EGU2020-13306
CMIP model evaluation with the ESMValTool v2.0

Axel Lauer[1], Lisa Bock[1], Veronika Eyring[1,2], Fernando Iglesias-Suarez[1], and the ESMValTool development team

[1]Deutsches Zentrum für Luft- und Raumfahrt (DLR), Institut für Physik der Atmosphäre, Oberpfaffenhofen, Germany
[2]University of Bremen, Institute of Environmental Physics (IUP), Bremen, Germany

8 May 2020

DLR

ESMValTool – aim and motivation

Facilitate the evaluation of complex Earth System Models, e.g.

- quick look at **standard diagnostic plots & output diagnostic variables**
- **easy comparison of new simulations** (e.g., sensitivity runs or runs with new model versions) with existing runs and with observations

Raise the standard for model evaluation

- include additional diagnostics of ongoing evaluation activities **so that it is not needed to start from scratch each time**
- implement **more observations, account for observational uncertainties**
- Quick assessment of where a new set of model simulations stands via **standard "recipes" that reproduce specific papers, reports,** etc.
- **Traceability and reproducibility**

Facilitates analysis of and participation in Model Intercomparison Projects

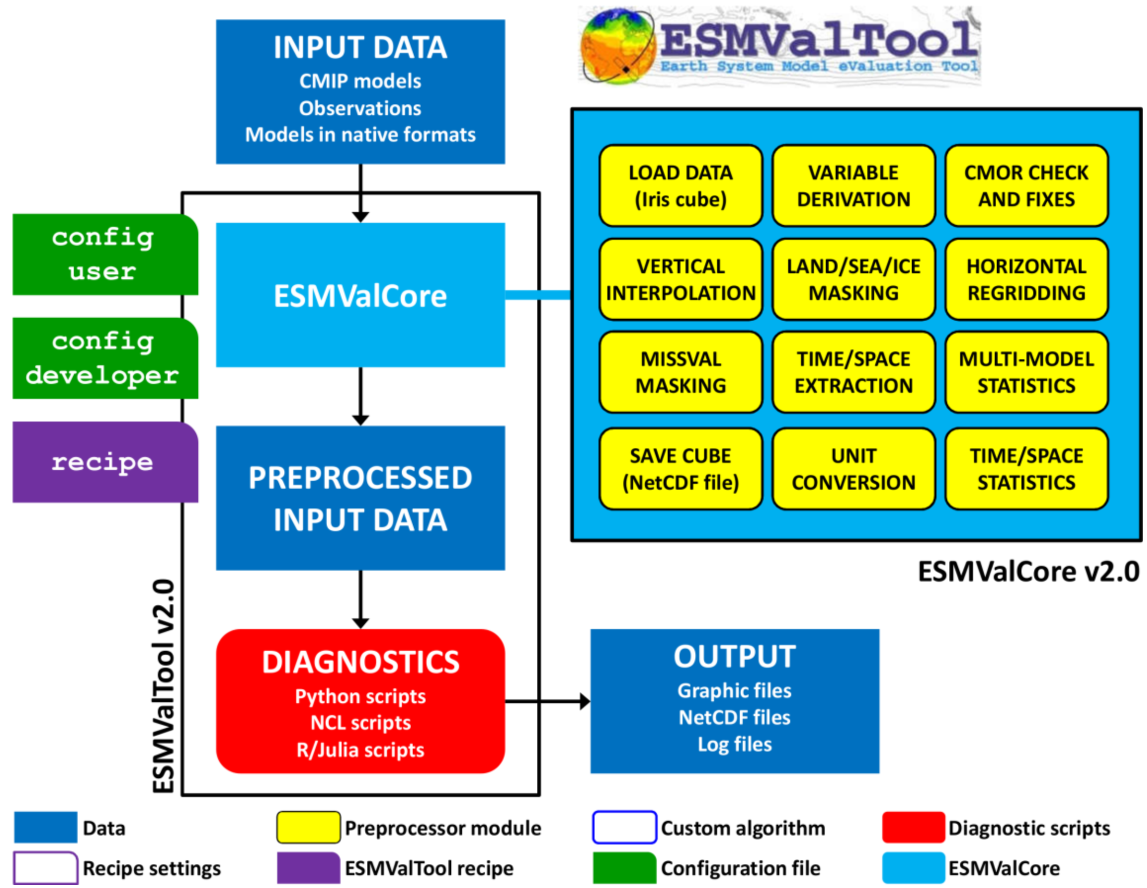
- easily comparison of models participating in **CMIP** and **CMIP6-Endorsed MIPs**

Easy expandability

- synergies with ongoing projects to expand the tool (e.g., NCAR CVDP)
- useful for **model groups & those analyzing models**
- useful for **model development**

