REGULARISATION OF A CARBON CYCLE MODEL-DATA FUSION PROBLEM

S. Delahaies*, I. Roulstone* AND N. K. Nichols†

Abstract. The problem of constraining variables and parameters in a model of a terrestrial ecosystem, using observations of the net ecosystem exchange of CO2, is studied within the framework of four-dimensional variational data assimilation (4DVAR). 4DVAR combines the observations, the nonlinear model and our prior knowledge of the ecosystem states and parameters, by optimizing a cost function. At the heart of the assimilation process lies a simple but non-trivial inverse problem involving a linear operator. Using a singular value decomposition of the linear operator, we show that the inverse problem is ill-posed and we use the truncated singular value decomposition to find a regularized solution.

Key words. carbon cycle, model-data fusion, variational data assimilation, inverse problem, ill-posedness, regularisation, singular value decomposition

AMS subject classifications. 37N25, 49K40, 49N45

1. Introduction. Model-data fusion, or inverse modelling, is the process of best combining our understanding of the dynamics of a system, observations and our prior knowledge of the state of the system. Improving our understanding of the carbon cycle is an important component of modelling climate and the Earth system, and a variety of inverse modelling techniques have been used to combine process models with different types of observational data.

Eddy covariance measurements, see [1], of net ecosystem exchange of CO_2 (NEE) have been used intensively for over a decade to confront the Data Assimilation-Linked Ecosystem model (DALEC) with real data. DALEC is a simple box model for terrestrial ecosystems simulating a large range of processes occurring at different time scales from days to millennia. The work of Williams et al. (2005), see [13], established the benefit of using NEE measurements in a Bayesian framework to improve estimates of carbon stocks and fluxes for ecosystem models, and to quantify uncertainties. Following this work, the REgional FLux Estimation eXperiment (REFLEX), reported in [5], compared the strengths and weaknesses of various inverse modelling strategies to estimate parameters and initial stocks for DALEC. Nine participants were asked to use DALEC together with NEE measurements with the inverse modelling technique of their choice and variants of the Ensemble Kalman filter and Monte Carlo methods were preferred choices; most results agreed on the fact that parameters and initial stocks directly related to fast processes were best estimated with narrow confidence intervals, whereas those related to slow processes were poorly estimated with very large uncertainties. While other studies have tried to overcome this difficulty by adding complementary data streams, see [11], or by considering longer observation windows, see [10], no systematic analysis has been carried out so far to explain the large differences among results of REFLEX.

DALEC is a simple model; it represents the basic processes at the heart of more sophisticated models of the carbon cycle. A qualitative analysis of its dynamical behaviour has been carried out in [2]. It is not our intention to arrive at a conclusion about the viability or otherwise of DALEC as a tool for data-fusion studies: rather, its merit is that its simplicity facilitates close mathematical scrutiny. The analysis we carry out here may be performed on more complex models (e.g. JULES), but it is imperative to understand the nature of the inverse problem for a simple, yet non-trivial, example first. Our analysis is carried out on the tangent linear approximation of DALEC. We consider a simple inverse problem involving an overdetermined linear system with synthetic observations. Using a singular value decomposition we show that the problem is seriously ill-posed, and we study

*Department of Mathematics, University of Surrey, Guildford, UK (s.b.delahaies@surrey.ac.uk).
†Department of Mathematics, University of Reading, Reading, UK.

the effect of noise in the observation on the relative error in the solution. Finally we propose a regularization of the solution using a truncated singular value decomposition.

The paper is organized as follows. In Section 2 we present the DALEC evergreen model and its tangent linear approximation. In section 3, we consider an inverse problem for the tangent linear model of DALEC. Using singular value decomposition we show that the problem is ill-posed, that is, the solution is sensitive to small errors. In section 4 we use the truncated singular value decomposition to regularize the inverse problem. Finally we draw conclusions in section 5.

