7 Global exemplar

An exemplar which demonstrated the potential value of the EVO approach to a global-scale science and policy question was developed. It addressed the sensitivity of carbon, water and energy fluxes to current uncertainty in climate projections from the 22 GCMs used in the 4th IPCC Assessment. The IT problems being tackled related to:

- increased accessibility to a large climate change impact model by the terrestrial climate change impacts community;
- increased speed of computing;
- elasticity and auto-scalability.

This global exemplar was intended to develop a third approach to that being developed for the local and national exemplars thus providing an independent exploration of possible futures for EVO. It demonstrates the potential benefits of exploiting developments in other scientific communities (i.e. bioinformatics) and addresses particular challenges in cloud-enabling large, complex models and managing the associated costs.

7.1 Science background

General Circulation Models (GCMs) have been the main tool to address issues of climate change. The complexity of these come with two penalties. First, even with the fastest computers, it remains only possible to make a few very selected simulations to investigate a given problem, thus restricting the number of global emissions scenarios that can be tested, and accessibility by the impacts community. Second, there remains a great deal of uncertainty surrounding climate projections. This is in part due to their complexity which needs to be explored and communicated by the impacts community. This is a particular concern for modelling land surface response, given its dual role as a fundamental component of the climate-carbon cycle system, and its direct link to food and water security. For this reason, an intermediate methodology was developed which

i. emulates climate models, and
ii. captures the full complexity of land surface models.

This system provides an ability to assess terrestrial ecosystem changes that might be expected in a warming world and of great value to the climate change impacts community. More specifically the "pattern-scaling" methodology is utilised, with the concept being that changes in surface climate, for each month and at each geographical position, are linearly proportional to the amount of average warming over land. In addition, a global thermal energy balance model was built which maps from different pathways in atmospheric greenhouse concentrations to mean global temperature increase over land, which allows scaling of the patterns. This structure is called the "GCM analogue model", (Huntingford and Cox, 2000) where "analogue" means replication (of the GCM). The analogue model system was developed for Version 3 of the Met Office Hadley Centre GCM (HadCM3; Gordon et al., 2000) linked to the Met Office Surface Exchange Scheme (MOSES; Cox et al., 1998) and the Dynamic Global Vegetation Model (DGVM), called TRIFFID (e.g. Clark et al., 2011; Huntingford et al., 2000). The full scheme was called IMOGEN ("Integrated Model Of Global Effects of climatic aNomalies") (Huntingford et al., 2010). This system operates relatively quickly, and a full transient simulation between modelled years of 1860 (for pre-industrial times) and year 2100, and for all "land points", can be undertaken in around 24 hours on the latest processors. However, to sample full climate uncertainty, this work needed to be repeated for climate patterns derived from the full 22 GCMs that contributed to the 4th IPCC assessment and using the enhanced land-atmosphere model which will contribute to the next Unified Model of the UK, namely JULES.

The science aims of the EVOp global exemplar were:

- To deliver a version of IMOGEN-JULES framework with inputs/outputs that are in a format that could be easily operated from a web interface, and driving its operation in a compute "cloud" environment.
- To define a set of scalable climate patterns representing the 22 GCM simulations that contributed to the 4th IPCC assessment, all on a common grid of 2.5°x3.75° spatial resolutions.

- To develop a working prototype applying this IMOGEN-JULES framework, within the cloud, to explore the uncertainty bounds of current GCM predictions on change in soil carbon stocks in soil (but with outputs for a wide range of other carbon and water fluxes).

7.2 Steps taken in implementation

7.2.1 Step 1

The first step as in all exemplars concerned the identification of potential end-users to ensure a clear focus to the work carried out within the EVO Pilot. A storyboard was created (see online Annex) to ensure a logical structure to the problem would be completed which addressed a specific need. Two end-users were chosen;

i. A NERC climate change impact scientist interested in the impact of climate projects on carbon fluxes to and from the atmosphere and the implications of uncertainty in climate projections spatially at a global scale
ii. A government department (e.g. DECC) interested in exploring methods of communicating uncertainty in climate projections.

