Reliability Ensemble Averaging of 21st century projections of terrestrial net primary productivity reduces global and regional uncertainties

Citation for published version:

Digital Object Identifier (DOI):
10.5194/esd-9-153-2018

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published In:
Earth System Dynamics

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
Reliability ensemble averaging of 21st century projections of terrestrial net primary productivity reduces global and regional uncertainties

Jean-François Exbrayat$^1$, A. Anthony Bloom$^2$, Pete Falloon$^3$, Akihiko Ito$^4$, T. Luke Smallman$^1$, and Mathew Williams$^1$

$^1$School of GeoSciences and National Centre for Earth Observation, University of Edinburgh, Edinburgh, EH9 3FF, UK
$^2$Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California, USA
$^3$Met Office Hadley Centre, Fitzroy Road, Exeter, EX1 3PB, UK
$^4$National Institute for Environmental Studies, Tsukuba, Japan

Correspondence: Jean-François Exbrayat (j.exbrayat@ed.ac.uk)

Received: 14 September 2017 – Discussion started: 19 September 2017
Accepted: 16 January 2018 – Published: 21 February 2018

Abstract. Multi-model averaging techniques provide opportunities to extract additional information from large ensembles of simulations. In particular, present-day model skill can be used to evaluate their potential performance in future climate simulations. Multi-model averaging methods have been used extensively in climate and hydrological sciences, but they have not been used to constrain projected plant productivity responses to climate change, which is a major uncertainty in Earth system modelling. Here, we use three global observationally orientated estimates of current net primary productivity (NPP) to perform a reliability ensemble averaging (REA) method using 30 global simulations of the 21st century change in NPP based on the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP) “business as usual” emissions scenario. We find that the three REA methods support an increase in global NPP by the end of the 21st century (2095–2099) compared to 2001–2005, which is 2–3 % stronger than the ensemble ISIMIP mean value of 24.2 Pg C $^{-1}$. Using REA also leads to a 45–68 % reduction in the global uncertainty of 21st century NPP projection, which strengthens confidence in the resilience of the CO$_2$ fertilization effect to climate change. This reduction in uncertainty is especially clear for boreal ecosystems although it may be an artefact due to the lack of representation of nutrient limitations on NPP in most models. Conversely, the large uncertainty that remains on the sign of the response of NPP in semi-arid regions points to the need for better observations and model development in these regions.

1 Introduction

Anthropogenic emissions of carbon dioxide (CO$_2$) enhance the uptake of atmospheric carbon by terrestrial ecosystems through net primary productivity (NPP). This so-called CO$_2$ fertilization effect has helped offset 25–30 % of CO$_2$ emissions responsible for climate change in recent decades (Canadell et al., 2007; Le Quéré et al., 2009). There exists a large uncertainty as to whether this positive effect of CO$_2$ fertilization will be resilient to climate change, as shown by the spread between model projections from various intercomparison projects (Friedlingstein et al., 2006; Arora et al., 2013; Friend et al., 2014; Nishina et al., 2014, 2015), especially in highly productive tropical regions (Ramig et al., 2010; Cox et al., 2013). However, large ensembles of projections are challenging to interpret as they may include models with an opposite response to the same change in boundary conditions (Friedlingstein et al., 2006). Simulations from the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP; Warszawski et al., 2014) have shown that most of the uncertainty in 21st century projections of the terrestrial carbon cycle can be attributed to differences
among global vegetation models (GVMs; Friend et al., 2014; Nishina et al., 2014, 2015), although a non-negligible part of the uncertainty arises from differences in climate projections themselves (Ahlström et al., 2012).

In recent years multi-model averaging has been widely used to extract information from large ensembles of simulations in studies targeting climate change (Bishop and Abramowitz, 2012; Krishnamurti et al., 1999), rainfall–runoff processes (Georgakakos et al., 2004; Huisman et al., 2009; Shamseldin et al., 1997; Viney et al., 2009) and catchment-scale nutrient exports (Exbrayat et al., 2010, 2013b). These methods range from simple arithmetic means of model ensembles to more elaborate weighting schemes such as Bayesian that take model performance into account model averaging (Raftery et al., 2005). The underlying assumption is that a model that is better able to reproduce current conditions should be given more weight in the final projection than a poorly performing model. The more complex reliability ensemble averaging (REA; Giorgi and Mearns, 2002) approach takes into account a measure of convergence between projections to identify the most likely change: this way, the REA method avoids giving too much weight to an overfitted model that may accurately represent current conditions for the wrong reasons but predicts vastly different change than other ensemble members (Exbrayat et al., 2013b). Metrics measuring model independence (Bishop and Abramowitz, 2012) have also been introduced in weighting schemes to avoid pseudo-replication.

