Managing multi-model scenario ensembles for climate change & energy systems transition

Lessons & insights from supporting the IPCC process and Horizon 2020 projects at the research theme "Scenario Services & Scientific Software" of the IIASA Energy, Climate, and Environment program

This presentation is available at
https://doi.org/10.5281/zenodo.6361259

Daniel Huppmann
EERAdata workshop, March 16, 2022

Agenda

- IIASA ECE as a community data hub
- Tools and packages for scenario management & analysis
- Best-practice of FAIR & open science: The IPCC SR15
- Conclusions & Outlook: Scope for collaboration with EERAdata

Part 1

IIASA ECE as a community data hub

The IIASA as community data hub

Supporting the modelling community for more than a decade

The role of the IIASA Energy, Climate, and Environment program

- Hosting scenario databases to support model comparison projects
  e.g. Energy Modeling Forum (EMF) organized by Stanford University

- Contributing to community processes on data standards & formats
  e.g., Integrated Assessment Modeling Consortium (IAMC)
  ⇒ "WG on Data Protocols & Management" co-chaired by Dr. Volker Krey

- Capacity-building for national teams (e.g., Horizon 2020 "ENGAGE")

Selected funding sources for infrastructure development

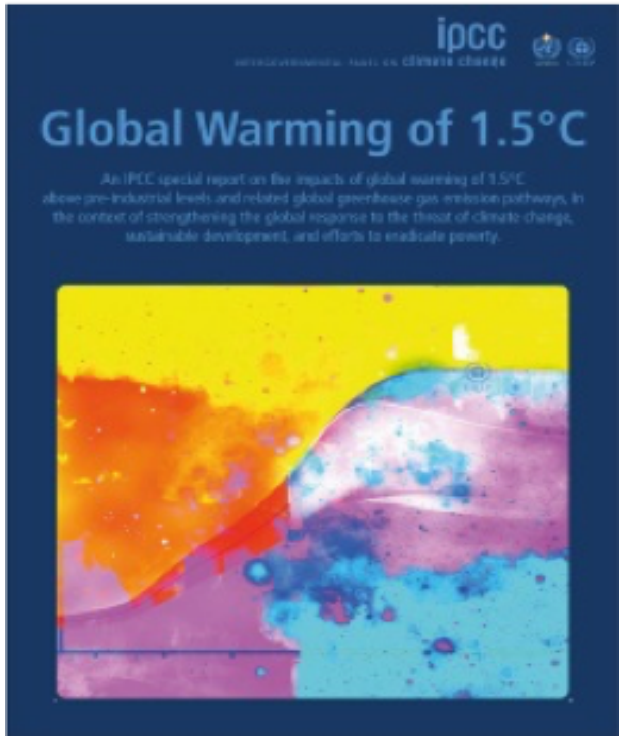
The IIASA Energy program as community data hub

Hosting community databases for dissemination of results

Selection of high-profile public scenario databases

- ⇒ Representative Concentration Pathways (RCPs, 2009)
- ⇒ IPCC AR5 Scenario Database (2014)
- ⇒ Shared Socio-economic Pathways (SSPs, 2018)
- ⇒ Horizon 2020 project "CD-LINKS" (2018-2019)
  Bringing together global & national modelling teams
- ⇒ IAMC 1.5°C Scenario Explorer supporting IPCC SR15

More information: https://data.ece.iiasa.ac.at



The IPCC *Special Report on Global Warming of 1.5°C* (SR15) was published in 2018. Many high-level statements are based directly on a scenario ensemble hosted by IIASA

The IIASA Energy program as community data hub

Ongoing efforts towards open & FAIR science

- Currently ongoing Horizon 2020 projects (selected)

  Developing new tools for dissemination, communication and stakeholder engagement

open ENTRANCE

ENGAGE
FEASIBILITY OF CLIMATE PATHWAYS

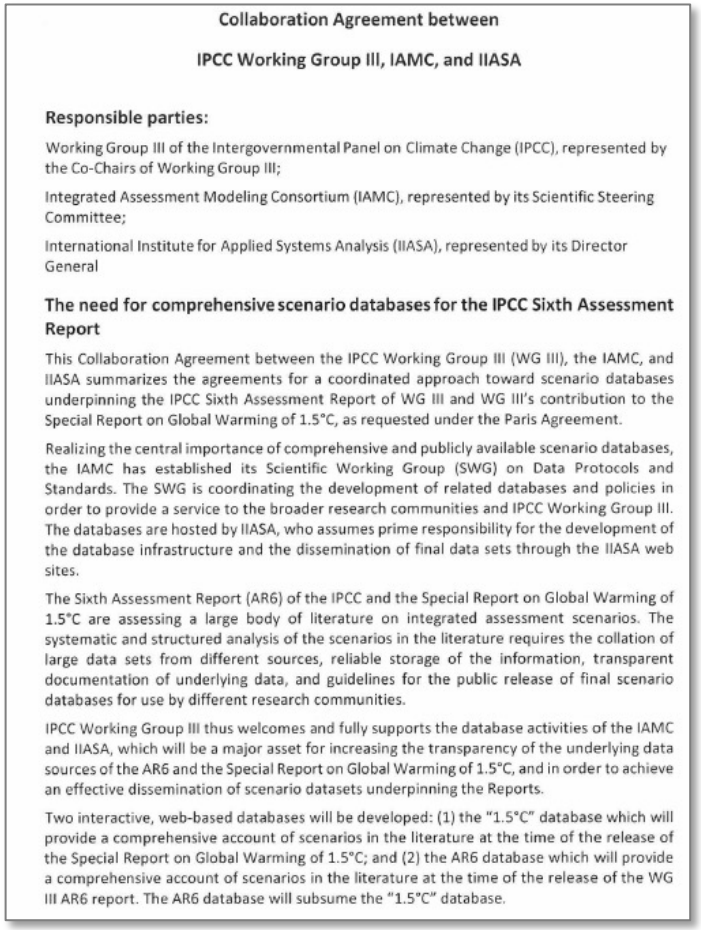
EUROPEAN CLIMATE+ENERGY MODELLING FORUM

IPCC
INTERGOVERNMENTAL PANEL ON climate change

- Collaboration with IPCC for 6th Assessment Report

  Researchers at the Energy program are currently compiling a scenario ensemble supporting the AR6

A collaboration agreement between the IPCC WGIII, the IAMC and IIASA sets the scope of cooperation for the sixth assessment cycle



Collaboration Agreement between

IPCC Working Group III, IAMC, and IIASA

Responsible parties:

Working Group III of the Intergovernmental Panel on Climate Change (IPCC), represented by the Co-Chairs of Working Group III;

Integrated Assessment Modeling Consortium (IAMC), represented by its Scientific Steering Committee;

International Institute for Applied Systems Analysis (IIASA), represented by its Director General

The need for comprehensive scenario databases for the IPCC Sixth Assessment Report

This Collaboration Agreement between the IPCC Working Group III (WG III), the IAMC, and IIASA summarizes the agreements for a coordinated approach toward scenario databases underpinning the IPCC Sixth Assessment Report of WG III and WG III's contribution to the Special Report on Global Warming of 1.5°C, as requested under the Paris Agreement.

Realizing the central importance of comprehensive and publicly available scenario databases, the IAMC has established its Scientific Working Group (SWG) on Data Protocols and Standards. The SWG is coordinating the development of related databases and policies in order to provide a service to the broader research communities and IPCC Working Group III. The databases are hosted by IIASA, who assumes prime responsibility for the development of the database infrastructure and the dissemination of final data sets through the IIASA web sites.

The Sixth Assessment Report (AR6) of the IPCC and the Special Report on Global Warming of 1.5°C are assessing a large body of literature on integrated assessment scenarios. The systematic and structured analysis of the scenarios in the literature requires the collation of large data sets from different sources, reliable storage of the information, transparent documentation of underlying data, and guidelines for the public release of final scenario databases for use by different research communities.

