

Mengis, Nadine; Keller, David P.; Rickels, Wilfried; Quaas, Martin; Oschlies, Andreas

Article — Accepted Manuscript (Postprint)

Climate Engineering-induced changes in correlations between Earth system variables - Implications for appropriate indicator selection

Climatic Change

Provided in Cooperation with:

Kiel Institute for the World Economy – Leibniz Center for Research on Global Economic Challenges

Suggested Citation: Mengis, Nadine; Keller, David P.; Rickels, Wilfried; Quaas, Martin; Oschlies, Andreas (2019) : Climate Engineering-induced changes in correlations between Earth system variables - Implications for appropriate indicator selection, *Climatic Change*, ISSN 1573-1480, Springer, Dordrecht, Vol. 153, Iss. 3, pp. 305-322, <https://doi.org/10.1007/s10584-019-02389-7>

This Version is available at:

<https://hdl.handle.net/10419/265097>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Climate Engineering-induced changes in correlations between Earth system variables - Implications for appropriate indicator selection

Nadine Mengis · David P. Keller · Wilfried Rickels · Martin Quaas · Andreas Oschlies

Received: date / Accepted: date

5 Abstract

Climate engineering (CE) deployment would alter prevailing relationships between Earth system variables, making indicators and metrics used so far in the climate change-assessment context less appropriate to assess CE measures. Achieving a comprehensive CE assessment requires a systematic and transparent reevaluation of the indicator selection process from Earth system variables. Here we provide a first step towards such a systematic assessment of changes in correlations between Earth system variables following simulated deployment of different CE methods. We therefore analyze changes in the correlation structure of a broad set of Earth system variables for two conventional climate change scenarios without CE and with three idealized CE model experiments: i) Solar Radiation Management, ii) Large-scale Afforestation, and iii) Ocean Alkalinity Enhancement. First, we investigate how the three CE scenarios alter prevailing correlations between Earth system variables when compared to an intermediate-high and a business-as-usual future climate change scenario. Second, we contrast the indicators identified for the non-CE climate change scenarios and the indicators identified when all five scenarios are considered. Finally, we use the identified indicator sets for an evaluation of the five climate change scenarios. We find that the additional indicators provide valuable information for the assessment of the CE measures, and their application hence allows for a more comprehensive and a comparative assessment of the mitigation and CE deployment scenarios.

Keywords Climate Engineering · Indicator selection · Correlation Changes

N. Mengis
Concordia University, Montreal, Quebec, Canada
E-mail: nadine.mengis@concordia.ca

D. P. Keller · A. Oschlies
GEOMAR Helmholtz Centre for Ocean Research Kiel, Kiel, Germany

W. Rickels · M. Quaas
Kiel Institute for the World Economy, Kiel, Germany

30

1 Introduction

Large-scale deliberate manipulations of the Earth's climate, so called Climate Engineering (CE) in the form of carbon dioxide removal or solar radiation management, are discussed as potential methods to counteract anthropogenic climate change. Assessment of the different CE methods is still in its early stages and so far, based on the same metrics and indicators as typically used in climate change assessment studies (Kravitz et al 2013; Oschlies et al 2017). However, in an engineered climate the prevailing relationships between Earth system properties may no longer persist (Klepper and Rickels 2014). The most prominent example is the impact on the transient climate response to CO₂ emissions (Irvine et al 2012): In a natural climate system temperature rises when atmospheric CO₂ increases, whereas under Solar Radiation Management deployment, temperatures can stabilize or even decrease while atmospheric CO₂ remains unmitigated. Accordingly, metrics and indicators applied in the climate mitigation context might no longer be appropriate for CE assessment, making in turn the selection of indicators for CE assessment one of the major challenges in the current CE debate (Oschlies et al 2017).

Previous studies on CE assessment apply indicators typical used for climate change impact and climate change mitigation studies (e.g., Kravitz et al 2013; Irvine et al 2016), such as global mean temperature, precipitation, atmospheric CO₂, and energy balance related variables (see Figure 1 from Oschlies et al 2017). The selection of these indicators was guided by their suitability to describe the state of the Earth system in a way assumed relevant for society, for example, by reporting the model outcome with respect to global mean surface air temperature or terrestrial precipitation (Stocker et al 2013). These indicators are deemed useful because they serve as a proxy for variables of more immediate social relevance, such as climate extremes or impacts. However, the selection of such indicators to summarize model output is, even though in a kind of scientific consensus, a normative choice in the assessment of CE, since no unambiguous rules for selecting indicators exist (Böhringer and Jochem 2007). In their literature review, Oschlies et al (2017) analyzed which indicators have, until now, been used for the assessment of Solar Radiation Management (SRM), terrestrial or marine Carbon Dioxide Removal (t- and m-CDR, respectively) and for comparative studies of CE methods. They found that the non-comparative studies mainly concentrate on assessing the effectiveness in the targeted Earth system component and ambiguously select side effects or other Earth system components to be assessed. They concluded that a comprehensive assessment of CE would require a reevaluation of the indicator selection process from Earth system variables.

The selection of indicators for CE assessment has so far not been based on any set of agreed guidelines or a discussion about the underlying normative choices (Oschlies et al 2017). Here, we propose a more systematic and transparent approach to select appropriate indicators for CE assessment, the Systematic Correlation Matrix Evaluation (SCoMaE) recently developed by Mengis et al (2018). Initially we will describe a natural-science led approach, however it can later be used in conjunction with expert or stakeholder input, thereby enabling the iterative approach as suggested by Oschlies et al (2017). Starting with a broad set of Earth System variables, the SCoMaE method is applied to select an indicator set for three idealized CE model experiments: i) Large-scale Afforestation, ii) Ocean Alkalinity Enhancement and iii) Solar Radiation Management in the form of solar dimming.

The SCoMaE method is a bottom-up approach for the systematic selection of a set of uncorrelated indicators that allow a detailed description of the state of the system. In this study, three CE model scenarios and two conventional climate change scenarios are implemented in the University of Victoria Earth System Climate Model of intermediate complexity (Eby et al 2013). In addition, for each scenario we ran 15 single perturbed parameter simulations (i.e., a total of 75 simulations) to obtain a multi-model like ensemble for our identification of correlated and uncorrelated indicators. Our approach provides therefore a first step towards i) a more comprehensive CE assessment, which considers a broad set of relevant Earth system components and processes and ii) a more comprehensive comparison of different CE methods that target different components of the Earth system, and hence need to consider a broad set of indicators to include changes in all of the components.

The article is structured as follows. In section 2 we describe the model and implemented scenarios, as well as the methods for indicator selection and scenario evaluation. In section 3.1 we identify changes in correlations between Earth system variables induced by the three idealized CE scenarios, relative to an intermediate-high and a high future emission scenario. In section 3.2 we identify additional indicators needed for a comprehensive comparison of the three CE methods to each other and the two reference scenarios. Finally, in section 3.3 we evaluate the three idealized CE scenarios based on the identified indicators within the limits of our approach. We conclude our findings in section 4.

2 Methods

90 2.1 The University of Victoria Earth System Climate Model

The model employed is version 2.9 of the University of Victoria Earth System Climate Model (UVic ESCM), an Earth system model of intermediate complexity (Weaver et al 2001; Eby et al 2013). It includes schemes for ocean physics based on the Modular Ocean Model Version 2 (MOM2) (Pacanowski, 1995), ocean biogeochemistry (Keller et al 2012), and a terrestrial component including soil and vegetation dynamics (Cox et al 1998; Cox 2001; Meissner et al 2003). The ocean model is coupled to a thermodynamic sea-ice model (Bitz and Lipscomb 1999) with elastic visco-plastic rheology (Hunke and Dukowicz 1997). The atmosphere is represented by a two-dimensional atmospheric energy moisture balance model (Fanning and Weaver 1996).