Until recently, applying these advanced multi-model averaging methods to simulations of the global carbon cycle has remained a challenge because of the lack of global observational datasets to constrain, for example the REA weighting scheme. Schwalm et al. (2015) have presented results of skill-based model averaging applied to an ensemble of 10 models from the Multi-scale synthesis and Terrestrial Model Intercomparison Project (MsTMIP; Huntzinger et al., 2013). This pixel-wise approach assigned weights to historical simulations based on their ability to simulate gross primary productivity and biomass stocks but did not consider future projections. Lovenduski and Bonan (2017) considered a single value of cumulative terrestrial carbon uptake for 1959–2005 to derive one global coefficient per model to produce new projections. However, we are not aware of any studies using these methods in the context of spatially-explicit projections of the terrestrial carbon cycle under climate change.

Here, we present the first example of a pixel-wise application of the REA approach to extract a best estimate of NPP change (ΔNPP) during the 21st century under a “business as usual” scenario of emissions from a large ensemble of projections. We perform the REA procedure three times using different observationally constrained estimates of current NPP: retrievals of the terrestrial carbon cycle with the CARbon DAta Model fraMework (CARDAMOM; Bloom et al., 2016) model–data fusion approach (Bloom and Williams, 2015a; Bloom et al., 2016), an approximation of NPP based on the up-scaled FLUXCOM gross primary productivity (GPP) datasets Jung et al., 2009, 2011, 2017; Tramontana et al., 2016) and the MOD17A3 MODIS NPP product (Running et al., 2004; Zhao et al., 2005; Zhao and Running, 2010). Based on optimally weighted model averages, we evaluate the impact of the REA method on 21st century projections of ΔNPP but also on the uncertainty in the future resilience of the CO₂ fertilization that exists among the models. We show that the REA procedure can help identify regions where uncertainties remain large and thereby inform the future development of models and observational networks needed to improve climate change projections.

2 Materials and methods

2.1 The ISIMIP ensemble

We used an ensemble of simulations of NPP from the ISIMIP. The ISIMIP simulations included here consist of six global vegetation models (GVM): HYBRID (Friend and White, 2000), JeDi (Pavlick et al., 2013), JULES (Clark et al., 2011), LPJmL (Sitch et al., 2003), SDGVM (Woodward et al., 1995) and VISIT (Ito and Inatomi, 2012). Each of these six GVMs was driven by bias-corrected output (Hempel et al., 2013) from five general circulation models (GCM): GFDL-ESM2M (Dunne et al., 2012), HadGEM2-ES (Collins et al., 2011), IPSL-CM5A-LR (Dufresne et al., 2013), MIROC-ESM-CHEM (Watanabe et al., 2011) and NorESM1-M (Bentsen et al., 2013), generating a total of 30 global simulations of NPP for the historical period and under the Representative Concentration Pathway 8.5 (RCP8.5). We chose the ISIMIP ensemble over other initiatives like C4MIP (Friedlingstein et al., 2006) or CMIP5 (Taylor et al., 2012) because the combination of multiple GVMs with multiple GCMs in ISIMIP allows a more comprehensive coverage of the uncertainty in the terrestrial carbon cycle and attribution of dominant factor in the uncertainty of the future (Friend et al., 2014; Nishina et al., 2014, 2015), although we note that these simulations omit feedbacks from the biosphere on weather and atmospheric CO₂ concentrations. As the ensemble integrates five representations of the same GVM, and six representations of the same GCM, we avoid issues related to model genealogy (Knutti et al., 2013) that could lead similar models to bias results of the averaging because of intrinsic lack of independence between the different ensemble members (Bishop and Abramowitz, 2012). We focus our approach on NPP projections under the RCP8.5 scenario of emissions for which more simulations were available (Nishina et al., 2015). Mean annual current NPP and projected changes are summarized in Table 1 and Supplement Fig. S1. We note a large spread in current global NPP simulated by the models from 51.7 to 77.8 Pg C y⁻¹ during 2001–2005, the last 5 years of the historical simulations, as well as ΔNPP in the last 5 years of the projections (2095–
2.2 Benchmark datasets of modern NPP