The navigation route they take to explore these questions were identified as being essentially the same.

7.2.2 Step 2

Translation of the storyboard navigation path into EVOp is currently accessed though a "Global" button within "Explore by Location". Eventually, other pathways from the portal home page would be activated (a “Climate Change” option within “Explore by Topic”, “Show me the Data” and “Show me the Resources”).

This takes the user to a welcome page which provides some information on the model (Figure 7.2):

7.2.3 Step 3

The end-users define an emission scenario through a web portal (Figure 7.3). Options include:

- Ecosystem Variables
  - ❑ Carbon: Soil Carbon, Vegetation Carbon, Net Primary Productivity
  - ❑ Water: Evaporation, Soil moisture, Runoff
  - ❑ Vegetation: Fractional cover of plant functional types (PFTs)
- Dates for model run
  - ❑ Absolute 2020, 2050 and/or 2100
  - ❑ Relative to pre-industrial 1860
- Output type
  - ❑ Data
  - ❑ Maps

If the emission scenario has already been run before, the user would be able to access maps held in the model library at no charge. Due to the large amount of data, data is not stored (Figure 7.4).

Currently, years are restricted to 2020, 2050 and 2100 due to resource limitations. A full implementation of the model would enable any year to be requested. After 3-5 days, an email is sent with the information requested demonstrating ~85-90% reduction in computing time which would be required without an EVO cloud solution (each one of the 22 GCM analogue models normally requires 3 days computing time = 66 days).

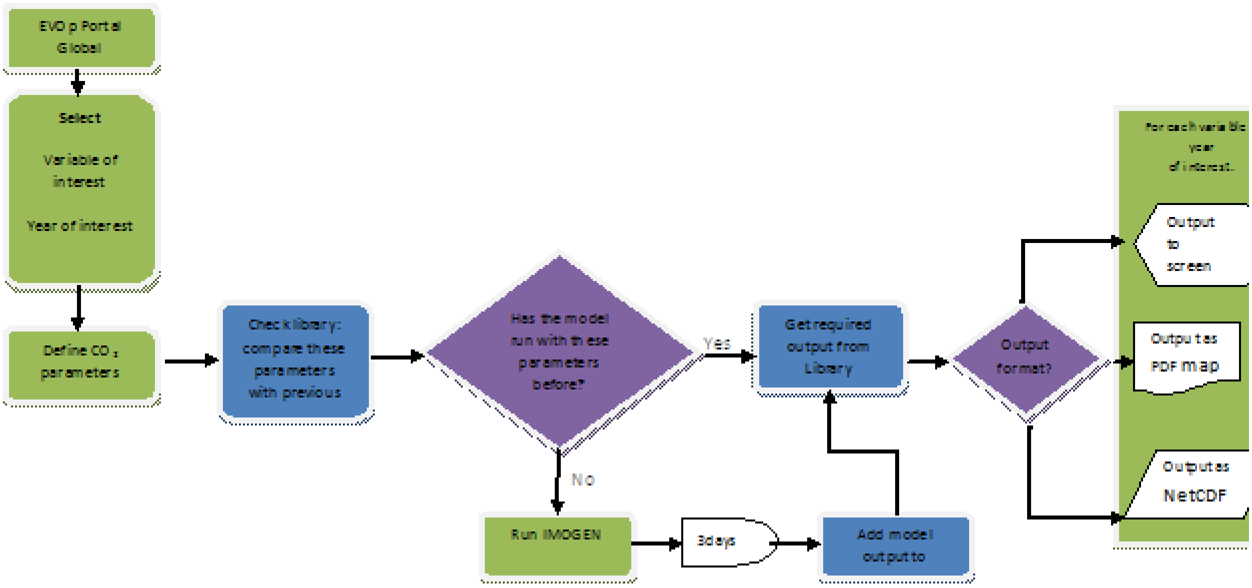


Figure 7.1 Schematic of accessing IMOGEN runs library or cloud resources using global exemplar. (See folder for original flowchart).

