Atmos. Chem. Phys. Discuss., 14, 11301–11342, 2014
www.atmos-chem-phys-discuss.net/14/11301/2014/
doi:10.5194/acpd-14-11301-2014

Atmospheric Chemistry and Physics Discussions

Greenhouse gas network design using backward Lagrangian particle dispersion modelling – Part 2: Sensitivity analyses and South African test case

A. Nickless[1,2], **T. Ziehn**[3], **P. J. Rayner**[4], **R. J. Scholes**[1], **and F. Engelbrecht**[5]

[1]Global Change and Ecosystem Dynamics, CSIR, Pretoria, 0005, South Africa
[2]Department of Statistical Sciences, University of Cape Town, Cape Town, 7701, South Africa
[3]Centre for Australian Weather and Climate Research, CSIRO Marine and Atmospheric Research, Aspendale, VIC 3195, Australia
[4]School of Earth Sciences, University of Melbourne, Melbourne, VIC 3010, Australia
[5]Climate Studies and Modelling and Environmental Health, CSIR, Pretoria, 0005, South Africa

Received: 24 March 2014 – Accepted: 7 April 2014 – Published: 7 May 2014

Correspondence to: A. Nickless (anickless@csir.co.za)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

This is the second part of a two-part paper considering network design based on a Lagrangian stochastic particle dispersion model (LPDM), aimed at reducing the uncertainty of the flux estimates achievable for the region of interest by the continuous observation of atmospheric CO_2 concentrations at fixed monitoring stations. The LPDM model, which can be used to derive the sensitivity matrix used in an inversion, was run for each potential site for the months of July (representative of the Southern Hemisphere Winter) and January (Summer). The magnitude of the boundary contributions to each potential observation site was tested to determine its inclusion in the network design, but found to be minimal. Through the use of the Bayesian inverse modelling technique, the sensitivity matrix, together with the prior estimates for the covariance matrices of the observations and surface fluxes were used to calculate the posterior covariance matrix of the estimated fluxes, used for the calculation of the cost function of the optimisation procedure. The optimisation procedure was carried out for South Africa under a standard set of conditions, similar to those applied to the Australian test case in Part 1, for both months and for the combined two months. The conditions were subtly changed, one at a time, and the optimisation routine re-run under each set of modified conditions, and compared to the original optimal network design. The results showed that changing the height of the surface grid cells, including an uncertainty estimate for the oceans, or increasing the night time observational uncertainty did not result in any major changes in the positioning of the stations relative to the basic design, but changing the covariance matrix or increasing the spatial resolution did. The genetic algorithm was able to find a slightly better solution than the incremental optimisation procedure, but did not drastically alter the solution compared to the standard case. Including correlation appeared to increase the spread in the layout of the stations. Increasing the surface resolution tended to clump the stations around areas of high activity. In conclusion, the specification used in an optimal network design should be chosen to best match the conditions under which an inversion would be run for the

ACPD
14, 11301–11342, 2014

region of interest. Increasing the spatial resolution beyond that which the given network size could reasonably resolve may lead to a network which would ignore small areas of high activity and reduce the capacity of the network to resolve fluxes for sub-regions in the domain of interest. Overall the results suggest that a good improvement in knowledge of South African fluxes is available from a feasible atmospheric network and that the general features of this network are robust to many reasonable choices in a network design study.

1 Introduction

It has become essential to accurately estimate the emission and uptake of CO_2 around the globe. Greenhouse gases affect the absorption, scattering and emission of radiation within the atmosphere and at the Earth's surface (Enting, 2002; Denman et al., 2007). Of these gases, CO_2 contributes the greatest forcing on the climate (Canadell et al., 2007). Monitoring CO_2 sources and sinks is necessary for determining the best course of action to mitigate Climate Change. The method of inverse modelling can be used to estimate the magnitude of sources and sinks of CO_2 at different temporal and spatial scales (Enting and Mansbridge, 1989; Rayner et al., 1999; Rödenbeck et al., 2003; Chevallier et al., 2010). This method relies on precision measurements of atmospheric CO_2 to refine the prior estimates of the fluxes. Using this theory, an optimal network of new measurement sites for atmospheric CO_2 concentrations from a list of potential sites can be derived.

Measurements over Africa are much sparser compared to other continents. Only the Cape Point Global Atmospheric Watch (GAW) station has a long term CO_2 concentration record, measuring since 1992. This tower was located at Cape Point (34.35° S, 18.49° E) predominantly to record baseline measurements of well-mixed, clean air originating over the Southern Ocean. A study by Whittlestone et al. (2009) demonstrated that it would be difficult to improve estimates of terrestrial CO_2 fluxes for southern Africa using the Cape Point station alone. In 2012, an atmospheric observatory was in-

ACPD
14, 11301–11342, 2014

stalled near the Gobabeb Training and Research Centre, on the west coast of Namibia (22.55° S, 15.03° E), which continuously measures trace gases, including CO_2 (Morgan et al., 2012). To build on this rudimentary existing network, and to improve estimates of CO_2 fluxes at least for South Africa, high precision instruments for measuring CO_2 concentrations have been purchased, and are to be installed at sites across South Africa. To maximise the impact of these stations on the estimation of CO_2 over South Africa, an optimal network design can indicate the best placement of new stations to reduce the uncertainty of the source and sink estimates. Part 1 of this paper conducted a similar study for Australia on how to improve its current observation network (Ziehn et al., 2014).

An optimal network design has two basic requirements: an inversion algorithm, which is used to calculate the quantity which is to be optimised, and an optimisation procedure, for selecting between possible elements in the network (Rayner, 2004; Lauvaux et al., 2012). Part 1 of this paper sought to reduce the uncertainty of Australian terrestrial fluxes by 50 %, and began by considering the addition of new stations to the existing base network (Ziehn et al., 2014). Similarly, the Cape Point and Gobabeb stations make up a base network of CO_2 monitoring stations for southern Africa, and this optimal network design will provide a theoretical study on the optimal locations of new stations within South Africa. The optimal network will add five measurement stations to best reduce the uncertainty in flux estimates across the country, and under the assumption of continuous, hourly measurements of CO_2 concentrations.

The score index for the optimisation procedure is based on the posterior covariance matrix estimated through the inversion. The calculation of the posterior covariance matrix, which describes the uncertainty of the estimated fluxes from the Bayesian inversion procedure, does not require any knowledge of the measured concentrations or of the prior fluxes, only on the prior covariance matrix of the fluxes, the covariance matrix of the concentrations, and the transport matrix, which are all determined independently. By basing the cost function of the optimisation procedure on the result of the posterior covariance matrix for a given network, this score can be optimised such that the uncer-

ACPD
14, 11301–11342, 2014

tainty in the estimated fluxes is reduced. Two non-parametric optimisation procedures will be considered to optimise the cost function, namely the genetic algorithm (GA) and the incremental optimisation (IO) procedure.

In addition to providing an optimal network design for South Africa, this paper will consider a number of sensitivity analyses of the parameters and choices which are necessary in running an optimal network design. In order to run the transport model to generate the influence function required in the inversion, a number of arbitrary choices in parameter values of the Lagrangian Particle Dispersion Model (LPDM) are necessary, such as the choice of the height of the surface layer and the size of the surface grids during the post-processing of the LPDM results (Lauvaux et al., 2012). In the construction of the covariance matrix for the prior fluxes, just as for a normal inversion procedure, choices need to be made regarding the size of the variances and the correlation lengths (Wu et al., 2013). Optimal network designs will be run under subtle changes to these choices to determine the impact this has on the choice of placement of new measurement stations.

2 Methods and the South African test case

2.1 Surface flux inversion

The Bayesian synthesis inversion method, first proposed by Tarantola (1987) and used for the network design in this paper, is the most commonly used formulation in the literature (Rayner et al., 1996, 1999; Bousquet et al., 1999; Kaminski et al., 1999; Gurney et al., 2002, 2003; Peylin et al., 2002; Law et al., 2003; Baker et al., 2006; Ciais et al., 2010; Enting, 2002). The inversion method is explained in more depth in the first part of the paper (Ziehn et al., 2014) The observed concentration (c) at a measurement station at a given time can be expressed as the sum of different contributions from the surface fluxes, from the boundaries and from the initial concentration at the site. For the purposes of the network design, the initial concentrations are ignored since it is

ACPD
14, 11301–11342, 2014

assumed that these concentrations are well constrained by the observations and therefore contribute very little to the flux uncertainty. The contribution from the boundaries has to be assessed and if the influence on the flux uncertainties is not negligible then the boundary conditions have to be included in the network design process.

A simple linear relationship can be used to describe the relationship between the modelled concentrations and the contribution from the sources (surface fluxes and boundary inflow):

$$\boldsymbol{c}_{\mathrm{mod}} = \mathbf{T}\boldsymbol{f} \tag{1}$$

The vector of the modelled concentrations $\boldsymbol{c}_{\mathrm{mod}}$ is a result of the contribution from the sources $\boldsymbol{f}$, described by the transport or sensitivity matrix $\mathbf{T}$. In earlier inversion exercises, $\mathbf{T}$ was referred to as Green's function and was the Jacobian matrix representing the first derivative of the modelled concentration at the observational site and dated with respect to the coefficients of the source components (Enting, 2002). If we assume a Gaussian error distribution for the surface fluxes and concentrations we obtain the following cost function for our least square problem:

$$J(\boldsymbol{f}) = \frac{1}{2}\left((\boldsymbol{c}_{\mathrm{mod}} - \boldsymbol{c})^T \mathbf{C}_{\mathrm{c}}^{-1}(\boldsymbol{c}_{\mathrm{mod}} - \boldsymbol{c}) + (\boldsymbol{f} - \boldsymbol{f}_0)^T \mathbf{C}_{f_0}^{-1}(\boldsymbol{f} - \boldsymbol{f}_0)\right) \tag{2}$$

where $\mathbf{C}_{\mathrm{c}}$ is the error covariance matrix of the observations, $\boldsymbol{f}_0$ is the vector of prior flux estimates, $\boldsymbol{f}$ is the vector of predicted fluxes and $\mathbf{C}_{f_0}$ is the prior error covariance matrix of the surface fluxes. The Bayesian cost function minimizes both the difference between modelled concentrations and measurements and the difference between prior flux estimates and predicted fluxes.

