https://doi.org/10.5194/egusphere-2025-2284
Preprint. Discussion started: 2 June 2025

Validation of climate mitigation pathways

Pascal Weigmann[1,*], Rahel Mandaroux[1,*], Fabrice Lecuyer[1], Anne Merfort[1], Tabea Dorndorf[1], Johanna Hoppe[1], Jarusch Muessel[1], Robert Pietzcker[1], Oliver Richters[1], Lavinia Baumstark[1], Elmar Kriegler[1], Nico Bauer[1], Falk Benke[1], Chen Chris Gong[1], and Gunnar Luderer[1]

[1]Potsdam Institute for Climate Impact Research (PIK), Member of the Leibniz Association, Potsdam, Germany
[*]These authors contributed equally to this work.

Correspondence: Pascal Weigmann (pascal.weigmann@pik-potsdam.de) and Rahel Mandaroux (rahel.mandaroux@pik-potsdam.de)

Abstract. Integrated assessment models (IAMs) are crucial for climate policymaking, offering climate mitigation scenarios and contributing to IPCC assessments. However, IAMs face criticism for lack of transparency and poor capture of recent technology diffusion and dynamics. We introduce the Potsdam Integrated Assessment Modeling validation tool, *piamValidation*, an open-source R package for validating IAM scenarios. The *piamValidation* tool enables systematic comparisons of variables from extensive IAM datasets against historical data and feasibility bounds, or across scenarios and models. This functionality is particularly valuable for harmonizing scenarios across multiple IAMs. Moreover, the tool facilitates the systematic comparison of near-term technology dynamics with external observational data, including historical trends, near-term developments, and stylized facts. We apply the tool to the integrated assessment model REMIND for near-term technology trend validation, demonstrating its potential to enhance transparency and reliability of IAMs.

1 Introduction

Integrated assessment models (IAMs) play a prominent role for science-based and climate policy advice. Early IAM applications date back to the late 1990s (Cointe et al., 2019) and evolved towards the formulation of mitigation targets and the monitoring of political ambition (Van Beek et al., 2020). IAM scenarios are also an important pillar of the assessments conducted by the Intergovernmental Panel on Climate Change (IPCC), in particular with regard to transformation pathways towards achieving climate policy goals. Therefore, they are a central component of climate change mitigation oriented policies. Edenhofer and Minx (2014) also point out that the Summary for Policymakers (SPM) established a collaborative environment where political discussions link with relevant scientific material. Relating the Conference of the Parties (COP) through a metaphor, "the COP operates as navigator, navigating a terrain charted by the IPCC" (Beck and Oomen, 2021, p. 172).

Despite the prominent role of IAMs in the climate science-policy interface, IAMs have faced increasing scrutiny. Particularly in policy advice, central criticism ranges from shortcomings regarding model representation (e.g., heterogeneous actors and capital markets) to their ability to capture technology diffusion and dynamics (Keppo et al., 2021). On the one hand, global and national models have often underestimated the rapid technological change in novel renewable power and demand electrification technologies: for instance, the cost of solar electricity and battery storage has declined by almost 90% (Creutzig et al., 2023)

https://doi.org/10.5194/egusphere-2025-2284
Preprint. Discussion started: 2 June 2025

and renewable electricity generation in Asia has more than doubled (IRENA, 2024) in the last decade, far exceeding model results. National models face the additional challenge of making assumptions about technology innovation and cost declines, which are largely driven by global developments and trends. Accurately capturing recent trends in technology capital costs is therefore crucial, as they play a predominant role in the sensitivity of model outputs (Giannousakis et al., 2021).

On the other hand, near-term IAM scenarios results exceeded the rate at which economies are transitioning away from fossil fuels: for example, due to factors beyond pure market-based energy economics, coal power capacities are still being developed in some regions despite the declining renewable energy costs. In particular, China and India alone account for 86 percent of worldwide coal power capacity under development (Carbon Brief, 2024).

Not only techno-economic parameters but also the technology representations more generally impact technology trends in IAM scenarios and vary strongly between IAMs, illustrated by Krey et al. (2019) for electricity generation. They first highlight the need for caution in harmonizing techno-economic assumptions, as they must fit model-specific technology representations and associated projection methods. Second, Krey et al. (2019) participate in the transparency debate (Stanton et al., 2009) by urging the IAM community to publish and openly discuss techno-economic parameters, their definitions and model technology representations.

This paper introduces the Potsdam Integrated Assessment Modeling validation tool for IAMs, *piamValidation*. Our objective is to enable a broad use of that tool in the IAM community to enhance transparency and reliability. By aligning with the key evaluation criterion outlined by Wilson et al. (2021), the *piamValidation* tool can contribute to increasing IAMs near-term realism. This tool and the related input data are meant to be a community resource, where the distribution of the *piamValidation* configuration files can serve as a knowledge sharing platform and as a performance metric, offering continuous feedback on modeling progress.

In section 2, we present the *piamValidation* open-source R package. The tool enables users to systematically compare variables of large IAM datasets with historical data and feasibility bounds, or across scenarios and models. This is primarily useful for harmonizing national and global IAMs in the short and medium term, particularly for current policy scenarios that capture implemented government policies. In addition, the tool allows for systematic comparisons of near-term technology dynamics with external observational data as historical, near-term trends and stylized facts. The *piamValidation* tool is particularly user-friendly and requires only a single command to generate the full HTML report featuring validation heat maps.

In section 3, we apply the *piamValidation* tool to validate near-term technology dynamics in the IAM REMIND (REgional Model of Investment and Development, see Baumstark et al. (2021) for details). Within our application case, we generate interactive heat maps presenting technology trend capture performance in the form of a traffic light evaluation. The heat maps, combined with the tool's configuration file, contain central information regarding observational data, defined bounds and scenario deviations.

We select central variables related to CO2 transport and storage, electric vehicles and offshore wind power. The choice of these technologies is twofold: first, the feasibility of their near-term scaling has severe implications for scenarios with high ambition for emissions reduction (Brutschin et al., 2021; Bertram et al., 2024); second, the *piamValidation* tool highlights

https://doi.org/10.5194/egusphere-2025-2284
Preprint. Discussion started: 2 June 2025

deviations of REMIND scenarios from historical data and current technology trends. We then present strategies that have been implemented to improve REMIND, and we demonstrate that subsequent scenarios improve their near-term realism.

2 PIAM validation tool

2.1 Overview

The package *piamValidation* provides validation tools for the Potsdam Integrated Assessment Modeling environment. It is written in R, with the source code being openly available under the LGPL-3 license on GitHub[1].

Users provide the tool with IAM scenario data and relevant reference data for historical or future time periods. Criteria for checks are defined in a configuration file to perform a vetting. This configuration file serves as a flexible interface, which allows for various use cases:

- Deviations of scenario to reference data
- Deviations between:
  - Models (same scenario, same periods);
  - Scenarios (same model, same periods);
  - Periods (same model, same scenario).
- Conformity to scalar thresholds, such as an absolute value or growth rate.

Deviations can be defined as either absolute or relative deviations. The function `validateScenarios` takes the inputs mentioned above and evaluates the validation checks. In the standard setting, this results in a dataset that contains an evaluation of either 'green' (test passed), 'yellow' (warning), or 'red' (fail) for each data point to be evaluated. Optionally, additional colors can be used to differentiate violating a lower bound from violating an upper bound. In that case, "blue" and "cyan" are used to hint toward values below thresholds, and "yellow" and "red" remain to capture upper exceedances.

