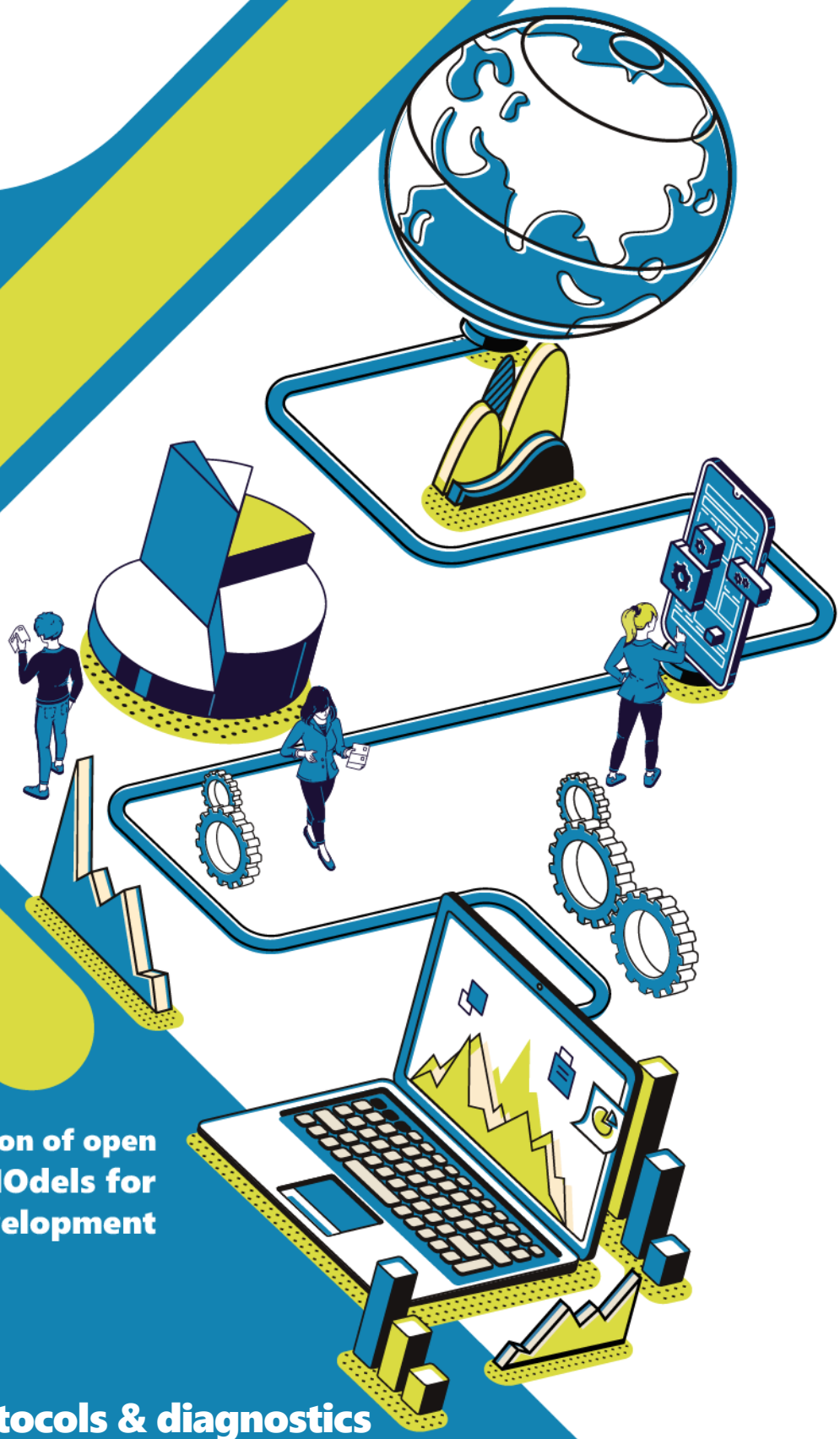
DIAMOND

www.climate-diamond.eu



Delivering the next generation of open Integrated Assessment MOdels for Net-zero, sustainable Development

D6.7 Open science protocols & diagnostics

WP6 – Open

30/11/2025

Funded by
the European Union

Disclaimer

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them.

Copyright Message

This report, if not confidential, is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0); a copy is available here: https://creativecommons.org/licenses/by/4.0/. You are free to share (copy and redistribute the material in any medium or format) and adapt (remix, transform, and build upon the material for any purpose, even commercially) under the following terms: (i) attribution (you must give appropriate credit, provide a link to the license, and indicate if changes were made; you may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use); (ii) no additional restrictions (you may not apply legal terms or technological measures that legally restrict others from doing anything the license permits).

Grant Agreement Number	101081179		Acronym	DIAMOND
Full Title	Delivering the next generation of open Integrated Assessment MOdels for Net-zero, sustainable Development			
Topic	HORIZON-CL5-2022-D1-02			
Funding scheme	HORIZON EUROPE, RIA – Research and Innovation Action			
Start Date	December 2022		Duration	48 Months
Project URL	http://www.climate-diamond.eu/			
EU Project Advisor	Silvia Vaghi			
Project Coordinator	Institute of Communications and Computer Systems - ICCS			
Deliverable	D6.7 – Open science protocols & diagnostics			
Work Package	WP6 – Open			
Date of Delivery	Contractual	30/11/2025	Actual	30/11/2025
Nature	Report		Dissemination	Public
Lead Beneficiary	Institute of Communications and Computer Systems - ICCS			
Responsible Author	Konstantinos Koasidis (ICCS)		Email	kkoasidis@epu.ntua.gr
	Georgios Xexakis (HOLISTIC)		Email	gxexakis@holisticsa.gr
Contributors	Stratis Alexandrou, Alexandros Nikas (ICCS)			
Reviewer(s)	Jan Ivar Korsbakken (CICERO); Jon Sampedro, Clàudia Rodés Bachs (BC3)			
Keywords	Open science; open pipelines; open licenses; model development; validation; vetting; diagnostics; ontologies; stakeholder validation			

EC Summary Requirements

1. Changes with respect to the DoA

No changes with respect to the work described in the DoA.

2. Dissemination and uptake

This report will be of primary interest to the IAM community, including members of the Integrated Assessment Modeling Consortium (IAMC) and modelling teams preparing scenarios for IPCC AR7, researchers developing or evaluating energy–climate–economy models, and scientific institutions advancing open science standards. Policy organisations, research institutions and initiatives developing scenario databases will benefit from the harmonised protocols and diagnostics. Stakeholders engaged in climate policy design, transparency initiatives, and FAIR data implementation can use the framework to strengthen trust, reproducibility, and societal relevance of model results.

3. Short summary of results (<250 words)

This report presents a comprehensive framework to enhance openness, transparency, and reliability in Integrated Assessment Models (IAMs) through dedicated open science protocols and diagnostics. Despite their central role in informing global climate policy, IAMs have long faced criticism over often appearing as "black boxes". Building on recent community efforts to improve openness of IAMs, standardise data, and align with FAIR principles, this deliverable aims to consolidate, strengthen, and operationalise best practices for open science within DIAMOND. The framework spans the full IAM value chain—from model development to model use and the release and validation of results. In the development phase, it integrates the 18 best practices for software engineering established in D6.6 and evaluates the openness of IAMs featured in the IPCC AR6. Despite recent progress, many models remain closed. DIAMOND addresses this gap by adopting fit-for-purpose licensing choices for its six IAMs, ensuring all are open-source and publicly available. For the model use phase, the report sets out a three-step pipeline: validation of input data and structures, vetting against global benchmarks and extended to EU-27 and member-state levels, and diagnostics to identify model behaviours and fingerprints across scenarios. Preliminary examples illustrate how these processes enhance robustness and comparability. The framework is accompanied by dedicated open tools that enhance automation of such processes. Finally, the framework emphasises open dissemination of tools, datasets, ontologies, and publications, alongside stakeholder engagement activities designed to strengthen societal trust. Together, these protocols aim to standardise and advance open science practices across the IAM community.

4. Evidence of accomplishment

This report, along with the validation tool (including both the validation and the vetting process) which is available on GitHub: GitHub - i2amparis/validation: An app for validating the results of climate-economy models, and the DIAMOND nomenclature also available on GitHub: https://github.com/CLIMATE-DIAMOND/diamond-nomenclature.

Preface

DIAMOND will update, upgrade, and fully open six Integrated Assessment Models (IAMs) that are emblematic in scientific and policy processes, improving their sectoral and technological detail, spatiotemporal resolution, and geographic granularity. It will further enhance modelling capacity to assess the feasibility and desirability of Paris-compliant mitigation pathways, their interplay with adaptation, circular economy, and other SDGs, their distributional and equity effects, and their resilience to extremes, as well as robust risk management and investment strategies. This will be done via integration of tools and insights from psychology, finance research, behavioural and labour economics, operational research, and physical science. The project will develop a transdisciplinary scientific approach to legitimise the implementation process and co-create research questions that stretch the frontiers of climate science, as well as establish vibrant communities of practice to transparently open model enhancements and to develop capacities, thereby lowering the entrance barriers to the established IAM community.

ICCS	INSTITUTE OF COMMUNICATIONS AND COMPUTER SYSTEMS	EL
BC3	ASOCIACION BC3 BASQUE CENTRE FOR CLIMATE CHANGE - KLIMA ALDAKETA IKERGAI	ES
CESAR	KRATENA KURT	AT
CICERO	CICERO SENTER FOR KLIMAFORSKNING	NO
CYI	THE CYPRUS INSTITUTE	CY
E4SMA	ENERGY ENGINEERING ECONOMIC ENVIRONMENT SYSTEMS MODELING AND ANALYSIS SRL	IT
HOLISTIC	HOLISTIC IKE	EL
COMILLAS	UNIVERSIDAD PONTIFICIA COMILLAS	ES
ISINNOVA	ISTITUTO DI STUDI PER L'INTEGRAZIONE DEI SISTEMI (I.S.I.S) - SOCIETA'COOPERATIVA	IT
SEURECO	SEURECO SOCIETE EUROPEENNE D'ECONOMIE SARL	FR
UM	UNIVERSITEIT MAASTRICHT	NL
ESMIA	ESMIA CONSULTANTS INC.	CA
USMF	THE UNIVERSITY OF MARYLAND FOUNDATIION INC	US
UMD	UNIVERSITY OF MARYLAND	US
EPFL	ECOLE POLYTECHNIQUE FEDERALE DE LAUSANNE	CH
ETH	EIDGENOESSISCHE TECHNISCHE HOCHSCHULE ZUERICH	CH
UNIBAS	UNIVERSITAT BASEL	CH
Imperial	IMPERIAL COLLEGE OF SCIENCE TECHNOLOGY AND MEDICINE	UK
Oxford	THE CHANCELLOR, MASTERS AND SCHOLARS OF THE UNIVERSITY OF OXFORD	UK
UCL	UNIVERSITY COLLEGE LONDON	UK

Executive Summary

Integrated assessment models (IAMs) and other energy–climate–economy tools have become essential components of global climate research, providing long-term scenarios that inform major assessments such as those of the IPCC and supporting climate mitigation strategies and policy design. At the same time, the complexity of IAM structures and assumptions has sometimes contributed to the perception that they function as "black boxes", underscoring the value of continued efforts to strengthen transparency, accessibility, and clarity in how scenarios are generated, communicated, and evaluated. Recently, the community has made notable progress in openness through expanded documentation, FAIR data practices, and the introduction of tools that support reproducibility. Within this context, the DIAMOND project sets out a comprehensive and practical framework for open science protocols, validation and vetting procedures, diagnostics, and shared ontologies designed specifically for IAMs.

The DIAMOND project's approach is positioned within the broader principles of open science, understood as the commitment to make scientific resources, workflows, and outputs as accessible and reusable as possible. In the context of IAMs, this includes integrating the 18 best practices for open software pipelines developed in Deliverable D6.6. Here, we examine in detail the practicalities of applying different open licences to the project's six IAMs. A review of IAMs featured through at least one scenario in AR6 shows an encouraging trajectory toward openness, albeit with slow improvements, variation across modelling groups, and significant areas for further progress. DIAMOND contributes meaningfully to advancing openness by ensuring that all models are publicly available, with licences tailored to the technical and institutional circumstances of each model and team. The rationale for each licensing choice is clearly documented, providing a practical reference for other groups navigating similar decisions.

A major contribution of the framework lies in its structured process for model use, comprising validation, vetting, and diagnostics. First, the validation process aims to ensure internal consistency by checking units, naming conventions, aggregation rules, and other structural aspects of model output. The IAMC data template and ontology, enhanced through the DIAMOND project's own nomenclature, play a central role in this step, enabling more efficient communication and reducing the likelihood of errors arising from inconsistent terminology. Second, the vetting process draws on benchmark ranges and historical values to assess whether model outputs fall within expected bounds. This process follows the IPCC AR6 vetting protocol, with our project's main contribution being the establishment of a detailed procedure for vetting at the EU27 and member-state (MS) level. We developed benchmarks based on Eurostat data, specified acceptability ranges, and elaborated on refinements needed relative to the IPCC approach. Illustrative examples using three DIAMOND models showed that the models fell well within the defined ranges, providing an additional layer of validation and demonstrating how this pipeline identifies issues and supports model improvement, particularly at the MS level. Third, the diagnostics process complements validation and vetting by analysing model behaviour across baseline and policy scenarios. This step highlights characteristic tendencies that arise from structural differences between models. These insights support clearer interpretation of scenario outcomes and strengthen the transparent communication of modelling assumptions. To assist automatic processing of scenarios along this pipeline, the open "IAM data validation app" has been developed and is continuously updated.

The framework concludes by describing how DIAMOND openly releases its datasets, tools, ontologies, and publications in line with the project's machine-actionable data management plan. The emphasis on openness also extends to stakeholder engagement, recognising that transparency involves not only technical accessibility but also clear communication and dialogue with policymakers, civil society, and other scientific communities.

Funded by
the European Union

Contents

1 Introduction 1
2 Framework for Open Science Protocols and Diagnostics 2
3 Developing open models 5
3.1 The 18 practices for open-source software development pipelines 5
3.2 A deep dive in the choice of licenses for open models 6
3.2.1 Common available open licenses 6
3.2.2 Openness of commonly used IAMs 8
3.2.3 Licenses selected for the six DIAMOND IAMs 10
3.3 Harmonisation of model development 12
4 Transparent Model Use 14
4.1 Harmonisation of Model Use & Ontologies 14
4.2 Automated Validation Routines 16
4.2.1 Methodological approach 16
4.2.2 Software Tool 18
4.2.3 Indicative Results 20
4.3 Global and Regional Vetting Process 22
4.3.1 Theory and methodological approach 22
4.3.2 Software Tool 25
4.3.3 Indicative Results 25
4.4 Model Diagnostics 29
4.4.1 Theory, methodological approach and stylised algorithms 29
4.4.2 Indicative Results 34
5 Scientific Outreach & Stakeholder Validation 38
5.1 Sharing resources with the scientific community 38
5.2 Open access scientific publications and data sharing 38
5.3 Stakeholder validation 39
6 Conclusions and next steps 41
References 42
Annex 46

Table of Figures

Figure 1. The three stages of the Framework for Open Science Protocols and Diagnostics 2
Figure 2. Screenshot from choosealicense.com presenting common use cases for software licenses 7
Figure 3. Categorisation of models featured in IPCC AR6 based on openness 10
Figure 4. A part of the YAML file tree describing DIAMOND variable nomenclature 16
Figure 5. A part of the OWL file tree describing DIAMOND variable nomenclature visualised using the Protégé software. 16
Figure 6. Screenshot from the first version of the validation tool developed in 2023 19
Figure 7. Screenshot from the second version of the validation tool developed by CICERO for the IAM COMPACT project 20

Figure 8. Screenshot of the validation tool with a test dataset uploaded and checked for IAMC template validity 21
Figure 9. Screenshot of the validation tool with the available checks for name consistency 22
Figure 10. Screenshot of an indicative check of variable names in the validation tool 22
Figure 11. Performance of each model in the eight variables selected for the regional vetting process. 28
Figure 12. Results for the RAI indicator 35
Figure 13. Combined evolution of carbon-energy intensity reduction (left) and results from the ERT indicator (right) 35
Figure 14. Results from the FFR indicator based on GCAM-Europe (left) and GEMINI-E3-EU (right) 36
Figure 15. Results from the TI indicator based on GCAM-Europe (left) and GEMINI-E3-EU (right) 36
Figure 16. Policy costs vs. Relative Abatement until 2030 in the GEMINI-E3-EU model 37
Figure 17. The APOLLO-Live open-source webtool for decision making and consensus support in energy and climate policy. 40

Table of Tables

Table 1. Software engineering practices for five different phases of IAM development and release 5
Table 2. Permissions, conditions, and limitations of licenses used in DIAMOND models 7
Table 3. Scoping the openness of models included in the IPCC AR6 8
Table 4. Selection of licenses for the six DIAMOND IAMs 11
Table 5. Summary of vetting variables, reference values and criteria in the IPCC AR6. 23
Table 6. Summary of vetting variables, reference values, sources and criteria for the regional EU27 vetting 24
Table 7. Vetting results for the EU27 region. 27
Table 8. Description of indicators used in the diagnostics protocol 29
Table 9. GEMINI-E3-EU vetting at a higher regional granularity within the EU 46

Disclaimer on the use of AI

ChatGPT (versions 4 and 5) was used only to refine language and improve the clarity of text originally produced by the authors. All content, data, and conclusions were produced, reviewed, and after language refinement verified by the authors. Use of AI did not influence the reasoning, substantive analysis, or findings presented in this work.

1 Introduction

Integrated assessment (IAMs) and other energy-climate-economy models have been at the forefront of climate science, underpinning major scientific assessments—such as those produced by the Intergovernmental Panel on Climate Change (IPCC)—by developing scenarios of plausible futures to assist climate change mitigation efforts and climate policy design and implementation (Weyant, 2017). However, IAMs have often been criticised of risking being seen as "black boxes", due to their complexity and occasionally inaccessibility and limited transparency in terms of input parameters (Wilson et al., 2021; Pfenninger et al., 2018; Gambhir et al., 2019). In recent years, the community has taken significant steps in "opening these black boxes" (Skea et al., 2021), with scientific teams opening the source code of many emblematic models, while developing new ones following open science principles, releasing them under open-source licenses, providing extensive documentation, offering training webinars, as well as embracing the use of Findability, Accessibility, Interoperability, and Reuse (FAIR) data. Yet, recent surveys on the utilisation of open science practices during model use (Rodés-Bachs et al., 2025) and development (see Deliverable D6.6) highlight significant challenges and barriers for the community to fully embrace "open science".

Although there is no formal definition of the term "open science" (Vicente-Saez and Martinez-Fuentes, 2018), it can broadly be viewed as the process of sharing resources and ideas in a way that makes them publicly and freely available for future use (Levin et al., 2016). Delving into and synthesising the assumptions and goals of the different schools of thought, we can view open science as a process towards making knowledge creation more efficient by working together (i.e., open code, data, cooperation networks etc.), creating openly available tools and platforms (see also Curley et al., 2025), making science accessible (including to non-scientists, such as citizens), making knowledge freely available, and revisiting the metrics used to measure scientific impact (Fecher and Friesike, 2013).

Offering guidelines towards embracing open science practices is heavily dependent on the research field of application (Crüwell et al., 2019; Levin et al., 2016). For DIAMOND, contextualising the notion of open science provides significant opportunities to both harmonising and even advancing the latest standards on openness within the field of IAMs. With the project's goal being to update, upgrade, and enhance core models of the community, ensuring the development of models that are "as open as possible, and as closed as necessary" (Landi et al., 2020; Schmitz and Seidl, 2023) is a core priority. At the same time, harmonising and expanding on processes established by the community, such as validating and vetting of global scenarios (Kikstra et a., 2022; Weigmann et al., 2025) to ensure consistency with historical values, as well as performing diagnostic exercises (Harmsen et al., 2021) to understand model fingerprints in produced pathways (i.e., how differences in modelling theories can impact results; Dekker et al., 2023), requires producing a dedicated framework for open science protocols and diagnostics that governs these activities from model development, to model use and result release.

