COOPEUS

WP3: Carbon Observations

Model Data Fusion Workshop

JD Paris, N Schneider (CEA-ICOS)

H Loescher, L Powers (NEON inc.)

with contribution from rapporteurs

Deliverable:	**D3.5, Report on the Model data fusion workshop**
Date:	**15/06/2015**
Lead Partner:	**CEA**
Document Issue:	**Final, v1.0**
Dissemination Level:	**Public**
Contact:	**Nadine Schneider, LSCE, France** **nadine.schneider@lsce.ipsl.fr** **Tel +33 - 69 08 - 71 24**

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement n°312118.

TABLE OF CONTENTS

ICOS and NEON

ICOS[1] and NEON[2] research infrastructures are in-situ observation networks providing research data on greenhouse gas fluxes from ecosystems to the atmosphere. Together, ICOS and NEON aim to make these data available without technical, scientific or political barrier. These data typically include greenhouse gas (GHG) concentration, carbon and energy flux observations, and the surface micrometeorology surrounding these measurements.

ICOS Atmospheric Thematic Center (ATC)
Laboratoire des Sciences du Climat et de l'Environnement (LSCE)[3]
related to the French National Center for Scientific Research (CNRS), the Atomic Energy Commission (CEA) and the University of Versailles Saint-Quentin (UVSQ).
Orme des Merisiers; 91191 Gif sur Yvette, France

NEON, inc.
National Ecological Observatory Network
1685 38th Street, Suite 100
Boulder, CO 80301

[1] www.icos-infrastructure.eu
[2] www.neoninc.org
[3] www.lsce.ipsl.fr

1 INTRODUCTION

1.1 Introduction and rationale of the workshop

COOPEUS[4] partners ICOS and NEON convened a *Data assimilation workshop* at the Cité Universitaire, Paris/France, April 20-21, 2015. The initial objective, from the COOPEUS DoW, Task 3.4 "Carbon Observations – Interoperability and interworkability analysis", is to '*compare experiences and identify issues encountered by the users'*, more specifically '*model-data fusion experts'.* We brought together a diverse group of greenhouse gas modeling scientists, from ocean, surface flux and atmospheric expertise, working at the forefront of data model fusion research also working across scales, disciplines, and involved in education, communication and community engagement. We set two primary objectives to the workshop, i) to present and discuss recent progress of the state of the art of data assimilation approaches - since data assimilation is considered as a large, if not the largest, potential use of data originating from various networks and hence a proxy for interoperability requirement of C networks-, and ii) to gather input on the data management, informatics, and workflow needs for this research community.

1.2 State of the Science.

Data Assimilation (DA) and data model fusion (MDF) has proved to be a robust methodology advancing our ability to estimate carbon exchange across spatial and temporal scales. However, progress in improving model uncertainty has been hard earned and there is a need expressed by the community to accelerate this process. Improving data access and interoperability will greatly facilitate decreasing model results' uncertainty, by allowing incorporation of larger, quality controlled observational data sets.

Over the past decade DA has been applied in many new and different fields of carbon cycle research, including emissions inventory, terrestrial and ocean surface processes, and incorporating social phenomena as drivers of emissions. Frontiers in the approach are numerous, and include, as highlighted during the workshop, i) developing novel methodologies for assessing model uncertainty, resulting in development and adoption of best community practices, ii) establishing emission inventories (temporal exchange rates), development of emission audit processes, and a process understanding for surface flux estimates, iii) constraining CH4 surface flux measurements (a poorly constrained but important GHG), iv) working towards common model architecture to optimize processing- and spin-up time, and v) an appreciation for the need to combine both fluxes and pools of carbon (not just fluxes) to increase certainty, accuracy and precision of terrestrial surface flux estimates.

Given the utility, advances and frontiers, all participants acknowledged the need to further build a community forum to facilitate further interactions and promote the access to new and existing observational data.

[4] www.coopeus.eu

1.3 Agenda

Day 1: 20/04/2015, Monday		
10.00-11.00	Registration with Coffee	
11.00-11.40	**Opening and welcome**	
20 mins	Introduction and scientific objectives	W Kutsch
10 mins	The COOPEUS perspective: ICOS and NEON	JD Paris, H Loescher
10 mins	Logistics	N Schneider
11.40-12.40 **SESSION 1** **2h presentations** **1h discussion**	**Observing systems (general introduction)** ***Chair: A Vermeulen*** ***Rapporteur: L Rivier***	
40 mins	Towards a global bottom up estimate of regional C fluxes: implications for systematic carbon observations	P Ciais
	LUNCH (invited)	
14.00-16.00 **SESSION 1** **(continued)**	**Elements of a global observing system** ***Chair: A Vermeulen*** ***Rapporteur: JD Paris***	
20 mins	Inventories and emission target compliance verification – the role of the Edgar database	G Maenhoud *(webex)*
20 mins	In-situ atmosphere observations	L Rivier
20 mins	In-situ ecosystem observations	H Loescher
20 mins	In-situ ocean observations	D Bakker
45 mins	Discussion	Moderator: W Kutsch
	BREAK	
16.30-18.30 **Session 2** **4h presentations** **1.5h discussion**	**Model-data fusion applications:** **Toward compliance verification and improved climate models** ***Chair: H Loescher*** ***Rapporteur: JD Paris***	
40 mins	Are atmospheric measurements suited to MRV in existing carbon pricing mechanisms?	V Bellassen

