

Incorporating land cover uncertainty into the existing uncertainty analysis of the biospheric carbon flux for England and Wales

Keith Harris

31st January 2010

1 A linear combination of code outputs with uncertain coefficients

Gosling and O'Hagan (2007) considered the linear combination

$$Y = \sum_{i=1}^n \alpha_i f_i(\mathbf{x}_i), \quad (1)$$

where α_i was a known scalar and $f_i(\mathbf{x}_i)$ was the scalar output of a computer code run for inputs specified by \mathbf{x}_i for each i . They considered the case where the computer code inputs \mathbf{x}_i were uncertain, the functions f_i were uncertain and there was no training data for most of the functions f_i . We will now consider the case where the scalars α_i are also uncertain.

1.1 General results

Before considering the special case of the uncertainty analysis for the Sheffield Dynamic Global Vegetation Model (SDGVM), we will give results for the general case described above. Here, we will consider any kind of computer model that produces a scalar output $f_i(\mathbf{x}_i)$ from some vector of inputs \mathbf{x}_i . We will also suppose that the sensitivity of the model output can be broken down in any way we like according to the different components of uncertainty in the model output. As mentioned above, in general this will include uncertainty about the inputs \mathbf{x}_i and uncertainty about the functions f_i (due to code uncertainty and possibly also interpolation uncertainty).

We would like to know what value we expect Y to take in general. From equation (1) we see that if we knew the uncertain scalars α_i , then the expectation of Y would be

$$E_{f_i(\mathbf{x}_i)}(Y|\{\alpha_i\}) = \sum_{i=1}^n \alpha_i E_{f_i(\mathbf{x}_i)}(f_i(\mathbf{X}_i)), \quad (2)$$

since $f_i(\mathbf{X}_i)$ doesn't depend on α_i . Thus, our expectation of Y is now given by

$$\begin{aligned} E_{\{\alpha_i\}}(E_{f_i(\mathbf{x}_i)}(Y|\{\alpha_i\})) &= E_{\{\alpha_i\}}\left(\sum_{i=1}^n \alpha_i E_{f_i(\mathbf{x}_i)}(f_i(\mathbf{X}_i))\right) \\ &= \sum_{i=1}^n E_{\alpha_i}(\alpha_i) E_{f_i(\mathbf{x}_i)}(f_i(\mathbf{X}_i)). \end{aligned} \quad (3)$$

We would also like to know the uncertainty around our expectation for Y induced by the uncertainty in the scalars α_i . From equation (2), we see that this is given by

$$\begin{aligned} \text{Var}_{\{\alpha_i\}}(E_{f_i(\mathbf{x}_i)}(Y|\{\alpha_i\})) &= \text{Var}_{\{\alpha_i\}}\left(\sum_{i=1}^n \alpha_i E_{f_i(\mathbf{x}_i)}(f_i(\mathbf{X}_i))\right) \\ &= \sum_{i=1}^n \text{Var}_{\alpha_i}(\alpha_i) (E_{f_i(\mathbf{x}_i)}(f_i(\mathbf{X}_i)))^2 \\ &\quad + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{Cov}_{\alpha_i, \alpha_j}(\alpha_i, \alpha_j) E_{f_i(\mathbf{x}_i)}(f_i(\mathbf{X}_i)) E_{f_j(\mathbf{x}_j)}(f_j(\mathbf{X}_j)). \end{aligned} \quad (4)$$

Finally, we would like to know the uncertainty around our expectation for Y induced by each different component of uncertainty in the model output $f_i(\mathbf{x}_i)$. Whatever source of model output uncertainty we are interested in, whether it be input uncertainty, code uncertainty, interpolation uncertainty or something else, our estimate of the amount of uncertainty induced in Y by this source can always be expressed as

$$E_{\{\alpha_i\}}(E_{S_2}(\text{Var}_{S_1}(E_{S_{\text{rest}}}(Y|S_1, S_2, \{\alpha_i\})|S_2, \{\alpha_i\})|\{\alpha_i\})), \quad (5)$$

where S_1 is the source of model output uncertainty we are interested in, S_2 is another source of model output uncertainty and S_{rest} is the remain-

ing sources of model output uncertainty. From equation (1) and recalling that $f_i(\mathbf{X}_i)$ doesn't depend on α_i we see that the above expression can be simplified to the following form

$$\begin{aligned}
& E_{\{\alpha_i\}} \left(\sum_{i=1}^n \alpha_i^2 E_{S_2}(\text{Var}_{S_1}(E_{S_{\text{rest}}}(f_i(\mathbf{X}_i)|S_1, S_2)|S_2)) + \right. \\
& \left. 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n \alpha_i \alpha_j E_{S_2}(\text{Cov}_{S_1}(E_{S_{\text{rest}}}(f_i(\mathbf{X}_i)|S_1, S_2), E_{S_{\text{rest}}}(f_j(\mathbf{X}_j)|S_1, S_2)|S_2)) \right) \\
& \quad = \sum_{i=1}^n E_{\alpha_i}(\alpha_i^2) E_{S_2}(\text{Var}_{S_1}(E_{S_{\text{rest}}}(f_i(\mathbf{X}_i)|S_1, S_2)|S_2)) + \\
& 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n E_{\alpha_i, \alpha_j}(\alpha_i \alpha_j) E_{S_2}(\text{Cov}_{S_1}(E_{S_{\text{rest}}}(f_i(\mathbf{X}_i)|S_1, S_2), E_{S_{\text{rest}}}(f_j(\mathbf{X}_j)|S_1, S_2)|S_2)).
\end{aligned} \tag{6}$$

1.2 Applying these results to the Sheffield Dynamic Global Vegetation Model

We will now consider the special case of equation (1) where $f_i(\mathbf{x}_i)$ is the scalar output of the SDGVM and use the general results of equations (3), (4) and (6) to derive equivalent results for the SDGVM.

From equation (12) of Gosling and O'Hagan (2007) and our equation (3) we see that the expectation of Y is now given by

$$\begin{aligned}
& E_{\{\alpha_i\}}(E_{\mu_1}(E_{\{f_i\}}(M|\{\mathbf{D}_i\})|\mathcal{M}_E, C(\mathcal{M}_E))) = \\
& E_{\{\alpha_i\}} \left(\sum_{i=1}^n \alpha_i E_{\mu_1}(E_{f_i}(E_{\mathbf{X}_i}(f_i(\mathbf{X}_i)|f_i)|\mathbf{D}_i)|\mathcal{M}_E, C(\mathcal{M}_E)) \right) = \\
& \sum_{i=1}^n E_{\alpha_i}(\alpha_i) E_{\mu_1}(E_{f_i}(E_{\mathbf{X}_i}(f_i(\mathbf{X}_i)|f_i)|\mathbf{D}_i)|\mathcal{M}_E, C(\mathcal{M}_E)). \tag{7}
\end{aligned}$$

The uncertainty in Y due to uncertainty in the scalars α_i , from equa-

tion (4), is given by

$$\begin{aligned}
& \text{Var}_{\{\alpha_i\}}(E_{\mu_1}(E_{\{f_i\}}(M|\{\mathbf{D}_i\})|\mathcal{M}_E, C(\mathcal{M}_E))) = \\
& \text{Var}_{\{\alpha_i\}}\left(\sum_{i=1}^n \alpha_i E_{\mu_1}(E_{f_i}(E_{\mathbf{X}_i}(f_i(\mathbf{X}_i)|f_i)|\mathbf{D}_i)|\mathcal{M}_E, C(\mathcal{M}_E))\right) = \\
& \sum_{i=1}^n \text{Var}_{\alpha_i}(\alpha_i) (E_{\mu_1}(E_{f_i}(E_{\mathbf{X}_i}(f_i(\mathbf{X}_i)|f_i)|\mathbf{D}_i)|\mathcal{M}_E, C(\mathcal{M}_E)))^2 \\
& + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{Cov}_{\alpha_i, \alpha_j}(\alpha_i, \alpha_j) E_{\mu_1}(E_{f_i}(E_{\mathbf{X}_i}(f_i(\mathbf{X}_i)|f_i)|\mathbf{D}_i)|\mathcal{M}_E, C(\mathcal{M}_E)) \\
& \quad E_{\mu_1}(E_{f_j}(E_{\mathbf{X}_j}(f_j(\mathbf{X}_j)|f_j)|\mathbf{D}_j)|\mathcal{M}_E, C(\mathcal{M}_E)). \quad (8)
\end{aligned}$$

The uncertainty in Y due to interpolation, from equation (14) of Gosling and O'Hagan (2007) and our equation (6), is

$$\begin{aligned}
& E_{\{\alpha_i\}}(\text{Var}_{\mu_1}(E_{\{f_i\}}(M|\{\mathbf{D}_i\})|\mathcal{M}_E, C(\mathcal{M}_E))) = \\
& E_{\{\alpha_i\}}\left(\sum_{i=1}^n \alpha_i^2 \text{Var}_{\mu_1}(E_{f_i}(E_{\mathbf{X}_i}(f_i(\mathbf{X}_i)|f_i)|\mathbf{D}_i)|\mathcal{M}_E, C(\mathcal{M}_E)) + \right. \\
& \left. 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n \alpha_i \alpha_j \text{Cov}_{\mu_1}(E_{f_i}(E_{\mathbf{X}_i}(f_i(\mathbf{X}_i)|f_i)|\mathbf{D}_i), E_{f_j}(E_{\mathbf{X}_j}(f_j(\mathbf{X}_j)|f_j)|\mathbf{D}_j)|\mathcal{M}_E, C(\mathcal{M}_E))\right) \\
& = \sum_{i=1}^n E_{\alpha_i}(\alpha_i^2) \text{Var}_{\mu_1}(E_{f_i}(E_{\mathbf{X}_i}(f_i(\mathbf{X}_i)|f_i)|\mathbf{D}_i)|\mathcal{M}_E, C(\mathcal{M}_E)) + \\
& 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n E_{\alpha_i, \alpha_j}(\alpha_i \alpha_j) \text{Cov}_{\mu_1}(E_{f_i}(E_{\mathbf{X}_i}(f_i(\mathbf{X}_i)|f_i)|\mathbf{D}_i), E_{f_j}(E_{\mathbf{X}_j}(f_j(\mathbf{X}_j)|f_j)|\mathbf{D}_j)|\mathcal{M}_E, C(\mathcal{M}_E)). \quad (9)
\end{aligned}$$

The uncertainty in Y due to emulation, from equation (16) of Gosling and O'Hagan (2007) and our equation (6), is

$$\begin{aligned}
& E_{\{\alpha_i\}}(E_{\mu_2}(\text{Var}_{\{f_i\}}(M|\{\mathbf{D}_i\})|\mathcal{M}_V, C(\mathcal{M}_V))) = \\
& E_{\{\alpha_i\}}\left(\sum_{i=1}^n \alpha_i^2 E_{\mu_2}(\text{Var}_{f_i}(E_{\mathbf{X}_i}(f_i(\mathbf{X}_i)|f_i)|\mathbf{D}_i)|\mathcal{M}_V, C(\mathcal{M}_V))\right) = \\
& \sum_{i=1}^n E_{\alpha_i}(\alpha_i^2) E_{\mu_2}(\text{Var}_{f_i}(E_{\mathbf{X}_i}(f_i(\mathbf{X}_i)|f_i)|\mathbf{D}_i)|\mathcal{M}_V, C(\mathcal{M}_V)). \quad (10)
\end{aligned}$$