The function `validationHeatmap` makes these results accessible by arranging them for one variable on an interactive heat map using the R library *ggplotly*. Depending on the category of checks that are being performed, data underlying the evaluation is available for each data point by hovering over the respective square in the heat map. Finally, `validationReport` can be used to perform the validation calculation and plot all heat maps in one HTML report, which is automatically rendered from an RMarkdown file provided by the package.

After successfully installing the *piamValidation* R package in an appropriate R environment, the tool structure can be described along three dimensions: Input, data processing, and output, as illustrated in Figure 1.

[1] https://github.com/pik-piam/piamValidation

https://doi.org/10.5194/egusphere-2025-2284
Preprint. Discussion started: 2 June 2025

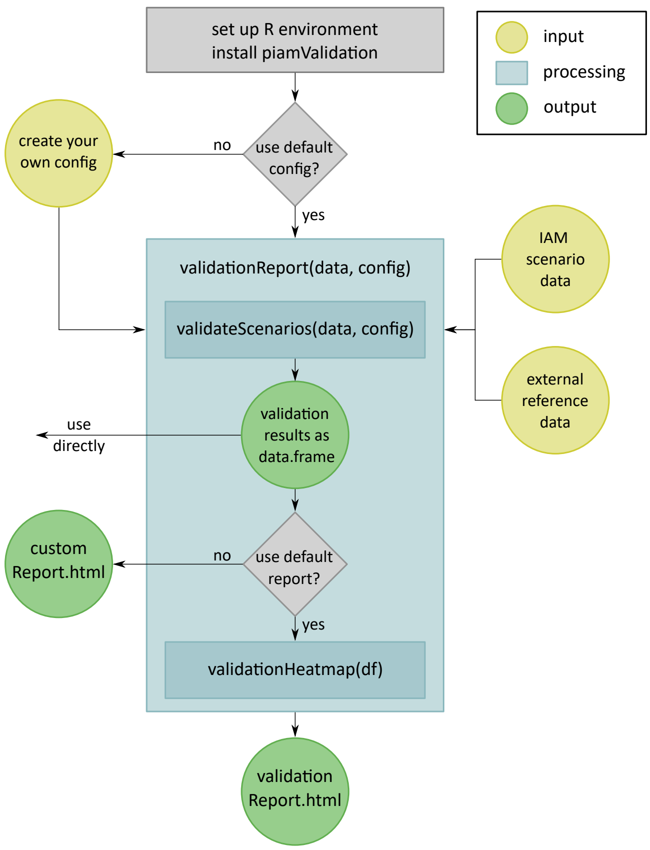

Figure 1. Validation workflow with respective inputs, processing steps and outputs.

2.2 Data Input

The input data for the *piamValidation* tool comprises three key components. First, the IAM scenario data to be validated. Second, the reference data, which is used to compare the initial IAM scenario data. This reference data may include outputs from other IAM models or scenarios, as well as third-party external datasets, such as historical, observational or other prediction data. Third, a configuration file is necessary to specify the details of the validation. The *piamValidation* R package provides a standard set of configuration files for general use. However, users seeking greater flexibility can create and customize their own configuration files to address specific validation requirements.

To facilitate use by different IAM communities and external users, the standardized way of organizing IAM data following the Integrated Assessment Consortium (IAMC) format [2] is used. It organizes data along the dimensions of "model", "scenario", "variable", "region" and "period". Thus, the validation process works with 5-dimensional data objects, expecting a data point to be uniquely defined as $x(m,s,v,r,p)$, with the indices referring to these five dimensions.

[2]https://www.iamconsortium.org/scientific-working-groups/data-protocols-and-management/iamc-time-series-data-template/

https://doi.org/10.5194/egusphere-2025-2284
Preprint. Discussion started: 2 June 2025

2.2.1 Configuration File

The *piamValidation* framework allows for a wide range of different validation approaches. They are defined via a configuration file with predefined columns as described in Tab.1. This is intended to enable users with minimal programming knowledge to use the tool, as the structure of the validation configuration file follows the intuitive idea of a validation process, answering the following three guiding questions:

1. Which type of validation should be performed using which metric?
2. Which data should be validated?
3. What is the reference being used for the validation?

These questions are answered by filling in the columns of the configuration file as described in Tab.1.

https://doi.org/10.5194/egusphere-2025-2284
Preprint. Discussion started: 2 June 2025

Column	Description
metric	Decide which type of comparison will be performed, currently supported: "relative", "difference", "absolute", "growthrate".
critical	Is it considered to be a critical check? The function `validationPass` will report failed checks only if this is set to “yes”.
variable	Choose one or multiple variables to be checked: Define multiple variables via “*” (one sub-level) or “**” (all sub-levels), e.g. "Final Energy|*" for "Final Energy|Industry" and "Final Energy|Transportation", etc. and "Final Energy|**" for "Final Energy|Industry|Electricity", "Final Energy|Industry|Electricity|Cement" and so on.
unit	This is an optional column that allows to verify that scenario - and reference data are consistent in their use of units.
model	Choose one or multiple (comma-separated) models, or leave empty to choose all.
scenario	Choose one or multiple (comma-separated) scenarios, or leave empty to choose all.
region	Choose one or multiple (comma-separated) regions, or leave empty to choose all.
period	Choose one or multiple (comma-separated) periods, or leave empty to choose all (generally "all" defaults to =< 2100 in this case, for historical checks 2005-2020). A range of periods can be defined via "yyyy-yyyy".
min/max_yel/red	Define minimum and maximum thresholds which decide whether a check is passed (green), produces a warning (yellow) or is failed (red). If reference data is missing, the result will be gray. Each line needs at least one threshold. Relative thresholds can be given either as percentage (e.g., 20%) or decimal (e.g., 0.2).
ref_model	For model intercomparison, set one or multiple reference model here, all models chosen in the column "model" will be compared to it.
ref_scenario	This column can be used in two ways - either compare two scenarios produced by the same model to one-another (similar to the model intercomparison), or set it to *historical* to compare model data to observational data provided by one or multiple external sources. It is recommended to choose a historical reference source explicitly in the “ref_model” column for historical comparisons.
ref_period	Compare data between different periods, set one or multiple reference periods here.
notes	An optional column that allows leaving comments about the thresholds or checks in this row.

Table 1. Column description of configuration file. Note: Each described column must be filled in accordance to the use case to ensure successful operation of the *piamValidation* tool.

https://doi.org/10.5194/egusphere-2025-2284
Preprint. Discussion started: 2 June 2025

2.2.2 Validation types and metrics

A common validation procedure consists in comparing IAM results to benchmarks, such as observational data from external data sources. Such reference data can be handed to the validation tool in large quantities by supplying a CSV, Excel or MIF file that is structured along the same five dimensions as the IAM data, with the specific requirement that the "scenario" column reads "historical" for all data points. This way, the tool can also differentiate between a model intercomparison exercise, where the same scenario is compared between two or more models and a comparison of multiple scenarios to one or multiple reference model scenarios, which don't require the same scenario name. In terms of data dimensions, a relative deviation can thus be expressed as

$$\text{rel\_deviation} = \frac{x(m_i, s_j, v_k, r_l, p_m) - x(m_{\text{ref}}, s_{\text{hist}}, v_k, r_l, p_m)}{x(m_{\text{ref}}, s_{\text{hist}}, v_k, r_l, p_m)} \quad ,$$

where $x(m_i, s_j, v_k, r_l, p_m)$ are the IAM values using the notation introduced in 2.2 and $x(m_{\text{ref}}, s_{\text{hist}}, v_k, r_l, p_m)$ are the observational data values from one or multiple reference models. In an analogous way, a “difference to historical” check can be performed by defining an absolute deviation to the reference data as

$$\text{abs\_deviation} = x(m_i, s_j, v_k, r_l, p_m) - x(m_{ref}, s_{hist}, v_k, r_l, p_m).$$