In this context, the purpose of this deliverable is to develop a transparent framework for new protocols, ontologies, and diagnostics that will be used in DIAMOND, and that can be shared with the modelling community towards a standardisation of open science in IAMs and broader climate change mitigation studies. First, we revisit the 18 practices for open-source software development pipelines (see Deliverable D6.6) and examine license choices for the DIAMOND IAMs, reflecting on existing practices and gaps in the broader IAM field. Second, we present the validation, vetting, and diagnostic processes that will be followed in DIAMOND, updating established practices to fit the project's goals, such as extending global checks to region-specific analyses, and provide preliminary examples from ongoing stylised model runs. Finally, we discuss the development of ontologies, dissemination activities (including open resources, FAIR data, open access publications), and opportunities for stakeholder engagement throughout the project.

Funded by
the European Union

2 Framework for Open Science Protocols and Diagnostics

This section presents the Framework for Open Science Protocols and Diagnostics, which is designed to ensure transparency, consistency, and credibility throughout all modelling activities that will take place in DIAMOND (Figure 1). The framework structures the modelling process and workflow into three highly interconnected stages, namely Open Model Development, Transparent Model Use and Validation, and Scientific Outreach and Stakeholder Validation. These stages together provide a systematic approach to open and reproducible integrated assessment modelling. The protocol builds on principles from open-source software development (see D6.6), detailed diagnostic processes, FAIR data practices, and participatory scientific validation. Systematic operationalisation of this framework within the upcoming project activities aims to foster an environment where models and tools, data, and insights can be understood, scrutinised, and reused by the scientific community including communities of practice for each model (see D6.4), policymakers, as well as broader stakeholder communities (potentially including citizens). The following paragraphs provide a description of the theoretical framework, hinting at how it is expected to be operationalised in upcoming DIAMOND activities, on top of indicative yet extensive implementations within this report.

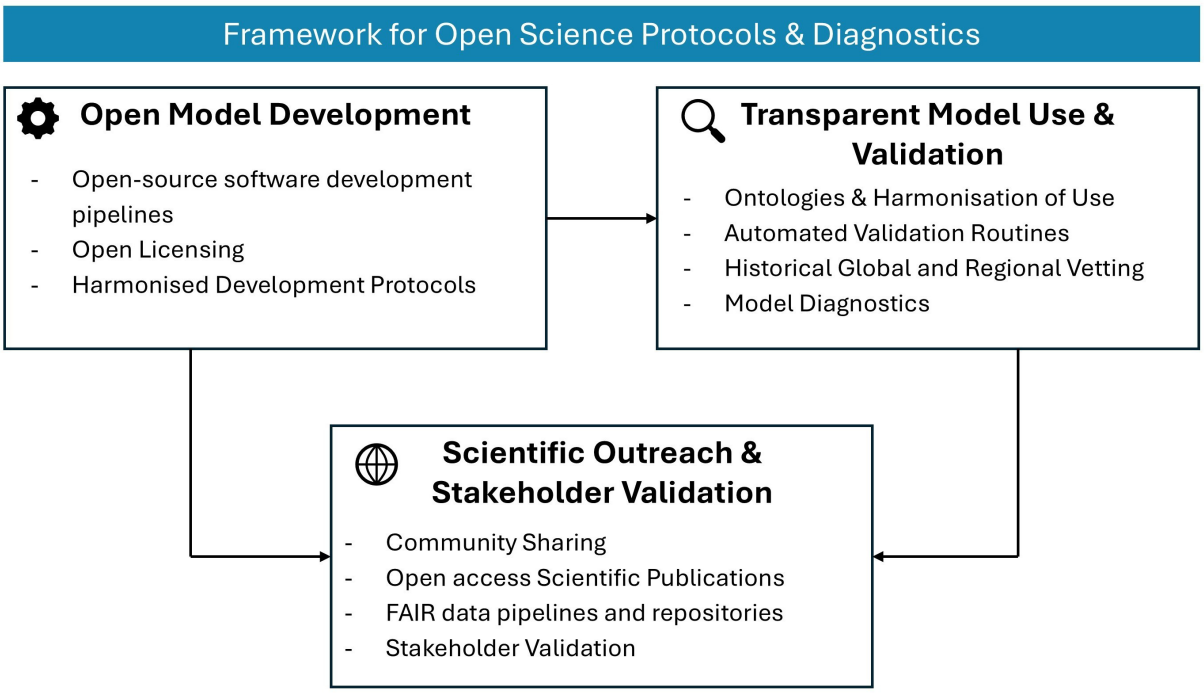


Figure 1. The three stages of the Framework for Open Science Protocols and Diagnostics

When describing the framework, the first stage focuses on how models are typically created, documented, and shared. For this, it draws on the elaboration of modern software engineering practices to the field of IAMs described in D6.6, and notably the 18 practices forming a framework of five critical phases in the lifecycle of IAM development. This open-source software development pipeline includes inter alia aspects related to version control systems, continuous integration (e.g., use of Git or the GitHub Actions), peer review of code changes, and extensive documentation of modelling assumptions and methods (for detailed descriptions see D6.6). Following these practices not only can improve technical robustness but also ensure that models can evolve through community input, while maintaining traceability of every modification and also delivering on the material that could potentially help reduce barriers to entry for less experienced modellers—especially from developing countries—at the early stages of their careers (Cannone et al., 2024; Lo Giudice et al., 2025; Howells et al., 2017; DeCarolis et al., 2012). A key enabler of openness is the use of open licensing. The presented framework requires

Funded by the European Union

that all model codes, documentation, and auxiliary tools developed in DIAMOND are released under open licenses that allow inspection, reuse, and adaptation, ensuring proper attribution, while also respecting the particularities of each model needs and requirements (i.e., related to intellectual property rights). Such selections promote interoperability across modelling teams and encourage the broader scientific community to engage with, test, and improve upon the models. Finally, harmonised development protocols provide a shared foundation across the project's modelling activities. These include common naming conventions, unit systems, data structures, and metadata standards, and also crucially the use of common authoritative datasets for key input assumptions (for more details see D3.2). Harmonising model development early in the process enables model coupling and comparison, and at the same time enables consistent interpretation of results when the models are used in multi-model intercomparison exercises (more details on the soft harmonisation process of model development in DIAMOND are presented in Section 3.3). Through these measures, the development phase ensures that models are not isolated artefacts but part of a coherent, accessible, and evolving modelling ecosystem.

The second stage of the framework governs how models can be used in practice and notably, how their outputs are checked and validated. To achieve this, the project employs shared ontologies and harmonised templates for model use based on the IAMC template, defining consistent representations of variables and parameters across all modelling activities (IAMC, 2024). These ontologies ensure semantic interoperability, facilitating a streamlined process to compare results within and across models. The framework introduces automated validation routines that systematically test model outputs for internal consistency (i.e., basic sum checks between aggregated and disaggregated variables) and numerical errors (i.e., missing values, proper checks of N/A versus 0 results etc.). A core part of these routines also includes historical vetting, a process to compare simulated values with recent observed data to ensure consistency at multiple regional scales. In particular, since DIAMOND focuses on developing models of higher spatial resolution, this process aims to go beyond global vetting established in the IPCC (Kikstra et al., 2022), to include vetting for the EU (at the aggregated, and depending on the model, at the member state level). Deviations are not intended to be used to discard results, but to prompt diagnostic investigation, helping researchers understand the sources of divergence and improve model reliability. Complementing validation, the framework outlines a process for performing model diagnostics to assess how models respond to specific climate policies. These diagnostic tests help identify structural sensitivities and behavioural differences based on the modelling theory followed by each model (e.g., partial vs general equilibrium, stochastic dynamic vs perfect foresight, etc.), providing interpretational insights into the strengths and limitations of each model configuration. Together, validation and diagnostics can establish a scientific quality-assurance layer that underpins the trustworthiness of the models and their outputs.

The final stage of the framework aims to ensure that the modelling process and its results are transparent and meaningful beyond the project. Openness at this stage concerns not only the dissemination of results but also the inclusion of external perspectives in assessing their plausibility and relevance. The framework promotes community sharing through open repositories and FAIR data principles. All models, ontologies, validation tools, and diagnostic routines are made openly available or are planned for an open release, enabling other researchers to replicate the processes to their own models and scenarios, thus extending project findings. Open-access scientific publications report on both the technical and conceptual advances achieved through the framework, while FAIR data pipelines guarantee that datasets used in modelling are traceable, properly annotated, and freely accessible (Nishikawa and Murakami, 2025). Crucially, the framework prescribes stakeholder validation as a structured feedback loop drawing from the transdisciplinary process established for stakeholder engagement (see D2.3 and D2.5), involving a wide range of actors from academia, policymaking, industry, and civil society (also including citizens). Stakeholders may assess whether model outputs are plausible, relevant, and comprehensible from their respective perspectives. Their input can add a vital layer of real-world validation that complements technical diagnostics, ensuring that the models remain policy-relevant and socially desirable.

Funded by
the European Union

The three stages should not be viewed as a linear process, but as an iterative loop that integrates and operationalises the principles of open, transparent, and participatory science. Notably, models are developed openly, used and tested transparently, and validated both by the scientific community and broader stakeholder groups. This process may require returning to the "model development and use" canvas to further enhance capacities and revisit assumptions. This cyclical process can foster continuous improvement, strengthen reproducibility, and build trust in the models and their results, ensuring that the project's modelling efforts contribute credibly to evidence-based policymaking and broader scientific progress.

In the following sections, we provide elaborations on the details of each part of the framework described here, which aim to guide operationalisation of the pipeline throughout the project activities. Specific numerical examples are also offered based on existing, yet tentative, scenario results, to highlight how the framework can work in practice.

3 Developing open models

3.1 The 18 practices for open-source software development pipelines

Many existing initiatives to improve the transparency and reproducibility of IAMs are inspired by good practices in the software engineering (SE) community, such as opening up models and creating vibrant communities around them (e.g., GCAM and MESSAGEix), formalising methods for evaluating modelling results (C. Wilson et al., 2021), establishing consistent result templates (Gidden & Huppmann, 2019), and improving data management (Stockhause et al., 2024). However, other practices have been less widespread in model development. Despite the ubiquity of agile software development in the industry, user-centric methods are rare in IAMs and relevant models (e.g., energy system models), as their development is often driven by modellers based on rigid planning rather than on frequent stakeholder interactions (Knapen et al., 2010). Similarly, other established SE practices are not the norm in IAM development such as automated code testing (Iwanaga et al., 2018), validation (Rodés-Bachs et al., 2025), or API methods for data exchange, despite many proposed methodologies for model linking (Keppo et al., 2026; Belete et al., 2017).

Deliverable D6.6 synthesised pertinent SE practices with the potential to improve the efficiency, transparency, and credibility of model development, especially on IAMs. The practices were collected from the literature on scientific software engineering (Heaton & Carver, 2015; Sandve et al., 2013; G. Wilson et al., 2017) and are adapted to modelling needs and requirements, building on previous efforts to distil best practices (DeCarolis et al., 2012, 2017). Overall, 18 practices have been identified and categorised into five phases within the lifecycle of IAM development (Table 1).

Table 1. Software engineering practices for five different phases of IAM development and release

Model development	Documentation	Evaluation	Versioning and collaboration	Deployment
Requirements Engineering	Conceptual documentation	Code testing	Version control	Dependency management
Practices for data processing	Code documentation	Model validation	Issue tracking	System requirements management
Programming paradigms, patterns, and standards	Data documentation	Continuous integration	Open-source development	Continuous deployment
Development lifecycle models	Tutorials and case studies			
Practices for refactoring and adding new features				

Source: Reproduced from DIAMOND Deliverable D6.6

The model development phase includes data processing methods (e.g., automated processing scripts under version control), methods for eliciting stakeholder requirements (requirements engineering) and refactoring model code, procedures for managing the lifecycle of development from project conception to delivery (e.g., agile development), as well as programming paradigms (e.g., object-oriented programming), design patterns (e.g., iterators in Python), and coding standards (e.g., PEP 8 in Python). The documentation phase includes various types of documentation used in modelling projects, e.g., conceptual descriptions of models, code comments, metadata,

and tutorials, while the evaluation phase includes code testing techniques (e.g., unit testing), processes for continuous integration (e.g., automated code testing), and methods for model validation (e.g., historical simulations such as hindcasting experiments). The framework also includes practices related to version control and issue tracking to support collaborative development as well as practices related to open-source development, e.g., proper licensing. Finally, we identify practices on the documentation, management, and automation of third-party dependencies and system requirements as well as practices related to the so-called continuous deployment, i.e., automation processes to reduce the time between development and release of a new software version. It is noted that the categorisation of the practices into five phases is indicative and mostly aims to help readers understand where these practices fit into the modelling lifecycle. For instance, practices such as code testing can equally be included under the Model Development phase.

More details on these practices as well as an evaluation survey on the current use of these practices among DIAMOND modelling teams can be found in Deliverable D6.6.

3.2 A deep dive in the choice of licenses for open models

The open science community has made significant progress in the development of a wide range of open licenses that are fit-for-purpose, depending on the needs of each research activity. In terms of data and documents, the different variants of the Creative Commons[1] licenses (CC) have been dominant, providing a standardised way to grant the public permission to use their creative work under copyright law. However, CC licenses—despite their widespread use—are not recommended for software, as they typically lack specific terms dedicated to software (e.g., distribution, modification, etc.). For this reason, the Open Source Initiative (OSI) has approved a set of licenses that are compatible with use for software and that comply with the Open Source Definition[2], offering a wide range of options in terms of permissions and restrictions.

Understanding the optimal choice for each DIAMOND IAM requires understanding the particularities of each case. The term "model" may imply that IAMs can be classified purely as software, yet the reality is that different models exhibit markedly different configurations. In most cases, models feature a model engine (the source code and part responsible for producing a solution), underlying input data, and supporting tools for different processes (i.e., data handling). As such different licenses may apply to different stages, depending also on the specific requirements of the model. For instance, GCAM-based models are largely integrated systems and can be more straightforwardly classified as software, whereas OMNIA, which is built on the TIMES model generator, is more data-centric in nature.

In the next sub-sections, we review the provisions of key OSI-approved open licenses and discuss current licensing practices within the IAM community, before presenting the final choices for license selection in DIAMOND.

3.2.1 Common available open licenses

By November 2025, OSI listed 122 licenses that comply with the Open Source Definition[3]. Among the most popular ones and/or with the strongest community using them are the Apache License, the Common Development and Distribution License, the Eclipse Public License, different versions of the GNU General Public License, the Mozilla Public License, the BSD Licenses, and the MIT License. The main differences between these licenses relate to the permissions and limitations that they convey to the software, along with specific conditions for its use. For instance, some of these licenses require users to disclose the original source of a software while others do not. One key difference between the licenses relates to whether the user is required to use the same license when releasing a new version of the software (or a derivative software) or they are free to choose another license. Based on this

[1] https://creativecommons.org/share-your-work/cclicenses/
[2] https://opensource.org/osd
[3] https://opensource.org/licenses

distinction, licenses are divided into two broad categories: copyleft licenses (requiring the same license in new versions) and permissive licenses (no condition on new versions). The final selection for a license depends on the software type and the use case. Figure 2 shows common use cases, including the need for a software license that will be relevant for a specific community, the need for a simple and permissive license, and the need for a license to ensure that new versions will be open (i.e., copyleft license).

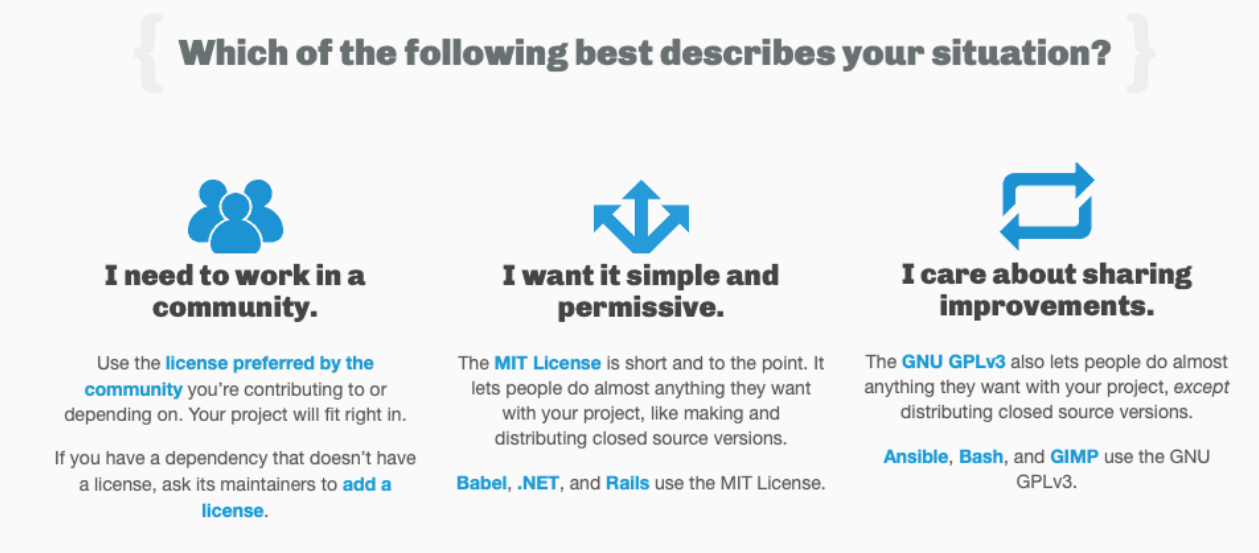


Figure 2. Screenshot from choosealicense.com presenting common use cases for software licenses

Table 2 shows the characteristics of common licenses that can be used in DIAMOND. These licenses are a subset of the common licenses presented before, selected based on an initial scoping with modelling teams. All these licenses allow commercial use, distribution, modification, and private use, require the users to use a license and copyright notice, and include a limitation of liability and warranty for the software. They vary mostly on conditions, with the GNU AGPL and GNU LGPL requiring next software versions to use the same licenses, while all the rest are permissive licenses. A comprehensive comparison of all common licenses can be found in choosealicense.com[4].

Table 2. Permissions, conditions, and limitations of licenses used in DIAMOND models

	Apache 2.0	ECL-2.0	GNU AGPLv3	GNU LGPLv2.1	MIT License
Permissions					
Commercial use	x	x	x	x	x
Distribution	x	x	x	x	x
Modification	x	x	x	x	x
Patent use	x	x	x		
Private use	x	x	x	x	x
Conditions					
Disclose source			x	x	
License and copyright notice	x	x	x	x	x
Network use is distribution			x		
Same license			x	x	
State changes	x	x	x	x	
Limitations					
Liability	x	x	x	x	x
Trademark use	x	x			
Warranty	x	x	x	x	x

Note: The table has been adapted from https://choosealicense.com/appendix/. The "x" indicates that the license includes the respective permission, condition, or limitation.

[4] https://choosealicense.com/appendix/

While these licenses cover most users, some use case require adaptations of a specific license. One such case is the distinction between commercial use and the right to sell a software that is provided in an open access license. To emphasise this distinction, the Commons Clause[5] can be added on top of an open-source license to ensure that a future user of the software cannot sell it. All other permissions, conditions, and limitations of the original license stay in place. In that way, a software can be used by a company but cannot be sold as a software itself. The drawback is that the new license is not considered as open source based on the OSI definition but rather "source-available" and upholds most of the other elements of the Open Source Definition.

Apart from software licenses, modelling applications often need other types of licenses to cover aspects beyond code, such as databases that are used as model inputs, for which (unlike databased related to outputs) choice of licensing may not always lie on model developers. Similar to software, many licenses exist to ensure that a dataset stays open access, with the most popular being the ones in the family of Creative Commons licenses[6]. This family includes six licenses in total with different permissions, conditions, and limitations. Starting from the most permissive one, CC BY, this license allows any type of use and adaptation as long as there is attribution to the creator. As variations of this license, the CC BY-SA stipulates that adaptations must be shared under the same terms (i.e., copyleft), the CC BY-NC restricts non-commercial use, while CC BY-ND does not permit derivatives or adaptations of a work. The remaining CC BY-NC-ND and CC BY-NC-SA are combinations of some of the other licenses.

3.2.2 Openness of commonly used IAMs

Since the inception of IAMs in the 1970s and their proliferation in the 1990s (van Beek et al., 2020), hundreds of models have been developed, featuring a wide range of characteristics and different methodological/theoretical approaches (Nikas et al., 2019; Belete et al., 2019). Yet, despite the more than 2,000 scenarios submitted in the IPCC AR6 (Kikstra et al., 2022), only a handful of models and/or modelling frameworks managed to produce at least one scenario that was included in the final report (Sognnaes and Peters, 2025).

