# Information content for observations of forest carbon stocks and fluxes when assimilated with the DALEC carbon balance model

E. Pinnington

October 23, 2014

## 1 Introduction

A large amount of data is currently being gathered that is relevant to the carbon balance of forests, with much of this data coming from Eddy covariance flux towers [1, 2, 10, 13]. Attempts are also being made to combine this data with models of forest carbon stocks and fluxes, such as the Data Assimilation Linked Ecosystem Carbon model (DALEC), in a data assimilation scheme [7, 14]. Currently, however, there are limitations with such schemes as there is a lack of understanding about the additional information provided by different observations. Better understanding of the information content of carbon balance observations will help inform measurement campaigns of when and which observations to take in order to gain the most possible information about the system. In this report we will look at different information content measures which have been used in meteorological data assimilation [5, 9, 11] and apply these to carbon balance observations assimilated with DALEC. Although we use DALEC in this report the results will be similar for other carbon balance models which use similar driving data and equations. We begin by introducing the DALEC model which will initially be used to look at the information content in different observations.

## 2 The DALEC Model

The DALEC model is a simple process-based model describing the carbon balance of an evergreen forest ecosystem [14]. The model is constructed of five carbon pools (foliage ($C_f$), fine roots ($C_r$), woody stems and coarse roots ($C_w$), fresh leaf and fine root litter ($C_l$) and soil organic matter and coarse woody debris ($C_s$)) linked via fluxes. The gross primary production function ($GPP$) uses meteorological driving data and the site's leaf area index (a function of $C_f$) to calculate the total amount of carbon to be allocated at a daily time step.

The model equations for the carbon pools at day $t+1$ are as follows:

$$C_f(t+1) = (1-p_5)C_f(t) + p_3(1-p_2)GPP(C_f(t),\phi), \tag{1}$$
$$C_r(t+1) = (1-p_7)C_r(t) + p_4(1-p_3)(1-p_2)GPP(C_f(t),\phi), \tag{2}$$
$$C_w(t+1) = (1-p_6)C_w(t) + (1-p_4)(1-p_3)(1-p_2)GPP(C_f(t),\phi), \tag{3}$$
$$C_l(t+1) = (1-(p_1+p_8)T(t))C_l(t) + p_5C_f(t) + p_7C_r(t), \tag{4}$$
$$C_s(t+1) = (1-p_9T(t))C_s + p_6C_w(t) + p_1T(t)C_l(t), \tag{5}$$

where $T(t) = \frac{1}{2}exp(p_{10}T_m(t))$, $T_m$ is daily mean temperature, $p_1,\ldots,p_{10}$ are rate parameters and $\phi$ represents the meteorological driving data used in the $GPP$ function. The full details

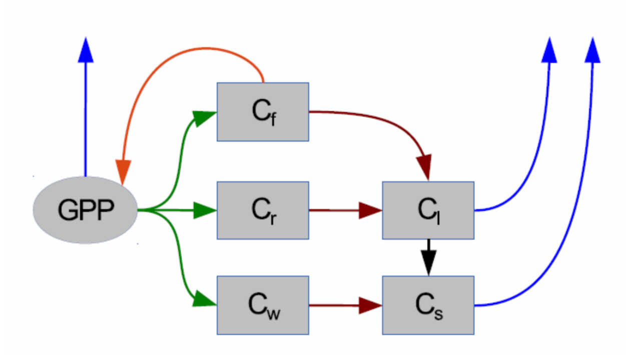



Figure 1: Representation of the carbon fluxes in the DALEC carbon balance model. Green arrows represent C allocation, dark red and black arrows represent litterfall and decomposition fluxes, blue arrows represent respiration fluxes and the light red arrow represents the feedback of foliar carbon to the $GPP$ function. [3]

of this version of DALEC can be found in [14], it is paramterized for data from a young pine stand in Ponderossa, Oregon, we include values of the model parameters and the equations used to calculate $GPP$ in the appendix. We now see how DALEC can be implemented in a four-dimensional variational data assimilation (4D-Var) framework.

### 2.1 DALEC in a variational assimilation framework

In 4D-Var we aim to maximise the probabiliy of our initial state $\mathbf{x}_0$ (for DALEC our state $\mathbf{x}_0$ corresponds to the initial values of the five carbon pools, $\mathbf{x}_0 = (C_f(t_0), C_r(t_0), C_w(t_0), C_l(t_0), C_s(t_0))^T$) given a set of observations $\mathbf{y}$, $P(\mathbf{x}_0|\mathbf{y})$, over some time window, $N$. We do this by minimising a cost function $J(\mathbf{x})$ derived from Baye's Theorem [8]. The cost function is given as,

$$J(\mathbf{x}_0) = \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}_b)^T\mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}_b) + \frac{1}{2}\sum_{i=0}^{N}(\mathbf{y}_i - h_i(\mathbf{x}_i))^T\mathbf{R}_i^{-1}(\mathbf{y}_i - h_i(\mathbf{x}_i)), \tag{6}$$

