

Informing data models in ecology: Which layers should I use in my biodiversity model?

Kristen J Williams^{1,3}, Michael P Austin¹, Simon Ferrier¹, Janet Stein², Lee Belbin³
¹CSIRO Ecosystem Sciences, ²ANU Fenner School, ³Atlas of Living Australia

ESA2011 Hobart, 21 November 2011, 13:30-13:45



Informing data models in ecology: outline

1

- Why we need to clarify which layers are useful

2

- Outlining the data model and case study

3

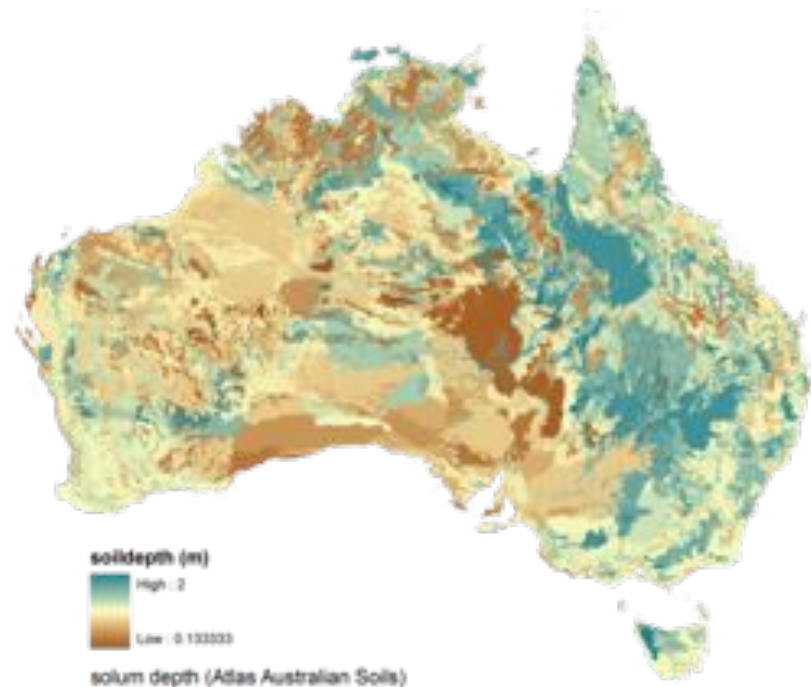
- Principles for selecting environmental layers

Why we need to clarify which layers are useful

- Digital environmental data is a necessary input to a wide range of ecological modelling approaches
- Deciding on which layers to include is an ongoing challenge
- Open access to models and data, such as through the Atlas of Living Australia (www.ala.org.au), highlight the need for more general guidance on which layers to use for a given purpose



Eucalyptus delegatensis canopy



Online data portals provide ready access to data for research and planning: <http://www.ala.org.au/>

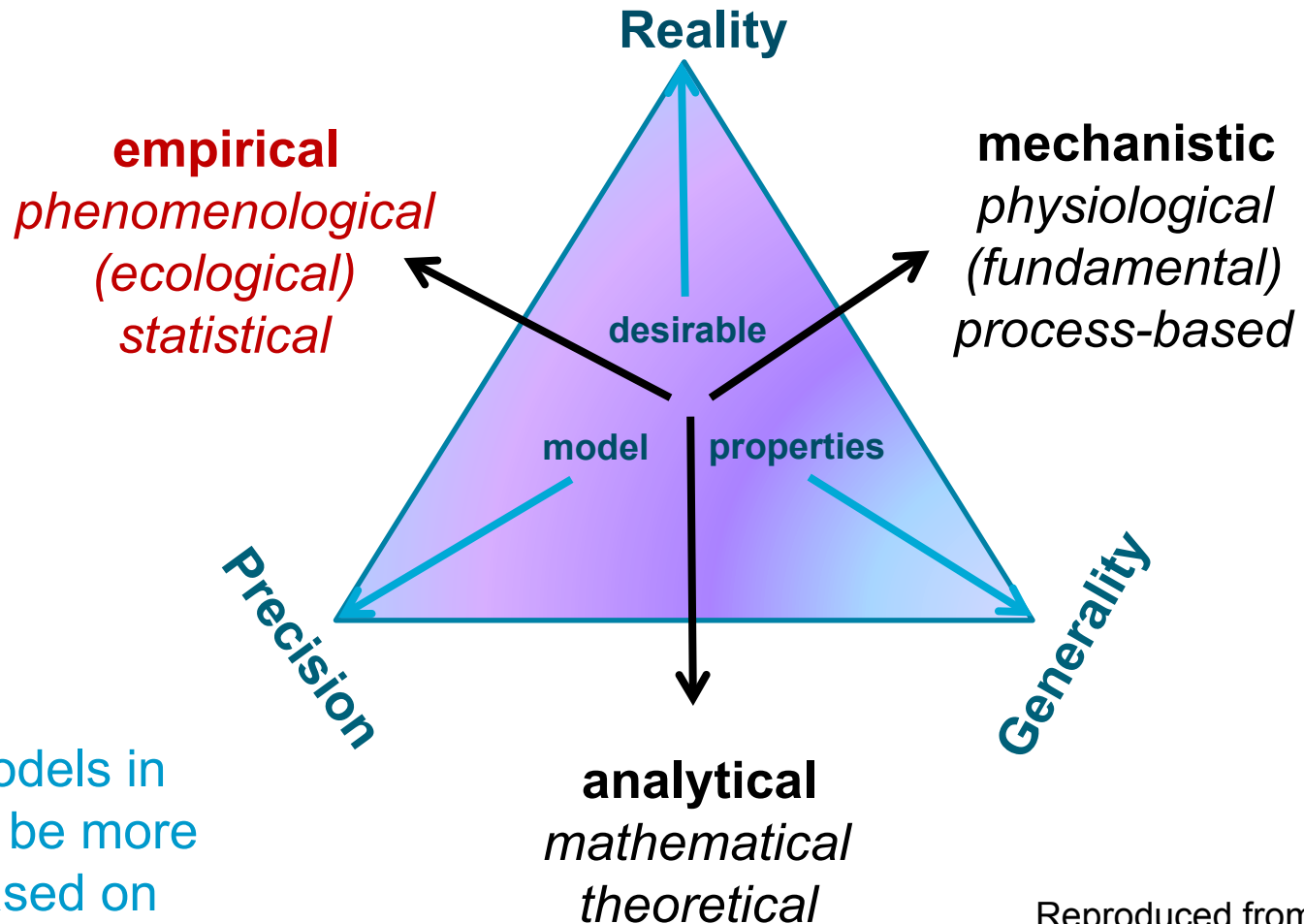
The screenshot shows the Atlas of Living Australia website. At the top left is the logo for 'ATLAS of LIVING AUSTRALIA' with the tagline 'sharing biodiversity knowledge'. The navigation menu includes 'Home', 'Explore', 'Tools', 'Share', 'Support', 'Contact Us', 'About the Atlas', 'My Profile', and 'Log in'. A search bar labeled 'Search the Atlas' is positioned in the top right. Below the navigation is a main heading: 'Explore information on species, maps, collections and regions'. On the left is a sidebar menu with 'Explore' selected, containing links for 'Your Area', 'Regions', 'Natural History Collections', 'Species Maps', and 'Themes'. The main content area features a 'Flora & fauna' section with another 'Search the Atlas' search bar and the text 'Find information on Australia's biodiversity'. Below this are four interactive cards: 'Your area' (Discover what lives around you), 'Regions' (Explore a state or territory's biodiversity), 'Map Species' (Map organisms and environmental data for the complete picture), and 'Collections' (Explore Australia's natural history collections and view digitised records). The 'Map Species' card is highlighted with a red border. Below these cards is a 'Themes' section with a paragraph explaining that themes provide stories of general interest to the Australian public about particular groups of organisms. Underneath, it lists 'Current themes within the Atlas of Living Australia' and provides a link to 'Continue reading'. A row of seven small images represents these themes: 'Wattles', 'Iconic Species', 'Shorebirds', 'Biodiversity Events', 'Thrips', 'Arts', and 'Biodiversity Case Studies'. The footer contains a navigation bar with 'Home | Explore | Tools | Share | Support | Contact Us | About the Atlas | Log in', a Creative Commons Attribution 3.0 Australia License logo, and links for 'Terms of Use', 'Citing the Atlas', 'Privacy Policy', and 'Site Map'. A note at the bottom states: 'This site is licensed under a Creative Commons Attribution 3.0 Australia License. Provider content may be covered by other Terms of Use'.