Table 3 presents the models that were included in the IPCC AR6, along with their economic theory, and crucially the openness and availability of each model (e.g., license). Out of the 13 models with at least one scenario present in the AR6 of IPCC, only 4 models are based on OSI-approved licenses and can therefore be considered as open source, namely REMIND, MESSAGE, WITCH and GCAM. This indicates that most models underpinning major scientific assessments are still based on mostly fully closed proprietary models (Figure 3). At the same time, even among most open-source models, the execution environment remains closed, as it relies on proprietary solvers to operate (e.g., GAMS/CONOPT).

Table 3. Scoping the openness of models included in the IPCC AR6

Model	Economic Theory	License	Open Source
AIM	Computable general equilibrium	N/A	No
COFFEE	Linear programming optimization model	N/A	No
C-ROADS	System dynamics simulation	N/A	No (freely available only as an online simulator)

[5] https://commonsclause.com
[6] https://creativecommons.org/share-your-work/cclicenses/

EPPA	Computable general equilibrium	N/A	No
GCAM	Partial Equilibrium recursive dynamic	Educational Community License Version 2.0 (ECL)	Yes**
GEM-E3	Recursive dynamic computable general equilibrium	Currently developed version is based on AGPL-3.0 with Commons Clause	GEM-E3 was closed source, but the current OPEN-GEM version is source-available; reliance on GTAP
IMAGE	Partial Equilibrium; Simulation solution	N/A	No
MERGE-ETL	General Equilibrium; Intertemporal optimisation	N/A	No
MESSAGE	5-module modelling framework with an energy system basis, and a general equilibrium approach; optimisation solution	Apache-2.0 license	Yes (with dependency on GAMS, i.e., closed execution environment)
POLES	Global sectoral simulation model	N/A	No
REMIND	Energy-economy general equilibrium (Ramsey-type growth model coupled with an energy system model); cost optimisation solution	GNU Affero General Public License v3.0 (AGPL-3.0) with the REMIND Exception	Yes (with dependency on GAMS, i.e., closed execution environment)
TIAM*	Partial Equilibrium welfare cost-optimization	GPL-3.0 license (only for the TIMES code)	The TIMES model generator source code is open; dependencies on GAMS and the VEDA software for input/output handling, i.e., closed execution environment
WITCH	Inter-temporal optimal growth model; general equilibrium; optimisation solution	Apache-2.0 license	Yes (with dependency on GAMS)

Note:

*There are many versions of the TIAM model based on the use by each institute. Here we present the disaggregation of regions in the TIAM-ECN version (see Kikstra et al., 2022 for the breakdown of models in the IPCC), acknowledging that different versions might have some different levels of detail.

**Proprietary data are used, but are aggregated in chunks, enabling the model to be used irrespective of access to the source data

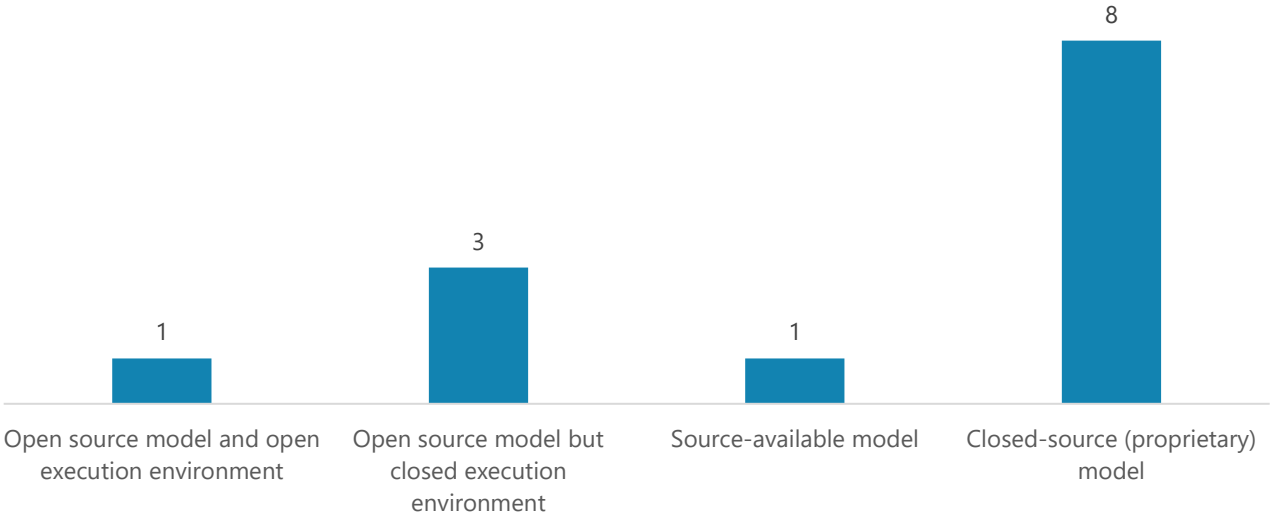


Figure 3. Categorisation of models featured in IPCC AR6 based on openness

At the same time, the OpenMod Initiative[7] (short for Open Energy Modelling Initiative) has played a pivotal role in promoting openness and transparency within the energy modelling community, currently featuring just under one hundred models aligned with its guiding principles. Established as a grassroots network of researchers, developers, and practitioners, OpenMod has long advocated for open data, open-source models, the use of OSI-approved licenses, and open collaboration practices across the energy systems domain—ensuring that models can be shared, scrutinised, and improved collectively. This community-driven effort has substantially influenced the broader shift towards open science in integrated assessment and energy modelling, especially through modelling frameworks such as OSeMOSYS and the broader family of models (e.g., CLEWs) that have been pivotal in the activities of the community. However, since most models of the OpenMod initiative are mostly energy or even electricity system models, it is difficult to be able to directly contribute through scenarios in major scientific assessments, highlighting a persistent gap between the open community and the integrated assessment modelling domain.

3.2.3 Licenses selected for the six DIAMOND IAMs

By recognising the opportunities created through open licensing and addressing the existing openness gaps in IAMs while also respecting the unique requirements of each model, DIAMOND seeks to enhance openness practices within the field. The selection of licenses to accompany the models, as well as supporting tools developed, is at the forefront of creating IAMs that are both transparent and widely available for use. We acknowledge that not all limitations can be addressed within the timeframe of DIAMOND. However, the purpose of the project is to ensure significant improvements in terms of openness, and as explained in the introduction to harmonise with the motto "as open as possible, and as closed as necessary". In this section, we transparently discuss the selection of licenses, improvements made from the current status quo in terms of openness, as well as limitations persisting. Table 4 summarises the current selection of licenses and the improvements achieved in terms of openness.

[7] https://openmod-initiative.org/

Table 4. Selection of licenses for the six DIAMOND IAMs

Updated Model in DIAMOND	Existing Model	Status Quo of existing model	License Selection of the updated model	Openness Classification	Supporting Tools	Limitations
CLEWs-EU	CLEWs	Fully open source (MIT)	MIT license	Open source model, and open execution environment	otoole (not developed as part of DIAMOND): MIT license	N/A
GCAM-Europe	GCAM	Fully open source (ECL 2.0)	Educational Community License, Version 2.0 (ECL-2.0)	Open source model, and open execution environment	gcamreport: MIT license	Proprietary data are used in chunks, but the model can run without independently purchasing said data.
GEMINI-E3-EU	GEMINI-E3	Closed Source (proprietary)	GNU Lesser General Public License (LGPL) version 2.1	Open source model, closed execution environment	N/A	Use of GAMS is required for running the model. Reliance on the GTAP database.
NEMESIS-World	NEMESIS	Closed Source (proprietary)	Apache 2.0 with the "Commons Clause" restriction, and Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International for the core NEMESIS-World Database	Source-available model	IODE (not developed as part of DIAMOND): GPL-3.0 pymrio (not developed as part of DIAMOND): LGPL-3.0	The "Commons Clause" restriction renders the model not fully open, but availability is ensured and academic use is unrestricted (i.e., source-available).
OMNIA	TIMES family of models (e.g., TIAM etc)	Only source code of the TIMES model generator is open (GLP v3)	Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International	Open-data model generator, closed execution environment	N/A	Use of GAMS and VEDA for running the model.
OPEN-PROM	PROMETHEUS	Closed Source (proprietary)	GNU Affero General Public License v3.0 (AGPL-3.0) with "Commons Clause" restriction	Source-available model	mrprom: GPL-3.0 postprom: GPL-3.0	The "Commons Clause" restriction renders the model not fully open, but availability is ensured and academic use is unrestricted (i.e., source-available). Use of GAMS is required for running the model.

Funded by the European Union

The first step was to ensure that models with existing vibrant communities reinforce existing good practices and ensure that choices made are harmonised with implemented standards, provided these follow open science principles. This was indeed the case for both CLEWs-EU and GCAM-Europe, two models with strong existing communities and among the most widely regarded open models existing in the field of IAMs and energy-climate-economy modelling. In the case of CLEWs-EU, models within the OSeMOSYS community typically follow the Apache 2.0 (e.g., OSeMOSYS, GENeSYS-MOD) or the MIT license (CLEWs). For CLEWs-EU the MIT license was selected to harmonise with the global version of the model. In the case of GCAM-Europe, all core and regional (e.g., GCAM-USA) versions follow the Educational Community License version 2.0 (ECL 2.0), which was also the license of choice for our version.

Similarly, OMNIA is also part of the active TIMES community. TIMES is considered a model generator, with the source code being available under the GNU General Public License v3 (GPL v3). As such, variants of the TIMES family of models are by nature more data-driven and can scarcely be viewed as software like other traditional models. In this case, OSI-approved licenses are less relevant, with data-focused licenses such as Creative Commons being more appropriate. This approach was followed by TIMES-GEO[8], the predecessor of OMNIA, which is also based on TIMES, and also on the GLUCOSE model[9], based on OSeMOSYS. As such, the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International Public License was selected for OMNIA.

The three other models, namely GEMINI-E3-EU, NEMESIS-World and OPEN-PROM, are based on models that were initially closed source. Usually being single institution-based models, no communities of practice exist that would set specific guidelines on the selection of license. As such, choices rested in the requirements and needs of each modelling team.

GEMINI-E3-EU is a general equilibrium model that heavily relies on the GTAP database, which is a proprietary data source. In this case the GNU Lesser General Public License (LGPL) version 2.1 was selected as more appropriate, as it ensures openness, modifications of the model core code (the part under LGPL) and distributions of it must be under the same LGPL but ensuring that the model can be linked or integrated with proprietary/closed components.

In the case of NEMESIS-World and OPEN-PROM, hybrid licensing approaches were adopted: the former combines the Apache License, Version 2.0 with the "Commons Clause", while the latter combines the GNU Affero General Public License v3.0 (AGPL-3.0) with the same "Commons Clause" restriction. Both models were originally developed by Small and medium-sized enterprises (SMEs), creating a legitimate need to safeguard intellectual property (IP) rights and commercial interests. Although these choices render NEMESIS-World and OPEN-PROM source-available rather than fully open-source, they enable the transparent sharing of source code while maintaining appropriate safeguards on its commercial use. This approach supports academic collaboration and public scrutiny of each model's structure and assumptions without limiting non-commercial research use. At the same time, it prevents the unauthorised commercial repackaging of the models' functionality, ensuring fair-use conditions consistent with nuanced interpretations of openness that emphasise the complementarity of Open Science and IP protection (see, for example, European IP Helpdesk, 2024).

The license used for each model may be updated based on requirements arising during the final stages of development. These will be reported in the final documentation of each model separately.

3.3 Harmonisation of model development

Different IAMs incorporate diverse economic theories and structural frameworks, which can substantially influence their produced outcomes. To ensure that uncertainties arising from such model heterogeneity are appropriately

[8] https://github.com/MaREI-EPMG/TIMES-GEO
[9] https://github.com/KTH-dESA/GLUCOSE

captured when IAMs are used to inform policy evaluation or design, multi-model intercomparison exercises are commonly employed (Sognnaes et al., 2021). These exercises provide insights into the potential ranges of key output variables. However, to ensure these ranges primarily reflect divergences in model theories and behaviours rather than inconsistencies in inputs, it is necessary to harmonise common input variables, namely, by adopting similar data sources and assumptions according to dedicated protocols (Giarola et al., 2021; Gidden et al., 2018).

Such harmonisation efforts have traditionally focused on the use of models, given that the IAMs involved are typically pre-existing and ready for application. Within the DIAMOND project, however, the objective to update and upgrade six core IAMs creates opportunities not only to continue the established practice of harmonising model use but also to initiate harmonisation activities earlier in the model development process. The development trajectories of these six IAMs within DIAMOND are not fully aligned. In several cases, the objective is to enhance spatial granularity by disaggregating European regions. Nevertheless, the extent of this disaggregation varies across models, while in the case of NEMESIS-World, development proceeds in the opposite direction, i.e., transitioning from an EU-level to a global model.

Consequently, D3.2, which outlines the model development strategy, established a set of soft harmonisation recommendations—instead of a rigid protocol. These guidelines are intended to support modelling teams during development, ensuring an adequate degree of harmonisation (i.e., standard default data sets), while accommodating the specific needs and design pathways of each model, as well as ensuring that models are developed to be as flexible as possible with respect to choice of input data and assumptions (so that they can be used in various contexts and for various reasons). Here, we summarise key decisions in this context, framed within the open science protocols that govern all project-related activities (for more details, see D3.2).

Historical Data (within the framing of the models):

- Socioeconomic variables based on the updated Shared Socioeconomic Pathways (SSPs).
- For EU regions, EUROSTAT Energy Balances.
- Emissions based on EDGAR (Crippa et al, 2024) or CEDS (Hoesly et al, 2018), depending on model needs.

Future Projections:

- Short-term GDP growth projections based on International Monetary Fund (IMF) projections (IMF, 2024)
- Long-term socioeconomic projections and qualitative assumptions based on SSPs, with a recommendation to at least calibrate the use of SSP2.
- Technology costs and fossil fuel price projections based on the International Energy Agency (IEA) Global Energy and Climate Model database (IEA, 2024), depending on technology representation in each model.
- Suggested adaptations to the abovementioned data for EU regions:
    - EU ageing report projections (European Commission, 2023) for socioeconomics
    - EU Reference Scenario (European Commission, 2021) for emissions and technology assumptions

The choice between EU-specific and global data is based on the scope of model use (see Section 4.1).

In most cases, the recommended data sources are open access. However, depending on each model's need for detailed data, the IMF and IEA projections include components that are licensed under Creative Commons or other open terms, while some parts require special agreements. Similarly, for the EU Reference Scenario, the summary tables are openly available, but access to the full model datasets may require a request or institutional permission. Finally, the updated SSPs are released under a custom license based on the Creative Commons Attribution 4.0 International Public License, with additional restrictions on the redistribution of "substantial portions".

4 Transparent Model Use

Moving beyond model development, the framework established in Section 2 outlines four steps towards open model use. The first step focuses on activities taking place before modelling, most importantly the harmonisation of key parameters according to the needs of each scenario exercise, and the description of the ontologies used to ensure consistent variable definitions. A central element of this step is the design of a common, project-wide, policy-based baseline scenario to guarantee comparability across different exercises.

The next three steps described in the framework focus on activities taking place after modelling—namely validation, vetting, and diagnostics—to ensure the robustness and appropriateness of the developed scenarios delivered by each model (i.e., scenario/model combinations). The validation step in our framework includes checking each scenario's results for common errors between aggregated and disaggregated variables (e.g., basic sum checks), as well as for missing values or inconsistencies between N/As and zero values. Similarly, the vetting step in the framework closely follows the protocol established in IPCC AR6 (Guivarch et al., 2023; Kikstra et al., 2022) and suggest its application to assess the alignment of key variables with historical data. Given that the DIAMOND models substantially expand the representation of EU regions, the novelty of this deliverable lies in adapting the vetting process to the regional, in addition to the global, level. Finally, during the diagnostics step of the framework, specific indicators (Harmsen et al., 2021) are used to explore how model behaviour may influence or aid the interpretation of results.

This sequential process ensures that the DIAMOND scenarios can be rigorously validated, meet the highest scientific standards, and are accompanied by information that strengthens the transparency of outcomes.

The following sections elaborate on each step, describing the specific methodologies applied, the supporting tools developed to assist modelling teams in implementing the framework, and examples from existing modelling runs that demonstrate its functionalities. Given that DIAMOND is still in the early stages of delivering extensive thematic scenarios—expected primarily through the activities of WP5—the results presented here should be interpreted as numerical examples illustrating the pipeline, rather than as fully implemented exercises.

4.1 Harmonisation of Model Use & Ontologies

Building on top of the soft recommendations on harmonising model development, three main cases can be distinguished when shifting to model use. The detailed approach to be followed should be specified based on the needs of the multi-model intercomparison exercise (within the context of DIAMOND, this refers to the activities of each task under WP5).

For models participating in global studies, harmonisation involves calibrating at least the SSP2 scenario—ideally all SSPs—across as many input parameters as possible (at minimum, population and income), with GDP growth rates up to 2029 aligned with IMF projections. Technology costs and fossil fuel prices are recommended to follow the latest IEA assumptions, that are assumed to be resembling the SSP2 principles, while regional projections beyond the USA, EU, China, and India can be mapped to these key regions or combinations thereof.

For models used in EU-specific studies, harmonisation entails calibrating to the 2024 Ageing Report projections for population and GDP, again overwriting GDP growth rates with IMF projections up to 2029. Technology costs and other relevant inputs should align with the Reference Scenario projections, using either current or updated values depending on data availability in 2025.

For models engaged in both global and EU-specific studies, three approaches are possible: (i) maintaining separate versions of the model calibrated to SSPs and EU-specific data and toggling between them depending on the study focus; (ii) applying EU-specific projections for EU countries and SSP2 for the rest of the world, given their close

alignment; or (iii) relying exclusively on SSP-based socioeconomic assumptions and IEA techno-economic projections, even for EU countries, when cross-narrative comparability is a priority.

On top of input parameter calibration during harmonisation implies setting up a common baseline scenario to ensure comparability across the different thematic scenarios developed as part of WP5. To ensure realism (Hausfather and Peters, 2020), a protocol is developed to implement a policy baseline scenario instead of the existing counterfactual no-policy references of each model. The details of the protocol will be described in D3.3, and D5.2, based on choices made during implementation. However, the core aspects of this scenario will include following assumptions based on the currently updated NDCs for non-EU countries, and emissions targets/caps for EU countries based on the updated National Energy and Climate Plans (NECPs), aggregated to the level of representation in each model. More explicit policies may be used for EU regions (i.e., renewables, energy use, coal phase-out, etc.), when an equivalency with emissions targets can be assumed (Frilingou et al., 2025).

To further improve transparency, consistency, and collaboration across the harmonisation and scenario development work between different models and teams, a structured ontology has been developed to document all underlying modelling dimensions such as variables, concepts, and assumptions. The ontology creates a shared vocabulary that helps all teams consistently harmonise models and data following commonly accepted definitions—such as those established in the 2006 IPCC guidelines for emission inventories (Eggleston et al., 2006)—enabling meaningful comparison of results and ensuring that modelling choices can be clearly traced.

The ontology is built on well-established community standards, and notably the IAMC variable template, along with the established common definitions[10] and the respective nomenclature[11] (Huppmann et al., 2021) from the IAMC community. On top of the definitions and naming conventions of common variables, the ontology can be potentially populated with additional layers focusing on the harmonisation framework, assumptions for scenario development and the diagnostic and vetting processes, so that assumptions, calibration data, and policy representations are consistently presented, along with the indicators used for cross-model evaluation. Additionally, the ontology is expected to assist modelling teams automatically run the open-science pipeline (i.e., diagnostics-vetting-diagnostics) presented in this deliverable.

The ontology is implemented in machine-readable formats (i.e., OWL), using tools such as the python libraries rdflib[12] and owlready2[13], as well as the open-source ontology editor Protégé[14]. The development process of setting and expanding the ontology follows two steps: initially, a table of variables and parameters (in Excel or CSV) was created based on the IAMC template and expanded based on feedback from the modelling teams (e.g., for variables missing in the initial list); the table is then converted into a YAML file for harmonised naming (Figure 4), and finally into an OWL ontology (Figure 5) that allows more detailed relationships and reasoning. The YAML file is intentionally structured so that it can be read by software tools, ensuring consistent variable names, hierarchies, and units across models. The OWL ontology complements this by providing a richer semantic layer that captures relationships and constraints which YAML alone cannot express. This combined approach helps ensure that these YAML definitions can be used as inputs to the nomenclature package and allow automated data pipelines for calibration, validation, vetting, and diagnostics to rely on a stable and machine-interpretable vocabulary.