As described in (Ziehn et al., 2014), for the network design approach we are only interested in the posterior covariance matrix of the predicted fluxes, since our aim is to obtain a network that reduces the CO_2 flux uncertainties. The posterior covariance

ACPD
14, 11301–11342, 2014

matrix can be calculated as follows (Tarantola, 1987):

$$\mathbf{C}_f = \left(\mathbf{T}^T \mathbf{C}_c^{-1} \mathbf{T} + \mathbf{C}_{f_0}^{-1}\right)^{-1} \quad (3)$$

$$= \mathbf{C}_{f_0} - \mathbf{C}_{f_0} \mathbf{T}^T \left(\mathbf{T}\mathbf{C}_{f_0}\mathbf{T}^T + \mathbf{C}_c\right)^{-1} \mathbf{T}\mathbf{C}_{f_0} \quad (4)$$

The use of the posterior covariance matrix is possible because it is obtained without the vector of observed concentrations $\boldsymbol{c}$ or the vector of prior fluxes $\boldsymbol{f}_0$, which means that it is possible to assess the contribution that a hypothetical station can have on the reduction of the flux uncertainty without the need to generate synthetic data or make unnecessary assumptions about the measurements. It only depends on the transport model, the prior flux uncertainties and observational uncertainties.

2.2 Prior covariance matrices

The elements of the prior covariance matrix need to be constructed from the best available knowledge of sources and sinks on the surface and at the boundaries. Lauvaux et al. (2008) carried out a mesoscale inversion on synthetic data and their approach was to obtain the elements of the observational covariance matrix by comparing modelled values of CO_2 to aircraft and tower data that was available for the four day period under assessment. They found the largest difference to be approximately 3 ppm, and set the diurnal observation error to be 4 ppm, taking into account uncertainty in LPDM. For the prior covariance matrix of the fluxes, the error was set at $2\,\mathrm{g\,C\,m^{-2}\,day^{-1}}$ for the surface and 4 ppm for the boundaries, and they assumed uncorrelated flux errors on the domain (no spatial correlation). This was further developed by Wu et al. (2013) who used available data to fit hyperparameters, which were the variance and correlation lengths of the prior flux and observational error matrices.

The approach of Chevallier et al. (2010) to obtain the elements of the prior flux covariance matrix was to set the standard deviations of the fluxes proportional to the hetrotrophic respiration flux of land-surface model ORCHIDEE. This is the approach

ACPD
14, 11301–11342, 2014

adopted for the variance elements of the covariance matrix in the case of the South African optimal network design, where we use a recent carbon assessment study by Scholes et al. (2013) which produced monthly maps of gross primary productivity (GPP), net primary productivity (NPP), hetrotrophic respiration (Rh), autotrophic respiration (Ra) and net ecosystem productivity (NEP) for South Africa. Of these products, most confidence lay in the GPP product. Since there was more confidence in the NPP estimates than in the Rh estimates, and since NEP = NPP – Rh and in a balanced system NEP should be a small flux (Lambers et al., 2008), NPP was used rather than Rh. Following Chevallier et al. (2010), the biosphere flux uncertainties for a particular month are estimated using the following simple relationship:

$$\sigma_{\mathrm{NEP}} = \begin{cases} \min(28\,\mathrm{g\,C\,m^{-2}\,week^{-1}}, \mathrm{NPP}) & \text{if South Africa} \\ \min(28\,\mathrm{g\,C\,m^{-2}\,week^{-1}}, \max(\mathrm{NPP})) & \text{if not South Africa} \end{cases} \quad (5)$$

As a conservative assumption, areas outside of South Africa, where there were no estimates available for NPP from the carbon assessment product, were assigned the maximum value obtained in South Africa for that particular month. This would also have the added benefit of increasing the importance of measurements near the terrestrial borders of South Africa so that the final network can give an adequately resolved solution for South Africa. The carbon assessment product produced monthly outputs for all the products. These products were converted into daily values. Since Ra and GPP were also available, and NPP = GPP – Ra, day time NPP and night time Ra were obtained by assuming that all the GPP took place during the day, and half of the Ra took place during the day and half at night. This meant that the day time NPP values tended to be larger in magnitude than the night time Ra values, which is what we would expect for the South African systems. The daily values are accumulated to one week to give the final uncertainty values used to construct the prior flux covariance matrix. The day time NPP and night time Ra values used to obtain the NEP uncertainties are plotted for July and Januray (Fig. 1). In South African systems it is expected that much more

ACPD
14, 11301–11342, 2014

activity occurs during the Summer months compared to the Winter months, with the consequence that the uncertainty during the Summer months is considerably larger.

Chevallier et al. (2010) assumed that the fluxes were not independent, but included temporal correlations which decayed exponentially with a length of one month, and that spatial correlations followed an e-folding length of 1000 km. As in Part 1 (Ziehn et al., 2014), the covariance matrices for the observations and prior fluxes are assumed to be diagonal for the basic network design, but during the sensitivity analysis, the impact of the off-diagonal elements is considered.

Since the domain of the network design includes the fossil fuel sources of South Africa, fossil fuel uncertainties need to be derived as well. As for the Australian test case (Ziehn et al., 2014), these uncertainties were derived from the Fossil Fuel Data Assimilation System (FFDAS) (Rayner et al., 2010). Ten realisations for the year 2010 were obtained from the FFDAS product at the original resolution of $0.1^\circ \times 0.1^\circ$. The fluxes were aggregated to our network design resolution of $1.2^\circ \times 1.2^\circ$ and then the variance calculated for each pixel. Since the emissions from fossil fuels are usually localised, such as those at power plant locations, the variability in the fossil fuel emissions is quite large. But, as shown by Ziehn et al. (2014), the effect of aggregating the data smooths the fossil fuel emissions over the network design domain, and this leads to a reduction in the variability between the different realisations of the FFDAS. Figure 2 shows that the uncertainties for the ten realisations based on the original $0.1^\circ \times 0.1^\circ$ resolution have much larger maximums for individual grid cells than the uncertainties calculated for the aggregated fluxes. The effect of using a higher spatial resolution for the surface grids is considered in the sensitivity analyses.

Finally, for the basic network design, the prior error covariance matrix of the land surface fluxes are estimated as the diagonal matrix, where the diagonal elements are the sum of the variances of the biospheric fluxes and the fossil fuel emissions for that pixel, multiplied by the fraction of the pixel covered by land, separately for day and night. By multiplying with the land fractions we guarantee that the prior uncertainties for coastal grid cells are scaled accordingly and ocean only grid cells are set to zero,

ACPD
14, 11301–11342, 2014

since the NEP and fossil fuel products only apply to the land surface. In the network design under the standard case, we keep the uncertainties of the ocean-only pixels set to zero, since our focus is on reducing the flux uncertainty over land.

Observational uncertainties are set to 2 ppm for all existing and potential stations. The actual measurement uncertainty at the sites has a much smaller uncertainty, but this is conservatively increased to account for uncertainty in the transport model. Again, the uncertainties are specified in terms of their standard deviation and we assume no correlations among the uncertainties of different observations. In this way $\mathbf{C}_c$ also becomes a diagonal matrix. In the sensitivity analysis, a doubling of the night time uncertainty in the observations is considered.

2.3 Lagrangian Particle Dispersion Model (LPDM)

To determine "where" and "how much of" a particular source a measurement site sees at a given moment, the influence function or transport matrix **T** is required. This matrix can be directly obtained from running a Lagrangian Particle Dispersion Model (LPDM) in backward mode. An LPDM simulates the release of a large number of particles from arbitrary emissions sources by tracking the motion of the particles (Uliasz, 1993, 1994). The model can be run backward in time, in receptor-orientated mode, to calculate the influence functions for a given receptor, as described in Ziehn et al. (2014). In this mode, the particles are released from the measurement locations and travel to the surface and the boundaries (Lauvaux et al., 2008; Seibert and Frank, 2004).

LPDM is driven by mean horizontal winds (u, v), potential temperature, and turbulent kinetic energy (TKE). In the case of the South African network design, these variables are produced by the CSIRO Conformal-Cubic Atmospheric Model (CCAM), a global circulation model. CCAM is a two time-level semi-implicit hydrostatic primitive equation developed by McGregor (1987) and later described by McGregor et al. (2001), which utilises semi-Lagrangian horizontal advection with bi-cubic spatial interpolation, with total-variation-diminishing vertical advection (Engelbrecht et al., 2009; Whittlestone et al., 2009). In the South African case, the CCAM has been setup so

ACPD
14, 11301–11342, 2014

that it produces output at an hourly time step and at a 0.15° spatial resolution (approximately 15 km by 15 km) over South Africa. The domain extends far beyond the South African border, from 10° S to 40° S and from 0° W to 60° E. Meteorological inputs for LPDM are generated for two months; January in Summer and July in Winter. For a four week period during each of these months, the LPDM is run for each of the hypothetical measurement sites.

We use the LPDM originally proposed by Uliasz (1994), which we run in reverse mode for each hypothetical measurement station we would like to include in the network design process. In our setup for the model, particles are released every 20 s for a total of four weeks for two example months and the particles position is recorded in 15 min intervals. Particles that are near the surface are counted for each grid cell to determine the surface influence or sensitivity. The determination of when a particle is near the ground will form part of the sensitivity analyses conducted for this paper. These particle counts are used to calculate the source–receptor (s–r) relationship which forms the transport matrix **T**. Here, we follow Seibert and Frank (2004) to derive the elements of that matrix. As described in Ziehn et al. (2014), we modify the approach of Seibert and Frank (2004) to account for the particle counts which are produced by our LPDM model as opposed to the mass concentrations which are output by the LPDM used by Seibert and Frank (2004). The resulting s–r relationship between the measurement site and source i at time interval n, which provide the elements of the matrix **T**, is:

$$\frac{\partial \bar{\chi}}{\partial \dot{q}_{\text{in}}} = \frac{\Delta T \, \text{g}}{\Delta P} \overline{\left(\frac{N_{\text{in}}}{N_{\text{tot}}}\right)} \frac{29}{12} \times 10^6, \tag{6}$$

where $\bar{\chi}$ is a volume mixing ratio (receptor) expressed in ppm and $\dot{q}_{\text{in}}$ is a mass flux density (source), N_{in} the number of particles in the receptor surface grid from source grid i released at time interval n, ΔT is the length of the time interval, ΔP is the pressure difference in the surface layer, g is the gravity of Earth, and N_{tot} the total number of particles released during a given time interval.