Mathematically similar, checks with the metric “relative” or “difference” can also be performed against subsets of the scenario data itself, which then functions as reference data. This can in turn be performed in three different ways - comparing to a selected period, scenario or model:

$$\text{abs\_deviation\_period} = x(m_i, s_j, v_k, r_l, p_m) - x(m_i, s_j, v_k, r_l, p_{ref}) \quad ,$$
$$\text{abs\_deviation\_scen} = x(m_i, s_j, v_k, r_l, p_m) - x(m_i, s_{ref}, v_k, r_l, p_m) \quad or$$
$$\text{abs\_deviation\_model} = x(m_i, s_j, v_k, r_l, p_m) - x(m_{ref}, s_j, v_k, r_l, p_m).$$

Relative comparisons again divide this difference by the reference value. Note, that all other dimensions are the same between scenario and reference data with only one dimension (period, scenario or model) acting as validation dimension (it is not possible to compare scenario s_j of model m_i with scenario s_{ref} of model m_{ref}). The reference of this dimension can be either a single period/scenario/model or a comma-separated list of them. Furthermore, when choosing multiple sources the user can decide whether the validation thresholds should be based on the *range* of the reference values, where the maximum value of the group is used as the reference for upper and the minimum as the reference for lower thresholds or their *mean* will serve as reference for all thresholds.

For instance, in a near-term validation study, multiple data sources can be chosen as references by selecting the following expression in the config file as ref_model[3]:

$$range(m_i, m_j, ..., m_x).$$

[3] Note: The same holds for ref_scenario and ref_period.

https://doi.org/10.5194/egusphere-2025-2284
Preprint. Discussion started: 2 June 2025

The metric "absolute" does not require any reference data, as it contains the absolute values of the thresholds directly in the threshold columns of the configuration file. This means that no calculations are required.

Lastly, the "growthrate" metric can be used to perform validation checks with respect to average yearly growth rates. For a time step of dt between periods, the growth rate is calculated as

$$\text{growthrate} = \left(\frac{x(m_i, s_j, v_k, r_l, p_m)}{x(m_i, s_j, v_k, r_l, p_{m-dt})} \right)^{\frac{1}{dt}} - 1 \ .$$

Depending on the metric of the validation check and whether it is a comparison to observational data, different columns of the configuration file need to be filled as shown in Figure 2.

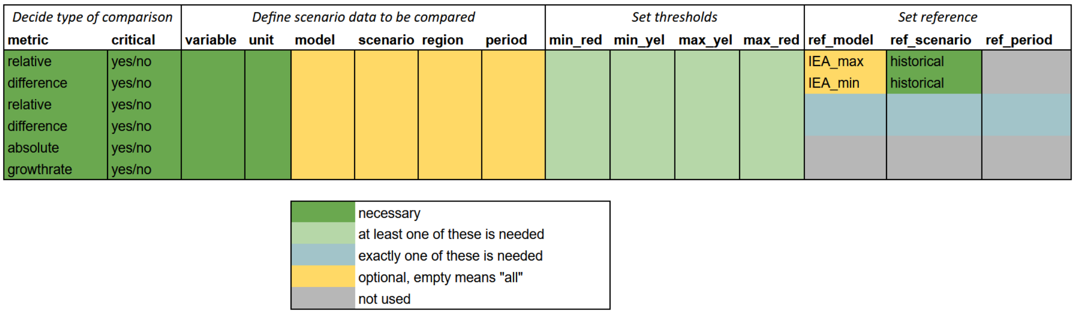

Figure 2. Overview of required and optional columns of the configuration file, depending on metric.

2.3 Data Processing

The validation configuration file allows users to define a wide range of validation checks, many of which might not be anticipated during creation of the validation tool. On the one hand, this requires the data processing and evaluation to be consistent and flexible at the same time, e.g., if only some checks require an intermediate calculation step but the output needs to be a single harmonized data object containing all checks. On the other hand, users might supply the validation tool with data and configuration files that are either internally inconsistent or exceed the scope of the currently supported use cases. In such cases, it is crucial to ensure transparency by allowing the user to comprehend the issue through clear mechanisms and the inclusion of essential information in error messages.

These requirements are met by the modular structure of the tool, where each data processing step is carefully separated from the next. The structure of the central data processing function of the validation tool, `validateScenarios`, is sketched using pseudocode in Alg.1.

At the beginning of a validation process using the `validateScenarios` function, the configuration file and scenario and reference data are imported and checked for consistency (see Algorithm 1). The loop starting in line 4 takes information from one row of the configuration file and assembles the scenario and reference data needed for the validation check defined. It connects the thresholds that are relevant for each data point to the scenario data. In case a validation check is performed against

https://doi.org/10.5194/egusphere-2025-2284
Preprint. Discussion started: 2 June 2025

a reference value (either historical or period/scenario/model intercomparison), also this value is appended for each data point. The resulting data slices for all rows are combined and checked for any duplicates that can result from overlapping definitions in the configuration file. Lines 11 and 12 contain the calculation and evaluation of the actual validation check, according to the respective metrics. Finally, data can be either exported as CSV or returned as an R data.frame.

Algorithm 1 validateScenarios is the core of the piamValidation framework.

Input: scenarioData, referenceData, validationConfig

Output: data.frame or CSV file containing original data, reference values and and validation result

function `validateScenarios()`

1: data = read(scenarioData, referenceData)
2: cfg = read(validationConfig)
3: cfg = check, clean and expand cfg
4: **for all** row in cfg **do**
5:     slice data according to filters in row
6:     left join config thresholds to data slice
7:     left join reference values to data slice (if applicable)
8:     valiData = bind all data slices
9: **end for**
10: valiData = removeDuplicateRows(valiData)
11: calculate value to be checked against threshold (depending on metric)
12: evaluate check and append validation result (colour) to valiData
13: (optional: export valiData as CSV file)
14: return(valiData)

2.4 Output and Visualization

While the data output of `validateScenarios` can be used in any preferred way, two specific visualization approaches are offered to users of *piamValidation*. The function `validationHeatmap` has the goal of providing a quick overview of the general outcome of a validation by laying out the validation result colors across up to four dimensions. This means that the five-dimensional IAM data objects need to be reduced to a single element in only one dimension. By default, the "variable" dimension is chosen for this. An example can be seen in Fig. 3 where "region" and "period" are chosen for the x- and y-axis respectively and "scenario" and "model" are plotted as x- and y-facets.

Alternatively, the function `linePlotsThresholds` offers an easy way to plot the thresholds as colored background areas behind line plots of the scenario data and, if applicable, the corresponding reference data. For this type of plot one additional

https://doi.org/10.5194/egusphere-2025-2284
Preprint. Discussion started: 2 June 2025

data dimension needs to be singular, by default this falls to the "region" dimension.