The resulting datasets, scripts, and documentation will be shared through dedicated public repositories in the project's GitHub and Zenodo pages to support transparency and reproducibility following FAIR and other open-

[10] https://github.com/IAMconsortium/common-definitions/tree/main/definitions/variable
[11] https://github.com/IAMconsortium/nomenclature/tree/main/nomenclature
[12] https://rdflib.readthedocs.io/en/stable/
[13] https://owlready2.readthedocs.io/en/v0.48/
[14] https://protege.stanford.edu/

science principles, as well as to allow other researchers and modelling teams to reuse and expand them. A first version already developed in collaboration with other projects (such as TRANSIENCE[15]) to ensure consistency between research projects focuses on describing the variables, parameters, and metadata that support scenario creation and model calibration. Later versions will include the diagnostic indicators and model-specific settings used for the vetting and intercomparison assessments.

```
- Expenditure|Household:
    description: Total household expenditure
    unit: billion US$2010/yr
- Export:
    description: total exports measured in monetary quantities.
    unit: billion US$2010/yr
- Import:
    description: total imports measured in monetary quantities.
    unit: billion US$2010/yr
- Capital Stock:
    description: Macroeconomic capital stock
    unit: billion US$2010/yr
- Interest Rate|Real:
    description: Real interest rate or return on capital that is relevant for energy
      system investments
    unit: '%'
- Emissions|CO2:
    description: total CO2 emissions
    unit: Mt CO2/yr
- Emissions|CO2 (w/o bunkers):
    description: 'total CO2 emissions excluding emissions from international bunkers
      (IPCC category 1A3ai and 1A3di) '
    unit: Mt CO2/yr
- Gross Emissions|CO2:
    description: Gross CO2 emissions, not accounting for negative emissions e.g.
      from BECCS or AFOLU
    unit: Mt CO2/yr
```

Figure 4. A part of the YAML file tree describing DIAMOND variable nomenclature

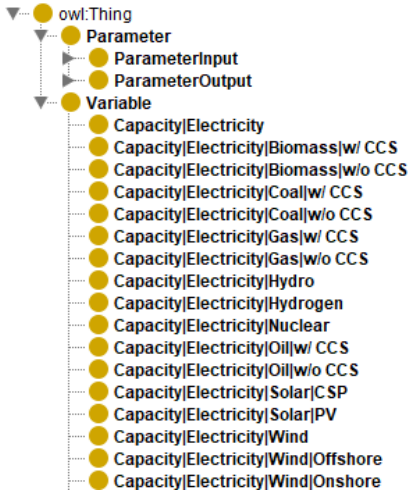


Figure 5. A part of the OWL file tree describing DIAMOND variable nomenclature visualised using the Protégé software.

4.2 Automated Validation Routines

4.2.1 Methodological approach

The validation process in the context of DIAMOND's open science protocol requires the provision of a dataset from a participating model that is formatted based on the IAMC template for yearly data[16]. The template is well established in the integrated modelling community and has been also used in the assessment reports of the IPCC. Based on the template, a dataset of modelling results should include five columns named Model, Scenario, Region,

[15] https://www.transience.eu/

[16] https://pyam-iamc.readthedocs.io/en/stable/#yearly-data

Variable, and Unit (i.e., the index columns) plus any number of additional columns (i.e., content columns) whose headers are years (e.g., 2010, 2015, 2020, etc.). The data below these content columns should be timeseries results for the model runs that are described in the index columns and form the main body of the dataset. Based on these assumptions, the following checks are included in the validation process and implemented in the software described in Section 4.2.2:

- **IAMC template validity**: This part of the process assesses whether a dataset has five columns that are named Model, Scenario, Region, Variable, Unit and whether all other columns use an integer as a column name (indicating years). Additionally, this part checks whether the timeseries data in the dataset are given in a float or integer format underneath the year columns. An empty cell under a content column will be interpreted as unavailable data, a 0-character will be interpreted as a value of zero, while alphabetic characters are not allowed.
- **Model, Scenario, Region, Variable, and Unit consistency checks**: The next part of the process assesses whether the data below the index columns and the index columns conform to some expected values; for instance, whether the scenario names in the dataset match the scenarios that were agreed between modelling teams in a model intercomparison exercise or within a specific project. The purpose of this check is to establish consistency between modelling results, e.g., avoiding using different names for the same variable. This ensures that results from different models and scenario runs can be easily compared to each other, facilitating their usefulness in modelling studies and increasing the comprehension of modelling results by end users (e.g., policymakers, other researchers).
- **Duplicate rows**: The dataset is checked whether there are more than one rows with the same index columns, i.e., having the same model, scenario, region, variable, and unit. In that case, the rows are flagged and one of them has to be deleted in order for the dataset to pass the validation.
- **Consistency between disaggregated and aggregated variables**: Modelling results are often provided in a disaggregated form across variables. For instance, results for energy capacity additions are provided for each energy source (e.g., solar, gas, nuclear) and results for energy demand are usually provided for each of the main demand sectors (e.g., buildings, transportation, industry). In case that both aggregated and disaggregated results are provided, this check ensures that the values of aggregated variables correspond to the sum of their respective disaggregated variables under an allowed error margin. For DIAMOND, a 2% error margin is used.

The consistency checks can be also supported by the functionalities provided by the pyam[17] and nomenclature[18] Python packages that have been developed for the IAMC community. Pyam includes tools and functions for analysing and visualising model inputs and results while nomenclature provides functions for validating and processing results. Specifically, nomenclature is based on definitions of data structures that include lists with allowed values for the index columns of the IAMC template, most typically for the allowed names of variables, region, and units. The IAMC has also collected a list of the most commonly used variables and regions[19] (defined as "Tier 1") and released them as suggested common definitions for model intercomparison projects. Nevertheless, each project is able to expand on or edit these definitions and create a project-specific nomenclature that can be used for validation (so-called "Tier 2" or "Tier 3" variables). In the case of DIAMOND, a specific nomenclature is developed (see Section 4.2.2).

[17] https://pyam-iamc.readthedocs.io
[18] https://nomenclature-iamc.readthedocs.io
[19] https://github.com/IAMconsortium/common-definitions/

Ensuring consistency across model results is particularly difficult for regional data, since models often use different regional aggregations. For instance, a model can output results specifically for the EU (i.e., EU27 countries), while another may output results for Europe as a whole, including both EU and non-EU countries. The IAMC nomenclature package offers a functionality to address this issue by aggregating model-specific regions to common regions with recognised region names (as included in the common definitions package). Based on this approach, the validation process of DIAMOND will first check the names of the regions included in each results dataset for consistency with the common definitions, as suggested above. If the region refers to a single country or to a multi-country region that exists in the common definitions, then it will stay as is (or slightly renamed to match the common definitions). If the region does not exist there, then it is assumed to be model-specific and it is renamed by adding the model's name as a prefix. Additionally, the results for this region are aggregated in the best way available to a larger region that is part of the common regions. Similarly, we will aim to adopt more advanced functionalities of the nomenclature package on aggregation checking such as calculating the mean value of specific sub-variables or using a weighting pattern.

4.2.2 Software Tool

The process described in the previous section has been operationalised in an online tool that was co-developed by the DIAMOND and IAM COMPACT projects. The first version of the tool was developed in 2023 and had the form of a web application based on the Streamlit library in Python[20]. The tool's user interface (UI) was structured as a series of pages to guide the user through the validation process while providing instructions in a sidebar (Figure 6). The user first uploaded a dataset that was then read and checked for consistency with the IAMC format. On failure, the user was prompt to fix any potential mistakes and upload the updated dataset, while, on success, the application moved to the second page where the user could choose the validation checks that they wanted to perform from the list provided in Section 4.2.1 along with the global vetting checks that are described in Section 4.3. The tool was deployed online in the i2am-paris.eu domain[21] and was publicly available for anyone to try, while it has been especially promoted to modellers in the DIAMOND and IAM COMPACT teams. The tool's code is provided as open source on GitHub[22] under the GPL-3.0 license.

[20] https://streamlit.io
[21] https://validation.i2am-paris.eu
[22] https://github.com/i2amparis/validation

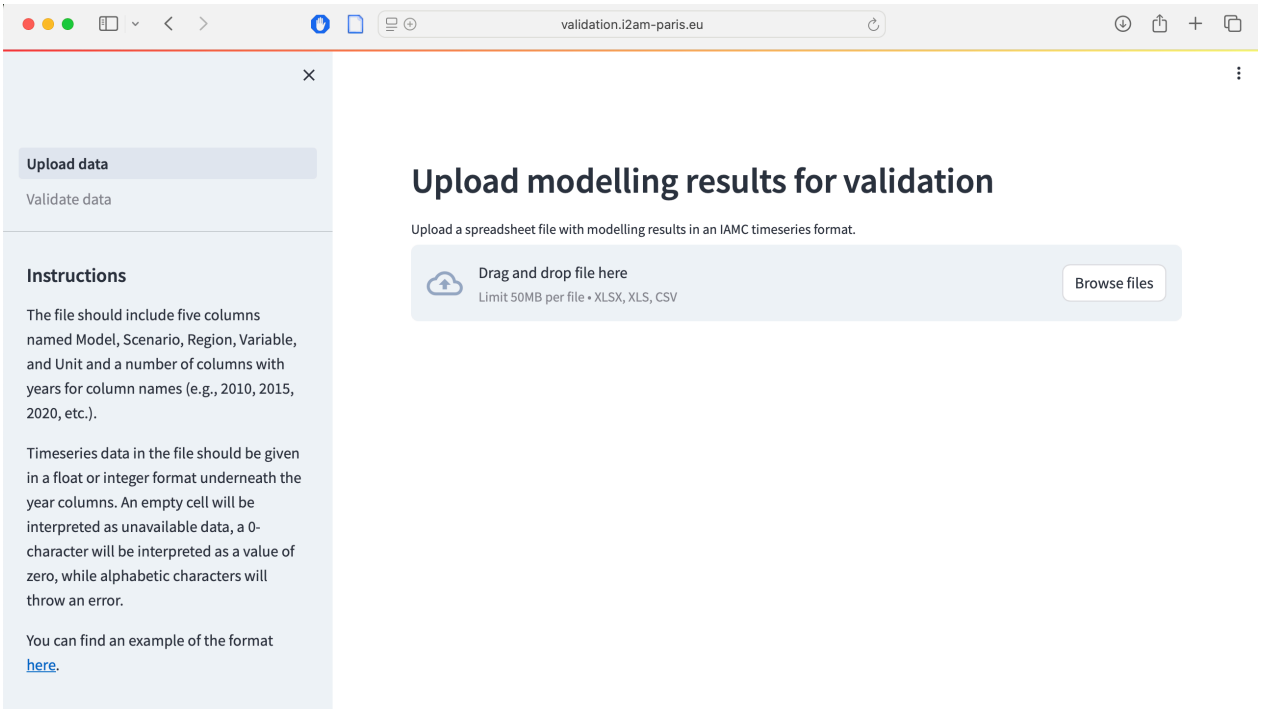


Figure 6. Screenshot from the first version of the validation tool developed in 2023

A second version of the tool was developed by CICERO Center for International Climate Research for the IAM COMPACT project in 2025. This version mainly incorporated the functionalities of pyam, nomenclature, and common definition packages from the IAMC to streamline the validation checks and ensured that the tool was in line with the community efforts to standardise validation processes. In addition, the tool's repository was split into two, to allow for a distinction between the input data for the validation[23] and the UI itself[24]. The input data followed the format suggested by the nomenclature package, using the common definitions as a basis and adding more variables and region names specific to IAM COMPACT models. The UI followed the same structure and appearance as the one from the first version (Figure 7), while creating separate pages to show the results of each check to the user. In addition, this version incorporated the region mapping functionality provided by the nomenclature. The tool is publicly available on the iam-compact.eu domain[25] and has been used by IAM COMPACT modellers to validate their results.

[23] https://github.com/ciceroOslo/iamcompact-nomenclature
[24] https://github.com/ciceroOslo/iamcompact-validation-ui
[25] https://validation.iam-compact.eu/

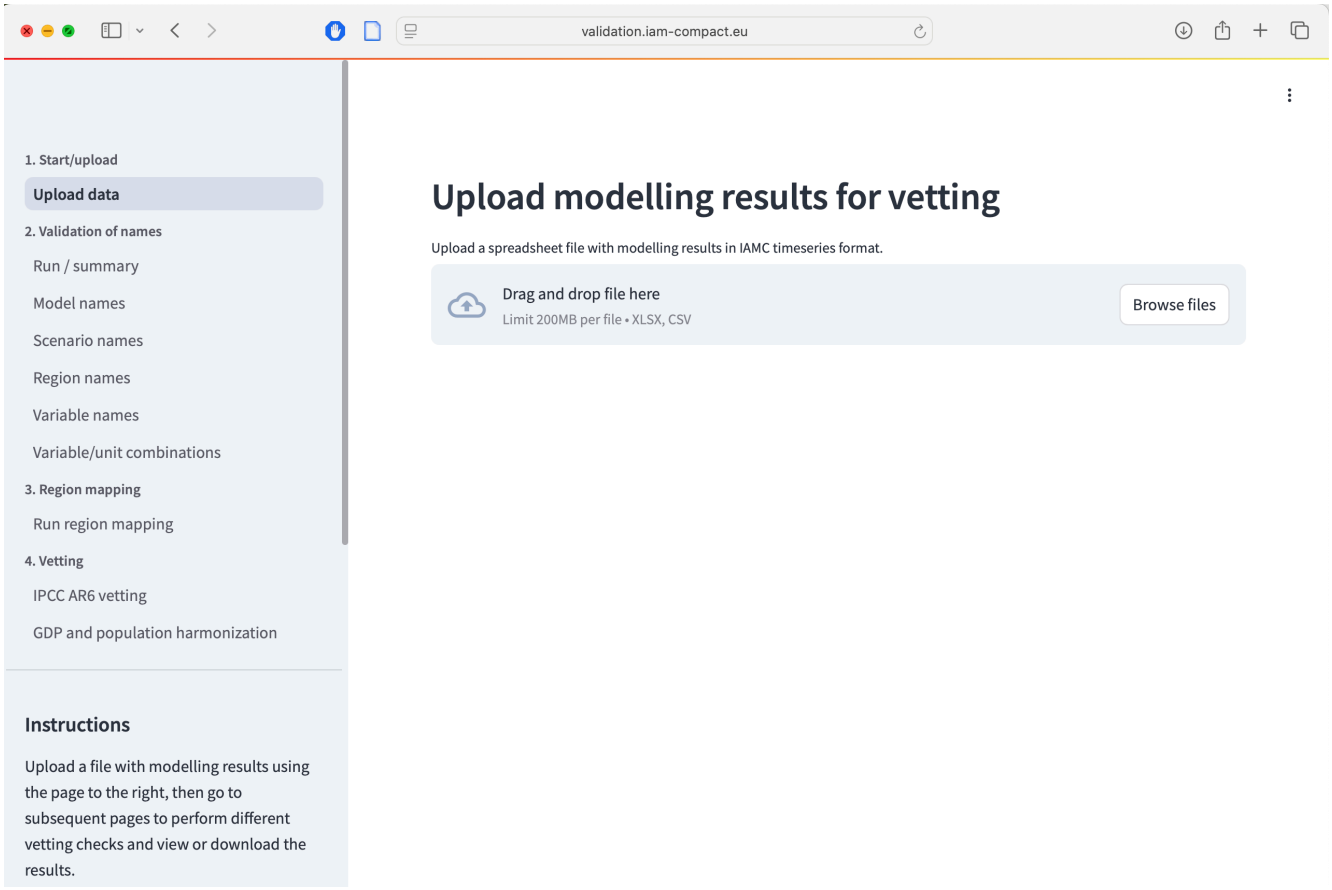


Figure 7. Screenshot from the second version of the validation tool developed by CICERO for the IAM COMPACT project

A third version of the validation tool is currently under development in the context of DIAMOND[26]. This version incorporates the improvements introduced by IAM COMPACT and slightly modifies the UI in order to re-establish the tool as project-agnostic and thus available to use for the whole community. This interaction between the two projects is in the spirit of established open-source software projects and ensures that the development done in IAM COMPACT will not be lost now that the project has ended but will continue through DIAMOND, avoiding duplicated work and extra costs. Similarly, critical dependencies such as the nomenclature library will be automatically updated to the newest version available to ensure that it stays in line with the latest updates of the IAMC developing team. In addition, a dedicated nomenclature repository is created for DIAMOND[27], listing the names of each specific model that participate in the project and, eventually, the scenarios, variables, and regions that will be accepted in the model results. This will help project modellers to easily validate their modelling results, which is especially important for DIAMOND as all the participating models are new and modelling development is still ongoing. Finally, the third version will incorporate the regional vetting methodology that is developed in the context of this deliverable (Section 4.3).

4.2.3 Indicative Results

Three indicative screenshots from running the third version of the validation tool are shown in Figures 8-10. In the process presented in the first figure (Figure 8), a test dataset is uploaded to the interface and is then evaluated to ensure consistency with the IAMC template. The user then proceeds to the next section (Figure 9) where the validation of names is performed for all index columns, and a summary of the validation outcomes is shown.

[26] https://github.com/i2amparis/new-validation-ui
[27] https://github.com/CLIMATE-DIAMOND/diamond-nomenclature

Finally, Figure 10 presents the outcomes for the validation of variable names as an indicative example of the detailed validation pages, listing all variables that have not passed validation as well as all variable names that are allowed by the project's nomenclature.

