# A Methodology and Implementation of Automated Emissions Harmonization for Use in Integrated Assessment Models

Matthew J. Gidden[a,*], Shinichiro Fujimori[b], Maarten van den Berg[c], David Klein[d], Steven J. Smith[e], Detlef P. van Vuuren[c], Keywan Riahi[a]

[a] *International Institute for Applied Systems Analysis, Schlossplatz 1, A-2361 Laxenburg, Austria*
[b] *Center for Social and Environmental Systems Research, National Institute for Environmental Studies, 16-2 Onogawa, Tsukuba, Ibaraki 305-8506, Japan*
[c] *PBL Netherlands Environmental Assessment Agency, Postbus 30314, 2500 GH The Hague, Netherlands*
[d] *Potsdam Institute for Climate Impact Research (PIK), Member of the Leibniz Association, P.O. Box 60 12 03, D-14412 Potsdam, Germany*
[e] *Joint Global Change Research Institute, 5825 University Research Court, Suite 3500, College Park, MD 20740*

**Abstract**

Emissions harmonization refers to the process used to match greenhouse gas (GHG) and air pollutant results from Integrated Assessment Models (IAMs) against a common source of historical emissions. To date, harmonization has been performed separately by individual modeling teams. For the hand-over of emission data for the Shared Socio-economic Pathways (SSPs) to climate model groups, a new automated approach based on commonly agreed upon algorithms was developed. This work describes the novel methodology for determining such harmonization methods and an open-source Python software library implementing the methodology. Results are shown for two example scenarios (with and without climate policy cases) using the MESSAGE-GLOBIOM IAM that satisfactorily harmonize over 96% of the total emissions trajectories while having a negligible effect on key long-term climate indicators. This new capability enhances the comparability across different models, increases transparency and robustness of results, and allows other teams to easily participate in intercomparison exercises

*Corresponding author
*Email address:* gidden@iiasa.ac.at (Matthew J. Gidden)

by using the same, openly available harmonization mechanism.

*Keywords:* Integrated Assessment Models, Climate Change, Harmonization, Air Pollution

---

**Software Availability**

`aneris`, first made available in 2017, is available online at `https://github.com/gidden/aneris` as a free and open-source Python software library (approximately 2000 lines of code). The `aneris` software was developed by the lead author, Matthew J. Gidden, Ph.D., whose contact information is shown on the title page of this manuscript. Documentation for the `aneris` Python package, including software requirements, is available online.

## Introduction

Integrated Assessment Models (IAMs) are tools used to understand the complex interactions between energy, economy, land use, water, and climate systems. IAMs provide global projections of systemic change by dividing the world into a number of representative regions (typically 10 to 30), the definition of which is distinct for each model [1]. Results from IAMs are integral in a number of international studies, which notably include projections of climate and economic futures. Recently, the IAM community has developed scenarios based on the Shared Socio-economic Pathways (SSPs) [2, 3, 4, 5, 6] which quantify a variety of potential global futures. The SSPs are designed to be used in research that include earth system model (ESM) simulations, climate impact, adaptation and climate mitigation studies [7].

While IAMs are implemented in myriad ways, including simulation and optimization, the core inputs and outputs are similar across different models. Modeling teams incorporate data on energy systems, land use, economics, demographics and emissions sources and concentrations, among other data, in order to provide consistent existing trajectories of modeled variables. The models then provide estimates of future trajectories of these variables under various socio-economic and technological assumptions as well as proposed policy constraints, e.g., targets for future Greenhouse Gas (GHG) emissions.

The emissions trajectories calculated by IAMs are critical inputs for ongoing, worldwide scientific community efforts in the Coupled Model Intercomparison Project Phase 6 (CMIP6) [8], which is utilizing a number of marker SSP scenarios developed by the IAM community (ScenarioMIP) [9]. These trajectories are endogenously calculated by modeling the individual technologies and sectors that contribute towards the emissions of different air pollutants and GHGs as well as various mitigation technologies. However, the historical emissions starting points of models can differ by large amounts depending on the region, sector, and emissions species.

In practice, IAMs calculate the total source intensity of emitting technologies,

for example the total activity of coal power plants in China, and incorporate emissions-intensity factors for individual gas species, for example the quantity of sulfur emissions from coal plants per megawatt-hour of production. Models are generally *calibrated* to historical data sources in one or more base years. Results in the historical period may differ between models as a result of the sometimes large uncertainties in historical data sets. Models can also differ in their choice of base-year, which may lag behind available inventory data. In addition, models have varying sectoral, regional, and fuel aggregations.

The global climate change community has recently developed a new global historical emissions data set for both anthropogenic emissions (i.e., the Community Emissions Data System (CEDS) [10] and open-burning land-use and land-use change (LULUC) emissions [11]) which, in conjunction with the SSP IAM trajectories, will be used for climate-related modeling exercises of CMIP6.

When participating in intercomparison exercises in which a consistent historical starting point is required (e.g., in CMIP6), model teams incorporate a single, common historical data set through *harmonization*. Harmonization refers to the process of adjusting model results to match a selected historical time series such that the resulting future trajectories are also consistent with the original modeled results. In the emissions context, this means that each individual combination of model region, model sector, and emissions species must be harmonized. Depending on the total number of model regions, sectors, and emissions species, this can require the selection of thousands to tens-of-thousands of harmonization methods.

Harmonization has been addressed in previous studies as it is a common practice in the IAM and climate change communities. For example, [12] describes the use of scaling routines for the 5 regions used in the Special Report on Emissions Scenarios (SRES) [13]; however, only total emissions were harmonized in the exercise, thus there is no sectoral dimension. Further, [14] describes the impacts of choosing various harmonization routines on future trajectories. During the evaluation of the Representative Concentration Pathways (RCPs), IAM results have been harmonized by sector and the 5 RCP global regions [15].

Importantly, the choice of harmonization method to date has been determined by individual experts and has generally been applied to all trajectories for a given class of emissions species.

Climate modeling efforts have continued to progress, demanding increased spatial and sectoral resolution from IAMs. Furthermore, a new generation of climate scenarios which combines aspects of both the RCPs and SSPs have been developed in order to incorporate both physical and socio-economic detail. In order to address the growing dimensionality of model outputs and support ongoing scenario generation and analysis efforts while still providing a consistent and scientifically rigorous harmonization procedure, an automated process for determining harmonization methods is preferred. The use of an automated, documented, and openly available harmonization mechanism additionally allows for full procedural reproducibility and for direct participation by additional modeling teams not involved in the original exercise.

The remainder of this paper describes the methodology and implementation of the harmonization software `aneris` [16], written in the Python programming language (detailed documentation available online). Section 2 provides a detailed description of the underlying mathematical components of `aneris` as well as the procedural workflow. The results of applying the automated harmonization mechanism on two example IAM scenarios, one with emissions growth and another with emissions mitigation, is presented in Section 3. Finally, the general effectiveness and potential future improvements on the automated methodology is discussed in Section 4.

## Methodology & Implementation

*Harmonization Methods*

IAM emission results are provided along temporal (normally half decade or decade), spatial (i.e., model regions), gas species, and sectoral dimensions. Each individual temporal trajectory, i.e., unique combinations of regions ($r$), species ($g$), and sectors ($s$), must be harmonized to the initial modeling period. Given a model trajectory, $m_{r,g,s}(t)$, historical values, $h_{r,g,s}(t)$, and model base year, $t_i$, a harmonized trajectory needs to be calculated. The *harmonization quality* of a trajectory, i.e., how well a given harmonization algorithm performs, depends on a number of factors. Of chief import is the faithful representation of the original, unharmonized, trajectory as well as the representation of negative trajectories (i.e., if a trajectory becomes negative, both the timing and total magnitude should be as close as possible) which are of critical importance for cumulative $CO_2$ calculations.