The uncertainty in Y due to input uncertainty, from equation (17) of Gosling and O'Hagan (2007) and our equation (6), is

$$\begin{aligned}
E_{\{\alpha_i\}}(E_{\mu_3}(E_{\{f_i\}}(V|\{\mathbf{D}_i\})|\mathcal{V}_E, C(\mathcal{V}_E))) = \\
E_{\{\alpha_i\}}\left(\sum_{i=1}^n \alpha_i^2 C_{ii} + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n \alpha_i \alpha_j C_{ij}\right) = \\
\sum_{i=1}^n E_{\alpha_i}(\alpha_i^2) C_{ii} + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n E_{\alpha_i, \alpha_j}(\alpha_i \alpha_j) C_{ij}, \quad (11)
\end{aligned}$$

where

$$C_{ii} = E_{\mu_3}(E_{f_i}(\text{Var}_{\mathbf{X}_i}(f_i(\mathbf{X}_i)|f_i)|\mathbf{D}_i)|\mathcal{V}_E, C(\mathcal{V}_E))$$

and

$$C_{ij} = E_{\mu_3}(E_{f_i, f_j}(\text{Cov}_{\mathbf{X}_i, \mathbf{X}_j}(f_i(\mathbf{X}_i), f_j(\mathbf{X}_j)|f_i, f_j)|\mathbf{D}_i, \mathbf{D}_j)|\mathcal{V}_E, C(\mathcal{V}_E)).$$

This completes the extended uncertainty analysis where the scalars α_i are also uncertain. To use these extended results we need to be able to calculate: $E_{\alpha_i}(\alpha_i)$, $\text{Var}_{\alpha_i}(\alpha_i)$, $\text{Cov}_{\alpha_i, \alpha_j}(\alpha_i, \alpha_j)$,

$$E_{\alpha_i}(\alpha_i^2) = \text{Var}_{\alpha_i}(\alpha_i) + (E_{\alpha_i}(\alpha_i))^2$$

and

$$E_{\alpha_i, \alpha_j}(\alpha_i \alpha_j) = \text{Cov}_{\alpha_i, \alpha_j}(\alpha_i, \alpha_j) + E_{\alpha_i}(\alpha_i) E_{\alpha_j}(\alpha_j).$$

For the case of analysing the uncertainty in the total biospheric carbon flux for England and Wales in the year 2000 these quantities have already been calculated by Cripps *et al.* (2008).

2 Aggregation for NBP

As noted by Gosling and O'Hagan (2007), in Kennedy *et al.* (2008), the linear combination of computer code outputs was

$$Y = \sum_{k=1}^{707} \sum_{t=1}^4 a_k \gamma_t(\mathbf{y}_k) f_{tk}(\mathbf{x}_{tk}),$$

where k was the site number, each site's coordinates were given by the vector \mathbf{y}_k , t was the plant functional type (PFT) index, a_k was the known area of site k , $\gamma_t(\mathbf{y}_k)$ was the proportion of PFT t at site k , which was assumed known, and $f_{tk}(\mathbf{x}_{tk})$ was the computer code NBP output of PFT t at site k for input \mathbf{x}_{tk} . We will consider the same sum, but where only one set of weights, the site areas, are assumed known. The PFT proportions at each site $\gamma_t(\mathbf{y}_k)$ will be considered unknown and their expectations, variances and covariances were calculated as in Cripps *et al.* (2008). Here, we have a double summation, but we could represent it as a single summation as considered in Section 1, with index i of equation (1) running over all 2828 pairs (t, k) and each α_i is the product of a known site weight a_k and an uncertain PFT weight $\gamma_t(\mathbf{y}_k)$. However, although the sites have a spatial structure that could be exploited for interpolation as before, the (t, k) pairs do not. Furthermore, there is interest not just in the grand total Y but also in the site NBP aggregated over PFTs,

$$Y_{.k} = \sum_{t=1}^4 \gamma_t(\mathbf{y}_k) f_{tk}(\mathbf{x}_{tk}),$$

and in each PFT's contribution to the total NBP, aggregated over England and Wales,

$$Y_{t.} = \sum_{k=1}^{707} a_k \gamma_t(\mathbf{y}_k) f_{tk}(\mathbf{x}_{tk}).$$

Although the methods developed earlier do not apply to the double summation, they apply to a single summation over t or over k , and this gives us two ways to quantify uncertainty in Y . We can first analyse the uncertainty in each $Y_{.k}$ and then in

$$Y = \sum_{k=1}^{707} a_k Y_{.k}.$$

Alternatively, we could first analyse the uncertainty in each $Y_{t.}$ and then in

$$Y = \sum_{t=1}^4 Y_{t.}.$$

In this report, we used the latter approach to quantify the uncertainty in the grand total Y .

2.1 Aggregating over PFTs

2.1.1 General results

We can analyse the uncertainty in each Y_k using the earlier results of Section 1.1 for a single summation, but with index i of equations (1) to (6) replaced by t , n replaced by 4, α_i replaced by $\gamma_t(\mathbf{y}_k)$ and $f_i(\mathbf{x}_i)$ replaced by $f_{tk}(\mathbf{x}_{tk})$.

2.1.2 Application to the SDGVM

Thus, from equation (7) we see that the expectation of Y_k is given by

$$\sum_{t=1}^4 E_{\gamma_t(\mathbf{y}_k)}(\gamma_t(\mathbf{y}_k)) E_{\mu_1}(E_{f_{tk}}(E_{\mathbf{X}_{tk}}(f_{tk}(\mathbf{X}_{tk})|f_{tk})|\mathbf{D}_{tk})|\mathcal{M}_E, C(\mathcal{M}_E)).$$

From equation (8) we see that the uncertainty in Y_k due to uncertainty in the scalars $\gamma_t(\mathbf{y}_k)$ is given by

$$\begin{aligned} & \sum_{t=1}^4 \text{Var}_{\gamma_t(\mathbf{y}_k)}(\gamma_t(\mathbf{y}_k)) (E_{\mu_1}(E_{f_{tk}}(E_{\mathbf{X}_{tk}}(f_{tk}(\mathbf{X}_{tk})|f_{tk})|\mathbf{D}_{tk})|\mathcal{M}_E, C(\mathcal{M}_E)))^2 \\ +2 & \sum_{t=1}^3 \sum_{t'=t+1}^4 \text{Cov}_{\gamma_t(\mathbf{y}_k), \gamma_{t'}(\mathbf{y}_k)}(\gamma_t(\mathbf{y}_k), \gamma_{t'}(\mathbf{y}_k)) E_{\mu_1}(E_{f_{tk}}(E_{\mathbf{X}_{tk}}(f_{tk}(\mathbf{X}_{tk})|f_{tk})|\mathbf{D}_{tk})|\mathcal{M}_E, C(\mathcal{M}_E)) \\ & \quad E_{\mu_1}(E_{f_{t'k}}(E_{\mathbf{X}_{t'k}}(f_{t'k}(\mathbf{X}_{t'k})|f_{t'k})|\mathbf{D}_{t'k})|\mathcal{M}_E, C(\mathcal{M}_E)). \end{aligned}$$

From equation (9) we see that the uncertainty in Y_k due to interpolation is given by

$$\begin{aligned} & \sum_{t=1}^4 E_{\gamma_t(\mathbf{y}_k)}(\gamma_t(\mathbf{y}_k)^2) \text{Var}_{\mu_1}(E_{f_{tk}}(E_{\mathbf{X}_{tk}}(f_{tk}(\mathbf{X}_{tk})|f_{tk})|\mathbf{D}_{tk})|\mathcal{M}_E, C(\mathcal{M}_E)) \\ +2 & \sum_{t=1}^3 \sum_{t'=t+1}^4 E_{\gamma_t(\mathbf{y}_k), \gamma_{t'}(\mathbf{y}_k)}(\gamma_t(\mathbf{y}_k)\gamma_{t'}(\mathbf{y}_k)) \text{Cov}_{\mu_1}(E_{f_{tk}}(E_{\mathbf{X}_{tk}}(f_{tk}(\mathbf{X}_{tk})|f_{tk})|\mathbf{D}_{tk}), \\ & \quad E_{f_{t'k}}(E_{\mathbf{X}_{t'k}}(f_{t'k}(\mathbf{X}_{t'k})|f_{t'k})|\mathbf{D}_{t'k})|\mathcal{M}_E, C(\mathcal{M}_E)), \quad (12) \end{aligned}$$

where

$$E_{\gamma_t(\mathbf{y}_k)}(\gamma_t(\mathbf{y}_k)^2) = \text{Var}_{\gamma_t(\mathbf{y}_k)}(\gamma_t(\mathbf{y}_k)) + (E_{\gamma_t(\mathbf{y}_k)}(\gamma_t(\mathbf{y}_k)))^2$$

and

$$E_{\gamma_t(\mathbf{y}_k), \gamma_{t'}(\mathbf{y}_k)}(\gamma_t(\mathbf{y}_k)\gamma_{t'}(\mathbf{y}_k)) = \text{Cov}_{\gamma_t(\mathbf{y}_k), \gamma_{t'}(\mathbf{y}_k)}(\gamma_t(\mathbf{y}_k), \gamma_{t'}(\mathbf{y}_k)) + E_{\gamma_t(\mathbf{y}_k)}(\gamma_t(\mathbf{y}_k))E_{\gamma_{t'}(\mathbf{y}_k)}(\gamma_{t'}(\mathbf{y}_k)).$$

Since kriging is done on each PFT separately (see Gosling and O'Hagan (2007) Section 3.1), expression (12) in this case simplifies to

$$\sum_{t=1}^4 E_{\gamma_t(\mathbf{y}_k)}(\gamma_t(\mathbf{y}_k)^2) \text{Var}_{\mu_1}(E_{f_{tk}}(E_{\mathbf{X}_{tk}}(f_{tk}(\mathbf{X}_{tk})|f_{tk})|\mathbf{D}_{tk})|\mathcal{M}_E, C(\mathcal{M}_E)).$$

From equation (10) we see that the uncertainty in Y_k due to emulation is given by

$$\sum_{t=1}^4 E_{\gamma_t(\mathbf{y}_k)}(\gamma_t(\mathbf{y}_k)^2) E_{\mu_2}(\text{Var}_{f_{tk}}(E_{\mathbf{X}_{tk}}(f_{tk}(\mathbf{X}_{tk})|f_{tk})|\mathbf{D}_{tk})|\mathcal{M}_V, C(\mathcal{M}_V)).$$

