Department of Mathematics and Statistics

Preprint MPS-2013-10

27 August 2013

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	

[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* H_i *for* $i = 1, \dots, 2000$*. 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 H_i using basic linear algebra and analysis.

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

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

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

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

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

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

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

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

where U is a $N \times N$ unitary matrix,

	= 10 ^{7}	= 10 ^{2}	= 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
	1.35e-08	1.64e+01	2.31+05

Table 3.1: *The change in the relative errors and $_i$ de ned in equations (3.7) and (3.8) as functions of .*

error increases drastically as increases, and for a realistic level of noise (= 0:5) the solution is not reliable. When = 10 ^{7} all variables are correctly estimated with at least six digits accuracy but yet we can see di erences among variables. With a standard deviation = 10 ^{2} the parameters p_1, p_6, and the carbon pool C_w are far from their true value. More

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 ϵ with standard deviation $\sigma = 0.5$. 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

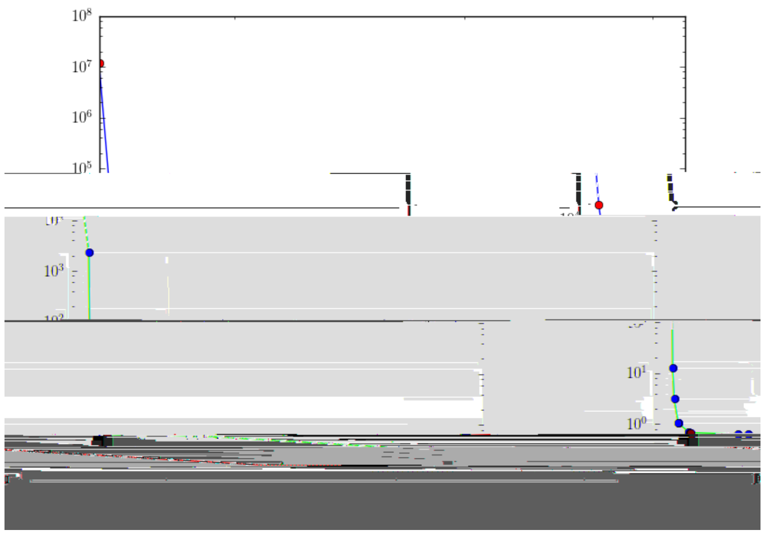

Fig. 4.1: *L-curve: log-log plot of the norm of the solution $\|\mathbf{x}_k\|$ against the norm of the residual $\|\mathbf{H}\mathbf{x}_k - (\mathbf{y} + \epsilon)\|$ parametrized by the regularisation parameter k. The blue curve shows an interpolation of the discrete L-curve (red points); the green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

	x	ν	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.*

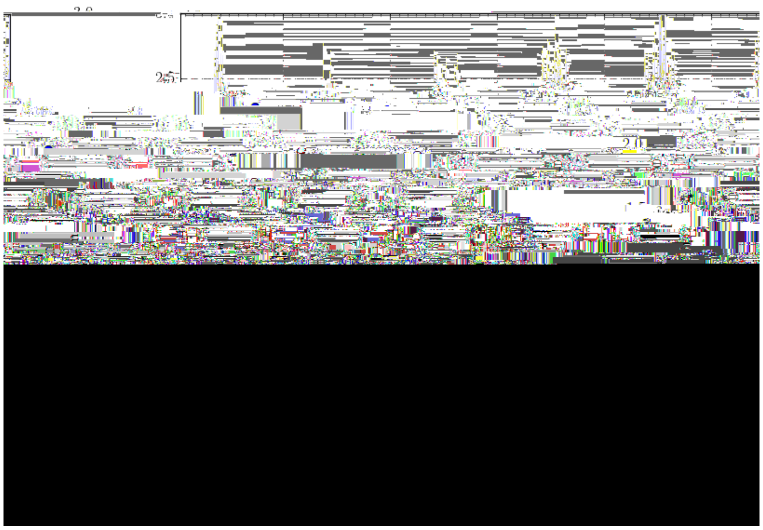

Fig. 4.2: *NEE time series: true trajectory (red curve), NEE observations (red points), trajectory obtained with the TSVD solution (blue), the blue shaded area is the 95% con dence interval for the regularized solution.*

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 lter 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 di erent 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*