Volume XVIII Number 3, 2026

Feeding Simulation Model: Data Acquisition to Achieve Climate Neutrality Goals

Mairita Zaķe, Andris Lapāns, Anda Mežgaile, Oskars Java

Vidzeme University of Applied Sciences, Valmiera, Latvia

Abstract

In line with the European Union's development plans, preserving a healthy environment for future generations is a key priority. Progress towards climate neutrality requires reliable, coherent, high-quality and interoperable data that capture historical trends and emerging societal dynamics while reflecting ecological, economic and technological systems. This paper addresses data acquisition and integration across five sectors: energy, agriculture, land use, land-use change and forestry, industry and waste management. The proposed framework supports Latvia's national climate neutrality simulation model with harmonised data. The analysis reveals substantial differences in data accessibility, formats, update regularity, and sectoral methodologies. Critical datasets are often available only in unstructured formats, such as PDFs of infographics, limiting automation in data collection. The developed modular workflow and harmonisation system improve model reliability and support evidence-based long-term climate policy design at national and global levels.

Keywords

Data mapping, data acquisition, climate neutrality, data model, data exchange.

Zaķe, M., Lapāns, A., Mežgaile, A. and Java, O. (2026) "Feeding Simulation Model: Data Acquisition to Achieve Climate Neutrality Goals", *AGRIS on-line Papers in Economics and Informatics*, Vol. 18, No. 3, pp. 137-146. ISSN 1804-1930. DOI 10.7160/aol.2026.180310.

Introduction

The European Union's climate policy aims to achieve climate-neutral Europe by 2050, which means the involvement of all member states. To monitor the progress of this process, several different models are developed, often covering one specific sector or a few of the sectors (for example, energy). It is emphasized that the need to work on harmonizing input data for better model use and provisions for cross-sector and cross-border integration (Mikropoulos et al., 2025; Samarasinghe et al., 2025). Assessing the impact of different drivers therefore requires climate-neutrality models capable of representing interactions across multiple sectors and system components. Such models are inherently complex, as they integrate diverse processes, assumptions, and data flows with a single analytical framework (Randall et al., 2007). As climate information may be collected from a variety of resources, based on different assumptions and different methodological limitations, many challenges for regional and national climate change, thus climate neutrality, may occur, such as quality of data and quality control, homogenization, data scarcity, gridding and structural uncertainty (IPCC, 2023).

Data acquisition challenges for climate neutrality simulation models manifest across multiple interconnected dimensions. The primary challenges span five domains: (1) data availability issues including inadequate collection and monitoring systems (Ulpiani et al., 2023) and inaccessible data formats (Zargar et al., 2024), (2) quality problems such as inaccurate emissions baselines (Ulpiani et al., 2023) and negative emission balance errors (O'Keeffe et al., 2023), (3) standardization gaps with inconsistent emission factor definitions across countries (Skujevska et al., 2024) and overly simplistic land-use representation (Prestele et al., 2017), (4) integration difficulties including incompatible data sources (Ghent et al., 2011) and lack of fully coupled models (Prestele et al., 2017), (5) and spatial-temporal resolution mismatches between model requirements and available data (Ghent et al., 2011; Prestele et al., 2017). These challenges stem from (1) methodological inconsistencies between countries (Prestele et al., 2017; Skujevska et al., 2024), (2) technical limitations in measurement systems (Schnase et al., 2016; Ulpiani et al., 2023), (3) institutional barriers in reporting frameworks (O'Keeffe et al., 2023; Zargar et al., 2024), (4) scientific uncertainties in process understanding (Figueroa & Farias, 2022; Ghent et al., 2011), and (5) resource constraints affecting infrastructure and capacity (Krysovatyy et al., 2024; Ulpiani et al., 2023). The long-term impacts of policies are often uncertain, making it challenging to design effective, far-reaching measures. Policy modelling tools can help to address all barriers to reduce the efforts of decision-making, but it is not possible to run a mathematical model without trusted data.

This publication is a part of a larger study that aims to develop a framework of a unified national policy decision-making support tool to simulate the impact of policy decisions on Latvia's progress toward climate neutrality by 2050. The project develops a dynamic simulation model using a system dynamics approach to interconnect key economic sectors and simulate policy impacts. Through multi-objective differential optimization, the model integrates technological, economic, social, and environmental dimensions to identify optimal pathways toward achieving climate neutrality. In this paper, we focus on data acquisition and analysis for five predefined economic sectors: (1) energy, (2) industry, (3) agriculture, (4) land use, land-use change and forestry (LULUCF), and (5) waste management.

