The power of model validation: Why, how and what for?

Tom Anderson[1] Maike Sonnewald[1,2] Martin Wood[1,2]

[1]National Oceanography Center, Southampton, University of Southampton, UK
[2]Institute for Complex Systems Simulation, University of Southampton, UK

National Oceanography Centre, Southampton
UNIVERSITY OF SOUTHAMPTON AND NATURAL ENVIRONMENT RESEARCH COUNCIL

ICSS

19th August 2014

Outline

1. Motivation
   - What we want you to take away from this...
2. Tom Anderson
   - Model Validation
3. Examples
   - Making “Daisyworld” useful
4. Resources!
   - Natural
   - Sociological
5. And now on to some handy stats...

Motivation

- Motivation of the “Complex Systems” approach
  - Keep it **simple** stupid...
- Examples of simple impact boosting model modifications
  - Keep it **relevant** stupid...
- Brief tour of available resources
  - Keep it **real** stupid...
- Demonstration of 1D and 2D timeseries analysis
  - Keep it **significant** stupid...

Take home message

Consider the data; Talk to people; Do awesome science!

Motivation

- Motivation of the “Complex Systems” approach
  - Keep it **simple** stupid...
- Examples of simple impact boosting model modifications
  - Keep it **relevant** stupid...
- Brief tour of available resources
  - Keep it **real** stupid...
- Demonstration of 1D and 2D timeseries analysis
  - Keep it **significant** stupid...

Take home message

Consider the data; Talk to people; Do awesome science!

Motivation

- Motivation of the “Complex Systems” approach
  - Keep it **simple** stupid...
- Examples of simple impact boosting model modifications
  - Keep it **relevant** stupid...
- Brief tour of available resources
  - Keep it **real** stupid...
- Demonstration of 1D and 2D timeseries analysis
  - Keep it **significant** stupid...

Take home message

Consider the data; Talk to people; Do awesome science!

Motivation

- Motivation of the “Complex Systems” approach
  - Keep it **simple** stupid...
- Examples of simple impact boosting model modifications
  - Keep it **relevant** stupid...
- Brief tour of available resources
  - Keep it **real** stupid...
- Demonstration of 1D and 2D timeseries analysis
  - Keep it **significant** stupid...

Take home message

Consider the data; Talk to people; Do awesome science!

Motivation

- Motivation of the “Complex Systems” approach
  - Keep it **simple** stupid...
- Examples of simple impact boosting model modifications
  - Keep it **relevant** stupid...
- Brief tour of available resources
  - Keep it **real** stupid...
- Demonstration of 1D and 2D timeseries analysis
  - Keep it **significant** stupid...

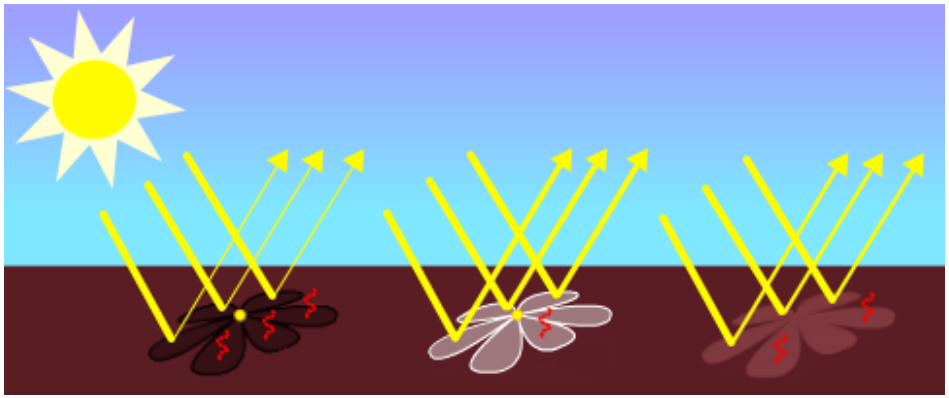
Take home message

Consider the data; Talk to people; Do awesome science!

