In this section we describe the main results from the Carbon Environmental Service (CARBONES) project. The main objective of CARBONES was to provide a state of the art multi-decadal reanalysis of the global carbon cycle and a demonstrator for a service to provide a rolling update of this analysis. Intermediate objectives of the CARBONES project included:

- the consolidation of the Carbon Cycle Data Assimilation System (CCDAS),
- preparation of high quality data streams for assimilation, model forcing and validation,
- development of a user friendly portal for displaying the results of the analysis and for comparison with independent products,
- development of a business model for the CARBONES service.

The above objectives have been achieved and the main developments and results are described in the subsections below.

1  CCDAS consolidation and generation of 20 year reanalysis

1.1 Evolution of the overall CCDAS

1.1.1 General description

The Carbon Cycle Data Assimilation System (CCDAS) used in CARBONES was designed to i) estimate the net CO$_2$ fluxes over land and ocean from regional to global scales for the past 20 years and ii) provide the estimates of the process-based ecosystem parameters (i.e. parameters of the underlying ecosystem equations) for improving the future simulation of carbon stocks and fluxes and for attributing the land carbon sink to the major processes. The generated CCDAS system relies on:

- Three model components:
  
  i) a global process-based vegetation model, ORCHIDEE, that can be applied at various spatial (local, regional, or global) and temporal (going from hours to centuries) scales;
  
  ii) a global statistical ocean model, OCVR (developed specifically for CARBONES project by ClimMod partner);
  
  iii) a tracer transport model, LMDz, to relate the surface fluxes to measured atmospheric CO$_2$ concentrations.
- Four main data streams to be assimilated:
  i) *in situ* eddy-covariance flux measurements of net carbon and water exchanges at more than 60 sites (FluxNet database);
  ii) satellite measurements of vegetation activity (NDVI from MODIS sensor);
  iii) surface ocean partial pressure of CO$_2$ (from Takahashi/SOCAT databases);
  iv) atmospheric CO$_2$ concentration measurements at more than 80 surface stations, from several data bases.

The approach relies on the minimization of a misfit function $J(x)$ that measures the mismatch between 1) all observation datasets $y$ and corresponding model outputs $H.x$, and 2) the values $x$ of the parameters (to optimize) and some prior information on them $x_b$, weighted by the prior error covariance matrices on observations $R$ and parameters $B$:

$$J(x) = \frac{1}{2} (H.x - y)' R^{-1} (H.x - y) + \frac{1}{2} (x - x_b)' B^{-1} (x - x_b)$$  \hspace{1cm} \text{Eq. 1}

We explicitly account for uncertainties regarding the model and the observations (summed within $R$), and the prior parameters (through $B$), assuming that the errors on prior parameters and observations follow Gaussian distributions.

### 1.1.2 Stepwise versus simultaneous multi-data optimization

In the CARBONES project, two different assimilation strategies were considered:

- a simultaneous assimilation of the different data streams in a single step assimilation, i.e. simultaneous minimization of a cost function quantifying the differences between modelled quantities and three of the observed data streams listed above, the ocean data stream being assimilated independently.

- a simpler-to-implement approach that relies on a stepwise (or sequential) assimilation of the different data streams. This approach was proposed and implemented given the scientific and technical issues concerning the simultaneous assimilation approach.

The different sensitivity tests and performance evaluations conducted through the CARBONES project led us to reconsider, from a scientific point of view, the “practical approach” versus the “ideal approach” for a multiple data streams assimilation system and in particular when using atmospheric observations. The main outcomes can be summarised as follow.

Given the following issues:

1. Even with their current complexity, global land ecosystem models still do not account for some crucial processes that control the net annual carbon balance of terrestrial ecosystems. For instance, in the version of ORCHIDEE that we used, we do not explicitly represent the nitrogen cycle and the limitation of the maximum photosynthetic capacity it may impose on high latitude ecosystems. We also do not account for all ecosystems management practices crucial for crops but also for...
forests. Moreover, soil processes are probably poorly represented given the complexity of soil ecosystems;

2. Atmospheric CO$_2$ observations provide unique constraint on the total net land and ocean carbon surface exchanges, including all processes; MODIS-NDVI provide only some information on the timing of the vegetation activity and not directly on the net CO$_2$ fluxes; FluxNet data provide direct constraints on the “rapid” carbon exchanges (from half-hourly to synoptic and inter-annual variations) at the plant level which may not include all ecosystem management practices at field or larger scales;

We thus proposed two different approaches depending on the “scientific objectives”:

1. **Objective 1: producing the best possible carbon fluxes re-analysis.** In that case the optimization of model parameters using atmospheric CO$_2$ observations might not be adapted given current ecosystem model errors. The best solution is probably to optimize the ORCHIDEE model parameters using MODIS-NDVI and FLUXNET data in a first step and then in a second step to correct the fluxes themselves with atmospheric CO$_2$ observations and specific constraint on the a priori flux error correlations.

