Climate Informatics Opportunities for Earth system model evaluation

Veronika Eyring

Deutsches Zentrum für Luft- und Raumfahrt e.V. (DLR), Institute of Atmospheric Physics, Oberpfaffenhofen, Germany

20 June 2017

1st Workshop on Environmental Informatic Challenges

Jena, Germany

DLR

Outline

1. Introduction

- Multi-model climate projections
- How to gain confidence in Earth system models (ESMs)?

2. Overview Coupled Model Intercomparison Project Phase 6 (CMIP6)

3. Climate Informatics Opportunities for Earth system model evaluation

- Data Management for efficient & more routine ESM evaluation with observations
- Data Analysis with data science methods

4. Summary

1. Introduction

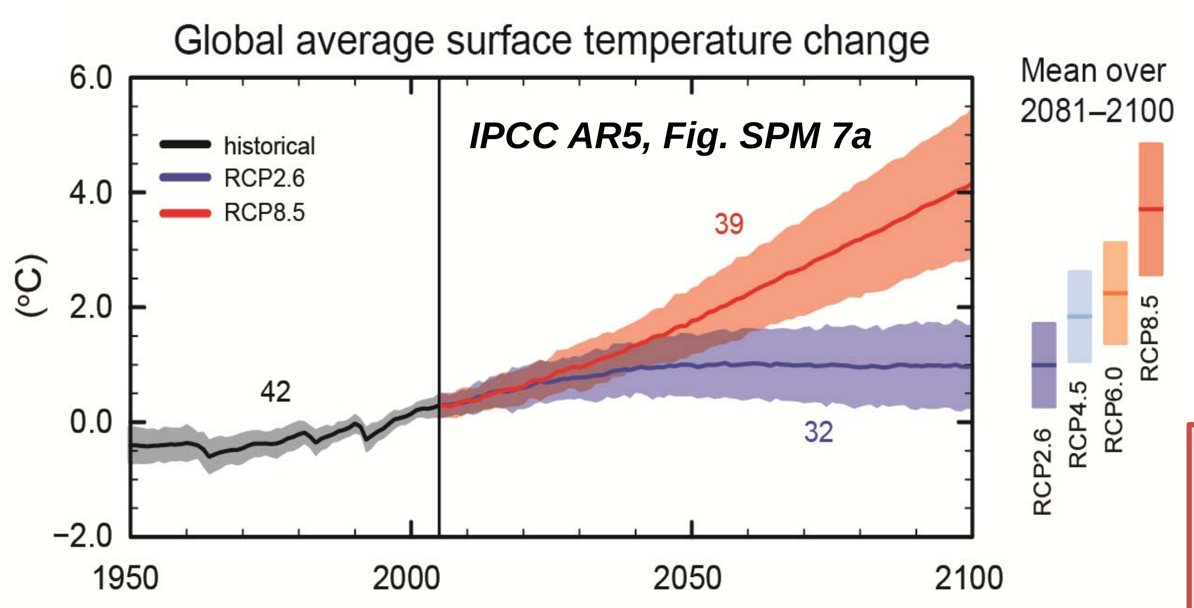
Climate Model Projections

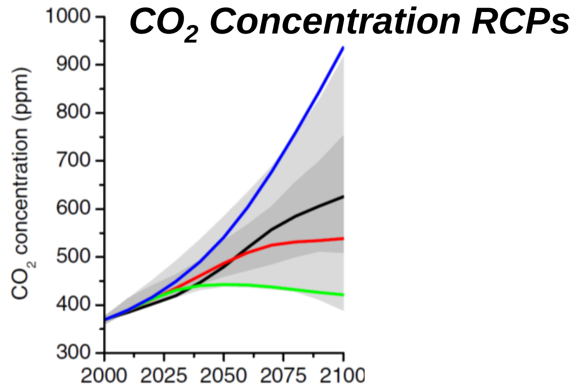
➢ Climate projections in the **IPCC Fifth Assessment Report (AR5)** largely based on climate model output coordinated by the World Climate Research Programme (WCRP) **Coupled Model Intercomparison Project Phase 5 (CMIP5)**.

➢ **The objective of WCRP's CMIP** is to better understand past, present and future climate change in a multi-model context by defining common experiment protocols, forcings and output.

➢ **Multi-model mean and spread is commonly used for climate projections**

1. Spread of model ensemble often used as first-order estimate of projection uncertainty.
2. Used to determine why similarly forced models produce a range of responses.





Distillation of robust information from multi-model output is needed for science and as evidence for policy-making

Climate Model Projections WG I AR5

- Largely based on Coupled Model Intercomparison Phase 5 (CMIP5) simulations -

Relative to the 1986-2005 average

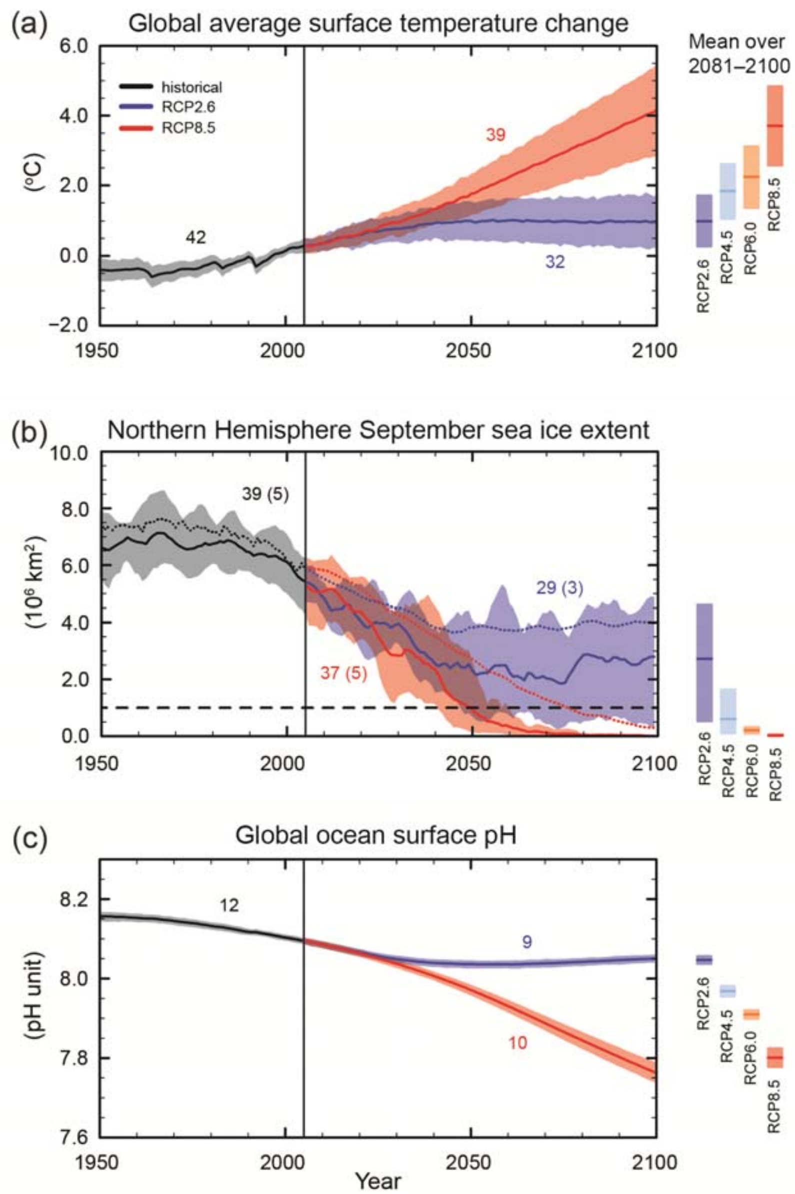


Figure SPM.7

WG I

Paleoclimate Archives (Chapter 5)

Process Understanding

- Chapter 6: Carbon and other Biogeochemical Cycles
- Chapter 7: Clouds and Aerosols

From Forcing to Attribution of Climate Change

- Chapter 8: Anthropogenic& Natural Radiative Forcing
- Chapter 9: Evaluation of Climate Models
- Chapter 10: Detection and Attribution of Climate Change: from Global to Regional

Future Climate Change and Predictability

- Chapter 11: Near-term Climate Change
- Chapter 12: Long-term Climate Change: Projections, Commitments and Reversibility

Integration

- Chapter 13: Sea Level Change
- Chapter 14: Climate Phenomena and their Relevance for Future Regional Climate Change

Atlas of Global and Regional Climate Projections

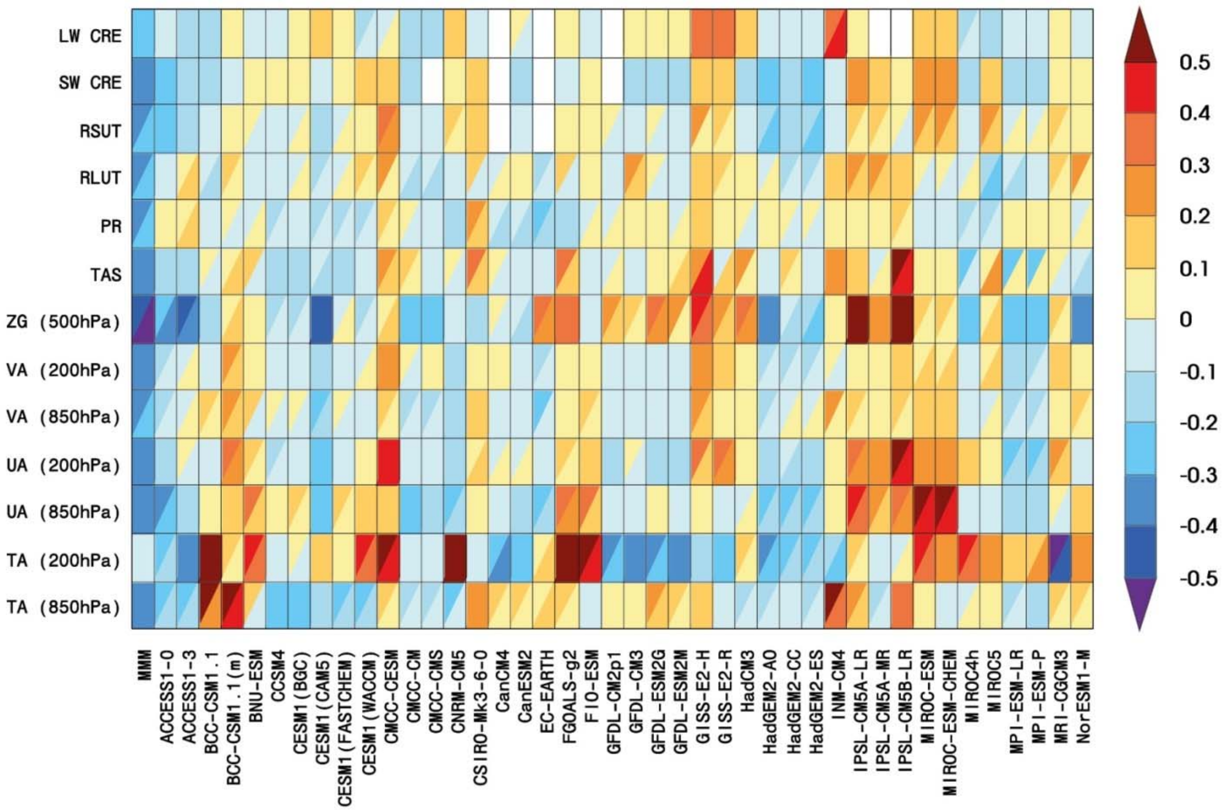
Process understanding and projections including uncertainty estimates also relevant for

WG II and III

INTERGOVERNMENTAL PANEL ON climate change

How do we gain confidence in climate model projections?