We use three different estimates of current NPP: (a) an observationally bound terrestrial carbon cycle analysis estimate, (b) an estimate based on up-scaled eddy-covariance CO₂ flux measurements and (c) an estimate based on satellite measurements of absorbed photosynthetically active radiation. To harmonize the approach, we re-gridded all observationally constrained NPP datasets to the lowest dataset resolution (1° × 1°) and confined our analysis to the overlap period 2001–2005 for which benchmark datasets and ISIMIP models were available. Mean annual NPP and variability for each dataset is presented in Figs. S2 and S3.

2.2.1 CARDAMOM retrievals

CARDAMOM produces spatially explicit retrievals of the global terrestrial carbon cycle following a model–data fusion procedure. In each 1° × 1° pixel, the Data Assimilation Linked Ecosystem Carbon version 2 (DALEC2; Bloom and Williams, 2015a; Williams et al., 2005) is driven by ERA-Interim reanalysis climate data (Dee et al., 2011) and burned area from the Global Fire Emission Database version 4 (Giglio et al., 2013). A Bayesian Markov chain Monte Carlo approach is implemented to constrain DALEC2 according to observations of MODIS leaf area index (Myneni et al., 2002), tropical biomass (Saatchi et al., 2011), soil carbon content from the Harmonized World Soil Database (HWSD; FAO, 2012) and a set of ecological and dynamic constraints (Bloom and Williams, 2015a). Through this Bayesian procedure, CARDAMOM provides an explicit estimation of the uncertainty in model parameters, and hence in land-atmosphere carbon fluxes such as NPP from site to global-scale (Bloom et al., 2016; Smallman et al., 2017). However, as not all the other datasets (see Sects. 2.2.2 and 2.2.3) provide a measure of the parametric uncertainty, in this study we rely on CARDAMOM’s highest confidence estimates of a mean annual NPP of 49.9 Pg C yr⁻¹. More details on the framework can be found in the Supplement of Bloom et al. (2016).

2.2.2 FLUXCOM

The FLUXCOM project uses machine-learning methods (Tramontana et al., 2016) to up-scale global datasets from eddy-covariance measurements of CO₂ and energy fluxes from the FLUXNET network (Baldocchi et al., 2001). In a first step, a machine-learning algorithm is used to extract a relationship between local environmental drivers and ecosystem fluxes (Jung et al., 2009). Then, the trained algorithm is used in combination with gridded climate data and remote sensing observations to produce a global estimate of monthly ecosystem fluxes at a 0.5° × 0.5° spatial resolution. In its first dataset, FLUXCOM products relied on a random forest method (Breiman, 2001) but newly available datasets have been produced using additional machine-learning methods (Tramontana et al., 2016; Jung et al., 2017).

Here, we use the average of an ensemble of six FLUXCOM GPP datasets to derive an estimate of annual NPP for 2001–2005. These datasets were created using three machine-learning methods: random forest, artificial neural networks and multivariate regression splines. Each machine-learning method was used to produce two GPP datasets corresponding to two partitioning methods of net ecosystem exchange (see Reichstein et al., 2005 and Lasslop et al., 2009). Then, we used CARDAMOM’s retrievals of carbon use efficiency (Bloom et al., 2016), the ratio of NPP to GPP, to derive a current value of NPP of 52.7 Pg C yr⁻¹ for the first 5 years of the 21st century from the 126.9 Pg C yr⁻¹ FLUXCOM-estimated GPP.

