International Institute for Applied Systems Analysis
IIASA www.iiasa.ac.at

TU WIEN

Lecture 1:
Principles of open-source and collaborative scientific programming for energy modelling

Open-Source Energy System Modeling
TU Wien, VU 370.062 – summer term 2023

Dipl.-Ing. Dr. Daniel Huppmann

Background: Climate change mitigation and energy system transformation

Following the approval of the IPCC Special Report on Global Warming of 1.5°C, media & newspapers widely quoted required system transformations

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?

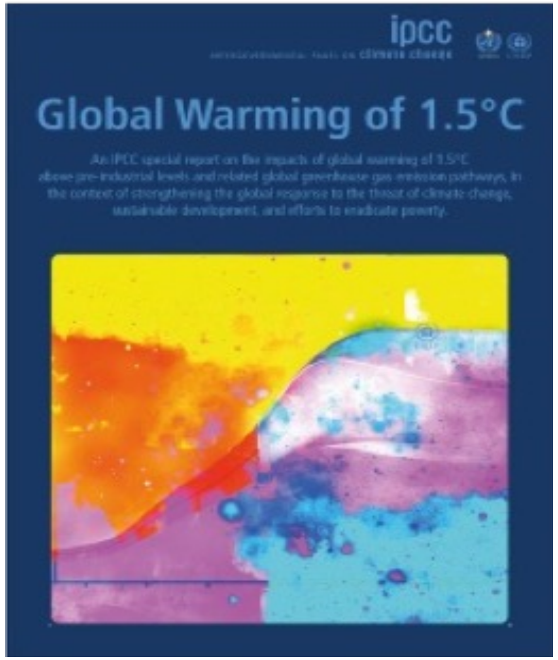
The New York Times

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 2018.
www.ipcc.ch/sr15



About myself: education and research career

From mathematics to energy economics and climate policy

- Dipl.-Ing. (MSc) in Mathematics at TU Wien, specialization *Mathematics in Economics*
- Researcher at the “German Institute for Economic Research” (DIW Berlin)
- Doctorate at TU Berlin in Operations Research, Game Theory and Energy Economics
- Postdoctoral Fellowship at Johns Hopkins University, Baltimore
- Research Fellow at “Resources for the Future” (think-tank in Washington D.C.)
- Research Scholar (since October 2015) at the Energy Program, International Institute for Applied Systems Analysis, Laxenburg
- Contributing Author and Chapter Scientist of the IPCC’s *Special Report on Global Warming of 1.5°C* (SR15) published in October 2018

Overview of the lecture

We will dive into the assessment of energy system transformation pathways while discussing the key concepts of collaborative scientific programming

Content and teaching goals:

- Introduction to scientific programming and open-source software/data (**Lectures 1 &2**)
  ⇒ What is it, why do we it, how do we do it?

- Integrated assessment of climate change & sustainable development (**Lecture 3**)
  ⇒ How can scenarios from these models be used in scientific assessment like the IPCC AR6?
  Using *Jupyter* notebooks and the *pyam* package for scenario analysis

- Development of a national energy system model for policy evaluation (**Lectures 4 & 5**)
  ⇒ How can we develop energy transition scenarios to analyse climate policy measures?
  Using an open-source energy modelling framework

Course structure and lecture content subject to change depending on feedback and interest!

Overview of the lecture (II)

The correct use of collaborative tools and workflows will be as important as the application to a problem and correct interpretation of the results

Requirements:

⇒ A good understanding of energy systems and climate policy

⇒ Experience with at least one scientific programming language

Mode of exercises:

⇒ Submit assignments via GitHub pull requests and Scenario Explorer workspaces

Grade:

⇒ Submitted assignments (50%)

⇒ Written questions and oral discussion of submitted exercises and related questions (30%)

⇒ Active participation in class – feel free to ask questions any time (20%)

About you...

What is your background and experience level with (scientific) programming?

Microsoft Excel as a programming language?

People tend to have strong feelings about Excel...

