LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Stefanie Falk (stefanie.falk@lmu.de)

WiSe 2023/24

Unsicherheiten bei der Umweltsystem-modellierung

Sitzung 13

- Recap

Course content

Why study uncertainty?
Earth-systems & Developing environmental models
Sources of uncertainty in environmental modelling
Review of statistical concepts & Data uncertainty
Gap-filling & Error propagation: Analytic approaches
Error propagation: Analytic approaches & Numerical methods
Uncertainty in praxis : Excursion to LRZ
Non-linear systems
Model parameter uncertainty
Model evaluation
Large ensembles/Scenario uncertainty
Downscaling & Data assimilation & Emergent constraints

DISCLAIMER

This lecture is based on the WS2019/20 course by **Dr. Ana Bastos**, now professor at university of Leipzig.

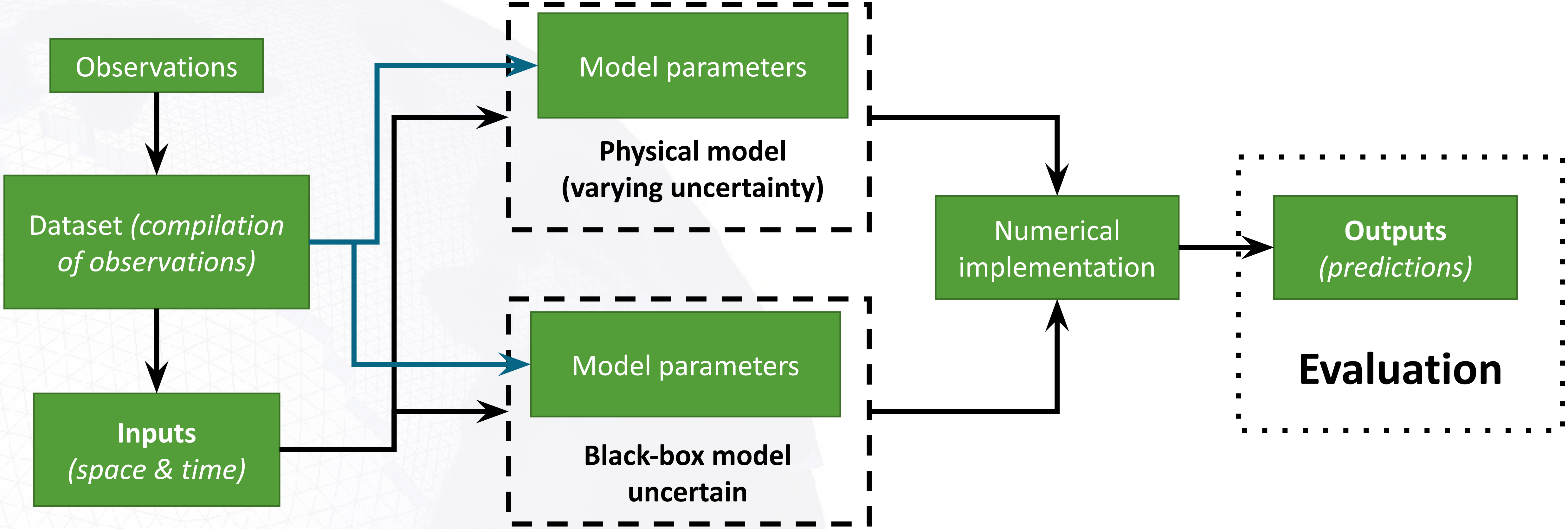
Co-developed with **Dr. Tammas Loughran** (VL+P), expert on massive climate modeling, extreme events, climate variability.

Additional input stems from ETH colleagues **Prof. Dr. Reto Knutti** (climate physics) and **Prof. Dr. David Bresch** (weather and climate risks).

Latest add-ons by **Prof. Julia Pongratz** (WS20/21) and **Dr. Stefanie Falk** (WS21/22-WS23/24).

In WS24/25 this lecture is given by **Dr. Sabine Egerer**.

Model uncertainties



World of Ideas

World of ideas

Original - truth
(Urbild)


equineguelph.ca

Reality

Imitation - realizations
(Abbild)