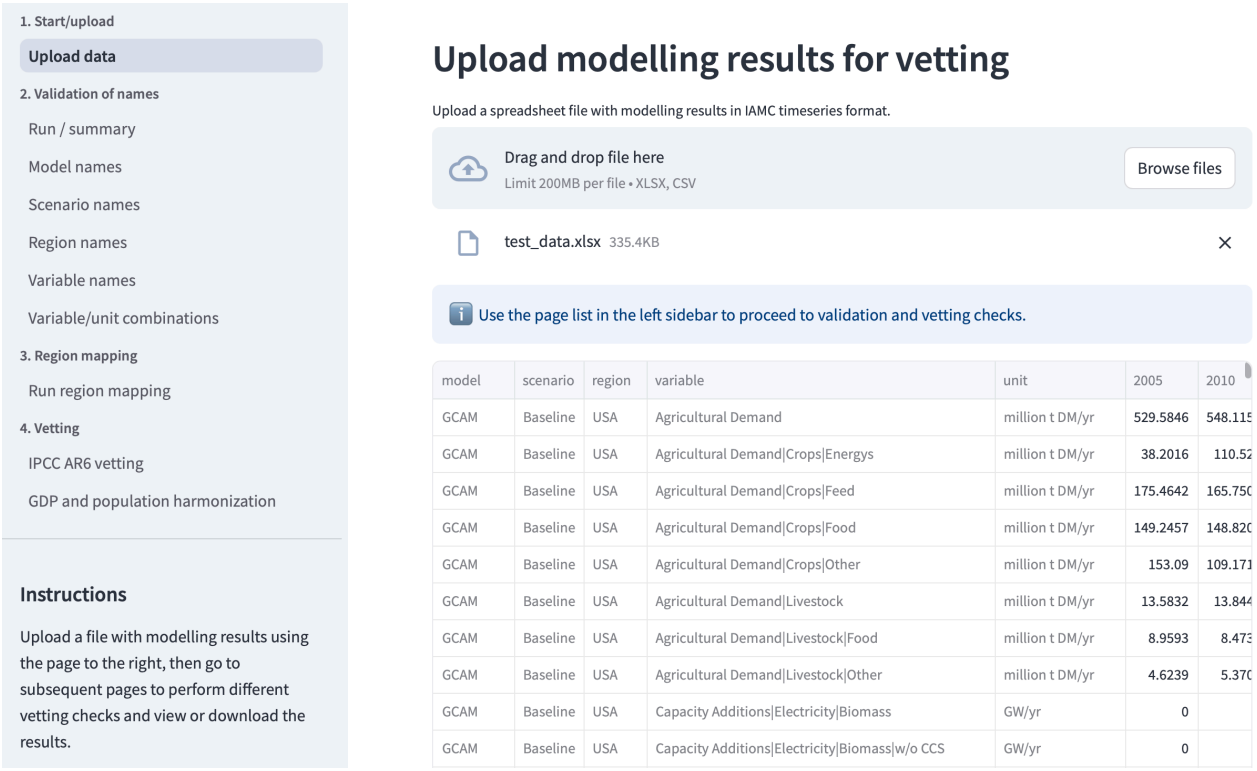


Figure 8. Screenshot of the validation tool with a test dataset uploaded and checked for IAMC template validity

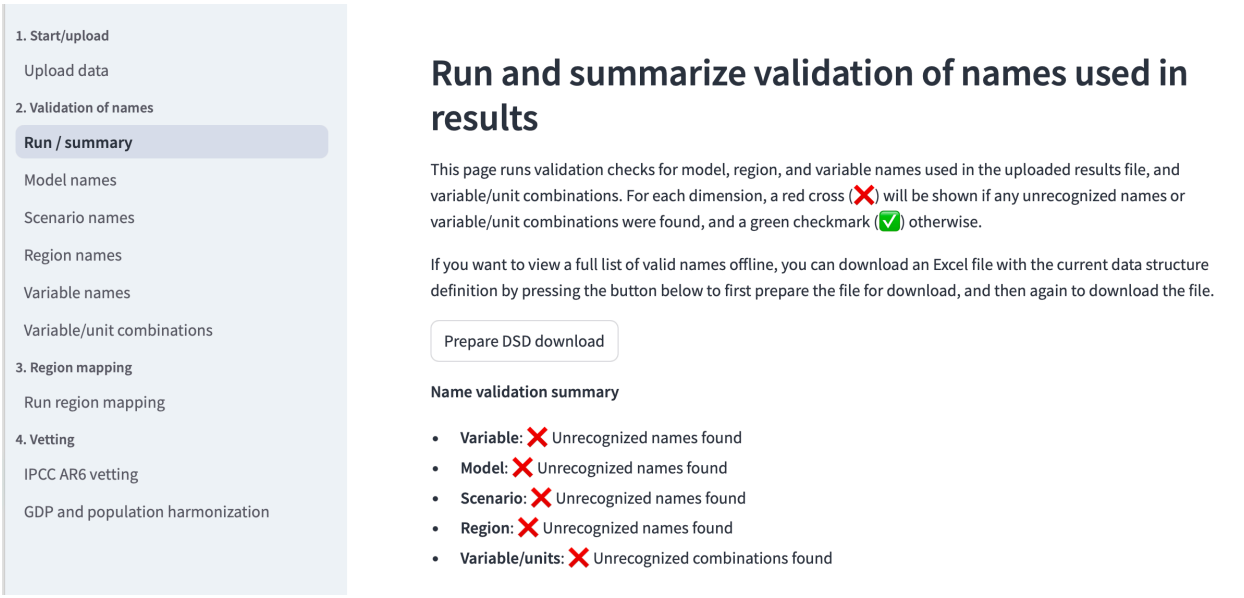


Figure 9. Screenshot of the validation tool with the available checks for name consistency

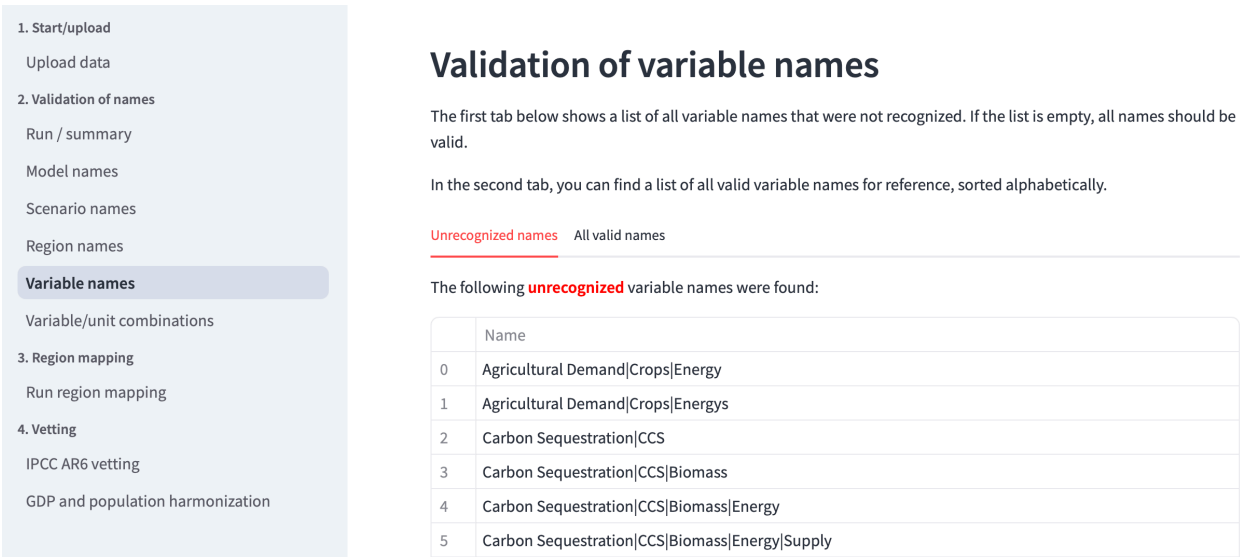


Figure 10. Screenshot of an indicative check of variable names in the validation tool

4.3 Global and Regional Vetting Process

4.3.1 Theory and methodological approach

The curation of the AR6 scenario database for the IPCC WGIII report featured multiple steps towards delivering climate assessments for each included scenario (Kikstra et al., 2022). Within the workflow followed, a core step was the vetting process, which aimed to verify the quality and completeness of data, by ensuring that modelling results for core global emissions and energy variables fall within reasonable ranges of true historical values. The vetting stage was a novel step introduced in AR6 compared to past climate assessments in IPCC reports, and it was a useful process to ensure historical alignment and realism, while at the same time providing flexibility with respect uncertainty and ensuring that scenarios are internally consistent.

The detailed list of variables used, including acceptability ranges and reference values can be seen in Table 5. More details on the sources used can be found in Annex III of IPCC AR6 WGIII (Guivarch et al., 2023), including adaptations performed to ensure that COVID-19 distortions are taken into consideration. The basis on the vetting

process focused on a comparison between true historical and model values in 2020. However, for a number of scenarios future values were also checked (2030 for main CO_2 and energy variables and 2040 for CH_4), not as exclusion criteria, but rather to raise potential flags.

Table 5. Summary of vetting variables, reference values and criteria in the IPCC AR6.

Variable (in 2020)	Reference Value	Acceptability Range
Total CO_2 emissions	44.25 GtCO2/yr	±40%
CO_2 emissions from energy and industrial processes	37.65 GtCO2/yr	±20%
Percentage Change of CO_2 emissions from energy and industrial processes between 2010 and 2020	N/A	0 to +50%
Total CH_4 emissions	379 MtCH4/yr	±20%
CCS from energy	N/A	0-250 MtCO2/yr
Primary Energy	578 EJ	±20%
Electricity production from nuclear	9.77 EJ	±30%
Electricity production from solar and wind	8.51 EJ	±50%

The binary approach adopted (i.e., scenarios either passed or failed vetting) raised discussions on potential biases that this process may introduce (Peters et al., 2023). These debates focused on the potentially reduced representativeness of the database as a result of over-emphasising scenarios from a subset of models, notably excluding sectoral models that may not include all requested variables despite offering nuanced insights on their respective areas of application (Koasidis et al., 2023), as well as potentially removing thematic extreme scenarios that may still be useful to analyse outliers.

At the same time, merely checking historical values may not be enough for models with a dedicated focus on specific regions, such as many DIAMOND models that offer a nuanced spatial representation for the EU, on top of the global dimension. In such cases, developing a novel process for regional vetting becomes a necessity. The goal of such a process is not to exclude scenarios, but rather to add another generalisable layer of validation and transparency to existing efforts from modelling teams, and crucially—similar to the IPCC—potentially raise flags in case large deviations are observed. Here, we present a process for regional vetting focusing on the EU, by establishing similar acceptability ranges for common energy and emission variables as with IPCC, but this time at the EU27 level. These ranges are based on the Eurostat databases, expanding the calculations at the member state level, to discuss potential deviations. The detailed protocol is presented in Table 6.

Funded by
the European Union

Table 6. Summary of vetting variables, reference values, sources and criteria for the regional EU27 vetting.

Variable (EU27 in 2020)	Reference Value (2019-2020 average)	Eurostat Database	Sectoral codes included	Acceptability Range
Total CO_2 emissions	2.877 $GtCO_2$/yr	env_air_gge	[CO2] TOTXMEMO, CRF1D1	±40%
CO_2 emissions from energy and industrial processes	3.074 $GtCO_2$/yr	env_air_gge	[CO2] CRF1a, CRF1b, CRF1D1, CRF2A, CRF2B, CRF2C, CRF2E	±20%
Percentage Change of CO_2 emissions from energy and industrial processes between 2010 and 2020	*(-16.1%)*	env_air_gge	[CO2] CRF1a, CRF1b, CRF1D1, CRF2A, CRF2B, CRF2C, CRF2E	-30 to 0 %
Total CH_4 emissions	15.12 $MtCH_4$/yr	env_air_gge	[CH4] CRF1-6	±20%
CCS from energy	N/A	N/A	N/A	0-20 $MtCO_2$/yr
Primary Energy	55.88 EJ	sdg_07_10	N/A	±20%
Electricity production from nuclear	2.94 EJ	nrg_ind_peh	[GEP] N9000	±30%
Electricity production from solar and wind	1.86 EJ	nrg_ind_peh	[GEP] RA300, RA400	±50%

Efforts have been made to ensure that the classifications between the IPCC/IAMC common definitions (e.g., IPCC 2006 guidelines) and the Eurostat codes are consistent. However, as with the IPCC, any potential deviation may be indirectly captured through the assigned acceptability ranges (e.g., in the case of total CO_2 emissions this is reflected in the higher allowed range).

After the publication of AR6 and as we move towards AR7, several advancements have already been made by the community on vetting. However, given that the submission of scenarios in the AR7 may be follow different processes than the AR6, different vetting criteria has appeared in the literature. For example, the piamValidaton uses the criteria set by Weigmann et al. 2025, while the Scenario Compass Initiative (SCI) is undergoing an external review process to set a protocol for inclusion the scenario database built[28]. Deviations between vetting approaches may include the variables included, benchmark values, as well as the colour coding (e.g., the SCI follows a colour coding on top of the exclusion criteria, focusing on sustainability and feasibility concerns). At the same time, most protocols would still need to undergo significant revisions to expand the criteria to more recent years (e.g., 2025). With these caveats in mind, we opted to retain the protocol from AR6—including variables and ranges—for compatibility with the latest authoritative dataset, except where regional dynamics warranted revision. The following deviations from the IPCC global vetting process were needed:

- For the reference values, we interpolated between the 2019 and 2021 values from Eurostat to account for distortions due to COVID-19.
- In the percentage change of CO_2 emissions from energy and industrial processes (EIP) between 2010 and 2020 (in real values this is adapted for COVID-19), the global trend of historical values saw an increase by roughly 11% according to IEA (IEA, 2025). This prompted the IPCC to adopt a flexible positive range, as presented in Table 5. However, in the EU a reverse trend is observed, with emissions reducing from 2010

[28] https://scenariocompass.org/news/vetting-criteria-review

by roughly 16%. As such both a different sign is followed in our protocol, as well as a stricter range, given that the base year in our models is more recent than the one in the models included in the IPCC.

- The check of the CCS variable in the IPCC aims to ensure that models do not deploy technologies that are not yet operational at a large scale. Accordingly, a small indicative range was adopted for global CO_2 sequestration in energy from CCS (250 $MtCO_2$). We then calibrate this range to reflect the EU27's share of global CO_2 EIP real historical values, based on adjusted 2020 emissions.

The protocol can be extended to non-EU countries, as well as other regions/countries of the world provided that data can be collected and are aligned with IPCC definitions. In case of misalignment, the acceptability ranges should be updated accordingly to reflect this uncertainty. At the same time, this process can indicatively be extended to future years, similar to the process followed in IPCC. In any case, once full 2025 data is available, it is advisable to update this protocol and the reference values, to perform vetting in 2025 instead of 2020.

4.3.2 Software Tool

The validation tool presented in Section 4.2.2 also provides an interface to perform the global vetting checks shown in the previous section. The vetting process can be run after uploading a dataset in the IAMC format, provided that the dataset includes results for the variables shown in Tables 5 and 6. This check uses the formalised variable names that are shown in the common definitions package and searches the dataset for their existence. If any of them do not exist, the tool informs the user that the vetting process can partially run for the variables that are found. The tool then checks whether the values of the selected variables are between the vetting limits and informs the user of the results.

4.3.3 Indicative Results

As part of the open science pipeline, scenarios developed in DIAMOND will undertake vetting tests based on the parameters defined in Section 4.3.1. Here, we present indicative examples from models that have already developed tentative scenarios with the beta versions of the models. These should neither be interpreted as finalised scenarios, nor the results of the vetting as an authoritative assessment of the validity of the models or scenarios, but rather serve as numerical examples of the process, and assist modelling teams during the final stages of model development.

Three models have been primarily used in this exercise, namely, GCAM-Europe, GEMINI-E3-EU, and OPEN-PROM, representing a range of different scenario implementations focusing on the EU27, which is the core novelty introduced here. When full scenarios are implemented, the vetting process will be implemented at the global level as well. In the case of GCAM-Europe, a tentative protocol based on the updated National Energy and Climate Plans for 2030 and long-term net-zero targets for 2050 has been implemented, with policies being modelled at the member state level. For more details see Frilingou et al. (2025). For GEMINI-E3-EU, the Fit-for-55 targets are modelled; indicatively see Perdana and Vielle (2025). Finally, for OPEN-PROM, a no policy reference scenario has been used. Considering that 2020 is a model year in most IAMs, i.e., models solve for 2020 instead of relying on real historical data, results within this year depend to some extent on the policies implemented. For example, high ambition policies may also lead to fast mitigation early in the solution period. At the same time, models may not represent certain variables exactly as these appear in Eurostat. Hence, perfect alignment is near impossible. However, the reference values provided, along with the acceptability intervals, offer a reasonable range that the models should aim to achieve irrespective of scenario implementation.

We observe that across all variables represented within each model, scenarios pass the vetting criteria for the EU27, with different error margins from the central value of the acceptability range (Figure 11 and Table 7). For example, in GCAM-Europe, both emission-related variables are within the lower end of the range, coupled with a

higher wind and solar penetration and lower primary energy consumption, which could potentially reflect the high level of ambition of today's NECPs compared to the policies back in 2020. We also note that although neither the scenarios nor the models are finalised, the results provide a significant validation level, ensuring the realism of the process, even in variables that have not been explicitly calibrated yet—e.g., in GCAM-Europe, the next steps of the calibration process include focusing on non-CO_2 and AFOLU emissions, yet the respective variables (e.g., CH_4 and total emissions) are already within the calculated range of Eurostat.

In Table 7 we observe that the only missing variable that could not be directly calculated by a model is "changes in EIP between 2010 and 2020" from GEMINI-E3-EU. This limitation can impact many other models within or outside DIAMOND, especially those mostly consisting of an energy-system core (e.g., OMNIA, CLEWs-EU), since past emissions are not represented or calculated by such models. We observe that calculating the emissions reduction trend from the 2010 historical value up to the 2020 model value highlights that the model can adequately capture the real trend. Assuming that the model fails vetting because of the lack of representation of 2010 emissions would then be an extremely harsh decision. Moving to AR7, which could potentially include an updated vetting process, it is important to offer contingencies for such cases, and transparently describe all the vetting criteria early on, so that modelling teams can remedy such limitations, that do not significantly impact the quality of the scenarios.

In Table 9 of the annex, present an indicative result from the vetting exercise at a detailed EU level for the GEMINI-E3-EU model for the CO2|EIP variable. The example also serves to showcase that the pipeline established can apply different regional masks to ensure that the regional disaggregation of each model is correctly compared to the appropriate EUROSTAT dataset. In principle, this exercise does not form a binary process that models/scenarios either pass or fail, but rather an additional level of validation, that offers the chance to raise flags, on aspects that models may improve in their representation. As such, similar tables to the one presented for GEMINI-E3-EU have been sent to modelling teams performing vetting to assist them during calibration. Based on this preliminary data collection—pending the implementation of finalised scenarios—we observed that the scenarios in the majority of cases are within range even at the member state level. Expectedly, larger deviations are observed in smaller regions/countries (i.e., Malta, Luxembourg, etc.), since small differences may disproportionally increase the calculated error, even though the absolute values are minorly affected. Other large errors may be traced to the limited representation of Land Use, Land Use Change, and Forestry (LULUCF) emissions in OPEN-PROM and GEMINI-E3-EU, which although does not have a significant impact on aggregated EU results, it has a large influence on countries with a large forestry sector (e.g., Sweden). These limitations should be taken into consideration when interpreting results. The detailed vetting process with a high level of spatial detail followed here, offers an important layer of transparency, which is necessary for potential users of the model and the full scenarios that will be developed. At the same time, these results are communicated with the modelling teams to assist with the final stages of development, and the link of the WP4 modules. For example, the deviations in GCAM-Europe at the member state level on CH_4 may assist in the final calibration, while the link between OPEN-PROM and MAgPIE may be informed by the results in total emissions.

Funded by
the European Union

Table 7. Vetting results for the EU27 region.

Variable (EU27 – 2020)	Unit	Eurostat	Criterion	GCAM-EU			GEMINI-E3			OPEN-PROM		
				NECP-LTT Scenario			FIT FOR 55 Scenario			No Policy Reference Scenario		
				Value	Deviation (%)	Pass	Value	Deviation (%)	Pass	Value	Deviation (%)	Pass
Emissions|CO2	MtCO2	2,876.9	±40%	2,699.79	-6.15	✓	3140.14	9.15	✓	2805.90	-2.47	✓
Emissions|CO2|Energy and Industrial Processes	MtCO2	3,073.74	±20%	2,856.89	-7.06	✓	3137.53	2.08	✓	2655.26	-13.61	✓
Emissions|CH4	MtCH4	15.12	±20%	13.21	-12.61	✓	15.08	-0.30	✓	14.43	-4.61	✓
Emissions|CO2|Energy and Industrial Processes 2010-2020 % change	%	*(-16.1%)*	[-30%, 0%]	-17.92	Within range	✓	*-14.4**	*Within range**	✓*	-21.54	Within range	✓
Carbon Capture|Energy	MtCO2	N/A	[0, 20]	0.63	Within range	✓	0	Within range	✓	0	Within range	✓
Primary Energy	EJ	55.88	±20%	48.35	-13.48	✓	52.84	-5.44	✓	55.78	-0.18	✓
Secondary Energy|Electricity|Nuclear	EJ	2.69	±30%	2.62	-2.66	✓	2.57	-4.60	✓	2.46	-8.68	✓
Secondary Energy|Electricity|Solar and Wind	EJ	1.86	±50%	2.37	27.62	✓	1.69	-8.95	✓	1.97	5.86	✓

* GEMINI-E3-EU does not calculate emissions in 2010. The calculation presented is based on the change between the 2010 values from Eurostat and the 2020 value calculated in the model.