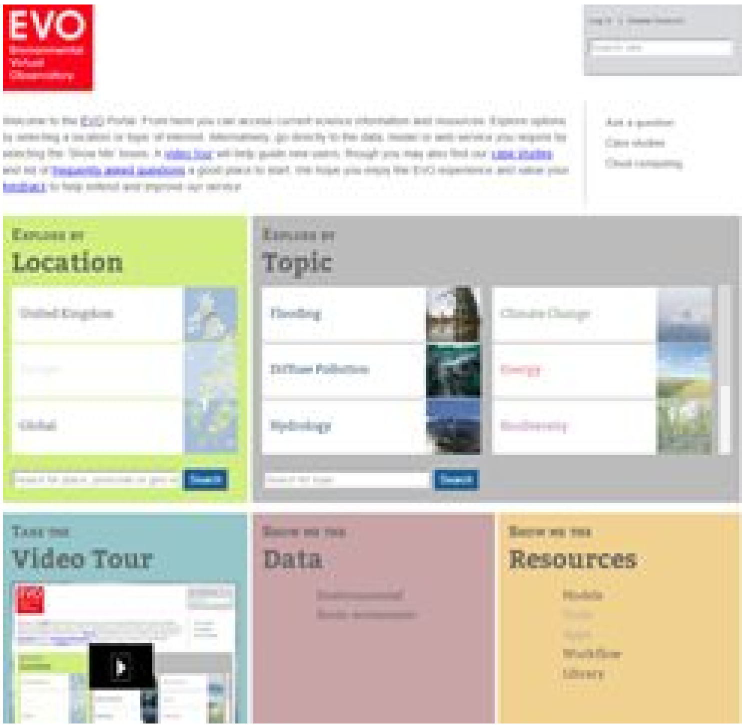
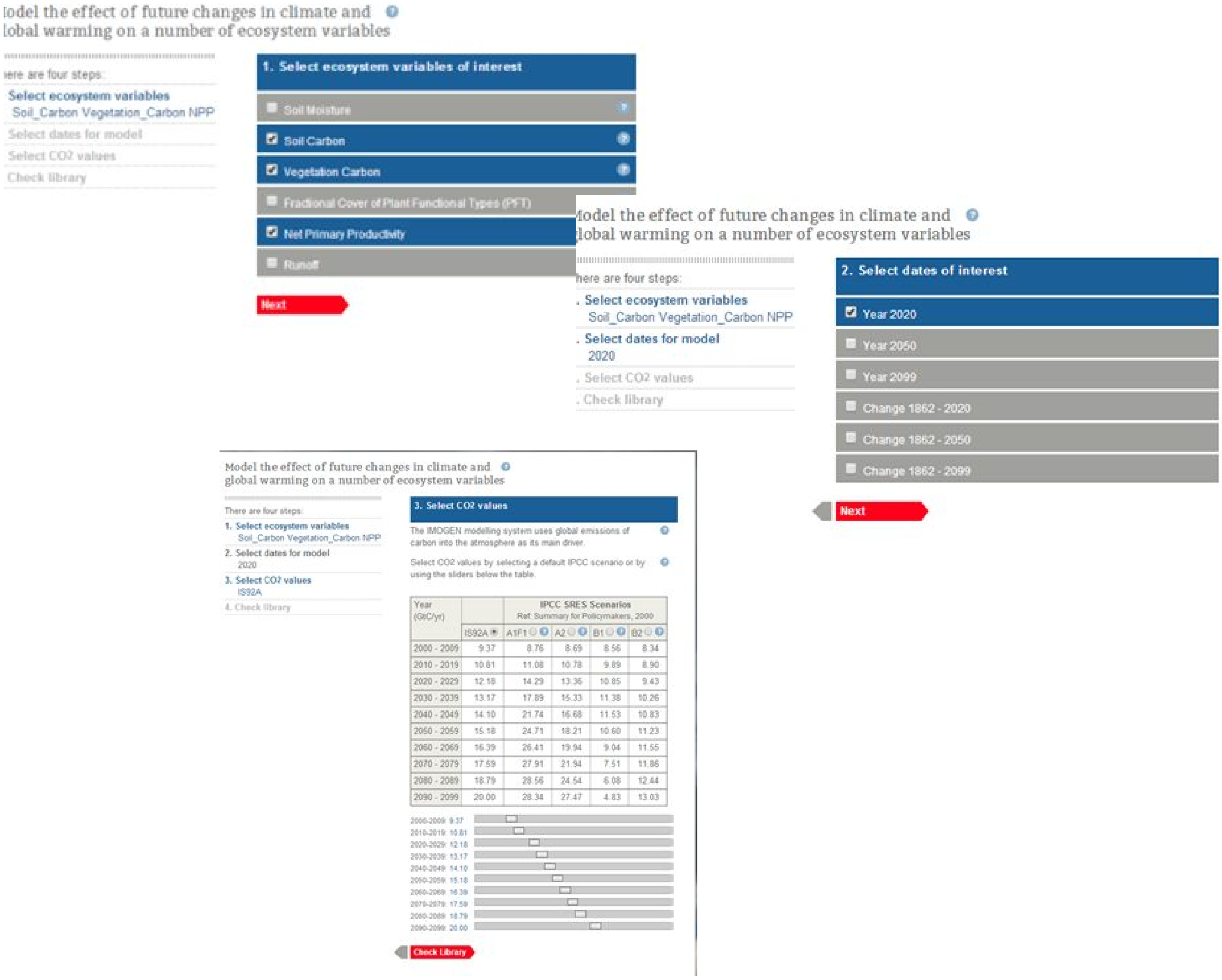