In previous studies [12, 14], two *families* of methods have been used: those that operate on the ratio of base year values (i.e., $\frac{h(t_i)}{m(t_i)}$) and those that operate on the offset of base year values (i.e., $h(t_i) - m(t_i)$). A number of the classic methods are implemented in `aneris` including ratio-convergence shown in Equation 2, offset-convergence shown in Equation 3, and interpolation shown in Equation 4. The convergence factor, $\beta$, scales linearly from 1 to 0 over $[t_i, t_f)$ and is shown in Equation 1. In all equations the region, species, and sector indices have been dropped for clarity. Each equation is a function of time, model trajectory, historical trajectory, base year ($t_i$), and a convergence year ($t_f$), at which point the harmonized trajectory converges to the unharmonized trajectory. `aneris` provides a number of methods to choose from for each of the harmonization families. A summary of all available methods is provided in Table 1.

$$\beta(t, t_i, t_f) = \begin{cases} 1 - \frac{t - t_i}{t_f - t_i}, & \text{if } t < t_f \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

$$m^{rat}(t, m, h, t_i, t_f) = [\beta(t, t_i, t_f)(\frac{h(t_i)}{m(t_i)} - 1) + 1]m(t) \tag{2}$$

$$m^{off}(t, m, h, t_i, t_f) = \beta(t, t_i, t_f)(h(t_i) - m(t_i)) + m(t) \tag{3}$$

$$m^{int}(t, m, h, t_i, t_f) = \begin{cases} \beta(t, t_i, t_f)(h(t_i) - m(t_f)) + m(t_f), & \text{if } t < t_f \\ m(t), & \text{otherwise} \end{cases} \tag{4}$$

Table 1: All Harmonization Methods Provided in `aneris`

| Method Name | Harmonization Family | Convergence Year |
|---|---|---|
| `constant_ratio` | ratio | $t_f = \infty$ |
| `reduce_ratio_<year>` | ratio | $t_f =$`<year>` |
| `constant_offset` | offset | $t_f = \infty$ |
| `reduce_offset_<year>` | offset | $t_f =$`<year>` |
| `linear_interpolate_<year>` | interpolation | $t_f =$`<year>` |

*Default Method Decision Tree*

A *decision tree* approach has been implemented in `aneris` which provides a systematic and documented decision-making process to determine the preferred harmonization algorithm. In order to provide reasonable *default* methods, the historical trajectory, unharmonized model trajectory, and relative difference between history and model values in the harmonization year are analyzed. The decision tree used in this analysis is a result of collaborative efforts between IAM teams and is shown graphically in Figure 1.

A number of characteristics impact the decision of which default method to select based on the effect of the characteristic on the potential harmonized trajectory. For example, it is possible for models to report zero values in the harmonization year in situations in which technologies are introduced in future

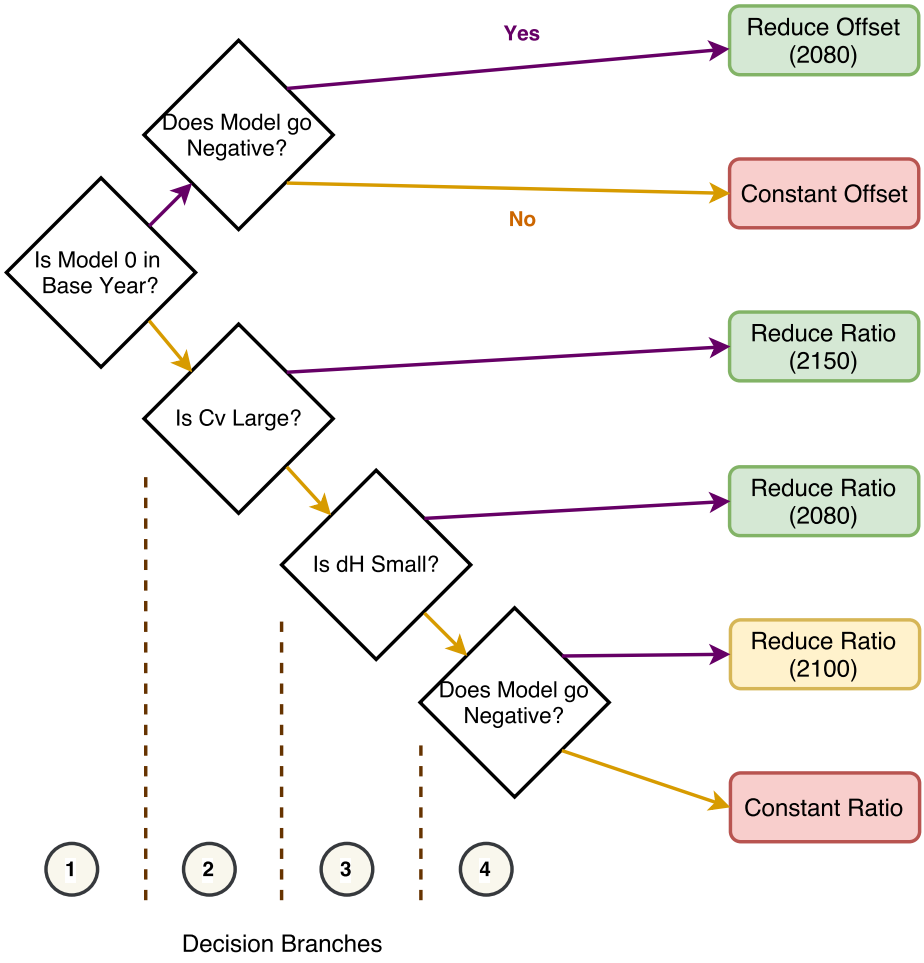


Figure 1: The default method decision tree used in the `aneris` software library. For all decisions, upper (purple) branches represent a "yes" response and lower (orange) branches represent a "no" response. The coefficient of variation, $c_v$, is defined in Equation 5, $dH$ is defined as $\left|\frac{h(t_i)-m(t_i)}{h(t_i)}\right|$, and decision-making thresholds for $c_v$ and $dH$ are described below. Methods labeled in *green* are likely to closely match unharmonized results, methods in *yellow* will likely somewhat match unharmonized results, and methods in *red* can be expected to have a large relative difference between harmonized and unharmonized results.

time periods in regions or for sectors which produce an emissions species that is absent in the initial modeling period. In such cases, an offset method is required as a ratio method would mask future emissions and erroneously harmonize the trajectory.

In most cases, however, models do report values in the harmonization year. Figure 2 displays a number of example trajectories which highlight the possible issues resulting from harmonizing model results in different contexts. When model and historical values are relatively close, a convergence method is chosen in order to be as representative as possible to the underlying unharmonized model results (Figure 2, Panel `a`). If values are not close, the constant ratio method is chosen in order to provide reasonable trajectories that still incorporate modeled effects (Figure 2, Panel `b`).

Models can additionally report negative emissions in certain contexts which must be taken into account during harmonization. Such a case is possible for gas species which can be extracted from the environment and stored, as is the case for $CO_2$ in future scenarios with climate mitigation. If a model provides a trajectory that transitions from positive to negative emissions and base year results are similar, then a convergence method is used in order to guarantee capture of this transition in a representative fashion (Figure 2, Panel `c`). If the discrepancy in base year results is large, it is possible for a negative trajectory to be inappropriately harmonized to a positive, but decreasing, trajectory. As such, the constant ratio method is chosen (Figure 2, Panel `d`).