- Based on physical understanding of the climate system and its representation in climate models, and
- On a demonstration of how well models represent a wide range of processes and climate characteristics on various spatial and temporal scales



Relative error measures of CMIP5 model performance (normalized by the median error of all model results), based on the global seasonal-cycle climatology (1980–2005)

> ➢ **Climate models have continued to be developed and improved since the AR4.**

IPCC AR5 Chapter 9, Fig. 9.7

Equilibrium Climate Sensitivity Remains Uncertain

Defined as the **change in global mean surface temperature at equilibrium that is caused by a doubling of the atmospheric CO_2 concentration**.

	TAR	AR4	AR5
ECS	Likely range: 1.5 to 4.5°C	*likely* range: 2.0 to 4.5°C *very unlikely* <1.5°C – best estimate about 3°C	***likely* range: 1.5 to 4.5°C** ***extremely unlikely* <1.0°C** ***very unlikely* >6.0°C** –

The model spread in ECS ranges from **2.1°C to 4.7°C** and is very similar to the assessment in AR4 (IPCC AR5, Chapter 9).

Large Uncertainty Remains in Some Projected Variables

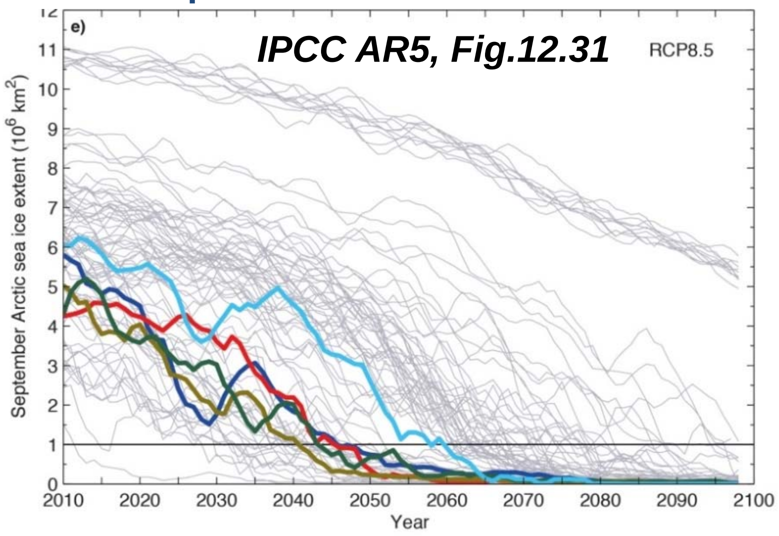
Is the multi-model mean always the best measure?

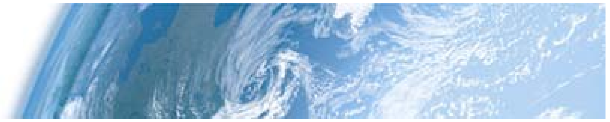
The spread of an ensemble of models is often used as a first-order estimate of projection uncertainty

– Despite the fact that models differ in terms of resolution, processes and components included, and agreement with observations.

– Despite there is inter-model dependence

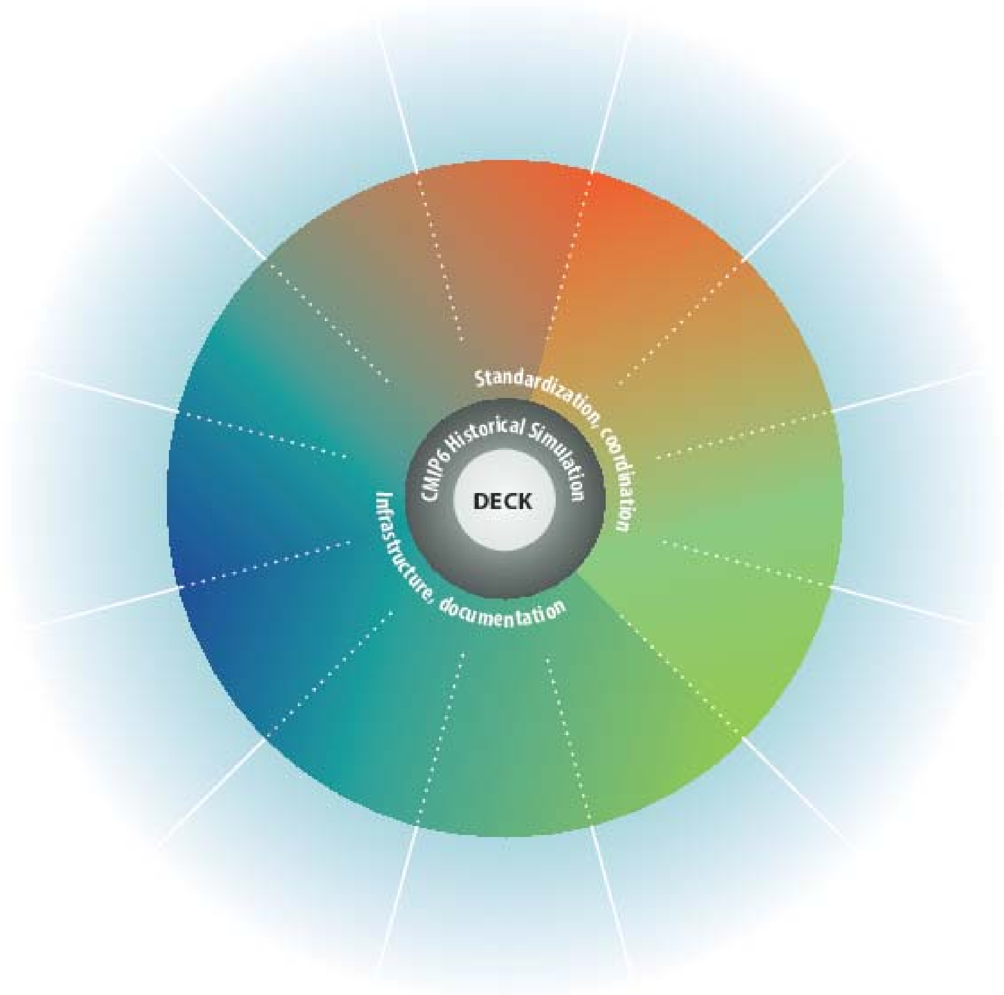
September Arctic sea ice extent





2. CMIP6 Organization and Design

CMIP: a More Continuous and Distributed Organization



(1) A handful of common experiments

DECK (entry card for CMIP)
For enhanced model evaluation and evaluation consistency

i. AMIP simulation (~1979-2014)
ii. Pre-industrial control simulation
iii. 1%/yr CO_2 increase
iv. Abrupt 4xCO_2 run

CMIP6 Historical Simulation (entry card for CMIP6)

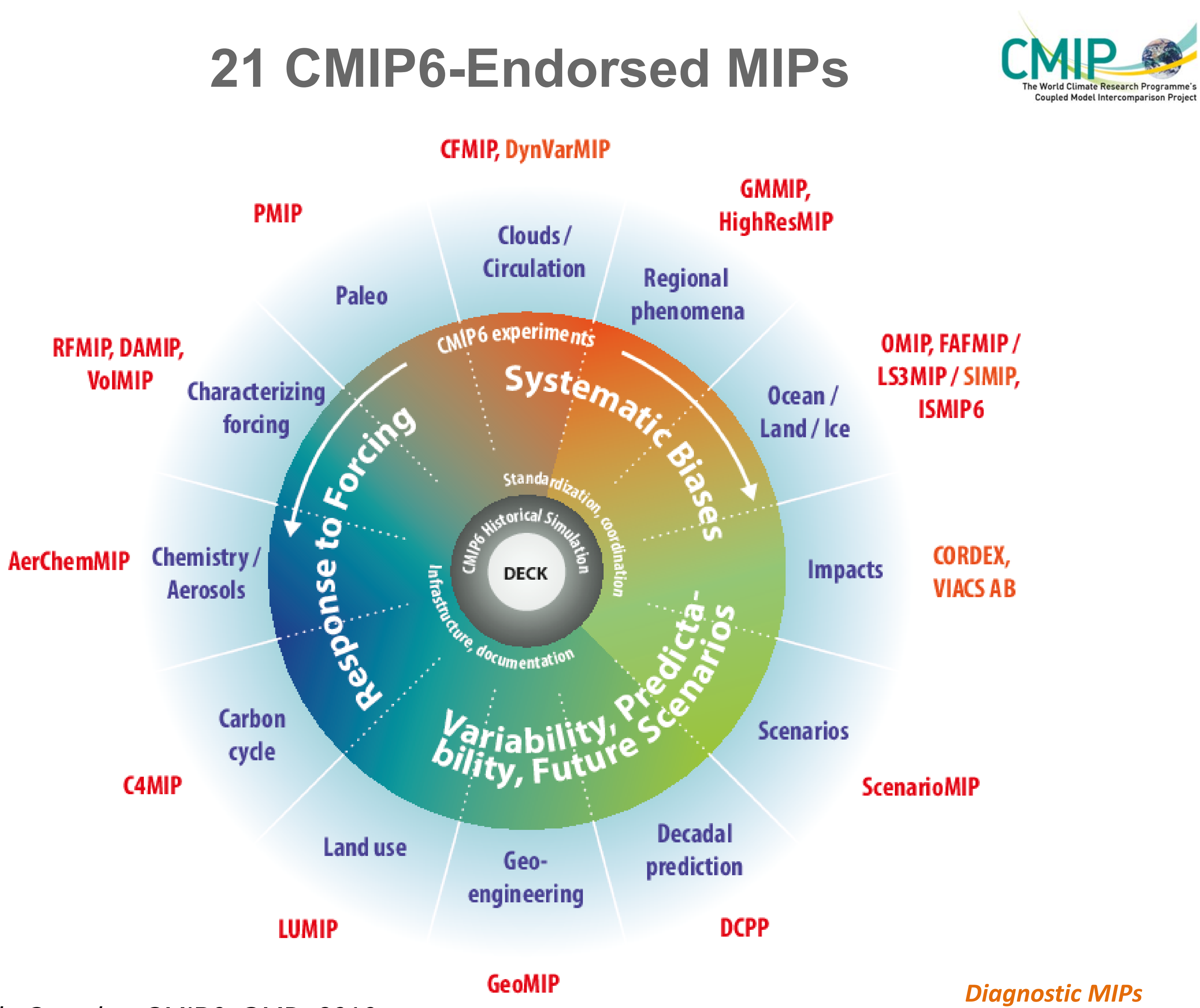
v. Historical simulation using CMIP6 forcings (1850-2014)