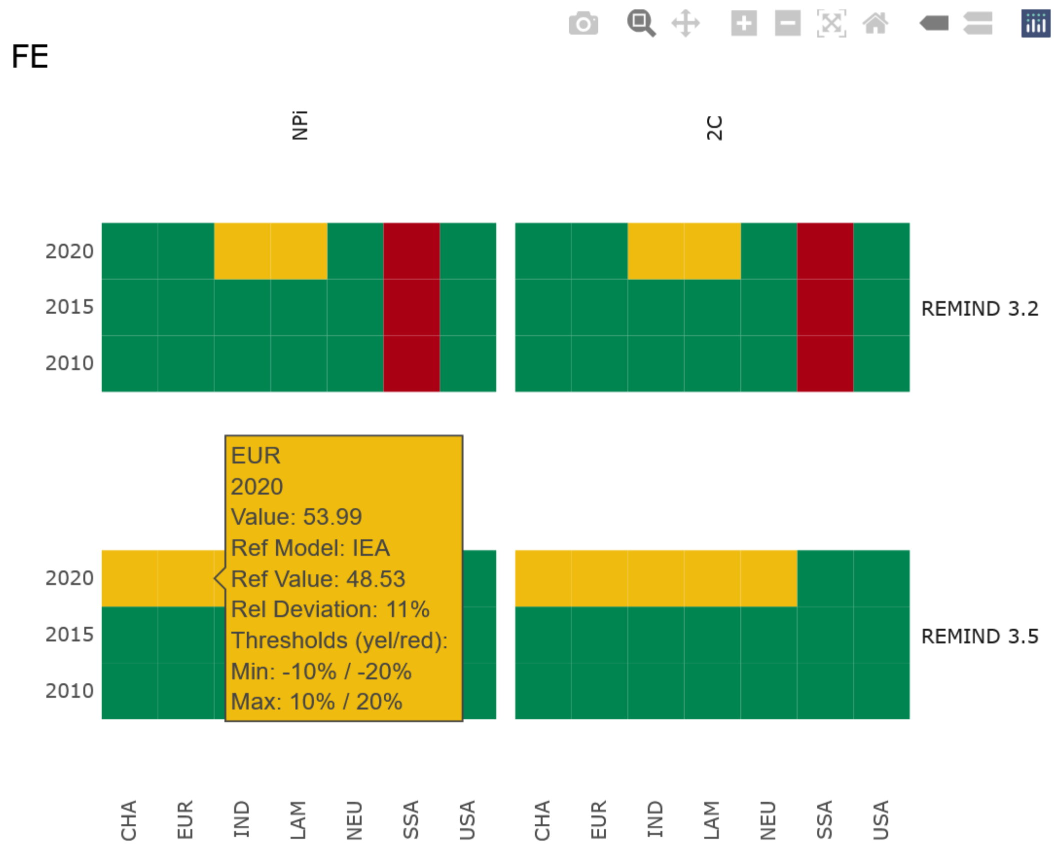

Figure 3. Example of an interactive validation heat map. Data along four dimensions. Hovering over a data point shows additional information in a tooltip. In this example, the Final Energy (FE) in two versions of the REMIND model for two different policy scenarios (National Policies implemented (NPi) and "well below" 2 degrees (66% likelihood) of temperature rise by the end of the century (2C) are compared with the historical IEA data.

The heat map plotting routine makes use of the *ggplot2* library using the `geom_tile` function in combination with *plotly* which enables interactive access to additional information by showing tooltips when hovering over a tile. This allows users to first get a quick overview of the general validation outcome by scanning the colors of the heat map and identifying areas of interest (e.g., a region with many red tiles). In a second step, the exact value of a data point as well as the corresponding thresholds and reference values, including their origin, are shown in the tooltip. This two-step process is a particular strength of the chosen visualization approach, as it makes validation results of large datasets accessible without overcrowding the plot or distributing the data over many different plots (line plots are usually not useful for visualizing data along four dimensions, often it is the "region" dimension which is spread over multiple plots).

https://doi.org/10.5194/egusphere-2025-2284
Preprint. Discussion started: 2 June 2025

The heat map tiles are chosen to always be squares, which has the side effect that the plot layout is dependent on the lengths of the data dimensions. In cases where a high number of scenarios or models is being validated, this can lead to gaps between heat map facets and inefficient use of space in the validation document. To alleviate this, `validationHeatmap` allows the user to manually specify the arrangement of the data dimensions by defining the function arguments "x_plot", "y_plot", "x_facet" and "y_facet". If they are not manually set, the function tries to find the optimal layout by analyzing the length of dimensions within the data passed to it.

2.5 Accessibility

One of the primary objectives of the *piamValidation* package is to enable broad application of the tool and support IAM capacity building for developers and users of scenarios. To achieve this, the tool is published under an open-source license on GitHub and designed with user-friendliness as a central design criterion. For basic usage, a validation process can be executed with a single command, ensuring accessibility and ease of use for a broad range of users.

For the first application, users can follow four straightforward steps: First, ensure a working R environment. If it is not already installed, one possibility could be to install R and the integrated development environment RStudio [4].

Second, install the *piamValidation* package by running the following R command:

```
install.packages("piamValidation", repo = "https://rse.pik-potsdam.de/r/packages")
```

The third steps consist of the data preparation. Verify that the IAM data that needs to be validated as well as the reference data follow the IAMC standard format. In addition, prepare the configuration file by defining the validation checks to be performed as described in Section 2.2.1 and Section 2.2.2.

In the final step, the following R command must be executed to generate the HTML validation report including interactive heat maps. This command consolidates the file paths for the IAM data to be validated, the reference data, and the appropriate configuration file into a single operation.

```
validationReport(c("path_to_IAM_data","path_to_ref_data"), "path_to_your_config")
```

For advanced users, the post-processing and visualization can be modified according to individual validation requirements by using the `R data.frame` containing the validation results directly. In addition, we encourage users to participate in ongoing discussions and package developments on GitHub:

https://github.com/pik-piam/piamValidation/issues.

3 Application case: Validate short-term technology trends

This section presents an application of the *piamValidation* tool to the Integrated Assessment Model REMIND (see Baumstark et al. (2021) for a detailed description). To illustrate the validation process, we validate model scenarios against historical

[4] Download the freeware here: R https://www.r-project.org/ and RStudio https://posit.co/products/open-source/rstudio/

https://doi.org/10.5194/egusphere-2025-2284
Preprint. Discussion started: 2 June 2025

data and observational or third-party technology trend data until 2030, focusing on three key technology trends: capacity of CO2 transport and storage (CTS), sales and stock of battery electric vehicle (BEV), and capacity of offshore wind power. These technology trajectories represent some of the key energy system transformations that Luderer et al. (2022) describe: the potential for carbon management, the electrification of end-use sectors, and the composition of the electricity mix.

In scenarios of the REMIND 3.2 release (Luderer et al., 2023), all three technologies exhibit significant deviations compared to historical and reference data. Figure 4 shows a heat map with yellow blocks for moderate deviations, red blocks for large upper deviations and blue blocks indicating large lower deviations. Upon identifying significant deviations, the focus shifts to improving the model and refining its underlying assumptions. The key strategies for achieving these improvements can be summarized as follows:

1. Updating the model input data;
2. Validating socio-economic assumptions to ensure alignment with observed trends;
3. Revising and verifying associated cost estimates;
4. Revising the representation of the inertia in up-scaling new technologies.

All adjustments have been incorporated into the REMIND 3.5 release (Luderer et al., 2025). We select two climate policy scenarios spanning a wide range of policy ambition: National Policies implemented (NPi) and "well below" 2 degrees (66% likelihood) of temperature rise by the end of the century (2C)[5]. Both scenarios follow the Shared Socioeconomic Pathways 2 called Middle of the Road. The 2°C stylized climate policy scenario assumes a peak budget of 1000 Gt CO2 on total CO2 emissions from 2020 to 2100. Figure 4 illustrates, how the model enhancements improve the near-term realism measured as performance against the reference data within the *piamValidation* tool.

