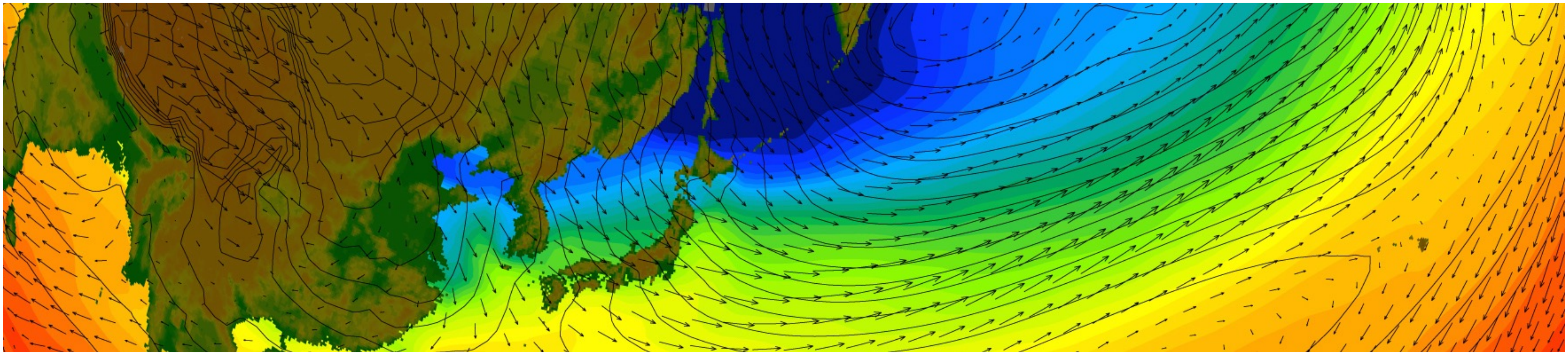


# Models, Data, and Wisdom:
## How do we know when to trust a climate model?


Steve Easterbrook, University of Toronto

Email: sme@cs.toronto.edu
Blog: www.easterbrook.ca/steve
Mastodon: @steve@fediscience.org

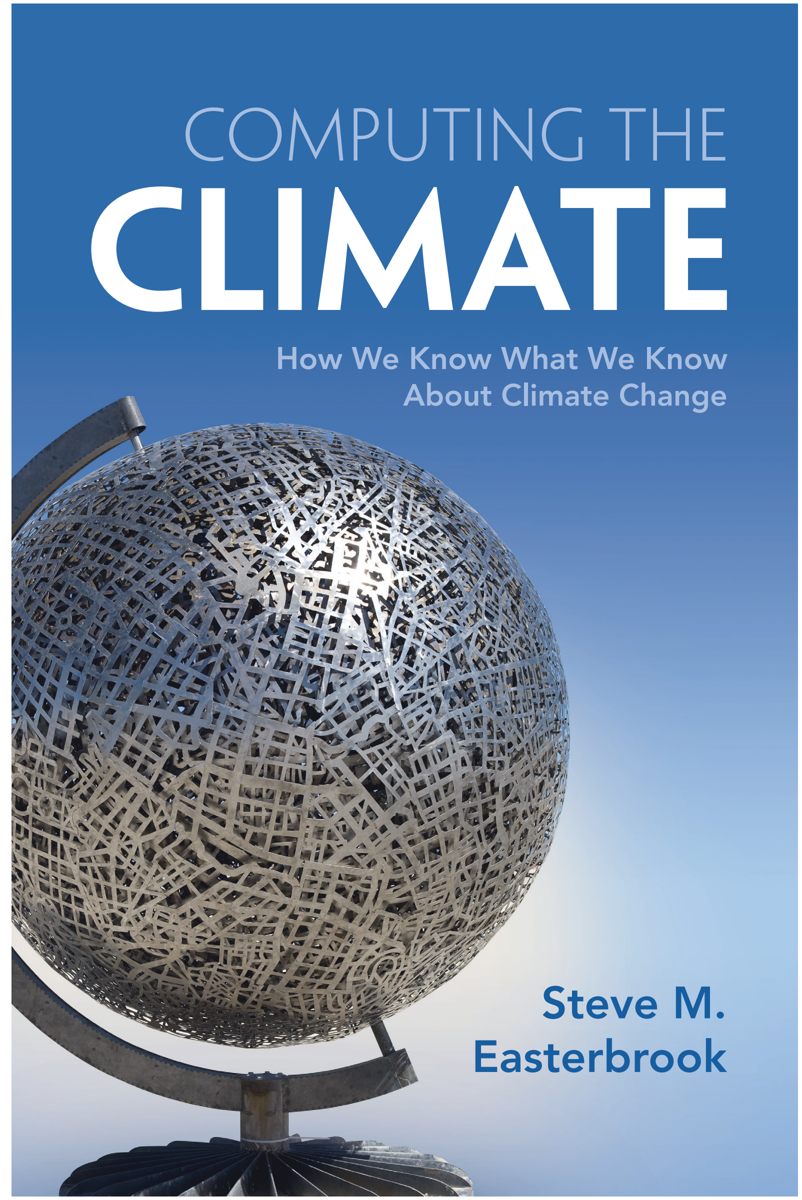
# Buy my book!

"This engaging, beautifully written book brings alive the scientists who created climate models, how they did it, and what the models can (and cannot) tell us - all in straightforward, nontechnical language and enlightening illustrations.

If you want to understand how modern climate science works, start here."

-- Paul N. Edwards, Stanford University. Author of *A Vast Machine: Computer Models, Climate Data, and the Politics of Global Warming*

Order directly from Cambridge University Press and use the code COMC2023 at checkout for a 20% discount!



# Talk Outline

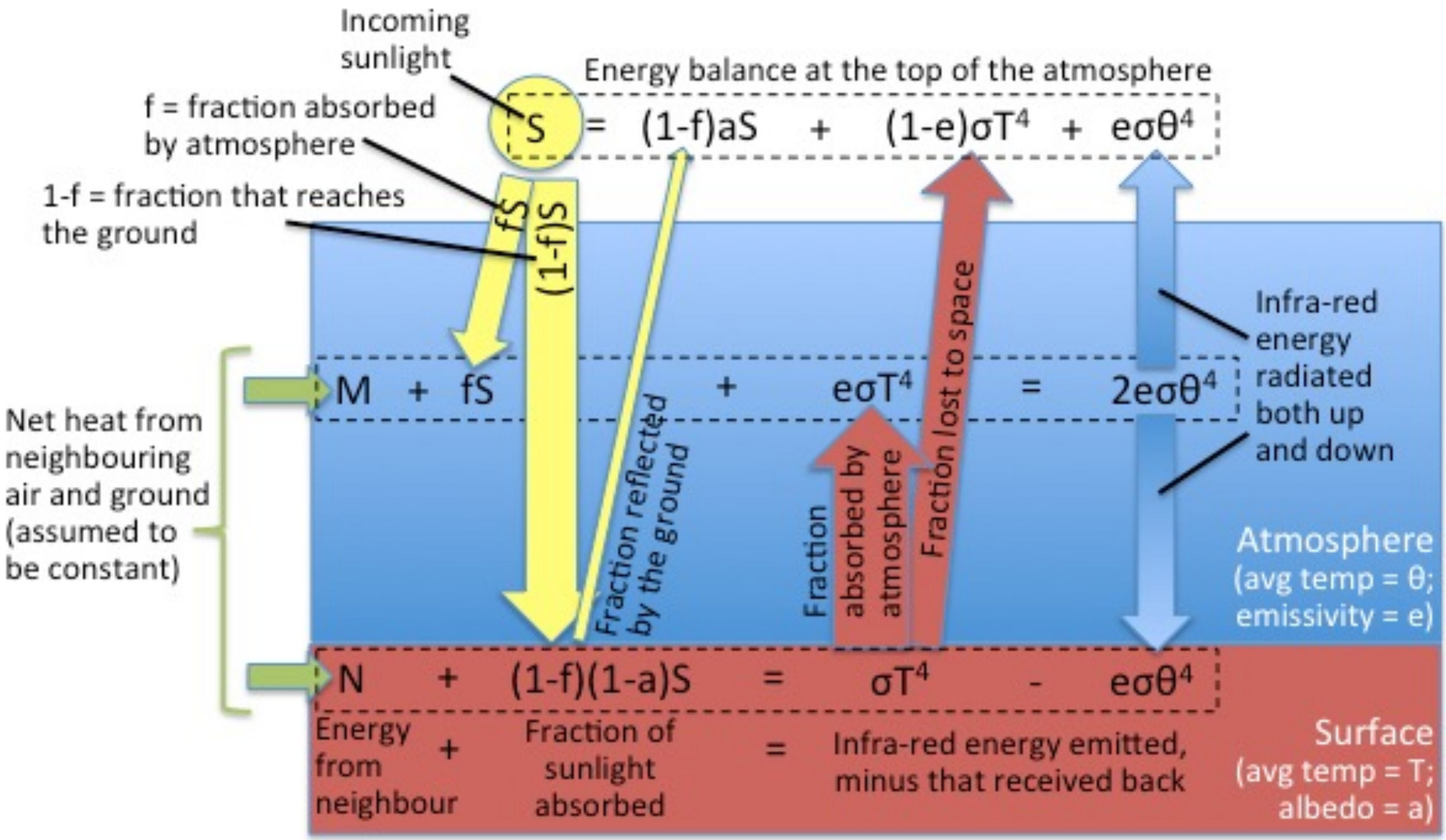
1. Prelude
   - What Arrhenius got wrong…
   - Systemic approaches for protection from errors
2. How good are today's models?
   - Engineering View: Are they well constructed?
   - Philosophical View: Are the models valid?
   - Empirical View: Do they match observations?
   - Sociological View: Are the results peer reviewed and replicated?
3. Extracting wisdom from models

# The First Computational Climate Model

1895: Svante Arrhenius constructs an energy balance model to test his hypothesis that the ice ages were caused by a drop in $CO_2$;

