

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

E. Pinnington

October 14, 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]. 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) [5], in a data assimilation scheme. 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. 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 [5]. 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:

$$C_f(t + 1) = (1 - p_5)C_f(t) + p_3(1 - p_2)GPP(C_f(t), \phi), \quad (1)$$

$$C_r(t + 1) = (1 - p_7)C_r(t) + p_4(1 - p_3)(1 - p_2)GPP(C_f(t), \phi), \quad (2)$$

$$C_w(t + 1) = (1 - p_6)C_w(t) + (1 - p_4)(1 - p_3)(1 - p_2)GPP(C_f(t), \phi), \quad (3)$$

$$C_l(t + 1) = (1 - (p_1 + p_8)T(t))C_l(t) + p_5C_f(t) + p_7C_r(t), \quad (4)$$

$$C_s(t + 1) = (1 - p_9T(t))C_s + p_6C_w(t) + p_1T(t)C_l(t), \quad (5)$$

where $T(t) = \frac{1}{2}exp(p_{10}T_m(t))$, T_m is daily mean temperature, p_1, \dots, p_{10} are rate parameters and ϕ represents the meteorological driving data used in the GPP function. The full details of this version of DALEC can be found in [5]. We now introduce Shannon Information Content as one method to assess the information content in different carbon balance observations.

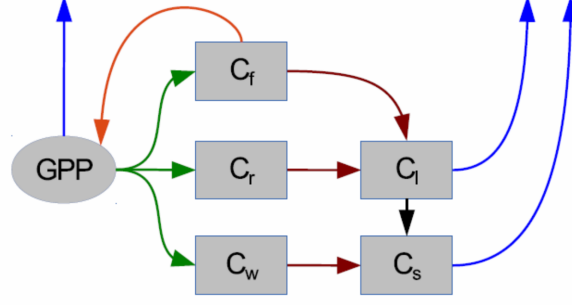


Figure 1: Representation of the carbon fluxes in the DALEC carbon balance model. Green arrows represent C allocation, dark red and black arrows represent litterfall and decomposition fluxes, blue arrows represent respiration fluxes and the light red arrow represents the feedback of foliar carbon to the *GPP* function. [2]

3 Introduction to Variational Assimilation

3.1 Baye's Theorem and 3D-Var

3.2 4D-Var

4 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 [4]. Assuming all pdfs are Gaussian we have,

$$SIC = \frac{1}{2} \ln \frac{|\mathbf{B}|}{|\mathbf{A}|},$$

where \mathbf{B} is the background error covariance matrix and \mathbf{A} is the analysis error covariance matrix. For a larger reduction in uncertainty in our analysis we have a larger value of *SIC*. I began by using *SIC* to understand the information content for different sets of observations at one time when being assimilated with the DALEC model. We specify the state vector for the assimilation as,

$$\underline{x}_b = (C_f, C_r, C_w, C_l, C_s)^T,$$

where the elements of the state vector have variances, $\sigma_{cf,b}^2, \dots, \sigma_{cs,b}^2$, respectively. We then have the following background error covariance matrix,

$$\mathbf{B} = \begin{pmatrix} \sigma_{cf,b}^2 & 0 & 0 & 0 & 0 \\ 0 & \sigma_{cr,b}^2 & 0 & 0 & 0 \\ 0 & 0 & \sigma_{cw,b}^2 & 0 & 0 \\ 0 & 0 & 0 & \sigma_{cl,b}^2 & 0 \\ 0 & 0 & 0 & 0 & \sigma_{cs,b}^2 \end{pmatrix},$$

here we assume a diagonal background error covariance matrix. In order to calculate the *SIC* we need $|\mathbf{A}^{-1}|$ we have,

$$\mathbf{A}^{-1} = \mathbf{J}'' = \mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H},$$

where \mathbf{J}'' is the Hessian of \mathbf{J} , the cost function to be minimized in Three-Dimensional Variational Data Assimilation (3D-Var), and \mathbf{H} is the linearized observation operator.

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 \underline{x}) an analytical expression for the *SIC* can be derived using,

$$\mathbf{H}_0 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \end{pmatrix},$$

where $\mathbf{H}_0 = \frac{\delta C_f(t_0)}{\delta \underline{x}}$ is the linearized observation operator at time t_0 . As we have a single observation at one time our observation error covariance matrix, \mathbf{R} , is just the variance of our observation of C_f , $\sigma_{cf,o}^2$, at time t_0 . Therefore,

$$\mathbf{R} = \sigma_{cf,o}^2$$

and

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

We then have,

$$SIC = \frac{1}{2} \ln \frac{|\mathbf{B}|}{|\mathbf{A}|} = \frac{1}{2} \ln |\mathbf{B}| |\mathbf{J}''|.$$

Hence,

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

Here we see the SIC is dependant on the ratio between the two variances and this will be the same for all other carbon pool observations. Something about better depending on pool observed.