John Oliver, *Last Week Tonight,* June 5, 2016. Meme from [memegenerator.net](), Clip on [youtube.com]()

Part 1

An introduction to open, collaborative scientific research

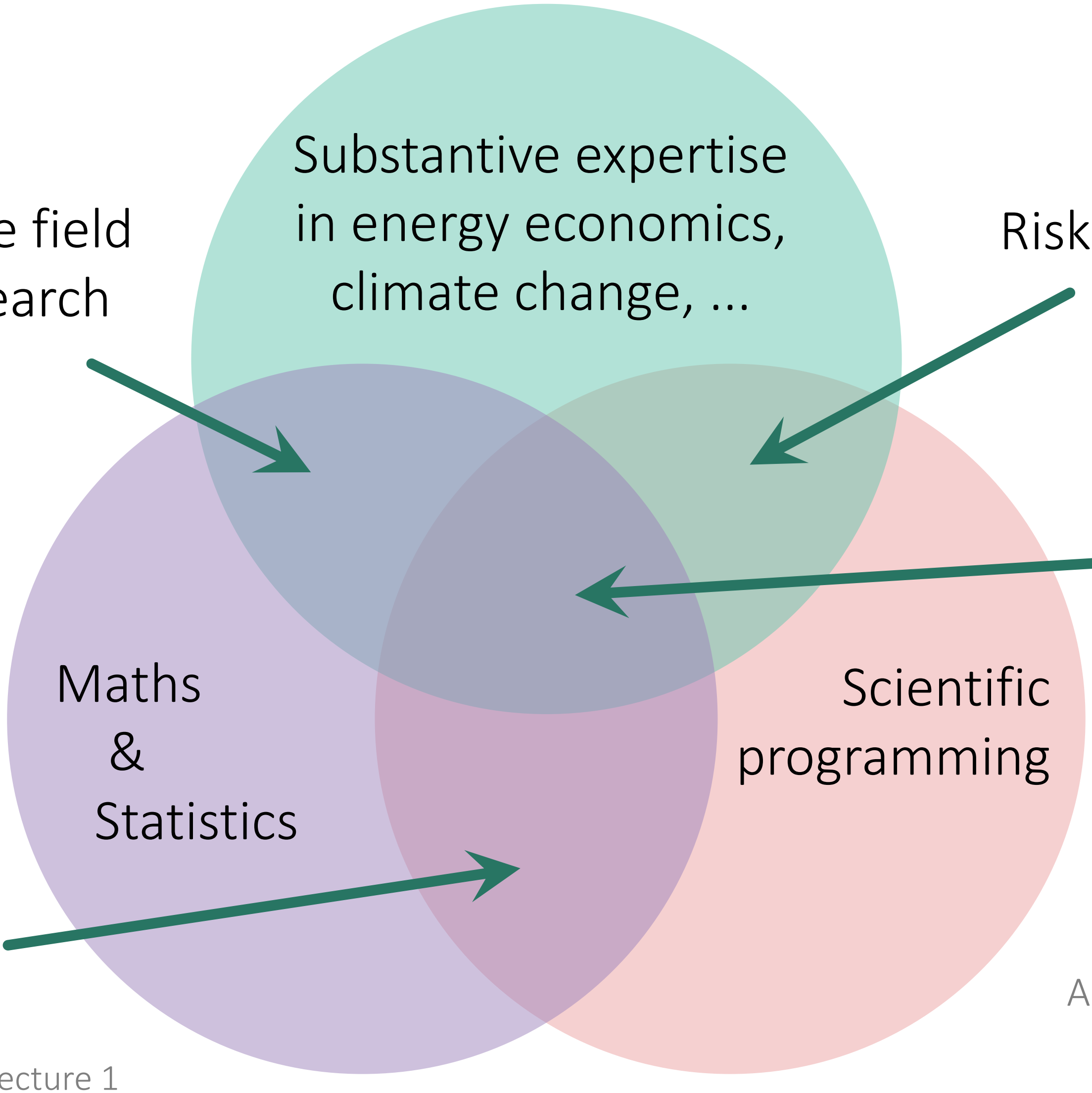
Based on material by Matthew Gidden (@gidden) and Paul Natsuo Kishimoto (@khaeru)

The intersection between energy economics and mathematics

Current research requires substantial expertise in scientific programming

“Traditional” research in the field of applied Operations Research

Risk of invalid conclusions!



