MEP: ECSC10040

Forest biomass dynamics

Session 2: 26th October 2022

Time series analysis, forecasting, dynamic models carbon fluxes

John L. Godlee

Continue with empirical models, this time focusing on incorporating time into our models and using models which incorporate dynamic processes.

Today…

1. Follow-up - last week's practical
2. Recap - forest growth dynamics
3. Stochastic vs. deterministic, static vs. dynamic
4. Time series decomposition and forecasting
5. Dynamic models

Objectives:

- Gain experience fitting ARIMA and ETS time series models.
- Gain experience forecasting time series data
- Understand the features of a dynamic model
- Use a dynamic model in the scenario of savanna tree-grass coexistence

First, follow up to last week's practical. Got to the end of section 6 in the practical worksheet, so wanted to go through those extra bits we didn't finish in class

Then, a short recap on forest growth dynamics and the carbon cycle

A bit more model theory, contrasting stochastic vs. deterministic, and static vs. dynamic models

Then, time series decomposition and forecasting time series data

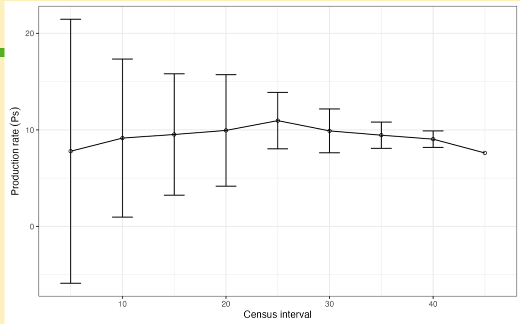
And finally some preparation for the practical which is going to be mostly about using dynamic models.

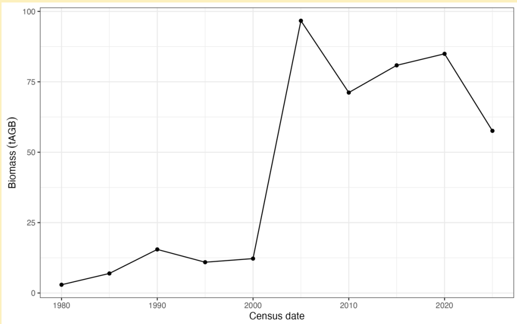
By the end of this you should have some experience using time series analysis to gain insight about time series data, and you should have an idea of how dynamic models work and what their pros and cons are.

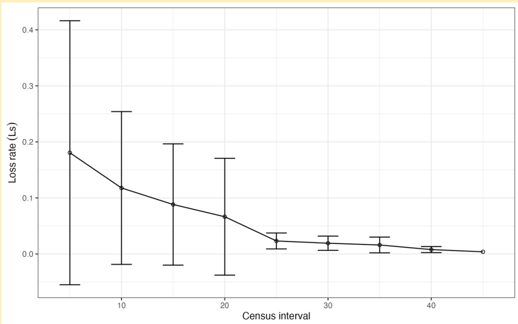
Last week's practical

How does time period affect estimates of productivity and loss?

- Increased variability in rates captured
- Under-estimate of loss rate at longer census intervals







In the practical we looked at modelling rates of productivity and loss using field data from forest plots.

I asked how the length of the time period affects our estimates of productivity and loss rates.

I conducted a small experiment with some fake data to illustrate what happens when we increase the census interval length.

Plot on the left shows the change in biomass over time in this fake site. There is a big jump in the middle, and it starts to decline at the end of the census period.

I calculated the rates using every pairwise combination of censuses, so we had census interval lengths of 5, 10, 15, 20 years etc.

Increased variability in the estimated rates is observed when you have shorter census period. Also, our estimate of loss decreases as the census interval length increases. This is because we are missing lots of growth, recruitment, and subsequent mortality with a longer census period.

Ultimately, the census period you choose should be dependent on what you are looking to research, either the variability in the system, or the long term behaviour of the system.

Last week's practical

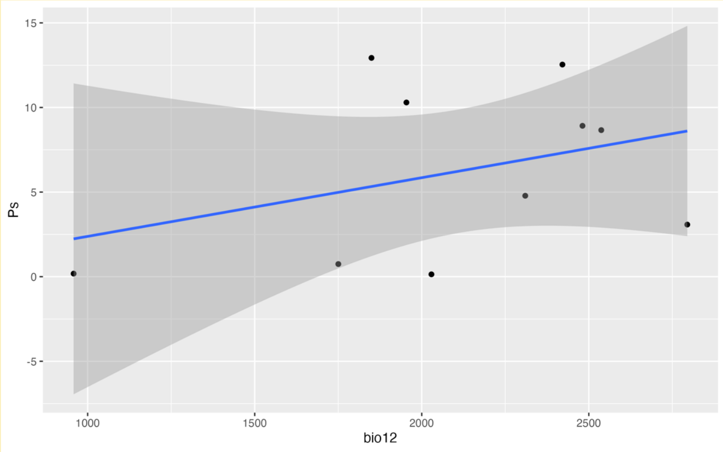
Predicting mean period biomass from climate data

bio12 = Mean Annual Precipitation

Expect that precipitation correlates with water availability.

Expect that increased water availability allows increased productivity.

What other variables do we expect to predict productivity of loss rates?



In section 7 we looked at whether we could predict productivity, loss or mean period biomass using environmental data, from the BioClim dataset.

I chose bio12, mean annual precipitation as an example. I'd expect that precipitation correlates with water availability, and that water availability allows increased productivity, as photosynthesis can occur at a higher rate if not limited by water, but I didn't find that.