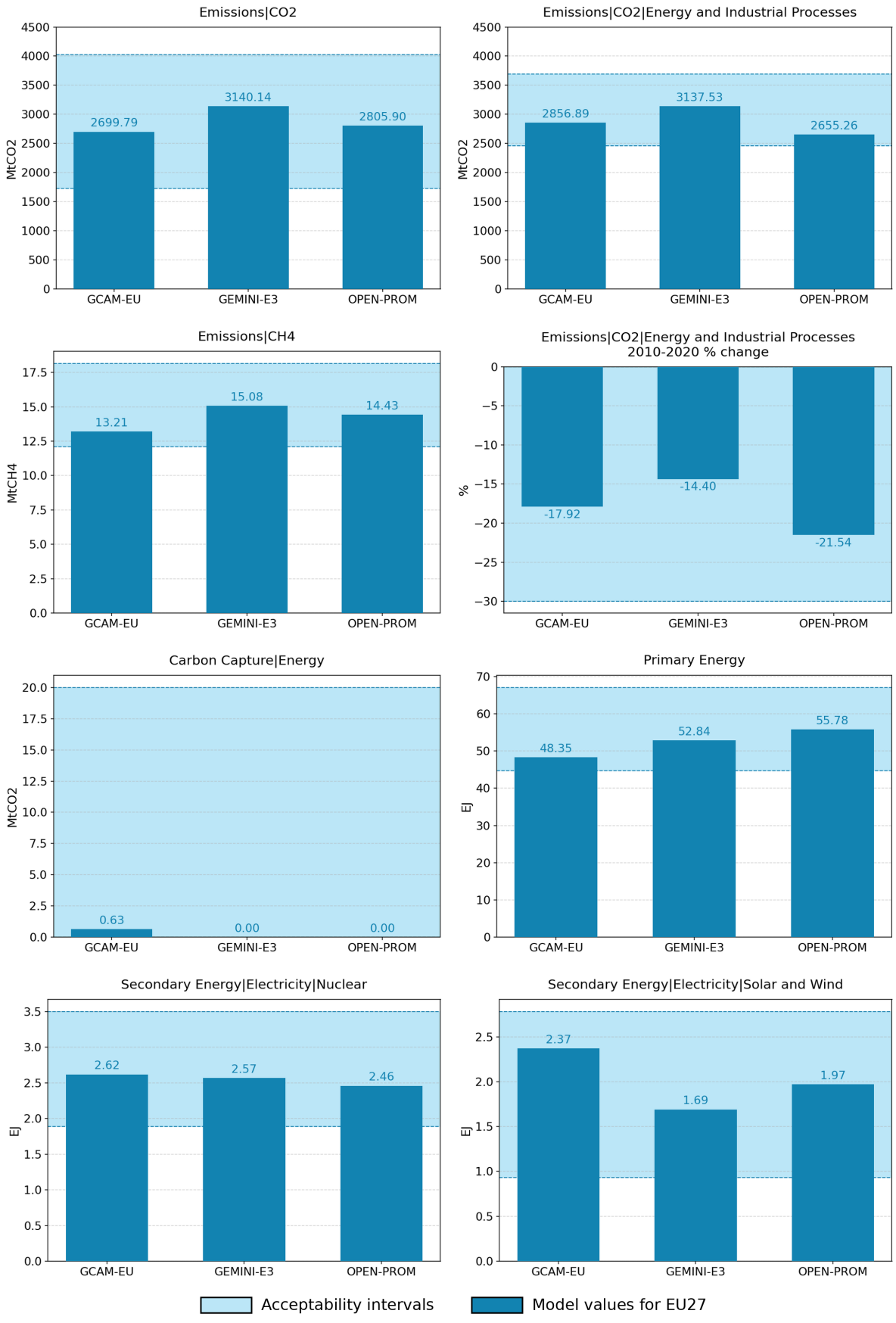


Figure 11. Performance of each model in the eight variables selected for the regional vetting process.

4.4 Model Diagnostics

4.4.1 Theory, methodological approach and stylised algorithms

Model diagnostics have become an increasingly important component of multi-model research in the fields of energy and climate policy analysis. They provide a structured approach for evaluating and comparing IAMs, not only in terms of their quantitative outcomes but also in terms of their internal logic and behavioural consistency. As the complexity of IAMs has grown, with more detailed representations of energy systems, technologies, and macroeconomic feedbacks, understanding the structural sources of divergence among models has become essential. Diagnostic frameworks respond to this need by translating rich model outputs into a small set of informative indicators that describe model responsiveness, transformation capacity, inertia, and cost characteristics in a comparable way. These indicators allow for the identification of systematic differences in model behaviour, revealing how policy shocks affect overall system responses such as the levels and speed of technology deployment, energy substitution, and costs. For example, under harmonised assumptions some models may be quicker to respond to a carbon (or other) tax, while others quicker to react to specific constraints (e.g., coal phase out). In that sense, model diagnostics complement scenario intercomparison exercises by focusing on the mechanisms that generate outcomes rather than the outcomes themselves. This adds a layer of transparency, since model results are accompanied by explanations related to model behaviours, hence allowing informed and more accurate interpretations of outcomes, such as policy recommendations.

The diagnostic indicators applied in this work are based on the IAM diagnostic framework developed by Harmsen et al. (2021), which was designed to reveal structural differences among IAMs and to provide systematic, quantitative insights into how models respond to mitigation policies. The approach builds upon earlier diagnostic work (Kriegler et al., 2015) and offers a set of stylised tests that quantify key aspects of model behaviour, including mitigation responsiveness, transformation pathways, and system inertia. It also constitutes a rather comprehensive and comprehensible list of indicators that could be applied in our case as part of a pipeline instead of a standalone diagnostics exercise. Although these indicators have been primarily used at the global level, the aim is to also calculate such variables at the EU27 level following the pipeline introduced here, to shed light on nuanced behaviours in the IAMs updated and expanded in the project. In particular, the indicators of the diagnostics protocol are presented in Table 8.

Table 8. Description of indicators used in the diagnostics protocol
Source: Harmsen et al. (2021)

Indicator	Brief description
Relative Abatement Index (RAI)	Emission response to mitigation policies
Emission Reduction Type (ERT)	Characterisation of emissions reductions as mainly supply- or demand-driven
Transformation Index (TI)	Structural change in the composition of primary energy
Fossil Fuel Reduction (FFR)	Reduction in fossil primary energy use
Inertia Timescale (IT)	Long term implications and dynamics as a result of systems' inertia in response
Cost per Abatement Value (CAV)	Economic cost of emissions reduction efforts

Funded by the European Union

Each indicator isolates a particular mechanism of model response, while, when used together, they provide a multidimensional view of model behaviour. In the following sections, we provide details on the rationale and the calculation of each indicator, providing pseudo-codes that can be quickly used by prospective users.

4.4.1.1 The Relative Abatement Index (RAI)

RAI measures the sensitivity of modelled CO_2 emissions from energy and industrial processes to the enforcement of mitigation policies. By comparing emissions in a policy and a baseline scenario, it reflects the overall strength of a model's mitigation response. It can be considered the main measure of policy sensitivity, as it summarises how effectively and quickly a model translates policy constraints into emission reductions.

$$RAI(t) = \frac{CO_2E\&I\ Baseline(t) - CO_2E\&I\ Policy(t)}{CO_2E\&I\ Baseline(t)},$$

where $CO_2E\&I(t)$ is the Emissions|CO2|Energy and Industrial Processes IAMC variable at the timeslice t, while *Baseline* and *Policy* refer to the baseline scenario and the assessed policy (i.e. carbon tax) scenario respectively.

```
# import values from dataset as same length vectors:
# CO2_EIP_baseline = …
# CO2_EIP_policy = …

RAI = (CO2_EIP_baseline - CO2_EIP_policy) / CO2_EIP_baseline
```

4.4.1.2 The Emission Reduction Type (ERT)

ERT shows whether emission reductions mainly come from cleaner energy supply (decarbonisation) or from lower energy use (efficiency). It measures the share of total reductions that result from supply-side actions such as renewable energy deployment and fuel switching. The remaining part, i.e., 1 minus ERT, reflects reductions from demand-side changes, mainly improved efficiency and lower final energy demand. ERT values above 0.5 mean that most reductions can be attributes to energy supply side measures, while values below 0.5 point to stronger demand-side effects. The indicator should not be interpreted as binary, but rather a relative comparison between the two contributing factors.

Calculating ERT entails first calculating the Carbon Intensity (CI) and Energy Intensity (EI) indicators for both a baseline and a policy scenario:

$$CI(t) = \frac{CO_2E\&I(t)}{Final\ Energy(t)}$$

$$EI(t) = \frac{Final\ Energy(t)}{GDP(t)}$$

Then, we calculate the CI and EI reduction rates:

$$CI_{red}(t) = \frac{CI_{base}(t) - CI_{policy}(t)}{CI_{base}(t)}$$

$$EI_{red}(t) = \frac{EI_{base}(t) - EI_{policy}(t)}{EI_{base}(t)}$$

Finally:

$$ERT(t) = \frac{CI_{red}(t)}{CI_{red}(t) + EI_{red}(t)}$$

Where $CO_2E\&I(t), Final\ Energy(t)$ and $GDP(t)$ correspond to Emissions|CO2|Energy and Industrial Processes, Final Energy and GDP|MER (or PPP) from the IAMC variables respectively.

```
# Import values from dataset as same length vectors:
# CO2_EIP_baseline = …
# CO2_EIP_policy = …
# Final_energy_baseline = …
# Final_energy_policy = …
# GDP_baseline = …
# GDP_policy = …

# Calculate Carbon Intensity (CI)
CI_baseline = CO2_EIP_baseline / Final_energy_baseline
CI_policy = CO2_EIP_policy / Final_energy_policy

# Calculate Energy Intensity (EI)
EI_baseline = Final_energy_baseline / GDP_baseline
EI_policy = Final_energy_policy / GDP_policy

# Calculate Carbon Intensity reduction rate (CIred)
CIred = (CI_baseline - CI_policy) / CI_baseline

# Calculate Energy Intensity reduction rate (EIred)
EIred = EI_baseline - EI_policy / EI_baseline

# Calculate ERT
ERT = CIred / (CIred + EIred)
```

4.4.1.3 Transformation Index (TI)

TI captures the level of structural change in the composition of primary energy in a given scenario. It is defined as a normalised distance measure between the shares of aggregated energy categories (fossil, bioenergy, non-bio renewables, nuclear) in the base year and in the target year. A TI value approaching zero denotes minimal transformation, whereas higher values (max = 2) indicate a greater reconfiguration of the energy system towards non-fossil sources.

$$TI(t) = \sum_{i=1}^{n} |S_i(t) - S_i(t_0)|$$

Where S_i is the share of energy source contributed by source in primary energy system and t_0 is the base year of the model. So S_i represents S_{fossil}, $S_{non_biomass}$, $S_{biomass}$ and $S_{nuclear}$ which correspond to Primary Energy|Fossil, Primary Energy|Non-Biomass Renewables, Primary Energy|Biomass and Primary Energy|Nuclear IAMC variables respectively.

```
# Import values from dataset as same length vectors
# of whom the first element is the base year value:

# S_fossil = …
# S_non_bio = …
# S_bio = …
# S_nuclear = …

TI = abs(S_fossil - S_fossil[0]) + abs(S_non_bio - S_non_bio[0]) + \
     abs(S_bio - S_bio[0]) + abs(S_nuclear - S_nuclear[0])
```

4.4.1.4 Fossil Fuel Reduction (FFR)

FFR measures the reduction in fossil primary energy use relative to the base year, thereby quantifying the pace and magnitude of fossil fuel phase-out under a scenarios' climate policy.

$$FFR = \frac{Primary\ Energy\ fossil(t_0) - Primary\ Energy\ fossil(t)}{Primary\ Energy\ fossil(t_0)}$$

Where $Primary\ Energy\ fossil(t)$ is the Primary Energy|Fossil IAMC variable at the time slice t and t_0 is the baseyear.

```
# Import values from dataset as a vector
# of which the first element is the base year value:
# Primary_energy_fossil = …

FFR = (Primary_energy_fossil[0] - Primary_energy_fossil) / Primary_energy_fossil[0]
```

4.4.1.5 Inertia Timescale (IT)

IT measures how slowly a model adjusts to the introduction of a climate policy. It quantifies the time required for a delayed-policy scenario to catch up in cumulative CO_2 emissions with a scenario where mitigation begins earlier. The indicator is derived from a dedicated diagnostic experiment in which the introduction of a carbon price policy is delayed, allowing inertia in the energy–economic system to be observed. The difference in emissions between the delayed-policy (shock) scenario and the default early-policy scenario around t_{shock} year (i.e., when policy is implemented in the delayed scenario as well) represents the emissions gap that results from delayed action. After the t_{shock} year, both scenarios are subject to the same carbon price trajectory raising questions on whether the scenarios will gradually converge. Typically, the persistence of higher emissions in the delayed-policy case during this transition period reflects the magnitude of system inertia.

$$IT = \frac{Cumulative\ emissions\ diff.(t_{shock} - t)}{Emissions\ gap(t_{shock})}$$

Where $Cumulative\ emissions\ diff.(t_{shock} - t)$ is the sum of every year's difference between the year that the mitigation policy is enforced (t_{shock}) and the target year for assessed policy and baseline scenarios, using the values of Emissions|CO2|Energy and Industrial Processes IAMC variable. Similarly, the $Emissions\ gap(t_{shock})$ is the point difference of the same variable in year t_{shock} between the two scenarios.

In standard multi-model studies, implementing an additional scenario purely for diagnostic purposes may become too labour-intensive. As such, the indicator can be calculated based on a simpler comparison that draws from a common baseline scenario (e.g., current policies), and a more ambitious one (not necessarily implemented

quicker), usually reflecting long terms dynamics of a system, i.e., how a certain policy may influence future observations even when the effect of the policy has essentially faded out (e.g., learning).

```
# Import values from dataset as same length vectors from year t_shock to t_target:
# of which the first element is the shock year value:

# CO2_EIP_baseline = …      #(from t_shock to t_target)
# CO2_EIP_policy = …        #(from t_shock to t_target)

cummulative_diff = sum(CO2_EIP_policy - CO2_EIP_baseline)
emission_gap_t_shock = CO2_EIP_policy[0] - CO2_EIP_baseline[0]

IT = cummulative_diff / emission_gap_t_shock
```

4.4.1.6 Cost per Abatement Value (CAV)

The CAV indicator expresses the economic implications of emissions reduction efforts at a given level of climate policy. It is a dimensionless ratio between the total policy costs and the marginal abatement costs (MACs), providing a measure of how costly additional emission reductions become as mitigation deepens. In partial-equilibrium (PE) models, CAV reflects the curvature of the implicit MAC curve: values close to one indicate a concave shape and higher overall policy costs, while lower values point to greater mitigation potential at lower carbon prices (Harmsen et al., 2021). In general-equilibrium (GE) models, macroeconomic feedbacks are also included, and values above 1 suggest that these feedbacks play a major role in driving total costs.

$$CAV(t) = \frac{Mitigation\ Costs(t)}{GHG\ Reduction(t) \cdot Carbon\ Price(t)}$$

$Mitigation\ Costs$ refers to costs arising from the implementation of a policy. Although there is no "Policy Cost" definition in the IAMC common-definitions database, there is an ensemble of "Policy Cost" definitions in the AR6 explorer[29]. The indicator may be calculated differently based on the model category. Partial Equilibrium (PE) models typically calculate climate mitigation policy costs as first order sector costs, such as the area under the MAC curve for reducing greenhouse gases. Contrary, general equilibrium (GE) models typically express policy costs in terms of consumption losses or GDP losses (Harmsen et al., 2021). Also, $GHG\ Reduction$ refers to the difference of the Emissions|Kyoto Gases IAMC variable between policy and baseline scenario. Moreover $Carbon\ Price$ corresponds to the Price|Carbon IAMC variable.

```
# Import values from dataset as same length vectors:

# policy_cost = …
# GHG_baseline = …
# GHG_policy = …
# carbon_price = …

GHG_reduction = GHG_policy - GHG_baseline

CAV = policy_cost / (GHG_reduction * carbon_price)
```

[29] https://data.ece.iiasa.ac.at/ar6/#/docs

4.4.2 Indicative Results

The strength of a diagnostics protocol is twofold. First, a comparison between two scenarios from the same model allows cross-validating that the model works as intended (e.g., whether there is indeed a plausible response to a given policy shock). Second, it allows comparing multiple harmonised models under a common scenario protocol, to highlight differences tracing back to model behaviour (e.g., faster response to a policy etc.). Hence, it is evident that performing a full diagnostics exercise requires that two scenarios are implemented for each model, and that these scenarios are harmonised across the models, to meaningfully flesh out the differences. Given the early stages of scenario development in DIAMOND, here we present a numerical example based on models with tentative policy runs focusing on the EU.

In particular, in the present application, the diagnostic procedure is implemented for two IAMs with improved spatial resolution at the EU level, namely GCAM-Europe and GEMINI-E3-EU, drawing from the scenarios presented in Section 4.3. For GCAM-Europe, diagnostics were run on two scenarios, namely a "No Policy Reference" (baseline) and the "NECP-LTT" (policy) scenario. For GEMINI-E3-EU, the baseline corresponds to a "Pre-Fit-for-55 Policies" (baseline) and a "Fit-for-55" (policy) scenario.

All calculations were primarily performed for the aggregated EU27 region. In addition, the indicators were computed for each of the twenty-seven EU Member States, providing a more detailed view of intra-European heterogeneity in model responses. In the case of GEMINI-E3-EU the regional structure is defined through sub-aggregations that merge certain countries into composite regions rather than reporting them individually. The diagnostic results for this model therefore reflect those sub-regional definitions, while maintaining full comparability at the EU27 aggregate level.

Five indicators were calculated for GCAM-Europe except CAV, since the tentative policy runs were not performed with the macro-economic module activated, and hence with GDP calibrated to exogenous variables. However, the model has a functionality to derive the necessary variables for calculating CAV. Also, the IT indicator in GCAM-Europe was adapted, using 2030 as the shock year and 2050 as the target year, consistent with the temporal resolution of the available outputs (2020–2050). It should be noted, however, that this IT calculation serves primarily as a demonstration of the diagnostic workflow, since neither of the two GCAM-Europe scenarios includes an explicit delayed-policy implementation. For GEMINI-E3-EU, the IT indicator was not computed because the scenario was calculated only up to 2030, which is insufficient to capture post-shock adjustment dynamics. The emissions dimension of the CAV indicator was calculated for GEMINI-E3-EU based on the variable Emissions|CO_2|Energy and Industrial Processes. The scope of the temporal coverage of GEMINI-E3-EU in the Fit-for-55 scenario also means that other indicators reflect mostly short-term structural changes (2020–2030).

All diagnostics were implemented using a harmonised computational framework in Python, expected to be integrated in the tool described in Section 4.2.1, following the stylised algorithmic definitions in Harmsen et al. (2021). In the following sub-sections, we present indicative results from this application.

4.4.2.1 Relative Abatement Index

Figure 12 shows the evolution of the RAI over time for GCAM-Europe and GEMINI-E3-EU, for the years available in each model. As explained in the previous section, the RAI can be interpreted as an indicator of the strength of the model response to policy implementation. The clear upward trend indicates that higher carbon prices are consistently associated with stronger emission reductions, as reflected by higher RAI values. The roughly parallel trajectories of the two models suggest that, despite differences in scale and temporal coverage, their relative responsiveness to carbon pricing evolves in a similar way. We should note however that the higher RAI values in GCAM-Europe are not necessarily evidence of stronger mitigation responsiveness, as they largely reflect the fact

that the GEMINI-E3-EU baseline already includes policy measures before the Fit-for-55 implementation, which narrows the difference between the baseline and policy cases, i.e., reducing the RAI marginal difference.

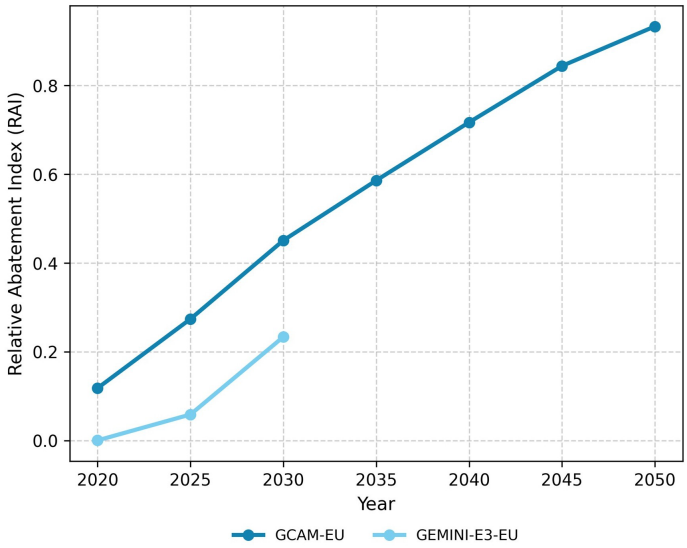


Figure 12. Results for the RAI indicator

4.4.2.2 Emission Reduction Type

In Figure 13 we can observe the relative relationship of energy vs carbon intensity reduction, as well as the evolution of the score of the ERT indicator in the two models. We can observe that GEMINI-E3-EU presents balanced mitigation between supply- and demand-side mitigation, with both intensity reductions increasing symmetrically, while the ERT score remains around the 0.5 value for both the 2025 and 2030 timesteps. GCAM-Europe introduce more dynamic behaviours, driven by the longer time horizon of the analysis. Notably, we observe that the model initially relies heavier on demand-side mitigation until 2030, while after 2035 supply-side measures start to dominate in the emissions reduction observed, with CI reductions coming much faster.