Materials and methods

This study collects and integrates statistical data from Latvia, drawing on diverse datasets from various sources (institutional, regulatory, and statistical) for the period 2015-2023. The methodological approach, starting from the data availability analysis and finishing with the main principles of the data exchange prototype, is described in this paragraph.

Analysis of data availability

Data availability analysis includes both data sources mapping and evaluating whether specific data sources are suitable for use in modelling. It is determined whether the available data provides the necessary details, accuracy, and time coverage. The format of data availability is determined, for example, whether it is necessary to gather information from different reports, graphs, or tables, whether the data is available free of charge, for a license fee, or whether the data is classified as restricted access data. An analysis and description of what kind of data processing will be required is carried out.

Data collection and source verification

Data is collected from various official and reliable sources, such as the Central Statistical Bureau of the Republic of Latvia, the cadastral register of State Land Service of the Republic of Latvia, industry reports, and scientific studies. At this stage, the reliability of the sources, the relevance of the data, and their relevance to the requested modelling context are checked.

Metadata review is primarily performed to ensure that the data contains the information required for the modelling process.

Data categorisation and standardisation

The resulting data is often available in different classifications or formats. For example, one source may provide data on "healthcare buildings" while another source breaks them down as "buildings of ambulatory or inpatient medical facilities" and "hospitals". In this step, the data is transformed and combined to ensure its consistency.

Lastly, unit standardization is conducted (e.g., if different sources use different units, such as square meters or hectares, they are converted to a single unit suitable for the modelling process).

Data consolidation

If the data is divided into different categories or time periods, it is combined. For example, if the time dimensions in different sources do not match (e.g., one source offers data yearly, and another quarterly), data transformation is performed to ensure a uniform time dimension.

Combining data from multiple sources into one unified data set is performed. This can include both statistical summation and averaging to obtain representative data.

Data acquisition, transformation, and compilation methodology

If the data is available as a downloadable dataset – usually in CSV, JSON or Excel formats – importing into a database is accomplished as the first step. The workflow involves importing the data stored "as is". The next step is transformation, which provides data conversion for consolidation purposes. The values are not changed, rather measurement units are changed. If necessary, the recalculation based on measurement units is performed, or aggregate functions and grouping are performed.

Manual data entry into the database is conducted if the data is available in read-only mode as PDF reports or online publications.

Transformations are implemented using data processing tools (Python, SQL). Afterwards, integration and storage in a structured data model in the database takes place, which is followed by export and transfer for further use in modelling tools (individual CSV format file for each sector). Each CSV file is supplemented by metadata – detailed information about the used data sources, measurement units, time period, as well as the necessary actions to integrate the available output information into the common data model.

Data storage and exchange prototype

A data storage and exchange prototype is implemented to facilitate straightforward data consumption by end users and the simulation environment.

To ensure consistent access and interoperability, all cleaned and processed data is stored in a central database for long-term maintenance and controlled version management.

Results and discussion

Analysis of data availability

The sub-model for greenhouse gas (GHG) inventory is developed in accordance with the Intergovernmental Panel on Climate Change (IPCC) Guidelines and Standards (IPCC, 2019). The study primarily uses Latvia's officially calculated, internationally reported, and projected greenhouse gas (GHG) data, prepared in accordance with national and international legislation and the Cabinet on Ministers of Latvia Regulation No. 675 of 25 October 2022, "Procedures for the Establishment and Maintenance of the Greenhouse Gas Inventory, Projection, and Climate Change Adaptation Reporting Systems" (Cabinet of Ministers, 2022). The data acquisition is implemented in accordance with the Order of the Cabinet of Ministers of Latvia No. 610, Section 6.3 (Cabinet of Ministers, 2023). The sub-model identifies the data required for the operation of the decision-support system, with particular attention to the integration of new regulatory provisions and the analysis of policy proposals.