2. **Objective 2: producing an optimized model for future simulation:** In that case we can argue that a simultaneous optimization of ORCHIDEE parameters using all datasets together is the best approach. The use of atmospheric CO$_2$ observations will be beneficial under the hypothesis that missing processes will lead to parameter corrections that do not degrade the predicting capability of the model. In other words, over tuning parameters to compensate for missing processes should not bias the model response to climate change. Although this is not guarantied, such hypothesis is probably reasonable and at least we can expect that the optimization with the global atmospheric constraint should produce a better model in terms of predicting long term mean fluxes.

1.1.3 Stepwise CCDAS approach

This section describes the evolution of the stepwise approach and the different implementations of such approach that have been applied.

1.1.3.1 General scheme description

Figure 1 shows the different steps of the stepwise approach. We can describe these steps as:

**Step 1:** Assimilation of the remotely sensed products of vegetation greenness (NDVI) derived from MODIS into ORCHIDEE; the prior parameters including values and error covariance (\(X_0^{\text{ORCH}}\) and \(P_0^{\text{ORCH}}\)) are optimized to produce a first set of optimized parameters \(X_1^{\text{ORCH}}\) with error covariance \(P_1^{\text{ORCH}}\), as well as the corresponding global optimized net CO$_2$ fluxes \(X_1^{\text{NEE}}\) and their associated uncertainty \(P_1^{\text{NEE}}\). For this first step we restricted the parameters to four phenology parameters for each Plant Functional Type (PFT).
Step 2: Assimilation of in situ flux measurements: the parameters $X_i^{ORCH}$ and $P_i^{ORCH}$ are used as input to the optimization system together with new parameters controlling the major “rapid” processes in ORCHIDEE (from diurnal to seasonal and partly inter-annual time scales). These parameters are further optimized to produce both:

i) a second set of optimized parameters ($X_2^{ORCH}$ and their error covariance $P_2^{ORCH}$)

ii) a set of optimized NEE fluxes and their uncertainties to be also used as prior input in step 4 ($X_2^{NEE}$ and their error covariance $P_2^{NEE}$).

The second option has been added as a complementary approach. As described above it prevents from relying too strongly on ORCHIDEE’s equations with the possibility to correct independently the land fluxes for all pixels in step 4.

Step 3: Assimilation of ocean pCO$_2$ measurements into a statistical model (neural network) to produce a priori air-sea fluxes.

Step 4: Final assimilation step using the atmospheric CO$_2$ measurements as a global constraint. It uses as input the results of step 2 for the land part and step 3 for the ocean part. We considered two complementary approaches for the land component. The optimized parameter (and attached uncertainty) will either consist in:

- approach 4.a: the parameters $X_i^{ORCH}$ of ORCHIDEE and their error covariance $P_i^{ORCH}$ from step 2. The approach then consists in readjusting the parameters that have been previously pre-optimized using FluxNet and satellite observations. Such approach led to version V3 of the system.

- approach 4.b: the net CO$_2$ fluxes over land $X_i^{NEE}$ and their errors $P_i^{NEE}$ from step 2. They will then be further optimized pixel by pixel.

The result of this last optimization step consists of i) either final parameter values for ORCHIDEE and errors or directly the net land carbon fluxes and ii) the optimized ocean fluxes. For the land, in the case of the flux optimization, some inconsistencies will still remain between the land carbon stocks and the land fluxes. However a final adjustment (post-processing) has been performed to adjust the gross fluxes (i.e. respiration and photosynthesis) so that they match the estimated net flux. The adjustment is based on an equal split of the net flux correction between photosynthesis and respiration.
1.1.3.2 Parameters to be optimized in each step

Figure 2 summarizes the number of ORCHIDEE parameters that are used in the different steps. The objective was to restrict as much as possible the number of parameters for each data stream to the most influencing ones. The objective is to expose only the main parameters controlling the dynamic of a given type of observations. We thus selected for each step:

- Step 1 (MODIS-NDVI): only 4 parameters controlling the onset and senescence of the leaf for the deciduous PFTs. This gives us around 40 parameters.