What other variables did you choose to investigate as drivers of productivity or loss? What was your ecologically relevant hypothesis? Did you find anything?

Last week's practical

How can the model of forest growth be improved?

Currently, our model of forest biomass dynamics does not self-regulate (static).

Our empirical model of biomass dynamics, using estimates of productivity and loss, currently is a static model, i.e. it doesn't include any terms or interactions which alter the behaviour of the system based on the state of the system itself. We estimate productivity and loss and biomass for a given period, and we can use those estimates to model biomass into the future, assuming a linear process, but this is unrealistic beyond a certain point, because the biomass will just continue to increase until the forest is a single block of wood. This isn't what happens in reality.

I asked whether there was any improvements we could make to the model so that it is more dynamic.

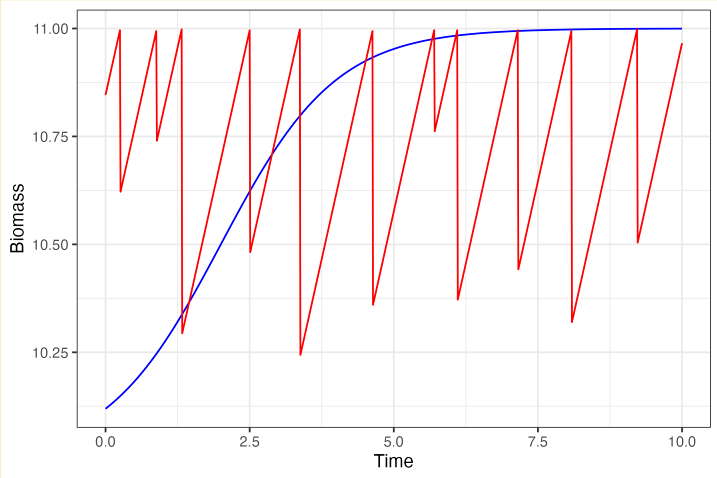
Last week's practical

How can the model of forest growth be improved?

Currently, our model of forest biomass dynamics does not self-regulate (static).

- Disturbance events (red)
- Negative density dependence (blue)

Any other methods?

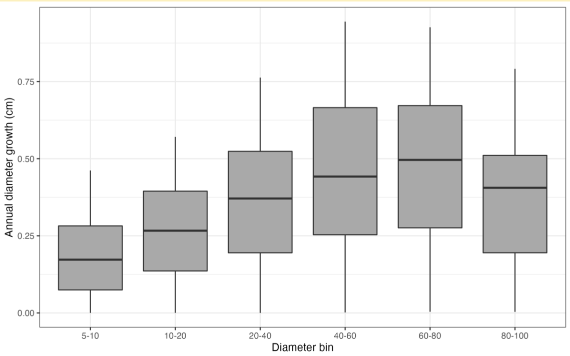


I came up with two methods. The first is to introduce some kind of disturbance which reduces biomass in the system. This is represented by the red line in the plot on the right. This might be fire, drought, or even tree harvesting by humans. The second is to introduce some kind of negative density dependence between the trees in the system so the productivity of the system saturates at higher biomass. E.g. competition for light in the canopy. Another tree can only recruit and grow once a previous tree falls and creates a gap in the canopy.

Last week's practical

Size-class and species specific growth rates

```
              Df Sum Sq Mean Sq F value Pr(>F)    
diam_cut       5     46    9.19     290 <2e-16 ***
Residuals   8242    261    0.03                   
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



	diff	lwr	upr	p adj
10-20-5-10	0.08088	0.06231	0.09945	0.0000
20-40-5-10	0.17689	0.15744	0.19634	0.0000
40-60-5-10	0.26426	0.23833	0.29019	0.0000
60-80-5-10	0.29146	0.24752	0.33540	0.0000
80-100-5-10	0.18946	0.06865	0.31026	0.0001
20-40-10-20	0.09601	0.08314	0.10887	0.0000
40-60-10-20	0.18338	0.16194	0.20481	0.0000
60-80-10-20	0.21058	0.16913	0.25202	0.0000
80-100-10-20	0.10857	-0.01135	0.22850	0.1021
40-60-20-40	0.08737	0.06517	0.10957	0.0000
60-80-20-40	0.11457	0.07272	0.15642	0.0000
80-100-20-40	0.01257	-0.10750	0.13263	0.9997
60-80-40-60	0.02720	-0.01802	0.07242	0.5222
80-100-40-60	-0.07480	-0.19608	0.04648	0.4932
80-100-60-80	-0.10200	-0.22836	0.02436	0.1937

In section 9 I showed how we could improve the generality of our model by calculating productivity and loss rates for different size classes and different species of tree.

I showed how to calculate diameter growth rates for different groups and then asked you to perform an ANOVA to see if they differed significantly.

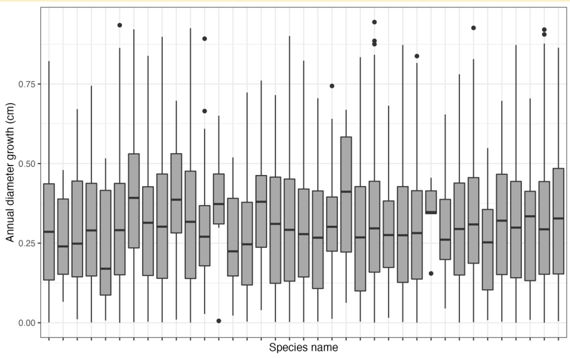
The boxplot shows that size class does appear to influence growth rates, indicating that this would be a valuable addition to our model of biomass dynamics if we wanted to model a novel system.

The Tukey's tests on the bottom right show that the smallest trees have a significantly lower growth rate than the larger trees. Also that the largest trees tend to have lower growth rates than mid-large trees.

Last week's practical

Size-class and species specific growth rates