where $\mathbf{x}_b$ is our background and acts as our initial guess to our state $\mathbf{x}_0$, $\mathbf{B}$ is the background error covariance matrix and quantifies our knowledge of the error in our background, $h_i$ is our observation operator at time $t_i$ and maps our state vector evolved by our nonlinear model ($\mathbf{x}_i$) to the observations at this time $\mathbf{y}_i$ and $\mathbf{R}_i$ is the observation error covariance matrix and represents our knowledge of the uncertainty in the observations. The state that minimises the cost function is called the analysis and is denoted as $\mathbf{x}_a$, this state is found using a minimisation routine that takes the cost function, our initial guess ($\mathbf{x}_b$) and also the gradient of the cost function defined as,

$$\nabla J(\mathbf{x}_0) = \mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}_b) - \sum_{i=0}^{N}\mathbf{M}_{i,0}^T\mathbf{H}_i^T\mathbf{R}_i^{-1}(\mathbf{y}_i - h_i(\mathbf{x}_i)), \tag{7}$$

where $\mathbf{H}_i = \frac{\delta h_i(\mathbf{x}_i)}{\delta \mathbf{x}_i}$ is our linearized observation operator, $\mathbf{M}_{i,0} = \mathbf{M}_{i-1}\mathbf{M}_{i-2}\cdots\mathbf{M}_0$ is our tangent linear model with $\mathbf{M}_i = \frac{\delta \mathbf{m}_i(\mathbf{x}_i)}{\delta \mathbf{x}_i}$. We can calculate the linearized model for DALEC from equations

1 to 5 as,

$$\mathbf{M}_i = \begin{pmatrix} (1-p_5)+p_3(1-p_2)\zeta_i & 0 & 0 & 0 & 0 \\ p_4(1-p_3)(1-p_2)\zeta_i & (1-p_7) & 0 & 0 & 0 \\ (1-p_4)(1-p_3)(1-p_2)\zeta_i & 0 & (1-p_6) & 0 & 0 \\ p_5 & p_7 & 0 & (1-(p_1+p_8)T_i) & 0 \\ 0 & 0 & p_6 & p_1 T_i & (1-p_9 T_i) \end{pmatrix}, \tag{8}$$

where $\zeta_i = GPP'(C_f(t_i), \phi)$ and $T_i = T(t_i)$.

Once we have performed the minimisation of the cost function and determined $\mathbf{x}_a$ we can also calculate the analysis error covariance matrix, $\mathbf{A}$, to quantify the uncertainty in our new estimate of the state. We can define the analysis error covariance matrix as,

$$\mathbf{A} = \left(\mathbf{J}''\right)^{-1} = \left(\mathbf{B}^{-1} + \hat{\mathbf{H}}^T \hat{\mathbf{R}}^{-1} \hat{\mathbf{H}}\right)^{-1}, \tag{9}$$

where $\hat{\mathbf{H}}$ is the matrix of linearized observation operators evolved by the tangent linear model and $\hat{\mathbf{R}}$ is the block diagonal matrix of observation error covariance matrices,

$$\hat{\mathbf{H}} = \begin{pmatrix} \mathbf{H}_0 \\ \mathbf{H}_1\mathbf{M}_0 \\ \vdots \\ \mathbf{H}_N\mathbf{M}_{N,0} \end{pmatrix} \quad \text{and} \quad \hat{\mathbf{R}} = \begin{pmatrix} \mathbf{R}_0 & 0 & 0 & 0 \\ 0 & \mathbf{R}_1 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \mathbf{R}_N \end{pmatrix}. \tag{10}$$

## 3 Information Content Measures

Information content measures are already being used to quantify the different levels of information provided by observations in the development of satellite instruments [4, 12] and in operational data assimilation schemes [5, 11]. In these fields two of the more widely used measures are Shannon Information Content (also known as entropy reduction) and the degrees of freedom for signal. We will apply both methods for observations assimilated with DALEC.

### 3.1 Shannon Information Content

In DA Shannon Information Content ($SIC$) is a measure of the reduction in entropy given a set of observations. Entropy physically corresponds to the volume in state space taken up by the probability density function (*pdf*) describing the knowledge of the state. When a measurement is made the volume of this *pdf* decreases, the information content of the measurement is a measure of the factor by which it decreases [9]. If $P_b(x)$ is our knowledge of the state before an observation and $P_o(x|y)$ is our knowledge after an observation then we have entropies,

$$S[P_b(x)] = -\int P_b(x)\log_2[P_b(x)]dx \quad \text{and} \quad S[P_o(x|y)] = -\int P_o(x|y)\log_2[P_o(x|y)]dx.$$

