Data-driven models for vegetation time series analysis

James Hitchcock & Juan C. Suárez (Forest Research)

A common application of satellite remote sensing is **anomaly detection**. In the context of **vegetation change analysis**, this is complicated by complex signals in the data associated with 1) *plant phenology* and 2) *systematic noise* associated with instrumental and atmospheric effects.

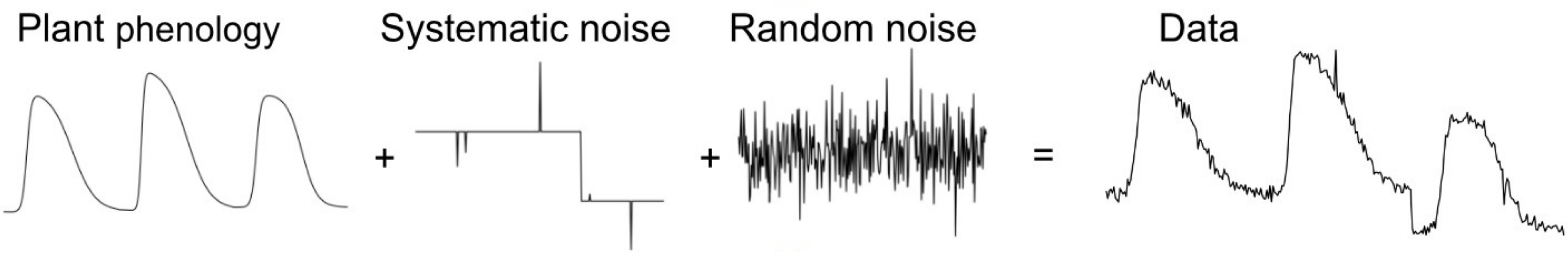
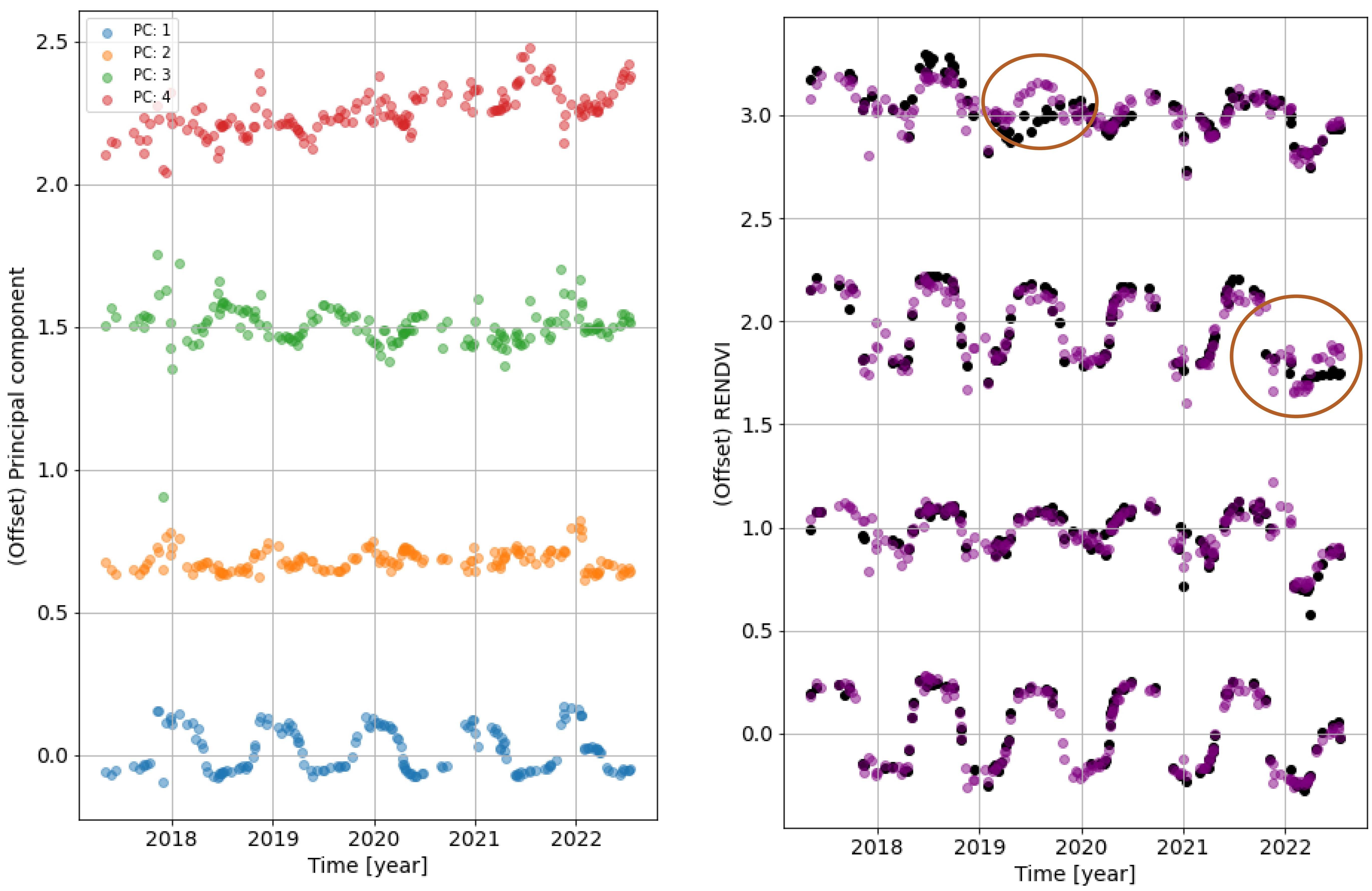


Figure 1—A generative model for optical satellite remote sensing data of vegetation

Because of these complexities, the typical approach to anomaly detection and/or change analysis is to generate **seasonal composites** by averaging over many images to remove systematic artefacts, and then comparing composites in the same season of different years to heuristically adjust for plant phenology. Alternatively, various **data smoothing** procedures are applied to the noisy observations to remove undesirable artefacts prior to any further analysis.

For many anomaly detection applications, *we do not care about the underlying phenological or systematic noise signals that generate the observed data*. We propose a simple **data-driven** (i.e. unsupervised) approach to modelling these signals via **Weighted Principal Component Analysis** (WPCA). A subset of the components with the highest explanatory power likely captures phenological and systematic signals *common to all the data within an area-of-interest,* to greater or lesser extents. Having captured these signals within these WPCA components, predictions can be made for the observations and subtracted from the data, and reveal anomalies within the **residuals**. The WPCA components are learned from very many pixel time series observations associated with vegetation within a particular area-of-interest. Because of localised artefacts associated with e.g. cloud and shadow masking in parts of an image, there will be inconsistent gaps in the pixel time series. By using a weighted PCA approach, we can neatly ignore missing observations by giving them zero weight.



Anomalies are flagged for pixel time series where the observations deviate from the WPCA model predictions. These can highlight felling, or diseases affecting the canopy of trees etc. Conversely, observations for healthy vegetation should show good agreement between the model and data.

Figure 2—(*Left*) The top 4 WPCA components learned from many 1000s of Sentinel-2 10x10 metre Red edge Normalised Difference Vegetation Index (RENDVI) pixel time series of a UK forest. One can see the phenological signal common to all pixel time series has been captured in PC1, while the other highest components have likely captured some combination of phenological and systematic signals. (*Right*) A small sample of the RENDVI pixel time series (black) and their WPCA model predictions (purple). Deviations between the model and data are indicative of change in the observed vegetation.