

# Observability and information content **DRAFT**

Ewan Pinnington

November 2, 2016

## 1 Introduction

In data assimilation it is important to understand if the observations available to us provide us with enough information to find a meaningful description of our studied system. Measurements of forest carbon balance are now routinely made in forests across the world using micrometeorological techniques, with many other relevant observations such as leaf area index and standing biomass also available [Baldocchi, 2008]. Many efforts have been made to combine this data with models of forest carbon balance using data assimilation techniques in order to improve modelled estimates [Fox et al., 2009, Niu et al., 2014, Quaife et al., 2008, Richardson et al., 2010, Zobitz et al., 2011, 2014]. Currently, however, the relative levels of information from different data types is not well understood.

In numerical weather prediction many measures of observation information content have been defined [Cardinali et al., 2004, Fisher, 2003, Rodgers et al., 2000]. These measures can be used to identify how information content might vary both temporally and spatially, when observations are made at different times or in different locations. It is not necessary to have made a physical observation in order to estimate its information content. It is enough to have an accurate estimate to the observation error at the specified time and location. It is therefore possible to use these measures to define target observations or design new observing systems [Eyre, 1990, Palmer et al., 1998]. Often we are required to know the derivative of our model in order to implement these measures. This can prove difficult to implement. However, techniques such as automatic-differentiation [Renaud, 1997] can reduce the time taken to find the derivative of a model.

In this chapter we aim to analyse the information content in the observations used for assimilation with the DALEC1 and DALEC2 models of ecosystem carbon balance. We begin by considering the observability of our system given a set of observations. Observability is a mathematical concept from control theory. Applied to data assimilation a system is defined as observable if for a given set of observations we can uniquely define the initial state for our model. This allows us to determine if from current observations used in carbon balance data assimilation, we have enough information to find a unique model state. In practice we include a background term in our assimilation (**ref. DA section**) to ensure we can always find a locally unique solution. However, it is informative to show that observations alone provide us with enough information to find a unique solution.

We then consider different information content measures applied to our system in order to show how the information content varies for the different observation types available to us for DALEC1. We then extend these results to the DALEC2 model and investigate how the same set of observations can have a different level of information depending on the type of ecosystem being observed. Using these measures also allows us to consider the effect of including error correlations in our data assimilation algorithm, as explored in section (**ref. 1st results chapter**), on the information content in the observations.

## 2 Background material

### 2.1 Metolius forest site

In this chapter we use meteorological driving data taken from the Metolious forest site, a temperate coniferous forest in Oregon, Northwestern US. The site has been studied extensively [Law et al., 2001a], and has also been the subject of data assimilation studies [Quaife et al., 2008, Williams et al., 2005]. The site has a semi-arid climate with a dominant canopy of ponderosa pine (*Pinus ponderosa*) and an understory of bitterbush (*Purshia tridentata*) and manzanita (*Arctostaphylos patula*) [Law et al., 2001b]. The forest stand was felled in 1978, having previously been a mature forest. It was then allowed to regrow naturally, with some areas of older growth forest still being left post-felling [Williams et al., 2005].

### 2.2 Observability

Observability is a mathematical concept from control theory. A system is said to be observable if it is possible to determine the state by measuring only the output. The following definition is taken from Barnett and Cameron [1985]: for the linear time varying system defined as,

$$\dot{\mathbf{x}} = \mathbf{A}(t)\mathbf{x}(t) + \mathbf{B}(t)\mathbf{u}(t) \quad (1)$$

$$\mathbf{y} = \mathbf{C}(t)\mathbf{x}(t) \quad (2)$$

where  $\mathbf{A}$  is  $n \times n$ ,  $\mathbf{B}$  is  $n \times m$  and  $\mathbf{C}$  is  $r \times n$  is *completely observable* if for any  $t_0$  and any initial state  $\mathbf{x}(t_0) = \mathbf{x}_0$  there exists a finite time  $t_i > t_0$  such that knowledge of  $\mathbf{u}(t)$  and  $\mathbf{y}(t)$  for  $t_0 \leq t \leq t_i$  suffices to uniquely determine  $\mathbf{x}_0$ . There is no loss of generality in assuming  $\mathbf{u}(t)$  is identically zero throughout the whole interval; this is the case for data assimilation. So that in data assimilation our system is *completely observable* if knowledge of the observations  $\mathbf{y}(t)$  allows us to uniquely determine the initial state  $\mathbf{x}_0$ .

**Theorem 2.1.** *When  $\mathbf{A}$ ,  $\mathbf{B}$  and  $\mathbf{C}$  are time-invariant the system is completely observable if and only if the  $nr \times n$  observability matrix*

$$\mathbf{V} = \begin{pmatrix} \mathbf{C} \\ \mathbf{CA} \\ \mathbf{CA}^2 \\ \vdots \\ \mathbf{CA}^{n-1} \end{pmatrix} \quad (3)$$

*has rank  $n$ .*

This result can be applied to the data assimilation problem [Johnson et al., 2005], where for 4D-Var the observability matrix corresponds to

$$\hat{\mathbf{H}} = \begin{pmatrix} \mathbf{H}_0 \\ \mathbf{H}_1\mathbf{M}_0 \\ \vdots \\ \mathbf{H}_N\mathbf{M}_{N,0} \end{pmatrix} \quad (4)$$

as defined in section (ref. DA section). In Appendix B of Zou et al. [1992] it is shown that for the linear data assimilation problem it is possible to obtain a unique analysis state over a specific assimilation window with no background term if the rank of  $\hat{\mathbf{H}}$  is equal to  $n$ , the size of  $\mathbf{x}_0$ . For the non-linear data assimilation problem the rank of  $\hat{\mathbf{H}}$  being equal to  $n$  ensures a locally unique analysis can be found without including a background term. In practice the cost function for 4D-Var data assimilation typically contains a background term which regularises the problem and means that we always have a unique solution.

## 2.3 Information content measures

In data assimilation we combine prior estimates with observations to improve our knowledge of the state of a system. In this process some observations will have a greater impact on our results than others. Many measures exist for understanding observation impact on the assimilation results.

Information content measures have been used to quantify the different levels of information provided by observations in the development of satellite instruments [Engelen and Stephens, 2004, Stewart et al., 2008] and in operational data assimilation schemes [Fisher, 2003, Singh et al., 2013]. In Fowler and Van Leeuwen [2013] it is discussed that in these operational schemes information content measures have been used for

- Removing observations with a lesser impact in order to improve the efficiency of the assimilation process [Rabier et al., 2002, Rodgers, 1998, Singh et al., 2013].
- Diagnosing erroneous observations and assumed statistics [Desroziers et al., 2009].
- Improving data assimilation results by adding observations which theoretically have a high impact. This can mean defining target observations [Palmer et al., 1998] or even designing new observing systems [Eyre, 1990, Wahba, 1985].

For the following measures the data assimilation problem is assumed to be Gaussian with a linear function mapping the state to observation space ( $\mathbf{H}$ ), following the derivation from Kalnay [2003] we have,

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{K}(\mathbf{y} - \mathbf{H}\mathbf{x}_b), \quad (5)$$

where  $\mathbf{K}$  is the Kalman gain matrix,

$$\mathbf{K} = (\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} + \mathbf{B}^{-1})^{-1} \mathbf{H}^T \mathbf{R}^{-1}. \quad (6)$$

In order to consider observations over a 4D-Var time window we rewrite equation (5) as,

$$\mathbf{x}_a = \mathbf{x}_b + \hat{\mathbf{K}}(\hat{\mathbf{y}} - \hat{\mathbf{H}}\mathbf{x}_b), \quad (7)$$

using the defined matrices in section (ref DA section), with  $\hat{\mathbf{K}} = (\hat{\mathbf{H}}^T \hat{\mathbf{R}}^{-1} \hat{\mathbf{H}} + \mathbf{B}^{-1})^{-1} \hat{\mathbf{H}}^T \hat{\mathbf{R}}^{-1}$ .

Making the assumption of a linear and Gaussian data assimilation problem is clearly a limitation. These measures are therefore limited to a period where the forecast model remains reasonably linear. The implications of assuming Gaussian error statistics are discussed in Fowler and Van Leeuwen [2013].

### 2.3.1 Sensitivity of analysis to observations

The influence matrix measures the sensitivity of the analysis in observation space to the observations [Cardinali et al., 2004] and is defined by,

$$\mathbf{S} = \frac{\partial \mathbf{H}\mathbf{x}_a}{\partial \mathbf{y}}. \quad (8)$$

From equation (5) we see that,

$$\begin{aligned} \mathbf{S} &= \frac{\partial \mathbf{H}(\mathbf{x}_b + \mathbf{K}(\mathbf{y} - \mathbf{H}\mathbf{x}_b))}{\partial \mathbf{y}} \\ &= \mathbf{K}^T \mathbf{H}^T, \end{aligned} \quad (9)$$

here  $\mathbf{S}$  will be a  $p \times p$  matrix, where  $p$  is the number of observations. The diagonal elements of  $\mathbf{S}$  are  $\mathbf{S}_{i,i} = \frac{\partial (\mathbf{H}\mathbf{x}_a)_i}{\partial y_i}$  and represent the ‘self-sensitivity’ of the  $i^{th}$  modelled observation to the  $i^{th}$  observation. The

off-diagonal elements of  $\mathbf{S}$  represent the ‘cross-sensitivity’ and are given by  $\mathbf{S}_{i,j} = \frac{\partial(\mathbf{H}\mathbf{x}_a)_i}{\partial\mathbf{y}_j}$ . If we wish to consider the influence matrix for observations over a 4D-Var time window we can re-write equation (8) as,

$$\mathbf{S} = \frac{\partial\hat{\mathbf{H}}\mathbf{x}_a}{\partial\hat{\mathbf{y}}} = \hat{\mathbf{K}}^T \hat{\mathbf{H}}^T. \quad (10)$$

The Kalman gain matrix  $\hat{\mathbf{K}}$  can be re-written as,

$$\hat{\mathbf{K}} = \mathbf{A}\hat{\mathbf{H}}^T \hat{\mathbf{R}}^{-1}, \quad (11)$$

where  $\mathbf{A}$  is the analysis error covariance,

$$\mathbf{A} = (\hat{\mathbf{H}}^T \hat{\mathbf{R}}^{-1} \hat{\mathbf{H}} + \mathbf{B}^{-1})^{-1}. \quad (12)$$

Inserting equation (11) into (10) we find,

$$\mathbf{S} = \hat{\mathbf{R}}^{-1} \hat{\mathbf{H}} \mathbf{A} \hat{\mathbf{H}}^T. \quad (13)$$

We can therefore see the sensitivity of the analysis to observations is inversely proportional to the observation error and proportional to the analysis error. This means that the most influential observations are those with the smallest error variance providing information about regions of state space with the largest prior error [Cardinali et al., 2004]. It is possible to identify the observations that have the greatest influence over the length of the window by summing the absolute values of the columns of the influence matrix.

### 2.3.2 Degrees of freedom for signal

The degrees of freedom for signal (*dfs*) indicates the number of elements of the state that have been measured by the observations. If we consider a state vector  $\mathbf{x}$  with  $n$  elements (or  $n$  degrees of freedom) then the maximum value the *dfs* could obtain would be  $n$ , in this case all elements of the state would have been measured. Conversely if *dfs* = 0 then no elements of the state would have been measured by our observations [Fowler and Van Leeuwen, 2013].

For a symmetric positive definite prior and posterior error covariance matrices  $\mathbf{B}$  and  $\mathbf{A}$ , we can define the degrees of freedom for signal by means of a transform  $\mathbf{L}$  that reduces the prior error covariance matrix,  $\mathbf{B}$  to the  $n \times n$  identity [Fisher, 2003]. Each diagonal element of the transformed matrix  $\mathbf{B}$  then corresponds to a single degree of freedom with the trace being equal to  $n$ , the total degrees of freedom.