```
                Df Sum Sq Mean Sq F value Pr(>F)
species_name    36    1.8  0.0507    1.36  0.073 .
Residuals     8211  305.5  0.0372
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```



	diff	lwr	upr	p adj
Macrolobium acaciifolium-Machaerium acutifolium	-0.11279	-0.26064	0.03507	0.5613
Macrolobium acaciifolium-Licania kunthiana	-0.06434	-0.14995	0.02126	0.5993
Panopsis rubescens-Machaerium acutifolium	-0.12437	-0.29164	0.04291	0.6270
Licania kunthiana-Hevea brasiliensis	0.04436	-0.01546	0.10418	0.6336
Machaerium acutifolium-Euterpe precatoria	0.11242	-0.04171	0.26656	0.6739
Matayba macrostylis-Cecropia polystachya	-0.18755	-0.45089	0.07579	0.7285
Matayba macrostylis-Machaerium acutifolium	-0.15282	-0.36832	0.06268	0.7378
Panopsis rubescens-Cecropia polystachya	-0.15910	-0.38469	0.06649	0.7495
Macrolobium acaciifolium-Cecropia polystachya	-0.14752	-0.35911	0.06408	0.7731
Machaerium acutifolium-Hevea brasiliensis	0.09280	-0.04179	0.22739	0.7944
Machaerium acutifolium-Guarea macrophylla	0.10861	-0.04954	0.26676	0.8019
Cecropia polystachya-Bowdichia virgilioides	0.21238	-0.09737	0.52213	0.8048
Euterpe precatoria-Cecropia polystachya	-0.14715	-0.36318	0.06887	0.8165
Licania kunthiana-Euterpe precatoria	0.06398	-0.03206	0.16002	0.8528
Spondias mombin-Cecropia polystachya	-0.17063	-0.42960	0.08833	0.8689
Machaerium acutifolium-Bowdichia virgilioides	0.17765	-0.09260	0.44789	0.8721

On the other hand, species doesn't seem to have a significant effect on growth rates.

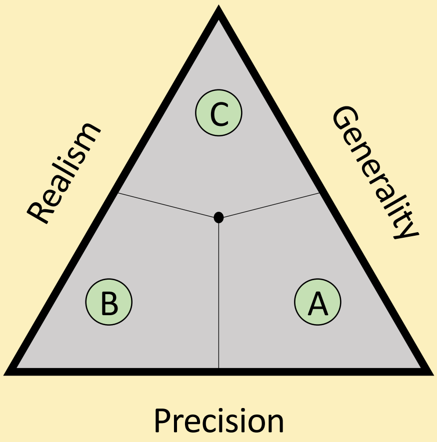
But we expect that tree species should differ in their niche if they are to coexist in a single plot. Why might we not have found any evidence of niche differentiation here?

Possibly because in a hyper-diverse rainforest like in the Amazon, there are many tree species which differ not in their growth rates and mortality rates, but instead according to other things like resistance to pests or mutualistic relationships.

Generality - Precision - Reality

- **Generality**:
  - Does the model function correctly under a diverse range of conditions?
- **Realism**:
  - Does the model realistically simulate processes?
- **Precision**:
  - Does the model provide precise numeric outputs?

A. General and Precise
   - Good for: describing systems
B. Precise and Realistic
   - Good for: understanding system behaviour
C. Realistic and General
   - Good for: theory development

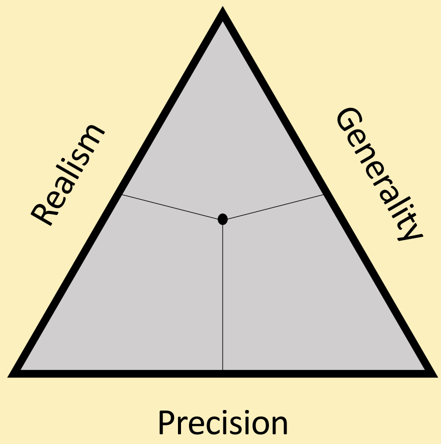


Levins et al. (1966)

Remember the triangle of generality, precision and realism?

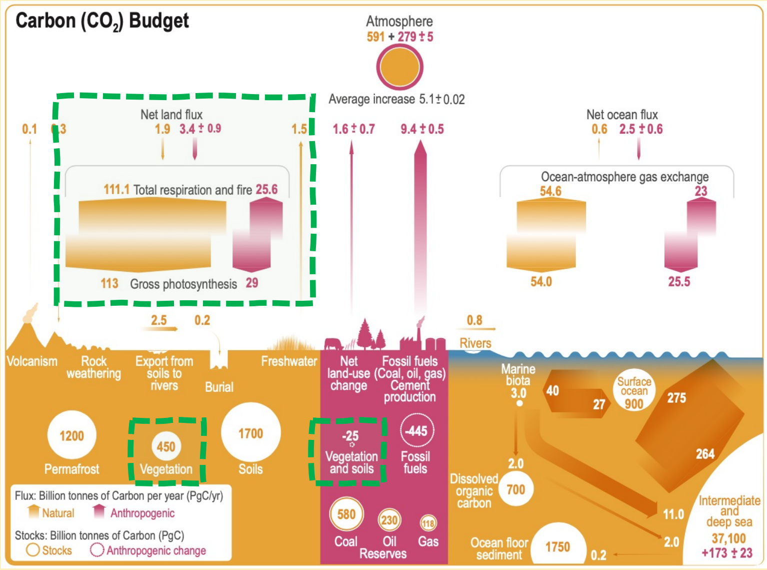
Exercise: Pin the model on the PGR triangle