IPCC Working Group III thus welcomes and fully supports the database activities of the IAMC and IIASA, which will be a major asset for increasing the transparency of the underlying data sources of the AR6 and the Special Report on Global Warming of 1.5°C, and in order to achieve an effective dissemination of scenario datasets underpinning the Reports.

Two interactive, web-based databases will be developed: (1) the "1.5°C" database which will provide a comprehensive account of scenarios in the literature at the time of the release of the Special Report on Global Warming of 1.5°C; and (2) the AR6 database which will provide a comprehensive account of scenarios in the literature at the time of the release of the WG III AR6 report. The AR6 database will subsume the "1.5°C" database.

A new research theme

To streamline our efforts towards supporting scenario analysis, IIASA established a theme „Scenario Services & Scientific Software“

Mission of the research theme:

- ⇒ Develop the infrastructure for working with scenario ensembles related to integrated assessment, energy systems, and climate change
- ⇒ Support researchers at the Energy, Climate, and Environment program in adopting better practices

  e.g., tutorials, code review, pair-programming, …
- ⇒ Serve as a support team for the wider community

More information: https://software.ece.iiasa.ac.at/

Part 2

Tools and packages for scenario management & analysis

The IIASA Scenario Explorer

An interactive, versatile web user interface for model comparison projects and dissemination of results to researchers, policymakers & stakeholders

Scope and features

- Make scenario results accessible including documentation
- Manage scenario results in model comparison projects
- Facilitate "post-processing" of scenario results

Currently used in various projects

open ENTRANCE

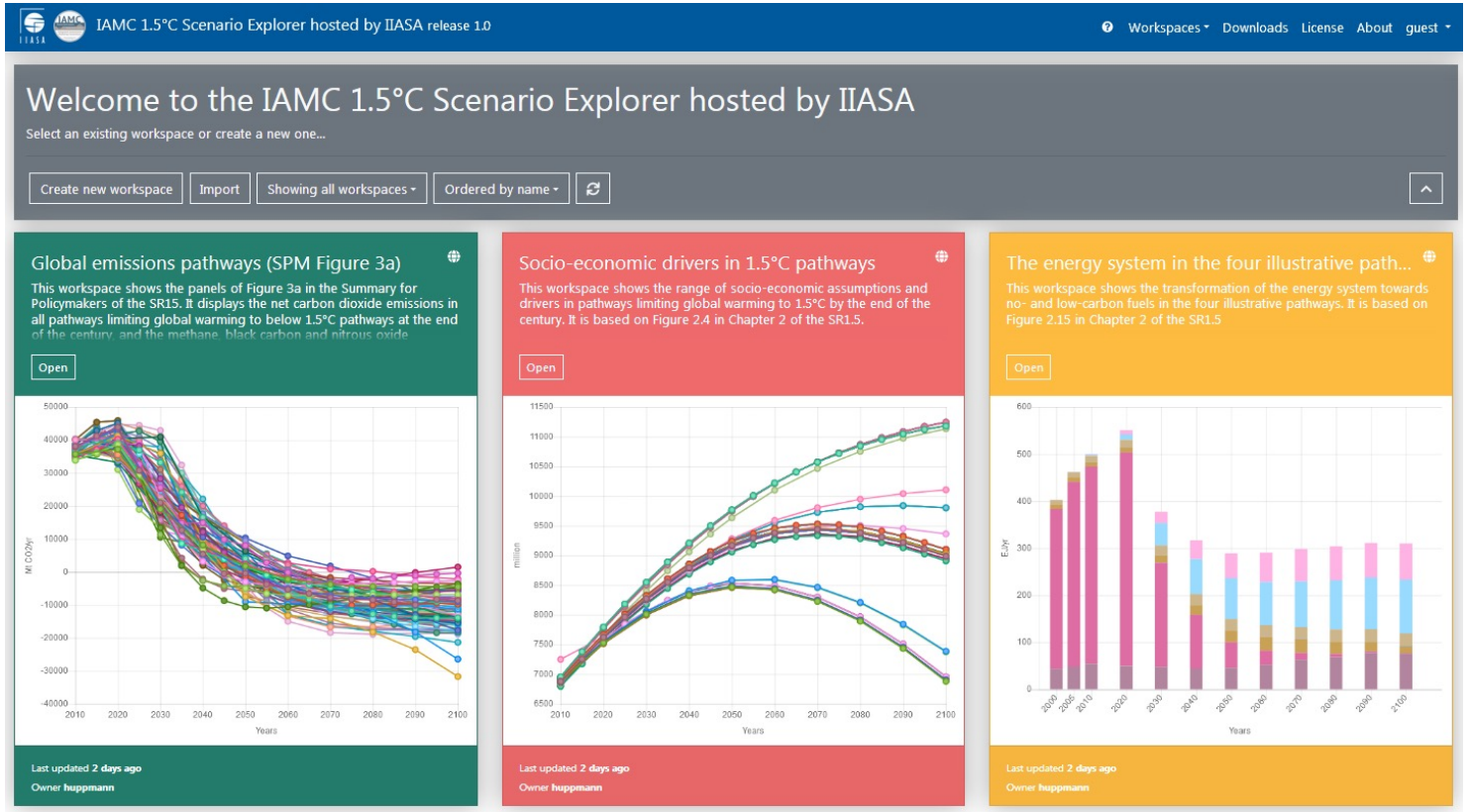
ENGAGE FEASIBILITY OF CLIMATE PATHWAYS

NAVIGATE

IPCC INTERGOVERNMENTAL PANEL ON climate change

EMF

KOPERNIKUS PROJEKTE Ariadne Die Zukunft unserer Energie

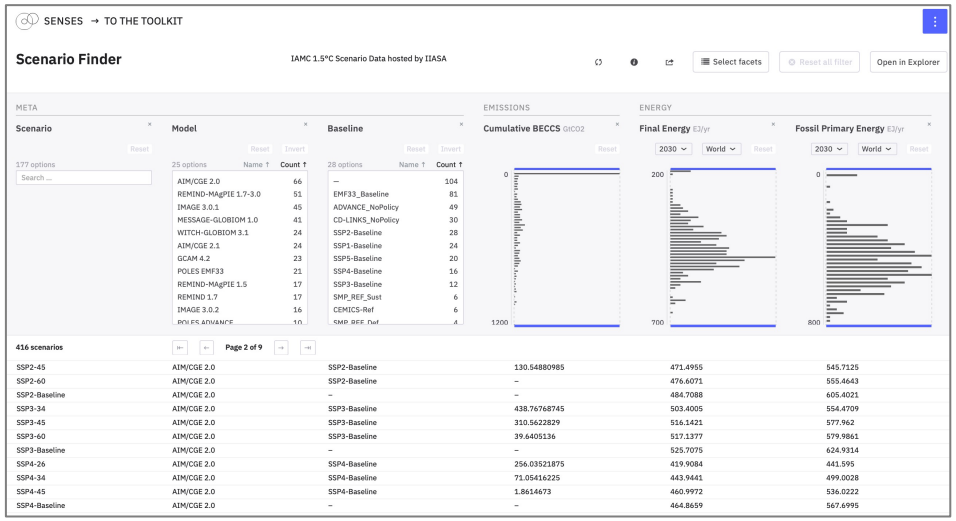


The Scenario Explorer was initially developed for the IPCC's Special Report on 1.5°C. Visit the IAMC 1.5°C Scenario Explorer https://data.ece.iiasa.ac.at/iamc-1.5c-explorer

The IIASA database API

The IIASA database infrastructure already provides an open API that enables third parties to develop tools and services

- Programming language connections via a RestAPI
  - ⇒ Python package "pyam" (see later slide)
  - ⇒ Matlab app (https://github.com/mathworks/)
  - ⇒ R package in development
- The RestAPI can be used to develop own visualization websites & tools
- Documentation of current implementation
  - ⇒ http://software.ece.iiasa.ac.at/ixmp-server



The "Scenario Finder" was implemented by FH Potsdam in the ERA4CS SENSES project to make data in the IAMC 1.5°C Scenario Explorer more easily accessible.
https://climatescenarios.org/finder/

The IAMC template for timeseries data

A community standard for compiling scenario results

IAMC
Integrated Assessment Modeling Consortium
Founded 2007

The integrated-assessment community (IAMC) developed a tabular scenario data format

- ⇒ Used in IPCC Reports (AR5, SR15), Horizon 2020 projects, …
- ⇒ Adopted by ~50 teams globally
- ⇒ Details in the pyam docs

EMF NAVIGATE ENGAGE FEASIBILITY OF CLIMATE PATHWAYS

	A	B	C	D	E	F	G	H	
1	**Model**	**Scenario**	**Region**	**Variable**	**Unit**	**2005**	**2010**	**2015**	
2	MESSAGE	CD-LINKS 400	World	Primary Energy	EJ/y	462.5	500.7	...	