zooroyal.de


(cc) Oliver Seltz


© Helmut Hartkopf

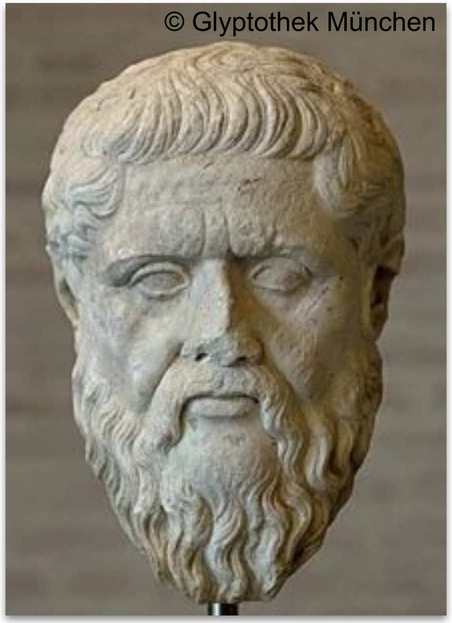
Symbiosis

Empirical -
Generalization

Observe *all horses in the world*

Knowledge

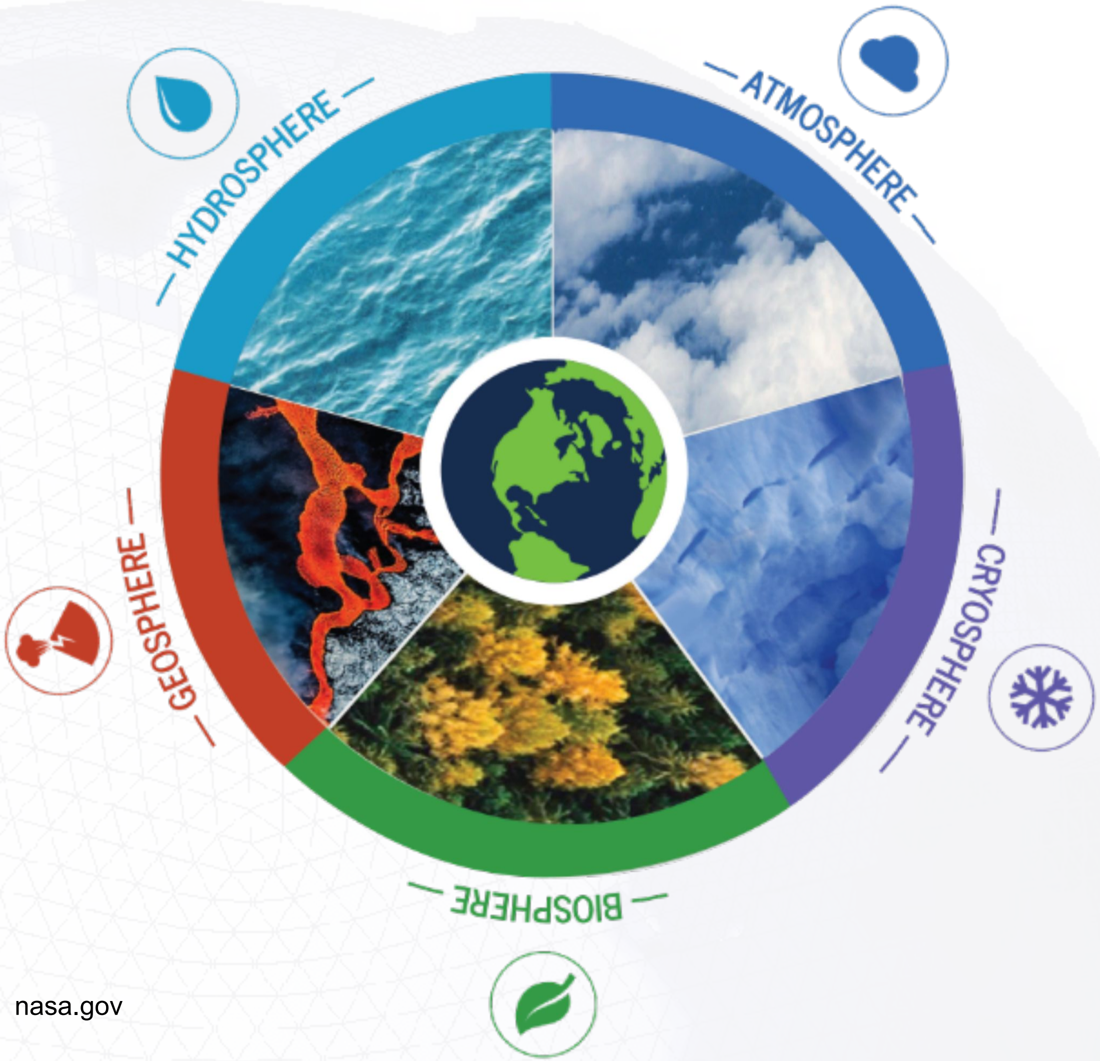
Approximation of truth


© Glyptothek München

Platon 427 - 423 BCE

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

What is “the environment”?



nasa.gov

1. (environment)

The surroundings or conditions in which a person, animal, or plant lives or operates.

“survival in an often hostile environment”

2. (the environment)

The natural world, as a whole or in a particular geographical area, especially as affected by human activity.

“the impact of pesticides on the environment”

Model types

Empirical (“black-box”) → “Gray box” model ← **Process-based** (“white box”)

Empirical (“black-box”)

Pros:

- **simplicity** (processes, mathematics)
- less data/computing intensive
- user-friendliness
- fast

Cons:

- **no process understanding**
- missing processes
- too crude approximations
- strong need for calibration/validation
- generalisation might be dangerous

Input → ??? → Result

Process-based (“white box”)

Pros:

- **realistic representation**
- parameters can be measured or derived from experiments
- obey fundamental physics
- theoretically do not need calibration
- can be generalised (“laws”)

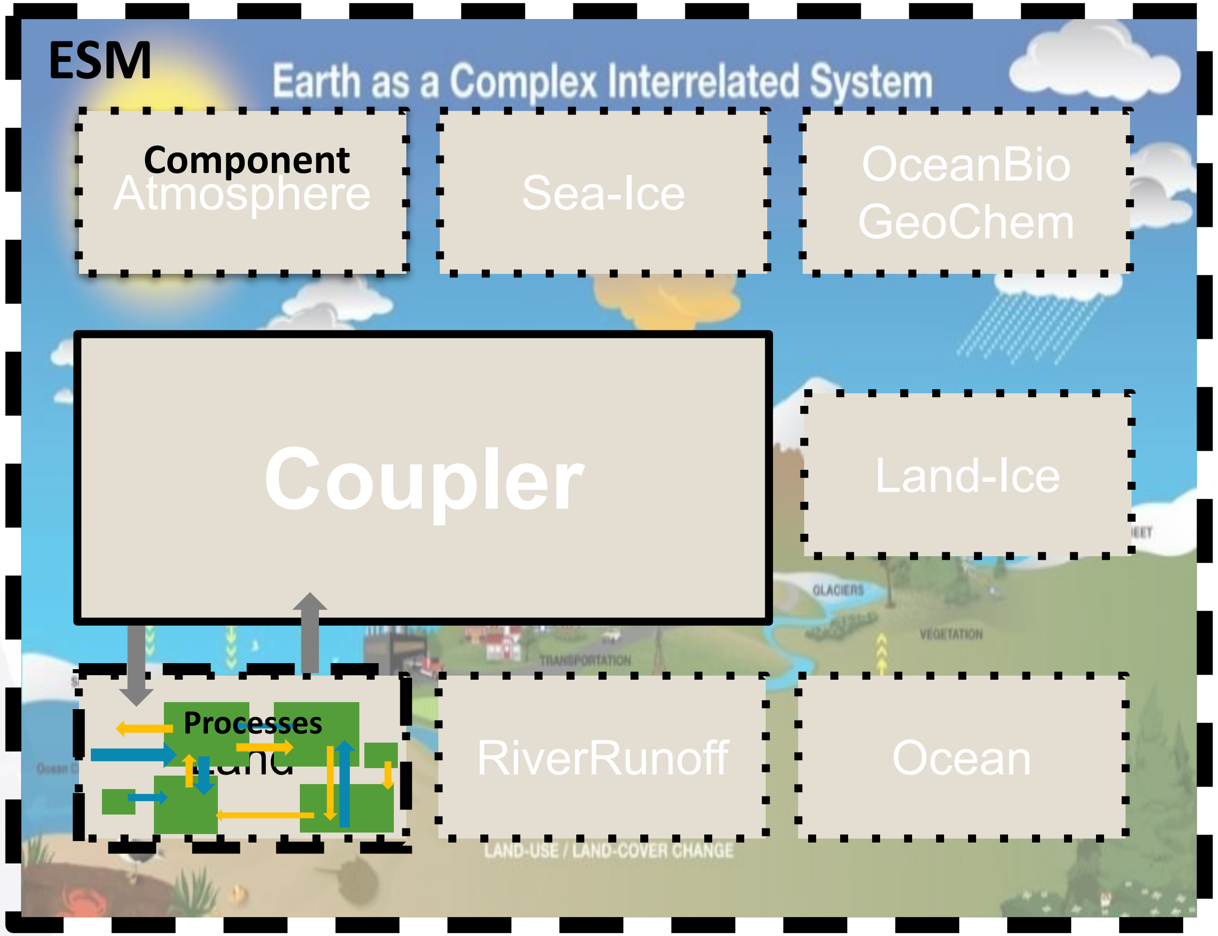
Cons:

- **mathematically complex**
- data/computing intensive
- non-linearity

Input → *f*(input) → Result

Earth System Models

- Models are always **simplifications** of the processes occurring in nature
- Model development necessarily involves making **decisions / choices**;
- Model choice needs to consider **trade-offs** (e.g. computing power, spatial scale…)
- **Uncertainty quantification** (if possible) is fundamental to guide model choice and development.



nasa.gov, cesm.ucar.edu adapted

Sources of uncertainty

- From errors in **input data**:
  - Measurements
  - Production of datasets
  - Processing for model input
- From calibration of **model parameters**
- From numerical **implementation**
- From the interactions between the above terms

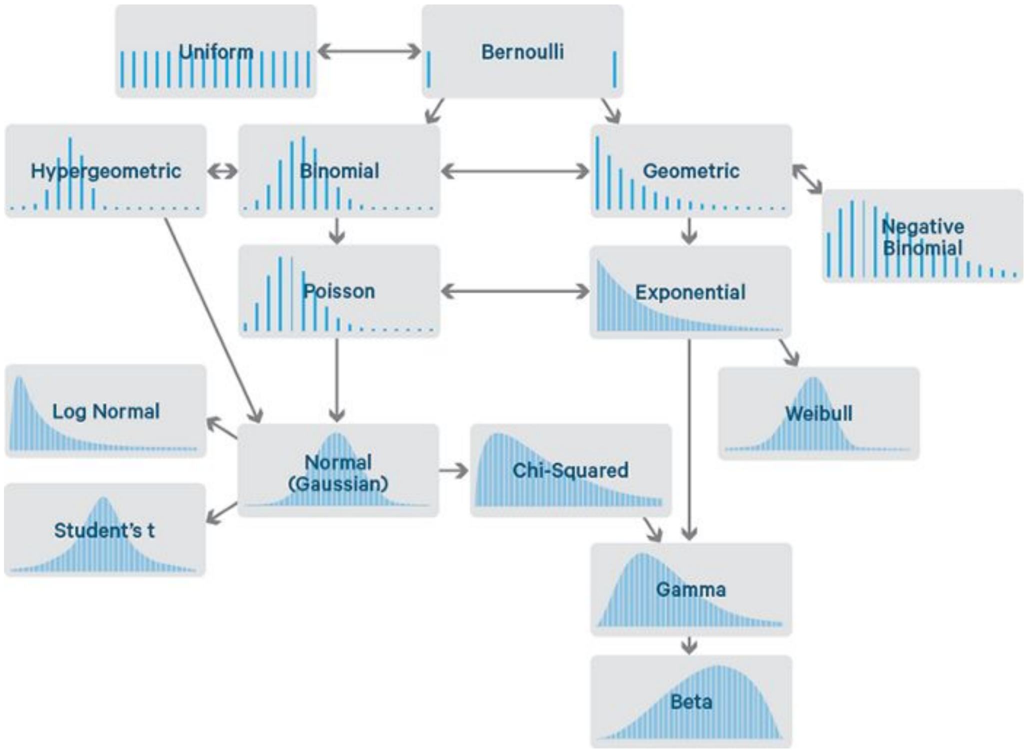
In principle quantifiable
(but not always)

- But also from system simplification, approximations and missing processes.

In principle <u>not</u> quantifiable

Statistical concepts

- Statistics allow characterising distributions of variables, and making **inferences about large populations from smaller samples** (e.g. election polls)
- Statistic analysis of measurements / observations allows **quantifying uncertainty** and **identifying** potential **measurements errors**
- **Bayes' Theorem** allows quantifying **how much we learn** about a given process by a **given information/observation**.