- Step 2 (FluxNet data): we used the same four parameters as in step 1 plus between 12 and 18 new parameters (depending on the set up) controlling photosynthesis, respiration, and energy and water balances, for each PFTs. Note that some parameters are common to all PFTs so that in total we have around 100 parameters. Also we consider specific parameters for each site to scale the initial soil carbon pools. These parameters are not used for global simulations.

- Step 4 (CO₂ data): We only used 2 or 3 parameters for each PFT (main test with 3 parameters) that control photosynthesis (the values and errors for these parameters come from step 2). Additionally, for the respiration, we introduce new parameters to scale the initial carbon pool size for 50 “eco-regions” and we also further optimized the Q10 parameter (sensitivity of the respiration to temperature) using for the prior and errors the mean across all PFTs from step 2. This led us to consider around 80 parameters in total. The scalar of the initial carbon pool size is crucial to account for all past disturbances that have brought the soil carbon pool in disequilibrium compared to the input of carbon from the current vegetation (due for instance to a transition from forest to grassland).
1.1.3.3. Summary of the different stepwise versions

Several different reanalysis corresponding to different setups of the CCDAS have been generated. They include four versions based on the stepwise approach.

- **Version V0.5**: The system relied on a standard atmospheric inversion based on Chevallier *et al.* 2010, using only few specific CARBONES inputs (namely the land fluxes from the ORCHIDEE model where the phenology had been updated).

- **Version V1**: Stepwise approach as described in the report D410. In this version, only the information content from MODIS NDVI data, from ocean pCO$_2$ data, and from atmospheric CO$_2$ data is assimilated in the CCDAS. The eddy-covariance data were not accounted in this version, but we used the new fossil fuel emissions (from IER).

- **Version V2**: This is a version of the stepwise approach with all four datasets (satellite NDVI, FluxNet, Ocean pCO$_2$, and atmospheric CO$_2$ data) assimilated. A new treatment of biomass burning emissions has been included as well as a revised version of fossil fuel fluxes and of the ocean component. For this version, the last step with the assimilation of the atmospheric CO$_2$ concentrations (step 4, see Figure 1) corresponds to a correction of the fluxes produced by the ORCHIDEE model over land from step 2. This choice follows from the objective to produce the best carbon flux reanalysis as described above. Version V2 is displayed on the web portal.

- **Version V3**: This is a second version of the stepwise approach (equivalent of version V2) where now the final step with the assimilation of the atmospheric data is performed through the optimization of the ORCHIDEE model parameters. This choice corresponds to the objective to produce the best land surface model to predict the fate of terrestrial carbon balances, as explained above.
1.1.4 Simultaneous CCDAS approach

This approach defines the ultimate goal of the CCDAS system: to combine the three observational data streams (except the surface pCO2) in order to estimate the parameters of a process-based ecosystem model. This will ensure a complete consistency between the optimized land carbon stocks and the gross/net CO2 fluxes. Over ocean, the approach is currently first based on a flux correction and not on parameter optimization. The simultaneous approach is conceptually simpler than the stepwise approach (only a single inversion), however, in its implementation a number of scientific and technical issues need to be addressed.

Given the complexity (both technical and scientific) of the construction of the simultaneous version, a single version, V4, has been run near the end of the project over a restricted period (3 years) to generate a corresponding reanalysis:

- **Version V4**: This is a first version of the simultaneous (or joint) approach with the optimization of the model parameters with three data streams together. It is based on monthly mean CO2 concentration for the atmospheric data. However this version constitutes only a demonstrator that has been run over a limited period of three years (2002 to 2004) in order to establish the benefit of simultaneous versus stepwise optimization. Several sensitivity tests for the stepwise and simultaneous approaches with more or fewer parameters to be optimized have been performed in order to establish a thorough analysis of the strength and weaknesses of each approach.

1.2 Updates in the underlying models

Important developments concerning the underlying physical models of the CCDAS were realized in the framework of the CARBONES project. They are discussed separately for the three concerned models below.

1.2.1 Ocean component

A new approach was developed in CARBONES to compute the air-sea CO2 fluxes by using ocean observations (ocean surface partial pressure of CO2 (pCO2oce)) within a statistical model (referred to as OCVR) coupled to the global CCDAS.

The air-sea CO2 flux is computed from the following expression:

$$F_{CO2} = K_{ex}(pCO2_{oce} - pCO2_{atm})$$  \hspace{1cm} (eq.13)

where pCO2 is the partial pressure respectively in the sea surface water (oce) and the atmosphere at the interface to the water (atm). K_{ex} stands for the exchange coefficient and F_{CO2} the flux from the water to the atmosphere.