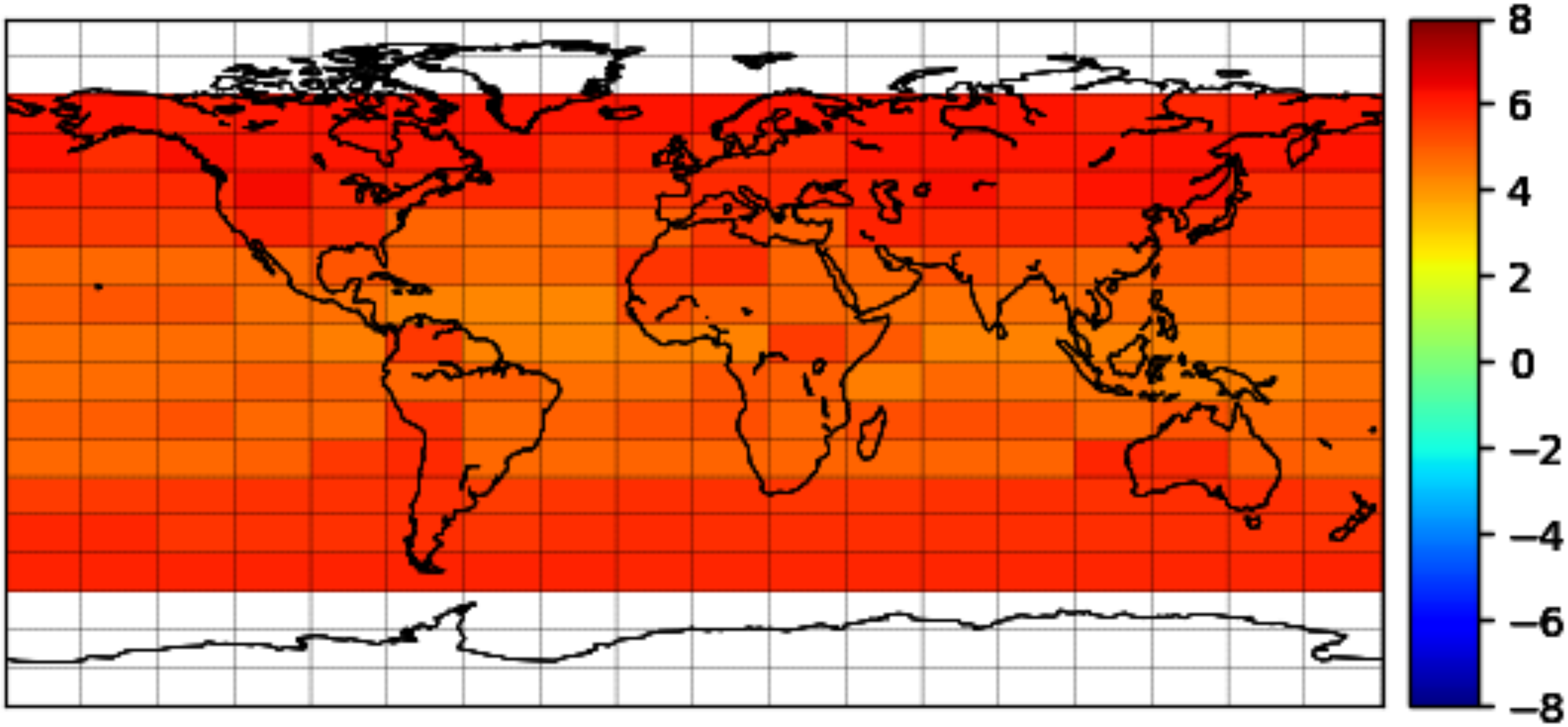
(Predicts global temperature rise of 5.7$^{\circ}$ C if we double $CO_2$)