20 mins	National inventory and uncertainties	J Ometto *(webex)*
20 mins	Ecosystem manipulation experiments	A Chabbi
40 mins	Discussion	Moderator: W Peters
19.30	Conference Dinner (tbc)	

Day 2: 21/04/2015, Tuesday		
9.00-10.30 SESSION 2 (continued)	***Chair: A Vermeulen*** ***Rapporteur: H Loescher***	
20 mins	Process modeling	N MacBean (P Peylin)
20 mins	Data assimilation for GHGs in ECMWF	A Agusti-Panareda
20 mins	Inverse modeling for CO2 and MAC III (global)	F Chevallier
20 mins	Regional-scale inverse modeling	U Karstens
	BREAK	
11.00-13.00 SESSION 2 (continued)		
20 mins	Data assimilation in Carbontracker	W Peters
20 mins	CCDAS	M Scholze
20 mins	FFDAS	A McRobert (P Rayner)
60 mins	Discussion	Moderator: W Peters
	LUNCH (invited)	
14.30-14.50 SESSION 2 (continued)		
20 mins	Data assimilation to enhance predictive ability of Earth system models	Yiqi Luo *(webex)*
14.50-16.00 **Session 3** 1h presentations 0.5h discussion	**Outreach: Greenhouse gas research in the wider context** ***Chair: L Powers*** ***Rapporteur: W Peters***	
20 mins	Communicating toward a decision-making audience and the general public, the ICOS Carbon Portal	A Vermeulen
20 mins	Global Carbon Project	P DeCola *(webex) TBC*

30 mins	Discussion	Moderator: A Vermeulen
	Break	
16.30-18.00	**Future and Closing**	
45 mins	Discussion: Future directions on MDF	Moderator: W Kutsch
45 mins	Conclusion, Wrap-up	H Loescher, JD Paris
18.00	*Meeting adjourned*	

2 CONTENT OF WORKSHOP SESSIONS

2.1 Observing system (general introduction) (rapporteur: L Rivier, JD Paris)

Towards a global bottom up estimate of regional C fluxes: implications for systematic carbon observations (P Ciais)

PC shows a bottom up computation of global carbon budget doing a synthesis of the RECAP papers completed by other recent studies. Care has been put on consistency issues in accounting the various fluxes including eg fires, fresh water fluxes, harvest import/export wood products etc. The take home message is that the heterotrophic respiration (HR) is much smaller than previously estimated. HR in carbon models often 95% of NPP. In reality it is more around 80%. Ratio of stored carbon to NEE about 0.8 instead of 0.95 as in C models, explained by lateral fluxes and exports.

PC then derives carbon storage efficiency (excl. LUC, related to NPP) (13% in Russia, small in dry region and surprisingly fairly large for managed regions). Lateral fluxes impact future prediction of C Cycle : increasing lateral fluxes, increasing CO2atm.

The bottom up sum is compatible with atmospheric derived numbers.

Standing question: what is the significance of the fuel wood? How do NEE compares to inversions?

The implication for future scenarios is discussed using the compact system model OSCAR. One conclusion is that, at the end of the century a higher effect of both lateral fluxes and climate feedback is to be expected.

Q&A: Soil heterotrophic respiration has to be estimated as a residual because no global coverage dataset. Even the UK 5-km grid is criticized for not accounting for erosion.

Need to improve the priors used in atmospheric inversion to have specific information from top down approaches.

Eco fluxes observations are biased on natural ecosystems by interest of scientists, but ignore too much of crops especially biofuels crops. RCP8.5 and RCP2.6 address this question from the scenario perspective.

Simulationof correct timescales and residence time remains an issue for carbon in rivers especially in arctic regions.

2.2 Elements of a global observing system (rapporteur: L Rivier, JD Paris)

Inventories and emission target compliance verification – the role of the Edgar database (G Maenhout)

GM listed the available inventories both global and regional, warned us about some downscaling issues (temporal resolution). Reported inventories updates usually correct previous years. Missing sources are still are a limitation of wider usage. In the EU, electronic reporting should improve the intercomparaison of the national reporting. Validation of inventories takes place by doing intercomparaison. Sometimes comparison with inversion can correct for errors (See for ex, UK methane declaration increased by 30% in 2014, ref Bergamaschi et al., 2013).