2. DALEC evergreen: mathematical formulation. The DALEC model is a simple process-based model describing an evergreen forest ecosystem as a set of five carbon pools linked via fluxes. The carbon pools represent foliage (C_{f}), woody stems and coarse roots

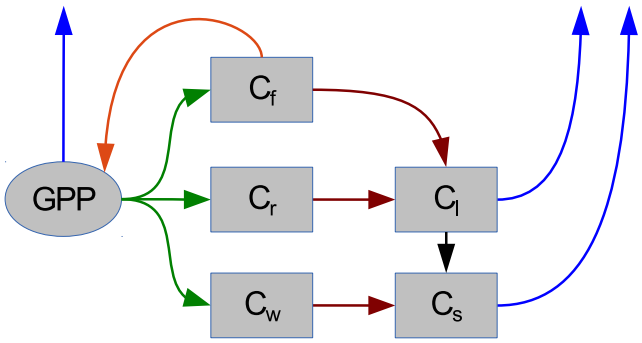

Fig. 2.1: *DALEC evergreen model:GPP is linked to the carbon pools (C) via allocation fluxes (green arrows), litterfall fluxes (red arrows), decompostion (black arrow). Respiration is represented by the blue arrows. The orange arrow represents the feedback of foliar carbon to GPP. Each arrow can be seen as a process controlled by one parameter.*

(C_{w}), fine roots (C_{r}), fresh leaf and fine root litter (C_{l}) and soil organic matter and coarse woody debris (C_{s}). The gross primary production (GPP), denoting the total amount of carbon to be allocated, is calculated at a daily time step using the aggregated canopy model (ACM) as a function of the site's leaf area index and meteorological drivers (total daily irradiance, minimum and maximum daily temperature, atmospheric CO_2 concentration). GPP is then allocated to carbon pools and respirations (autotrophic and heterotrophic) via fluxes as depicted on Figure 2.1. The allocation processes are controlled by ten parameters. An additional parameter, the foliar nitrogen content, is used in ACM to calculate the GPP. The *active variables* are then the dynamic variables $\mathbf{C} = (C_{\mathrm{f}}, C_{\mathrm{r}}, C_{\mathrm{w}}, C_{\mathrm{l}}, C_{\mathrm{s}})^T$ together with the parameters $\mathbf{p} = (p_1, \ldots, p_{11})^T$. The complete description of DALEC and the GPP function can be found in [13]. A qualitative analysis of the dynamical behaviour of the system was conducted in [2]. The evolution of the carbon pools at day $t+1$ is given by

$$C_{\mathrm{f}}(t+1) = (1-p_5)C_{\mathrm{f}}(t) + p_3(1-p_2)\,\mathrm{GPP}(C_{\mathrm{f}}(t), p_{11}, \Phi), \tag{2.1}$$

$$C_{\mathrm{r}}(t+1) = (1-p_7)C_{\mathrm{r}}(t) + p_4(1-p_3)(1-p_2)\,\mathrm{GPP}(C_{\mathrm{f}}(t), p_{11}, \Phi), \tag{2.2}$$

$$C_{\mathrm{w}}(t+1) = (1-p_6)C_{\mathrm{w}}(t) + (1-p_4)(1-p_3)(1-p_2)\,\mathrm{GPP}(C_{\mathrm{f}}(t), p_{11}, \Phi), \tag{2.3}$$

$$C_{\mathrm{l}}(t+1) = (1-(p_1+p_8)T(t))C_{\mathrm{l}}(t) + p_5 C_{\mathrm{f}}(t) + p_7 C_{\mathrm{r}}(t), \tag{2.4}$$

$$C_{\mathrm{s}}(t+1) = (1-p_9 T(t))C_{\mathrm{s}}(t) + p_6 C_{\mathrm{w}}(t) + p_1 T(t) C_{\mathrm{l}}(t), \tag{2.5}$$

Pool	label	description	range	initial state
C_1	C_f	foliar C mass	0/500	332
C_2	C_r	fine root C mass	0/500	313
C_3	C_w	wood C mass	0/30000	13121
C_4	C_l	fresh litter C mass	0/500	52
C_5	C_s	soil organic matter and woody matter C mass	0/15000	10024

Table 2.1: *DALEC evergreen carbon pools with their respective range ($gCm^{-2}day^{-1}$).*