# Schematic of the model equations



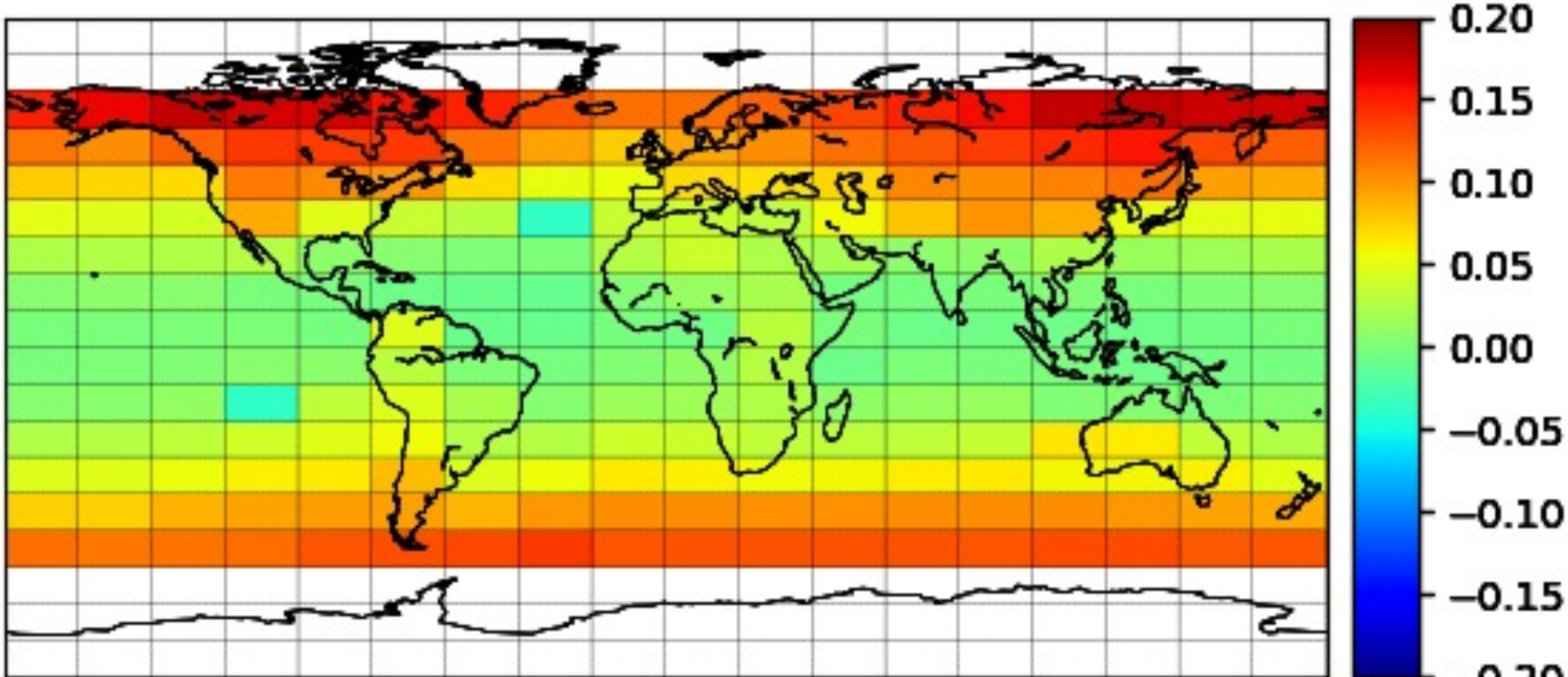
# A reimplementation

$\Delta$T for Doubled CO2 –
Using Arrhenius's radiative absorption data



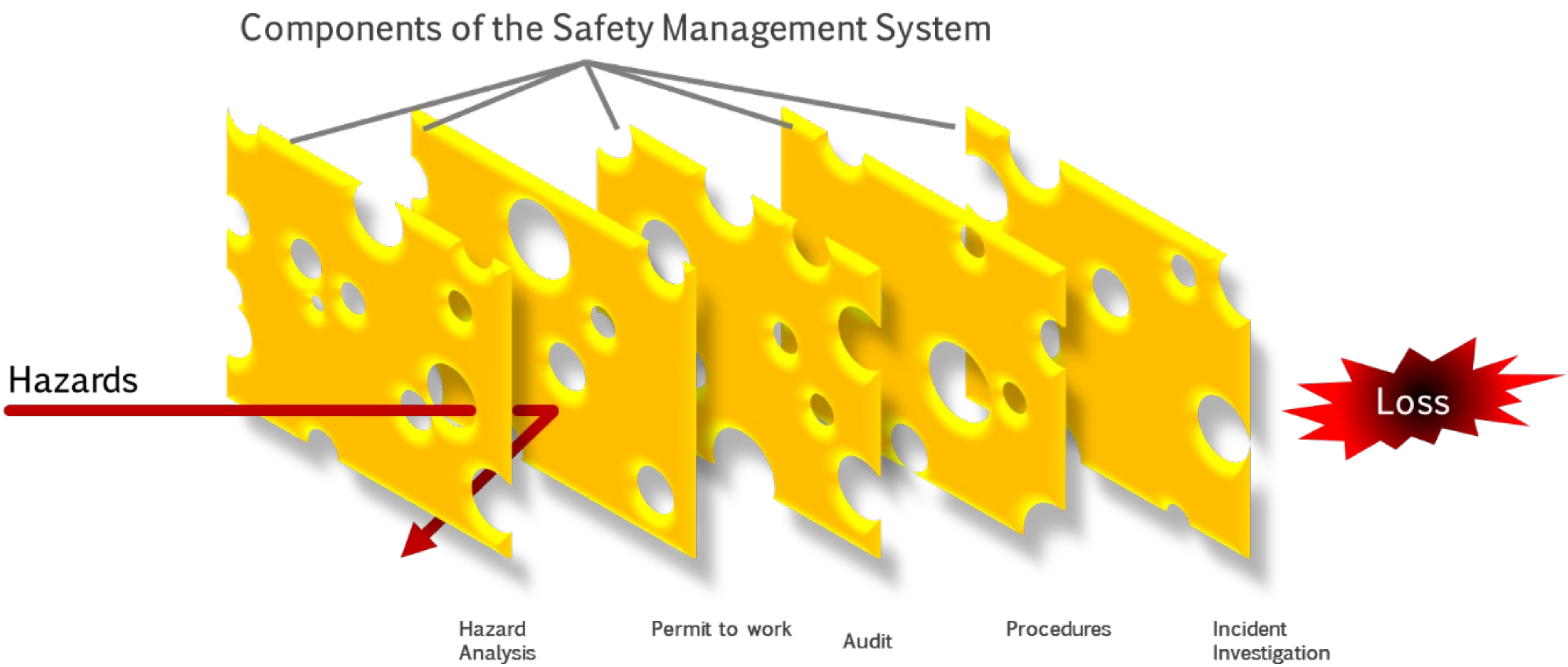
# Now with modern data…

ΔT for Doubled CO2 –
Using Lowtran Radiation data



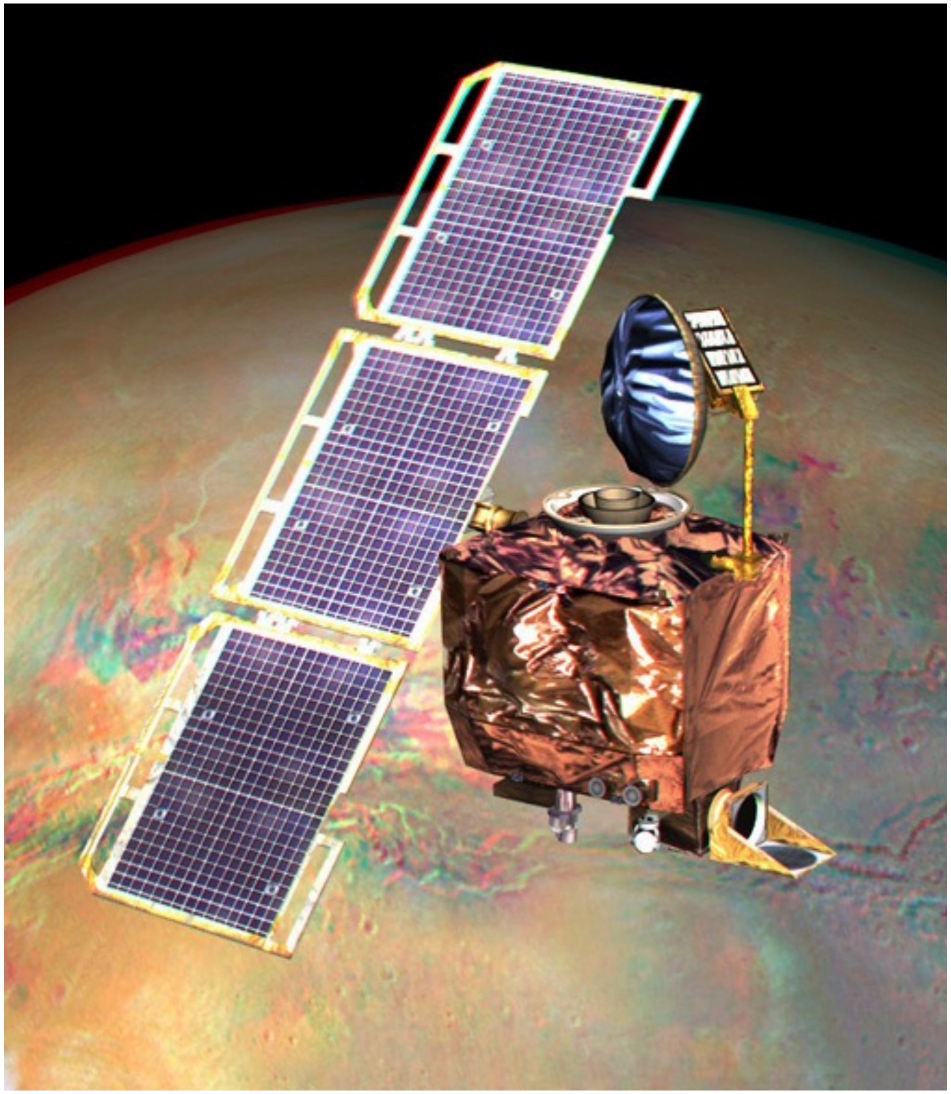

For more, see: Dufresne, J.-L. (2009). *L'effet de serre: sa découverte, son analyse par la méthode des puissances nettes échangées et les effets de ses variations récentes et futures sur le climat terrestre*. Habilitation Thesis, Université Pierre et Marie Curie, Paris.

# Swiss Cheese model of fault protection



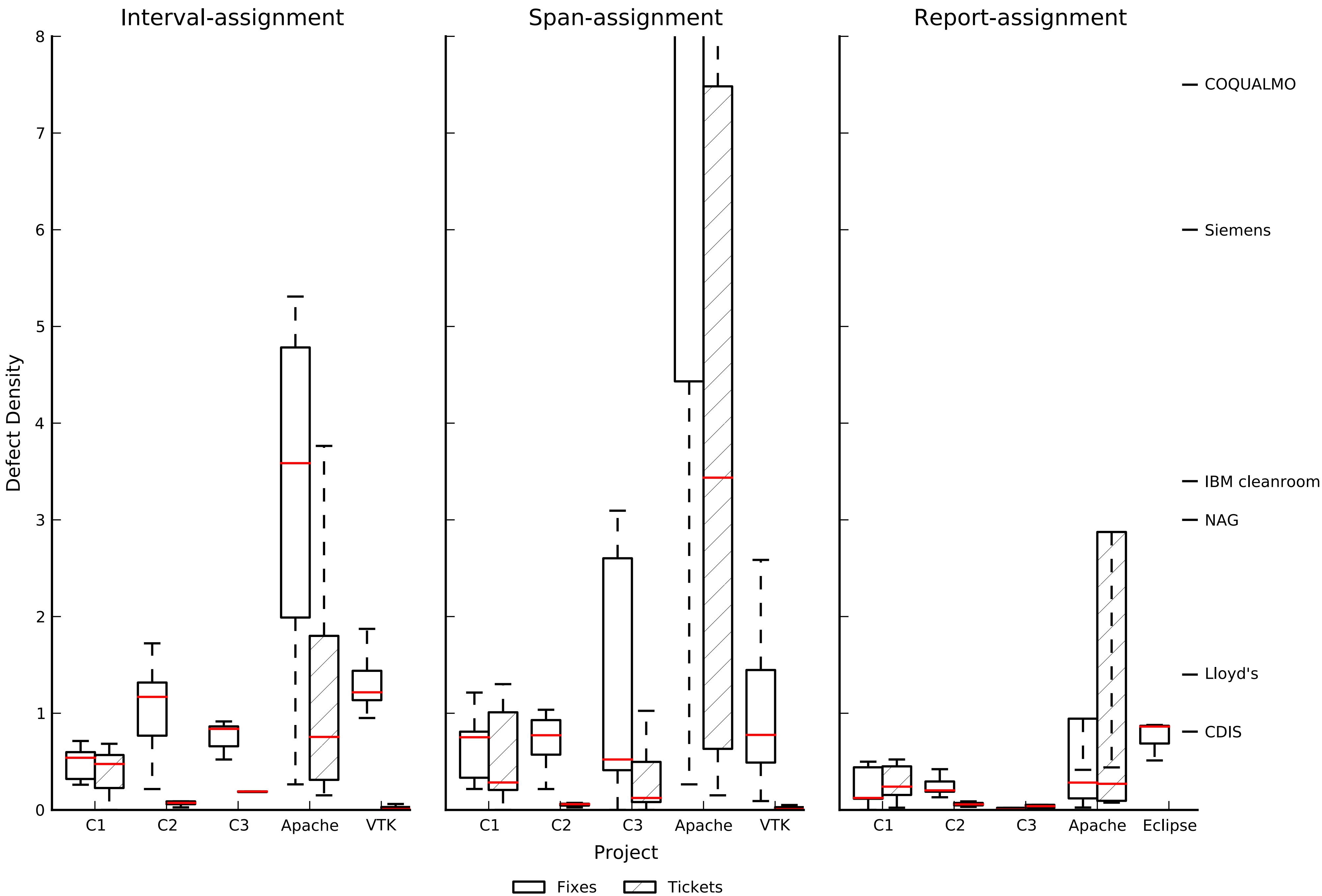
# Mars Climate Orbiter

- Launched
  - 11 Dec 1998
- Mission
  - interplanetary weather satellite
  - communications relay for Mars Polar Lander
- Fate:
  - Arrived 23 Sept 1999
  - No signal received after initial orbit insertion
- Cause:
  - Faulty navigation data caused by failure to convert imperial to metric units



# Assessing Model Quality

1. Engineering quality:
   - How many errors in the code?
   - Is it tested to industry standards?
2. Philosophically speaking:
   - Popper: Are they refutable?
   - Lakatos: Is the field progressing?
3. Empirically speaking:
   - Do the models match observations?
   - Have the models made successful predictions?
4. Sociologically speaking:
   - Are the models and results independently replicated?
   - Is all the data and code freely available?




Pipitone, J., Easterbrook, S. (2012). Assessing climate model software quality: a defect density analysis of three models. *Geoscientific Model Development*, 5(4), 1009–1022.

# Hypotheses for low defect rates

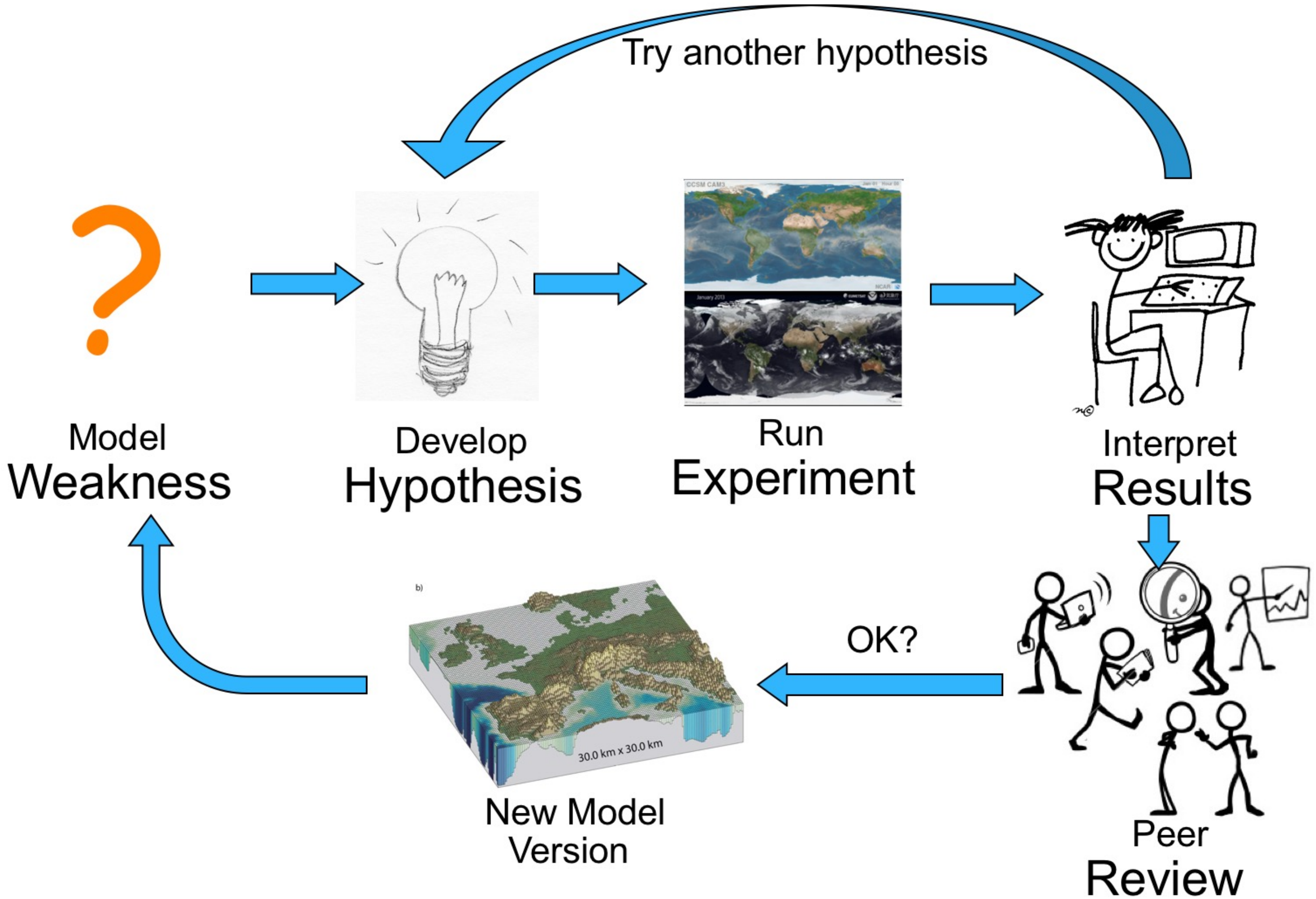
- Domain Expertise
  - Developers are users and experts
- Rigorous Development Process
  - Code changes as scientific experiments, with peer review
- Slow, cautious development process
- Narrow Usage Profile
  - And hence potential for brittleness
- Intrinsic Defect Sensitivity / Tolerance
  - Bugs are either obvious or irrelevant
- Successful Disregard (and hence higher technical debt)
  - Scientists tolerate poor code & workarounds, if they don’t affect the science

Pipitone, J., Easterbrook, S. (2012). Assessing climate model software quality: a defect density analysis of three models. *Geoscientific Model Development*, 5(4), 1009–1022.