ALA “Map Species” Spatial Analysis Tools:

<http://www.ala.org.au/explore/species-maps/>

- **Sampling:** attribute species occurrence data with values of environmental and contextual data layers
- **Filtering:** use values of environmental or contextual layers to define an envelope and subset species occurrence data
- **Scatter plot:** view the species occurrence data in environmental space for any two variables and an area of interest
- **Prediction:** use MaxEnt to model the relationship between species occurrence data and selected environmental layers
- **Classification:** classify selected environmental layers into domains (the ALOC algorithm from PATN)

Desirable properties and types of models



Empirical models in ecology can be more general if based on meaningful parameters

Reproduced from Fig 2 in
Guisan & Zimmermann
(2000) *Ecological Modelling*
135:147-186.

Three components of a biodiversity model

Ecological Model

- What ecological theory is assumed or tested?

Data Model

- Are there any limitations imposed by the nature of the data used?

Statistical Model

- Are the statistical procedures and methods used compatible with ecological theory?

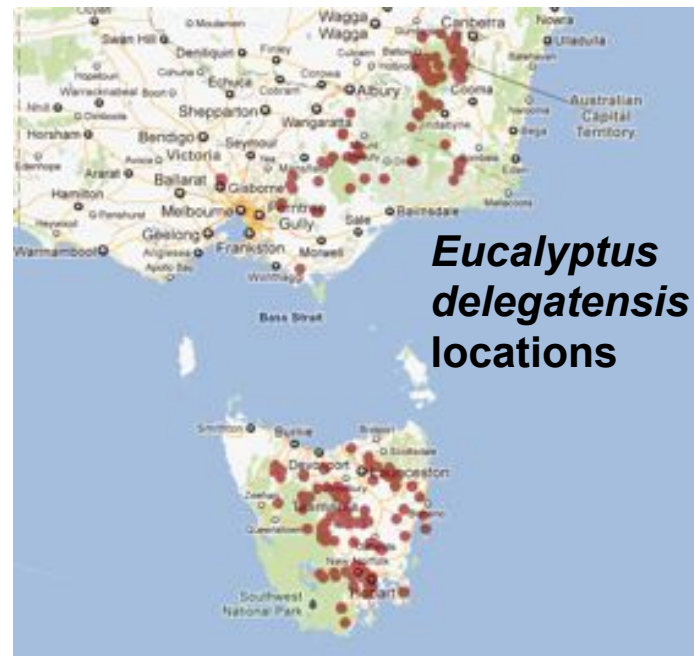
Austin, M. P. 2002. *Ecological Modelling*, 157: 101-118.

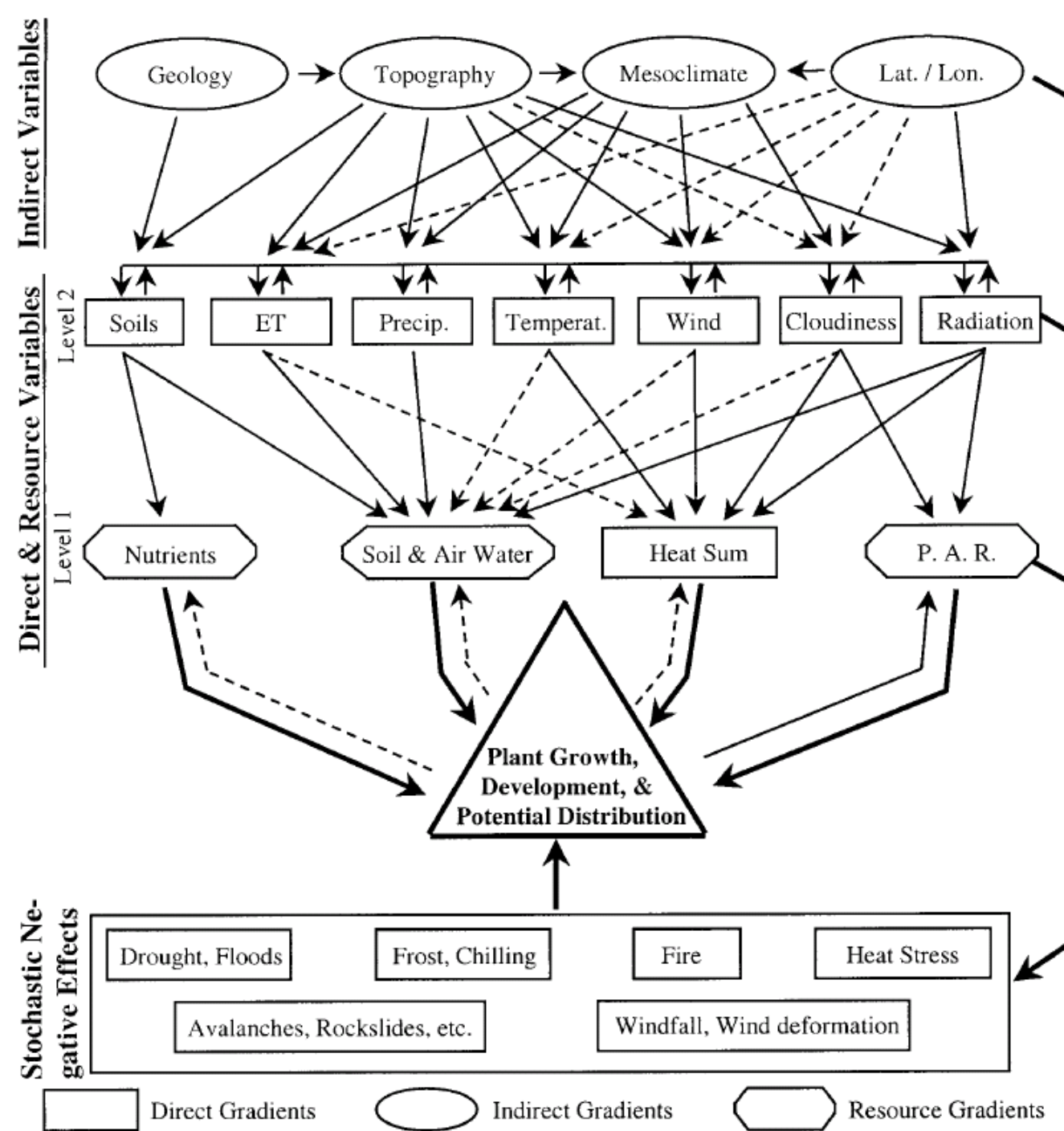
Austin, M. 2007. *Ecological Modelling*, 200: 1-19.

