Key themes
ONU	Global	231	15	Governance; Risk; Ecology; Housing; Mobility
EUROSTAT	European	102	9	Housing; Risk; Mobility; Ecology
ISTAT	National / Regional	342	32	Ecology; Risk; Housing; Mobility; Climate Change; Governance
ASVIS	National / Regional	168	6	Mobility; Ecology; Housing
NSDS	National	245	16	Ecology; Governance; Mobility; Risk; Housing quality

Upon comparing the five analysed sets, four key dimensions emerge as critical for comprehending the concept of a sustainable city: Housing, Ecology, Mobility, and Risk.

Underlining the universal significance of ensuring decent and secure housing for all inhabitants, each source incorporates indicators about housing conditions, such as the percentage of the population residing in urban slums, informal settlements or substandard housing, as well as those living in dwellings afflicted by structural deficiencies, dampness, or overcrowding. Certain indicators, such as the incidence of urban green areas in the urbanised area or the exceedance of daily thresholds for pollutants, underscore the imperative of sustainable management of natural resources and the urban environment. This entails advocating for sustainable consumption and production practices and enhancing the valorization of urban ecosystems. Numerous indicators, including the accessibility to public transport, passenger-km offered by public transport, and the prevalence of car usage for passenger transit, underscore the importance of fostering efficient and accessible public transportation infrastructure and services. Additionally, data highlights the necessity of implementing preventive and mitigation measures, as well as promoting safety and inclusivity within urban areas. This is evidenced by indicators such as the population exposed to landslide and flood risks and fatalities resulting from road accidents.

2.2 SDG 11 Indicators Within the Italian Regional Sustainable Development Strategy (Phase 2)

The NSDS - established as the national framework by Article 34 of Legislative Decree 152/2006 - provides for the formulation of strategies at the regional level to ensure monitoring of the contribution to the achievement of NSOs by territorial policies.

In this regard, regional and provincial administrations have established steering committees composed of internal administrative structures with competencies in social, economic, and environmental issues for sustainability monitoring. These entities, either through working groups or coordination boards, are engaged in the formulation of Regional Sustainable Development Strategies (RSDSs). In certain instances, the responsibility for coordination is directly assigned to the Regional Presidency or the General Secretariat.

Furthermore, numerous regions engage external entities, including provinces, associations, regional agencies, and other public or private organizations, to foster an integrated and participatory approach to urban sustainability. Moreover, by setting up Forums for Sustainable Development, coordination and monitoring efforts seek to engage a diversified range of stakeholders from the local community to boost RSDSs.

Within this intricate multi-level governance framework, there is a pressing need to establish an integrated monitoring system for the NSDS. The main effort involves identifying a common set of indicators, adaptable to different territorial levels, to gauge the contribution of regions towards NSDS implementation and SDG attainment.

While ASviS proposes a comprehensive set of 168 indicators to assess regions' performance in advancing the SDGs across various sectors, a unified national regulatory framework for such monitoring has yet to be defined. An analysis of the 17 RSDSs approved by MASE reveals significant variation in the indicators selection process among Italian regions due to the absence of a standardised tool. Some regions rely on the 44 indicators devised by the Working Group while others utilise indicators from ISTAT SDGs or the Fair and Sustainable Well-being (BES) framework. However, the majority of regions emphasise the selection of indicators based on local relevance and data availability, often integrating national indicators with regional or local ones.

Moreover, different approaches to monitoring progress in attaining the Sustainable Development Goals (SDGs) have been observed. While some regions have devised assessments that cut across their strategic initiatives, constructing composite indicators to assess progress towards the SDGs, others have chosen to evaluate their strategic initiatives in alignment with the NSDS model. In this latter approach, regions prefer to assess their strategic initiatives using indicators linked to the SDGs, rather than directly measuring the achievement of individual Goals. Consequently, the evaluation is not focused on the achievement of the SDGs themselves but on assessing the appropriateness of actions concerning the SDGs' targets.

In the pursuit of identifying a potential set of shared indicators at the regional level, the research compares the indicators either explicitly stated or inferred from the RSDSs for monitoring SDG 11 (Fig. 2). Through an examination of the various datasets of indicators provided at the regional level, notable discrepancies emerge in the approach adopted by each region.