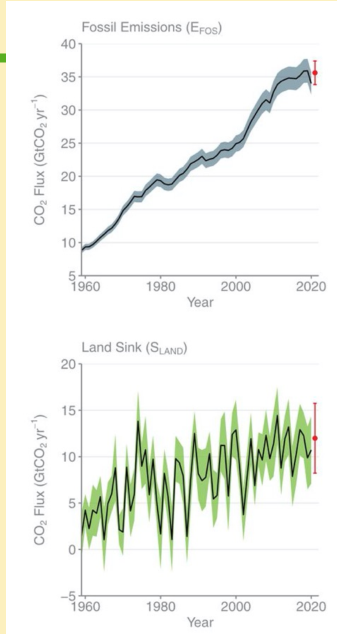
1. Read the scenarios (10 minutes)
2. Mark where the model in the scenario sits on the PGR triangle
3. Explain why…
4. Describe the model in other ways:
    - Spatial and temporal limits
    - Empirical, mechanistic, conceptual
    - Stochastic or deterministic
    - What is the purpose of the model?



Levins et al. (1966)

Recap: forest growth dynamics





Canadell et al. (2021)

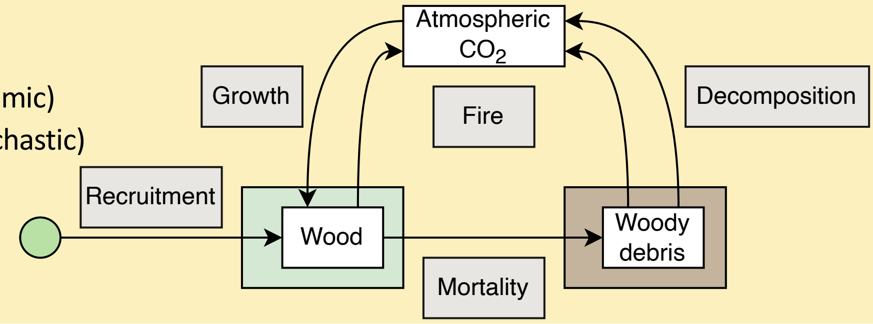
We are looking fundamentally at the carbon cycle, carbon sinks, and the ability for the terrestrial land sink to offset anthropogenic carbon emissions.

Static vs. dynamic … Deterministic vs. stochastic

- Goal: understand variation in forest biomass dynamics
- What drives the variation?
  - Internal - species composition, size class distribution
  - External - environment, biogeography(?)

- Static model (not dynamic)
- Deterministic (not stochastic)

$$NBP = G + R - M$$



Last week we used a simple box model of tree growth and mortality to generate estimates of net biome production, i.e. the net biomass produced by the system minus all the mortality and other losses.

But as I discussed in reference to the practical: static models are only parameterized for a single time point or period. The parameters of the model don't interact with each other to change model behaviour over time. We don't include any mechanistic understanding of the system in our model.

Our model previous model was also deterministic. It uses fixed inputs to calculate model outputs. It doesn't consider a range of starting values, and it doesn't include stochastic processes in the model such as disturbance events.

This limits the realism of these models, as we know that there is inherent variation in ecosystems and their extrinsic drivers.

Abrupt vs. gradual drivers of forest dynamics

Abrupt

- Fire
- Cyclones
- Flooding
- Land use change
  - Clear-felling, tree planting

Gradual

- Climate change
  - Temperature
  - Precipitation
- Nitrogen deposition
- Atmospheric CO_2

- Biogeography (species function)
- Succession
- Size class structure

A quick mention that drivers of biomass change operate at different temporal scales, so we need models which can incorporate both of these.

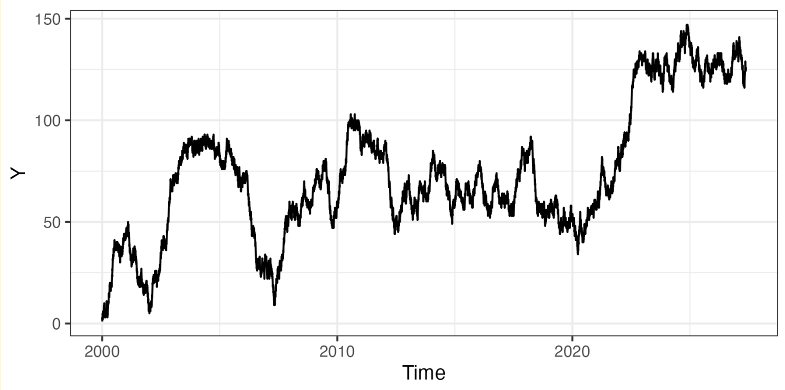
Depends on temporal resolution what we classify as abrupt vs. gradual.

Also depends on how we structure our model, which ultimately depends on the research question we want to answer

We are going to spend some time in today's session thinking about how to model the effect of these different kinds of drivers on forest biomass dynamics.