Basic components informing the choice of environmental layers for a data model

- **Purpose** (location, study species, response variable) – informs the extent and resolution of environmental layers to compile
- **Applicable ecological theory** – informs the scope and type of environmental layers to compile and evaluate
- **Applicable statistical model** – informs the format of layers and procedures for selecting environmental layers to include/test

Best demonstrated using a case study, plant species





An attempt to structure the problem for plants

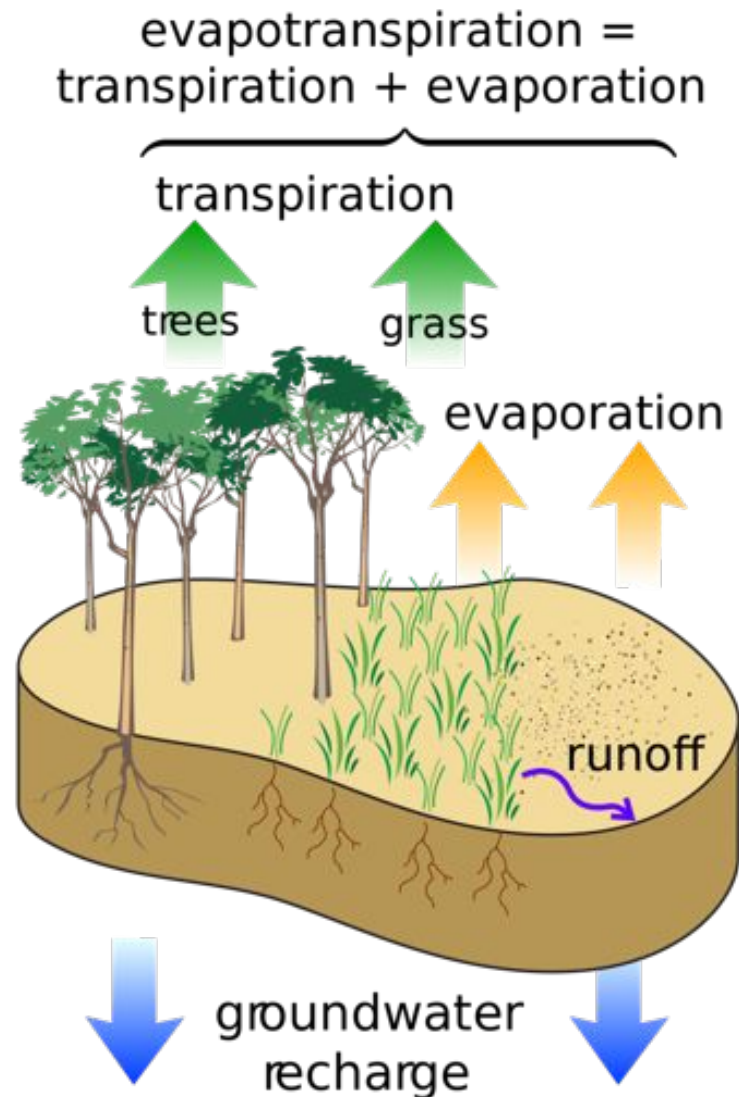
But lots of potential factors or their proxies to compile and consider in a model, including outputs from biophysical process models, with variable quality and correlations

Over 100 layers to choose from in the ALA

A wide range of National 1km gridded environmental variables in the ALA spatial portal

- **Climate variables**
 - Mean conditions of evaporation, precipitation, minimum and maximum temperature, wind, humidity, solar radiation (ANUCLIM derivatives)
- **Disturbance regimes**
 - Fire frequency and mean climatic extremes (proxies for drought, flood, frost, heat, etc)
- **Soil variables**
 - Soil depth, clay%, bulk density, hydraulic conductivity, structure, water holding capacity, calcrete, nutrient status, etc (derivatives of the Atlas of Australian soils)
- **Geoscience variables**
 - Geological age and inherent fertility (derivatives of the 1:1M National Geology), and geophysics - gravity and magnetics
- **Terrain variables (DEM derivatives)**
 - Slope, aspect, elevation diversity, topographic position, wetness indices, multi-resolution indices, etc