More dynamic insights on model to trend deviation are presented in Figure 5. The key developments in the updated model are described below.

[5]The exact original scenario names are "SSP2EU-NPi" and "SSP2EU-PkBudg1150" in REMIND 3.2 and "SSP2-NPi2025" and "SSP2-PkBudg1000" in REMIND 3.5 for NPi and 2C respectively. They have been renamed for clarity and brevity.

https://doi.org/10.5194/egusphere-2025-2284
Preprint. Discussion started: 2 June 2025

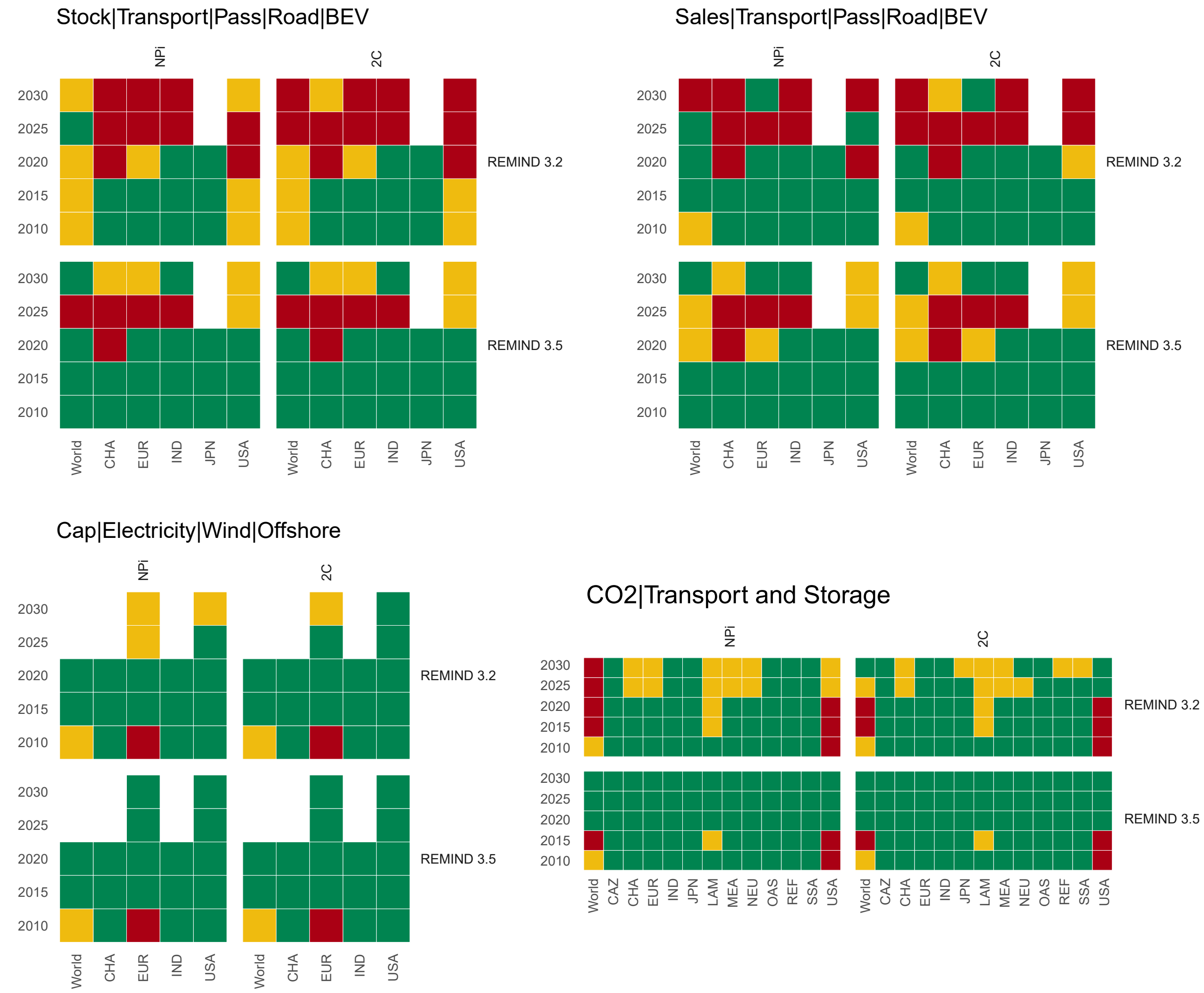

Figure 4. ***piamValidation*** **heat map of selected technology trends.** Each heat map presents the performance of REMIND 3.2 (top) compared to updated version REMIND 3.5 (bottom) for the two climate policy scenarios: National Policies implemented (NPi, left) and "well below" 2 degrees (2C, right). For BEV we validate historical modeling values against IEA Global Electric Vehicle Outlook (GEVO) Historical data. For the outlook, the modeling values are validated against the IEA GEVO Announced Pledges Scenario (APS) and Stated Policies Scenario (STEPS). Similarly, wind offshore capacity is historically compared to International Renewable Energy Agency (IRENA) data and the outlook validated Global Wind Energy Council (GWEC) data. Carbon management capacity is validated again IEA CCUS data. An overview of all thresholds is provided in the validation config file (https://github.com/pik-piam/piamValidation). In addition, a detailed description and community discussion on the thresholds they can be accessed here: https://github.com/pik-piam/mrremind/discussions/544.

https://doi.org/10.5194/egusphere-2025-2284
Preprint. Discussion started: 2 June 2025