(2) Standardization, coordination, infrastructure, documentation

DECK (Diagnosis, Evaluation, and Characterization of Klima) & CMIP6 Historical Simulation to be run for each model configuration used in CMIP6-Endorsed MIPs

Eyring et al., Overview CMIP6, GMD, 2016

21 CMIP6-Endorsed MIPs



Eyring et al., Overview CMIP6, GMD, 2016

CMIP6: Participating Model Groups

	Institution	Country		Institution	Country		Institution	Country
1	AWI	Germany	12	DOE	USA	23	MRI	Japan
2	BCC	China	13	EC-Earth-Cons	Europe	24	NASA-GISS	USA
3	BNU	China	14	FGOALS	China	25	NCAR	USA
4	CAMS	China	15	FIO-RONM	China	26	NCC	Norway
5	CasESM	China	16	INM	Russia	27	NERC	UK
6	CCCma	Canada	17	INPE	Brazil	28	NIMS-KMA	Republic of Korea
7	CCCR-IITM	India	18	IPSL	France	29	NOAA-GFDL	USA
8	CMCC	Italy	19	MESSY-Cons	Germany	30	NUIST	China
9	CNRM	France	20	MIROC	Japan	31	TaiESM	Taiwan, China
10	CSIR-CSIRO	South Africa	21	MOHC	UK	32	THU	China
11	CSIRO-BOM	Australia	22	MPI-M	Germany			

New in CMIP:

2 new model groups from Germany (AWI, MESSY-Consortium)
4 new model groups from China (CAMS, CasESM, NUIST, THU)
1 new model group from Brazil (INPE)
1 new model group from India (CCCR-IITM)
1 new model group from Taiwan, China (TaiESM)
1 new model group from USA (DOE)
1 new model group from Republic of Korea (NIMS-KMA)
1 new model group from South Africa / Australia (CSIR-CSIRO)
==

⇒ **12 new model groups so far**

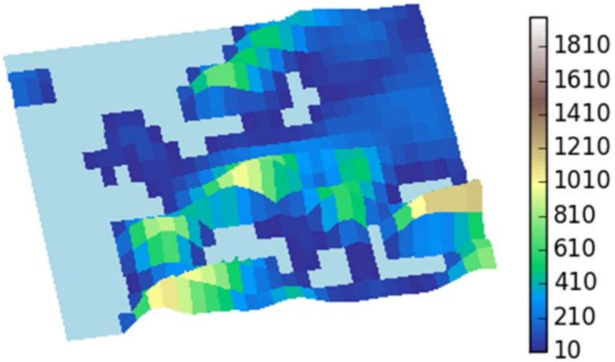
More models (>70)
New models
> 180 CMIP6 experiments

* Other models can join providing DECK and historical simulations are submitted

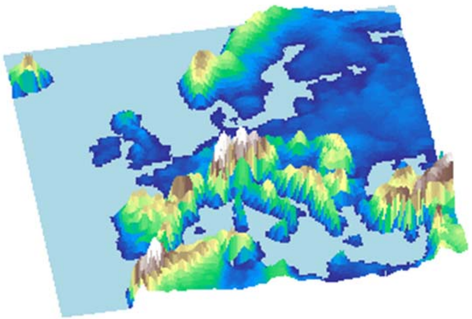
Models are Increasing in Complexity and Resolution

From AOGCMs to Earth System Models with biogeochemical cycles, from lowres to highres

130 km resolution orography

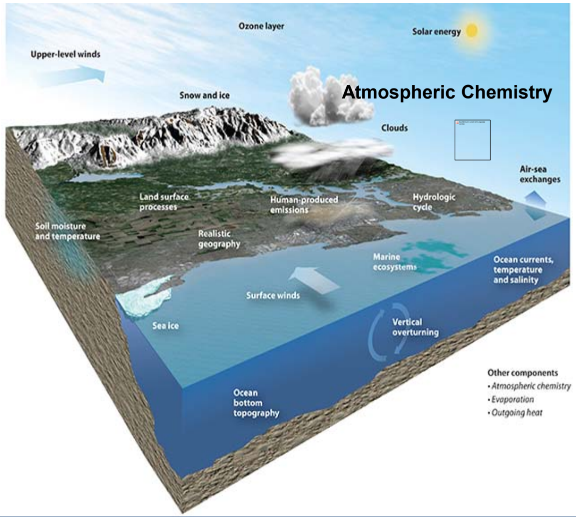


25 km resolution orography



I. Allows to study processes as horizontal resolution is increased to “weather-resolving” global model resolutions (~25km or finer)

II. Allows to study new physical & biogeochemical processes & feedbacks (e.g., carbon cycle, chemistry, aerosols, ice sheets)

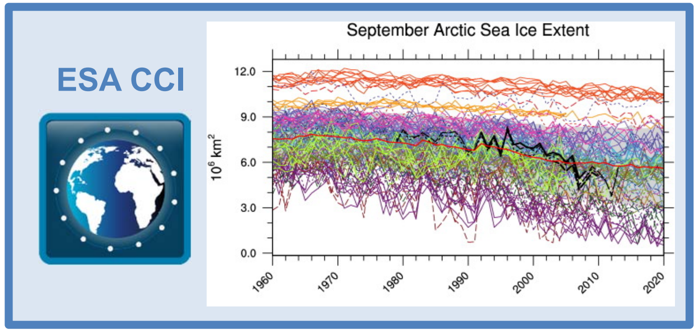


Increase in complexity and resolution
More (and new) models participating in CMIP6

- Increase in data volume (from ~2PB in CMIP5 to ~20-40 PB in CMIP6)
- Archiving, documenting, subsetting, supporting, distributing, and analysing the huge CMIP6 output will challenge the capacity and creativity of the largest data centres and fastest data networks, and the analysis of the data.

Earth System Observations

- **Earth System Observations**
  - Satellite measurements
  - Insitu data
  - Meteorological reanalyses



- The recent progress in climate science is producing **an unprecedented amount of data from climate models and observations**, giving us a unique opportunity to develop and apply methods for machine learning and data mining.
- The increasing desire for **operational analysis** means that a system has to be set in place that allows for an efficient and comprehensive analysis of the large volume of data from models and observations.

Climate Informatics

- Over the last decades, **the magnitude of climate data from satellite sensors and climate models has substantially increased**. This is starting **to overwhelm the relatively simple tools and methods** currently used to analyse the data.
- It is therefore essential to develop **an innovative and efficient computational approach** to address these analysis challenges.
- This relatively new field is called ***Climate Informatics*** and represents a promising and growing path of research, which could contribute to substantially enhance understanding of the Earth system and confidence in future climate projections.
  - The complexity and vast amount of Earth system data together with the pressing scientific and societal demand for more accurate climate projections, makes Earth system data an ideal candidate for **developing and applying computational intelligent analysis techniques to quickly sort through the data** from models and observations.
  - **New methods of Data Science** (e.g., data mining, artificial intelligence, machine learning) could potentially help finding new ways of analysing the data.
  - These techniques have been successfully applied in other natural sciences (e.g., biology), yet they are not fully exploited in climate research.

see also Monteleoni et al. (2013)

3. Climate Informatics Opportunities for Earth system model evaluation

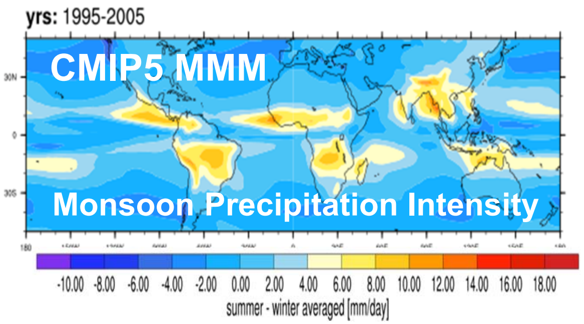
(A) Data Management for efficient and more routine ESM evaluation with observations

How to characterize the wide variety of models in CMIP6?