PCO2oce are computed by using an artificial neural network that relates pCO2 to explanatory variables (Sea surface temperature, Sea surface Salinity, Mix layer Depth, Chlorophyll content observed from space) derived partly from MERCATOR physical ocean re-analysis over the 20 years. The inferred pCO2oce data from OCVR are validated by using data not assimilated in the neural network process. A good agreement was obtained when comparing the inferred PCO2oce with Takahashi climatology data (Takahashi et al. 2009). The
atmospheric partial pressure of CO$_2$, referred as pCO$_2$$_{atm}$, is given by the surface concentration of CO$_2$ from a global transport simulation of previously optimized surface fluxes (Chevallier et al. 2007) corrected i) for surface pressure using the IERA analysed field and ii) by the saturation vapour pressure at the water temperature (following the OCMIP protocol).

The performance of the OCVR model for surface pCO$_2$ has been evaluated by comparing the differences between model and observed pCO$_2$ for 25% of the raw data not used in the learning phase of the neural network. The performances are compared to the climatology of Takahashi extrapolated in time with a linear trend. Overall the performance of OCVR significantly exceeds that of the climatology.

The final air-sea flux product used in the CCDAS (in step 4 of the stepwise approach) provides a global air-sea exchange. The major global features are:

- A mean sink over the 20 years around 1.6 PgC/year,
- A decrease of the sink from 1992 to 2000 of nearly 1 PgC/year,
- A strong increase of the ocean carbon uptake after 2000 that reach in 2009 around 2.7 PgC/year.

1.2.2 LMDz transport model

The Carbon Cycle Data Assimilation System (CCDAS) that has been developed strongly relies on the atmospheric transport model, LMDz, to relate surface fluxes to the observed atmospheric CO$_2$ concentrations. The main characteristics of that model have been described in the report D410. We only remind that:

- The model has a typical horizontal resolution of 3.75°(longitudes) x 2.5°(latitudes) and 19 sigma-pressure layers.
- The winds (u and v) are relaxed to ECMWF analyzed meteorology (Uppala et al., 2005) with a relaxation time of 2.5h (nudging) in order to realistically account for large scale advection.
- The advection of tracers is calculated based on the finite-volume, second-order scheme as described by Hourdin and Armengaud [1999].
- Deep convection is parameterized according to the scheme of Tiedtke [1989].

Recent and on-going developments at the LMD laboratory have led to an improved version (LMDz version 4), compared to the first version that was used, with the following changes:

- A more sophisticated parametrization of the planetary boundary layer has been implemented by Hourdin et al. (2002) based on the representation of a dry thermal in terms of mass fluxes, as in deep convection schemes.
- Several small inconsistencies have been cleared for the convection and advection schemes.
The updated version has led to significant improvement in terms of simulated atmospheric 
CO\textsubscript{2} concentrations.

1.2.3 ORCHIDEE Global Vegetation Model
Several major additions were done in the framework of CARBONES on the ORCHIDEE 
model. They included the improvement of the crop module with the inclusion of specific crop 
plantation and harvesting dates across Europe. Moreover, the Peking University partner also 
developed specifically a parameterisation of the rice crop for the tropical regions.

1.3 Error estimation and error propagation
Providing errors on the estimated carbon fluxes is crucial for potential users of the 
CARBONES products. The optimization problem has been formulated following the 
Bayesian probability theory. In this framework, several hypothesis were made to solve the 
inversion problem and further estimate the model parameter uncertainties:
- All errors (parameters, model, measurements) were assumed Gaussian,
- The land ecosystem model was linearized at the minimum of the cost function in 
  order to compute the posterior uncertainties on the parameter following the linear 
  assumption.

With the linear assumption, the posterior error variance-covariance matrix \( P_b' \) on the 
estimated parameters is expressed as a function of the a priori error variance-covariance 
matrices on the parameters \( P_b \) and on the observations \( R \), as well as a function of the Jacobian 
matrix of the model \( H \) describing the sensitivity of the different outputs to the parameters that 
are optimized:

\[
P'_b = \left[ H' R^{-1} H + P_b^{-1} \right]^{-1} \tag{1}
\]

Such formulation may involve a potentially very large matrix \( H \) that has the dimension: \( \text{nobs} \times \text{ npar} \) with \( \text{nobs} \) being the number of observations and \( \text{ npar} \) being the number of parameters. 
Depending on the number of observations or parameters the size of \( H \) may become 
prohibitive.