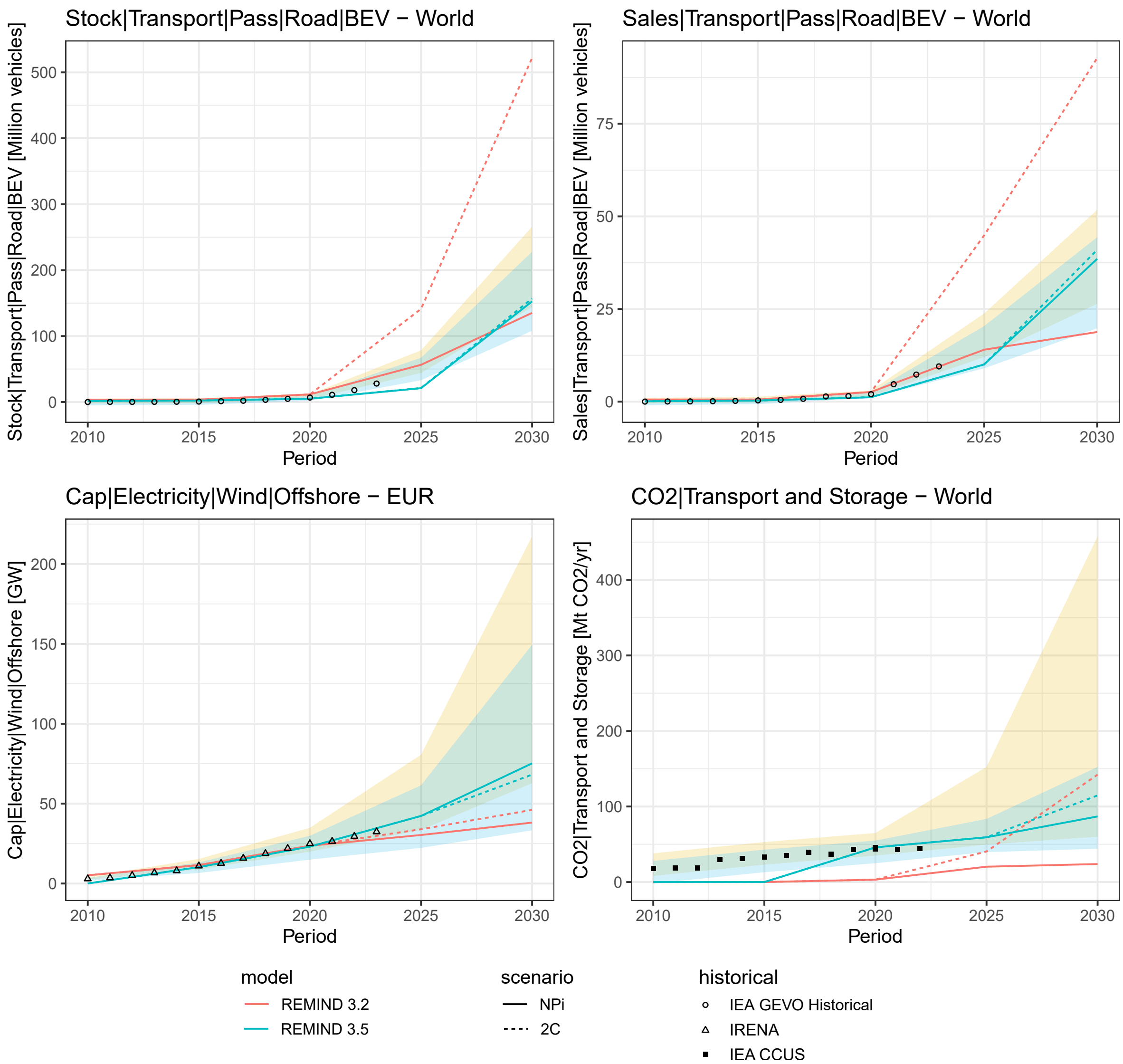

Figure 5. Historical and near-term realism. The REMIND scenarios and thresholds presented are identical with those in Figure 4. Scenarios updated refer to the improvements in REMIND 3.5.

3.1 CO2 transport and storage

CO2 transport and storage comprises the capture, transport and storage or use of CO2 from different CO2 sources. All CO2 that is captured, transported and stored is accounted in the variable *"CO2|Transport and Storage"*. The carbon transport and storage infrastructure is represented as a single technology in REMIND. The construction of geologic carbon storage infrastructure is associated with long lead times, as the development of geologic storage typically takes 7 to 10 years from exploration to industrial-scale injection (Bui et al., 2018). Upscaling may be further limited by the availability of geo-engineers and drilling

https://doi.org/10.5194/egusphere-2025-2284
Preprint. Discussion started: 2 June 2025

rigs (Budinis et al., 2018). In addition, a wide range of cost estimates have been documented. While studies on technical cost range between 3 and 20 USD/tCO2 (Budinis et al., 2018; Ipcc, 2014), cost announced for projects reach up to 80 USD/tCO2 (Bellona, 2022; Jakobsen et al., 2017; IOGP, 2023).

There are thus two trends relevant for model improvement. First, storage potentials until 2030 are limited by projects announced today due to the long lead times. We thus refer to the database on Carbon Capture, Utilization and Storage (CCUS) projects, published by the International Energy Agency (IEA) in 2023 that is also used for the validation reference values. To improve the near-term realism of geologic storage in the model, we introduce regional lower and upper bounds. The lower bounds in 2025 and 2030 are based on the announced capacities of storage projects that are either operational or under construction (with announced start of operation before 2025 and 2030 respectively). The upper bounds for 2025 and 2030 additionally include 40% of the capacity of all announced projects that report a start of operation by 2025 or 2030 respectively. Note that for these, the final investment decision is often pending, thus a certain share of unrealized project capacities is assumed based on historic delay and failure rates. Furthermore, as the EU27 pursues a transnational, Europe-wide CCS infrastructure, the capacity of a given CCS project is not fully attributed to the respective member state within the model. Instead, the EU27 CCS capacities are pooled (including 50% of Norway's and 50% of UKs near-term potential) and redistributed across member states based on GDP. The second improvement concerns the costs of carbon transport and storage. The technical cost estimate in REMINDv3.2.0 of 7.5 USD/tCO2 appears too low, despite REMIND modeling the cost of upscaling through adjustment cost. We thus increase our capital cost estimate to reach 12 USD/tCO2 informed by the average in Budinis et al. (2018).

3.2 Electric vehicle stock and sales

REMIND's transportation sector is modeled by the simulation model "Energy Demand Generator-Transport" (EDGE-T) (Rottoli et al., 2021) that is integrated into the broader sociotechnical framework of REMIND. EDGE-T simulates technology and transport mode decisions relying on a multinomial logit approach. The decision tree, the logit, comprises a nested structure of subsequent choices among comparable alternatives for modes and technologies. This established approach is handy for the modeling of end-consumer decisions under probabilistic conditions and allows for including non-economic factors, such as inconvenience costs, in the decision simulation.

In the transport sector, a prominent example is the increase in electric vehicles (EVs), which is not mainly driven by CO2 prices and, consequently, fuel costs, but rather by technical innovations leading to up-front investment cost reductions, policy incentives, and shifts in how consumers assess this technology alternative. Model parameters were thus updated to reflect technology advancement and better understanding of the uptake dynamics of electric vehicles. Building on the Global EV Outlook (IEA, 2024) and its information about the shares of electric vehicles in car sales and car fleet in different world regions, the model representation of consumer attitudes (so-called inconvenience costs) towards electromobility was updated. Furthermore, input data processing was critically assessed, e.g. by screening for implausible outliers in input data, and correction routines were implemented. Given the scarcity of high-quality transport data such as total energy service demand in passenger-kilometers and ton-kilometers per transport mode, size distribution of the vehicle fleet, annual mileage of the vehicle

https://doi.org/10.5194/egusphere-2025-2284
Preprint. Discussion started: 2 June 2025

fleet, investment and operation cost for vehicles for most countries of the world, checking input data plausibility by comparing data across countries, technologies and vehicle sizes is a crucial step for improving model and scenario quality.

3.3 Wind offshore capacity

The competition between onshore and offshore wind power is largely determined by factors like grid integration costs or the matching between generation profiles and demand. REMIND 3.2 – like other long-term global integrated assessment models – represents these drivers in a very aggregated and parameterized way, since it does not explicitly resolve the relevant spatial and temporal scales. Consequently, the model limits the share of offshore wind generation in total wind energy to a certain range. This share is constrained in the near-term to reflect the historical lag between onshore and offshore deployment.

In REMIND 3.2, historical capacities are available until 2020, at which point wind offshore is still a nascent technology with around 35GW of installed capacity world-wide. To prevent the model from deploying offshore earlier than observed in the real-world, only a portion of the total offshore wind potential is available: before 2010, none of the potential is available, then the available portion increases gradually until 2050. Conversely, specific equations ensure that the development of offshore capacity is not too slow compared to onshore after 2025: in a given region, if the available offshore potential is four times as high as onshore, the equations constrain the model to deploy each year at least twice as much offshore as onshore new capacity.

The *piamValidation* tool has revealed that these constraints were overly pessimistic: wind offshore deployment in REMIND 3.2 is slower than observed in real data, particularly in Europe. To address this discrepancy, we updated the model to match the real data more closely. REMIND 3.5 uses newly released data by IRENA (2024) to compute consolidated historical capacities for 2020 and a lower bound for 2025 based on 2023 capacity. Moreover, the model allows for an accelerated phase-in of wind offshore by making the full potential available by 2030 instead of 2050. The equations linking offshore to onshore are maintained after 2030, in order to follow the rising trend of wind offshore despite its higher investment costs.