Time series

- Data points in time order
- Natural order to data
- Temporal autocorrelation
- Lag effects



Time series data are not like other 1D data.

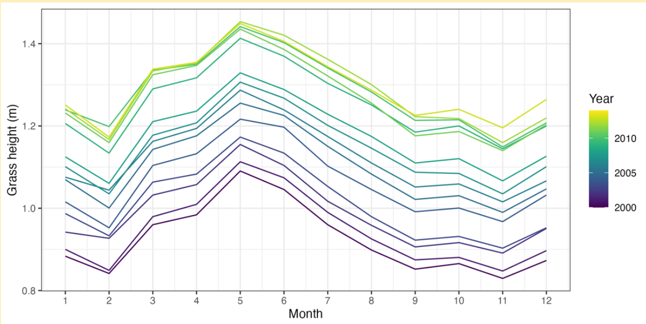
The data points have a natural order to them and often previous values affect the current value. Additionally, there might be lag effects, i.e. response to driver is not immediate.

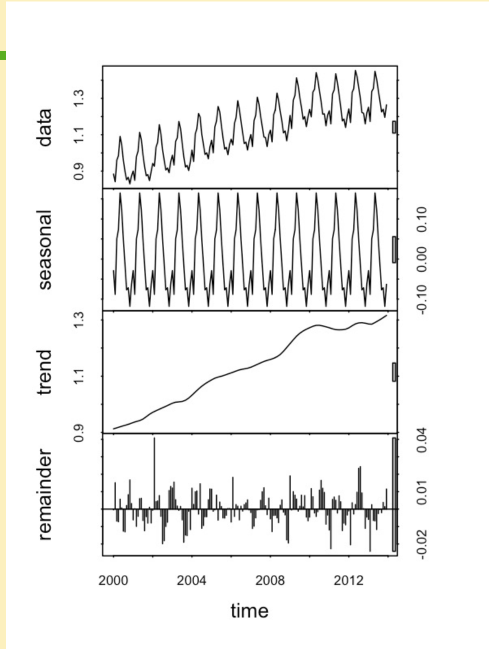
Because of this, we can't use many normal parametric statistical analyses to investigate patterns in time series data.

Decomposing time series

- Trend component
- Seasonality / cyclical behaviour
- Remainder (noise?)

Can subtract components from data to adjust time series





Time series data can have nested patterns within them. We might have an overall trend in the data, but this trend can be obscured if we don’t also account for seasonal or other cycles of change. We can do an analysis called time series decomposition to partition the trend from seasonal patterns, and the noise in the data.

We will have a go at this in the practical.

Forecasting and interpolating time series - ETS

ETS = Error, Trend Seasonality models = **Exponential Smoothing** State Space models

How does a single variable change over time?

Provides an estimate of uncertainty.

Appropriate if data is not stationary and if there is seasonality

Weights influence of previous points depending on how much time has passed.

Can specify:
- Error type
- Trend type
- Season type

- None
- Additive
- Multiplicative
- Auto

Hyndman and Athanasopoulos (2021)

Often, we want to use a time series to figure out what the value of the data might be at some point in the future, this is called forecasting. We might also want to fill in gaps in the time series where data has been lost or wasn't collected, this is called interpolation.

We are going to look at two main forms of time series model which can be used for forecasting and interpolation. We will spend more time on this in the practical. For now just making you aware that they exist.

The first is ETS - Error,Trend,Seasonality models.

They weight the influence of previous points depending on how much time has passed and use decomposition to determine the error, trend and seasonal cycles in the data, then forecast those forwards.

ETS analysis is only really applicable when there is both a trend and some seasonality in the data, otherwise it models the noise and gets confused.

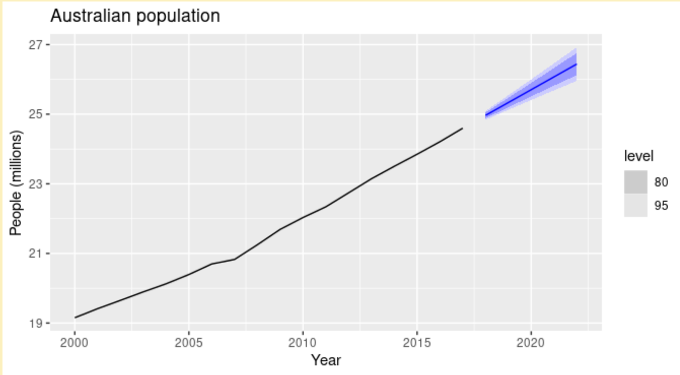
Much more in the referenced book below.

Forecasting and interpolating time series - ARIMA

ARIMA = Auto-Regressive Integrated Moving Average models

Appropriate if data is stationary or not stationary

Constructs model using autocorrelation of data



Hyndman and Athanasopoulos (2021)

ARIMA models = Auto-Regressive Integrated Moving Average models can be used whether the data is stationary or not. In contrast to ETS models, ARIMA models focus on using the autocorrelation structure of the data to forecast into the future. They can also be used for interpolation. ARIMA models are more commonly used in ecological research nowadays.