From equation (11) we see that the uncertainty in Y_k due to input uncertainty is given by

$$\sum_{t=1}^4 E_{\gamma_t(\mathbf{y}_k)}(\gamma_t(\mathbf{y}_k)^2) C_{tt}^{(k)} + 2 \sum_{t=1}^3 \sum_{t'=t+1}^4 E_{\gamma_t(\mathbf{y}_k), \gamma_{t'}(\mathbf{y}_k)}(\gamma_t(\mathbf{y}_k)\gamma_{t'}(\mathbf{y}_k)) C_{tt'}^{(k)},$$

where

$$C_{tt}^{(k)} = E_{\mu_3}(E_{f_{tk}}(\text{Var}_{\mathbf{X}_{tk}}(f_{tk}(\mathbf{X}_{tk})|f_{tk})|\mathbf{D}_{tk})|\mathcal{V}_E, C(\mathcal{V}_E))$$

and

$$C_{tt'}^{(k)} = E_{\mu_{tt'}}(E_{f_{tk}, f_{t'k}}(\text{Cov}_{\mathbf{X}_{tk}, \mathbf{X}_{t'k}}(f_{tk}(\mathbf{X}_{tk}), f_{t'k}(\mathbf{X}_{t'k})|f_{tk}, f_{t'k})|\mathbf{D}_{tk}, \mathbf{D}_{t'k})|\mathcal{C}_{E_{tt'}}, C(\mathcal{C}_{E_{tt'}})).$$

2.2 Aggregating over sites

2.2.1 General results

Similar to analysing the uncertainty in each Y_k , we can analyse the uncertainty in each Y_t using the earlier results for a single summation, but with index i of equations (1) to (6) replaced by k , n replaced by 707, α_i replaced by $a_k\gamma_t(\mathbf{y}_k)$ and $f_i(\mathbf{x}_i)$ replaced by $f_{tk}(\mathbf{x}_{tk})$. We then just need to simplify

the resulting expressions slightly by using the fact that

$$E_{\alpha_i}(\alpha_i) = E_{\gamma_t(\mathbf{y}_k)}(a_k \gamma_t(\mathbf{y}_k)) = a_k E_{\gamma_t(\mathbf{y}_k)}(\gamma_t(\mathbf{y}_k)),$$

$$\text{Var}_{\alpha_i}(\alpha_i) = \text{Var}_{\gamma_t(\mathbf{y}_k)}(a_k \gamma_t(\mathbf{y}_k)) = a_k^2 \text{Var}_{\gamma_t(\mathbf{y}_k)}(\gamma_t(\mathbf{y}_k)),$$

and

$$\begin{aligned} \text{Cov}_{\alpha_i, \alpha_j}(\alpha_i, \alpha_j) &= \text{Cov}_{\gamma_t(\mathbf{y}_k), \gamma_t(\mathbf{y}_{k'})}(a_k \gamma_t(\mathbf{y}_k), a_{k'} \gamma_t(\mathbf{y}_{k'})) \\ &= a_k a_{k'} \text{Cov}_{\gamma_t(\mathbf{y}_k), \gamma_t(\mathbf{y}_{k'})}(\gamma_t(\mathbf{y}_k), \gamma_t(\mathbf{y}_{k'})). \end{aligned}$$

2.2.2 Application to the SDGVM

Thus, from equation (7) we see that the expectation of Y_t is given by

$$\begin{aligned} & \sum_{k=1}^{707} E_{\gamma_t(\mathbf{y}_k)}(a_k \gamma_t(\mathbf{y}_k)) E_{\mu_1}(E_{f_{tk}}(E_{\mathbf{X}_{tk}}(f_{tk}(\mathbf{X}_{tk})|f_{tk})|\mathbf{D}_{tk})|\mathcal{M}_E, C(\mathcal{M}_E)) \\ &= \sum_{k=1}^{707} a_k E_{\gamma_t(\mathbf{y}_k)}(\gamma_t(\mathbf{y}_k)) E_{\mu_1}(E_{f_{tk}}(E_{\mathbf{X}_{tk}}(f_{tk}(\mathbf{X}_{tk})|f_{tk})|\mathbf{D}_{tk})|\mathcal{M}_E, C(\mathcal{M}_E)). \end{aligned} \tag{13}$$

From equation (8) we see that the uncertainty in Y_t due to uncertainty in the scalars $\gamma_t(\mathbf{y}_k)$ is given by

$$\begin{aligned} & \sum_{k=1}^{707} \text{Var}_{\gamma_t(\mathbf{y}_k)}(a_k \gamma_t(\mathbf{y}_k)) (E_{\mu_1}(E_{f_{tk}}(E_{\mathbf{X}_{tk}}(f_{tk}(\mathbf{X}_{tk})|f_{tk})|\mathbf{D}_{tk})|\mathcal{M}_E, C(\mathcal{M}_E)))^2 \\ &+ 2 \sum_{k=1}^{706} \sum_{k'=k+1}^{707} \text{Cov}_{\gamma_t(\mathbf{y}_k), \gamma_t(\mathbf{y}_{k'})}(a_k \gamma_t(\mathbf{y}_k), a_{k'} \gamma_t(\mathbf{y}_{k'})) E_{\mu_1}(E_{f_{tk}}(E_{\mathbf{X}_{tk}}(f_{tk}(\mathbf{X}_{tk})|f_{tk})|\mathbf{D}_{tk})|\mathcal{M}_E, C(\mathcal{M}_E)) \\ & \quad E_{\mu_1}(E_{f_{tk'}}(E_{\mathbf{X}_{tk'}}(f_{tk'}(\mathbf{X}_{tk'})|f_{tk'})|\mathbf{D}_{tk'})|\mathcal{M}_E, C(\mathcal{M}_E)) \\ &= \sum_{k=1}^{707} a_k^2 \text{Var}_{\gamma_t(\mathbf{y}_k)}(\gamma_t(\mathbf{y}_k)) (E_{\mu_1}(E_{f_{tk}}(E_{\mathbf{X}_{tk}}(f_{tk}(\mathbf{X}_{tk})|f_{tk})|\mathbf{D}_{tk})|\mathcal{M}_E, C(\mathcal{M}_E)))^2 \\ &+ 2 \sum_{k=1}^{706} \sum_{k'=k+1}^{707} a_k a_{k'} \text{Cov}_{\gamma_t(\mathbf{y}_k), \gamma_t(\mathbf{y}_{k'})}(\gamma_t(\mathbf{y}_k), \gamma_t(\mathbf{y}_{k'})) E_{\mu_1}(E_{f_{tk}}(E_{\mathbf{X}_{tk}}(f_{tk}(\mathbf{X}_{tk})|f_{tk})|\mathbf{D}_{tk})|\mathcal{M}_E, C(\mathcal{M}_E)) \\ & \quad E_{\mu_1}(E_{f_{tk'}}(E_{\mathbf{X}_{tk'}}(f_{tk'}(\mathbf{X}_{tk'})|f_{tk'})|\mathbf{D}_{tk'})|\mathcal{M}_E, C(\mathcal{M}_E)). \end{aligned}$$

From equation (9) we see that the uncertainty in Y_t due to interpolation

is given by

$$\begin{aligned}
& \sum_{k=1}^{707} E_{\gamma_t(\mathbf{y}_k)} \left((a_k \gamma_t(\mathbf{y}_k))^2 \right) \text{Var}_{\mu_1} \left(E_{f_{tk}} \left(E_{\mathbf{X}_{tk}} (f_{tk}(\mathbf{X}_{tk}) | f_{tk}) | \mathbf{D}_{tk} \right) | \mathcal{M}_E, C(\mathcal{M}_E) \right) \\
& + 2 \sum_{k=1}^{706} \sum_{k'=k+1}^{707} E_{\gamma_t(\mathbf{y}_k), \gamma_t(\mathbf{y}_{k'})} (a_k \gamma_t(\mathbf{y}_k) a_{k'} \gamma_t(\mathbf{y}_{k'})) \text{Cov}_{\mu_1} \left(E_{f_{tk}} \left(E_{\mathbf{X}_{tk}} (f_{tk}(\mathbf{X}_{tk}) | f_{tk}) | \mathbf{D}_{tk} \right), \right. \\
& \quad \left. E_{f_{tk'}} \left(E_{\mathbf{X}_{tk'}} (f_{tk'}(\mathbf{X}_{tk'}) | f_{tk'}) | \mathbf{D}_{tk'} \right) | \mathcal{M}_E, C(\mathcal{M}_E) \right) \\
& = \sum_{k=1}^{707} a_k^2 E_{\gamma_t(\mathbf{y}_k)} (\gamma_t(\mathbf{y}_k)^2) \text{Var}_{\mu_1} \left(E_{f_{tk}} \left(E_{\mathbf{X}_{tk}} (f_{tk}(\mathbf{X}_{tk}) | f_{tk}) | \mathbf{D}_{tk} \right) | \mathcal{M}_E, C(\mathcal{M}_E) \right) \\
& + 2 \sum_{k=1}^{706} \sum_{k'=k+1}^{707} a_k a_{k'} E_{\gamma_t(\mathbf{y}_k), \gamma_t(\mathbf{y}_{k'})} (\gamma_t(\mathbf{y}_k) \gamma_t(\mathbf{y}_{k'})) \text{Cov}_{\mu_1} \left(E_{f_{tk}} \left(E_{\mathbf{X}_{tk}} (f_{tk}(\mathbf{X}_{tk}) | f_{tk}) | \mathbf{D}_{tk} \right), \right. \\
& \quad \left. E_{f_{tk'}} \left(E_{\mathbf{X}_{tk'}} (f_{tk'}(\mathbf{X}_{tk'}) | f_{tk'}) | \mathbf{D}_{tk'} \right) | \mathcal{M}_E, C(\mathcal{M}_E) \right),
\end{aligned}$$

where

$$E_{\gamma_t(\mathbf{y}_k), \gamma_t(\mathbf{y}_{k'})} (\gamma_t(\mathbf{y}_k) \gamma_t(\mathbf{y}_{k'})) = \text{Cov}_{\gamma_t(\mathbf{y}_k), \gamma_t(\mathbf{y}_{k'})} (\gamma_t(\mathbf{y}_k), \gamma_t(\mathbf{y}_{k'})) + E_{\gamma_t(\mathbf{y}_k)} (\gamma_t(\mathbf{y}_k)) E_{\gamma_t(\mathbf{y}_{k'})} (\gamma_t(\mathbf{y}_{k'})).$$

From equation (10) we see that the uncertainty in Y_t due to emulation is given by

$$\begin{aligned}
& \sum_{k=1}^{707} E_{\gamma_t(\mathbf{y}_k)} \left((a_k \gamma_t(\mathbf{y}_k))^2 \right) E_{\mu_2} \left(\text{Var}_{f_{tk}} \left(E_{\mathbf{X}_{tk}} (f_{tk}(\mathbf{X}_{tk}) | f_{tk}) | \mathbf{D}_{tk} \right) | \mathcal{M}_V, C(\mathcal{M}_V) \right) \\
& = \sum_{k=1}^{707} a_k^2 E_{\gamma_t(\mathbf{y}_k)} (\gamma_t(\mathbf{y}_k)^2) E_{\mu_2} \left(\text{Var}_{f_{tk}} \left(E_{\mathbf{X}_{tk}} (f_{tk}(\mathbf{X}_{tk}) | f_{tk}) | \mathbf{D}_{tk} \right) | \mathcal{M}_V, C(\mathcal{M}_V) \right).
\end{aligned}$$

