Predictors in plant ecology analysis

Lae, 19th-20th February 2018

Presented by:
Fabio Attorre
University “Sapienza” of Rome

www.fao.org/forestry
Important Predictors for plants

Introduction

Typologies of predictors
- Regulators, Disturbances, Resources
- Proximal vs Distal
- Relationships between predictors

Important Predictors
- Temperature
- Water
- Nutrient substrate
- Light
- Biotic interactions
- Disturbances
- Topography and Land Use

Database of predictors
- WorldClim
- Digital Elevation Model
- Soil Grid
- Envirem

Collinearity
Ecologically sound reasoning for the choice of predictors in the Ecological modeling and analysis should be common practice, and the models and predictions should always be interpreted through the perspective of the set of predictors used.

A good strategy is to select variables that:

a) *are ecologically relevant*;

b) *are feasible to collect data on*;

c) *are closer to the mechanism (in the sequence resource-direct-indirect-proxy variables).*
Regulators, Disturbances, Resources

Typologies of predictors

- **REGULATORS** (modulating the organism physiology, gradual response) or **LIMITING FACTORS** (causing linear or step responses): factors controlling a species’ metabolism (e.g. low temperatures);

- **DISTURBANCES**: perturbations affecting environmental system (natural or human-induced);

- **RESOURCES**: compounds that can be consumed by organisms (e.g. radiations, water or nutrients for plants, prey for animals)
Typologies of predictors

**Proximal vs Distal**

Predictors can be

**Proximal:** direct physiological effect on plant growth or competition

**Distal:** no direct physiological effect on plant growth or competition

- e.g. Temperature, pH...
- e.g. altitude, longitude, latitude
Typologies of predictors

Relationships between predictors

From Mod et al. 2016
Important Predictors for plants

Most important Predictors for plants
(Mod et al. 2016; Körner 2014; Austin 2002)

- Temperature
- Water
- Nutrients
- Light
- Disturbances
- Biotic interactions

<table>
<thead>
<tr>
<th>Categories</th>
<th>Temperature</th>
<th>Water</th>
<th>Substrate</th>
<th>Radiation</th>
<th>Biotic Interactions</th>
<th>Disturbance</th>
<th>Topography</th>
<th>Land Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classes</td>
<td>Mean (annual, seasonal, monthly) temperature</td>
<td>Mean/summed (annual, seasonal, monthly) precipitation</td>
<td>pH, bedrock</td>
<td>Radiation, clouds</td>
<td>Variables related to other organisms</td>
<td>Geomorphological processes, fire</td>
<td>Slope, aspect, elevation,</td>
<td>Anthropogenic variables</td>
</tr>
<tr>
<td>Extreme temperatures</td>
<td>Extreme precipitation</td>
<td>Seasonality</td>
<td>Water balance</td>
<td>Soil moisture</td>
<td>Nutrients</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Important predictors 1/13
Important Predictors for plants

**TEMPERATURE**

Determines the speed of growth and, in case of strong seasonality, defines the growing season length. Additionally, minimum and maximum temperatures can reflect physiological thresholds for plants by frost or heat resistance.

*But be careful!*

- large variety of temperature data products, with different impacts on model performance;

- their resolution and accuracy can be coarse compared with the species data
Important Predictors for plants

**Temperature**

**Temperature variables**

**Mean temperature**
- (annual / monthly) mean temperature (of coldest / warmest / driest / wettest quarter / summer / winter)
- soil temperature
- warmth index (the annual sum of positive differences between monthly mean temperatures and e.g. 5 degrees, i.e. a measure of the effective warmth for plants)

**Extreme temperature**
- (annual) min / max temperature (of coldest / warmest driest / wettest quarter / month / season)
- mean temperature of coldest / warmest / driest / wettest month
- mean daily max / min temperature (for DJF / MAM / JJA / SON)

**Temperature seasonality**
- seasonality, annual / diurnal range
- growing degree days (all thresholds) / freezing degree days (FDD) (soil / air) / non-FDD / chilling degree days
- isothermality
- heat units (annual sum of daily temperatures exceeding X degrees)
- frost duration
- winter / summer cold / heat wave duration
Important Predictors for plants

**WATER**

photosynthesis, cooling by transpiration and maintaining turgor.

In ecological models “water” is usually reflected by either precipitation alone or in combination with evapotranspiration (e.g. water balance). These environmental variables are considered a proxy for plant available water.