ACPD
14, 11301–11342, 2014

For the network design we are interested in weekly fluxes of carbon divided into day and night time contributions, which means that we have to provide the particle count N_{in} as the sum over one week (ΔT = 1 week (day/night)). Therefore, the mass flux density $\dot{q}_{in}$ in Eq. (6) has units of $kg\,C\,m^{-2}\,week^{-1}$ for the day and similarly for the night.

For the basic network design, the surface layer height is set to 50 m which corresponds to approximately 595 Pa (ΔP). If we assume well mixed conditions, then the source–receptor relationship should be independent of the thickness of the surface layer as long as the layer is not too deep (Seibert and Frank, 2004).

2.4 Influence from outside the modelled domain

Since the surface sources are expressed as fluxes in carbon, the contribution to the concentration at the measurement site is expressed in the amount of carbon seen at the measurement site from a particular source. In the case of the boundary sources (or contributions from outside of the domain) which are given as concentrations, their contributions to the concentration at the measurement site are expressed as a proportion of their concentration, dependent on their influence at the receptor site. Part 1 (Ziehn et al., 2014) shows that by calculating the Jacobian which provides the sensitivities of observed concentrations to boundary concentrations, the boundary contribution can then be written as:

$$\boldsymbol{c}_b = \mathbf{M}_B \boldsymbol{c}_B \tag{7}$$

where $\mathbf{M}_B$ is the Jacobian and $\boldsymbol{c}_b$ the boundary concentrations. If the elements of $\mathbf{M}_B$ are large enough it may be necessary to include the boundary conditions in the network design.

For the network design, four boundaries (north, south, east and west) were used and we calculated the sensitivity of hourly observed concentrations to weekly boundary concentrations. To determine if the boundary influence should be included in the network design, we need to know whether the uncertainty contributed by the boundary concentrations is significant compared to other contributions. To see this we calculate

ACPD
14, 11301–11342, 2014

$\mathbf{M_B}$ for each station. Assuming concentration uncertainties of 1 ppm at the boundary (reasonable for the Southern Hemisphere) this yields:

$$\mathbf{C_b} = \mathbf{M_B C_I M_B^T} \tag{8}$$

where $\mathbf{C_I}$ is the identity matrix. The diagonal elements of $\mathbf{C_b}$ provide us with the uncertainty contribution of the boundary concentrations to the uncertainty of simulated concentration at each station. If they are much smaller than the observational uncertainty we can neglect boundary influences in the network design.

Using the same idea, but calculating the sensitivities in $\mathbf{M_B}$ as described in Sect. 2.1, a test can be carried out to determine the influence on a station from ocean pixels or terrestrial pixels outside of South Africa.

2.5 Optimisation and ground based measurement stations

The aim of the network design is to find the set of stations that minimizes the flux uncertainty over a certain region. Currently the network for southern Africa consists of just two stations, one at Cape Point and the other at Gobabeb. Hypothetical stations were selected from a regular grid over South Africa, resulting in 36 equally spaced locations (Fig. 3). Five new instruments are potentially available for deployment. Ultimately, the exact location of the stations will be determined by practical considerations, such as making use of existing infrastructure and manpower, the relative safety of the instruments, and the accessibility of the sites.

The optimisation of the network is achieved by calculating the transport matrix $\mathbf{T}$ for the hypothetical stations and constructing the posterior covariance matrix $\mathbf{C_f}$ as described above. Then based on the results of an optimisation routine, the set of five stations which best reduces the posterior uncertainty are selected as the optimal network. Two alternative optimisation routines have been used in the literature. The incremental optimisation (IO) routine, as used for the Australian network design (Ziehn et al., 2014), is used for the basic network design. To determine the impact of the choice of

ACPD
14, 11301–11342, 2014

optimisation routine, the genetic algorithm (GA) is also considered during the sensitivity analyses.

The overall uncertainty in fluxes can be expressed in two different metrics. Either through obtaining the trace of $\mathbf{C}_f$ (cost function J_{Ct}) or by summing over all the elements of $\mathbf{C}_f$ (cost function J_{Ce}):

$$J_{Ct} = \sqrt{\sum_{i=1}^{n} C_{f_{ii}}} \tag{9}$$

$$J_{Ce} = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n} C_{f_{ij}}} \tag{10}$$

where n is the number of elements in the diagonal of $\mathbf{C}_f$. In the first case we consider only the uncertainty of the fluxes estimated at the source regions, aiming to minimise the average uncertainty across source regions. In the second case, the uncertainty of the total flux estimate of the target region is considered, since the variance of the sum of correlated variables is equal to the sum of the covariances and variances of these variables. There is no clear answer as to which of these is the best cost function for the determination of overall uncertainty reduction, so as for Part 1 (Ziehn et al., 2014), we use $\mathbf{J}_{Ce}$ as the cost function for the basic design, and we then have a case under the sensitivity tests where $\mathbf{J}_{Ct}$ is used instead. In the South African test case, because the domain contains terrestrial regions outside of South Africa, we only include the elements of $\mathbf{C}_f$ which are within South Africa.

In the IO scheme we obtain the s–r relationship for each of the hypothetical stations, which represents the sensitivity of hourly observations to weekly fluxes, and then construct $\mathbf{C}_f$. We add one station at a time from the candidate list to our base network of two stations and calculate $\mathbf{C}_f$. We choose the station that gives us the smallest cost function value and add it to the network, simultaneously removing it from the candidate

ACPD
14, 11301–11342, 2014

list. We then repeat the process until we reach the number of instruments we have available (five).

For the basic network design, we use a resolution of $1.2^\circ \times 1.2^\circ$ (i.e. resolution for the flux uncertainties) and consider weekly fluxes divided into day time (84 h a week) and night time (84 h a week). We also assume that observations (CO_2 concentration measurements) will be available from all stations from the candidate list and the base network at an hourly time scale.

We evaluate the different networks in terms of the uncertainty reduction:

$$U_R = 1 - \frac{\hat{J}_{\mathrm{Ce}}}{J_{\mathrm{Ce\ base}}} \tag{11}$$

where $\hat{J}_{\mathrm{Ce}}$ is the optimal cost function value and $J_{\mathrm{Ce\ base}}$ the cost function value based on the posterior uncertainties if just the base stations are included.

Although IO is expected to be more computationally efficient, optimisation through a GA would also be well suited for this kind of problem, considering that this network design for South Africa is starting essentially from scratch. The GA is highly parallel and we can therefore take advantage of high performance computing.

The approach used to run the GA during the sensitivity analyses is adopted from Rayner (2004). GA procedures are a class of stochastic optimisation procedures for any numerical algorithm which calculates a score based on a function of inputs. In this case the algorithm calculates a score based on the uncertainty in the estimated fluxes, given a set of stations. A genetic algorithm does not optimise based on a single solution, but rather by improving a population of solutions, from which a final solution is selected. New members are added to the population through a process of loss of members which are not sufficiently fit (culling), pairwise combination of previous members (cross-over), and random changes to members (mutation). This represents "survival of the fittest" and pairwise reproduction and mutation in biological populations.

In this implementation of the genetic algorithm elitism is maintained by keeping the best solution from the previous population, without making any changes through cross-

ACPD
14, 11301–11342, 2014

over or mutation on this member. The algorithm converges once a given number of iterations is reached, or once a convergence criterion is met. The solution with the best fitness criterion is then selected from this population, where the fitness F is calculated as:

$$F = 1 - \frac{r - 0.5}{N} \quad (12)$$

where r is the ordinal ranking of the member and N is the population size, which in the South African test case was taken to be 50 members. A pseudorandom number x is generated and members are then deleted, or culled, if the value of F is less than the x. The culling process will remove about 50 % of the population members. These need to be regenerated to get the population back to the required size. Members are selected at random from the remaining population, and based on new pseudorandom number, members are duplicated if their fitness score is above this random number. Sampling is with replacement, so the members with the highest fitness have a good chance of being duplicated more than once. This continues until all the culled members have been replaced and the population size is back to N.

The genetic algorithm requires a trade-off between the diversity in the solutions, ensuring that the algorithm does not get stuck in local extrema, and strong enough selection to ensure that the population moves towards the optimum solution. This is achieved by adjusting the mutation rate – high enough to produce diversity in the solutions, but low enough to ensure that members with high fitness persist and so ensure a tendency towards the optimum solution. From previous work (Rayner, 2004) a good mutation rate for network design was found to be 0.01.

The population size and number of iterations affects the computation time of the algorithm. A large population size is favourable because this ensures diversity in the solutions. The more iterations that take place, the more solutions the algorithm can assess and the better the chance of finding the global minimum. High values for both of these parameters results in long computation times.

ACPD
14, 11301–11342, 2014

3 Sensitivity analysis

Since the output of the network design could be significantly altered by making different choices in the setup of the transport model, inversion and prior information, the second part of this paper considers the sensitivity of the network design to alternative choices.

3.1 Surface height

In the post-processing step of the LPDM runs, the particle counts are determined for a given spatial and temporal resolution. In addition to the *x y* dimensions of the surface grid boxes, the height of the surface box (or thickness of the surface layer) also needs to be specified. In previous uses of this LPDM model (Lauvaux et al., 2008, 2012), the height has been set at 50 m. Under well mixed conditions, the height of the surface grid is not important, as the particle count will be adjusted proportional to the volume of the grid box, but under stable conditions, this may not be the case. To investigate the importance of this choice on the optimal network design, the LPDM output for each station was reprocessed under a surface height of 60 m and 75 m. The optimisation routine was run again for both cases, holding all other choices the same as for the basic network design.

