Fluid: Towards Transparent Science Communication

Institute of Computing for Climate Science

intelligible openness

"Open" should mean **accessible**, **intelligible** and **contextualised**, not just available for download

Today, outputs are disconnected from the pipelines used to create them

Notebooks help, but push the burden onto authors

AI can help reconstruct things after the fact, but how can we rely on this?

Can we help **authors** tell stories able to reveal their own supporting evidence?

For:

- Data journalism
- Publishing
- Distillation

disconnect between **"data"** and **"discourse"**

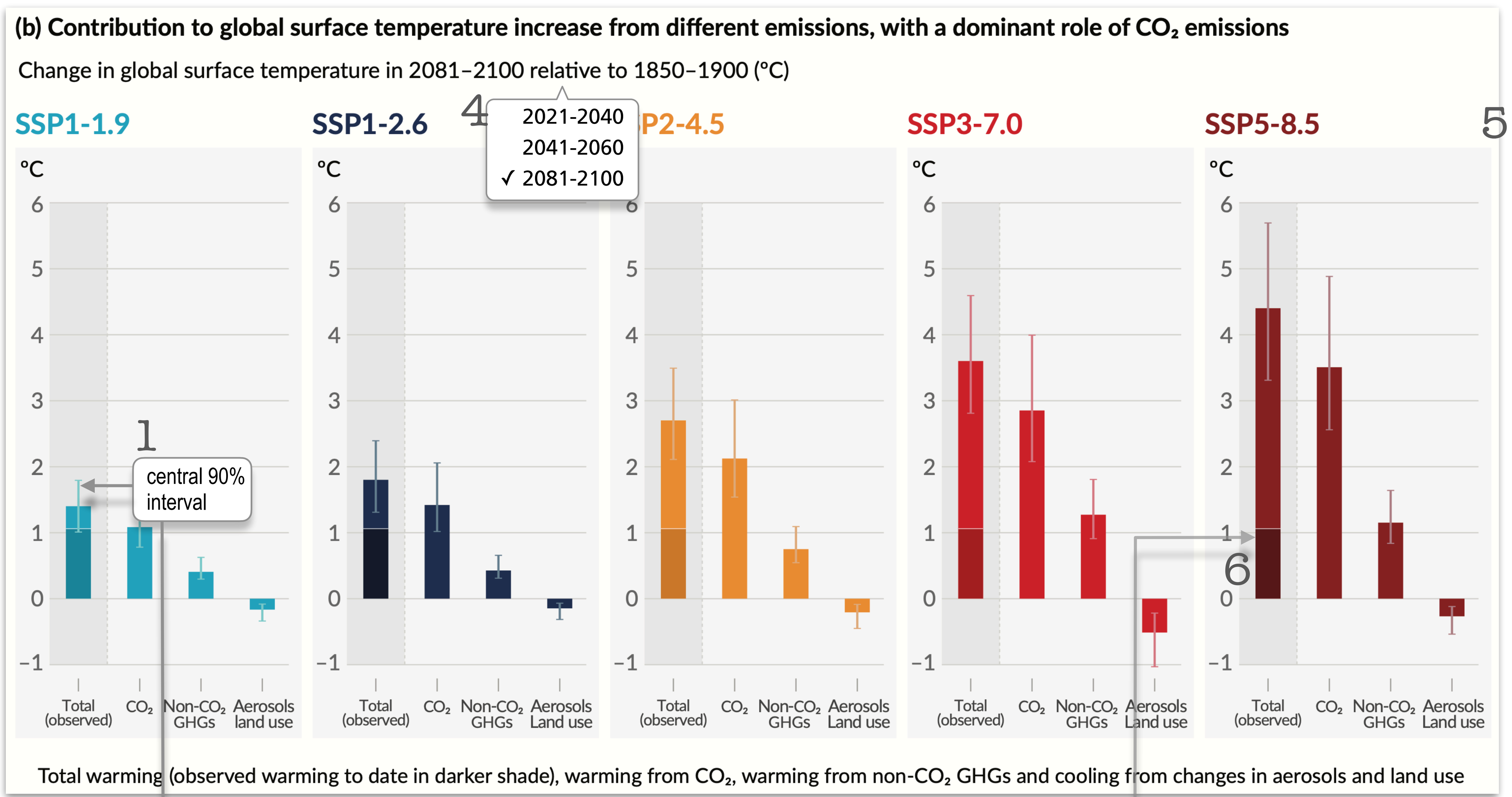
a **software infrastructure** problem!