- Routine Benchmarking and Evaluation Central Part of CMIP6 -

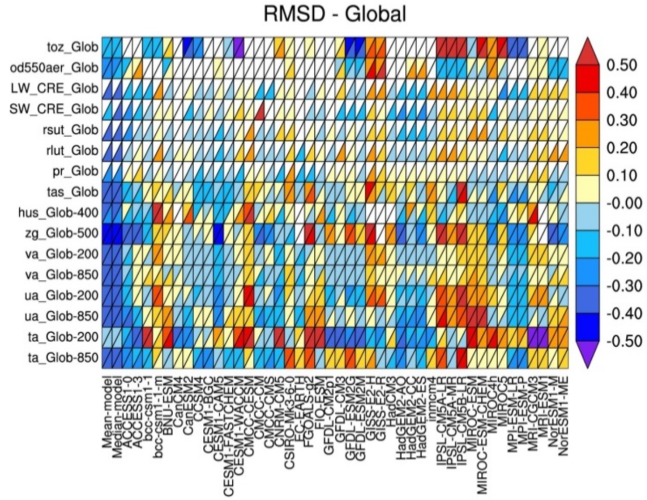
Earth System Model Evaluation Tool (ESMValTool) developed as community tool at DLR-IPA (PI) and many other institutes to produce well-established analyses as soon as CMIP model output is submitted

Similar to ***Figure 9.7*** *of AR5*



Broad Characterization of Model Behavior
(incl. IPCC AR5 Chap 9 & 12 diagnostics in ESMValTool)

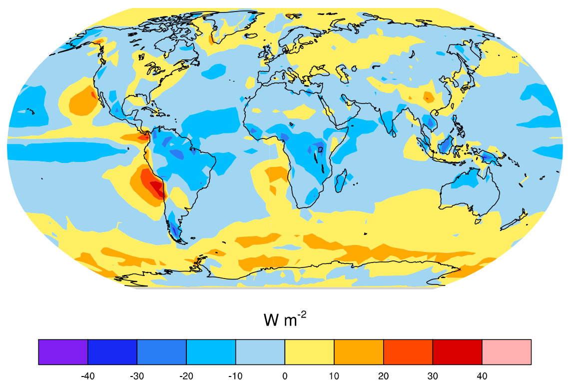
Running alongside the ESGF



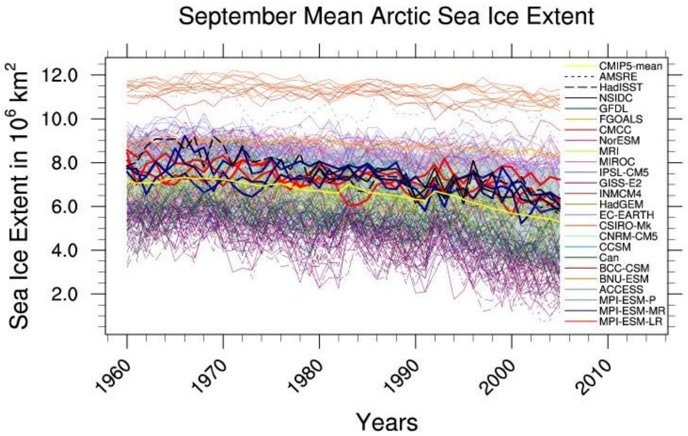
Link to projections

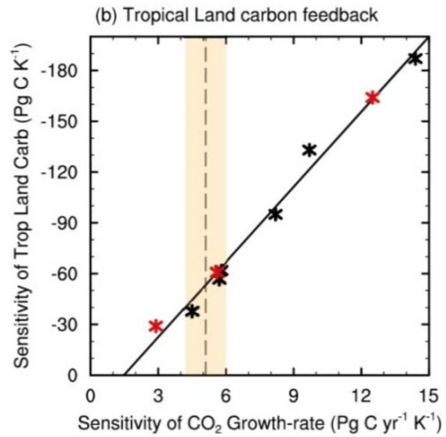
Similar to ***Figure 9.5*** *of AR5*

Net Cloud radiative effect against CERES EBAF



Similar to ***Figure 9.24*** *of AR5*





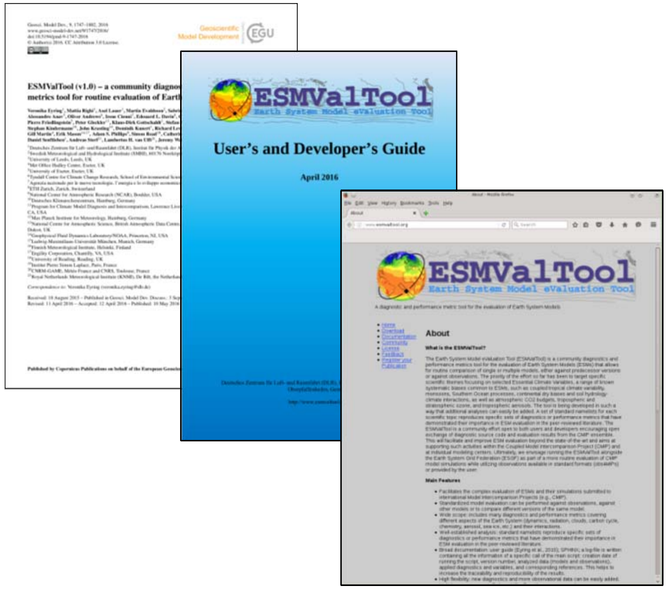
Similar to ***Figure 9.24*** *of AR5*

Eyring et al., ESMValTool version 1.0, GMD (2016)

ESMValTool version 1.0 released as open source software

http://www.esmvaltool.org/

Eyring et al., GMD, ESMValTool v1.0, 2016



GitHub public

GitHub private

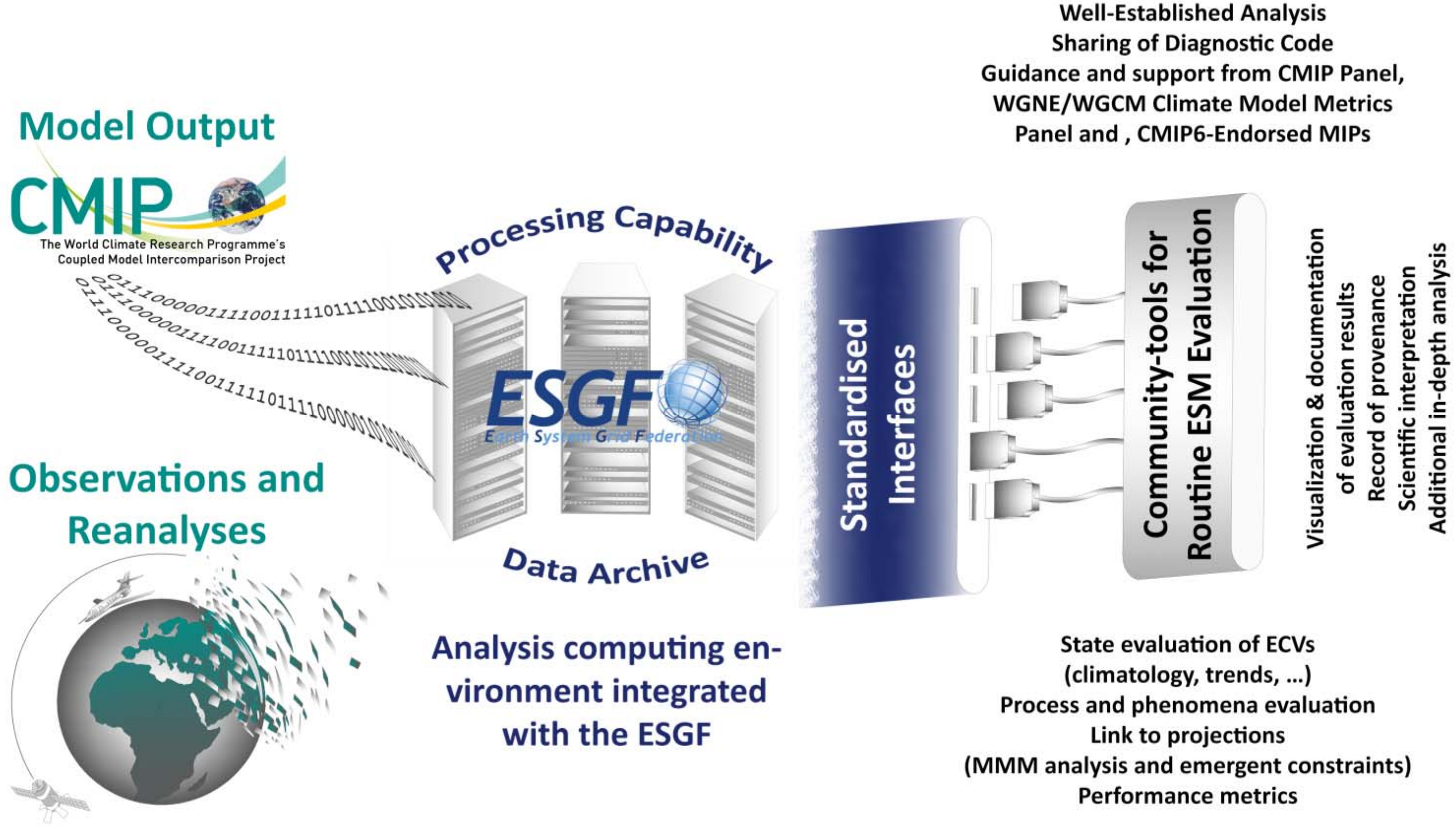
ESMValTool
Earth System Model eValuation Tool

- **Community diagnostics and performance metrics tool** for the evaluation of Earth System models
- **Open source code** based on python, NCL, R etc.
- **Standardized model evaluation** can be performed against observations, against other models or to compare different versions of the same model
- Many diagnostics and performance metrics covering **different aspects of the Earth System** (dynamics, radiation, clouds, carbon cycle, chemistry, aerosol, sea-ice, etc.) and their interactions
- Well-established analysis from **peer-reviewed literature**
- Ensuring **traceability and provenance** (e.g. input data, metadata, diagnostics (incl .citation), tool version, doi)
- Currently **≈ 80 scientist** from >30 institutions part of the development team and **> 120 users**
- **Development in several projects** (e.g. APPLICATE, CRESCENDO, C3S-MAGIC, ESA CMUG, PRIMAVERA)
- **Rapidly expanding**