The role of biophysical process models: e.g., Environmental water predictors

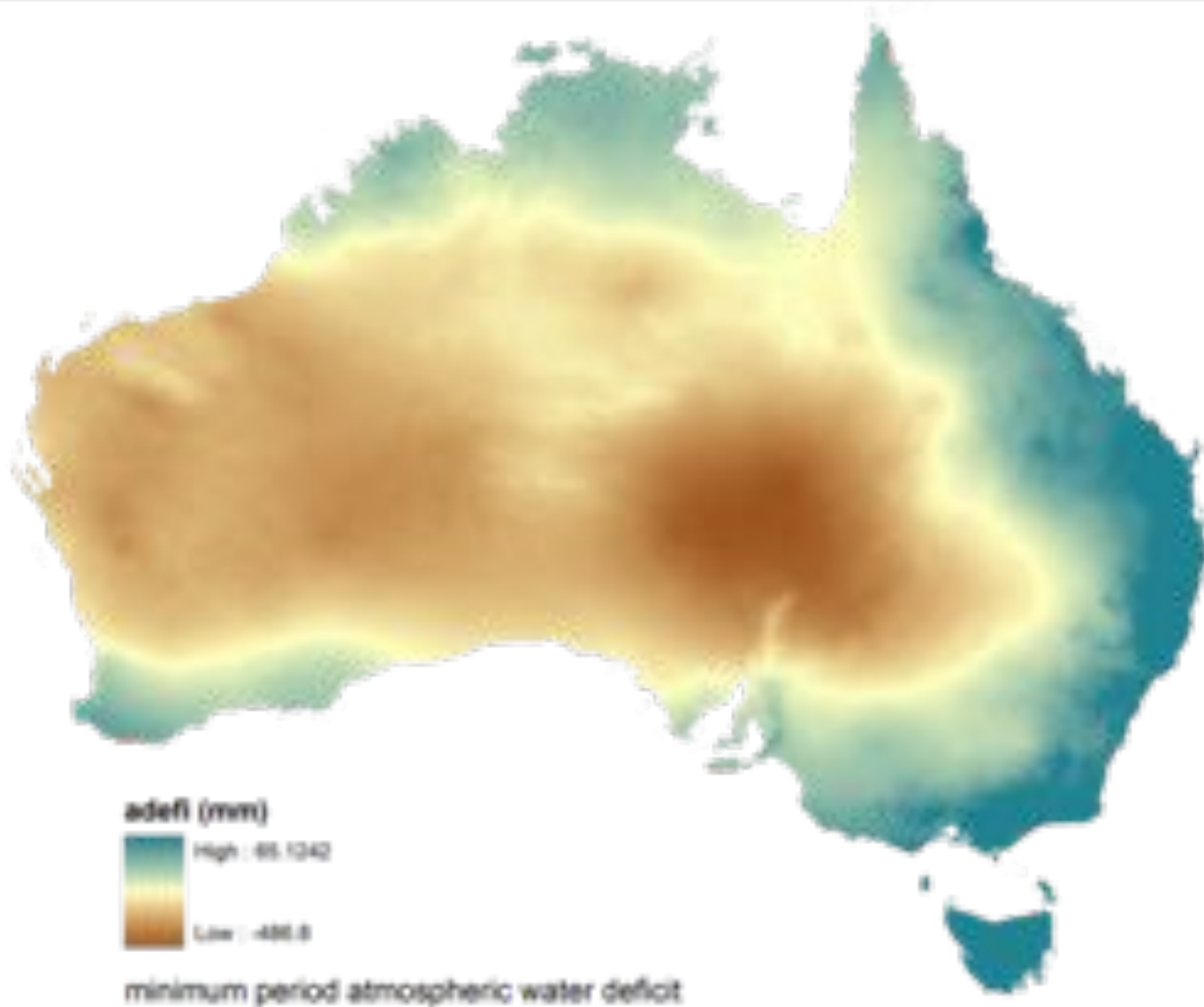


It is good practice to develop physiologically-relevant predictors

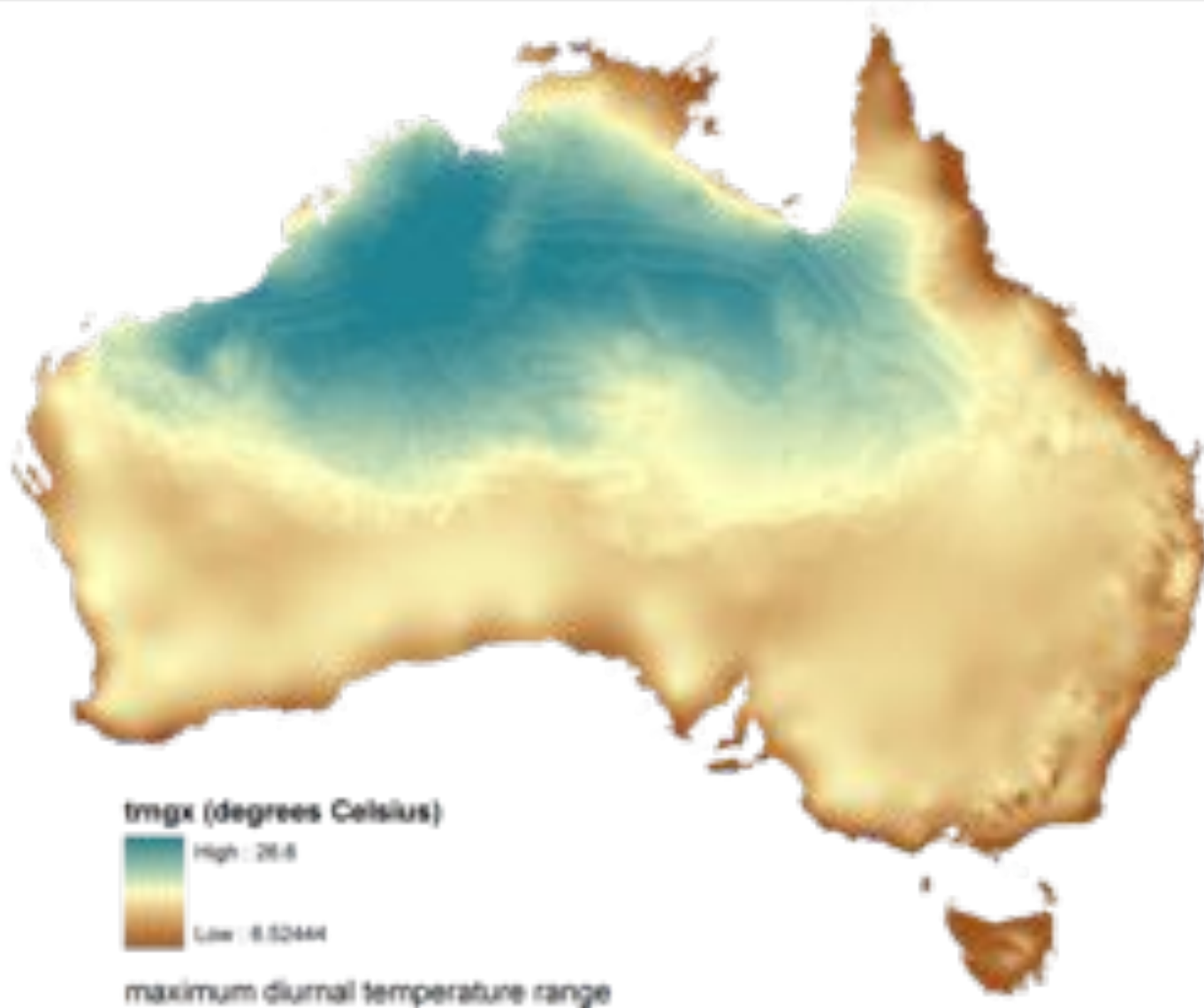
Water balance models potentially reduce a large number of predictors to just a few

But the outputs can be confounded by poor resolution and inaccuracy of soil parameters

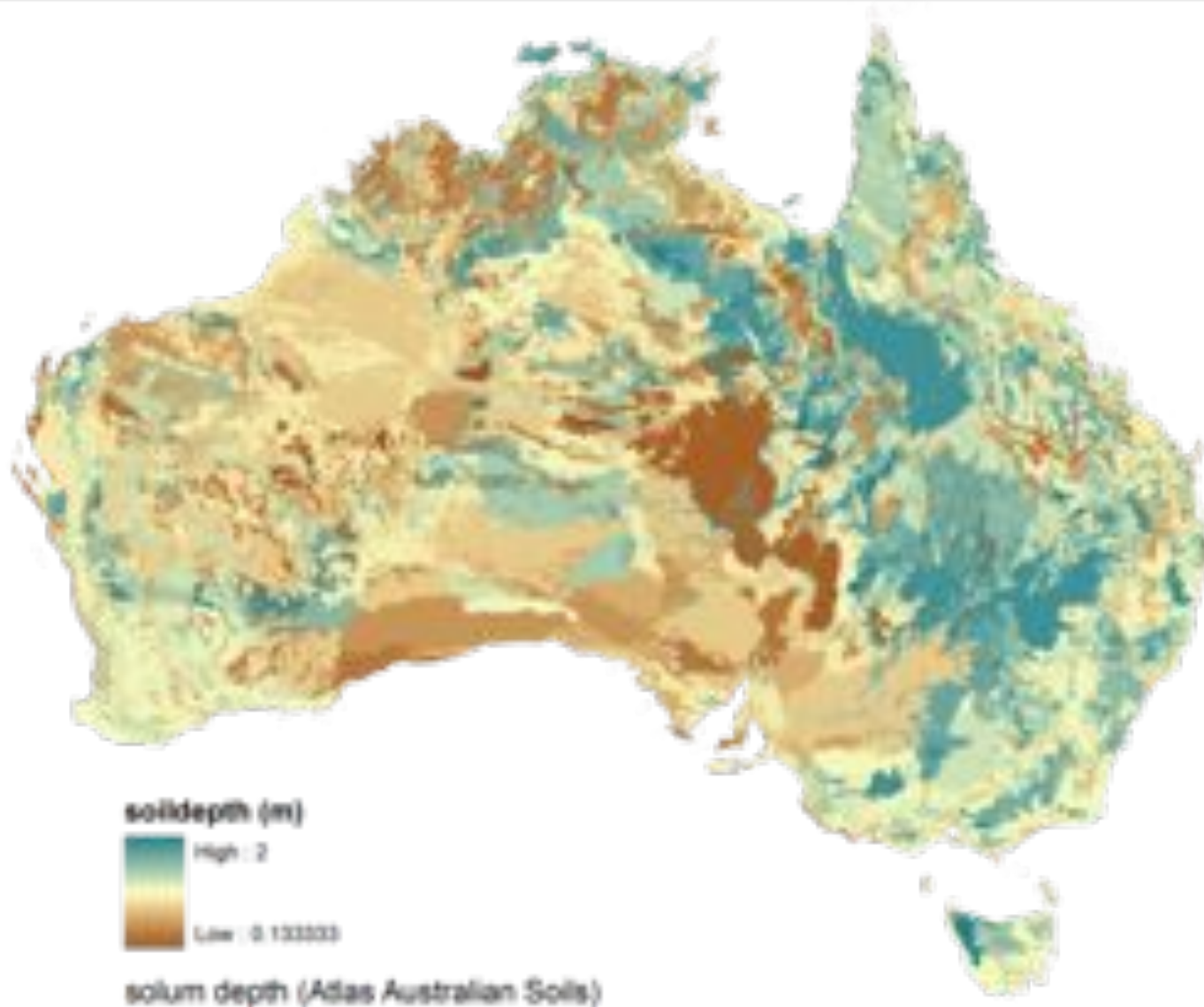
Example variables: water (P-E)