This is where the magic happens!

Current data science, machine learning, ChatGPT

Adapted from drewconway.com

Key misconceptions about best practice in open scientific programming

If you think that this topic is of no concern to you, you're probably wrong

- Who is your main (and usually worst) collaborator?
  - ⇒ Yourself from six months ago!
    And you probably didn't write enough documentation and don't respond to emails
- Why is it a bad idea to use data or software that does not have an open license?
  - ⇒ Bad karma!
  - ⇒ Are you intending to distribute your work?
    How are you planning to deal with non-open contributions that your project depends on?
- Why should you share data and code under an open-source license?
  - ⇒ Good karma!
  - ⇒ Standard licenses have a disclaimer of liability, so you cannot be accountable for problems
  - ⇒ There is probably a growing expectation from your (potential) collaborators
  - ⇒ *Treat your GitHub, etc. profile as your "business card" similar to your list of publications*

Background to copyright and license

Every creative work has copyright!

- Per default, any creative work (including software code) attracts copyright
  ⇒ The authors (or the employer) retains all rights on how the work may be used be others

- Exemptions from copyright apply in some cases (journalism, scientific use, teaching)
  ⇒ But in case of doubt, it is the responsibility of a user to argue why an exception applies

- A *patent* grants anyone the right to inspect the work, but not to use it
  ⇒ The "openness" in the sense of inspection is not what we mean by "open-source"

- A *license* grants the right to
  ⇒ Use the work to for specified applications
  ⇒ By a group of users (possibly restricted)
  ⇒ Under certain conditions (including revocation of the right)

Licensing – free and/or open-source software

Freedom in science is not about the price – it's about what you're allowed to do

- *Free software* is not quite the same as *open-source* in the literal sense (inspection)

  defined by "Four Freedoms" →

  ⇒ In practice, the terms are used interchangeably
- An *open-source license* grants the four freedoms to any potential user without restrictions
- Only licenses approved by the Open Source Initiative or Free Software Foundation are considered "open"
- Two classes of free/open software licenses:

  ⇒ *Copyleft*: All modifications must be redistributed under the same open license

  ⇒ *Permissive*: No restriction on redistribution, including the right *not to share* derivative work

Freedom 0: To ***run*** the program for any purpose.
Freedom 1: To ***study*** how the program works, and change it to make it do what you wish.
Freedom 2: To ***redistribute*** and make copies so you can help your neighbour.
Freedom 3: To ***improve*** the program, and release your improvements/modifications to the public.

The first formal definition of free software was written by Richard Stallmann for the *Free Software Foundation*. GNU's Bulletin 1(1):8, February 1986. Via Wikipedia.

To find out which license is appropriate for your project: choosealicense.com

Licensing for (non-software) creative works

Creative Commons provides a suite of licenses for any purpose

„A nonprofit organization that helps overcome legal obstacles to the sharing of knowledge and creativity to address the world's pressing challenges.“ (CC website, What We Do)

- Free cultural works – equivalent to "open-source", attribution required (BY)
  - ⇒ CC-BY – permissive re-use
  - ⇒ CC-BY-SA (share-alike) – copyleft-type license
- Public-domain dedication (no attribution required)
  - ⇒ CC0 (zero) – this is a waiver of all rights
- Additional specifications of CC licenses (not equivalent to "open-source")
  - ⇒ ...-NC – license restricted to non-commercial usages
  - ⇒ ...-ND – no permission to share modified versions of the work (no derivative)

Visit https://creativecommons.org/choose/ to select which license is right for you!