The entropy reduction, or $SIC$, due to the observation is then,

$$SIC = S[P_b(x)] - S[P_o(x|y)]. \tag{11}$$

If we assume all *pdfs* are Gaussian and use the natural logarithm as opposed to $\log_2$ (for algebraic convenience) [9] the entropy of a multivariate Gaussian distruibution for a vector $\mathbf{x}$ with $n$ elements (before and after observations) can be derived as,

$$S[P_b(\mathbf{x})] = n\ln(2\pi e)^{\frac{1}{2}} + \frac{1}{2}\ln|\mathbf{B}| \tag{12}$$

and

$$S[P_o(\mathbf{x}|\mathbf{y})] = n\ln(2\pi e)^{\frac{1}{2}} + \frac{1}{2}\ln\left|\mathbf{A}\right| \tag{13}$$

where $\mathbf{B}$ is the background error covariance matrix and $\mathbf{A}$ is the analysis error covariance matrix. Combining equations 11, 12 and 13 we can write the $SIC$ as,

$$SIC = \frac{1}{2}\ln\frac{\left|\mathbf{B}\right|}{\left|\mathbf{A}\right|}. \tag{14}$$

From equation 9 we can see the $SIC$ can also be written as,

$$SIC = \frac{1}{2}\ln\left|\mathbf{B}\right|\left|\mathbf{J}''\right|. \tag{15}$$

### 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 [6].

We have symmetric positive definite background and analysis error covariance matrices $\mathbf{B}$ and $\mathbf{A}$, the eigenvalues of each matrix gives a representation for the uncertainty in the direction of the associated eigenvector, thus by comparing the eigenvalues of both matrices we can determine the reduction in uncertainty given a set of observations [12].

In order to do this we take $\mathbf{B}^{\frac{-1}{2}}$ such that $\mathbf{B}^{-1} = \mathbf{B}^{\frac{-1}{2}}\mathbf{B}^{\frac{-1}{2}}$. We now take $\mathbf{Q}$ to be the orthogonal matrix composed of the eigenvectors of $\mathbf{B}^{\frac{-1}{2}}\mathbf{A}\mathbf{B}^{\frac{-1}{2}}$ we have,

$$\mathbf{Q}^T\left(\mathbf{B}^{\frac{-1}{2}}\mathbf{A}\mathbf{B}^{\frac{-1}{2}}\right)\mathbf{Q} = \mathbf{\Lambda}, \tag{16}$$

$$\mathbf{Q}^T\left(\mathbf{B}^{\frac{-1}{2}}\mathbf{B}\mathbf{B}^{\frac{-1}{2}}\right)\mathbf{Q} = \mathbf{I}_{nxn} \tag{17}$$

where $\mathbf{\Lambda}$ is a diagonal matrix. Each diagonal element of our transformed $\mathbf{B}$ is equal to one and corresponds to one degree of freedom. The diagonal elements of $\mathbf{\Lambda}$ correspond to the matrixs eigenvalues and can be interpreted as the relative reduction in variance for each of the $n$ degrees of freedom [11]. We can then define the $DFS$ as,

$$\begin{aligned} DFS &= \mathrm{trace}(\mathbf{I}_{nxn} - \mathbf{\Lambda}) \\ &= n - \mathrm{trace}(\mathbf{\Lambda}) \\ &= n - \mathrm{trace}(\mathbf{B}^{\frac{-1}{2}}\mathbf{A}\mathbf{B}^{\frac{-1}{2}}) \\ &= n - \mathrm{trace}(\mathbf{B}^{-1}\mathbf{A}). \end{aligned} \tag{18}$$

## 4 Shannon Information Content for DALEC

We begin by using $SIC$ to understand the information content for different observations when being assimilated with the DALEC model. For these experiments the model is set up as in section 2.1 where the elements of the state vector have standard deviations, $\sigma_{cf,b}, \dots, \sigma_{cs,b}$, respectively. These standard deviations represent the uncertainty in our inital background estimate and are taken as a percentage of the initial carbon pools (these values can be found in the appendix). The background error covariance matrix is then taken as the diagonal matrix of the variances of the carbon pools,

$$\mathbf{B} = \begin{pmatrix} \sigma^2_{cf,b} & 0 & 0 & 0 & 0 \\ 0 & \sigma^2_{cr,b} & 0 & 0 & 0 \\ 0 & 0 & \sigma^2_{cw,b} & 0 & 0 \\ 0 & 0 & 0 & \sigma^2_{cl,b} & 0 \\ 0 & 0 & 0 & 0 & \sigma^2_{cs,b} \end{pmatrix}. \tag{19}$$

Our inital experiements look at the $SIC$ in observations taken at single time, the 4D-Var data assimilation then becomes three-dimensional variational assimilation (3D-Var) as we are not summing over a time window and have no time component.