Temporal variability of the historical trajectory is also an important characteristic when considering the choice of harmonization method. Emissions from forest and grassland fires, for example, vary from year to year due to a combination of meteorological conditions and anthropogenic drivers. Land use emissions in many IAMs are modeled using average emission factors and do not capture conditions in a specific year. A longer convergence horizon is thus desired in order to incorporate highly variable historical data with modeled results as is consistency in harmonization method because the effects are modeled similarly across regions and species. In order to detect emissions with a high amount of

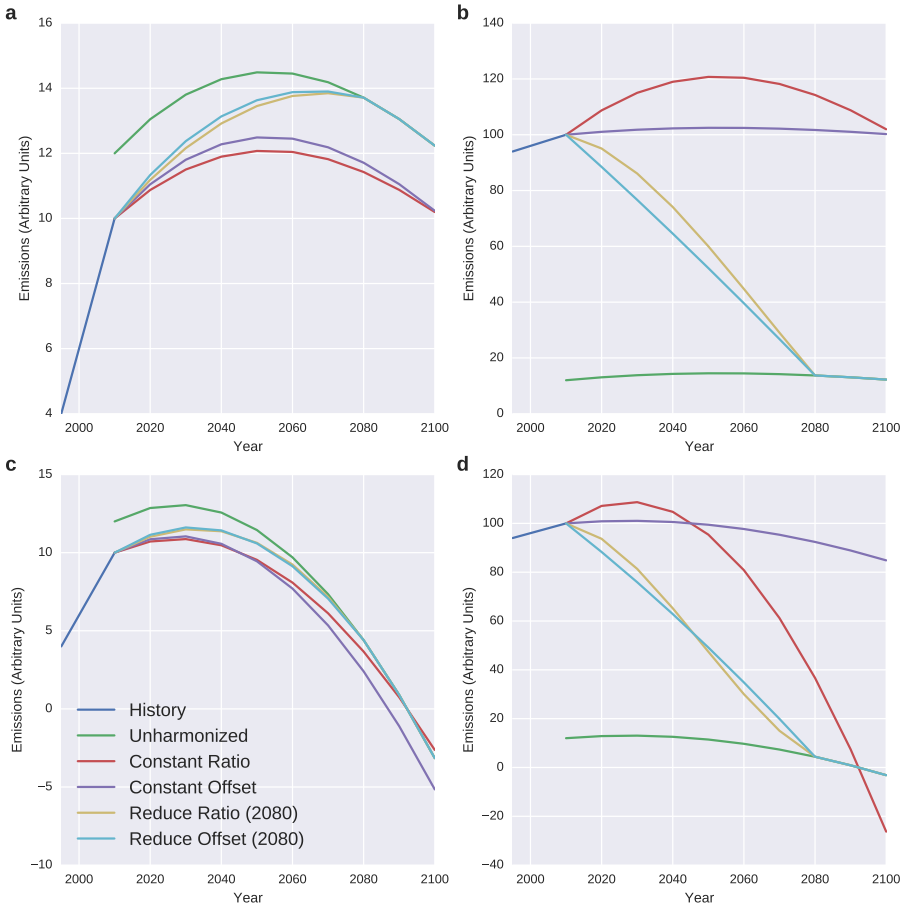


Figure 2: The effect of different harmonization routines on model trajectories under "normal" circumstances (Panel `a`), when there is a large difference between historical and model values in the harmonization year (Panels `b` and `d`), and when model trajectories result in negative emissions by the end of the modeling time horizon (Panels `c` and `d`). Identical model trajectories are used in each row (Panels `a`, `b`; `c`, `d`). In each column, historical values are increased in the base year by an order of magnitude (from 10 to 100). In each Panel, a subset of the potential routines provide a better harmonization quality than others as described in the text.

variation, a measure of the coefficient of variation, $c_v$, of the first derivative of the historical trajectory is calculated using the standard deviation, $\sigma$, and the mean, $\mu$, as shown in Equation 5.

The value of $c_v$ is then tested against a threshold, $\tau_{c_v}$. To determine this threshold, an analysis of the recent CEDS and LUC historical data has been performed. Figure 3 shows the distribution of LUC $c_v$s and non-LUC $c_v$s as determined for historical data aggregated to the model regions of 5 different IAMs involved in the SSP process. A threshold value of $\tau_{c_v} = 20$ has been chosen based on these observations as it optimally divides the two distributions. Importantly, tails of the LUC and non-LUC overlap, thus there are both false positives ( 7% of non-LUC trajectories) and false negatives ( 10% of LUC trajectories). However, as any regional definition is model dependent and thus any regional aggregation is possible an automated detection mechanism is necessary.

$$c_v = \frac{\sigma(h'(t))}{\mu(h'(t))} \tag{5}$$

Finally, consideration is taken with respect to the relative difference between the historic and model values in the harmonization time period. In order to investigate the possible values that these relative differences can take, the IAM values used in the SSP and (ongoing) CMIP6 inter-comparison exercises are used. A distribution of these differences for all models in the study is presented in Figure 4. Given the available data, a threshold value of $\tau_{dH} = 50\%$ was chosen to be used as a default in `aneris`.

### *`aneris` Workflow*

The full `aneris` workflow is comprised of a number of components shown graphically in Figure 5. Unharmonized model data and a run-control configuration are read in via an Excel spreadsheet. Data is assumed to be in the IAMC format, i.e., using `Model`, `Scenario`, `Region`, `Variable`, and `Unit` columns in addition to columns representing each modeled time period.

Users are able to control the harmonization process via a number of options. The primary mechanism by which users control the process is by providing

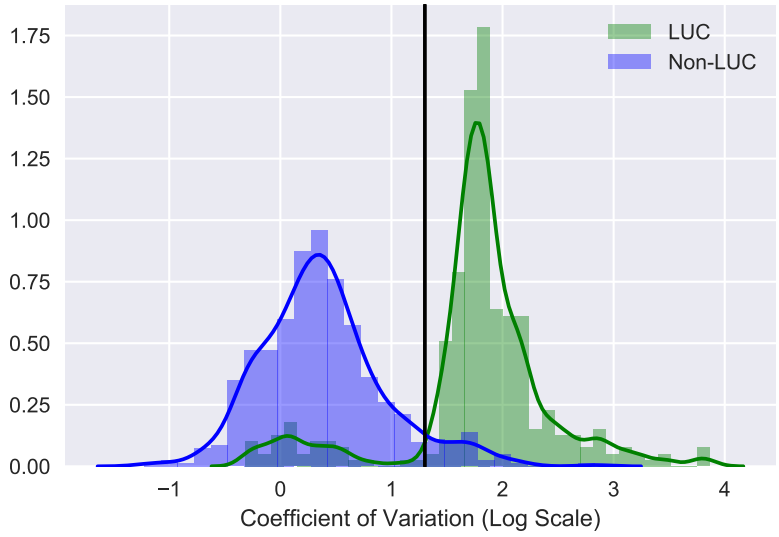


Figure 3: The distribution of $c_v$ values for LUC and non-LUC historical trajectories is shown. CEDS historical data [10] is used for non-LUC data and [11] is used for LUC data. All historical data has been aggregated from countries to IAM model regional definitions, and all gas species included in the historical data sets are included in the analysis. The solid black line indicates the threshold value, $\tau_{c_v}$, used by default in `aneris`.

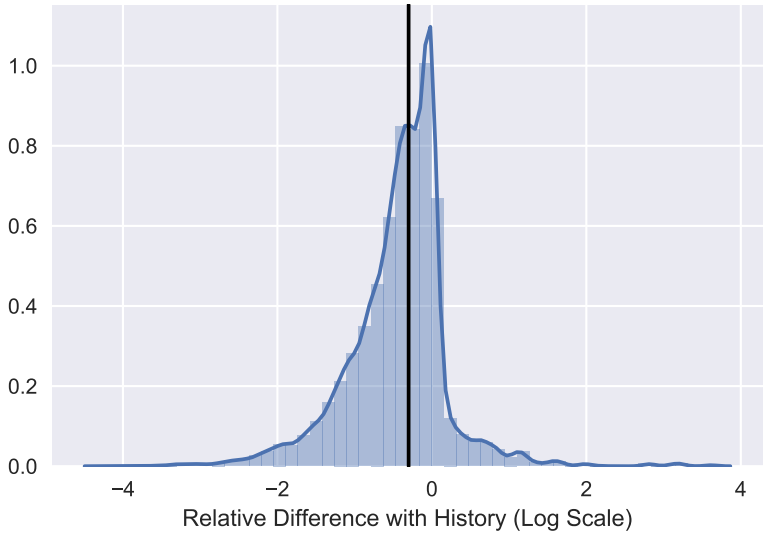


Figure 4: The distribution of relative differences between model and historical values in the harmonization year is shown. The solid black line indicates the 50% threshold value, $\tau_{dH}$, used by default in `aneris`.