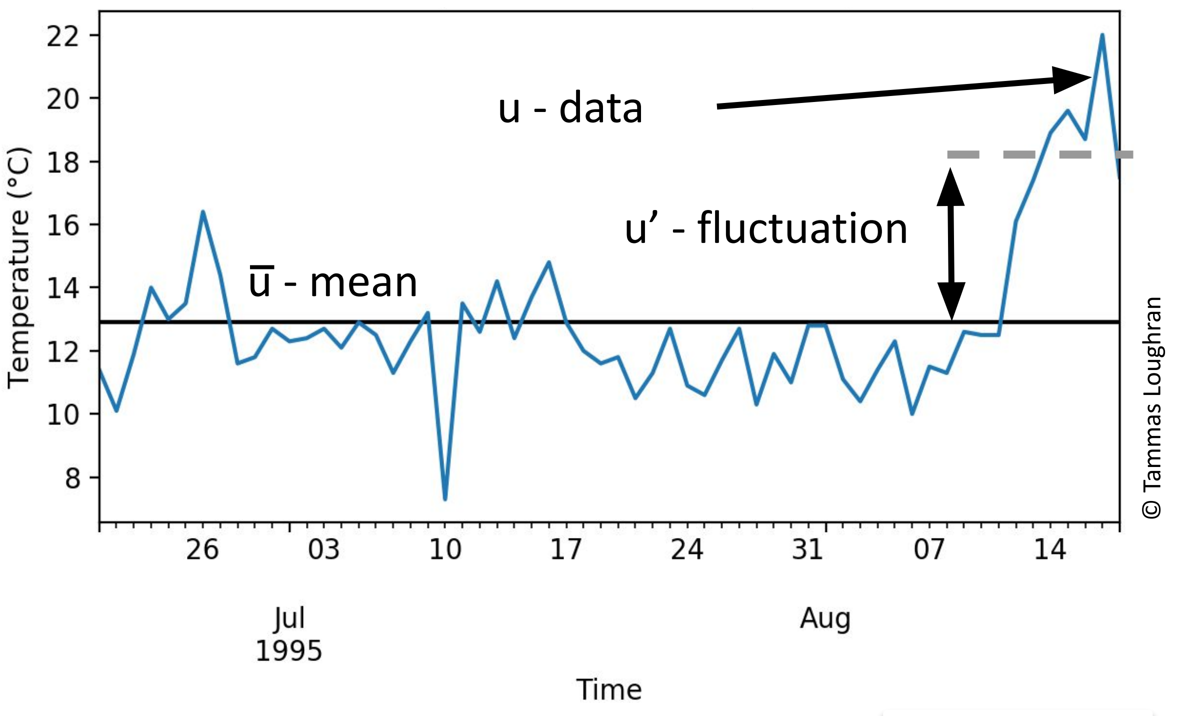


towardsdatascience.com

Gap filling

- Gaps exist because of
  - System / instrument malfunction
  - Data corruption / poorly recorded data
  - Social / financial limitations
  - (industrial strike/war)
- Gaps should be filled because
  - Some analysis methods require **complete data**
  - Gaps reduce **robustness** of analysis
  - Easier **comparison** to dynamical model data
  - **Forcing** dynamical models
- Simple gap filling methods
  - **Single imputation**: replace missing values with single value
  - **Multiple imputation**: replace missing values with multiple values
- Reynold's decomposition for reconstruction of ozone concentrations

Reynold's decomposition



$u = u + u'$

Non-linear systems

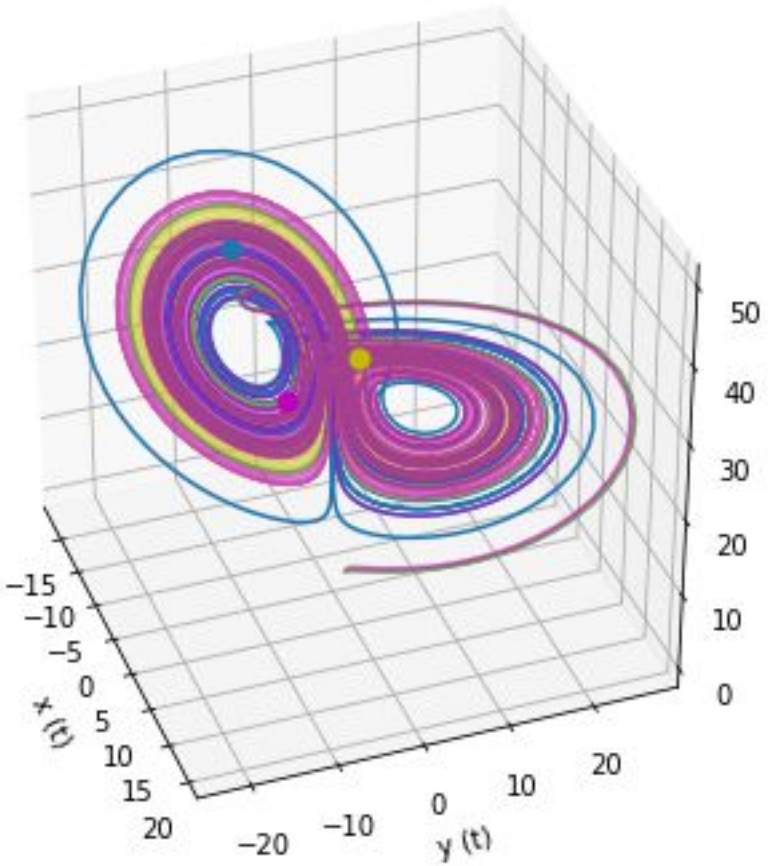
- **Boundary conditions** and **initial conditions** define the chaotic behavior of a system.
- **Initial conditions** are most effective on **short and medium time** scales.
- Over **longer time boundary conditions** will define the development of the system.