ESMValTool version 2.0 (Righi et al., GMD)

- Development with national and international partners on GitHub
- Core functions completely rewritten in Python 3
- Division into „core“ and „diagnostics“ to allow for easy contributions while maintaining high level of code quality for core functions
- Improved performance and efficiency using state-of-the-art Python libraries (Iris, Dask)
- Greatly improved user friendliness (installation, configuration, documentation)



From Righi et al. (2020)

DLR

New large-scale diagnostics (Eyring et al., GMDD)

Integrative measures of model performance

- Performance metrics for essential climate variables for the atmosphere, ocean, sea ice and land
- Centered pattern correlations for different CMIP ensembles
- Single model performance index
- Auto-Assess diagnostics

Atmosphere

- Multi-model mean bias for temperature and precipitation
- Precipitation quantile bias
- Atmospheric dynamics
- Thermodynamics of the climate system
- Natural modes of climate variability and weather regimes

Ocean and cryosphere

- Physical ocean
- Southern Ocean
- Arctic Ocean
- Sea Ice

Land processes

- Land Cover
- Albedo changes associated to land cover transitions

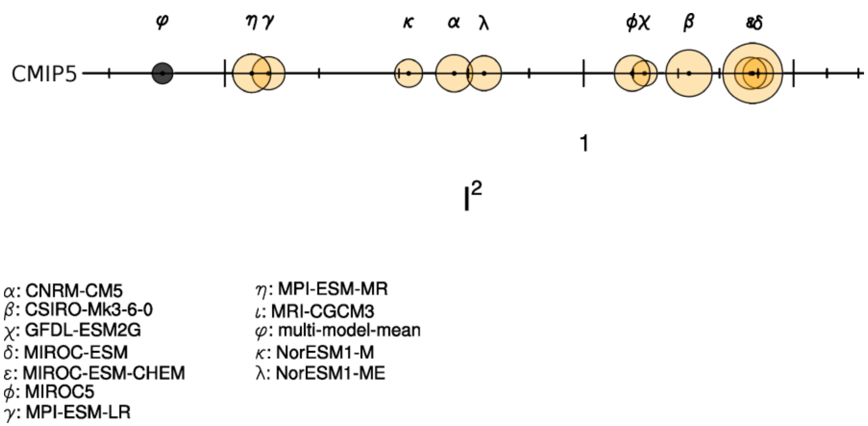
Biogeochemical processes

- Terrestrial biogeochemistry
- Ecosystem Turnover Times of Carbon
- Marine biogeochemistry
- Stratospheric temperature and trace species influencing stratospheric ozone chemistry

From Eyring et al. (in review)

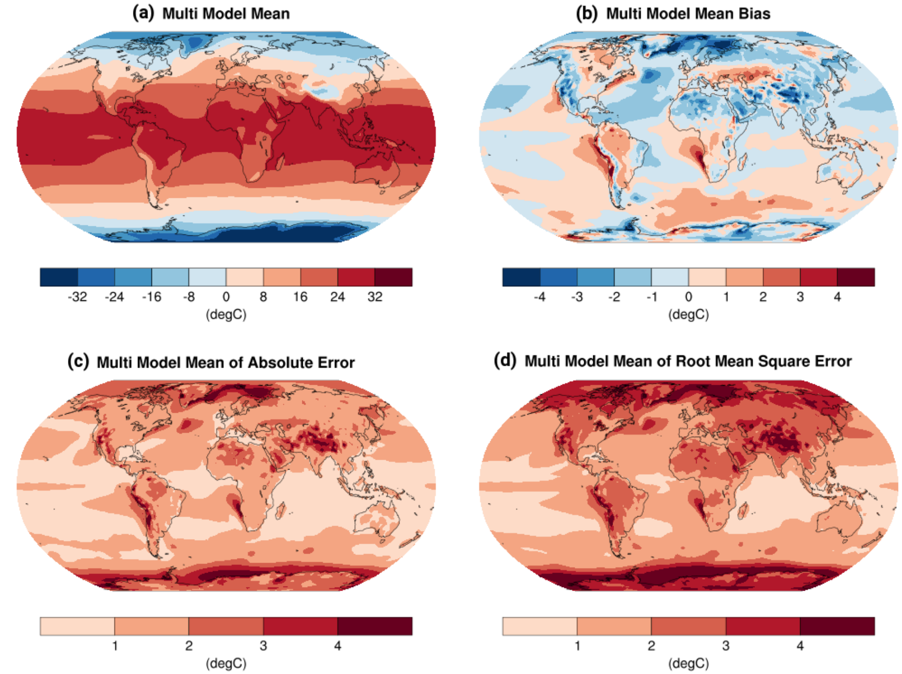
Examples of new large-scale diagnostics

Single Model Performance Index (SMPI)



Single Model Performance Index I2 for individual models (orange circles). The size of each circle represents the 95% confidence interval of the bootstrap ensemble. The black circle indicates the I2 of the CMIP5 multi-model mean. The I2 values vary around one, with underperforming models having a value greater than one, while values below one represent more accurate models. This allows for a quick estimation which models are performing the best on average across the sampled variables and in this case shows that the common practice of taking the multi-model mean as best overall model is accurate. Similar to Reichler and Kim (2008) Fig. 1 and produced with *recipe_smpi.yml*.