*override* methods for any combination of region and variable (i.e., sector and gas species). In practice, it may be possible that not all default methods chosen will provide robust harmonized trajectories, especially if there is a significant difference between historical and model values in the harmonization year, if there is significant upward or downward movement in the model trajectory, or if there are known discrepancies in sectoral definition between the IAM and historical data source. In such cases, override methods can ameliorate the issues associated with the default method choice. In order to help identify these cases, harmonization *diagnostics* are provided which analyze the relative difference between harmonized and unharmonized trajectories at their mid (if $> 400\%$) and end-points (if $> 200\%$). If override methods are provided, they are used instead of the default methods as determined by the default method decision tree. Furthermore, users can set the above-mentioned thresholds as well as the LUC method used in the decision tree (see Figure 1). Further detail of input parameters can be found online.

Input data then undergoes a cleaning operation, which adds (null) model trajectories that exist in the historical data set but are not provided by the model input and detects any issues that would cause the harmonization process to fail. The methods used to harmonize the data are then determined and the harmonization process is executed. Upon completion of the harmonization process, aggregation of common analysis regions is performed. A common regional aggregation used in the IAM community was defined in the Representative Concentration Pathways (RCPs) [15], shown in Figure 6. Finally, any exogenous trajectories the user provides are added. Exogenous trajectories are normally provided for unmodeled gases with well-accepted scenario trajectories, e.g., chlorofluorocarbons provided by WMO [17]. Upon completion, the harmonized trajectories and meta data regarding the harmonization process are returned. A description of all returned meta data is provided in Table 2.

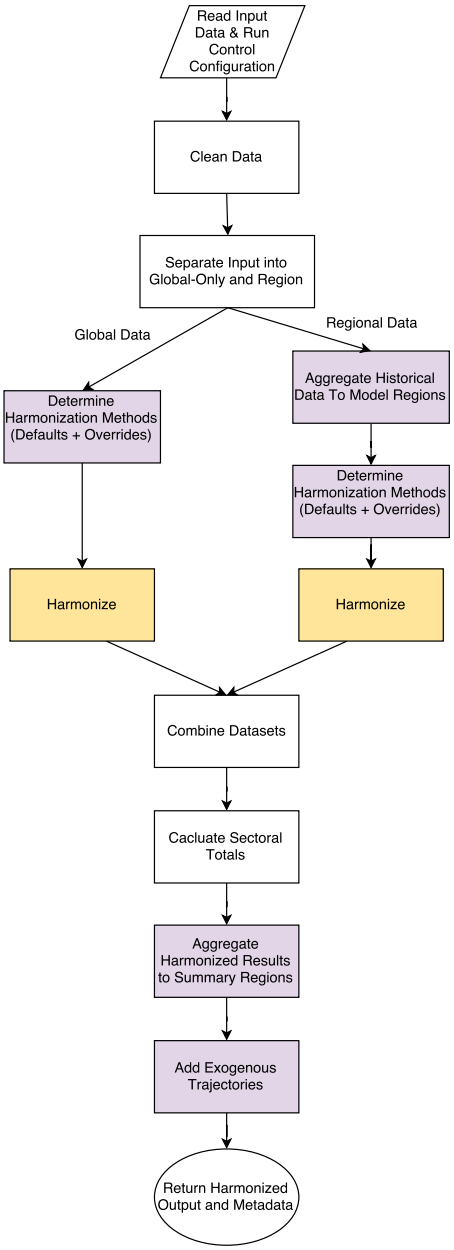


Figure 5: The full harmonization process as executed by `aneris`. Operations that can be configured with user-based input configurations are shown in purple. The core harmonization process is shown in yellow.

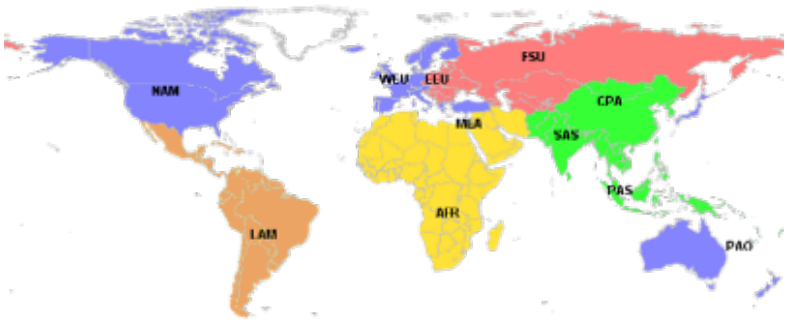


Figure 6: The 5 regions used in the RCPs with their 11-region constituents: Asia (Central Asia, South Asia, Pacific Asia), Latin America, the Middle East and Africa, the OECD (North America, Western Europe, and Pacific OECD), Reforming Economies (Eastern Europe and Former Soviet Union).

Table 2: Meta data provided by the `aneris` harmonization routine. This meta data is provided for every combination of region, sector, and emissions species.

| **Column** | **Description** |
|---|---|
| method | The harmonization method used. |
| default | The default harmonization method as determined by the default decision tree. |
| override | The method provided as an override (if any). |
| offset | The offset value between history and model in the harmonization year. |
| ratio | The ratio value between history and model in the harmonization year. |
| cov | The coefficient of variation value of the historical trajectory. |
| unharmonized | The unharmonized value in the harmonization year. |
| history | The historical value in the harmonization year. |
| harmonized | The resulting harmonized value in the harmonization year. |

## Results

In order to show a representative cross section of the performance of the `aneris` harmonization procedure, we focus on the harmonization of results of the IAM MESSAGE-GLOBIOM [3]. Two scenarios from the SSP scenario library [18, 19] are presented. The SSP2, or "middle of the road", scenario (referred to as SSP2-Ref) is chosen to be analyzed because MESSAGE-GLOBIOM is the marker scenario for this SSP. We additionally present the results for the SSP2-based mitigation scenario leading to a radiative forcing of 4.5 $\frac{\mathrm{W}}{\mathrm{m}^2}$ (referred to as SSP2-4.5). The SSP2-45 scenario is chosen because mitigation technologies and policies are enacted causing a general reduction in pollutants and GHGs, including (eventual) negative CO2 emissions in some regions and sectors due to carbon capture and sequestration and afforestation. Figure 7 shows the different trends of Kyoto Gases in each scenarios.

MESSAGE-GLOBIOM includes a representation of 11 distinct regions which can be mapped directly to the 5-region definition used in the RCPs. Historical data is taken from previously described LUC and anthropogenic sources, which comprise 10 separate pollutant and GHG species and 12 sectors shown in Table 3. A total of 970 distinct trajectories were harmonized for each scenario, and therefore 1940 trajectories were harmonized in total. $NO_x$ generated from the Energy sector provides an example of an emissions species and sector in which all regions were satisfactorily harmonized with the default methods. Figure 8 shows the results of harmonization in Asia, and Table 4 describes the parameters that underlie the choice of method for each harmonized trajectory.

