We don’t need any more open-source models!

Energy Modelling Seminar Series, October 21, 2020
Aalto University

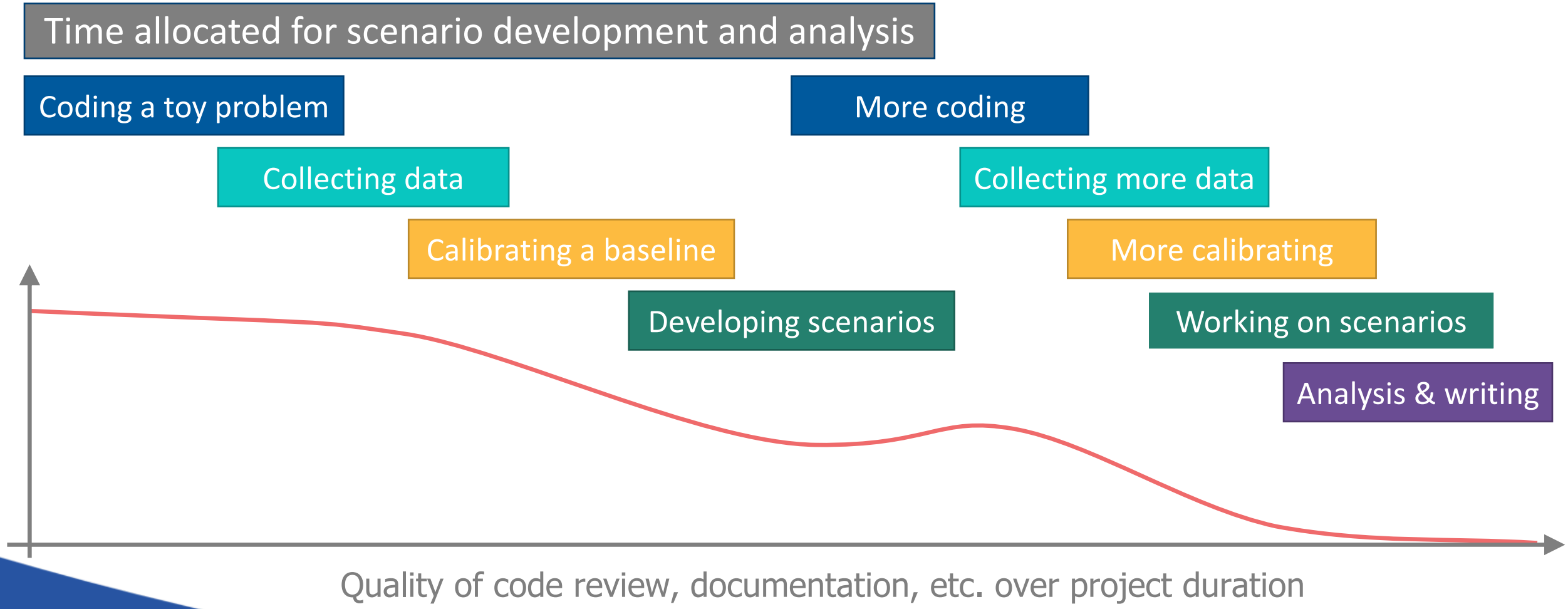
Daniel Huppmann

Please consider the environment before printing this slide deck

Icon from all-free-download.com, Environmental icons 310835, by BSGstudio, license CC-BY

Introduction: a typical modelling project

Open-source tools (can) increase the efficiency of modelling, scenario development, analysis, and writing



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:

- …?
- …?
- …?

⇒ It's just a question of committed resources…

⇒ Overall, the downsides & risks are (pretty much) the same as a close-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

- ⇒ In particular for early-career researchers, how to get recognition for contributions to other projects?
- ⇒ 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

A one-slide guide to open & FAIR research

Even accomplished researchers aren't always up to speed…

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

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?

