ENVIRONMENTAL INFORMATION: PLACING BIODIVERSITY PHENOMENA IN AN ECOLOGICAL AND ENVIRONMENTAL CONTEXT

ARTHUR D. CHAPMAN,

Australian Biodiversity Information Services, PO Box 7491, Toowoomba South, Qld 4352, Australia

MAURO E. S. MUÑOZ AND INGRID KOCH

Centro de Referência em Informação Ambiental (CRIA), Av. Romeu Tórtima, 388, Barão Geraldo, 13083-885 Campinas, SP, Brazil

Abstract.—Environmental niche models are increasingly being used to outline species' distributions for a range of uses. This use has become an important component of the recent science known as biodiversity informatics. Because of the nature of species' occurrence data, considerable effort has often been spent in assessing their quality, but less attention has been paid to determining the quality of environmental data used to model species' distributions. This paper examines a range of environmental data, and evaluates how they are prepared, their quality and use, and some commonly encountered pitfalls and problems in using environmental data in species' distribution modeling.

Key words.—species modeling, environmental data, environmental modeling, climate data, data quality, scale.

The world faces a challenge to manage its biodiversity resources in a sustainable manner, while conserving as much of it as possible for future generations. The study and conservation of biodiversity are not easy tasks, and the past 300 years of scientific endeavor has only just scraped the surface as far as knowing what biodiversity exists on Earth, and how it functions. No amount of biological survey can adequately sample the whole Earth, one country, or even a part of one country. So, in order to gain some understanding of which species occur in areas not yet surveyed or are under-surveyed, smart technologies need to be employed. There are a number of technologies available that allow the estimation of spatial distribution patterns of species (Nix 1986, Austin et al. 1990, Margules and Redhead 1995). By using environmental parameters such as climate, soils and vegetation, and knowledge of where species have been found to occur in the past, likely occurrences of species can be modeled, both now and into the future.

ENVIRONMENTAL DATA AND THEIR USE

Data are the essential starting point for all environmental management processes. This paper will concentrate on the non-biotic or environmental data used in biodiversity informatics along with ecological data. Species' occurrence data will not be covered here.

Non-biotic environmental data are increasingly being used in analyses aimed at estimating biodiversity and modeling distribution patterns of species or populations using point records obtained from collection data (Faith and Walker 1996, Ferrier and Watson 1997, Williams et al. 2002). Often, biological data have been collected opportunistically and thus, for large areas, it is often difficult to determine whether particular species actually occur there or not. The interrelationship between biological data and environmental data, and the knowledge of where those environments occur, can be used to fill in the gaps in the biological data. Environmental data play an important role in biodiversity informatics, as all biological events are directly or indirectly related to environmental conditions. The theory behind ecological niche models is that species have certain habitat preferences that have an environmental basis (Nix 1986). Many models use climatological information such as temperature, rainfall, radiation, evaporation, soil moisture, etc. as the basis on which to define the habitat or ecological niche. Environmental data are also generally more widely available, and generally exist in a more consistent form, than most biological data (Williams et al. 2002).

Some ecological niche models also use classified vegetation maps, detailed habitat information, ranges of interacting species, and soil types. These data are often less accessible, in inappropriate formats, or at inappropriate scales

for use in many biodiversity informatics studies. All too often, the data are of a categorical nature, making use in statistical models where continuous data are required difficult. These data may include both polygon-based vegetation information and pre-classified remotely sensed (RS) raster data.

Data exist in two basic formats: (1) Primary data, such as individual point-referenced meteorological data, and (2) secondary or derived data such as climate surfaces and vegetation classifications. Primary data, which are collected and referenced to individual points, largely eliminate problems of scale and categorization. Categorized information commonly used to produce natural resource maps (e.g., soil types, vegetation categories, tree height classes and species – i.e. a collection of individual specimens – as well as climate layers) is problematic in a number of ways, but is essential for information presentation. One problem with pre-classified data occurs when the concepts underpinning the classification change, and thus the underlying data may become unusable. Data stored as primary attributes (e.g., individual specimens, actual tree heights, etc.), in contrast, can be used to produce classified entities for display and communication while remaining available for use in alternative classifications and for use as individual data points (Chapman and Busby 1994).

Environmental Data

Terrestrial environmental data fall into three basic categories: terrain, climate, and substrate. Too often, these data are used in environmental modeling uncritically and without consideration of the error contained within, leading to erroneous results, misleading information, and even unwise decisions.

Terrain refers to surface morphology, and includes parameters such as elevation, slope, relief, and aspect. Digital Elevation Models (DEMs) are representations of surface morphology, and can be developed at varying scales. The development of DEMs allows for consistent and repeatable interpolation across whole regions and constitutes the necessary first step in generating many climate surfaces (Hutchinson 1991). Construction of a DEM is time-consuming and technically demanding (Hutchinson 1991), but, once created, it doesn't generally need to be developed again for a long time. Errors in this type of data can arise in many ways, and the method used to create the DEM can be important in determining both the type and dimension of likely errors.

Climate data are generally available from national meteorological agencies, but may have to be digitized and interpolated spatially for use in biodiversity modeling programs. Spatial interpolation of climate data can be carried out with the aid of DEMs by fitting surfaces as smooth tri-variate functions of latitude, longitude and elevation (Hutchinson 1995). These interpolations are usually developed at the scale of the DEM, and, when done correctly, involve a lot of data cleaning and quality control. Development of appropriately scaled climate surfaces is essential for modeling species' distributions if models are to have any environmental meaning at scales required for management or decision-making.

Substrate data, both physical and chemical, can be the most difficult to obtain, and quality from one layer to another can be variable. Mapping has been done in most regions of the world, but at varying scales and levels of completeness. Substrate layers include soils, lithology, surficial geology, hydrology, and landform. These data are generally in the form of polygons, and are usually of a categorical nature; although in some cases continuous data may be derivable (e.g., soil texture).

Preparation of environmental layers is one of the most time-consuming, and computer-intensive areas of modeling. Fortunately, it only has to be done occasionally; once surfaces are prepared, they can be used for many models. Climate surfaces, for example, have been prepared for much of the world's land surface, and are available for use by researchers either for free or at nominal charge. These data sets, however, are at varying scales, and surfaces at suitable scales have not been available for some parts of the world until recently (e.g., South America). Recent work has lead to release of globally-consistent climate layers at 30" (c. 1 km) resolution, with derived layers at 2.5', 5', and 10' resolution released in early 2004 (Hijmans et al. 2004a, b). These layers are at ideal scales for modeling for both local area (30") and continental (2.5' and 5') analysis, and provide a major advance over layers previously available. Release of these layers allows for consistent modeling to be carried out across and between continents.

Ecological Data

Ecological data can be of many forms: from point-based biological data and polygon-based

vegetation data through to RS raster data. There are a number of issues associated with accuracy and error with each type of data, and types of error may vary depending on whether data are raster, polygon, or point.

Ecological data, such as vegetation and soils, are often categorical, and boundaries, although appearing discrete in the database, are seldom discrete in nature. For example, vegetation is usually stored as vegetation classes, and when mapped it is shown as polygons with distinct lines between one class and the next. In reality, distinctions between classes are not always clear, and mapped boundaries are usually subjective and quite artificial. In most ecological niche models, differences between classes have to be regarded as equal, whereas in reality some classes may be very close ecologically and others quite distant. This variability can lead to distinctions in the model output that may not exist in nature, and thus use of categorical data requires considerable precautions. As a result, it is often better to use categorical data as overlays in a geographic information system (GIS) to refine the modeled distribution, after modeling is completed, rather than as a layer within the model itself. For example, if a species' distribution model is obtained using climate, it can then be overlayed on vegetation types to exclude areas on unlikely vegetation types, and thus define the niche of the species more finely.

Ecological data are not always categorical, and continuous layers such as soil texture, pH, and water-holding capacity can be derived from them, and used in models as continuous data.

MANAGING ENVIRONMENTAL DATA LAYERS

An environmental data layer refers to a data set describing a characteristic of the environment that varies over a particular geographical region. Environmental data, although sharing common attributes such as georeferencing, can be categorized into different data types. Each data type has its own method of storage, with some data types having more than one file format. A brief introduction to the various environmental data types is given below, with information on how they are georeferenced, stored, and used.

Data Types

Environmental data layers must be composed of compatible elements if they are to be manipulated individually. Geographic information systems include two layer types: vector and raster. In shape files, all layer components are described geometrically, including points, lines, and polygons. In raster files, each grid square stores all of the available information for that square.