100 All model components have a common horizontal resolution of 3.6° longitude and 1.8° latitude and the oceanic component has a vertical resolution of 19 levels, with vertical thickness varying between 50 m near the surface to 500 m in the deep ocean. Wind velocities are prescribed as monthly climatological wind fields from NCAR/NCEP reanalysis data (Eby et al 2013) and are used to calculate i) the advection of atmospheric heat and moisture and ii) the air-sea-ice fluxes of surface momentum, heat, and water fluxes. Wind anomalies, which are determined from surface pressure anomalies with respect to pre-industrial surface air temperature, are added to the prescribed wind fields (Weaver et al 2001). For the model simulations, the UVic ESCM was spun up with pre-industrial (year 1765) seasonal forcing for over ten thousand years. All simulations were integrated from 850 until 2005 using historical fossil-fuel emissions and land-use changes (LUC), as well as radiative forcing from solar variability and volcanic activity following Eby et al (2013). Following Keller et al (2014), continental ice sheets were held constant to facilitate the experimental setting and analyses. Warming from black carbon, indirect ozone effects, and cooling from indirect sulfate aerosol effects were not included.

105 The UVic ESCM is a well-established model of intermediate complexity and has shown to perform well in reconstructing past historical trends (Eby et al 2013) and has previously been used in simulations of hypothetical CE deployments (Keller et al 2014; Mengis et al 2016).

115 2.2 Perturbed Parameter Simulations

In addition to the model's default setting we performed 14 perturbed parameter simulations, to obtain a multi model-like ensemble. Due to the low internal variability in the UVic ESCM, we use this perturbed parameter ensemble to perform the correlation analysis as described in Section 2.4. The parameter perturbations address in particular the parameter uncertainty regarding the CO₂ and temperature sensitivity of the terrestrial and marine primary production (see Table S2 for a list of the simulations and Section S1 for a detailed explanation of the implementation).

120 The selected parameters were varied given observational constraints. The resulting simulations were verified to all lie within the range of annual global mean temperatures of the models from the 5th coupled model inter-comparison project and within natural variability of historic temperature anomalies (see Figure S1). The perturbed parameter simulations thus provide a multi-model like ensemble for the analysis of correlations among Earth system variables in non-CE and CE scenarios.

2.3 Simulated RCP Climate Change and Climate Engineering Scenarios

130 For the RCP climate change scenarios, the forcing from the Representative Concentration Pathway (RCP) 4.5 and 8.5 scenarios (Meinshausen et al 2011) were implemented. This includes CO₂ emissions as well as LUCs emissions and non-CO₂ climate forcing from 2005 onward until 2100.

For the CE scenarios, we simulate three idealized CE methods following Sonntag et al (2018):

- 135 i) Solar radiation management (SRM), directly affecting the radiation balance of the atmosphere (i.e. solar dimming),
- ii) large-scale afforestation (LAF), directly influencing primary production on land, and
- iii) ocean alkalinity enhancement (OAE) directly affecting ocean biogeochemistry.

140 All three CE methods were simulated using a business as usual climate change scenario (RCP8.5). The CE measures are then assumed to be substitutes for mitigation and are designed to target an intermediate-high climate change scenario (RCP4.5), with respect to one specific Earth system variable. We can then assess how well mitigation action can be substituted by an implementation of CE methods. Following the implementation of Sonntag et al (2018), all three CE methods were simulated from 2005 onwards (i.e., with the start of the RCP scenarios) to avoid abrupt changes in the system. In the following we will briefly describe the implementation of the three simulated CE scenarios.

145

150 In the SRM scenario, we reduced the amount of top of the atmosphere (TOA) incoming shortwave radiation in the model, mimicking a dimming of the sun, which could be achieved for example by successful deployment of sunshades (Lunt et al 2008; Kravitz et al 2011; Keller et al 2014). The reduction of incoming shortwave radiation was scaled such that TOA radiative forcing of the RCP4.5 scenario of the respective perturbed parameter set is reached (as in Sonntag et al 2018).

155 In the LAF scenario, we reduced the area of agriculturally used land following the RCP4.5 LUC scenario and increased in turn the area which is naturally reforested (as in Sonntag et al 2018). Such an unmanaged afforestation scheme assumes that the naturally occurring vegetation has a higher carbon storage capacity than agriculturally used land. In our model, the main transition occurred in the tropical areas, where the tropical grass lands, representing agriculture lands, were reduced in favor of broad-leaf trees, which indeed have a higher carbon content per square meter (Cox 2001).

160 In the OAE scenario, we enhanced ocean surface alkalinity to increase oceanic carbon uptake. The enhancement was scaled such that the atmospheric CO₂ concentration matches the concentration in the RCP4.5 scenario of the respective perturbed parameter set. Following González and Ilyina (2016) and Sonntag et al (2018), we accordingly calculated a multiplication factor as the ratio of simulated atmospheric CO₂ concentration to the target atmospheric CO₂ concentration. If the simulated concentration deviated by more than 1% from the target, we applied this factor (larger one) to the simulated surface ocean alkalinity, and thereby increased the ocean carbon uptake and reduced atmospheric CO₂.

2.4 Indicator Selection

170 Based on the review of CE studies with respect to commonly used assessment indicators by Oschlies et al (2017), we defined a set of 46 Earth system variables as the basis for the indicator selection process (see Table S1 for the full list). For each of the 46 variables simulated by the UVic ESCM, we calculated the change between the ending (2090-2100) and starting (2005-2015) periods of our scenarios (i.e., the long-term difference), across the perturbed parameter simulations. We calculated correlations of the changes for each pair of the 46 model output variables across the 15 parameter simulations and stored this information in a so-called correlation matrix, following the procedure and nomenclature of Mengis et al (2018) (see Figure S2 for a visualization of the correlation matrix construction). We then calculated Pearson correlation coefficients and define established statistical significance of the correlations after performing a two-sided t-test, on a 5% significance level.

180 Repeating this procedure for each of the five scenarios, we can then assess if the significance of correlations between model output variables has been altered by the imposed CE methods. For an easier visual evaluation of these changes we constructed common correlation matrices (Figures S4-S8). At first, we regard changes in correlations for each CE scenario with respect to the baseline RCP8.5 and the target RCP4.5 (Figure S4 – S6). Finally, we constructed common correlation matrices for 1) the two RCP climate change scenarios only, i.e., the RCP4.5 and RCP8.5, and then 2) for all five scenarios, the RCP4.5, RCP8.5, SRM, LAF and OAE (Figure S7 and S8, respectively).

185 Following the indicator selection of Mengis et al (2018), the significantly correlated variables of the matrices are assigned to clusters of variables that provide similar information (Figure S3 illustrates this process). This allows us to identify correlations between Earth system variables that are significant for all considered scenarios. The comparison of the non-CE and CE indicator sets show us the added value of this systematic approach over previous attempts in which existing climate change metrics were applied.