One of the main observations made of the carbon balance of a forest at 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,

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

For a single observation of *NEE* at one time, t_0 , an analytical expression for the *SIC* can be derived using,

$$\mathbf{H}_0 = \begin{pmatrix} -(1 - p_2)\zeta_0 & 0 & 0 & p_8 T_0 & p_9 T_0 \end{pmatrix},$$

where $\zeta_0 = GPP'(C_f(t_0), \phi)$, $T_0 = T(t_0)$ and $\mathbf{H}_0 = \frac{\delta NEE(t_0)}{\delta \underline{x}}$ is the linearized observation operator at time t_0 . Again our observation error covariance matrix, \mathbf{R} , is just the variance of our observation of *NEE*, $\sigma_{nee,0}^2$, at time t_0 . Therefore,

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

and

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

We then have,

$$SIC = \frac{1}{2} \ln \frac{|\mathbf{B}|}{|\mathbf{A}|} = \frac{1}{2} \ln |\mathbf{B}| |\mathbf{J}''|.$$

Hence,

$$SIC = \frac{1}{2} \ln \frac{(p_2 - 1)^2 \zeta_0^2 \sigma_{cf,b}^2 + \sigma_{nee,0}^2 + T_0^2 (p_8^2 \sigma_{cs,b}^2 + p_8^2 \sigma_{cl,b}^2)}{\sigma_{nee,0}^2}.$$

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 , ζ_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 of carbon fluxes***. By plotting the SIC for a single observation of NEE varying with three years of meteorological driving data and the temperature term for the same period of the same data we can see that both are closely linked in figure 2.

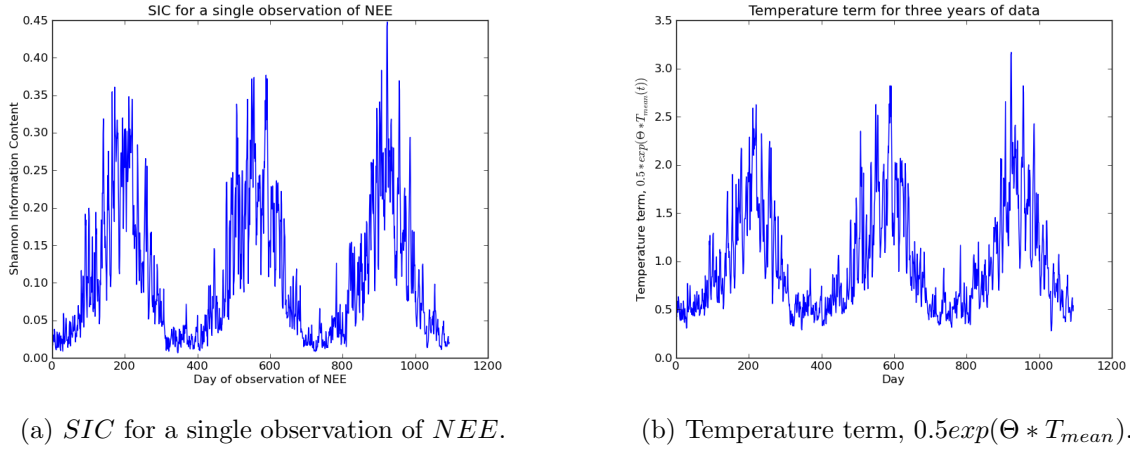


Figure 2: SIC and temperature varying over three years using driving data from Oregon pine forest.

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 show 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.

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, \underline{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 [3]. Firstly the tangent linear model for DALEC was calculated analytically as $\mathbf{M}_i = \frac{\delta \underline{m}_i}{\delta \underline{x}_i}$.