Point.—The point element is composed of a pair of georeferencing codes (e.g., longitude and latitude, UTM), along with additional optional attributes associated with the locality the point represents. Examples are gazetteer locations, specimen location data, and meteorological stations. For the latter, each station has its own georeferencing information (longitude, latitude) and additional attributes such as measured precipitation and temperature, the station's name, the station's responsibility, a textual description of location, etc.

Line.—A line data element is a set of connected linear segments. Each segment has a pair of georeferencing codes (e.g., longitude, latitude) representing the beginning and the end of the segment, plus the line's attributes. Examples are roads, rivers, transect survey data, etc. In the river example, sequential georeferenced line segments describe its location; additional attributes may include name, flow direction, etc.

Polygon.—A polygon element is composed of a set of ordered georeferenced points such that the first and the last points are coincident (thus the shape is closed and defined) (Noonan 2003). The points represent the polygons' vertices, which can be ordered clockwise or counterclockwise. Polygons are used to delimit geographical regions and each also has its own attributes. Examples are cities, conservation areas, vegetation and soil classes, and rivers and roads (when their widths are relevant). In the city example, the polygons can delineate city limits and the attributes can be the city's name, population size, etc.

Grid.—The grid element is a georeferenced matrix of cells. Usually the grid represents a rectangular region, as does each cell. A rectangular grid is defined by the four georeferenced points that represent the corners of the rectangle, cell width and height, attributes of the grid, and individual cell values. Examples of grid include climate grids, DEMs, and land use summaries.

Grids are used to represent phenomena that vary continuously over a geographic area. The phenomena can be discrete like soil type and land use, or continuous like temperature and elevation. When representing continuous phenomena, the grid stores samples of the

phenomenon and not the phenomenon function itself. The information accuracy therefore depends on cell dimensions, because each cell holds one value to represent the phenomenon over the total of its region.

Some phenomena cannot be measured, are not important, or do not make sense for certain areas. For example: soil type in water, water pH in land, political divisions, etc. To represent the information in cells where the phenomenon value is not known the grid element is given a special value called "novalue". The value actually used to represent "novalue" can be different from one grid to another, and is defined in the grid metadata stored in the file's header.

Georeferencing

Georeferencing is a simple concept, but is a difficult task. The concept of georeferencing is related to locating or positioning something on the Earth or relating it to the "real world." Problems arise in trying to define the Earth's surface mathematically, because the Earth has a highly irregular surface. The solution is to represent the surface by its ellipsoid of revolution, and to use a geographic coordinate system (GCS) to locate points on the surface.

The GCS is a spherical coordinate system aligned with the spin axis of the Earth. It defines two angles measured from the center of the Earth. One angle (latitude) measures the angle between any point and the equator line. The other angle, called longitude, measures the angle along the Equator from an arbitrary point on the Earth. Greenwich, England, is the accepted zero-longitude point (Sobel 1995, Wikipedia 2004).

A problem arises when trying to fit the GCS, which is spherical, to an ellipsoid. To solve this problem, the concept of a 'datum' was created. A datum is a set of points used to reference the GCS position in the sphere to the ellipsoid of revolution. Depending on the region of the Earth being georeferenced, different datums are used so that the GCS better approaches the ellipsoid in the target region.

To define a coordinate system one thus needs to know:

- Ellipsoid of revolution
- Datum
- GCS spheroid radius (tied to datum definition)
- Origin meridian (usually Greenwich)

Examples of ellipsoids are:

- Intl – International 1909 (Hayford)
- New Intl – New international 1967
- WGS84 - World Geodetic System 1984
- Evrst69 – Everest 1969

Examples of datums are:

- WGS84 – World Geodetic System 1984[1]
- GDA94 – Geocentric Datum of Australia 1994
- NAD83 – North American Datum 1983
- SAD69 – South American Datum 1969

Usually the datum name is sufficient to define the coordinate system (e.g., WGS84).

Projection is another aspect that influences the way maps are georeferenced. The Earth is approximately a 3D sphere, but the usual way to communicate visual or graphical information is as a bi-dimensional or flat surface, which has lead to a proliferation of different projections. Many projections are regional, while others are historical. Projections or representations that function well at equatorial latitudes do not always function well at high latitudes, and vice-versa. Examples of projections are:

- Universal Traverse Mercator (UTM)
- Albers Equal Area
- Azimuthal Equidistant
- Equidistant Cylindrical Projection
- Hammer-Aitoff Equal Area Projection
- Geographic coordinates

Some projections preserve distances between points, so that one can measure the distance and multiply it by the map scale to obtain the real distance (e.g., Lamberts Conical Projection). Other projections preserve areas, but not distances (e.g., Albers Equal Area).

Choice of the GCS and/or projection is determined by the location and extent of the region to be mapped and the way the map is to be presented and used. Poor choices will lead to inaccurate or distorted data. A consequence is that when different sources of data are used, coordinate system transformations can become an important task for the environmental data user, and especially when working over large regions or areas. In general, analyses over large regions or globally use WGS84, as it is generally regarded as the best datum for relating one continent with another.

Working with Grid Layers

Grid layers are the most commonly used data

---

[1] http://www.wgs84.com.

type for ecological niche modeling. They also generally cause the greatest difficulties for users. Different geographic information systems (GIS) use different approaches to deal with grid layers.

File formats.—Grid layers are stored in many different formats. While some formats are open, others are restricted. Open file formats have an internal organization that is publicly available, and relevant software applications know how to read and write them. On the other hand, restricted grid file formats are generally not known, and thus cannot be read or written by different software applications. Examples of open grid file formats are: GeoTiff (Geographic Tagged Image File[2]) and Arc/Info ASCII Grid (from ESRI®). Fortunately for software developers, many computer libraries (e.g., GDAL) provide implementers with uniform ways to access different formats.

Data related to georeferencing are stored in a grid's header file. The header information varies with each file format. Despite differences between formats, there is common information used by almost all formats: region extent, number of cell rows and columns, cell dimensions, georeferencing system, projection, content unit (e.g., km, m, kg), and "no data" value. Some grid formats include additional software-specific information to speed up or simplify specific software tasks, such as maximum, minimum and average cell values; how cell values should be translated into colors; and cell value storage type (e.g., byte, word, floating point, etc). The user of a single application does not have to worry about file formats, as the application will handle reading and writing. On the other hand, if the user wants to use the grid in different applications, it is necessary to ensure that all applications can correctly handle the chosen file format(s).

Generating grid information.—Another important issue with grid layers is the way information is read and written. As mentioned above, each cell in a grid layer stores a value associated with the region it represents and the grid is used to store a phenomenon over a whole region. Due to the discrete nature of the cell and cell region of the grid, the phenomenon information needs to be discretized to fit in the grid. If the phenomenon is continuous (e.g., temperature), the cell can store a sample (e.g., the value at the center of the region) or the average value for the cell region. If the

phenomenon is discrete (e.g., soil category), the cell can store the most important, most abundant, or even a random category found within the region. Choice of what is stored in a cell is determined during grid creation, and can depend on the way the phenomenon is captured.

Reading grid layer information.—The reading process presents the inverse problem. The grid is a discrete matrix of samples (or averages) and a map of continuous values (e.g., temperature) and/or continuous regions (e.g., temperature or soil categories on a shore) may be required. For example, Figure 1 shows four cells (A, B, C, and D) of a grid that stores a phenomenon (annual average temperature). Each cell region encompasses 1 x 1° of longitude and latitude, and has an associated temperature value. According to knowledge of the way the grid was originated, the value at (x, y) can be estimated using one of several methods. Resampling of the grid values can assist in making this estimation, and is used when one wants to read a value of a grid layer assuming that the information is continuous within the region.
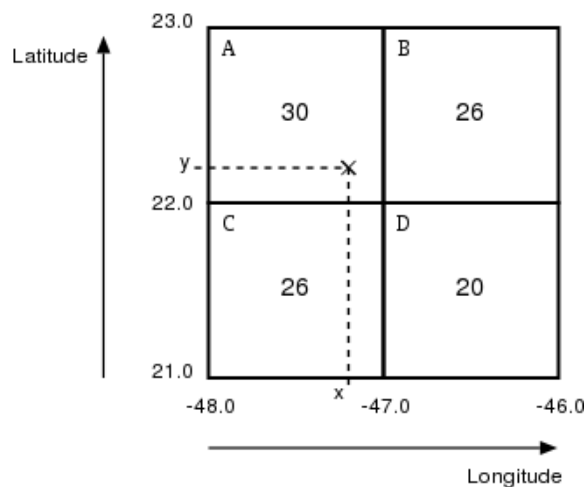

Figure 1. A 4-cell grid with a representative point (x,y).