2.5 CE Scenario Performance Assessment

195 For the assessment of the CE scenarios against each other, in this example, instead of selecting single Earth system variables (and thereby assigning weights of zero to all the disregarded variables), we used the identified clusters to calculate composite indicators. We therefore aggregate the deviation of the climate state of all clustered variables at the end of the simulations relative to the 2005-2010 reference climate state. This means, our CE assessment supposes that the 2005-2010 climate state represents a desirable climate. It is however noteworthy, that staying close to today's climate may not be considered as appropriate reference state by those countries which are expected to benefit from increased warming (consider for example the identified climate change winners in Burke et al (2015)).
200 As a first step, we calculated the absolute deviation (dV) of all 46 Earth system variables (V) for each scenario ($scen$) and each parameter perturbation (pp) as:

$$dV(scen, pp) = |V_{t=2100}(scen, pp) - V(pp)_{t=2005}| \quad (1)$$

205 The fact that we use the absolute changes assumes that deviations from the reference state in both directions are considered to be equally 'bad' and scale linearly with the magnitude of the deviation. Obviously, using the absolute deviation is only one possible metric and using, for example, quadratic deviations instead, like Moreno-Cruz et al (2012) did in their assessment of SRM, would have shifted more weight to Earth system variables with a higher deviation.

210 To achieve comparability of the scenarios' performance and to allow for a meaningful aggregation of the clustered variables into composite indicators a normalization is required (Ebert and Welsch 2004). We normalized the absolute deviations of the variables by the observed maximum deviation across scenarios and parameter perturbations:

$$215 \quad dV_{norm}(scen, pp) = \frac{dV_{max} - dV(scen, pp)}{dV_{max}} = 1 - \frac{dV(scen, pp)}{dV_{max}} \quad (2)$$

220 Implicitly, this corresponds to a Min-Max normalization where we presume that a desired minimum deviation of zero is possible (i.e. $dV_{min} = 0$). By definition for this normalization a score close to 1 (0) indicates a small (large) deviation from the reference climate state and accordingly corresponds to a 'good' (poor) performance of the CE measure in staying close to it. As discussed above, our normalization reflects the assumption that the 2005-2010 climate state represents a desirable climate and reflects this normative judgement.

225 Having dimensionless and comparable variables, the composite indicators (I) were calculated as generalized means with equal weights, α_V for all variables in one cluster:

$$230 \quad I(scen, pp) = \left(\sum_{V=1}^N \alpha_V * V_{norm}^{\frac{\sigma-1}{\sigma}}(scen, pp) \right)^{\frac{\sigma}{\sigma-1}} \quad (3)$$

235 Again, calculating the generalized mean includes the normative decisions about the weighting factors, and about the specification of the substitution elasticity, σ . The substitution elasticity is a measure of how sensitive the score of the composite indicator is towards an extreme performance of single variables. In case $\sigma \rightarrow \infty$, the generalized mean simply becomes an arithmetic mean, implying that the deviation in one variable can be completely offset by an equal deviation in the other direction of another variable. Facing a complex system like the climate system with various not fully understood relationships, we propose using a substitution possibility below 1 to allow for some degree of substitution but with a significant influence of the overall performance from the indicator with poorest individual performance, accounting therefore for unbalanced performances across indicators (Rickels et al. 2014). We decided to use a substitution elasticity of $\sigma=0.5$, as used in previous studies concerning climate change (Stern and Persson 2008, see section S5 for sensitivity analysis on the substitution elasticity (σ) concerning our results).

240 These composite indicators take the information provided by all the variables within one cluster into account and were used to evaluate the performance of the five climate change scenarios against each other. Considering the information about changed correlations due to the scenario implementation, this method hence allowed us to comparatively assess deviations against the defined 2005-2010 reference climate state. Note, that the derived metric and its judgement of the performance of CE scenarios crucially depends on the abovementioned assumptions going into the selection of a reference climate state and the normalization.

245 **3 Results and Discussion**

3.1 Changes in correlations between Earth system variables induced by single CE scenarios

250 In the following section, we will shortly summarize most important changes in correlations between Earth system variables, that have been caused by single CE scenario implementations, and explain which processes caused the respective change in correlation. The obtained information provides insights into changes of model inherent processes in the case of CE implementation. This can guide future process understanding for model development, and should inform future indicator selection for assessment studies.

255 For the SRM scenario, we find the correlations between many temperature- and CO₂ related variables to become insignificant (Figure S4), which agrees with previous findings (Irvine et al 2012). Furthermore, the correlations between the land and ocean carbon fluxes towards the total ocean carbon become insignificant. This decoupling between the surface carbon fluxes and the total ocean carbon is caused by different time scales relevant for the reaction of these two processes: On the decadal to centennial time scale that we are assessing here, the surface carbon flux into the ocean reacts to the CE perturbation. It is increased by SRM, due to lower atmospheric and sea-

260 surface temperatures, while atmospheric CO₂ concentrations remain high. This signal however accounts only for a
small percentage change of the total ocean carbon over the period of 80 years, and hence only has a minor effect on
total ocean carbon. For the same reason, the ocean heat flux and TOA net radiation are no longer correlated to
neither the total ocean carbon, nor the total land carbon reservoir in case of SRM implementation.

265 For the LAF scenario, we do not find large deviations from the high RCP climate change scenario (RCP8.5)
concerning correlations between variables. This is somewhat expected because the LAF scenario assumed here has a
rather low carbon removal potential and stays therefore close to RCP8.5. However, we find that correlations
between ocean surface biogeochemistry and northern hemisphere sea ice and correspondingly ocean albedo become
insignificant because of the LAF implementation (Figure S5). This can be explained by the manipulation of the
270 vegetation and the resulting higher runoff rate, which alters the dilution rate of the surface ocean, and therefore the
biogeochemistry. In contrast to a non-LAF climate, where the surface water dilution is often correlated with the
fresh water input from melting sea ice, this is now no longer the case if the runoff becomes predominant. The same
response explains, why mean ocean phytoplankton is no longer correlated to surface properties of nitrate and
phosphate. The higher rate of fresh water input and hence the dilution of surface properties disrupts the relationship
of those two surface variables to the global mean phytoplankton growth.

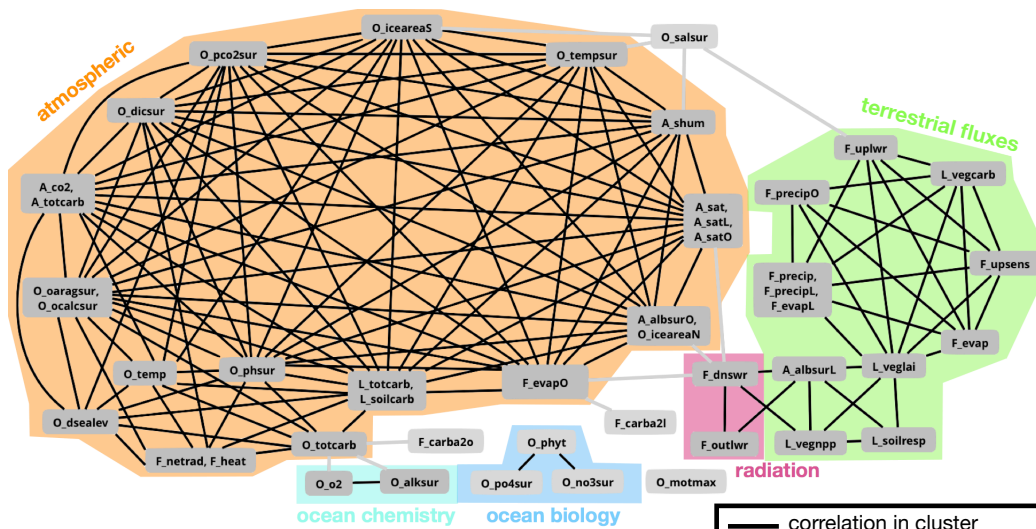
275 For the OAE scenario, we find changes in correlations for many ocean carbon-related variables (Figure S6). Under
non-OAE climate conditions, total ocean carbon and total land carbon are connected to each other, but if the ocean is
forced to take up relatively more of the atmospheric carbon, the relationship between the ocean, land, and
atmospheric carbon has changed, affecting in turn the carbon fluxes between these reservoirs (Keller et al 2018). In
280 the same way, the correlations between land carbon-related variables towards temperature and atmospheric carbon
related variables are changed in the OAE scenario: The land starts releasing carbon in this scenario, but the release
does not ultimately affect atmospheric carbon concentration and in turn temperature because the OAE scenario was
designed such that any deviation of atmospheric carbon from the reference scenario is addressed by further
increasing alkalinity (i.e., any increase in atmospheric CO₂ is countered with OAE).