The transform  $\mathbf{L}$  can also be represented by  $\mathbf{Q}^T \mathbf{L}$ , where  $\mathbf{Q}$  is an orthogonal matrix. So that  $\mathbf{Q}^T \mathbf{L} \mathbf{B} \mathbf{L}^T \mathbf{Q} = \mathbf{Q}^T \mathbf{Q} = \mathbf{I}_{n \times n}$ . By defining  $\mathbf{Q}$  to be the matrix of the eigenvectors of  $\mathbf{L} \mathbf{A} \mathbf{L}^T$ , we reduce  $\mathbf{B}$  to the identity and  $\mathbf{L} \mathbf{A} \mathbf{L}^T$  to the diagonal matrix of its eigenvalues,  $\mathbf{\Lambda}$ . The eigenvalues  $\lambda_i$  of  $\mathbf{L} \mathbf{A} \mathbf{L}^T$  can be interpreted as the fractional reduction in uncertainty for the  $n$  state members. If an eigenvalue is close to zero the corresponding state member has been well observed, if it is close to one the corresponding state member has not been constrained by the assimilated observations [Stewart et al., 2008]. We then define the degrees of freedom for signal as,

$$\begin{aligned} dfs &= \text{trace}(\mathbf{Q}^T \mathbf{L} \mathbf{B} \mathbf{L}^T \mathbf{Q} - \mathbf{Q}^T \mathbf{L} \mathbf{A} \mathbf{L}^T \mathbf{Q}) \\ &= \text{trace}(\mathbf{I}_{n \times n} - \mathbf{\Lambda}) \\ &= n - \text{trace}(\mathbf{\Lambda}) \\ &= n - \text{trace}(\mathbf{L} \mathbf{A} \mathbf{L}^T) \\ &= n - \text{trace}(\mathbf{B}^{-1} \mathbf{A}). \end{aligned} \quad (14)$$

In Rodgers et al. [2000] it is shown that the  $dfs$  can also be calculated as the trace of the influence matrix  $\mathbf{S}$  (defined in section 2.3.1) with,

$$dfs = \text{trace}(\mathbf{S}) = \sum_i \lambda_i, \quad (15)$$

where  $\lambda_i$  is the  $i^{th}$  eigenvalue of  $\mathbf{S}$ .

### 2.3.3 Shannon information content

Shannon Information Content (SIC) is a measure of the reduction in entropy (uncertainty) given a set of observations. When a measurement is made, the entropy or uncertainty in our state decreases. The SIC of an observation is a measure of the factor by which the uncertainty decreases [Cover and Thomas, 1991]. We can define this using the prior,  $p(\mathbf{x})$ , and posterior,  $p(\mathbf{x}|\mathbf{y})$ , distributions. From Rodgers et al. [2000], for the Gaussian case SIC unsurprisingly becomes a function of the prior and posterior error covariance matrices with,

$$\text{SIC} = \frac{1}{2} \ln \frac{|\mathbf{B}|}{|\mathbf{A}|}. \quad (16)$$

The SIC can also be defined in terms of the eigenvalues of the influence matrix  $\mathbf{S}$  with,

$$\text{SIC} = -\frac{1}{2} \sum_i \ln |1 - \lambda_i| \quad (17)$$

where  $\lambda_i$  is the  $i^{th}$  eigenvalue of  $\mathbf{S}$ . In Eyre [1990] using SIC is shown to be beneficial over solely measuring the change in error variances before and after assimilation as the SIC also uses information about the change in error covariances. This is also true for the  $dfs$ .

## 3 Results

### 3.1 DALEC1 observability

DALEC1 is the original version of the DALEC2 model introduced in section (ref. D2 section). At the start of the PhD project work was undertaken with DALEC1 before the DALEC2 model was released. The version of DALEC1 used was an evergreen only model; further details of the model can be found in section (ref. D1 section) and Williams et al. [2005].

We initially consider observability of the DALEC1 state estimation system. DALEC1 is a smaller model and allows us to understand the concept of observability before moving onto work with the more complicated DALEC2 joint state and parameter estimation system. DALEC1 was implemented in a 4D-Var data assimilation scheme for state estimation, with the tangent linear model being computed analytically. Using this analytic implementation of the tangent linear model we can compute the observability of the model for differing sets of observations. We have the tangent linear model,

$$\mathbf{M}_i = \frac{\partial \mathbf{m}_{i-1 \rightarrow i}(\mathbf{x}_i)}{\partial \mathbf{x}_i} \quad (18)$$

$$= \begin{pmatrix} (1 - \theta_{fol}) + f_{fol}(1 - f_{auto})\zeta^i & 0 & 0 & 0 & 0 \\ f_{roo}(1 - f_{fol})(1 - f_{auto})\zeta^i & (1 - \theta_{roo}) & 0 & 0 & 0 \\ (1 - f_{roo})(1 - f_{fol})(1 - f_{auto})\zeta^i & 0 & (1 - \theta_{woo}) & 0 & 0 \\ \theta_{fol} & \theta_{roo} & 0 & (1 - (\theta_{min} + \theta_{lit})\chi^{i-1}) & 0 \\ 0 & 0 & \theta_{woo} & \theta_{min}\chi^{i-1} & (1 - \theta_{som}\chi^{i-1}) \end{pmatrix},$$

where  $\mathbf{x}_i = (C_{fol}^i, C_{roo}^i, C_{woo}^i, C_{lit}^i, C_{som}^i)^T$ ,  $\zeta^i = \partial GPP^i(C_{fol}^{i-1}, \Psi) / \partial C_{fol}^{i-1}$  and  $\chi^{i-1} = e^{\Theta T^{i-1}}$  with the parameters and symbols having the same meaning as in section (ref. D2 section).

We can use the linearised model with the linearised observation operator  $\mathbf{H}_i$  to form the matrix in equation (4) and compute the observability for a specific set of observations over a finite window. We will need at least 5 observations of any type for the system to be observable as the state  $\mathbf{x}_0$  is of size 5 in the DALEC1 state estimation case. We first consider the observability for 5 observations of LAI. For DALEC1 LAI takes the form

$$LAI^i = \frac{C_{fol}^i}{c_{lma}}. \quad (19)$$

We then have the linearised observation operator

$$\mathbf{H}_i = \frac{\partial LAI^i}{\partial \mathbf{x}_i} = \left( \frac{1}{c_{lma}} \quad 0 \quad 0 \quad 0 \quad 0 \right). \quad (20)$$

Using the linearised observation operator and the linear model from equation (18) we can compute  $\hat{\mathbf{H}}$  for 5 observations of LAI on consecutive time steps

$$\begin{aligned} \hat{\mathbf{H}} &= \begin{pmatrix} \mathbf{H}_0 \\ \mathbf{H}_1 \mathbf{M}_0 \\ \vdots \\ \mathbf{H}_4 \mathbf{M}_{3,0} \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{c_{lma}} & 0 & 0 & 0 & 0 \\ \frac{1}{c_{lma}} ((1 - \theta_{fol}) + f_{fol}(1 - f_{auto})\zeta^0) & 0 & 0 & 0 & 0 \\ \frac{1}{c_{lma}} \prod_{i=0}^1 ((1 - \theta_{fol}) + f_{fol}(1 - f_{auto})\zeta^i) & 0 & 0 & 0 & 0 \\ \frac{1}{c_{lma}} \prod_{i=0}^2 ((1 - \theta_{fol}) + f_{fol}(1 - f_{auto})\zeta^i) & 0 & 0 & 0 & 0 \\ \frac{1}{c_{lma}} \prod_{i=0}^3 ((1 - \theta_{fol}) + f_{fol}(1 - f_{auto})\zeta^i) & 0 & 0 & 0 & 0 \end{pmatrix}, \end{aligned} \quad (21)$$

so that no matter how many observations of LAI we add, our system will not be observable as the rows of  $\hat{\mathbf{H}}$  are all linearly dependant, so that  $\hat{\mathbf{H}}$  in this case has rank 1. We can repeat this for different observations to see for which observation types our system is observable.

From figure 1 we can see that our system is observable for 5 observations of the soil and organic matter carbon pool  $C_{som}$ . In figure 1 we have shown results for the rank of  $\hat{\mathbf{H}}$  when we have 5 observations in each case; this has also been tested with increasing numbers of observations being added to the system with the results from figure 1 remaining unchanged.

The system being observable for observations of  $C_{som}$  physically makes sense as all the carbon in the system that is not respired to the atmosphere eventually ends up in the soil and organic matter carbon pool,  $C_{som}$ , so that by taking observations of this pool we observe all the others. In a similar way  $\hat{\mathbf{H}}$  is also full rank for observations of NEE and ground respiration. We can see from the form of these observations in DALEC1 that they both contain indirect observations of  $C_{som}$  with NEE taking the form

$$NEE^i = -(1 - f_{auto})GPP^i(C_{fol}^{i-1}, \Psi) + \theta_{lit}C_{lit}e^{\Theta T^i} + \theta_{som}C_{som}e^{\Theta T^i} \quad (22)$$

with a corresponding linearised observation operator

$$\mathbf{H}_i = \frac{\partial NEE^i}{\partial \mathbf{x}_i} = \left( -(1 - f_{auto})\zeta^i \quad 0 \quad 0 \quad \theta_{lit}e^{\Theta T^i} \quad \theta_{som}e^{\Theta T^i} \right), \quad (23)$$

and for ground respiration

$$G_{resp}^i = \frac{1}{3}f_{auto}GPP^i(C_{fol}^{i-1}, \Psi) + \theta_{lit}C_{lit}e^{\Theta T^i} + \theta_{som}C_{som}e^{\Theta T^i} \quad (24)$$

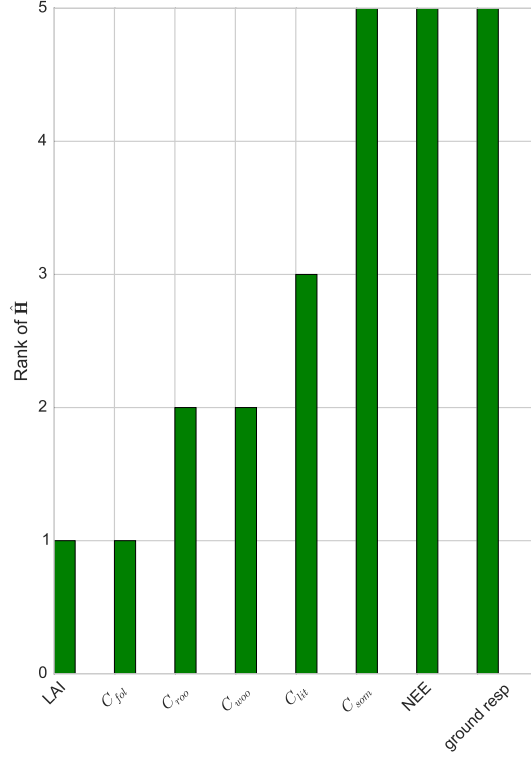


Figure 1: Rank of the observability matrix  $\hat{\mathbf{H}}$  for 5 observations of different types. The ranks shown here are computed analytically using SymPy [Joyner et al., 2012].

(here we have assumed the fraction of total autotrophic respiration from below ground to be  $\frac{1}{3}$ ) with a corresponding linearised observation operator

$$\mathbf{H}_i = \frac{\partial G_{resp}^i}{\partial \mathbf{x}_i} = \begin{pmatrix} \frac{1}{3}f_{auto}\zeta^i & 0 & 0 & \theta_{lit}e^{\Theta T^i} & \theta_{som}e^{\Theta T^i} \end{pmatrix}. \quad (25)$$

At flux tower sites NEE is the most observed quantity, these results give us confidence that we can construct a unique solution when working with flux tower data. We will further explore the concept of observability for the joint parameter and state estimation case with DALEC2 in section 3.2.