Annual mean 2-m air temperature



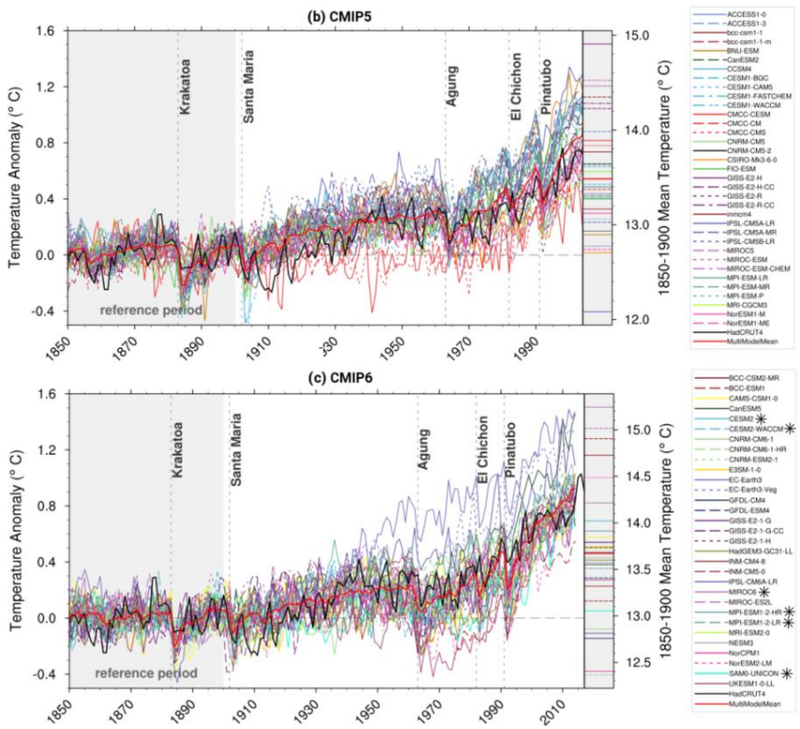
Annual-mean 2-m air temperature (°C) for the period 1980-2005. (a) Multi-model (ensemble) mean constructed with one realization of all available models used in the CMIP5 historical experiment. (b) Multi-model mean bias as the difference between the CMIP5 multi-model mean and the climatology from ECMWF reanalysis of the global atmosphere and surface conditions (ERA)-Interim (Dee et al., 2011). (c) Mean absolute model error with respect to the climatology from ERA-Interim. (d) Mean root mean square error of the seasonal cycle with respect to the ERA-Interim. Updated from Fig. 9.2 of Flato et al. (2013) and produced with *recipe_flato13ipcc.yml*.

DLR

Examples of new large-scale diagnostics

From Bock et al. (in review)

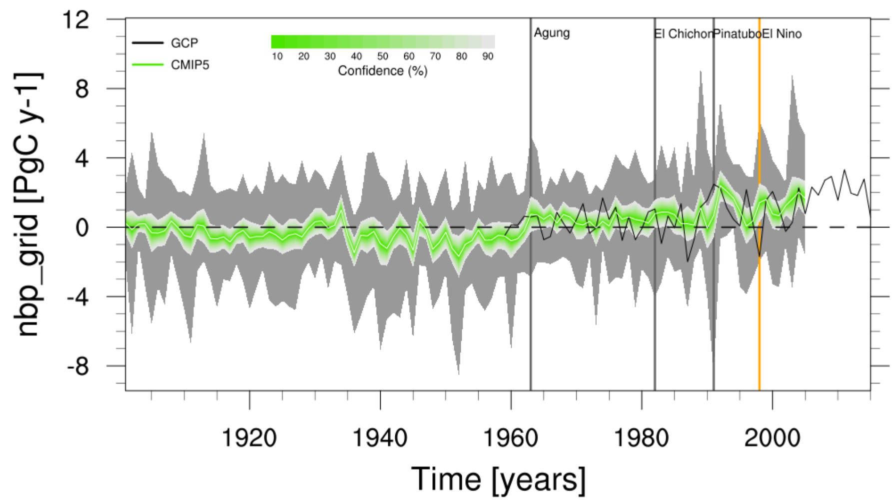
Global average 2-m temperature anomalies