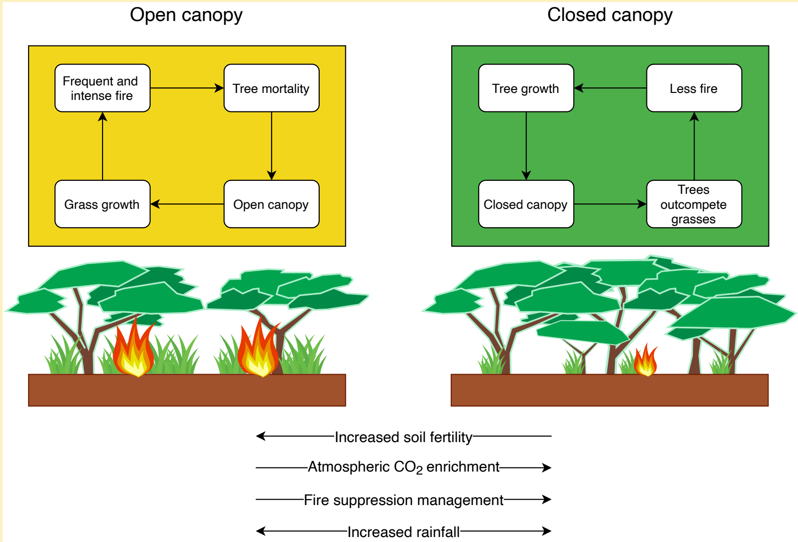
Savanna tree-grass coexistence

Savannas are a mosaic of closed and open canopy patches.

"Alternative stable states"

Positive feedbacks of fire/herbivory on grass fuel load.

Will the balance tip due to climate change?



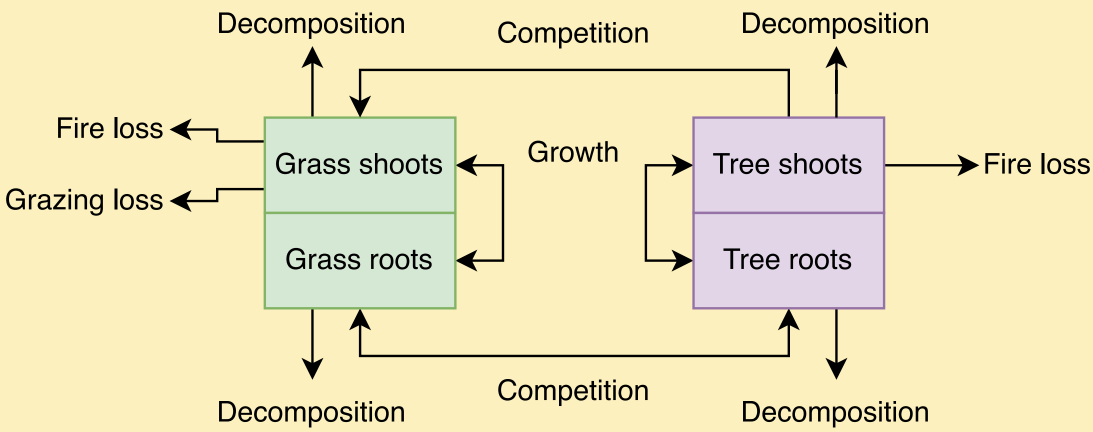
Staver et al., (2011)

Moving on from time series, let's talk more about dynamic models.

These are models where the state of the system influences the behaviour of the system at a given time point.

In the practical we are going to use the coexistence of trees and grasses in African savannas as an example system to study dynamic modelling.

A dynamic model of tree-grass coexistence



Higgins et al. (2010)

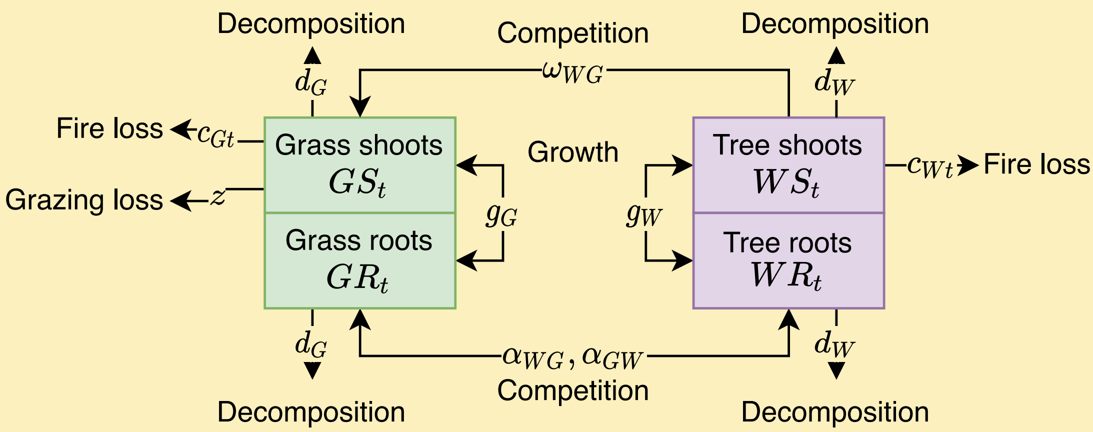
This conceptual model was created by Steve Higgins about 15 years ago.

Assumes that trees and grasses compete with each other above and below ground. Though above-ground competition is only from trees to grasses, grasses can't shade out trees above ground.

Includes stochastic processes in the form of fire loss, and proportional losses in the form of herbivory of grass shoots.

"Decomposition" in this sense really means mortality, or any loss of living biomass from the system.

A dynamic model of tree-grass coexistence



Higgins et al. (2010)

They parameterized the model.