# E.g. Testing strategy for ICON

- Simple tests with a known solution
  - Shallow water test
  - Baroclinic wave test (runs automatically for ICON)
- Bit-level reproducibility tests
  - Compare restarted run with uninterrupted run
  - Compare parallel vs sequential configurations
- Comparison with reference model
- Aquaplanet tests
- Hindcasts for the fully coupled model
  - $20^{th}$ Century
  - Pre-industrial
  - Paleoclimate

# Every code change is hypothesis testing



# Acknowledge Model Errors

See: Stevens, B., et al. (2013). Atmospheric component of the MPI-M Earth System Model: ECHAM6. *Journal of*

# “All models are wrong…”

## Karl Popper

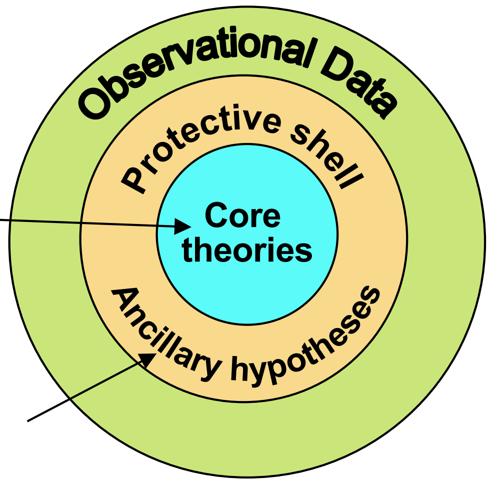
- A theory is scientific if it can be refuted
- In practice, you don’t throw out a theory at the first failed test…
- Science evolves through “survival of the fittest”:
  - many competing theories, discard the most problematic

## Imre Lakatos

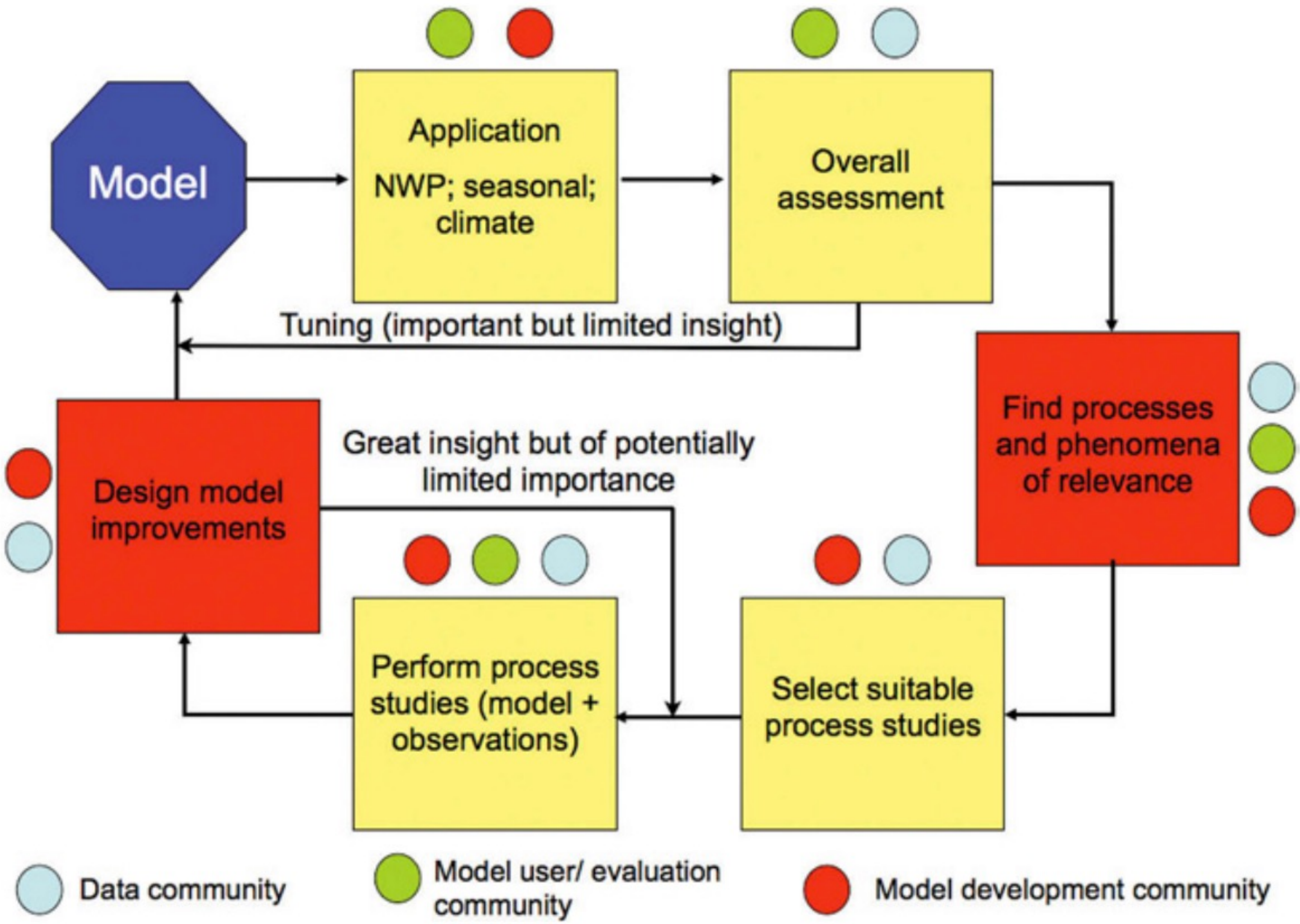
- A program of research is scientific if it makes progress = more successful predictions over time
- Hard core of established theory + a protective shell of ancillary hypotheses
  - Adjust these to explain more and more of the world

