20200501 Presentation

Jon Page

May 2020

Contents

1 Comments on Previous Work

Following chat with Martin, below suggestions were made:

- Remove gridlines from plots - DONE
- Look into certain sites to see if why there is no correlation with precipitation and GPP - DONE (See section below on Fluxnet sites)
- Order plots by MAP or coefficient of variation of precipitation - DONE
- Add temperature to the model
- Compare GPP/Precip for the sites from Thanos' paper, both observations and then the model outputs

1.1 Fluxnet Sites

The majority of the Fluxnet sites that show no correlation between yearly GPP and precipitation are very clearly temperature limited. Therefore introducing temperature into the model will hopefully show this. My hypothesis is "The influence of antecedent temperature on productivity is not significant at yearly timescales".

- AT_Neu - "Highly variable summer", alpine, 970 mASL
- CH_Fru - 1000 mASL, possible fertilisation (it is a forage site), VERY wet
- DK_ZAH - In Greenland, very much Arctic, average July high is 7 °C, year average is -6.6 °C
- IT_MBo - 1550 mASL, year average is 5.1 °C
- DE_Gri - extensively managed (but not fertilised)
- US_Var - grows in wet season (Oct-May). Mediterranean climate.

2 Reading and Thoughts

The below thoughts are mostly inspired from reading Ryan's 2015 paper on antecedent temperature and moisture:

Plotting monthly weights as error bars (with 95% confidence interval) is probably a bit more informative. Although 2.5% quantiles appear to never equal 0, maybe there is some way to say whether it's different enough to be "significant"?

Checking the confidence interval for the alphas is also SUPER informative - is the antecedent precipitation even significant? Or could the covariate be 0/negative/positive?

Is it worthwhile/informative/not-completely-illogical to change the model into

$$\mu_{GPP}[i] = \alpha_1 + \alpha_2 * PPT[i] + \alpha_3 * PPT_{ant}[i]$$

with the current year's precipitation also included separately (and then probably removed from the antecedent term so it isn't used twice)? Would this allow us to look at α_k and say "Current precipitation is significant but antecedent isn't" in a more informed manner?

Could we look at different time scales to allow us to use more parameters in our model without overfitting? For instance, if we looked at GPP as a monthly sum, our time series would be twelve times as long. HOWEVER, due to a "growing season", this may cause issues as precipitation in e.g. December is less likely to affect productivity but for GPP in Jan, this is lag 1, but lag 2 for Feb, etc. which I believe would "confuse" the model i.e. it would prevent the model from saying "Rainfall received in the previous 3 months is important" because, while this might be true for growing seasons, it wouldn't be for outside the growing seasons. Is there some way to account for this?

It would be interesting to check the following scenario: A simple linear regression on P shows that ANPP is highly correlated with current P but not with P for y-1. Would SAM with a lag of 1 have a significant α for PPT_{ant} and then use monthly weights to show the dependence on current year? Or would it struggle to show this?

