Chapter 2: Observability and Information Content DRAFT

Ewan Pinnington

June 12, 2016

1 Introduction (to be included in literature review chapter)

1.1 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,

$$\dot{\mathbf{x}} = \mathbf{A}(t)\mathbf{x}(t) + \mathbf{B}(t)\mathbf{u}(t) \tag{1}$$
$$\mathbf{y} = \mathbf{C}(t)\mathbf{x}(t) \tag{2}$$

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.

Theorem 1.1. *When $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$ are time-invariant the system is completely observable if and only if the $nr \times n$ observability matrix*

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

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

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

as defined in section REF. 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 a background term is typically included in the cost function for 4D-Var data assimilation which regularises the problem and means that we always have a unique solution.

2 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 analysis.

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. Defining target observations [Palmer et al., 1998]. 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,

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

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

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

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

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

using the defined matrices in section REF, 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].

2.1 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,

$$\mathbf{S} = \frac{\partial \mathbf{H}\mathbf{x}_a}{\partial \mathbf{y}}. \tag{8}$$

From equation (5) we see that,

$$\mathbf{S} = \mathbf{K}^T\mathbf{H}^T, \tag{9}$$

here $\mathbf{S}$ will be a $p \times p$ matrix, where p is the number of observations. The diagonal elements of $\mathbf{S}$ are $\mathbf{S}_{i,i} = \frac{\partial(\mathbf{H}\mathbf{x}_a)_i}{\partial \mathbf{y}_i}$ 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,

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

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

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

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

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

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

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

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].

2.2 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 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 λ_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,

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

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 2.1) with,

$$dfs = \mathrm{trace}(\mathbf{S}) = \sum_i \lambda_i, \tag{15}$$

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

2.3 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 as,

$$SIC = \int p(\mathbf{x})\ln p(\mathbf{x})d\mathbf{x} - \int p(\mathbf{y}) \int p(\mathbf{x}|\mathbf{y})\ln p(\mathbf{x}|\mathbf{y})d\mathbf{x}d\mathbf{y}. \tag{16}$$

From Rodgers et al. [2000], for the Gaussian case *SIC* becomes a function of the prior and posterior error covariance matrices with,

$$SIC = \frac{1}{2}\ln\frac{|\mathbf{B}|}{|\mathbf{A}|}. \tag{17}$$

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

$$SIC = -\frac{1}{2}\sum_i ln|1-\lambda_i| \tag{18}$$

where λ_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 Chapter overview

Draft

In this chapter we aim to analyse the information content in the observations used for assimilation with the DALEC and DALEC2 model. We begin by considering the observability of our system to see if we have enough information from the observations available to us in order to construct a unique solution to our data assimilation problem. 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 both DALEC and DALEC2. Using these measure also allows us to consider the effect of including error correlations in our data assimilation algorithm (see previous section) on the information content in the observations.

3.2 DALEC1 observability

DALEC1 is the original version of the DALEC2 model introduced in section REF. 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 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 by hand. 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,

$$\mathbf{M}_i = \frac{\partial \mathbf{m}_{i-1\to i}(\mathbf{x}_i)}{\partial \mathbf{x}_i} \tag{19}$$
$$= \begin{pmatrix} (1-\theta_{fol}) + f_{fol}(1-f_{auto})\zeta^i & 0 & 0 & 0 & 0 \\ f_{roo}(1-f_{fol})(1-f_{auto})\zeta^i & (1-\theta_{roo}) & 0 & 0 & 0 \\ (1-f_{roo})(1-f_{fol})(1-f_{auto})\zeta^i & 0 & (1-\theta_{woo}) & 0 & 0 \\ \theta_{fol} & \theta_{roo} & 0 & (1-(\theta_{min}+\theta_{lit})\chi^{i-1}) & 0 \\ 0 & 0 & \theta_{lit} & \theta_{min}\chi^{i-1} & (1-\theta_{som}\chi^{i-1}) \end{pmatrix},$$

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

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. 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

$$LAI^i = \frac{C^i_{fol}}{c_{lma}}. \tag{20}$$

We then have the linearised observation operator

$$\mathbf{H}_i = \frac{\partial LAI^i}{\partial \mathbf{x}_i} = \begin{pmatrix} \frac{1}{c_{lma}} & 0 & 0 & 0 & 0 \end{pmatrix}. \tag{21}$$