The uncertainty for country CO_2 emission ranges from 15 to 35% depending on the sector (energy, mining, industry, transport …).
Reported emissions to UNFCCC and CLRTAP for CH_4 are surprisingly different, esp. for US. Coal mining and shale gas extraction source of high uncertainty for CH_4.
Inverse modelling useful for comparing with inventories
Warning not to use grid masks for calculating country/region total because large point sources can be on borders and wrongly attributed.
Emissions drivers show high year to year variations: energy intensity of industry, heating-degree-day, incomes, …
SO_2 emissions diminish with GDP, but CO_2 does not! Therefore the Kuznet curve (where income drive emissions, with a bell-shape) is to be questioned.
To summarize key limitations are completeness (perimeter) issue, downscaling issue, and need to learn about details on drivers, country-specific. The 2014 inventory will be issued soon.
Some downscaled inventory down to 0,01 deg where you then see point sources exit for ex for Paris and Milan region. Downscaling implies strong time variation (eg a plant closing) that were averaged out at lower resolutions. Issues at country boarders where large point sources could be attributed to the wrong country when discretizing.
Global trends show that Chinese Carbon efficiency (C/GDP) is decreasing however staying higher than for Europe. Note that decreased emission with higher income is not seen for CO_2 (works however for SO_2 for ex). More generally, to understand drivers of trends you need long records.
Beware using population as a proxy at regional scale, it doesn't always work.

In-situ atmosphere observations (L Rivier)

Link bias to applied water correction and instrumental calibration in Picarro, Gerbig et al. INWIRE, Masarie et al. ,
Frontier: assimilation of boundary layer height? The data will be made available by ICOS ATC. When going to higher resolution, do we need to keep the 0.1 ppm accuracy target? 14C: gives $FFCO_2$, $13CO_2$, isotopic signature for different sources,
small sensors for dense city networks
Q&A How ICOS can accommodate for numerous, small sensors with lower accuracy? This is rather emerging topic at the moment, no ready instrument, no plan to integrate in ICOS immediately

In-situ ecosystem observations (H Loescher)

'number'/budget approach vs process understanding, allow prognostic ability.
No direct measurement of GPP, NPP interest scientists as a proxy but various definitions/perimeters, relevant time scales to be considered.
C cycle approaches are shaped by the intention of the scientists who design them: implicit models of C cycle. Can be designed for ecosystem inventories, direct measurements, based on physiological models, or optical derived proxies.
Ecosystem states, equilibrium, resilience, are important for prognostic capability

In-situ ocean observations (D Bakker)

Surface and interior ocean data is made available through various projects of data processing and synthesis: GLODAP v2 (global data analysis project, include interior ocean), SOCAT, SOCOM (Surface Ocean CO2 Mapping intercomparison). Now various mapping methods inform global carbon budget. Models will be useless in 100 years but correctly sampled observations will be reused indefinitely.
Variables that constitute ocean observations include: DIC, TA, pH, fCO2 (fugacity) = gamma*pCO2 (partial pressure), gamma = 0.97
Flux= k K' (fCO2water-fCO2air)
CH4 and N2O are also measured
Volunteer observing ships have the possibility to observe: fCO2, DIC, TA,
Deep sections: interior DIC, TA CH4, N2O
Drifters: pH, fCO2 , …
GLODAP 2 integrates GLODAP, PACIFICA, CARINA (Tanhua).
Core data is Temp, salinity, DOC, alkalinity. Includes also oxygen, nutrients, freons,. Will be released in 2015
SOCAT V3: covers 1957-2013, published in 2015
Systematically referenced in ESSD
Riverine carbon input is main uncertainty in marine C budget .

Discussion (Moderator: W Kutsch)

Werner: It is desirable to issue a call for more observations and sensors, but we're not structured in appropriate way to sustain this. DB: for example Glodap lives on a showstring .
Open questions include:
- How does a portal ensure interoperability ?
- Data availability and accessibility beyond a specific community?

AAP: ECMWF Copernicus atmospheric services: finds data useful but not enough stations to feed a global model. Need more data globally.
P Ciais: could shut down funding when data not available. Need to unlock old data as well. WK: ICOS not yet running formally, and old data not in the standard of ICOS. Although some data used in publication are held from Fluxnet. Hank: NEON system will start running, data available summer 2017
PC: Users not limited by tools, have their own FTPs etc, but need for versioning of data after each update/recalculation/recalibration.
WP: package data for users who prefer it, but also direct transmission of data e.g. obspcak. Carbon Portal could repackage the old CarboEurope (legacy) data.
Abad: is legacy data still usable? Need repurposing, recalculation? -> Is no data better than bad (poorly characterized, inadequately sampled, not compatible) data? E.g. obspack is not perfect but useful, and on the opposite Fluxnet data seems to be kept private (la thuile) behind the notion of perfect data.
People designing observation system focus on future: technology, interoperability, discoverability, while the user are currently experiencing problem with getting existing data (being kept locked), using legacy data, etc.

