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

Effective strategies for multi-sectoral research using large-scale models

Dr. Paul Natsuo Kishimoto
<paul.kishimoto@iiasa.ac.at>

ISSST 2021 / §ESST6 Integrated assessment and energy system modeling
Wednesday, 23 June 2021 paul.kishimoto.name/2021/06/issst

Energy, Climate, and Environment (ECE) Program
International Institute for Applied Systems Analysis (IIASA)

Outline

- Motivation and concepts
  - Three perspectives on models
  - Validity, reproducibility, interoperability, reuse
  - Costs, resources
- Practices for model-based research
  - Use version control; write documentation
  - Write modular code for reuse
  - Write tests (=internal validity)
  - Automate, towards “continuous reproduction”
- Practices for multi-sector research
  - Separate model-building components
  - Be precise about metrology; use “data interfaces”
- Conclusions

Motivation and concepts

Motivation

- Pursuit of climate change mitigation and other SDGs entails changes in systems—complex, large, interconnected, open, sociotechnical.[1]
- Quantitative computer models[2] are used to study these systems.
- Changes trouble the boundaries between 'sectors' of human/economic activity:
  - Changes large enough that feedbacks from other sectors are non-negligible.
  - New technologies establish new interactions, e.g. electric vehicle–grid interoperation.
- We are motivated to connect sets ($N \geq 2$) of models or increase complexity to study changes in multiple subsystems at once.

[1] or "CLIOS", per Mostashari and Sussman 2009.

[2] aka. analyses, workflows, tools, scripts.

Internal vs. external validity

Concerns for scientific modeling & scenario research

Internal validity. Research is free of errors:

- Correctly implements theory w/o conceptual errors.
- Confounding variables addressed to identify relationships between independent and dependent variables.
- Alternative hypotheses can be rejected.

External validity. Research is generalizable to other conditions:

- Research can be replicated or reproduced in a different context.
- Research is robust to differences between the study context and other contexts to which conclusions are applied.
- Research is robust to plausible alternatives to key assumptions.

What is a model? I

Three perspectives and resulting insights

A knowledge object that embodies or represents a theory or understanding of some real-world phenomenon.

- Theories often causal.
- Relationships expressed quantatively: equations connecting variables representing concepts measured in certain, systematic ways.
- In large-scale integrated assessment, systematized concepts often aggregate: GDP, country, sector.

What is a model? II

Three perspectives and resulting insights

A scientific instrument[3] that is used to perform experiments: "What would be the outcome (effect on quantity Y) if X were changed from x_1 to x_2?"

- Another instrument: the Large Hadron Collider (LHC).
  - EUR 7.5 billion budget; labour from many specialized roles.
  - Components for preparing the experiment, running it, measuring outcomes are carefully designed, constructed, tested.
- Instruments require meticulous attention to detail.
- Description of methods includes description of instruments, so the experiment can be reproduced.

What is a model? III

Three perspectives and resulting insights

A software project in which people in organizations create code that is run on computer systems.

- All software has bugs; all organizations have politics.
- Software is constantly evolving and never complete.
- Tendency to overinvest time in new code vis-à-vis quality & docs.
- "Technical debt": code grows stale over time.

But! good software development practices exist, and are widely used to ensure that software meets needs.

[3] O'Malley 2019.

Validity and reproducibility

Since the model is not the real world, implications drawn from modeling results must be externally valid. Specific threats, as forms of uncertainty:

Structural Is the theory a correct description of the phenomena?
$\rightarrow$ Response: alternate model formulations.

Measurement uncertainty of input data and parameters.
$\rightarrow$ Sensitivity analyses, large ($> 10^3$) ensembles of model runs.

Epistemic uncertainty in conditions (e.g. future policy) that are unknowable, or whereof uncertainty cannot be quantified.
$\rightarrow$ Alternate scenarios.

All require a quality instrument that can be reused in an easy, automated manner, giving the same results every time—a **reproducible** model.

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

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

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

"… Interoperable & Reusable"

Are these true *in principle* or *in practice*?

- Easy to claim that others could, in principle, re-use elements of a model-based research workflow.
- Much harder: make this an actual practice, i.e. feasible with resources ≥1 others have.

Even achievable reproduction is not true reusability.

- Equity & inclusion require that analytical tools and capabilities be broadly distributed.
- Not adequate that researchers from LMICs join well-resourced incumbent modeling teams, if these remain central.
- Urgency of climate change & SDGs requires we bring more hands to the work.