Using the linearised observation operator and the linear model from equation 19 we can compute $\hat{\mathbf{H}}$ for 5 observations of LAI with

$$\hat{\mathbf{H}} = \begin{pmatrix} \mathbf{H}_0 \\ \mathbf{H}_1\mathbf{M}_0 \\ \vdots \\ \mathbf{H}_4\mathbf{M}_{3,0} \end{pmatrix}$$
$$= \begin{pmatrix} \frac{1}{c_{lma}} & 0 & 0 & 0 & 0 \\ \frac{1}{c_{lma}}((1-\theta_{fol}) + f_{fol}(1-f_{auto})\zeta^0) & 0 & 0 & 0 & 0 \\ \frac{1}{c_{lma}}\prod_{i=0}^{1}((1-\theta_{fol}) + f_{fol}(1-f_{auto})\zeta^i) & 0 & 0 & 0 & 0 \\ \frac{1}{c_{lma}}\prod_{i=0}^{2}((1-\theta_{fol}) + f_{fol}(1-f_{auto})\zeta^i) & 0 & 0 & 0 & 0 \\ \frac{1}{c_{lma}}\prod_{i=0}^{3}((1-\theta_{fol}) + f_{fol}(1-f_{auto})\zeta^i) & 0 & 0 & 0 & 0 \end{pmatrix}, \tag{22}$$

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 C_{som}, so that by taking observations of

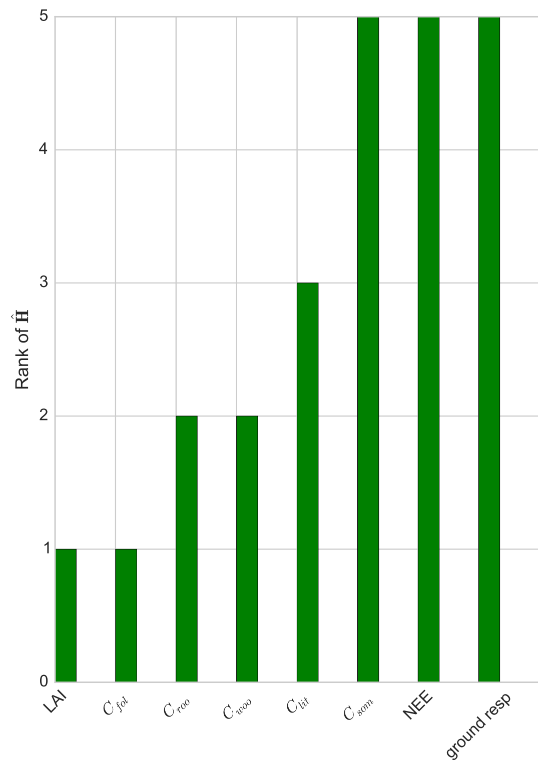


Figure 1: Rank of the observability matrix $\hat{\mathbf{H}}$ for 5 observations of different types. The ranks shown here are computed analytically using SymPy [Joyner et al., 2012].

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

$$NEE^i = -(1 - f_{auto})GPP^i(C_{fol}^{i-1}, \Psi) + \theta_{lit}C_{lit}e^{\Theta T^i} + \theta_{som}C_{som}e^{\Theta T^i} \tag{23}$$

with a corresponding linearised observation operator

$$\mathbf{H}_i = \frac{\partial NEE^i}{\partial \mathbf{x}_i} = \begin{pmatrix} -(1 - f_{auto})\zeta^i & 0 & 0 & \theta_{lit}e^{\Theta T^i} & \theta_{som}e^{\Theta T^i} \end{pmatrix}, \tag{24}$$

and for ground respiration

$$G_{resp}^i = \frac{1}{3}f_{auto}GPP^i(C_{fol}^{i-1}, \Psi) + \theta_{lit}C_{lit}e^{\Theta T^i} + \theta_{som}C_{som}e^{\Theta T^i} \tag{25}$$

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

$$\mathbf{H}_i = \frac{\partial G_{resp}^i}{\partial \mathbf{x}_i} = \begin{pmatrix} \frac{1}{3}f_{auto}\zeta^i & 0 & 0 & \theta_{lit}e^{\Theta T^i} & \theta_{som}e^{\Theta T^i} \end{pmatrix}. \tag{26}$$

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 3.4.

For the state and parameter estimation case we will not be able to 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. In table 1 we show that for both numeric and analytic implementations we have the same results for the rank of $\hat{\mathbf{H}}$.

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.