Figure 7.2 Portal welcome page.



Year (GtC/yr)	IS92A	A1F1	A2	B1	B2
2000 - 2009	9.37	8.76	8.69	8.56	8.34
2010 - 2019	10.81	11.08	10.78	9.89	8.90
2020 - 2029	12.18	14.29	13.36	10.85	9.43
2030 - 2039	13.17	17.89	15.33	11.38	10.26
2040 - 2049	14.10	21.74	16.68	11.53	10.83
2050 - 2059	15.18	24.71	18.21	10.60	11.23
2060 - 2069	16.39	26.41	19.94	9.04	11.55
2070 - 2079	17.59	27.91	21.94	7.51	11.86
2080 - 2089	18.79	28.56	24.54	6.08	12.44
2090 - 2099	20.00	28.34	27.47	4.83	13.03

Figure 7.3 Select options.

7.2.4 Step 4

Unlike the other EVO exemplars, for this exemplar the emission scenarios to be selected by the users sit on an Amazon EC2 compute node initialised via the 'Biocloud Central' web server. The benefits of this are that the amount of continual logging steps and authentication steps are significantly reduced relative to stepping through the Amazon EC2 console.

This significant reduction in computing time was made possible through the implementation of Steps 5 -8 .

7.2.5 Step 5

Developing a set of scalable climate patterns representing the 22 GCM simulations that contributed to the 4th IPCC assessment, all on a common grid of 2.5°x3.75° spatial resolutions. This was a major piece of work as all GCMs do not use a common grid and there are many complexities e.g. resolving issues relating to coastal grid squares.

7.2.6 Step 6

These patterns together with JULES and IMOGEN modelling systems were ported to a custom virtual image, CBL-Imogen, based on the CloudBioLinux virtual image.

The Amazon EC2 platform was selected as the biggest supporter of open source software. There are costs associated with the use of the Amazon cloud but the computer usage would have exceeded those provided for academic use and there is a cost verse time trade-off. Benefits relative to the use of HPC is greater accessibility and flexibility.

CloudBioLinux was used to enable the efficient utilisation of the Amazon cloud resources and exploit the extensive experience from the bioinformatics community using this cloud-managing interface. CloudBioLinux incorporates CloudMan, allowing it to be deployed as clusters of compute nodes. The benefits are an accessible and reproducible cloud environment that enables decentralisation of services and realises a scalable model. A fully functional computer cluster is created on demand that can be scaled depending on the requirements defined by the user (fast but expensive or slow and cheaper). Computers are run in parallel thus increasing speed. Once the model runs are complete, the cluster created terminates and the instances are stopped automatically.