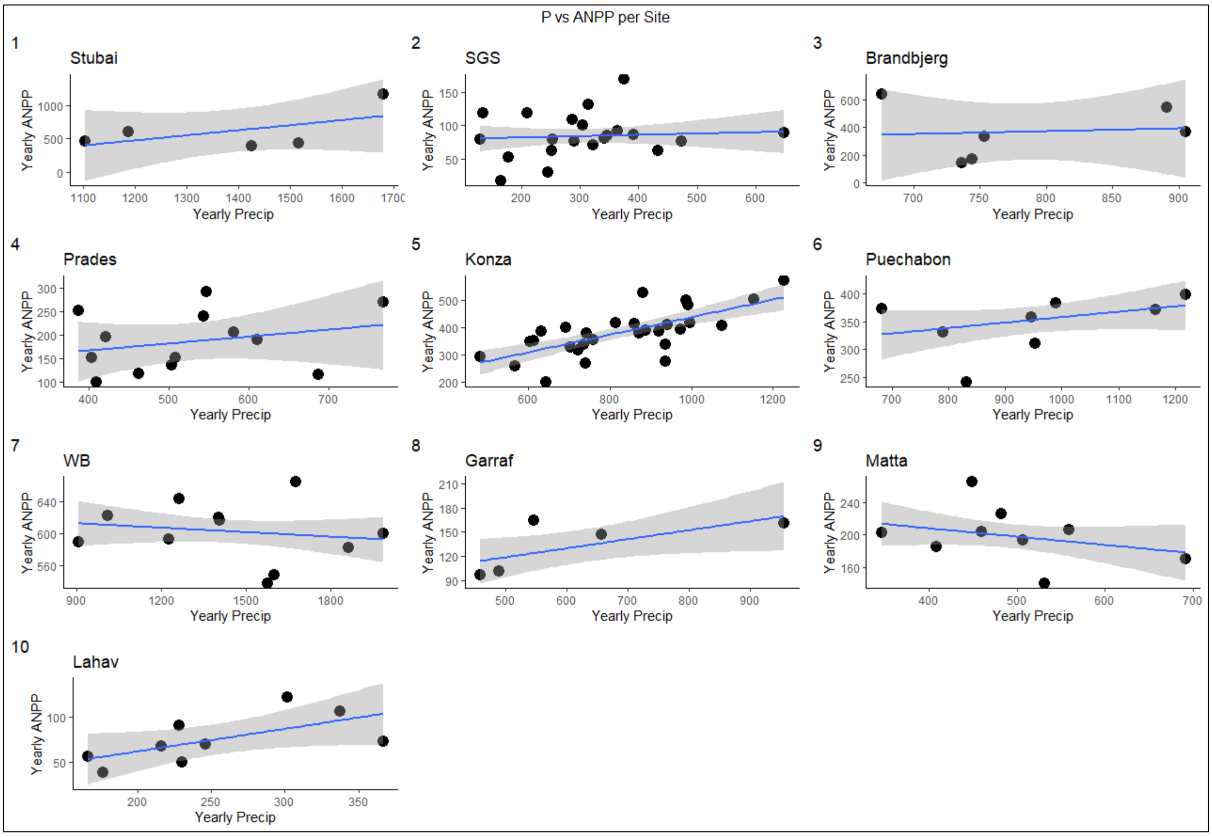

Figure 1: Precipitation against ANPP for each site on a yearly basis. Blue line is linear regression fitted by geom_smooth from ggplot2 package. Shaded area is 95% confidence interval.

3 New Work - Model Comparison

3.1 Observations

All work has been produced under the folder "GPPCorrelationModelCheck".

The first step was to extract the observations for each site and check whether these showed any possible correlations between P, T and ANPP.

It took a while due to each model output having a different format, but I have produced the below, which combine to plot precipitation vs GPP for each of the 10 sites:

- DataExtraction.R - a function that extracts and organises the meteo data and model outputs for each site
- MainScript.R - a script that plots the required graph

Ordering the plots in 1 by mean annual precipitation, mean annual temperature or the coefficient of variation of precipitation (a measure of the variability of rainfall) does not reveal any patterns in the correlations. I believe this is mostly due to the various ecosystems being studied:

Grasslands:

- Konza
- SGS

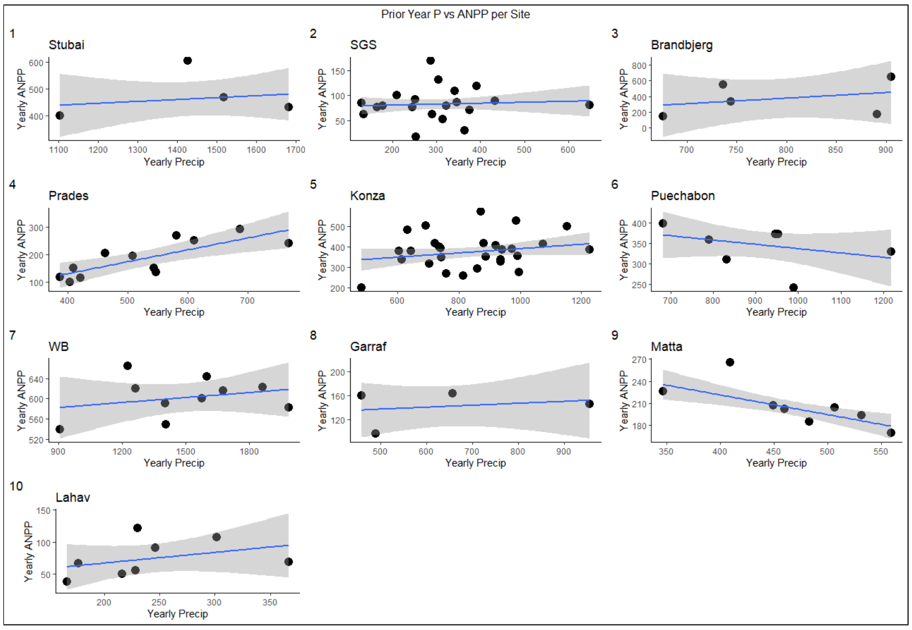

Figure 2: Previous year's precipitation against ANPP for each site on a yearly basis. Blue line is linear regression fitted by geom_smooth from ggplot2 package. Shaded area is 95% confidence interval.

- Stubai (fertilised)
- Brandbjerg (heath/grassland)

Forests:

- WB
- Puechabon
- Prades

Shrub/Heathlands:

- Garraf
- Matta
- Lahav

However, within the grassland sites, Brandbjerg, SGS and Stubai have low MAT (¡10°C) and show no real correlation between precipitation and productivity (and hence are likely temperature limited). Note also that Stubai is incredibly wet (1382 mm/yr) and so is unlikely to be water-limited even in dry years. Meanwhile Konza, which has a higher MAT (12.8°C), has a very strong positive correlation.

Interestingly, SGS has low precipitation but also no correlation. With a MAT of 8.5°C, one might expect more dependence on precipitation.

Konza is a prairie site in Kansas with a substantial data record. From http://lter.konza.ksu.edu/sites/default/files/DC. 75% of rainfall is experienced in the growing season which would explain the correlation. I removed one "outlier" and checked correlation again but it made no difference. If it had made a difference, I would have investigated the outlier to determine why it was an outlier.

Further plots were made of P_{i-1} $ANPP_i$ and T_i $ANPP_i$. A couple of sites exhibit strong correlations for these variables but, for the grasslands, mostly no real relationships are evident. Konza has a negative