### 4.1 $SIC$ for a single observation at one time

If we first consider one observation of $C_f$ (the first element of our state vector $\mathbf{x}$) at time $t_0$, we can derive an analytical expression for the $SIC$ using,

$$\mathbf{H}_0 = \frac{\delta C_f(t_0)}{\delta \mathbf{x}} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \end{pmatrix}, \tag{20}$$

where $\mathbf{H}_0$ is our linearized observation operator at time $t_0$. As we have a single observation at one time our observation error covariance matrix, $\mathbf{R_0}$, is just the variance of our observation of $C_f$ at time $t_0$ ($\sigma^2_{cf,o}$). Therefore,

$$\mathbf{R}_0 = \sigma^2_{cf,o}. \tag{21}$$

We then have from equation 9,

$$\begin{aligned} \mathbf{J}'' &= \mathbf{B}^{-1} + \hat{\mathbf{H}}^T \hat{\mathbf{R}}^{-1} \hat{\mathbf{H}} \\ &= \mathbf{B}^{-1} + \mathbf{H}_0^T \mathbf{R}_0^{-1} \mathbf{H}_0 \\ &= \begin{pmatrix} \sigma^{-2}_{cf,b} + \sigma^{-2}_{cf,o} & 0 & 0 & 0 & 0 \\ 0 & \sigma^{-2}_{cr,b} & 0 & 0 & 0 \\ 0 & 0 & \sigma^{-2}_{cw,b} & 0 & 0 \\ 0 & 0 & 0 & \sigma^{-2}_{cl,b} & 0 \\ 0 & 0 & 0 & 0 & \sigma^{-2}_{cs,b} \end{pmatrix}. \end{aligned} \tag{22}$$

We can now derive the $SIC$ using equation 15 as,

$$SIC = \frac{1}{2} ln \left| \mathbf{B} \right| \left| \mathbf{J}'' \right| = \frac{1}{2} ln \frac{(\sigma^2_{cf,o} + \sigma^2_{cf,b})}{\sigma^2_{cf,o}} = \frac{1}{2} ln \left( 1 + \frac{\sigma^2_{cf,b}}{\sigma^2_{cf,o}} \right). \tag{23}$$

We see the SIC for an observation of a single observation of $C_f$ is dependant on the ratio between the observation and background variances, the $SIC$ will have the same form for all other direct observations of carbon pools contained in the state. As our background standard deviations are set to being the same percentage for each carbon pool (SEE APPENDIX), the carbon pool observation which will give us the highest $SIC$ is the pool that we can measure most acurately as this will maximise the ratio $\frac{\sigma_{c,b}^2}{\sigma_{c,o}^2}$ by minimising $\sigma_{c,o}^2$.

One of the main carbon balance observations made at forest flux tower sites is the net ecosystem exchange ($NEE$) of $CO_2$, which can be estimated by DALEC as the difference between $GPP$ and the respiration of $C_l$ and $C_s$, giving,

$$NEE(t) = -(1-p_2)GPP(C_f(t), \phi) + p_8 C_l T(t) + p_9 C_s T(t). \tag{24}$$

For a single observation of $NEE$ at one time, $t_0$, we can again derive an analytical expression for the $SIC$ using,

$$\mathbf{H}_0 = \frac{\delta NEE(t_0)}{\delta \mathbf{x}} = \begin{pmatrix} -(1-p_2)\zeta_0 & 0 & 0 & p_8 T_0 & p_9 T_0 \end{pmatrix}, \tag{25}$$

where $\zeta_0 = GPP'(C_f(t_0), \phi)$, $T_0 = T(t_0)$ and $\mathbf{H}_0$ is the linearized observation operator at time $t_0$. Again our observation error covariance matrix, $\mathbf{R_0}$, is just the variance of our observation of $NEE$, $\sigma_{nee,o}^2$, at time $t_0$. Therefore,