Observed and simulated time series of the anomalies in annual and global mean surface temperature. All anomalies are differences from the 1850–1900 time-mean of each individual time series. The reference period is indicated by gray shading. The thin lines show individual climate model simulations from (top) CMIP5 and (bottom) CMIP6, the thick red lines show the multi-model means. The observational data (thick black lines) are the Hadley Centre/Climatic Research Unit gridded surface temperature data set version 4 (HadCRUT4; Morice et al., 2012). All models have been subsampled using the HadCRUT4 observational data mask (see G. S. Jones, Stott, & Christidis, 2013). Inset: the global mean surface temperature for the reference period 1961–1990 of the subsampled fields.

From Eyring et al. (in review)

Global land-atmosphere CO_2 flux



Time series plot of the global land-atmosphere CO_2 flux (*nbp*) for CMIP5 models compared to observational estimates by GCP, Le Quere et al. (2018) (black line). Gray shading represents the range of the CMIP5 models, green shading shows the confidence interval evaluated from the CMIP5 ensemble standard deviation assuming a *t*-distribution centered at the multi-model mean (white line). Vertical lines indicate volcanic eruptions (gray) and El Niño events (orange). As positive values correspond to a carbon uptake of the land, similar to Fig. 5 of Anav et al. (2013) and produced with *recipe_anav13jclim.yml.*

DLR

New diagnostics for extreme events, regional model and impact evaluation and analysis of ESMs (Weigel et al., in prep)

Hydrological cycle

- Hydroclimatic intensity and related indices
- Droughts

Extreme events

Evaluation for impact assessments

- Heat and cold wave duration
- Combined Climate Extreme Index
- Daily temperature range variation
- Capacity factor

Regional features

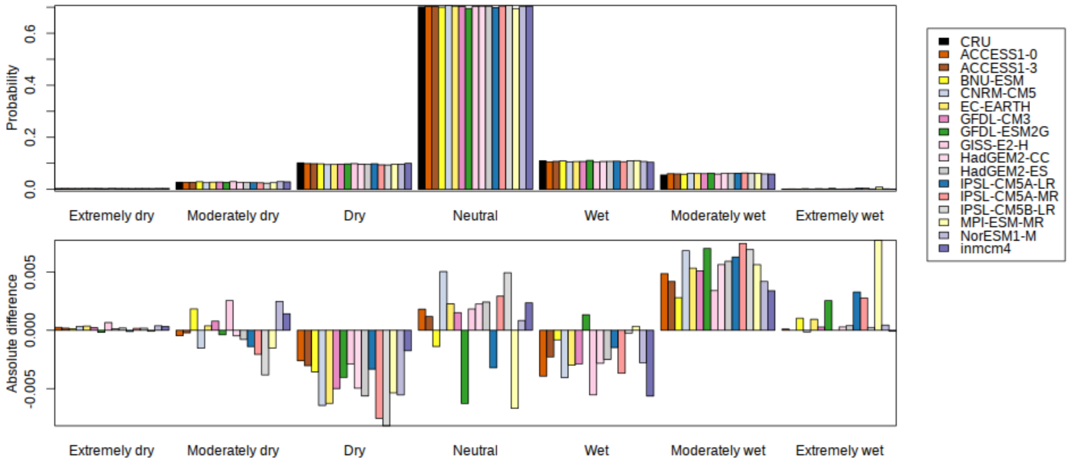
- Evaluation of global climate models for selected regions
- Stochastic Downscaling

Multi-model ensemble member sub-selection

DLR

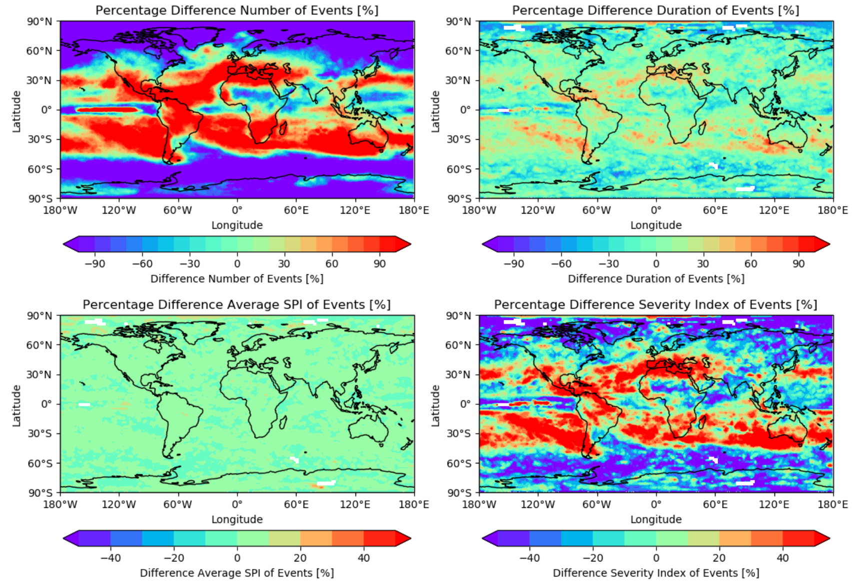
Examples of new diagnostics for extreme events, regional model and impact evaluation and analysis of ESMs