The FAIR Guiding Principles

Existing digital ecosystem of scholarly data publication prevents us from extracting maximum benefit from our research investments

- Good data management and stewardship is not a goal in itself
  ⇒ Rather, it's a pre-condition supporting knowledge discovery and innovation
- Increasingly, science funders, publishers and governmental agencies require data management and stewardship plans for publicly funded research projects
- Digital research objects should be available for **transparency**, **reproducibility** and **reusability**
  ⇒ This includes data as well as algorithms, tools and workflows to compile and assess data
- Data management must be geared towards human readers *and* machine processing
  ⇒ Humans have an intuitive sense of 'semantics' (the meaning or intent of a digital object)
  ⇒ But humans are not able to operate at the scope, scale, and speed required for the scale of contemporary scientific data and complexity

Mark Wilkinson et al. *Scientific Data* 3:160018 (2016) doi: 10.1038/sdata.2016.18

The distinction between FAIR for humans vs. machines

Humans are good at making sense of visual elements…

International Institute for Applied Systems Analysis
IIASA www.iiasa.ac.at

IAMC
Physical Earth Systems
Human Earth Systems
Integrated Assessment Modeling Consortium

IAMC 1.5°C Scenario Explorer and Data hosted by IIASA

Daniel Huppmann, Elmar Kriegler, Volker Krey, Keywan Riahi, Joeri Rogelj, Steven K. Rose, John Weyant, Nico Bauer, Christoph Bertram, Valentina Bosetti, Katherine Calvin, Jonathan Doelman, Laurent Drouet, Johannes Emmerling, Stefan Frank, Shinichiro Fujimori, David Gernaat, Arnulf Grubler, Celine Guivarch, Martin Haigh, Christian Holz, Gokul Iyer, Etsushi Kato, Kimon Keramidas, Alban Kitous, Florian Leblanc, Jing-Yu Liu, Konstantin Löffler, Gunnar Luderer, Adriana Marcucci, David McCollum, Silvana Mima, Alexander Popp, Ronald D. Sands, Fuminori Sano, Jessica Strefler, Junichi Tsutsui, Detlef Van Vuuren, Zoi Vrontisi, Marshall Wise, Runsen Zhang. **IAMC 1.5°C Scenario Explorer and Data hosted by IIASA**. *International Institute for Applied Systems Analysis & Integrated Assessment Modeling Consortium*. (2018) 10.22022/SR15/08-2018.15429

Item Type: Dataset

Archive
iamc15_scenario_data_all_regions_r1.1.xlsx - Published Version (Release 1.1)
iamc15_scenario_data_all_regions_r1.xlsx - Published Version (Release 1.0)

Please access this resource from the **Scenario Explorer Website**. For usage rights please see our license here.

Version History:

Release 1.1 (February 7, 2019)
This release includes additional timeseries data to increase reproducibility of the figures and tables in the SR15, and it corrects a number of data issues identified since Release 1.0. None of the changes have any impact on the assessment in the SR15.

Release 1.0 (October 15, 2018)
Scenario ensemble release for the soft launch of the IPCC SR15 following the approval plenary in Incheon, Republic of Korea.

Please view the About page for details.

Rendered version of the landing page for doi 10.22022/SR15/08-2018.15429

The distinction between FAIR for humans vs. machines

Computers require additional structure to parse information

```
<h1>IAMC 1.5°C Scenario Explorer and Data hosted by IIASA</h1>
<Creator:Author>Daniel Huppmann</Creator:Author>
...
<CreationName:Title><b>IAMC 1.5°C Scenario Explorer and Data
hosted by IIASA</b></CreationName:Title>
<Agent:Publisher><i>International Institute for Applied Systems Analysis
& Integrated Assessment Modeling Consortium</i></Agent:Publisher>.
<DateOfPublishing>(2018)</DateOfPublishing>
<Name:Identifier:DoiName>
<a target="_blank" href="https://doi.org/10.22022/SR15/08-2018.15429">
10.22022/SR15/08-2018.15429</a></Name:Identifier:DoiName>
Item Type: <Type>Dataset</Type>
Please access this resource from the <Digital:Website><a target="_blank”
href="https://data.ene.iiasa.ac.at/ iamc-1.5c-explorer">
<b>Scenario Explorer Website</b></a></Digital:Website>.
```

Source code of landing page for doi 10.22022/SR15/08-2018.15429

The FAIR Guiding Principles (II)

Scientific work should be Findable, Accessible, Interoperable and Reusable

Data and/or metadata...

