

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

E. Pinnington

September 23, 2014

## Section One

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 [1]. 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. 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,

$$\mathbf{H}_0 = (1 \ 0 \ 0 \ 0 \ 0),$$

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,

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

One of the main observations made of the carbon balance of a forest at flux tower sites is the net ecosystem exchange ( $NEE$ ) of  $\text{CO}_2$ , 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  $\mathbf{H}_0 = \frac{\delta NEE(t_0)}{\delta \mathbf{x}}$  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  $NEE$ ,  $\sigma_{nee,0}^2$ , 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,

$$SIC = \frac{1}{2} \ln \frac{(p_2 - 1)^2 \zeta_0^2 \sigma_{cf,b}^2 + \sigma_{nee,0}^2 + T_0^2 (p_9^2 \sigma_{cs,b}^2 + p_8^2 \sigma_{cl,b}^2)}{\sigma_{nee,0}^2}.$$

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  $GPP$ ,  $\zeta_0^2$ .

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 of carbon fluxes\*\*.

## References

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