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,

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

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,

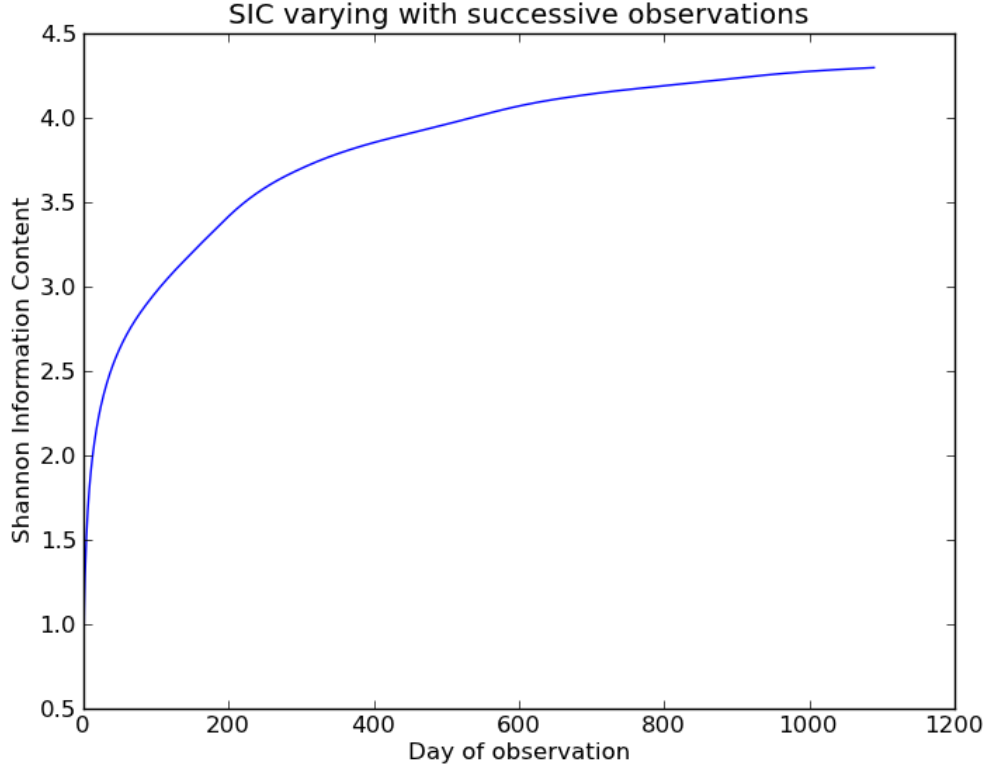


Figure 3: SIC varying as successive observations of Cf are added using driving data from Oregon pine forest.

5 Degrees of freedom for signal

$$DOFS = n - \text{trace}(\mathbf{B}^{-1}\mathbf{A})$$

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] 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.
- [3] John M Lewis, Sivaramakrishnan Lakshmivarahan, and Sudarshan Dhall. *Dynamic data assimilation: a least squares approach*, volume 13. Cambridge University Press, 2006.
- [4] Clive D Rodgers et al. *Inverse methods for atmospheric sounding: Theory and practice*, volume 2. World scientific Singapore, 2000.