2.2.3 MODIS NPP

The MOD17 MODIS GPP/NPP dataset provides 8-day estimates of GPP and annual NPP at a 1 km spatial resolution since the year 2000. Therefore, GPP is calculated as the product of the amount of absorbed photosynthetically active radiation (estimated from the MOD15 MODIS LAI/FPAR product; Myneni et al., 2002) and a biome-specific radiation use efficiency that is adjusted as a function of air temperature and vapour pressure deficit. Land cover classification is derived from MODIS using the MCD12Q1 product (Friedl et al., 2012) while meteorological data are taken from the National Centers for Environmental Prediction (NCEP)/Department of Energy (DOE) Reanalyses II. Then, annual maintenance respiration is estimated using a temperature-acclimated Q₁₀ relationship (Tjoelker et al., 2001) while growth respiration is assumed to be a fixed fraction of NPP. The MODIS NPP dataset has been used to quantify the impact of droughts (Zhao and Running et al., 2010) and the El Niño–Southern Oscillation on global terrestrial ecosystem productivity (Bastos et al., 2013). We re-gridded the annual NPP data to a 1° × 1° spatial resolution for the reference years 2001–2005 from which we derived a 53.4 Pg C yr⁻¹ mean annual value.

2.3 Reliability ensemble averaging

The REA method was developed to assign coefficients to models in the context of future projections. In addition to using a measure of model performance to reproduce historical conditions, the REA weighting scheme implements a measure of model convergence to penalize models that do not predict the same response to changes (Exbrayat et al., 2013b).
In each $1^\circ \times 1^\circ$ pixel, each model projection $i$ of the 30 GVM–GCM ensemble is assigned a reliability factor $R_i$ that is calculated such as

$$R_i = R_{B,i} \times R_{D,i} = \left(\frac{\varepsilon}{|B_i|}\right) \times \left(\frac{\varepsilon}{|D_i|}\right).$$

(1)

where $\varepsilon$ represents the variability in observations expressed as the difference between the largest and smallest values of annual NPP in each pixel (Fig. S3; Giorgi and Mearns, 2002), while $B_i$ and $D_i$ correspond to a measure of model $i$’s performance and convergence, respectively. The performance coefficient $R_{B,i}$ ranges from 0, for a poorly performing model, to 1 if the absolute value of $B_i$ is smaller than the variability $\varepsilon$. Similarly, the convergence coefficient $R_{D,i}$ ranges from 0 for outlier projections to 1 if the absolute value of $D_i$, the difference between the projection and the REA mean, is smaller than $\varepsilon$. As a result, the final model weight $R_i$ also considers values ranging from 0 (excluded) to 1 (included). We produce three REA estimates based on CARDAMOM, FLUX-COM and MODIS NPP, further referred to as $\text{REA}_C$, $\text{REA}_F$ and $\text{REA}_M$, respectively. For each REA application, terms $\varepsilon$, $B_i$, $D_i$, and hence $R_i$ (Eq. 1) are recalculated based on the particular observational dataset to produce three independent sets of model coefficients.

Here, we apply the REA method to the ensemble of 30 ISIMIP simulations of 21st century $\Delta$NPP under the RCP8.5 emission scenario. We first re-gridded the ISIMIP data using the “remapcon” function of the Climate Data Operators version 1.6.9 to match the $1^\circ \times 1^\circ$ spatial resolution of the observationally constrained datasets (see Sect. 2.2) and performed the procedure in each land pixel to create maps of REA averages. We then apply the REA method three times ($\text{REA}_C$, $\text{REA}_F$ and $\text{REA}_M$) to evaluate the current performance of the ISIMIP simulations of NPP.

For each 30 simulations of the ISIMIP ensemble, we calculated $B_i$ in each pixel as

$$B_i = \text{NPP}_i - \text{NPP}_{\text{obs}},$$

(2)

where $\text{NPP}_i$ is the mean annual NPP predicted by model $i$ during the last 5 years of the historical simulations and $\text{NPP}_{\text{obs}}$ corresponds to either of the observational datasets’ mean annual NPP. Then for each model the value of $D_i$ was calculated in each pixel as the difference between the change predicted by model $i$ and the REA average as

$$D_i = \Delta \text{NPP}_i - \frac{\sum_{i=1}^{N} R_i \cdot \Delta \text{NPP}_i}{\sum_{i=1}^{N} R_i},$$

(3)

where $\Delta \text{NPP}_i$ is the change in mean NPP in the last 5 years of the RCP8.5 simulation (2095–2099) compared to the last five years of the historical simulations (2001–2005) predicted by the ensemble member $i$, and $N$ is the total number of ensemble members. The REA average is not known beforehand and weights $R_{D,i}$ are calculated iteratively (Giorgi and Mearns, 2002).