All code is open-course and available in github :

- Imogen Infrastructure Portal: https://github.com/afgane/imogen
- GCM analogue portal: https: //github.com/afgane/ghem
- CloudBioLinux: https: //github.com/chapmanb/cloudbiolinux

7.2.7 Step 7

Unlike the other EVO exemplars, a free and open source post-processing tool called GrADS to create maps from model output was used in the cloud. This is important for cloud computing as licensing may restrict this kind of deployment. This is also a major step for the climate change modelling scientists who have traditionally used expensive proprietary software approaches. Unfortunately, resources were not available to fully implement this in a dynamic way but the principal has been demonstrated.

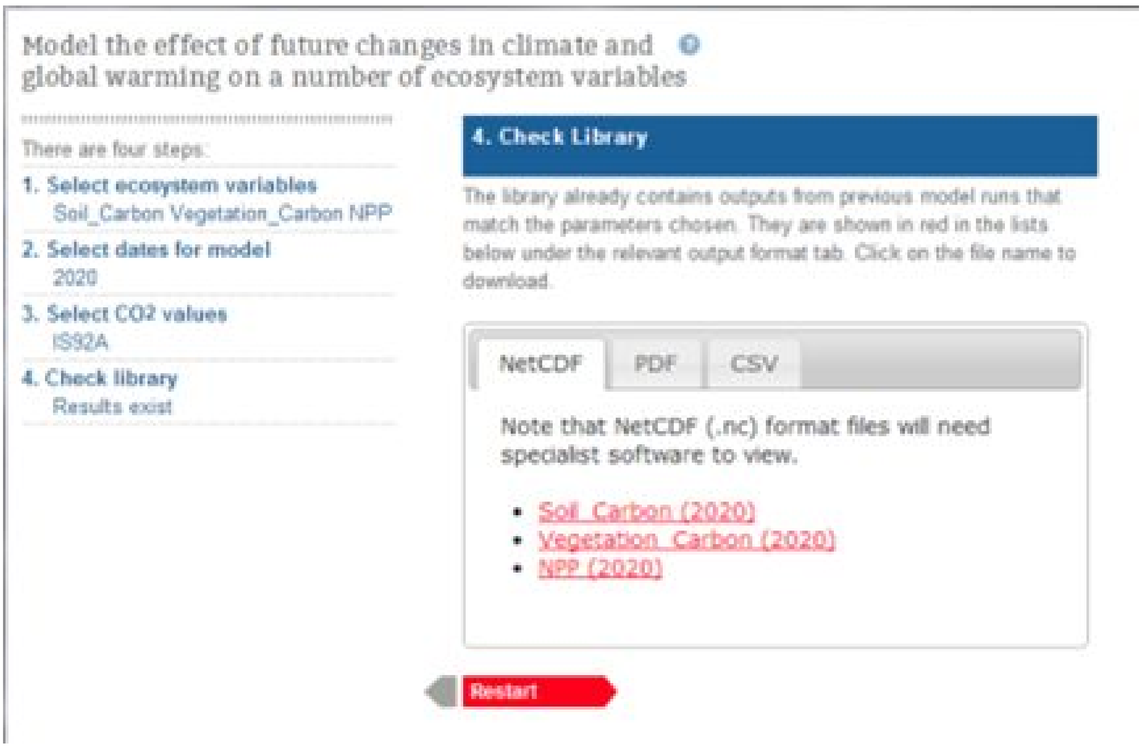


Figure 7.4 Access data (for demonstration purposes data is currently provided from the library).