3.2 Night time observations

It is a known issue in atmospheric transport modelling that the transport at night has greater uncertainty compared to the day time. Since the transport model is not assigned a covariance matrix, the uncertainty is transferred to the observations. To account for larger uncertainties at night, the standard deviation value for the night time observations was doubled to 4 ppm.

ACPD
14, 11301–11342, 2014

3.3 Cost function

This sensitivity test compares the results of the network under the $\mathbf{J}_{Ce}$ cost function against those under the $\mathbf{J}_{Ct}$ cost function. The sum of all of the covariance terms represents the variance of the total flux estimate for South Africa (the sum of the fluxes from all the surface grids across South Africa). The sum of the diagonal elements implies that we're seeking to optimise the network based on improvement in the estimation of fluxes from individual source grids independently from one another. Either case should result in an optimal network for improved flux estimation across South Africa, but the results could potentially be quite different, particularly if the off-diagonal covariance terms are large.

3.4 Prior flux covariance matrix

In the basic network design, we assumed no correlation in the prior covariance matrix of the fluxes. In reality, grid cells with similar biota and under similar climate will have correlated fluxes. Similarly, fluxes from the same source which occur close in time will also be correlated (Chevallier et al., 2010; Wu et al., 2013). To determine what impact correlation lengths in the prior flux covariance matrix could have on the optimal network design, we used the results from Chevallier et al. (2012), and put together a simple correlation structure where it was assumed that temporal correlations for the same pixel one week apart had a correlation of 0.5 (independent for day and night), decaying to 0.3 at two weeks apart and 0.1 at three weeks apart. Pixels adjacent to each other had a correlation of 0.3. The interaction between time and space correlations was determined by the product of the respective correlations (e.g. two pixels adjacent to each other but one week apart would have a correlation of 0.3×0.5).

ACPD
14, 11301–11342, 2014

3.5 Surface flux spatial resolution

For the sake of saving on computational time and on computing memory, the spatial resolution of the surface flux grid boxes was set so that the domain was divided into 50 by 25 grid boxes (a resolution of approximately $1.2^{\circ} \times 1.2^{\circ}$). If the surface grid boxes were set at a higher resolution, this would result in higher variability in the biospheric and fossil fuel fluxes estimates between grid boxes. As this would clearly affect the values in the prior flux covariance matrix, it could impact on the optimal network design as well. As a sensitivity test, the resolution of the surface grid boxes was reset so that there were 72 by 36 grid boxes (a resolution of approximately $0.8^{\circ} \times 0.8^{\circ}$), and the prior flux estimates adjusted accordingly.

3.6 Variance of the ocean sources

The standard design assumed that there was zero variance from the ocean sources as we wished to emphasize the importance of the terrestrial uncertainties in the network design. A sensitivity test was conducted whereby 10 % of the maximum land NEP uncertainty was allocated to the ocean grid cells. This uncertainty represents the uncertainty in the ocean productivity models which would be used to obtain prior estimates of ocean fluxes during an inversion.

3.7 Optimisation scheme

As an alternative to the IO routine, the GA was considered in this sensitivity test. If a network is already available, then it makes sense to add one station at a time to this existing structure (Ziehn et al., 2014), as is done in the IO procedure. In the South African case, there are already five instruments available for deployment, and therefore theoretically these instruments could be deployed simultaneously. The GA operates by optimising the five member network simultaneously, and therefore may be more suited to the case where there are multiple deployments, because it could be conceived that

ACPD
14, 11301–11342, 2014

the best solution for a five member network in terms of reducing the overall uncertainty for South Africa, may not include the one station which on its own reduces the uncertainty more than any other station.

Holding all other choices the same, the optimisation procedure used in the standard network design was replaced by the GA, and the resulting network compared to the basic optimal network design.

4 Results and discussion

As for Part 1 (Ziehn et al., 2014), the LPDM model was run for each station, including the two stations in the existing network. The influence functions for each station for the months of January and July were calculated. The particle counts used to calculate the influence functions can be summed over the month in order to obtain a footprint of each station. To illustrate this, plots of the influence footprint in January (Fig. 4) are provided for Cape Point and three other candidate stations numbered 28 (near Potchefstroom), 18 (near Mthatha), and 4 (near Port Elizabeth) as examples. For both January and July, the influence footprints show that the three candidate stations, to the north and to the east of Cape Point, have far more influence from terrestrial South Africa than Cape Point has.

With the influence functions available for each station, the test on the influence from the boundary could be conducted. Given the large domain over which LPDM was run, it was not surprising that the boundaries had minimal influence. The average value for the square root of the sum of all the diagonal elements of $\mathbf{C}_b$ for all stations was only 0.073 with a standard deviation of 0.026 in January, and 0.070 with a standard deviation of 0.031 in July.

The tests on the influence of sources outside of South Africa were also conducted. Since we decided to exclude the uncertainty of the ocean fluxes from our basic network design, the above influence test was modified to determine the strength of the influence from the ocean pixels on the observations at each measurement site. As South Africa

ACPD
14, 11301–11342, 2014

is surrounded by ocean to the west, south and east, it is expected that there would be a significant ocean influence, particularly on those stations near the coast. Assuming that each source emitted 1 kg C m^{-2} $week^{-1}$ (in order to easily compare between source regions), the values for the test were on average 37.12 ppm with a standard deviation of 52.09 during January and 29.17 with a standard deviation of 59.58 during July. Stations 12, 18, 25, and the Cape Point station (36) all had influence values over 100 during either or both January and July. The first three stations are all stations near the east coast of South Africa.

The influence test was further modified to determine the influence from land pixels which fall outside of South Africa. We decided it was important to take into account uncertainty contributed by these pixels, as we would have relatively poor observational constraints from these locations. But in determining the cost function for the network design, we would not include the uncertainty explicitly from these pixels; only those falling into South Africa. Therefore it would be of interest to find out if there are any stations which are strongly influenced by the non-South African land pixels. The stations with the strongest land influence from outside of South Africa were 30, 32, 33, 34 and 35, all of which are located in the northern parts of South Africa, near the border of Botswana. The Gobabeb station (37) had a surprisingly low influence from other African pixels, but higher compared to stations in the southern parts of South Africa.

The next sections present the results of the optimal network design; first under the basic parametrisations as used in Ziehn et al. (2014), and then under the different parametrisations for the influence function, for the covariance matrix, and under a different optimisation routine.

4.1 Basic network design

The results under the standard conditions used in the basic network design for the month of July reveal that the best set of stations to add to the current network would include a station near the south west coast, a station near the southern coast, a station in the western interior of the country, and two stations in the north eastern part of the

ACPD
14, 11301–11342, 2014

country near the borders of Botswana and Zimbabwe (Table 1). The base network consisting of Cape Point and Gobabeb GAW stations reduces the uncertainty of the flux estimates by 37.1% relative to the prior flux uncertainty during the month of July. The best station to add to this network would be station 4, located in the southern coast of South Africa, which reduces the overall uncertainty by 60.0% with its addition to the network (Fig. 5). The second best station during the month of July is station 28, in the North West provice of South Africa, which is an industry intensive province. This station reduces the uncertainty by an additional 27.8%. Since the optimal network includes a station near Cape Point during July, it supports the conclusions by Whittlestone et al. (2009) that Cape Point is not sufficient to estimate fluxes for the Western Cape region. The final reduction in uncertainty by the addition of the three other stations to the network is only a 3.8% drop in uncertainty. Therefore the returns on uncertainty reduction for additional stations to the network is reduced during the Winter months. This large reduction in uncertainty from only two additional stations during the Winter months is due to the localised fossil fuel emissions, and the small biospheric fluxes which occur during the Winter months, mainly on the western coast of South Africa.

In January the picture changes somewhat, with more stations located towards the interior of the country, and no stations positioned on the western side of South Africa (Fig. 5). The uncertainty reduction attributable to the base network is lower in January, at 26.7%. The best performing station in the network for January is station 18, located near the border of Lesotho and the eastern coast of South Africa, which reduced the uncertainty by an additional 30.7% relative to the base network. The next best performing station was station 28, as for the July optimal network, which reduced the uncertainty by an additional 18.4%. An additional 11.7% increase in uncertainty reduction is attained from adding the last three stations to the first and second added stations in the network. The return on additional stations to the network is therefore higher in January, when the uncertainty from biospheric fluxes is higher.

The network for the combined months of January and July result in a very similar positioning of stations compared to January (Fig. 5), as well as a very similar reduction

ACPD
14, 11301–11342, 2014

in uncertainty at 87.2%. This is due to the much lower NEP uncertainty estimates for the Winter months across South Africa compared to the Summer months, with the optimal network for the combined seasons being dominated by the need to reduce these larger uncertainties. The area near which station 0 is located in the Western Cape falls in the fynbos biome, which is under a Winter rainfall regime. Therefore productivity during the Winter months is expected to be higher in this area (Fig. 1a). In contrast, activity over much of South Africa during the Winter months, when water availability is reduced, is expected to be low to almost entirely dormant. It is likely that this resulting increased uncertainty in NEP in the fynbos regions during the Winter months, as well as the proximity to the City of Cape Town, would result in the need for a station in this area to reduce the overall uncertainty of South Africa. In contrast, the NEP uncertainty during Summer is much higher on the eastern side of the country compared to the mid interior or the west of the country (Fig. 1c), resulting in a need to concentrate the new measurement sites in this area during the Summer months.