The Horizon 2020 project openENTRANCE is implementing an extension to cover sub-annual time resolution.

- ⇒ Check out github.com/openENTRANCE/openentrance for details!

open ENTRANCE

The *pyam* package

A community package for scenario processing, analysis & visualization following best practice of collaborative scientific software development

Use cases and features

- ⇒ **Data processing** Aggregation, downscaling, unit conversion, I/O to xlsx, csv & frictionless datapackage…
- ⇒ **Validation** Checks for completeness of data, internal/external consistency, numerical plausibility …
- ⇒ **Analysis & visualization** Categorization and statistics of scenario ensembles, plotting library, …

Huppmann et al., 2021. *Open Research Europe*, 1:74 https://doi.org/10.12688/openreseurope.13633.2

Repository hosted on GitHub

Community supported by Groups.io slack

Documentation hosted by Read the Docs

#pyam_iamc

pyam-iamc.readthedocs.io

The design principles of our software stack

Our vision is a versatile and modular infrastructure for scenario management and analysis

- Project specifications: codelists, regions, …
  ⇒ Use yaml files hosted on GitHub
  e.g., github.com/openENTRANCE/openentrance

```
- Primary Energy:
    definition: Total primary energy use
    unit: EJ/yr
- Primary Energy|Coal:
    definition: Primary use of coal
    unit: EJ/yr
```

- Nomenclature
  ⇒ An open-source Python utility for validation and region-processing
- Pyam
  ⇒ Functions and methods for scenario processing (specific to the IAMC format)
- Pandas
  ⇒ The full-fledged solution for data operations in Python

Improve findability of research

We manage two communities on Zenodo

Zenodo is the ideal place for storing data and packages in line with the FAIR principles

The research theme manages two communities:

- ⇒ For the work done at our research program: https://zenodo.org/communities/iiasa-ece
- ⇒ For the wider research domain (on behalf of the IAM consortium) https://zenodo.org/communities/iamconsortium

A one-slide guide to open & FAIR research

Making it easy for researchers to adopt FAIRer practices

DOI:
10.22022/ene/04-2020.16404

More about the FAIR principles:
www.go-fair.org/fair-principles/

Please cite as: Daniel Huppmann et al., 2020
Five best-practice steps to make your research open & FAIR v1.0
doi: 10.22022/ene/04-2020.16404 | url: openENTRANCE.eu

Five best-practice steps to make your research open & FAIR v1.0

open ENTRANCE

You may think that putting your work* on a website already makes it free & open.
But that's not quite true – follow these steps to implement best practice of ***#openscience***!

* data sets, text, tables, figures & illustrations, source code, scientific software, ... even #Horizon2020 deliverables

1. Open	If you want your ***work to be read, used & shared by others***, be explicit about it: For text, data, figures, ... – use the CC-BY license | For code, visit choosealicense.com
2. Findable	To make it easy for others to find and cite your work, get a digital object identifier (DOI) and add a ***recommended citation***
3. Accessible	Depositing your work in an institutional repository or a service like zenodo ensures that your work is still ***available even after the end of the project***
4. Interoperable	Using established community standards, data formats and software packages lets others ***quickly understand and use your work***
5. Reusable	To make it easy for others to ***build on your work***, make sure to assign a version number and relevant (machine-readable) metadata

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 835896

Part 3

Best-practice of FAIR & open science: The IPCC SR15

A Special Report on Global Warming of 1.5°C

Analyzing impacts of climate change in the context of the SDGs

Harry Taylor, 6, played with the bones of dead livestock in Australia, which has faced severe drought.

Brook Mitchell/Getty Images

Where do these numbers come from?

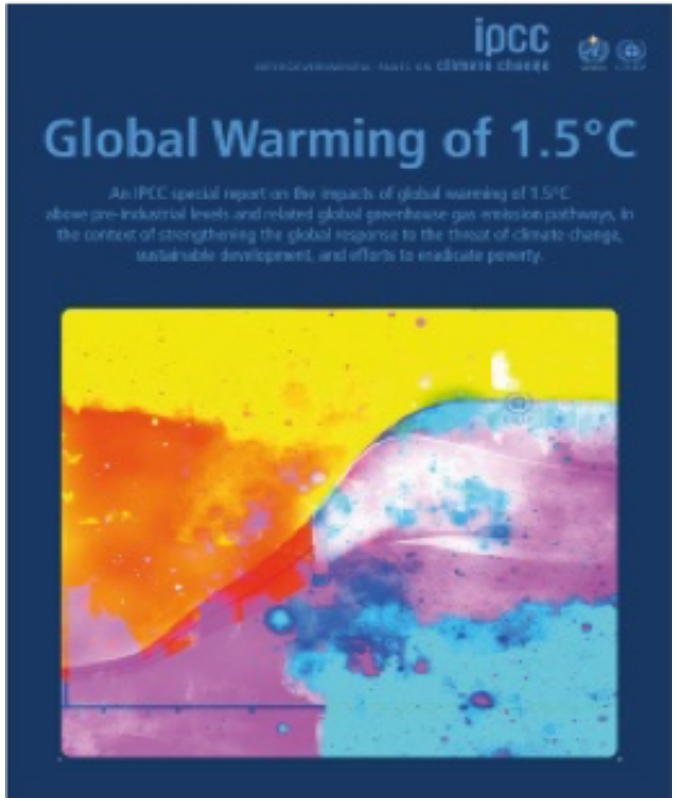
The New York Times

Major Climate Report Describes a Strong Risk of Crisis as Early as 2040

[…] To prevent 2.7 degrees of warming, the report said, greenhouse pollution must be reduced by 45 percent from 2010 levels by 2030, and 100 percent by 2050. It also found that, by 2050, use of coal as an electricity source would have to drop from nearly 40 percent today to between 1 and 7 percent. Renewable energy such as wind and solar, which make up about 20 percent of the electricity mix today, would have to increase to as much as 67 percent. […]