For each step of the stepwise approach we have:
- Step 1 and step 2: in these cases the number of parameters and the number of 
  observations were relatively small (few hundreds) so that we could easily estimate \( P_b' \) 
  with equation (1)
- Step 4 (with atmospheric CO\textsubscript{2} data): In this case \( H \) becomes the product of a Jacobian 
  matrix for the ORCHIDEE land surface model with the Jacobian matrix from the 
  transport model. Given the size of these intermediate matrices we were only able to 
  derive the uncertainty for a restricted period (one year).

After the computation of the error variance covariance matrix on the parameters, we can 
directly derive the errors on the output variables due to parameter errors (error propagation) 
and especially characterize the errors on the net CO\textsubscript{2} fluxes following the matrix product:

\[
R' = H P_b' H' 
\]

with \( R' \) being the error variance covariance matrix on the outputs. Note that in this case \( H \) 
will only correspond to the Jacobian of the ORCHIDEE land surface model.
1.4 Validation and benchmarking of CARBONES products

1.4.1 Validation of CARBONES products against independent data

We performed a first evaluation of CARBONES products against other independent products. This comparison was done to highlight the information content brought by CARBONES on the carbon budgets at continental to hemispheric scales. This comparison should still be considered as preliminary as it will be further updated after the duration of the project. Overall, CARBONES performs as well as the other products for the version that was first compared (V1) and bring new features on the carbon cycle, that need further investigations:

- Fossil fuel emissions with hourly temporal variations appear to be significantly larger that those from other products
- Ocean carbon fluxes show new Inter-annual Variability (IAV) with a pronounced increased ocean uptake after 2002 that still needs to be validated or confirmed with other proxies.
- Net land carbon fluxes follow most standard atmospheric inversion results but with slight differences: the IAV over the tropic is slightly lower in CARBONES than the other products; the net land carbon uptake in the northern continental regions is slightly lower than most inversions, especially for North America; the southern land IAV tend to be larger than in the other products. All these features will be summarized in paper focussing on the analysis of the 20-year CARBONES reanalysis.
- A favourable comparison against biomass forest data from FAO statistics appears with lower biases after the assimilation of MODIS-NDVI and FluxNet data.

1.4.2 Comparison of CARBONES reanalysis against results from ESM models

In conclusion, our evaluation has shown that CARBONES is clearly improved from either stand-alone ORCHIDEE or the use of ORCHIDEE within the IPSL coupled ESM. However, the CARBONES version (V1) that is used in this comparison is not yet “out of the pack” of ESMs – other ESMs are comparable to CARBONES and in some metrics and regions outperform it. Nevertheless we expect that the subsequent release of CARBONES product (even beyond the end of the project) will constantly improve and thus may outperform the other ESMs. Note also that such comparison is based on a comparison with “data-oriented products” and not direct observation which are not available at the continental to global scale and that these product may be associated with large uncertainties. Further assessment of observational uncertainty is required before we can make more quantitative statements about CARBONES.

Overall, this reveals the clear success of the CARBONES data assimilation work to improve and optimise the parameters in ORCHIDEE. It also suggest that the structure of the ORCHIDEE model is probably as or even more crucial than the parameter values that are optimized within CARBONES. Indeed, the above exercise shows that CARBONES is not, yet, at a level of quality where it can be seen to be clearly better than all other ESMs or to be used to evaluate the performance of ESMs.
Research into evaluating and constraining ESMs has shown a strong relationship between inter-annual response and long term response of tropical carbon to climate change. This work has highlighted the importance of evaluating not just climatologies of ESMs but their response to perturbations, especially in the tropics.

1.4.3 Benchmarking of GMES core products against CARBONES products

We performed a first comparison of CARBONES products against the MACC-2 net ecosystem carbon exchange and the GEOLAND-2 LAI product. This comparison was performed to highlight the information content brought by CARBONES and can be taken as a possible benchmarking of the GMES core product. First one should resume the main differences between the systems that could explain the product differences:

- **MACC2**: Such product is similar to CARBONES in the sense that it uses the same transport model (LMDz) and the same set of atmospheric observations and inverse set up to optimize the fluxes from atmospheric CO2 gradients. However, MACC-2 did not use the information content from the FluxNet sites (net ecosystem carbon exchange and latent heat flux) and the satellite information (in this case MODIS-NDVI) to constrain the prior land fluxes from ORCHIDEE (i.e., time of the photosynthesis activity); it also had a simpler fossil fuel emissions and only a climatological prior ocean flux whereas CARBONES uses inter-annually varying fluxes derived from pCO2 measurements.

- **GEOLAND2**: We only used the Leaf Area Index from that project given that the net land carbon uptake was not the primary objective of GEOLAND2. In the comparison we considered that GEOLAND2 provides one of the most advanced product for LAI currently available.