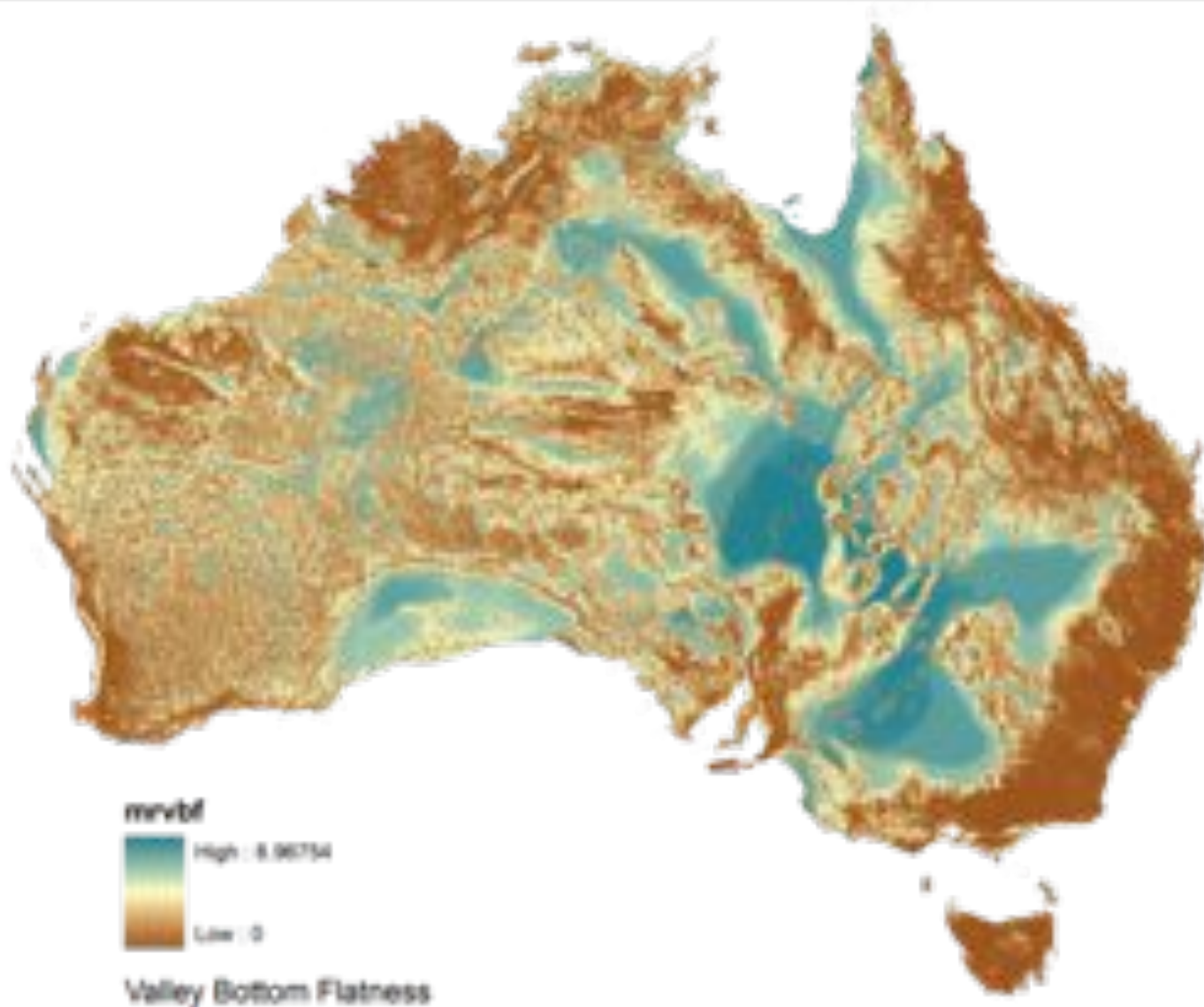
Example variables: diurnal temperature range



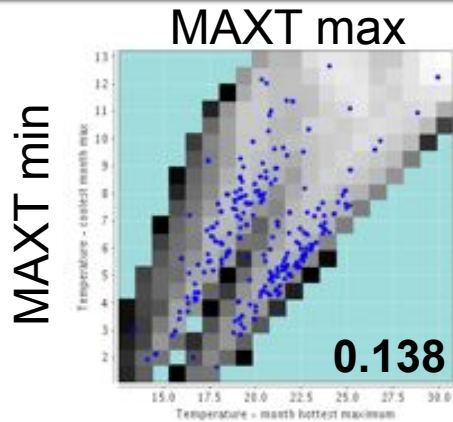
Example variables: soil depth



Example variables: DEM multi-resolution indices

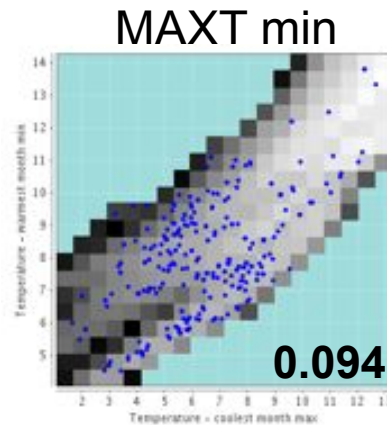
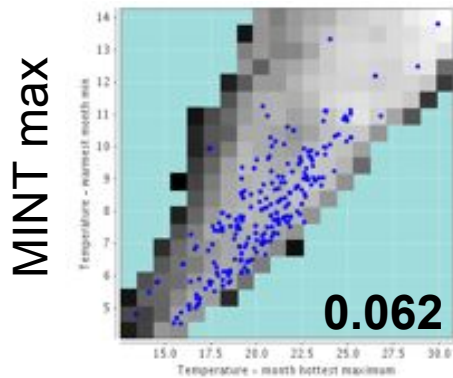


Example scatter plots from ALA spatial portal – minimum and maximum temperatures



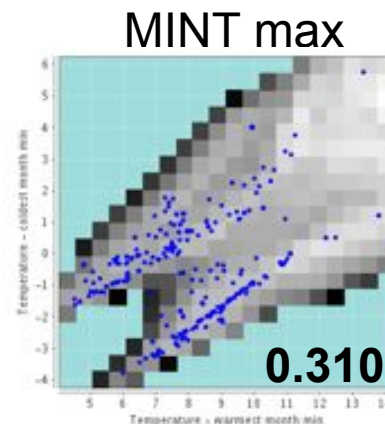
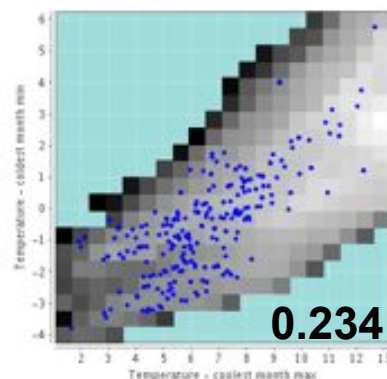
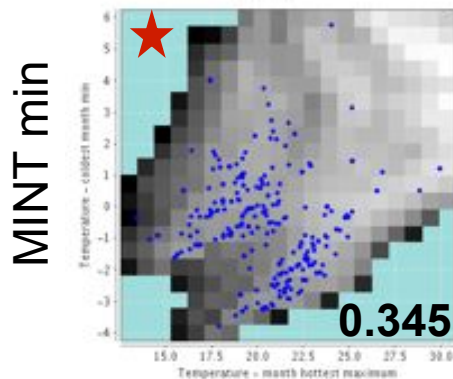
Blue points: species (*Eucalyptus delegatensis*)

Grey scale: extent of environment in Australia
(white grids correspond to a large area, black grids represent a small area)



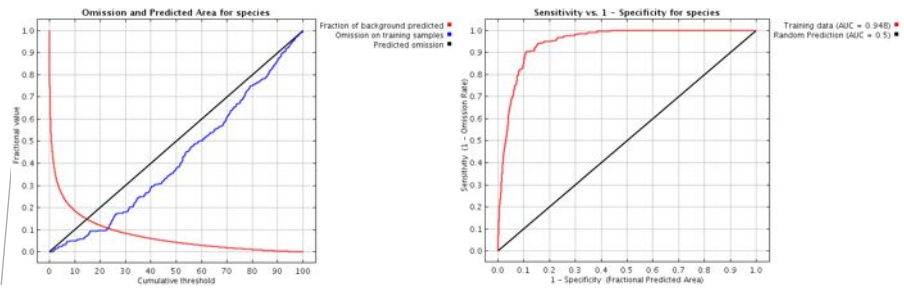
Blue space: beyond background environments