DLR

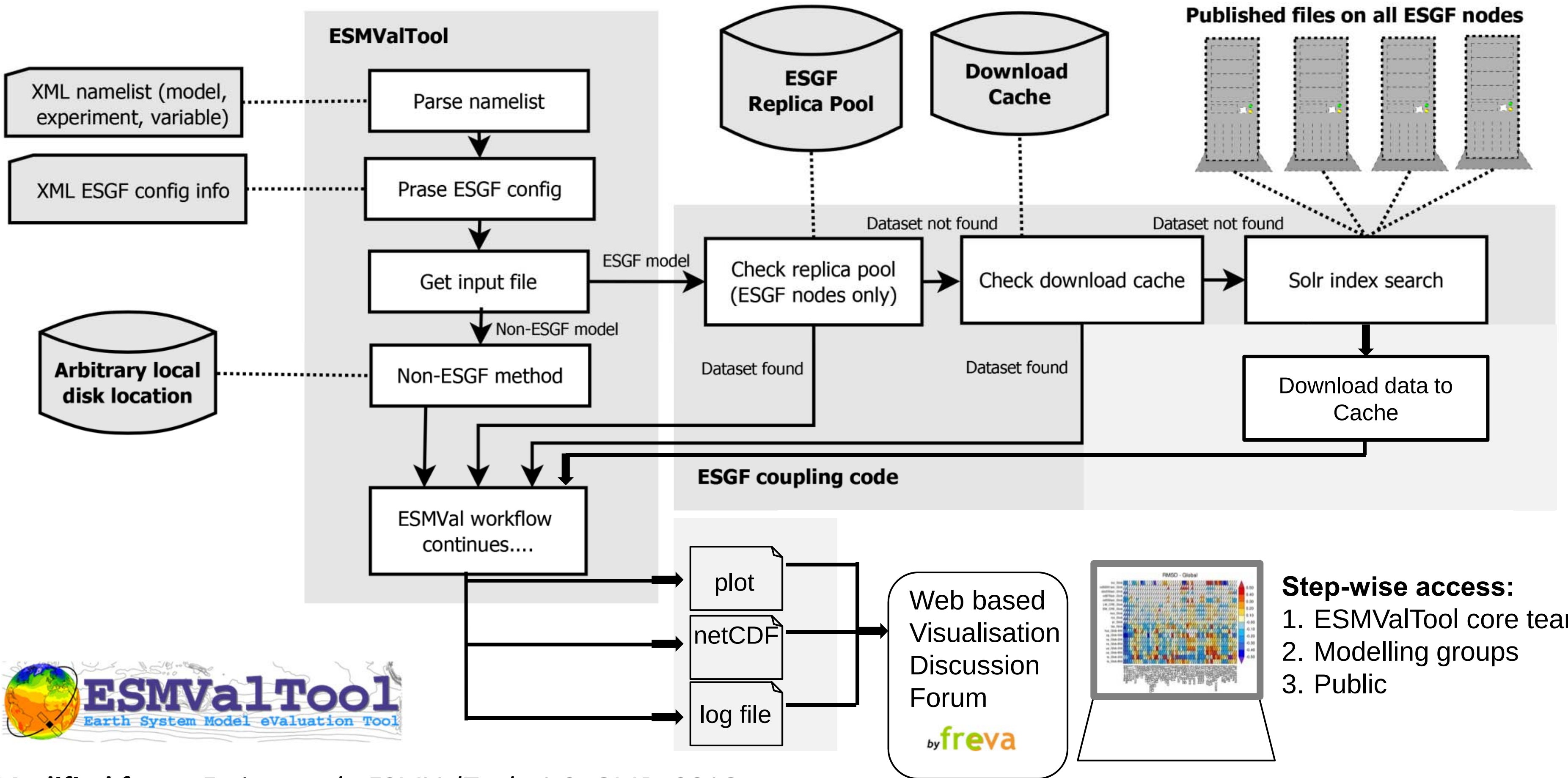
Envisaged Workflow for Model Evaluation in CMIP

- We argue that the community has reached a critical juncture at which many baseline aspects of ESM evaluation need to be performed much more efficiently
- The resulting, increasingly systematic characterization of models will, compared with early phases of CMIP, more quickly and openly identify strengths & weaknesses of the simulations.
- This activity also aims to assist modelling groups in improving their models
- Running alongside the ESGF, as soon as the output is published



Eyring et al., ESD (2016)

Envisioned ESMValTool Workflow for routine evaluation at the ESGF (CMIP6-DICAD)



Step-wise access:
1. ESMValTool core team
2. Modelling groups
3. Public

Modified from: *Eyring et al., ESMValTool v1.0, GMD, 2016*

Reproducibility & Provenance of evaluation results

Namelist

Evaluation analysis is controlled by the namelist file that defines the internal workflow for the desired analysis.

It defines:

- **Input datasets** (observations, models)
- **Regridding** operation (if needed)
- Set of **diagnostics**
- Misc. (output formats, output folder, etc…)

Output files (NetCDF, png)

Contain meta data from input files and meta data generated by ESMValTool

Observational data

- Well defined processing chain
- creation of metadata

Logfile

At each execution of the tool a log file is automatically created

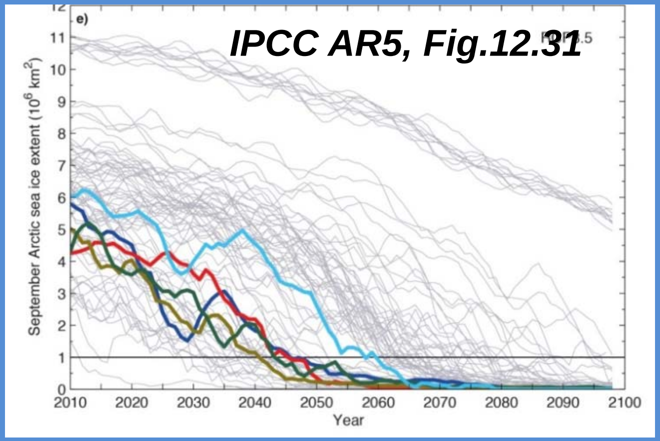
The log file contains:

- The list of **all input data** which have been used (version, data source, etc.)
- The list of **variables** that have been processed
- The list of **diagnostics** that have been applied
- The list of **authors and contributors** to the given diagnostic, together with the relevant references and projects
- Software **version** of ESMValTool that was used

3. Climate Informatics Opportunities for Earth system model evaluation

(B) Data Analysis with data science methods

	TAR	AR4	AR5
ECS	Likely range: 1.5 to 4.5°C	*likely* range: 2.0 to 4.5°C *very unlikely* <1.5°C – best estimate about 3°C	***likely* range: 1.5 to 4.5°C** ***extremely unlikely* <1.0°C** ***very unlikely* >6.0°C** –



IPCC AR5, Fig.12.31

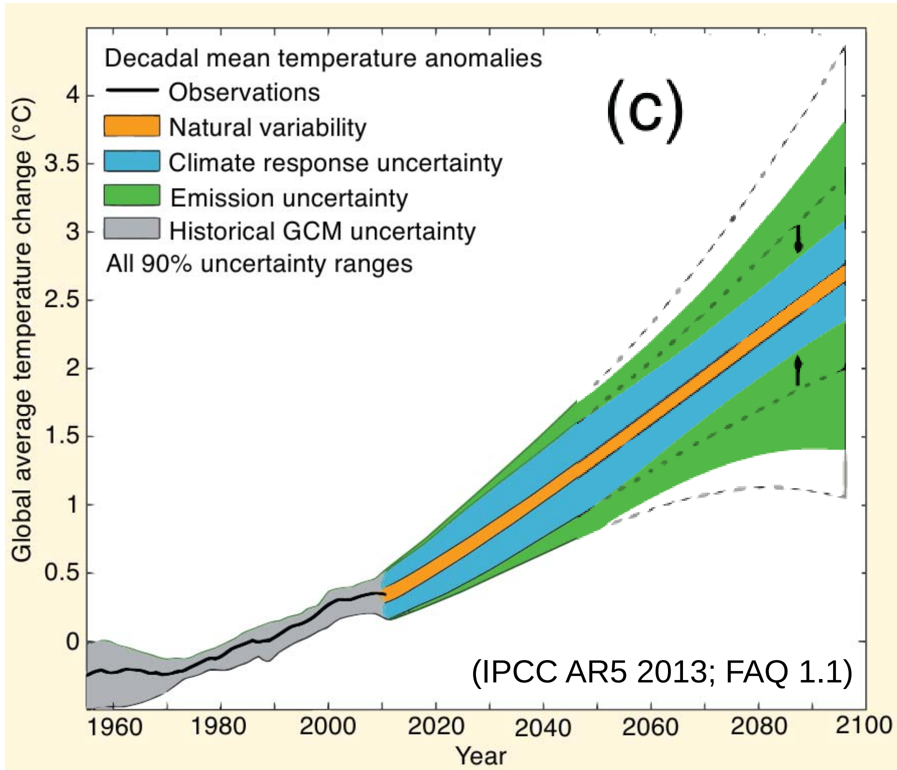
Promising Examples Climate Informatics Analysis

1. **Emergent constraints**
   - Emergent constraints between observable aspects of variability and Earth system sensitivities offer the possibility to reduce uncertainties in climate projections.
   - Finding such correlations is a challenge and data mining could help identifying a comprehensive list of strong correlations that could used as starting point for further analysis.
2. **Improving multi-model ensembles of climate projections**
3. **Anomaly detection for abrupt climate change and extreme events**
4. **Multivariate process evaluation**
   - Model evaluation approaches are often limited to the performance assessments of multiple single variables but to identify specific events requires looking across variables in space and time.
   - New algorithms could help in clustering and detecting such patterns and climate networks constructed from observations might help identifying dependencies between climate variables and processes.

