Meeting Notes:

Current issues:

1. Annual fluctuation
   - All bands and, by extension, indices contain a component which is periodic over an annual cycle.
   - This annual fluctuation is the major component of the signal—dominating the correlations.
2. Lag
   - It may be that satellite bands/indices lag behind the JULES signal (or vice versa).
   - This *may* make things appear less correlated than they are in reality.

Solutions

1. Annual Fluctuations

The time-series for a particular band/index output b, in a particular greenbelt g, is given by some unknown function $f_{b,g}$:

$$\mathbf{y}_{b,g} = f_{b,g}(\mathbf{t})$$

where

$$\begin{aligned} \mathbf{t} &= \{t_y\}_{y=2016,...,2025} \\ \mathbf{t}_y &= \{t_m\}_{m=1,...,12} \end{aligned}$$

and for a single year,

$$\mathbf{y}_{b,g,y} = f_{b,g}(\mathbf{t}_{\mathbf{y}})$$

We will compute the average time-series for a particular band in a particular greenbelt:

$$\bar{\mathbf{y}}_{b,g} = \frac{1}{(2025-2016)} \sum_{y=2016}^{2025} \mathbf{y}_{b,g,y}$$

Then calculate the deviation from this mean $\hat{\mathbf{y}}_{b,g,y}$ as:

$$\hat{\mathbf{y}}_{b,g,y} = \mathbf{y}_{b,g,y} - \bar{\mathbf{y}}_{b,g}$$

Similarly, JULES predictions are given by known functions:

$$\begin{aligned} \mathbf{y}_{\text{soil}} &= h_{s,g}(\mathbf{t}) \\ \mathbf{y}_{\text{veg}} &= h_{v,g}(\mathbf{t}) \end{aligned}$$

We can compute their means:

$$\bar{\mathbf{y}}_{\text{soil}} = \frac{1}{(2025 - 2016)} \sum_{y=2016}^{2025} h_{s,g}(\mathbf{t}_\mathbf{y})$$
$$\bar{\mathbf{y}}_{\text{veg}} = \frac{1}{(2025 - 2016)} \sum_{y=2016}^{2025} h_{v,g}(\mathbf{t}_\mathbf{y})$$

And their deviations:

$$\hat{\mathbf{y}}_{\text{soil}} = \mathbf{y}_{\text{soil}} - \bar{\mathbf{y}}_{\text{soil}}$$
$$\hat{\mathbf{y}}_{\text{veg}} = \mathbf{y}_{\text{veg}} - \bar{\mathbf{y}}_{\text{veg}}$$

We then compute the Pearson correlations of $\hat{\mathbf{y}}_{b,g}$ with $\hat{\mathbf{y}}_{\text{soil}}$ and $\hat{\mathbf{y}}_{\text{veg}}$, respectively:

$$c_{b,g}^{\text{soil}} = \text{Corr}(\hat{\mathbf{y}}_{b,g}, \hat{\mathbf{y}}_{\text{soil}})$$
$$c_{b,g}^{\text{veg}} = \text{Corr}(\hat{\mathbf{y}}_{b,g}, \hat{\mathbf{y}}_{\text{veg}})$$

where Corr is the Pearson Correlation Coefficient, given by:

$$\text{Corr}(\mathbf{x}, \mathbf{y}) = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2}\sqrt{\sum(y_i - \bar{y})^2}}$$

Output Matrices:

We can then compute and visualize the following matrices:

$$\begin{array}{c} M_{\text{soil}} \\ \begin{bmatrix} c_{b_1 g_1}^{\text{soil}} & c_{b_1 g_2}^{\text{soil}} & \cdots & c_{b_1 g_m}^{\text{soil}} \\ c_{b_2 g_1}^{\text{soil}} & c_{b_2 g_2}^{\text{soil}} & \cdots & c_{b_2 g_m}^{\text{soil}} \\ \vdots & \vdots & \ddots & \vdots \\ c_{b_n g_1}^{\text{soil}} & c_{b_n g_2}^{\text{soil}} & \cdots & c_{b_n g_m}^{\text{soil}} \end{bmatrix} \end{array} \qquad \begin{array}{c} M_{\text{veg}} \\ \begin{bmatrix} c_{b_1 g_1}^{\text{veg}} & c_{b_1 g_2}^{\text{veg}} & \cdots & c_{b_1 g_m}^{\text{veg}} \\ c_{b_2 g_1}^{\text{veg}} & c_{b_2 g_2}^{\text{veg}} & \cdots & c_{b_2 g_m}^{\text{veg}} \\ \vdots & \vdots & \ddots & \vdots \\ c_{b_n g_1}^{\text{veg}} & c_{b_n g_2}^{\text{veg}} & \cdots & c_{b_n g_m}^{\text{veg}} \end{bmatrix} \end{array}$$

where n is the number of bands/indices and m is the number of greenbelts.

---

2. Lag

For each band/index, we compute the correlation at different offsets (δt) and compare the results:

Algorithmically:

$$
\begin{aligned}
&\mathbf{c}^{\text{soil}} \leftarrow \text{initialized as empty list/vector}\\
&\mathbf{c}^{\text{veg}} \leftarrow \text{initialized as empty list/vector}\\
\\
&\textbf{for } \delta t \in \{-l, \ldots, l\} \textbf{ do}\\
&\quad \mathbf{y'}_{b,g} = f_{b,g}(\mathbf{t} + \delta t)\\
&\quad \bar{\mathbf{y}}'_{b,g} = \tfrac{1}{(2025-2016)} \textstyle\sum_{y=2016}^{2025} \mathbf{y'}_{b,g,y}\\
&\quad \hat{\mathbf{y}}'_{b,g,y} = \mathbf{y'}_{b,g,y} - \bar{\mathbf{y}}'_{b,g}\\
&\quad c^{\text{soil}}_{b,g,\delta t} = \mathrm{Corr}(\hat{\mathbf{y}}'_{b,g}, \hat{\mathbf{y}}_{\text{soil}})\\
&\quad c^{\text{veg}}_{b,g,\delta t} = \mathrm{Corr}(\hat{\mathbf{y}}'_{b,g}, \hat{\mathbf{y}}_{\text{veg}})\\
&\quad \mathbf{c}^{\text{soil}} \leftarrow \mathbf{c}^{\text{soil}} \cup \{c^{\text{soil}}_{b,g,\delta t}\}\\
&\quad \mathbf{c}^{\text{veg}} \leftarrow \mathbf{c}^{\text{veg}} \cup \{c^{\text{veg}}_{b,g,\delta t}\}\\
&\textbf{end for}
\end{aligned}
$$

and then analyse $\mathbf{c}^{\text{soil}}$ and $\mathbf{c}^{\text{veg}}$ for an optimal δt and check the consistency across years/greenbelts.