Numbers: inter-layer association
0 = identical, 1 = entirely different

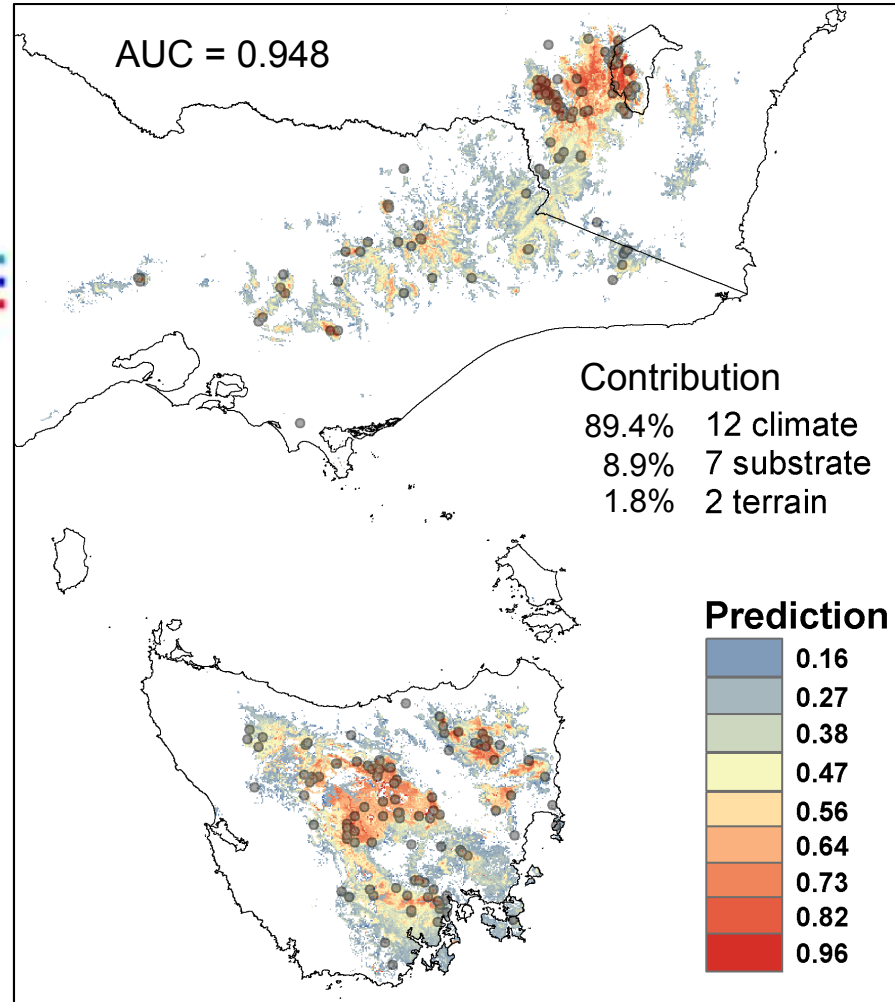


* Extreme ends of the gradient are more independent

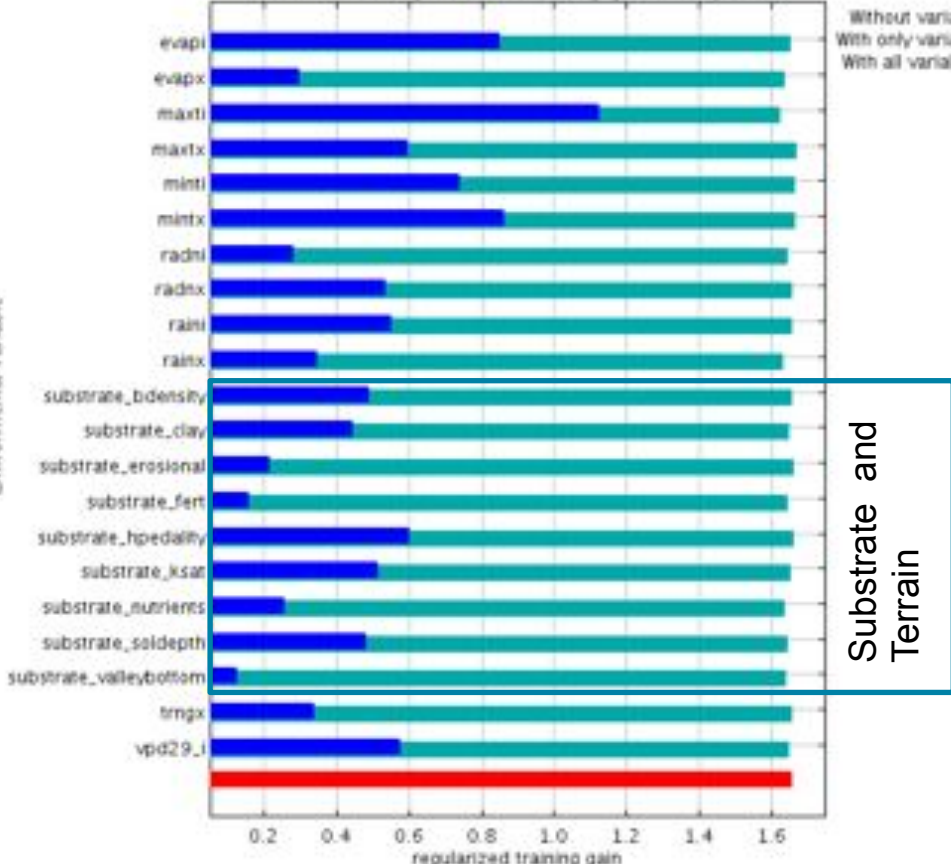
Example ALA output: MaxEnt model prediction



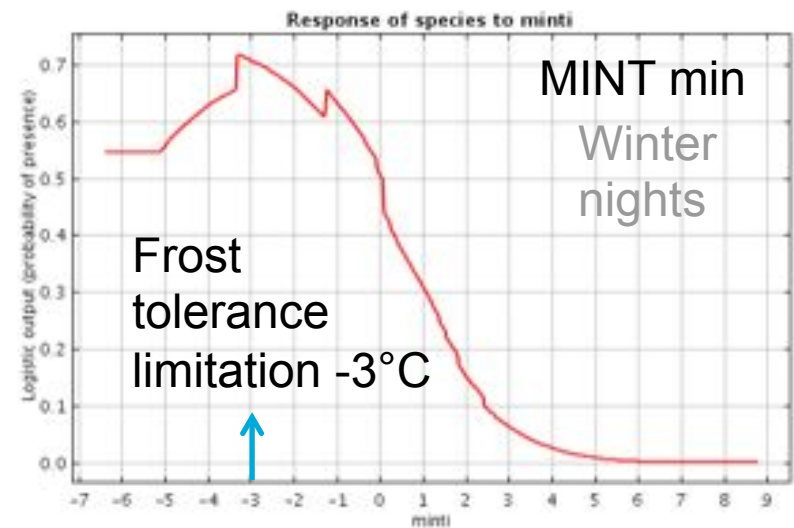
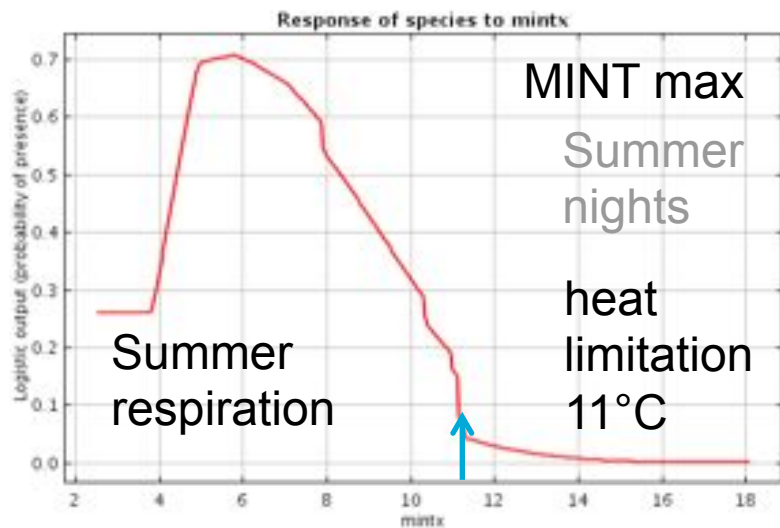
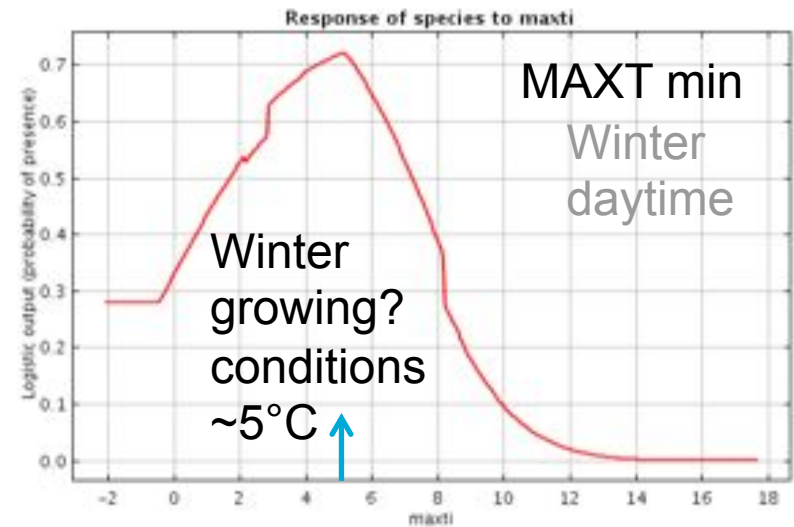
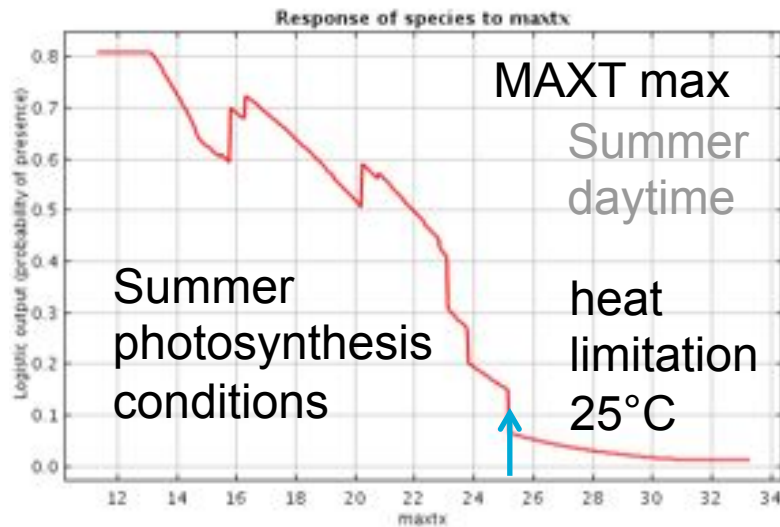
Maxent model: *E. delegatensis*