### 3.2 DALEC2 observability

For DALEC2 we perform joint parameter and state estimation and have an augmented state of size  $n = 23$ . The augmented state is made up of the 6 carbon pool state members and 17 model parameters as described in section (ref. D2 section). As we are also estimating the parameters of DALEC2 the concept of observability for our system is closely linked to the concept of identifiability [Navon, 1998]. A system is identifiable if given observations of the state variables and knowledge of the model dynamics it is possible to obtain a unique deterministic set of model parameter values [Ljung, 1998]. If a model parameter is not observable it will not be identifiable [Jacquez and Greif, 1985]. It is therefore useful to compute the observability of the DALEC2 joint parameter and state estimation system.

We compute observability in the same way as in section 3.1 by finding the rank of  $\hat{\mathbf{H}}$  for a given set of observations. For the state and parameter estimation case we cannot compute the observability of the system

analytically, it is therefore important to check that the numerical calculation of the rank of  $\hat{\mathbf{H}}$  for DALEC1 is equal to the rank when calculated analytically. This will give us confidence that our implementation of the numeric rank is correct for DALEC2 when applied to a well-conditioned problem as the implementation is the same in both cases. We have tested our numeric implementation for the state estimation case with DALEC1 and find the same results for the rank of  $\hat{\mathbf{H}}$  as for the analytic case, as shown in table 1. We calculate the rank of the  $\hat{\mathbf{H}}$  matrix using a singular value decomposition (SVD) which can have issues if the condition number of  $\hat{\mathbf{H}}$  is large [Paige, 1981]. This is a problem we encounter in the DALEC2 case when trying to calculate the rank of  $\hat{\mathbf{H}}$  directly.

Observation	Rank of $\hat{\mathbf{H}}$ (numeric)	Rank of $\hat{\mathbf{H}}$ (analytic)
LAI	1	1
$C_{fol}$	1	1
$C_{roo}$	2	2
$C_{woo}$	2	2
$C_{lit}$	3	3
$C_{som}$	5	5
NEE	5	5
$G_{resp}$	5	5

Table 1: Rank of  $\hat{\mathbf{H}}$  for 5 observations of different types for both numeric and analytic implementations with DALEC1.

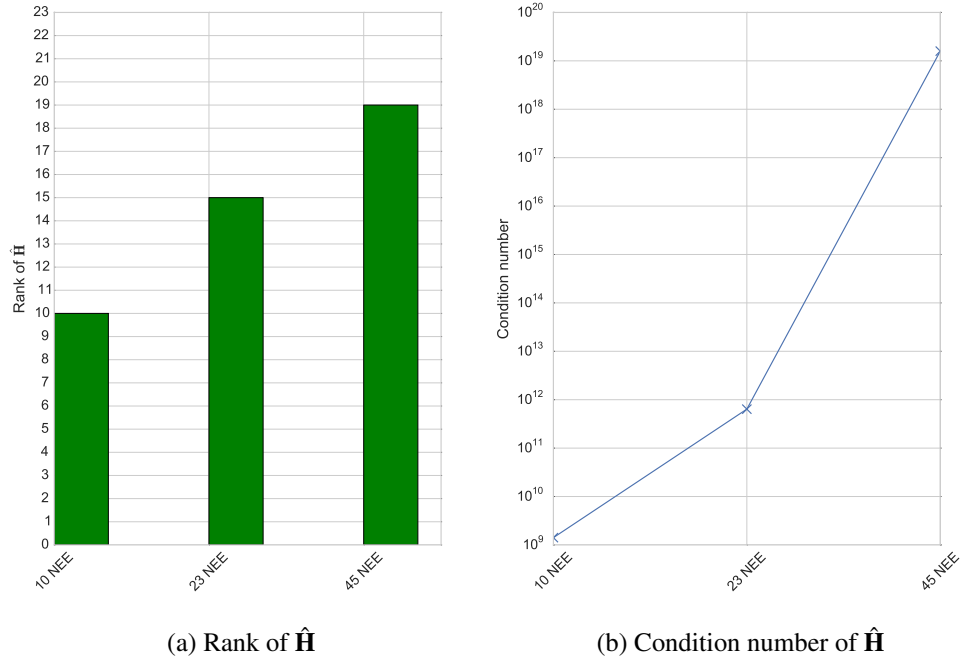


Figure 2: Observability of DALEC2 for  $\hat{\mathbf{H}}$  with an increasing number of NEE observations displayed alongside the condition number for the  $\hat{\mathbf{H}}$  matrices.

Figure 2 highlights the problems we have calculating the rank of the  $\hat{\mathbf{H}}$  matrix for the DALEC2 joint



parameter and state estimation case. In figure 2a we see that for 23 observations of NEE our system is unobservable as we have a rank deficient  $\hat{\mathbf{H}}$ . However, we cannot trust the rank calculation of  $\hat{\mathbf{H}}$  in this case. Figure 2b shows that for 23 observations of NEE,  $\hat{\mathbf{H}}$  has a condition number in the order of  $10^{19}$ . The condition number of a matrix corresponds to the ratio of the largest to the smallest singular values. A condition number of this size means that we have very small singular values. In the calculation of the rank of a matrix using an SVD we define the rank to be the number of singular values greater than the threshold  $\text{tol} = \max(S) * \max(n, m) * \text{eps}$  [Press et al., 2007], where  $S$  is the vector of singular values,  $n$  and  $m$  are the rows and columns of the matrix whose rank we wish to calculate and  $\text{eps}$  is the machine accuracy for the datatype of  $S$  (In this case a double-precision float with  $\text{eps} = 2.22\text{e-}16$ ). For 23 observations of NEE,  $\hat{\mathbf{H}}$  is classed as being rank deficient as  $\text{tol} = 1.02\text{e-}10$  and the three smallest singular values of  $\hat{\mathbf{H}}$  are  $[1.39\text{e-}11, 7.84\text{e-}15, 1.46\text{e-}15]$  but here we are working past the accuracy of the computer and so cannot have confidence that  $\hat{\mathbf{H}}$  is rank deficient in this case.

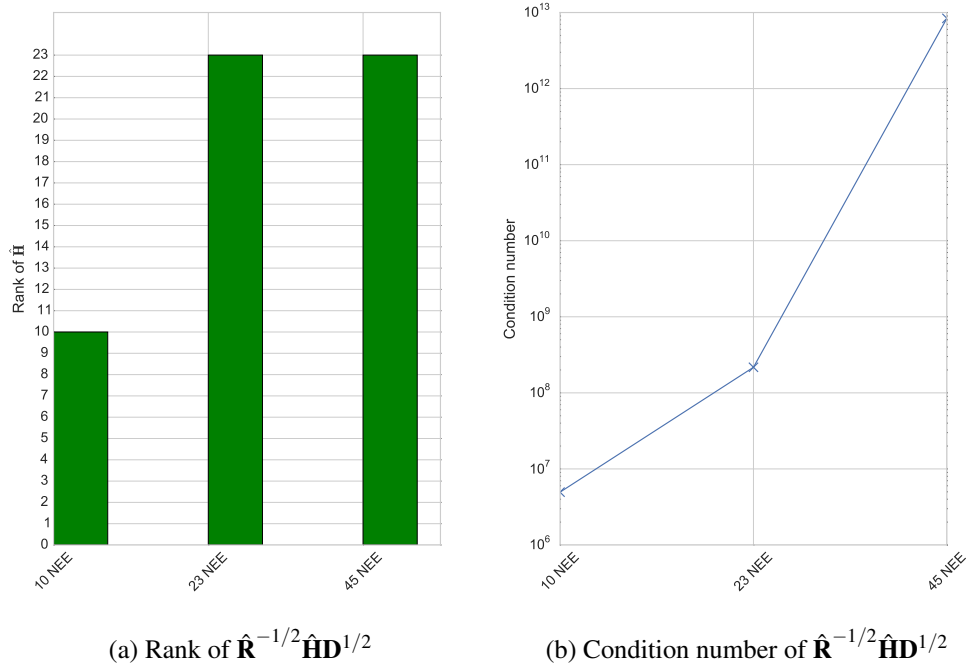


Figure 3: Observability of the CVT DALEC2 for  $\hat{\mathbf{R}}^{-1/2} \hat{\mathbf{H}} \mathbf{D}^{1/2}$  with an increasing number of NEE observations displayed alongside the condition number for the  $\hat{\mathbf{R}}^{-1/2} \hat{\mathbf{H}} \mathbf{D}^{1/2}$  matrices.

In order to address the problem of ill-conditioning of the  $\hat{\mathbf{H}}$  matrix we can instead calculate the rank of the control variable transform observability matrix,  $\hat{\mathbf{R}}^{-1/2} \hat{\mathbf{H}} \mathbf{D}^{1/2}$ , where the symbols have the same meaning as in section (ref DA CVT section,  $\mathbf{D} = \text{diag}\{\mathbf{B}\}$ ). The rank of  $\hat{\mathbf{R}}^{-1/2} \hat{\mathbf{H}} \mathbf{D}^{1/2}$  and  $\hat{\mathbf{H}}$  are the same since  $\hat{\mathbf{R}}$  and  $\mathbf{D}$  are both full rank matrices. The results using this new better conditioned matrix are shown in Figure 3. From Figure 3b we can see this matrix is much better conditioned than  $\hat{\mathbf{H}}$ , and for 23 observations of NEE we now have an observable system. Although the condition numbers here are still large we can have more confidence in these results as we are working within the precision of the computer.

### 3.2.1 Observability for observations randomly distributed through time

In the previous experiments we have considered increasing numbers of NEE observations taken on adjacent days. It is also useful to consider the observability of the system when we have a number of observations randomly distributed throughout a time window. This is more consistent with what we expect from the real data we have to work with.

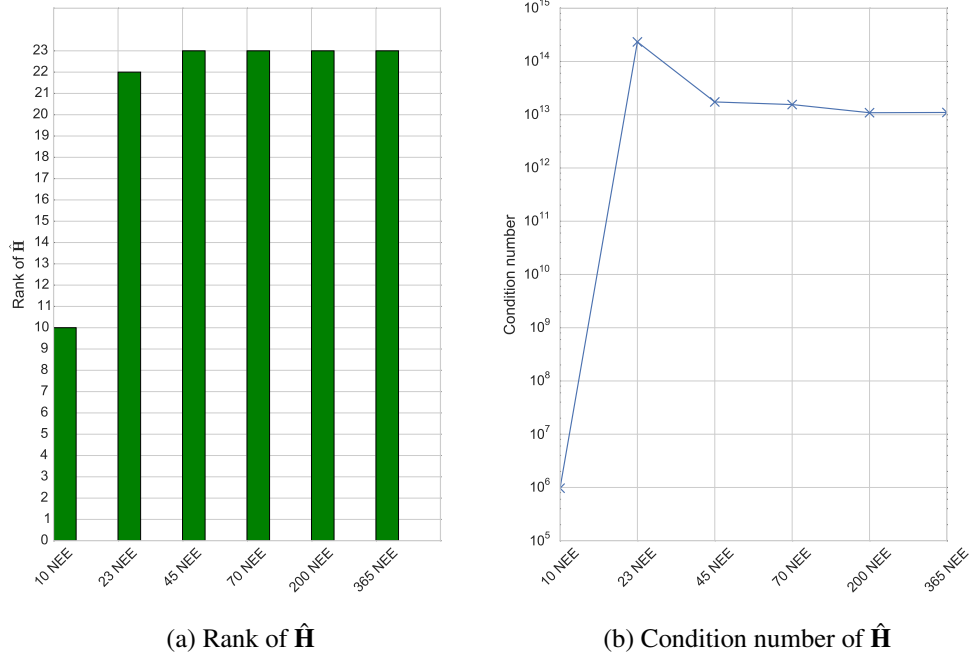


Figure 4: Observability of DALEC2 for a  $\hat{\mathbf{H}}$  with an increasing number of NEE observations randomly distributed through a 1 year assimilation window (left). Condition number for the  $\hat{\mathbf{H}}$  matrices (right).