The harmonization of emissions pathways is performed in order to accurately represent new or updated data sets of historical emissions inventories while also maintaining consistency with the original, unharmonized pathway. As such, when the default methods as provided by the harmonization procedure distort or otherwise sufficiently misrepresent the underlying unharmonized results, an override method is required to be provided for the trajectory of the region, sector, and species in question. Of the 970 trajectories, approximately 10%

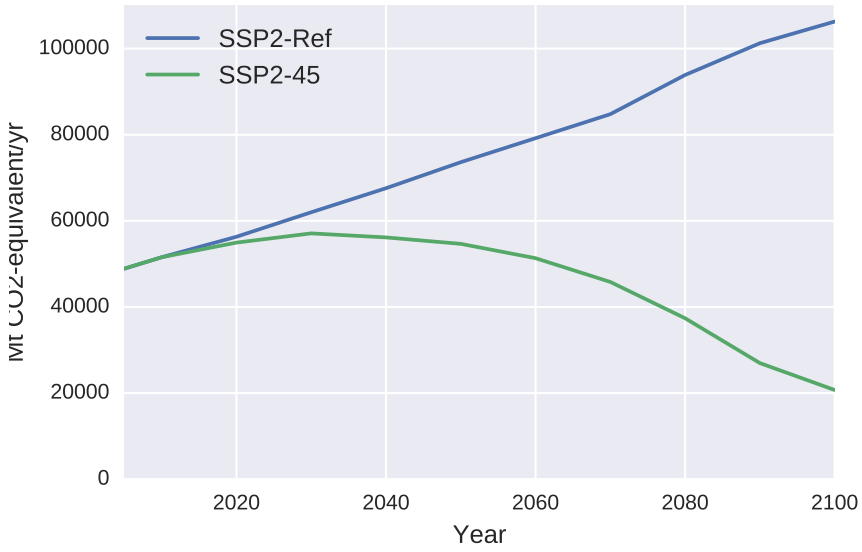


Figure 7: Unharmonized Kyoto gas emissions for SSP2-Ref, a scenario with generally increasing global emissions trends, and SSP2-45, a scenario with generally decreasing global emissions trends.

Table 3: Harmonized Species and Sectors

| **Emissions Species** | **Sectors** |
|---|---|
| Black Carbon (BC) | Agricultural Waste Burning |
| Hexafluoroethane ($C_2F_6$) [a] | Agriculture |
| Tetrafluoromethane ($CF_4$) [a] | Aircraft [b] |
| Methane ($CH_4$) | Energy Sector |
| Carbon Dioxide ($CO_2$) [c] | Forest Burning |
| Carbon Monoxide (CO) | Grassland Burning |
| Hydrofluorocarbons (HFCs) [a] | Industrial Sector |
| Nitrous Oxide ($N_2O$) [a] | International Shipping[b] |
| Ammonia ($NH_3$) | Residential Commercial Other |
| Nitrogen Oxides ($NO_x$) | Solvents Production and Application |
| Organic Carbon (OC) | Transportation Sector |
| Sulfur Hexafluoride ($SF_6$) [a] | Waste |
| Sulfur Oxides ($SO_x$) | |
| Volatile Organic Compounds (VOCs) | |

[a] Global total trajectories are harmonized due to lack of detailed historical data.

[b] Global sectoral trajectories are harmonized due to lack of detailed historical data.

[c] A global trajectory for land-use $CO_2$ is used; non-land-use sectors are harmonized for each model region.

Table 4: Key Parameters for Deciding Harmonization Methods for $NO_x$ Emissions in the Energy Sector in Asia

| **Region** | **dH** | **$c_v$** | **Decision Tree Traversal (Branch and Direction)** | **Default Method Chosen** |
|---|---|---|---|---|
| CPA | 0.35 | 2.26 | 1 (no), 2 (no), 3 (yes) | reduce_ratio_2080 |
| PAS | 0.14 | 1.24 | 1 (no), 2 (no), 3 (yes) | reduce_ratio_2080 |
| SAS | 0.56 | 0.58 | 1 (no), 2 (no), 3 (no), 4 (no) | constant_ratio |

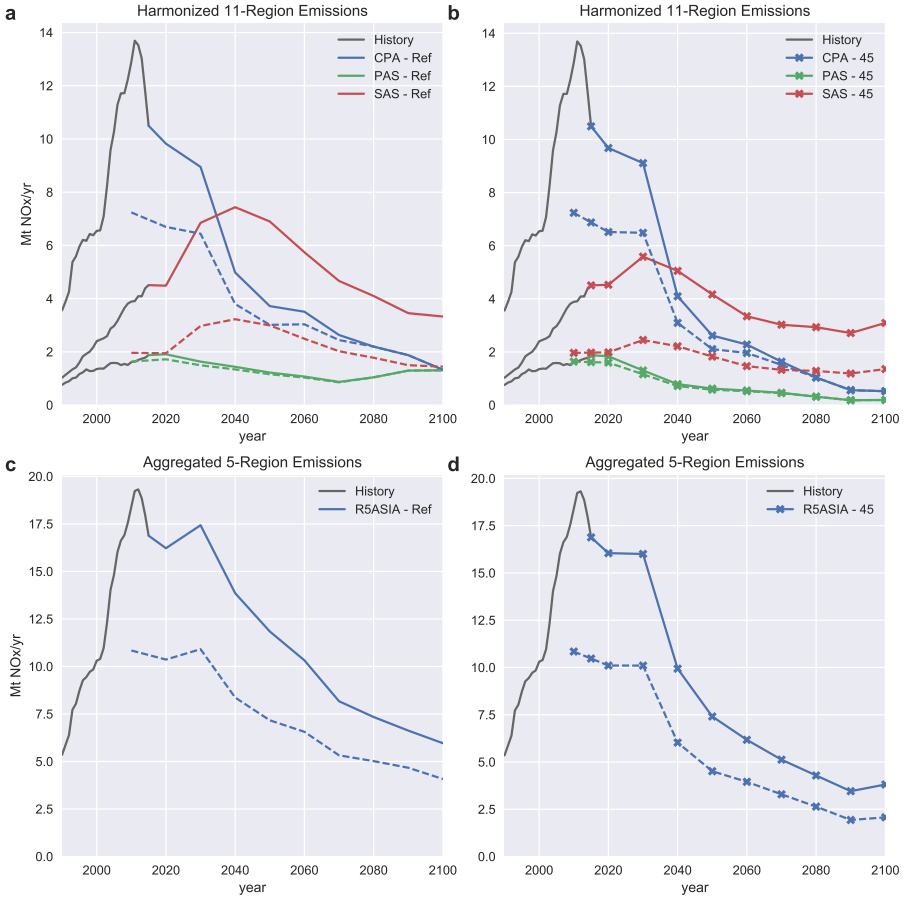


Figure 8: $NO_x$ Energy Sector harmonized (solid lines) and unharmonized (dashed lines) trajectories for SSP2 and SSP2-45 with historical trajectories (grey lines) are presented. The SSP2 reference scenario is shown in Panels **a** and **c**; the SSP-45 scenario is denoted with "x" markers in Panels **b** and **d**. The upper panels show the results for endogenously modeled and harmonized regions in Asia while the lower panels display the aggregate region results.

were reported as a diagnostic (see Section 2.3) of which 3.5% required the use of harmonization overrides after an initial investigation; thus, 96.5% of all trajectories were satisfactorily harmonized using the default methods. The trajectories that required overrides clustered into two classifications: regional trajectories whose *magnitude* was overly distorted and regional trajectories whose *shape* was overly distorted.

Figure 9 presents a case in which the magnitude of a trajectory is distorted. A large discrepancy (∼300% relative difference) is observed in the harmonization year for carbon monoxide (CO) emissions in the industrial sector specifically for the South Asia (SAS) MESSAGE-GLOBIOM region, which comprises most of the emissions of the Asian subcontinent. The default method chosen (`constant_ratio`) maintains model trends for the region; however, overall model results are distorted. By applying a `constant_offset` override, the regional trend and magnitude is maintained. With the new harmonization method for the SAS region, the global trajectory for industrial CO also is representative the trends seen in the unharmonized trajectory and the relative importance of the underlying regional trajectories is maintained.

In certain circumstances, the application of the default harmonization methods can affect not only the magnitude but also the shape of regional trajectories. Figure 10 shows an example case of emissions trajectories for ammonia ($NH_3$) from the agriculture sector in Asia. Again, the SAS region shows a large discrepancy in the harmonization year (>150% in this case). The resulting trajectory harmonized with the default method (`constant_ratio`) provides a large increase after 2080 in the SSP2 reference scenario. Notably, the SSP2-45 scenario is not affected to the same degree. While this distortion changes the magnitude of the SAS trajectory, it largely affects the post-2080 shape of the global trajectory (see Figure 10, panel **c**). By using a `constant_offset` method as an override, this distortion is addressed and more accurately reflects unharmonized results in the SAS region, the relative importance between regions, and global results for agricultural ammonia emissions, each of which contributes to a better harmonization quality for the harmonized SAS trajectory.