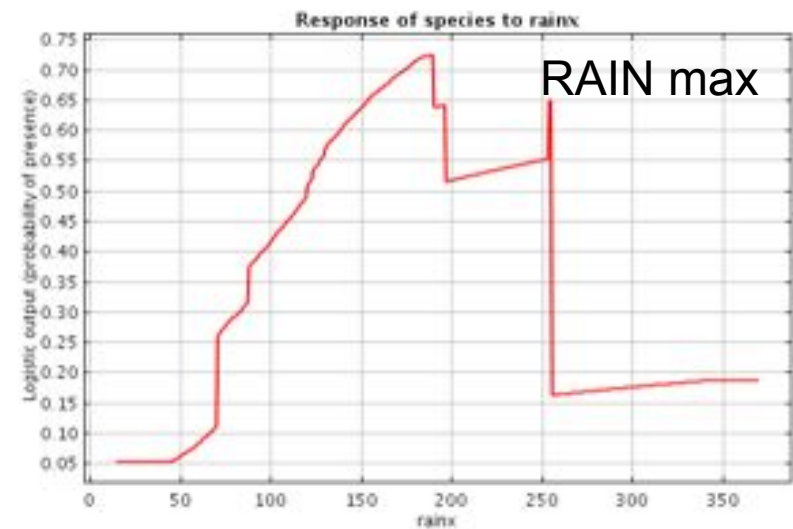
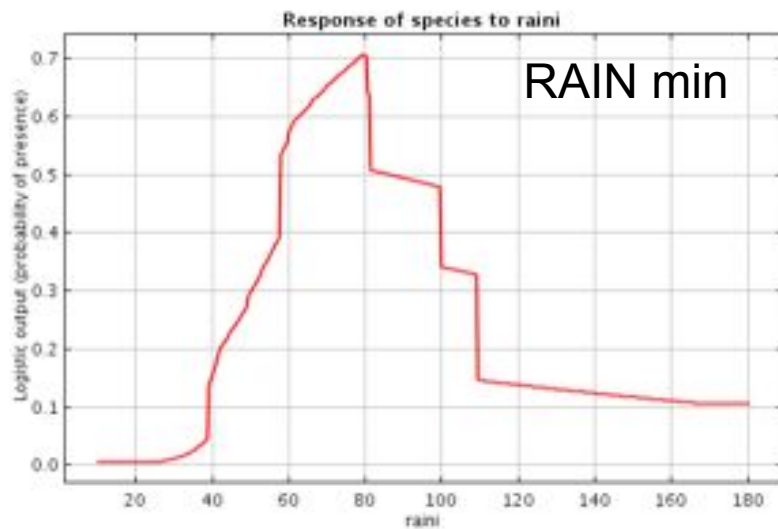
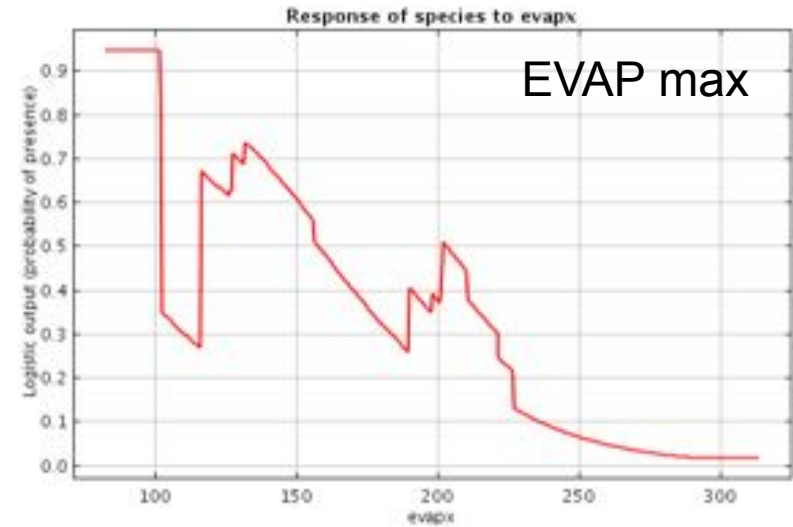
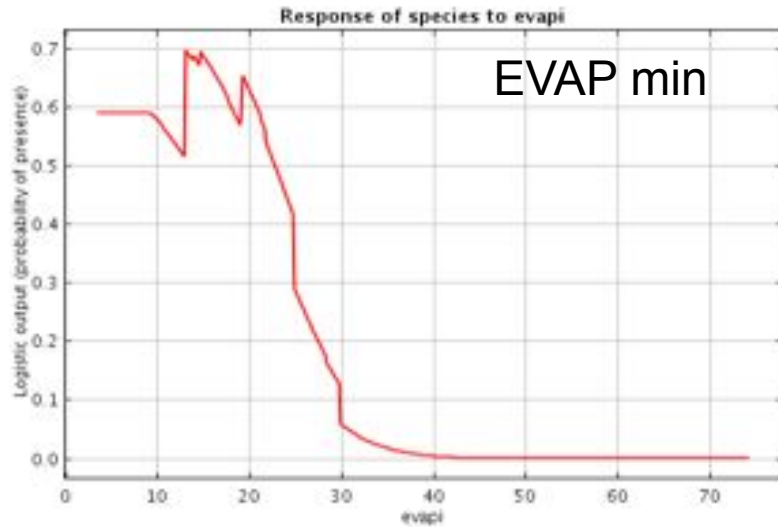
Jackknife of regularized training gain for species