Data collection and source verification

The main data sources include the Latvian Environment, Geology and Meteorology Centre, the Central Statistical Bureau of the Republic of Latvia, and the Ministry of Climate and Energy of the Republic of Latvia, ensuring alignment with the IPCC Guidelines for National Greenhouse Gas Inventories. Table 1 summarizes the main data sources for climate neutrality monitoring in the analyzed sectors, indicating the data sources, time coverage, data format, and the most significant challenges in their use. The table provides an overview of data availability and identified limitations that need to be considered in climate policy modelling and monitoring.

The information summarized in the table shows that a wide range of official statistics and administrative data is available in Latvia for the purposes of monitoring climate neutrality. At the same time, common problems have been identified in several sectors, including data fragmentation between institutions, insufficient availability of machine-readable data, different time coverage, limited metadata and insufficient data interoperability between sectors. These challenges limit the use of data in integrated climate policy modelling and in assessing progress towards climate neutrality.

The metadata analysis led to the systematic identification of key data sources within each sector, accompanied by an assessment of their availability, level of detail, and compatibility with other datasets. The analysis revealed substantial inter-sectoral variation in data structure and degree of standardization. A unique identifier was assigned to each data source, enabling its consistent link to the metadata catalogue and the integrated database. This configuration supports precise attribution of data origin and ensures full traceability across all stages of data processing and integration. Consequently, a unified and structured overview of sector-specific data sources was established.

Sector	Main data sources	Temporal coverage	Data format	Key challenges
Energy	CSP statistical databases, State Construction Control Office energy performance data, State Land Service cadastral data, Public Utilities Commission heat supply reports and tariff database	Mostly annual time series; coverage varies by dataset, generally available for multiple years	CSV, Excel, online statistical tables; some data available only as PDF reports or administrative files	Fragmented data sources; some data available only in PDF format; limited access to confidential data; absence of systematic distinction between renovated and non-renovated buildings; inconsistent metadata and time-series documentation
Agriculture	CSP agricultural statistics, Rural Support Service administrative data, SUDAT/FADN data, agricultural annual reports, European Commission FADN/FSDN database	Varies across indicators; some key variables available only since 2019	CSV, Excel, online databases, PDF reports, administrative datasets	Incomplete long-term time series; missing indicators for manure management and enteric fermentation; limited methodological harmonisation; data dispersed across institutions; insufficient integration of economic and biophysical indicators
LULUCF	CSP forestry and land-use statistics, National Forest Register (State Forest Service), administrative datasets, EU climate finance databases, Eurostat fisheries statistics	Long-term statistical series available for major indicators; coverage varies for administrative datasets	CSV, Excel, online statistical tables, administrative databases, geospatial datasets	Limited access to detailed administrative and geospatial data; insufficient harmonisation of spatial datasets; methodological changes affecting time-series consistency; lack of unified access mechanisms for modelling purposes
Waste management	CSP waste statistics, Latvian Environment, Geology and Meteorology Centre waste databases, municipal waste management data, State Environmental Service registers, wastewater utility reports	Annual reporting cycles with varying historical coverage	Excel, CSV, databases, PDF reports	Data stored in multiple systems; limited automated access (lack of APIs); inconsistent integration with energy sector data; insufficient technical detail for modelling waste treatment processes and methane recovery
Industry	CSP industrial production statistics, energy balance statistics, fuel and biogas consumption data	Long-term annual statistical series	CSV, Excel, online statistical tables	Strong coverage of economic indicators but limited process-level activity data; insufficient detail on energy use and raw material consumption for emissions modelling; need for higher sectoral granularity
Macro-economics	CSP national accounts, demographic statistics, energy price statistics, regional GDP indicators	Long-term annual and periodic statistical series	CSV, Excel, online statistical tables	Missing or incomplete metadata; some data gaps and duplicate indicators; insufficient coherence with data structures of other sectors; limited possibilities for cross-sectoral integration

Source: Authors

Table 1: Main data sources for climate neutrality monitoring by sector.

Data categorisation and standardisation

Data quality was assessed with respect to completeness, consistency, comparability, and temporal coverage. Completeness was assessed by checking whether the dataset contained sufficient information for all required time periods and sectors, thereby eliminating incomplete or fragmentary information that could affect the results of the analysis. Consistency was analyzed across different data series and sources to identify potential discrepancies. Comparability was ensured by examining the alignment of the data with other national and international datasets, enabling reliable and verifiable analyses.