Parameter	label	description	range	initial state
p_1	T_d	decomposition rate	$10^{-6}/0.01$	4.41×10^{-6}
p_2	F_g	fraction of GPP respired	0.2/0.7	0.52
p_3	F_nf	fraction of NPP1 allocated to foliage	0.01/0.5	0.29
p_4	F_nrr	fraction of NPP2 alocated to roots	0.01/0.5	0.41
p_5	T_f	turnover rate of foliage	$10^{-4}/0.1$	2.8×10^{-3}
p_6	T_w	turnover rate of woods	$10^{-6}/0.01$	2.06×10^{-6}
p_7	T_r	turnover rate of roots	$10^{-4}/0.1$	3.0×10^{-3}
p_8	T_l	mineralisation rate of C_l	$10^{-5}/0.1$	2.0×10^{2}
p_9	T_s	mineralisation rate of C_s	$10^{-5}/0.1$	2.65×10^{-6}
p_{10}	E_t	temperature dependant rate parameter	0.05/0.2	6.93×10^{-2}
p_{11}	P_r	Nitrogen use efficiency parameter in ACM	5/20	7.4

Table 2.2: *DALEC evergreen parameters with their respective range.*

where T is defined by

$$T(t) = \frac{1}{2}\exp(p_{10}T_m(t)), \tag{2.6}$$

where T_m is daily mean temperature. Φ denotes the climate drivers: total daily irradiance, minimum and maximum daily temperature, atmospheric CO_2 concentration. The net ecosystem exchange (NEE) of CO_2 is estimated using DALEC as the difference between GPP and respirations. It can be written as

$$\mathrm{NEE}(t) = (1-p_2)\,\mathrm{GPP}(C_\mathrm{f}(t), p_{11}, \Phi) + p_8 C_\mathrm{l} T(t) + p_9 C_\mathrm{s} T(t). \tag{2.7}$$

The definition and acceptable range of the different carbon pools and allocation parameters are summarized in Table 2.1 and 2.2. DALEC is designed to model a large range of processes with very different time scales; this is reflected in the different scales of the carbon pools and parameters. In order to avoid the computational problems caused by the different scales we transform the active variables $(\mathbf{C}^T, \mathbf{p}^T)^T \in \mathbb{R}^n$ to normalized variables $\mathbf{z} = (\mathbf{z}_1^T, \mathbf{z}_2^T)^T \in \mathbb{R}^n$, where $\mathbf{z}_1 = \log(\mathbf{C})$ and $\mathbf{z}_2 = \log(\mathbf{p})$. We denote by H_i the nonlinear function that maps the active variable $\mathbf{z}$ at day t_0 to the NEE at day t_i

$$H_i : \mathbf{z} \mapsto \mathrm{NEE}(t_i). \tag{2.8}$$

Let $\mathbf{x} \in \mathbb{R}^n$ be a perturbation. We denote by $\mathbf{H}_i$ the tangent linear operator defined by

$$H_i(\mathbf{z}+\mathbf{x}) = H_i(\mathbf{z}) + \mathbf{H}_i\mathbf{x} + O(\|\mathbf{x}\|^2), \tag{2.9}$$

where $\|\|$ denotes the euclidean norm, and formally given as a row vector of size n by

$$\mathbf{H}_i = \left[\frac{\partial \mathrm{NEE}_i}{\partial \mathbf{z}}\right]^T. \tag{2.10}$$

DALEC, in its original form, is given as a FORTRAN code of roughly one hundred lines of code. The tangent linear operator $\mathbf{H}_i$ can be derived using well documented techniques (cf

[6]) or using automatic differentiation software (OpenAD, see [12]). Both approaches were tested producing similar results up to machine precision. Here we work with a code derived by hand.