Prof. Tom Anderson

Expert in ecosystem modeling

Introducting: Daisyworld



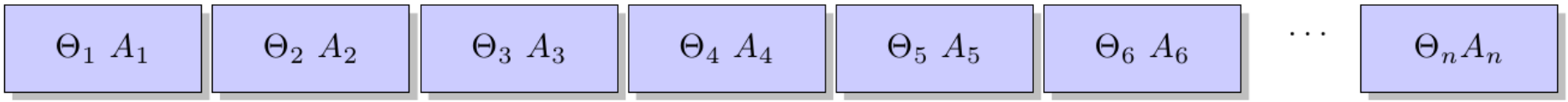
In a nutshell

The effect of albedo on regulating temperature

Introducting: Daisyworld

$$\frac{dR}{dt} = \alpha \sum_i \Theta_i + \beta(P - R)$$

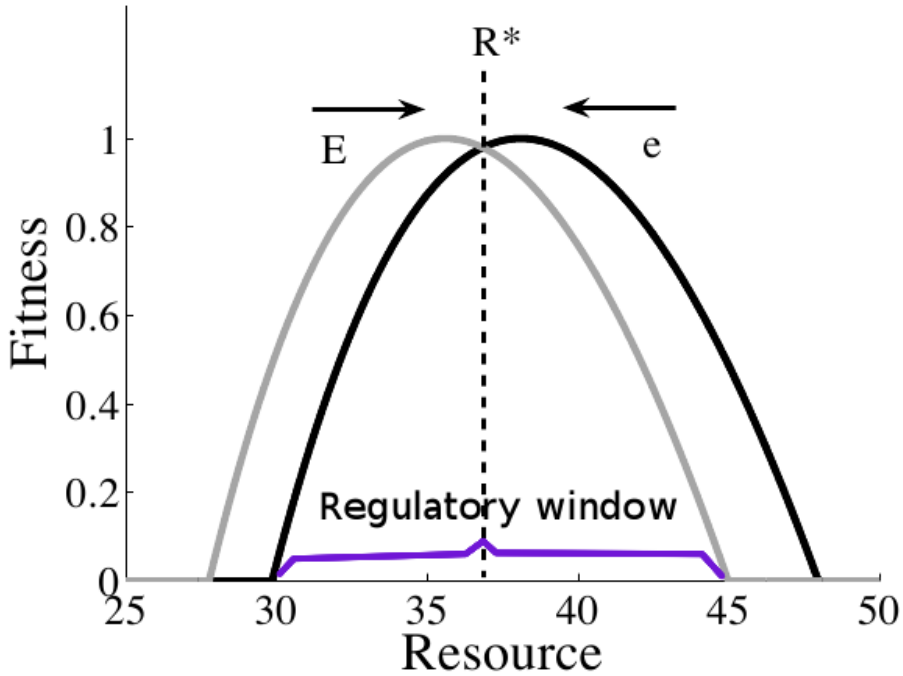
- R=Resource
- α=population influence on R
- β=External forcing influence on R
- Θ=Effect of individual on R
- P=External forcing



Array of individual daisies given by: A, the optimal R value and Θ

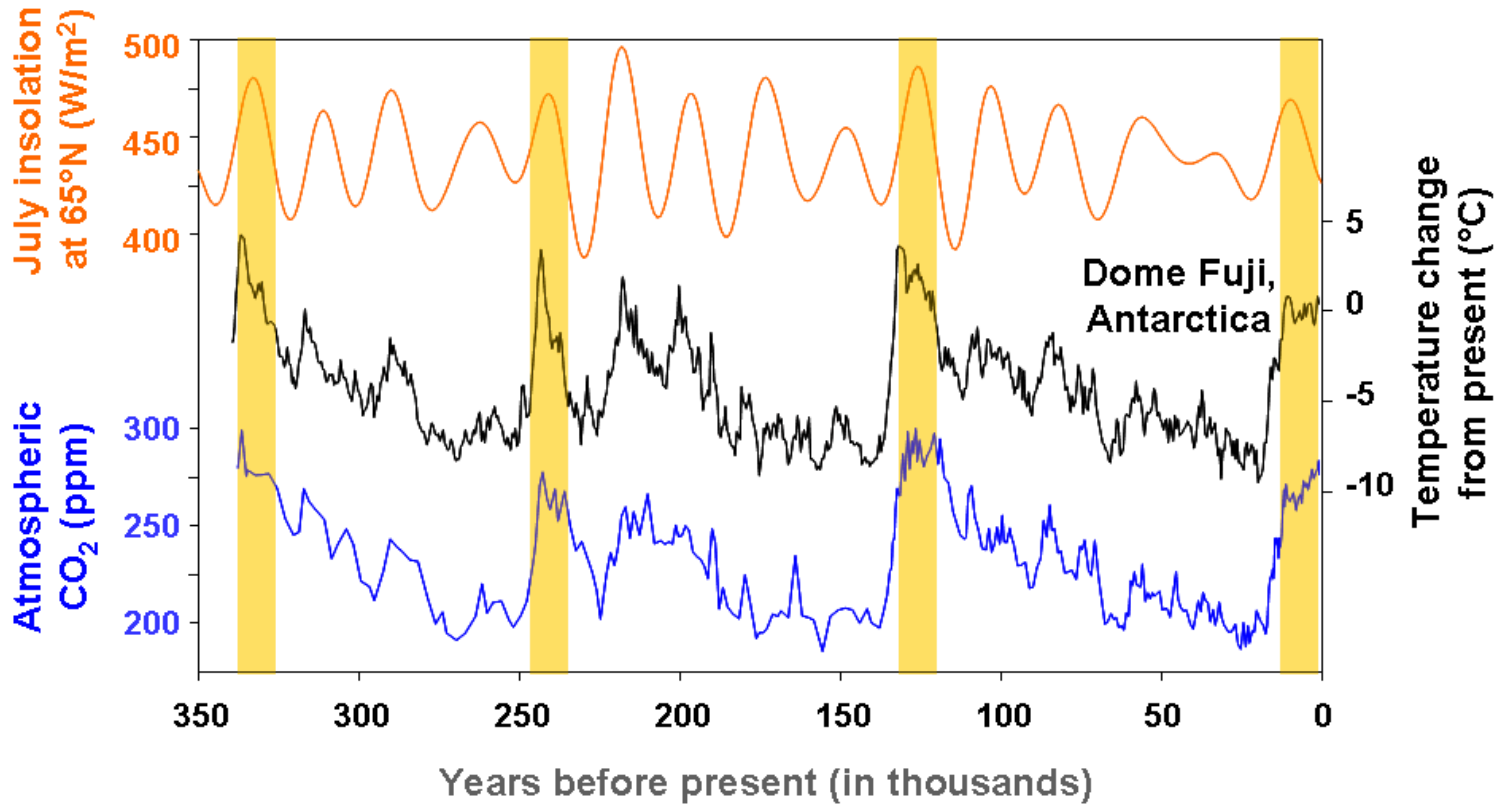
How the regulation works...