SOCAT's most used product is synthesis, people less interested in gridded product and by fancy website's access capability. SOCAT was designed for legacy data.
Difference between EU and US is EU may have easier repository for PI data. We encourage journals and funding agencies to 'force' open access to data and good referencing. Other bad practice: PI need to give access to the good data, not hide the good and provide the bad.

2.3 Model-data fusion applications: Toward compliance verification and improved climate models (Rapporteur: JD Paris)

Are atmospheric measurements suited to MRV in existing carbon pricing mechanisms? (V Bellassen)

Carbon pricing economy, unperfect but being used in current mecanisms.
Theory of the commons: individual cost < individual benefit < global cost => market alone do not maximize welfare.
Regulation: Pigou (1920): tax marginal cost of pollution. Other: attribute property right on pollution, then trade to reach maximized welfare; but cost of transaction. However intervention of regulator is necessary. Other: in case of uncertainty in benefits and costs, better to regulate quantity or prices
Continent" of compliance: out of a potential of 20-100GER in 2013 (60% EU ETS, 10% Australia, 7% China, 5% California) gets offsets of 200MEUR. On the other hand "island" of voluntary offsets gets 250MEUR.
Toolbox: quantity (market) or price (tax). Application: national scale (national MRV), installation/company scale (e.g. EU ETS), or project scale (CDM). Industry scale is more coactive than national scale (Kyoto is optional)
Theory on which case should regulator care about uncertainty. If monitoring is biased, regulated price is not set optimally. But how much are you ready to pay for better knowledge? If monitoring is unprecise the unprecision is averaged over the number of agents. So same welfare. If regulator is risk averse, large number of sources allow better precision; 50% imprecision on 11000 EU ETS installations = 0.5% over total emissions. Precision of inventories (excl. AFOLU): France 2.8%, US 5% uncertainty on emission factor, it applies equally to all, not regarding quantities of emissions. Not obvious that reducing uncertainty makes economic sense.
If asymmetry in information, possibly interest to reduce uncertainty. Then If scheme voluntary, could care about precision, and apply abatement related to the uncertainty. If not volunteer scheme, reduce uncertainty by prescribing it (e.g. default emissions factors to be imposed)
What do regulators ask for in practice? What are MRV requirements and costs, and trade-offs between both? Is requirement adapted to the emissions considered?
Forestry projects can be reduced in impact (discounted) if uncertainty in sampling (ie. Not if uncertainty from emission factors). Voluntary project: deduction if uncertainty >30%. Uncertainty requirements can be : none, qualitative (categories to be used), quantitative on sampling error, qualitative on total (e.g. EU ETS). But then incentives to reduce uncertainty can be only guidelines or indirect.

The largest the source, the cheapest it is to monitor the source (Bellassen et al , 2015), but cost is also lower for companies than for jurisdictions, due to the fact that participation is mandatory => simpler schemes put into place. So difficult to impose costly MRV on small projects.
Accuracy can be assessed by direct measurements. Atmospheric inversions are not able to reduce uncertainty in the current requirements at relevant cost. The need for priors => extra cost for the inversion itself. Less uncertainty for more money is not desirable.
But possible to change the rules. Easier for projects than for UNFCCC than for ETS. But UNFCCC possibly more influenced by IPCC scientists than ETS. Projects are less and less funded. The issue is that no information asymmetry means that the regulator does not need independent measurements.
Best approach: sectors not included in CPMs: agriculture, waste. Uncertainty of French inventory with AFOLU = 19.8%

Can atmospheric measurements reduce the uncertainty on emissions factors? National inventory and uncertainties? (J Ometto)

Brazil 62% forest, >99% natural. Latin America+ Carriberan: 47% of emissions are land/forestry, 20% agriculture. Brazil reduced drastically emissions by forest and land uyse but continued increase in industry. Emisions from amazon reduce, but increase in Cerrado. Many plots were used to calculate AGB (above ground biomass). Selection of allometric equation crucial
INPES launches program to simulate the evolution of forest.

Ecosystem manipulation experiments (A Chabbi)

Question of ecosystems under changing climate: changes in ecosystems, soils? Soils are generally sampled only in surface. Unclear effect of co2 on canopy temperature and transpiration and heat stress (White et al., 2011). Simulating heat stress in crops improve simulated vs. observed yields agreement. Growing relevance of extreme events. ANAEE (currently preparatory phase) aims at dealing with these challenges.
Case of ecotron in Montpellier, demonstrate the impact of extreme events combined with elevated CO_2 is neutral on carbon storage on this particular ecosystem (Roy et al., in prep for BG).
Science questions emerging: how much CO_2 is produced, stored? Sources of uncertainties?
Experimentations needed to complete ICOS data. Missing sink of 25% of anthropogenic emissions of CO_2? No adequate experimental methodology to explain discrepancy. (?) Chemoautotrophic processes as a sink for CO_2 in the C cycle? What drives water use by ecosystems?