www.nytimes.com/2018/10/07/climate/ipcc-climate-report-2040.html

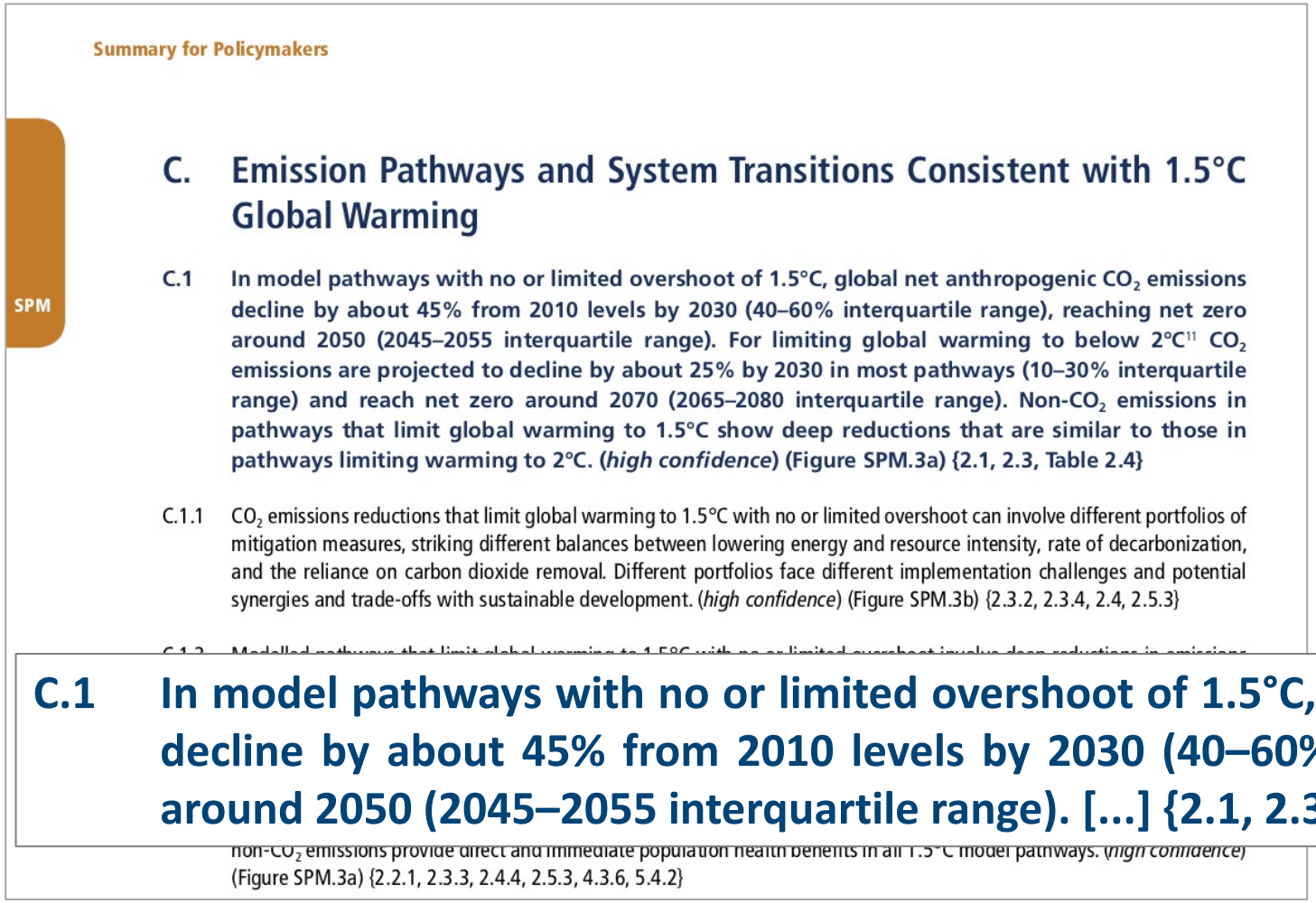
The IPCC *Special Report on Global Warming of 1.5°C* (SR15) was published in the fall of 2018.



www.ipcc.ch/sr15

Diving into the 'Summary for Policymakers'

The IPCC assessed a large ensemble of emissions pathways



Summary for Policymakers

SPM

C. Emission Pathways and System Transitions Consistent with 1.5°C Global Warming

C.1 In model pathways with no or limited overshoot of 1.5°C, global net anthropogenic CO_2 emissions decline by about 45% from 2010 levels by 2030 (40–60% interquartile range), reaching net zero around 2050 (2045–2055 interquartile range). For limiting global warming to below 2°C[11] CO_2 emissions are projected to decline by about 25% by 2030 in most pathways (10–30% interquartile range) and reach net zero around 2070 (2065–2080 interquartile range). Non-CO_2 emissions in pathways that limit global warming to 1.5°C show deep reductions that are similar to those in pathways limiting warming to 2°C. (*high confidence*) (Figure SPM.3a) {2.1, 2.3, Table 2.4}

C.1.1 CO_2 emissions reductions that limit global warming to 1.5°C with no or limited overshoot can involve different portfolios of mitigation measures, striking different balances between lowering energy and resource intensity, rate of decarbonization, and the reliance on carbon dioxide removal. Different portfolios face different implementation challenges and potential synergies and trade-offs with sustainable development. (*high confidence*) (Figure SPM.3b) {2.3.2, 2.3.4, 2.4, 2.5.3}

C.1.2 Modelled pathways that limit global warming to 1.5°C with no or limited overshoot involve deep reductions in emissions

non-CO_2 emissions provide direct and immediate population health benefits in all 1.5°C model pathways. (*high confidence*) (Figure SPM.3a) {2.2.1, 2.3.3, 2.4.4, 2.5.3, 4.3.6, 5.4.2}

C.1 In model pathways with no or limited overshoot of 1.5°C, global net anthropogenic CO_2 emissions decline by about 45% from 2010 levels by 2030 (40–60% interquartile range), reaching net zero around 2050 (2045–2055 interquartile range). […] {2.1, 2.3, Table 2.4}

The Summary for Policymakers of the IPCC *Special Report on Global Warming of 1.5°C* (SR15).