$$F_i = \begin{cases} 1 - \lambda(A_i - R)^2 & \text{if } |A_i - R| < \lambda^{-\frac{1}{2}} \\ 0 & \text{otherwise} \end{cases} \quad (1)$$



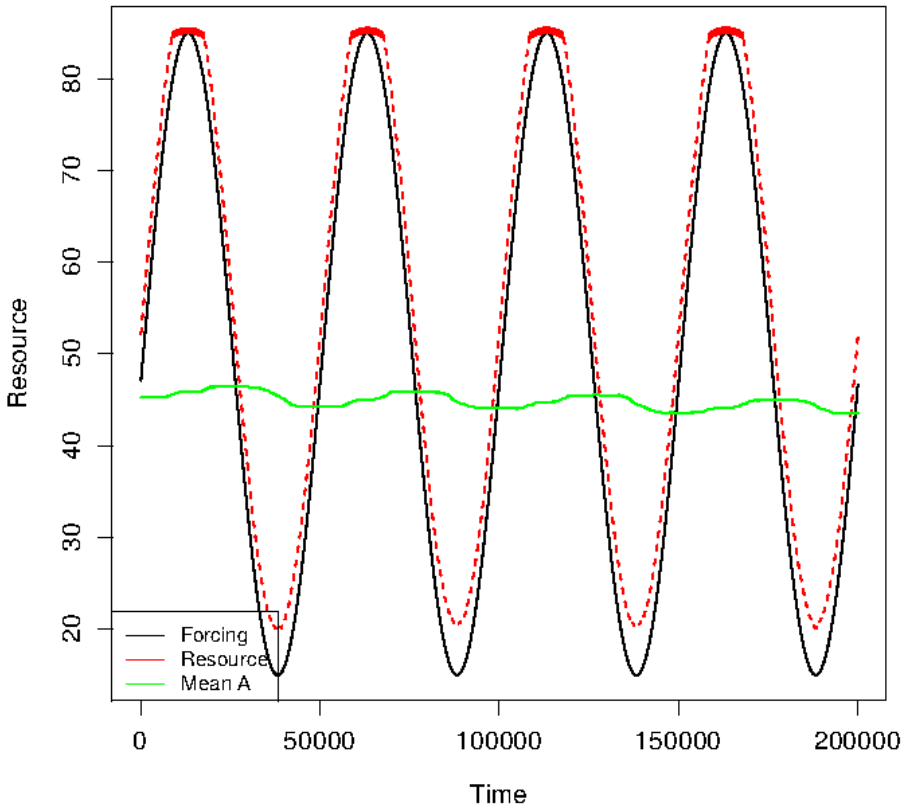