The tangent linear model can be used to perform forward and backward sensitivity analysis. Figure 2.2 shows the Jacobian matrix for the NEE for a time window of two thousand days. This picture gives the sensitivity with respect to the normalized variables. We see some differences among columns which represent the active variables. The columns

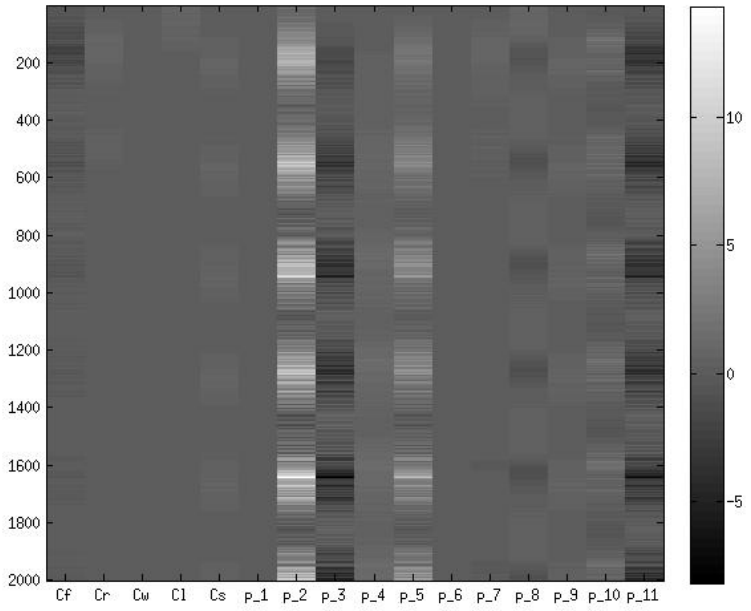

Fig. 2.2: *Jacobian matrix for the NEE for 2000 days: each line represents the operator* $\mathbf{H}_i$ *for* $i = 1, \ldots, 2000$. *The matrix is scaled to enhance the different magnitudes.*

corresponding to C_w, p_1 and p_6 keep the same gray color during all the time window; this indicates that NEE is very weakly sensitive with respect to these variables. On the contrary for C_f, p_2, p_3, p_5 and p_{11} we see periodic oscillations showing a larger sensitivity of the NEE with respect to those variables. The periodicity of the signal corresponds to the seasonal variability of the climate drivers (temperature, solar irradiance). Further analysis of the sensitivity of DALEC can be found in [3]. We will see in the next sections how these features of the Jacobian matrix affect the model-data fusion problem.

3. An ill-posed inverse problem. The aim of data fusion is to determine the model trajectory that best fits the observed data. The best fit minimizes the errors between the observations and the model predictions of the observations. We study the simplest case that exhibits the difficulties inherent to fusing NEE observations with DALEC in order to demonstrate and investigate the nature of the problem and propose simple methods to overcome the difficulties. To do so we focus on the tangent linear operator $\mathbf{H}_i$ using basic linear algebra and analysis.

We start with a perturbation $\mathbf{x}_0 \in \mathbb{R}^n$, hereafter called the truth, and we generate N, $N > n$, exact observations $\mathbf{y} = (y_1, \ldots, y_N)^T$ uniformly distributed in time $\{t_1, \ldots, t_N\}$ by

$$\mathbf{y} = \mathbf{H}\mathbf{x}_0, \tag{3.1}$$

where $\mathbf{H}$ denotes the observability matrix, that is the $N \times n$ matrix defined by

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_{t_1} \\ \vdots \\ \mathbf{H}_{t_N} \end{bmatrix}. \tag{3.2}$$

Let $\boldsymbol{\epsilon} \in \mathbb{R}^N$ be a discrete white noise with variance ν^2. We study the effect of the noise on the least square solution

$$\mathbf{x}_{\mathrm{LS}} = \operatorname{argmin} \|\mathbf{H}\mathbf{x} - (\mathbf{y} + \boldsymbol{\epsilon})\|, \tag{3.3}$$

of the overdetermined linear system $\mathbf{H}\mathbf{x} = \mathbf{y} + \boldsymbol{\epsilon}$. We consider a singular value decomposition of $\mathbf{H}$ of the form

$$\mathbf{H} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T, \tag{3.4}$$

where $\mathbf{U}$ is a $N \times N$ unitary matrix, $\mathbf{V}$ is a $n \times n$ unitary matrix and $\boldsymbol{\Sigma}$ is the $N \times n$ diagonal matrix whose diagonal elements are the singular values $\sigma_1 \geq \cdots \geq \sigma_n \geq 0$. The solution $\mathbf{x}_{\mathrm{LS}}$ can be written as