Standardized Precipitation Index (SPI)



Output from SPI (standardized precipitation index) diagnostic in *recipe_spei.yml* with globally averaged histogram of SPI over land areas, weighted by the cosine of latitude for a selection of CMIP5 models and using gridded observations from CRUts4.01. (top) Absolute values, and (bottom) bias of all models compared to CRUts4.01.

Drought Events



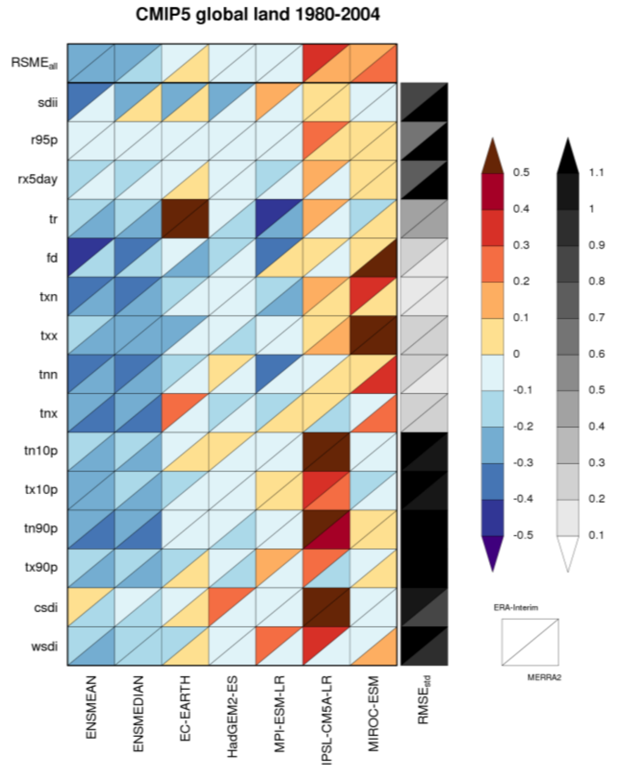
Difference in number (top left), duration (top right), average SPI (bottom left), and severity index (bottom right) of drought events between the RCP8.5 (2050-2100) and historic (1950 to 2100) multi-model mean of 13 CMIP5 models. Here, a drought event is defined as any number of consecutive months with an SPI < -2. It shows an increase in the number of drought events, the severity index and to a smaller amount the duration of drought events in the RCP8.5 scenario compared to the historical model runs especially in the subtropical areas. The figure is similar to Fig. 3a-d of (Martin 2018) and produced with *recipe_drought_events.yml*.

DLR

From Weigel et al. (in prep.)

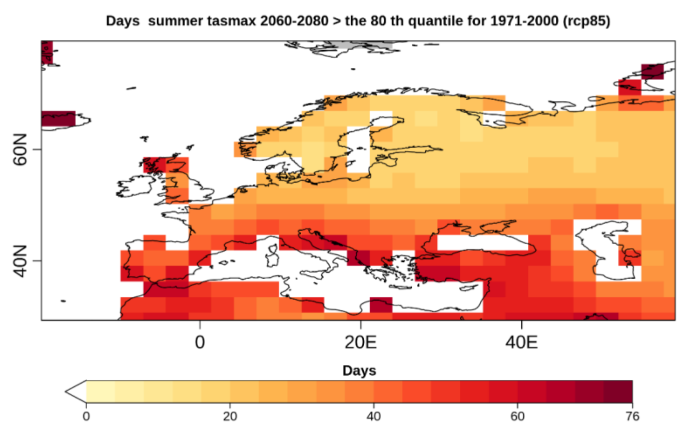
Examples of new diagnostics for extreme events, regional model and impact evaluation and analysis of ESMs