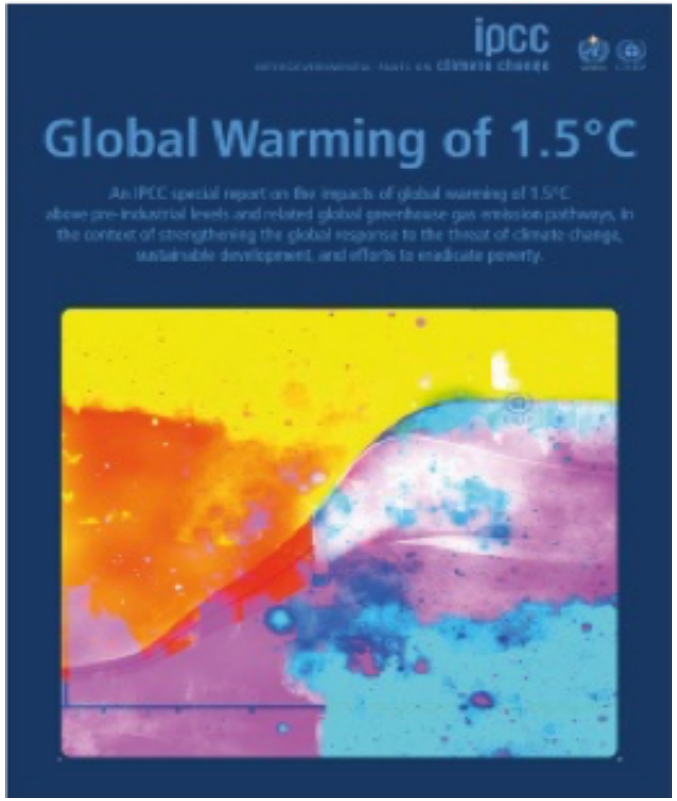
The New York Times

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

An example of open & FAIR science