4.2 Sensitivity analysis

Results from the sensitivity tests are presented in Fig. 6 and Table 2. The optimal networks under the cases where the maximum height for the surface grid is changed, increasing from 60 m and then to 75 m, when the increased uncertainty of night time transport is accounted for, and the case when the variance in fluxes over the ocean are accounted for, obtain the same optimal network for a particular month. The resulting networks are also identical to the base network. This implies that the optimal network design is robust to changes in the height of the surface grid cell, provided it is a reasonable estimate of the surface height. The change to the night time observational uncertainty was applied across the entire domain in the same way, and therefore this is likely why it did not affect the final outcome of the optimal network design. Ziehn et al. (2014) showed that changing the observational uncertainty so that it was different between the measurement sites would result in a different network compared to the

ACPD
14, 11301–11342, 2014

case of equivalent uncertainty at all sites. Applying an equivalent, non-zero value for the variance of the ocean pixels did not affect the solution for the optimal network.

When the trace of the posterior covariance matrix was used as the cost function, this resulted in very similar reductions in uncertainty for July, January and the combined months, and almost identical networks. The network optimised for July shared four stations relative to the basic design, but excluded station 7 in favour of station 9. In January, the optimal network is the same as for the standard case. The network for the combined months of July and January results in a very similar selection of station compared to the January trace case, but favouring station 4 in place of station 9. Three stations are in common with the standard case for July and January, which are stations 18, 11, and 29, but station 28 is favoured over station 27, and station 4 over station 9. This shows that the cost function can potentially have an impact on the final design, but is less so affected if the variance values are large, as for the month of January.

The case where correlation was assumed to exist between surface fluxes in time and in space resulted in a network for July with the three stations shared with the standard case for July, having station 4 and 28 as the best performing stations, but a preference for station 6 over station 0, and for 24 over station 7. The January network for the correlation case was very different to the basic January network, sharing only one station, with stations concentrated more towards the central interior of the country. The reductions in uncertainty were approximately 4.5% less in July compared to the basic case, and 4.6% less reduction in January compared to the basic case. The combined case of January and July for the first time resulted in a fairly different network to the January and July cases, introducing stations 11 and 30 which were neither in the July nor the January optimal networks. The reduction in uncertainty for the combined months was less than for the basic design, by 9.6%. Having correlation terms in the prior covariance matrix means that over an area where the high uncertainty pixels are relatively sparse, there is more memory in the covariance matrix of these uncertainties. The observation network is therefore assumed to be seeing more than just the pixels for which there are

ACPD
14, 11301–11342, 2014

sensitivities, providing more information to reduce the total uncertainty in the domain of interest.

The network obtained under the high resolution surface grids is the most different compared to the original network. As shown in (Ziehn et al., 2014), the aggregation of pixels smooths out the fossil fuel and biospheric fluxes. By using a higher resolution for aggregation, this smoothing is reduced (compare Figs. 7–1c and 2b). In July, three stations are shared with the basic July network (stations 0, 28 and 31). The best station in the July network is station 31, reducing the uncertainty relative to the base network by 58%. The final uncertainty reduction is 82%, which is approximately 10% lower compared to the original network. The January network for the high resolution case shares three stations with the original January network, and reduces the uncertainty by 79%, which is 8% lower compared to the original network. The location of the stations is very similar to the July case with stations concentrated towards the east of the country, near the areas of higher biospheric fluxes and the locations of power stations and industry in South Africa. The combined months of July and January result in the same optimal network as for January, sharing four stations with the standard case for July and January, with an overall uncertainty reduction of 79%, 8% lower compared to the standard case network.

The final sensitivity analysis was the use of a different optimisation routine, which was in the form of the GA. The resulting optimal network design for July was identical to the basic network design under the IO method. But although the GA did find the same solution as derived by the IO method for the month of January in its population of solutions, it was able to find an even better solution overall. The optimal solution for January from the GA method shared two stations with the basic network (stations 29 and 11), but replaced station 9 with 23, favoured station 27 to station 28 (just to the east of station 28) and station 12 to station 18 (just south of the station 18). The network for the combined months of July and January shared two stations with the standard case, replacing stations in the original network with adjacent stations. The final uncertainty reduction in the case of the combined months was slightly larger,

ACPD
14, 11301–11342, 2014

indicating that the GA did not find the global maximum in this particular run. The GA networks were very similar in nature to the original network obtained from the basic design case, and therefore, for the sake of efficiency and a reduction in computation time, the IO method can be implemented for a network design, as the results suggest that very similar conclusions would be drawn from both optimisation methods.

A disadvantage of a network design that concentrates stations on the eastern side of the country, where there is higher biospheric activity, is that, although the highest areas of uncertainty are well constrained, the western side, which has relatively smaller areas of high biological activity but does contain a major city, is less well constrained. Therefore the capacity to break the country down into subregions and solve for the higher resolution fluxes is reduced.

5 Summary and conclusions

Under a reference set of conditions, an optimal network design was obtained for South Africa for two representative months of the year. The resulting designs reduced the uncertainty of carbon fluxes from South Africa compared to the base network by 92 % in July and 87 % in January. These large reductions in uncertainty are due to the lack of coverage by the current network, which only reduces the uncertainty of fluxes from South Africa by 26 % and 37 % for July and January respectively, as well as the localisation of fossil fuel emission sources and the reduced biospheric activity in the Winter months. In all cases the first station has the greatest impact on uncertainty reduction, with lower returns yielded from additional stations to the network. The concentration of stations suggested by all networks is towards the central interior, near the North West Province of South Africa and in the eastern parts of the country. These represent the areas where the largest biospheric fluxes as well as fossil fuel emissions in the country take place, therefore location of the stations in these areas is expected. These areas also border with other African countries where there is greater uncertainty under the current carbon assessment results used in this study, and therefore, where pixels over-

ACPD
14, 11301–11342, 2014

lap with South Africa's neighbours, a larger resulting uncertainty in biospheric fluxes is obtained.

The sensitivity analysis demonstrated that for many of the choices required for the optimal network design, such as the height of the surface grid cells or the choice of whether to inflate night time observations relative to day time observations have a negligible impact on the final network design. Substituting the trace for the sum of the covariance elements, or using a different optimisation routine resulted in very similar solutions, giving confidence to the results for the optimal network, but both of these changes have the potential to result in larger changes to a network than observed in this case. The use of the trace leads to a more detail-focused solution, whereas the solution based on the sum of the covariance elements is more concerned with seeing the solution for the whole of South Africa, even if it's less detailed. The GA has the potential to capitalise on all the information available in the posterior covariance matrix, allowing it to find the network solution which in combination offers the best coverage of the domain of interest. The cost in efficiency and computation time relative to IO is high.

Choices which affect the prior covariance matrix, such as including correlation or increasing the resolution of the surface grid cells, result in an notable change in the location of stations in the network. Including correlation in the prior covariance results in a more spread out network design and higher value placed on observing smaller areas with high activity. The increase of spatial resolution on the other hand lead to higher concentration of sites near the larger areas of high activity. This implies that if the objective of the network is to reduce the overall uncertainty for a large area, like South Africa, having a high spatial resolution for the network may result in an over-concentration of sites in high activity areas, leaving large parts of the country undersampled. Such a network would reduce the capacity of the observation network to estimate fluxes in subregions across South Africa.

Therefore, in conclusion, it is best to run an optimal network design with specifications as close as possible to the that which would be used under the actual inversion

ACPD
14, 11301–11342, 2014

procedure for estimating surface fluxes from concentration observations, including taking correlation between pixels into account. The resolution of the spatial grids should be in line with the number of stations added to the network and the size in subregion for which fluxes could be estimated over the domain of interest given the available number of stations. Wu et al. (2013) demonstrated an algorithm for optimising the regions for estimation as well as the network. Overall the results suggest that a good improvement in knowledge of South African fluxes is available from a feasible atmospheric network and that the general features of this network are robust to many reasonable choices in a network design study.

Acknowledgements. Peter Rayner is in receipt of an Australian Professorial Fellowship (DP1096309). This worked was supported by parliamentary grant funding from the Council of Scientific and Industrial Research. The authors would like to thank the helpful commentary from Thomas Lauvaux on the implementation and post processing of the LPDM model.

References

Baker, D. F., Law, R. M., Gurney, K. R., Rayner, P., Peylin, P., Denning, A. S., Bourquet, P., Bruhwiler, L., Chen, Y., Ciais, P., Fung, I. Y., Heimann, M., John, J., Maki, T., Maksyutov, S., Masarie, K., Prather, M., Pak, B., Taguchi, S., Zhu, Z.: TransCom 3 inversion intercomparison: impact of transport model errors on the interannual variability of regional CO_2 fluxes, 1988–2003, Global Biogeochem. Cy., 20, GB1002, doi:10.1029/2004GB002439, 2006. 11305

Bousquet, P., Ciais, P., Peylin, P., Ramonet, M., and Monfray, P.: Inverse modeling of annual atmospheric CO_2 sources and sinks: 1. Method and control inversion, J. Geophys. Res., 104, 26161–26178, 1999. 11305

Canadell, J. G., Le Quéré, C., Raupach, M. R., Field, C. B., Buitenhuis, E. T., Ciais, P., Conway, T. J., Gillett, N. P., Houghton, R. A., and Marland, G.: Contributions to accelerating atmospheric CO_2 growth from economic activity, carbon intensity, and efficiency of natural sinks, P. Natl. Acad. Sci. USA, 104, 18866–18870, doi:10.1073/pnas.0702737104, 2007. 11303

Chevallier, F., Ciais, P., Conway, T. J., Aalto, T., Anderson, B. E., Bousquet, P., Brunke, E. G., Ciattaglia, L., Esaki, Y., Fröhlich, M., Gomez, A., Gomez-Pelaez, A. J., Haszpra, L., Krummel, P. B., Langenfelds, R. L., Leuenberger, M., Machida, T., Maignan, F., Matsueda, H.,

ACPD
14, 11301–11342, 2014

Morgui, J. A., Mukai, H., Nakazawa, T., Peylin, P., Ramonet, M., Rivier, L., Sawa, Y., Schmidt, M., Steele, L. P., Vay, S. A., Vermeulen, A. T., Wofsy, S., and Worthy, D.: CO_2 surface fluxes at grid point scale estimated from a global 21 year reanalysis of atmospheric measurements, J. Geophys. Res., 115, D21307, doi:10.1029/2010JD013887, 2010. 11303, 11307, 11308, 11309, 11318