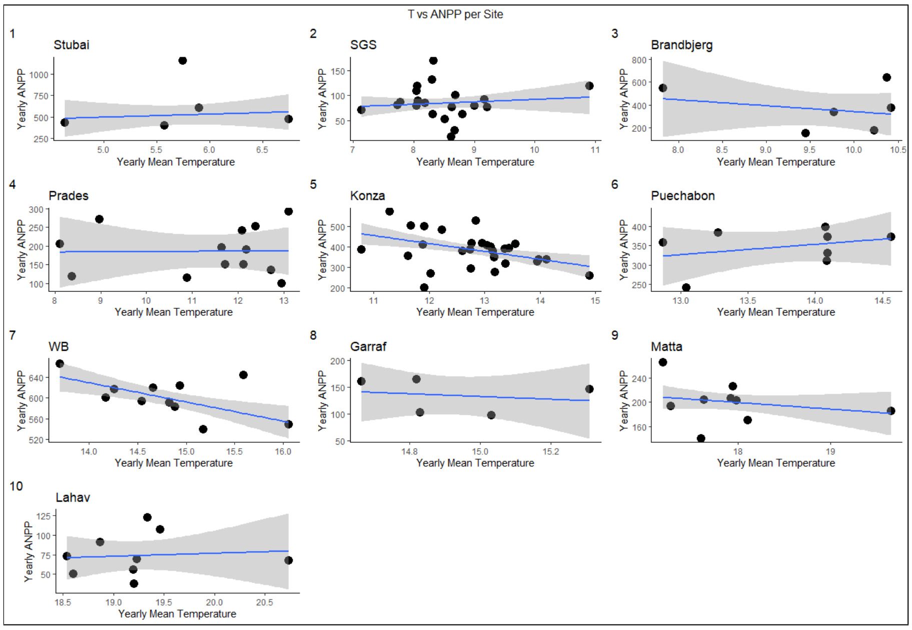

Figure 3: Mean temperature against ANPP for each site on a yearly basis. Blue line is linear regression fitted by geom_smooth from ggplot2 package. Shaded area is 95% confidence interval.

relationship between Tair and ANPP - I would hypothesise that that is because higher Tair tends to coincide with lower rainfall.

A correlation test between Konza Tair and PPT holds this to be true. We obtain r = -0.44 with a p-value of 0.011.

3.2 Possible Next Steps

I can continue this in a variety of ways. I am considering the below methods:

- Take one (or more) sites and apply the SAM-P model to it
- Extend the SAM-P model to include T
- Compare how the SAM-PT model performs for observations vs each of the models

SGS and Konza both have long time series which makes them good candidates for this next step.

3.3 SAM Modelling

I have written/modified the below R scripts:

- TimeSeries.R - moved the creation of monthly and yearly time series into one function
- OrderSites.R - a function that orders sites based on certain characteristics from their yearly data
- SAMFunction.R and Model.R - modified slightly to run with these data
- SAMPlot.R - modified slightly to run with these data

I also added some checks to MainScript.R to check whether data files exist before re-running certain files - no point in doubling up work. One downside to this is if the functions are still being tweaked. Therefore, these checks are definitely something to add in once a function has been finalised.

The plots have been slightly updated as well - just aesthetically.

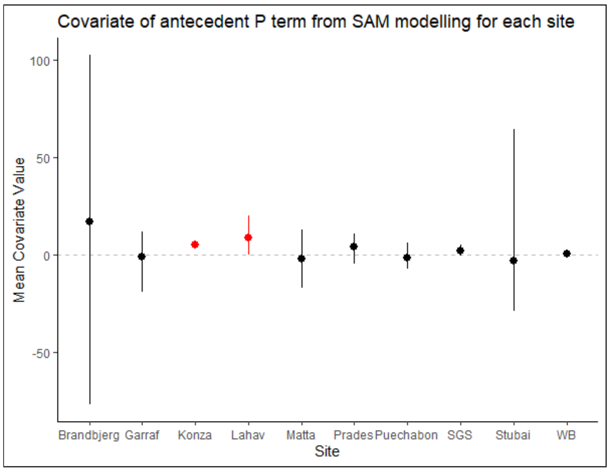

Figure 4: Mean value of α_2 from the SAM model for each site. Error bars are 95% confidence interval. Red values indicate that α_2 is significantly different from 0 at these sites.

4 Week Commencing 11/05

Following a chat with Martin on Friday, I am to try and do the following this week (or as and when I need a break)

- Learn to connect RStudio to git
- Make a website
- Learn some fortran

4.1 SAM Modelling continued...

I have got the SAMPlot function to work and have also created a plot which shows the value of α_2 and whether this is significant or not. I believe that this is the KEY plot - can the influence of PPT_{ant} be ruled significantly different from 0 in a linear model?

In this case, only 2 sites have a significant α_2 as seen in Figure 4 - Konza and Lahav. These also had relatively strong correlations between P and ANPP. As such, these are the only two sites for which the monthly weights are informative.