From equation (11) we see that the uncertainty in Y_t due to input uncertainty is given by

$$\begin{aligned}
& \sum_{k=1}^{707} E_{\gamma_t(\mathbf{y}_k)} \left((a_k \gamma_t(\mathbf{y}_k))^2 \right) C_{kk}^{(t)} + 2 \sum_{k=1}^{706} \sum_{k'=k+1}^{707} E_{\gamma_t(\mathbf{y}_k), \gamma_t(\mathbf{y}_{k'})} (a_k \gamma_t(\mathbf{y}_k) a_{k'} \gamma_t(\mathbf{y}_{k'})) C_{kk'}^{(t)} \\
& = \sum_{k=1}^{707} a_k^2 E_{\gamma_t(\mathbf{y}_k)} (\gamma_t(\mathbf{y}_k)^2) C_{kk}^{(t)} + 2 \sum_{k=1}^{706} \sum_{k'=k+1}^{707} a_k a_{k'} E_{\gamma_t(\mathbf{y}_k), \gamma_t(\mathbf{y}_{k'})} (\gamma_t(\mathbf{y}_k) \gamma_t(\mathbf{y}_{k'})) C_{kk'}^{(t)},
\end{aligned} \tag{14}$$

where

$$C_{kk}^{(t)} = E_{\mu_3} (E_{f_{tk}} (\text{Var}_{\mathbf{X}_{tk}} (f_{tk}(\mathbf{X}_{tk}) | f_{tk}) | \mathbf{D}_{tk}) | \mathcal{V}_E, C(\mathcal{V}_E))$$

and

$$C_{kk'}^{(t)} = E_{\mu_3} (E_{f_{tk}, f_{tk'}} (\text{Cov}_{\mathbf{X}_{tk}, \mathbf{X}_{tk'}} (f_{tk}(\mathbf{X}_{tk}), f_{tk'}(\mathbf{X}_{tk'}) | f_{tk}, f_{tk'}) | \mathbf{D}_{tk}, \mathbf{D}_{tk'}) | \mathcal{V}_E, C(\mathcal{V}_E)).$$

Kennedy *et al.* (2008) approximated $C_{kk'}^{(t)}$ by $\rho_t \sqrt{C_{kk}^{(t)} C_{k'k'}^{(t)}}$, where $\rho_1 = 0.051$, $\rho_2 = 0.012$, $\rho_3 = 0.006$ and $\rho_4 = 0.182$ for grassland, crop, DBL and ENL, respectively.

2.3 Aggregating over sites and PFTs

2.3.1 General results

In the previous section, we gave results that allow us to quantify the uncertainty in each Y_t . We now need to aggregate over PFTs to obtain the corresponding analysis for $Y = \sum_{t=1}^4 Y_t$.

Firstly, we would again like to know what value we expect Y to take in general and an expression for this value that we can compute is given below.

$$\begin{aligned} E_{\{\gamma_t(\mathbf{y}_k)\}} (E_{f_{tk}(\mathbf{x}_{tk})} (Y | \{\gamma_t(\mathbf{y}_k)\})) &= E_{\{\gamma_t(\mathbf{y}_k)\}} \left(\sum_{t=1}^4 E_{f_{tk}(\mathbf{x}_{tk})} (Y_t | \{\gamma_t(\mathbf{y}_k)\}) \right) \\ &= \sum_{t=1}^4 E_{\{\gamma_t(\mathbf{y}_k)\}} (E_{f_{tk}(\mathbf{x}_{tk})} (Y_t | \{\gamma_t(\mathbf{y}_k)\})), \end{aligned} \tag{15}$$

where the expectations in the above sum are given by equation (3).

Secondly, we would again like to know the uncertainty around our expectation for Y induced by the uncertainty in the scalar weights $\{\gamma_t(\mathbf{y}_k)\}$

and an expression for this value that we can compute is given below.

$$\begin{aligned}
\text{Var}_{\{\gamma_t(\mathbf{y}_k)\}}(E_{f_{tk}(\mathbf{x}_{tk})}(Y|\{\gamma_t(\mathbf{y}_k)\})) &= \text{Var}_{\{\gamma_t(\mathbf{y}_k)\}}\left(\sum_{t=1}^4 E_{f_{tk}(\mathbf{x}_{tk})}(Y_t|\{\gamma_t(\mathbf{y}_k)\})\right) \\
&= \sum_{t=1}^4 \text{Var}_{\{\gamma_t(\mathbf{y}_k)\}}(E_{f_{tk}(\mathbf{x}_{tk})}(Y_t|\{\gamma_t(\mathbf{y}_k)\})) + \\
2 \sum_{t=1}^3 \sum_{t'=t+1}^4 \text{Cov}_{\{\gamma_t(\mathbf{y}_k)\},\{\gamma_{t'}(\mathbf{y}_{k'})\}} &(E_{f_{tk}(\mathbf{x}_{tk})}(Y_t|\{\gamma_t(\mathbf{y}_k)\}), E_{f_{t'k'}(\mathbf{x}_{t'k'})}(Y_{t'}|\{\gamma_{t'}(\mathbf{y}_{k'})\})),
\end{aligned} \tag{16}$$

where the variance terms in the above expression are given by equation (4) and the covariance terms simplify as follows.

$$\begin{aligned}
&\text{Cov}_{\{\gamma_t(\mathbf{y}_k)\},\{\gamma_{t'}(\mathbf{y}_{k'})\}}(E_{f_{tk}(\mathbf{x}_{tk})}(Y_t|\{\gamma_t(\mathbf{y}_k)\}), E_{f_{t'k'}(\mathbf{x}_{t'k'})}(Y_{t'}|\{\gamma_{t'}(\mathbf{y}_{k'})\})) \\
&= \text{Cov}_{\{\gamma_t(\mathbf{y}_k)\},\{\gamma_{t'}(\mathbf{y}_{k'})\}}\left(E_{f_{tk}(\mathbf{x}_{tk})}\left(\sum_{k=1}^{707} a_k \gamma_t(\mathbf{y}_k) f_{tk}(\mathbf{X}_{tk}) \middle| \{\gamma_t(\mathbf{y}_k)\}\right), \right. \\
&\quad \left. E_{f_{t'k'}(\mathbf{x}_{t'k'})}\left(\sum_{k'=1}^{707} a_{k'} \gamma_{t'}(\mathbf{y}_{k'}) f_{t'k'}(\mathbf{X}_{t'k'}) \middle| \{\gamma_{t'}(\mathbf{y}_{k'})\}\right)\right) \\
&= \text{Cov}_{\{\gamma_t(\mathbf{y}_k)\},\{\gamma_{t'}(\mathbf{y}_{k'})\}}\left(\sum_{k=1}^{707} a_k \gamma_t(\mathbf{y}_k) E_{f_{tk}(\mathbf{x}_{tk})}(f_{tk}(\mathbf{X}_{tk})), \sum_{k'=1}^{707} a_{k'} \gamma_{t'}(\mathbf{y}_{k'}) E_{f_{t'k'}(\mathbf{x}_{t'k'})}(f_{t'k'}(\mathbf{X}_{t'k'}))\right) \\
&= \sum_{k=1}^{707} \sum_{k'=1}^{707} a_k a_{k'} E_{f_{tk}(\mathbf{x}_{tk})}(f_{tk}(\mathbf{X}_{tk})) E_{f_{t'k'}(\mathbf{x}_{t'k'})}(f_{t'k'}(\mathbf{X}_{t'k'})) \text{Cov}_{\gamma_t(\mathbf{y}_k),\gamma_{t'}(\mathbf{y}_{k'})}(\gamma_t(\mathbf{y}_k), \gamma_{t'}(\mathbf{y}_{k'})).
\end{aligned} \tag{17}$$

Finally, we would again like to know the uncertainty around our expectation for Y induced by each different component of uncertainty in the model output $f_i(\mathbf{x}_i)$. Similar to equation (5), this can always be expressed as

$$E_{\{\gamma_t(\mathbf{y}_k)\}}(E_{S_2}(\text{Var}_{S_1}(E_{S_{\text{rest}}}(Y|S_1, S_2, \{\gamma_t(\mathbf{y}_k)\})|S_2, \{\gamma_t(\mathbf{y}_k)\})|\{\gamma_t(\mathbf{y}_k)\})),$$

which, similar to equation (6), can be simplified to the following form

$$\begin{aligned}
& \sum_{t=1}^4 E_{\{\gamma_t(\mathbf{y}_k)\}} (E_{S_2}(\text{Var}_{S_1}(E_{S_{\text{rest}}}(Y_t | S_1, S_2, \{\gamma_t(\mathbf{y}_k)\})) | S_2, \{\gamma_t(\mathbf{y}_k)\}) | \{\gamma_t(\mathbf{y}_k)\})) + \\
& 2 \sum_{t=1}^3 \sum_{t'=t+1}^4 E_{\{\gamma_t(\mathbf{y}_k)\}} (E_{S_2}(\text{Cov}_{S_1}(E_{S_{\text{rest}}}(Y_t | S_1, S_2, \{\gamma_t(\mathbf{y}_k)\}), \\
& \quad E_{S_{\text{rest}}}(Y_{t'} | S_1, S_2, \{\gamma_{t'}(\mathbf{y}_k)\}) | S_2, \{\gamma_t(\mathbf{y}_k)\}) | \{\gamma_t(\mathbf{y}_k)\})), \quad (18)
\end{aligned}$$

where the expected variances in the above expression are given by equation (6) and the expected covariances are often 0 due to independence.

2.3.2 Application to the SDGVM

The estimate of Y from equation (15) is of course just the sum of the estimates of Y_t , which are given by equation (13).

From equation (16), the variance in Y due to uncertainty in the scalars $\gamma_t(\mathbf{y}_k)$ is given by

$$\begin{aligned}
& \text{Var}_{\{\gamma_t(\mathbf{y}_k)\}} (E_{\mu_1} (E_{\{f_{tk}\}} (E_{\mathbf{X}}(Y | \{f_{tk}\}) | \{\mathbf{D}_{tk}\}) | \mathcal{M}_E, C(\mathcal{M}_E))) \\
& = \text{Var}_{\{\gamma_t(\mathbf{y}_k)\}} (E_{\mu_1} (E_{\{f_{tk}\}} (E_{\mathbf{X}}(Y_1 + \dots + Y_4 | \{f_{tk}\}) | \{\mathbf{D}_{tk}\}) | \mathcal{M}_E, C(\mathcal{M}_E))) \\
& = \text{Var}_{\{\gamma_t(\mathbf{y}_k)\}} \left(\sum_{t=1}^4 E_{\mu_1} (E_{\{f_{tk}\}} (E_{\mathbf{X}}(Y_t | \{f_{tk}\}) | \{\mathbf{D}_{tk}\}) | \mathcal{M}_E, C(\mathcal{M}_E)) \right) \\
& = \sum_{t=1}^4 \text{Var}_{\{\gamma_t(\mathbf{y}_k)\}} (E_{\mu_1} (E_{\{f_{tk}\}} (E_{\mathbf{X}}(Y_t | \{f_{tk}\}) | \{\mathbf{D}_{tk}\}) | \mathcal{M}_E, C(\mathcal{M}_E))) \\
& + 2 \sum_{t=1}^3 \sum_{t'=t+1}^4 \text{Cov}_{\{\gamma_t(\mathbf{y}_k)\}, \{\gamma_{t'}(\mathbf{y}_{k'})\}} (E_{\mu_1} (E_{\{f_{tk}\}} (E_{\mathbf{X}}(Y_t | \{f_{tk}\}) | \{\mathbf{D}_{tk}\}) | \mathcal{M}_E, C(\mathcal{M}_E)), \\
& \quad E_{\mu_1} (E_{\{f_{t'k'}\}} (E_{\mathbf{X}}(Y_{t'} | \{f_{t'k'}\}) | \{\mathbf{D}_{t'k'}\}) | \mathcal{M}_E, C(\mathcal{M}_E))).
\end{aligned}$$