Resampling.—The first step in resampling a grid layer is to find a function that represents its phenomenon. This function is usually piecewise, using a combination of values corresponding to cells in the neighborhood of the point being resampled. The phenomenon value is then calculated as the function at the desired coordinate.

The three most common methods used are nearest neighbor, bilinear interpolation, and cubic convolution. With nearest neighbor, the value returned is the value of the cell at the given

---

[2] http://remotesensing.org/geotiff/geotiff.html.

point (*x*, *y*). In other words, for all points (*x*, *y*) in a cell region, the resulting value is the value of that cell. This method is the fastest, and is the best for use with categorical data, because it assures that the returned value is present in the original grid. For example, in Figure 1, the returned value for (*x*, *y*) is the value of cell A, 30.

Bilinear interpolation is a weighted average of the values of four nearest cells (cells A, B, C and D in Figure 1). This method should not be used for categorical grids, as the result is not guaranteed to be a valid category. This method guarantees that the resulting value is always within the range of the 4 nearest cell values.

In cubic convolution the 16 nearest cells values are used to find a smooth surface (usually with cubic splines) that passes through all its centers. The resultant value is not guaranteed to be within the values of the cells used. This method should not be used with categorical grids, as the result can be different from those defined in the grid. Although computationally intensive, it gives good results when grid phenomena are continuous.

DIGITAL ELEVATION MODELS
Use in Biodiversity Informatics

Digital elevation models (DEM) or digital terrain models (DTM) form one important element in ecological niche modeling. They are used directly in providing data on elevation, slope, and aspect, and indirectly in development of climate layers.

Elevation can be an important environmental variable in determining niches of species. It is well known that some species grow at high altitudes, and others at low altitudes. The reasons are often climatic (temperature, occurrence of frost and snow, etc.), and for this reason elevation is important in derivation of climate layers. Likewise, slope and aspect are important driving characteristics for species' niches. Some species grow preferentially on northern (Southern Hemisphere) or southern (Northern Hemisphere) aspects to obtain more of the sun's warmth during the day. Other species are the opposite, and prefer cooler daytime temperatures. Some species prefer to grow in areas of high slope where water may not accumulate and pool and frost slides off, while others prefer flat areas where soil water content may be more consistent and less variable.

Methods of Development
Stereo photogrammetry.—DEMs have been

around for many decades. Traditionally, they have been created using a combination of spot height information determined through field surveys and stereo-photogrammetry using aerial photographs. These methods are time-consuming, laborious, and costly, and are really only suitable over small areas. Data from these methods, especially the spot-height information, provide valuable information for use with the other techniques mentioned below.

ANUDEM.—The most common method for developing DEMs in recent years has been that used in the ANUDEM program developed at the Centre for Resource and Environmental Studies in Canberra (Hutchinson 1989), a version of which is now included in ArcInfo as TOPOGRID. The method iteratively applies a spline interpolation algorithm to calculate values on a regular grid using irregularly spaced elevation data points from contour line data, streamline data and individual spot heights. The strength of the ANUDEM method over most other methods is that it imposes a global drainage condition through an approach known as drainage enforcement, to produce elevation models that represent more closely the actual terrain surface and which contain fewer artifacts than those produced with more general-purpose surface interpolation routines (USGS 2003).

Remotely sensed imagery.—Several remote-sensing (RS) techniques can also be used to create DEMs, such as use of Radarsat or JERS images like the Synthetic Aperture Radar (SAR). These methods can be used directly to create a DEM using interferogram (phase difference) techniques that use two (or more) images of the same area taken from different angles (Rao and Rao 1999). It is predicted that these methods, possibly in conjunction with some others, could lead to accuracies as small as 1 mm (Arora et al. 2002) but that is only likely to be over very small areas, and would be costly in both resources and computing power.

Combination of methods.—More recently, techniques have been developed that use the results derived from Radarsat imagery in conjunction with the ANUDEM software to provide a DEM of much greater accuracy. This approach has been especially valuable in areas with little terrain variability, and has recently been used to create a DEM of the Antarctic (Liu et al. 2001).

Accuracy and Sources of Error
DEMs can be quite variable in accuracy,

depending on their development method and the availability of suitable data from which to derive them. The elevation error at a single point in a DEM depends on the resolution (cell size) and the roughness of the surface being modeled (Hutchinson 1996, 2003b). Imposed global drainage as used in ANUDEM, has been found to increase DEM accuracy significantly, especially in terms of drainage properties (Hutchinson 2004). The USGS 1 km DEM Version 1.0, for example, is variable in accuracy across different areas, and is known to have considerable error in some areas of South America (NGDC 2000).

Recent use of ANUDEM has led to development of DEMs with much less error than previous methods (Hutchinson 1996) and these methods are now being used across much of the world to develop accurate, continental-scale DEMs at scales as fine as 80 m or 2" of latitude-longitude. An example is the Global 30" DEM--GTOPO30--produced through a collaborative effort coordinated by the USGS's EROS Data Centre (USGS 2003). Although output grids are improved over previous versions, their accuracy is still limited by the accuracy of the source materials used to create them (Olsen and Bliss 1997). Ideally, using ANUDEM, accuracy of the interpolation should approach one-half of the contour intervals of the source data.

CLIMATE DATA

Climate data, as used in ecological niche modeling, are interpolated surfaces developed from information such as temperature and rainfall from weather stations at known locations, and integrated with terrain data (usually a DEM) to form a smooth coverage over the Earth's surface. These data allow a user to estimate climate conditions at any point on the surface. This information is important in modeling because weather stations are not sufficiently abundant to permit accurate determination of climate at points where weather stations do not exist, and thus it is essential to use such derived surfaces.

Use in Biodiversity Informatics

Climate information has formed the basis of most ecological niche modeling applications over the past 20 years. It has been used in models associated with biogeographic studies (Longmore 1986, Peterson et al. 1999), conservation planning (Faith et al. 2001), reserve selection (Margules and Pressey 2000), development of environmental regionalizations

(Thackway and Cresswell 1995), climate change studies (Chapman and Milne 1998, Peterson et al. 2002), agriculture and forestry production (Booth 1996, Nicholls 1997, Cunningham et al. 2001), species translocation studies (Mackey 1996, Soberón et al. 2000, Peterson and Vieglais 2001), etc.

In many cases, climate layers have been used as fairly raw layers, such as maximum and minimum temperatures in certain months of the year. One of the most important factors in determining where a plant may grow, however, at least at the macro-level, is the relationship between rainfall and temperature. Agronomists have long relied on this knowledge to plan summer and winter plantings of different crops. Some species respond to rainfall at certain times of the year, and others at other times. Rainfall in the middle of winter, for example, may have little or no effect on species that are dormant during that period. In more tropical areas, however, this season may be the key growing period, as many plants reduce transpiration over summer when it is too hot, and most of the growing is done during the cooler period of the year. Alternatively, a long dry period in the middle of a hot summer may have more detrimental effects on a plant than the same long dry period during the middle of winter. Long experience of modeling in Australia has determined that this combination of layers has more environmental relevance, and generally produces better models, than just raw monthly values (Nix 1986). Layers such as mean temperature of the wettest and driest, warmest and coolest quarters, and mean precipitation of the warmest and coolest, wettest, and driest quarters relate well to the environmental conditions that determine where a plant or animal is likely to occur. We recommend that consideration be given to greater use of such synthetic layers in future modeling efforts.

Methods of Development

The distribution of meteorological stations around the world is very uneven. Traditionally, stations have been established in urban areas and areas of high agricultural production, with predominantly natural areas having very poor coverage. As such, important areas such as mountaintops, and wilderness areas have very little detailed climate information. Because of the sparseness of this information, algorithms have had to be developed to fill in gaps and to produce

a 'blanket' of climate information to cover all areas. Rainfall and temperature patterns are heavily reliant upon altitude, slope, aspect, positioning of hills and mountains, and relationship to large water bodies. Several algorithms have been developed to fit climate surfaces using known meteorological data points in conjunction with surface topology to create climate layers for use in ecological niche modeling. The underlying DEM and the meteorological station data, therefore, form the basis of most (if not all) of these surface-fitting algorithms. Different methods and algorithms used for interpolation produce layers with important differences. A few of the more important differences are discussed briefly below. All of these methods rely on the underlying DEM to create the surfaces, and the finer and more accurate the DEM, the better are the resultant surfaces. Some algorithms (e.g., ANUSPLIN) place greater reliance on the DEM than other methods and are thus more likely to result in more robust surfaces, but again that is dependent on the accuracy of the DEM, and many have problems where relief is subtle.