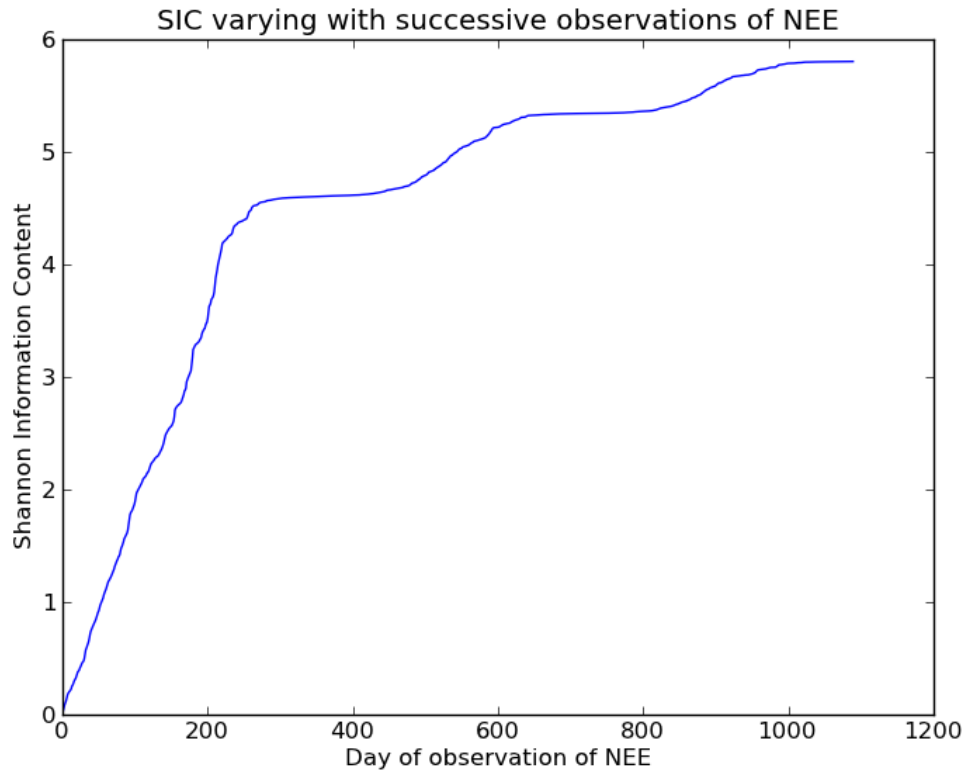


Figure 4: SIC varying as successive observations of NEE are added using driving data from Oregon pine forest.

- [5] 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.

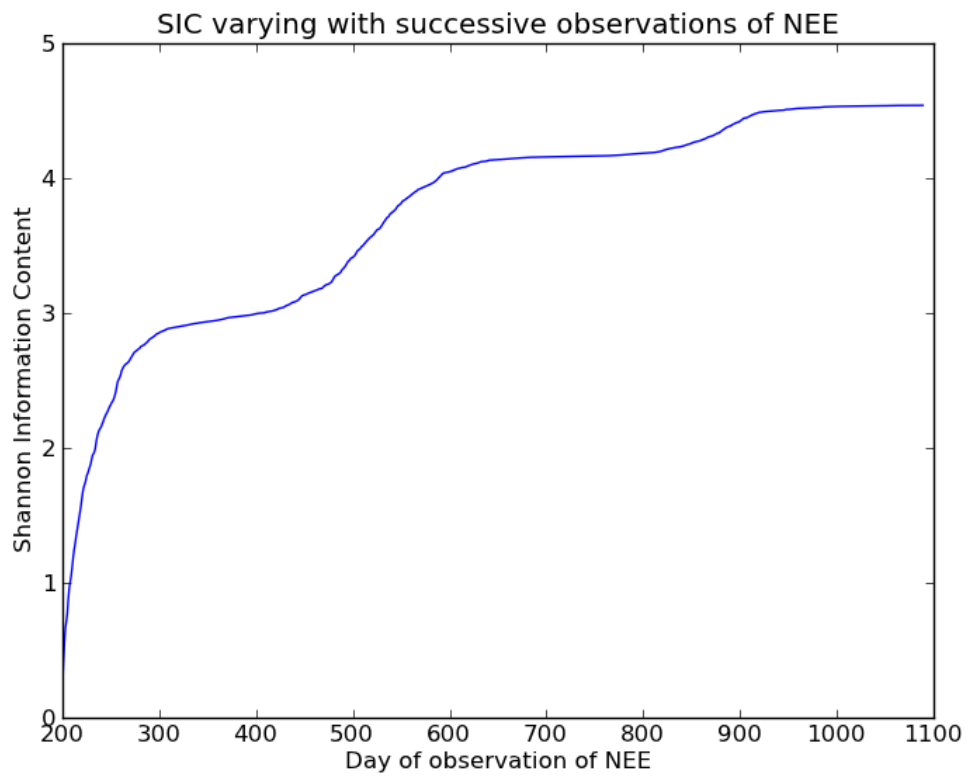


Figure 5: SIC varying as successive observations of NEE are added using driving data from Oregon pine forest. Starting at day 200.