Modified from Dyke *et al.* 2007

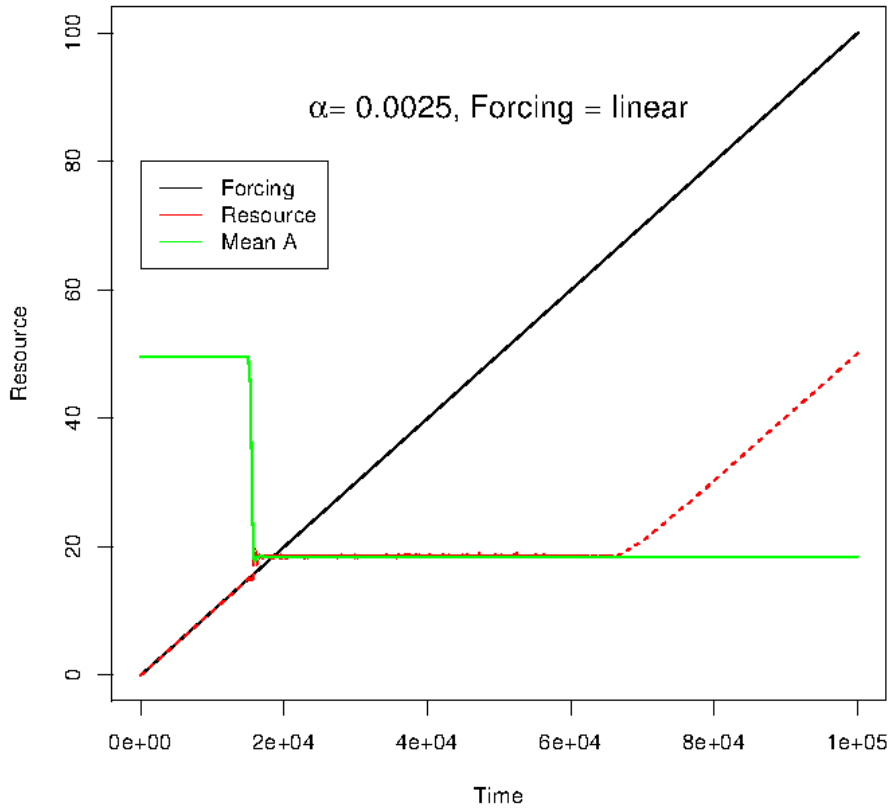
Now to the real world!



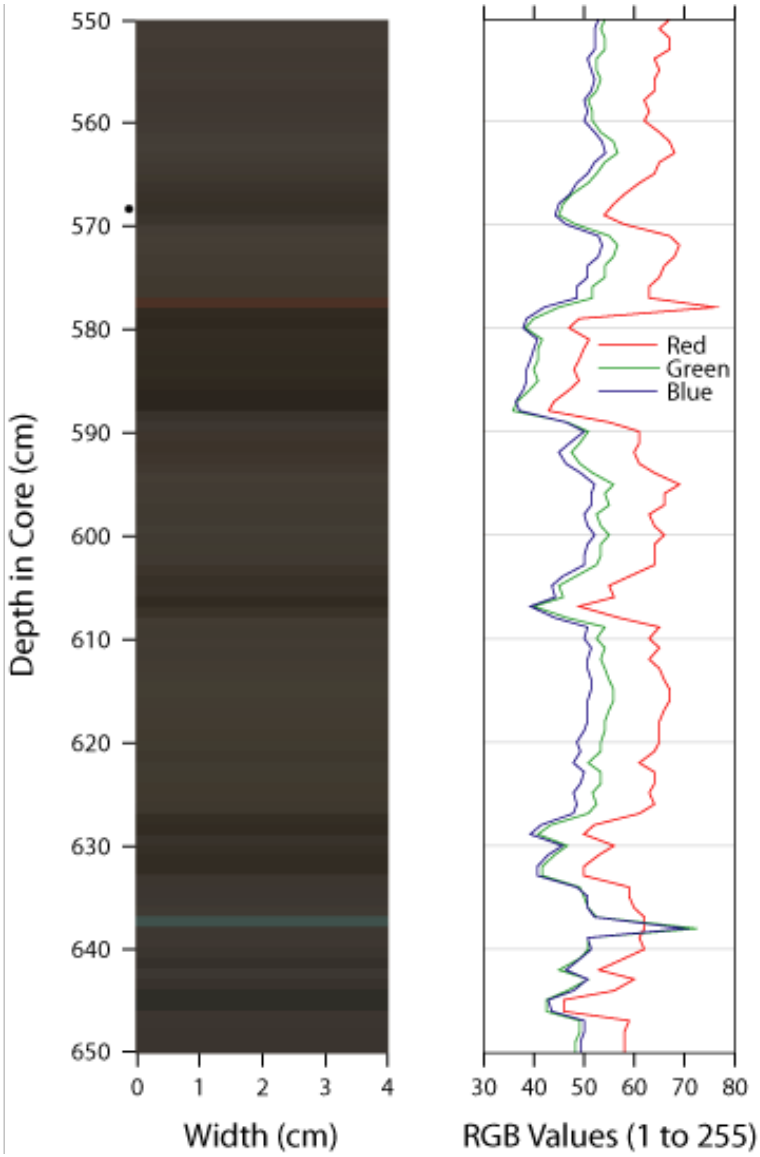
ncdc.noaa.gov (2008)

Lets look at the difference...