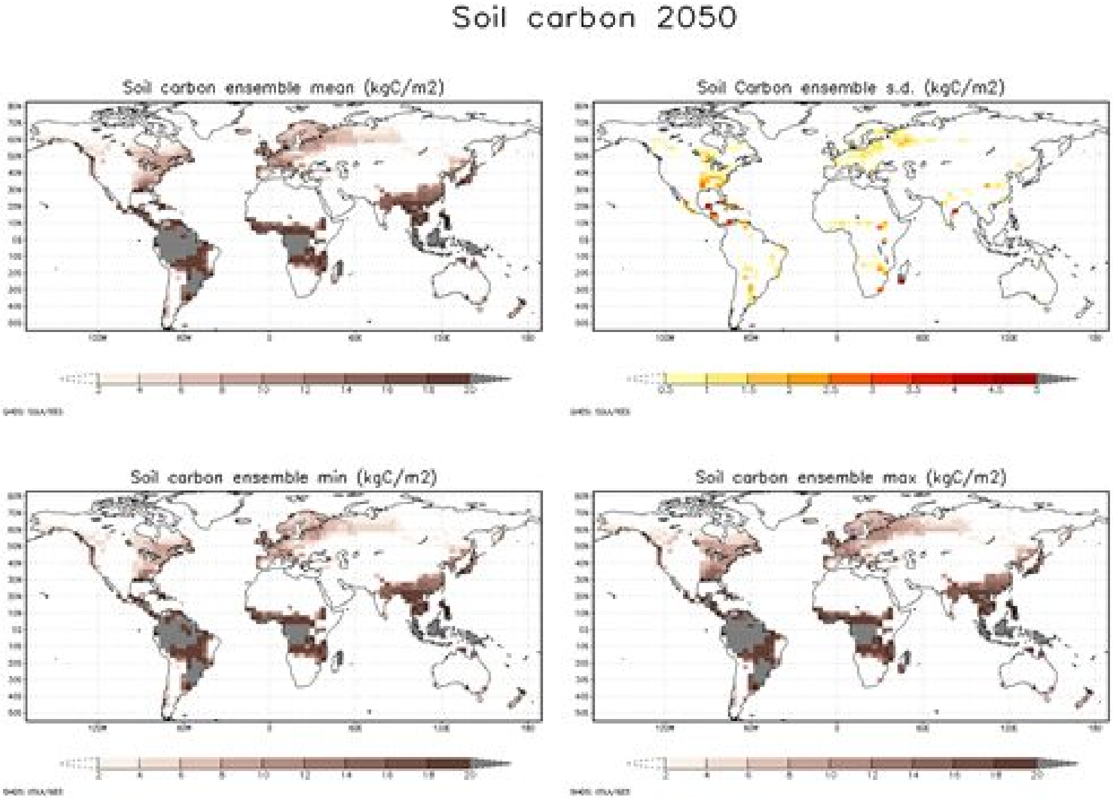


Figure 7.5 Mean soil carbon estimate (kgC/m2), the standard deviation from the ensemble of the 22 GCM analogue models, the minimum and the maximum. Maps indicate the impact of uncertainty in climate projections for soil carbon across the Amazon, Central and West Africa and SE Asia.