285 **3.2 Comparing Correlation Cluster across Scenarios and Identification of additional indicators**

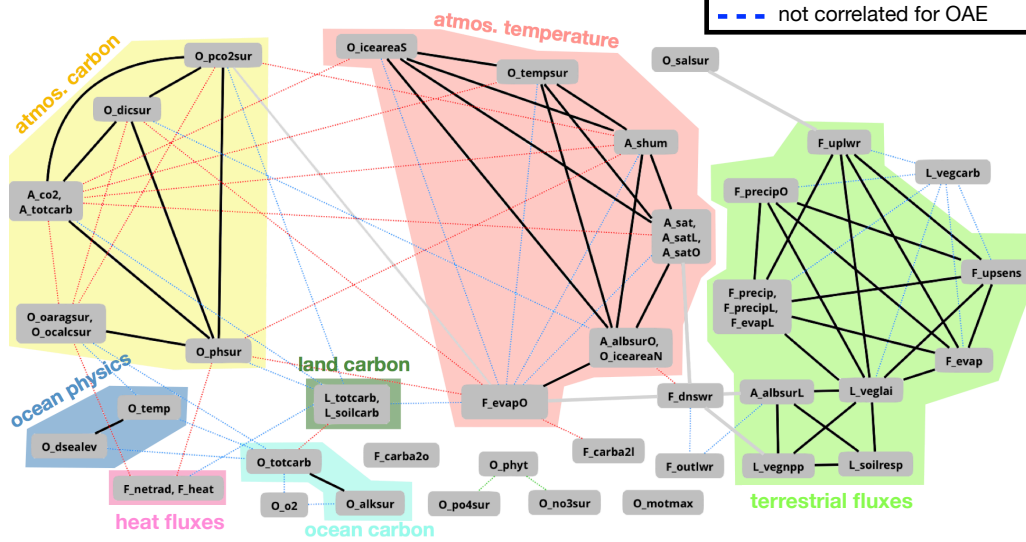
Comparing the correlation clusters of the two RCP climate change scenarios to the clusters identified for all five
scenarios (SRM, LAF, OAE, RCP4.5 and RCP8.5), clearly identifies the CE forcing-induced changes in correlations
(Figure 1).

290 The large atmospheric cluster initially identified from the correlation matrix of the two reference scenarios (Fig. 1a)
is divided into a total of 6 smaller clusters when all CE scenarios are included (Fig. 1b), which represent 1)
atmospheric temperature related variables, 2-4) atmospheric, land and ocean carbon related variables, 5) ocean
physics related variables, and 6) radiative fluxes. The vegetation indicator remains almost the same for both
295 analyses, with the exception of the vegetation carbon variable, which becomes a single indicator when CE scenarios
are included. Finally, in addition to the four single indicators identified for the two reference scenarios (Fig. 1a),
ocean surface phosphate and nitrate, ocean phytoplankton and oxygen, vegetation carbon, and outgoing longwave
radiation become single indicators identified from the correlation matrix of the five scenarios (Fig. 1b).

a) Identified clusters based on the correlation matrix of the GHG driven scenarios



b) Identified clusters based on the correlation matrix of all five climate change scenarios



300

305

310

315

Figure 1 Indicator clusters identified from the systematic correlation matrix evaluation (SCoMaE) of the common correlation matrix of the target and baseline RCP climate change scenarios (RCP4.5 and RCP8.5, respectively) (a), (correlation matrix shown in Figure S7), and of the common correlation matrix of the solar radiation management (SRM), large-scale afforestation (LAF), ocean alkalinity enhancement (OAE), target (RCP4.5) and baseline (RCP8.5) scenarios (b), (correlation matrix shown in Figure S8). Black lines indicate correlations significant for all scenarios within a cluster, shading indicates identified clusters based on the correlation information, light grey lines, indicate significant correlations outside of an identified cluster. Colored dotted lines indicate which CE scenario is causing the correlation to be insignificant, i.e., the correlation is significant in all scenarios but the SRM (red dotted), LAF (green dotted), or OAE (blue dotted) scenario. See Table S1, for the abbreviation explanations.

The separation of the large atmospheric cluster was mostly caused by changes in correlations due to either the SRM or the OAE scenarios (red and blue dotted lines in Figure 1b, respectively, compare figure S8). The separation of variables representing atmospheric carbon and atmospheric temperature into two clusters is caused by the implementation of SRM scenario. The increasing atmospheric CO₂ concentration is no longer causing an atmospheric temperature increase, because by definition SRM manipulates the radiation balance to cool the global mean temperature.

320 The separation of the atmospheric carbon cluster depicted in Fig. 1a into the two other carbon reservoirs in Fig. 1b is
caused by the OAE implementation, but the separation between the land and the ocean carbon reservoir is caused by
the SRM scenario forcing. Under the OAE forcing scenario the ocean's surface chemistry, namely the surface
alkalinity, is manipulated so that it is taking up a disproportional amount of carbon from the atmosphere, no longer
connected to the atmosphere to ocean gradient. The land in this scenario reacts in the same way as always and
reduces its carbon uptake in the case of reduced atmospheric CO₂. In contrast to this, in the SRM forcing scenario,
325 both the land and the ocean carbon uptake are increased due to the coincidence of high atmospheric CO₂ and lower
temperature. Accordingly, for a comparative assessment of both these scenarios, we henceforward must consider
changes in all the three carbon reservoirs. The consideration of the ocean physics cluster and of ocean oxygen as a
single indicator, due to the separation from the ocean carbon cluster, is caused by the OAE forcing. Again, the large
increase in total ocean carbon in this scenario is no longer linked to ocean physics or the ocean oxygen
concentration. Lastly, the radiation cluster concerning ocean heat fluxes (Fig. 1 b) is separated from the big
330 atmosphere cluster (Fig. 1a) due to the SRM and the OAE forcing implementation. The OAE forcing decouples the
ocean heat flux from the ocean carbon fluxes, and accordingly from the carbon reservoirs. In addition, the SRM
forcing decouples the ocean heat flux from the ocean surface properties connected to atmospheric carbon.