Data consolidation

Sectoral consistency was maintained by confirming that the classification and structuring of the data adhered to industry standards, ensuring compatibility with professionally used datasets. Additionally, time coverage was reviewed during the sourcing process to guarantee continuity throughout the entire period. Based on this assessment, an improvement plan was developed to address data gaps, refine inconsistent series, and define additional verification procedures. The resulting dataset and workflow are designed to remain interoperable with the Climate Policy and GHG Methodology Analysis modules.

Data acquisition, transformation, and compilation methodology

A dedicated analysis tier produces JSON-formatted output, which is processed in Python for data analysis, anomaly detection, and missing data imputation (Lapāns, 2025). For this purpose, a Row-Wise Z-Score Filtering method was developed. Row-level outliers are identified and removed using a configurable z_threshold, after which missing values are imputed column by column based on multivariate patterns across rows, treating each column with missing data as a target and the remaining columns as predictors. This procedure requires that rows contain compatible and comparable values, avoiding inconsistent units (e.g. mixing MW with MWh or EUR/MWh with kWh/m²).

As shown in the Table 2, the sectoral distribution of missing data reveals substantial heterogeneity in both completeness and temporal stability over the 2015–2023 period.

Sector	Range of missing data between 2015-2023 (%)	Average z-score (2015-2023)
Energy	27.45 - 73.86	1.66
Agriculture	5.30 - 64.80	-0.22
LULUCF	2.22 - 8.89	-0.52
Waste management	19.95 - 27.32	0.39
Industry	0.00 - 4.17	-0.76

Source: Authors

Table 2: Summary of missing data and z-score.

The industrial sector exhibits the most robust reporting profile, maintaining very low missingness throughout the study period and achieving complete data coverage in several years. The transport sector also shows a favourable trajectory, with missing data steadily decreasing from approximately 10% in the earlier years to full completeness in 2022, followed by only a minimal increase in 2023. In contrast, the waste management sector is characterised by persistently high missingness, remaining consistently above 19% and indicating long-term structural limitations in data collection or reporting. The LULUCF sector remains comparatively stable at a low to moderate level of missingness, although a gradual upward trend suggests a modest erosion in data completeness over time. The agricultural sector displays a markedly different pattern: while missingness remains moderate and relatively stable through 2022, the abrupt rise to 64.8% in 2023 constitutes a clear structural discontinuity and the most severe single-year deterioration among the analysed sectors. The energy sector represents the most consistently incomplete dataset overall, with exceptionally high missingness across all years; however, unlike the agricultural sector, it shows a gradual improvement over time, declining from 73.86% in 2015 to 27.45% in 2023. Taken together, these findings indicate that data quality is strongly sector-dependent, with the industrial sector and the transport sector demonstrating reliable reporting, the waste management and the energy sectors reflecting persistent systemic incompleteness, and the agricultural sector emerging as a critical outlier in the most recent year.

When using z-score to measure how much a certain score differs from the typical score (Berendrecht et. al., 2022), it indicates such large differences between years that additional attention should be paid to why there are inconsistencies in the data only for the energy sector, which is the only one where the z-score exceeds the threshold of 1.

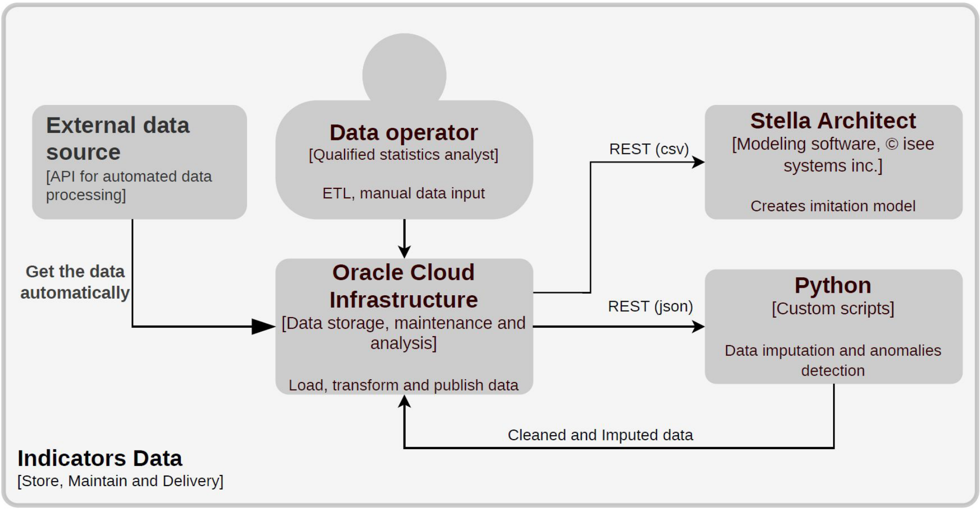
To handle heterogeneous data sources, both ETL (Extract, Transform, Load) and ELT (Extract, Load, Transform) workflows are employed. ETL is applied when sources are noisy or inconsistently structured, with transformations performed in Python before loading. ELT is used for well-structured sources, where data are first loaded into the database and then transformed using SQL and PL/SQL. Combining ETL and ELT ensures an adaptable and efficient pipeline for generating model-ready inputs.