Discussion (Moderator: W Peters)

WP: speakers highlighted the existence of 3 types of potential super users: MRV, biomass, and ecosystem changes.
VB: ICOS data not useful for verification of ETS (large installations). Sectors generating CH_4 and N_2O with high uncertainty may use atmospheric data (inversions estimate), but need it very well pre-digested for non-scientists.
JO: more information at regional scale in open access data sources and capacity building efforts. (issues with access to datasets?) INPE in charge of deforestation analysis disclose all its data to the general public. Open data enhances research.

Would bottom up estimates at relatively high precision like Vulcan be useful for MRV? Again, if cheap it is possible.
Felix: bias could be a problem in the future, we need to have system anticipating this possibility. VB: yes but beware, agents interested in information not asked by regulator might be misled.
Bias can be corrected but this component of uncertainty is difficult to integrate in economic models

2.4 Model-data fusion applications: Toward compliance verification and improved climate models (Rapporteur: H Loescher)

Process Modeling (N MacBean)

Presentation: Data assimilation within the framework of the Orchidee (model). The aim is to optimize the parameterization of the model to be able to deliver prediction of the carbon model along with the uncertainties. Utilize multiple types of data sources to optimize the model parameterization (tower-, Meteorological-, Biomass-, ATM CO2-, Modis NDVI Satellite- etc.) linked with an atmospheric transport model. Also execute MIPs. Use ***cost function*** as the optimization scheme. Local-to-global scale. Optimization is 'best' when using one variable, re, Flux, APAR, FPAR, atmospheric CO2 concentration, rather than some combination thereof. Depending on how the model is optimized, it can provide different predictions, which seem to be a function of soil carbon pools rather than above ground vegetation. Assimilation of above ground biomass reduces the model fit to in-situ measures.
Discussion: (W Peters) to what degree are you allowing the parameterization to change over time? ***A:*** in theory they should be fixed, however it did not improve the optimization, which means that there is a process that is not captured by the model. (D. Bakker) Does this mean that you do not let the model vary with temperature. ***A:*** not explicitly, but temperature variation is expressed at the process level (at several places in the model).

Data Assimilation for GHGs (A. Agusti-Panareda)

Presentation: Use the same weather forecasting data and models to parameterize the transport of the atmospheric CO2 concentrations along with GOSAT (mean total column) and OCO-2 for the CO2 concentrations (Carbon Tracker like model and animation approach). Vegetation, fires, oceans, and anthropogenic sources adjust CO2 surface fluxes. Because they do not use fluxes to parameterize, the model overestimates and is biased. So, then the schema is to re-scale from the flux datasets, by comparing model fluxes to the measured fluxes, re. flux adjustment partition GPP/Re and assume that Re > NEE the dominant partitioned flux in winter and switches seasonally. Has the capability to forecast on the 1-10 day cycle.
Discussion: (A. Vermeulen) what is the status of the anthropogenic fluxes. A: not available from EDGAR in real time, because we do not have attribution parameterization in real time. (W. Kutsch) providing real time data does not mean we can provide the highest quality data. Highest quality data in ICOS will be in 6 months. Can you run a comparison of provisional data verses high quality data? ***A:*** yes. (L. Powers) what have you learned about the biases, given that you had good fit, w/ the exception of Mediterranean sites. ***A:*** The biases are primarily due to poor vegetation model fits. We need to better parameterize the vegetation models to apply to the location and ecosystem. (H.

Loescher) the bias that is examined by the GOSAT, northern altitudes and Amazon is that due to low view angle in the higher latitudes, and cloud cover over the amazon? ***A:*** winter data= snow at higher latitudes and cloud cover over the Amazon.

Inverse modeling for CO2 and MAC III global (F. Chevalier)

Presentation: MACC portfolio (LSCE Atmospheric Inversion). Process and transport models, and Atmospheric CO2 as prior information>> develop a probability distribution of the fluxes, the maximum of the probability distribution = optimization. Spatial resolution 3.75° x 1.9° per week over 35 y. As a group, Transcom does not even agree on the latitudinal distribution of the surface fluxes (noodle plot, Peylin et al. BG 2013). Lessons learned, i) small spatial error correlations, and ii) large temporal error correlations== need better spatial density. We hope that the new satellites will provide the spatial density problem. Data from GOSAT and recent data form OCO2>> preliminary analyses suggest i) 1 Pg emission from Africa which we know is unlikely, ii) source sink status over N America and Europe seem reasonable. The fluxes are driven by the changed retrieval product and the inversion system itself (model bias). Misfits between prior retrieval increments. Technical challenges have buried basic statistical principles. As such, this flaw is exacerbated by the lack of ground truth, and this question has not motivated the inversion community so far, that in turn receive little consideration for its statistical expertise/justification.
Discussion: (W. Peters) We could be doing a lot better on the a priori statistics that inform the retrieval, but should we really do this? A: To get them correct at all levels in the models it is very difficult. Each prior flux that we use has it own specificities. Need to tailor statistics for each a priori condition and model. (A. Vermeulen) What do you think the OCO-2 to improve? ***A:*** better quality of measurements, and higher density of the measurements, overall better products. We will all be on a new learning curve. (A. Vermeulen) what about the ground based measurements? ***A:*** we have TCCON and the issue is around gap filling, and lack of measurements in S. America. We are often trying to examine the smallest of signals from the biosphere, which is very difficult. Q: How do we increase the spatial coverage. ***A:*** We cannot rely on spatial correlations to understand the biases due to spatial coverage alone. Maybe better transport models, but if you have +/- a certain level, re. 0.5 ppm, there are other sources of biases that become more important.