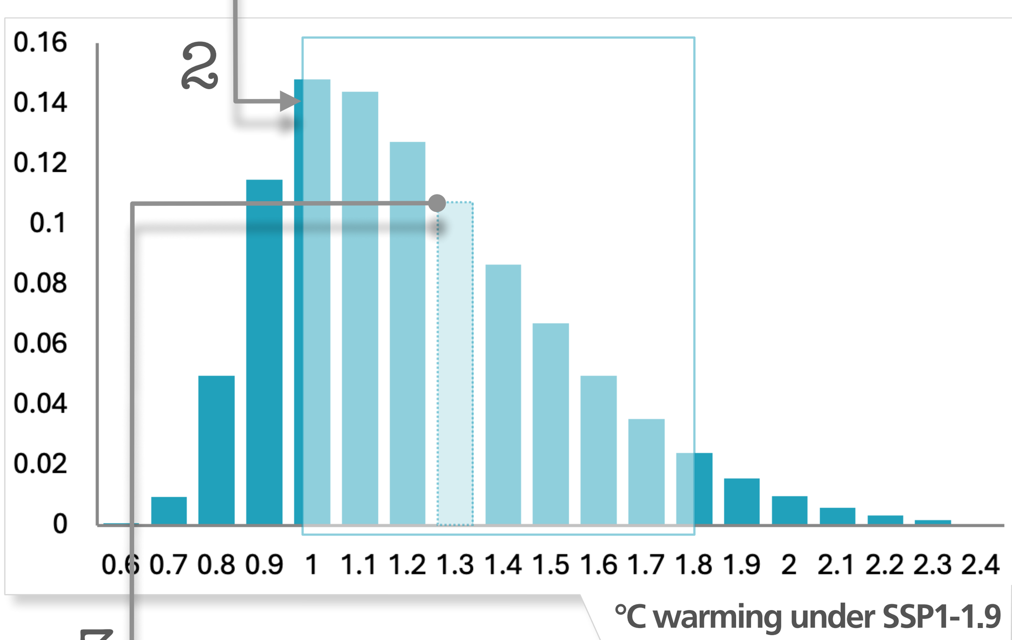
Can we help **readers** engage, critique and reach their own conclusions?

For:

- Sense making
- Policy decisions
- Peer review

implemented / work-in-progress / future work





Panel (b). Bars and whiskers represent median values and the *very likely* range, respectively. Within each scenario bar plot, the bars represent: total global warming (°C); warming contributions (°C) from changes in CO_2 and from non-CO_2 greenhouse gases (comprising well-mixed greenhouse gases and ozone); and net cooling from other anthropogenic drivers (aerosols and land use). The best estimate for observed warming in 2010–2019 relative to 1850–1900 is indicated in the darker column in the 'total' bar.

Run	°C warming	Time period	Scenario
12	1.39	2018-2100	SSP1-1.9
28	1.40	2018-2100	SSP1-1.9
29	1.36	2018-2100	SSP1-1.9
44	1.31	2018-2100	SSP1-1.9
52	1.37	2018-2100	SSP1-1.9
56	1.37	2018-2100	SSP1-1.9
58	1.40	2018-2100	SSP1-1.9
76	1.32	2018-2100	SSP1-1.9
81	1.36	2018-2100	SSP1-1.9
89	1.33	2018-2100	SSP1-1.9

satisfies $1.3 < x \leq 1.4$

SSP1-1.9 ensemble runs (10 of 1380)

virtually certain	≤	99%	100%
extremely likely	≤	95%	100%
very likely	≤	90%	100%
likely	≤	66%	100%
more likely than not	<	50%	100%
about as likely as not	≤	33%	66%
unlikely	≤	0%	33%
very unlikely	≤	0%	10%
extremely unlikely	≤	0%	5%
exceptionally unlikely	≤	0%	1%