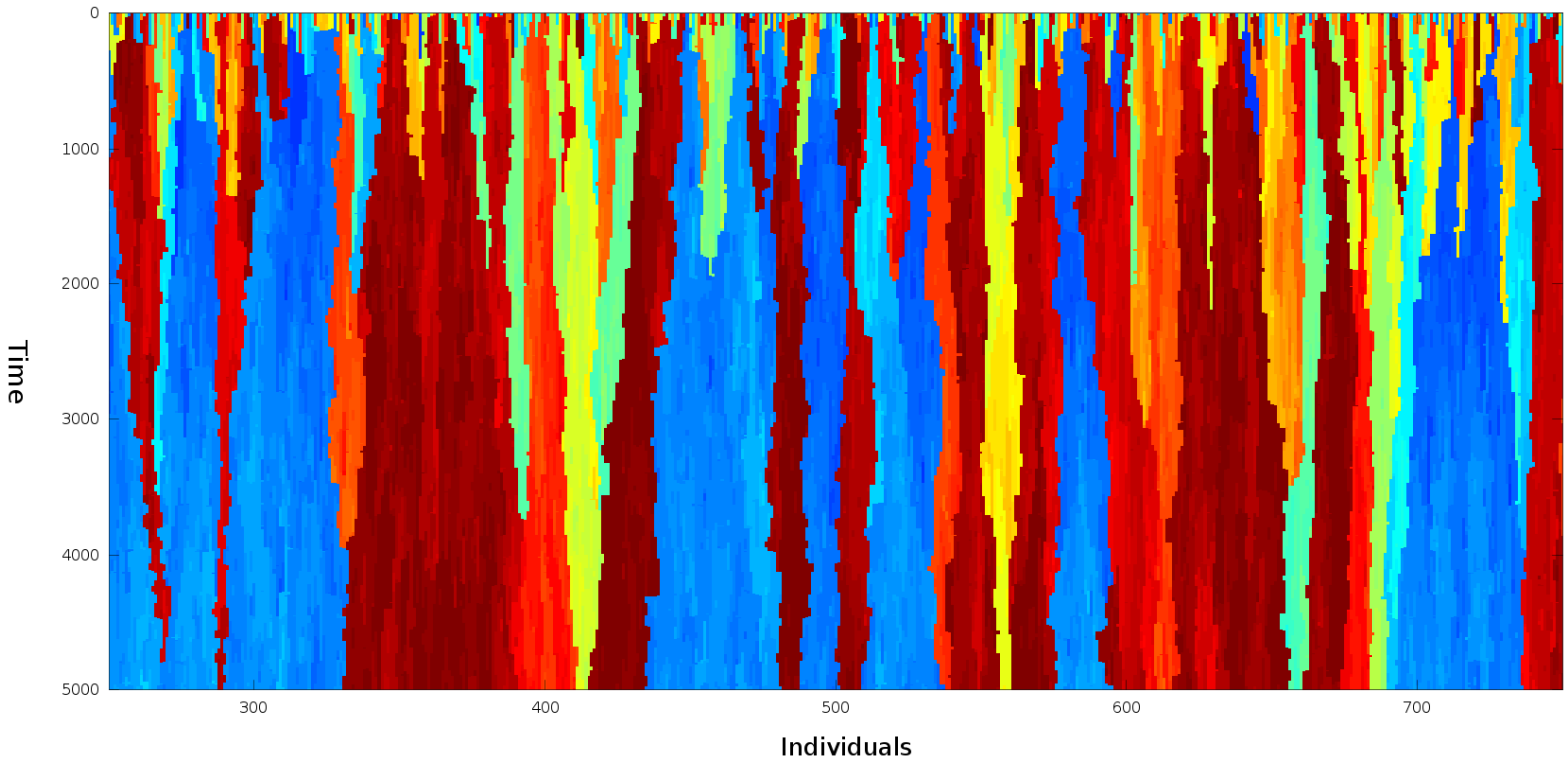


Lets look at some real world data...