Centro Internacional de Agricultura Tropical (CIAT, Columbia).—The CIAT method uses a simple interpolation algorithm based on the inverse square of the distance between the station and the interpolated point of the nearest five stations (CIAT n.d.). This method has the advantage of speed and ease of use for large data sets where computational capacity is limited (Booth and Jones 1996). The influence of a bad data point can be significant and can cause significant circling in the resultant surface. Because it is using only five data points, it relies less on the underlying DEM than other methods.

ANUSPLIN (Australian National University).—ANUSPLIN (Hutchinson 2001) is a technique that uses partial thin-plate splines to interpolate multivariate data. It is made up of nine programs (Kesterven and Hutchinson 1996), was developed in the 1980s, and has been refined extensively since. It is a proven methodology, with wide acceptance across the world. Recent modifications now allow for simultaneous analysis of several surfaces (the concept of "surface independent variables") (Kesterven and Hutchinson 1996). Cost to the user is not high, and most of the work can be carried out "in-house". The advantage in using ANUSPLIN is its heavy reliance on the underlying DEM, and thus its tendency to produce more finely resolved surfaces. Because most major climate-summary

efforts around the world are using ANUSPLIN, surfaces created are likely to be consistent with surfaces in other parts of the world. This method has recently been used to develop globally consistent (as far as the meteorological data permits) 30" climate surfaces covering most areas of the Earth's terrestrial surface (Hijmans et al. 2004a, b).

Parameter-evaluation Regressions on Independent Slopes Model (PRISM).—PRISM is a knowledge-based interpolation system developed at Oregon State University in the 1990s. It has been used principally for developing climate layers for the United States (Daly et al. 1994; Gibson et al. 2004). The method incorporates a number of spatial interpolation quality control measures in similar ways to ANUSPLIN. It is being used in conjunction with automated data collection in the USA (Daly et al. 2004).

Other methods.—Several other methods have been used recently for estimating climates. These methods include regression (Zheng and Basher 1996), inverse distance (Goovaerts 2000), first detrend for elevation, exposure, orographic influences, then kriging (Holdaway 1996), cokriging with elevation (Phillips et al. 1992), and gradient plus inverse-distance-squared (GIDS; Nalder and Wein 1998)

Summary

Several studies have compared methods for fitting climate surfaces. In nearly all cases, the conclusions have favored partial thin-plate spline techniques over others. For example, one study compared methods for interpolating climate surfaces using Mexican data (Hartkamp et al. 1999); the authors concluded: "taking in account error prediction, data assumptions, and computational simplicity, we would recommend use of thin-plate smoothing splines for interpolating climate variables." Another study (Booth and Jones 1996) suggested that more complex interpolation algorithms such as Laplacian splines are better interpolators than most of the simpler methods, but need much more computing power. The same study states "the major difference between the techniques used by CIAT (Jones et al. 1990) and CRES [=ANUSPLIN, above] (Hutchinson 1989) is that the CIAT method uses a standard lapse rate applied over the whole dataset; the CRES method uses a 3-dimensional spline algorithm to determine a local lapse rate from the data." They concluded that ANUSPLIN provides a powerful

set of programs for climate analysis. A comparison between GIDS and ANUSPLIN (Price et al. 2000) concluded that thin-plate splines generally produced better smoothing and better gradients at high elevations and in areas where climate station coverage was poor, and in predicting climate variables at points withheld at random from the source datasets. A study in New Zealand (Barringer and Lilburne 2000) looked at methods for determining soil surface temperatures and also concluded that partial thin-plate spline surfaces had the lowest residuals for long-term mean monthly and specific month/year soil temperature surfaces, and that multiple linear regression provided a simple and robust method for soil temperature interpolation where the data were not strongly spatially dependent.

Time Period

One issue often not taken into account is the time period from which the climate surfaces being used were developed. For many climate stations, it is difficult to obtain runs of consistent climate data for periods greater than 10 years. To prepare robust climate surfaces and ease out short-term variation, one should seek to obtain at least 30-year runs of data. Wherever possible, for modeling with older species' occurrence information, climate layers should be prepared for runs prior to at least 1990, and preferably as far back as 1970 (i.e., before recent climate change effects became noticeable).

Accuracy and Sources of Error

Similar to species' occurrence data, mis-location of weather stations, or errors in readings at those stations, can create errors in resultant climate surfaces. Georeferencing weather stations can be just as tedious and error-prone as georeferencing species' occurrence data, but at least there are fewer weather stations than species collections! One common problem is that the geocode refers to the center of the town, when the meteorological station may be kilometers away (Busby 1991). When prepared critically, and once individual meteorological data records have been cleaned, climate layers will have significantly lower levels of error.

Errors are of two major types--positional and attribute. Positional error depends largely on the accuracy of the underlying DEM. The DEM for South Africa, for example, at 10' has a standard error of between 20-150 m (Hutchinson 2003a). The attribute error for the climate data, on the other hand, is different for temperature and

rainfall. Because ANUSPLIN depends on elevation, it is significantly more accurate than methods that use bivariate functions of longitude and latitude only (Margules and Redhead 1995). Standard errors of temperature data are of the order of 0.5° and of rainfall about 5-15%, depending on data density and spatial variability of the actual monthly mean rainfall (Margules and Redhead 1995; Hutchinson 1996, 2003a).

One of the greatest sources of error in climate surface development is lack of reliable meteorological stations across large areas of the world. The overall accuracy of climate surfaces derived, therefore, depends largely on the ability of different methods to handle this scarcity of data, and to interpolate best into areas where meteorological data are scarce.

EVAPOTRANSPIRATION

Evapotranspiration (ET) is the term used for the transfer of water, as water vapor, from land surfaces (both vegetated and non-vegetated) to the atmosphere by evaporation and transpiration. ET depends on the energy supply (mainly direct solar radiation), vapor pressure, and winds, and is affected by climate, water availability, and vegetation cover. It is difficult to obtain accurate field measurements of physical parameters necessary to measure evapotranspiration, so procedures have been developed to assess ET from meteorological data to produce continuous maps of ET or potential evapotranspiration (PET).

Wang et al. (2004) adapted the complementary relationship areal ET model (Morton 1983) to estimate and map evapotranspiration as 'areal actual,' 'areal potential,' and 'point potential' using modified Priestley-Taylor and energy transfer and balance equations. The Food and Agriculture Organization (FAO), on the other hand, uses the Penman-Monteith equation, which they found best for use at global scales (Allen et al. 1998). A third method often used is that of Hargreaves and Samani (1982), which estimates potential evapotranspiration as a function of solar radiation and air temperature.

Accuracy and Sources of Error

Mapping ET is affected by the spatial coverage of available climate stations, accuracy of interpretation of vegetation cover, and by interpolation and mapping techniques used. Errors appear greater at high latitudes owing to unreliability of solar radiation estimates in those

areas. In a comparison of the Penman-Monteith and Hargreaves methods, Reynolds et al. (2000) recommended that the Penman-Monteith method was more data intensive and the most reliable method where accurate data was available, but that the Hargreaves method was a preferred alternative where accurate data collection is less certain. ET can be valuable in ecological niche modeling, however, it "is almost impossible to measure or observe directly at a meaningful scale in space or time" (Wang 2004).

TOPOGRAPHY

Topography is another set of layers derived from DEMs. The two most common derived layers are slope and aspect, although specific catchment area and contour curvature are also sometimes used for determining ET and soil moisture (Gallant and Hutchinson 1996). The fineness of the underlying DEM is key in determining the accuracy of the derived slope and aspect layers (see, e.g., Figure 5). Secondary layers such as solar radiation are based on slope and aspect and modified by topographic shadowing (Moore et al. 1993; Gessler et al. 1995).

Use in Biodiversity Informatics

Several process-based landscape-scale ecological niche models have included topographic data in their development (Gallant and Hutchinson 1996).

Accuracy and Sources of Error

Topographic layers are sensitive to the resolution of the source from which they were derived, the underlying DEM. Surface reconstruction using contouring from DEMs has been shown to be largely dependent on the scale of the DEM, which can lead to large variation and error in derived layers such as slope, aspect, solar radiation, catchment area, soil moisture, and contour curvature (Gallant and Hutchinson 1996).