Overall, the comparison highlighted several significant differences that should (and will) be further investigated before drawing specific conclusion about the strengths and weaknesses of each approach and the land/ocean carbon budgets. As for the net carbon fluxes (comparison with MACC-2) the ocean year-to-year flux variations estimated from the surface pCO2 measurements (Takahashi database) with a statistical model (Neural network) in CARBONES and further corrected by the atmospheric CO2 data significantly differ from those obtained in MACC-2. Although they are driven by pCO2 measurements the variations in CARBONES need to be further confirmed by independent proxies before drawing strong conclusion about possible weaknesses of MACC-2 for the air-sea flux inter-annual variations. Over land, the CARBONES product provides net flux variations that are coherent with MACC-2 with some latitudinal/continental differences. These differences, although larger over the southern and tropical regions also concern the northern continental regions. They are tightly linked to the constraints brought by the FluxNet and MODIS-NDVI data, with a direct impact on the year-to-year flux variations and to a lower extend the long-term trend, through the adjustment of the land model sensitivity to climate anomalies. These changes are currently under deeper investigation, to specifically draw the recommendation about the future model multi-data fusion schemes.

As for the LAI product of GEOLAND2, the comparison was only preliminary and will be extend over a longer period (several years corresponding to GEOLAND2 product). Overall, we obtain higher LAI values in CARBONES compared to GEOLAND2 during the peak of
the growing season (for northern and southern ecosystems) that do not however bias the simulated photosynthesis fluxes (the net land carbon fluxes weakly depend on the maximum LAI at high values). More importantly the timing of the LAI appears to be rather similar between the two products. This feature will be investigated in more details using “10-day” mean LAI values as originally produced in GEOLAND2 in order to analyse in details possible small phase shift between the two products.

1.4.4 Impact of CARBONES on climate predictions

The initial objective was to assess the impact of CARBONES on climate predictions in the 21st century. The approach was to use the ORCHIDEE land surface model optimized through the assimilation of MODIS-NDVI, FluxNet CO₂ and latent heat fluxes, and atmospheric CO₂ concentrations. ORCHIDEE is indeed the land surface component of an Earth System Model (ESM) developed at IPSL (IPSL-CM5 ESM). Note that the version of ORCHIDEE used for CARBONES project corresponds to that used in the recent CMIP5 climate simulations for the IPCC.

However, due to scientific and technical complications, the final CARBONES product, i.e. an optimized land surface model, was not available until the end of the project. With this constraint, we could therefore not run the full ESM over the next century and instead adopted a simplified approach to assess the potential impact of model data fusion on climate simulations. The main objective is to assess the impact of the model parameter optimization on the fate of the global carbon cycle, and in particular the evolution of the land carbon uptake in the year 2100. Differences in land carbon uptake will indeed be one of the major drivers of climate change in 50 to 100 years.

- To achieve this goal, we performed two forward simulations of ORCHIDEE over the next century using a given climate scenario, one with the standard parameters of the model that were used for the CMIP5 climate simulations, and one with the parameters optimized in the CARBONES project.

- We then compared the global carbon budget of the two simulations in terms of cumulated fluxes over the next century as well as continental partition. A final objective was also to assess which gross carbon fluxes (photosynthesis versus respiration) lead to the largest difference between the optimized and the standard simulation.

This simplified approach neglects potential feedbacks between the climate system and the carbon cycle, which are usually positive: warmer climate tends to decrease the land carbon sink and consequently further enhances the greenhouse effect. However, this approach provides the first order effect of the parameter adjustment with current carbon cycle observations. The overall approach relies on the Inter-Sectoral Impact Model Intercomparison Project (ISI-MIP) protocol for the climate forcing. They took the output of Global Circulation Models (GCMs) used in the context of CMIP5, with historical forcings and Representative Concentration Pathway (RCP) scenarios taking into account greenhouse gases, short-lived gases and aerosols and land use. The modelled climate was first downscaled at 0.5 degree and further bias corrected using observed WATCH Forcing Data (WFD, Weedon et al., 2011) over the land surface.
The optimized ORCHIDEE version predicts a substantial decrease in the capacity of the terrestrial biosphere to absorb carbon, resulting in an increase in atmospheric CO$_2$. This therefore suggests that there will be an additional warming at the end of the 21st century. This is mainly attributable to a change in the soil and litter C pools to absorb the increase in atmospheric CO$_2$. We are currently investigating which parameter change is responsible for the strong soil carbon stock decrease, among the 70 parameters that were optimized. Several parameters controlling the sensitivity of the heterotrophic respiration to temperature (Q10) and to soil moisture are deeply involved.