$$\mathbf{x}_{\mathrm{LS}} = \sum_{i=1}^{n} \frac{\mathbf{u}_i^T(\mathbf{y} + \boldsymbol{\epsilon})}{\sigma_i} \mathbf{v}_i, \tag{3.5}$$

where $\mathbf{u}_i$ and $\mathbf{v}_i$ are the left and right singular vectors, namely the column vectors of the matrices $\mathbf{U}$ and $\mathbf{V}$ respectively. The solution corresponds to the maximum likelihood estimator and its variance, used later in the construction of confidence intervals, is given by

$$\mathrm{Cov}(\mathbf{x}_{\mathrm{LS}}) = (\mathbf{H}^T\mathbf{H})^{-1}. \tag{3.6}$$

We now examine the maximum relative change in the solution, or relative error, due to noise with standard deviation ν and defined by

$$\eta(\nu) = \frac{\|\mathbf{x}_{\mathrm{LS}}(\nu) - \mathbf{x}_0\|}{\|\mathbf{x}_0\|}. \tag{3.7}$$

The relative error in the i-th component of the solution is given by

$$\eta_i(\nu) = \left| \frac{\mathbf{x}_{\mathrm{LS},i}(\nu) - \mathbf{x}_{0,i}}{\mathbf{x}_{0,i}} \right|, \tag{3.8}$$

for $i = 1, \ldots, n$. It is well known, see [4, 7], that the relative error is bounded by

$$\frac{\|\mathbf{x}_{\mathrm{LS}} - \mathbf{x}_0\|}{\|\mathbf{x}_0\|} \leq \kappa(\mathbf{H}) \frac{\|\boldsymbol{\epsilon}\|}{\|\mathbf{y}\|}, \tag{3.9}$$

where $\kappa(\mathbf{H})$ is the condition number of $\mathbf{H}$ defined by $\kappa(\mathbf{H}) = \sigma_1/\sigma_n$. When the condition number is large the matrix is said to be ill-conditioned, the problem is ill-posed and the solution (3.5) is unstable: small perturbations to the system can lead to very large perturbations in the solution.

The ill-posedness of the problem and its effect on the solution $\mathbf{x}_{\mathrm{LS}}$ is illustrated in Table 3.1. As described above we choose a true perturbation $\mathbf{x}_0$, we generate $N = 200$ true observations uniformly distributed in days $\{t_1, \ldots, t_N\}$ and we form the observability matrix $\mathbf{H}$. The condition number of $\mathbf{H}$ is $\kappa(\mathbf{H}) = 1.6 \times 10^9$. Three different values for the noise variance ν are considered: 10^{-7} corresponding to machine epsilon for single precision, 10^{-2}, and 0.5 corresponding to a realistic level of noise for NEE measurements. The values for η_i and η are reported in Table 3.1. As shown in the last row of the Table 3.1 the relative

	$\nu = 10^{-7}$	$\nu = 10^{-2}$	$\nu = 0.5$
C_{f}	3.96e-14	2.87e-04	1.15e+00
C_{r}	1.07e-09	1.21e+00	1.58e+04
C_{w}	1.87e-08	2.29e+01	3.21e+05
C_{l}	4.67e-10	5.86e-01	7.65e+03
C_{s}	2.06e-10	4.45e-01	6.56e+02
p_1	2.30e-06	2.65e+03	3.46e+07
p_2	1.97e-14	1.11e-04	7.83e-02
p_3	5.46e-14	9.83e-04	1.26e-01
p_4	9.79e-10	1.17e+00	1.52e+04
p_5	2.34e-14	8.43e-04	5.38e-01
p_6	1.36e-08	1.41e+01	2.50e+05
p_7	3.62e-11	4.03e-02	5.51e+02
p_8	4.84e-10	5.86e-01	7.64e+03
p_9	2.13e-10	4.42e-01	7.06e+02
p_{10}	2.39e-14	1.38e-06	1.25e-01
p_{11}	1.72e-14	1.03e-04	1.03e+00
η	1.35e-08	1.64e+01	2.31+05