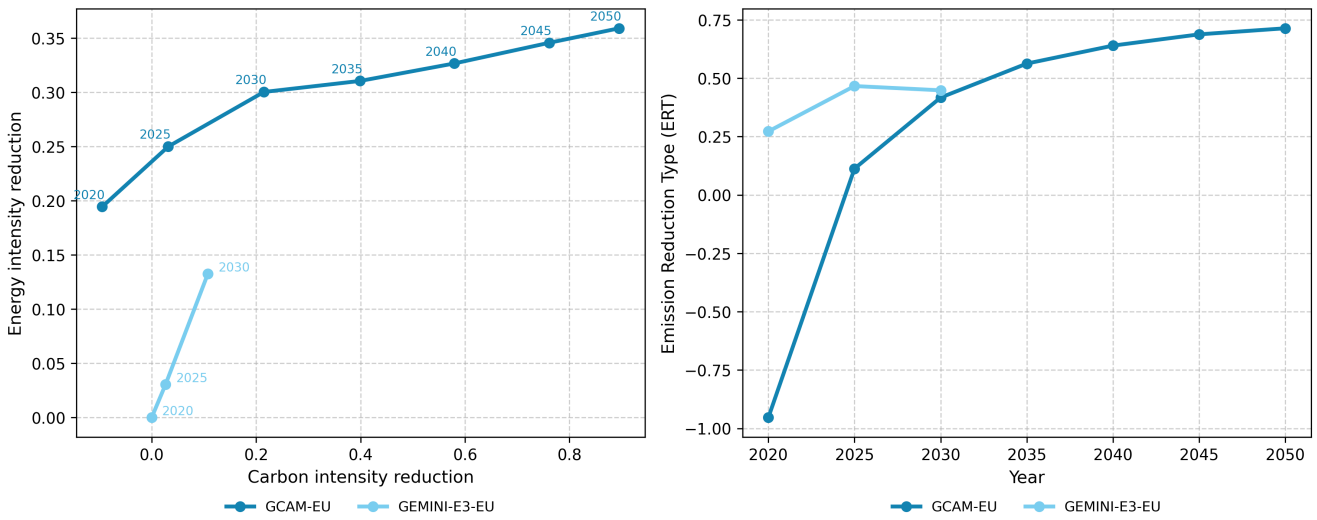


Figure 13. Combined evolution of carbon-energy intensity reduction (left) and results from the ERT indicator (right)

4.4.2.3 Transformation index, fossil fuel reduction, and inertia timescale

The TI and FFR indicators together illustrate how the pace and depth of the energy system transformation differ between the two models. In Figure 14, the FFR trajectories reveal distinct mitigation patterns. In GCAM-Europe,

the policy scenario shows a steady and continuous decline in fossil fuel use, reaching a significant reduction by 2050, while its no-policy baseline scenario remains largely unchanged and even suggests a slight increase in fossil fuel use. GEMINI-E3-EU, in contrast, shows that since both scenarios feature policies there is significant fossil fuel reduction in the two scenarios. Given that the Fit-for-55 package mainly increased emission reduction ambition related to the 2030 target, it is also expected that the two scenarios start diverging after 2025.

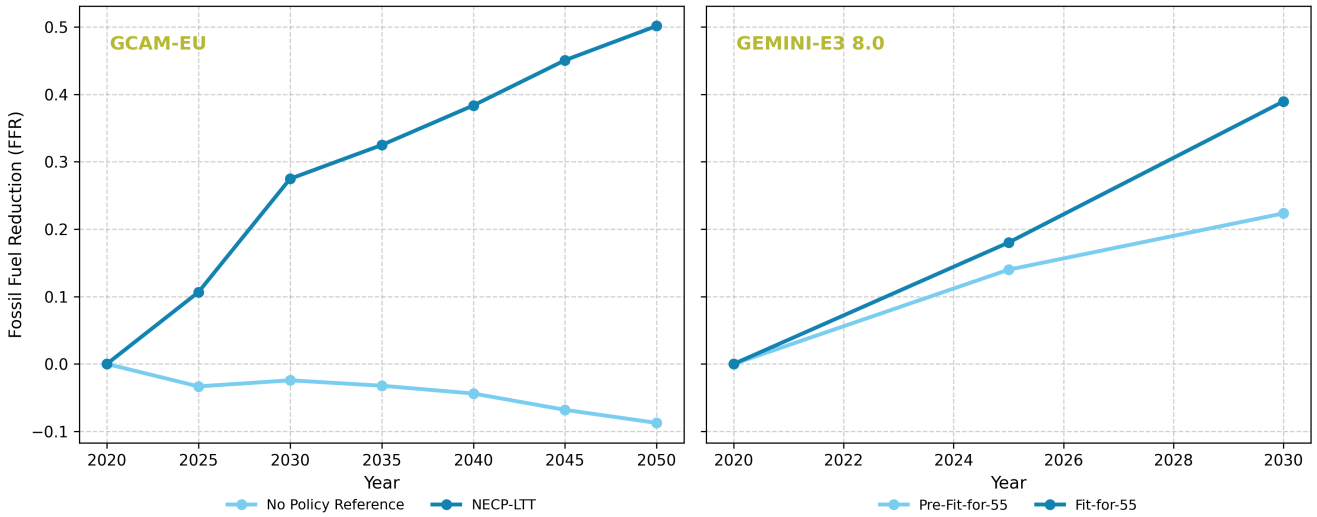


Figure 14. Results from the FFR indicator based on GCAM-Europe (left) and GEMINI-E3-EU (right)

Figure 15 complements these findings by showing the evolution of the TI indicator, which captures the overall transformation of the energy system beyond the fossil fuel sector. GCAM-Europe shows a steady and consistent increase in transformation up to 2050, indicating structural change in the wider energy system, whereas the no policy baseline indicates a much more stable energy system. GEMINI-E3-EU presents the same trend as with the FFR indicator, i.e., convergence until 2025, and then a revamping of transformation in the Fit-for-55 scenario.

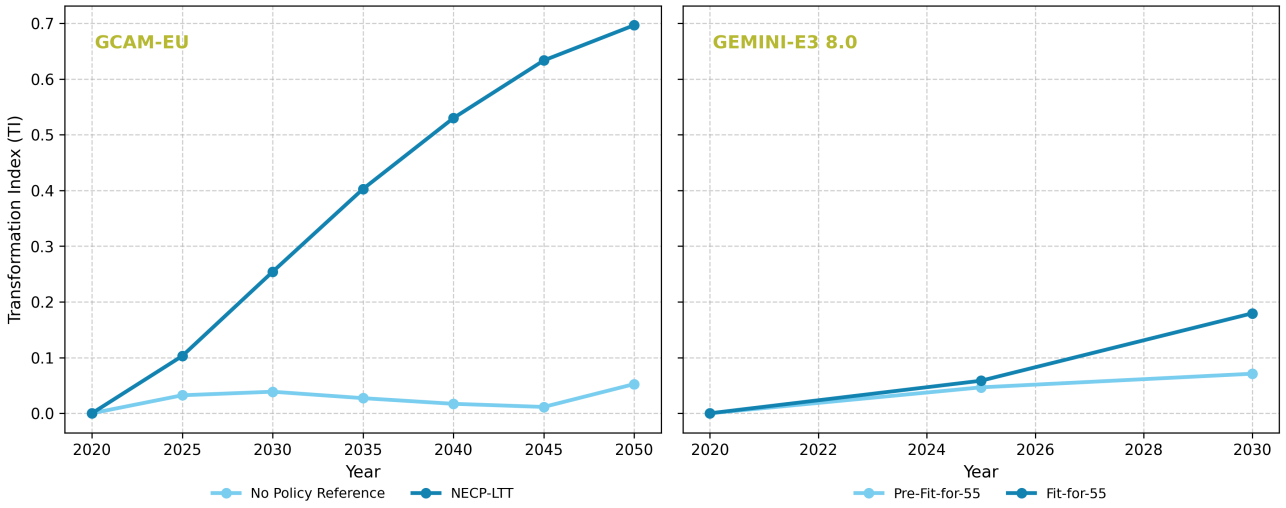


Figure 15. Results from the TI indicator based on GCAM-Europe (left) and GEMINI-E3-EU (right)

For GCAM-Europe, the IT indicator was calculated in 2050, reaching a value of 8.95. Although it is difficult to provide significant commentary without a comparative assessment as a result of using multiple models, this result implies that the policies' effects compound over time rather than fading quickly, and the system doesn't revert to the baseline, which is expected in an NECP-LTT-based scenario compared to a no-policy one. Calculating the

indicator until 2100 to allow the system to show the underlying inertia, as well as providing results from more models, could flesh out deeper insights on the long-term impact of policies.

4.4.2.4 Cost per Abatement Value (CAV)

Figure 16 shows the relationship between the Relative Abatement Index (RAI) and the projected policy cost, with the Cost per Abatement Value (CAV) indicated for each year. The CAV is a dimensionless indicator expressing the ratio between policy costs and marginal abatement costs, providing a measure of the economic efficiency of mitigation. Plotting the RAI against policy cost allows for a direct comparison of the model's emission abatement response with the associated economic burden, while the CAV values highlight how efficiently this abatement is achieved. In the case of GEMINI-E3-EU, the results show that Consumption loss rises sharply (from almost $20b to $100b) between 2025 and 2030, but at the same time emissions reductions are also much larger as a result of the more ambitious policy in 2030. This is driven by a much larger carbon price that drives higher emissions reductions that consumption losses (i.e., lower CAV values), indicating that policy becomes more cost-effective relative to its ambition.

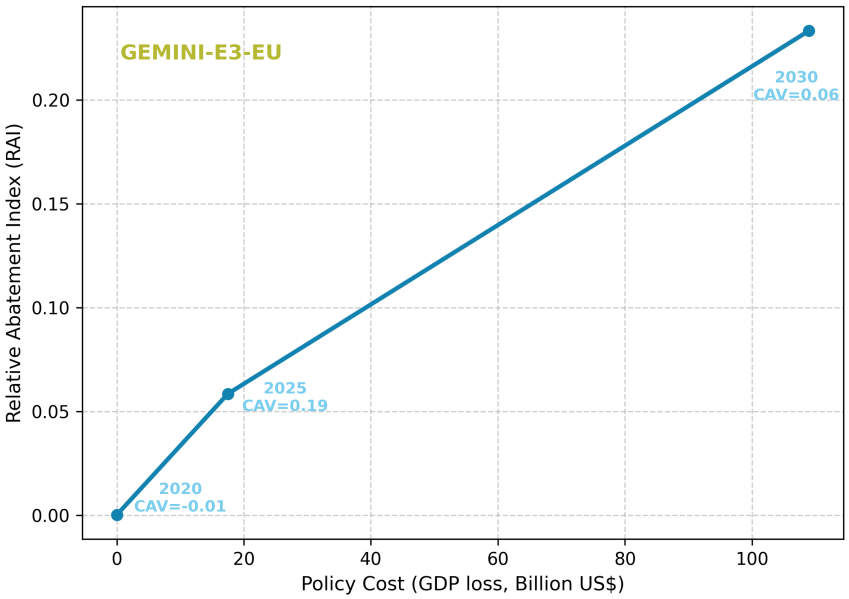


Figure 16. Policy costs vs. Relative Abatement until 2030 in the GEMINI-E3-EU model

5 Scientific Outreach & Stakeholder Validation

5.1 Sharing resources with the scientific community

DIAMOND actively promotes the open exchange of knowledge, methods, and data within the integrated assessment and energy-climate modelling communities. All supporting material, including data processing scripts, diagnostic tools, and validation routines developed under the project are made or will be made available through public repositories on GitHub and Zenodo, ensuring traceability, interoperability, and reusability. These resources are complemented by detailed documentation, and example workflows, allowing external researchers to reproduce the results and adapt the tools to their own modelling contexts.

Beyond code and data sharing, DIAMOND contributes to the broader modelling ecosystem through the IAM PARIS platform[30], which serves as a central hub for model documentation, scenario visualisation, and interactive analysis. The platform hosts detailed metadata for each model, linking results to corresponding datasets and publications, and connecting to the machine-actionable DMP (maDMP) maintained through ARGOS (see D6.1-3). Through this architecture, DIAMOND strengthens the findability and accessibility of modelling resources in line with FAIR principles.

To maximise scientific impact, project partners engage actively with communities of practice such as the Integrated Assessment Modelling Consortium, the OpenMod Initiative, and other open modelling networks, especially ones that are dedicated to activities of specific models (e.g., the Annual GCAM meetings etc.). DIAMOND outputs, including harmonisation protocols, diagnostic scripts, and ontology schemas, are shared and discussed in dedicated scientific workshops, training sessions, and conferences. These activities support transparency and knowledge diffusion and foster collaborations with other initiatives that develop complementary modelling tools or scenario frameworks.

By embedding its open-science practices within established community infrastructures, DIAMOND ensures that its innovations contribute directly to collective progress in climate-economy modelling and support the reproducibility and legitimacy of future integrated assessments.

5.2 Open access scientific publications and data sharing

All DIAMOND scientific publications, datasets, and accompanying materials are disseminated under open-access and open-data principles. Publications are released either in fully open-access journals (such as inter alia Open Research Europe) or under gold open-access options in case institutional agreements exist between partners and publishers, while publications are mirrored in the DIAMOND's Zenodo community to ensure long-term preservation and discoverability. Additionally, the option to publish early pre-prints of the publications is pursued when available. Datasets underpinning these publications, together with model versions, and input and output data, are archived in Zenodo and linked to the project's maDMP in ARGOS, complete with persistent identifiers (DOIs) and rich metadata following the Horizon Europe template.

The project applies a consistent licensing framework to all outputs, through the use of variants of the Creative Commons Attribution (CC BY 4.0) licence for documents and datasets, and appropriate licences for model code (e.g., MIT, Apache 2.0, GPL, ECL-2.0, depending on model needs). This ensures legal clarity for reuse and adaptation while respecting intellectual-property considerations.

To enhance accessibility and interoperability, DIAMOND uses widely supported, non-proprietary file formats (e.g., CSV, TXT, PDF, JSON) and provides APIs or standardised templates for data exchange. The project's FAIR data

[30] https://iamparis.eu/

pipeline connects Zenodo—which also is aligned with TRUST (Transparency, Responsibility, User focus, Sustainability, and Technology) principles (Lin et al., 2020)—ARGOS, and the IAM PARIS platform, enabling automatic cross-referencing between datasets, publications, and repositories.

Through these practices, DIAMOND operationalises open science not merely as a compliance exercise but as an integrated scientific ethos by linking open modelling, transparent data workflows, and open communication of results. In doing so, it contributes to a lasting cultural shift toward openness, reproducibility, and inclusiveness in the integrated assessment modelling community.

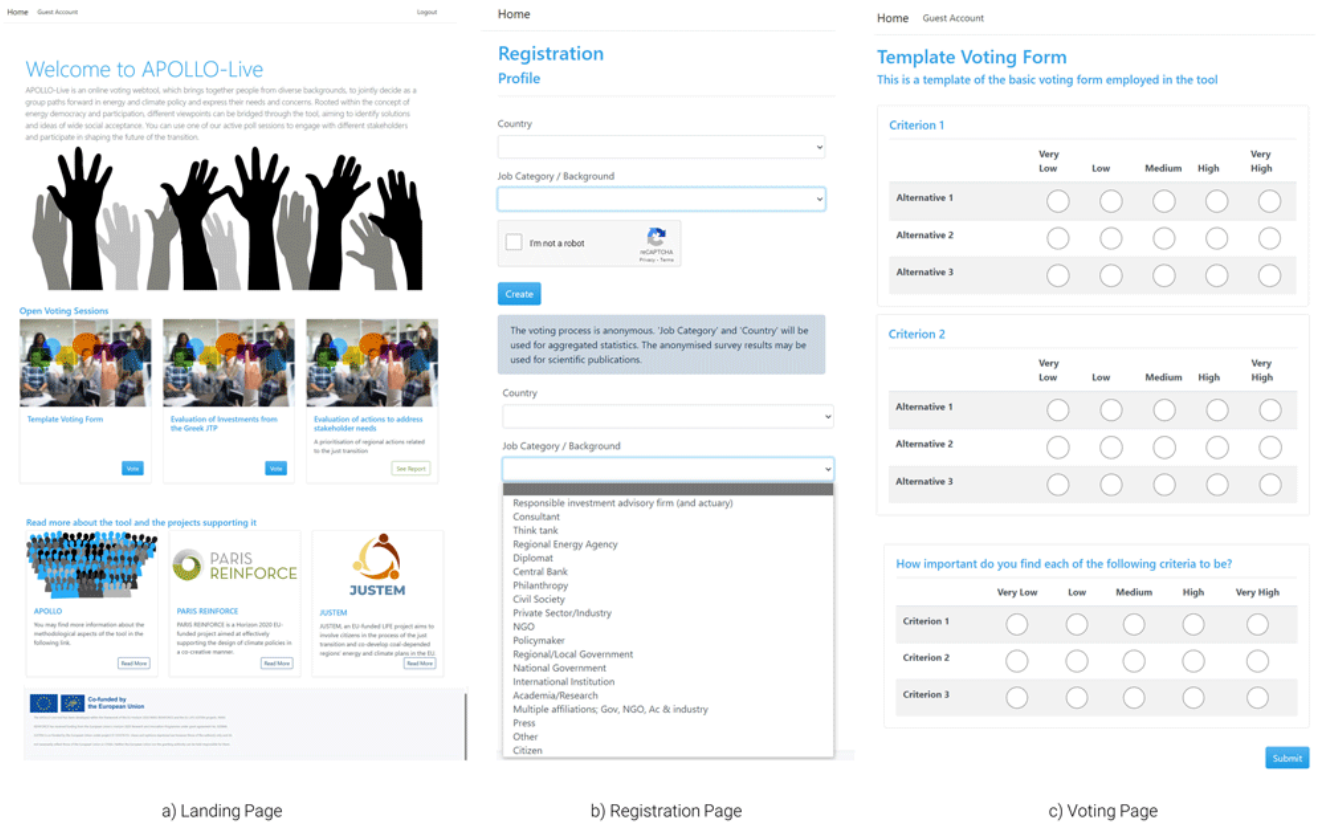
5.3 Stakeholder validation

While technical validation routines and diagnostic checks are essential to ensure internal consistency and empirical credibility of model outputs, they alone cannot guarantee that model results are meaningful or desirable in a broader societal context (Nikas et al., 2020). The final layer of the framework for open science protocols and diagnostics focuses on establishing a stakeholder validation stage, which is deeply rooted within the transdisciplinary (TD) process established and implemented in DIAMOND (see D2.3 and D2.5). Such processes have been identified as crucial good practices in participatory modelling (McGookin et al., 2024). This final step moves beyond technical soundness to assess whether the model's narratives, assumptions, and crucially, outcomes in the form of scenarios align with the values, priorities, and realities of diverse stakeholder groups from academia, policymaking, industry, and civil society. In doing so, the framework bridges analytical robustness with real-world legitimacy, rendering DIAMOND's open IAMs meaningful tools that can inform scientific, policy, and broader societal deliberations on climate action.

As part of the TD process, numerous engagement activities have so far been organised. A series of foresight workshops have been implemented focusing on understanding stakeholder requirements when developing the new and updated modelling capacity, reflecting on high-level insights on model development, as well as flesh out citizen perceptions and priority areas for modelling (more details on the organisation and results from these workshops are presented in MS6 and D2.6). At the same time, a series of multi-stakeholder discussion groups have been organised, deep diving into specific thematic areas of DIAMOND (such as circular economy, climate impacts, finance, sustainable development etc.), to allow the co-creation of research questions between modelling scientists and broader stakeholders that will be pursued in the upcoming activities of the project (more details on the organisation and results from these workshops are presented in D2.5). Phase 3 of the TD process entitled "Towards model co-ownership" will aim to present scenarios and other results from model runs to relevant stakeholders (ideally those who were part of the co-productive process) and invite their feedback on how to make the results most useful in policy and practice. As implementation of scenarios and modelling exercises is expected to intensify in the upcoming period, dedicated activities related to stakeholder validation will take place as part of the TD process to improve true co-ownership of both models and results.