Figure 4 shows the observability for an increasing number of observations distributed through a 1 year assimilation window. In this case we are using the matrix  $\hat{\mathbf{H}}$  and not the CVT observability matrix. In figure 4 we see that having the observations randomly distributed throughout a 1 year assimilation window has improved the conditioning of  $\hat{\mathbf{H}}$  in comparison to figure 2. This is due to the observations being randomly distributed rather than adjacent. The rows of  $\hat{\mathbf{H}}$  are more distinct when being evolved to different times in the year by the tangent linear model rather than evolved to adjacent days only. However, we still have a rank deficient  $\hat{\mathbf{H}}$  for the 23 NEE observation case. From figure 4b we see that this is the case where the condition number peaks. As we add more randomly distributed observations the condition number of  $\hat{\mathbf{H}}$  is reduced by an order of  $10^2$  and we have a full rank  $\hat{\mathbf{H}}$ .

In figure 5 we again see that using the CVT observability matrix has much improved the conditioning of the problem in comparison to figure 4. We now see that the DALEC2 system is observable when we have 23 observations of NEE randomly distributed throughout the 1 year assimilation window. We have more confidence that this is the case as the condition numbers for the CVT observability matrix are almost half the values of those for  $\hat{\mathbf{H}}$ . We again see a similar pattern in figure 5 for the condition numbers with a peak for 23 NEE observations and then a reduction of order  $10^2$  when more observations are added.

We have tested the observability of the system for observations of NEE when we have different driving data, linearising around different states and with different distributions of observations throughout our assimilation window and in every case we have an observable system given an adequate number of NEE

observations (at least 23). We can therefore have confidence that for the available data, typically 60-80 observations of daily NEE for any year's window, we can construct a unique solution with the observations alone.

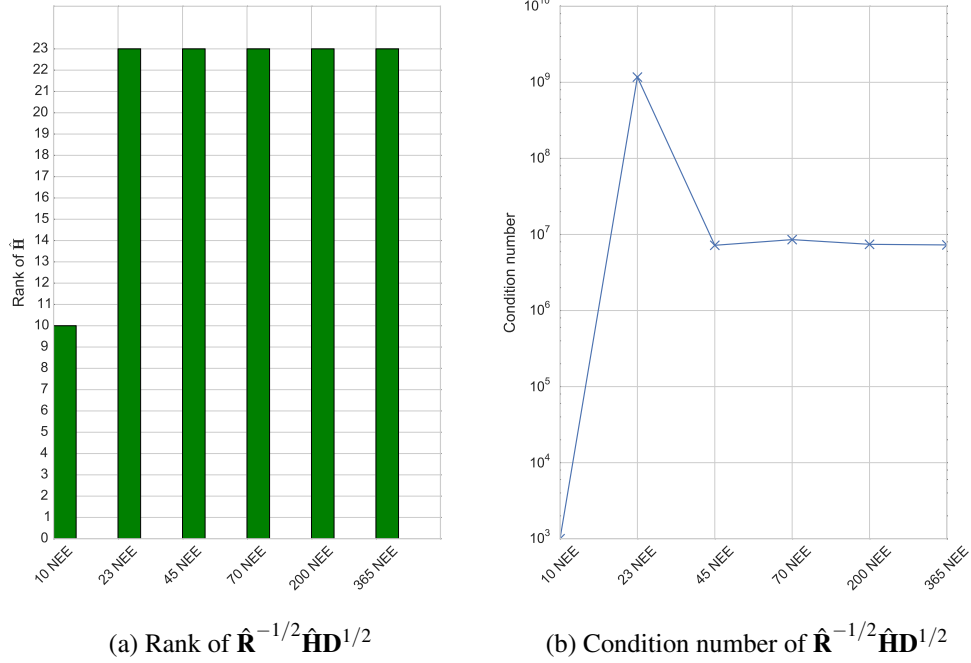


Figure 5: Observability of the CVT DALEC2 system for  $\hat{\mathbf{R}}^{-1/2} \hat{\mathbf{H}} \mathbf{D}^{1/2}$  with an increasing number of NEE observations randomly distributed through a 1 year assimilation window (left). Condition number for the  $\hat{\mathbf{R}}^{-1/2} \hat{\mathbf{H}} \mathbf{D}^{1/2}$  matrices (right).

### 3.3 DALEC1 information content

#### 3.3.1 Information content for a single observation

For the DALEC1 state estimation we can calculate the analytic representation of the information content measures discussed in section 2.3. This will allow us to understand how the information content changes for differing numbers of observations, different observation types and the effect of including observation error correlations in the assimilation scheme, before moving onto work with the larger DALEC2 joint parameter and state estimation case. For these experiments the elements of the state vector have corresponding background standard deviations  $\sigma_{cfol,b}$ ,  $\sigma_{croo,b}$ ,  $\sigma_{cwoo,b}$ ,  $\sigma_{clit,b}$ ,  $\sigma_{csom,b}$ . We then have

$$\mathbf{B} = \begin{pmatrix} \sigma_{cfol,b}^2 & 0 & 0 & 0 & 0 \\ 0 & \sigma_{croo,b}^2 & 0 & 0 & 0 \\ 0 & 0 & \sigma_{cwoo,b}^2 & 0 & 0 \\ 0 & 0 & 0 & \sigma_{clit,b}^2 & 0 \\ 0 & 0 & 0 & 0 & \sigma_{csom,b}^2 \end{pmatrix}. \quad (26)$$

We begin by considering the Shannon Information Content (SIC) and degrees of freedom for signal

(*dfs*) for a single observation of LAI. We have the linearised observation operator

$$\mathbf{H}_i = \frac{\partial LAI^i}{\partial \mathbf{x}_i} = \frac{\partial}{\partial \mathbf{x}_i} \left( \frac{C_{fol}^i}{c_{lma}} \right) = \begin{pmatrix} \frac{1}{c_{lma}} & 0 & 0 & 0 & 0 \end{pmatrix}. \quad (27)$$

As we have a single observation at one time, our observation error covariance matrix,  $\mathbf{R}$ , is just the variance of our observation of LAI at time  $t_0$  ( $\sigma_{LAI,o}^2$ ). Therefore,

$$\mathbf{R}_i = \sigma_{LAI,o}^2. \quad (28)$$

We then have from equation (ref. A matrix eqn),

$$\begin{aligned} \mathbf{A} &= (\mathbf{J}'')^{-1} \\ &= (\mathbf{B}^{-1} + \hat{\mathbf{H}}^T \hat{\mathbf{R}}^{-1} \hat{\mathbf{H}})^{-1} \\ &= (\mathbf{B}^{-1} + \mathbf{H}_0^T \mathbf{R}_0^{-1} \mathbf{H}_0)^{-1} \\ &= \begin{pmatrix} \frac{c_{lma}^2 \sigma_{LAI,o}^2 \sigma_{cfol,b}^2}{\sigma_{cfol,b}^2 + c_{lma}^2 \sigma_{LAI,o}^2} & 0 & 0 & 0 & 0 \\ 0 & \sigma_{croo,b}^2 & 0 & 0 & 0 \\ 0 & 0 & \sigma_{cwo,b}^2 & 0 & 0 \\ 0 & 0 & 0 & \sigma_{clit,b}^2 & 0 \\ 0 & 0 & 0 & 0 & \sigma_{csom,b}^2 \end{pmatrix}. \end{aligned} \quad (29)$$

We can now derive the SIC and *dfs* using equation (16) and (14) as,

$$\text{SIC} = \frac{1}{2} \ln \frac{|\mathbf{B}|}{|\mathbf{A}|} = \frac{1}{2} \ln \frac{(c_{lma}^2 \sigma_{LAI,o}^2 + \sigma_{cfol,b}^2)}{c_{lma}^2 \sigma_{LAI,o}^2} = \frac{1}{2} \ln \left( 1 + \frac{\sigma_{cfol,b}^2}{c_{lma}^2 \sigma_{LAI,o}^2} \right) \quad (30)$$

and

$$dfs = n - \text{tr}(\mathbf{B}^{-1} \mathbf{A}) = 5 - \left( \frac{c_{lma}^2 \sigma_{LAI,o}^2}{(c_{lma}^2 \sigma_{LAI,o}^2 + \sigma_{cfol,b}^2)} + 4 \right) = 1 - \frac{c_{lma}^2 \sigma_{LAI,o}^2}{(c_{lma}^2 \sigma_{LAI,o}^2 + \sigma_{cfol,b}^2)}. \quad (31)$$

We see that in general for a direct observation of any of the carbon pools  $C$  we have

$$\text{SIC} = \frac{1}{2} \ln \left( 1 + \frac{\sigma_{c,b}^2}{\sigma_{c,o}^2} \right) \quad (32)$$

and

$$dfs = 1 - \frac{\sigma_{c,o}^2}{(\sigma_{c,o}^2 + \sigma_{c,b}^2)}, \quad (33)$$

where  $\sigma_{c,o}$  and  $\sigma_{c,b}$  are the observation and background standard deviations respectively, corresponding to any of the 5 carbon pools. We see the SIC for a single observation of one of the carbon pools is dependent on the ratio between the observation and background variances. The carbon pool observation which will give us the highest SIC is the observation with the largest ratio  $\frac{\sigma_{c,b}^2}{\sigma_{c,o}^2}$ . This is also the case for *dfs*. Assuming a fixed background standard deviation, the carbon pool observation which will give us the highest information content is the pool which we can measure most accurately, as expected. From equations (30) and (31) for an observation of LAI the information content is also dependent on  $c_{lma}$  the parameter describing leaf mass area.

Next we consider the information content in a single observation of NEE. We have

$$\mathbf{H}_i = \frac{\partial NEE^i}{\partial \mathbf{x}_i} = \begin{pmatrix} -(1 - f_{auto}) \zeta^i & 0 & 0 & \theta_{lit} e^{\Theta T^i} & \theta_{som} e^{\Theta T^i} \end{pmatrix} \quad (34)$$

and

$$\mathbf{R}_i = \sigma_{NEE,o}^2. \quad (35)$$

We then find

$$\text{SIC} = \frac{1}{2} \ln \left( 1 + \frac{(f_{auto} - 1)^2 (\zeta^i)^2 \sigma_{cfol,b}^2 + (e^{\Theta T^i})^2 (\theta_{som}^2 \sigma_{csom,b}^2 + \theta_{lit}^2 \sigma_{clit,b}^2)}{\sigma_{NEE,o}^2} \right) \quad (36)$$

and

$$dfs = 1 - \frac{\sigma_{NEE,o}^2}{(f_{auto} - 1)^2 (\zeta^i)^2 \sigma_{cfol,b}^2 + (e^{\Theta T^i})^2 (\theta_{som}^2 \sigma_{csom,b}^2 + \theta_{lit}^2 \sigma_{clit,b}^2) + \sigma_{NEE,o}^2}. \quad (37)$$

We see that Equations (36) and (37) have a similar form to Equations (32) and (33). The information content is again dependent on the ratio between the observation and background variances. The information content for the observations of NEE is also dependent on the magnitude of the first derivative of GPP with respect to  $C_{fol}$  and the magnitude of the exponential function of temperature controlling the rate of heterotrophic respiration,  $e^{\Theta T^i}$ . Both the first derivative of GPP and  $e^{\Theta T^i}$  will be of greater magnitude when we have higher mean daily temperatures. This means that observations of NEE made at times with higher temperatures will have higher information content and more of an impact on data assimilation results.