3.3 DALEC1 information content

For the DALEC1 state estimation we can calculate the analytic representation of the information content measures discussed in section 2. 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

$$\boldsymbol{B} = \begin{pmatrix} \sigma^2_{cfol,b} & 0 & 0 & 0 & 0 \\ 0 & \sigma^2_{croo,b} & 0 & 0 & 0 \\ 0 & 0 & \sigma^2_{cwoo,b} & 0 & 0 \\ 0 & 0 & 0 & \sigma^2_{clit,b} & 0 \\ 0 & 0 & 0 & 0 & \sigma^2_{csom,b} \end{pmatrix}. \tag{27}$$

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

$$\mathbf{H}_i = \frac{\partial LAI^i}{\partial \mathbf{x}_i} = \begin{pmatrix} \frac{1}{c_{lma}} & 0 & 0 & 0 & 0 \end{pmatrix}. \tag{28}$$

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 ($\sigma^2_{LAI,o}$). Therefore,

$$\mathbf{R}_0 = \sigma^2_{LAI,o}. \tag{29}$$

We then have from equation **??**,

$$\begin{aligned}
\mathbf{A} &= (\mathbf{J}'')^{-1} \\
&= (\mathbf{B}^{-1} + \hat{\mathbf{H}}^T \hat{\mathbf{R}}^{-1} \hat{\mathbf{H}})^{-1} \\
&= (\mathbf{B}^{-1} + \mathbf{H}_0^T \mathbf{R}_0^{-1} \mathbf{H}_0)^{-1} \\
&= \begin{pmatrix} \frac{c_{lma}^2 \sigma_{LAI,o}^2 \sigma_{cfol,b}^2}{\sigma_{cfol,b}^2 + c_{lma}^2 \sigma_{LAI,o}^2} & 0 & 0 & 0 & 0 \\ 0 & \sigma_{croo,b}^2 & 0 & 0 & 0 \\ 0 & 0 & \sigma_{cwoo,b}^2 & 0 & 0 \\ 0 & 0 & 0 & \sigma_{clit,b}^2 & 0 \\ 0 & 0 & 0 & 0 & \sigma_{csom,b}^2 \end{pmatrix}.
\end{aligned} \tag{30}$$

We can now derive the SIC and DFS using equation **??** and **??** as,

$$SIC = \frac{1}{2}\ln\frac{|\mathbf{B}|}{|\mathbf{A}|} = \frac{1}{2}\ln\frac{(c_{lma}^2 \sigma_{LAI,o}^2 + \sigma_{cfol,b}^2)}{c_{lma}^2 \sigma_{LAI,o}^2} = \frac{1}{2}\ln\left(1 + \frac{\sigma_{cfol,b}^2}{c_{lma}^2 \sigma_{LAI,o}^2}\right) \tag{31}$$

and

$$DFS = n - tr(\mathbf{B}^{-1}\mathbf{A}) = 5 - \left(\frac{c_{lma}^2 \sigma_{LAI,o}^2}{(c_{lma}^2 \sigma_{LAI,o}^2 + \sigma_{cfol,b}^2)} + 4\right) = 1 - \frac{c_{lma}^2 \sigma_{LAI,o}^2}{(c_{lma}^2 \sigma_{LAI,o}^2 + \sigma_{cfol,b}^2)}. \tag{32}$$

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

$$SIC = \frac{1}{2}\ln\left(1 + \frac{\sigma_{c,b}^2}{\sigma_{c,o}^2}\right) \tag{33}$$

and

$$DFS = 1 - \frac{\sigma_{c,o}^2}{(\sigma_{c,o}^2 + \sigma_{c,b}^2)}, \tag{34}$$

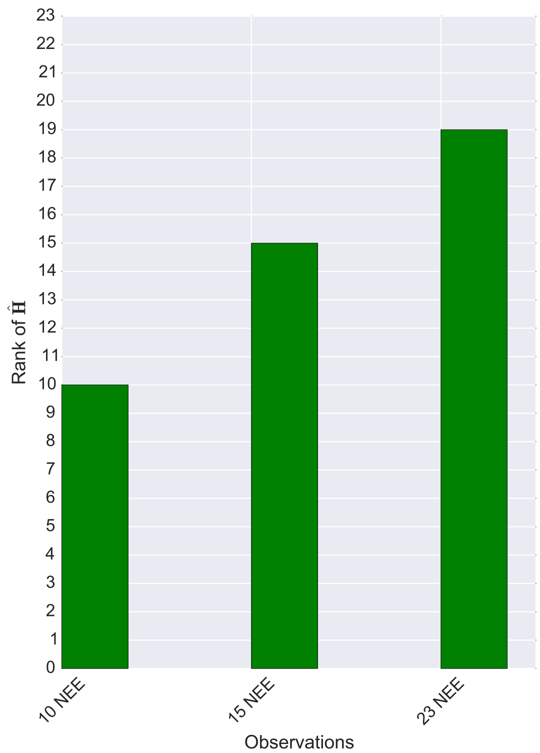
where $\sigma_{c,o}$ and $\sigma_{c,b}$ are the observation and background standard deviations respectively, corresponding to any of the 5 carbon pools. We see the SIC for an observation of a single observation of one of the carbon pools is dependent on the ratio between the observation and background 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 (31) and (32) 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