The IPCC SR15 as a case study of open & FAIR scenario analysis

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

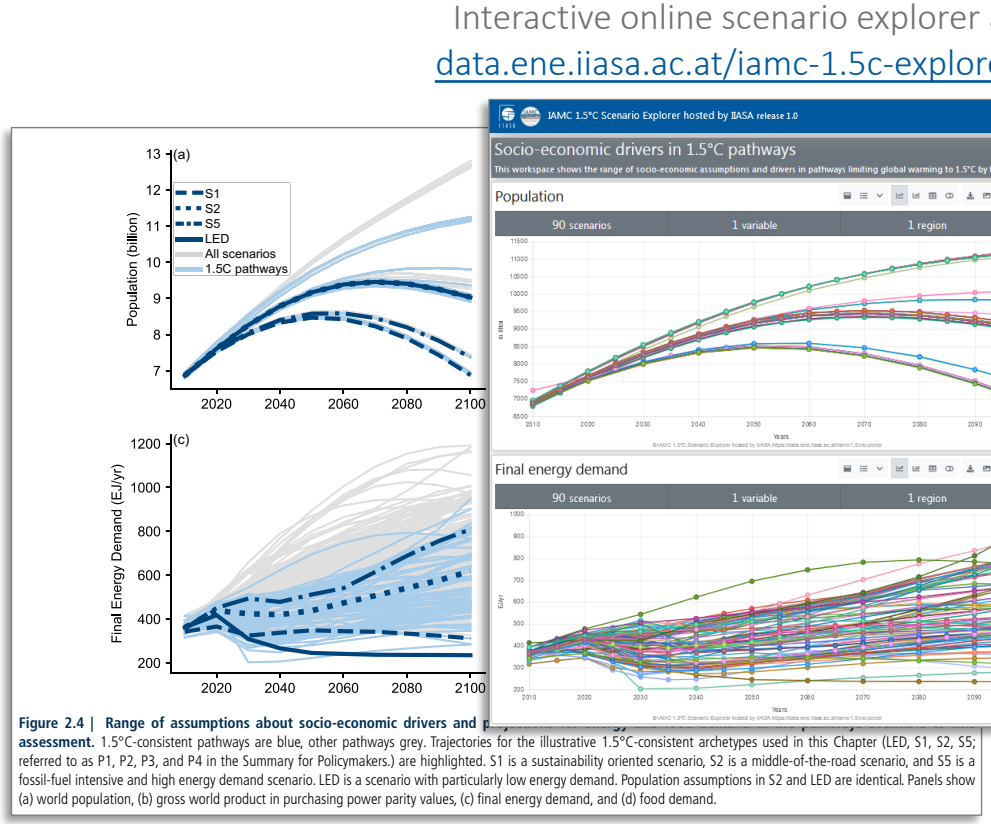
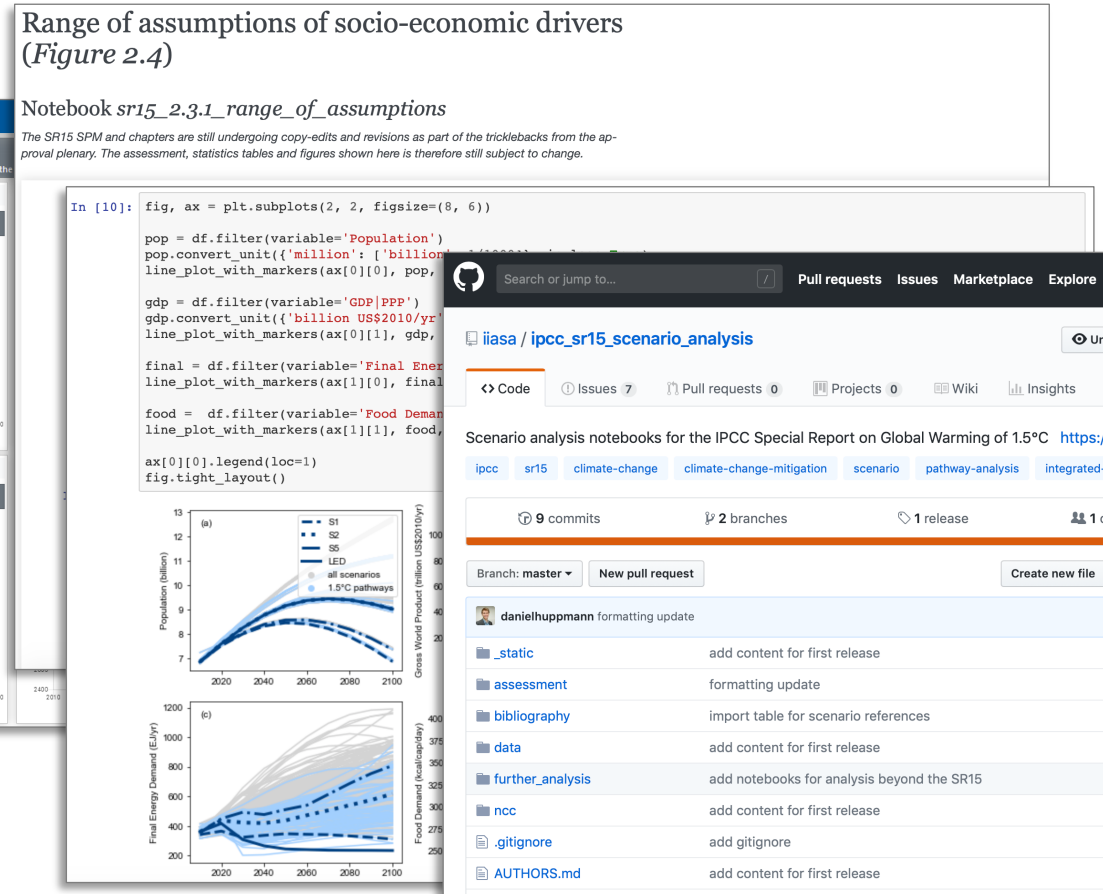