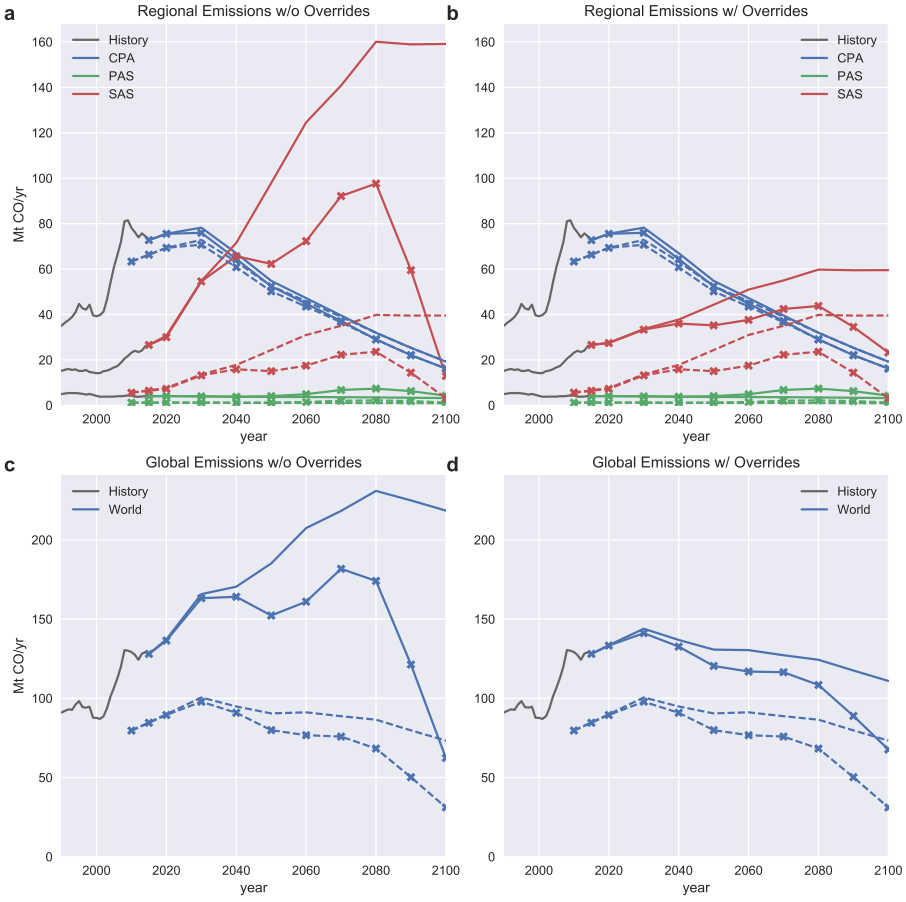


Figure 9: CO Industrial Sector harmonized and unharmonized emissions are presented for SSP2 and SSP2-45 scenarios. Scenarios as denoted identically to Figure 8. Panels **a** and **b** show harmonized and overridden-harmonized (respectively) regional trajectories for the 3 MESSAGE-GLOBIOM regions that comprise the R5ASIA region: Centrally Planned Asia (CPA), Other Pacific Asia (PAS), and South Asia (SAS). Notably, the SAS regional trajectory displays a distorted trajectory due to the harmonization-year difference between history and model results in both scenarios. The distortion is large enough to affect global results, as shown in Panels **c** and **d**.

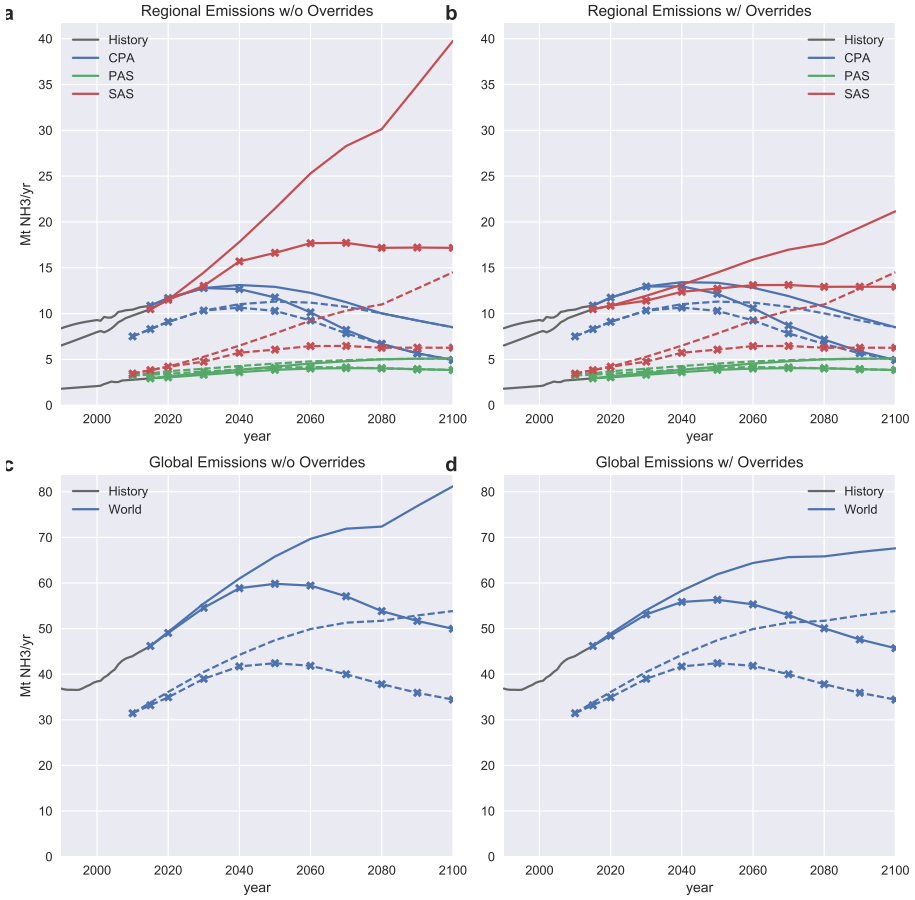


Figure 10: $NH_3$ agricultural harmonized and unharmonized emissions are presented for SSP2 and SSP2-45 scenarios. Scenarios and panel layouts are identical to Figure 9. In this case, the SAS trajectory again shows not only a magnitude distortion, but also a shape distortion at the tail of the trajectory. Additionally, global trajectories are greatly affected by the harmonization method choice (there is ∼20% relative difference between trajectories in the reference scenario in 2100). Override methods have been applied to correct the distortion.

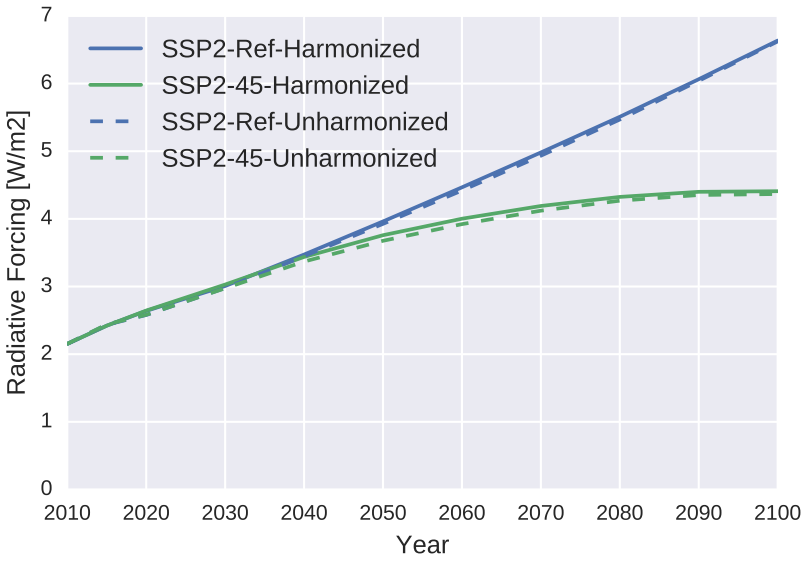


Figure 11: The results of the simple climate model, MAGICC6, forced with the SSP2-Ref (blue) and SSP2-4.5 (green) harmonized and unharmonized scenarios is presented. The radiative forcing trajectories of harmonized and unharmonized scenarios are shown in solid lines and dashed lines, respectively.