A dynamic model of tree-grass coexistence

$$GS_{t+1} = GS_t + g_G GR_t(1 - GS_t - \omega_{WG} WS_t) - c_{Gt} GS_t - d_g GS_t - zGS_t$$

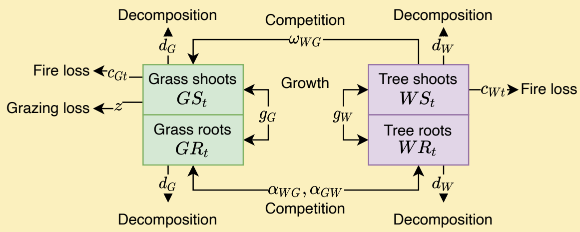
Grass shoot biomass =

+ Existing grass shoot biomass
+ growth by grass roots modulated by competition and existing grass shoot biomass
– loss by fire – loss by "decomposition" (mortality) – loss by herbivory

$$GR_{t+1} = GR_t + g_G GS_t(1 - GR_t - \alpha_{WG} WR_t)$$
$$-d_G GR_t$$

$$WS_{t+1} = WS_t + g_W WR_t(1 - WS_t)$$
$$-c_{Wt} WS_t - d_w WS_t$$

$$WR_{t+1} = WR_t + g_W WS_t(1 - WR_t - \alpha_{GW} GR$$
$$-d_w WR_t$$



Higgins et al. (2010)

And wrote the model as four linked equations which describe each of the four biomass components at time t+1.

Using the first as an example:

The grass shoot biomass at t+1 is determined by the existing grass shoot biomass, the growth by the grass roots modulated by competition and the existing grass shoot biomass, minus the loss by fire, loss by mortality, and loss by herbivory.

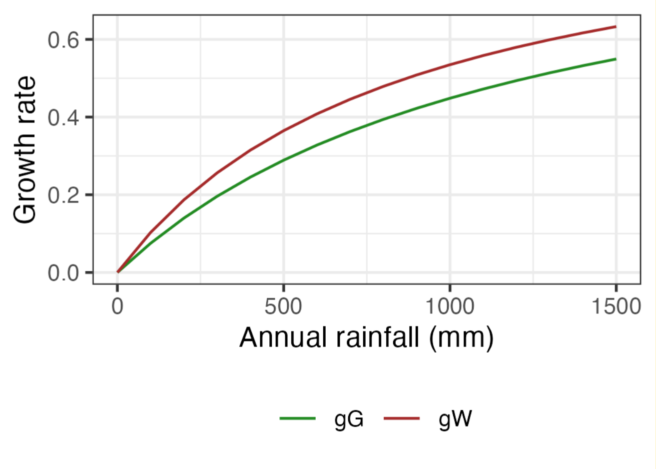
A dynamic model of tree-grass coexistence

Growth rates are dependent on rainfall:

$$g = \frac{1}{1+(\tilde{g}/R)}$$

Separate $\tilde{g}$ for trees and grasses

Saturating relationship (density dependent)



Higgins et al. (2010)

The growth rates in the model are determined separately for grasses and trees, and are related to rainfall. There is a saturating relationship, so growth rate increases less at higher rainfall.

The g-tilde term determines the shape of that saturating relationship.

Grass growth rate lower, and saturates earlier, due to shallower roots and other limits on growth.

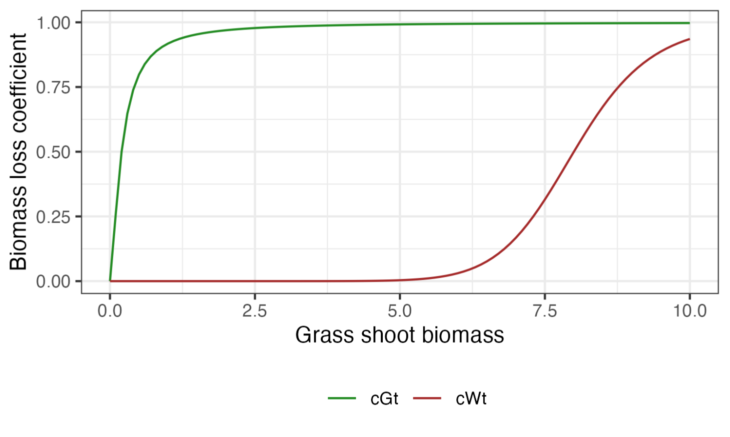
A dynamic model of tree-grass coexistence

Proportion of biomass loss due to fire are dependent on existing grass shoot biomass

$$C_t = \frac{GS_t^b}{a^b + GS_t^b}$$

Separate a and b for trees and grasses.

Asymptotic (sigmoidal) relationship



Higgins et al. (2010)

The amount of biomass lost to fire is dependent on the grass fuel load. More grass results in more intense fires.

A sigmoidal relationship describes the relationship between grass shoot biomass and the biomass loss coefficient ct. It takes more grass fuel to remove trees than grasses. After around 6 tons the fuel load is high enough to kill trees.

”a” and “b” coefficients describe the sigmoidal curve shape.