Data storage and exchange prototype

Oracle Cloud Infrastructure (OCI) with an Oracle APEX 24.2.10 cloud-based database was selected as the integrated environment for pre-loading, processing, and preparing results for all indicator categories. This setup provides a unified, systematically organized data storage solution and supports simultaneous access by multiple users.

The integrated data retrieval mechanism (see Figure 1) offers two main outputs tailored to different use cases. First, a downloadable CSV file is generated and specifically formatted for direct use n the STELLA® modelling software, following the import and export specifications provided by Isee Systems (isee systems, 2025) to ensure error-free integration with the simulation model. Second, a visually rich HTML page presents the same data in a structured, user-friendly format, enabling interactive browsing and visual inspection of data quality. This interface allows users to assess completeness, identify anomalies, and explore trends across datasets.

Both outputs are delivered via RESTful web services, enabling online access and seamless integration into external systems or analytical workflows. This dual-output design maintains a clear separation between storage, processing, and application layers - data are securely stored in a central database, while REST-based interfaces



Source: Authors

Figure 1: A graphical overview of the end-to-end workflow for data collection, maintenance, and delivery, summarizing the processes described above in a single integrated scheme.

provide standardized, system-independent access for external tools. This architecture supports reliable data integration, scalability, and interoperability across the modelling workflow.

Key findings and data challenges

The findings of this study confirm that data acquisition for climate neutrality modelling in Latvia is a complex and resource-intensive task. This is mainly due to (1) fragmented data ownership, (2) heterogeneous formats, (3) limited availability of key datasets, and (4) restricted accessibility for researchers. Although various institutions collect important indicators, data resolution is often insufficient for comprehensive analysis. For example, total building floor area may be reported, but systematic data on technical condition, renovation status or energy performance class are not publicly available. Many relevant data sources are also published in PDF or Word formats (e.g., inventory reports) or as narrative text and infographics on institutional websites, which are unsuitable for automated extraction and must largely be processed manually. Automated web scraping is further constrained by dynamically rendered content and graphical images that hinder optical character recognition.

Additional challenges arise from data made available in forms that are not amenable to systematic, automated processing. For instance, information on renovated apartment buildings is often provided only as visual annotations on public GIS maps rather than downloadable structured files. Extracting this data requires bespoke scripts and introduces uncertainty, as points may be inaccurately positioned, lack metadata or don't correspond to actual building addresses, thus complicating spatial analysis and linkage to other datasets.

Heterogeneous formats, outliers, and incomplete records during transformation required careful handling. Nevertheless, cleaning and imputation significantly improved usability by standardizing input, reducing noise, and producing a dataset further suitable for modelling.

The overall quality of the processed data remains mixed. Some variables show low levels of missing values and anomalies, whereas others require extensive imputation and are therefore less robust. These issues reveal persistent structural weaknesses in Latvia's data governance environment, including the absence of unified national standards for statistical metadata and online publication formats, and limited integration between sectoral and cross-sectoral information systems. Although the data ecosystem has improved in recent years, substantial methodological and institutional advances are still needed to achieve fully interoperable, high-resolution datasets capable of supporting complex climate policy modelling.

Implications for climate neutrality modelling and policy

The structured data mapping system and metadata catalogue developed in this study provide a systematic overview of data inputs for climate neutrality modelling. By documenting (1) source institution, (2) temporal coverage, (3) territorial scope, (4) data format, (5) accessibility, and (6) methodological notes, of each dataset, the framework strengthens transparency and facilitates future extensions of the data infrastructure. This level of documentation is essential for reproducible modelling and for institutional learning about data governance.