I am tempted to remove the monthly weight plots for anything that isn't significant - they are not important.

It is done.

This makes the figure much easier to read for the sites where the SAM results matter at a monthly level (otherwise SAM basically says "We are unable to determine whether this site depends on antecedent rainfall" and therefore the monthly results are pointless.)

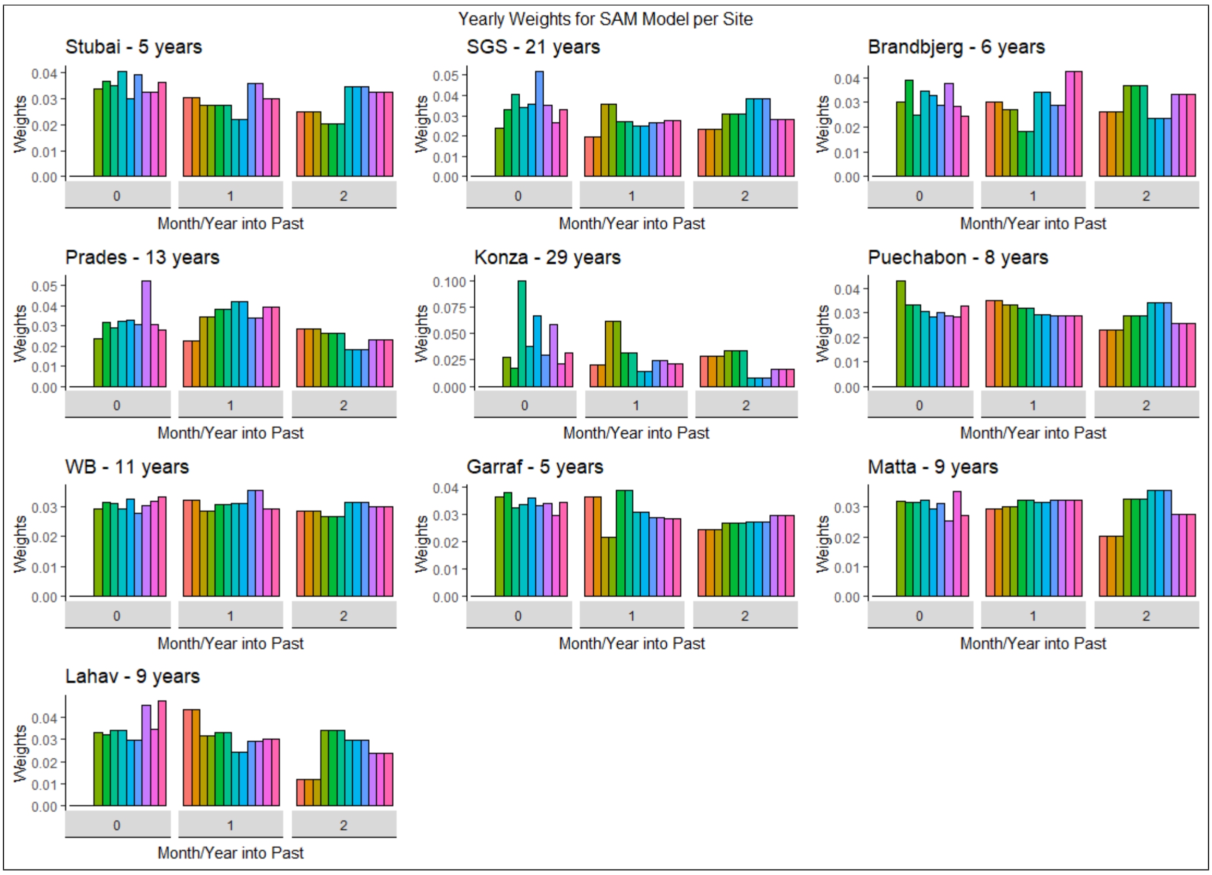

Figure 5: Weight applied to each month as per SAM modelling for each site. The plots extend into the past from left to right - i.e. the first column is December of the current year, the last is January of 2 years ago. Note that Oct-Dec of current year are set to 0.