Chevallier, F., Wang, T., Ciais, P., Maignan, F., Bocquet, M., Arain, M. A., Cescatti, A., Chen, J., Dolman, A. J., Law, B. E., Margolis, H. A., Montagnani, L., and Moors, E. J.: What eddy-covariance measurements tell us about prior land flux errors in CO_2-flux inversion schemes, Global Biogeochem. Cy., 26, GB1021, doi:10.1029/2010GB003974, 2012. 11318

Ciais, P., Rayner, P., Chevallier, F., Bousquet, P., Logan, M., Peylin, P., and Ramonet, M.: Atmospheric inversions for estimating CO_2 fluxes: methods and perspectives, Climatic Change, 103, 69–92, 2010. 11305

Denman, K. L., Brasseur, G., Chidthaisong, A., Ciais, P., Cox, P. M., Dickinson, R. E., Hauglustaine, D., Heinze, C., Holland, E., Jacob, D., Lohmann, U., Ramachandran, S., da Silva Dias, P. L., Wofsy, S. C., and Zhang, X.: Couplings between changes in the climate system and biogeochemistry, in: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K. B., Tignor, M., and Miller, H. L., Cambridge University Press, Cambridge, UK and New York, NY, USA, 499–587, 2007. 11303

Engelbrecht, F. A., McGregor, J. L., and Engelbrecht, C. J.: Dynamics of the Conformal-Cubic Atmospheric Model projected climate-change signal over southern Africa, Int. J. Climatol., 29, 1013–1033, 2009. 11310

Enting, I. G.: Inverse Problems in Atmospheric Constituent Transport, Cambridge Univ. Press, New York, 2002. 11303, 11305, 11306

Enting, I. G. and Mansbridge, J. V.: Seasonal sources and sinks of atmospheric CO_2: direct inversion of filtered data, Tellus B, 41, 111–126, 1989. 11303

Gurney. K. R., Law, R. M., Denning, A. S., Rayner, P. J., Baker, D., Bousquet, P., Bruhwiler, L., Chen, Y., Ciais, P., Fan, S., Fung, I. Y., Gloor, M., Heimann, M., Higuchi, K., John, J., Maki, T., Maksyutov, S., Masarie, K., Peylin, P., Prather, M., Pak, B. C., Randerson, J., Sarmiento, J., Taguchi, S., Takahashi, T., and Yuen, C.: Towards robust regional estimates of CO_2 sources and sinks using atmospheric transport models, Nature, 405, 626–630, 2002. 11305

ACPD
14, 11301–11342, 2014

Gurney, K. R., Law, R. M., Denning, A. S., Rayner, P. J., Baker, D., Bousquet, P., Bruhwiler, L., Chen, Y., Ciais, P., Fan, S., Fung, I. Y., Gloor, M., Heimann, M., Higuchi, K., John, J., Kowalczyk, E., Maki, T., Maksyutov, S., Peylin, P., Prather, M., Pak, B. C., Sarmiento, J., Taguchi, S., Takahashi, T., and Yuen, C.: TransCom 3 CO_2 inversion intercomparison: 1. Annual mean control results and sensitivity to transport and prior flux information, Tellus B, 55, 555–579, 2003. 11305

Kaminski, T., Heimann, M., and Giering, R.: A coarse grid three dimensional global inverse model of the atmospheric transport, 2. Inversion of the transport of CO_2 in the 1980s, J. Geophys. Res., 104, 18555–18581, 1999. 11305

Lambers, H., Chapin, F. S., and Pons, T. L.: Plant Physiology Ecology, Springer Science+Business Media L. L. C., New York, USA, 634 pp., 2008. 11308

Lauvaux, T., Uliasz, M., Sarrat, C., Chevallier, F., Bousquet, P., Lac, C., Davis, K. J., Ciais, P., Denning, A. S., and Rayner, P. J.: Mesoscale inversion: first results from the CERES campaign with synthetic data, Atmos. Chem. Phys., 8, 3459–3471, doi:10.5194/acp-8-3459-2008, 2008. 11307, 11310, 11317

Lauvaux, T., Schuh, A. E., Uliasz, M., Richardson, S., Miles, N., Andrews, A. E., Sweeney, C., Diaz, L. I., Martins, D., Shepson, P. B., and Davis, K. J.: Constraining the CO_2 budget of the corn belt: exploring uncertainties from the assumptions in a mesoscale inverse system, Atmos. Chem. Phys., 12, 337–354, doi:10.5194/acp-12-337-2012, 2012. 11304, 11305, 11317

Law, R. M., Chen, Y., Gurney, K. R., and Transcom 3 Modellers: TransCom 3 CO_2 inversion intercomparison: 2. Sensitivity of annual mean results to data choices, Tellus B, 55, 580–595, 2003. 11305

McGregor, J. L.: Accuracy and initialization of a two-time-level split semi-Lagrangian model, in: Collection of Papers Presented at WMO/IUGG NWP Symposium, Tokyo, 4–8 August 1987, 233–246, 1987. 11310

McGregor, J. L. and Dix, M. R.: The CSIRO conformal-cubic atmospheric GCM, in: IUTAM Symposium on Advances in Mathematical Modelling of Atmosphere and Ocean Dynamics, Limerick, Ireland, 2–7 July 2000, edited by: Hodnett, P. F., Kluwer, Dordrecht, 197–202, 2001. 11310

Morgan, E., Lavrič, J., Seely, M., and Heimann, M.: Establishment of an atmospheric observatory for trace gases and atmospheric oxygen in Namibia, Geophys. Res. Abstr., 14, 5122–5122, 2012. 11304

ACPD
14, 11301–11342, 2014

Peylin, P., Baker, D., Sarmiento, J., Ciais, P., and Bousquet, P.: Influence of transport uncertainty on annual mean and seasonal inversions of atmospheric CO_2 data, J. Geophys. Res., 107, 4385, doi:10.1029/2001JD000857, 2002. 11305

Rayner, P. J.: Optimizing CO_2 observing networks in the presence of model error: results from TransCom 3, Atmos. Chem. Phys., 4, 413–421, doi:10.5194/acp-4-413-2004, 2004. 11304, 11315, 11316

Rayner, P. J., Enting, I. G., and Trudinger, C. M.: Optimizing the CO_2 observing network for constraining sources and sinks, Tellus B, 48, 433–444, 1996. 11305

Rayner, P. J., Enting, I. G., Francey, R. J., and Langenfelds, R. L.: Reconstructing the recent carbon cycle from atmospheric CO_2, $\delta^{13}C$ and O_2/N_2 observations, Tellus B, 51, 213–232, 1999. 11303, 11305

Rayner, P. J., Raupach, M. R., Paget, M., Peylin, P., and Koffi, E.: A new global gridded data set of CO_2 emissions from fossil fuel combustion: mehodology and evaluation, J. Geophys. Res., 115, D19306, doi:10.1029/2009JD013439, 2010. 11309

Rödenbeck, C., Houweling, S., Gloor, M., and Heimann, M.: CO_2 flux history 1982–2001 inferred from atmospheric data using a global inversion of atmospheric transport, Atmos. Chem. Phys., 3, 1919–1964, doi:10.5194/acp-3-1919-2003, 2003. 11303

Scholes, R. J., von Maltitz, G. P., Archibald, S. A., Wessels, K., van Zyl, T., Swanepoel, D., and Steenkamp, K.: National Carbon Sink Assessment for South Africa: First Estimate of Terrestrial Stocks and Fluxes, CSIR Technical Report, Pretoria, South Africa, CSIR/NRE/GC/ER/2013/0056/B, 2013. 11308

Seibert, P. and Frank, A.: Source-receptor matrix calculation with a Lagrangian particle dispersion model in backward mode, Atmos. Chem. Phys., 4, 51–63, doi:10.5194/acp-4-51-2004, 2004. 11310, 11311, 11312

Tarantola, A.: Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation, Elsevier, Amsterdam, 1987. 11305, 11307

Uliasz, M.: The atmospheric mesoscale dispersion modeling system, J. Appl. Meteorol., 31, 139–149, 1993. 11310

Uliasz, M.: Lagrangian particle modeling in mesoscale applications, in: Environmental Modelling II, Computational Mechanics Publications, Southampton, UK, 71–102, 1994. 11310, 11311

ACPD
14, 11301–11342, 2014

Whittlestone, S., Kowalczyk, E., Brunke, E. G., and Labuschagne, C.: Source Regions for CO_2 at Cape Point Assessed by Modelling 222Rn and Meteorological Data, Technical Report for the South African Weather Service, Pretoria, South Africa, 2009. 11303, 11310, 11322

Wu, L., Bocquet, M., Chevallier, F., Lauvaux, T., and Davis, K.: Hyperparameter estimation for uncertainty quantification in mesoscale carbon dioxide inversions, Tellus B, 65, 20894, doi:10.3402/tellusb.v65i0.20894, 2013. 11305, 11307, 11318, 11328

Ziehn, T., Nickless, A., Rayner, P. J., Law, R. M., Roff, G., and Fraser, P.: Greenhouse gas network design using backward Lagrangian particle dispersion modelling – Part 1: Methodology and Australian test case, Atmos. Chem. Phys. Discuss., 14, 7557–7595, doi:10.5194/acpd-14-7557-2014, 2014. 11304, 11305, 11306, 11309, 11310, 11311, 11312, 11313, 11314, 11319, 11320, 11321, 11323, 11325

ACPD
14, 11301–11342, 2014

Table 1. Ranking of the new stations added to the base network for two seasons (Summer and Winter) represented by January and July, as well as the integrated months. The cummulative reduction of uncertainty relative to the base uncertainty is provided in brackets.