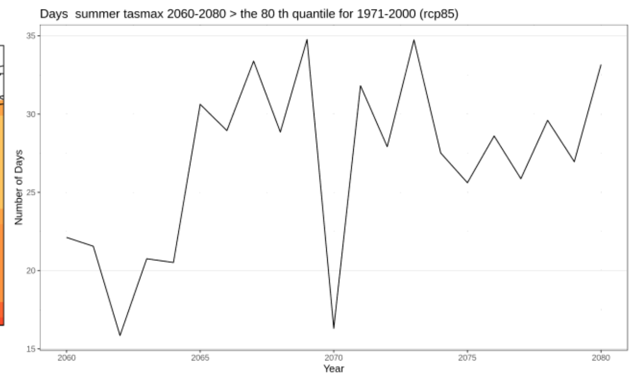
Portrait diagram – temperature and precipitation indices



“portrait” diagram showing relative spatially averaged root mean square errors (RMSE) in the 1981–2000 climatologies of 12 temperature and 3 precipitation indices simulated by CMIP5 models (x-axis) with respect to the two reanalyses ERA-Interim (upper triangle) and MERRA2 (lower triangle). The RMSEs are spatially averaged over all land grid points. The top row ($RMSE_{all}$) indicates the mean relative RMSE across all indices for the CMIP5 ensemble mean (first column) and median (second column) and each model individually. Blue (red) colors indicate that a model performs better (worse) than the ensemble mean error compared to the respective reanalysis dataset. The gray shaded column at the right-hand side indicates the median RMSE normalized by the spatial standard deviation of the index climatology in the reanalyses ($RMSE_{std}$). Similar to Fig. 9.37a of Flato et al. (2013) and produced with *recipe_extreme_events.yml*.

Heatwaves





(left) Average number of summer days during the time period 2060-2080 when the daily maximum near-surface air temperature exceeds the 80th quantile of the 1971-2000 reference period. (right) Yearly number of summer days when the daily maximum near-surface air temperature exceeds the 80th quantile of the 1971-2000 reference period averaged over the region shown in a). Results shown are for the RCP 8.5 scenario simulated by BCC-CSM1-1 and produced with *recipe_heatwaves_coldwaves.yml*.

DLR

New diagnostics for emergent constraints and analysis of future projections from ESMs (Lauer et al., GMDD)

Calculations of multi-model products

Effective climate sensitivity (ECS) and transient climate response (TCR)

Emergent constraints

- Effective climate sensitivity
- Carbon cycle
- Year of disappearance of September Arctic sea ice
- Snow-albedo effect
- Hydrological cycle

Climate model projections

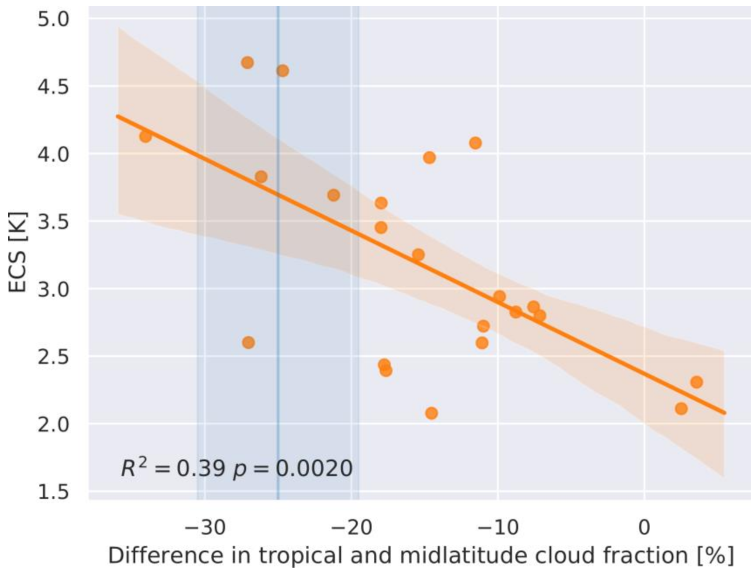
- MDER to constrain future austral jet position
- Toy model
- Climate projection chapter of IPCC WGI AR5
- Sea ice

DLR

From Lauer et al. (in review)