Figure 2.4 | Range of assumptions about socio-economic drivers and p… assessment. 1.5°C-consistent pathways are blue, other pathways grey. Trajectories for the illustrative 1.5°C-consistent archetypes used in this Chapter (LED, S1, S2, S5; referred to as P1, P2, P3, and P4 in the Summary for Policymakers.) 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
```

A toolbox for scenario analysis & visualization

The pyam package enables streamlined processing of results as well as versatile exploration, analysis and visualization tools

Requirement 1:

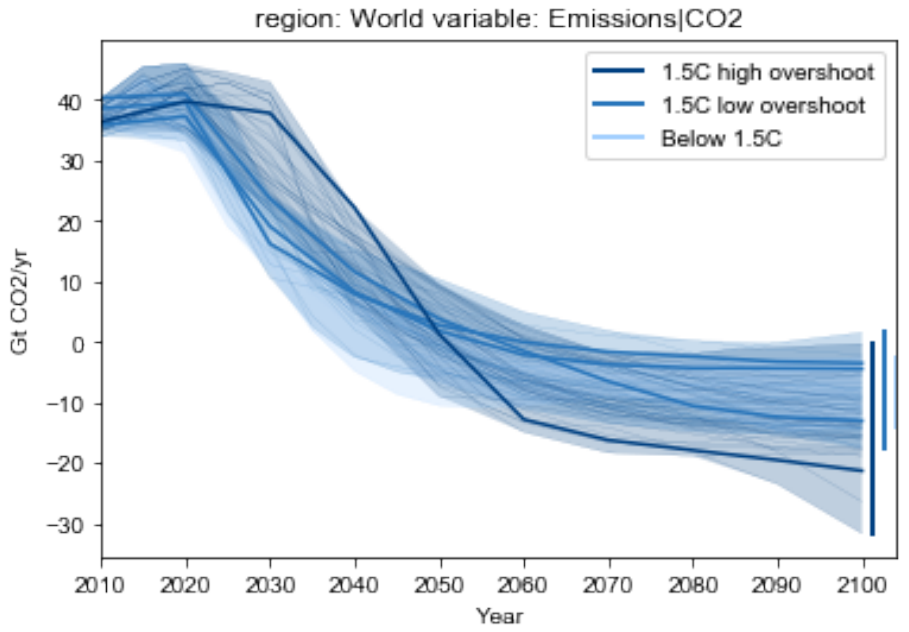
⇒ Standardized, scripted, version-controlled, unit-tested workflow to efficiently get from raw model *output* to processed *results*

e.g., aggregate over sectors & regions using context-specific methods (sum, weighted average, min/max)

⇒ Validation and consistency checking

Requirement 2:

⇒ Exploration and analysis of results in a *reproducible & transparent* manner



Carbon dioxide emissions across 1.5°C pathways
SPM 3a, IPCC SR15 | doi: 10.22022/SR15/08-2018.15428
url: https://data.ene.iiasa.ac.at/sr15_scenario_analysis/

pyam: a Python package for scenario analysis

To facilitate the validation and assessment of scenario results, we developed a dedicated open-source package

Aim: develop a package for scenario analysis & visualization
following best practice of collaborative scientific software development

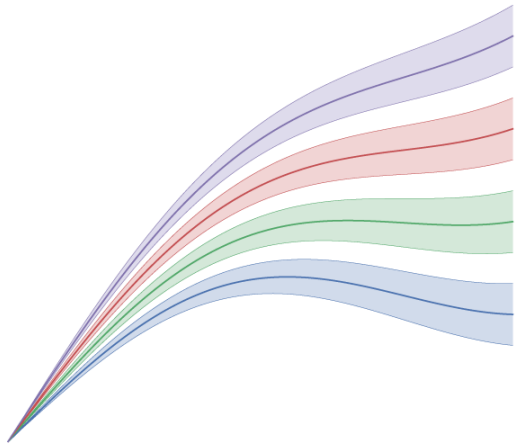
Features:

⇒ Analysis & validation

⇒ Categorization & indicators