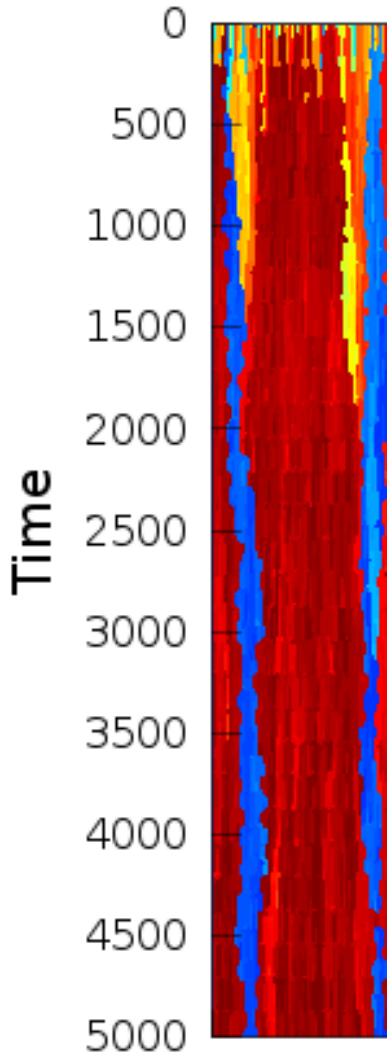


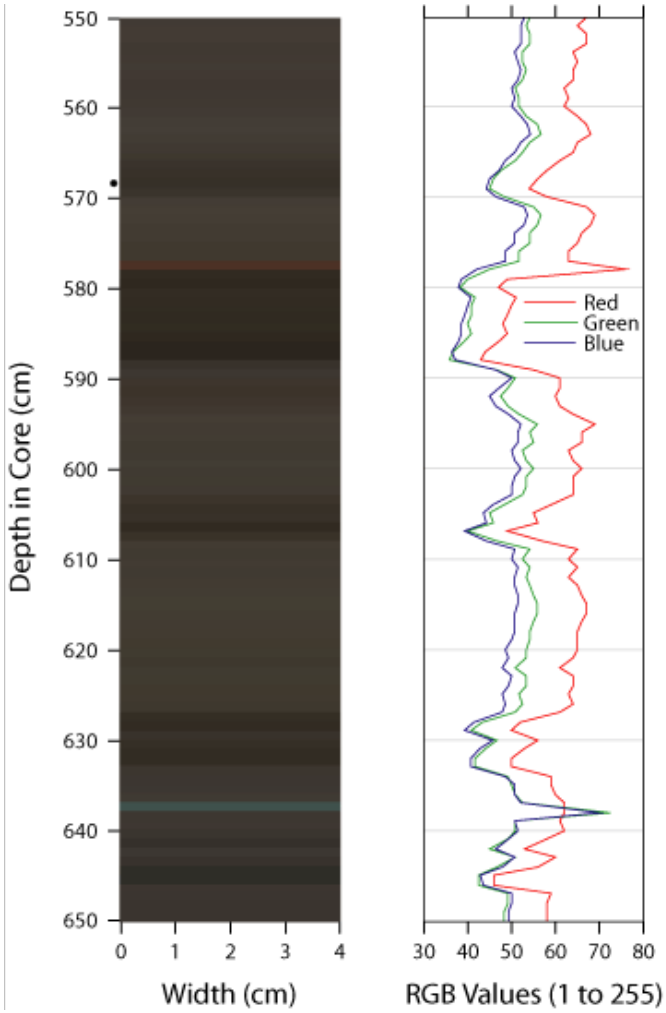
- Core samples of oceanic sediment
  - Global coverage: species migration with changing climate/resource
  - Used for temperature reconstruction etc.
- Can we see similar behaviour in our model?

Lets look at the difference...
System memory and niche construction!



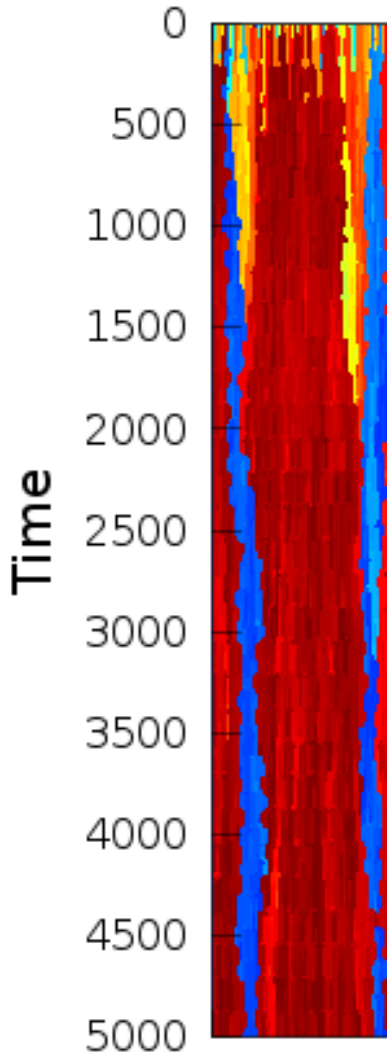
Lets look at the difference... System stabilisation!