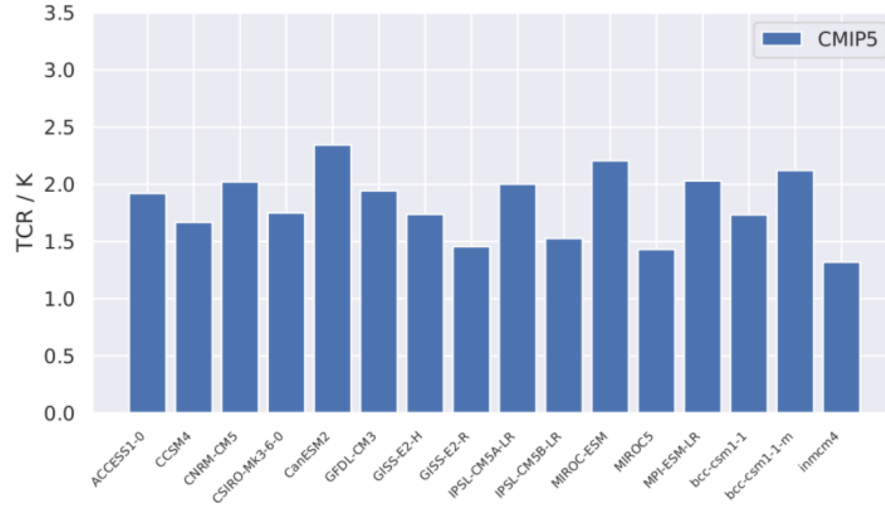
Examples of new diagnostics for emergent constraints and analysis of future projections from ESMs

Emergent Constraint for ECS (Volodin, 2008)



Effective climate sensitivity (ECS) vs. difference in total cloud cover between the tropics (28°S-28°N) and southern mid-latitudes (56°S-36°S) for CMIP5 models (orange dots). The orange line and shaded area show the linear regression line and its 95% uncertainty range (estimated via bootstrapping). Together with the observational estimate (vertical blue line and shaded area), this can be used as an emergent constraint for ECS (Volodin, 2008). The observational range is based on ISCCP-D2 data (Rossow and Schiffer, 1991) and taken from Volodin (2008). Similar to Fig. 3a of Volodin (2008) and produced with *recipe_ecs_multivariate_constraint_cmip5.yml*.

Transient Climate Response (TCR)



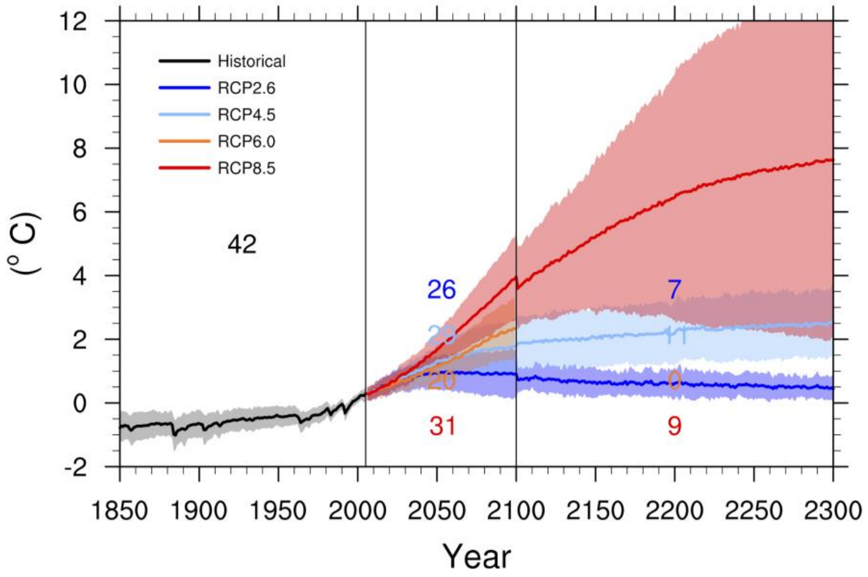
Transient climate response (in K) for CMIP5 models calculated with the method by Gregory and Forster (2008). Produced with *recipe_tcr.yml*.

DLR

From Lauer et al. (in review)

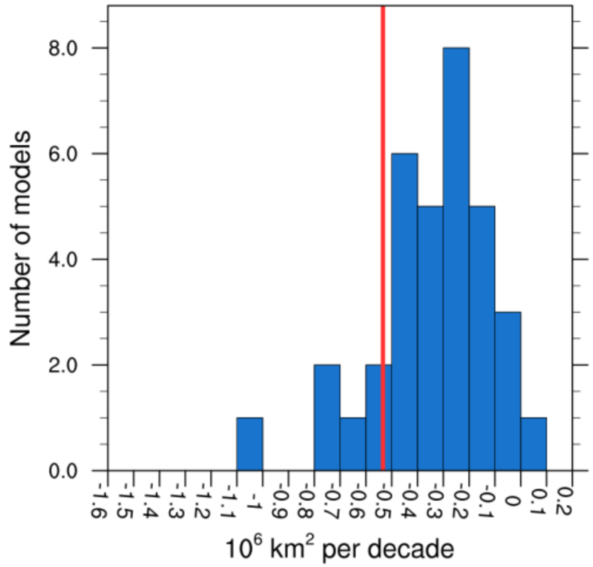
Examples of new diagnostics for emergent constraints and analysis of future projections from ESMs

Global 2-m temperature change



Time series of global annual mean surface air temperature anomalies (relative to 1986-2005) fro CMIP5 models and RCP2.6, 4.5, 6.0, and 8.5 scenarios. The solid lines show the multi-model mean, the shading shows the 5 to 95% range (±1.64 standard deviations). The numbers indicate the number of models these estimates are based on. Similar to Collins et al. (2013) Fig. 12.5 and produced with *recipe_collins13ipcc.yml*.