*But be careful!*

- this might not be the case if soils and topography are heterogeneous, as available water is strongly influenced by both soil type and topographic position.
- The seasonality of available water/precipitation might lead to temporal flooding, drought or snow cover and thus requires special adaptations by plant species
Important Predictors for plants

**WATER variables**

**Mean precipitation**
- (annual / monthly) mean / summed precipitation (of coldest / warmest / driest / wettest quarter / season)
- rainfall intensity

**Extreme precipitation**
- mean / summed / min / max precipitation of coldest / warmest / driest / wettest month
- highest 5-day precipitation

**Precipitation seasonality**
- seasonality, annual range
- snow (cover duration, annual snowfall)
- dry / wet season /day length / intensity / frequency
- % of annual precipitation falling during the growing season
- average flood duration
- the standard deviation of hydrographs

**Water-balance**
- (annual / seasonal / monthly) water balance
- (annual / seasonal) evapo-transpiration, vapour pressure
- (mean / annual / seasonal / soil) water / moisture deficit / surplus / availability /stress
- (annual / seasonal / plant available) water/ wetness / moisture / aridity index
- water content
- flow accumulation
- average water level
- soil moisture (days; days when soil moisture - air temperature ratio is favourable for plant growth)
- waterlogging index

**Soil water capacity**
- soil water capacity, measured soil moisture
- soil drainage class
- hydraulic soil presence class
Important Predictors for plants

They are taken up with water by roots (often with the help of mycorrhiza). Many *micronutrients* are essential for plant survival including *potassium*, *calcium*, *magnesium*, *sulphur*, *boron*, *chlorine*, *manganese*, *molybdenum* and *zinc* but most significant for productivity are usually the contents of *nitrogen* and *phosphorus*.

Nutrients in a wider sense can also influence the pH of the soils, whereas *bedrock* together with living organism are the primary regulators of available nutrients in soils. Therefore, while deriving nutrient content of the soils might not be effective, bedrock, soil pH and soil texture are often used as surrogates in the ecological models.
Important Predictors for plants

**Substrate**
- bedrock, lithology, rock type
  - pH
  - surface geology, geological substrate

**Nutrients**
- nutrients, fertility, Cation-exchange capacity, calcareous
- soil material / depth / order / quality / texture / type
- organic matter, loaminess, alluvial, clay / silt / sand content, salt, gypsum
- soil grain size, bulk density
- FAO soil group
- remote sensed Normalized difference soil index, soil production index
- water regime (ordered classes from dry to waterlogged)
Important Predictors for plants

- Global radiation and therefore energy (W/m²) driving temperature (air, leaf, and soil) and evapotranspiration.
- Photo active radiation (PAR) and is thereby directly related to photosynthesis.
- Light might also contain important signals for plant development (e.g. germination and photoperiodism).

While radiation can be easily modelled and is relatively independent of the vegetation, PAR is strongly affected by the canopy structure of the vegetation.

Variables

- solar radiation (daily, annual, seasonal)
- most / least radiated quarter
- mean hours of sunshine
- clouds
They have **intra and interspecific effects** and have both **positive and negative impacts** by prohibiting or ameliorating growth. Impact of other species can be direct (e.g. competition, herbivory) or indirect (e.g. ameliorating harsh microclimatic conditions, shading, nutrient addition by manure). Biotic interactions have been included to the ecological models as e.g. presence or cover of dominant species, remote sensed vegetation index or interaction matrices for multispecies co-occurrence datasets.

While radiation can be easily modelled and is relatively independent of the vegetation, PAR is strongly affected by the canopy structure of the vegetation.
Important Predictors for plants

**BIOTIC INTERACTIONS variables**

- NDVI, Landsat bands, Enhanced Vegetation Indices, remote sensed vegetation (indices / classes)
- vegetation height / density / volume / cover
- canopy / forest / tree cover
- productivity, Net Primary Production
- ecological classification, succession time
- pollinators
- litter
- distance to moorland, moorland presence / absence
- stand basal area
- % of sparsely / dense vegetated brownfield
- % of brownfield with low / high vegetation
The type and necessity of including disturbance variables in models are highly environment-specific.

**Natural**
- fire, volcanic ash
- geomorphological disturbance
- trampling, grazing
- % area of disturbed terrain

**Anthropogenic**
- population / settlement / building density
- distance to urban areas / roads / harbour / roads
- agriculture, afforestation, soil drainage, roads, human perturbation, forest / etc. management
- human footprint, anthropization degree
- brick rubble
- ownership status (measure of land management)
- predominance of exotic species
They improve ecological models and are easy to get

But be careful!!