Rank	July	January	July + January
1	4 (60.0 %)	18 (57.4 %)	18 (54.7 %)
2	28 (87.8 %)	28 (75.6 %)	29 (71.1 %)
3	7 (89.3 %)	11 (81.4 %)	11 (80.1 %)
4	31 (90.7 %)	29 (84.9 %)	27 (83.5 %)
5	0 (91.6 %)	9 (87.3 %)	9 (87.2 %)

ACPD
14, 11301–11342, 2014

Table 2. Ranking of the new stations added to the base network under eight different sensitivity tests for the combined months of July and January. The tests include the cases when the surface grid height is set at 60 m, the surface grid height is set at 75 m, the trace of the posterior covariance is used in the cost function, the standard deviations of the night time observations are doubled, a correlation structure is implemented in the prior covariance of the fluxes, the spatial resolution is increased to 0.8°, the ocean sources are assigned a variance, and the genetic algorithm is used for optimisation. The percentage cumulative reduction of uncertainty relative to the base uncertainty is provided in brackets.

Rank	Height 60 m	Height 75 m	Trace	Double Night Cov	Correlation	High Res	Ocean Variance	GA
1	18 (53.5)	18 (51.7)	18 (40.3)	18 (51.5)	18 (43.1)	18 (41.8)	18 (54.4)	24
2	29 (69.4)	29 (67.1)	28 (54.7)	29 (67.6)	27 (57.7)	29 (61.5)	29 (74.4)	27
3	11 (79.0)	11 (77.5)	4 (72.5)	11 (77.5)	7 (68.4)	11 (69.4)	11 (80.9)	10
4	9 (82.8)	9 (81.7)	11 (76.8)	27 (81.3)	11 (74.4)	30 (74.4)	9 (84.7)	29
5	27 (86.6)	27 (85.6)	29 (80.7)	9 (85.3)	30 (77.6)	27 (78.9)	27 (86.9)	5 (87.6)

ACPD
14, 11301–11342, 2014

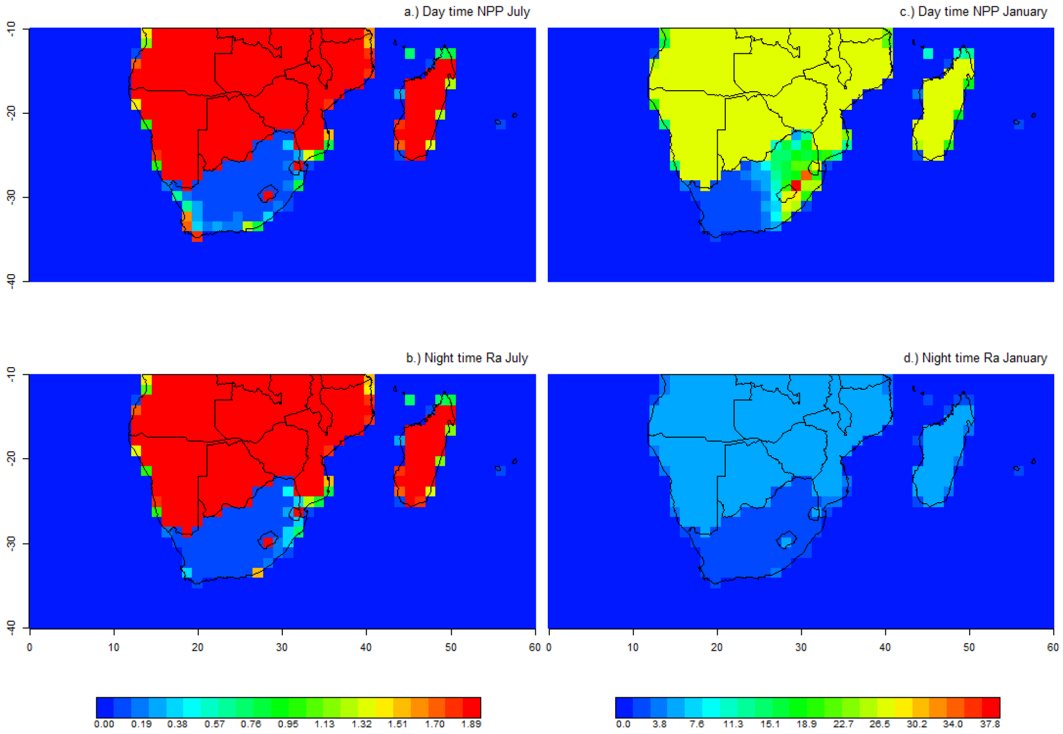


Fig. 1. The day time net primary productivity (NPP) and night time autotrophic respiration (Ra) data used as standard deviations of net ecosystem productivity (NEP) at the resolution of 1.2° expressed in $g C m^{-2} week^{-1}$ for July (left) and January (right). Values for the standard deviation are capped at $28 g C m^{-2} week^{-1}$. The maximum value (separately for day and night) is assigned to the non-South African land surface, or set at $28 g C m^{-2} day^{-1}$ if the maximum exceeds this value.

ACPD
14, 11301–11342, 2014

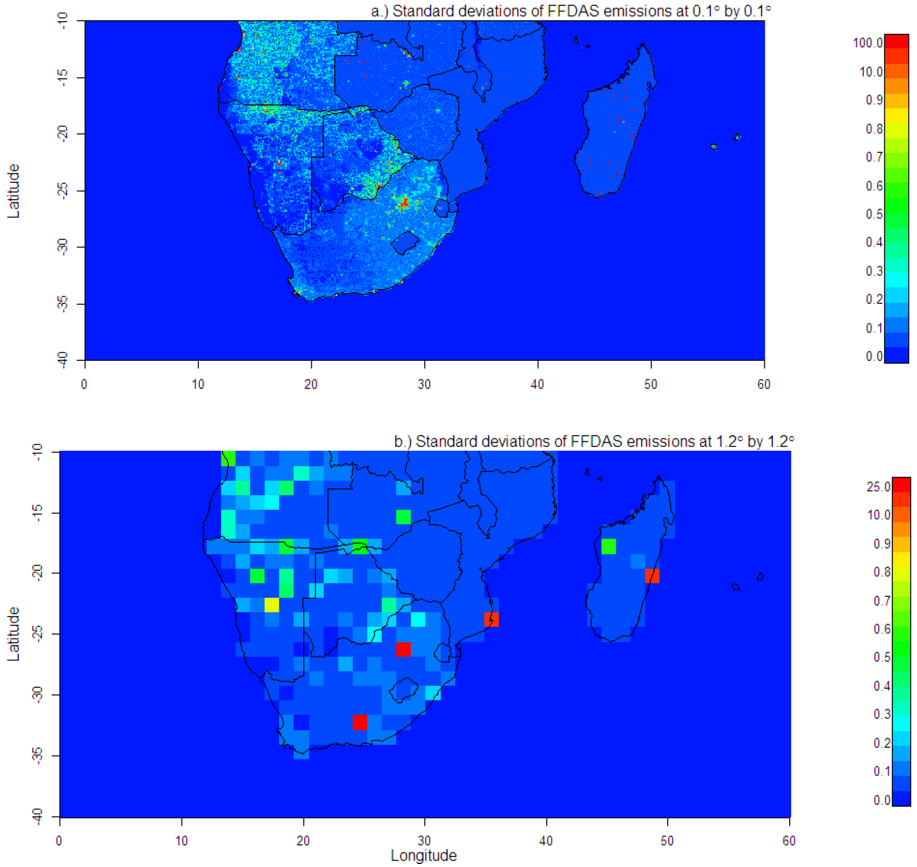


Fig. 2. The standard deviations of ten realisations (top) of the Fossil Fuel Data Assimilations System (FFADS) at the original 0.1° resolution in $gCm^{-2}week^{-1}$. The standard deviations of the aggregated fluxes (bottom) (1.2° resolution) showing significant smoothing of the fossil fuel fluxes over the lower resolution.

ACPD
14, 11301–11342, 2014

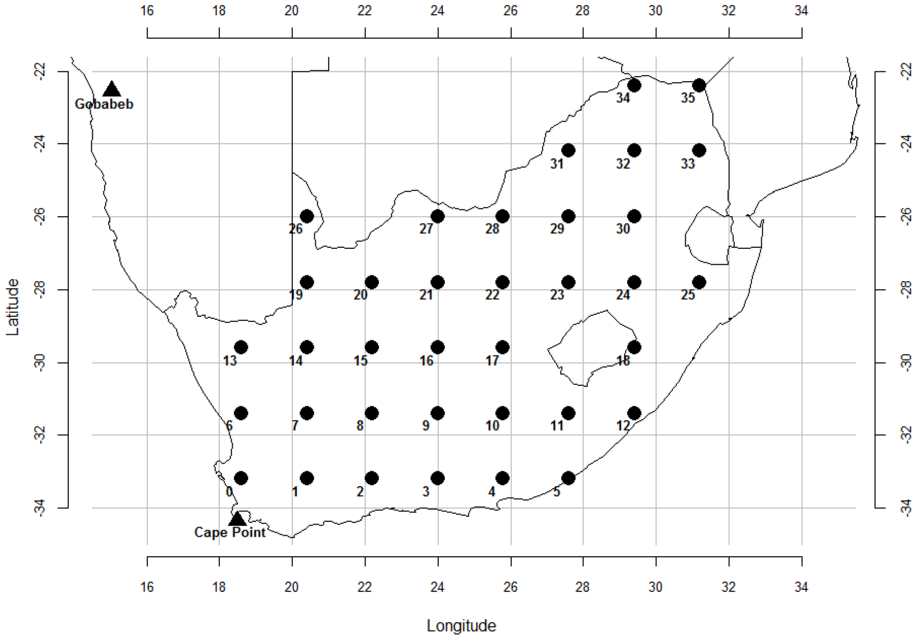


Fig. 3. The 36 potential locations of the new stations in the optimal network design. The locations were spaced on a regular grid over the surface of South Africa. The existing Cape Point and the Gobabeb GAW stations are marked by the triangles.