Regional Scale Inverse Modeling (U. Karstens)

Presentation: Showing two examples of regional scale modeling>> re. having higher spatial and temporal scale than that in a global scale model, and hope to resolve small scale variability and reduce model uncertainty and transport error. This will lead to national scale emission inventories for specific countries and used for MRV reporting functions. First example: top down estimation of the EU CH4 (regional inversion) emission. Use of sensitivity studies with different specification of a priori and a posterior error provides future guidance and interpretive power. Comparisons with these inversion results compared to UNFCCC, the latter data was more uncertain over a much broader range. These inversion results (here) are smaller and assist in constraining the emission budget. Second example: Atmos GHG network. Spatial correlation = 50 km, temporal correlation is ~30 days. Calls for more data, more validation, get more realistic data inversions, etc.. How good is good enough?

Discussion: (M. Scholze) question amount the two model comparison, did they have the same error correlations, ***A:*** overall and generally yes, yet there is still some small difference, need to examine…. ? (W. Peters) All are a bit disappointing. Given that we have a large distribution of measurements, and we have 0.5 ppm accuracy, how do we use all these data ***A:*** still have the problem of high resolution models still behave poorly with nighttime data. Currently we focus only on daytime data. (B Dekker) how do you include CH4into national inventories? ***A:*** ?? (L. Rivier) how do you discriminate against other models. Given that different countries are increasing (Italy) and decreasing (Germany) in CH4, how do you discriminate? ***A:*** need more data, possibly adding a radon measurement.

Data Assimilation in Carbontracker (W. Peters)

Presentation: issues include: tropical carbon balance, Vegetation drought responses, gross carbon fluxes, fossil fuel attribution, but given the 20 min time constraint, will focus on vegetation drought responses. O18 seasonal cycles observed at Mauna Loa. If the O18 is lighter, year-to-year, the biosphere is not as robust. Re. if the vegetation is stressed, it is less likely to take up the heavier isotope, and hence discriminate against the heavier isotope. Stomatal closure > changing in WUE, and fractionation changes. Can we see the global C13 estimates as a signal of large-scale changes in WUE? Use Carbontracker (based on flask network), and use the accepted CO2 global budget terms, and add the fractionation derivation. (Assuming C3 plants).
Discussion: (?) what is parameterization in in CASA for plant stress? ***A:*** is wilting point to field capacity, beta of 0.2 to that you have to go very close to witling point before you see a stomata effect. (?) can also demonstration a respiration to VPD fractionation response.

Carbon Cycle Data assimilation (M Scholze)

Presentation: MIP (model inter-comparison) project. Rationale is that there is a large uncertainty in predicative capability form these land carbon data assimilation models. Given the future capability of more flask samples, surface fluxes, remote sensing, and ocean data, we are re-visiting our schema (make use and constrain our process model). Misfit from the observations > transport model> net fluxes (not used)> ingest models that use the flux data to describe the process rates (EC derived functional responses, w/ > 100 parameters in total). Optimized though a cost function. The approach is comparable to N MacBean's CCDAS model/discussion earlier today. CCDAS is a two-step procedure for inferring diagnostics and prognostics of model performance. CCDAS is part of the GEO Carbon project. Summary: Can use CCDAS to evaluate complimentary networks e.g. ecosystem and atmospheric etc. QA: Tabled for discussion after.
Discussion: none

Fossil Fuel data assimilation system (A. McRobert, Peter Rayner)

Presentation: City and country scale anthropogenic CO2 emissions are necessary at high resolution to work toward mitigation. FFDAS uses an ad joint data assimilation method that allows partitioning of cost functions resulting in a stronger statistical fit. DA allows the option to optimally include all sources of data. Data sets: Satellite nighttime lights. World's population data, difference between these two data sets is particularly apparent in developing countries. CO2 = f(CO2 emissions, energy use, gross domestic product, and population).