4 Conclusions

Integrated Assessment Models are central to the climate negotiations and play a key role in the IPCC assessments and thus at the interface between climate science and policy. However, they face challenges and criticism regarding their transparency (Robertson, 2021), their ability to capture technology trends (Keppo et al., 2021; Anderson, 2019) and their overall reliability (Wilson et al., 2021).

We introduce the open-source R package *piamValidation*, a versatile tool designed to address multiple use cases in the validation of IAM scenario data. This package facilitates validation across IAM scenarios, models, and temporal periods. Furthermore, it supports validation against historical, observational, and benchmark datasets, using time series, thresholds, or growth rates as a reference. To enhance reliability, *piamValidation* can be seamlessly integrated into the regular development workflow of IAMs, enabling comprehensive sanity checks and performance evaluations.

We outline the structure of the *piamValidation* package, emphasizing its user-friendly design to facilitate widespread adoption within the AIM community. The aim is to enhance the overall validity and reliability of IAMs through the systematic

https://doi.org/10.5194/egusphere-2025-2284
Preprint. Discussion started: 2 June 2025

application of the tool. To support this objective, we provide a comprehensive and structured guideline, along with a standard use case demonstrating the application of the *piamValidation* tool through a single-command execution. We demonstrate the application of the *piamValidation* tool to assess the near-term realism of technology trends in the integrated assessment model REMIND. This analysis spans a policy range from nationally implemented policies to scenarios aligned with a well-below 2°C target. Our case study shows how near-term realism has been improved for wind offshore capacity, carbon management and electric cars in REMIND by using the *piamValidation* tool to first spot the deviations to historical and outlook data and then demonstrate performance improvements. Our application case reveals the application value of the tool for historical and benchmark validation critically depends on the validation settings and availability of reliable reference data. Consequently, the utility of the *piamValidation* tool would be enhanced by broader adoption and the systematic documentation of configuration files within the IAM community.

The primary benefits of the *piamValidation* tool lies in its ability to efficiently provide an overview of variable deviations within large-scale datasets. It also facilitates systematic validation across IAM scenarios and supports sanity checks throughout model development processes. Furthermore, the tool serves as a potential platform for the IAM community to collaboratively build a repository of knowledge on reasonable validation thresholds.

The *piamValidation* tool is subject to several limitations. First, the identification of meaningful validation cases, reference data and thresholds is the responsibility of the user and substantially influences the results. Special attention is required in selecting the reference data regarding the projection method as well as technology resolution of the model (Krey et al., 2019). In other words, the usefulness of the validation test depends on the quality of the reference data. Since the tool identifies discrepancies between scenario and reference data, the user always needs to question the sources and reliability of the reference data as well as its meaning. Finally, the *piamValidation* tool is a performance measure that may motivate IAM improvements, but implementing them remains an additional modeling effort, as demonstrated in the REMIND application case.

Ongoing improvements are being made to the *piamValidation* tool, with continuous development efforts focused on enhancing its capabilities, including three central development features. First, within our application case, we used relative as well as absolute metrics to correct for scale-dependence of percentage deviation, especially for percentage deviations near zero. To reduce this scale dependency and increase the user-friendliness of the *piamValidation* tool, we are working on a new feature that includes a general tolerance. Our plan is that within the processing we include an optional lump sum buffer as a fraction of the largest reference value. Second, we would like to include a feature that enables the *piamValidation* tool to differentiate between different validation dimensions, for instance, feasibility, sustainability or equity. A third key development area centers around the identification ability of high-quality scenarios from an extensive scenario database. These improvements are designed to increase the versatility and usefulness of the tool within the IAM community.

Code availability. The REMIND-MAgPIE code is implemented in GAMS, whereas code and data management is done using R. The REMIND 3.2 code is archived at Zenodo (https://zenodo.org/records/7852740) Luderer et al. (2023) and REMIND 3.5 at (https://zenodo.org/records/7852740). The technical model documentation is available on the common integrated assessment model documentation website

https://doi.org/10.5194/egusphere-2025-2284
Preprint. Discussion started: 2 June 2025

(https://www.iamcdocumentation.eu/Model_Documentation_-_REMIND-MAgPIE), and is published as open-source code (https://github.com/remindmodel/remind.). A repository for the source code of the *piamValidation* package is available via GitHub at https://github.com/pik-piam/piamValidation, the reproduction of the plots in this publication can be achieved with the RMarkdown file at https://github.com/pik-piam/piamValidation/blob/main/inst/markdown/validationReport_publication.Rmd. The validation in this study is performed with the configuration file at https://github.com/pik-piam/piamValidation/blob/main/inst/config/validationConfig_publication.csv.

Author contributions. GL, LB and EK supervised the development of the *piamValidation* package. The *piamValidation* package is initiated and led by PW, where FL and OR contributed to its development. FL, AN, TD, JH, JM, RP, OR, LV, NB, FB, CC, RM and GL contributed to the REMIND model development. All authors commented, reviewed and contributed to the final manuscript.

Competing interests. One of the authors, Gunnar Luderer, serves as a Topic Editor for this journal. All other authors declare no competing interests.

Acknowledgements. The authors are grateful to the attendees of the Seventeenth Assessment Modeling Consortium (IAMC) Annual Meeting 2024 for their valuable feedback and reflections. This work received support from the European Commission Directorate-General of Climate Action (DG CLIMA) under Service Contract No. 14020241/2022/884157/SER/CLIMA.A.2 CLIMA/2022/EA-RP/0007 and the Horizon Europe Project PRISMA, under the grant agreement number 1010816.

https://doi.org/10.5194/egusphere-2025-2284
Preprint. Discussion started: 2 June 2025

References

Anderson, K.: Wrong tool for the job: Debating the bedrock of climate-change mitigation scenarios, Nature, 573, 348–348, 2019.

Baumstark, L., Bauer, N., Benke, F., Bertram, C., Bi, S., Gong, C. C., Dietrich, J. P., Dirnaichner, A., Giannousakis, A., Hilaire, J., et al.: REMIND2. 1: transformation and innovation dynamics of the energy-economic system within climate and sustainability limits, Geoscientific Model Development Discussions, 2021, 1–50, 2021.

Beck, S. and Oomen, J.: Imagining the corridor of climate mitigation–What is at stake in IPCC's politics of anticipation?, Environmental Science & Policy, 123, 169–178, 2021.

Bellona: Norway's Longship CCS project, https://network.bellona.org/content/uploads/sites/3/2020/10/Longship-Briefing_Bellona-1.pdf, 2022.

Bertram, C., Brutschin, E., Drouet, L., Luderer, G., van Ruijven, B., Aleluia Reis, L., Baptista, L. B., de Boer, H.-S., Cui, R., Daioglou, V., et al.: Feasibility of peak temperature targets in light of institutional constraints, Nature Climate Change, 14, 954–960, 2024.

Brutschin, E., Pianta, S., Tavoni, M., Riahi, K., Bosetti, V., Marangoni, G., and Van Ruijven, B. J.: A multidimensional feasibility evaluation of low-carbon scenarios, Environmental Research Letters, 16, 064 069, 2021.

Budinis, S., Krevor, S., Mac Dowell, N., Brandon, N., and Hawkes, A.: An assessment of CCS costs, barriers and potential, Energy strategy reviews, 22, 61–81, 2018.