Costs & resources

We have finite resources (time etc.) with which to conduct research. Work to create and use models should spend resources efficiently.

Search & information

- How do I run the model? What does this line of code do?
- What about student S, who did … 2 years ago—where is that?
- What version of the model produced results for this 1 y/o manuscript?

Quality control & enforcement

- When/why did our reference forecast shift in region *r* & sector *g*?
- Who broke the model so Policy Z no longer has a feasible solution?

Recovery/disruption

- If colleague C left tomorrow, could we continue our work?[4]

[4] aka. the bus factor or truck number.

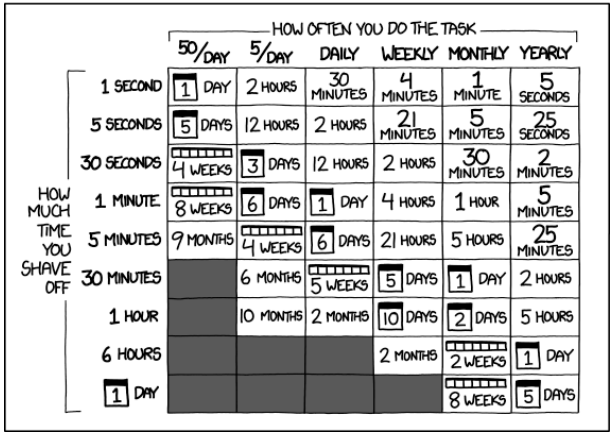
None of this is new

or, standing on the shoulders of giants

Reproducibility crisis in quantitative social sciences, e.g. psychology.

Computing as fundamental to valid research: atmospheric & climate sciences, engineering disciplines (cf. Barba et al.; see appendix), basic sciences.

Most practices from software industries; minor adjustments.



https://xkcd.com/1205/

Practices for model-based research

Examples from MESSAGE*ix*-GLOBIOM

MESSAGE*ix* Generalized formulation for an energy-economic/integrated assessment LP model.

`https://docs.messageix.org`

`https://github.com/iiasa/message_ix`

MESSAGE*ix*-GLOBIOM A family of global and single-country models and variants built in this framework.

`ixmp` Data storage backend & solver interface.

… related tools for data, etc.

Caveats:

- We aren't perfectly 'OFAIR' yet. This talk = mix of status & goals.
- We celebrate continual improvement & efforts of others to improve.

Strategy and practice for modeling

Start with organizational culture:

- Discuss and identify in your team where resources are spent.
- Look for possible improvements in practice.
- Agree that there is a positive return on investment.

A collection of interlinked and mutually-reinforcing practices.

- "A truly remarkable variety of practices, but these slides are too few to contain them."
- Can be adopted separately, incrementally.
- Also corresponding skills $\rightarrow$ support skills development in your team.

Motivation and concepts
- Three perspectives on models
- Validity, reproducibility, interoperability, reuse
- Costs, resources

Practices for model-based research
- Use version control; write documentation
- Write modular code for reuse
- Write tests (=internal validity)
- Automate, towards “continuous reproduction”

Practices for multi-sector research
- Separate model-building components
- Be precise about metrology; use “data interfaces”

Conclusions

The basics: version control, docs

Use version control.

- `https://github.com/iiasa/message_ix`
- Use fewer, larger, better-organized repositories.
- Learn and use the GitHub workflow.

Write (and read) documentation.

- The #1 audience for this is *you*, or your closest collaborators.
- Rubber duck & pair program: explain to a duck what the code does.
- Use services like Read The Docs: automate build & publish steps for every change to the code.[5]

[5] Recent builds of the MESSAGE*ix* docs.

Write modular code for reuse

Common to have a variety of tasks in one (very long) "script" (or a few):

- Input data processing, assumptions, bits of methods adopted from literature, core methods/workflow, post-processing/plotting, output, logging…

Instead, and from the start:

- Separate concerns: 1 task per code object; files group related tasks.
- DRY: don't repeat yourself. Write & reuse functions and classes $\rightarrow$ fewer occasions for error.
- Smaller, atomic functions & classes are easier to document, understand, and validate.[6]
- New data, methods, etc. can be easily swapped-in.

[6] Often can be discarded in favour of high-quality, performant ones from popular libraries; read the docs!