$$\mathbf{R}_0 = \sigma_{nee,0}^2 \tag{26}$$

and again,

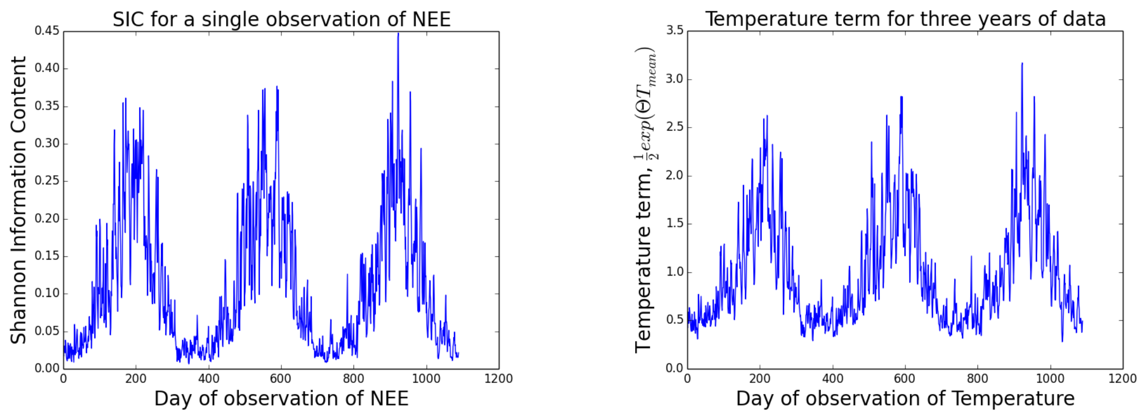
$$\begin{aligned}\mathbf{J}'' &= \mathbf{B}^{-1} + \mathbf{H}_0^T \mathbf{R}_0^{-1} \mathbf{H}_0 \\ &= \begin{pmatrix} \sigma_{cf,b}^{-2} + \sigma_{nee,0}^{-2}(1-p_2)^2\zeta_0^2 & 0 & 0 & \sigma_{nee,0}^{-2}(1-p_2)\zeta_0 p_8 T_0 & \sigma_{nee,0}^{-2}(1-p_2)\zeta_0 p_9 T_0 \\ 0 & \sigma_{cr,b}^{-2} & 0 & 0 & 0 \\ 0 & 0 & \sigma_{cw,b}^{-2} & 0 & 0 \\ \sigma_{nee,0}^{-2}(1-p_2)\zeta_0 p_8 T_0 & 0 & 0 & \sigma_{cl,b}^{-2} + \sigma_{nee,0}^{-2} p_8^2 T_0^2 & \sigma_{nee,0}^{-2} p_8 p_9 T_0^2 \\ \sigma_{nee,0}^{-2}(1-p_2)\zeta_0 p_9 T_0 & 0 & 0 & \sigma_{nee,0}^{-2} p_8 p_9 T_0^2 & \sigma_{cs,b}^{-2} + \sigma_{nee,0}^{-2} p_9^2 T_0^2 \end{pmatrix}.\end{aligned} \tag{27}$$

We then have,

$$SIC = \frac{1}{2} ln \left|\mathbf{B}\right| \left|\mathbf{J}''\right| = \frac{1}{2} ln \frac{(p_2-1)^2 \zeta_0^2 \sigma_{cf,b}^2 + \sigma_{nee,0}^2 + T_0^2(p_9^2 \sigma_{cs,b}^2 + p_8^2 \sigma_{cl,b}^2)}{\sigma_{nee,0}^2}. \tag{28}$$

If we assume that the variances and parameters here are fixed we can see that the size of the $SIC$ is dependent on the temperature term, $T_0$, and the square of the first derivative of $GPP$, $\zeta_0^2$. Generally, the value of $GPP$ (and its first derivative) is highest in summer with higher total daily irradiance and higher temperatures. We therefore have that there will be more information content in observations that are taken when temperatures are higher. Physically this makes sense as more $NEE$ takes place when temperatures are higher (to a point) so measurements are of greater magnitude and give us more information about carbon fluxes. By plotting the SIC for a single observation of $NEE$ varying with three years of meteorological driving data and the temperature term ($T(t_i)$) for the same period of the same data we can see that both are closely linked in figure 2.

However the relationship is not linear as the magnitude of $GPP$'s first derivative is also dependent on daily irradiance and the value of the foliar carbon pool ($C_f$). This shows that observations of $NEE$ made in the summer are much more valuable than those made in the winter assuming warmer temperatures, higher daily irradiance and a higher amount of foliar carbon in the summer. In the next section we consider a series of observations over some time window. We will see that it takes 14 observations of $NEE$ in winter to gain the same level of information as 1 observation of $NEE$ in the winter.




(a) $SIC$ for a single observation of $NEE$. (b) Temperature term, $\frac{1}{2}exp(\Theta T_{mean})$.

Figure 2: $SIC$ and temperature varying over three years using driving data from Oregon pine forest.

## 4.2 $SIC$ for successive observations over a time window

Following the results for $SIC$ based at a single time, we now consider the $SIC$ when successive observations are added over a period of time. The DALEC model is now built into a Four-Dimensional Variational Data Assimilation (4D-Var) framework where our observation operator, $\mathbf{H}$, and observation error covariance matrix, $\mathbf{R}$, are now,

$$\mathbf{H} = \begin{pmatrix} \mathbf{H}_0 \\ \mathbf{H}_1\mathbf{M}_0 \\ \vdots \\ \mathbf{H}_n\mathbf{M}_{n,0} \end{pmatrix} \quad \text{and} \quad \mathbf{R} = \begin{pmatrix} \mathbf{R}_0 & 0 & 0 & 0 \\ 0 & \mathbf{R}_1 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \mathbf{R}_n \end{pmatrix},$$

where $\mathbf{H}_i$ is our linearized observation operator at time $t_i$, $\mathbf{M}_{i,0} = \mathbf{M}_{i-1}\mathbf{M}_{i-2}\cdots\mathbf{M}_0$ is our linearized model evolving the state vector, $\mathbf{x}_b$, at time $t_0$ to time $t_i$ and $\mathbf{R}_i$ is the observation error covariance matrix corresponding to $\mathbf{H}_i$ at time $t_i$ [8]. Firstly the tangent linear model for DALEC was calculated analytically as $\mathbf{M}_i = \frac{\delta \mathbf{m}_i}{\delta \mathbf{x}_i}$.