⇒ Visualization features

⇒ Statistics package



pyam: analysis and visualization of integrated assessment scenarios

License Apache 2.0 | pytest passing | docs passing | codecov 95%

DOI 10.5281/zenodo.1470400 | JOSS 10.21105/joss.01095

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

pyam-iamc.readthedocs.io

More information:

Documentation: pyam-iamc.readthedocs.io

Scientific reference: M. Gidden and D. Huppmann (2019).
Journal of Open Source Software 4(33):1095.
doi: 10.21105/joss.01095

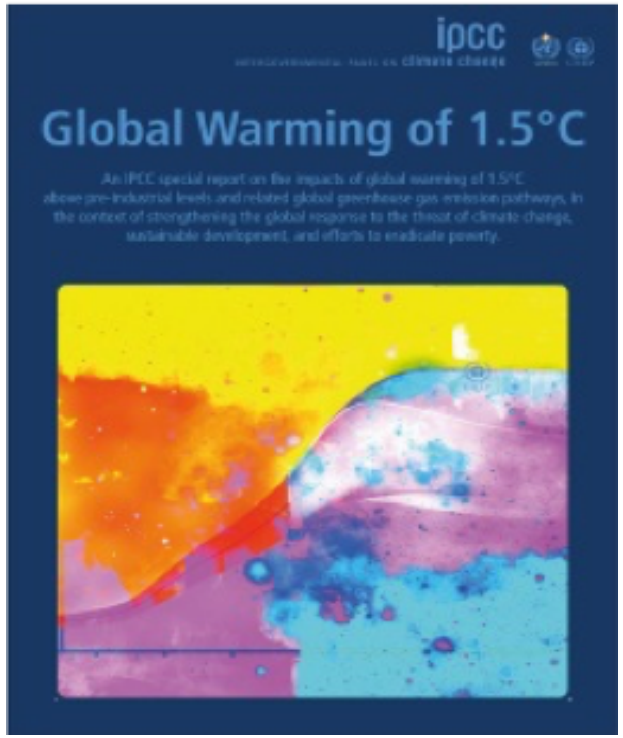
#pyam_iamc

pyam: supporting the IPCC SR15 assessment

Many figures & tables in the SR15 were created with pyam

Check out the open-source notebooks – transparent & reproducible!

The IPCC *Special Report on Global Warming of 1.5°C*



www.ipcc.ch/sr15

Range of assumptions about socio-economic drivers and projections for energy and food demand

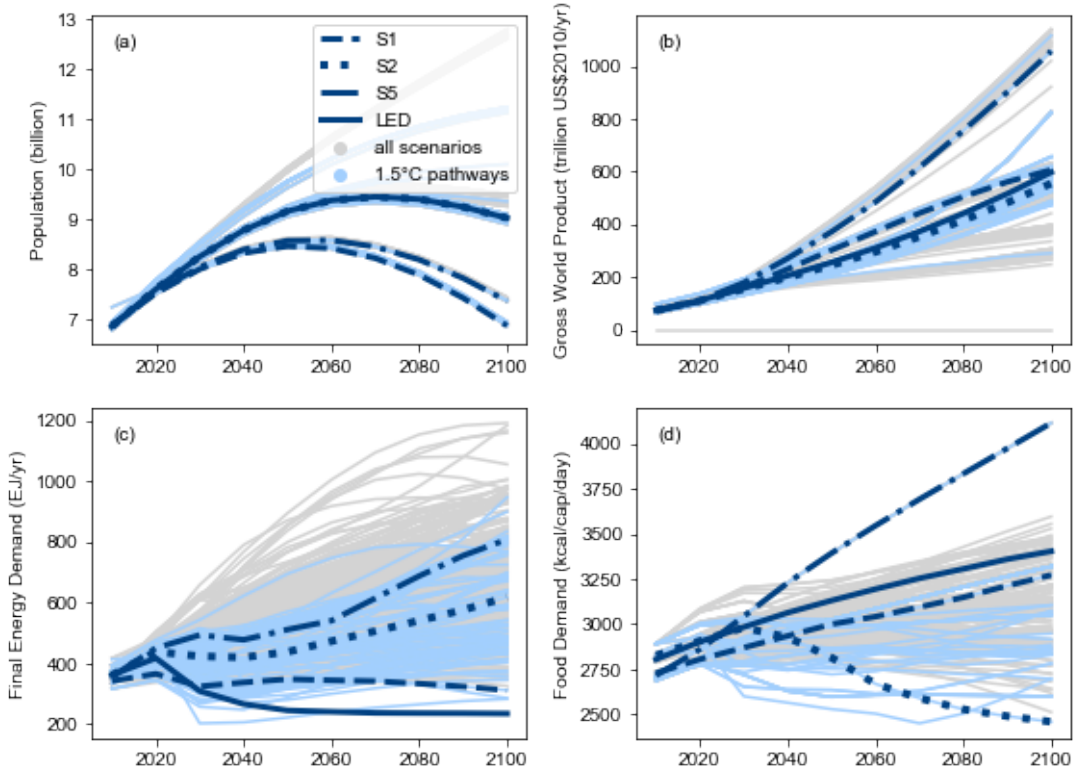


Figure 2.4, page 111, SR15

More information on open tools & resources supporting the IPCC SR15

The key components for model integration

In the openENTRANCE project, we aim to develop an ecosystem of models & tools for decarbonization pathways

Model integration requires a common nomenclature

i.e., naming conventions, definitions, list of regions

⇒ We started a collaborative process on GitHub to facilitate an open discussion and a clear history of changes.

⇒ Implemented as yaml-format dictionary files to balance (human) readability vs. machine processability