Immune to refutation

Adjusted to improve applicability

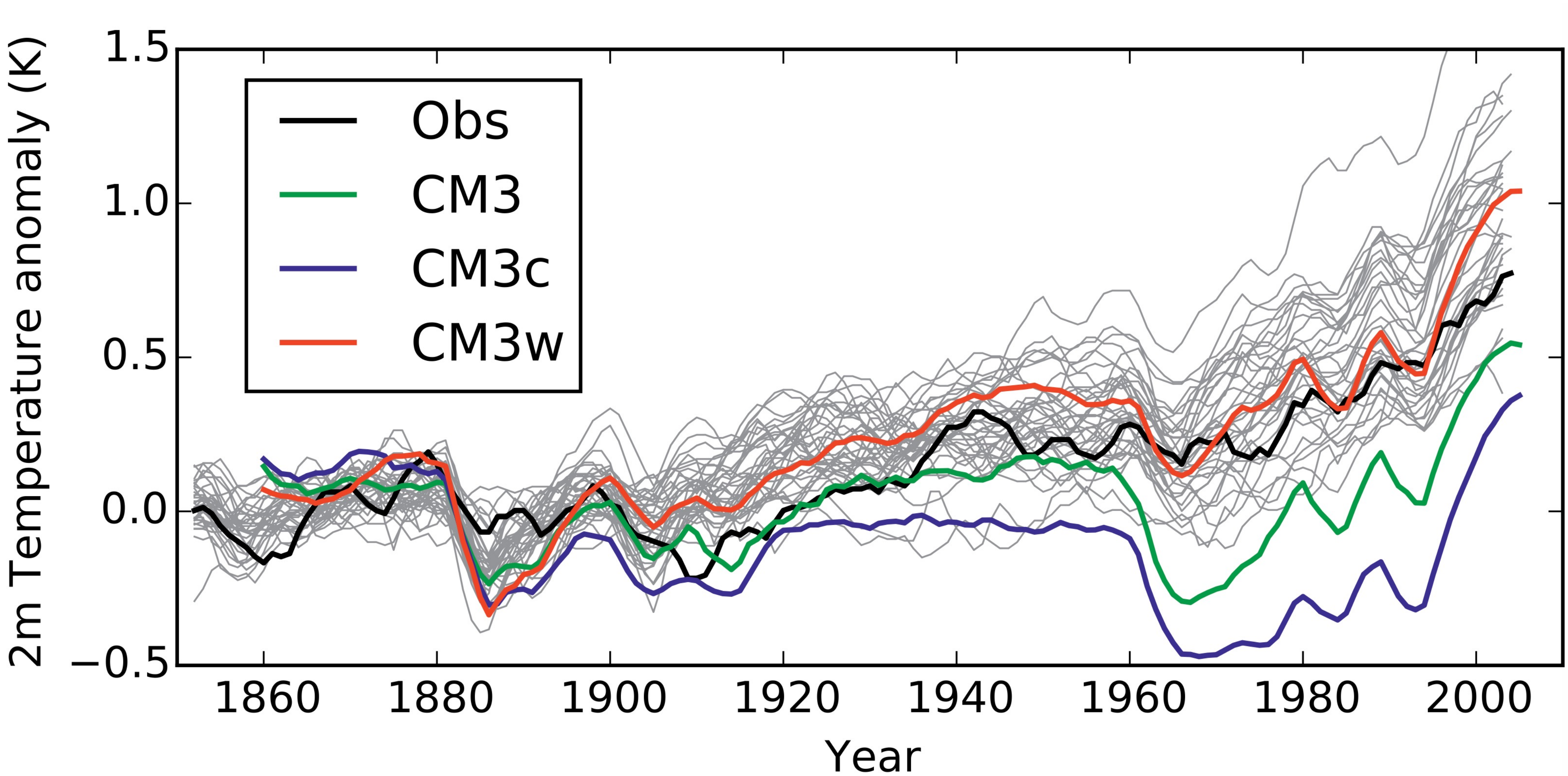


# Models and Process Studies



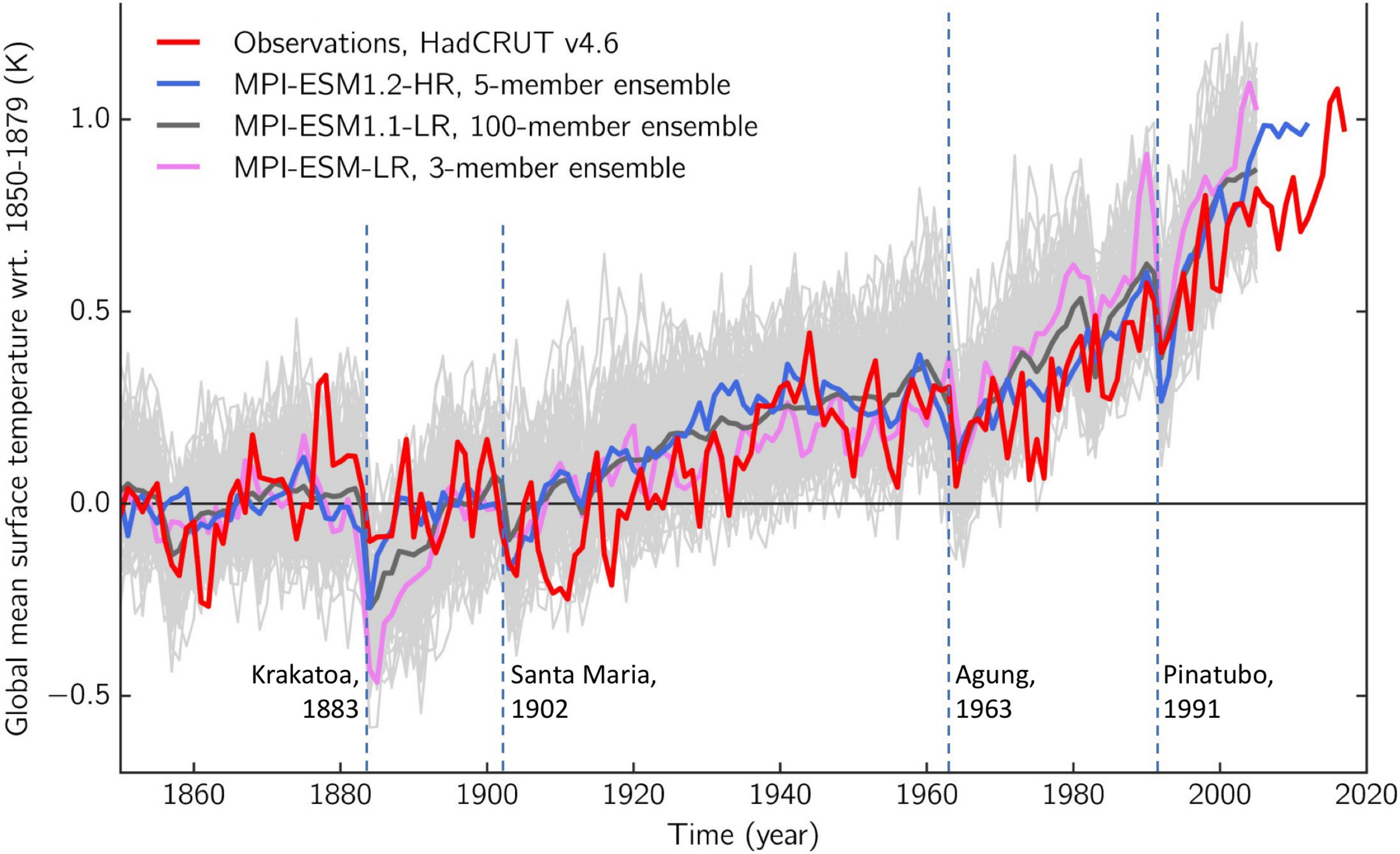

Jakob, C. (2010). Accelerating Progress in Global Atmospheric Model Development through Improved Parameterizations. *Bulletin of the American Meteorological Society*, *91*(7), 869–876.

# Model Tuning - example

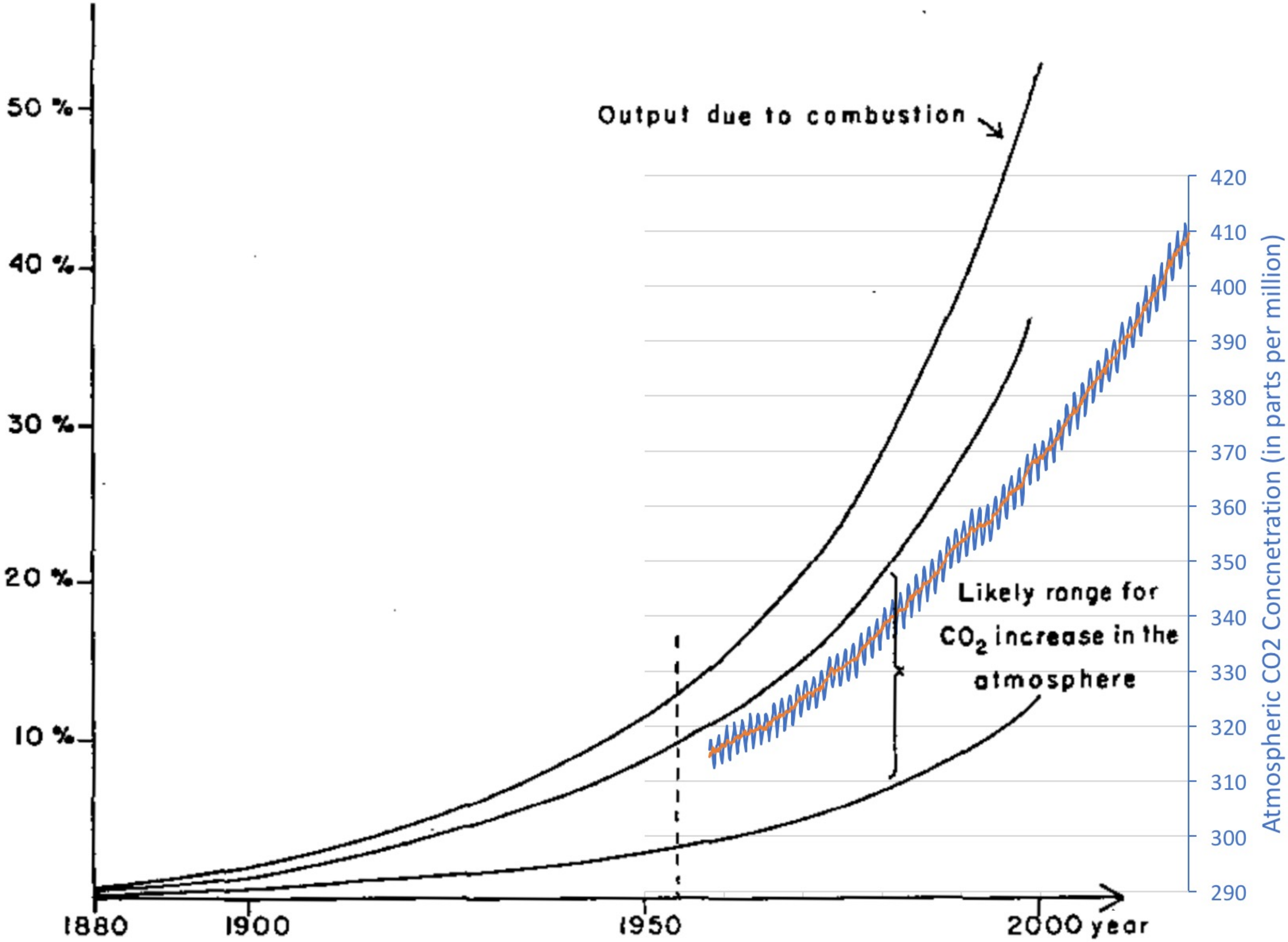



Hourdin, F., Mauritsen, T., Gettelman, A., Golaz, J.-C., Balaji, V., Duan, Q., ... Williamson, D. (2017). The Art and Science of Climate Model Tuning. *Bulletin of the American Meteorological Society*, *98*(3), 589–602

# Model Ensembles (varied initial conditions)



# Successful Predictions



# First computer prediction of climate change

1967: Syukuro Manabe builds a computer model of the vertical structure of the atmosphere.

Predicts doubling CO2 would raise surface temperature by 2°C

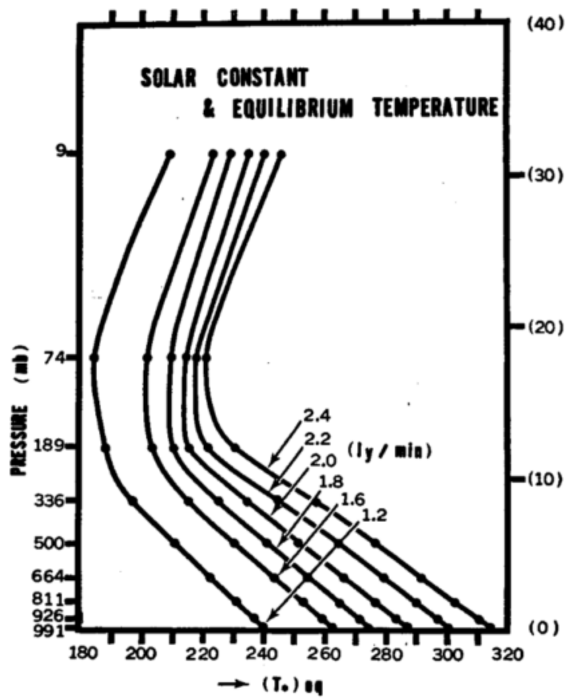



FIG. 8. Vertical distribution of radiative convective equilibrium temperature of the atmosphere with a given distribution of relative humidity for various values of the solar constant.

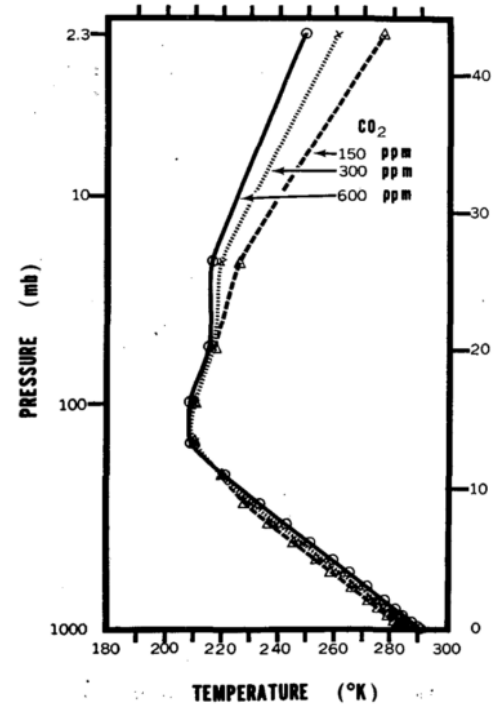
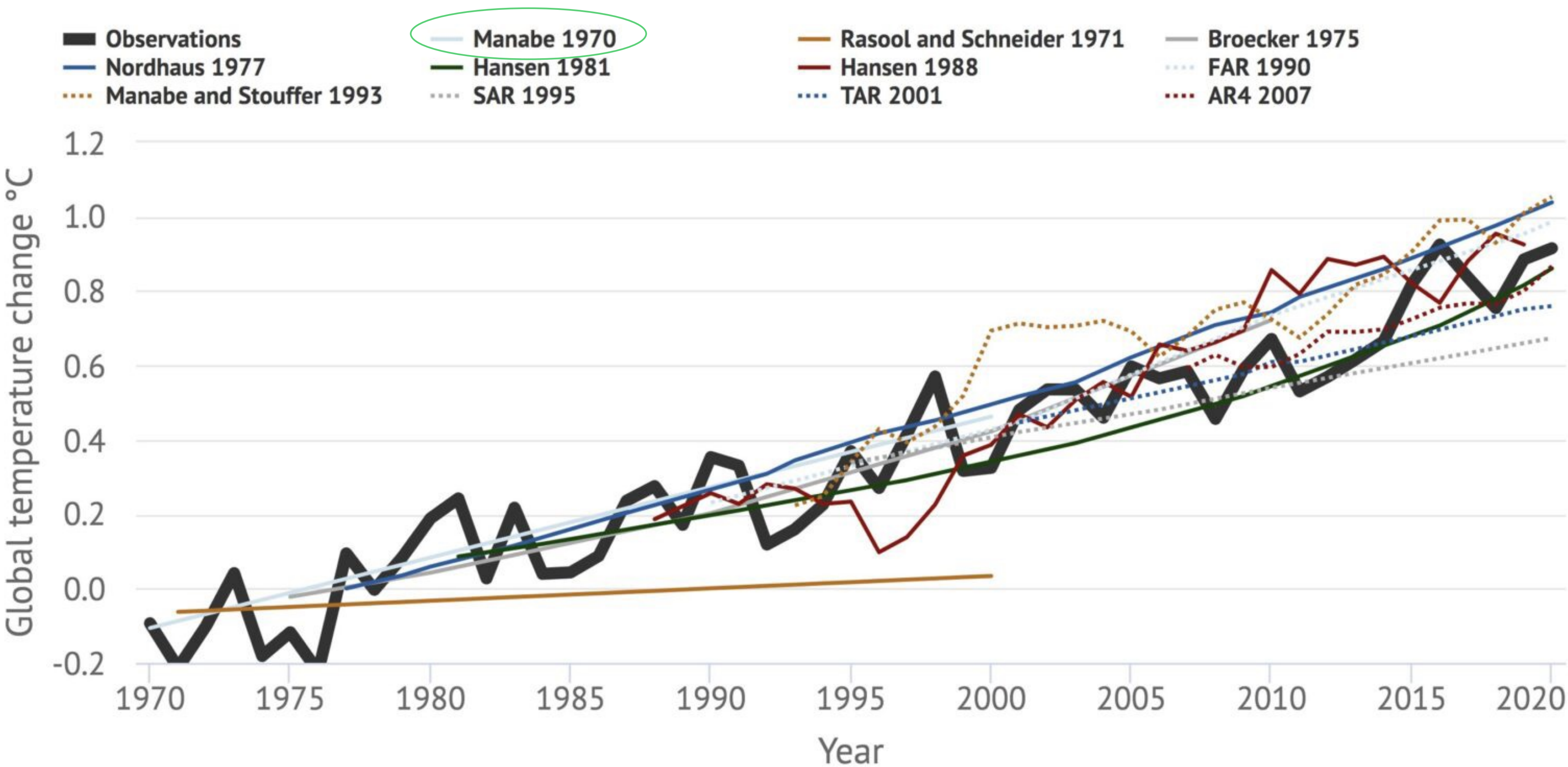