Table 3.1: *The change in the relative errors η and η_i defined in equations (3.7) and (3.8) as functions of ν.*

error η increases drastically as ν increases, and for a realistic level of noise ($\nu = 0.5$) the solution is not reliable. When $\nu = 10^{-7}$ all variables are correctly estimated with at least six digits accuracy but yet we can see differences among variables. With a standard deviation $\nu = 10^{-2}$ the parameters p_1, p_6, and the carbon pool C_{w} are far from their true value. More generally we find an agreement with the results of REFLEX: parameters directly linked to foliage and gross primary productivity are better estimated than parameters related to allocation to and turnover of fine root/wood. The results of Table 3.1 reflect the sensitivity analysis discussed in the previous section. The variables with respect to which NEE is the most (resp. least) sensitive are the less (resp. more) affected by the noise. In the next section we consider a well established method to reduce the impact of the noise on the solution of the problem.

4. Regularization. Several methods exist to regularize ill-posed inverse problems; their performance depends on the nature of the ill-posedness. The truncated singular value decomposition (TSVD) is a popular method for regularization. TSVD consists in truncating the sum in equation (3.5) in order to remove the smallest singular values, the most affected by the noise. The solution is then given by

$$\mathbf{x}_{\mathrm{TSVD}} = \sum_{i=1}^{k} \frac{\mathbf{u}_i^T(\mathbf{y} + \boldsymbol{\epsilon})}{\sigma_i} \mathbf{v}_i, \tag{4.1}$$

where k is the truncation rank. The covariance of the solution is given by

$$\mathrm{Cov}(\mathbf{x}_{\mathrm{TSVD}}) = \mathbf{V}\boldsymbol{\Sigma}_k^{-2}\mathbf{V}^T, \tag{4.2}$$

where $\boldsymbol{\Sigma}_k$ is the matrix of singular values with the $n-k$ smallest singular values set to zero. The truncation rank k can be chosen using the L-curve method. The L-curve is a log-log plot of the norm of the solution $\|\mathbf{x}_k\|$ against the norm of the residual $\|\mathbf{H}\mathbf{x}_k - (\mathbf{y} + \boldsymbol{\epsilon})\|$ parametrized by the regularisation parameter k. The optimal parameter corresponds to the point of maximum curvature of the L-curve. Further details on the L-curve method can be found in [9].

We now apply TSVD to our inverse problem. As previously we choose a perturbation $\mathbf{x}_0$ and we generate the $N = 200$ true observations $\mathbf{y}$ uniformly distributed in time $\{t_1, \ldots, t_N\}$ where $\mathbf{y} = (y_1, \ldots, y_N)^T$ is given by

$$y_i = \mathbf{H}_{t_i}\mathbf{x}_0, \quad i = 1, \ldots, N, \tag{4.3}$$

and finally we add a white noise $\boldsymbol{\epsilon}$ with standard deviation $\nu = 0.5$. We used Hansen's regularization tools [7] to perform the TSVD method. The truncation rank $k = 7$ is found using the L-curve shown on Figure 4.1. Table 4.1 shows the regularized solution, the standard

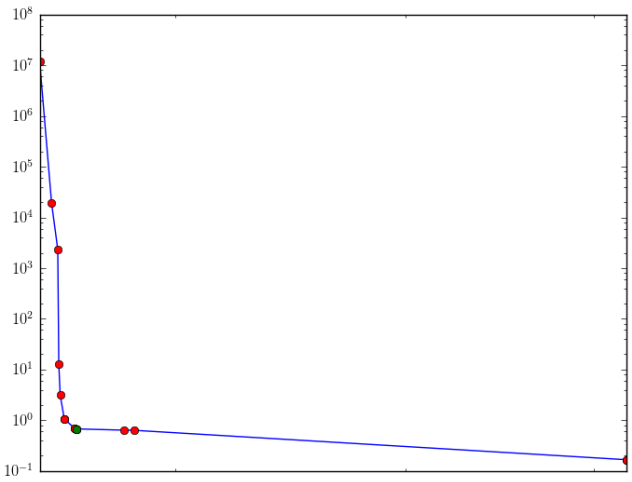

Fig. 4.1: *L-curve: log-log plot of the norm of the solution* $\|\mathbf{x}_k\|$ *against the norm of the residual* $\|\mathbf{H}\mathbf{x}_k - (\mathbf{y} + \boldsymbol{\epsilon})\|$ *parametrized by the regularisation parameter* k*. The blue curve shows an interpolation of the discrete L-curve (red points); the green point corresponding to* $k = 7$ *is the corner of the curve.*