$$\frac{dx}{dt} = \sigma(y - x),$$

$$\frac{dy}{dt} = x(\rho - z) - y,$$

$$\frac{dz}{dt} = xy - \beta z.$$

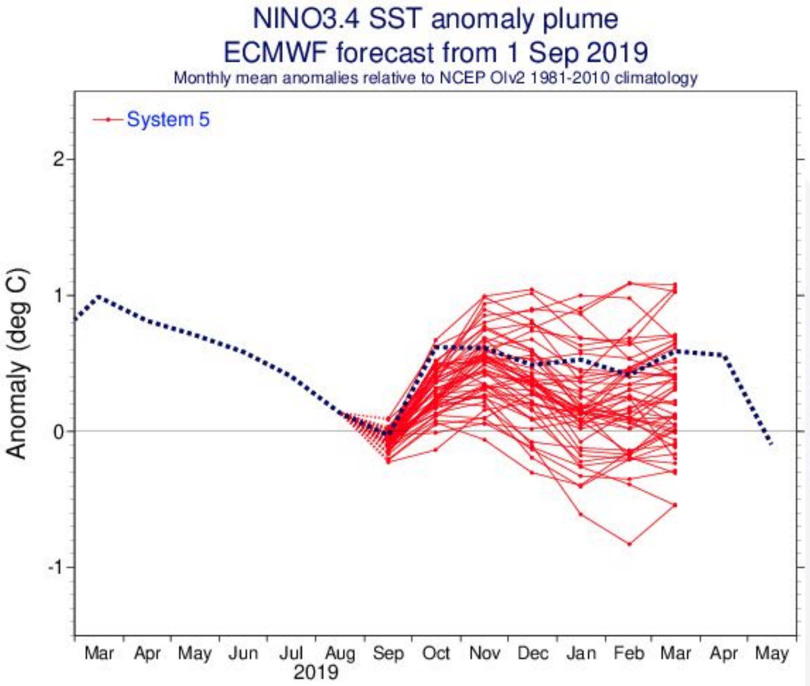
Lorenz Attractor



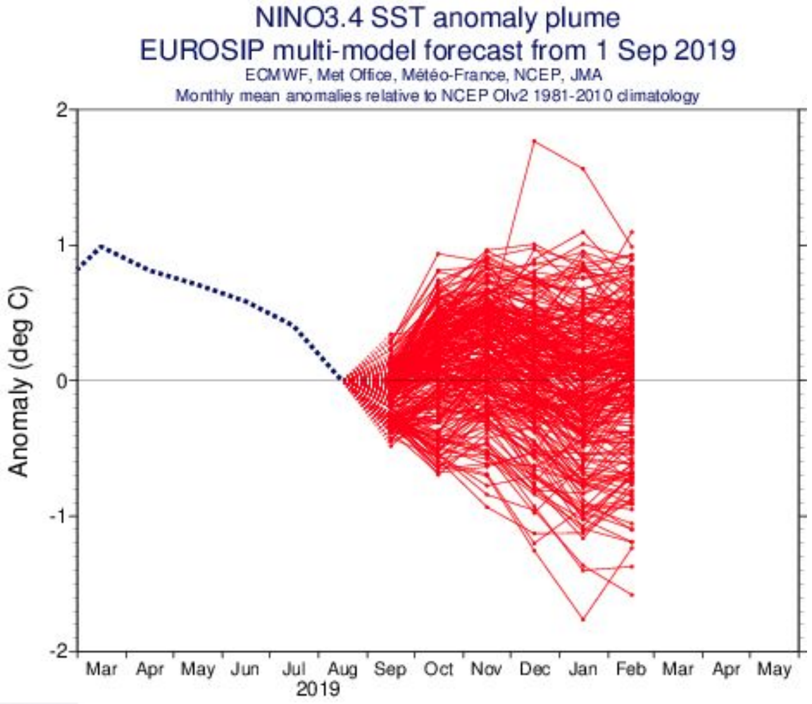
Zachary G. Nicolaou et al. (arXiv, 2023)

Model ensembles

An ensemble is **a group of model integrations** of the **<u>same model</u>** with ***different initial*** AND/OR ***boundary conditions*** OR of **<u>different models</u>**.



ECMWF



ECMWF

+ UK Met Office + NCEP
+ Météo-France + JMA

Copernicus - climate change service: ECMWF SST indices

Long range (seasonal) forecast

Error propagation: Analytical

More generally, we can define a **multivariate distribution**:

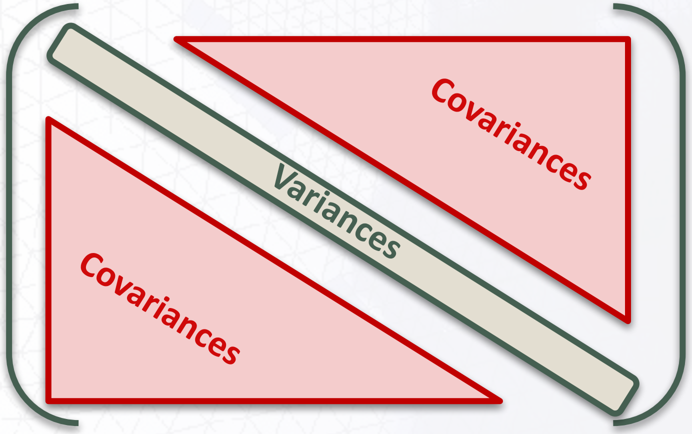
$$Q = f(\underbrace{x, y, z, \dots, n}_{S})$$

$$Q \equiv f(\mu_x, \mu_y, \mu_z, \dots, , \mu_n)$$

If errors are correlated:

$$\sigma_Q^2 \approx \boxed{\left(\frac{\partial Q}{\partial x}\right)^2_{\mu_x} \cdot \sigma_x^2 + \left(\frac{\partial Q}{\partial y}\right)^2_{\mu_y} \cdot \sigma_y^2 + \left(\frac{\partial Q}{\partial z}\right)^2_{\mu_z} \cdot \sigma_z^2} + \cdots + 2\sigma_{xy}\left(\frac{\partial Q}{\partial x}\right)\left(\frac{\partial Q}{\partial y}\right)$$
$$+2\sigma_{xz}\left(\frac{\partial Q}{\partial x}\right)\left(\frac{\partial Q}{\partial z}\right) + \dots$$

cov(***X***, ***Y***) =



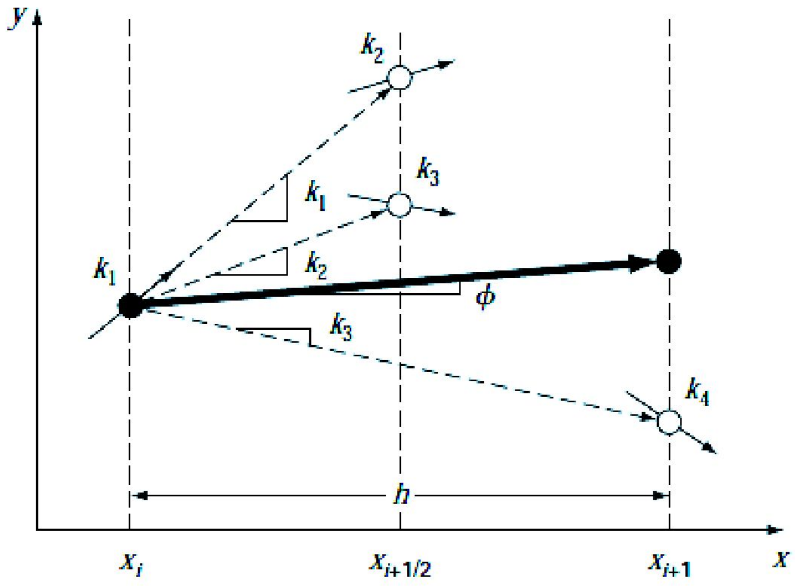
Variances: σ^2_x, σ^2_y, σ^2_z, …, σ^2_n