Many regions (such as Puglia, Abruzzo, Molise, Lazio, Friuli Venezia Giulia, Liguria, Piemonte, and Valle d'Aosta) stand out for their extensive use of indicators, surpassing the selected set of 17 indicators. These regions, characterised by high population densities and intricate urban challenges, decided to adopt a comprehensive and multifaceted approach to evaluating urban sustainability. Conversely, some regions have opted for a limited number of indicators, which could indicate a focus on key aspects or resource constraints for developing a comprehensive monitoring strategy. For instance, Umbria and Tuscany utilise 3 and 6 indicators respectively, suggesting a more targeted approach or a territorial context requiring less detailed analysis. Certain regions prioritise indicators from the ISTAT SDGs database (such as Campania, Puglia, Sardegna, Sicilia, Marche, and Lazio), potentially driven by a desire for national and international comparability as well as limited monitoring capabilities at the local level.

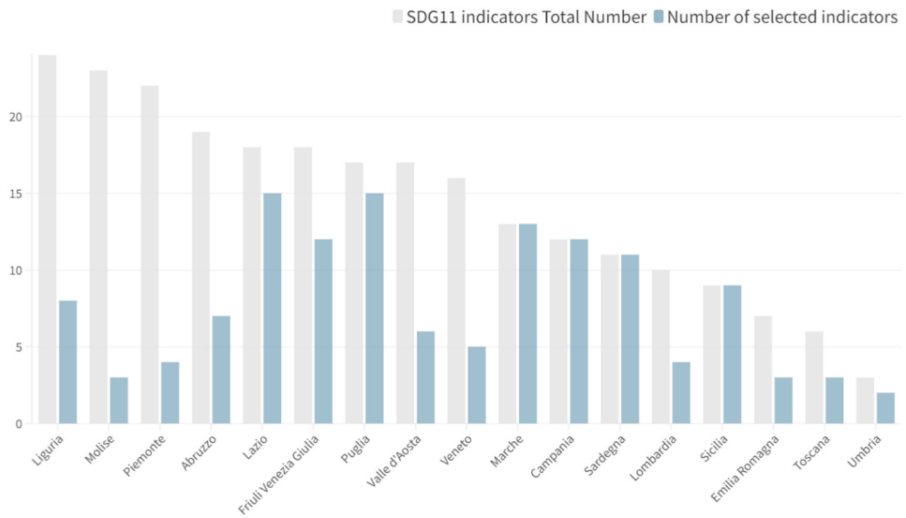

Fig. 2. Column chart of indicator frequency across Italian regions for SDG 11 assessment.

Similarly, other regions (including Abruzzo, Molise, Emilia Romagna, Friuli Venezia Giulia, Liguria, Piemonte, and Valle d'Aosta) favour territorial indicators developed at the local level. Collaboratively constructed with local authorities and territorial organizations, these indicators offer a more nuanced and context-specific perspective on urban sustainability. They may encompass data related to urban infrastructure, environmental quality, and community engagement.

2.3 The Scientific Landscape (Phase 3)

SDG 11, as is known, aims to make cities and human settlements inclusive, safe, resilient, and sustainable. Achieving this goal requires effective measurement and assessment frameworks to monitor progress and guide policy interventions at the regional level. Composite indicators, which integrate multiple indicators into a single metric, offer a comprehensive approach for assessing complex phenomena like urban sustainability. Statistical and multi-criteria methods play a crucial role in the construction of such composite indicators. This literature review critically evaluates the suitability of different statistical and multi-criteria methods for constructing composite indicators to assess SDG 11 in regional planning contexts.

The methodology involved a systematic literature review across academic databases, including Scopus and Web of Science, and using predefined keywords and logical operators as follows: "SDG 11" AND ("Composite indicator" OR "indicator" OR "indicator-based method" OR "aggregation method" OR "multicriteria decision analysis") AND ("sustainable development strategies" OR "regional planning" OR "policy" OR "scenarios assessment" OR "scenario planning" OR "scenario analysis"). The inclusion criteria encompass peer-reviewed articles, book chapters, and reviews published in the last decade, written in English, and focusing on the application of statistical and multi-criteria methods to effectively construct composite indicators for SDG 11 assessment in regional planning contexts.

25 relevant articles published between 2016 and 2024 were identified and screened for inclusion based on their relevance to the discussed topics and keywords. The results show that the most recurring research fields, according to WoS categories, are Environmental Sciences, Environmental Studies, Green Sustainable Science Technology and Urban Studies.