deviations and the relative errors. The last column of Table 4.1, presenting the relative error in the regularized solution, can be compared with the last column of Table 3.1 which shows the relative error of the unstable solution with the same level of noise. Whereas the relative errors in the unstable solution range from 7.83×10^{-2} to 3.46×10^7 the relative errors in the regularized solution range from 2×10^{-2} to 1. The standard deviations are of the same magnitude as the variables, but considering the large ranges for the variables (see Table 2.1 and 2.2) they nevertheless provide relatively narrow confidence intervals. We see that TSVD has the effect of keeping small the variables that cannot be estimated correctly: for the variables C_s, p_1 and p_6, for which the relative errors were the biggest in Table 3.1, the relative error is close to 1. As previously stated the results of the regularization should to be related to the sensitivity analysis depicted on Figure 2.2: TSVD performs better on the variables with respect to which NEE is the most sensitive and prevents the variables with respect to which NEE is the least sensitive from growing unboundedly.

Figure 4.2 shows the noisy observations (red points) together with the true trajectory for the NEE (red curve) and the trajectory obtained with the regularized solution (blue). Finally the 95% confidence area (blue shaded area) is constructed using a χ^2 test. We see that the regularized solution matches the truth with a narrow confidence interval in the three-year observation period. Moreover the regularized solution remains close to the truth with still a narrow confidence interval in the next three-year period.

	$\mathbf{x}$	$\boldsymbol{\nu}$	η_i
C_{f}	53.2	56.9	0.080
C_{r}	34.7	67.3	0.357
C_{w}	1.9	2.7	0.999
C_{l}	6.5	16.0	0.689
C_{s}	1739.3	778.2	0.172
p_1	-5.8E-10	4.2E-10	1.000
p_2	0.12	0.056	0.217
p_3	0.04	0.048	0.141
p_4	0.11	0.13	0.346
p_5	3.6E-4	5.4E-4	0.352
p_6	3.7E-10	5.5E-10	0.999
p_7	4.6E-4	7.3E-4	0.220
p_8	4.8E-3	0.3E-3	0.204
p_9	4.3E-6	1.8E-6	0.185
p_{10}	1.3E-2	1.0E-2	0.020
p_{11}	1.6	1.2	0.133

Table 4.1: *TSVD solution: solution, standard deviation and relative error.*

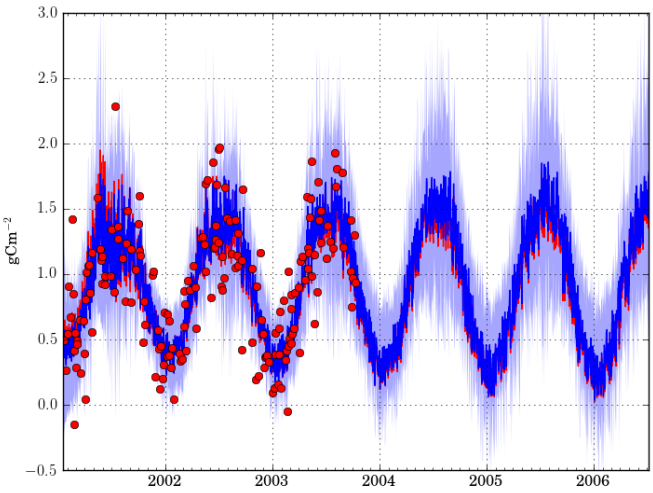

Fig. 4.2: *NEE time series: true trajectory (red curve), NEE observations (red points), trajectory obtained with the TSVD solution (blue), the blue shaded area is the 95% confidence interval for the regularized solution.*