The testing of anomaly detection and missing-value imputation frameworks using Row-Wise Z-score filtering and an iterative Random Forest–based imputer (Stekhovn and Bühlmann, 2012) demonstrates both the potential and the limits of automated quality enhancement. While historical patterns and cross-source correlations can improve completeness, their effectiveness is restricted by short and fragmented time series. For policy-relevant modelling, this implies that hybrid approaches—combining automated methods with rule-based corrections and expert judgement—are necessary to balance completeness and reliability. Importantly, gaps without sufficient contextual information were deliberately left flagged rather than filled, which supports honest communication of uncertainty.

The results show that re-imputed values are relatively robust for variables with limited data loss, but much less so for variables with high proportions of missing or flagged values. For climate neutrality modelling and policy analysis, this means that some indicators can be used with moderate confidence, while others should be treated as exploratory and subjected to sensitivity analysis or explicit caveats in scenario interpretation. Selective and data-conscious cleaning is therefore a critical component of responsible model building.

By compiling source-level and project-level metadata and linking them to a unified data model, the study creates a dataset that is (1) traceable, (2) interpretable, and (3) auditable. The integrated time series for 2015–2023 enables cross-source comparisons and cross-sectoral analyses and reveals an improving trend in metadata completeness over time. This strengthens the empirical foundation for an integrated national climate neutrality model.

Finally, the harmonized data from the five sectors—energy, industry, agriculture, land use, LULUCF, and waste management—provide the necessary inputs to parameterize and run the system dynamics (SD) model, capturing sectoral interactions and temporal dynamics. Implemented as a decision-impact support system with a user-friendly interface, the SD model will enable users to design potential climate policies and evaluate their effects on GHG emission reductions, required investments, and other key impacts. At the same time, the documented data limitations underscore the need for cautious interpretation of model outputs and for continuous improvement of the underlying data systems.

Looking ahead, the data framework and integrated dataset developed in this study provide a scalable foundation for further enhancement of Latvia's climate neutrality modelling capacity. Future work should focus on institutionalizing data mapping and quality assessment procedures, expanding machine-readable publication of key indicators, and progressively reducing reliance on manually processed sources. In parallel, iterative coupling between the SD model and the evolving data infrastructure can support continuous improvement of both modelling accuracy and the evidence base for climate policy design.

Conclusion

The results of this study highlight that improving data infrastructure for climate neutrality modelling in Latvia requires both institutional and technical measures. Based on the identified challenges and the experience gained in building the integrated dataset, the following recommendations are proposed.

Improve data standards and publication formats:

- Develop national standards for statistical metadata, units, and indicator definitions relevant to climate neutrality
- Require key datasets to be published in machine-readable, structured formats (e.g. CSV, JSON, GIS vector files) in addition to reports in PDF/Word
- Ensure better integration between sectoral and cross-sectoral systems through harmonized classifications and exchange formats

Implement automated and modular data ingestion:

- Begin with automated API ingestion using scheduled scripts or ETL tools to enable real-time or periodic updates from primary providers
- For structured files (CSV, XLSX, GIS exports), implement batch processing pipelines that validate, normalize, and convert data into a schema aligned with the integrated database model

- For unstructured sources, apply specialized parsing, OCR and, where appropriate, NLP for PDFs and reports; use web scraping frameworks with content extraction logic for web pages, subject to legal and ethical constraints
- Each ingestion module should include metadata tagging, error logging, and version control to ensure traceability and support quality control

Centralize and harmonize storage:

- Store all processed data in an integrated database, using standardized transformation rules and mapping logic to ensure coherence across formats and sources
- Maintain a clear separation between storage, processing, and application layers, allowing different modelling and analytical tools to access a single authoritative data source
- Strengthen quality management and transparency
- Apply a formal data quality assessment framework (completeness, consistency, comparability, accuracy) and clearly flag variables with extensive imputation or uncertainty
- Combine automated anomaly detection and imputation with expert review, specifically for critical indicators and fragmented time series.

Facilitate access for modelling and policy analysis:

- Provide model-ready outputs via RESTful web services and preformatted CSVs for specific tools (e.g. STELLA®)
- Offer simple web-based interfaces for exploring completeness, anomalies, and trends to support both modelers and policy users.