Through the network visualisation of the scientific landscape, as shown in Fig. 3, seven clusters related to composite indicators for SDG 11 assessment have been identified and referred to as: 1. Environmental footprint (red); 2. Energy and water management (green); 3. SDGs assessment (blue); 4. Network analysis (yellow); 5. Environmental management (purple); 6. Wellbeing (light blue); 7. WEFE nexus (orange). The weight of clusters indicates the number of term occurrences in the revised literature. It can be noted that the term "sustainability" clearly cuts across the clusters, which are characterised, on the one hand, with reference to the environment and natural resources and, on the other hand, in terms of assessment methods and sustainable management of resources, including a focus on the economic dimension.

Focusing on the sustainability assessment methods emerged from the revised literature, Multi-Criteria Decision Analysis (MCDA) provides several aggregation techniques, through compensatory, partially compensatory or non-compensatory approaches, to

Fig. 3. The scientific landscape.

measure city progress towards SDGs [15]. Among different approaches, Ricciolini et al. (2022) suggest using Multiple Reference Point Weak-Strong Composite Indicators (MRP-WSCI) and its partially compensatory version (MRP-PCI) to incorporate indicator reference levels, enabling the use of both absolute and relative benchmarks. The advantage linked to the hybridisation of the two methods results in a composite indicator with varying degrees of compensation, which, on the one hand, empowers DMs to evaluate overall performance, on the other hand, to identify areas where improvement is possible [16].

Other assessment methods proposed by UN Inter-Agency and Expert Group on SDGs (IAEG-SDGs) Working Group on Geospatial Information highlight the geostatistical approach as crucial for harmonising indicators and enabling detailed SDGs evaluation at the most possible detailed geographical regions. According to this approach Benedek et al. (2021) compute an overall index to measure SDG progress locally and regionally in Romania using a proposed integrated approach with 90 indicators stored in a PostgreSQL database [17].

In Italy, the Italian Alliance for Sustainable Development (ASviS) has proposed a composite indicator for each of the 17 SDGs at the regional level using the Adjusted Mazziotta Pareto Index (AMPI), a methodology adopted by ISTAT to compare performances of administrative units over time and space [18]. AMPI is based on a mathematical function considering the arithmetic mean of normalised indicators, introducing a penalty for units with unbalanced indicator values, thus reflecting the multidimensionality of the phenomenon under consideration. ASviS chooses the indicator values - normalised to 100 - in 2010 as reference points to compare the regional performances for assessing progress towards SDGs.

Preliminary results from the literature review showed evidence supporting the feasibility of an integrated approach to policy evaluation through SDGs. Traditionally, such policies have primarily focused on economic metrics, while emerging research suggests a more holistic perspective, which considers economic, social, and environmental dimensions, offering a more nuanced understanding of effectiveness and broader societal implications.

3 Normative, Assessment, and Methodological Gaps

Three types of gaps in indicators-based assessments have emerged in answering the previous research questions concerning SDG11 monitoring framework from global to national level (RQ1), comparison of Italian Regions according to a standard set of indicators (RQ2), and methodological divergences and limitations in scientific literature (RQ3).

A significant concern is inherent to inconsistencies stemming from different assessment frameworks employed at different levels to monitor progress towards SDGs. Each framework, indeed, prioritises different urban dimensions, resulting in divergent conclusions even when analysing similar phenomena. While the four prevailing dimensions in sustainability assessment of cities - Housing, Ecology, Mobility, and Risk - recur in all the analysed frameworks at all governance levels, two infrequently selected dimensions emerge from the analysis of indicators, related to Governance and Climate Change.

Another challenge arises from the process of downscaling data from global to local levels. While global-level data offer a broad understanding of trends and patterns, its relevance and accuracy at local scales still remains compromised without guidelines and consistent criteria to validate indicators. Downscaling methodologies introduce errors or overlook critical nuances, leading to discrepancies in localised assessments. Consequently, decision-makers struggle to derive actionable insights tailored to specific geographical and social contexts.

On the other hand, the analysis of regional indicators highlighted a diversification of monitoring strategies for SDG11, influenced by territorial, economic, and socio-cultural factors. While some regions prioritise a broader and standardised perspective, others opt for a more targeted and contextualised approach, reflecting the complexity and diversity of urban and territorial challenges in Italy. These inconsistencies prevent DMs to cross-compare regions at a national level and, also, raise questions regarding the reliability of assessment outcomes.

Moreover, the choice of aggregation rules to obtain composite indicators poses another relevant gap. Aggregation methods vary in their complexity and assumptions, influencing the interpretation of the final measures. Decisions regarding aggregation favour certain perspectives or biases, potentially hindering important issues present in the underlying data. As a result, composite indices and indicators could oversimplify complex phenomena or fail to capture essential variations, limiting the comprehensiveness of evaluations.