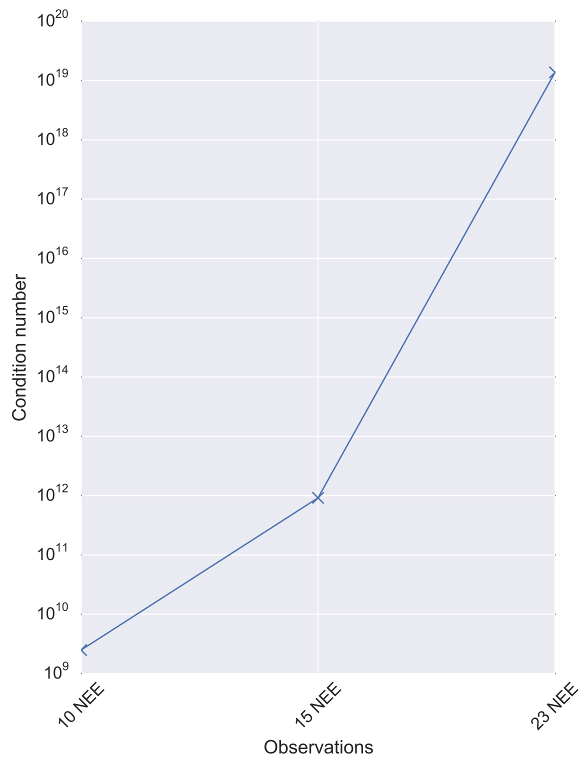
3.4 DALEC2 observability

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. 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 3.2 by finding the rank of $\hat{\mathbf{H}}$ for a given set of observations. As discussed in section 3.2 we cannot compute the observability analytically for the DALEC2 joint parameter and state estimation scheme. 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. 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.