The "line of sight" of the SR15 scenario ensemble

We developed a suite of open tools to dive into the SR15 analysis

Interactive online scenario explorer at
data.ene.iiasa.ac.at/iamc-1.5c-explorer

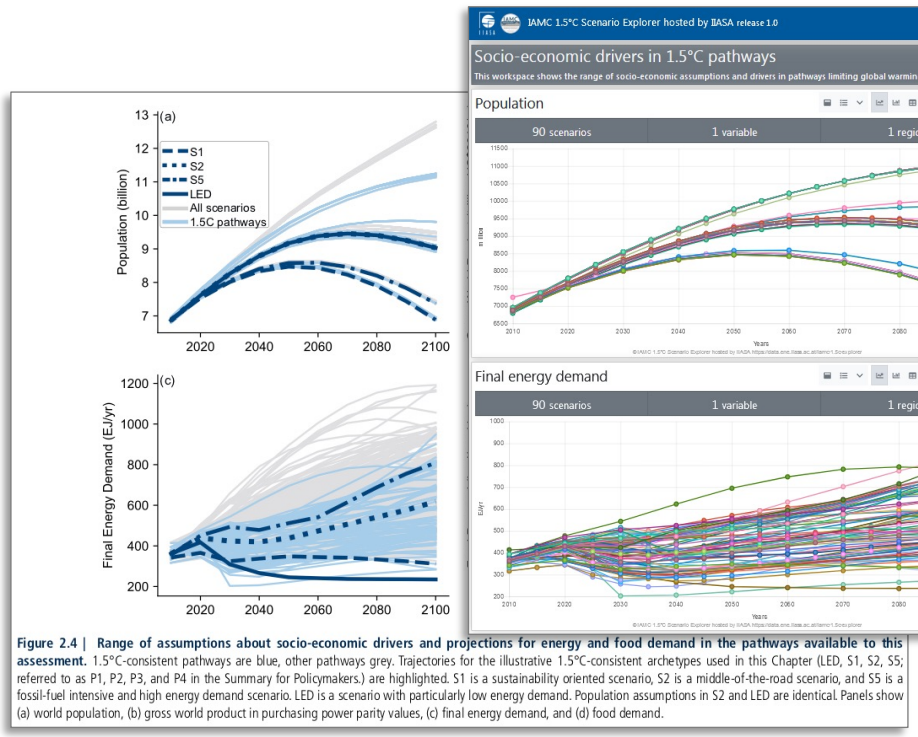
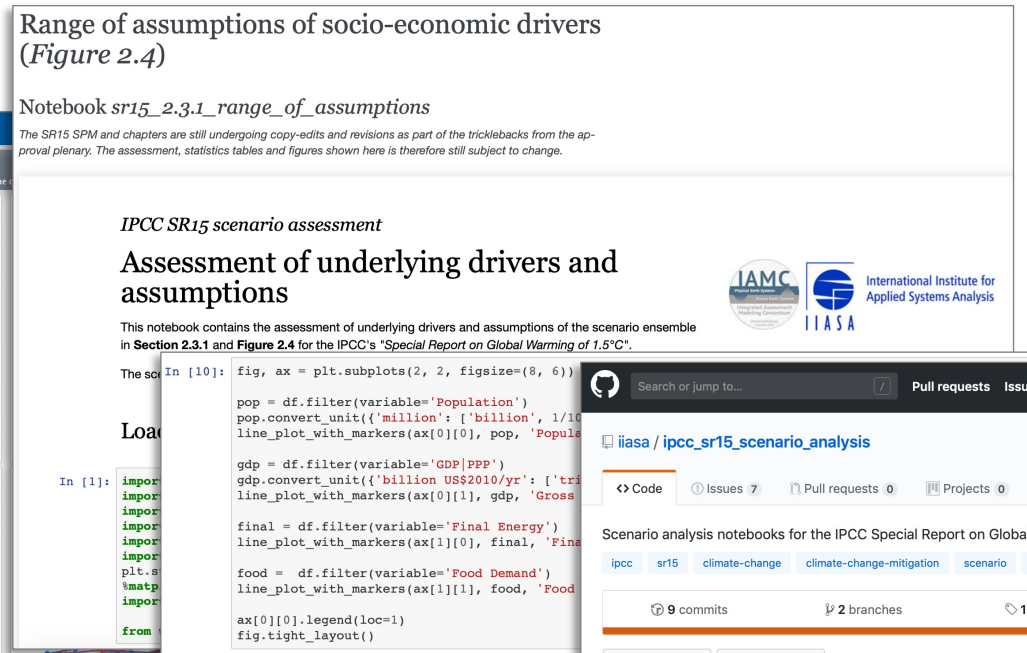
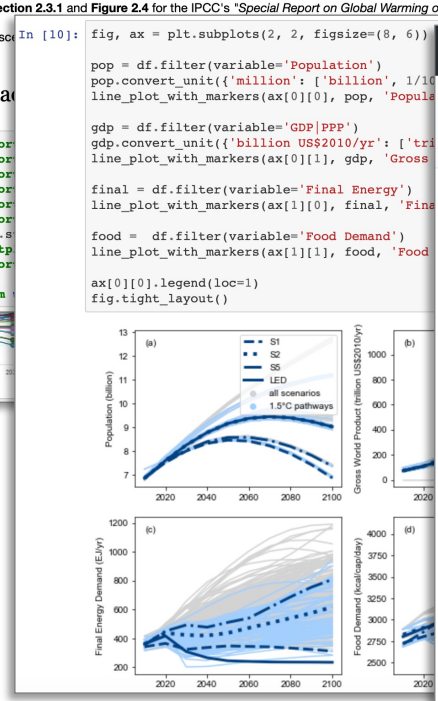
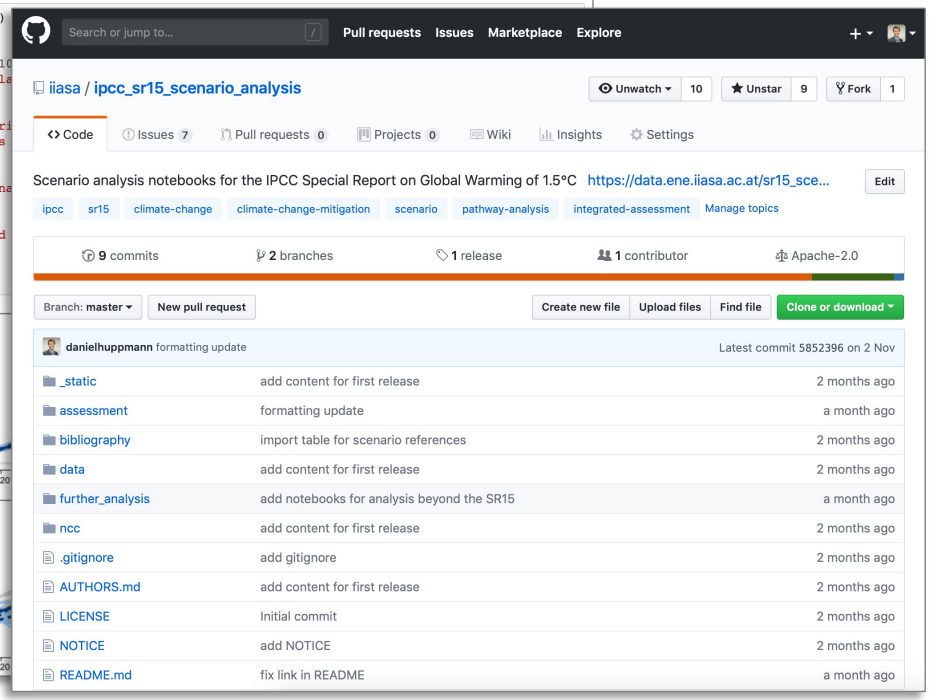


Figure 2.4 | Range of assumptions about socio-economic drivers and projections for energy and food demand in the pathways available to this assessment. 1.5°C-consistent pathways are blue, other pathways grey. Trajectories for the illustrative 1.5°C-consistent archetypes used in this Chapter (LED, S1, S2, S5; referred to as P1, P2, P3, and P4 in the Summary for Policymakers.) are highlighted. S1 is a sustainability oriented scenario, S2 is a middle-of-the-road scenario, and S5 is a fossil-fuel intensive and high energy demand scenario. LED is a scenario with particularly low energy demand. Population assumptions in S2 and LED are identical. Panels show (a) world population, (b) gross world product in purchasing power parity values, (c) final energy demand, and (d) food demand.

Figure 2.4 as printed in the SR15 (www.ipcc.ch/sr15)







Rendered notebooks to generate figures and tables at
data.ene.iiasa.ac.at/sr15_scenario_analysis

```
$ git clone git@github.com:iiasa/ipcc_sr15_scenario_analysis.git
```

Increasing the "FAIRness" of the IPCC assessment

Going beyond efforts in AR5, we followed the FAIR principles to increase transparency and reproducibility of the scenario assessment

Goal	Implemented measures
Findable	Use proper recommended references including DOIs for data and notebooks
Accessible	Make data and notebooks available for multiple levels of user sophistication as well as via common machine-readable API's
Interoperable	Use common data template developed by the IAMC Analysis using open-source Python package pyam
Reusable	Data and assessment notebooks released under licenses that enable follow-up research

Wilkinson, M. D., et al. (2016). Scientific Data 3:160018. doi: 10.1038/sdata.2016.18

Findable

Use appropriate references & metadata for each item

Separate treatment for distinct pieces of the scientific "supply chain"

- Scientific assessment: Chapter 2 of the SR15 and Annex
- Scenario ensemble (data)
- Notebooks for scenario assessment
- Scientific software package
- Journal manuscript on scenario ensemble compilation and user guidelines