335 Apart from the separation of the large atmospheric cluster into 6 smaller ones, the SRM, OAE and LAF forcing
scenarios cause separation of several variables to become single indicators (Figure 1b). While outgoing longwave
radiation is usually connected to the terrestrial surface albedo (Fig. 1a), or the downward shortwave radiation, this is
no longer true in case of the OAE implementation. In the OAE scenario the outgoing longwave radiation is
connected to atmospheric carbon-related variables (Figure S6), which are however already considered within
another cluster. In addition, the surface shortwave radiation in the SRM scenario is disconnected from the sea ice
340 concentration and is linked to the albedo on land (Figure S4), a variable which is again already considered in another
cluster. This causes the surface shortwave radiation and the top of the atmosphere longwave radiation to become
single indicators. The SCoMaE method hence, singles out these two variables to be considered as single indicators.
Furthermore, the OAE forcing scenario causes the vegetation carbon variable to become a single indicator, by
changing its correlations towards the moisture and radiative fluxes. This disruption is caused by the fact, that the
345 OAE scenario is the only scenario in which deforestation (as in the RCP8.5) coincides with low atmospheric carbon
concentrations (see more detailed explanation in section S3.3). Finally, the LAF forcing scenario causes the mean
ocean phytoplankton variable to be considered separately from the surface nitrate and phosphate variables, causing
all three to become single indicators. As explained earlier (section 3.1) the manipulation of the vegetation and the
resulting higher runoff rate alter the dilution rate of the surface ocean, and therefore the biogeochemistry. This
350 explains, why mean ocean phytoplankton is no longer correlated to surface properties of nitrate and phosphate.

In conclusion, each of the implemented CE scenarios causes changes in correlations between Earth system variables
relative to the two RCP climate change scenarios. These are reflected in the selection of this uncorrelated, non-
redundant indicator set shown in Fig.1b, including composite and single indicators, which now allow for a
355 comparative assessment of the state of the Earth system for all three CE and the two RCP climate change scenarios.

Assessment framework for indicator selection	
RCP climate change scenarios (RCP4.5, RCP8.5)	all climate change scenarios (RCP4.5, RCP8.5, SRM, LAF, OAE)
Atmospheric CI (23)	Atmospheric temperature CI (9)
	Atmospheric carbon CI (7)
	Land carbon CI (2)
	Ocean physics CI (2)
	Heat flux CI (2)
Terrestrial fluxes CI (12)	Ocean carbon CI (2)
	Terrestrial fluxes CI (11)
Ocean biology CI (2)	Vegetation carbon SI (1)
	Ocean Phytoplankton SI (1)
	Ocean surface nitrate SI (1)
Radiation CI (2)	Ocean surface phosphate SI (1)
	Surface shortwave radiation SI (1)
Ocean chemistry CI (2)	TOA outgoing longwave radiation SI (1)
	Ocean carbon CI (2)
Atmos. to ocean carbon flux SI (1)	Ocean oxygen SI (1)
Atmos. to land carbon flux SI (1)	Atmos. to ocean carbon flux SI (1)
Sea surface salinity SI (1)	Atmos. to land carbon flux SI (1)
Ocean overturning SI (1)	Sea surface salinity SI (1)
	Ocean overturning SI (1)

Table 2 List of calculated and selected composite indicators (CI) and single indicators (SI for the two different assessment frameworks). Composite indicators are given in bold, the number of aggregated variables is given in brackets.

360

3.3 Evaluation of the CE model experiments based on the identified indicator set

365

Based on the previous analysis we are able to identify clusters of variables which are robust across all climate change scenarios (shading in Figure 1b). These clusters are used to calculate composite indicators, and the remaining un-clustered variables become single indicators (section 2.5). This approach does eliminate the expert-guess-driven selection of single assessment indicators, since we here take the information from all 46 variables into account. Note, that we are not evaluating the CE measures based on their correlation strength. Correlation strength ‘only’ informs the indicator selection process to obtain a more comprehensive indicator set. The CE scenario is then evaluated according to the deviation of the indicators from the 2005 reference state. Hereafter, this is referred to as “performance” when comparing the scenarios. In the following, we will contrast the results of two such assessments based on

370

- i) an indicator set given the correlation structure of climate change scenarios only (i.e., applying the indicator set identified in Fig. 1a and Table 2 (left column) to the CE simulations even though these clearly do not represent the correlation changes in the CE simulations)
- ii) an indicator set given the correlation structure we identified from all five scenarios (i.e., climate change and CE; the indicators depicted in Fig. 1 b and listed in Table 2 (right column)).

375

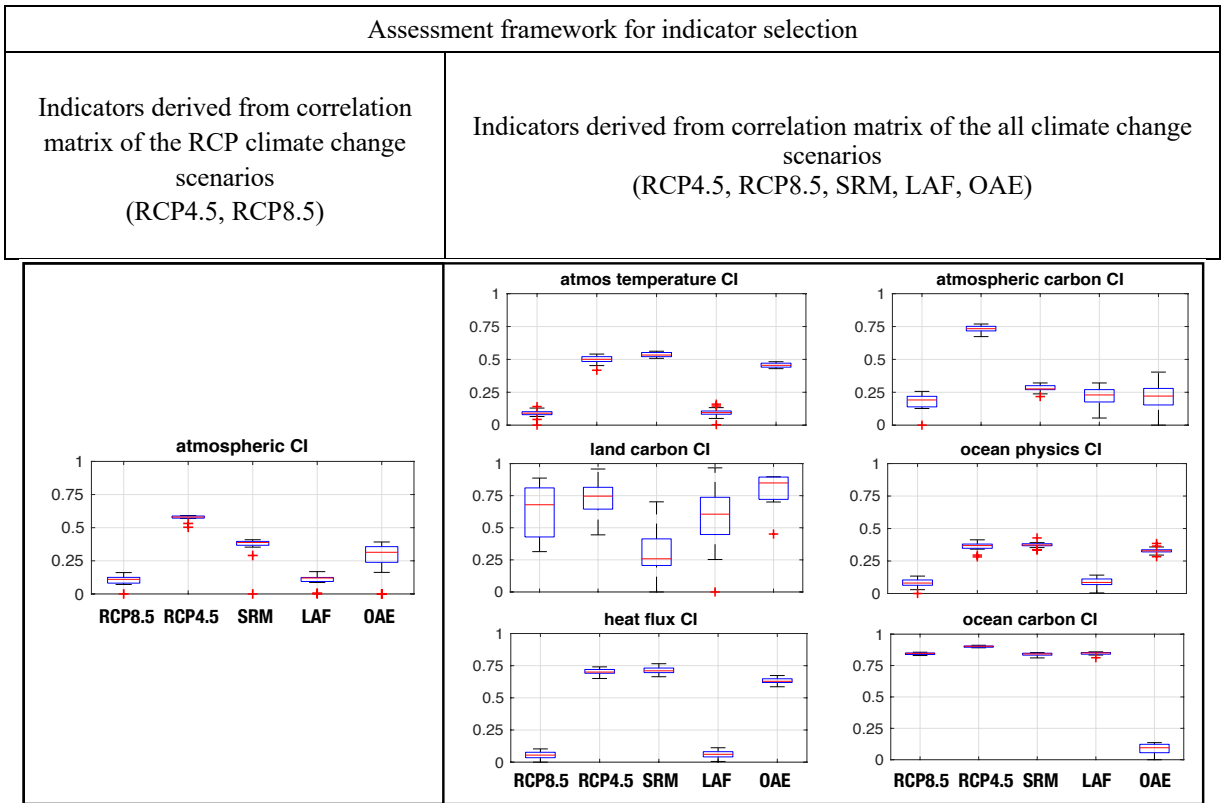
Comparing these two assessment frameworks will illustrate the added value of an adjustment of indicator sets for a comprehensive and comparative assessment of CE scenarios.