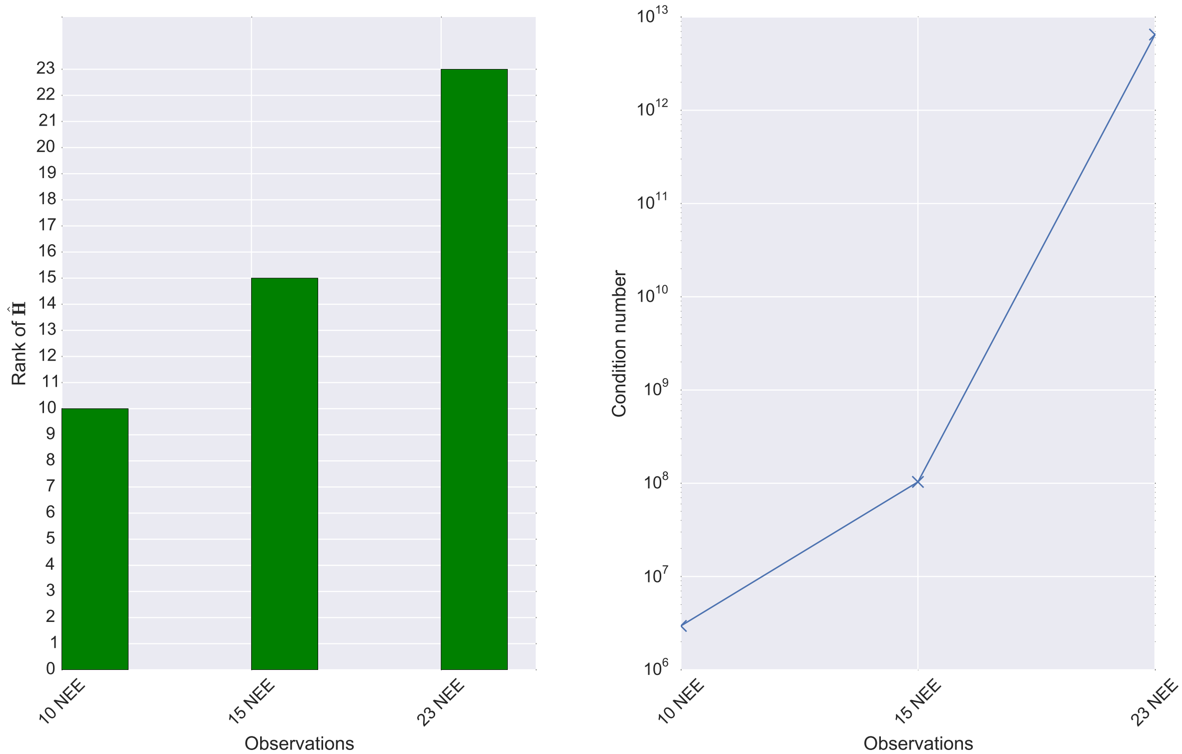
(a) Rank of $\hat{\mathbf{H}}$



(b) Condition number of $\hat{\mathbf{H}}$

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.

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 `tol = max(S) * max(n, m) * 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 `eps` is the machine accuracy for the datatype of `S` (In this case a double-precision float with `eps = 2.22e-16`). For 23 observations of NEE $\hat{\mathbf{H}}$ is classed as being rank deficient as `tol = 1.02e-10` and the three smallest singular values of $\hat{\mathbf{H}}$ are `[1.39e-11, 7.84e-15, 1.46e-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.



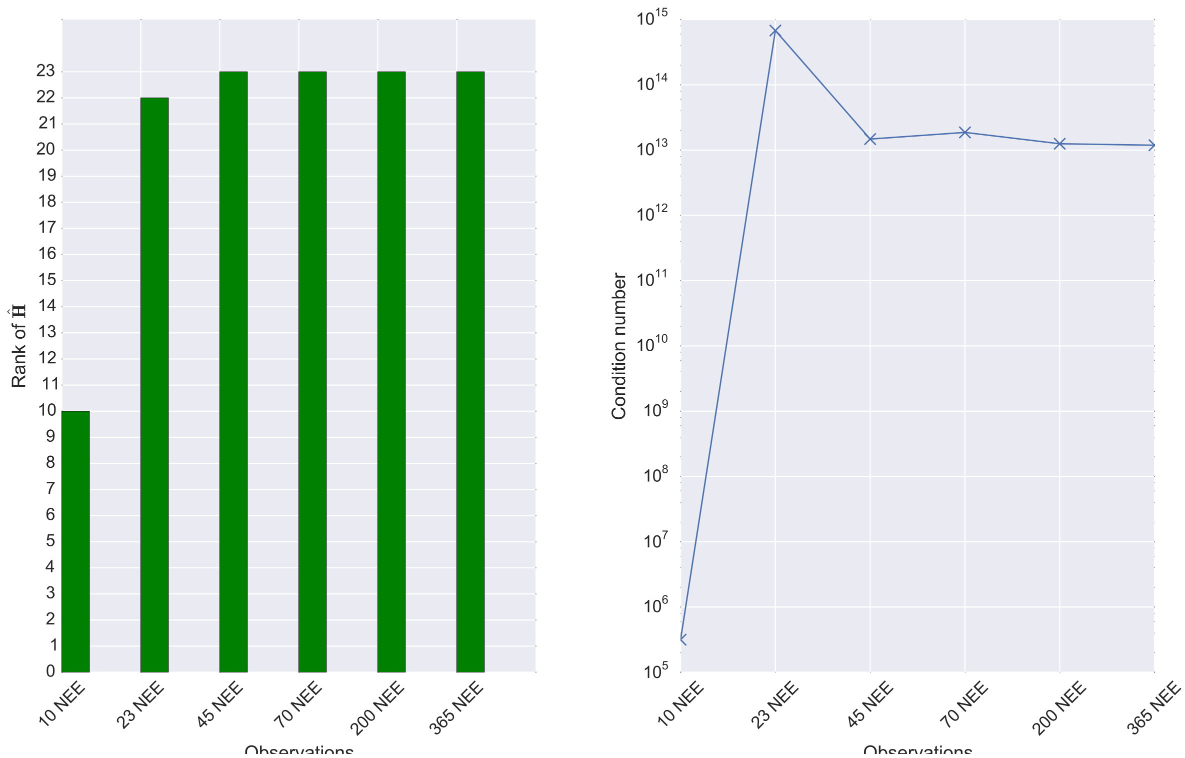
(a) Rank of $\hat{\mathbf{R}}^{-1/2}\hat{\mathbf{H}}\mathbf{D}^{1/2}$ (b) Condition number of $\hat{\mathbf{R}}^{-1/2}\hat{\mathbf{H}}\mathbf{D}^{1/2}$

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 section where CVT is described, $\mathbf{D} = diag\{\mathbf{B}\}$. 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. 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.