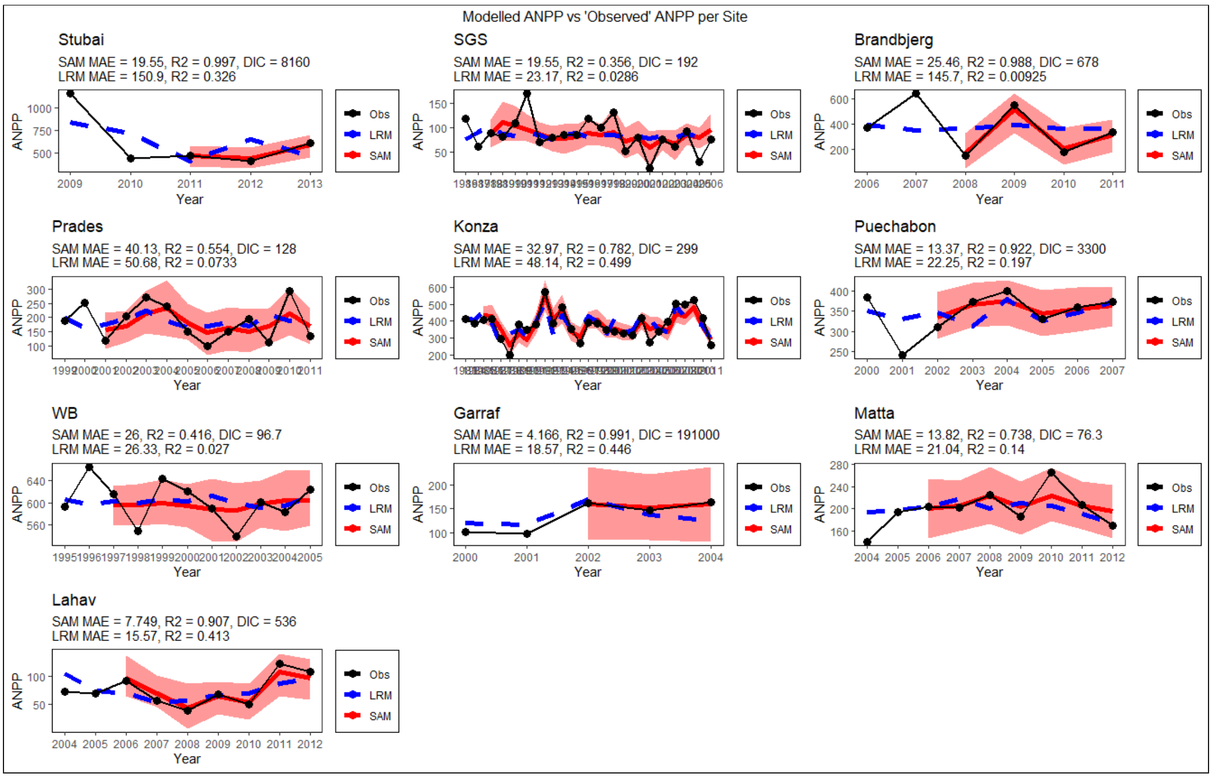

Figure 6: Comparison of modelled and observed ANPP. Two models were used - SAM (red, with 95% confidence interval) and a simple linear regression (blue). Observations are in black. Selected performance metrics are included in subheading.