For further examples, see also Monteleoni et al. (2013)

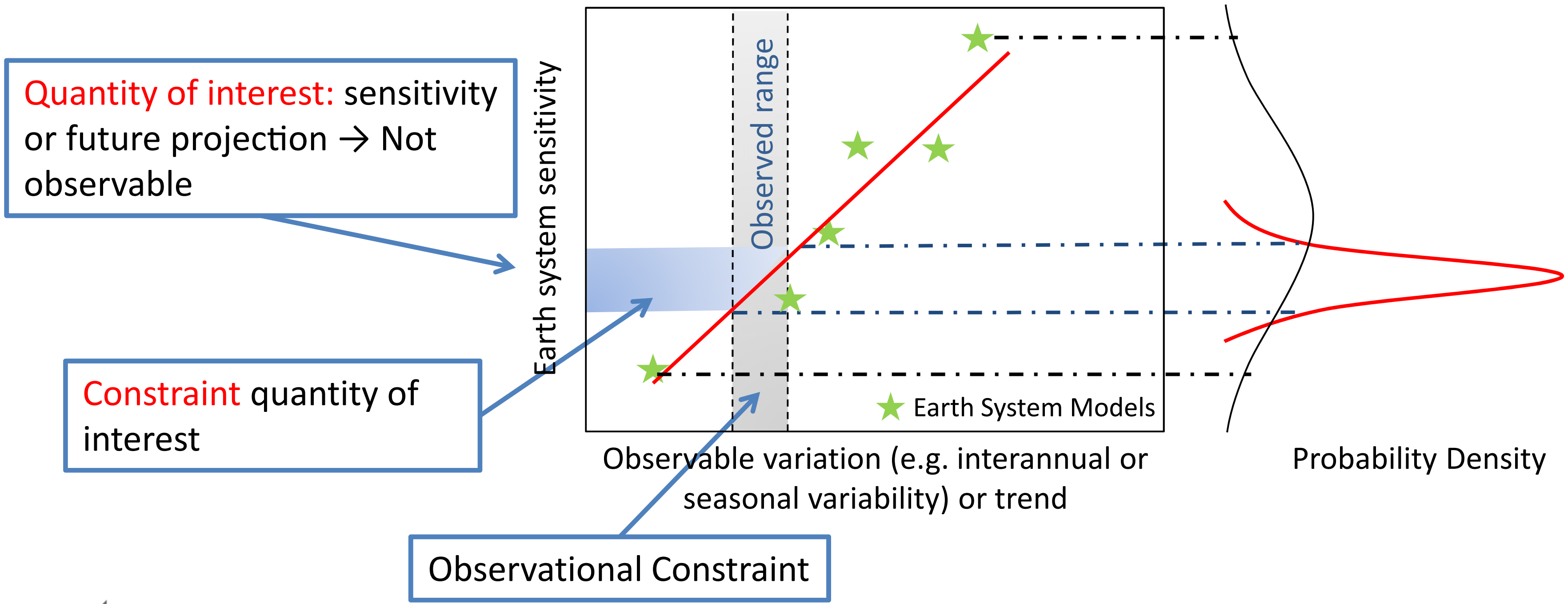
Example 1: Emergent Constraints
Uncertainties in Projections of Future Climate

1. **Internal Variability**
   - Due to the chaotic nature of climate system
   - Noise of climate record is constant with time
2. **Emission Uncertainty**
   - Dominant uncertainty for long term projections estimated as mean of different scenarios
   - Varying greenhouse gas emissions
   - Land use change
3. **Climate Response Uncertainty**
   - Models are build on same principles but parametrizations are needed
   - Increases when process become more relevant
   - Decreases with **model improvements** and **observational constraints**



Emergent Constraints (ECs)

- **ECs** are a relationship across an ensemble of models, between some aspect of **Earth system sensitivity and** an **observable trend or variation** in the current climate
  - ➢ Emergent relationship because it emerges from the ensemble of ESMs.
  - ➢ Constraint
    - – because it enables an observation to constrain the estimate of the Earth System sensitivity in the real world.
    - – The goal is to find a observable physical explanation to constrain the unobservable Earth system sensitivity

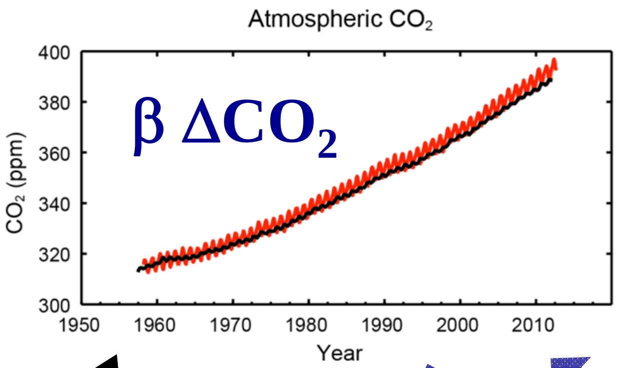


Emergent Constraints: Carbon Cycle Feedbacks

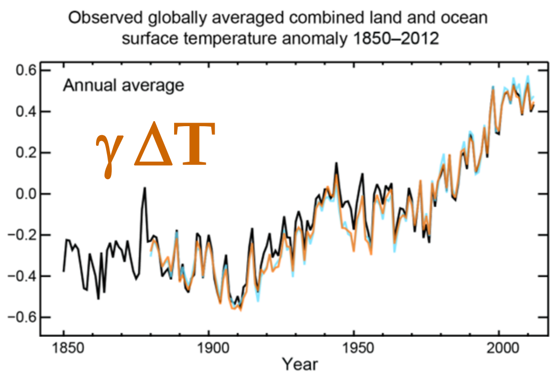
Change land carbon uptake: ΔC_{Land} [GtC] $=$

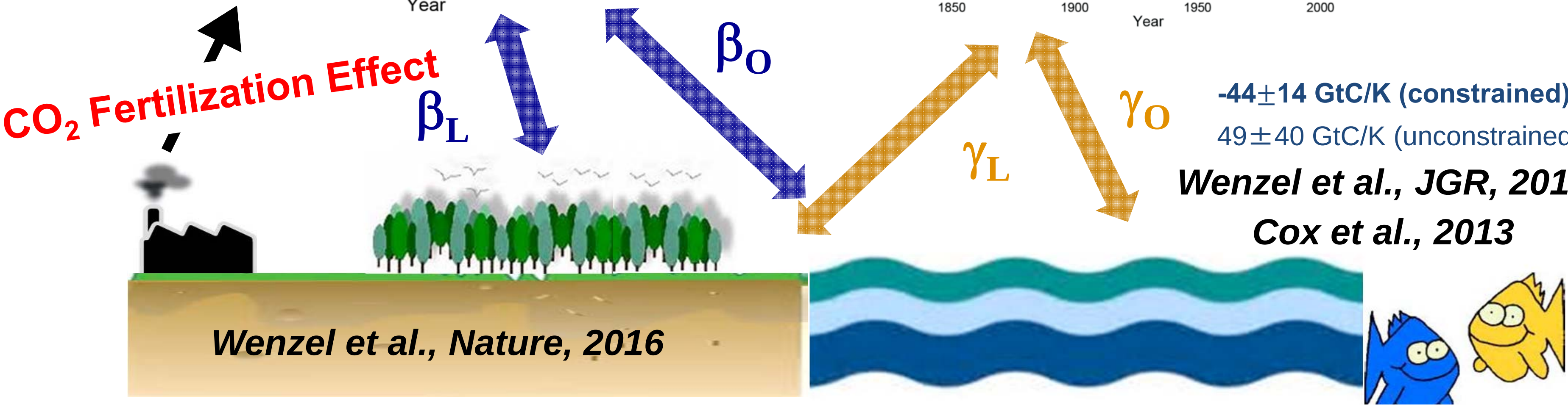
Plants take up CO_2 via photosynthesis when they grow, so CO_2 is removed from the atmosphere and is stored as organic carbon in the plants. This flux is the **Gross Primary Productivity (GPP).**

Climate warming reduces the efficiency of CO_2 absorption by the land and ocean => more emitted carbon stays in the atmosphere leading to additional warming



+





-44±14 GtC/K (constrained)

49±40 GtC/K (unconstrained)

Wenzel et al., JGR, 2014

Cox et al., 2013

Wenzel et al., Nature, 2016

β : Carbon cycle - CO_2 concentration Feedback – **Negative Feedback**

γ : Carbon cycle - Climate Feedback – **Positive Feedback**

Projected land photosynthesis constrained by changes in the seasonal cycle of atmospheric CO_2

Why should we care?

- Photosynthesis is the ultimate source of energy for life on Earth.
- Vegetation and soil are currently **slowing down global warming** by absorbing about 20% of our CO_2 emissions. This land carbon sink is believed to be in part due to increases in photosynthesis.
- But how will photosynthesis change?
- Climate-carbon models agree that elevated atmospheric CO_2 concentrations will enhance terrestrial GPP but the magnitude of this CO_2 fertilization effect varies.

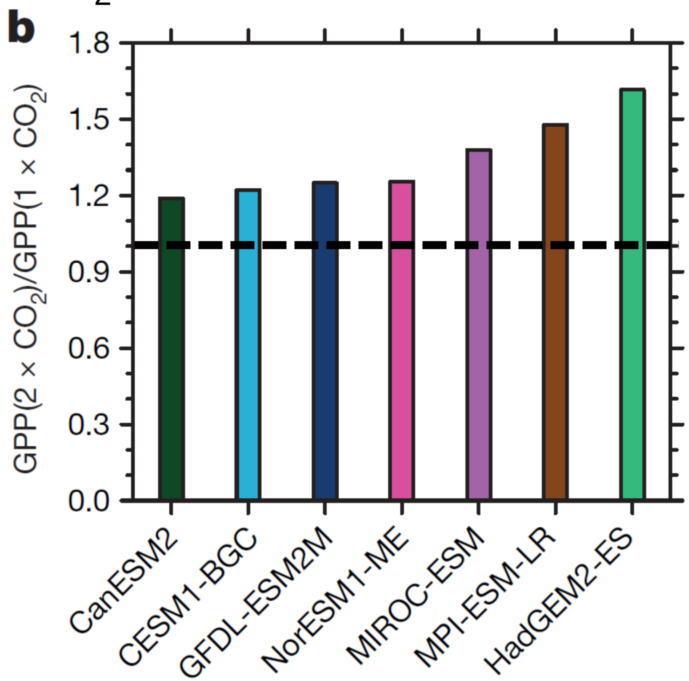
The unknown:

CO_2 Fertilization of Photosynthesis

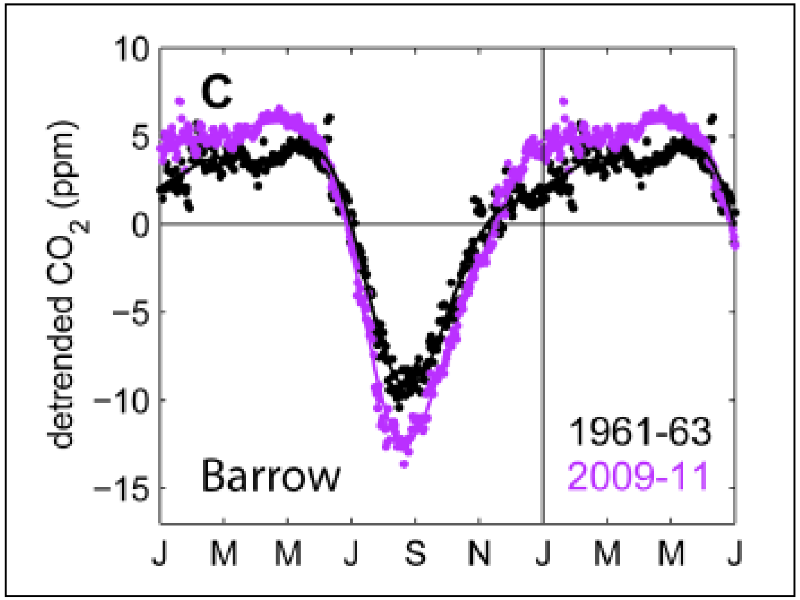
Here: relative change in high-latitude (60-90°N) GPP due to a doubling of atmospheric CO_2 in the 1%BGC simulations