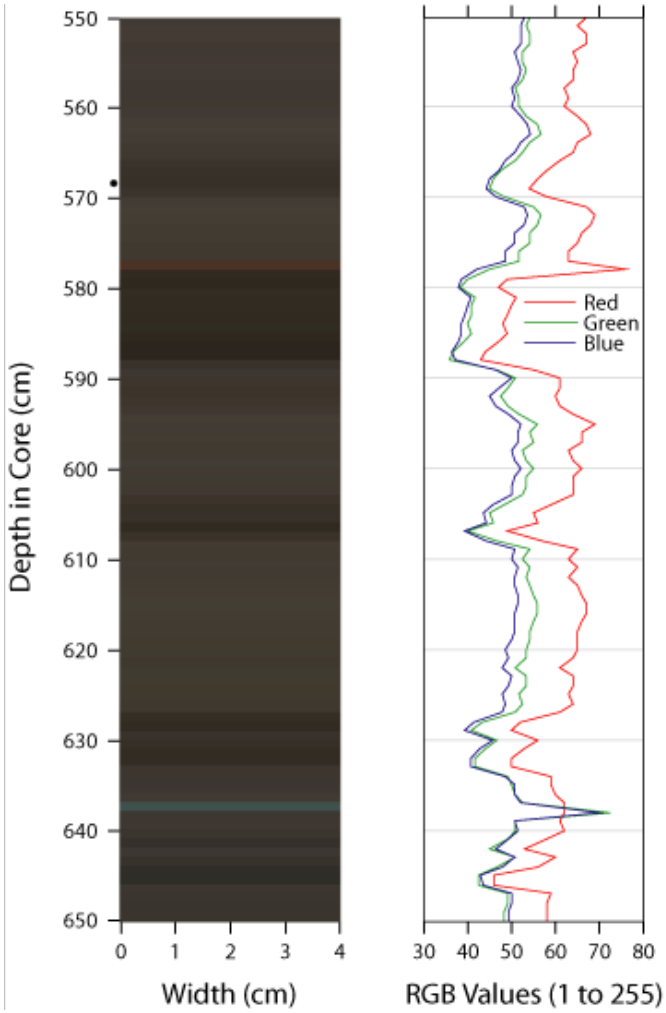




Abstract theories about niche construction suddenly become usefull!

Lets look at the difference... System stabilisation!





Abstract theories about niche construction suddenly become usefull!

Science is much more fun if you talk to people

Importance of collaboration: Talk to people!
You can't be an expert on everything...

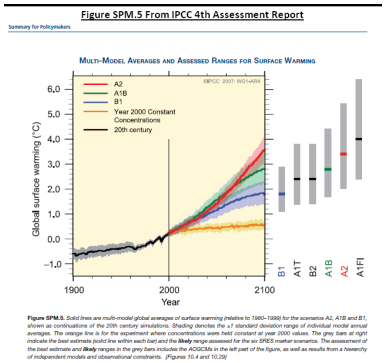
Hint:

You're at a **multi-disciplinary** conference; There will be wine later...

Examples of datasets from the natural world

NOAA NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
UNITED STATES DEPARTMENT OF COMMERCE

AVISO



NOAA

The Paleobiology Database
Revealing the history of life

Science 27 April 2001:
Vol. 292 no. 5517 pp. 686-693
DOI: 10.1126/science.1059412

‹ Prev | Table of Content

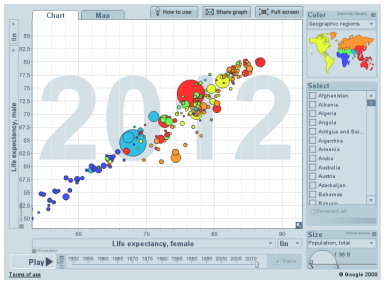
CHRONOS

REVIEW

Trends, Rhythms, and Aberrations in Global Climate 65 Ma to Present

James Zachos[1,*], Mark Pagani[1], Lisa Sloan[1], Ellen Thomas[2,3], Katharina Billups[4]

Examples of datasets from the rest of the world



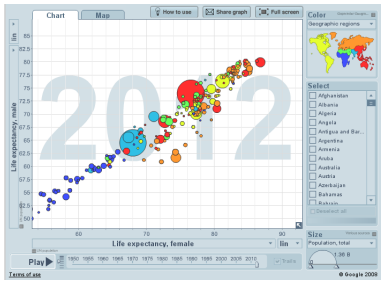
Economic and Social Research Council
Shaping Society

Others: Financial times and Bloomberg

Take home message

Consider the data; Talk to people; Do awesome science!