Furthermore, the issue of missing data records for specific indicators at different time spans complicates the assessment process. Data gaps can arise due to incomplete data collection, methodological changes, resource constraints, monitoring lack in a certain time span. These gaps undermine the continuity of monitoring as well as introduce uncertainties in decision-making. The following Sections highlight possible solutions for setting a standard in the assessment of Italian Regions according to a common set of SDG 11 indicators and propose the use of ML to predict null-values in large datasets through the comparison of different methods.

4 Comparing Italian Regions Against SDG 11 Indicators: An Assessment Framework

A comparison of 17 Italian RSDSs has been performed to identify the most commonly adopted SDG 11 indicators. After determining the frequency of occurrence, indicators that appeared at least three times across the regional datasets were selected and presented in Table 2, thus ensuring a substantial representation of the issues addressed by Italian regions.

Table 2. The selected indicators framework.

Category	ID	Indicator	Source	Frequency	Polarity
Ecology	ECO_01	Incidence of urban green areas on the urbanised surface of cities	ISTAT	11	+
	ECO_02	Disposal of urban waste in landfills	ISPRA	9	–
	ECO_03	Soil sealing and land consumption per capita	ISTAT	7	–
	ECO_04	Urban air quality - PM10	ISTAT	7	–
	ECO_05	Air quality - PM2.5	ISTAT	6	–
	ECO_06	Urban air quality - Nitrogen dioxide	ISTAT	5	–
	ECO_07	Separate collection of urban waste	ISPRA	3	+
Housing	HOU_01	Percentage of people living in housing with structural problems or moisture problems	ISTAT	10	–
	HOU_02	Percentage of people living in overcrowded housing	ISTAT	10	–
	HOU_03	Building abuses	CRESME	10	–
	HOU_04	Percentage of people living in housing with noise from neighbours or the street	ISTAT	3	–
Mobility	MOB_01	Families reporting difficulty connecting with public transportation in the area where they reside	ISTAT	10	–
	MOB_02	People who routinely travel to work only by private means	ISTAT	8	–
	MOB_03	Students who habitually commute to study places only by public transport	ISTAT	7	+
	MOB_04	Passenger-km offered by public transport	ISTAT	5	+

(*continued*)

Table 2. (*continued*)

Category	ID	Indicator	Source	Frequency	Polarity
	MOB_05	Regular users of public transport	ISTAT	4	+
Risk	RISK_01	Population exposed to landslide risk	ISPRA	8	–
	RISK_02	Population exposed to flood risk	ISPRA	8	–

Hence, out of a total of 114 indicators, 18 were identified, with their frequency of use ranging from 3 to 11, of which only 3 had a value lower than 5. Regarding the sources, 5 indicators originate from ISPRA, 1 from CRESME, and the rest from ISTAT. Consistent with findings from the analysis of indicator sets provided by national and international agencies for monitoring SDG 11, the indicators chosen by the RSDSs encompass the four dimensions of urban sustainability previously defined.

Selected as the most frequent, these indicators shed light on various aspects of urban sustainability, facilitating a holistic and thorough evaluation of SDG11. The dataset is largely dominated by ecological and mobility indicators, reflecting the emphasis placed on environmental well-being and the imperative to mitigate detrimental emissions [19, 20]. Additionally, the significance of quality of life, defined as the right to reside in healthy and secure environments, is underscored across all RSDSs.

5 Results

The indicators were transferred into a bi-dimensional matrix through Python 3.0, where the sample refers to 19 years in the time span 2003–2022 and the features represent the indicator's values per region. The final matrix is, thus, shaped by 19 rows and 359 columns. The shaped matrix has approximately 2000 null values (NaN) corresponding to the third part of the total records. A mapping of the main statistics – e.g. mean, standard deviation, percentiles, frequency, etc. – has allowed exploring the overall dataset and single features related to indicators values per region.

5.1 A Predictive Model Based on the KNN ML Algorithm

KNN was used to predict null values in the dataset. The algorithm was set to select 2 similar points within known data representing a feature (column) of the dataset. Before implementing KNN, the matrix was split into training and testing data, hiding 50% of known values in the training data. An accuracy test was produced to check the consistency of the model comparing predicted values to hiding data. Several experimental tests have been performed to increase the model accuracy by changing parameters related to the number of K-points and the test size. Nevertheless, the accuracy test reached the limit of approximately 33% of correct predictions, as shown in Table 3.