The covariance terms in the above expression simplify like in equation (17):

$$\begin{aligned}
& \text{Cov}_{\{\gamma_t(\mathbf{y}_k)\}, \{\gamma_{t'}(\mathbf{y}_{k'})\}} \left(E_{\mu_1} \left(E_{\{f_{tk}\}} \left(E_{\mathbf{X}} \left(\sum_{k=1}^{707} a_k \gamma_t(\mathbf{y}_k) f_{tk}(\mathbf{X}_{tk}) \middle| \{f_{tk}\} \right) \middle| \{\mathbf{D}_{tk}\} \right) \middle| \mathcal{M}_E, C(\mathcal{M}_E) \right) \right), \\
& E_{\mu_1} \left(E_{\{f_{t'k'}\}} \left(E_{\mathbf{X}} \left(\sum_{k'=1}^{707} a_{k'} \gamma_{t'}(\mathbf{y}_{k'}) f_{t'k'}(\mathbf{X}_{t'k'}) \middle| \{f_{t'k'}\} \right) \middle| \{\mathbf{D}_{t'k'}\} \right) \middle| \mathcal{M}_E, C(\mathcal{M}_E) \right) \Big) \\
&= \text{Cov}_{\{\gamma_t(\mathbf{y}_k)\}, \{\gamma_{t'}(\mathbf{y}_{k'})\}} \left(\sum_{k=1}^{707} a_k \gamma_t(\mathbf{y}_k) E_{\mu_1} (E_{f_{tk}} (E_{\mathbf{X}_{tk}} (f_{tk}(\mathbf{X}_{tk}) | f_{tk}) | \mathbf{D}_{tk}) | \mathcal{M}_E, C(\mathcal{M}_E)), \right. \\
& \left. \sum_{k'=1}^{707} a_{k'} \gamma_{t'}(\mathbf{y}_{k'}) E_{\mu_1} (E_{f_{t'k'}} (E_{\mathbf{X}_{t'k'}} (f_{t'k'}(\mathbf{X}_{t'k'}) | f_{t'k'}) | \mathbf{D}_{t'k'}) | \mathcal{M}_E, C(\mathcal{M}_E)) \right) \\
&= \sum_{k=1}^{707} \sum_{k'=1}^{707} a_k a_{k'} E_{\mu_1} (E_{f_{tk}} (E_{\mathbf{X}_{tk}} (f_{tk}(\mathbf{X}_{tk}) | f_{tk}) | \mathbf{D}_{tk}) | \mathcal{M}_E, C(\mathcal{M}_E)) \\
& E_{\mu_1} (E_{f_{t'k'}} (E_{\mathbf{X}_{t'k'}} (f_{t'k'}(\mathbf{X}_{t'k'}) | f_{t'k'}) | \mathbf{D}_{t'k'}) | \mathcal{M}_E, C(\mathcal{M}_E)) \text{Cov}_{\gamma_t(\mathbf{y}_k), \gamma_{t'}(\mathbf{y}_{k'})} (\gamma_t(\mathbf{y}_k), \gamma_{t'}(\mathbf{y}_{k'})).
\end{aligned}$$

From equation (18), the variance due to input uncertainty is

$$\begin{aligned}
& \text{Var}_{\mathbf{X}}(Y | \{f_{tk}\}, \{\gamma_t(\mathbf{y}_k)\}) = \text{Var}_{\mathbf{X}}(Y_1 + \dots + Y_4 | \{f_{tk}\}, \{\gamma_t(\mathbf{y}_k)\}) \\
&= \sum_{t=1}^4 V_t + 2 \sum_{t=1}^3 \sum_{t'=-t+1}^4 \text{Cov}_{\mathbf{X}_t, \mathbf{X}_{t'}}(Y_t, Y_{t'} | \{f_{tk}\}, \{f_{t'k}\}, \{\gamma_t(\mathbf{y}_k)\}, \{\gamma_{t'}(\mathbf{y}_k)\}),
\end{aligned}$$

where

$$V_t = \text{Var}_{\mathbf{X}_t}(Y_t | \{f_{tk}\}, \{\gamma_t(\mathbf{y}_k)\}).$$

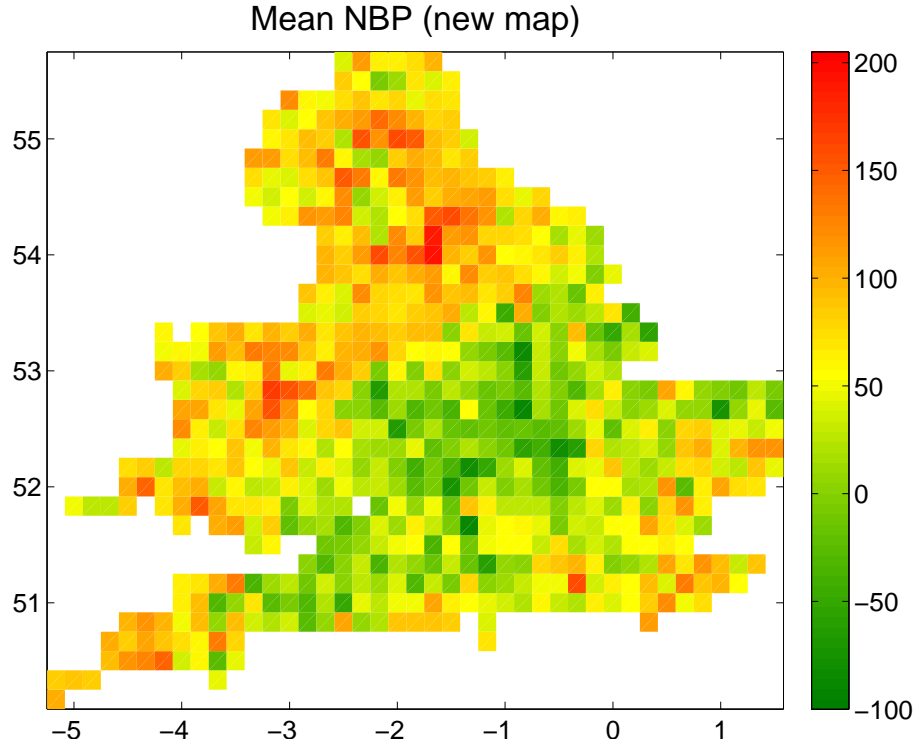
The formula for $E_{\{\gamma_t(\mathbf{y}_k)\}}(E_{\mu_3}(E_{\{f_{tk}\}}(V_t | \{\mathbf{D}_{tk}\}) | \mathcal{V}_E, C(\mathcal{V}_E)))$ is given by equation (14). We must now consider the covariance between Y_t and $Y_{t'}$. From Gosling and O'Hagan (2007) this covariance, assuming that emulators at different sites are independent, is given by the following equation:

$$\begin{aligned}
& \text{Cov}_{\mathbf{X}_t, \mathbf{X}_{t'}}(Y_t, Y_{t'} | \{f_{tk}\}, \{f_{t'k}\}, \{\gamma_t(\mathbf{y}_k)\}, \{\gamma_{t'}(\mathbf{y}_k)\}) = \\
& \sum_{k=1}^{707} a_k^2 \gamma_t(\mathbf{y}_k) \gamma_{t'}(\mathbf{y}_k) \text{Cov}_{\mathbf{X}_{tk}, \mathbf{X}_{t'k}}(f_{tk}(\mathbf{X}_{tk}), f_{t'k}(\mathbf{X}_{t'k}) | f_{tk}, f_{t'k}).
\end{aligned}$$

The expected value of this covariance is thus:

$$E_{\{\gamma_t(\mathbf{y}_k)\}, \{\gamma_{t'}(\mathbf{y}_k)\}} (\text{Cov}_{\mathbf{X}_t, \mathbf{X}_{t'}}(Y_t, Y_{t'} | \{f_{tk}\}, \{f_{t'k}\}, \{\gamma_t(\mathbf{y}_k)\}, \{\gamma_{t'}(\mathbf{y}_k)\})) = \sum_{k=1}^{707} a_k^2 E_{\gamma_t(\mathbf{y}_k), \gamma_{t'}(\mathbf{y}_k)} (\gamma_t(\mathbf{y}_k) \gamma_{t'}(\mathbf{y}_k)) \text{Cov}_{\mathbf{X}_{tk}, \mathbf{X}_{t'k}}(f_{tk}(\mathbf{X}_{tk}), f_{t'k}(\mathbf{X}_{t'k}) | f_{tk}, f_{t'k}).$$