The uncertainty around the REA average change is calculated as the weighted root-mean square difference (RMSD) calculated following

$$\text{RMSD} = \left(\frac{\sum_{i=1}^{N} R_i \cdot (\Delta \text{NPP}_i - \Delta \text{NPP}_{\text{REA}})^2}{\sum_{i=1}^{N} R_i}\right)^{1/2},$$

(4)

where $\Delta \text{NPP}_{\text{REA}}$ is the REA average change. Assuming that the error distribution is somewhere between uniform and Gaussian, the 60–70% confidence interval of the REA is represented by $\Delta \text{NPP}_{\text{REA}} \pm \text{RMSD}$ (Giorgi and Mearns, 2002).

Giorgi and Mearns (2002) further introduced a quantitative measure of the collective model reliability $\rho$, based on $R_i$, where

$$\rho = \frac{\sum_{i=1}^{N} R_i^2}{\sum_{i=1}^{N} R_i},$$

(5)

which will vary pixel-wise based on each model’s performance with respect to the mean and variability represented.
in each observational dataset as well as the convergence to the REA average. The reliability measure \( \rho \) can be further decomposed in \( \rho_B \) and \( \rho_D \), as

\[
\rho_B = \frac{\sum_{i=1}^{N} R_{B,i}}{N},
\]

\[
\rho_D = \frac{\sum_{i=1}^{N} R_{D,i}}{N},
\]

where \( \rho_B \) and \( \rho_D \) correspond to the ensemble reliability with respect to model biases and model convergence, respectively. \( \rho, \rho_B \) and \( \rho_D \) all take values ranging from 0, indicating a lack of agreement between models, to 1, indicating a consensus between models in terms of performance and projected changes.

### 3 Results

The REA method yields a global increase in NPP of 24.6 ± 8.5 Pg C yr\(^{-1}\) (REA average ± RMSD) using CARDAMOM in REAC, 24.8 ± 9.5 Pg C yr\(^{-1}\) using FLUXCOM in REAF and 25.0 ± 14.4 Pg C yr\(^{-1}\) using MODIS NPP inREAM. As the ISIMIP ensemble mean indicated a ΔNPP of 24.2 Pg C yr\(^{-1}\), these results represent a ~2 % increase in the mean for both REAC and REAF and 3 % forREAM. The pixel-wise 1 SD uncertainty in the ISIMIP ensemble was 26.3 Pg C yr\(^{-1}\) and the REA results indicate strong reduction of 68 % for REAC, 64 % for REAF and 45 % forREAM. These results further indicate that in all three cases, the REA method reduces the uncertainty of the ensemble spread toward an agreement on a future increase in the global land carbon uptake.

Zonal means (Fig. 1) indicate that the ISIMIP ensemble mean and all three REAC, REAF and REAM averages estimate an increase in NPP across all latitudes. All three REA averages predict a weaker increase in NPP at high latitudes of the Northern Hemisphere and Southern Hemisphere. They also agree on a stronger increase in NPP than the ISIMIP ensemble mean for tropical regions between 15° S and 10° N but also between 20 and 25° N and temperate regions around 55 to 65° N. REAC and REAF indicate a weaker increase in NPP than ISIMIP around 20° S while the REAM average is similar to the ISIMIP ensemble mean in these regions. The uncertainty around each of the REA averages is smaller than the uncertainty around the ISIMIP ensemble mean across all latitudinal zones. Furthermore, while the very large uncertainty around the ISIMIP ensemble mean does not provide confidence on the sign of ΔNPP across most regions, the uncertainty around all three REA averages is constrained toward an increase in NPP across all regions, except around 20° S.

The spatial distribution of the ISIMIP ensemble mean ΔNPP contrasts with that of the three REA averages with noticeable differences across all regions of the globe (Fig. 2).