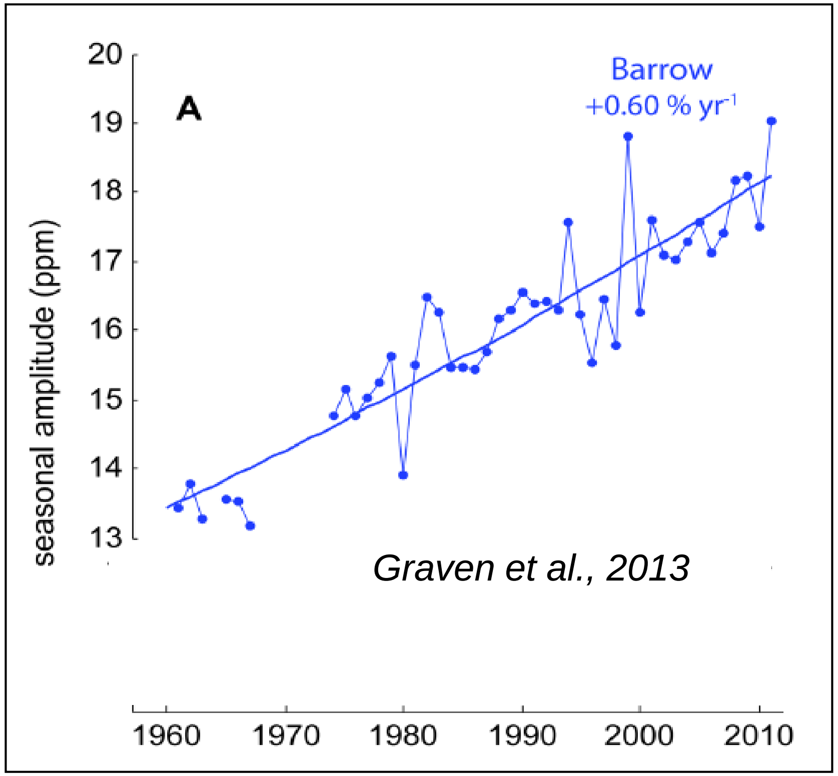
20% to 60% projected increase of GPP due to doubling of CO_2

Wenzel et al., Nature, 2016



The Observations: Increasing Seasonal Amplitude of Atmospheric CO_2





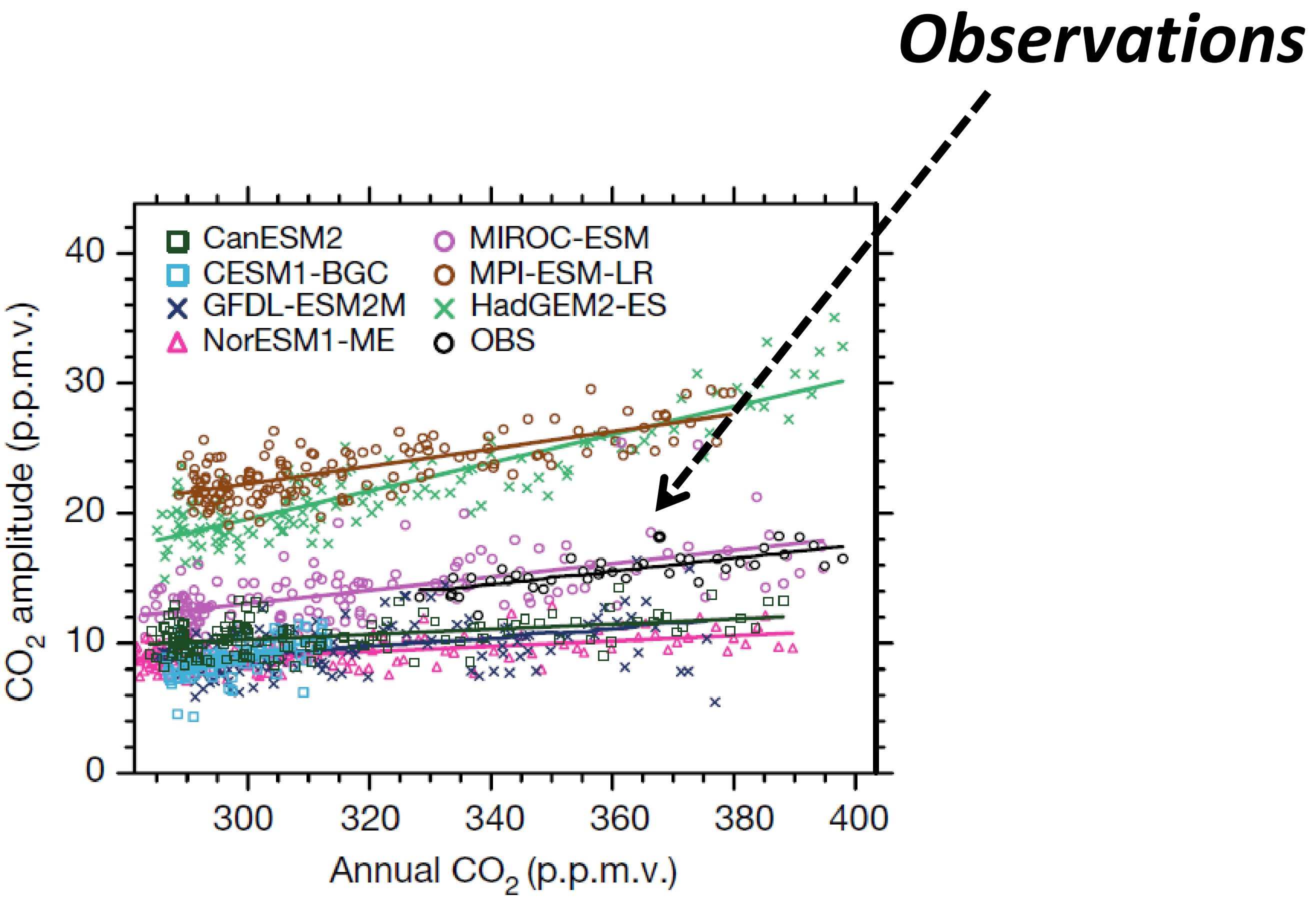
- CO_2 concentrations measured for many decades on Hawaii and Alaska show characteristic cycles, with **lower values in the summer** when strong photosynthesis causes plants to absorb CO_2, and **higher values in the winter** when photosynthesis stops.
- The peak-to-trough amplitude of the seasonal cycle therefore depends on the strength of the summer photosynthesis and the duration of the growing season.

The measurements made on Hawaii and in Alaska show an increasing amplitude of the seasonal cycle.

But what does this mean for the future?

DLR CRESCENDO

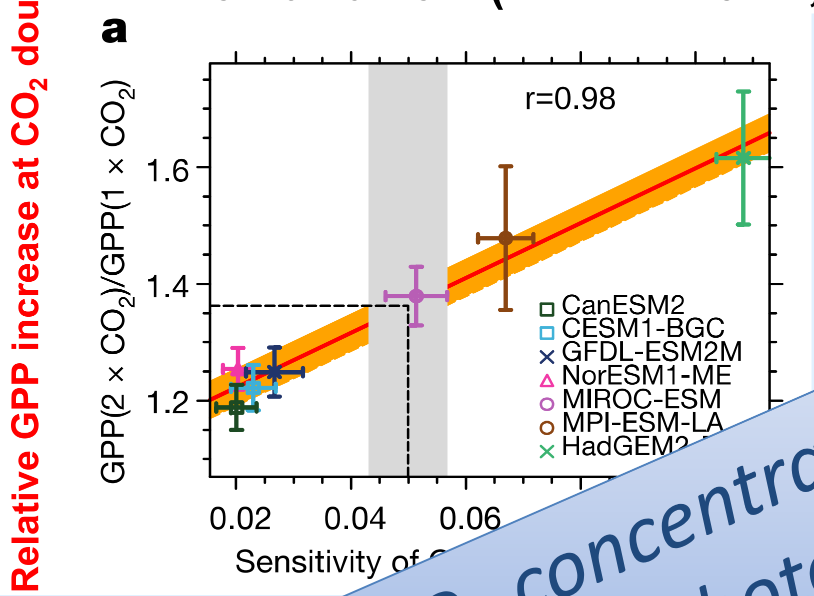
Comparison of CO_2 seasonal amplitudes against annual mean CO2 for CMIP5 historical simulations and observations at BRW



Wenzel et al., Nature, 2016

Emergent Constraints on CO_2 Fertilization

Point Barrow (BRW: 71.3° N, 156.6° W), Alaska (high-latitude)

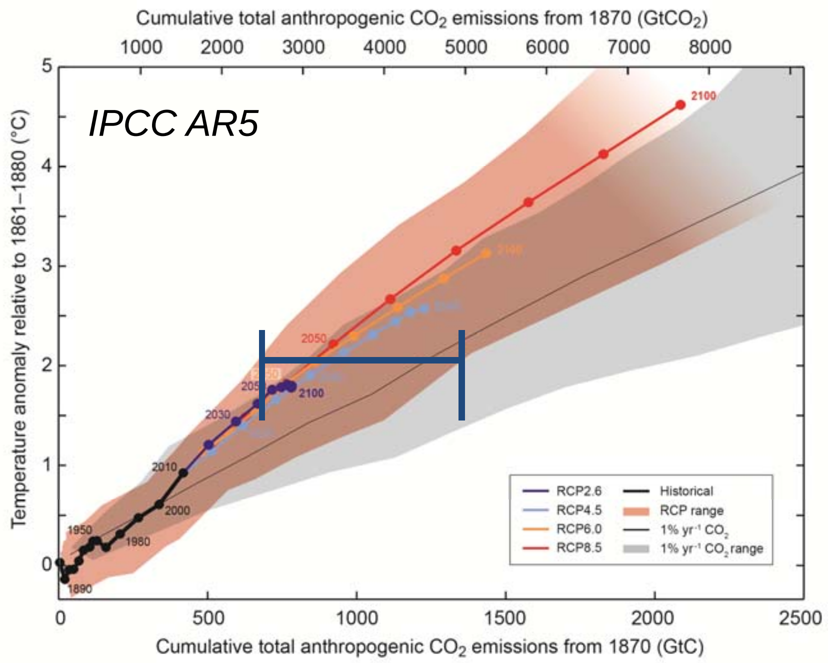


Doubling CO_2 concentration (alone) will lead to an increase in land photosynthesis of about a third. Earth System Model projections can be more than the sum of the parts!