Trends in September Arctic sea ice extent

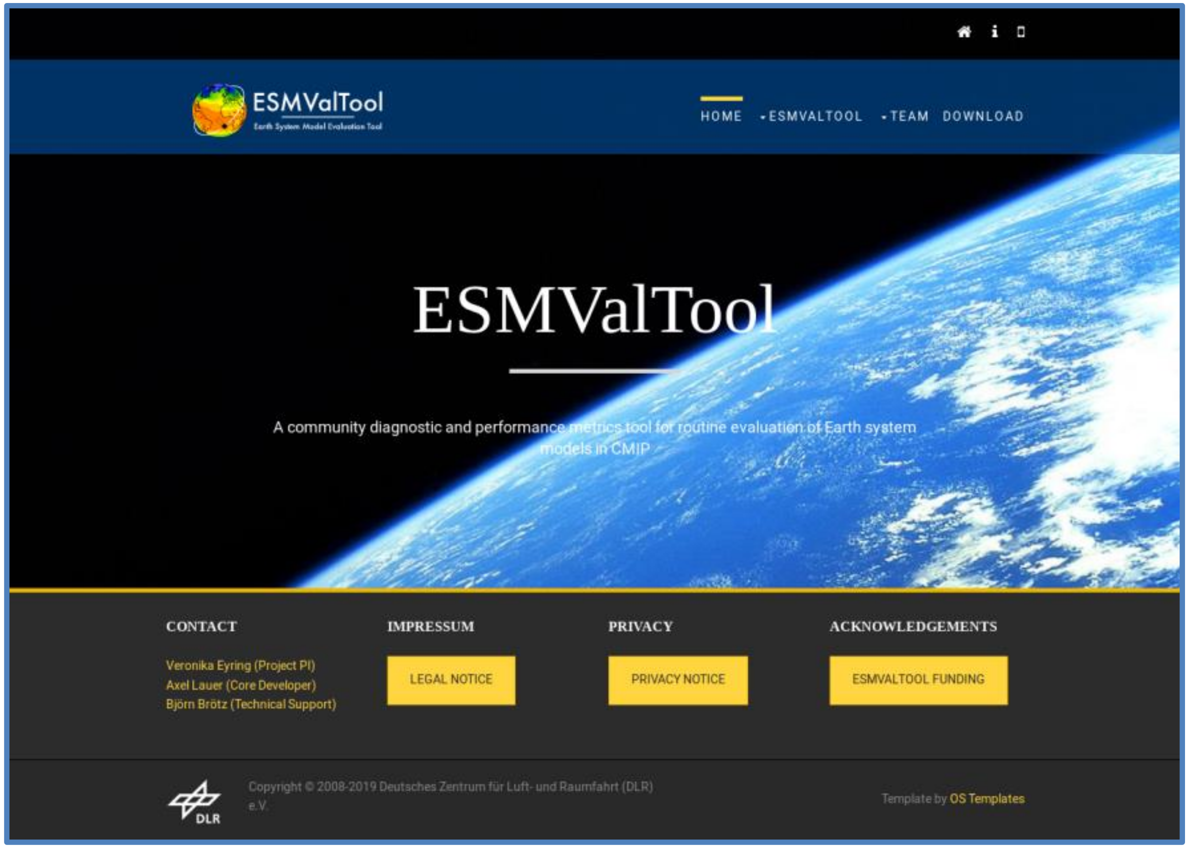


Distribution of trends in Arctic September sea ice extent calculated from the historical simulations (1960-2005) of 26 CMIP5 models (similar to Flato et al. (2013), Fig. 9.24c). An observational estimate of the trend in summer sea ice extent from HadISST (Rayner et al., 2003) over the same time period is shown by the vertical red line. Produced with *recipe_seaice.yml*.

DLR

ESMValTool Version 2.0

https://www.esmvaltool.org/



https://github.com/ESMValGroup/ESMValTool

References

1. **Technical overview**: Righi et al., Geosci. Model Dev., 2020
2. **Large-scale diagnostics**: Eyring et al., Geosci. Model Dev. (in review)
3. **Diagnostics for extreme events and regional/impact evaluation:** Weigel et al., Geosci. Model Dev. (in prep.)
4. **Emergent constraints and diagnostics for future projections:** Lauer et al., Geosci. Model Dev. (in review)