380

For the RCP climate change assessment, we accordingly have four single indicators and aggregate the remaining 42 variables into five composite indicators. For the assessment based on all five scenarios, we obtain eleven single indicators and aggregate the remaining 35 variables into seven composite indicators. The selected and calculated single and composite indicators are summarized in Table 2. Due to the CE-induced changes in correlations, we obtain a higher number of indicators, and hence more detailed information about the performance of the single CE

385 methods. Note, that at this stage the indicators do not necessarily all hold impact-relevant information. Such an impact-relevant indicator set could be determined in an iterative process with stakeholders. The framework presented here informs about the changes in correlations induced by CE, and is the first step to identifying a set of indicators for a more comprehensive comparative assessment of different CE methods. For the atmosphere indicator of the RCP climate change driven scenarios, we find the SRM scenario to perform best followed by the OAE and the LAF scenario (Figure 2, left panel). In our indicator set for all scenarios (Figure 2, right panel), the Earth system changes captured by the atmosphere indicator are now represented by 6 different indicators (due to changes in correlations induced by the SRM and OAE scenarios). And in fact, the additional indicators provide more differentiated information.

390 The SRM scenario performs well for the atmospheric temperature, ocean physics, radiation and ocean carbon indicator, but has low scores for the atmospheric and land carbon indicators (Figure 2, right panel). The relatively poor performance (i.e., strong deviation from the reference state, indicated by a value close to zero) in the land carbon indicators is caused by the decoupling of atmospheric temperature and carbon content. In the RCP8.5 scenario the increase in atmospheric CO₂ acts to increase the vegetation carbon and hence the total land carbon. However, in contrast to the SRM scenario, in the RCP8.5 scenario the increasing temperatures at the same time act to decrease soil carbon by enhancing soil respiration. This effect is much smaller in the SRM scenario because it is cooler, which causes the total land carbon to strongly increase in this scenario (Figure S9). Note, that a deviation in any direction (positive or negative) from the reference state lowers the performance of the indicators.



405 **Fig. 2** Assessment of deviations from the climate reference state for all five scenarios concerning atmospheric processes. The left panel shows the boxplot for the atmospheric composite indicator (CI) which was derived using the RCP climate change scenarios only. The right panel shows the boxplots for the indicators, which are derived using correlation information from all scenarios and correspond to the previously identified atmospheric CI. Both indicator sets comprise the same Earths system variables, but their aggregation differs due to the considered correlation information and according clustering as shown in Figure 1. The indicators are normalized and all composite indicators (CI) are aggregated based on the identified variable clusters. The performance of the various indicators is assessed as their respective deviations from the 2005–2010 reference state, after normalization (i.e., best performance if close to 1, bad performance if close to zero), as explained in section 2.4. The boxplots include information of all the 15 parameter perturbation simulations (i.e., the box represents the 25th and 75th percentile, and red crosses show outliers) and show the range for the two reference scenarios, RCP8.5 and RCP4.5, and each of the three CE scenarios, SRM, LAF, and OAE.

415 The OAE scenario performs well according to the atmospheric temperature, land carbon, ocean physics and radiation indicators, but less well for atmospheric and ocean carbon (Figure 2, right panel). It is surprising that the OAE does not show a good performance in the atmospheric carbon, since this is a variable that is targeted by the CE measure. But, the atmospheric carbon indicator comprises not only information of the atmospheric CO₂ concentration variable, for which the OAE simulations stays close to the target scenario, but also some ocean surface carbon variables (e.g., ocean surface pH and surface dissolved inorganic carbon). For these variables, the OAE scenario causes a comparatively large deviation and hence these variables cause a lower performance of this composite indicator (Figure S9). The relatively poor performance with respect to the ocean carbon indicators is caused by the manipulation of the model's surface alkalinity to increase the oceans carbon uptake in this scenario (Figure S9).

420

425 According to the atmospheric indicators, the LAF scenario shows overall low performances, except for the land and ocean carbon indicators (Figure 2, right panel). But in general, this scenario does not show a high potential in mitigating atmospheric climate changes, i.e., the climate state is rather close to that of the RCP 8.5 scenario.

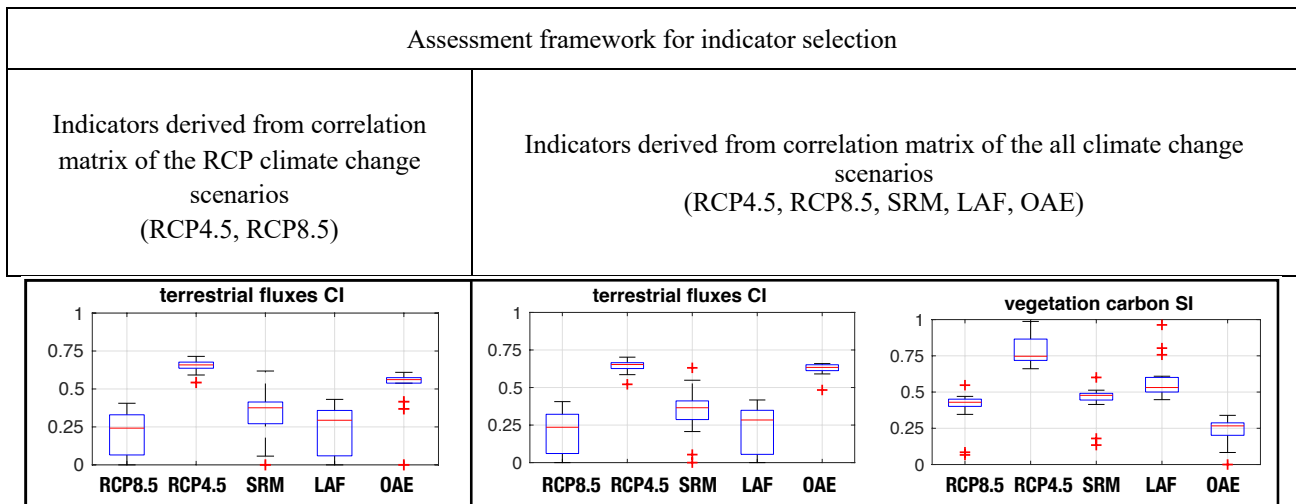


Fig. 3 Same as Figure 2, but for the terrestrial flux composite indicator (CI), for which we obtain an addition single indicator (SI), namely the vegetation carbon SI.