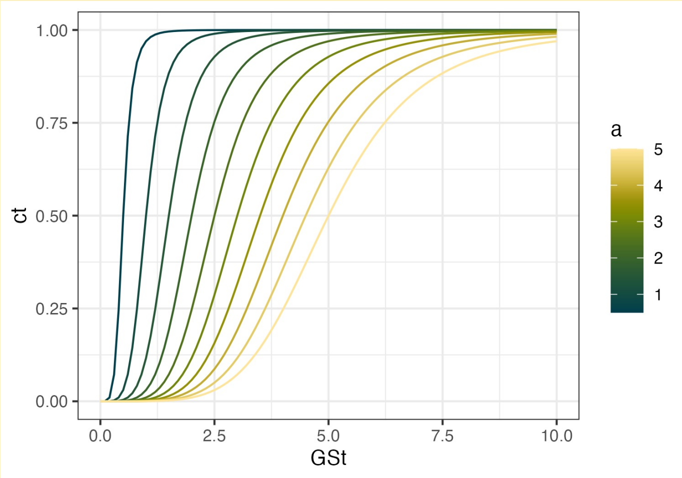
A dynamic model of tree-grass coexistence

Proportion of biomass loss due to fire are dependent on existing grass shoot biomass

$$C_t = \frac{GS_t^b}{a^b + GS_t^b}$$

Separate a and b for trees and grasses.

Asymptotic (sigmoidal) relationship



Higgins et al. (2010)

As a increases the the curve starts saturating lower, lowers steepness. As “b” increases the inflection point of the curve moves up and to the right.

Wood has a lower ”b” as it takes higher fire intensity to kill a tree

Wood has a higher “a”, as less will be consumed regardless of intensity.

What do we know?

- How to describe a model in terms of Precision, Generality, Realism
- Two important types of time series analysis (ARIMA, ETS)
  - Interpolation, forecasting, decomposition
- Dynamic models != Static models. Stochastic models != Deterministic models
  - Inclusion of stochastic events
  - More complex interactions between model state variables
- Savanna tree-grass coexistence is determined by many factors
  - This coexistence can be modelled using a dynamic model

Practical

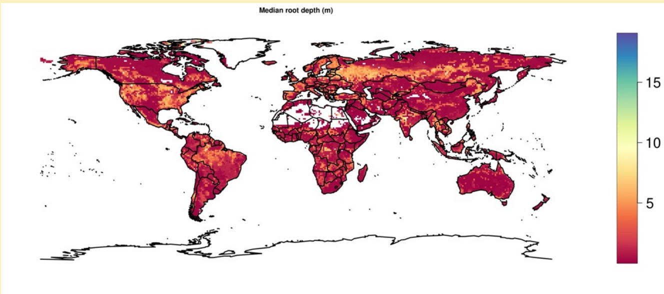
Go to Learn:

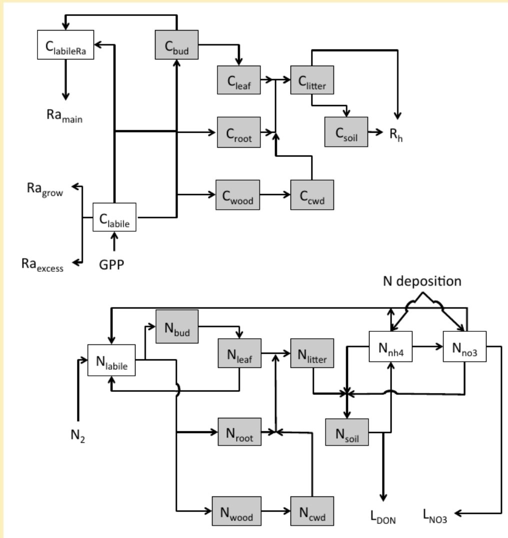
i. Modelling Ecosystem Processes
ii. Module 3 - Forest Biomass Dynamics
iii. Week 6

- Download all the files to a single folder
- Open Rstudio
- Work through the practical worksheet

Next time…

- Luke Smallman's module
- Process-based (mechanistic) models
- More on carbon cycling
- Complex interactions





MID-COURSE FEEDBACK!

References

Bloom, A. A., J. Exbrayat, I. R. van der Velde, L. Feng & M. Williams (2016). 'The decadal state of the terrestrial carbon cycle: Global retrievals of terrestrial carbon allocation, pools, and residence times'. In: Proceedings of the National Academy of Sciences 113.5, pp. 1285–1290. doi: 10.1073/pnas.1515160113.

Canadell, J., P. Monteiro, M. Costa, L. Cotrim da Cunha, P. Cox, A. Eliseev, S. Henson, M. Ishii, S. Jaccard, C. Koven et al. (2021). 'Global Carbon and other Biogeochemical Cycles and Feedbacks'. In: Climate Change 2021: The Physical Science Basis. Contribution of Working

Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Ed. by V. Masson-Delmotte, P. Zhai, A. Pirani, S. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M. Gomis et al. Cambridge, United Kingdom and New York, NY, USA, pp. 673–816. doi: 10.1017/9781009157896.007.

Fick, S. E. & R. J. Hijmans (2017). 'WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas'. In: International Journal of Climatology 37.12, pp. 4302–4315. doi: http://dx.doi.org/10.1002/joc.5086.

Higgins, S. I., S. Scheiter & M. Sankaran (June 2010). 'The stability of African savannas: insights from the indirect estimation of the parameters of a dynamic model'. In: Ecology 91.6, pp. 1682–1692. doi: 10.1890/08-1368.1.

Hyndman, R. J. & G. Athanasopoulos (2021). Forecasting: principles and practice, 3rd edition. OTexts. Melbourne, Australia.

Levins, R. (1966). 'The strategy of model building in population biology'. In: American Scientist 54.4, pp. 421–431.

Staver, A. C., S. Archibald & S. Levin (2011). 'Tree cover in sub-Saharan Africa: rainfall and fire constrain forest and savanna as alternative stable states'. In: Ecology 92.5, pp. 1063–1072. doi: http://dx.doi.org/10.1890/10-1684.1.