Examples of datasets from the rest of the world



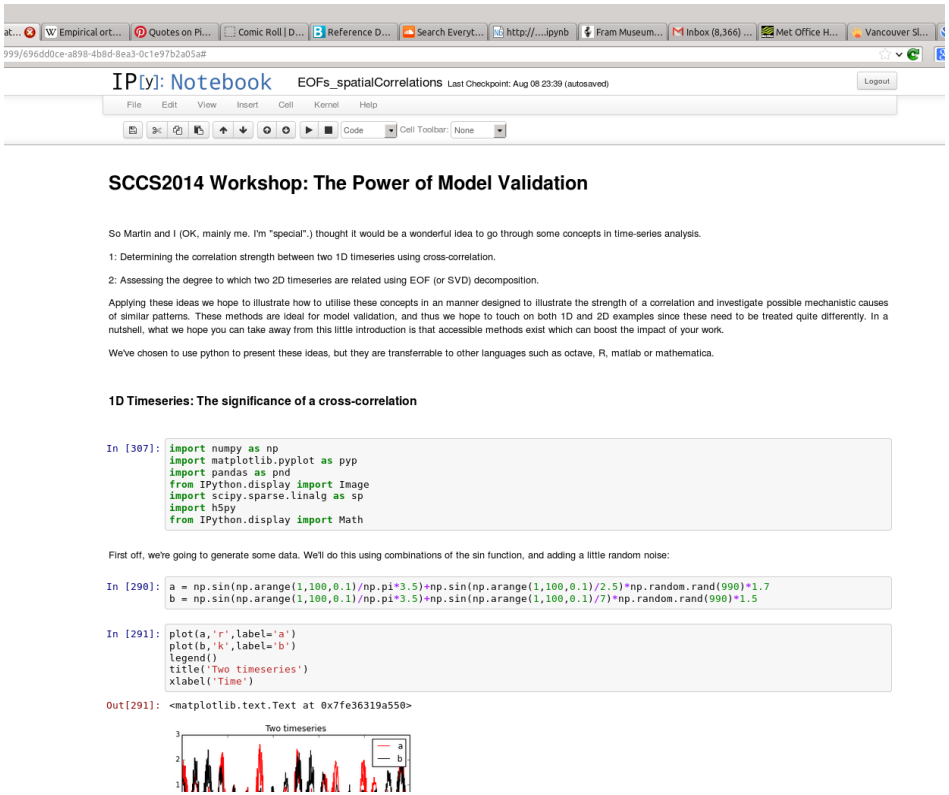
Economic and Social Research Council
Shaping Society

Others: Financial times and Bloomberg

Take home message

Consider the data; Talk to people; Do awesome science!

Now we have the idea and some data, what do we do...?!



IP[y]: Notebook EOFs_spatialCorrelations Last Checkpoint: Aug 08 23:39 (autosaved)

File Edit View Insert Cell Kernel Help

SCCS2014 Workshop: The Power of Model Validation

So Martin and I (OK, mainly me. I'm "special".) thought it would be a wonderful idea to go through some concepts in time-series analysis.

1: Determining the correlation strength between two 1D timeseries using cross-correlation.

2: Assessing the degree to which two 2D timeseries are related using EOF (or SVD) decomposition.

Applying these ideas we hope to illustrate how to utilise these concepts in an manner designed to illustrate the strength of a correlation and investigate possible mechanistic causes of similar patterns. These methods are ideal for model validation, and thus we hope to touch on both 1D and 2D examples since these need to be treated quite differently. In a nutshell, what we hope you can take away from this little introduction is that accessible methods exist which can boost the impact of your work.

We've chosen to use python to present these ideas, but they are transferrable to other languages such as octave, R, matlab or mathematica.

1D Timeseries: The significance of a cross-correlation

```
In [307]: import numpy as np
          import matplotlib.pyplot as pyp
          import pandas as pnd
          from IPython.display import Image
          import scipy.sparse.linalg as sp
          import h5py
          from IPython.display import Math
```

First off, we're going to generate some data. We'll do this using combinations of the sin function, and adding a little random noise:

```
In [290]: a = np.sin(np.arange(1,100,0.1)/np.pi*3.5)+np.sin(np.arange(1,100,0.1)/2.5)*np.random.rand(990)*1.7
          b = np.sin(np.arange(1,100,0.1)/np.pi*3.5)+np.sin(np.arange(1,100,0.1)/7)*np.random.rand(990)*1.5
```

```
In [291]: plot(a,'r',label='a')
          plot(b,'k',label='b')
          legend()
          title('Two timeseries')
          xlabel('Time')
```

```
Out[291]: <matplotlib.text.Text at 0x7fe36319a550>
```


Widely applicable Timeseries analysis in 1D and 2D in python: Significance of correlations and EOF analysis.