5. Concluding remarks. The problem of estimating parameters and initial stocks for the DALEC model using NEE observations has been the subject of many papers in recent years [5, 10, 11, 13]. Inverse modelling techniques such as Ensemble Kalman filter and Monte Carlo methods have proven their ability to correctly estimate some parameters and to reduce the uncertainty of the predicted carbon fluxes; however, to date there is no consensus on the reason why some parameters remain poorly estimated and how to address this issue. This paper adopts the formalism of variational data assimilation to try provide insights into these questions. We derived the tangent linear model of the DALEC evergreen

model and we considered a simple inverse problem for the linearisation of DALEC using synthetic observations. The small size of the problem allowed us to use basic linear algebra to show the ill-posedness of the problem. We then considered the truncated singular value decomposition and we showed that this method provides a robust solution.

Having found a regularization of this much studied model-data fusion problem, we are investigating other techniques, and studying their application to more sophisticated models of the carbon cycle. This work will be complemented by studies of the dynamical system aspects of these models (cf. Chuter et al. 2013), and analyses of the performance of data assimilation algorithms using eddy covariance measurements.

REFERENCES

[1] D. Baldocchi, *Assessing the eddy covariance technique for evaluating carbon dioxide exchange rates of ecosystems: past, present and future*, Glob. Chang. biol., 9 (2003), pp. 479–492.

[2] A. M. Chuter, P. J. Aston, A. C. Skeldon and I. Roulstone, *Tipping points for forests, using the data assimilation linked ecosystem carbon model*, submitted, 2013.

[3] A. M. Chutter, *A Qualitative Analysis of the Data Assimilation Linked Ecosystem Carbon Model, DALEC*, PhD Thesis, University of Surrey, Guildford, UK, 2012.

[4] G. H. Golub and C. F. Van Loan, *Matrix Computations*, Third ed., The Johns Hopkins University Press, Baltimore, MD, 1996.

[5] A. Fox, M. Williams, A. D. Richardson, D. Cameron, J. H. Gove, T. Quaife, D. Ricciuto, M. Reichstein, E. Tomelleri, C. M. Trudinger and M. T. Van Wijk, *The REFLEX project: Comparing different algorithms and implementations for the inversion of a terrestrial ecosystem model against eddy covariance data*, Agric. For. Meteorol., 149 (2009), pp. 1597–1615.

[6] R. Giering and T. Kaminski, *Recipes for Adjoint Code Construction*, ACM trans. Math. sof., 24 (), pp. 437–474.

[7] P. C. Hansen, *Discrete Inverse Problems: Insight and Algorithms*, SIAM 2010.

[8] P. C. Hansen, *Regularization tools: a MATLAB package for analysis and solution of discrete ill-posed problems*, Numer. Algorithms, 6 (1994), pp. 1–35.

[9] P. C. Hansen and D. P. OLeary, *The use of the L-curve in the regularization of discrete ill-posed problems*, SIAM J. Sci. Comput., 14 (1993), pp. 1487–1503.

[10] T. C. Hill, E. Ryan and M. Williams, *The use of CO2 flux time series for parameter and carbon stock estimation in carbon cycle research*, Glob. Chang. Biol., 18 (2012), pp. 179–193.

[11] A. D. Richardson, M. Williams, D. Y. Hollinger, D. J. P. Moore., D. B. Dail, E. A. Davidson, N. A. Scott, R. S. Evans, H. Hughes, J. T. Lee, C. Rodrigues, K. Savage, *Estimating parameters of a forest ecosystem C model with measurements of stocks and fluxes as joint constraints*, Oecologia, 164 (2010), pp. 25–40.

[12] J. Utke, U. Naumann and A. Lyons, *OpenAD/F: User Manual*, http://www.mcs.anl.gov/OpenAD/openad.pdf , 2010.

[13] M. Williams, P. Schwartz, B. Law, J. Irvine, and M. R. Kurpuis , *An improved analysis of forest carbon dynamics using data assimilation*, Glob. Chang. Biol., 11 (2005), pp. 89–105.