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.

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 $\hat{\mathbf{H}}$.



(a) Rank of $\hat{\mathbf{H}}$ (b) Condition number of $\hat{\mathbf{H}}$

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).

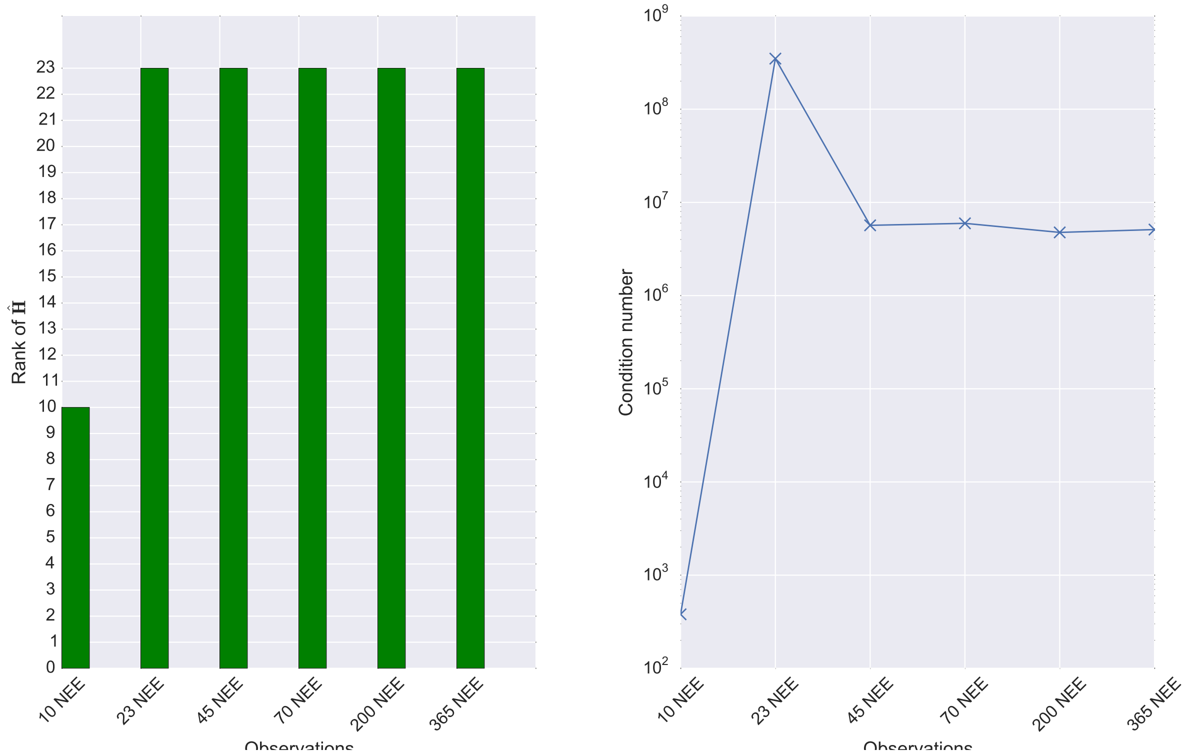
In figure 5 we again see that using the CVT observability matrix has much improved the conditioning of the problem in comparison to figure 4. We now have that the DALEC2 system is observable when we have 23 observations of NEE randomly distributed throughout the 1 year assimilation window. We have more 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 years window, we can construct a unique solution with the observations alone.

References

S. Barnett and R. Cameron. *Introduction to Mathematical Control Theory*. Oxford Applied Mathematics and Computing Science Series. Clarendon Press, 1985. ISBN 9780198596400.

C. Cardinali, S. Pezzulli, and E. Andersson. Influence-matrix diagnostic of a data assimilation system.



(a) Rank of $\hat{\mathbf{R}}^{-1/2}\hat{\mathbf{H}}\mathbf{D}^{1/2}$

(b) Condition number of $\hat{\mathbf{R}}^{-1/2}\hat{\mathbf{H}}\mathbf{D}^{1/2}$

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).

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.

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 co2. *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. 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.

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.

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.

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.

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.

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.

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.

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.