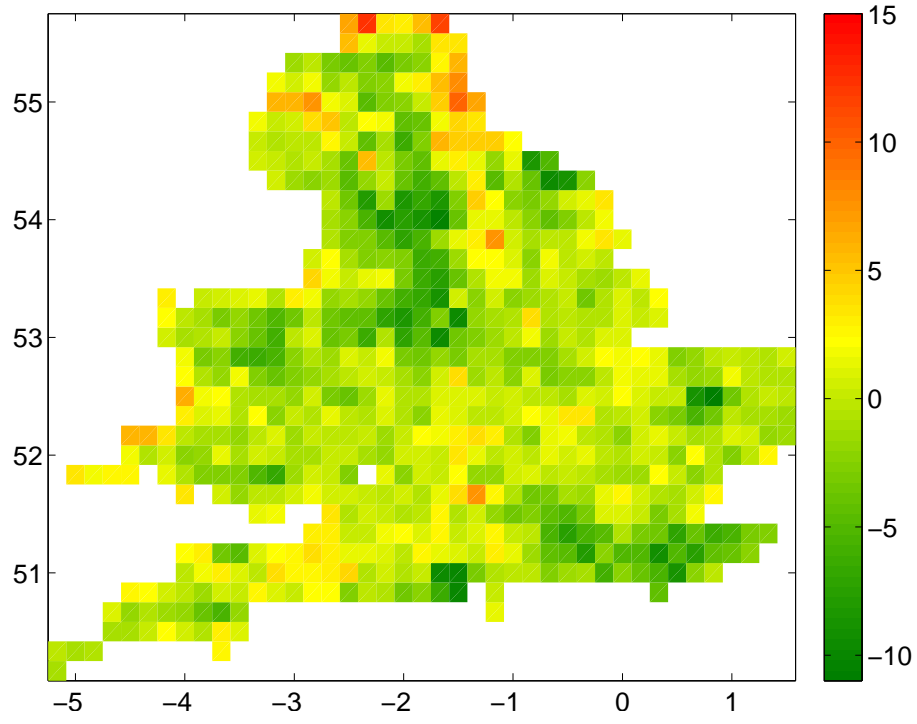
3 Results



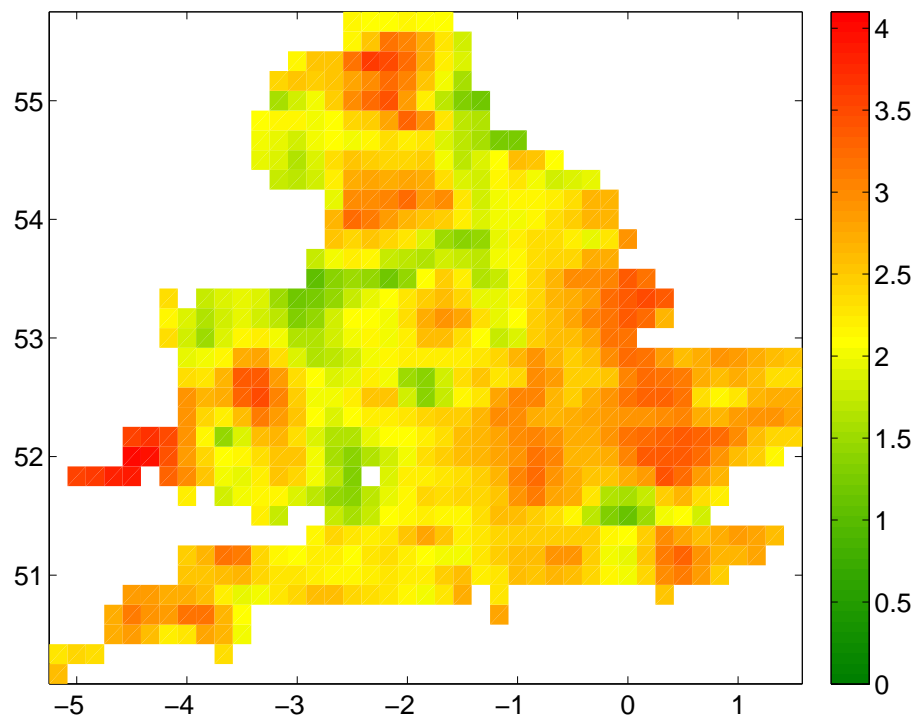
PFT	Mean	Var _{Lcp}	Var _{Em}	Var _{Int}	Var _{Inp}	Var _{Tot}
Grassland	4.3719	0.0076	8.2616×10^{-5}	0.0080	0.2296	0.2453
Crop	0.4260	0.0006	5.6355×10^{-5}	0.0084	0.0236	0.0327
DBL	1.7998	0.0071	4.7970×10^{-5}	0.0056	0.0095	0.0221
ENL	0.8633	0.0043	5.5886×10^{-8}	9.9047×10^{-6}	0.0006	0.0048
Covariances		-0.0091			0.0010	-0.0081
Total	7.4610	0.0105	0.0002	0.0220	0.2642	0.2968

Table 1: Contribution to the mean and variance of total NBP from different plant functional types and covariances between these types.

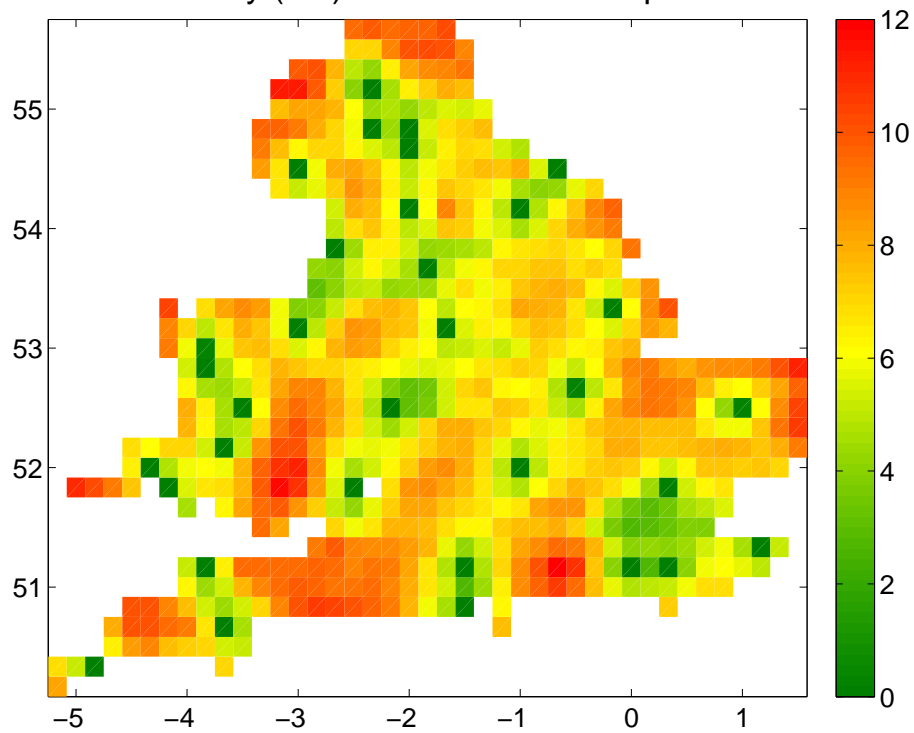
New map – Old map



Uncertainty (SD) in NBP due to emulation



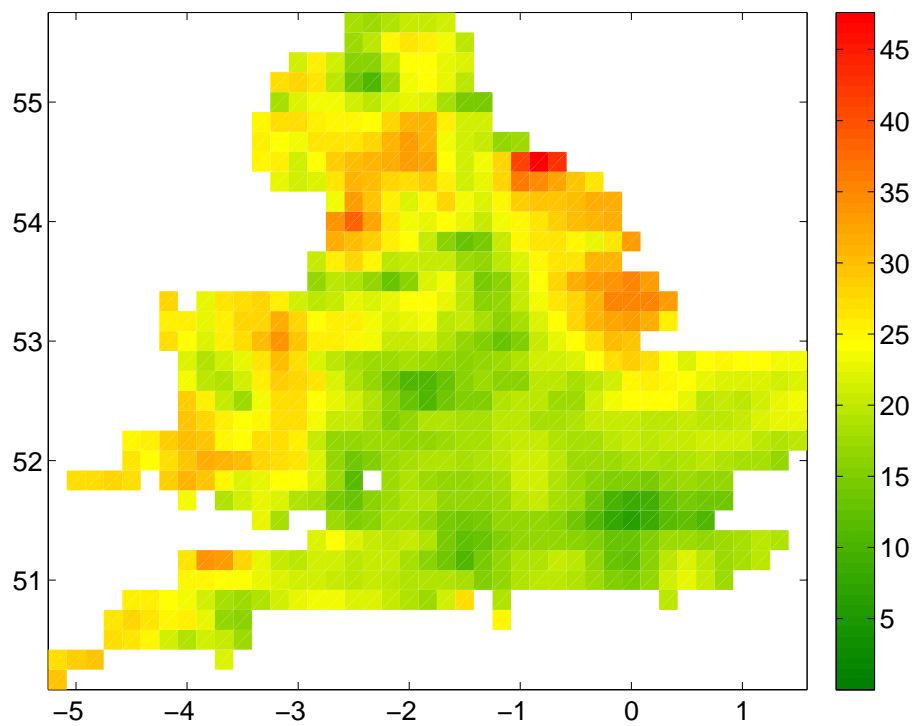
Uncertainty (SD) in NBP due to interpolation



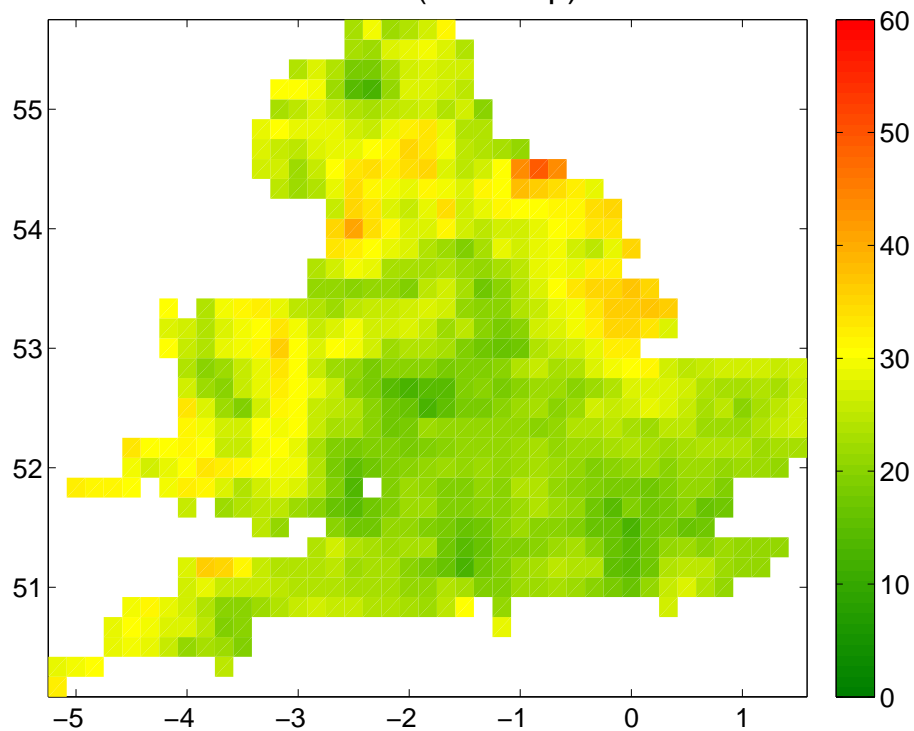
Uncertainty (SD) in NBP due to unknown land cover

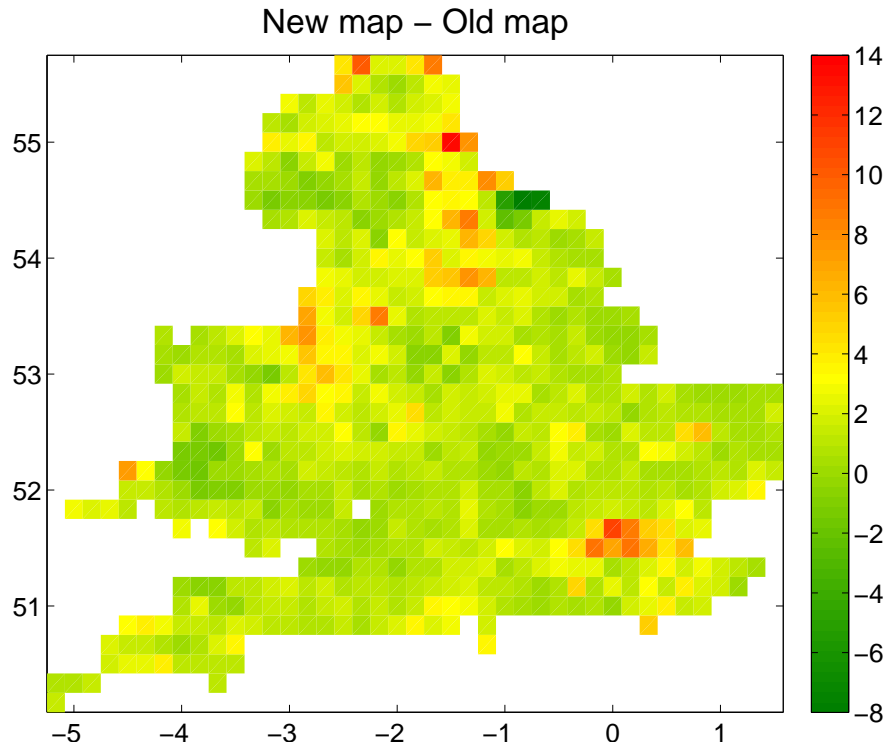


Uncertainty (SD) in NBP due to input uncertainty



SD NBP (new map)





PFT	Mean	Var _{Em}	Var _{Int}	Var _{Inp}	Var _{Tot}
Grassland	4.3719	7.7942×10^{-5}	0.0079	0.2291	0.2371
Crop	0.4260	5.4876×10^{-5}	0.0084	0.0235	0.0320
DBL	1.7998	3.5544×10^{-5}	0.0054	0.0088	0.0142
ENL	0.8633	4.1378×10^{-5}	9.6324×10^{-6}	0.0005	0.0006
Covariances				0.0010	0.0010
Total	7.4610	0.0002	0.0218	0.2630	0.2849

Table 2: The Kennedy *et al.* (2008) uncertainty analysis using the Cripps *et al.* (2008) point estimates of land cover.