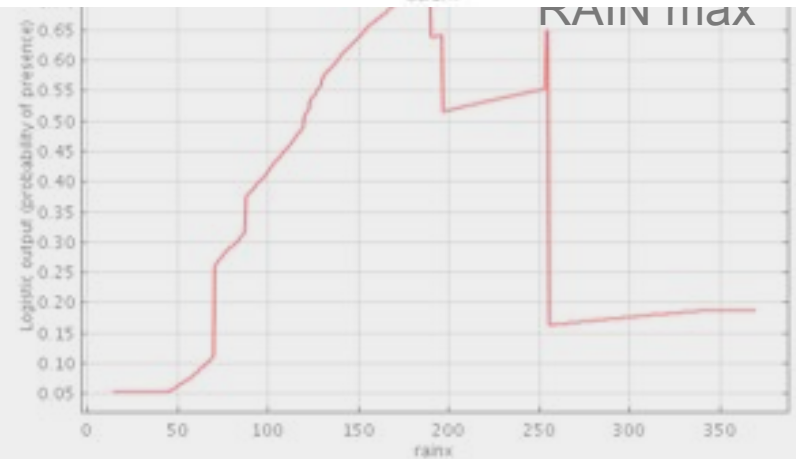
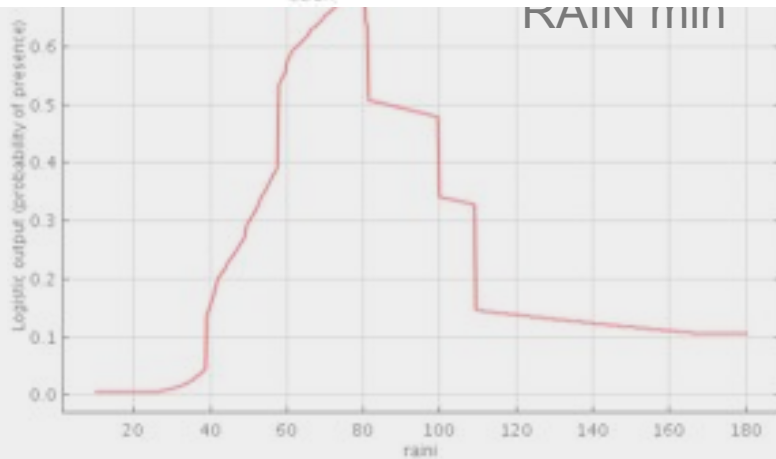
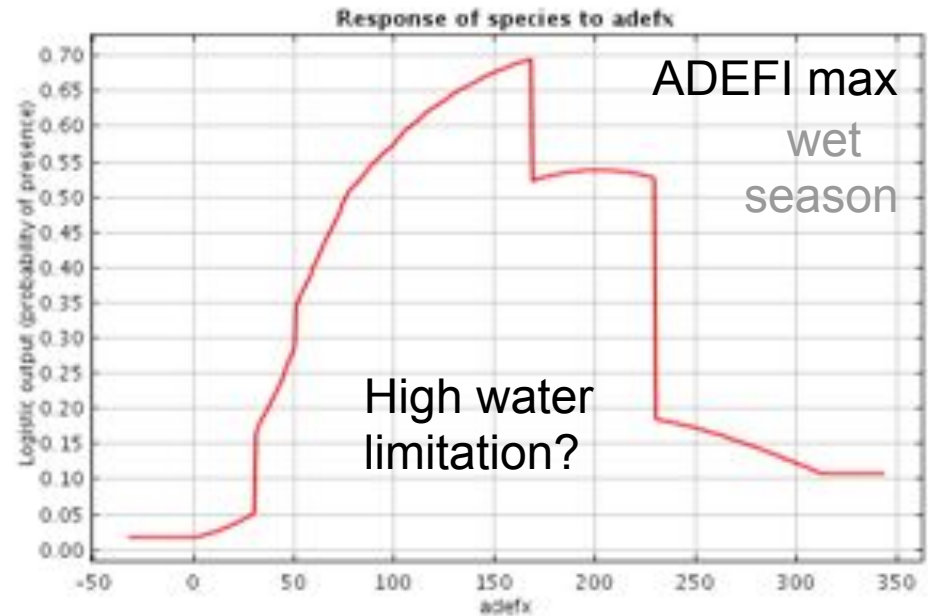
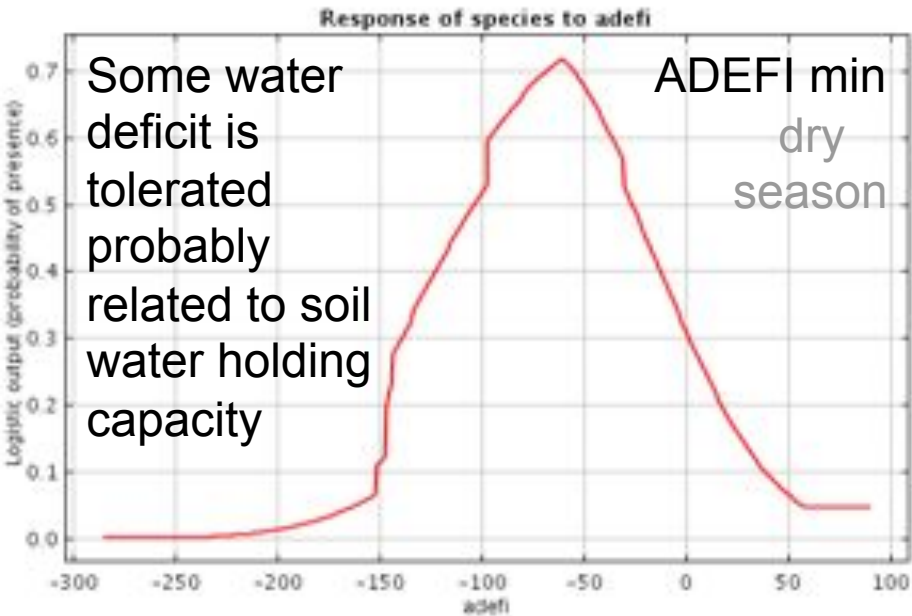
Indicative physiological parameters?



Rainfall and evaporation? Would atmospheric water deficit be more ecologically meaningful?



Rainfall and evaporation? Would atmospheric water deficit be more ecologically meaningful?



Principles for selecting environmental layers

- Adopt explicit ecological rationale based on theory (purpose)
- Note whether the variable is a direct or indirect driver of distribution patterns, proximal or distal to physiological process
- Use logic and scatter plots to explore correlation patterns between variables to understand origins and which are relatively independent
- Develop a hypothesis-driven framework for successively including variables in a model (initial set and supplementary to test residuals)
- Avoid combining variables that are self-excluding alternative sets (e.g. rainfall and evaporation, atmospheric water deficit)
- Iteratively revise your understanding of the environmental layers through the practice of model building, note effective combinations
- Visualise results in both environmental and geographic space and link to ecological rationale (potential physiological optima and limits)

Conclusions

- No simple solutions, but ecological theory adds structure to the process of identifying, evaluating and selecting layers
- Biophysical process models improve reality (more proximal predictors), but gains may be offset by inaccurate, incomplete or low resolution input parameters: use simple, interpretable variables
- Scatter plots and model testing with explicit subsets in a hierarchical or structured way, is an iterative learning process
- Above all – know how your layers were generated (metadata), critically evaluate their utility and encourage new developments
- This presentation has focused on terrestrial environments, but the same applies to other environments: marine and freshwater, etc

CSIRO Ecosystem Sciences

Kristen Williams

Ecological Geographer

Phone: 02 6246 4213

Email: kristen.williams@csiro.au

Web: www.csiro.au



Eucalyptus delegatensis canopy

www.csiro.au

Thank you



<http://www.ala.org.au/>

Contact Us

Phone: 1300 363 400 or +61 3 9545 2176

Email: enquiries@csiro.au Web: www.csiro.au