We begin by considering successive observations of $Cf$ in time. Here we again have,

$$\mathbf{H}_i = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \end{pmatrix} \quad \text{and} \quad \mathbf{R}_i = \sigma_{cf,o}^2.$$

The linearized model at time $t_i$ is given as,

$$\mathbf{M}_i = \begin{pmatrix} (1-p_5)+p_3(1-p_2)GPP'(C_f(t_i),\phi) & 0 & 0 & 0 & 0 \\ p_4(1-p_3)(1-p_2)GPP'(C_f(t_i),\phi) & (1-p_7) & 0 & 0 & 0 \\ (1-p_4)(1-p_3)(1-p_2)GPP'(C_f(t_i),\phi) & 0 & (1-p_6) & 0 & 0 \\ p_5 & p_7 & 0 & (1-(p_1+p_8)T(t_i)) & 0 \\ 0 & 0 & p_6 & p_1T(t_i) & (1-p_9T(t_i)) \end{pmatrix}.$$

Then for two successive observations of $Cf$ we have,

$$\mathbf{H} = \begin{pmatrix} \mathbf{H}_0 \\ \mathbf{H}_1\mathbf{M}_0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ (1-p_5)+p_3(1-p_2)GPP'(C_f(t_0),\phi) & 0 & 0 & 0 & 0 \end{pmatrix}$$

and

$$\mathbf{R} = \begin{pmatrix} \mathbf{R}_0 & 0 \\ 0 & \mathbf{R}_1 \end{pmatrix} = \begin{pmatrix} \sigma_{cf,o}^2 & 0 \\ 0 & \sigma_{cf,o}^2 \end{pmatrix}.$$

We then have,

$$SIC = \frac{1}{2} ln \left|\mathbf{B}\right| \left|\mathbf{J}''\right| = \frac{1}{2} ln \left(1 + \frac{\sigma_{cf,b}^2}{\sigma_{cf,o}^2} + \frac{\sigma_{cf,b}^2 \eta_0^2}{\sigma_{cf,o}^2}\right),$$

where $\eta_i = (1 - p_5) + p_3(1 - p_2)GPP'(C_f(t_i), \phi)$. We can continue adding more observations at successive times and we start to see a pattern. For three observations at successive times we have,

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

for four,

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

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

$$SIC \text{ for } n \text{ observations of } Cf \;= \frac{1}{2} ln \left(1 + \frac{\sigma_{cf,b}^2}{\sigma_{cf,o}^2} \big(1 + \sum_{k=0}^{n-2} \prod_{i=0}^{k} \eta_i^2 \big)\right)$$

We have plotted the $SIC$ for increasing numbers of observations of $Cf$ using three years of meteorological driving data from an Oregan pine forest as seen in figure 3.

As before with a single observation at one time we can repeat this with successive observations of $NEE$ instead of $Cf$ this is plotted in figure 4. Here we can see the seasonal cycle of information content as in figure 2 with little information being added during the winter months and more being added during summer.

# 5 Degrees of freedom for signal for DALEC

# 6 Conclusion

# References


[1] Dennis 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.

[2] Dennis Baldocchi, Eva Falge, Lianhong Gu, Richard Olson, David Hollinger, Steve Running, Peter Anthoni, Ch Bernhofer, Kenneth Davis, Robert Evans, et al. Fluxnet: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities. *Bulletin of the American Meteorological Society*, 82(11):2415–2434, 2001.

[3] Sylvain Delahaies, Ian Roulstone, and Nancy Nichols. A regularization of the carbon cycle data-fusion problem. In *EGU General Assembly Conference Abstracts*, volume 15, page 4087, 2013.

[4] RJ Engelen and GL Stephens. Information content of infrared satellite sounding measurements with respect to co2. *Journal of Applied Meteorology*, 43(2):373–378, 2004.