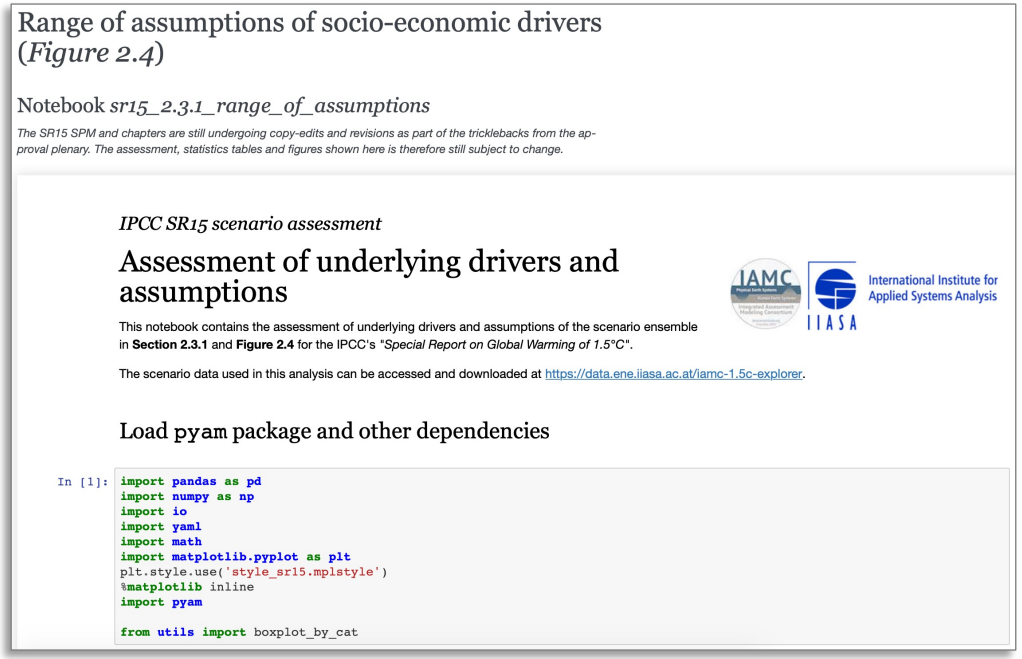
⇒ Each item has its own recommended citation and DOI

⇒ Use proper versioning for each item (data & software release cycle)

Accessible (I) – machine-readable formats

The infrastructure provides multiple entry points & interfaces

- Scenario ensemble data:
  - ⇒ Downloadable as `xlsx` and `csv`
  - ⇒ Accessible via a `RestAPI` from the Scenario Explorer backend
- Assessment notebooks
  - ⇒ Distributed via GitHub
  - ⇒ Also available as rendered notebooks
- Scientific software
  - ⇒ Maintained on GitHub
  - ⇒ Available via `conda` & `pypi`



Range of assumptions of socio-economic drivers (*Figure 2.4*)

Notebook *sr15_2.3.1_range_of_assumptions*

The SR15 SPM and chapters are still undergoing copy-edits and revisions as part of the tricklebacks from the approval plenary. The assessment, statistics tables and figures shown here is therefore still subject to change.

IPCC SR15 scenario assessment

Assessment of underlying drivers and assumptions

International Institute for Applied Systems Analysis

This notebook contains the assessment of underlying drivers and assumptions of the scenario ensemble in **Section 2.3.1** and **Figure 2.4** for the IPCC's "*Special Report on Global Warming of 1.5°C*".

The scenario data used in this analysis can be accessed and downloaded at https://data.ene.iiasa.ac.at/iamc-1.5c-explorer.

Load `pyam` package and other dependencies

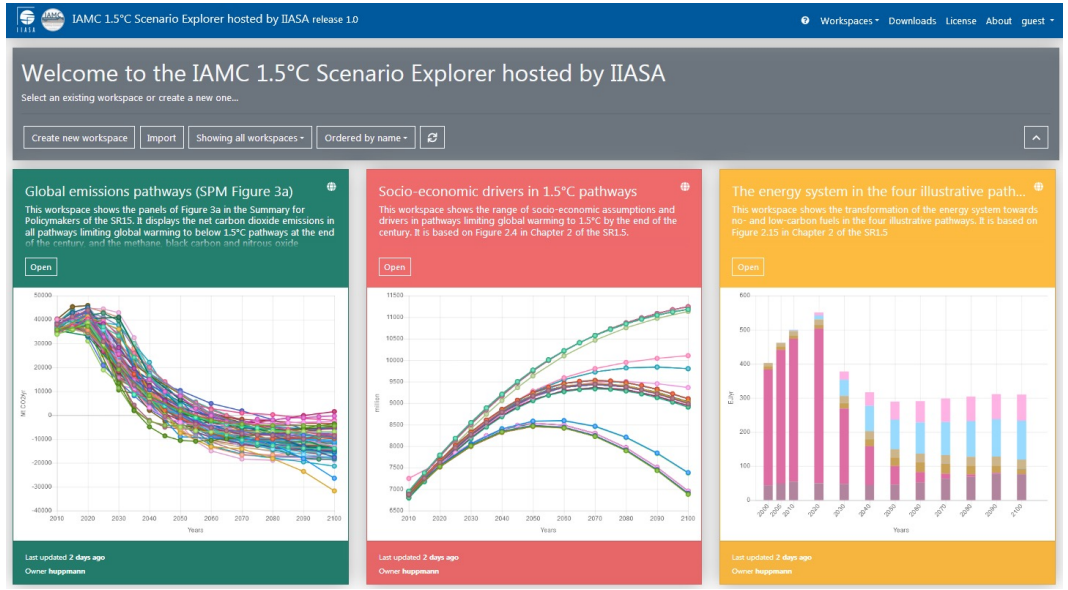
```
In [1]: import pandas as pd
        import numpy as np
        import io
        import yaml
        import math
        import matplotlib.pyplot as plt
        plt.style.use('style_sr15.mplstyle')
        %matplotlib inline
        import pyam

        from utils import boxplot_by_cat
```

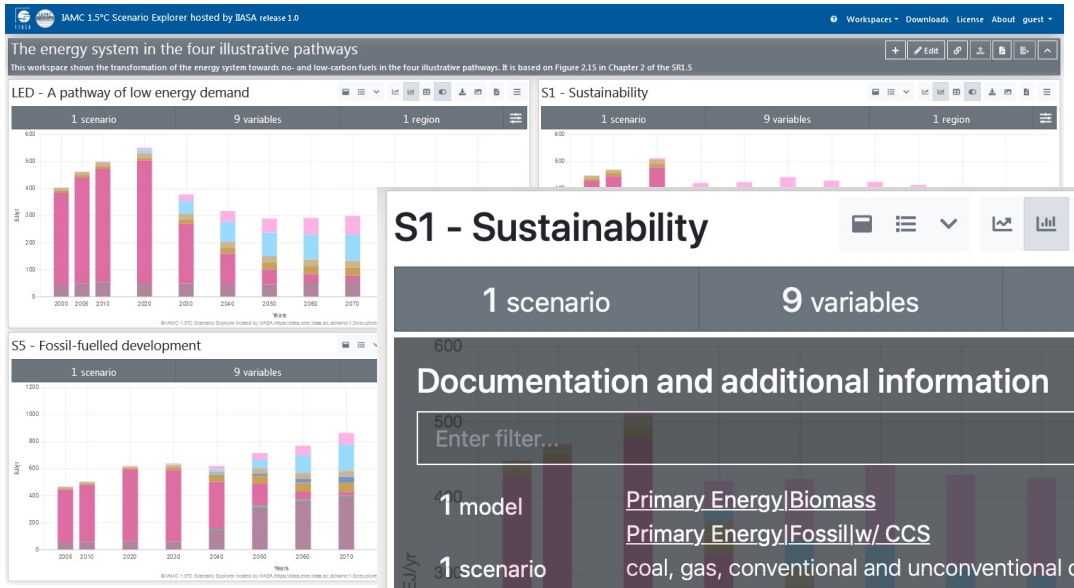
Rendered notebooks to generate figures and tables at https://data.ece.iiasa.ac.at/sr15_scenario_analysis