References

Cripps, E., O’Hagan, A., Quaife, T. and Anderson, C. W. (2008). Modelling uncertainty in satellite derived land cover maps. Submitted to Applied Statistics.

Fuller, R. M., Smith, G. M., Sanderson, J. M., Hill, R. A., Thomson, A. G., Cox, R., Brown, N. J., Clarke, R. T., Rothery, P. and Gerard, F. F. (2002). *Countryside Survey 2000 Module 7. Land Cover Map 2000. Final*

Report. Tech. rep., Centre for Ecology and Hydrology (Natural Environment Research Council).

Gosling, J. P. and O’Hagan, A. (2007). *Understanding the uncertainty in the biospheric carbon flux for England and Wales*. Tech. Rep. 567/06, Department of Probability and Statistics, University of Sheffield.

Haines-Young, R. H., Barr, C. J., Black, H. I. J., Briggs, D. J., Bunce, R. G. H., Clarke, R. T., Cooper, A., Dawson, F. H., Firbank, L. G., Fuller, R. M., Furse, M. T., Gillespie, M. K., Hill, R., Hornung, M., Howard, D. C., McCann, T., Morecroft, M. D., Petit, S., Sier, A. R. J., Smart, M., Smith, G. M., Stott, A. P., Stuart, R. C. and Watkins, J. W. (2000). *Accounting for nature: assessing habitats in the UK countryside*. Department of the Environment, Transport and the Regions, London.

Kennedy, M., Anderson, C. W., O’Hagan, A., Lomas, M., Woodward, F. I., Gosling, J. P. and Heinemeyer, A. (2008). Quantifying uncertainty in the biospheric carbon flux for England and Wales. *Journal of the Royal Statistical Society, Series A*, **171**, 109–135.

A A step by step guide to physically reproducing the results of this report from scratch

1. Collect training data $\mathbf{D}_{tk} = \{(\mathbf{x}_{tk}, f_{tk}(\mathbf{x}_{tk}))\}$ for the 33 representative sample sites and 4 PFTs by running the Sheffield Dynamic Global Vegetation Model daily version (SDGVMD) at these sites and PFTs for 200 randomly selected inputs \mathbf{x}_{tk} . The training data used in this report is given in the folder SDGVM data on the CD called Keith’s CTCD Work.

2. Use the training data collected in step 1 to build a statistical emulator for each function f_{tk} by running the GEM-SA software (see <http://www.tonyohagan.co.uk/academic/GEM/index.html>) for each data set \mathbf{D}_{tk} . Firstly, it calculates the mean of expected code output, which in our notation is $E_{f_{tk}}(E_{\mathbf{X}_{tk}}(f_{tk}(\mathbf{X}_{tk})|f_{tk})|\mathbf{D}_{tk})$. Secondly, it calculates the variance of expected code output, which in our notation is $\text{Var}_{f_{tk}}(E_{\mathbf{X}_{tk}}(f_{tk}(\mathbf{X}_{tk})|f_{tk})|\mathbf{D}_{tk})$. Finally, it calculates the meal of total variance in code output, which in our notation is $E_{f_{tk}}(\text{Var}_{\mathbf{X}_{tk}}(f_{tk}(\mathbf{X}_{tk})|f_{tk})|\mathbf{D}_{tk})$. Note that the GEM-SA source code would have to be altered in order to build emulators for training data

from the PFT evergreen needleleaf, as one of the input parameters for this PFT required special treatment and did not have a Normal prior distribution.

3. As we only have training data for 33 of the 707 sample sites in England and Wales, we need to interpolate the emulation results obtained in step 2. For each PFT t we obtained the following three sets of emulation results:

$$\mathcal{M}_E = \{E_{f_{tk}}(E_{\mathbf{X}_{tk}}(f_{tk}(\mathbf{X}_{tk})|f_{tk})|\mathbf{D}_{tk}) : k \in K\},$$

$$\mathcal{M}_V = \{\text{Var}_{f_{tk}}(E_{\mathbf{X}_{tk}}(f_{tk}(\mathbf{X}_{tk})|f_{tk})|\mathbf{D}_{tk}) : k \in K\},$$

and

$$\mathcal{V}_E = \{E_{f_{tk}}(\text{Var}_{\mathbf{X}_{tk}}(f_{tk}(\mathbf{X}_{tk})|f_{tk})|\mathbf{D}_{tk}) : k \in K\},$$

where K is the set of sites k for which we have training data \mathbf{D}_{tk} . Like Gosling and O'Hagan (2007), we use ordinary kriging techniques to interpolate these sets of results. In order to krige each set of results, it is necessary to specify a correlation structure between the unknown functions. We find a suitable correlation structure for each interpolation by fitting several different covariance models to the emulation results using the standard variogram fitting techniques of the S-PLUS package `gstat`. In our report the chosen correlation structures for \mathcal{M}_E , \mathcal{M}_V and \mathcal{V}_E were denoted by $C(\mathcal{M}_E)$, $C(\mathcal{M}_V)$ and $C(\mathcal{V}_E)$, respectively.

For $k \notin K$, we interpolate \mathcal{M}_E using the kriging function μ_1 to find values for $E_{f_{tk}}(E_{\mathbf{X}_{tk}}(f_{tk}(\mathbf{X}_{tk})|f_{tk})|\mathbf{D}_{tk})$. When kriging a set of results, we allow for uncertainty in the interpolation; hence, we can calculate a mean result for each unknown function and the uncertainty around that estimate. We interpolate \mathcal{M}_V using μ_2 and \mathcal{V}_E using μ_3 . Doing this for each PFT t by using the `gstat` command `krige` we obtain the following predictions and prediction variance:

$$E_{\mu_1}(E_{f_{tk}}(E_{\mathbf{X}_{tk}}(f_{tk}(\mathbf{X}_{tk})|f_{tk})|\mathbf{D}_{tk})|\mathcal{M}_E, C(\mathcal{M}_E)),$$

$$\text{Var}_{\mu_1}(E_{f_{tk}}(E_{\mathbf{X}_{tk}}(f_{tk}(\mathbf{X}_{tk})|f_{tk})|\mathbf{D}_{tk})|\mathcal{M}_E, C(\mathcal{M}_E)),$$

$$E_{\mu_2}(\text{Var}_{f_{tk}}(E_{\mathbf{X}_{tk}}(f_{tk}(\mathbf{X}_{tk})|f_{tk})|\mathbf{D}_{tk})|\mathcal{M}_V, C(\mathcal{M}_V)),$$

and

$$E_{\mu_3}(E_{f_{tk}}(\text{Var}_{\mathbf{X}_{tk}}(f_{tk}(\mathbf{X}_{tk})|f_{tk})|\mathbf{D}_{tk})|\mathcal{V}_E, C(\mathcal{V}_E)) = C_{kk}^{(t)}.$$

4. Use the results obtained in step 2 to calculate the following six inter-PFT expected covariances for the 33 sample sites we collected training data for in step 1: $E_{f_{tk}, f_{t'k}}(\text{Cov}_{\mathbf{X}_{tk}, \mathbf{X}_{t'k}}(f_{tk}(\mathbf{X}_{tk}), f_{t'k}(\mathbf{X}_{t'k}) | f_{tk}, f_{t'k}) | \mathbf{D}_{tk}, \mathbf{D}_{t'k})$. Gosling and O'Hagan (2007) showed that these expected covariances could be calculated using the following approximation:

$$\begin{aligned} E_{f_{tk}, f_{t'k}}(\text{Cov}_{\mathbf{X}_{tk}, \mathbf{X}_{t'k}}(f_{tk}(\mathbf{X}_{tk}), f_{t'k}(\mathbf{X}_{t'k}) | f_{tk}, f_{t'k}) | \mathbf{D}_{tk}, \mathbf{D}_{t'k}) \approx \\ \{ E_{f_{tk}+f_{t'k}}[\text{Var}_{\mathbf{X}_{tk}, \mathbf{X}_{t'k}}(f_{tk}(\mathbf{X}_{tk}) + f_{t'k}(\mathbf{X}_{t'k}) | f_{tk}, f_{t'k}) | \mathbf{D}_{tk}, \mathbf{D}_{t'k}] - \\ E_{f_{tk}}[\text{Var}_{\mathbf{X}_{tk}}(f_{tk}(\mathbf{X}_{tk}) | f_{tk}) | \mathbf{D}_{tk}] - E_{f_{t'k}}[\text{Var}_{\mathbf{X}_{t'k}}(f_{t'k}(\mathbf{X}_{t'k}) | f_{t'k}) | \mathbf{D}_{t'k}] \} / 2. \end{aligned} \quad (19)$$

Before we can use this approximation we need to calculate

$$E_{f_{tk}+f_{t'k}}[\text{Var}_{\mathbf{X}_{tk}, \mathbf{X}_{t'k}}(f_{tk}(\mathbf{X}_{tk}) + f_{t'k}(\mathbf{X}_{t'k}) | f_{tk}, f_{t'k}) | \mathbf{D}_{tk}, \mathbf{D}_{t'k}].$$

This can be done by building an emulator for each sum $f_{tk} + f_{t'k}$ using GEM-SA.

5. Estimate the six inter-PFT expected covariances for the other 674 sample sites in England and Wales by interpolation. For each pair of PFTs (t, t') we obtained the following set of expected covariances in step 4:

$$\mathcal{C}_{E_{tt'}} = \{ E_{f_{tk}, f_{t'k}}(\text{Cov}_{\mathbf{X}_{tk}, \mathbf{X}_{t'k}}(f_{tk}(\mathbf{X}_{tk}), f_{t'k}(\mathbf{X}_{t'k}) | f_{tk}, f_{t'k}) | \mathbf{D}_{tk}, \mathbf{D}_{t'k}) : k \in K \}.$$

We used the same approach as we did in step 3 to interpolate these six sets of expected covariances. In our report the chosen correlation structure for $\mathcal{C}_{E_{tt'}}$ was denoted by $C(\mathcal{C}_{E_{tt'}})$. For $k \notin K$, we interpolate $\mathcal{C}_{E_{tt'}}$ using the kriging function $\mu_{tt'}$ and thus obtain the following prediction for the expected covariance:

$$E_{\mu_{tt'}}(E_{f_{tk}, f_{t'k}}(\text{Cov}_{\mathbf{X}_{tk}, \mathbf{X}_{t'k}}(f_{tk}(\mathbf{X}_{tk}), f_{t'k}(\mathbf{X}_{t'k}) | f_{tk}, f_{t'k}) | \mathbf{D}_{tk}, \mathbf{D}_{t'k}) | \mathcal{C}_{E_{tt'}}, C(\mathcal{C}_{E_{tt'}})).$$

6. In step 3 it wasn't possible to calculate the following kriging covariances using the S-PLUS package `gstat`:

$$\text{Cov}_{\mu_1}(E_{f_{tk}}(E_{\mathbf{X}_{tk}}(f_{tk}(\mathbf{X}_{tk}) | f_{tk}) | \mathbf{D}_{tk}), E_{f_{t'k'}}(E_{\mathbf{X}_{t'k'}}(f_{t'k'}(\mathbf{X}_{t'k'}) | f_{t'k'}) | \mathbf{D}_{t'k'}) | \mathcal{M}_E, C(\mathcal{M}_E)).$$

Instead we calculated these kriging covariances using a program written in

C by John Paul Gosling to specifically do this task. This program is called `Krige.exe` and is given in the folder JP's Kriging Program on the CD of my work.