All three REA averages predict a weaker increase in NPP than the ISIMIP ensemble in Canada and Scandinavia, while they predict a stronger increase in NPP in Eurasia. Similarly, all three REA averages predict a stronger increase in NPP than the ISIMIP ensemble in the tropical rainforests of South America, Africa and Southeast Asia. The REA averages agree on a weaker ΔNPP in semi-arid regions of the Sahel, southern Africa, Australia and the Tibetan Plateau. Overall, the REAC, REAF and REAM values exhibit broadly similar patterns in the spatial distribution of ΔNPP differences with the ISIMIP ensemble mean that are confirmed by Pearson’s correlation coefficient of 0.63, between REAC and REAF, 0.61 between REAC and REAM, and 0.68 between REAF and REAM.

The uncertainty in ΔNPP is reduced across most regions of the globe for REAC, REAF and REAM (Figs. 1 and 3). This reduction of uncertainty leads to a confidence on the sign estimation of ΔNPP in 86, 80 and 76 % of all the land pixels for REAC, REAF and REAM, respectively, compared with 43 % for the ISIMIP ensemble. The average reduction in uncertainty is large in regions north of 40° N (Fig. 1), mostly...
corresponding to a reduction in uncertainty in boreal Eurasia (Fig. 3) that provides better confidence in an increase in NPP (Fig. 2). We note that the uncertainty in the REA_M remains similar to the uncertainty around the ISIMIP ensemble mean for large portions of the Southern Hemisphere such as southern Africa. However, all three REA_C, REA_F and REA_M cannot provide confidence on the sign of \( \Delta\text{NPP} \) for southern Africa and Australia.

The zonal means of the mean values of the three coefficients \( R_i \), \( R_{B,i} \) and \( R_{D,i} \) (Fig. 4) show that a MODIS-based REA_M yields larger values of all coefficients compared to REA_C and REA_F. We note strong inter-model similarities in the spatial distribution of model weights (\( R_i \); Fig. 4a–c), biases (\( R_{B,i} \); Fig. 4d–f) and convergence of the projected \( \Delta\text{NPP} \) (\( R_{D,i} \); Fig. 4g–i). Only the HYBRID models are almost systematically assigned lower \( R_i \) values as a result of lower values for both \( R_{B,i} \) (i.e. a larger bias than the other models) and \( R_{D,i} \) (i.e. a divergence in projected \( \Delta\text{NPP} \)). This is especially obvious in boreal regions north of 60\(^\circ\) N where HYBRID is assigned values significantly closer to 0 in REA_C, REA_F and REA_M.

The collective model reliability measure \( \rho \) provides a quantification of the spread of model weights determined through the REA method (Fig. 5). Regions where \( \rho \) is close to 1 show a strong consensus between models on the current NPP and on 21st century \( \Delta\text{NPP} \). There are large differences in \( \rho \) depending on the NPP observational datasets used to constrain the REA (Fig. 5). Indeed, while the average value of \( \rho \) is 0.29 for REA_C and 0.32 for REA_F, it is 0.62 for REA_M. REA_C and REA_F yield very low values of \( \rho \) in boreal regions (Fig. 5) while REA_M leads to values of \( \rho \) close to 1 in many regions south of 60\(^\circ\) S. The measure of reliability \( \rho \) can be further decomposed into two components \( \rho_B \) and \( \rho_D \) (Fig. 5d–i, Eqs. 6 and 7). Results indicate that \( \rho_D \) is consistently greater than \( \rho_B \) for REA_C, REA_F and REA_M. This result means that model convergence in the simulation of \( \Delta\text{NPP} \) is greater than the model ability in reproducing current NPP. In other words, the model performance evaluated against the three current NPP datasets contributes the most to decreasing the ensemble reliability \( \rho \). Values of \( \rho_B \) are lower than 0.10 in boreal regions for REA_C and REA_F, indicating that model bias is greater than the variability in NPP \( \varepsilon \) estimated from the CARDAMOM retrievals and the FLUXCOM-based NPP by a factor of 10. Conversely, regions where \( \rho_B \) is close to 1 for REA_M indicate that the variability in the MODIS NPP observations is larger than model biases.


Figure 4. Zonal mean $R_i$, $R_{R,i}$ and $R_{D,i}$ (row-wise) in each REA$_C$, REA$_F$ and REA$_M$ (column-wise). Each line represents the average value obtained across the five simulations of each GVM.