- Distal effects: they do not directly impact plants, but they do alter light, moisture, temperature and nutrient conditions (Moeslund et al. 2013);

- It is difficult to interpret the causal relationships between these variables and the target species (Austin 2007)
Important Predictors for plants

Topography and Land Use

**TOPOGRAPHY AND LAND USE variables**

**Topography**
- altitude (range), terrain curvature, topographic position, slope, flatness, meso-topography, % of steep topography, slope type
- aspect, eastness, northness
- rockiness, ruggedness, topographic wetness index
- topographic diversity

**Land-Use**
- Corine, land-use classes (if only "biotic" land-use -> ‘biotic’ class)
- distance to potential forest, age of forest
WorldClim is a set of global climate layers (gridded climate data) with a spatial resolution of about 1 km$^2$. These data can be used for mapping and spatial modeling.

http://www.worldclim.org/

**Current** conditions (interpolations of observed data, representative of 1960-1990; new Version 2.0, 1970-2000, current climate only)

**Future** conditions: downscaled global climate model (GCM) data from CMIP5 (IPPC Fifth Assessment)

**Past** conditions (downscaled global climate model output)
### Database of Predictors

**CURRENT CONDITIONS – Version 2**

<table>
<thead>
<tr>
<th>variable</th>
<th>10 minutes</th>
<th>5 minutes</th>
<th>2.5 minutes</th>
<th>30 seconds</th>
</tr>
</thead>
<tbody>
<tr>
<td>minimum temperature (°C)</td>
<td>tmin 10m</td>
<td>tmin 5m</td>
<td>tmin 2.5m</td>
<td>tmin 30s</td>
</tr>
<tr>
<td>maximum temperature (°C)</td>
<td>tmax 10m</td>
<td>tmax 5m</td>
<td>tmax 2.5m</td>
<td>tmax 30s</td>
</tr>
<tr>
<td>average temperature (°C)</td>
<td>tavg 10m</td>
<td>tavg 5m</td>
<td>tavg 2.5m</td>
<td>tavg 30s</td>
</tr>
<tr>
<td>precipitation (mm)</td>
<td>prec 10m</td>
<td>prec 5m</td>
<td>prec 2.5m</td>
<td>prec 30s</td>
</tr>
<tr>
<td>solar radiation (kJ m(^{-2}) day(^{-1}))</td>
<td>srad 10m</td>
<td>srad 5m</td>
<td>srad 2.5m</td>
<td>srad 30s</td>
</tr>
<tr>
<td>wind speed (m s(^{-1}))</td>
<td>wind 10m</td>
<td>wind 5m</td>
<td>wind 2.5m</td>
<td>wind 30s</td>
</tr>
<tr>
<td>water vapor pressure (kPa)</td>
<td>vapr 10m</td>
<td>vapr 5m</td>
<td>vapr 2.5m</td>
<td>vapr 30s</td>
</tr>
</tbody>
</table>
## FUTURE CLIMATE DATA

### 2041-2060

<table>
<thead>
<tr>
<th>GCM</th>
<th>code</th>
<th>rcp26</th>
<th>rcp45</th>
<th>rcp60</th>
<th>rcp85</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACCESS1-0 (#)</td>
<td>AC</td>
<td>tn, tx, pr, bi</td>
<td>tn, tx, pr, bi</td>
<td>tn, tx, pr, bi</td>
<td>tn, tx, pr, bi</td>
</tr>
<tr>
<td>BCC-CSM1-1</td>
<td>BC</td>
<td>tn, tx, pr, bi</td>
<td>tn, tx, pr, bi</td>
<td>tn, tx, pr, bi</td>
<td>tn, tx, pr, bi</td>
</tr>
<tr>
<td>CCSM4</td>
<td>CC</td>
<td>tn, tx, pr, bi</td>
<td>tn, tx, pr, bi</td>
<td>tn, tx, pr, bi</td>
<td>tn, tx, pr, bi</td>
</tr>
<tr>
<td>CESM1-CAM5-1-FV2</td>
<td>CE</td>
<td>tn, tx, pr, bi</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CNRM-CM5 (#)</td>
<td>CN</td>
<td>tn, tx, pr, bi</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**rcp:** REPRESENTATIVE CONCENTRATION PATHWAYS

- **tn** - monthly average minimum temperature (degrees C * 10)
- **tx** - monthly average maximum temperature (degrees C * 10)
- **pr** - monthly total precipitation (mm)
- **bc** - *bioclimatic* variables
## Database of Predictors