FIG. 16. Vertical distributions of temperature in radiative convective equilibrium for various values of $CO_2$ content.

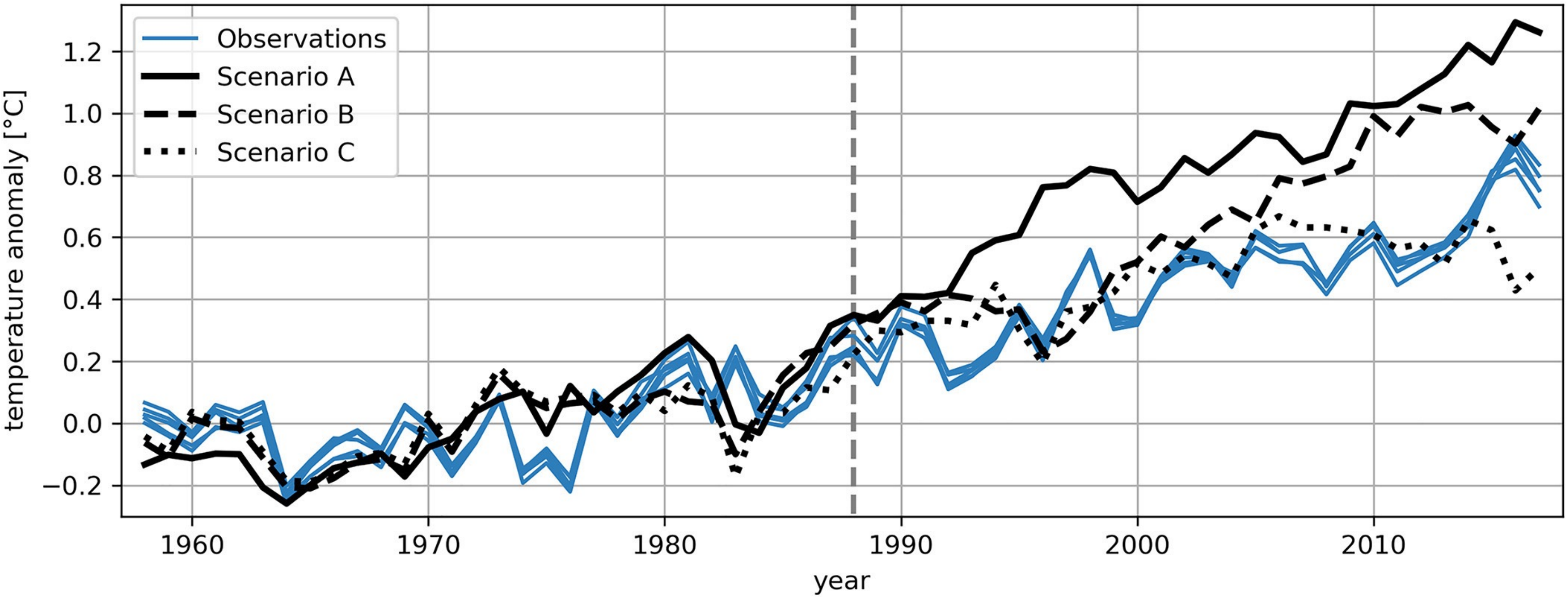
Manabe, S., & Wetherald, R. T. (1967). Thermal Equilibrium of the Atmosphere with a Given Distribution of Relative Humidity. Journal of the Atmospheric Sciences.

# Manabe’s prediction for 2000



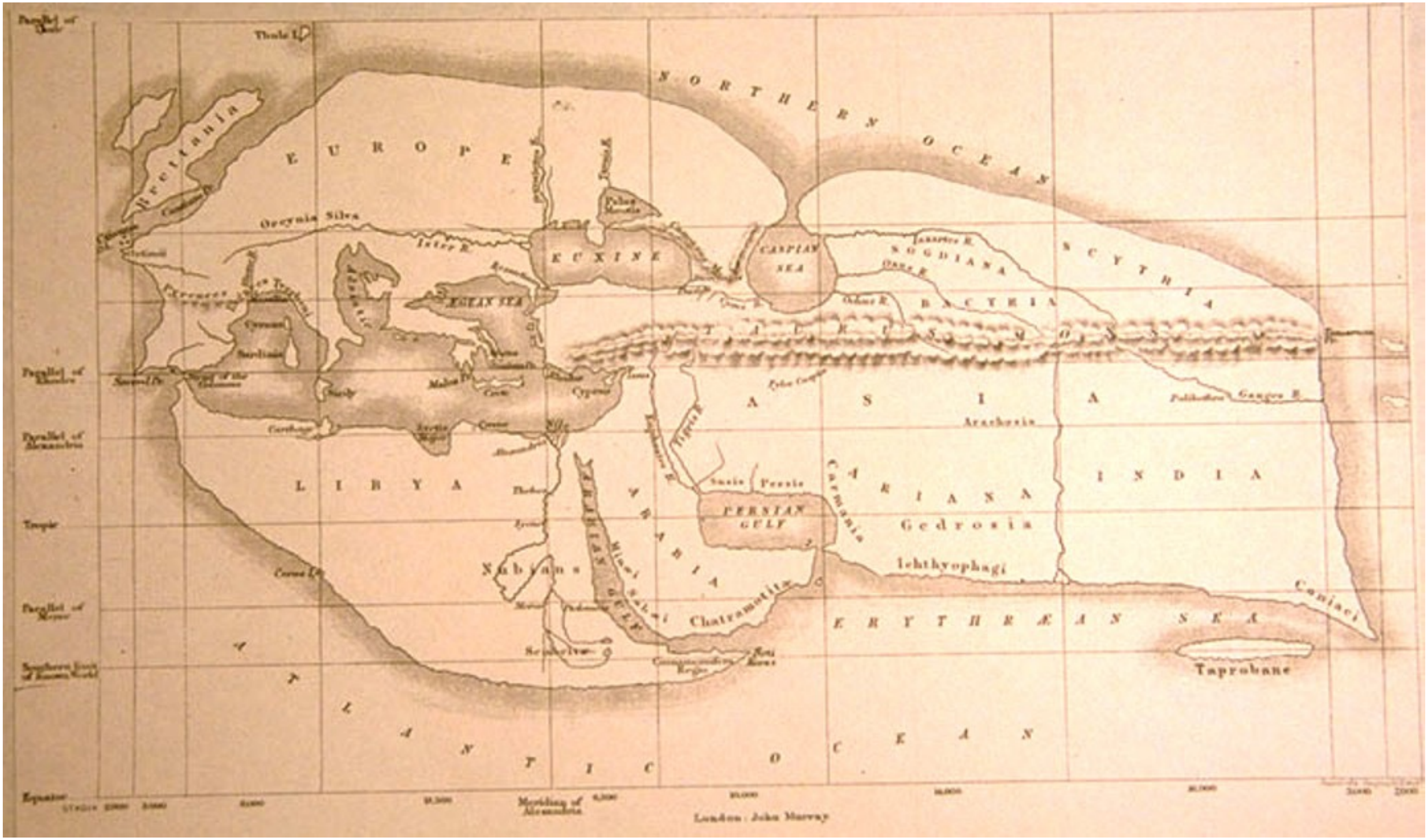

https://www.realclimate.org/index.php/archives/2021/10/a-nobel-pursuit/

# Hansen's 1988 projections

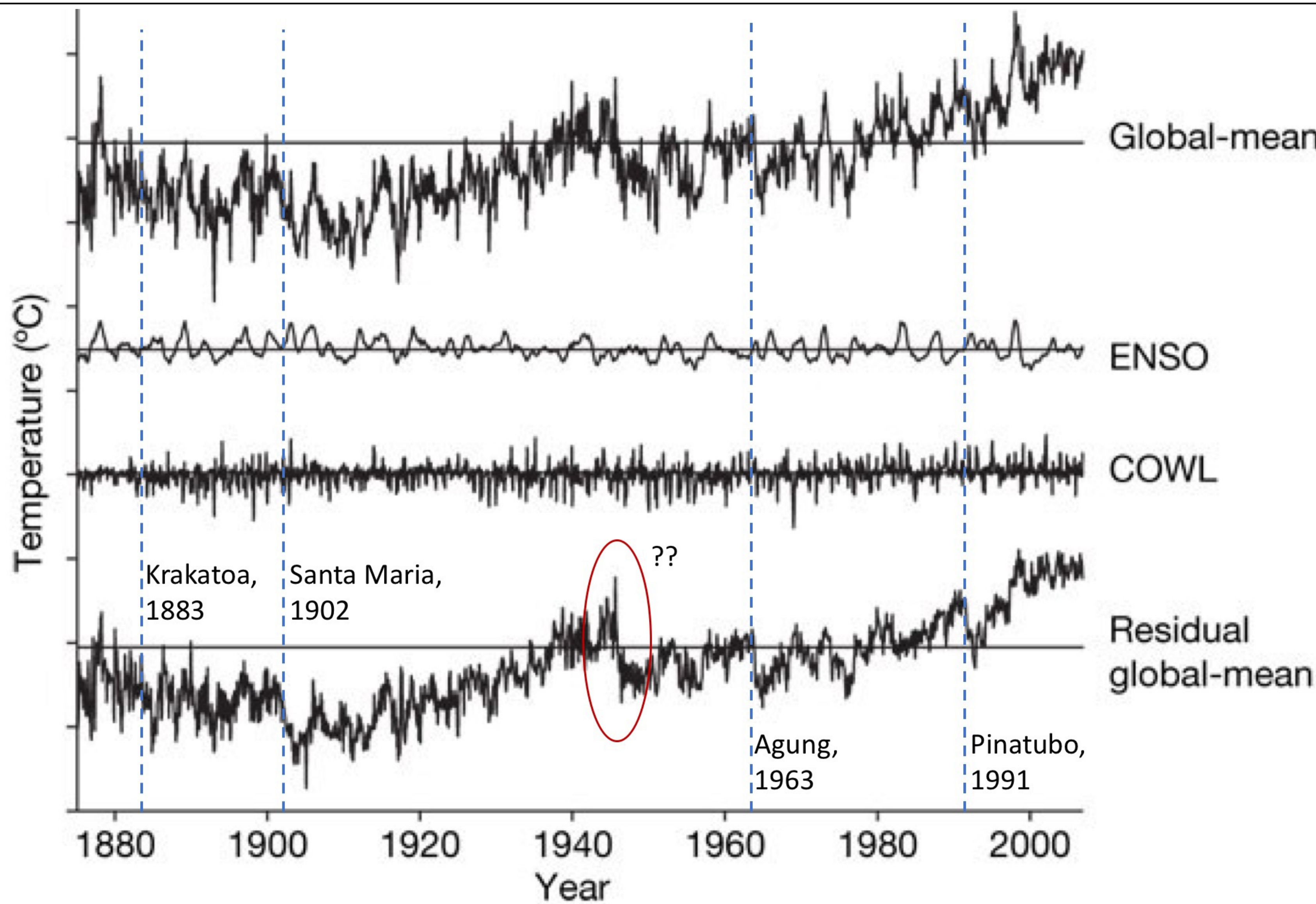



Hausfather, Z et al (2020) Evaluating the Performance of Past Climate Model Projections. Geophysical Research Letters, 47(1).