7. Use the results obtained in step 2 to calculate the following 528 inter-site expected covariances for the 33 sample sites we collected training data for in step 1: $E_{f_{tk}, f_{tk'}}(\text{Cov}_{\mathbf{X}_{tk}, \mathbf{X}_{tk'}}(f_{tk}(\mathbf{X}_{tk}), f_{tk'}(\mathbf{X}_{tk'}) | f_{tk}, f_{tk'}) | \mathbf{D}_{tk}, \mathbf{D}_{tk'})$. Analogous to equation (19) in step 4, we have that

$$\begin{aligned} E_{f_{tk}, f_{tk'}}(\text{Cov}_{\mathbf{X}_{tk}, \mathbf{X}_{tk'}}(f_{tk}(\mathbf{X}_{tk}), f_{tk'}(\mathbf{X}_{tk'}) | f_{tk}, f_{tk'}) | \mathbf{D}_{tk}, \mathbf{D}_{tk'}) \approx \\ \{ E_{f_{tk} + f_{tk'}}[\text{Var}_{\mathbf{X}_{tk}, \mathbf{X}_{tk'}}(f_{tk}(\mathbf{X}_{tk}) + f_{tk'}(\mathbf{X}_{tk'}) | f_{tk}, f_{tk'}) | \mathbf{D}_{tk}, \mathbf{D}_{tk'}] - \\ E_{f_{tk}}[\text{Var}_{\mathbf{X}_{tk}}(f_{tk}(\mathbf{X}_{tk}) | f_{tk}) | \mathbf{D}_{tk}] - E_{f_{tk'}}[\text{Var}_{\mathbf{X}_{tk'}}(f_{tk'}(\mathbf{X}_{tk'}) | f_{tk'}) | \mathbf{D}_{tk'}] \} / 2, \end{aligned}$$

where $E_{f_{tk} + f_{tk'}}[\text{Var}_{\mathbf{X}_{tk}, \mathbf{X}_{tk'}}(f_{tk}(\mathbf{X}_{tk}) + f_{tk'}(\mathbf{X}_{tk'}) | f_{tk}, f_{tk'}) | \mathbf{D}_{tk}, \mathbf{D}_{tk'}]$ is calculated by building an emulator for the sum $f_{tk} + f_{tk'}$ using GEM-SA.

8. Use the results obtained in steps 3 and 7 to estimate the following inter-site expected covariances for the other 674 sample sites in England and Wales:

$$C_{kk'}^{(t)} = E_{\mu_3}(E_{f_{tk}, f_{tk'}}(\text{Cov}_{\mathbf{X}_{tk}, \mathbf{X}_{tk'}}(f_{tk}(\mathbf{X}_{tk}), f_{tk'}(\mathbf{X}_{tk'}) | f_{tk}, f_{tk'}) | \mathbf{D}_{tk}, \mathbf{D}_{tk'}) | \mathcal{V}_E, C(\mathcal{V}_E)).$$

This can be done as follows. Let $\mathbf{C}^{(t)}$ be the matrix with elements $C_{kk'}^{(t)}$ and $\mathbf{C}_K^{(t)}$ be the submatrix corresponding to $k, k' \in K$. We need values for the off-diagonal elements $C_{kk'}^{(t)}$ of $\mathbf{C}^{(t)}$ outside $\mathbf{C}_K^{(t)}$. Like Kennedy *et al.* (2008), we approximate these terms using $C_{kk'}^{(t)} = \rho_t \sqrt{C_{kk}^{(t)} C_{k'k'}^{(t)}}$, where the value of ρ_t is chosen such that if the same approximation was used for the off-diagonal elements of $\mathbf{C}_K^{(t)}$ then the sum of all these elements would be unchanged. This process yielded the following estimated PFT correlations: $\rho_1 = 0.051$, $\rho_2 = 0.012$, $\rho_3 = 0.006$ and $\rho_4 = 0.182$ for grassland, crop, DBL and ENL, respectively.

9. Use the data from the land cover map of Haines-Young *et al.* (2000) and the confusion matrix of Fuller *et al.* (2002) to calculate $E_{\gamma_t(\mathbf{y}_k)}(\gamma_t(\mathbf{y}_k))$, $\text{Var}_{\gamma_t(\mathbf{y}_k)}(\gamma_t(\mathbf{y}_k))$, $\text{Cov}_{\gamma_t(\mathbf{y}_k), \gamma_{t'}(\mathbf{y}_k)}(\gamma_t(\mathbf{y}_k), \gamma_{t'}(\mathbf{y}_k))$, $\text{Cov}_{\gamma_t(\mathbf{y}_k), \gamma_t(\mathbf{y}_{k'})}(\gamma_t(\mathbf{y}_k), \gamma_t(\mathbf{y}_{k'}))$ and $\text{Cov}_{\gamma_t(\mathbf{y}_k), \gamma_{t'}(\mathbf{y}_{k'})}(\gamma_t(\mathbf{y}_k), \gamma_{t'}(\mathbf{y}_{k'}))$, where $\gamma_t(\mathbf{y}_k)$ is the proportion of PFT t at site k . This can be done using the approach of Cripps *et al.* (2008). A MATLAB program for calculating these quantities called `sim_spat_prop_main2.m` is given in the folder MatLab Code on the CD of my work.

10. Use the results obtained in steps 3 and 9 to calculate the expectation of Y_k for all 707 sample sites in England and Wales, plot a map of these site expectations to produce the first graph of Section 3, plot the old map of site expectations by Kennedy *et al.* (2008) for comparison and plot the difference between these two maps to produce the second graph of Section 3. A MATLAB program for doing this called `site_nbp2.m` is given in the folder MatLab Code on the CD of my work.

11. Use the results obtained in steps 3 and 9 to calculate the uncertainty in Y_k due to emulation for all 707 sample sites in England and Wales, and plot a map of these uncertainties to produce the third graph of Section 3. A MATLAB program for doing this called `site_nbp_em_var2.m` is given in the folder MatLab Code on the CD of my work.

12. Use the results obtained in steps 3 and 9 to calculate the uncertainty in Y_k due to interpolation for all 707 sample sites in England and Wales, and plot a map of these uncertainties to produce the fourth graph of Section 3. A MATLAB program for doing this called `site_nbp_int_var2.m` is given in the folder MatLab Code on the CD of my work.

13. Use the results obtained in steps 3 and 9 to calculate the uncertainty in Y_k due to uncertainty in the land cover proportions $\gamma_t(\mathbf{y}_k)$ for all 707 sample sites in England and Wales, and plot a map of these uncertainties to produce the fifth graph of Section 3. A MATLAB program for doing this called `site_nbp_lcp_var2.m` is given in the folder MatLab Code on the CD of my work.

14. Use the results obtained in steps 3, 5 and 9 to calculate the uncertainty in Y_k due to input uncertainty for all 707 sample sites in England and Wales, and plot a map of these uncertainties to produce the sixth graph of Section 3. A MATLAB program for doing this called `site_nbp_input_var2.m` is given in the folder MatLab Code on the CD of my work.

15. Use the results obtained in steps 11 to 14 to calculate the total uncertainty in Y_k for all 707 sample sites in England and Wales, plot a map of these uncertainties to produce the seventh graph of Section 3, plot the old map of site standard deviations by Kennedy *et al.* (2008) for comparison, and plot the difference between these two maps to produce the eighth and final graph of Section 3. A MATLAB program for doing this called `site_nbp_var2.m` is given in the folder MatLab Code on the CD of my work.

16. Use the results obtained in steps 3 and 9 to calculate the expectation of Y_t for all 4 PFTs and then add these 4 expectations together to obtain the expectation of the grand total Y . This produces the values in the second column of Table 1. A MATLAB program for doing this called `pft_nbp.m` is given in the folder MatLab Code on the CD of my work.

17. Use the results obtained in steps 3, 6 and 9 to calculate the uncertainty in Y_t due to interpolation for all 4 PFTs. A MATLAB program for doing this called `pft_nbp_int_var.m` is given in the folder MatLab Code on the CD of my work.

18. Use the results obtained in steps 3 and 9 to calculate the uncertainty in Y_t due to emulation for all 4 PFTs. A MATLAB program for doing this called `pft_nbp_em_var.m` is given in the folder MatLab Code on the CD of my work.

19. For each PFT, add the variance of Y_t due to interpolation obtained in step 17 with the variance of Y_t due to emulation obtained in step 18, to obtain the variance of Y_t due to emulation and interpolation for all 4 PFTs. Then add these 4 variances together to obtain the variance of the grand total Y due to emulation and interpolation. This produces the values in the fourth column of Table 1.

20. Use the results obtained in steps 3, 8 and 9 to calculate the uncertainty in Y_t due to input uncertainty for all 4 PFTs. This produces the first 4 values of the fifth column of Table 1. A MATLAB program for doing this called `pft_nbp_input_var.m` is given in the folder MatLab Code on the CD of my work.

21. Use the results obtained in steps 3, 6 and 9 to calculate the uncertainty in Y_t due to uncertainty in the land cover proportions $\gamma_t(\mathbf{y}_k)$ for all 4 PFTs. This produces the first 4 values of the third column of Table 1. A MATLAB program for doing this called `pft_nbp_lcp_var.m` is given in the folder MatLab Code on the CD of my work.

22. Use the results obtained in steps 3, 5, 9, 17, 18, 20 and 21 to calculate the expected covariances between Y_t and $Y_{t'}$ due to input uncertainty, the variance of the grand total Y due to input uncertainty, the covariances between the expectation of Y_t and the expectation of $Y_{t'}$ due to uncertainty in the land cover proportions $\gamma_t(\mathbf{y}_k)$, the variance of the grand total Y due to uncertainty in the land cover proportions $\gamma_t(\mathbf{y}_k)$ and the overall variance of the grand total Y . This produces the final 2 values of the fifth

column, the final 2 values of the third column and the final value of the sixth column of Table 1, respectively. A MATLAB program for doing this called `total_nbp_var.m` is given in the folder MatLab Code on the CD of my work.

23. Finally, complete Table 1 by adding together its third, fourth and fifth columns to obtain the missing values of the sixth column.