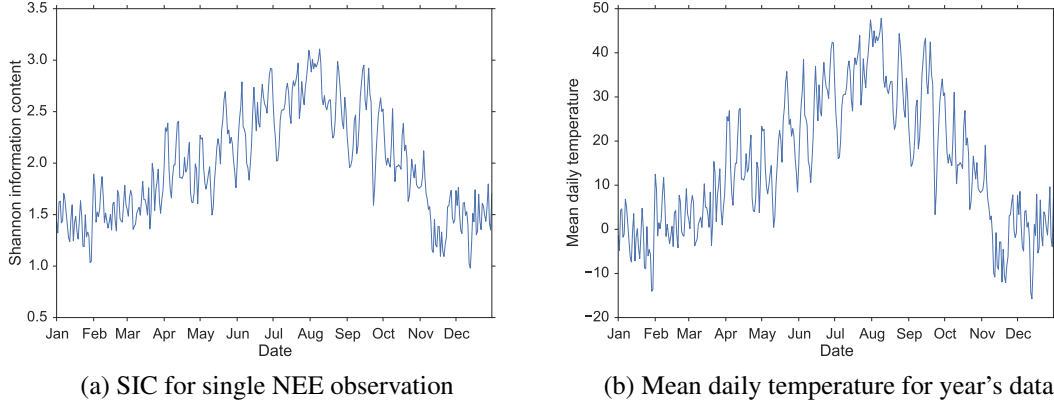


Figure 6: SIC for a single NEE observation changing throughout a year's window using driving data from a forest of ponderosa pine in Oregon taken in 2007 (left). Mean daily temperature for the same site and period (right).

In Figure 6 we show how closely SIC is related to mean daily temperature for NEE observations throughout a year's window using daily driving data from a forest of ponderosa pine in Oregon (as described in section 2.1). Higher information content in summer observations of NEE makes physical sense. In summertime fluxes of carbon through the forest ecosystem are of greater magnitude than in winter, with more photosynthesis and respiration occurring. This gives us more information about the fluxes of carbon through our system in summertime observations of NEE. It is important to consider this result when planning for down time or routine maintenance at flux tower sites measuring NEE. The temperature dependence of information content will also hold true for other observations whose observation operators include the nonlinear temperature term controlling heterotrophic respiration. These observations include ground respiration, measured using soil respiration chambers, and total ecosystem respiration, estimated from nighttime NEE measurements.

In Figure 6a we have assumed constant prior and observation standard deviations. This is an accurate assumption for our prior errors. However, it has been shown that NEE errors are heteroscedastic [Richardson

et al., 2008] and therefore scale with the magnitude of the flux. This would reduce the magnitude of the results shown in Figure 6a, as our standard deviation in observations of summer NEE would be larger, reducing the information content.

For Figure 6a we have used a numerical implementation in Python to calculate the SIC varying for 365 days of driving data. It is important to test our numerical implementation for correctness. In table 2 we show the SIC and  $dfs$  calculated both analytically and numerically. From this table we can see that both analytic and numerical implementations give us the same result to 15 or more significant figures. This gives us a degree of confidence that our implementation is also correct for DALEC2. In this table we have assumed constant prior and observation standard deviations for the carbon pools.

Obs.	SIC analytic value	SIC numeric value	$dfs$ analytic value	$dfs$ numeric value
NEE	0.0209343224569909	0.0209343224569913	0.0410042587324008	0.0410042587324008
$C_{fol}$	0.8047189562170501	0.8047189562170515	0.7999999999999998	0.7999999999999998
$C_{roo}$	0.1838623900626585	0.1838623900626572	0.3076923076923075	0.3076923076923083
$C_{woo}$	0.8047189562170501	0.8047189562170515	0.7999999999999998	0.7999999999999998
$C_{lit}$	0.1838623900626585	0.1838623900626572	0.3076923076923075	0.3076923076923074
$C_{som}$	0.1838623900626585	0.1838623900626572	0.3076923076923075	0.3076923076923074

Table 2: Correctness tests showing numeric and analytic values of information content calculated using 2007 driving data and parameter values from an Oregon ponderosa pine forest.

### 3.3.2 Information content for observations at a single time

We next consider the SIC when we have more than one observation at a single time. Here we will investigate the representation of information content when assimilating an observation of NEE with an observation of a carbon pool state member. We begin with a single observation of NEE and an observation of  $C_{fol}$ . We have the linearised observation operator,

$$\mathbf{H}_i = \frac{\partial}{\partial \mathbf{x}_i} (NEE^i, C_{fol}^i) = \begin{pmatrix} -(1-f_{auto})\xi^i & 0 & 0 & \theta_{lit}e^{\Theta T^i} & \theta_{som}e^{\Theta T^i} \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix} \quad (38)$$

and observation error covariance matrix

$$\mathbf{R}_i = \begin{pmatrix} \sigma_{NEE,o}^2 & 0 \\ 0 & \sigma_{C_{fol},o}^2 \end{pmatrix}. \quad (39)$$

We then find,

$$\text{SIC} = \frac{1}{2} \ln \left( 1 + \frac{\sigma_{C_{fol},b}^2}{\sigma_{C_{fol},o}^2} + \frac{\xi^i}{\sigma_{NEE,o}^2} + \frac{\sigma_{C_{fol},b}^2 (e^{\Theta T^i})^2 (\theta_{som}^2 \sigma_{csom,b}^2 + \theta_{lit}^2 \sigma_{clit,b}^2)}{\sigma_{NEE,o}^2 \sigma_{C_{fol},o}^2} \right) \quad (40)$$

where,  $\xi^i = (f_{auto} - 1)^2 (\xi^i)^2 \sigma_{C_{fol},b}^2 + (e^{\Theta T^i})^2 (\theta_{som}^2 \sigma_{csom,b}^2 + \theta_{lit}^2 \sigma_{clit,b}^2)$ . From equation (40) we can see that we have the first order terms for both NEE and  $C_{fol}$  as in equations (32) and (36). We also have a second order term for the combination of these observations. We can repeat this for the other carbon pools and find for  $\mathbf{H}_i = \frac{\partial}{\partial \mathbf{x}_i} (NEE^i, C_{roo}^i)$ ,

$$\text{SIC} = \frac{1}{2} \ln \left( 1 + \frac{\sigma_{C_{roo},b}^2}{\sigma_{C_{roo},o}^2} + \frac{\xi^i}{\sigma_{NEE,o}^2} + \frac{\sigma_{C_{roo},b}^2 ((f_{auto} - 1)^2 (\xi^i)^2 \sigma_{C_{fol},b}^2 + (e^{\Theta T^i})^2 (\theta_{som}^2 \sigma_{csom,b}^2 + \theta_{lit}^2 \sigma_{clit,b}^2))}{\sigma_{NEE,o}^2 \sigma_{C_{roo},o}^2} \right), \quad (41)$$

for  $\mathbf{H}_i = \frac{\partial}{\partial \mathbf{x}_i}(NEE^i, C_{woo}^i)$ ,

$$SIC = \frac{1}{2} \ln \left( 1 + \frac{\sigma_{cwoo,b}^2}{\sigma_{cwoo,o}^2} + \frac{\xi^i}{\sigma_{NEE,o}^2} + \frac{\sigma_{cwoo,b}^2((f_{auto} - 1)^2(\xi^i)^2\sigma_{cfol,b}^2 + (e^{\Theta T^i})^2(\theta_{som}^2\sigma_{csom,b}^2 + \theta_{lit}^2\sigma_{clit,b}^2))}{\sigma_{NEE,o}^2\sigma_{cwoo,o}^2} \right), \quad (42)$$

for  $\mathbf{H}_i = \frac{\partial}{\partial \mathbf{x}_i}(NEE^i, C_{lit}^i)$ ,

$$SIC = \frac{1}{2} \ln \left( 1 + \frac{\sigma_{clit,b}^2}{\sigma_{clit,o}^2} + \frac{\xi^i}{\sigma_{NEE,o}^2} + \frac{\sigma_{clit,b}^2((f_{auto} - 1)^2(\xi^i)^2\sigma_{cfol,b}^2 + (e^{\Theta T^i})^2\theta_{som}^2\sigma_{csom,b}^2)}{\sigma_{NEE,o}^2\sigma_{clit,o}^2} \right) \quad (43)$$

and for  $\mathbf{H}_i = \frac{\partial}{\partial \mathbf{x}_i}(NEE^i, C_{som}^i)$ ,

$$SIC = \frac{1}{2} \ln \left( 1 + \frac{\sigma_{csom,b}^2}{\sigma_{csom,o}^2} + \frac{\xi^i}{\sigma_{NEE,o}^2} + \frac{\sigma_{csom,b}^2((f_{auto} - 1)^2(\xi^i)^2\sigma_{cfol,b}^2 + (e^{\Theta T^i})^2\theta_{lit}^2\sigma_{clit,b}^2)}{\sigma_{NEE,o}^2\sigma_{csom,o}^2} \right). \quad (44)$$

Assuming constant prior and observation standard deviations across our carbon pool observations we see that the information content will be largest in equations (41) and (42). For both  $\mathbf{H}_i = \frac{\partial}{\partial \mathbf{x}_i}(NEE^i, C_{roo}^i)$  and  $\mathbf{H}_i = \frac{\partial}{\partial \mathbf{x}_i}(NEE^i, C_{woo}^i)$  we have an extra term in the numerator for our second order term corresponding to the combination of the two observations. If we consider the linearised observation operator for both these cases,

$$\mathbf{H}_i = \frac{\partial}{\partial \mathbf{x}_i}(NEE^i, C_{roo}^i) = \begin{pmatrix} -(1 - f_{auto})\xi^i & 0 & 0 & \theta_{lit}e^{\Theta T^i} & \theta_{som}e^{\Theta T^i} \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix} \quad (45)$$

and

$$\mathbf{H}_i = \frac{\partial}{\partial \mathbf{x}_i}(NEE^i, C_{woo}^i) = \begin{pmatrix} -(1 - f_{auto})\xi^i & 0 & 0 & \theta_{lit}e^{\Theta T^i} & \theta_{som}e^{\Theta T^i} \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix}, \quad (46)$$

we can see that these observations provide an orthogonal constraint to the observation of NEE. Neither of these pools are observed with a single observation of NEE. The information content being greater when assimilating  $C_{roo}$  or  $C_{woo}$  alongside NEE is therefore expected.

In practice we cannot assume constant prior and observation errors across the different carbon pools. Root carbon is hard to measure accurately [Brown, 2002]. However, woody biomass ( $C_{woo}$ ) is regularly measured using mensuration [Husch et al., 2002] or point-centred quarter methods [Dahdouh-Guebas and Koedam, 2006] at forest sites. Advancements in Light Detection And Ranging (LiDAR) scanning [Lefsky et al., 1999] mean that we have increasingly more accurate observations of woody biomass. The European Space Agency BIOMASS mission [Le Toan et al., 2011] will also provide a much more abundant source of woody biomass measurements in the future. If we consider NEE to be the main observation currently used in ecosystem data assimilation, then the increasing number of available woody biomass measurements will benefit assimilation schemes greatly.