IPCC confidence levels

has probability 1.0 of exceeding

in at least one of {2021-2040, 2041-2060, 2081-2100}

B.1.2 Global warming of 2°C, relative to 1850–1900, would be exceeded in the 21st century under the high and very high GHG emissions scenarios (SSP3-7.0 and SSP5-8.5). Global warming of 2°C would *extremely likely* be exceeded in the intermediate GHG emissions scenario (SSP2-4.5). Under the very low and low GHG emissions scenarios, global warming of 2°C is *extremely unlikely* to be exceeded (SSP1-1.9) or *unlikely* to be exceeded (SSP1-2.6).

computational transparency

"Computational transparency" preserves fine-grained provenance and execution history into a computed artifact, integrating the evidence base into the document. This **mock-up** illustrates a mixture of implemented features, WIP and future work.

Self-describing visual elements (1) surface first-class documentation into the UI.

Visual features linked to data (2) via fine-grained provenance. Here the UI can show that the whiskers correspond to a specific interval of an underlying distribution.

Nested provenance (3) allows readers to go back further in the pipeline and look at the samples aggregated into a specific bin.

Multiverse analyses (4) and **faceting** (5) complement this approach. Different time periods or scenarios might have different but analogous provenance stories.

References to visual elements (6) in natural language can be given precise meaning by interpreting them as visual queries.

Computational explanations (7) reveal how values were computed or why data elements are selected, and may have nested structure.

Quantitative natural language (8) can be assigned computational meaning and take advantage of the linked brushing and data provenance infrastructure.

technical approach

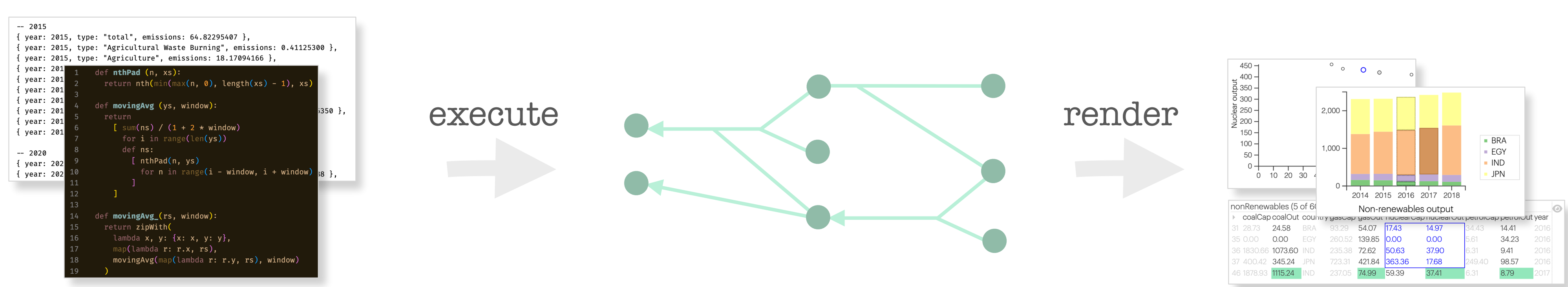
A **transparent programming language** called Fluid provides fine-grained provenance tracking via a dynamic dependence graph. Graph supports queries and serves as a **ground truth** for AI-generated natural language explanations.

A **web-based publishing front-end** enriches the final artifact with explorable provenance information, via additional interactions.

AI assistants will help **authors** create supporting natural language grounded in data, and help **readers** by translating query results into natural language.

AI **authoring** assistant to help underwrite fragments of natural language with code

AI **reading** assistant to assign natural language to provenance query results



Raw inputs
- declarative visualisation and analysis code
- data sources
- expository text

Dynamic dependence graph
- fragments of data as vertices
- computation steps as hyperedges
- cognacy queries reveal relationships between vertices

Interactive output
- web interface
- provenance queries
- why and how queries
- multiverses and facets

Graphics and text adapted from IPCC AR6 Summary for Policy Makers, 2021. Other images from Flaticon.com, Font Awesome.

https://f.luid.org