2 Preparation of high quality data constraining the global carbon cycle over the last 20 years

The quality and reliability of the CARBONES reanalysis is strongly dependent on the quality of the various input datastreams used for forcing, assimilation and also validation. A significant effort was extended during the project to prepare the high quality datastreams with, whenever possible, an estimation of corresponding uncertainty. Of major importance for the successful outcome of this task was the good collaboration between data providers and data users. This collaboration has been guaranteed by, among others, numerous meetings and telephone conferences.

For the needs of users, the description of the CARBONES datastreams has been included on the project portal (http://www.carbones.eu/wcmqs/project/ccdas/). For each datastream the following sections are provided: data description, data uncertainty, data policy and references.

The list of available products includes:

Data for model forcing:

- Fossil emissions: University of Stuttgart has prepared and delivered a global hourly data set at a 1°x1° resolution for the years 1990 to 2009, considering three emission heights. The data sets are based on EDGAR v4.2 emissions on a 0.1°x0.1° degree resolution. The final delivery considers completely updated time profiles for the globe considering temperature dependencies. The temperature dependent time profiles were added for the following sectors: energy production, residential, commercial and intuitional combustion, manufacturing and road transport. The derived profiles are based on European consumption and statistical data in combination with global temperature data. For the industry sector also UN monthly bulletin data were considered when available. Additionally, university of Stuttgart did develop a fossil fuel emissions model for Europe on a 1km by 1km resolution with the corresponding time profiles. The University of Stuttgart has created a data portal for the download of the hourly global fossil fuel emissions as well as the spatial distribution of the European data sets. The hourly global data sets are offered as monthly files distinct into three emissions heights and the spatial distribution of the European data sets can be downloaded in a 5km by 5km sector-wise for the year 2008.
- Biomass burning: First guess biomass burning data were gathered from GFEDv3 database and were separated into three categories, depending on the regrowth
properties (LSCE). We distinguished i) savannah fire with an emission of carbon during the burning period and a regrowth at the same place during the next growing season, ii) natural forest fire with emission of carbon due to forest burning but also uptake of carbon due to re-growing forests in the nearby region, both fluxes compensating at steady state, and iii) net deforestation (including peat land fire). These three categories are treated differently for the emission of carbon in the atmosphere.

- ECMWF weather data using the Intermediate re-analysis (ERA-I) (LSCE). The data were reloaded and reformatted at a new resolution (0.72 degree).
- Annual land cover maps were prepared by LSCE.
- Air-sea fluxes from the ocean inversion model.
- Soil texture data (%sand, clay, loam) has been derived from HWSD data and aggregated to the ORCHIDEE resolution to be used in ORCHIDEE runs at some point in the future.
- Crop management: PKU has compiled a database on crop managements including crop calendar, fertilizer application and irrigation area. The datasets were harmonized from several resources. The nitrogen fertilizer application and irrigation area were based on statistics at NUTS-2 level and interpolated into 1-10 km resolution according to crop distribution and growth needs. The crop calendar were harmonized from available field observations mainly from USDA and FAO (USDA, 2006; FAO, 2007) and interpolated into 5 minutes resolution. To provide drivers to ORCHIDEE with consistent spatial resolution, all data were harmonized to 10 minutes resolution. These data will be provided to drive the agricultural module of ORCHIDEE in the CCDAS.

**Data for assimilation:**

- pCO2 data from the Surface Ocean CO2 Atlas: The SOCCAT database has not been used in the statistical model for the learning phase (we kept the Takahashi database) but it has been used to evaluate the OCVR products.
- FLUXNET data (MPI-BGC): FLUXNET eddy covariance site data were collected for more than 80 sites for a large range of ecosystems. The FLUXNET data files were prepared in a harmonized format with same and consistent variable names, meta-information and uncertainty estimates for fluxes were included into the files.
- AVHRR and MODIS NDVI data (LSCE).
- Atmospheric flask measurements of CO2: Raw data from major networks: NOAA-CMDL (US), RAMCES (French), CSIRO (Australia) and few other site from WCGD.

**Data for validation:**

- Soil carbon (ABDN): HWSD soil data has been extrapolated down the whole profile using a soil depth database from Webb et al., (2000) according to an algorithms developed Nuno Carvalhais (MPI_BGC).
- Grassland biomass and soil carbon inventory for China: (PKU): More than 800 plots over China’s major grasslands (semi-arid and arid grasslands over Inner Mongolia and alpine meadow and tundra over Tibetan Plateau) has been collected and harmonized to build the grassland biomass and soil carbon dataset. These plot-
scale data have been upscale to provide a 1 km resolution grassland biomass and
soil carbon map over China. According to this dataset, China's grassland biomass
stores 2.0 Pg C, and the grassland soil stores 26.1 Pg C. This dataset is provided as
an independent benchmark of biomass and soil carbon estimates of the
CARBONES CCDAS.