Write tests for internal validity

Software tests = code that runs other code, giving a "pass" or "fail" result.

```
def test_stock(dummy_data):
    observed = compute_vehicle_stock(dummy_data)
    expected = 42.1
    assert observed == expected
```

Code that implements core theory/methods can be tested for a variety of inputs ≡ checks of internal validity.

MESSAGE*ix*: 100s of tests from basic (data I/O) to complex functionality (LP constraint relaxation; dynamic penetration of new technologies…)
→ github.com/iiasa/message_ix/tree/main/message_ix/tests

Automate

towards "continuous reproduction"

Continuous integration (CI) services:

- Watch a code repository, e.g. on GitHub, for changes.
- Automatically grab new versions.
- Perform certain actions, e.g. run a suite of tests.

Tests cover *all core methods* in a model → CI reduces work to guard against invalidity when improving models ('reggressions').

Code includes *all steps* in a model-based analysis → CI system can continuous confirm reproducibility.
Example: A tutorial notebook from MESSAGE*ix*.

- Constructs and solves a simple energy system model.
- Full-scale models currently private (proprietary data).

Practices for multi-sector research

Motivation and concepts

- Three perspectives on models
- Validity, reproducibility, interoperability, reuse
- Costs, resources

Practices for model-based research

- Use version control; write documentation
- Write modular code for reuse
- Write tests (=internal validity)
- Automate, towards “continuous reproduction”

Practices for multi-sector research

- Separate model-building components
- Be precise about metrology; use “data interfaces”

Conclusions

Separate model-building components

Separate code that prepares a "base" model from code that adds/alters detail & resolution related to a particular phenomenon or sector.

← instance of the global MESSAGE*ix*-GLOBIOM model

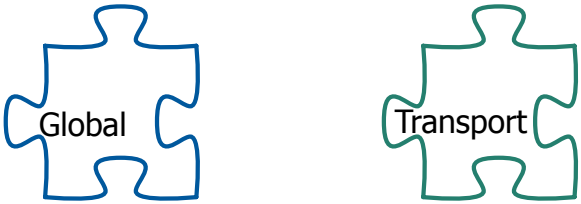
Separate model-building components

Separate code that prepares a "base" model from code that adds/alters detail & resolution related to a particular phenomenon or sector.

← instance of the global MESSAGE*ix*-GLOBIOM model

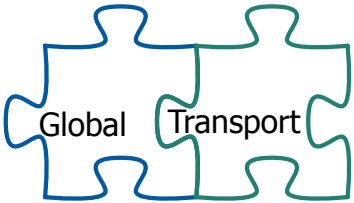
Separate model-building components

Separate code that prepares a "base" model from code that adds/alters detail & resolution related to a particular phenomenon or sector.



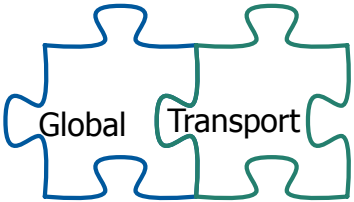
Separate model-building components

Separate code that prepares a "base" model from code that adds/alters detail & resolution related to a particular phenomenon or sector.



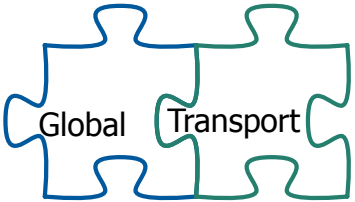
← global model + added transport-sector resolution

Separate model-building components



Each of these pieces is under continual development by separate teams of researchers. This *could* entail frequent and laborious adjustments.

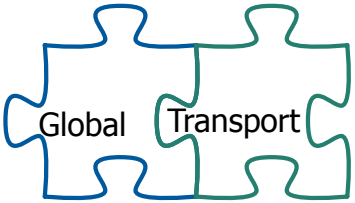
Separate model-building components



Each of these pieces is under continual development by separate teams of researchers. This *could* entail frequent and laborious adjustments.

Modularity + testing ensure that the "shape of the piece" (structure and data of the model prepared by some code) is stable.

Separate model-building components

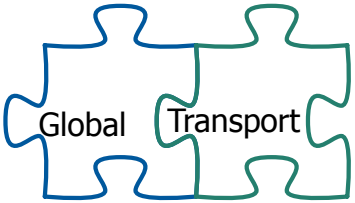


Each of these pieces is under continual development by separate teams of researchers. This *could* entail frequent and laborious adjustments.