Covariances: σ(x,y), σ(y,x), σ(x,z), σ(z,x), …, σ(x,n), σ(n,x), …

Error propagation: Numerical

- Developing “white-box” models often requires **solving differential equations**;
- **Numerical methods** can be used to **approximate** solutions for **initial value problems**, especially for complex equations (or systems of DEs);
- Numerical approximations always involve errors “**truncation errors**”;
- The **Euler Method** approximates the solution by a straight line at each point of a discrete interval, but the **error is high** for **sparse intervals**;
- The **RK-4 Method** requires more calculations at each point (weighted average of the slope between two points), but is **more accurate.**

Runge-Kutta Method



Chapra & Canale, Numerical Methods for Engineers, 6th ed, 2010, McGraw Hill

Down scaling

- Dynamic Downscaling + RCM simulations
- Empirical-Dynamic Downscaling and transfer functions
- Disaggregation Methods (spatial + temporal)
- Hybrid Downscaling Approaches

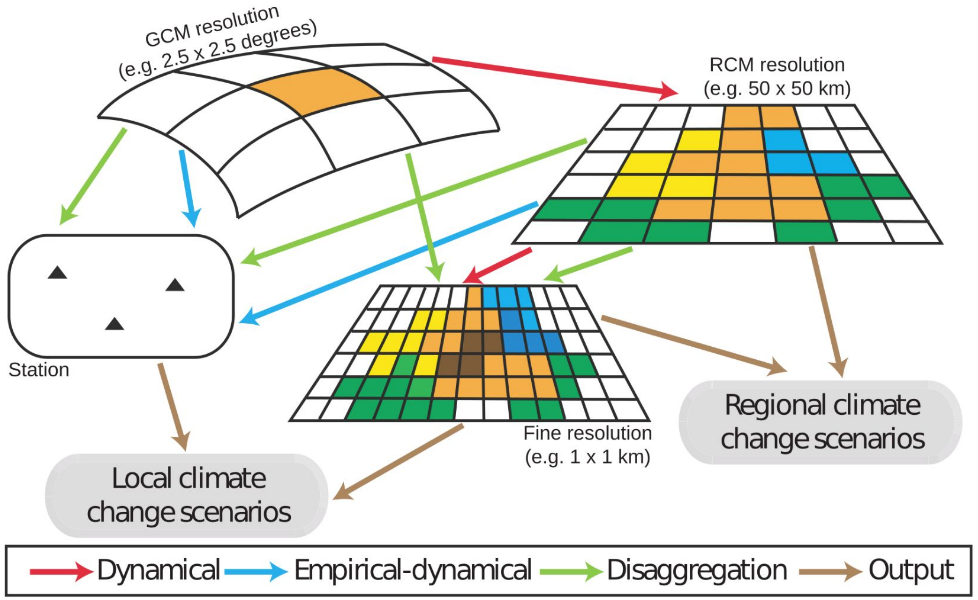


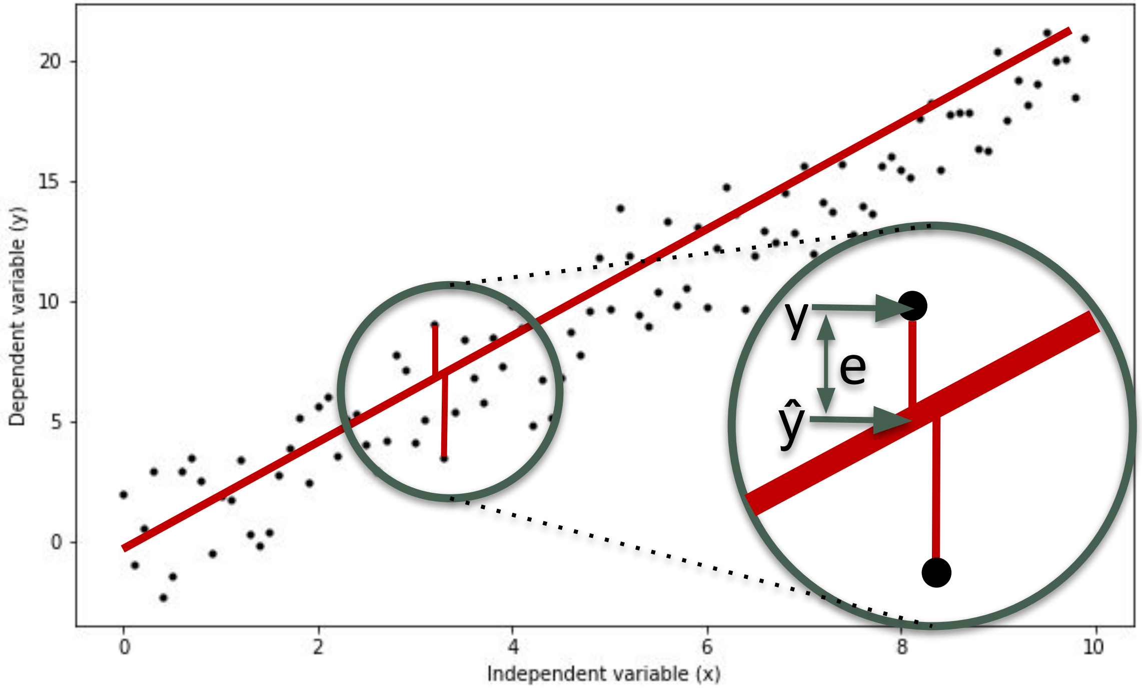
Fig. 2. Schematic of the outputs of dynamic downscaling, empirical-dynamic downscaling and disaggregation downscaling methods when applied to GCM simulations. The products from the downscaling can be gridded f elds of climate variables at a range of spatial scales or climate scenarios for specif c locations. Different approaches to downscaling can be applied, as shown by the colored arrows. Also, multiple downscaling steps can be used to obtain the desired spatial resolution (RCM, regional climate model; GCM, global climate model).