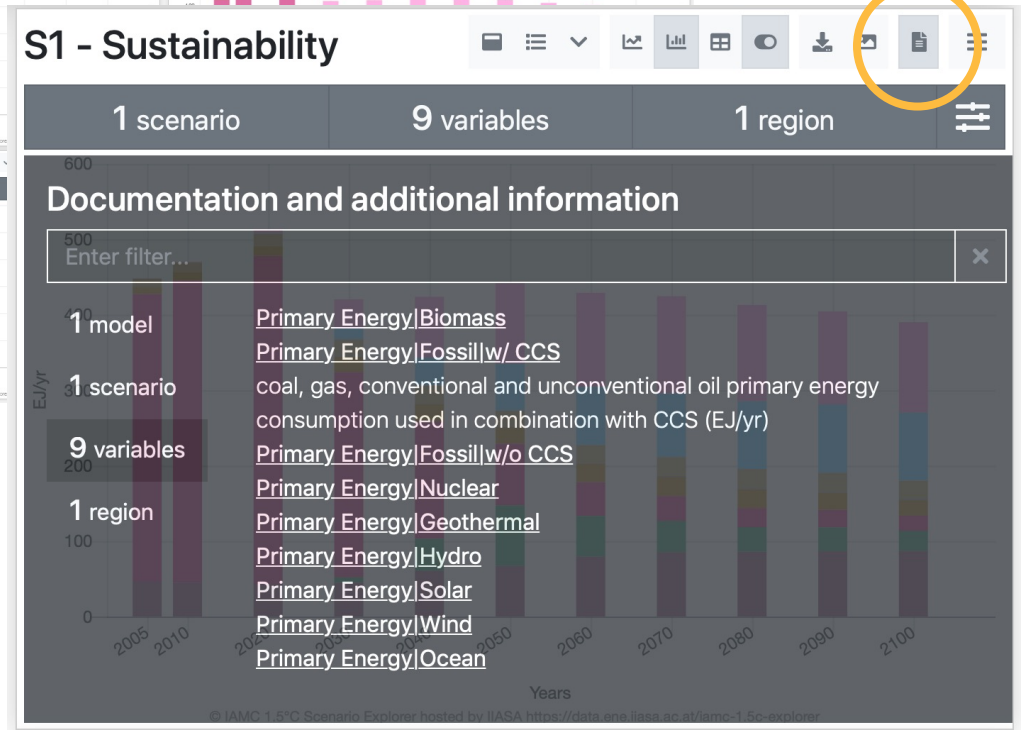
Accessible (II) – for human users

A new "IAMC 1.5°C Scenario Explorer hosted by IIASA"

Using "workspaces" to manage figures & data tables
including pre-defined panels replicating SR15 figures







The scenario explorer provides documentation
and references for models, scenarios & variables

Visit the IAMC 1.5°C Scenario Explorer at
https://data.ece.iiasa.ac.at/iamc-1.5c-explorer

Interoperable

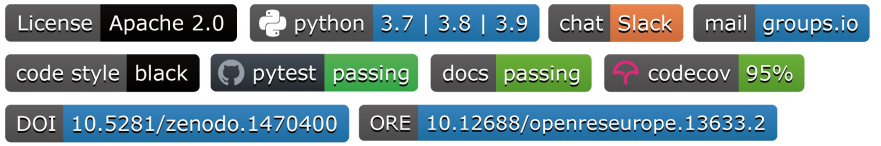
Apply common data standards and open-source packages

- Use common data template developed by the IAMC
  - ⇒ High-profile use case: IPCC Reports (AR5, SR15), EMF
  - ⇒ Used by ~50 research teams globally

IAMC
Integrated Assessment Modeling Consortium
Founded 2007

	A	B	C	D	E	F	G	H
1	**Model**	**Scenario**	**Region**	**Variable**	**Unit**	**2005**	**2010**	**2015**
2	MESSAGE	CD-LINKS 400	World	Primary Energy	EJ/y	462.5	500.7	...

- Assessment using the open-source Python package pyam
  - ⇒ Scenario analysis & visualization toolbox based on collaborative scientific-software practices
  - ⇒ Documentation: pyam-iamc.readthedocs.io



Repository hosted on GitHub
Community supported by Groups.io slack
Documentation hosted by Read the Docs

Reusable (I)

All items of the scientific supply chain are released under licenses that enable follow-up research and re-use

- Scenario ensemble data:
  - ⇒ Custom license modified from Creative Commons CC-BY 4.0
  - ⇒ Aim: allow re-use for scientific research and science communication but keep IAMC 1.5°C Scenario Explorer as "gateway" for entire dataset
  - ⇒ Why? anticipating updates, we want to avoid multiple out-of-sync versions
- Assessment notebooks:
  - ⇒ Licensed under Apache 2.0, distributed via GitHub
- Scenario ensemble manuscript:
  - ⇒ Bound by Springer-Nature policy
  - ⇒ But: distribute Readcube link for free access on personal website and social media, share post-print version on IIASA website after embargo period

Reusable (II)

The scenario set is an unstructured "ensemble of opportunity"

The data was compiled from studies & reports addressing various research questions and based on differing scenario designs and underlying assumptions.

A user's guide to the analysis and interpretation of scenario ensembles

Don't interpret the scenario ensemble as a statistical sample or as likelihood/agreement.

Don't focus only on the medians, but consider the full range over the scenario set.

Don't cherry-pick individual scenarios to make general conclusions.

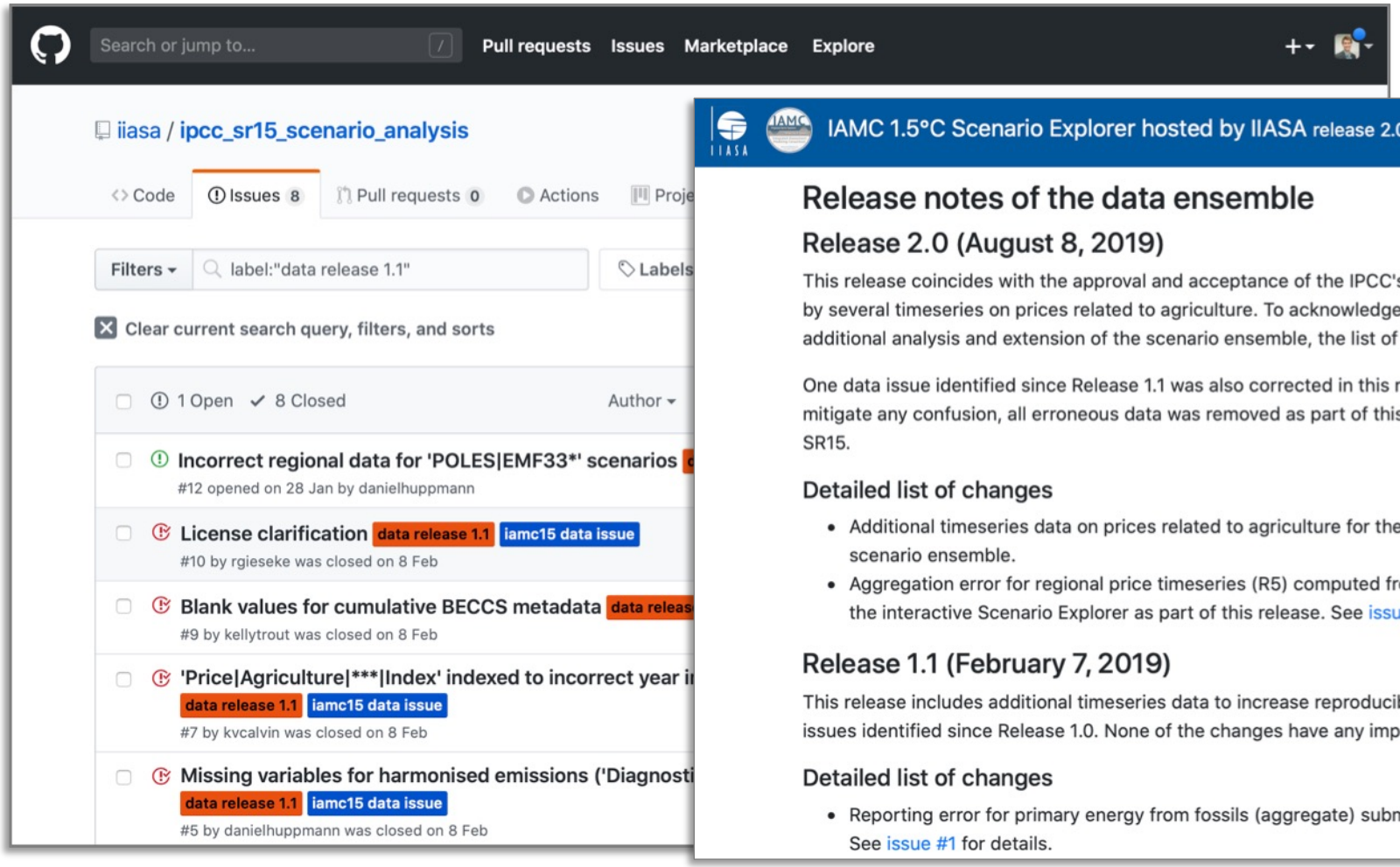
Don't over-interpret scenario results and don't venture too far from the original question.