### FUTURE CLIMATE DATA

<table>
<thead>
<tr>
<th>GCM</th>
<th>code</th>
<th>rcp26</th>
<th>rcp45</th>
<th>rcp60</th>
<th>rcp85</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACCESS1-0 (#)</td>
<td>AC</td>
<td>tn, tx, pr, bi</td>
<td>tn, tx, pr, bi</td>
<td>tn, tx, pr, bi</td>
<td>tn, tx, pr, bi</td>
</tr>
<tr>
<td>BCC-CSM1-1</td>
<td>BC</td>
<td>tn, tx, pr, bi</td>
<td>tn, tx, pr, bi</td>
<td>tn, tx, pr, bi</td>
<td>tn, tx, pr, bi</td>
</tr>
<tr>
<td>CCSM4</td>
<td>CC</td>
<td>tn, tx, pr, bi</td>
<td>tn, tx, pr, bi</td>
<td>tn, tx, pr, bi</td>
<td>tn, tx, pr, bi</td>
</tr>
<tr>
<td>CESM1-CAM5-1-FV2</td>
<td>CE</td>
<td>tn, tx, pr, bi</td>
<td>tn, tx, pr, bi</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CNRM-CM5 (#)</td>
<td>CN</td>
<td>tn, tx, pr, bi</td>
<td>tn, tx, pr, bi</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**tn** - monthly average minimum temperature (degrees C * 10)

**tx** - monthly average maximum temperature (degrees C * 10)

**pr** - monthly total precipitation (mm)

**bc** - *bioclimatic* variables
Database of Predictors

CHELSA – Free climate data at high resolution

CHELSA (Climatologies at high resolution for the earth’s land surface areas) is a high resolution (30 arc sec) climate data set for the earth land surface areas currently hosted by the Swiss Federal Institute for Forest, Snow and Landscape Research WSL (Dr. Dirk N. Karger, Prof. Dr. Niklaus Zimmermann), and has been developed in cooperation with climatologists from the Department of Geography of the University of Hamburg (Prof. Dr. Jürgen Böhm, Dr. Olaf Conrad, Tobias Kainz), the University of Zurich (Dr. Michael Kessler, Prof. Dr. Peter Linder), and the University of Göttingen (Prof. Dr. Holger Krift).

It includes monthly mean temperature and precipitation patterns for the time period 1975-2013. CHELSA is based on a quasi-mechanistical statistical downscaling of the ERA.

Selection of indicators 1/4

Background and rationale 2/2

Existing Environmental Database 1/13

WorldClim

CHELSA Bioclim

Bioclimatic variables are derived variables from the monthly mean, max, min temperature, and mean precipitation values. They are developed for species distribution modelling and related ecological applications. They represent annual trends (e.g., mean annual temperature, annual precipitation) seasonality (e.g., annual range in temperature and precipitation) and extreme or limiting environmental factors (e.g., temperature of the coldest and warmest month, and precipitation of the wet and dry quarters). A quarter is defined as a period of three months (1/4 of the year). The code follows those of Worldclim and ANUCLIM.

The layers can be downloaded here

Codes:

Bio1 = Annual Mean Temperature
Bio2 = Mean Diurnal Range
Bio3 = Isothermality
Bio4 = Temperature Seasonality
Bio5 = Max Temperature of Warmest Month
Bio6 = Min Temperature of Coldest Month
Bio7 = Temperature Annual Range
Bio8 = Mean Temperature of Wettest Quarter
Bio9 = Mean Temperature of Driest Quarter
Bio10 = Mean Temperature of Warmest Quarter
Bio11 = Mean Temperature of Coldest Quarter
Bio12 = Annual Precipitation
Bio13 = Precipitation of Wettest Month
Bio14 = Precipitation of Driest Month
Bio15 = Precipitation Seasonality
Bio16 = Precipitation of Wettest Quarter
Bio17 = Precipitation of Driest Quarter
Bio18 = Precipitation of Warmest Quarter
Bio19 = Precipitation of Coldest Quarter
DIGITAL ELEVATION DATA

http://www.cgiar-csi.org/data

Globally available 90m (30m for the US) dataset SRTM is a helpful data source, especially since it offers seamless availability for all areas between 60° N and 60° S.

http://asterweb.jpl.nasa.gov/gdem.asp

ASTER-DEM. 83° N-83° S, globally available, 30m resolution.
### Database of Predictors

#### Soil Grid

http://www.isric.org/explore/soilgrids

**A system for automated soil mapping based on global compilation of soil profile data and publicly available remote sensing data. Resolution 1km or 250m**