Discussion: (W. Kutsch) comment; great to have something in parallel to other efforts. How do you resolve the CO2 at a particular place? ***A:*** as a residual, and Energy per GDP, as part of the mitigation efforts are used to decouple these two products? ***A:*** if variable, can pull that out in post analyses diagnostics. (W. Peters) What do you do about validation? A: use the Kevin Gurney Vulcan database, but only achievable in the US. Validation is temporal, spatial and in magnitude. (?) what about planes, transport, emissions?, A: not included.

Session Discussion (moderator: W Peters)

(Y. Luo, question on last talk) In your presentation, you used different models for biogenic GPP (gross primary production) and NPP (net primary production), and show large difference, why is that?, what are the reasons for this difference? ***A:*** the difference systems/models used different data (some using only Atmospheric CO2) and were not constraining GPP.

(W. Peters): opening remarks issues at hand i) do we have the ability to merge weather / climate, CO2 signals and benefits?, ii) what is the future of the parallel MACC III / CAMS efforts?, iii) Is CAMS moving forward to address surface flux re-analyses?, iv) Flux is not correlated over space, but how to tackle the spatial resolution issue?, v) is the flux heterogeneity over scales of 10-100 km problem when we target 'large regions"? , re, are we properly addressing synoptic scales, vi) transport models, are the resolution and correlation structure matter for time, horizontal and vertical dimensions, vii) should we try harder as a community to separate out good models from other less desirable results?

(M Scholze) No data processing/analysis tools are needed, users actually want the data. The data needed should be standardized, quality checked, documented, need to discover the data easily and have information available about their quality.
(W. Kutsch) in GEO have been discussing open data. May need a flagship initiative. Organizing the data streams and optimizing metadata and making it more accessible rather than model development. Standardization. A GEO Carbon Flagship would be a large effort, do not know how to initiate it because it is a different 'team'.
First steps would be possible, developing the larger stakeholder community, connect to the community better, and develop the standardization.

(W. Peters) What are then the user requirements?
(JD Paris) Requirements for a GEO Flagship should involve building the carbon data stream and associated informatics surrounding the easy accessibility of data, attribution of data, persistent identifiers, etc.
(W. Peters) Inter-comparison, produce the results and show the uncertainties. Needs to be searchable, comparable, downloadable, properly documented, standard metadata, controlled vocabularies.
Could be needing 'a place' to do these large scale syntheses, may need a service to place all the data, test the data to meet the requirements, and then provide incentive for the submitter of the data, persistent identifiers (DOI) for broader use and traceability of the data, citable data, adds value, and

that concept should work for all the of the data streams that we have talked about over the past few years.
What would be the first idea to initiate GEO Flagship?
(M Scholze) limited ability for PI that are going from grant to grant, to be able to contribute toward longer term projects, Need operational support.
(W. Peters) World Data centers that may play a role as repositories
(D. Bakker) Repositories are one thing for data, but having a place for models and synthesis results is a very different challenge.
(H. Loescher) Operational support is at the funding agencies. We all have to educate and engage with our program officers to demonstrate the need. To be able to have the conversation with program officers, we have to develop the political will with our sponsors.
(L. Rivier) How do we trust and verify emissions in the context of long-term sustainable support. Particularly if our gap between expectations from supporting agencies and research's progress is not reduced by an improvement in uncertainty of these approaches.
(H. Loescher) Do not know if this is going to answer the question, but one of the facets of this issue is that the science community communicate very poorly (including about uncertainty). There has to be a cultural change. For example, how, we as a community, discuss uncertainty and risk, is very different HL communication and cultural change (uncertainty and risk argument). We need to be able to develop the skillsets to communicate what we mean by our data (incl risk and uncertainty) to sponsors, decision makers and the public, alike.
(JD Paris) Yes, agreed, but what we discuss here, is a very complex issue to TRY to communicate forward.
(F. Chevalier), How best populate the community b (what are the timescale lag time to produce the results for policy makers decision maker? Shorter lead time = less accuracy. Second question is how best validate and verify. I doubt very much that we could potentially verify all, we may be able to verify some parameters (that may or may not be used by a model) by independent satellite data. A) have very high standards, b) continue to bring people together to have as an ongoing discussion (data providers, and end users), c) more statistically rigorous approaches that are supported by the community.
(A. Vermeulen) Our methods would be more relevant in the future when verification by 2050 is 50% reduction, as opposed to only 8% today. We need to prepare a baseline for this, and prepare our stakeholders. The message to our stakeholders should not necessarily be that we can 'already do this'….
(W. Kutsch) maybe our view is too narrow on the analyses and the fluxes numbers, maybe our models can be used for different mitigation options.

3 CONCLUSIONS: SCIENCE AND LESSONS FOR COOPEUS

Model data fusion: CO2 forecast, reanalysis, regional simulations to urban scale fluxes, use of DA for drought impact research. To be appreciated in the light of final societal impact: emissions reduction mechanisms, refinement of coupled carbon earth system models.

All this research has strong need in in situ monitoring, ancillary data, satellite observations. Efforts are consistently difficult to sustain for research groups.