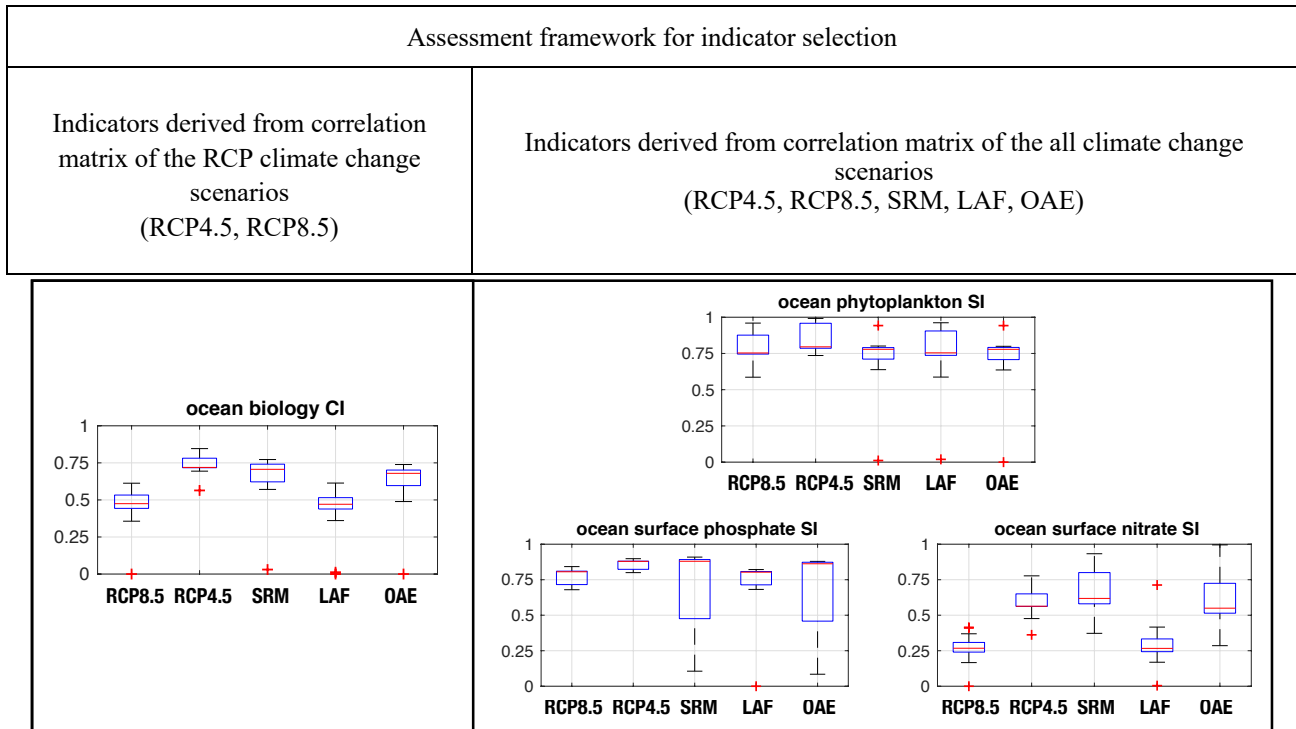
430 The terrestrial fluxes indicator, which comprises information about heat and moisture fluxes is very similar for both assessments, RCP climate change driven and all five scenarios (compare Figure 1). The only difference is the separation of the vegetation carbon variable into a single indicator due to correlation changes in the OAE scenario. Initially the terrestrial composite indicator shows a good performance of the OAE scenario, followed by the SRM and then the LAF scenario (Figure 3, left panel). The OAE yields a high score because both, atmospheric carbon and consequently atmospheric temperatures are reduced, i.e., the terrestrial moisture fluxes are staying close to the RCP45 reference scenario (Figure S9). For the SRM scenario the temperatures are reduced but CO₂ stays comparatively high, which favors a reduction in vegetation moisture fluxes (Mengis et al 2015), and hence causes a lower score. In the same way, the LAF scenario, is not able to mitigate atmospheric CO₂ concentrations, and in addition maintains high temperatures, causing the terrestrial indicator to score as low as the RCP8.5 scenario. This ranking for the terrestrial indicator is still true for the newly identified indicator based on all five scenarios (Figure 3, right panel). However, the additionally identified vegetation carbon indicator shows a different picture. This indicator shows a higher score for the LAF scenario, followed by the SRM and then the OAE scenario. This ranking is directly linked to the implemented LUC schemes in the scenario. The LAF scenario follows the RCP4.5 LUCs, causing the vegetation carbon to increase. The slight deviation relative to the RCP4.5 in the LAF scenario is caused by the higher CO₂ concentrations and the consequent CO₂ fertilization effect, causing an even higher increase in vegetation carbon (Figure S9). Again, note that deviations in both directions are penalized by the metric. The SRM scenario and the OAE scenario follow the LUCs of the RCP8.5 scenario, which further reduced global vegetation carbon. Meanwhile, the SRM scenario has a high atmospheric CO₂ concentration, increasing the CO₂ fertilization effect, and hence performs slightly better than the RCP8.5 scenario. In contrast the OAE scenario, combines deforestation (this is included in the RCP 8.5 LUC forcing), with low atmospheric CO₂ concentrations and hence shows the highest reduction in vegetation carbon, causing the low score for this indicator (Figure S9). This information would not have been obtained if we had simply used the indicator set derived from only the two RCP climate change scenarios.

435

440

445

450



455 **Fig. 4** Same as Figure 2, but for the ocean biology composite indicator (CI), and the three addition single indicators (SI).

The ocean biology indicator identified by the RCP climate change driven only indicator set, is separated into three single indicators caused by changes in correlations due to the LAF scenario (compare Figure 1). When indicators based on the RCP scenarios are used, we find a clear ranking, in which the SRM scenario has the highest score close to the RCP4.5, followed by the OAE and then the LAF scenario which performs slightly worse than the RCP8.5 scenario (Figure 4, left panel). If the three newly identified indicators are regarded, this ranking becomes less clear. It becomes evident, that the bad score of the LAF scenario is caused by the ocean surface nitrate and phosphate which are reduced in the LAF scenario due to high surface freshwater input (Figure S9). For the ocean phytoplankton, however the LAF scenario is even performing slightly better than the other two CE scenarios (Figure 4, right panel).

460

465

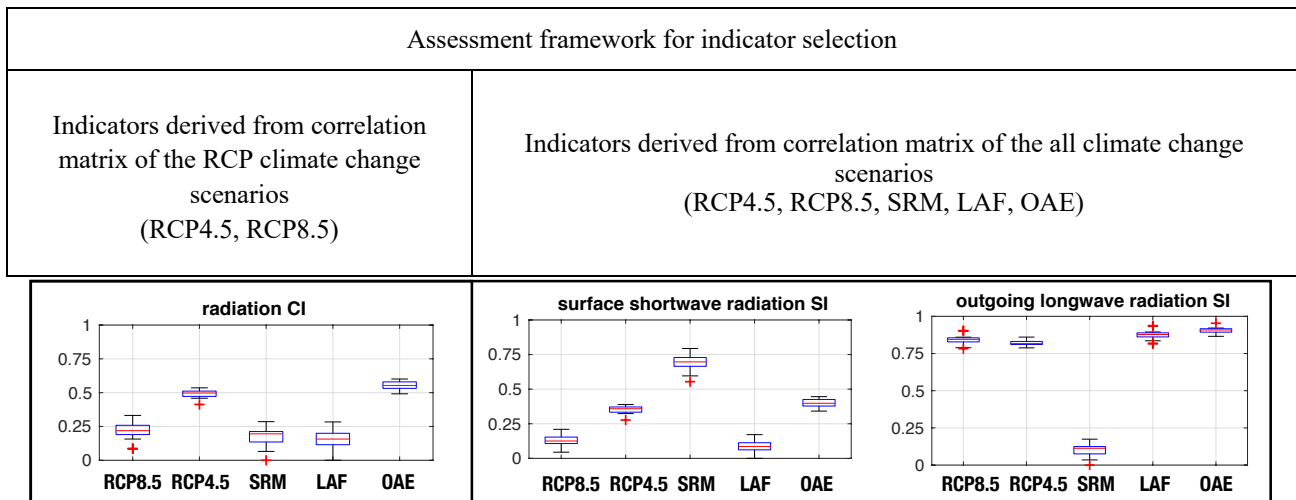


Fig. 5 Same as Figure 2, but for the radiation composite indicator (CI), for which we obtain two addition single indicators (SI).