# When the map and the territory disagree…

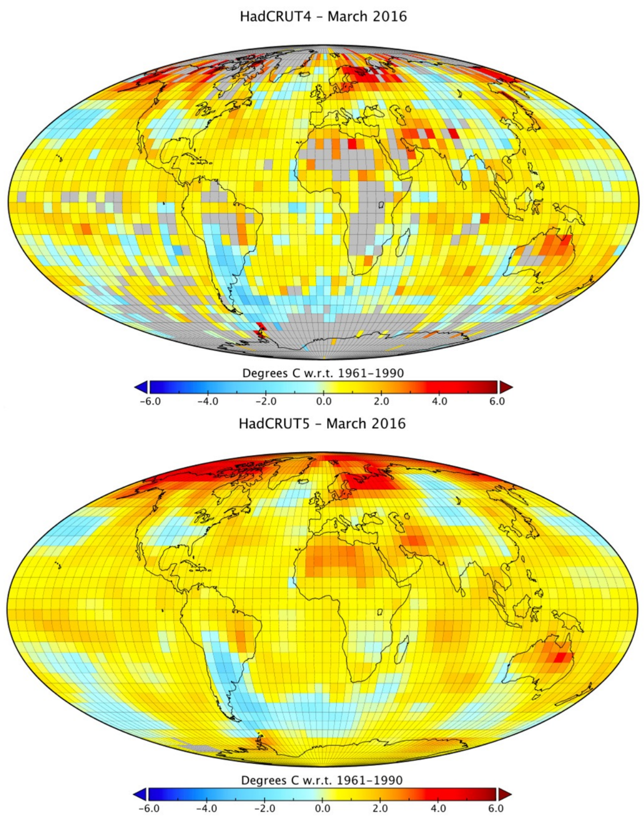


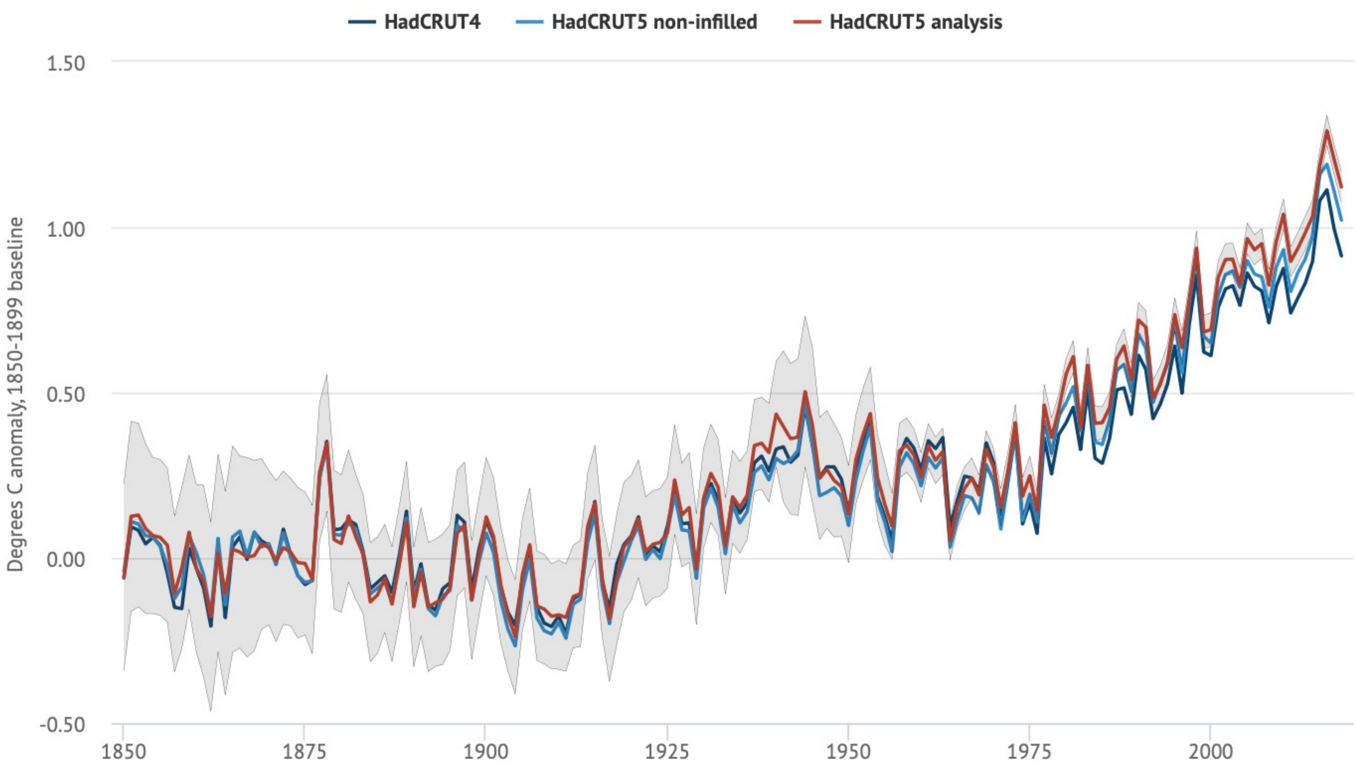
# Better than Data?




Thompson, D. W. J., Kennedy, J. J., Wallace, J. M., & Jones, P. D. (2008). A large discontinuity in the mid-twentieth century in observed global-mean surface temperature. *Nature*, *453*(7195), 646–649.

# Better than Data?



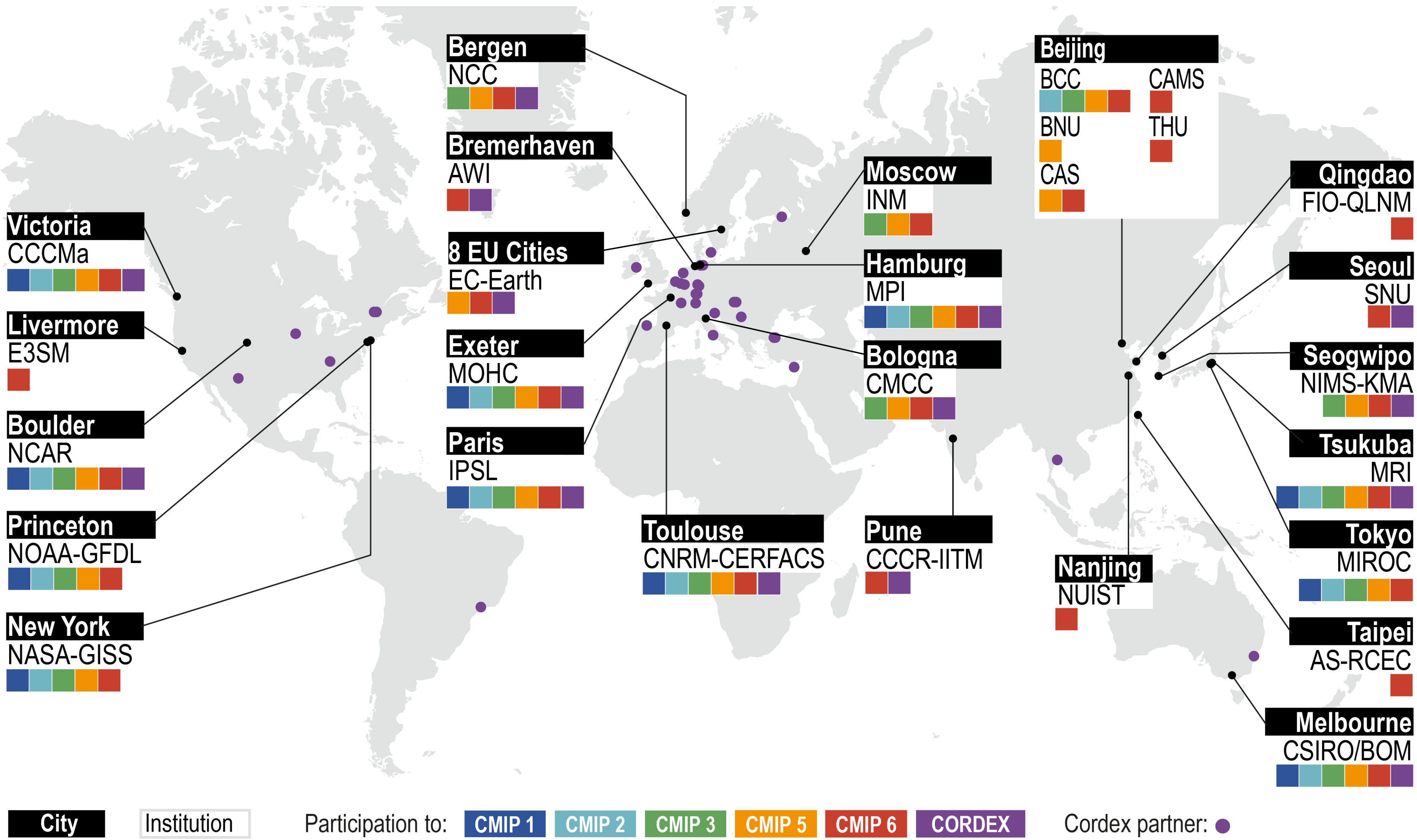

https://www.carbonbrief.org/analysis-why-the-new-met-office-temperature-record-shows-faster-warming-since-1970s