Despite the low accuracy percentage highlighted a limitation of this study, it has been detected that KNN provide better results compared to filling dataset with zero or mean values. Further parameters changes can improve the accuracy level of the test, allowing for more consistent predictions.

Table 3. Accuracy test related to 4 imputation methods of null-values.

K-points	Test Size	Accuracy (%)
1	30%	16,0%
2	**50%**	**33,3%**
3	40%	1,5%
7	30%	0,0%

5.2 Exploratory Data Analysis (EDA)

An EDA was performed to derive the main statistics linked to the indicators set. Figures 4, 5 and 6 show the correlogram of all the regional indicators and the bivariate distribution of the most correlated variables. The correlation coefficients, being just more than 90%, provide strong insights into the relationships between the variables, indicating both the strength and direction of the relationships.

In the Campania region, it can be noticed that 'HOU_04' is negatively correlated with 'ECO_07' at 91%. Considering the divergent polarity of these indicators, it highlights that as the percentage of people living in housing with noise from neighbours or the street increases, the value of separate waste collection tends to decrease accordingly. 'MOB_04' is strongly positively correlated with 'ECO_02' at 94% and negatively correlated with 'ECO_07' at 93%. 'ECO_01' is strongly positively correlated with 'ECO_03' and negatively correlated with the PM2 aerosol loading ('ECO_05') at 91%.

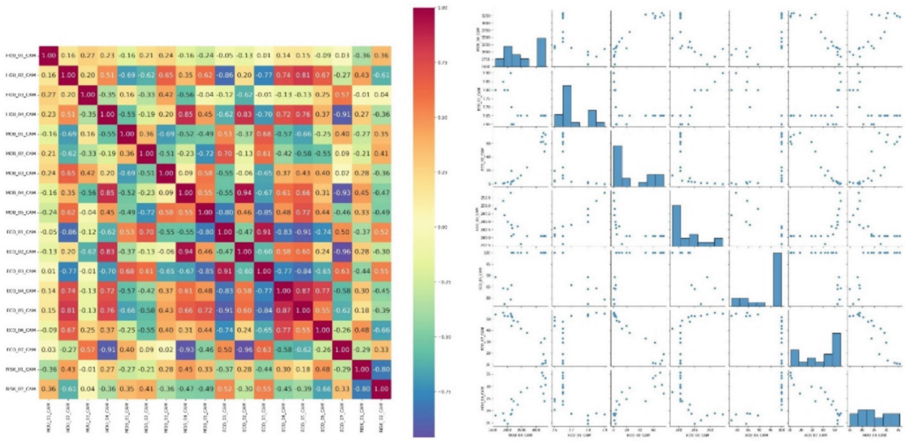

Fig. 4. The Campania region indicators correlation matrix (on the left) and a scatterplot of bivariate distribution (on the right).

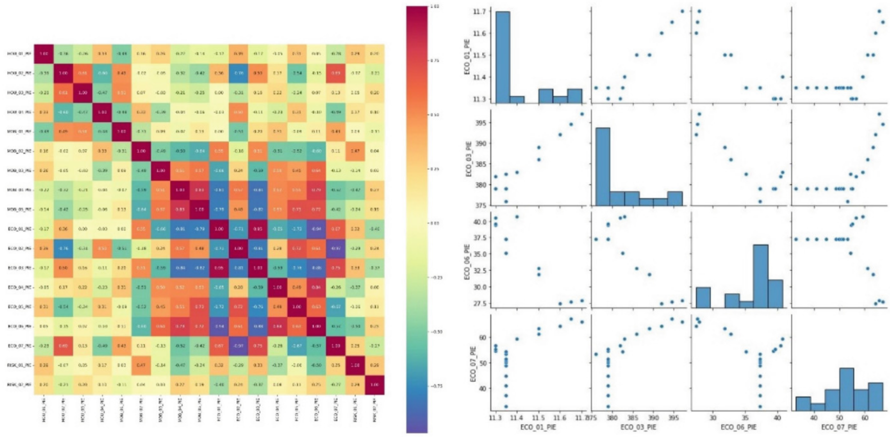

Fig. 5. The Piedmont region indicators correlation matrix (on the left) and a scatterplot of bivariate distribution (on the right).