Wenzel et al., Nature, 2016

Compatible CO_2 Emissions

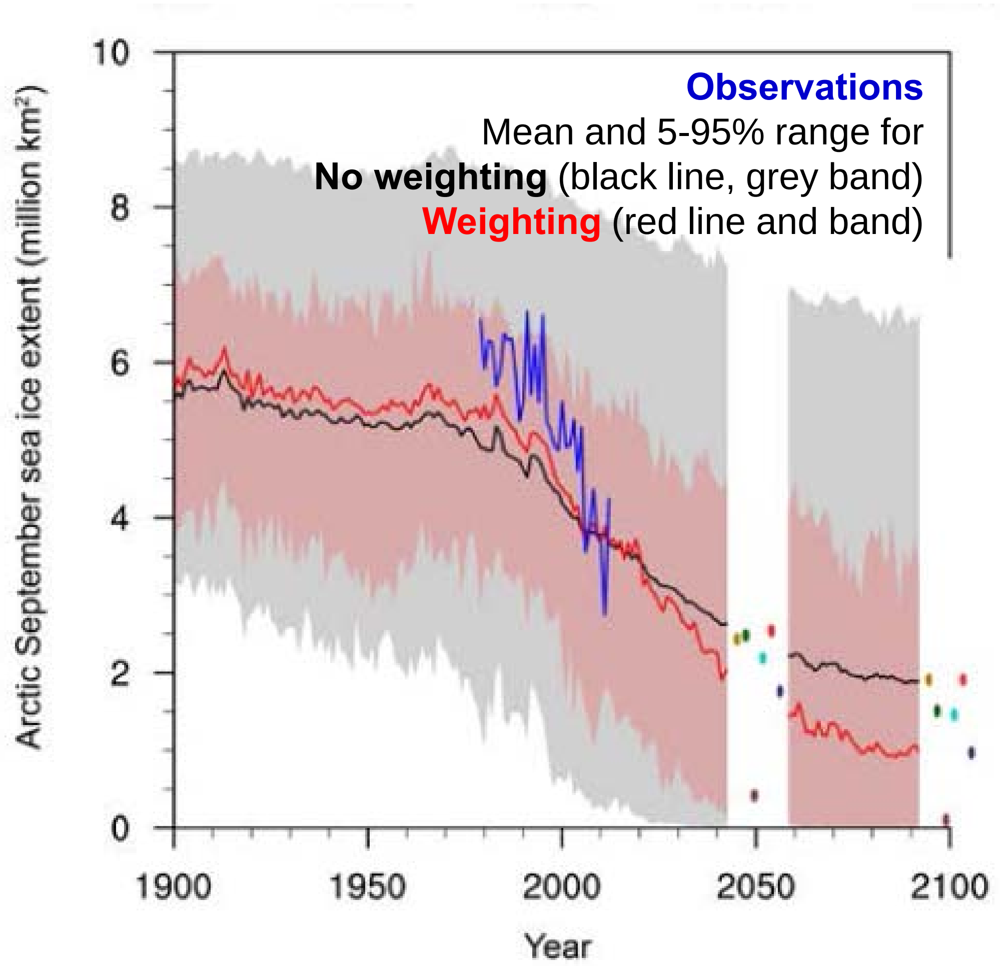
- Large uncertainty in CO_2 emissions compatible with a given climate target.
- Budget for the 2°C target is about 700 GtC to 1300 GtC.
- Given ~550 GtC emitted so far, that's 15 to 75 years of current anthrop. CO_2 emissions.



Next to work on: Development of data mining methods

- Automate the process of identifying current climate quantities with skill at predicting individual feedbacks, ECS and transient climate response to emissions (TCRE).

Example 2: Weighted sea-ice projections based on model-performance and interdependence



Knutti et al., GRL, 2017

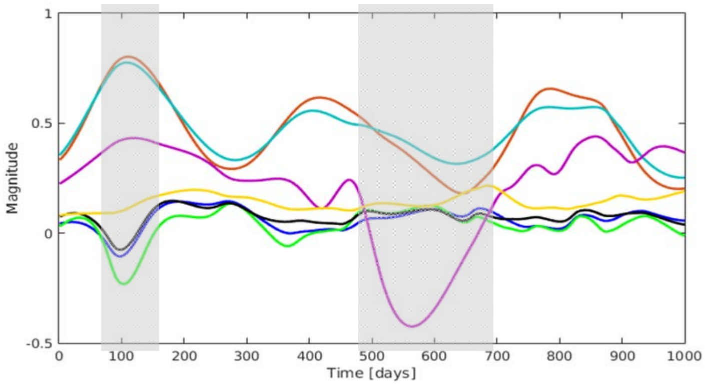
Diagnostics included in this example:

1. None (unweighted)
2. Climatological mean (1980-2013) Sep sea ice extent,
3. Sep sea ice extent trend 1980-2013,
4. Climatology of monthly surface temperature (1980-2013)
5. Interannual variability of monthly surface temperature,
6. All diagnostics 2-5

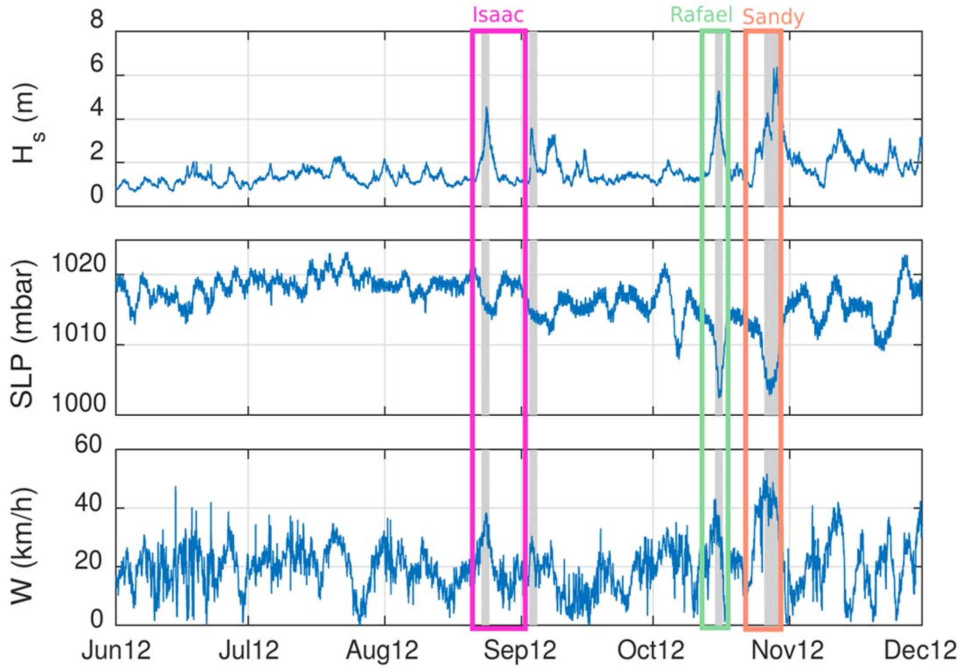
Development of machine learning techniques for model weighting

Are there weighting strategies that maximize predictive skill?

Example 3: Maximally Divergent Intervals:
Detection of Anomalies in Multivariate Time-Series Data



$$KL(p_I, p_\Omega) = \int p_I\,(x) \log\left(\frac{p_I(x)}{p_\Omega(x)}\right) dx$$



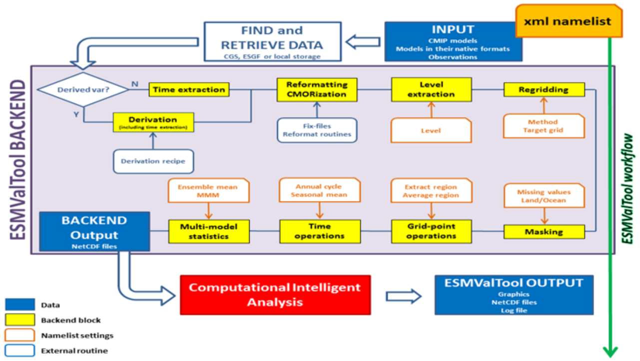
- Maximizing Kullback-Leibler divergence
- Example: Detection of hurricanes
- Cooperation in EU-H2020 Project BACI (FSU, MPI-BGC)

Rodner et al., Anomaly Detection Workshop (2016)

DLR

New Group *Climate Informatics* at the DLR Data Science Institute in Jena

- In cooperation with **DLR Institute of Atmospheric Physics** (Eyring), the official external collaboration partner of the group **FSU Computer Vision** (Denzler), and the **MPI Department of Biogeochemical Integration** (Reichstein).
- **Development of innovative and highly efficient data science methods** for Earth system data analysis.
- **Focus of research group on three broad scientific themes:**
  - **Basic research on graphical models**
  - **Data management for an efficient data analysis:** Development of innovative and highly efficient methods for data proecessing to be implemented in the ESMValTool.
  - **Data analyses:** Development of **innovative and highly efficient data science methods** for data analysis for the extraction of space-time characteristics in massive geoscientific data from Earth system models and observations; subsequent implementation in the ESMValTool.



Summary

- **Climate Informatics**
  - is a challenging and promising research field where little has been done so far and where a concentrated effort will have a high impact both to advance science and to address topics of critical importance for the society.
  - It bridges the gap between data and understanding through a strong collaboration between climate scientists with machine learning, data mining, and statistics researcher (see also Monteleoni et al., 2013).
- **New methods of Data Science**, such as data mining, artificial intelligence and machine learning techniques could potentially help finding new ways of analysing Earth system data and helping to ensure that results are robust.
- **The newly-developed data science methods** will be included in the ESMValTool so open a broad opportunity for other applications that analyze Earth observations or models (or both).
- **Efficiently provide a number of derived products such as for example**:
  - Multi-model means (operational service)
  - Evaluation of individual or multiple models with observations
  - Multiple observational datasets