# The Coupled Model Intercomparison Projects

| | CMIP (1996 on) | CMIP2 (1997 on) | CMIP3 (2005-2006) | CMIP5 (2010-2011) | CMIP6 (2017-9) |
|---|---|---|---|---|---|
| **Number of Experiments** | 1 | 2 | 12 | 110 | 305 |
| **Centres Participating** | 16 | 18 | 15 | 31 | 49 |
| **# of Distinct Models** | 19 | 24 | 21 | 59 | 109 |
| **# of Runs (≈ Models x Expts)** | 19 | 48 | 211 | 841 | >10K |
| **Total Dataset Size** | 1 Gigabyte | 500 Gigabyte | 36 Terabyte | 3.3 Petabyte | 50 Petabyte |
| **Total Downloads from archive** | ? | ? | 1.2 Petabyte | (still growing) | (still growing) |
| **Number of Papers Published** | 47 | | 595 | thousands | ?? |

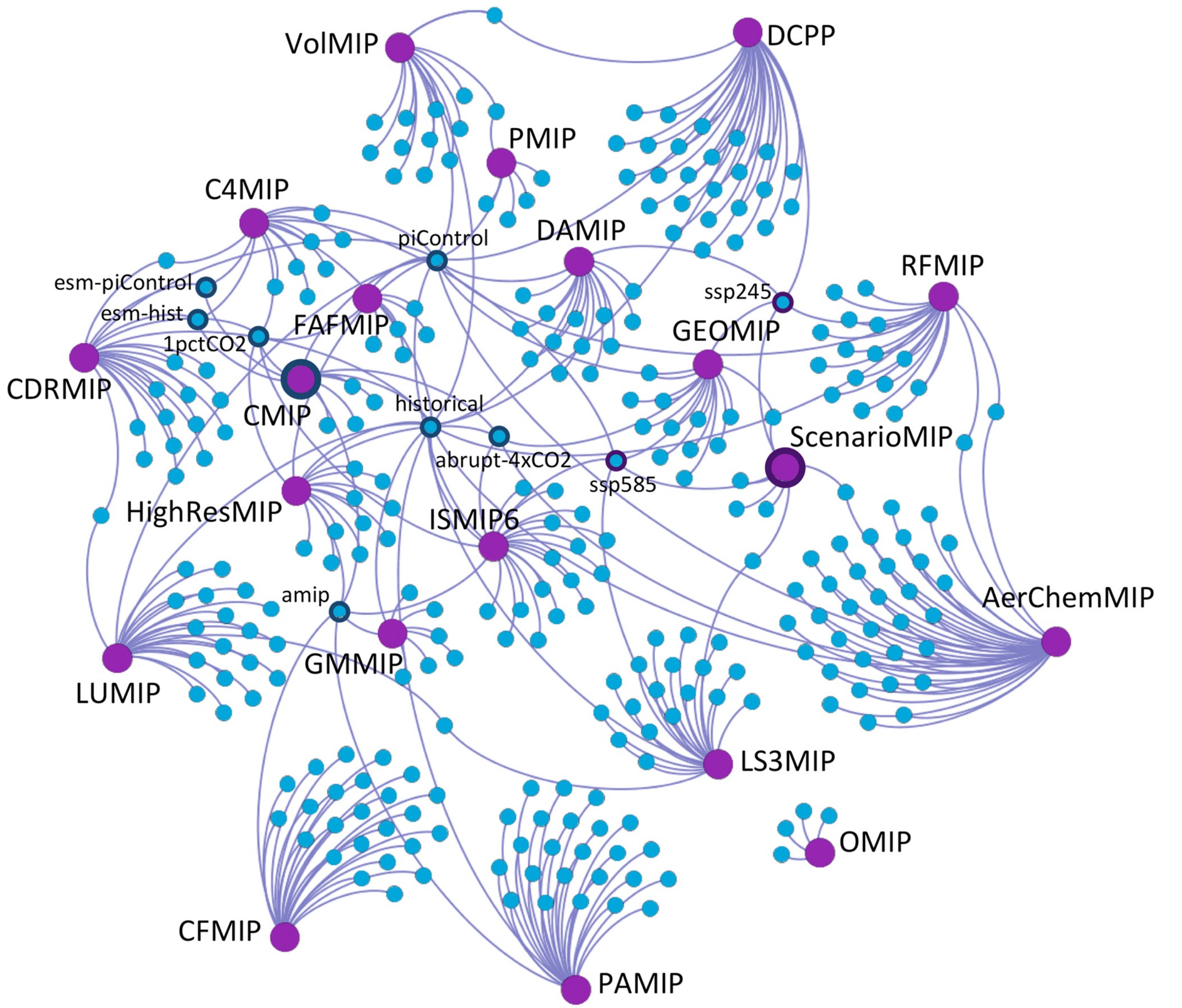
All data freely available on the Earth System Grid Federation
e.g. see: https://esgf-data.dkrz.de/projects/esgf-dkrz/

# Replicated Experiments




Source: IPCC AR6 WG1 Fig 1.20

# An ecosystem of shared experiments



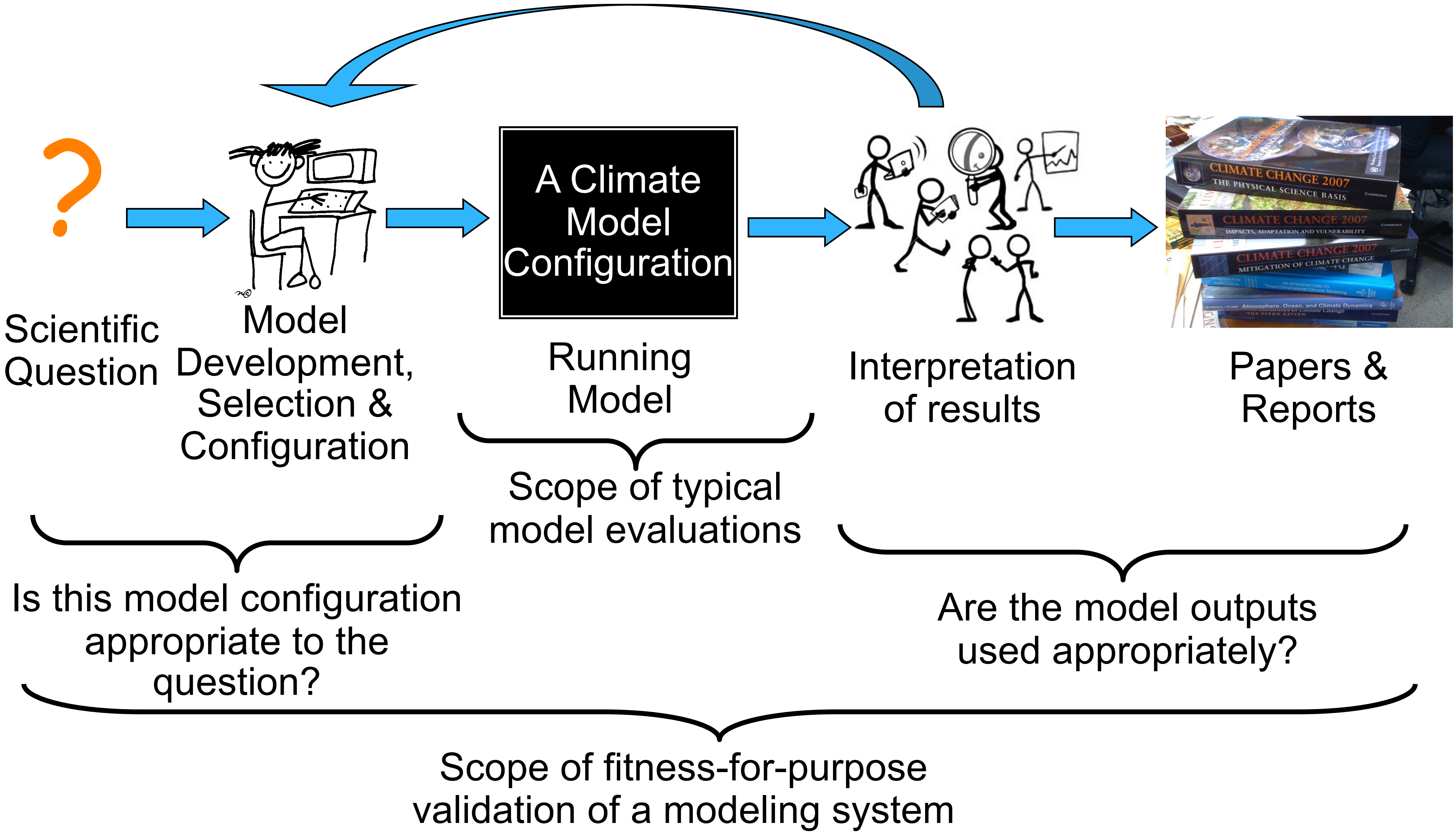

| CMIP6 (core MIPS recorded by ES-DOC) |
|---|
| **AerChemMIP**: Aerosols and Chemistry MIP - Collins et al. (2017) |
| **C4MIP**: Coupled Climate Carbon Cycle MIP - Jones et al. (2016) |
| **CDRMIP**: The Carbon Dioxide Removal Model Intercomparison Project - Keller et al. (2018) |
| **CFMIP**: Cloud Feedback Model Intercomparison Project - Webb et al. (2017) |
| **CMIP**: Climate Model Intercomparison Project - Eyring et al. (2016) |
| **DAMIP**: Detection and Attribution Model Intercomparison Project - Gillett et al. (2016) |
| **DCPP**: Decadal Climate Prediction Project - Boer et al. (2016) |
| **FAFMIP**: Flux-Anomaly-Forced Model Intercomparison Project - Gregory et al. (2016) |
| **GMMIP**: Global Monsoons Modeling Inter-comparison Project - Zhou et al. (2016) |
| **GeoMIP**: The Geoengineering Model intercomparison Project - Kravitz et al. (2015) |
| **HighResMIP**: High Resolution Model Intercomparison Project - Haarsma et al. (2016) |
| **ISMIP6**: Ice Sheet Model Intercomparison Project for CMIP6 - Nowicki et al. (2016) |
| **LS3MIP**: Land Surface, Snow and Soil Moisture MIP - van den Hurk et al. (2016) |
| **LUMIP**: Land-Use Model Intercomparison Project - Lawrence et al. (2016) |
| **OMIP**: Ocean Model Inter-comparison Project - Griffies et al. (2016) |
| **PAMIP**: Polar Amplification Model Intercomparison Project - Smith et al. (2019) |
| **PMIP**: Paleoclimate Modeling Intercomparison Project - Kageyama et al. (2018) |
| **RFMIP**: Radiative Forcing Model Intercomparison Project - Pincus et al. (2016) |
| **ScenarioMIP**: Scenario Model Intercomparison Project - O'Neill et al. (2016) |
| **VolMIP**: Model Intercomparison Project on the climatic response to Volcanic forcing - Zanchettin et al. (2016) |

Pascoe, C., et al. (2020). Documenting numerical experiments in support of the Coupled Model Intercomparison Project Phase 6 (CMIP6). *Geoscientific Model Development*, *13*(5), 2149–2167

# Talk Outline

1. Prelude
   - What Arrhenius got wrong…
   - Systemic approaches for protection from errors
2. How good are today's models?
   - Engineering View: Are they well constructed?
   - Philosophical View: Are the models valid?
   - Empirical View: Do they match observations?
   - Sociological View: Are the results peer reviewed and replicated?
3. Extracting wisdom from models

# From models to modeling systems



# Q&A