### 3.3.3 Information content in successive observations

In section 3.3.1 we investigate the information in observation for DALEC1 at a single time. In this section we will consider successive observations in time. It has been shown that the SIC in observations is additive with successive observations in time. The proof for this can be found in appendix A.1 of Fowler and Jan Van Leeuwen [2012]. We can see this if we calculate the SIC for successive observations of foliar carbon,  $C_{fol}$ . We have the linearised observation operator and observation error covariance matrix at time  $t_i$ ,

$$\mathbf{H}_i = \frac{\partial C_{fol}^i}{\partial \mathbf{x}_i} = (1 \ 0 \ 0 \ 0 \ 0) \quad \text{and} \quad \mathbf{R}_i = \sigma_{cfol,o}^2. \quad (47)$$

For two successive observations of  $C_{fol}$  we have,

$$\hat{\mathbf{H}} = \begin{pmatrix} \mathbf{H}_0 \\ \mathbf{H}_1 \mathbf{M}_0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ (1 - \theta_{fol}) + f_{fol}(1 - f_{auto})\zeta^0 & 0 & 0 & 0 & 0 \end{pmatrix} \quad (48)$$

and

$$\hat{\mathbf{R}} = \begin{pmatrix} \mathbf{R}_0 & 0 \\ 0 & \mathbf{R}_1 \end{pmatrix} = \begin{pmatrix} \sigma_{cfol,o}^2 & 0 \\ 0 & \sigma_{cfol,o}^2 \end{pmatrix}. \quad (49)$$

We then have,

$$SIC = \frac{1}{2} \ln \frac{|\mathbf{B}|}{|\mathbf{A}|} = \frac{1}{2} \ln \left( 1 + \frac{\sigma_{cfol,b}^2}{\sigma_{cfol,o}^2} + \frac{\sigma_{cfol,b}^2 \eta_0^2}{\sigma_{cfol,o}^2} \right), \quad (50)$$

where  $\eta_i = (1 - \theta_{fol}) + f_{fol}(1 - f_{auto})\zeta^i$ . We see this is similar to equation (32) for the SIC of a single carbon pool observation but with an added term evolved by the linearised model. Here the second term is multiplied by  $\eta_0^2$  which is the square of the first element of the linearised model  $\mathbf{M}_0$ . We can continue adding more observations at successive times. For three observations at successive times we have,

$$SIC = \frac{1}{2} \ln \left( 1 + \frac{\sigma_{cfol,b}^2}{\sigma_{cfol,o}^2} + \frac{\sigma_{cfol,b}^2 \eta_0^2}{\sigma_{cfol,o}^2} + \frac{\sigma_{cfol,b}^2 \eta_0^2 \eta_1^2}{\sigma_{cfol,o}^2} \right), \quad (51)$$

for four,

$$SIC = \frac{1}{2} \ln \left( 1 + \frac{\sigma_{cfol,b}^2}{\sigma_{cfol,o}^2} + \frac{\sigma_{cfol,b}^2 \eta_0^2}{\sigma_{cfol,o}^2} + \frac{\sigma_{cfol,b}^2 \eta_0^2 \eta_1^2}{\sigma_{cfol,o}^2} + \frac{\sigma_{cfol,b}^2 \eta_0^2 \eta_1^2 \eta_2^2}{\sigma_{cfol,o}^2} \right). \quad (52)$$

Using a simple proof by induction we find that for  $n$  successive observations we have,

$$SIC \text{ for } n \text{ successive observations of } C_{fol} = \frac{1}{2} \ln \left( 1 + \frac{\sigma_{cfol,b}^2}{\sigma_{cfol,o}^2} \left( 1 + \sum_{k=0}^{n-2} \prod_{i=0}^k \eta_i^2 \right) \right) \quad (53)$$

This demonstrates that SIC is additive for successive observations in time. In Figure 7 we have plotted the SIC and  $dfs$  for increasing numbers of observations of  $C_{fol}$ , using a year of meteorological driving data from a pine stand in Oregon. We see that as successive observations are added the information content tends to a limit where we are adding no new information with extra observations of  $C_{fol}$ . For  $dfs$  this limit is one as we are only observing a single degree of freedom so cannot constrain more than a single element of the state. For SIC we add a decreasing amount of information as observations are made further away from the initial state. We find similar results for all other carbon pools. This suggests making observations of any individual carbon pool for a forest site too often is not cost effective as after just a few observations the information you are adding to your system begins to decrease.

In section 3.3.1 it was shown that observations of NEE made during the summer had significantly higher information content than those made during winter for an evergreen forest site. In figure 8 we show that 27 days of successive winter NEE observations (made from January 1<sup>st</sup> 2007) are required to give the same information content as a single summer observation of NEE (taken on 22<sup>nd</sup> June 2007).

### 3.3.4 Effect of time correlations between observation errors on information content

In chapter (ref. 1st results chapter) we have investigated the effect of including correlations in time between observation errors on the results from data assimilation with DALEC2. We can see the effect on the analytic representation of information content for two successive observations of NEE when including



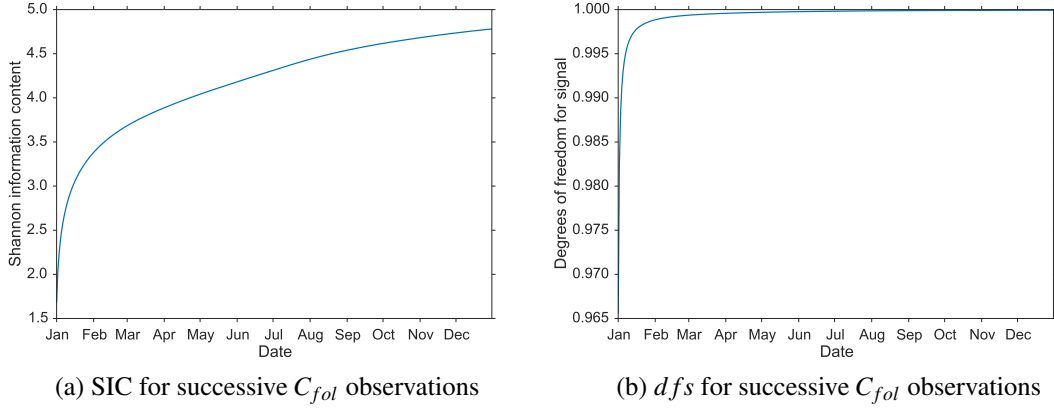


Figure 7: SIC and  $dfs$  for as successive  $C_{fol}$  observations are added throughout a year's window using driving data from a pine stand in Oregon taken in 2007.

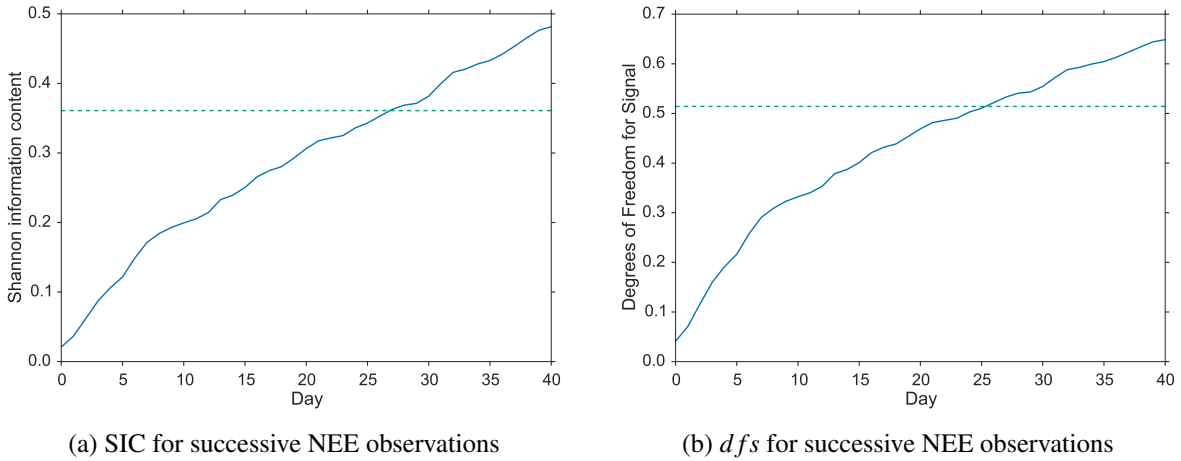


Figure 8: Blue line: SIC and  $dfs$  for as successive NEE observations are added for 40 days from the 1<sup>st</sup> January 2007 using driving data from a pine stand in Oregon, green dotted line: SIC and  $dfs$  for a single summer observation of NEE made on 22<sup>nd</sup> June 2007.

an off-diagonal correlation term in the matrix  $\hat{\mathbf{R}}$ . So that  $\hat{\mathbf{R}} = \hat{\mathbf{D}}\mathbf{C}\hat{\mathbf{D}}^T$ , where  $\hat{\mathbf{D}}$  is the diagonal matrix of observation standard deviations and  $\mathbf{C}$  is a correlation matrix of the same shape. We then have

$$\hat{\mathbf{R}} = \hat{\mathbf{D}}\mathbf{C}\hat{\mathbf{D}}^T = \begin{pmatrix} \sigma_{nee,o} & 0 \\ 0 & \sigma_{nee,o} \end{pmatrix} \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \begin{pmatrix} \sigma_{nee,o} & 0 \\ 0 & \sigma_{nee,o} \end{pmatrix} = \begin{pmatrix} \sigma_{nee,o}^2 & \rho\sigma_{nee,o}^2 \\ \rho\sigma_{nee,o}^2 & \sigma_{nee,o}^2 \end{pmatrix}, \quad (54)$$

with  $0 \leq \rho < 1$ .

We have not shown the analytic representation for the SIC here as it is too large. We instead use the symbolic Python package SymPy [Joyner et al., 2012] to plot the SIC for an increasing value of  $\rho$  in figure 9. Figure 9 shows that as the size of time correlation  $\rho$  approaches 1 the information content in the two observations of NEE decreases. This decrease in information content makes sense as including the correlation in time is decreasing the amount of independent information we are assimilating. This result is also seen in Järvinen et al. [1999] where including a serial correlations between observation errors is shown to reduce the weight given to the mean of the observations in the assimilation (equivalent to inflating the variance of the observations). This also supports the results found in section (ref. 1st results chapter)

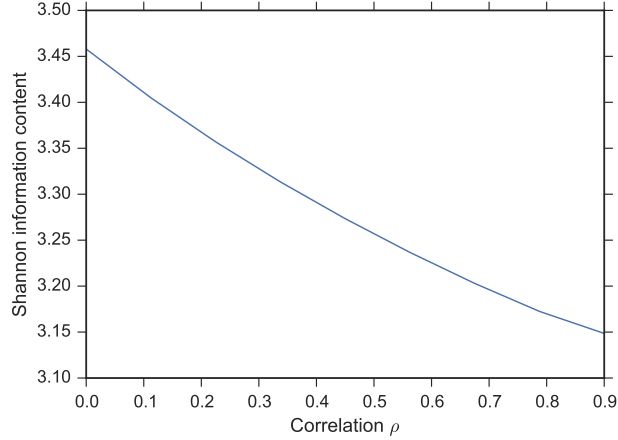


Figure 9: Shannon information content for two successive observations of NEE when a varying time correlation is included between observation errors.

where we see that including correlations in time between observation errors reduces the fit to the assimilated observations in the analysis window.

### 3.4 DALEC2 information content

#### 3.4.1 Information content in observations for DALEC2

In this section we repeat and extend some of the results we have found for information content with the DALEC1 state estimation case in section 3.3 to the DALEC2 joint parameter and state estimation case. This means we now have an augmented state of 23 elements (17 parameters and 6 state variables) as opposed to just the 5 state members for DALEC1. For this reason we no longer examine the analytic representations of information content but instead consider the information content calculated numerically for DALEC2.

In section 3.3.1 it was shown that for DALEC1 the information content for a single observation of NEE was dependent on temperature. From Figure 10 we can see that this is still the case for DALEC2. However the value of SIC is higher for DALEC2 in Figure 10 than for DALEC1 in Figure 6 as the augmented state for the DALEC2 case also includes the parameters. This means that a single observation of NEE is giving us information about more elements of the state than for the DALEC1 state estimation case.

In figure 10 we have shown the information content varying for an evergreen forest site. As DALEC2 can also be parameterised and run for deciduous sites (with much work in this thesis being undertaken at Forest Research’s deciduous study site, see section [REF](#)) it is important to investigate the difference in information content between these cases. In order to visualise this difference, in figure 11 we show the analysis sensitivity to observations or influence matrix [Cardinali et al., 2004] as described in section 2.3.1,  $\mathbf{S} = \mathbf{K}^T \mathbf{H}^T$ , for a year’s assimilation window with 365 observations of NEE. The influence matrix will depend on the initial augmented state we chose to linearise around, the driving data we use to run our model and the observations we specify for assimilation. In figure 11 we use an initial augmented state optimised for the Alice Holt deciduous forest and an initial augmented state optimised for an evergreen site in Oregon, we then use the same yearly driving data for both states so that it is only the difference between the initial augmented states of the sites effecting the difference between the influence matrices.