Interest for easy discoverability of diverse data rather than unique access. However users have limited need for informatics services and tools (they use FTP, advanced access SOCAT system less used than synthesis…) but require timely access to data that should be shared (no holding back data by PIs), need access to legacy data, need specification on the quality of the data that they can appreciate. Open data access need to be not only promoted but also forced by journals and funding agencies.

Economy, inventories: no economic value per se for verification for regulators

Outreach: Not verify emissions, but either verifies overall impact of policies or find niches. For example, agricultural emissions of CH4 and N2O, waste management. Set up dynamics toward extending the perimeter for verification where we can make a difference. Finally we need to elaborate a better message on uncertainty in our work that is compatible with non-scientific vocabulary.

Key data management and data policy related issues that were discussed include:

- **Legacy data strategy: how to agree on easy to use data management and policy? How to involve PIs of legacy data?**
- **Lack of systematically-identified, well-adapted data repositories**
- **Embargo (Even though everybody agrees on open data sharing principles, the actual data sharing is limited by PI embargo, explicit or implicit, in 'fear' of acknowledgement or competition issues.)**

Other key conclusions:

- Working groups should be continued and developed to advance data sharing and high level modeling data products
- RDA may be a forum to prototype these activities
- Input towards data sharing: there is a mismatch between informatics' potential for users and the actual practice of people who use our data, which includes issues and limitation related to data sharing and legacy datasets. Accessibility and discoverability of the data is considered as a valuable way forward. No need for high power computing capability, as this is user end, if needed.
- What we do well is to grab data and use it for our own projects. A useful progress would be to develop a real inventory (catalog) of the data that is currently out there.

- PI based data, curated, and lacks data quality, there should be a means to accommodate and make that data available, but flagged accordingly, such that it is distinguished from high quality, QA'd, NEON and ICOS data.
- How do we make our data usable for new users that are unaware of the nature (and uncertainty related to) of our data? This is still an unknown.
- It has been noted that students and early career scientists are thinking about the accessibility and discoverability in very new ways, not necessarily in the same way as their mentors and professors.
- Adoption of more open Data policy and harmonization of data policies still an ongoing issue (embargo is still a practice, not necessarily fought).
- Clarification of the nature of the relationship between ICOS (NEON) and fluxnet datasets.
- Tool Development to easily compare model results at the Carbon portal(s) is useful
- *Data sharing* - Data sharing limited by current practice. Needs to be fostered by journals/funders
- *Data workflow* - Thinking out of the box: reorganizing workflow

Follow-up Actions:

- ***User engagement*: organize dedicated session at either AGU or EGU meetings on global integration of GHG observations**
- ***Technical progress*: data providers to be involved at high level in RDA and in the GEO flagship**

4 ACKNOWLEDGEMENTS

We thank the speakers for contributing their time, efforts and expertise to this workshop. We acknowledge the crucial contribution of rapporteurs, chairs and moderators as well as support personnel (C. Milcent) for the overall success of this task.

Participant list

#	Name	Surname	Institute	Country
1	Agusta-Panareda	Anna	ECMWF	UK
2	Bakker	Dorothee	UEA	UK
3	Bellassem	Valentin	INRA	France
4	Breon	Francois-Marie	LSCE	France
5	Broquet	Gregoire	LSCE	France
6	Chabbi	Abad	INRA	France
7	Chevallier	Fred	MACC-III, LSCE	France
8	Ciais	Philippe	CEA/LSCE	France
9	DeCola	Phil	Sigma Space	US
10	Delahaies	Sylvain	University of Surrey	UK
11	Delire	Christine	CNRS / Meteo France	France
12	Gielen	Bert	University of Antwerp	Belgium
13	Henne	Stephan	EMPA	Switzerland
14	Jerkovic	Gorana	University of Helsinki	Finland
15	Karstens	Ute	MPG	Germany
16	Kutsch	Werner	ICOS HQ, University of Helsinki	Finland
17	Lo Monaco	Claire	LOCEAN IPSL	France
18	Loescher	Hank	NEON	US

19	Luo	Yiqi	University of Oklahoma	US
20	MacBean	Natasha	LSCE	France
21	Maenhuout	Greet	JRC, Edgar	Italy
22	McRobert	Andrew	University of Lund	Sweden
23	Milcent	Catherine	CEA/LSCE	France
24	Ometto	Jean	CCST/INPE	Brazil
25	Paris	Jean-Daniel	CEA/LSCE	France
26	Peters	Wouter	Wageningen University	The Netherlands
27	Powers	Lindsay	NEON	US
28	Ramonet	Michel	LSCE	France
29	Rivier	Leo	LSCE	France
30	Schneider	Nadine	CEA/LSCE	France
31	Scholze	Marko	University of Lund	Sweden
32	Vermeulen	Alex	University of Lund	Sweden
33	Vogel	Felix	LSCE	France