In the Piedmont region, indicators ‘ECO_01’, ‘ECO_03’, ‘ECO_06’, and ‘ECO_07’ are the most correlated within the dataset. Specifically, ‘ECO_01’ is positively correlated with ‘ECO_03’ at 95% as for the Campania region, while a lower negative correlation level can be observed comparing both the variables with ‘ECO_07’. Furthermore, indicator ‘ECO_01’ is strongly negatively correlated with ‘ECO_06’ at 94%.

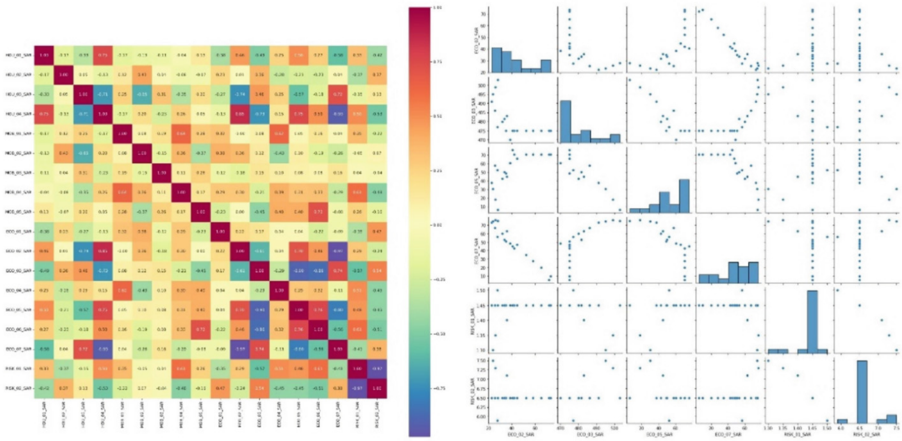

Fig. 6. The Sardinia region indicators correlation matrix (on the left) and a scatterplot of bivariate distribution (on the right).

In the Sardinia region, indicators ‘ECO_07’ and ‘ECO_02’ are strongly negatively correlated at 97%, as well as ‘ECO_03’ and ‘ECO_05’ at 90%. The indicators related to flooding ‘RISK_01’ and ‘RISK_02’ are negatively correlated as a decreased exposure to landslide risk due to depopulation of inland areas is expected. Indicators 'ECO_07' and 'HOU_04' are also negatively correlated. Overall, no positive correlations above 90% are found.

6 Discussion and Conclusion

The contribution has proposed an assessment framework to compare Italian regions according to a common set of indicators, with the aim of identifying normative, assessment, and methodological gaps in monitoring progress towards SDG 11 at the global, European, and Italian levels. While Machine Learning has not completely solved the null values issues, it serves as a powerful tool to predict data and provide analysts with complete dataset to be manipulated. On the other hand, EDA sparks a data-supported interpretation of the most interrelated phenomena concerning urban sustainability.

Results of EDA conducted on the three mentioned Italian regions have provided some meaningful insights for interpreting the correlation between variables. In particular, it can be noticed from the Campania region correlogram that whether the correlation levels among these indicators could be spurious, it could suggest that as the passenger-km offered by public transport increases, the amount of urban waste in landfills tends to decrease, probably due to greater awareness of environmental sustainability or the availability and functioning of public transport. The incidence of urban green areas on the urbanised surface of cities positively affects the soil sealing and land consumption per capita as well as the air quality in the cities.

The results for the Piedmont region, instead, have shown high levels of positive correlation between the two above-mentioned ecological indicators respectively referred to as 'Incidence of urban green areas on the urbanised surface of cities' and 'Soil sealing and land consumption per capita'. These variables equally increase in the considered time span. It could depend on several factors, such as: data inconsistencies linked to low accuracy of KNN predictions, unbalanced development of detected green areas in different regional zones as well as spurious correlations between indicators. On the other hand, the negative correlation between urban green areas and the percentage of nitrogen dioxide is evident, nevertheless it can be further investigated by introducing a third variable concerning the incidence of free traffic zones that contribute to reducing pollutants. Thus, a deeper exploration of strongly correlated measures to improve environmental sustainability is needed.

Furthermore, the Sardegna region correlogram highlights specific phenomena to be interpreted: the decrease in the population living in areas exposed to noise pollution is related to the increase in separate collection. In fact, noise pollution in urban areas is often related to commercial and nightlife areas, with heavy vehicular traffic or nearby transportation infrastructure such as highways, railways, ports, airports or industrial areas. In these areas it can be more difficult to achieve high levels of separate collection. The decrease in urban waste in landfills, as can be expected, appears to be linked to an increase in separate collection. Moreover, the decrease in population exposed to landslide risk is linked to an increase in population exposed to flood risk. This phenomenon could be explained by a gradual migration of inhabitants from the more mountainous and landslide-exposed inland areas downstream and into the coastal areas more exposed to flood risk.