Winkler et al (2011)

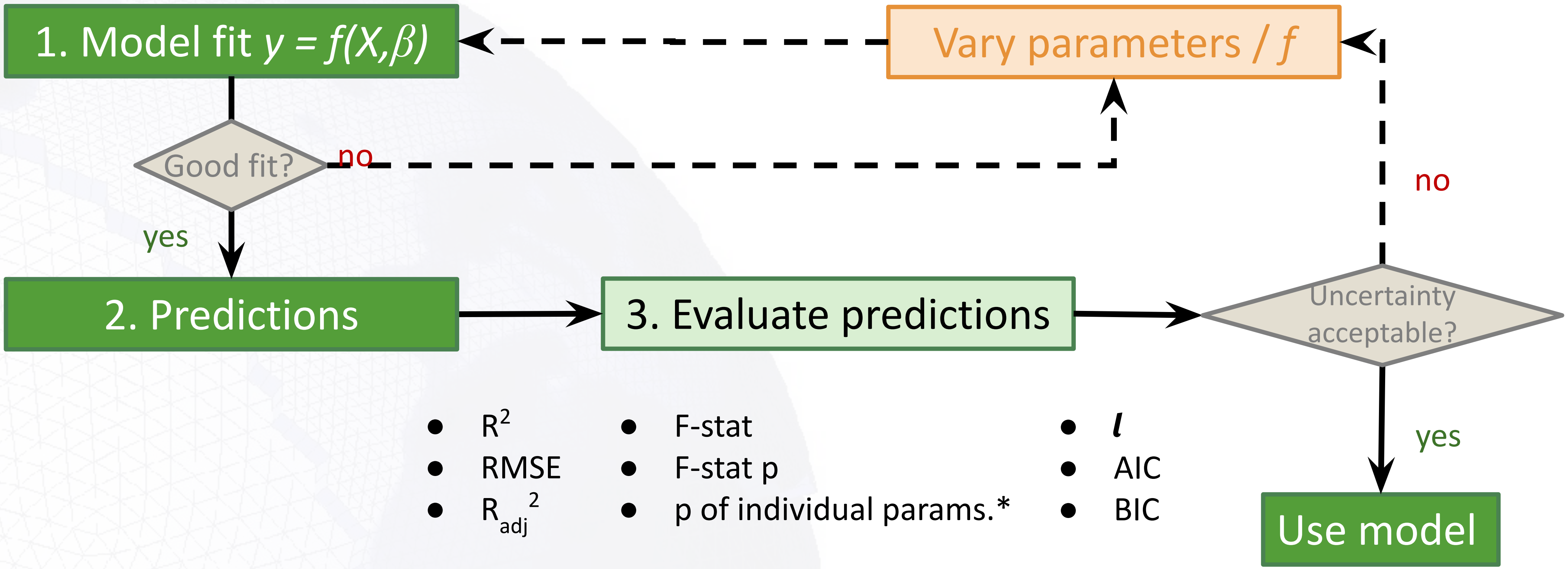
Model parameter errors

- **Black-box models** rely on empirical relationships between variables derived from observations;
- The required sample size to achieve a given uncertainty can be estimated;
- Coefficients of model fitting are always uncertain, since data is uncertain;
- OLS approaches work well for linear problems with normal errors;
- For more complex problems, Bayesian Regression is a powerful tool.

Linear regression



Model evaluation



- R^2
- RMSE
- R_{adj}^2
- F-stat
- F-stat p
- p of individual params.*
- ℓ
- AIC
- BIC

LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

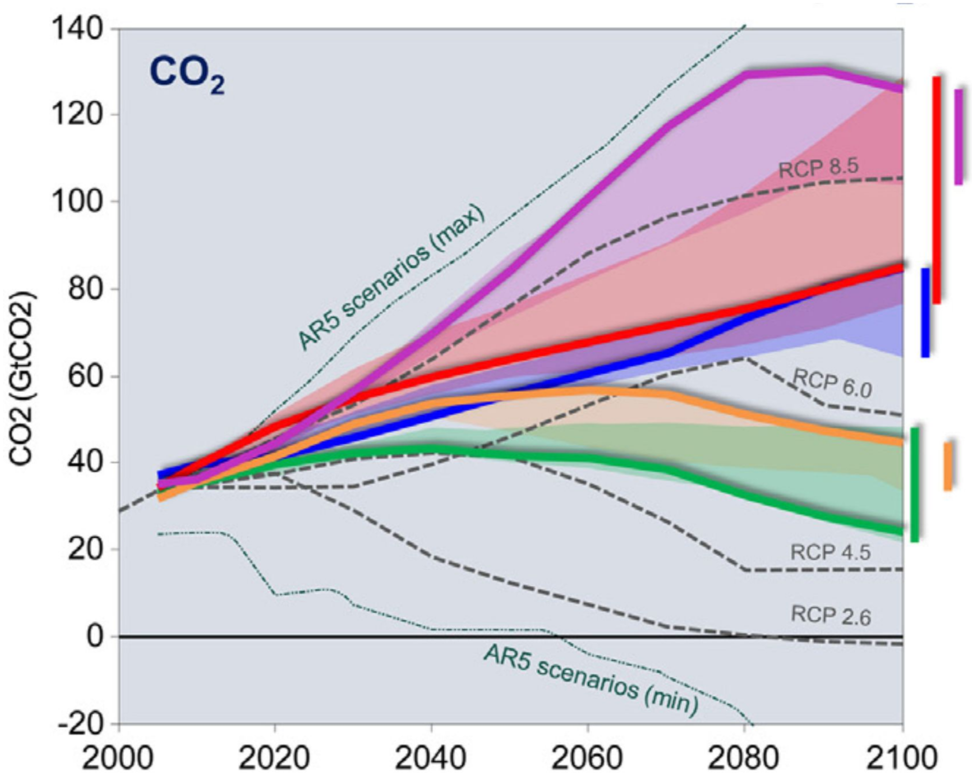
Scenario uncertainties

- 3 different qualities of uncertainties:

1) **Internal climate variability** that we cannot influence

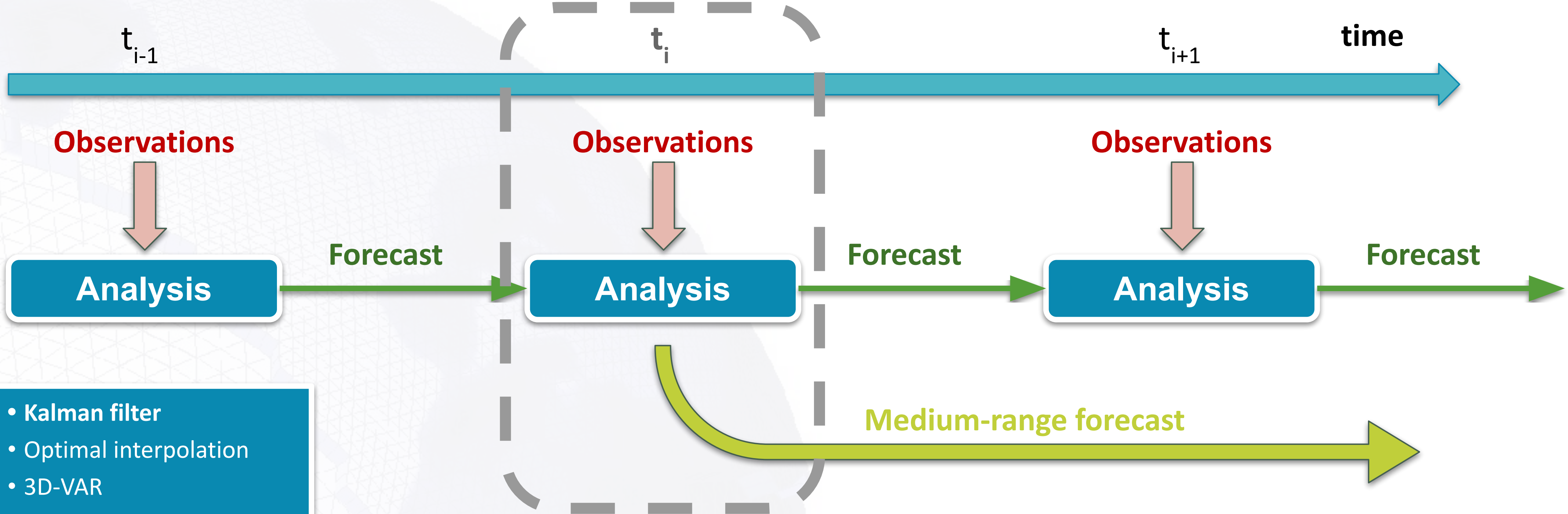
1) **Model** (ESM and IAM) **uncertainty** that we can try to reduce
(but unavoidable errors due to truncation etc will always persist)

1) **Scenario uncertainty**, likelihoods are harder to assign, prediction almost impossible
(Which technology will be available? How does the political landscape look like?).
This is ongoing field of research, but important for adaptation and mitigation.



Riahi et al. (2017)

Data assimilation



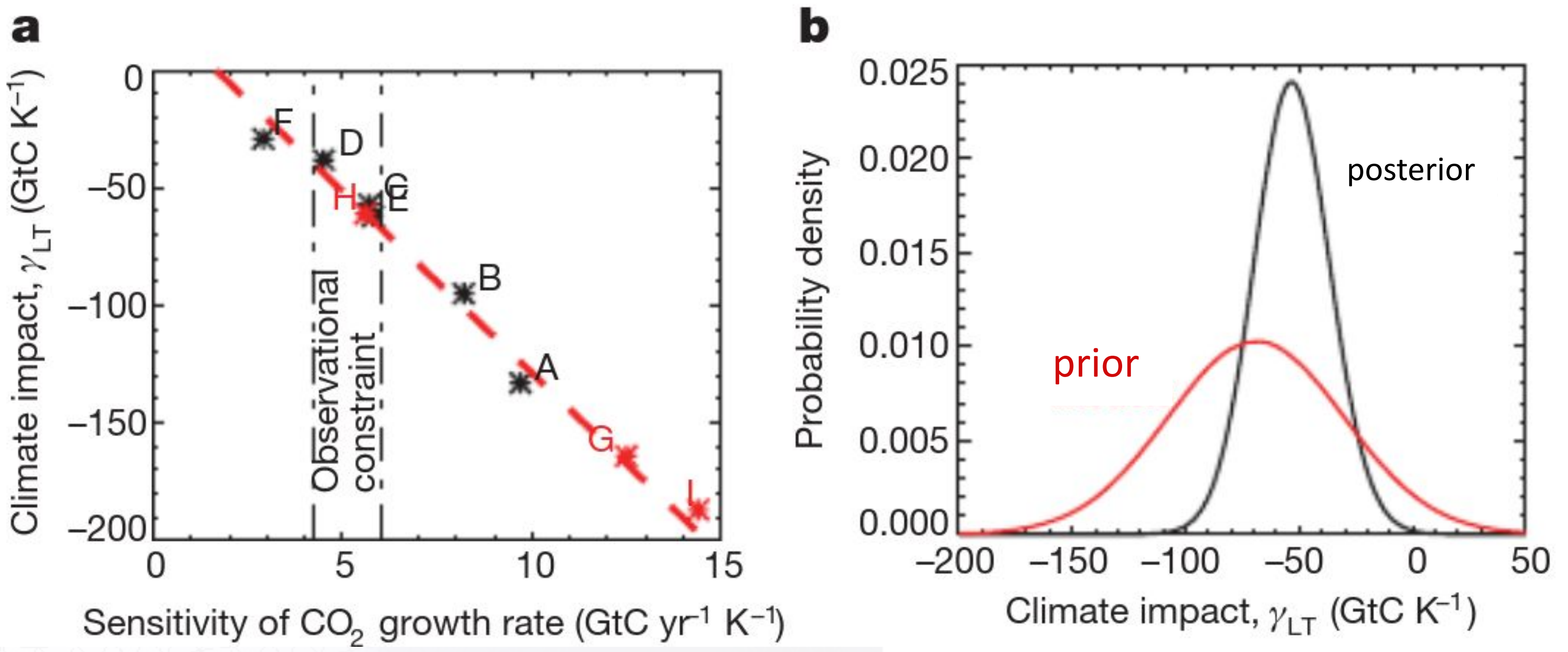
- **Kalman filter**
- Optimal interpolation
- 3D-VAR
- PSAS
- **4D-VAR**
- Extended Kalman Filtering

based on ECWMF - E-learning

Emergent constraints

*If observations are available, then future model projections might be weighted in order to favor those models better matching observations, and **penalize those further away**.*

This has been attempted for the global ecosystem C balance, by constraining models with sensitivity of CO_2 growth rate to tropical temperature



Cox et al. (Nature, 2023)