From figure 11 we can see that the influence of the assimilated observations of NEE is noticeably different between the deciduous and evergreen sites. However, in both cases at the beginning of the window there is a group of observations with similar influence. This makes sense as we are predicting the initial

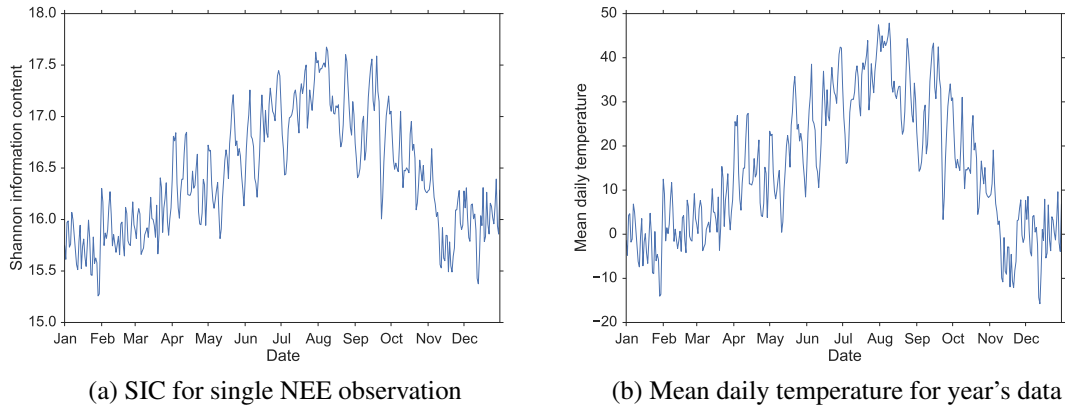


Figure 10: SIC a single NEE observation changing throughout a year's window using driving data from a pine stand in Oregon taken in 2007 (left). Mean daily temperature for the same site and period (right).

augmented state for DALEC2, so that observations closer to this initial state should have greater influence.

For the deciduous site in figure 11a we have groups of observations with high influence from around day 125 to day 175 and from day 250 to day 300. We also have some high influence observations between these two groups. High influence observations between these two groups would be consistent with the results showing that NEE observations have higher information content with higher temperatures, as the period between day 175 and 250 contains days with higher mean temperatures. For the evergreen site in figure 11b, although we have a group of observations at the beginning of the growing season with higher influence, we do not see a group of with the same high influence between day 250 to day 300 as with the deciduous case. We still see observations of high influence corresponding to times of higher temperatures for the evergreen case.

In order to further investigate these groups of high influence observations we show the phenology functions controlling the rate of leaf-on and leaf-off for the DALEC2 model in figure 12. The description of phenology is the main difference between the more simplistic, evergreen only, DALEC1 and DALEC2 which can be parameterised for both deciduous and evergreen sites. It is logical that this is what is causing the difference in information content between the models and between the different sites. In figure 12 we see that the function controlling leaf-off for the deciduous site has a far larger peak than that of the evergreen site. This is expected as the deciduous site will drop all of its leaves at the end of the season. In both cases the forest puts most effort into putting on new leaves at the start of the growing season. This highlights the fact that the NEE for a deciduous site is highly controlled by phenology, as the forest cannot photosynthesise without leaves. Therefore the observations of NEE that help to constrain the phenology of the site should have a higher influence, as seen in figure 11a. Conversely for an evergreen site NEE is driven less by phenology and more by the climatic driving data. Seeing a greater relationship between temperature and information content for an evergreen site consequently makes sense and this can be seen in figure 11b.

### 3.4.2 Effect of time correlations on observation information content

In section 3.3.3 it was shown that, for the analytic DALEC1 case, when assimilating two successive observations of NEE the SIC decreased when including a correlation in time between NEE observation errors. It was noted that this was consistent with results found in section (ref. 1st results chapter) where including correlations between observation errors in time reduced the weight of the observations in the assimilation, in turn reducing the issue of overfitting to the assimilated observations. In figure 13 we repeat the experiment in section 3.3.3 but for DALEC2 with the year's worth of NEE observations assimilated in section (ref. 1st

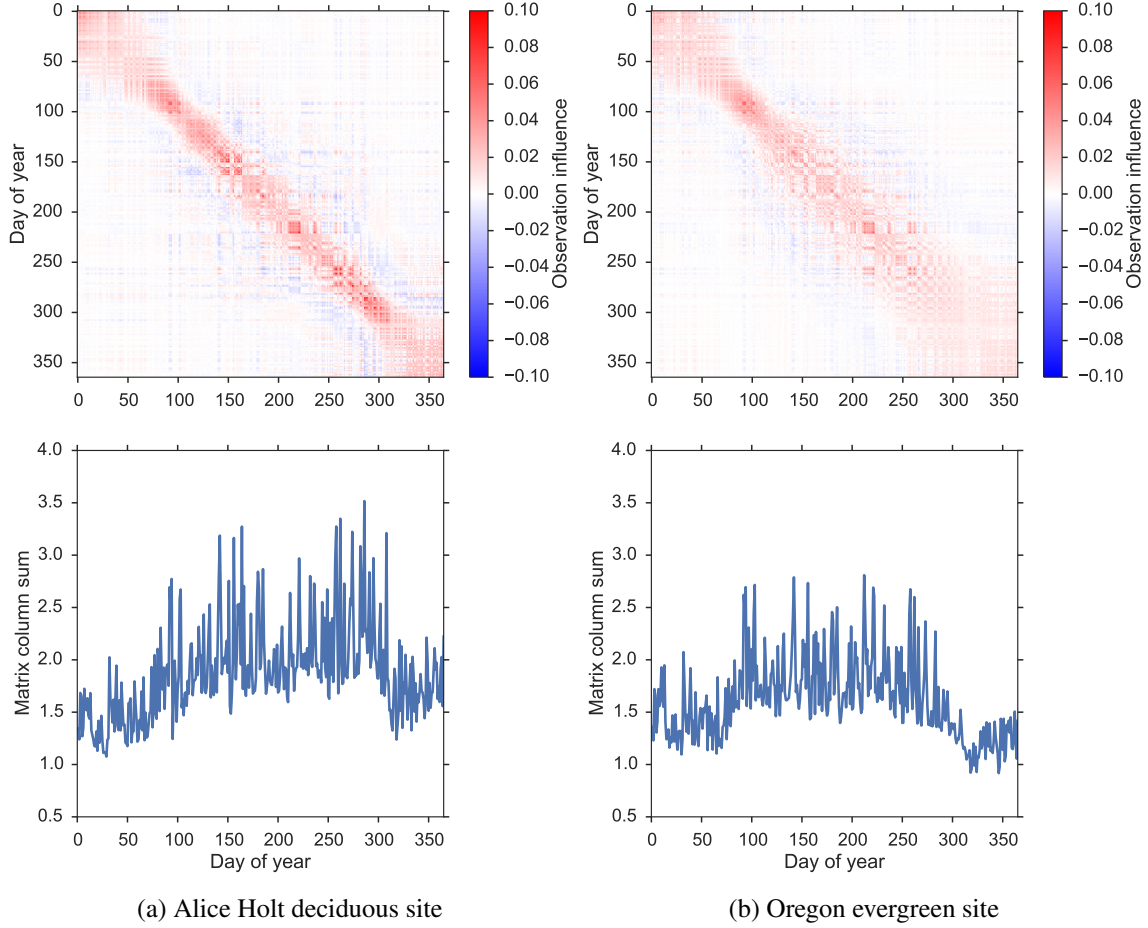


Figure 11: Influence matrices and column absolute value sums as described in section 2.3.1, showing the sensitivity of the modelled observations to the assimilated observations for a year’s assimilation window starting at the beginning of January with 365 observations of NEE.

results chapter), in order to verify that including a correlation in time reduces the information content in assimilated observations. From figure 13 we see that we have similar results as in figure 9 where the information content in our observations decreases as we increase the time correlation between the assimilated observation errors. However, in figure 13 we have a higher value of SIC as we are assimilating many more observations than in figure 9. In figure 13 we have used the same correlation function as in section (ref. 1st results chapter) to create a correlated matrix  $\hat{\mathbf{R}}$  and then varied the magnitude of the included correlation,  $\rho$ . The decreasing information content with an increasing correlation between observation errors in time supports the results in section 3.3.3 and section (ref. 1st results chapter). This is also consistent with the results of Järvinen et al. [1999] where including correlations between observation errors in time is shown to reduce the weight given to the mean of the observations in the assimilation (equivalent to inflating the variance of the observations).

## 4 Conclusions

In this chapter we have investigated both the observability and information content given the observations available to us. In section 3.1 and section 3.2 we have shown that for both DALEC1 and DALEC2 we

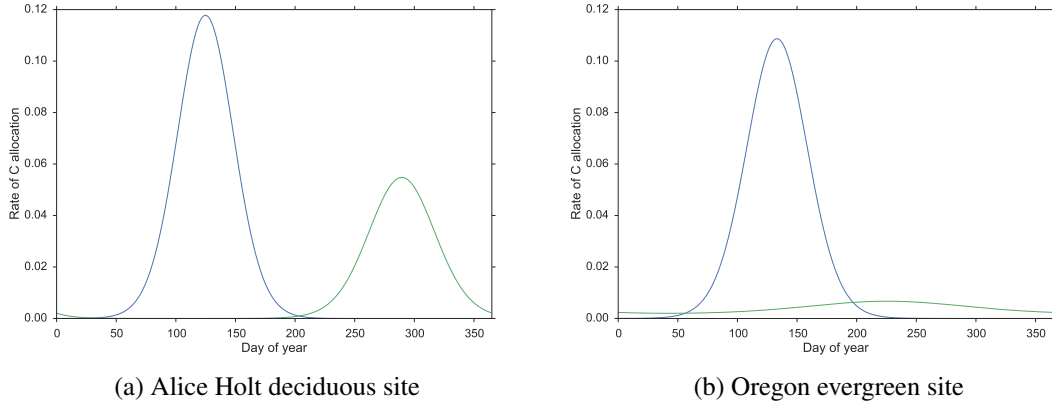


Figure 12: Phenology of DALEC2 model for a deciduous and evergreen forest. Blue line: function controlling rate of leaf-on ( $\phi_{on}$ ), green line: function controlling rate of leaf-off ( $\phi_{off}$ ).

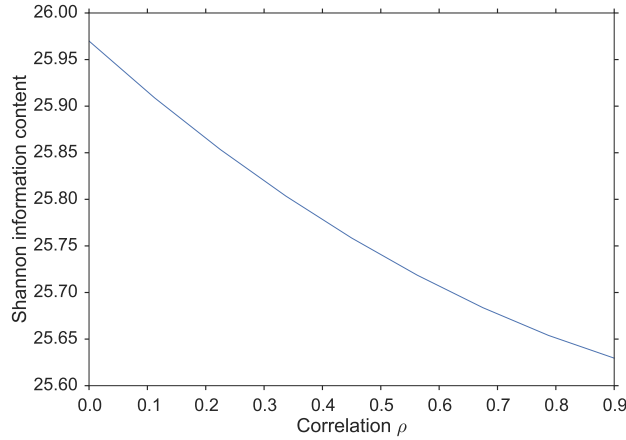


Figure 13: Shannon information content for 67 observations of NEE taken throughout a year's assimilation window when a varying time correlation is included between observation errors.

have an observable system with the available observations, in this case NEE. An observable system in this case means that for data assimilation we can construct a locally unique solution from the observational information alone.