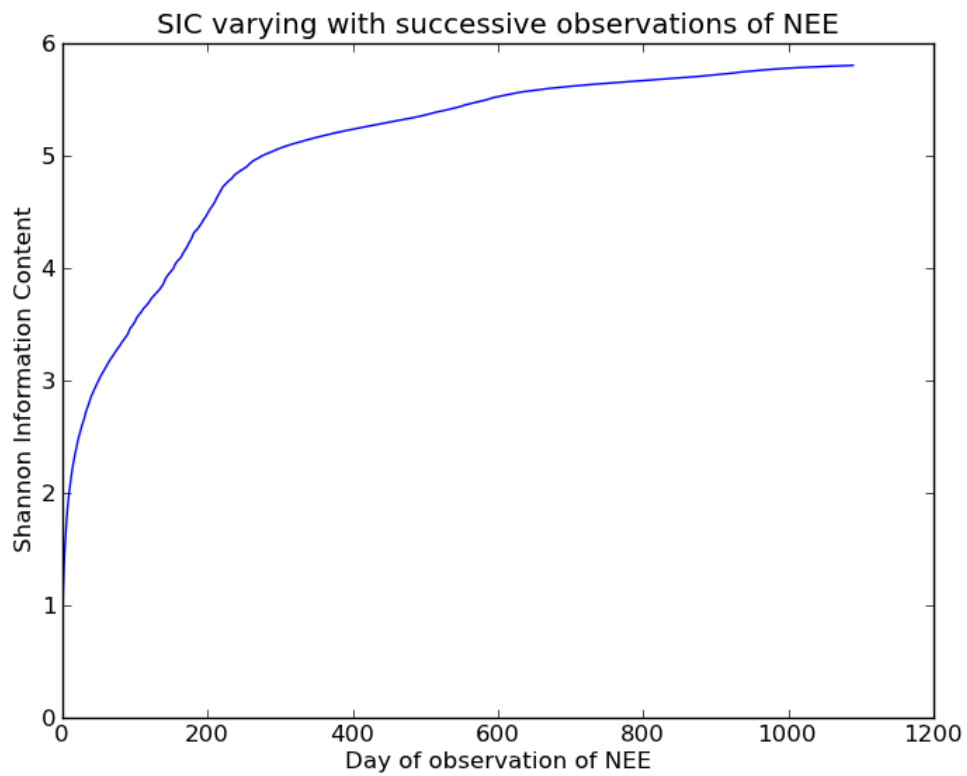


Figure 6: SIC varying as successive observations of NEE and Cf are added using driving data from Oregon pine forest.

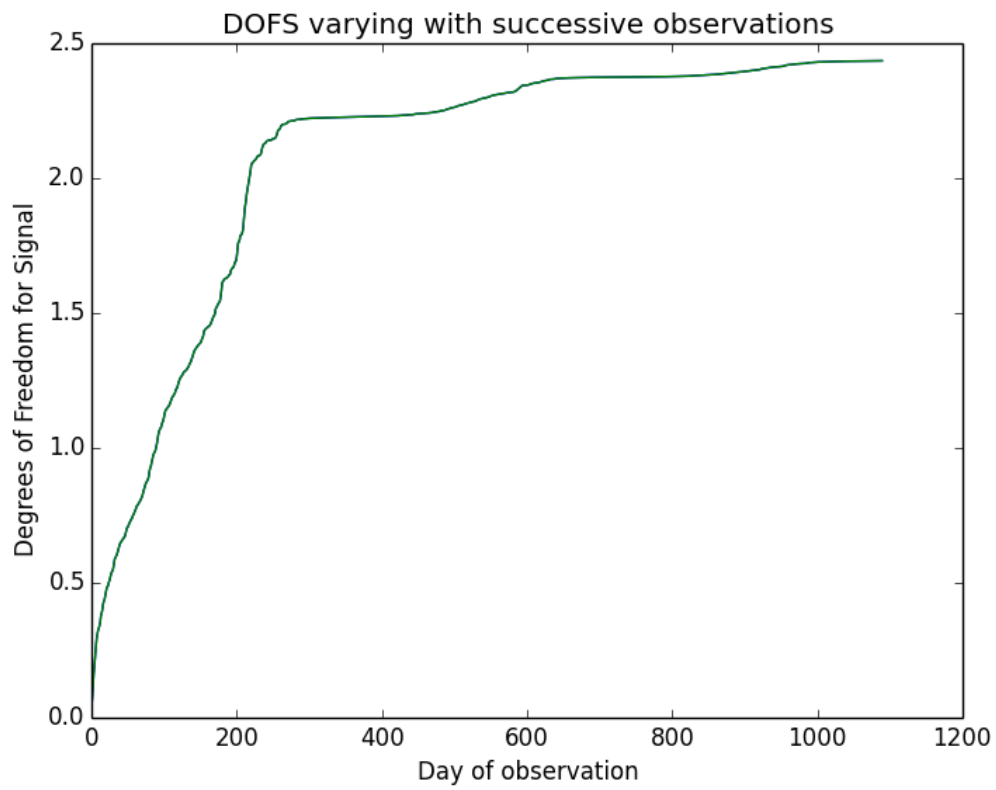


Figure 7: *DOFS* varying as successive observations of *NEE* are added using driving data from Oregon pine forest.