Figure 5. Collective model reliability $\rho$, model performance $\rho_B$ and model convergence $\rho_D$ (row-wise) for REA$_C$, REA$_F$ and REA$_M$ (column-wise).

4 Discussion

The globally integrated values of the REA average change (24.6 to 25.0 Pg C y$^{-1}$) and the ISIMIP ensemble mean (24.2 Pg C y$^{-1}$) are similar. This is in agreement with a previous multi-model approach that only found a 0.01 Pg C y$^{-1}$ difference in the historical mean annual net ecosystem exchange between a simple mean and a weighted average based on model performance (Schwalm et al., 2015). However, in contrast to this previous study, we find that in REA$_C$, REA$_F$ and REA$_M$ a large spatial variability in grid cell differences (Fig. 2) that compensate for each other to yield a relatively small global difference with the ISIMIP ensemble mean. The three REA averages indicate a stronger positive $\Delta$NPP than the ISIMIP ensemble mean for boreal Eurasia and tropical rainforests (Figs. 1 and 2) and a weaker but still positive $\Delta$NPP in northern Canada and semi-arid regions like the Sahel, the Tibetan Plateau, southern Africa and Australia.
The reduction in uncertainty arising from the REA method helps put a greater confidence in a sustained CO$_2$ fertilization effect throughout the 21st century, although these results may be influenced by model-wise differences in process representation. In both the ISIMIP ensemble mean and the three REA averages, NPP increases at high latitudes where nitrogen (N) limitation on NPP dominates (Zhang et al., 2011; Exbrayat et al., 2013a) but is only represented in the HYBRID and SDGVM models (Table 1; Nishina et al., 2014). The increase in NPP in these N-limited regions is in contrast with observations in Free-Air CO$_2$ Enrichment experiments that indicate a quick weakening of the CO$_2$ fertilization effect as soil N stores deplete (Norby et al., 2010). Models which integrate coupled C–N cycles generally predict the historical land carbon sink in good agreement with estimates from the Global Carbon Project (Huntzinger et al., 2017) and project a decrease in NPP throughout the 21st century (Thornton et al., 2002; Goll et al., 2012; Zhang et al., 2013; Wieder et al., 2015).

Similarly, recent observations have concluded a total absence of CO$_2$ fertilization effect under phosphorus-limited conditions (Ellsworth et al., 2017) which dominate in the tropics and lead to an additional reduction of NPP in model projections (Goll et al., 2012; Zhang et al., 2013; Wieder et al., 2015). Here, only the HYBRID and SDGVM models integrate the representation of N limitations on NPP (Nishina et al., 2014) and none of them represent phosphorous limitations. HYBRID is also the only model to predict a possible decrease in global NPP throughout the 21st century (Table 1 and Friend et al., 2014) because of a reduction in global NPP at high latitudes and in tropical rainforests (Supplement Fig. S1). Thus, HYBRID is assigned low $R_{D,i}$ weights in these regions (Fig. 4g–i and Supplement Figs. S4–12) and cannot influence the REA average and the calculation of its uncertainty (Eq. 4) despite integrating a more detailed representation of ecosystem processes. However, HYBRID also exhibits stronger differences from the observational datasets than the other models, especially at high latitudes (Fig. 4d–f), which may indicate a strong sensitivity to N limitations. Nevertheless, we note that all models’ performances tend to decrease in regions north of 60° N where their ΔNPP projections also diverge (Figs. 4d–f and 5d–f).

Overall, the promising REA results should be used carefully as they cannot correct for the omission of key processes by a large fraction of the ensemble members. Like in previous multi-model averaging studies focused on the carbon cycle (e.g. Schwalm et al., 2015; Lovenduski and Bonan, 2017) or climate (Krishnamurti et al., 1999; Giorgi and Mearns, 2002), we used available simulations in a post-processing procedure. We note, however, that the ratio of two out of six models including carbon–nutrient interactions in the ISIMIP ensemble is commensurate to other model inter-comparison projects: 3 out of 10 CMIP5 models (Exbrayat et al., 2014) or 2 out of 8 models in the new ISIMIP experiments presented by Chen et al. (2017). There is also considerable debate on how good large-scale NPP observational products are (Smith et al., 2015; de Kauwe et al., 2016), a problem that we address by performing the REA approach three times.