Findable	• F1. ... are assigned a unique and persistent identifier (Digital Object Identifier, DOI) • F2. ... are described with rich metadata (defined by R1 below) • F3. ... clearly and explicitly include the identifier of the data it describes • F4. ... are registered or indexed in a searchable resource (including Google)
Accessible	• A1. ... are retrievable by their identifier using a standardized protocol • A2. ... are accessible, even when the data are no longer available
Interoperable	• I1. ... use a formal, shared, applicable language for knowledge representation • I2. ... use vocabularies that follow FAIR principles • I3. ... include qualified references to other (meta)data
Reusable	• R1. ... are richly described with a plurality of accurate and relevant attributes: clear data license, detailed provenance, meet community standards

Adapted from Box 2: The FAIR Guiding Principles, Mark Wilkinson et al. *Scientific Data* 3:160018 (2016) doi: [10.1038/sdata.2016.18](https://doi.org/10.1038/sdata.2016.18)

A one-page guide to open & FAIR practices

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

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

Common misconceptions about open & collaborative scientific software

There are many arguments against open-source – almost none are valid

- "I put all my source code/data on my website, so it is open!"
  ⇒ This is only true if you added an approved open-source license
  ⇒ Otherwise, don't use the term *open*, because it can be (mis)understood as *open-source*
- "My code/data is open because I'll just send a copy to anyone who asks"
  ⇒ This is not *open* or *free* according to the common understanding in the community
- "If I make release my code/data under an open-source license, some people may misuse it!"
  ⇒ If you don't make it openly available, nobody is going to use it at all
- "My code/data can't have a DOI because there are proprietary data included..."
  ⇒ The DOI is only attached to the metadata of the object, so there is no problem
- "I can't release my code/data now because I have to clean it first and write documentation"
  ⇒ If that is your approach to scientific programming, you're doing it wrong...

Reproducibility is key to good scientific research (and your own sanity)

Some examples of what's reproducible… not!

Archiving

Definition: Permanent, incorruptible (as far as possible) storage of code, data or results

⇒ Data or results can be preserved, yet may be impossible to recreate (or just understand).

Version control

Definition: VC tracks changes to software source code or data over time.

⇒ VC can be used by one person and yet be unintelligible (i.e., not reproducible) to another.

Testing & quality control

Definition: Implementation of checks to verify that software and data behave as expected.

⇒ Reproducibility of the analysis for one research project doesn't prevent the next researcher from 'breaking' (de-calibrating, misusing) a model or piece of software.

Recommended further reading:

Barnes (2010). Publish your computer code: it is good enough. *Nature* 467(753):775. doi: 10.1038/467753a
Barba (2016). The hard road to reproducibility. *Science* 354(6308):142. doi: 10.1126/science.354.6308.142

The rationale for proper version control tools

In love and in scientific research, there is no such thing as "final"…

Adapted from "notFinal.doc" at "Piled Higher and Deeper" by Jorge Cham, http://phdcomics.com

Required reading and preparation

- Preparation for scientific programming exercises in this lecture:
  - ⇒ create a GitHub account
  - ⇒ know what it means to 'clone' a repository, make a 'commit' and 'push'
  - ⇒ either get familiar with 'git' using the command line
  - ⇒ or get familiar with a program for working with git repos (for novice users, try Gitkraken)
  - ⇒ install Python (for novice users, I recommend that you use Anaconda)
  - ⇒ get familiar with the basic Python syntax

- Required reading:
  - ⇒ The FAIR Guiding Principles, Mark Wilkinson et al. *Scientific Data* 3:160018 (2016) doi: 10.1038/sdata.2016.18
  - ⇒ Greg Wilson et al. Good enough practices in scientific computing. *PLOS Computational Biology* 13(6), 2017. doi: 10.1371/journal.pcbi.1005510

Thank you very much for your attention!

Many thanks to Matthew Gidden (@gidden) and Paul Natsuo Kishimoto (@khaeru)
for sharing their lecture material and experience with collaborative programming

Dr. Daniel Huppmann

Senior Research Scholar – Energy, Climate, and Environment Program

International Institute for Applied Systems Analysis (IIASA)
Schlossplatz 1, A-2361 Laxenburg, Austria

huppmann@iiasa.ac.at

@daniel_huppmann

@daniel_huppmann@mastodon.social

www.iiasa.ac.at/staff/daniel-huppmann