470 The radiation indicator identified by the RCP scenario analysis, shows a good performance of the OAE scenario and low performances by the other two CE scenarios (Figure 5, left panel). Changes in correlations induced by both, the OAE and the SRM scenario, cause this indicator to be separated into two single indicators, the surface shortwave radiation and the outgoing longwave radiation indicators. Regarding those two indicators separately it becomes evident, that the OAE scenario performs well according to both, while the SRM scenario shows a very high score for mitigating the surface shortwave radiation, but a very low score for the outgoing longwave radiation (Figure 5, right panel). This behavior is explained by the manipulation of the incoming shortwave radiation for this CE measure, which stays close to the reference climate state. In turn the outgoing longwave radiation is strongly reduced, since the atmospheric CO₂ concentration remains high, but temperatures become low (Figure S9).
475

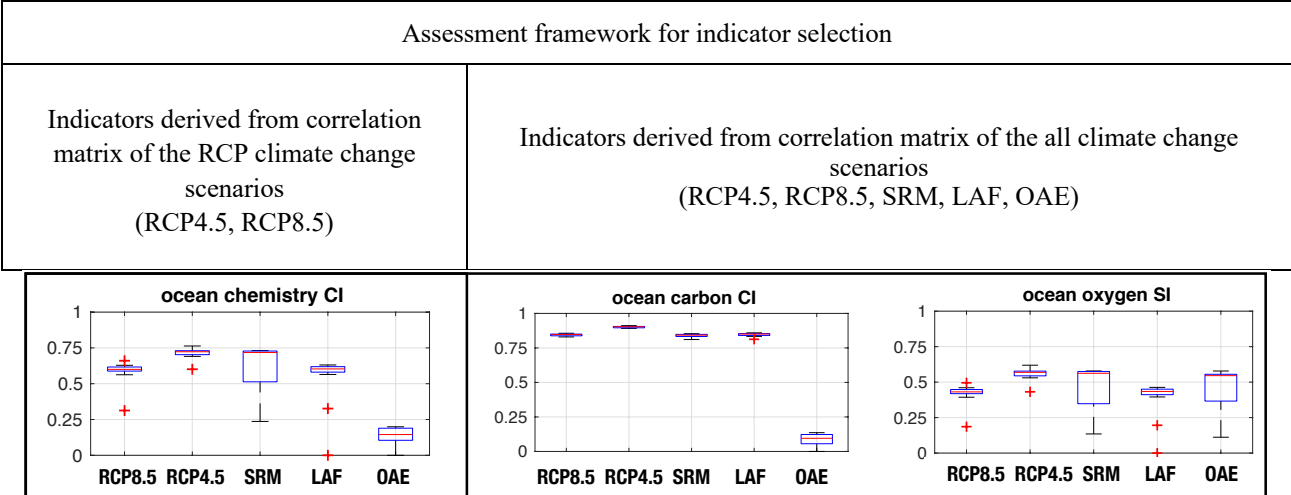


Fig. 6 Same as Figure 2, but for the ocean chemistry composite indicator (CI), which is changed into the ocean carbon CI and the ocean oxygen single indicator (SI).

480 The ocean chemistry indicator identified by the RCP scenario analysis, shows a good performance for the SRM and also the LAF scenario, but a very low performance for the OAE scenario (Figure 6, left panel). The changes in correlation induced by the OAE scenario causes this indicator to become an ocean carbon indicator and to regard the ocean oxygen as a separate indicator. From this separation, we can see that it is indeed only the ocean carbon that causes the bad performance of the OAE scenario, while the ocean oxygen indicator shows very similar performances for all three scenarios (Figure 6, right panel).
485

We can summarize that the indicators that are added due to the separation of the clusters by the induce correlation changes do provide new information valuable for the assessment of the CE measures.

4 Conclusion

490 This is the first study to systematically use CE induced changes in correlations across climate variables to comprehensively assess Climate Engineering scenarios against conventional climate change mitigation scenarios. We show that not only the implementation of SRM breaks down prevailing relationships in the Earth system as previously shown (Irvine et al 2012), but that adding large amounts of alkalinity to the ocean and large-scale afforestation would also change prevailing correlations in the Earth system. Although some changes may seem intuitive in hindsight, our methodology identifies all correlation and related process (identified by deriving indicators) changes. Such changes need to be considered in the selection of appropriate indicators or the development of metrics for a comprehensive assessment of any stand-alone, comparative and combined climate engineering scheme. In addition, this information should be considered in any future CE impact assessment study for which only a limited number of model output variables is chosen. Considering precipitation and temperature only (as indicators of how other climate system properties may change), might be appropriate for RCP driven climate change assessment, but does not suffice for a CE evaluation.
495
500

To enable a more comprehensive comparative assessment of the CE scenarios and the two RCP driven scenarios, we systematically evaluated the common correlation matrix of the five scenarios and compared the identified clusters to the RCP climate change driven only matrix. We can clearly see the added value of the additional indicators for the overall assessment. In the end, we could identify seven scientifically meaningful clusters, in addition to eleven
505

single indicators, which provide an initial framework, that can be used for a comprehensive and comparative assessment of the five climate change scenarios.

510 Our analyses based on the 18 indicators provides insights in the changes in the Earths system induced by CE, such
as unproportioned changes within one component of the Earth system, implementation specific problems (side
effects), or changes in the indicators induced by the changed correlations in the model itself. An evaluation of the
CE scenarios based on the 18 identified indicators shows a tendency for the SRM scenario to deviate the least from
515 the defined reference climate state for many indicators followed by the OAE and the relatively ineffective LAF
scenario. Overall, our analysis provides an assessment of which CE method performs “best” depending on the
selected indicators, and based on multiple normative decisions, which in this study were made as transparent as
possible. These include the selection of the CE scenario and its model implementation, the choice of relevant
variables (e.g. in the present study we only looked at global mean properties), the selection of the reference state, the
520 functional form of measuring the deviation of the reference state, the specific value for the elasticity of substitution,
and the selection of the underlying ESM. Clearly, these normative decisions require transparent and explicit
communication of the calculation, transformation, weighting, and aggregation schemes applied, supported by
scientific arguments. Our analysis aims for providing such an assessment framework, in which these choices and the
accordingly presented indicators are made transparent, and which can be applied to various CE scenarios. This is in
525 contrast to some earlier assessment studies on climate engineering in the literature, where the implicit value and
weighting decisions involved in the selection of the presented model results were not transparently discussed.

In agreement with Oschlies et al (2017), we suggest having a debate on indicator selection in an iterative process,
which is possible within the SCoMaE framework, since it allows for pre-selected indicators, chosen due to their
530 political or societal relevance. It is beyond the scope of this study to give a value judgment on which of the
indicators is of higher importance for society, or what is the desired climate state. Clearly, the presented evaluation
is restricted to natural-science considerations. A fully comprehensive assessment would require the inclusion of
socio-economic indicators, which may also require the inclusion of spatially disaggregated Earth System variables.
For such an analysis, however, the UVic ESCM multi model-like ensemble should be replaced by a higher
535 complexity Earth system model multi-model ensemble. Here we simply aim to provide the technical framework to
enable such a discussion. It is also important to keep in mind that side-effects of the different CE methods require a
more detailed assessment, particularly, an extensive risk analysis on regional scales. Our analysis can guide this
future assessment to be more comprehensive.

540 **Acknowledgements** The authors thank Wolfgang Koeve and Ulrike Loeptien for helpful comments and discussion,
as well as the participants of the Metrics Workshop of the SPP 1689 in March 2015, Hamburg, and in April 2016 in
Kiel, for their thoughts on metrics and indicators.

The data used to generate the figures will be made available at <http://thredds.geomar.de>.

545 This work was funded by the DFG Priority Program: Climate Engineering: Risks, Challenges, Opportunities? (SPP
1689). The authors declare that they have no competing financial interests.

N.M., D.P.K., W.R. M.Q. and A.O. conceived the experiment. N.M. and D.K. implemented and performed the
simulations. N.M. and W.R. analyzed the data, and wrote the manuscript with contributions from D.P.K. M.Q. and
A.O..