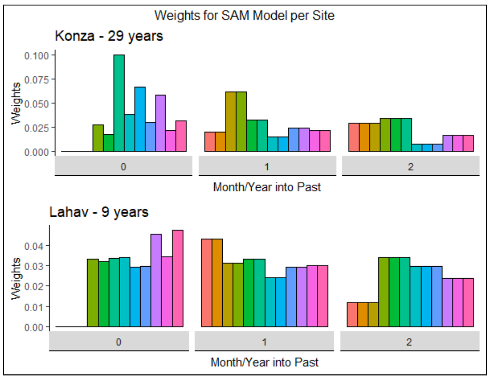

Figure 7: Weight applied to each month as per SAM modelling for each site where covariate of antecedent rainfall is significantly different from 0. The plots extend into the past from left to right - i.e. the first column is December of the current year, the last is January of 2 years ago. Note that Oct-Dec of current year are set to 0.

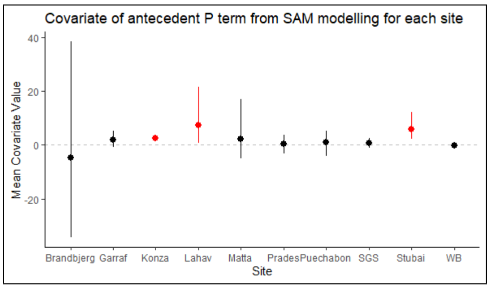

Figure 8: Covariates for Nlag = 1. Red indicates significance.

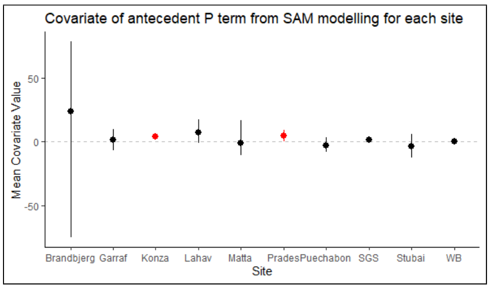

Figure 9: Covariates for Nlag = 2. Red indicates significance.

4.2 Questions and Thoughts

- Is there a more informative way to plot the month weights? Does mean-sd for each weight help us see which weights are also significantly different from 0? Note that the 2.5% quantile is always greater than 0 and therefore this is not very informative (or is it?) - **Ryan et al, 2015, has plots in the supplementary information that plot the quantile CIs - it appears that the important takeaway is to compare weights relative to each other and against the prior weight (which is approx. 1/Nblock)**
- Is running a comparison for 2-5 year lags to see whether the significant αs change and whether the weightings of each year change worthwhile? - **This has been done and is discussed below**
- Does the sum of weights within a year provide any info - e.g. that the previous year is more important than the current if their weights sum to 38% and 32% respectively?

I have run the SAM model for lags from 1-5 years and have received an interesting result - sites where antecedent rainfall is significant at a shorter lag are no longer significantly affected at antecedent rainfalls when a longer lag is considered. This is a unexpected result. One would assume that, if rainfall in years i and i-1 can predict productivity, introducing year i-2 should still be able to predict productivity by setting the weights for year i-2 to 0 (or close to) if necessary... See figures 8 and 9 for the plots...

Note how antecedent P at Lahav and Stubai is no longer significant when a lag of an additional year is considered. P_{ant} is significant at Prades for Nlag = 2 but not for any other Nlag (1-5).

5 Week Commencing 18/05

Work completed last week includes new functions for conducting the SAM approach but for a model of the form

$$\mu_{ANPP}[i] = \alpha_1 + \alpha_2 * PPT[i] + \alpha_3 * T_{ant}[i]$$

A new style of plotting monthly weights has also been introduced (although more work could be done) which shows the change from the prior.

This has raised an interesting question: With the Dirichlet prior, each weight is assigned an equal value in the prior. Is it logical/feasible to use a different prior, which assigns equal weights to the first year but then 0 to the rest of the weights? This is equivalent to (or is justified by) the argument "We hypothesise that production is only influenced by current climate" - as such, we want strong evidence before accepting non-zero weighting for antecedent precipitation. Would this work with the precisely zero covariates, or would this extra layer of non-influence mean that we are too demanding for proof?

Another interesting issue is sites that show significant P_{ant} in the P model do not always show this significance in the PT model.