In all three REA$_C$, REA$_F$ and REA$_M$ cases, the global uncertainty around the REA average is reduced compared to the uncertainty within the ISIMIP ensemble, which provides a higher degree of confidence in the resilience of the global CO$_2$ fertilization effect to warming. The reduction in uncertainty and the gain in confidence on the sign of ΔNPP, is especially obvious in boreal regions for all three REA cases (Fig. 3). Conversely, uncertainties on the sign of ΔNPP remain large for all REA cases in semi-arid regions of southern Africa and Australia. It is a non-trivial result as the response of these ecosystems to climate events like El Niño and La Niña drives the inter-annual variability and the trend of the global terrestrial carbon sink (Bastos et al., 2013; Poulter et al., 2014; Ahlström et al., 2015), while projections indicate a gain of forest ecosystems over savannahs in the future (Moncrieff et al., 2016).

Because of the way the REA method assigns coefficients to ensemble members with respect to the annual variability in the data $\varepsilon$ (Eq. 1), the final REA average and uncertainty are conditional on the variability represented in the current estimate of NPP. Figure 5, sections a–c show that the reliability of the ensemble measured by $\rho$ varies depending on which observational dataset is used, although generally lower values of $\rho_B$ and $\rho_D$ at high latitudes indicate that models disagree on the current NPP and future ΔNPP in these regions. Furthermore, high values of $\rho$ for REA$_M$ indicate a larger variability $\varepsilon$ in the MODIS dataset compared to CARDamom and the FLUXCOM-based NPP data (Fig. S3). This larger variability leads to more models being given a weight close to 1 in the averaging scheme because the variability is larger than their bias (Fig. 5f) or the predicted change (Fig. 5i). Conversely, the relatively smaller variability in CARDamom retrievals leads more models to be weighted poorly according to both their performance (Fig. 5d) and their convergence with other models (Fig. 5g). The variability $\varepsilon$ influences the final uncertainty and as a result the REA$_C$ has a smaller uncertainty because it is more penalizing on models, and vice versa with MODIS NPP.

5 Conclusions

We applied the REA method on a pixel-by-pixel basis to an ensemble of 30 simulations of historical and 21st century NPP from the ISIMIP project. Our results indicate that using either CARDamom retrievals, a FLUXCOM based estimate of current NPP or data from MODIS to constrain the REA scheme helps by halving the uncertainty in 21st century global ΔNPP. This process leads to a higher confidence in a sustained CO$_2$ fertilization effect. We nevertheless note that a large uncertainty remains in semi-arid regions that is mostly attributable to differences in process representation.
in global vegetation models. Furthermore, most models used here do not account for N limitations on NPP and this may have altered the outcome of the convergence coefficient used in REA.

Data availability. CARDAMOM output used in this study is available from Bloom and Williams (2015b) from the University of Edinburgh’s DataShare service at https://doi.org/10.7488/ds/316. FLUXCOM GPP estimates used in this study are available from the data portal of the Max Planck Institute for Biogeochemistry at https://doi.org/10.17871/FLUXCOM_RS_METEO_CRUNCEPv6_1980_2013_v1 (Jung, 2016). The MODIS MOD17 NPP product is available from the website of the Numerical Terradynamic Simulation Group at the University of Montana (http://files.ntsg.umt.edu/data/NTSG_Products/MOD17/GeoTIFF/MOD17A3/; Zhao et al., 2005; Zhao and Running, 2010). REA output (Exbrayat and Williams, 2018) is available from the University of Edinburgh’s DataShare service at https://doi.org/10.7488/ds/2304.

The Supplement related to this article is available online at https://doi.org/10.5194/esd-9-153-2018-supplement.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. This work was supported by the Natural Environment Research Council through the National Centre for Earth Observation. Part of this work was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration: Anthony Bloom was supported by NASA Earth Sciences grant (no. NNI16ZDA001N-IDS). Pete Falloon was supported by the Joint UK DECC/Defra MetOffice Hadley Centre Climate Programme (GA01101). For their roles in producing, coordinating, and making available the ISIMIP model output, we acknowledge the modelling groups and the ISIMIP coordination team.

Edited by: Somnath Baidya Roy
Reviewed by: two anonymous referees

References