ACPD
14, 11301–11342, 2014

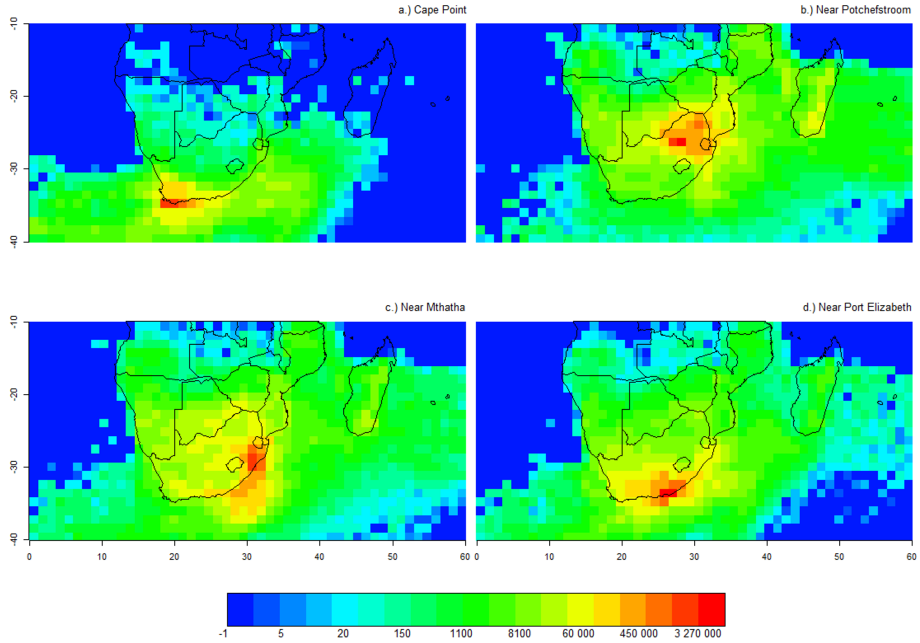


Fig. 4. The footprint of Cape Point, station 28 (top right), station 18 (bottom left), and station 4 (bottom right) relative to the surface grid cells at a resolution of 1.2° expressed as the count of particles over the month of January from a particular surface grid cell.

ACPD
14, 11301–11342, 2014

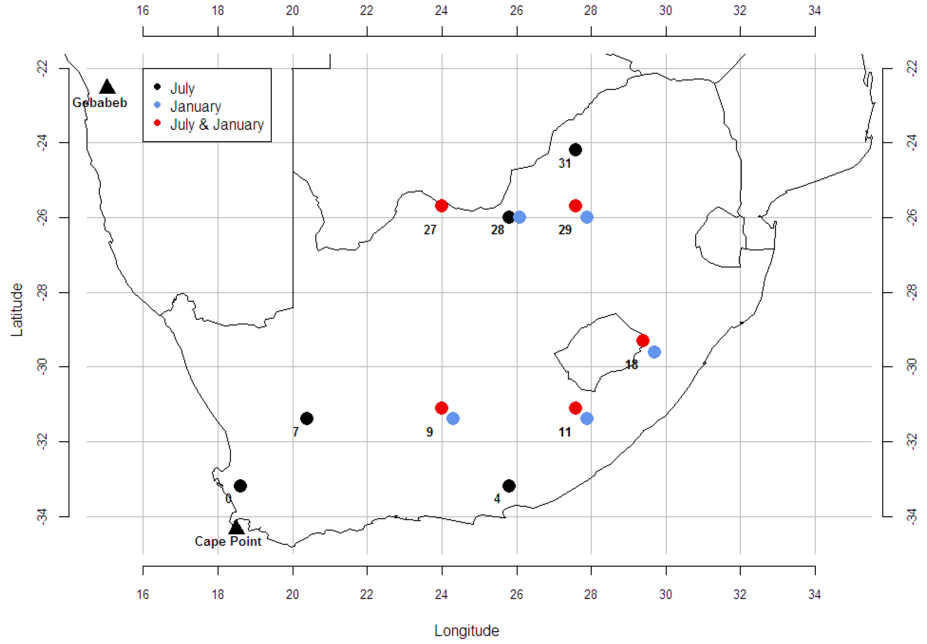
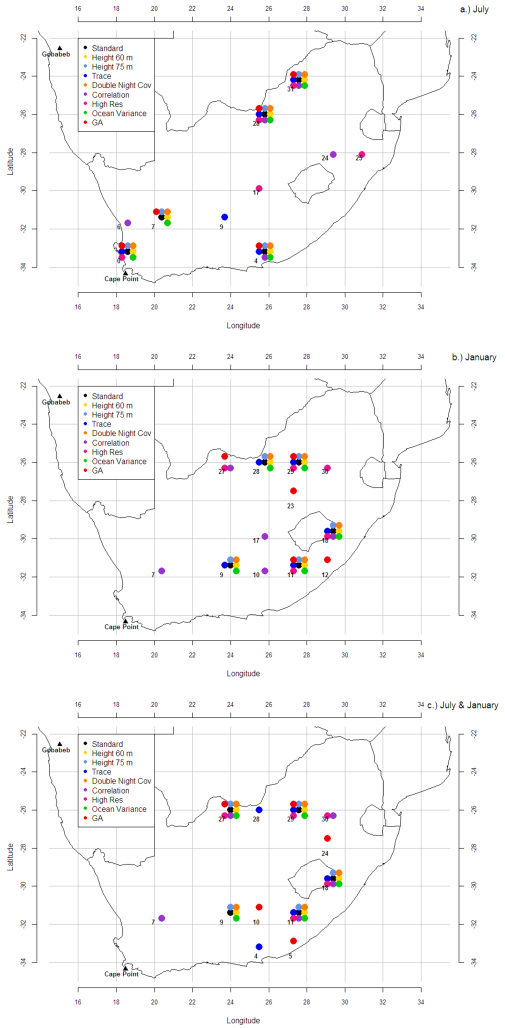


Fig. 5. Map of the locations of the optimal stations to add to the existing network to reduce of the overall uncertainty of fluxes in South Africa under the basic design of 50 m surface grid height, diagonal prior covariance, 2 ppm uncertainty in concentration observations, a 1.2° surface grid resolution, and using the sum of the elements of the posterior covariance matrix to calculate the cost function for the optimisation procedure for July, January, and the combined months of July and January.

ACPD
14, 11301–11342, 2014



ACPD
14, 11301–11342, 2014

Greenhouse gase network design South Africa

A. Nickless et al.

Fig. 6. Map of the locations of the optimal stations to add to the existing network to reduce the overall uncertainty of fluxes in South Africa under the nine different sensitivity cases for July (top), January (middle), and the combined months of July and January (bottom). The cases include the standard case (Standard), the surface grid height set at 60 m (height 60 m), the surface grid height set at 75 m (height 75 m), the use of the trace in the cost function (Trace), the doubling of the night time observation standard deviation (Double Night Cov), the addition of correlation between elements in the prior covariance matrix (Correlation), the increasing of the spatial resolution to 0.8° (High Res), the inclusion of variance of the ocean sources (Ocean), and the use of the genetic algorithm (GA).

ACPD
14, 11301–11342, 2014

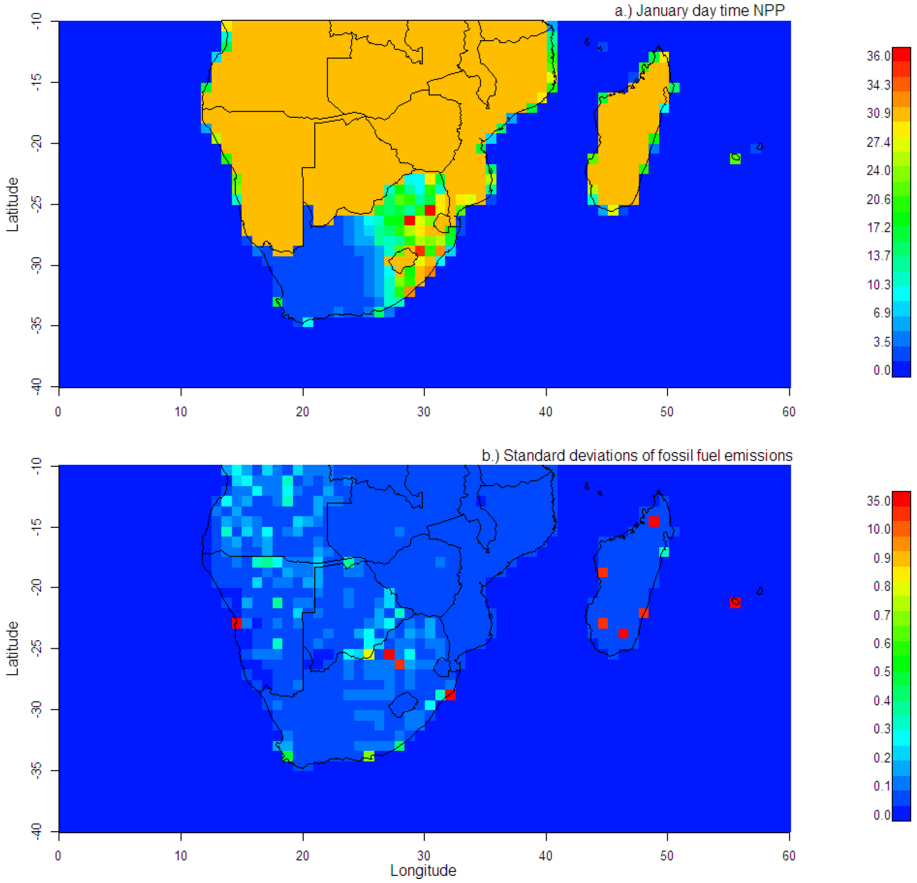


Fig. 7. The day time net primary productivity (NPP) data used as standard deviations of net ecosystem productivity (NEP) at the resolution of 0.8° expressed in $gCm^{-2}week^{-1}$ for January (top), and the Fossil Fuel Data Assimilation System standard deviations aggregated over a resolution of 0.8°, also expressed in $gCm^{-2}week^{-1}$ (bottom).

ACPD
14, 11301–11342, 2014