Modularity + testing ensure that the "shape of the piece" (structure and data of the model prepared by some code) is stable.

$\rightarrow$ base "Global" model presents the same shape.

Separate model-building components



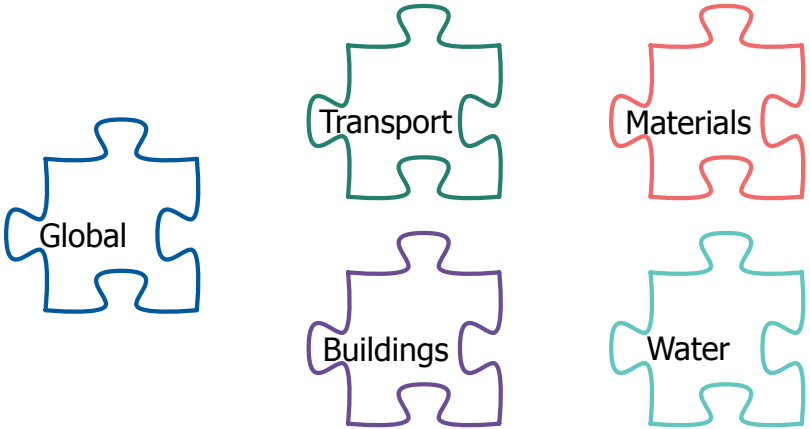
Each of these pieces is under continual development by separate teams of researchers. This *could* entail frequent and laborious adjustments.

Modularity + testing ensure that the "shape of the piece" (structure and data of the model prepared by some code) is stable.

→ base "Global" model presents the same shape.
→ code that configures the "Transport" variant works on anything that has this shape.

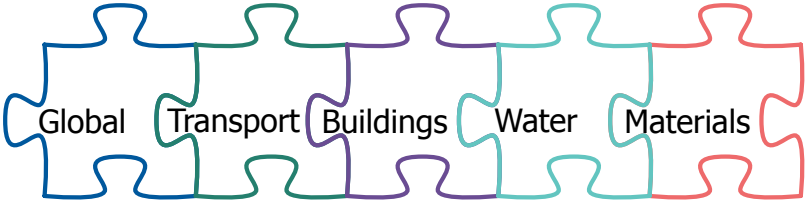
Separate model-building components

Pieces can "be plugged in" to any base or enhanced model, so long as it presents the right shape ≡ valid models can be composed with details required for particular studies.



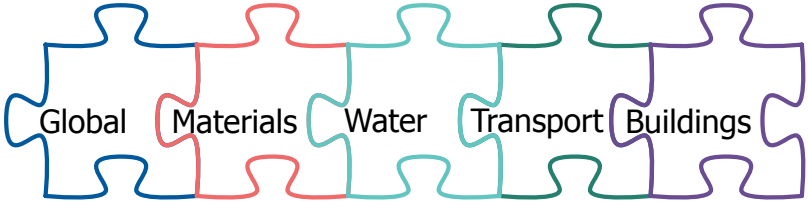
Separate model-building components

Pieces can "be plugged in" to any base or enhanced model, so long as it presents the right shape ≡ valid models can be composed with details required for particular studies.



Separate model-building components

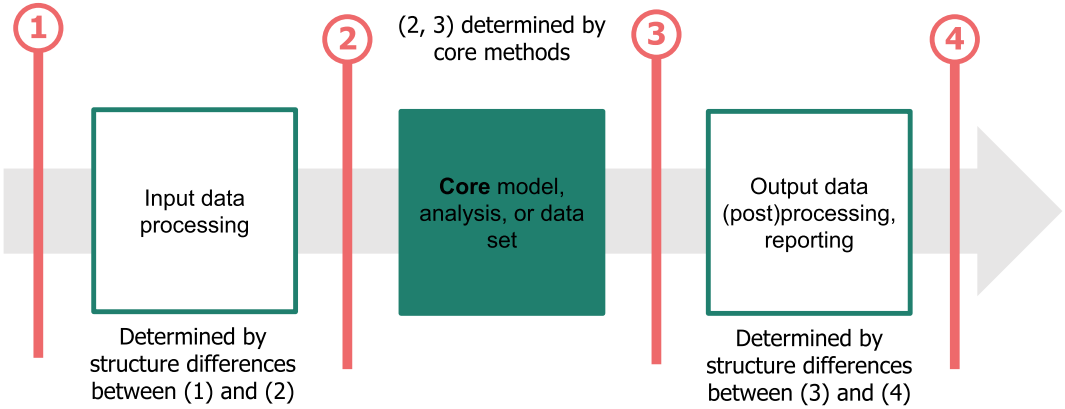
Pieces can "be plugged in" to any base or enhanced model, so long as it presents the right shape ≡ valid models can be composed with details required for particular studies.