We investigate the aggregate effect of harmonization with all necessary override methods on total anthropogenic radiative forcing projections with the simple carbon-cycle and climate model, MAGICC6 [20, 21], for each harmonized and unharmonized scenario respectively as shown in Figure 11. We find that the change due to harmonization is small, ranging between 1 and 2.5% over the model horizon. Relative near-term differences persist in the mitigation case (SSP-4.5) because differences in near-term emissions define to a larger degree the longer-term forcing outcome due to the cumulative nature of long-lived climate forcers like $CO_2$ . The resulting difference in forcing in 2100 is 0.04 $\frac{\mathrm{W}}{\mathrm{m}^2}$ for SSP2-4.5 and 0.01 $\frac{\mathrm{W}}{\mathrm{m}^2}$ for SSP2-Ref, both of which are well within acceptable tolerances (e.g., 0.75 $\frac{\mathrm{W}}{\mathrm{m}^2}$ defined for ScenarioMIP [9]). Thus harmonization is considered to have a negligible effect on key long-term climate indicators.

## Discussion & Future Work

This work presented a novel methodology and Python implementation of automated emissions harmonization for IAMs, `aneris`. An in-depth explanation of the processes and methods for determining the use of harmonization methods was provided in Section 2. `aneris` was able to satisfactorily harmonized over 96% of the 1940 individual trajectories that were analyzed in Section 3. Of the remaining trajectories, harmonization method overrides were applied, and example situations in overrides were needed were discussed.

The automated approach drastically reduces the need for expert opinion in determining harmonization methods for each individual combination of model region, sector, and emissions species while still providing a justifiable explanation for each automated choice of harmonization method based on both the historical and future emissions trajectories. Furthermore, the automated approach continues to scale well as models become more detailed in both the regional and sectoral dimensions. Finally, expert opinion is still allowed to trump the automated method as determined by the algorithm via method overrides; however, these cases are clearly documented via the meta data provided as an output of `aneris` and thus can be individually explained. This provides not only transparency and reproducibility, but also scientific integrity in the choice of harmonization methods.

The use of an open-source, automated harmonization process also provides benefits to the wider climate science and IAM communities. By providing a standard mechanism for harmonization, the IAM community can directly provide input into the harmonization algorithms and rules for their default selection. Additionally, modeling teams are easily capable of executing identical harmonization procedures in order to participate in ongoing and further iterations of intercomparison exercises and analysis. Future scenario analyses can also utilize this common harmonization approach such that they are consistent with prior efforts.

There are a variety of avenues for future improvement of both the `aneris`

software and underlying methodology. As with any software project, additional users will provide use cases for more robust handling of input/output concerns and corner cases. Further configuration parameters can also be added in the future in order to provide overrides for all gas species in a given sector or region. Perhaps the most fruitful investigation will involve further refinement of the default decision tree introduced in Section 2. A key aspect missing from the decision tree is input from models regarding whether missing sources or activity levels are the likely cause of a harmonization year discrepancy (suggesting the use of an offset method) or instead a significant difference in emissions factors (suggesting the use of a ratio method) [14].

This work provides a new direction and framework which the IAM and climate communities can build upon in order to reduce the necessity of consistent expert opinion and increase transparency and reproducibility of harmonization exercises. Furthermore, it provides an open-source, tested, and documented software library which can be used and improved upon by these communities. Both of these are clear steps in a positive direction for future climate and integrated assessment modeling exercises.

**Acknowledgements**

The authors would like to acknowledge a number of colleagues who helped contribute both discussion and feedback regarding this work including Drs. Elmar Kriegler, Gunnar Luderer, and Joeri Rogelj. This project has received funding from the European Unions Horizon 2020 research and innovation programme under grant agreement No 641816. The authors further wish to thank the Global Environment Facility for their generous financial support.

## References


[1] V. Krey, Global energy-climate scenarios and models: a review, Wiley Interdisciplinary Reviews: Energy and Environment 3 (4) (2014) 363–383. doi:10.1002/wene.98.
URL http://onlinelibrary.wiley.com/doi/10.1002/wene.98/abstract

[2] D. P. van Vuuren, E. Stehfest, D. E. H. J. Gernaat, J. C. Doelman, M. van den Berg, M. Harmsen, H. S. de Boer, L. F. Bouwman, V. Daioglou, O. Y. Edelenbosch, B. Girod, T. Kram, L. Lassaletta, P. L. Lucas, H. van Meijl, C. Mller, B. J. van Ruijven, S. van der Sluis, A. Tabeau, Energy, land-use and greenhouse gas emissions trajectories under a green growth paradigm, Global Environmental Change 42 (2017) 237–250. doi:10.1016/j.gloenvcha.2016.05.008.
URL http://www.sciencedirect.com/science/article/pii/S095937801630067X

[3] O. Fricko, P. Havlik, J. Rogelj, Z. Klimont, M. Gusti, N. Johnson, P. Kolp, M. Strubegger, H. Valin, M. Amann, T. Ermolieva, N. Forsell, M. Herrero, C. Heyes, G. Kindermann, V. Krey, D. L. McCollum, M. Obersteiner, S. Pachauri, S. Rao, E. Schmid, W. Schoepp, K. Riahi, The marker quantification of the Shared Socioeconomic Pathway 2: A middle-of-the-road scenario for the 21st century, Global Environmental Change 42 (2017) 251–267. doi:10.1016/j.gloenvcha.2016.06.004.
URL http://www.sciencedirect.com/science/article/pii/S0959378016300784

[4] S. Fujimori, T. Hasegawa, T. Masui, K. Takahashi, D. S. Herran, H. Dai, Y. Hijioka, M. Kainuma, SSP3: AIM implementation of Shared Socioeconomic Pathways, Global Environmental Change 42 (2017) 268–283. doi:10.1016/j.gloenvcha.2016.06.009.

URL http://www.sciencedirect.com/science/article/pii/S0959378016300838

[5] K. Calvin, B. Bond-Lamberty, L. Clarke, J. Edmonds, J. Eom, C. Hartin, S. Kim, P. Kyle, R. Link, R. Moss, H. McJeon, P. Patel, S. Smith, S. Waldhoff, M. Wise, The SSP4: A world of deepening inequality, Global Environmental Change 42 (2017) 284–296. doi:10.1016/j.gloenvcha.2016.06.010.
URL http://www.sciencedirect.com/science/article/pii/S095937801630084X

[6] E. Kriegler, N. Bauer, A. Popp, F. Humpender, M. Leimbach, J. Strefler, L. Baumstark, B. L. Bodirsky, J. Hilaire, D. Klein, I. Mouratiadou, I. Weindl, C. Bertram, J.-P. Dietrich, G. Luderer, M. Pehl, R. Pietzcker, F. Piontek, H. Lotze-Campen, A. Biewald, M. Bonsch, A. Giannousakis, U. Kreidenweis, C. Mller, S. Rolinski, A. Schultes, J. Schwanitz, M. Stevanovic, K. Calvin, J. Emmerling, S. Fujimori, O. Edenhofer, Fossil-fueled development (SSP5): An energy and resource intensive scenario for the 21st century, Global Environmental Change 42 (2017) 297–315. doi:10.1016/j.gloenvcha.2016.05.015.
URL http://www.sciencedirect.com/science/article/pii/S0959378016300711

[7] D. P. v. Vuuren, E. Kriegler, B. C. ONeill, K. L. Ebi, K. Riahi, T. R. Carter, J. Edmonds, S. Hallegatte, T. Kram, R. Mathur, H. Winkler, A new scenario framework for Climate Change Research: scenario matrix architecture, Climatic Change 122 (3) (2013) 373–386. doi:10.1007/s10584-013-0906-1.
URL http://link.springer.com/article/10.1007/s10584-013-0906-1

[8] V. Eyring, S. Bony, G. A. Meehl, C. A. Senior, B. Stevens, R. J. Stouffer, K. E. Taylor, Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, Geosci. Model Dev.