SOILS

Discrete soil measurements can be converted to continuous soil layers by environmental correlation with continuous spatial data sets (e.g., DEMs, RS imagery) using statistical, geostatistical, or numerical models. The results of these models produce soil attributes necessary to generate continuous data sets for use in niche modeling. Gessler et al. (1995) used the compound topographic index (cti) for allocating field sample locations and exploratory data analysis to search for useful relationships between modeled soil attributes and environmental attributes in a spatially continuous manner. These relationships were then confirmed and defined by statistical models, and improved by field verification. The methodology was improved on by Gessler and Chadwick (1997). Continuous soil data such as pH, soil water holding capacity, texture, chemical composition, etc., are layers that may be used to advantage in niche modeling.

REMOTELY SENSED DATA

Remotely sensed (RS) data represent a powerful resource where information about large geographic areas is required. RS data cover the energy captured from a sensor distant from the object or radiating phenomenon. Data captured from airplanes and satellites are the most common type of RS data for use in biodiversity informatics and are what most people understand by the term "remotely-sensed" data.

The two most common types of remote sensors used for the study of the Earth's surface are optical and radar. Optical sensors use the visible, near-infrared, and short-wave infrared parts of the spectrum to form images of the Earth's surface by detecting solar radiation reflected from objects on the ground. Different materials reflect and absorb radiation differently at different wavelengths, so objects such as bare rock, vegetation types, etc., can be differentiated by their spectral reflectance signatures. Radar sensors use radar or sonar to detect variations in terrain, including ocean depths, elevation, etc.

Different types of optical remote sensors include (Liew 2004):

- *Panchromatic* – a single channel sensor that covers one broad wavelength, and measures the apparent brightness of the object; the resulting image resembling a black-and-white photograph (e.g., SPOT HRV-PAN)
- *Multispectral* – a multi-channel sensor covers just a few spectral bands (usually 3-7), with each channel recording reflectance in a narrow spectral band; results in a multi-layer image that can be colored in various ways to emphasize particular characteristics (e.g., AVHRR, Landsat TM)
- *Superspectral* – has many more channels (usually >10), with each channel covering much narrower bandwidths, enabling finer characteristics of the environment to be detected (e.g., MODIS)
- *Hyperspectral* – instruments known as "imaging spectrometers" acquire information in 100+

contiguous spectral bands. These data are used for monitoring phytoplankton, pollution, etc., but have high potential for use in biodiversity informatics. Given the extraordinary amounts of data being reported, however, this data resource is beyond most existing project specifications (e.g., Hyperion on EO1).

Raw RS data, the radiometric information data resulting directly from the sensors, needs to be preprocessed to generate useable thematic information, such as sea surface temperature or vegetation types. These exercises can be time-consuming and computer-intensive, and are usually best carried out by specialists.

Use in Biodiversity Informatics

A main use for RS data in biodiversity informatics has been the derived Normalized Difference Vegetation Index (NDVI), or Greenness Index. It approaches a representation of 'greenness' or amount of photosynthetic mass. It is calculated from reflected solar radiation in the near infrared (NIR) and red (RED) wavelengths as:

$$NDVI = \frac{(NIR - RED)}{(NIR + RED)}$$

NDVI is correlated with photosynthesis through absorption of red light by plant chlorophyll and reflection of infrared radiation by water-filled cells in leaves. It is thus commonly used as an estimate or surrogate of green vegetation (DEH 2004). In biodiversity informatics, it has largely been used as an overlay in a GIS as a surrogate for vegetation following modeling (e.g., Chapman and Milne 1998). Because it is a continuous (non-linear) function that varies between –1 and +1, it could easily be used as a layer in niche modeling (Pereira 2002). Its biggest problem may be when RED and NIR are zero, as then NDVI is undefined.

The scale at which RS data are used determines their use and accuracy. The most commonly used images are from AVHRR images at about 1 km pixel size. Other commonly used images are from Landsat (25-100 m pixel size) and low-altitude Spot images (about 10 m pixel size). For fine-scale regional studies, images from sensors similar to those used in satellites can be acquired from airplanes, bringing pixel sizes down to ≤1 m. Such fine-scale data are expensive, however, and unlikely to be suitable in many modeling studies, other than as GIS overlays.

Accuracy and Sources of Error

Many sources of error enter into RS data. These sources include variation between sensors on different satellites, signal decay over time, angle of incidence between images, angle of the sun's rays when the image was taken, rectification of the image relative to the Earth's surface, and variation in screening procedures for cloud and pollutants in the atmosphere.

Several algorithms exist for calculating the magnitude of these errors. For example, calibration equations exist for rectifying data from various satellites (Rao and Chen 1995, 1996). These calibrations are usually calculated using reflectance in areas of low vegetation cover (e.g. deserts, water bodies) where vegetation cover (and hence reflectance) doesn't vary over time.

Low sun angle can lead to poor data quality owing to large angle corrections--shadows in steep terrain, increased reflectance off atmospheric pollutants, etc. In Australia, between April and September 1994, for example, NOAA 11 data deteriorated considerably as to be almost unusable owing to low sun angles (DEH 2004).

One key aspect of error in RS data is the accuracy with which a pixel can be located on the ground. The margin of error, depending upon the method of geometric registration, is generally accepted to be ~1 pixel. Thus, AVHRR, which has a pixel size of ~1 km, has an accuracy of about +/- 1 km (Mao et al. 1999). In many cases, however, especially in areas of the Earth with few identifiable registration points (e.g., Australian deserts, marine areas, the Amazon), pixel accuracy cannot be relied on at better than +/- 2 pixels (S. Cridland pers. comm. 2000).

To account for cloud cover, cloud screening masks have been developed (DEH 2004). These tools seek pixels that do not appear biologically consistent with images before and after the image of interest. Cloud screening usually includes subjective steps, and thus are a possible source of error.

VECTOR-BASED ECOLOGICAL DATA

Not all environmental data are raster-based. Terrain-based polygon data, for example, form the basis of many traditional paper-based maps, as well as many digital maps used in GIS. These maps can have quite varying levels of accuracy. For example, TOPO-250K for Australia

(1:250,000 topographic data) is a well-researched data set and its accuracy is described as "not more than 10% of well-defined points being in error by more than 160 meters; and in the worst case, a well defined point is out of position by 300 meters" (Geoscience Australia 2003). This accuracy can be quite important, for example, if one is trying to determine if a species (with accuracy ~1 km) occurs in a national park (with an accuracy of ~160 m). With paper topographic maps, drawing constraints may restrict accuracy with which lines are placed. A 1 mm wide line depicting a road on a 1:250,000 map represents 250 meters on the ground. To depict a railway running beside the road, a separation of 1-2 mm (250-500 meters) is needed, and then the line for the railway (another 1 mm or 250 meters) makes a total of 750-1000 m as a minimum representation. If one is using such features to determine an occurrence locality, for example, then maximum precision would be ~1 km. Accurate coastline representation (Figure 2) can also be a nightmare--does the map use the high-water mark or mean sea level, and have neap, spring, or king tides been taken into account (Bannerman 1999)? When it crosses the mouth of a river, does it take a direct line across, or does it follow the river upstream for a distance? Are rivers depicted by two lines (one for each bank), or by just a centerline? Similarly, are towns shown as points, and if so what part of the town does that point represent? If towns are represented by polygons, is it the municipal boundary or the boundary of outer development (Wieczorek 2001)? How is terrain represented--just by contours, thus excluding highest and lowest points, or does it show spot heights? How accurate are they? Does it also show low points? Many features (e.g., coastlines and rivers) change over time. In many tropical areas of the world, seasonal billabongs are formed--how are they represented (Bannerman 1999)? Again, the scale of the map can make a significant difference.

The depiction of phenomena that don't have discrete boundaries in nature (vegetation, soils, geology, etc.) as polygons is also a source of error (Burrough and McDonnell 1998; discussion under Ecological Data, above). Where lines are drawn can lead to major errors, both geographic and attribute-related. In some cases, a vegetation description may be a mosaic of vegetation types and not a single discrete type (Sattler and Williams 1999). If one is attempting to map a particular community (Community 'A', for example), one group of polygons may contain 5-45% Community 'A', one group may contain 50-95%, and one group of polygons may contain 100%. The polygon with 5-45% of Community 'A', doesn't mean that it has a cover of 5-45% of that community, but that there is 5-45% chance of finding it somewhere within that polygon. Depending on the scale of the mapping, there may be a small area of 100% of that community in one small area, and none in the rest. So, if you are talking about Community 'A', where do you make the cut off? Instead of using raw data, classification might be desirable, but can cause problems in interpretation and thus error.
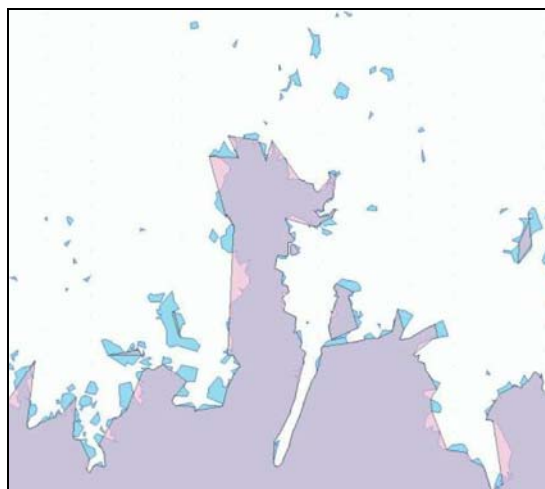

Figure 2. Example of coastlines at two different scales: 1:5,000,000 and 1:500,000.