In section 3.3 we have seen that for the DALEC1 evergreen case the information content in observations of NEE is largely dependent on temperature, with higher temperatures meaning higher information content. This is important for informing planned maintenance or down time at flux tower sites measuring NEE. This dependance of information content on temperature is also seen for observations of ground respiration and total ecosystem respiration. When assimilating additional observations at the same time alongside NEE we have found that most information is added when the additional observation provides an orthogonal constraint to that of NEE. This is the case for root carbon ( $C_{roo}$ ) and woody biomass carbon ( $C_{woo}$ ), with the European Space Agency BIOMASS mission being launched soon this should add valuable information to current data assimilation schemes. When using DALEC1 and assimilating successive observations in time it was shown that as observations are added further away from the initial state their impact is decreased. For two successive observation of NEE it was also shown that including a correlation in time between observation errors decreases the information content in the assimilated observations. This is consistent with results

found in (ref. 1st results chapter) where including correlations in time between observation errors reduces overfitting to the assimilated observations.

In section 3.4 we again see the temperature dependence of information content in observations of NEE for DALEC2. However, for DALEC2 we also have varying information content based on the type of ecosystem we are observing. For a deciduous forest site we see that as well as temperature the information content in observations of NEE is strongly dependent on the time of growing season. Observations made at the time of leaf-on and leaf-off have higher influence on the results of the assimilation. This makes physical sense as for a deciduous ecosystem NEE is highly controlled by phenology, as the forest cannot photosynthesise without leaves. Therefore the observations of NEE that help to constrain the phenology of the site should have a higher influence. For an evergreen forest site we see much less dependence on phenology and have a greater relationship between temperature and information content. We again see similar results as in section 3.3 when including correlations in time between observation errors, with an increasing correlation,  $\rho$ , reducing the information content in the assimilated observations.

## References

- D. Baldocchi. Turner review no. 15. ‘breathing’ of the terrestrial biosphere: lessons learned from a global network of carbon dioxide flux measurement systems. *Australian Journal of Botany*, 56(1):1–26, 2008.
- S. Barnett and R. Cameron. *Introduction to Mathematical Control Theory*. Oxford Applied Mathematics and Computing Science Series. Clarendon Press, 1985. ISBN 9780198596400.
- S. Brown. Measuring carbon in forests: current status and future challenges. *Environmental pollution*, 116(3):363–372, 2002.
- C. Cardinali, S. Pezzulli, and E. Andersson. Influence-matrix diagnostic of a data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 130(603):2767–2786, 2004. ISSN 1477-870X. doi: 10.1256/qj.03.205. URL <http://dx.doi.org/10.1256/qj.03.205>.
- T. M. Cover and J. A. Thomas. Elements of information theory. new york: J. Wiley and Sons, 5:5, 1991.
- F. Dahdouh-Guebas and N. Koedam. Empirical estimate of the reliability of the use of the point-centred quarter method (pcqm): Solutions to ambiguous field situations and description of the pcqm+ protocol. *Forest Ecology and management*, 228(1):1–18, 2006.
- G. Desroziers, L. Berre, V. Chabot, and B. Chapnik. A posteriori diagnostics in an ensemble of perturbed analyses. *Monthly Weather Review*, 137(10):3420–3436, 2009.
- R. Engelen and G. Stephens. Information content of infrared satellite sounding measurements with respect to  $\text{CO}_2$ . *Journal of Applied Meteorology*, 43(2):373–378, 2004.
- J. Eyre. The information content of data from satellite sounding systems: A simulation study. *Quarterly Journal of the Royal Meteorological Society*, 116(492):401–434, 1990.
- M. Fisher. *Estimation of entropy reduction and degrees of freedom for signal for large variational analysis systems*. European Centre for Medium-Range Weather Forecasts, 2003.
- A. Fowler and P. Jan Van Leeuwen. Measures of observation impact in non-Gaussian data assimilation. *Tellus A*, 64(0), apr 2012. ISSN 1600-0870. doi: 10.3402/tellusa.v64i0.17192. URL <http://www.tellusa.net/index.php/tellusa/article/view/17192>.

- A. Fowler and P. J. Van Leeuwen. Observation impact in data assimilation: The effect of non-gaussian observation error. *Tellus, Series A: Dynamic Meteorology and Oceanography*, 65(1):1–16, 2013. ISSN 02806495. doi: 10.3402/tellusa.v65i0.20035.
- A. Fox, M. Williams, A. D. Richardson, D. Cameron, J. H. Gove, T. Quaife, D. Ricciuto, M. Reichstein, E. Tomelleri, C. M. Trudinger, et al. The reflex project: comparing different algorithms and implementations for the inversion of a terrestrial ecosystem model against eddy covariance data. *Agricultural and Forest Meteorology*, 149(10):1597–1615, 2009.
- B. Husch, T. W. Beers, and J. A. Kershaw Jr. *Forest mensuration*. John Wiley & Sons, 2002.
- J. A. Jacquez and P. Greif. Numerical parameter identifiability and estimability: Integrating identifiability, estimability, and optimal sampling design. *Mathematical Biosciences*, 77(1-2):201–227, 1985. ISSN 00255564. doi: 10.1016/0025-5564(85)90098-7.
- H. Järvinen, E. Andersson, and F. Bouttier. Variational assimilation of time sequences of surface observations with serially correlated errors. *Tellus A*, 51(4):469–488, 1999.
- C. Johnson, B. J. Hoskins, and N. K. Nichols. A singular vector perspective of 4d-var: Filtering and interpolation. *Quarterly Journal of the Royal Meteorological Society*, 131(605):1–19, 2005.
- D. Joyner, O. Čertík, A. Meurer, and B. E. Granger. Open source computer algebra systems: SymPy. *ACM Commun. Comput. Algebra*, 45(3/4):225–234, Jan. 2012. ISSN 1932-2240. doi: 10.1145/2110170.2110185.
- E. Kalnay. *Atmospheric modeling, data assimilation, and predictability*. Cambridge university press, 2003.
- B. Law, P. Thornton, J. Irvine, P. Anthoni, and S. Van Tuyl. Carbon storage and fluxes in ponderosa pine forests at different developmental stages. *Global Change Biology*, 7(7):755–777, 2001a.
- B. Law, S. Van Tuyl, A. Cescatti, and D. Baldocchi. Estimation of leaf area index in open-canopy ponderosa pine forests at different successional stages and management regimes in oregon. *Agricultural and Forest Meteorology*, 108(1):1–14, 2001b.
- T. Le Toan, S. Quegan, M. Davidson, H. Balzter, P. Paillou, K. Papathanassiou, S. Plummer, F. Rocca, S. Saatchi, H. Shugart, et al. The biomass mission: Mapping global forest biomass to better understand the terrestrial carbon cycle. *Remote sensing of environment*, 115(11):2850–2860, 2011.
- M. A. Lefsky, D. Harding, W. Cohen, G. Parker, and H. Shugart. Surface lidar remote sensing of basal area and biomass in deciduous forests of eastern maryland, {USA}. *Remote Sensing of Environment*, 67(1):83 – 98, 1999. ISSN 0034-4257. doi: [http://dx.doi.org/10.1016/S0034-4257\(98\)00071-6](http://dx.doi.org/10.1016/S0034-4257(98)00071-6). URL <http://www.sciencedirect.com/science/article/pii/S0034425798000716>.
- L. Ljung. *System Identification: Theory for the User*. Pearson Education, 1998. ISBN 9780132440530.
- I. Navon. Practical and theoretical aspects of adjoint parameter estimation and identifiability in meteorology and oceanography. *Dynamics of Atmospheres and Oceans*, 27(1):55–79, 1998.
- S. Niu, Y. Luo, M. C. Dietze, T. F. Keenan, Z. Shi, J. Li, and F. S. C. Iii. The role of data assimilation in predictive ecology. *Ecosphere*, 5(5):art65, 2014. ISSN 2150-8925. doi: 10.1890/ES13-00273.1. URL <http://www.esajournals.org/doi/abs/10.1890/ES13-00273.1>.
- C. Paige. Properties of numerical algorithms related to computing controllability. *IEEE Transactions on Automatic Control*, 26(1):130–138, 1981. ISSN 0018-9286. doi: 10.1109/TAC.1981.1102563.

- T. Palmer, R. Gelaro, J. Barkmeijer, and R. Buizza. Singular vectors, metrics, and adaptive observations. *Journal of the Atmospheric Sciences*, 55(4):633–653, 1998.
- W. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes 3rd Edition: The Art of Scientific Computing*. Cambridge University Press, 2007. ISBN 9780521880688.
- T. Quaife, P. Lewis, M. De Kauwe, M. Williams, B. E. Law, M. Disney, and P. Bowyer. Assimilating canopy reflectance data into an ecosystem model with an Ensemble Kalman Filter. *Remote Sensing of Environment*, 112(4):1347–1364, 2008. ISSN 00344257. doi: 10.1016/j.rse.2007.05.020.
- F. Rabier, N. Fourrié, D. Chafäi, and P. Prunet. Channel selection methods for infrared atmospheric sounding interferometer radiances. *Quarterly Journal of the Royal Meteorological Society*, 128(581):1011–1027, 2002.
- J. Renaud. Automatic differentiation in robust optimization. *AIAA journal*, 35(6):1072–1079, 1997.
- A. D. Richardson, M. D. Mahecha, E. Falge, J. Kattge, A. M. Moffat, D. Papale, M. Reichstein, V. J. Stauch, B. H. Braswell, G. Churkina, B. Kruijt, and D. Y. Hollinger. Statistical properties of random  $\{CO_2\}$  flux measurement uncertainty inferred from model residuals. *Agricultural and Forest Meteorology*, 148(1):38–50, 2008. ISSN 0168-1923. doi: <http://dx.doi.org/10.1016/j.agrformet.2007.09.001>.
- A. D. Richardson, M. Williams, D. Y. Hollinger, D. J. Moore, D. B. Dail, E. A. Davidson, N. A. Scott, R. S. Evans, H. Hughes, J. T. Lee, et al. Estimating parameters of a forest ecosystem model with measurements of stocks and fluxes as joint constraints. *Oecologia*, 164(1):25–40, 2010.
- C. D. Rodgers. Information content and optimisation of high spectral resolution remote measurements. *Advances in Space Research*, 21(3):361–367, 1998.
- C. D. Rodgers et al. *Inverse methods for atmospheric sounding: Theory and practice*, volume 2. World scientific Singapore, 2000.
- K. Singh, A. Sandu, M. Jardak, K. Bowman, and M. Lee. A practical method to estimate information content in the context of 4d-var data assimilation. *SIAM/ASA Journal on Uncertainty Quantification*, 1(1):106–138, 2013.
- L. M. Stewart, S. Dance, and N. Nichols. Correlated observation errors in data assimilation. *International journal for numerical methods in fluids*, 56(8):1521–1527, 2008.
- G. Wahba. Design criteria and eigensequence plots for satellite-computed tomography. *Journal of Atmospheric and Oceanic Technology*, 2(2):125–132, 1985.
- M. Williams, P. A. Schwarz, B. E. Law, J. Irvine, and M. R. Kurpius. An improved analysis of forest carbon dynamics using data assimilation. *Global Change Biology*, 11(1):89–105, 2005.
- J. Zobitz, A. Desai, D. Moore, and M. Chadwick. A primer for data assimilation with ecological models using markov chain monte carlo (mcmc). *Oecologia*, 167(3):599–611, 2011.
- J. M. Zobitz, D. J. P. Moore, T. Quaife, B. H. Braswell, A. Bergeson, J. a. Anthony, and R. K. Monson. Joint data assimilation of satellite reflectance and net ecosystem exchange data constrains ecosystem carbon fluxes at a high-elevation subalpine forest. *Agricultural and Forest Meteorology*, 195-196:73–88, 2014. ISSN 01681923. doi: 10.1016/j.agrformet.2014.04.011. URL <http://dx.doi.org/10.1016/j.agrformet.2014.04.011>.
- X. Zou, I. Navon, and F. Le Dimet. Incomplete observations and control of gravity waves in variational data assimilation. *Tellus A*, 44(4):273–296, 1992.