```
Final Energy:
  description: Total final energy consumption by all end-use sectors and all fuels,
    excluding transmission/distribution losses
  unit: EJ/yr
```

More info github.com/openENTRANCE/nomenclature

My view on the road ahead for energy modelling

Open science has to go beyond open source…

In an IIASA nexus blog post published a few weeks ago…

⇒ I discuss how open-source scientific software and FAIR data can bring us a step closer to a community of open science

https://blog.iiasa.ac.at/

More information on my lectures at TU Wien (VU 370.062)

https://data.ene.iiasa.ac.at/teaching (including recording of some lectures)

My plea to the audience of this seminar

⇒ Contribute to other work rather than develop new models!

⇒ Follow best-practice of scientific software dev & FAIR principles!

Thank you very much for your attention!

Curious about the pyam package?

- Read the docs on pyam-iamc.readthedocs.io

Want to learn more about the IPCC SR15 scenario ensemble & assessment?

- Read our commentary in Nature Climate Change at https://rdcu.be/9i8a

Dr. Daniel Huppmann
Research Scholar – Energy Program
International Institute for Applied Systems Analysis (IIASA)
Laxenburg, Austria
huppmann@iiasa.ac.at
@daniel_huppmann
http://www.iiasa.ac.at/staff/huppmann

The IAMC template for timeseries data

A community standard for compiling scenario results

Over the past decade, the integrated-assessment community (IAMC) developed a tabular data format for model inter-comparison projects

IAMC
Integrated Assessment Modeling Consortium
Founded 2007

⇒ High-profile use case: IPCC Reports (AR5, SR15)

⇒ Used by ~50 research teams globally

EMF · CD-links · open ENTRANCE · ENGAGE FEASIBILITY OF CLIMATE PATHWAYS

	A	B	C	D	E	F	G	H	I
1	**Model**	**Scenario**	**Region**	**Variable**	**Unit**	**2005**	**2010**	**2015**	**2020**
2	MESSAGE	CD-LINKS 400	World	Primary Energy	EJ/y	454.5	479.6	...	...

It's not a great standard...

⇒ No direct metadata, no sub-annual time resolution (yet), scalability issues, ...

⇒ But it's easy to work with for non-experts, across platforms,
and there is plenty of existing infrastructure to work with this format

The “variable” column

The IAMC data format uses the “variable” column to implement a semi-hierarchical structure

The “variable” column can be used to implement a hierarchical tree

- ⇒ Aggregate: `Primary Energy`
- ⇒ Subcategory: `Primary Energy|Coal`
- ⇒ Further detail: `Primary Energy|Coal|w/CCS`

⇒ The pyam package offers many tools to work with such hierarchical trees

```
df.filter(variable='Primary Energy*', level=1)
df.aggregate(variable='Primary Energy')
```

Read the docs for more information:

⇒ https://pyam-iamc.readthedocs.io/en/stable/data.html

Good practice for scenarios ensemble analysis

As part of the effort supporting the IPCC SR15 assessment, we wrote a list of "do's and don'ts" for model/scenario comparison

A user's guide to the analysis and interpretation of (unstructured) 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 | open-access version: https://rdcu.be/9i8a