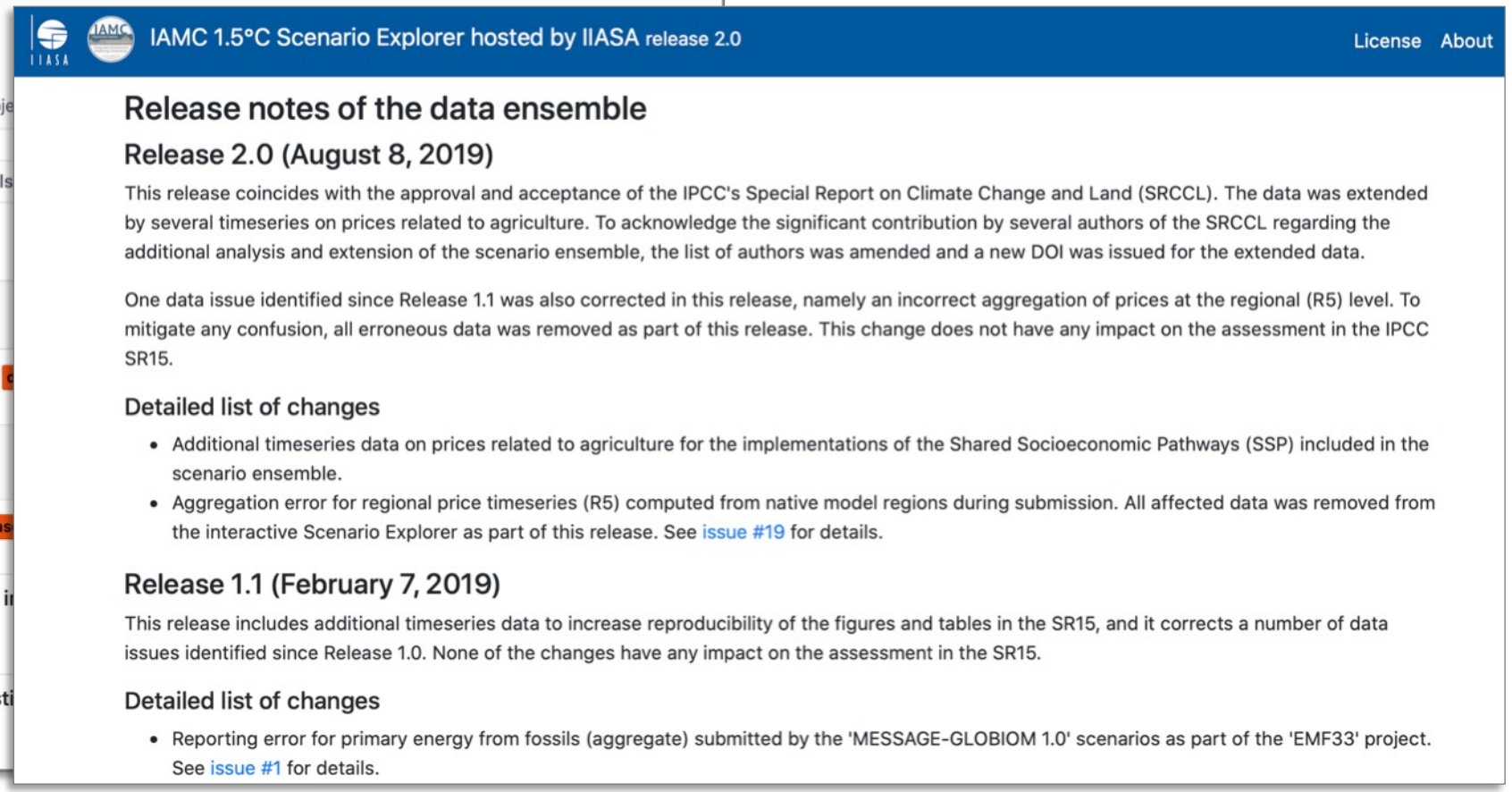
Don't conclude that the absence of a particular scenario (necessarily) means that this scenario is not feasible or possible.

Based on Box 1, Huppmann et al., *Nature Climate Change* 8:1027-1030 (2018).
doi: 10.1038/s41558-018-0317-4 | paywall-free access: rdcu.be/9i8a

Dealing with data errors (after publication)

Using GitHub "Issues" to track errors in the scenario ensemble





See github.com/iiasa/ipcc_sr15_scenario_analysis/issues and data.ene.iiasa.ac.at/iamc-1.5c-explorer/#/about for more information

Outlook AR6: Integration with stylized climate models

Make entire climate assessment workflow in AR6 open & FAIR

- In the IPCC SR15 process, results from integrated-assessment models were passed to stylized climate models to estimate the warming impact
  - ⇒ Scenarios categorized by end-of-century temperature and "overshoot"
- In the past, this was a "black box" for (energy+) modelling teams
  - ⇒ But stylized climate models are becoming open-source tools!
- Current discussions:
  - ⇒ Develop standardized connections to a suite of climate models
  - ⇒ Open the entire emissions harmonization and climate impact workflow
  - ⇒ Add provenance information to the workflow

Part 4

Conclusions & Outlook

Scope for collaboration with EERAdata

Problems with open-source scientific software

There are many concerns that open-source projects deliver sub-par quality compared to closed-source tools

List of drawbacks:

- …?
- …?
- …?

My personal assessment…

⇒ It's just a question of committed resources…

⇒ Overall, the downsides & risks are (pretty much) the same as in a closed-source (commercial or academic) project

Actual issues of open-source scientific software

If the quality of open-source projects depends on resources, how do we make sure that projects get adequate support?

A few ideas on how to improve collaboration:

- ⇒ Make open-source required by funding agencies
- ⇒ Change the expectation in the community
- ⇒ Look around for existing projects rather than start from scratch…

Challenges:

- ⇒ How to get recognition for contributions to other projects?
  This is particulary relevant for early-career researchers…
- ⇒ Open-source doesn’t mean high-quality scientific software

Rationale for best-practice scientific programming

Following best-practice principles in your work will give you more time to do better research

Modelling and scientific analysis is usually a "constant prototyping" exercise

- ⇒ "Just adding one more feature" often breaks existing functionality
- ⇒ Dependencies (open-source packages) change over time
- ⇒ Models and tools are too complex to immediately notice changed behaviour

Who has not yet experienced the panic & stress from a model not solving shortly before a deadline…?

Following best-practice principles…

- ⇒ Guards against models and tools failing to work (as expected)
- ⇒ Helps you to understand *your own thinking* a few months later

Scope for improvements

So far, we focused on (re)usability of scenario analysis
Going forward, we want to support FAIRer users

⇒ Better metadata (entire resource but also individual scenarios)

⇒ Provenance tracking (integration to the „databus" concept)

⇒ Integration & cross-referencing of our „codelists" with existing ontologies

⇒ Inclusion of our database resources in the EERAdata platform

⇒ Use the FAIRification tool developed in EERAdata

Thank you very much for your attention!

Dr. Daniel Huppmann
Research Scholar – Energy, Climate, and Environment Program (ECE)
International Institute for Applied Systems Analysis (IIASA)
Laxenburg, Austria
huppmann@iiasa.ac.at
@daniel_huppmann
https://www.iiasa.ac.at/staff/daniel-huppmann