Our implementation: in the `message-ix-models` package.

Use precise metrology I

Specify data flows separately from methods:



These form another kind of interface and help towards interoperability.

Use precise metrology II

At each interface (1) through (4) be precise about:

- Background vs. systematized concepts vs. specific measures.[7]
- Dimensions, and the specific codes[8] used along each.
- Units of measurement. (Check with Pint or similar.)

Treat all assumptions as input data → none in code.

Don't invent new data formats:

- Reuse existing formats and protocols for exchange
  e.g. SDMX (1, 2), NetCDF, Zarr, etc.
- Reuse existing (or shared) codes, categorizations, and labels
  e.g. ISO 3166-1; SDMX global registry.

[7] Adcock and Collier 2001.

[8] e.g. `Canada` vs. `CAN` vs. `CA`; read more.

Conclusions

Conclusion: back to costs I

Not mentioned earlier: cost of disobeying incentives.

Some incentives that can affect us as model-builders and -users:

- Publish; only work that can be claimed 'novel', and only when final.
- Signal compliance with disciplinary norms with minimal effort.
- Assist only collaborators / co-authors; neglect others.
- Don't budget for maintenance and support.
- Value 'Impressive' polish, GUIs, and ease of rudimentary use…
  …over 'mundane' validation and reducing I & R costs.

Conclusion: back to costs II

In contrast, free software—thus also open science—succeeds when:

- Communities work together to build a smaller number of higher-quality projects that are public goods.
- Innovation is planned & done out in the open.
- Support, documentation, and enabling others' contributions is first-class, valued work.

$\rightarrow$ not the same activities, shared *ex post*; but a dramatic change in norms.

Thank you!

Appendix

References & further reading I

- Adcock, R et al. (Sept. 2001). "Measurement Validity: A Shared Standard for Qualitative and Quantitative Research". In: *American Political Science Association* 95.3 (03), pp. 529–546. ISSN: 1537-5943. DOI: 10.1017/S0003055401003100.
- Huppmann, D et al. (Apr. 2020). "Five best-practice steps to make your research open & FAIR v1.0". Laxenburg, Austria. DOI: 10.22022/ene/04-2020.16404.
- Irving, D (2016). "A Minimum Standard for Publishing Computational Results in the Weather and Climate Sciences". In: *Bulletin of the American Meteorological Society* 97.7, pp. 1149–1158. DOI: 10.1175/BAMS-D-15-00010.1.
- Mostashari, A et al. (2009). "A Framework for Analysis, Design and Management of Complex Large-Scale Interconnected Open Sociotechnological Systems". In: *International Journal of Decision Support System Technology* 1.2, pp. 53–68. DOI: 10.4018/jdsst.2009040104.

References & further reading II

- O'Malley, M (May 23, 2019). "Is low carbon energy system research an excuse to solve bigger equations?" Energy Program seminar (International Institute for Applied Systems Analysis (IIASA)). Laxenburg, AT.
- Pauliuk, S et al. (2019). "A general data model for socioeconomic metabolism and its implementation in an industrial ecology data commons prototype". In: *Journal of Industrial Ecology* 23.5, pp. 1016–1027. DOI: 10.1111/jiec.12890.

- L. Barba group @ GWU SEAS: r13y syllabus w/readings on research group website; **barba-2017**.
- Other disciplines: Irving 2016, Pauliuk et al. 2019.
- Max Planck Institute for Meteorology "Good scientific practice" policy, rules, forms.
- Christensen & Miguel (2016), "Transparency, Reproducibility, and the Credibility of Economics Research" forthcoming in *JEL* — UC Berkeley Econ.

References & further reading III

- Nick Barnes: "Publish your computer code: it is good enough" in *Nature News* — Climate Code Foundation.
- 45+ more peer-reviewed articles and other resources.

Colophon

PDF and abstract: paul.kishimoto.name/2021/06/issst
LaTeX source, copyright, & license: github.com/khaeru/doc