In conclusion, SDG 11 stands out as a pivotal framework for addressing the diverse challenges faced by cities, which progress monitoring cannot be underestimated, as it provides crucial insights for informed decision-making and supports multidimensional transitions within European cities.

Although it has been noted that indicators may be an inadequate tool for assessing progress toward SDG 11 at the global scale due to "a lack of benchmarks, targets, and explicit measurement of equity considerations" [21], this contribution provides evidence regarding the effectiveness of the proposed framework in performing comparative assessments at the regional level, due to the potential of using indicators not only as monitoring tools but also as evaluation standards for city planning to accelerate multidimensional transition, to detect from the comparison whether and how to act [22], and to push regions to confront not considering poor performance as punitive measures but rather as a spur to improvement in the rationale that "no one should be left behind".

Moreover, data-driven approaches, including Exploratory Data Analysis (EDA) and Machine Learning algorithms, offer robust frameworks for measuring, analyzing, and interpreting indicators, thereby empowering stakeholders to track urban development indicators and tailor strategies for equitable and sustainable urbanization [23].

However, challenges such as data gaps and inconsistencies in assessment frameworks persist, hindering effective monitoring of SDG 11 progress. Overcoming these gaps is essential and it requires a concerted effort to enhance the consistency, accuracy, and comprehensiveness of SDG 11 indicators-based assessments, which is one of the main goal of the GLOSSA project. Standardizing assessment frameworks and methodologies, improving data collection and sharing mechanisms, and employing robust validation techniques are crucial steps towards mitigating these issues.

Acknowledgments. The contribution is part of the Research Project of National Relevance (PRIN) "GLOSSA - GLOcal knowledge System for Sustainability Assessment of urban projects", coordinated by POLITO, Principal Investigator: Francesca Abastante; UNINA research unit leader: Giuliano Poli; UNICA research unit leader: Francesco Piras.

Authors Contributions. G.P.: Writing – review & editing, Writing – original draft, Validation, Methodology, Conceptualization, Data curation, Visualization. S.C.: Writing – original draft, Data curation, Methodology, Visualization. E.M.: Writing – review & editing, Formal Analysis, Visualization.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. McArthur, J.W., Rasmussen, K.: Classifying sustainable development goal trajectories: a country-level methodology for identifying which issues and people are getting left behind. World Dev. **123**, 104608 (2019). https://doi.org/10.1016/j.worlddev.2019.06.031