• Forest cover and forest stocks (DLO): The global datasets on forest cover and
forest stock are improved and interpolated and global maps at 1°x1° are produced
at five year intervals from 1950. From these maps of forest biomass, forest carbon
and changes in forest carbon are derived. Uncertainties for forest area and forest
stock are estimated and mapped. The CARBONES global forest dataset is fully
documented and will be published in 2013.

• 30-year leaf area index (LAI) dataset from AVHRR satellites for China (PKU): AVHRR LAI dataset over China was compiled from LAI third generation dataset,
which has full spatial coverage of China with 10km spatial resolution and half-
month temporal resolution available from 1982-2011. The dataset is provided to
benchmark of LAI output of CARBONES CCDAS.

• FLUXNET-Upscaling fluxes of carbon and water fluxes (MPI-BGC): Global data-
oriented estimates of carbon and water fluxes are available. These datasets were
upscaled from FLUXNET eddy covariance site observation with remote sensing
and meteorological data based on a model tree ensemble algorithm (MTE). The
reliability of these datasets was evaluated through comparisons against FLUXNET
site data based on cross validation and through comparisons with satellite derived
fluorescence measurements. These comparisons identified mean annual gross
primary production (GPP) from MTE as a reliable benchmark dataset for the
CCDAS. Uncertainties of this dataset were derived. Additionally, a GPP estimate
from the LPJmL dynamic global vegetation model was provided as additional
evaluation dataset from an independent process model. The LPJmL model has
contrary to the CCDAS no optimized phenology and photosynthesis scheme while
it integrates satellite observations of land cover and fire disturbance. These datasets
of mean annual GPP from MTE and LPJmL were prepared to be used for
evaluating the CCDAS (see also WP 430).

• GEOLAND II data: The kilometric Leaf Area Index product over the earth has
been downloaded and processed for one particular year (2002) in order to average
the LAI at the 1x1 degree resolution of CARBONES. Similar treatment for other
years is on-going.

• MACC II data: The net ecosystem exchange and ocean fluxes have been obtained
and reformatted on the same grid as the CARBONES products to facilitate the
evaluation.
3 Development of the portal

One of the key elements of the CARBONES system was the development of a tailored, user-friendly web interface (www.carbones.eu), allowing climate modellers, other science communities, local or regional communities and the general public to understand and easily visualize the living carbon cycle over the past 20 years. The portal allows displaying the CARBONES reanalysis via a web site in which spatial data are displayed on a dynamic map, similar to Google Maps. It has been developed so that users can select the output according to area, time period and output variable. Web map elements are implemented as a thin client compatible with all major browsers without requiring any plug-in or specific setting (although JavaScript is required to view the maps).

The definition of the web portal’s functionalities has been developed through gathering and analysing user requirements from various partners CARBONES partners as well as external climate experts through dedicated meetings. The CARBONES portal allows displaying the products either as maps or as time series. The web-mapping includes highly interactive maps with zoom and pan functionality similar to that available in Google Maps. It presents the “raw” data set without post-processing as well as the results of post-processing, including temporal filtering. In addition the portal allows comparing the CARBONES products against predefined independent data. It also gives an overview of the project, explaining the motivation, objectives, approach and product characteristics and the dissemination.

4 Development of future business models for CARBONES services

An important and innovative work was performed within the CARBONES project to obtain a clear understanding of who are the possible end-users of CARBONES-like products, what are their needs and also what are their expectations as far as possible purchasing price of the CARBONES product. A complete market analysis, based on material issued from dedicated workshop and from interviews with climate experts, with representatives of other public and private sectors, and on desk research led to a complete and detailed market analysis. In parallel, a business plan was prepared with possible business models and ideas for potential applications.

The main conclusion of the Market Assessment report was that in the current context, where other competitor products are available free of charge, there is little revenue potential for the core CARBONES services, which were developed under the EU funded project. The Market Assessment includes recommendations for future activity beyond project completion. Based on the market assessment and the running costs of CARBONES provided by partners a number of different business models have been identified and discussed within the Business Plan. As the core services are unlikely to generate any revenue it has been recommended to diversify the services. Suggestions for a diversified business model have been made. However, due to the current development stage of these new services, it has been impossible to estimate the future market value of CARBONES.