<table>
<thead>
<tr>
<th>Physical Soil Property</th>
<th>Chemical soil property</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bulk density (fine earth) in kg/cubic–meter</td>
<td>Cation exchange capacity</td>
</tr>
<tr>
<td>Clay content (0–2 micro meter) mass fraction in %</td>
<td>Soil organic Carbon content</td>
</tr>
<tr>
<td>Coarse fragments volumetric in %</td>
<td>Soil pH</td>
</tr>
<tr>
<td>Silt content (2–50 micro meter) mass fraction in %</td>
<td>Soil Classification</td>
</tr>
<tr>
<td>Sand content (50–2000 micro meter) mass fraction in %</td>
<td>TAXNWRB</td>
</tr>
<tr>
<td></td>
<td>TAXOUSDA</td>
</tr>
</tbody>
</table>

Environmental variables that are thought to be relevant to species' ecology and geographic distribution are essential for applications such as species distribution modeling. However, the number of such variables that are available, that span multiple time periods, and that can easily be integrated with other datasets is very limited.

**ENVIREM** dataset provides a number of climatic and topographic variables that have been described in the literature, and make them available at the same resolutions as are available at [WorldClim](http://worldclim.org), and for both current and past time periods. Additionally, it provides an R package that makes it possible to generate these variables for other input datasets, such as future climate scenarios.
# Database of Predictors

**Envirem**

## Existing Environmental Database

<table>
<thead>
<tr>
<th>variable abbreviation</th>
<th>brief description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>annualPET</code></td>
<td>annual potential evapotranspiration: a measure of the ability of the atmosphere to remove water through evapotranspiration processes, given unlimited moisture</td>
</tr>
<tr>
<td><code>aridityIndexThornthwaite</code></td>
<td>Thornthwaite aridity index: Index of the degree of water deficit below water need</td>
</tr>
<tr>
<td><code>climaticMoistureIndex</code></td>
<td>a metric of relative wetness and aridity</td>
</tr>
<tr>
<td><code>continentality</code></td>
<td>average temp. of warmest month - average temp. of coldest month</td>
</tr>
<tr>
<td><code>embergerQ</code></td>
<td>Emberger's pluviothermic quotient: a metric that was designed to differentiate among Mediterranean type climates</td>
</tr>
<tr>
<td><code>growingDegDays0</code></td>
<td>sum of mean monthly temperature for months with mean temperature greater than 0°C multiplied by number of days</td>
</tr>
<tr>
<td><code>growingDegDays5</code></td>
<td>sum of mean monthly temperature for months with mean temperature greater than 5°C multiplied by number of days</td>
</tr>
</tbody>
</table>
Collinearity

Collinearity refers to the **non independence of predictor variables**, usually in a regression-type analysis. It is a common feature of any descriptive ecological data set and can be a problem for parameter estimation because it inflates the variance of regression parameters and hence potentially leads to the wrong identification of relevant predictors in a statistical model.

**In all real world data, there is some degree of collinearity between predictor variables.**

**Examples:** - we could try to explain the jumping distance of a collembolan by the length of its furca, its body length or its weight. Since they are all representations of body size, they will all be highly correlated;  
- **compositional** data (data where the whole set of information is described by relative quantities;  
- **incidental**, meaning that variables may be collinear by chance, for example when sample size is low, because not all combinations of environmental conditions exist;  
- ........
Correlation matrix of the following six bioclimatic variables (WorldClime): mean annual temperature, temperature seasonality, mean temperature of coldest quarter, annual precipitation, precipitation of driest month, precipitation seasonality.

Dormann, C. F. *et al.* Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography (Cop).* 36, 27–46 (2013)
What can we do about collinearity?

- **Know your data** well enough to be aware of patterns of collinearity in the data sets;

- **Quantify the collinearity** through correlation coefficient, condition index, variance inflation factors.....there are several statistical methods to identify clusters of collinear predictors.

**Methods for removing collinearity prior to analysis:**
1) Principal Component Analysis;
2) represent the cluster by the variable closest to the cluster centroid;
3) represent the cluster by the variables with highest univariate predictive value for the response;
4) Cluster independent methods
What can we do about collinearity?

**Modelling with latent variables**

Some methods are designed to incorporate collinear variables. They deal with collinearity by constructing so-called ‘latent’ variables, i.e. unobserved variables which underlie the observed collinear variables. As a result of the methods used, most variance in the observed explanatory variables is concentrated in the first few new latent variables and usually the less important latent variables are discarded, leading to a reduction in dimensions. Methods differ in how the latent variables are derived, whether the response variable is included in this derivation and how many latent variables are extracted.