SCALE

One very important consideration in choosing environmental layers is that of scale. Too fine a scale can often lead to errors due to mismatching with biological data. For example, historical biological data often do not have accuracy of better than 5-6 km, but are often used in models as single latitude/longitude points. If the climate grid is at 1 km resolution, the actual point represented by those coordinates could in reality be in any one of 25 grid squares, so just taking the one where the point appears to be located may lead to wrong assumptions in regard to the climate where the species actually occurs; such errors will be propagated throughout the model. Too coarse a scale may not adequately delineate niche dimensions. The type of biological data available should determine the scale of climate layers used in modeling; only where occurrence data are obtained using geographic positioning systems (GPS) should fine environmental layers be used. Too often, modelers give little consideration to

scale in their selection of environmental layers, although these choices are determined more by availability of layers. Another consideration is the computing power available. If one is modeling at a continental, or broad regional scale, it may not be practical to use environmental layers at 1 km resolution that may be very large and cumbersome for computing. A grid at 5 km (2.5') occupies 1/25 the storage space of a 1 km grid, and thus requires considerably less computing power to run. If such is the case, then it is important that it be noted, and assumptions arising elaborated. Use of poor environmental layers can produce models that are of very little practical value in understanding environmental factors that drive species' niche preferences.

Table 1. Results from modeling *Raulvolfia nitida* using climate data from distinct sources and the GARP algorithm. Pixel sizes; CIAT=10 minutes, WorldClim=30 seconds.

Prediction	Presence (pixels)	Absence (pixels)
CIAT (3-10 models)	466506	2012289
Worldclim (3-10 models)	433070	2045842
(3-10 models) common to both	226119	1806438

An example was developed to demonstrate the effects of modeling using two different sets of environmental layers--one the 10' (~18 km) grid-based climate layers derived from CIAT (Jones 1991), and the other the 30" WorldClim climate layers (Hijmans 2004a, b). We used *Rauvolfia nitida* (Apocynaceae), a shrub species that occurs in the islands of Central America. Model results from the Genetic Algorithm for Rule-set Prediction (Pereira 2002) showed little overall difference in the total extent of the modeled distribution. However, the finer scale climate layers allowed for improved delineation of the species' distribution (Figure 3C and D). Using the 10' grid, the total area identified was 466,506 km^2, compared with 433,070 km^2 using the 0.5' grid (Table 1). Models based on the finer grid tended to exclude areas where the species is unlikely to occur given the climate profile at the points of known occurrence, and identified other areas that the coarser layers missed.

The scale of environmental layers can also be a problem in assembling occurrence data before modeling. As an example, see the differences that can arise from the size of the grids alone in the coverage of the Caribbean islands (Figure 4A-C). Parts of some islands, and even some entire islands, have no climatic information at either 30' or 10' grid resolution. This effect is most evident in coastal areas, because the definition of the grids does not allow faithful overlapping of environmental information with boundaries of islands. These effects in coastal areas will strongly influence the results of modeling species occurring there, because the only environmental data used are those coinciding with occurrence points.
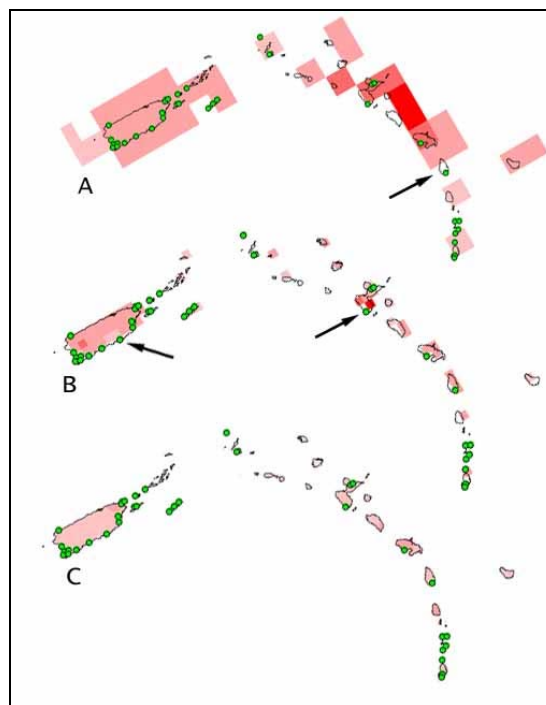

Figure 4. Annual precipitation information in the Caribbean, with occurrence points of species of *Rauvolfia* (Apocynaceae) overlain. Arrows indicate some points that fall outside of the layers. (A) IPCC dataset (30' resolution); (B) CIAT dataset (10' resolution); (C) WorldClim dataset (30" resolution).

In general, modeling uses mathematical procedures to analyze environmental parameters related to species' occurrence points, and the modeled output is produced when environmental parameters are projected onto a map showing where similar conditions are found--in other words, the potential occurrence areas of a species. Thus, to produce a good model, it is desirable to have a good sample of environmental data related with the occurrence points; in some situations, e.g., in coastal areas, a low-resolution environmental dataset will not allow large enough samples. This effect occurs

36

not only when some occurrence points have no environmental information, but also when occurrence points are highly clumped spatially. In that situation, common in the islands, many points will fall in the same grid squares, reducing the number of spatially unique points for use in analyses.

DEM – Slope and Aspect

The scale of the DEM used in a model is also very important, especially when using derived layers such as slope, aspect, and landform. If one has a DEM at a scale of 9" (~250 m), one can derive quite-valuable slope and aspect layers. If one then reclassifies upwards to a scale of, say, 1' (~2 km), then one has already reduced the meaningfulness of those two criteria—slope, for example, is unlikely to remain consistent over the 2 km. If resolution is then further eroded to 10' (~18 km), then these layers can become meaningless. For example, in Figure 5, the slope taken over 1 km is -10%, whereas over 10 km it is +2.5% and ignores major differences in terrain. Because DEMs are derived using complicated algorithms, they should be used only at the scale at which they were derived, rather than reclassified in any way.
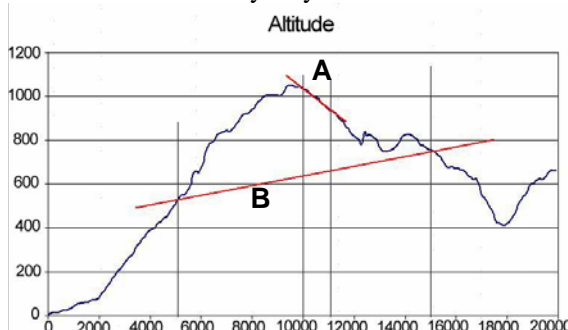

Figure 5. Graph showing problems with determining slope using different scales. **A**. shows slope across 1 km (between 10 and 11 km from coast) of –10%. **B**. shows slope over 10 km (between 5 and 15 km from coast) of +2.5%.

ISSUES, PITFALLS AND CHALLENGES IN USING ENVIRONMENT DATA

Many issues and pitfalls occur in using environmental data to model species' ecological and geographic distributions. It is just as important to evaluate the modeling approach applied as the characteristics of the data being used--both the occurrence information and the environmental data. The following are a few additional considerations of common problems and pitfalls related to environmental data that often occur in niche modeling.