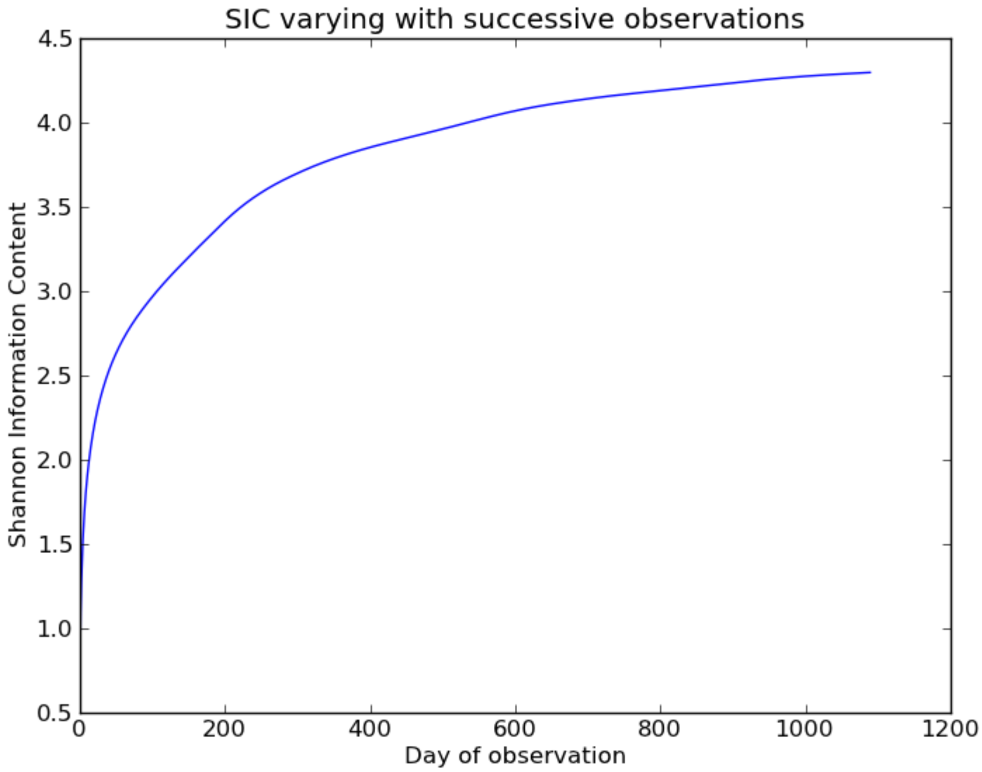



Figure 3: $SIC$ varying as successive observations of $Cf$ are added using driving data from Oregon pine forest.


[5] Michael Fisher. *Estimation of entropy reduction and degrees of freedom for signal for large variational analysis systems*. European Centre for Medium-Range Weather Forecasts, 2003.

[6] Alison Fowler and Peter Jan van Leeuwen. Measures of observation impact in gaussian data assimilation. 2011.

[7] Andrew Fox, Mathew Williams, Andrew D Richardson, David Cameron, Jeffrey H Gove, Tristan Quaife, Daniel Ricciuto, Markus Reichstein, Enrico Tomelleri, Cathy 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.

[8] John M Lewis, Sivaramakrishnan Lakshmivarahan, and Sudarshan Dhall. *Dynamic data assimilation: a least squares approach*, volume 13. Cambridge University Press, 2006.

[9] Clive D Rodgers et al. *Inverse methods for atmospheric sounding: Theory and practice*, volume 2. World scientific Singapore, 2000.

[10] SW Running, DD Baldocchi, DP Turner, ST Gower, PS Bakwin, and KA Hibbard. A global terrestrial monitoring network integrating tower fluxes, flask sampling, ecosystem modeling and eos satellite data. *Remote Sensing of Environment*, 70(1):108–127, 1999.