2. Fallah Shayan, N., Mohabbati-Kalejahi, N., Alavi, S., Zahed, M.A.: Sustainable development goals (SDGS) as a framework for corporate social responsibility (CSR). Sustainability. **14**, 1222 (2022). https://doi.org/10.3390/su14031222
3. Lami, I.M., Abastante, F., Gaballo, M., Mecca, B., Todella, E.: Fostering sustainable cities through additional SDG11-related indicators. Valori e Valutazioni. **32**, 45–61. https://doi.org/10.48264/vvsiev-20233205
4. OECD, Union, E., Commission, J.R.C.-E.: Handbook on Constructing Composite Indicators: Methodology and User Guide. OECD Publishing
5. Carrillo, M.: Measuring progress towards sustainability in the European Union within the 2030 agenda framework. Mathematics. **10**, 2095 (2022). https://doi.org/10.3390/math10122095
6. Kitchin, R., Lauriault, T.P., McArdle, G.: Knowing and governing cities through urban indicators, city benchmarking and real-time dashboards. Reg. Stud. Reg. Sci. **2**, 6–28 (2015). https://doi.org/10.1080/21681376.2014.983149
7. Claveria, O., Monte, E., Torra, S.: A data-driven approach to construct survey-based indicators by means of evolutionary algorithms. Soc. Indic. Res. **135**, 1–14 (2016). https://doi.org/10.1007/s11205-016-1490-3
8. Morgenthaler, S.: Exploratory data analysis. WIREs Comput. Stat. **1**, 33–44 (2009). https://doi.org/10.1002/wics.2
9. Pearson, R.K.: Exploratory data analysis: a second look. In: Exploratory Data Analysis Using R, pp. 357–458. Chapman and Hall/CRC (2018)
10. Glänzel, W., Thijs, B., Schubert, A., Debackere, K.: Subfield-specific normalized relative indicators and a new generation of relational charts: Methodological foundations illustrated on the assessment of institutional research performance. Scientometrics **78**, 165–188 (2008). https://doi.org/10.1007/s11192-008-2109-5
11. Cerreta, M., Inglese, P., Malangone, V., Panaro, S.: Complex values-based approach for multidimensional evaluation of landscape. In: Murgante, B., Misra, S., Ana, M.A., Rocha, C., Torre, C., Rocha, J.G., Falcão, M.I., Taniar, D., Apduhan, B.O., Gervasi, O. (eds.) Computational Science and Its Applications – ICCSA 2014: 14th International Conference, Guimarães, Portugal, June 30 – July 3, 2014, Proceedings, Part III, pp. 382–397. Springer International Publishing, Cham (2014). https://doi.org/10.1007/978-3-319-09150-1_28
12. Cerreta, M., Mazzarella, C., Spiezia, M., Tramontano, M.R.: Regenerativescapes: incremental evaluation for the regeneration of unresolved territories in east Naples. Sustainability. **12**, 6975 (2020). https://doi.org/10.3390/su12176975
13. Khalid, A.M., Sharma, S., Dubey, A.K.: Data gap analysis, indicator selection and index development: a case for developing economies. Soc. Indic. Res. **148**, 893–960 (2019). https://doi.org/10.1007/s11205-019-02225-6
14. Liu, J., et al.: Nexus approaches to global sustainable development. Nat. Sustain. **1**, 466–476 (2018). https://doi.org/10.1038/s41893-018-0135-8
15. Greco, S.: Multiple Criteria Decision Analysis: State of the Art Surveys. Springer Science & Business Media, New York (2006). https://doi.org/10.1007/b100605
16. Ricciolini, E., et al.: Assessing progress towards SDGs implementation using multiple reference point based multicriteria methods: the case study of the European countries. Soc. Indic. Res. **162**, 1233–1260 (2022). https://doi.org/10.1007/s11205-022-02886-w
17. Benedek, J., Ivan, K., Török, I., Temerdek, A., Holobâcă, I.: Indicator-based assessment of local and regional progress toward the Sustainable Development Goals (SDGs): an integrated approach from Romania. Sustain. Dev. **29**, 860–875 (2021). https://doi.org/10.1002/sd.2180
18. Mazziotta, M., Pareto, A.: Measuring well-being over time: the adjusted mazziotta–pareto index versus other non-compensatory indices. Soc. Indic. Res. **136**, 967–976 (2017). https://doi.org/10.1007/s11205-017-1577-5

19. Caropreso, C., Di Salvo, C., Botte, M., D'Acierno, L.: A long-term analysis of passenger flows on a regional rail line. Int. J. Transport Dev. Integr. **1**, 329–338 (2017). https://doi.org/10.2495/tdi-v1-n3-329-338
20. Somma, M.: New tools to analyse the wastescapes of the cities: the case study of the metropolitan city of Naples. In: La Rosa, D., Privitera, R. (eds.) Innovation in Urban and Regional Planning: Proceedings of the 11th INPUT Conference - Volume 1, pp. 171–179. Springer International Publishing, Cham (2021). https://doi.org/10.1007/978-3-030-68824-0_18
21. Thomas, R., Hsu, A., Weinfurter, A.: Sustainable and inclusive – evaluating urban sustainability indicators' suitability for measuring progress towards SDG-11. Environ. Plann. B: Urban Anal. City Sci. **48**, 2346–2362 (2020). https://doi.org/10.1177/2399808320975404
22. Lehtonen, M., Sébastien, L., Bauler, T.: The multiple roles of sustainability indicators in informational governance: between intended use and unanticipated influence. Curr. Opin. Environ. Sustain. **18**, 1–9 (2016). https://doi.org/10.1016/j.cosust.2015.05.009
23. Sacco, S., Di Martino, F., Cerreta, M.: Smart circular cities and stakeholders engagement: a literature review to explore the role of artificial intelligence. In: Gervasi, O., Murgante, B., Ana, M.A., Rocha, C., Garau, C., Scorza, F., Karaca, Y., Torre, C.M. (eds.) Computational Science and Its Applications – ICCSA 2023 Workshops: Athens, Greece, July 3–6, 2023, Proceedings, Part V, pp. 239–258. Springer Nature Switzerland, Cham (2023). https://doi.org/10.1007/978-3-031-37117-2_18