- Just because a model produces a map doesn't mean it is a *good* model. "One can lie with maps just as easily as one can lie with statistics" (Wein 2002), and probably have them believed easier.
- Many species (especially rare species) occur along transition zones between vegetation types. Often these transition zones are not very well delineated in environmental data layers, particularly if the scale is coarse.
- Scale is often thought of as a step-wise process, but it is continuous. One needs also to be careful mixing environmental layers at different scales within a single model.
- Some modeling methods require continuous data as the input layers, and use of categorical data can cause problems. This requirement can be especially problematic where error inherent in the occurrence data may make the category into which a locality falls uncertain.
- By restricting the geographic boundaries of a model, one is also restricting the possible environment available for the species' modeled niche. For example, if one is modeling a broadly distributed species within just one part of its range, the climatic range of the species may be underestimated, and the resulting modeled distribution under predicted.
- Most environmental layers used for terrestrial species modeling are not necessarily suitable for modeling aquatic species. The habitat requirements of a fish, for example, may be more related to water temperature, pH and stream flow than it is to the air temperature or precipitation in a local area. Indeed, precipitation *up-stream* may be more important to the species' distribution than precipitation at the actual location.
- Modeling marine organisms presents yet a different challenge. Most data available for marine ecosystems (other than depth and bathymetry) are measured in the top 30 cm of the ocean, whereas many marine organisms move through a range of environments and depths.

CONCLUSION

Selection of environmental data layers is among the most important aspects of ecological niche modeling. A model is, at best, only as good as the input data, be they the occurrence data or the environmental layers. Too often, models are run after hours and hours of data cleaning of the occurrence data, but with blind acceptance of the environmental layers used. The occurrence data may only be used once, but the environmental data layers are used over and over in model after model, and their accuracy and nature will

determine the reliability of all resulting models. It is important that these be evaluated critically for quality before use; otherwise, resulting models will not adequately reflect the true ecological niche or potential distribution of the species being modeled.

REFERENCES

Allen, R., G., L. S. Pereira, D. Raes, and M. Smith 1998. Crop evapotransporation – Guidelines for computing crop water requirements. FAO Irrigation and drainage paper 56. FAO, Rome. [cited 6 August 2003][3].

Arora, M. K., V. Patel, and M. L. Sharma. 2002. SAR Interferometry for DEM Generation [online]. GIS Development, Technology, Remote Sensing. [cited 13 January 2004][4].

Austin, M. P., A. O. Nicholls, and C. R. Margules. 1990. Measurement of the realised qualitative niche of plant species: examples of the environmental niches of five *Eucalyptus* species. Ecol. Monogr. 60:161-177.

Bannerman, B. S. 1999. Positional Accuracy, Error and Uncertainty in Spatial Information. Geoinnovations, Howard Springs, NT, Australia. [cited 13 January 2004][5].

Barringer, J. R. F., and L. R. Lilburne. 2000. Developing fundamental data layers to support environmental modeling in New Zealand: progress and problems *in* 4th International Conference on Integrating GIS and Environmental Modelling (GIS/EM4): Problems, Prospects and Research Needs. September 2-8, in Banff, Alberta, Canada 2000. [cited 26 May 2004].[6]

Booth, T. H. 1996. Matching trees and sites *in* Proceedings of an international workshop held in Bangkok, Thailand, 27-30 March 1995, ACIAR Proceedings No. 63.

Booth, T. H., and P. G. Jones. 1996. Climatic Databases for use in Agricultural Management and Research *in* Proceedings of the Arendal II Workshop on UNEP/GRID and CGIAR cooperation to meet requirements for the use of digital data in agricultural management and research. Arendal, Norway 9-11 May 1995. [cited 13 January 2004].[7]

Burrough, P. A., and R. A. McDonnell. 1998. Principles of Geographical Information Systems. Oxford University Press, Oxford.

Busby, J. R. 1991. BIOCLIM – a bioclimatic analysis and prediction system. Pp. 4-68 *in* C. R. Margules, and M. P. Austin (eds) Nature Conservation: Cost Effective Biological Surveys and data Analysis. CSIRO, Melbourne.

Chapman, A. D., and J. R. Busby. 1994. Linking plant species information to continental biodiversity inventory, climate and environmental monitoring. Pp. 177-195 *in* R. I. Miller (ed.). Mapping the Diversity of Nature. Chapman and Hall, London.

Chapman, A. D., and D. J. Milne. 1998. The Impact of Global Warming on the Distribution of Selected Australian Plant and Animal Species in relation to Soils and Vegetation. Environment Australia, Canberra.

CIAT n.d. FloraMap Users Manual. Centro Internacional de Agricultura Tropical (CIAT), Cali, Colombia.

Cunningham, D., K. Walsh, and E. Anderson. 2001. Potential for Seed Gum Production from *Cassia brewsteri*. RIRDC Project No. UCQ-12A. Rural Industries Research and Development Corporation, Kingston, ACT. [cited 10 February 2004].[8]

Daly, C., R. P. Nielson, and D. L. Phillips. 1994. A statistical-topographic model for mapping climatological precipitation over mountainous terrain. J. Appl. Meteor. 33:140-158.

Daly, C., W. Gibson, M. Doggett, J. Smith, and G. Taylor. 2004. A probabilistic-spatial approach to the quality control of climate observations. *In* Proceedings of the 14th AMS Conference on Applied Climatology, Amer. Meteorological Soc., Seattle, WA. January 13-16, 2004. [cited 26

---

[3] http://www.fao.org/docrep/X0490E/x0490e00.htm#Contents.
[4] http://www.gisdevelopment.net/technology/rs/techrs0021.htm.
[5] http://www.geoinnovations.com.au/posacc/patoc.htm.
[6] http://www.colorado.edu/research/cires/banff/pubpapers/221/.
[7] http://www.grida.no/cgiar/htmls/climpres.htm.
[8] http://www.rirdc.gov.au/reports/NPP/UCQ-12A.pdf.

May 2004].[9]

Department of Environment and Heritage (DEH). 2004. Normalized Difference Vegetation Index. [cited 26 February 2004].[10]

Faith, D. P., and P. A. Walker. 1996. Environmental diversity: on the best possible use of surrogate data for assessing the relative biodiversity of sets of areas. Biodiv. and Conserv. 5:399-415.

Faith, D. P., P. A. Walker, C. R. Margules, J. Stein, and G. Natera. 2001. Practical application of biodiversity surrogates and percentage targets for conservation in Papua New Guinea. Pacific Conservation Biology 6:289-303. [cited 10 February 2004].[11]

Ferrier, S., and G. Watson. 1997. An Evaluation of the Effectiveness of Environmental Surrogates and Modelling Techniques in Predicting the Distribution of Biological Diversity. Department of Environment, Sport and Territories, Canberra.

Gallant, J. C., and M. F. Hutchinson. 1996. Towards an understanding of landscape scale and structure *in* Proceedings of the Third International Conference Integrating GIS and Environmental Modeling. University of California, Santa Barbara, National Center for Geographic Information and Analysis: CD-ROM [cited 6 August 2003].[12]

Geoscience Australia. 2003. Geodata TOPO-250K (Series 1) Topographic Data [online]. Geoscience Australia, Canberra. [cited 14 January 2004].[13]

Gessler, P. E., I. D. Moore, N. J. McKenzie, and P. J. Ryan. 1995. Soil-landscape modelling in southeastern Australia *in* M. Goodchild et al. (eds) *GIS and* Environmental Modeling: Progress and Research Issues*. GIS World Books.

Gessler, P. E. and O. A. Chadwick. 1997. Quantitative Soil-landscape Modeling: A Key to Linking Ecosystem Processes on Hillslopes. Pedometrics 97. [cited 19 September 2004].[14]

Gibson, W., C. Daly, T. Kittel, D. Nychka, C. Johns, N. Rosenbloom, A. McNab, and G. Taylor. 2004. Development of a 103-year high-resolution climate data set for the conterminous United States *in* Proceedings 13th AMS Conference on Applied Climatology, Amer. Meteorological Soc., Portland, OR, May 13-16, 2002. pp. 181-183. [cited 26 May 2004].[15]

Goovaerts P. 2000. Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. J. Hydrol. 228:113-129.

Hargreaves, G. H., and Z. A. Samani. 1982. Estimating potential evapotranspiration. Tech. Note, J. Irrig. and Drain. Engrg., ASCE. 108:225-230.

Hartkamp, A. D., K. De Beurs, A. Stein, and J. W. White. 1999. Interpolation Techniques for Climate Variables. NRG-GIS Series 99-01. CIMMYT, Mexico, D.F. [cited 17 October 2003].[16]

Hijmans, R., J., S. Cameron, and J. Para. 2004a. WorldClim, version 1.2. A square kilometer resolution database of global terrestrial surface climate. [cited 26 May 2004].[17]

Hijmans, R., J., S. Cameron, and J. Parra. 2004b. WorldClim, a new high-resolution global climate database. Abstract: Inter-American Workshop on Environmental Data Access. Campinas, Brazil. Mar 2004. [cited 26 May 2004].[18]

Holdaway, M. R. 1996. Spatial modelling and interpolation of monthly temperature using kriging, Climate Research 6:215-225.