Numerous engagement formats, as well as qualitative and quantitative tools, can assist in this stakeholder validation process (McGookin et al., 2024). DIAMOND contributes to this toolset by explicitly customising the APOLLO-Live open source webtool (Koasidis et al., 2025) to deliver semi-quantitative assessments of model assumptions and results. APOLLO-Live is a stakeholder engagement webtool that can be used live in workshops to facilitate deliberations in energy and climate policy (Figure 17). The tool relies on linguistic variables, which are easily comprehensible by the participants, and employs the 2-tuple TOPSIS group decision making method to prioritise needs faced by different communities, as well as solutions that can be implemented to advance the energy transition. It also fleshes out differences in the voting patterns of different groups of voters and calculates a consensus metric to shed light on conflicts arising. Through tips and suggestions provided within the tool and based on multiple rounds of voting, supplemented by live discussions during the workshop, the tool can improve

consensus and synthesise multiple perspectives in the produced outcomes, assisting towards bridging the conflicts and producing solutions that are widely accepted. APOLLO-Live has been developed as part of the JUSTEM project of the LIFE Programme—building upon early versions of the tool developed under the H2020 PARIS REINFORCE project—and has been employed to contribute to discussions related to the just transition. DIAMOND contributed to integration of the APOLLO-Live webtool within the IAM PARIS ecosystem, enabling its use vis-à-vis IAMs. The flexibility offered in the configuration of stakeholder-informed evaluation workspaces will be useful in the validation exercises, enabling us to establish assessment criteria depending on the specifics of each scenario exercise, as well as priority areas already highlighted in previous exercises. Use of the tool will then inform on best ways forward to make the results more useful or even highlight areas for potential improvement.



a) Landing Page b) Registration Page c) Voting Page

Figure 17. The APOLLO-Live open-source webtool for decision making and consensus support in energy and climate policy.

Source: Koasidis et al., (2025)

6 Conclusions and next steps

In this deliverable we presented a framework for open science protocols and diagnostics, aiming to contribute to the transparency of IAMs by providing detailed processes and implementation tools to the modelling community in order to enhance open practices in the field. The framework spans the whole value chain of IAMs, from the model development phase to model use, as well as to openly providing research outputs and validating them with stakeholders. These distinct, yet highly intertwined, stages aim to provide all the necessary steps to deliver models that are open, transparent, and useful to respond to pertinent societal and policy needs.

During the model development phase, the framework integrates the 18 best practices from software engineering that have been developed during D6.6 and supplements them with an analysis of models that produced scenarios which were featured in the IPCC AR6 to understand the status quo in terms of openness. We highlight that although significant improvements have been made in the community, there is still significant room for improvement, as many IAMs are still closed source (i.e., more than half of the models included in the last IPCC report). By scoping available options from the Open Source Initiative and understanding the specific needs of each of the six IAMs updated and enhanced within the project, we discuss how the licensing choices made constitute major improvements in the openness of the field. Notably, we ensured that all IAMs are made publicly available, many without restrictions, and others with minimal provisions necessary to respect protection of intellectual property rights, without posing any restrictions on the use of the models for research and innovation purposes. As a result, two IAMs are fully open source, including the execution environment, two are open source, but rely on a closed execution environment, and two are provided as source-available tools.

Moving to the model use phase, the framework establishes three processes in the form of a pipeline to ensure that modelling results can be trusted. Notably, we first focused on validation processes to ensure that common errors are identified through simple processes such as inter alia basic sum checks, missing values and name consistency. A key vessel in this process is the IAMC template and the associated ontology to ensure ease of communication between the modelling teams, which are used to ensure consistency in the definition and naming convention of variables. Once validation is clear, a vetting check is performed mimicking the global vetting process established in the IPCC AR6. We move a step further from this application by adapting the process with dedicated benchmarks so that a vetting process can be followed for the EU27 as well, given that the improvements made by many models include enhancing spatial representation. The process also produces results at the member state level to identify issues that can be communicated with modelling teams. We provide a numerical example of this process with three IAMs from DIAMOND with tentative scenarios that were at an advanced stage, showcasing how each passes the vetting process. The final step of the pipeline includes a diagnostics exercise to identify specific model behaviours in the produced results by comparing policy scenarios and baselines across a series of well-defined indicators, so that any model outcome can be associated with a transparent reflection on model tendencies.

During the last stage of the framework, we discuss the delivery of tools, processes, scientific publications and datasets openly to the broader modelling community and scientists from other closely related scientific fields following FAIR principles established in the project's machine-actionable data management plan. Finally, towards increasing the trust of the broader society in the models, we discuss stakeholder validation and tools to support such processes as part of the transdisciplinary approach of DIAMOND to ensure that the models are societally relevant.

Funded by
the European Union

References

Belete, G. F., Voinov, A., & Laniak, G. F. (2017). An overview of the model integration process: From pre-integration assessment to testing. Environmental Modelling and Software, 87, 49–63. Scopus. https://doi.org/10.1016/j.envsoft.2016.10.013

Belete, G. F., Voinov, A., Arto, I., Dhavala, K., Bulavskaya, T., Niamir, L., ... & Filatova, T. (2019). Exploring low-carbon futures: A web service approach to linking diverse climate-energy-economy models. Energies, 12(15), 2880.

Cannone, C., Allington, L., de Wet, N., Shivakumar, A., Goyns, P., Valderrama, C., ... & Howells, M. (2024). clicSAND for OSeMOSYS: a user-friendly interface using open-source optimisation software for energy system modelling analysis. Energies, 17(16), 3923.

Crippa, M., Guizzardi, D., Pagani, F., Banja, M., Muntean, M., Schaaf, E., ... & Vignati, E. (2023). GHG emissions of all world countries. Publications Office of the European Union, Luxembourg, 10, 953322.

Crüwell, S., van Doorn, J., Etz, A., Makel, M. C., Moshontz, H., Niebaum, J. C., ... & Schulte-Mecklenbeck, M. (2019). Seven easy steps to open science. Zeitschrift für Psychologie.

Curley, A., Xexakis, G., Zuiderwijk, A., Minkman, E., & Okur, Ö. Policy Support Platforms on Climate Change Mitigation and Adaptation: An Assessment Framework. Energy Strategy Reviews, 60, 101812.

DeCarolis, J. F., Hannah Daly, Dodds, P., Keppo, I., Li, F., McDowall, W., Pye, S., Strachan, N., Trutnevyte, E., Usher, W., Winning, M., Yeh, S., & Zeyringer, M. (2017). Formalizing best practice for energy system optimization modelling. Applied Energy, 194, 184–198. https://doi.org/10.1016/j.apenergy.2017.03.001

DeCarolis, J. F., Hunter, K., & Sreepathi, S. (2012). The case for repeatable analysis with energy economy optimization models. Energy Economics, 34(6), 1845-1853.

DeCarolis, J. F., Hunter, K., & Sreepathi, S. (2012). The case for repeatable analysis with energy economy optimization models. Energy Economics, 34(6), 1845–1853. https://doi.org/10.1016/j.eneco.2012.07.004

Dekker, M. M., Daioglou, V., Pietzcker, R., Rodrigues, R., De Boer, H. S., Dalla Longa, F., ... & Van Vuuren, D. (2023). Identifying energy model fingerprints in mitigation scenarios. Nature Energy, 8(12), 1395-1404.

DIAMOND. (2025). D6.6 Open-source software development pipelines. DIAMOND project. https://doi.org/10.5281/zenodo.15592781

European Commission. (2021). EU reference scenario 2020: energy, transport and GHG emissions : trends to 2050. Publications Office. https://data.europa.eu/doi/10.2833/35750

European Commission. (2023). 2024 Ageing Report. Underlying Assumptions and Projection Methodologies, Institutional Paper 257. November 2023. Brussels. https://economy-finance.ec.europa.eu/publications/2024-ageing-report-underlying-assumptions-and-projection-methodologies_en

European IP helpdesk. (2024). Different concepts of "openness" in Horizon Europe, Publications Office of the European Union, 2024, https://data.europa.eu/doi/10.2826/921123

Fecher, B., & Friesike, S. (2013). Open science: one term, five schools of thought. In Opening science: The evolving guide on how the internet is changing research, collaboration and scholarly publishing (pp. 17-47). Cham: Springer International Publishing.

Frilingou, N., Van de Ven, D. J., Horowitz, R., Rodés Bachs, C., Mittal, S., Torne, A., ... & Nikas, A. (2025). Cost-optimal vs. policy-driven scenarios for a decarbonised European energy system. https://doi.org/10.21203/rs.3.rs-7847096/v1

Gambhir, A., Butnar, I., Li, P. H., Smith, P., & Strachan, N. (2019). A review of criticisms of integrated assessment models and proposed approaches to address these, through the lens of BECCS. Energies, 12(9), 1747.

Giarola, S., Mittal, S., Vielle, M., Perdana, S., Campagnolo, L., Delpiazzo, E., ... & Van De Ven, D. J. (2021). Challenges in the harmonisation of global integrated assessment models: A comprehensive methodology to reduce model response heterogeneity. Science of The Total Environment, 783, 146861.

Gidden, M. J., Fujimori, S., van den Berg, M., Klein, D., Smith, S. J., van Vuuren, D. P., & Riahi, K. (2018). A methodology and implementation of automated emissions harmonization for use in Integrated Assessment Models. Environmental Modelling & Software, 105, 187-200.

Gidden, M., & Huppmann, D. (2019). pyam: A Python Package for the Analysis and Visualization of Models of the Interaction of Climate, Human, and Environmental Systems. Journal of Open Source Software, 4(33), 1095. https://doi.org/10.21105/joss.01095

Guivarch, C., Kriegler, E., Portugal-Pereira, J., Bosetti, V., Edmonds, J., Fischedick, M., ... & Zhou, N. (2023). Annex III: Scenarios and modelling methods. IPCC 2022: Climate Change 2022: Mitigation of Climate Change. Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, 1842-1908.

Harmsen, M., Kriegler, E., Van Vuuren, D. P., van Der Wijst, K. I., Luderer, G., Cui, R., ... & Zakeri, B. (2021). Integrated assessment model diagnostics: key indicators and model evolution. Environmental Research Letters, 16(5), 054046.

Hausfather, Z., & Peters, G. P. (2020). Emissions–the 'business as usual' story is misleading. Nature, 577(7792), 618-620.

Heaton, D., & Carver, J. C. (2015). Claims about the use of software engineering practices in science: A systematic literature review. Information and Software Technology, 67, 207–219. https://doi.org/10.1016/j.infsof.2015.07.011

Hoesly, R. M., Smith, S. J., Feng, L., Klimont, Z., Janssens-Maenhout, G., Pitkanen, T., ... & Zhang, Q. (2018). Historical (1750–2014) anthropogenic emissions of reactive gases and aerosols from the Community Emissions Data System (CEDS). Geoscientific Model Development, 11(1), 369-408.

Howells, M., Holger Rogner, H., Mentis, D., & Broad, O. (2017). Energy access and electricity planning.

Huppmann, D., Gidden, M. J., Nicholls, Z., Hörsch, J., Lamboll, R., Kishimoto, P. N., ... & Smith, C. J. (2021). pyam: Analysis and visualisation of integrated assessment and macro-energy scenarios. Open Research Europe, 1, 74.

IAMC. (2024). Common definitions for model comparison projects. Retrieved from: https://github.com/IAMconsortium/common-definitions/tree/main

IEA. (2024). Global Energy and Climate Model Documentation 2024, IEA, Paris.

IEA. (2025). Global Energy Review 2025 - CO2 Emissions. Retrieved from: https://www.iea.org/reports/global-energy-review-2025/co2-emissions

IMF. (2024). World Economic Outlook: Policy Pivot, Rising Threats. Washington, DC. October.

Iwanaga, T., Rahman, J., Partington, D., Croke, B., & Jakeman, A. J. (2018). Software Development Best Practices in Integrated Environmental Model Development.

Keppo, I., Al Khourdajie, A., Gardumi, F., Holtz, G., Nikas, A., Xexakis, G., ... & Zamanipour, B. (2026). Model linking for low-carbon transitions: Technical and conceptual challenges and best practices. Renewable and Sustainable Energy Reviews, 226, 116384.

Kikstra, J. S., Nicholls, Z. R., Smith, C. J., Lewis, J., Lamboll, R. D., Byers, E., ... & Riahi, K. (2022). The IPCC Sixth Assessment Report WGIII climate assessment of mitigation pathways: from emissions to global temperatures. Geoscientific Model Development, 15(24), 9075-9109.

Knapen, R., Verweij, P. J. F. M., & Janssen, S. (2010, January 1). Agilists and the Art of Integrated Assessment Tool Development. 2010 International Congress on Environmental Modelling and Software.

Koasidis, K., Nikas, A., & Doukas, H. (2023). Why integrated assessment models alone are insufficient to navigate us through the polycrisis. One Earth, 6(3), 205-209.

Koasidis, K., Soursos, A., Xexakis, G., Labella, Á., Karamaneas, A., & Nikas, A. (2025). APOLLO-Live: A multi-criteria-based webtool for synchronous group decision making and consensus support in energy and climate policy deliberations. Open Research Europe, 5(88), 88.

Landi, A., Thompson, M., Giannuzzi, V., Bonifazi, F., Labastida, I., da Silva Santos, L. O. B., & Roos, M. (2020). The "A" of FAIR–as open as possible, as closed as necessary. Data Intelligence, 2(1-2), 47-55.

Levin, N., Leonelli, S., Weckowska, D., Castle, D., & Dupré, J. (2016). How do scientists define openness? Exploring the relationship between open science policies and research practice. Bulletin of science, technology & society, 36(2), 128-141.

Lin, D., Crabtree, J., Dillo, I., Downs, R. R., Edmunds, R., Giaretta, D., De Giusti, M., L'Hours, H., Hugo, W., Jenkyns, R., Khodiyar, V., Martone, M. E., Mokrane, M., Navale, V., Petters, J., Sierman, B., Sokolova, D. V., Stockhause, M., & Westbrook, J. (2020). The TRUST Principles for digital repositories. *Scientific Data*, *7*(1), Article 1. https://doi.org/10.1038/s41597-020-0486-7

Lo Giudice, C., Gardumi, F., Senaratne, C., Salehu, F., Abdulhalim, S., Tsipouridis, I., Jayasuriya, L., Dodonov, B., Kravtsev, A., & Piddubnyi, I. (2025). IAM_COMPACT_D6.8_Open_access_models_for_case_study_countries_Update. Zenodo. https://doi.org/10.5281/zenodo.17140940

McGookin, C., Süsser, D., Xexakis, G., Trutnevyte, E., McDowall, W., Nikas, A., ... & Gallachóir, B. Ó. (2024). Advancing participatory energy systems modelling. Energy Strategy Reviews, 52, 101319.

Nikas, A., Doukas, H., & Papandreou, A. (2019). A detailed overview and consistent classification of climate-economy models. Understanding risks and uncertainties in energy and climate policy, 1-54.

Nikas, A., Lieu, J., Sorman, A., Gambhir, A., Turhan, E., Baptista, B. V., & Doukas, H. (2020). The desirability of transitions in demand: Incorporating behavioural and societal transformations into energy modelling. Energy Research & Social Science, 70, 101780.

Nishikawa, K., & Murakami, A. (2025). Does open access foster interdisciplinary citations? Decomposing open access citation advantage. Scientometrics, 1-24.

Perdana, S., & Vielle, M. (2025). Industrial European regions at risk within the fit for 55: How far implementing CBAM can mitigate?. Renewable and Sustainable Energy Transition, 6, 100088.

Peters, G. P., Al Khourdajie, A., Sognnaes, I., & Sanderson, B. M. (2023). AR6 scenarios database: an assessment of current practices and future recommendations. npj Climate Action, 2(1), 31.

Pfenninger, S., Hirth, L., Schlecht, I., Schmid, E., Wiese, F., Brown, T., ... & Wingenbach, C. (2018). Opening the black box of energy modelling: Strategies and lessons learned. Energy Strategy Reviews, 19, 63-71.

Rodés-Bachs, C., Sampedro, J., Frilingou, N., Gardumi, F., & Giudice, C. L. (2025). Open Science Practices in Integrated Assessment Models. Open Research Europe, 5, 12.

Sandve, G. K., Nekrutenko, A., Taylor, J., & Hovig, E. (2013). Ten Simple Rules for Reproducible Computational Research. PLOS Computational Biology, 9(10), e1003285. https://doi.org/10.1371/journal.pcbi.1003285

Schmitz, L., & Seidl, T. (2023). As open as possible, as autonomous as necessary: Understanding the rise of open strategic autonomy in EU trade policy. JCMS: journal of common market studies, 61(3), 834-852.

Skea, J., Shukla, P., Al Khourdajie, A., & McCollum, D. (2021). Intergovernmental Panel on Climate Change: Transparency and integrated assessment modeling. Wiley Interdisciplinary Reviews: Climate Change, 12(5), e727.

Sognnaes, I., & Peters, G. P. (2025). Influence of individual models and studies on quantitative mitigation findings in the IPCC Sixth Assessment Report. Nature Communications, 16(1), 8343.

Sognnaes, I., Gambhir, A., van de Ven, D. J., Nikas, A., Anger-Kraavi, A., Bui, H., ... & Peters, G. P. (2021). A multi-model analysis of long-term emissions and warming implications of current mitigation efforts. Nature Climate Change, 11(12), 1055-1062.

Stockhause, M., Huard, D., Khourdajie, A. A., Gutiérrez, J. M., Kawamiya, M., Klutse, N. A. B., Krey, V., Milward, D., Okem, A. E., Pirani, A., Sitz, L. E., Solman, S. A., Spinuso, A., & Xing, X. (2024). Implementing FAIR data principles in the IPCC seventh assessment cycle: Lessons learned and future prospects. PLOS Climate, 3(12), e0000533. https://doi.org/10.1371/journal.pclm.0000533

Van Beek, L., Hajer, M., Pelzer, P., van Vuuren, D., & Cassen, C. (2020). Anticipating futures through models: the rise of Integrated Assessment Modelling in the climate science-policy interface since 1970. Global Environmental Change, 65, 102191.

Vicente-Saez, R., & Martinez-Fuentes, C. (2018). Open Science now: A systematic literature review for an integrated definition. Journal of business research, 88, 428-436.

Weigmann, P., Mandaroux, R., Lécuyer, F., Merfort, A., Dorndorf, T., Hoppe, J., ... & Luderer, G. (2025). Validation of climate mitigation pathways. EGUsphere, 2025, 1-20.

Weyant, J. (2017). Some contributions of integrated assessment models of global climate change. Review of Environmental Economics and Policy.

Wilson, C., Guivarch, C., Kriegler, E., van Ruijven, B., van Vuuren, D. P., Krey, V., Schwanitz, V. J., & Thompson, E. L. (2021). Evaluating process-based integrated assessment models of climate change mitigation. Climatic Change, 166(1–2). https://doi.org/10.1007/s10584-021-03099-9

Wilson, G., Bryan, J., Cranston, K., Kitzes, J., Nederbragt, L., & Teal, T. K. (2017). Good enough practices in scientific computing. PLOS Computational Biology, 13(6), e1005510. https://doi.org/10.1371/journal.pcbi.1005510

Annex

Table 9. GEMINI-E3-EU vetting at a higher regional granularity within the EU

Variable	Country	Year	Eurostat (MtCO2)	GEMINI-E3-EU (MtCO2)	Deviation	Flag
Emissions|CO2|EIP	EU1	2020	156.7	164.2	4.8	
Emissions|CO2|EIP	EU2	2020	79.5	81.1	2.0	
Emissions|CO2|EIP	EU3	2020	182.7	189.5	3.7	
Emissions|CO2|EIP	EU4	2020	133.8	147.5	10.2	
Emissions|CO2|EIP	EU5	2020	35.1	45.8	30.3	⚑
Emissions|CO2|EIP	EU6	2020	115.6	149.8	29.6	⚑
Emissions|CO2|EIP	Belgium	2020	127.0	108.7	-14.4	
Emissions|CO2|EIP	France	2020	334.4	339.8	1.6	
Emissions|CO2|EIP	Germany	2020	716.3	732.4	2.3	
Emissions|CO2|EIP	Italy	2020	351.1	347.0	-1.2	
Emissions|CO2|EIP	Netherlands	2020	191.7	197.5	3.1	
Emissions|CO2|EIP	Poland	2020	326.0	319.7	-1.9	
Emissions|CO2|EIP	Spain	2020	275.7	270.4	-1.9	
Emissions|CO2|EIP	Sweden	2020	48.2	44.0	-8.7	
Emissions|CO2|EIP	**EU27**	**2020**	**3073.7**	**3137.5**	**2.1**	

Notes:

- Aggregated regions:
  - EU1: Croatia, Cyprus, Greece, Portugal, Malta
  - EU2: Austria, Luxembourg
  - EU3: Czech Rep., Hungary, Slovak Rep.
  - EU4: Bulgaria, Romania, Slovenia
  - EU5: Estonia, Latvia, Lithuania
  - EU6: Denmark, Finland, Ireland
- Eurostat 2020 values are replaced by the average between 2019 and 2021 to avoid COVID-19 distortions

Funded by the European Union