9 (5) (2016) 1937–1958. doi:10.5194/gmd-9-1937-2016.
URL http://www.geosci-model-dev.net/9/1937/2016/

[9] B. C. O'Neill, C. Tebaldi, D. P. van Vuuren, V. Eyring, P. Friedlingstein, G. Hurtt, R. Knutti, E. Kriegler, J.-F. Lamarque, J. Lowe, others, The scenario model intercomparison project (scenariomip) for CMIP6, Geoscientific Model Development 9 (9) (2016) 3461.
URL http://search.proquest.com/openview/e2eae675069c0ff45c7f19861f7dfcca/1?pq-origsite=gscholar&cbl=105726

[10] R. M. Hoesly, S. J. Smith, L. Feng, Z. Klimont, G. Janssens-Maenhout, T. Pitkanen, J. J. Seibert, L. Vu, R. J. Andres, R. M. Bolt, T. C. Bond, L. Dawidowski, N. Kholod, J.-I. Kurokawa, M. Li, L. Liu, Z. Lu, M. C. P. Moura, P. R. O'Rourke, Q. Zhang, Historical (17502014) anthropogenic emissions of reactive gases and aerosols from the Community Emission Data System (CEDS), Geosci. Model Dev. Discuss. 2017 (2017) 1–41. doi:10.5194/gmd-2017-43.
URL http://www.geosci-model-dev-discuss.net/gmd-2017-43/

[11] M. J. E. van Marle, S. Kloster, B. I. Magi, J. R. Marlon, A.-L. Daniau, R. D. Field, A. Arneth, M. Forrest, S. Hantson, N. M. Kehrwald, W. Knorr, G. Lasslop, F. Li, S. Mangeon, C. Yue, J. W. Kaiser, G. R. van der Werf, Historic global biomass burning emissions based on merging satellite observations with proxies and fire models (1750-2015), Geosci. Model Dev. Discuss. 2017 (2017) 1–56. doi:10.5194/gmd-2017-32.
URL http://www.geosci-model-dev-discuss.net/gmd-2017-32/

[12] M. Meinshausen, S. J. Smith, K. Calvin, J. S. Daniel, M. L. T. Kainuma, J.-F. Lamarque, K. Matsumoto, S. A. Montzka, S. C. B. Raper, K. Riahi, A. Thomson, G. J. M. Velders, D. P. P. v. Vuuren, The RCP greenhouse gas concentrations and their extensions from 1765 to 2300, Climatic Change 109 (1-2) (2011) 213. doi:10.1007/s10584-011-0156-z.

URL https://link.springer.com/article/10.1007/s10584-011-0156-z

[13] N. Nakićenović, J. Alcamo, G. Davis, B. de Vries, J. Fenhann, S. Gaffin, K. Gregory, A. Grübler, T. Y. Jung, T. Kram, E. Lebre La Rovere, L. Michaelis, S. Mori, T. Morita, W. Pepper, H. Pitcher, L. Price, K. Riahi, A. Roehrl, H. H. Rogner, A. Sankovski, M. Schlesinger, P. Shukla, S. Smith, R. Swart, S. van Rooijen, N. Victor, Z. Dadi, IPCC Special Report on Emissions Scenarios (SRES), Cambridge University Press, UK, 2000.
URL http://www.ipcc.ch/ipccreports/sres/emission/index.php?idp=0

[14] J. Rogelj, W. Hare, C. Chen, M. Meinshausen, Discrepancies in historical emissions point to a wider 2020 gap between 2 C benchmarks and aggregated national mitigation pledges, Environmental Research Letters 6 (2) (2011) 024002. doi:10.1088/1748-9326/6/2/024002.
URL http://stacks.iop.org/1748-9326/6/i=2/a=024002

[15] D. P. v. Vuuren, J. Edmonds, M. Kainuma, K. Riahi, A. Thomson, K. Hibbard, G. C. Hurtt, T. Kram, V. Krey, J.-F. Lamarque, T. Masui, M. Meinshausen, N. Nakicenovic, S. J. Smith, S. K. Rose, The representative concentration pathways: an overview, Climatic Change 109 (1-2) (2011) 5–31. doi:10.1007/s10584-011-0148-z.
URL http://link.springer.com/article/10.1007/s10584-011-0148-z

[16] M. Gidden, aneris: Harmonization for Integrated Assessment Models (Jun. 2017). doi:10.5281/zenodo.802832.
URL https://doi.org/10.5281/zenodo.802832

[17] G. O. Research, M. Project, Scientific Assessment of Ozone Depletion: 2014, Tech. rep., World Meteorological Organization, Geneva, Switzerland (2014).

[18] K. Riahi, D. P. van Vuuren, E. Kriegler, J. Edmonds, B. C. ONeill, S. Fujimori, N. Bauer, K. Calvin, R. Dellink, O. Fricko, W. Lutz, A. Popp, J. C.

Cuaresma, S. KC, M. Leimbach, L. Jiang, T. Kram, S. Rao, J. Emmerling, K. Ebi, T. Hasegawa, P. Havlik, F. Humpender, L. A. D. Silva, S. Smith, E. Stehfest, V. Bosetti, J. Eom, D. Gernaat, T. Masui, J. Rogelj, J. Strefler, L. Drouet, V. Krey, G. Luderer, M. Harmsen, K. Takahashi, L. Baumstark, J. C. Doelman, M. Kainuma, Z. Klimont, G. Marangoni, H. Lotze-Campen, M. Obersteiner, A. Tabeau, M. Tavoni, The shared socioeconomic pathways and their energy, land use, and greenhouse gas emissions implications: An overview, Global Environmental Change 42 (2017) 153 – 168. doi:https://doi.org/10.1016/j.gloenvcha.2016.05.009.
URL http://www.sciencedirect.com/science/article/pii/S0959378016300681

[19] S. Rao, Z. Klimont, S. J. Smith, R. V. Dingenen, F. Dentener, L. Bouwman, K. Riahi, M. Amann, B. L. Bodirsky, D. P. van Vuuren, L. A. Reis, K. Calvin, L. Drouet, O. Fricko, S. Fujimori, D. Gernaat, P. Havlik, M. Harmsen, T. Hasegawa, C. Heyes, J. Hilaire, G. Luderer, T. Masui, E. Stehfest, J. Strefler, S. van der Sluis, M. Tavoni, Future air pollution in the shared socio-economic pathways, Global Environmental Change 42 (2017) 346 – 358. doi:https://doi.org/10.1016/j.gloenvcha.2016.05.012.
URL http://www.sciencedirect.com/science/article/pii/S0959378016300723

[20] M. Meinshausen, S. C. Raper, T. M. Wigley, Emulating coupled atmosphere-ocean and carbon cycle models with a simpler model, magicc6–part 1: Model description and calibration, Atmospheric Chemistry and Physics 11 (4) (2011) 1417–1456.

[21] M. Meinshausen, S. J. Smith, K. Calvin, J. S. Daniel, M. Kainuma, J. Lamarque, K. Matsumoto, S. Montzka, S. Raper, K. Riahi, et al., The rcp greenhouse gas concentrations and their extensions from 1765 to 2300, Climatic change 109 (1-2) (2011) 213.