Bui, M., Adjiman, C. S., Bardow, A., Anthony, E. J., Boston, A., Brown, S., Fennell, P. S., Fuss, S., Galindo, A., Hackett, L. A., et al.: Carbon capture and storage (CCS): the way forward, Energy & Environmental Science, 11, 1062–1176, 2018.

Carbon Brief: Just 15 countries account for 98% of new coal-power development, https://www.carbonbrief.org/guest-post-just-15-countries-account-for-98-of-new-coal-power-development/, accessed: 10/10/2024, 2024.

Cointe, B., Cassen, C., and Nadaï, A.: Organising policy-relevant knowledge for climate action: integrated assessment modelling, the IPCC, and the emergence of a collective expertise on socioeconomic emission scenarios, Science & Technology Studies, 2019.

Creutzig, F., Hilaire, J., Nemet, G., Müller-Hansen, F., and Minx, J. C.: Technological innovation enables low cost climate change mitigation, Energy Research & Social Science, 105, 103 276, 2023.

Edenhofer, O. and Minx, J.: Mapmakers and navigators, facts and values, Science, 345, 37–38, 2014.

Giannousakis, A., Hilaire, J., Nemet, G. F., Luderer, G., Pietzcker, R. C., Rodrigues, R., Baumstark, L., and Kriegler, E.: How uncertainty in technology costs and carbon dioxide removal availability affect climate mitigation pathways, Energy, 216, 119 253, 2021.

IEA: Global EV Outlook 2024, https://www.iea.org/reports/global-ev-outlook-2024, licence: CC BY 4.0, 2024.

IOGP: Creating a sustainable business case for CCS value chains – the needed funding and de-risking mechanisms, https://iogpeurope.org/wp-content/uploads/2023/11/Creating-a-Business-Case-for-CCS-Value-Chains-IOGP-Europe.pdf, 2023.

Ipcc, A.: Ipcc fifth assessment report—synthesis report, IPPC Rome, Italy, 2014.

IRENA: Renewable energy highlights, https://www.irena.org/-/media/Files/IRENA/Agency/Publication/2023/Jul/Renewable_energy_highlights_July_2023.pdf, accessed: 10/10/2024, 2024.

IRENA: Renewable capacity statistics, Tech. rep., International Renewable Energy Agency, 2024.

Jakobsen, J., Roussanaly, S., and Anantharaman, R.: A techno-economic case study of CO2 capture, transport and storage chain from a cement plant in Norway, Journal of cleaner production, 144, 523–539, 2017.

https://doi.org/10.5194/egusphere-2025-2284
Preprint. Discussion started: 2 June 2025

Keppo, I., Butnar, I., Bauer, N., Caspani, M., Edelenbosch, O., Emmerling, J., Fragkos, P., Guivarch, C., Harmsen, M., Lefevre, J., et al.: Exploring the possibility space: Taking stock of the diverse capabilities and gaps in integrated assessment models, Environmental Research Letters, 16, 053 006, 2021.

Krey, V., Guo, F., Kolp, P., Zhou, W., Schaeffer, R., Awasthy, A., Bertram, C., de Boer, H.-S., Fragkos, P., Fujimori, S., et al.: Looking under the hood: a comparison of techno-economic assumptions across national and global integrated assessment models, Energy, 172, 1254–1267, 2019.

Luderer, G., Madeddu, S., Merfort, L., Ueckerdt, F., Pehl, M., Pietzcker, R., Rottoli, M., Schreyer, F., Bauer, N., Baumstark, L., et al.: Impact of declining renewable energy costs on electrification in low-emission scenarios, Nature Energy, 7, 32–42, 2022.

Luderer, G., Bauer, N., Baumstark, L., Bertram, C., Leimbach, M., Pietzcker, R., Strefler, J., Aboumahboub, T., Abrahão, G., Auer, C., Benke, F., Bi, S., Dietrich, J., Dirnaichner, A., Fuchs, S., Führlich, P., Giannousakis, A., Gong, C. C., Haller, M., Hasse, R., Hilaire, J., Hoppe, J., Klein, D., Koch, J., Körner, A., Kowalczyk, K., Kriegler, E., Levesque, A., Lorenz, A., Ludig, S., Lüken, M., Malik, A., Mandaroux, R., Manger, S., Merfort, A., Merfort, L., Moreno-Leiva, S., Mouratiadou, I., Odenweller, A., Pehl, M., Pflüger, M., Piontek, F., Popin, L., Rauner, S., Richters, O., Rodrigues, R., Roming, N., Rottoli, M., Schmidt, E., Schötz, C., Schreyer, F., Schultes, A., Sörgel, B., Ueckerdt, F., Verpoort, P., and Weigmann, P.: REMIND - REgional Model of INvestments and Development, https://doi.org/10.5281/ZENODO.7852740, 2023.

Luderer, G., Bauer, N., Baumstark, L., Bertram, C., Leimbach, M., Pietzcker, R., Strefler, J., Aboumahboub, T., Abrahão, G., Auer, C., Bantje, D., Beier, F., Benke, F., Bi, S., Dietrich, J. P., Dirnaichner, A., Dorndorf, T., Duerrwaechter, J., Fuchs, S., Führlich, P., Giannousakis, A., Gong, C. C., Hagen, A., Haller, M., Hasse, R., Hayez, L., Hilaire, J., Hofbauer, V., Hoppe, J., Klein, D., Koch, J., Köhler-Schindler, L., Körner, A., Kowalczyk, K., Kriegler, E., Lécuyer, F., Levesque, A., Lorenz, A., Ludig, S., Lüken, M., Malik, A., Mandaroux, R., Manger, S., Merfort, A., Merfort, L., Moreno-Leiva, S., Mouratiadou, I., Müßel, J., Odenweller, A., Pehl, M., Pflüger, M., Piontek, F., Popin, L., Rauner, S., Richters, O., Rodrigues, R., Roming, N., Rosemann, R., Rottoli, M., Rüter, T., Salzwedel, R., Sauer, P., Schmidt, E., Schötz, C., Schreyer, F., Schultes, A., Sitarz, J., Sörgel, B., Ueckerdt, F., Verpoort, P., Weigmann, P., and Weiss, B.: REMIND - REgional Model of INvestments and Development, https://doi.org/10.5281/ZENODO.15147820, 2025.

Robertson, S.: Transparency, trust, and integrated assessment models: An ethical consideration for the Intergovernmental Panel on Climate Change, Wiley Interdisciplinary Reviews: Climate Change, 12, e679, 2021.

Rottoli, M., Dirnaichner, A., Kyle, P., Baumstark, L., Pietzcker, R., and Luderer, G.: Coupling a detailed transport model to the integrated assessment model REMIND, Environmental Modeling & Assessment, 26, 891–909, 2021.

Stanton, E. A., Ackerman, F., and Kartha, S.: Inside the integrated assessment models: Four issues in climate economics, Climate and Development, 1, 166–184, 2009.

Van Beek, L., Hajer, M., Pelzer, P., van Vuuren, D., and Cassen, C.: Anticipating futures through models: the rise of Integrated Assessment Modelling in the climate science-policy interface since 1970, Global Environmental Change, 65, 102 191, 2020.

Wilson, C., Guivarch, C., Kriegler, E., Van Ruijven, B., Van Vuuren, D. P., Krey, V., Schwanitz, V. J., and Thompson, E. L.: Evaluating process-based integrated assessment models of climate change mitigation, Climatic Change, 166, 1–22, 2021.