Taking together, these measures would address many of the structural and technical barriers identified in this study. Implementing them would move Latvia closer to having fully interoperable, high-resolution, and transparently documented datasets capable of supporting robust system dynamics modelling and evidence-based climate policy design.

Acknowledgements

This research was funded by the State Research Programme, Latvia project Climate Neutrality Decision Models in Action, VPP-KEM-Klimatneitralitāte-2023/1-0002.

Corresponding author:
Ph.D. Oskars Java
Vidzeme University of Applied Sciences, Tērbata street 10, Valmiera, LV-4202, Latvia
Phone: +37128678558, Email: oskars.java@va.lv

References

[1] Berendrecht, W., van Vliet, M. and Griffioen, J. (2022) "Combining statistical methods for detecting potential outliers in groundwater quality time series", *Environmental Monitoring and Assessment*, Vol. 195, No. 1, pp. 1-14. ISSN 1573-2959. DOI 10.1007/s10661-022-10661-0.

[2] Cabinet of Ministers. (2023) "*Par valsts pētījumu programmu "Klimatneitralitātes mērķu sasniegšanas lēmumu pieņemšanas atbalsta sistēma" 2023.– 2025*", gadam. Sep. 2023. [Online]. Available: https://likumi.lv/ta/id/345896-par-valsts-petijumu-programmu-klimatneitralitates-merku-sasniegsanas-lemumu-pienemsanas-atbalsta-sistema-2023-2025. [Accessed: Sep. 3, 2025]. (In Latvian).

[3] Intergovernmental Panel on Climate Change (IPCC). (2019) "*2019 refinement to the 2006 IPCC guidelines for national greenhouse gas inventories*". IPCC. [Online]. Available: https://www.ipcc.ch/report/2019-refinement-to-the-2006-ipcc-guidelines-for-national-greenhouse-gas-inventories/. [Accessed: Sep. 3, 2025].

[4] Intergovernmental Panel on Climate Change (IPCC). (2023) "Linking Global to Regional Climate Change", In *Climate Change 2021 – The Physical Science Basis: Working Group I Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, pp. 1363-1512). ISBN 9781009157896. DOI 10.1017/9781009157896.

[5] Figueroa, V. S. and Farias, C. B. (2022) "Carbon Neutrality Scenario Projection Model under Uncertainty", *2022 IEEE International Conference on Automation/XXV Congress of the Chilean Association of Automatic Control (ICA-ACCA)*, Vol. 2022, pp. 1-6. E-ISBN 978-1-6654-9408-3. DOI 10.1109/ICA-ACCA56767.2022.10005931.

[6] Ghent, D., Kaduk, J., Remedios, J. and Balzter, H. (2011) "Data assimilation into land surface models: The implications for climate feedbacks", *International Journal of Remote Sensing*, Vol. 32, No. 3, pp. 617-632. E-ISSN 1366-5901. DOI 10.1080/01431161.2010.517794.

[7] Isee systems. (2025) "*Creating Import and Export Files*". [Online]. Available: https://www.iseesystems.com/resources/help/v3/#03-buildingmodels/importing_and_exporting_data/creating_import_and_export_files.htm. [Accessed: Oct. 17, 2025].

[8] Krysovatyy, A., Maksymova, I., Kurilyak, V., Radin, M. and Kurilyak, M. (2024) "International convergence towards a climate-neutral economy: modeling the agricultural sector", *AgEcon SEARCH*, Vol. 10, No. 2, pp. 52-79. ISSN 2414-584X. DOI 10.22004/ag.econ.355962.

[9] Lapāns, A. (2025) "*Energy Data Imputation and Anomaly Detection System*", Nov. 2025. [Online]. Available: https://Github.Com/Ala-Hub-25/Vpp-Kem/. [Accessed: Nov. 12, 2025].

[10] Mikropoulos, E., Roelfsema, M., Chen, H. H., Staffell, I., Oreggioni, G., Hdidouan, D., Thellufsen, J. Z., Chang, M. A., Fragkos, P., Giannousakis, A., Chatterjee, S., Ürge-Vorsatz, D., Pfenninger, S., Pickering, B., Victoria, M., Brown, T. and van Vuuren, D. P. (2025) "Examining pathways for a climate neutral Europe by 2050; A model comparison analysis including integrated assessment models and energy system models", *Energy*, Vol. 319. pp. 1-18. E-ISSN 1873-6785. DOI 10.1016/j.energy.2025.134809.