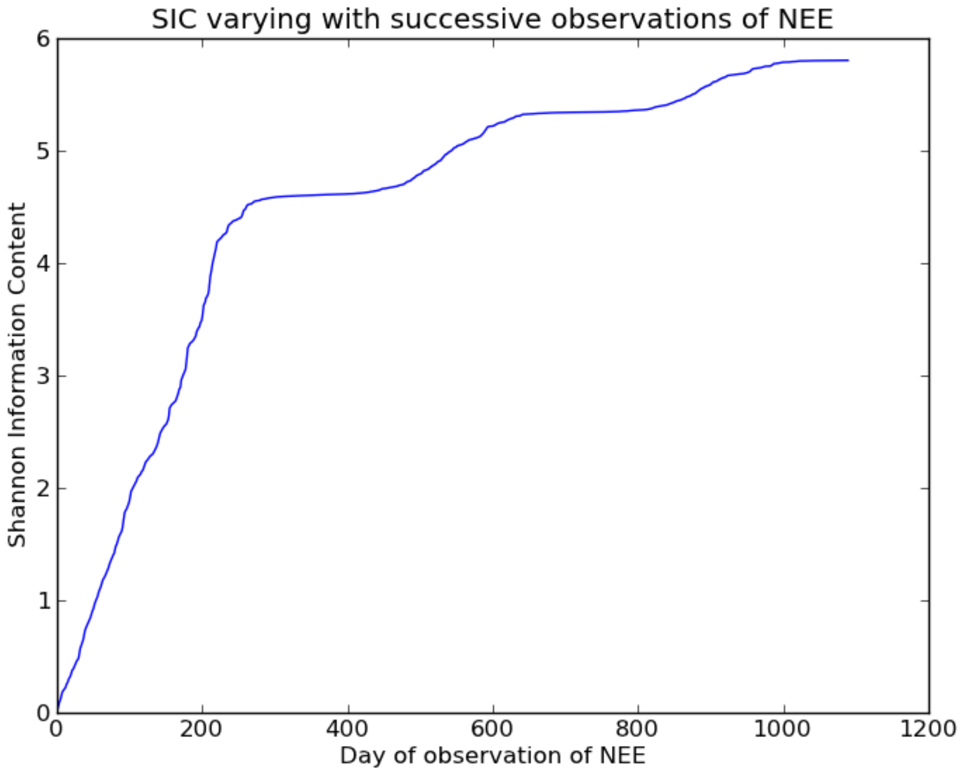



Figure 4: $SIC$ varying as successive observations of $NEE$ are added using driving data from Oregon pine forest.

[11] Adrian Sandu, Kumaresh Singh, Mohamed Jardak, Kevin Bowman, and Meemong Lee. A practical method to estimate information content in the context of 4d-var data assimilation. i: Methodology. 2012.

[12] Laura M Stewart, SL Dance, and NK Nichols. Correlated observation errors in data assimilation. *International journal for numerical methods in fluids*, 56(8):1521–1527, 2008.

[13] R Valentini, G Matteucci, AJ Dolman, E-D Schulze, CJMEAG Rebmann, EJ Moors, A Granier, P Gross, NO Jensen, K Pilegaard, et al. Respiration as the main determinant of carbon balance in european forests. *Nature*, 404(6780):861–865, 2000.

[14] Mathew Williams, Paul A Schwarz, Beverly E Law, James Irvine, and Meredith R Kurpius. An improved analysis of forest carbon dynamics using data assimilation. *Global Change Biology*, 11(1):89–105, 2005.

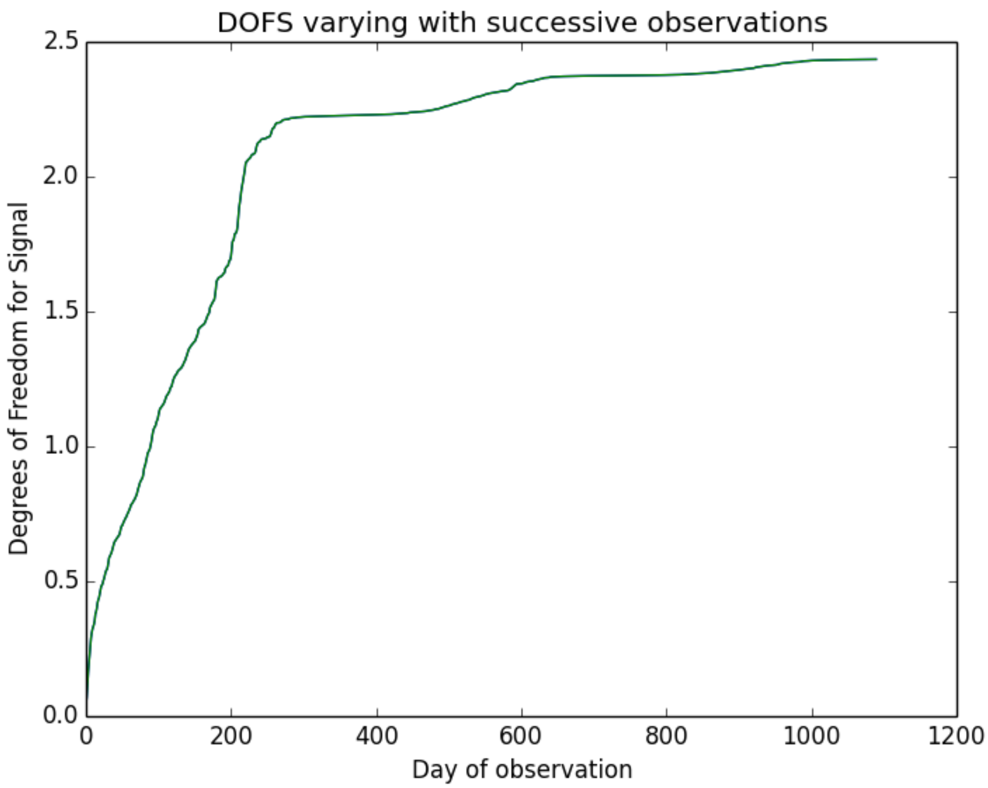



Figure 5: $DOFS$ varying as successive observations of $NEE$ are added using driving data from Oregon pine forest.