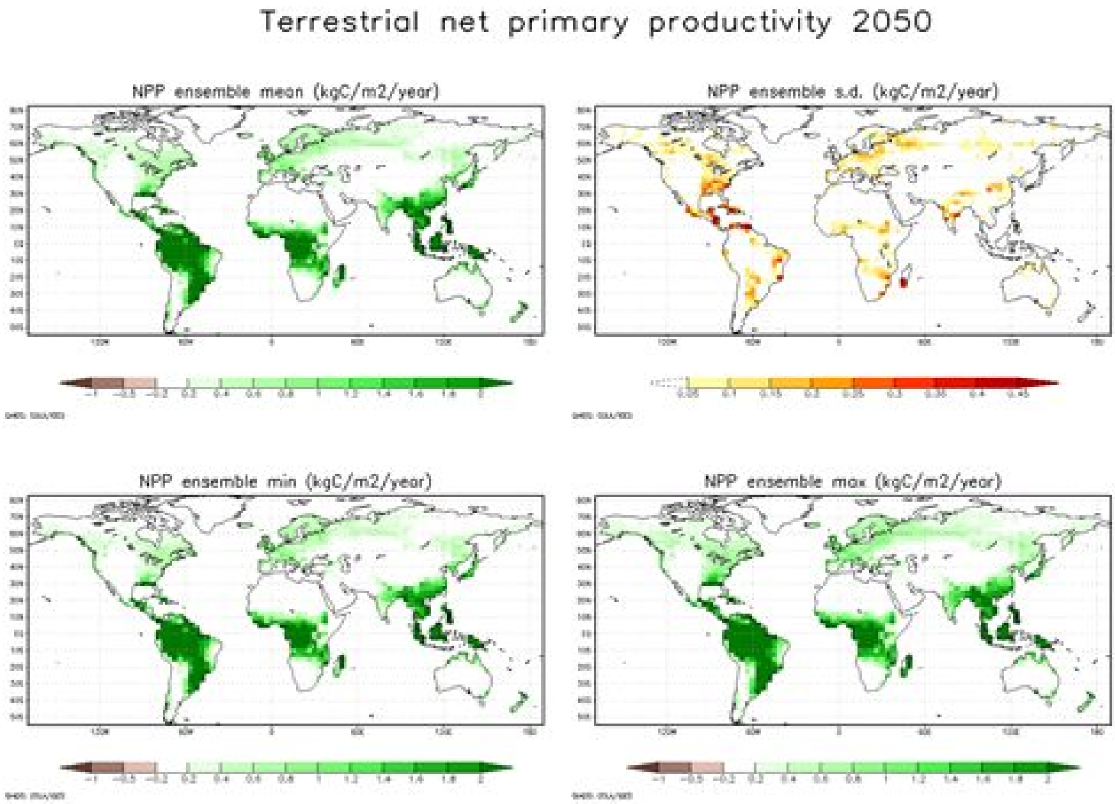


Figure 7.6 Mean net primary productivity (kgC/m2/year), the standard deviation from the ensemble of the 22 GCM analogue models, the minimum and the maximum. Again, maps indicate the impact of uncertainty in climate projections lie primarily in the Amazon, Central and West Africa and SE Asia.

7.2.8 Step 8

A series of output maps for emailing to users were designed. This is a cost-effective way of delivering results as maps and parameters are saved, but not the data. This is because storing large datasets in cloud and direct download off the cloud incur an ongoing cost but dispatching over e-mail is free (Figure 7.5).

7.3 Unique selling points of the global exemplar

The following aspects have been tested / explored in the EVOp Global exemplar:

- Cloud-enabling of a large, complicated model to increase accessibility and flexibility in use of computer resources with major reductions in run time from 66 days to 3 days.
- Demonstration of value of use of more commercial cloud providers. Usage would have exceeded public/academic resources and greater flexibility and accessibility than HPC solution.
- Cloud-based web portal using a solution which minimised authentication and verification needs, being the closest to our ideal user scenario.
- Cloud-based post-processing, non-proprietary map visualisation tool.
- Autoscable use of computing resources for cost efficient use of computing time.
- The benefits of using Open source solutions with all code developed for the project available on-line..

7.4 Future funding / next steps.

Funding to cloud-enable the JULES model under the NERC Big Data initiative has been secured. This builds on the interest raised through the success of this EVO global exemplar. Once completed, this new project will provide greater accessibility to this important model beyond its primary land-atmosphere community.

7.5 References

Clark, D.B. et al., 2011. The Joint UK Land Environment Simulator (JULES), model description - Part 2: Carbon fluxes and vegetation dynamics. Geoscientific Model Development, 4(3): 701-722.

Cox, P.M., Huntingford, C. and Harding, R.J., 1998. A canopy conductance and photosynthesis model for use in a GCM land surface scheme. Journal of Hydrology, 212(1-4): 79-94.

Gordon, C. et al., 2000. The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments. Climate Dynamics, 16(2-3): 147-168.

Huntingford, C. et al., 2010. IMOGEN: an intermediate complexity model to evaluate terrestrial impacts of a changing climate. Geoscientific Model Development, 3(2): 679-687.

Huntingford, C. and Cox, P.M., 2000. An analogue model to derive additional climate change scenarios from existing GCM simulations. Climate Dynamics, 16(8): 575-586.

Huntingford, C., Cox, P.M. and Lenton, T.M., 2000. Contrasting responses of a simple terrestrial ecosystem model to global change. Ecological Modelling, 134(1): 41-58.