[11] O'Keeffe, S., O'Sullivan, D. and Bruton, K. (2023) "Development of a modelling tool to aid the transition to carbon neutral industrial manufacturing", *Journal of Cleaner Production*, Vol. 425, pp. 1-19. ISSN 0959-6526. DOI 10.1016/j.jclepro.2023.138604.

[12] Prestele, R., Arneth, A., Bondeau, A., de Noblet-Ducoudré, N., M Pugh, T. A., Sitch, S., Stehfest, E. and Verburg, P. H. (2017) "Current challenges of implementing anthropogenic land-use and land-cover change in models contributing to climate change assessments", *Earth System Dynamics*, Vol. 8. No. 2, pp. 369-386. E-ISSN 2190-4987. DOI 10.5194/esd-8-369-2017.

[13] Cabinet of Ministers. (2022) "*Procedures for Establishing and Maintaining the System for Greenhouse Gas Inventories, the Projection System, and the System for Reporting on the Adaptation to Climate Change*". [Online]. Available: https://likumi.lv/ta/en/en/id/336733-procedures-for-establishing-and-maintaining-the-system-for-greenhouse-gas-inventories-the-projection-system-and-the-system-for-reporting-on-the-adaptation-to-climate-change. [Accessed: Nov. 10, 2025].

[14] Randall, D. A., Wood, R. A., Bony, S., Colman, R., Fichefet, T., Fyfe, J., Kattsov, V., Pitman, A., Shukla, J., Srinivasan, J., Stouffer, R. J., Sumi, A. and Taylor, K. E. (2007) "Climate models and their evaluation", In Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K. M. Tignor, M. and Miller, H. L. (Eds.) "*Climate change 2007: The physical science basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change",* pp. 589-662. Cambridge University Press. ISBN 978 0521 88009-1.

[15] Samarasinghe, S. L., Moghimi, M. and Kaparaju, P. (2025) "A review of modelling tools for net-zero emission energy systems, based on model capabilities, modelling criteria and model availability", *Renewable Energy Focus*, Vol. 53, pp. 1-24. ISSN 1755-0084. DOI 10.1016/j.ref.2024.100659.

[16] Schnase, J. L., Lee, T. J., Mattmann, C. A., Lynnes, C. S., Cinquini, L., Ramirez, P. M., Hart, A. F., Williams, D. N., Waliser, D., Rinsland, P., Webster, W. F., Duffy, D. Q., McInerney, M. A., Tamkin, G. S., Potter, G. L. and Carrier, L. (2016) "Big Data Challenges in Climate Science", *IEEE Geoscience and Remote Sensing Magazine*, Vol. 4, No. 3, pp 10-22. E-ISSN 2168-6831. DOI 10.1109/MGRS.2015.2514192.

[17] Skujevska, D., Feofilovs, M. and Romagnoli, F. (2024) "Challenges in Standardizing Global Emission Factors for Peatlands", *CONECT. International Scientific Conference of Environmental and Climate Technologies*, pp. 44-45. E-ISSN 2592-9704. DOI 10.7250/conect.2024.028.

[18] Stekhoven, D. J. and Bühlmann, P. (2012) "MissForest: non-parametric missing value imputation for mixed-type data", *Bioinformatics*, Vol. 28, No. 1, pp. 112-118. E-ISSN 1367-4811. DOI 10.1093/bioinformatics/btr597.

[19] Ulpiani, G., Vetters, N., Melica, G. and Bertoldi, P. (2023) "Towards the first cohort of climate-neutral cities: Expected impact, current gaps, and next steps to take to establish evidence-based zero-emission urban futures", *Sustainable Cities and Society*, Vol. 95. pp. 1-20. E-ISSN 2210-6715. DOI 10.1016/j.scs.2023.104572.

[20] Zargar, S., Kannangara, M., Gonzales-Calienes, G., Yang, J., Shadbahr, J., Decès-Petit, C. and Bensebaa, F. (2024) "Data Hub for Life Cycle Assessment of Climate Change Solutions—Hydrogen Case Study", *Data*, Vol. 9, No. 11. ISSN 2306-5729. DOI 10.3390/data9110129.