Hutchinson, M. F. 1989. A new procedure for gridding elevation and stream line data with automatic removal of spurious pits. J. Hydrol. 106:211-232.

Hutchinson, M. F. 1991. The application of thin plate smoothing splines to continent-wide data assimilation. Pp. 104-113 *in* J. D. Jasper (ed.), Data Assimilation Systems. Bureau of Meteorology Research Report No. 27, Bureau of Meteorology, Melbourne.

Hutchinson, M. F. 1995. Interpolating mean rainfall using thin plate smoothing splines. Int. J. GIS 9:305-403.

Hutchinson, M. F. 1996. A locally adaptive approach to the interpolation of digital elevation models *in* NCGIA (ed.), Proceedings of the Third International Conference Integrating GIS and Environmental Modeling. Santa Fe, New Mexico, 21-25 January, 1996. University of California, Santa Barbara, National Center for Geographic Information and Analysis: CD-ROM [cited 16 October 2003].[19]

Hutchinson, M. F. 2001. *ANUSPLIN Version 4.3*. [online]. Centre for Resource and Environmental Studies, Australian National University, Canberra. [cited 6 August 2004].[20]

Hutchinson, M. F. 2003a. *Topographic and Climate Database for Africa. Version 1.0*. [online]. Centre for Resource and Environmental Studies, Australian National University, Canberra. [cited 13 January 2004].[21]

[9] http://www.ocs.orst.edu/pub/prism/docs/applclim04-probalistic_spatial_approach-daly.pdf.
[10] http://www.deh.gov.au/erin/ndvi/ndvi.html.
[11] http://wwwscience.murdoch.edu.au/centres/others/pcb/toc/pcb_contents_v6.html.
[12] http://www.ncgia.ucsb.edu/conf/SANTA_FE_CD-ROM/sf_papers/gallant_john/paper.html.
[13] http://www.ga.gov.au/nmd/products/digidat/250k.htm.
[14] http://www.essc.psu.edu/pedometrics/abstracts/html/gessler.html.
[15] http://culter.colorado.edu/~kittel/pdf_Gibson02_AMS.pdf.

[16] http://www.cimmyt.org/Research/nrg/pdf/NRGGIS%2099_01.pdf.
[17] http://biogeo.berkeley.edu/.
[18] http://www.cria.org.br/eventos/iaed/rhijmans_pre.html.
[19] http://www.ncgia.ucsb.edu/conf/SANTA_FE_CD-ROM/sf_papers/hutchinson_michael_dem/local.html and <http://edcdaac.usgs.gov/gtopo30/papers/local.html.
[20] http://cres.anu.edu.au/outputs/anusplin.php.
[21] http://cres.anu.edu.au/outputs/africa.html.

Hutchinson, M. F. 2003b. *Upgrade of the 9 Second Australian Digital Elevation Model*. [online]. Centre for Resource and Environmental Studies, Australian National University, Canberra. [cited 13 January 2004].[22]

Hutchinson, M. F. 2004. *ANUDEM Version 5.1*. [online]. Centre for Resource and Environmental Studies, Australian National University, Canberra. [cited 6 August 2004].[23]

Jones, P. G. 1991. The CIAT Climate Database. Version 3.41. Machine readable dataset. CIAT, Cali, Colombia.

Jones, P. G., D. M. Robison, and S. E. Carter. 1990. A geographical information approach for stratifying tropical Latin America to identify research problems and opportunities in natural resource management for sustainable agriculture *in* Centro Internacional de Agricultura Tropical (CIAT). CIAT, Cali, Colombia.

Kesteven, J., and M. Hutchinson. 1996. Spatial modelling of climate variables on a continental scale. Proceedings of the Third International Conference Integrating GIS and Environmental Modeling. University of California, Santa Barbara, National Center for Geographic Information and Analysis: CD-ROM [cited 6 August 2003].[24]

Liew, S. C. 2001. Principles of Remote Sensing. Centre for Remote Imaging, Sensing and Processing, National University of Singapore [cited 8 June 2004].[25]

Liu, H. K., B. Li Jezek, and Z. Zhao. 2001. Radarsat Antarctic Mapping Project Digital Elevation Model Version 2. NDSIC, Boulder, Colorado. [cited January 2004].[26]

Longmore, R. (ed.) 1986. Atlas of Elapid Snakes of Australia. Australian Flora and Fauna Series No. 7. Australian Government Publishing Service, Canberra.

Mackey, A.P. (ed.) 1996. Prickly Acacia (*Acacia nilotica*) in Queensland. Queensland Department of Natural Resources and Mines, Brisbane. [cited 10 February 2004].[27]

Mao, Z., P. D. Huang, and H. Huang. 1999. Automatic registration of SeaWIFS and AVHRR Imagery *in* G. Xu and Y. Chen (eds). Towards Digital Earth. Proceedings of the International Symposium on Digital Earth. Science Press.

Margules, C. R., and T. D. Redhead. 1995. BioRap. Guidelines for using the BioRap Methodology and Tools. CSIRO, Canberra.

Margules, C. R., and R. L. Pressey. 2000. Systematic Conservation Planning. Nature 405:243-253.

Moore, I. D., P. E. Gessler, G. A. Nielsen, and G. A. Peterson. 1993. Soil attribute prediction using terrain analysis. Soil Sci. Soc. Am. J. 57:443-452.

Morton, F. I. 1983. Operational estimates of areal evapotranspiration and their significance to the science and practice of hydrology. J. Hydrol. 66:1-76.

Nalder, I. A., and R. W. Wein. 1998. Spatial interpolation of climatic normals: test of a new method in the Canadian boreal forest. Agric. For. Meteorol. 9:211-225.

NGDC. 2000. Global Land One-kilometer Base Elevation (GLOBE) Digital Elevation Data Version 1.0. [online]. NGDC, USA. [cited 14 January 2004].[28]

Nicholls, N. 1997. Increased Australian wheat yield due to recent climate trends. Nature 387:484-485.

Nix, H. A. 1986. A biogeographic analysis of Australian elapid snakes *in* R.C. Longmore (ed). Atlas of Australian elapid snakes. Australian Flora and Fauna Series No. 7:4-15. Australian Government Publishing Service, Canberra.

Noonan, G. R. 2003. GIS technology: A powerful tool for entomologists. Insight 1:1-98. [cited 27 May 2004].[29]

Olsen, L. M., and N. B. Bliss. 1997. Development of a 30-arc-second Digital Elevation Model of South America *in* ESRI International User Conference, San Diego, California July 8-11, 1997. [cited 8 January 2004].[30]

Pereira, R. S. 2002. Desktop Garp. University of Kansas Biodiversity Research Center, Lawrence, Kansas.

Peterson, A. T., and D .A. Vieglais. 2001. Predicting species invasions using ecological niche modeling. BioScience 51:363-371.

Peterson, A. T., J. Soberón, and V. Sánchez-Cordero. 1999. Conservatism of ecological niches in evolutionary time. Science 285: 1265-1267.

Peterson, A. T., M. A. Ortega-Huerta, J. Bartley, V. Sánchez-Cordero, J. Soberón, R .H. Buddemeier, and D. R. B. Stockwell. 2002. Future projections for Mexican faunas under global climate change scenarios. Nature 416: 626-629.

Phillips, D. L., J. Dolph, and D. Marks. 1992. A comparison of geostatistical procedures for spatial analysis of precipitation in mountainous terrain, Agric. For. Meteorol.. 58:119-141.

Price, D. T., D. W. McKenney, I. A. Nalder, M. F. Hutchinson, and J. L. Kesteven. 2000. A comparison of two statistical methods for spatial interpolation of Canadian monthly mean climate data. Agric. For. Meteorol. 101:81-94.

Rao, C. R. N., and J. Chen, 1995. Inter-satellite calibration linkages for the visible and near-infrared channels of the Advanced Very High Resolution Radiometer on the NOAA-7, -9, and -

[22] http://cres.anu.edu.aqu/dem/.
[23] http://cres.anu.edu.au/outputs/anudem.php.
[24] http://www.ncgia.ucsb.edu/conf/SANTA_FE_CD-ROM/sf_papers/kesteven_jennifer/jlkpaper.html.
[25] http://www.crisp.nus.edu.sg/~research/tutorial/rsmain.htm.
[26] http://nsidc.org/data/docs/daac/nsidc0082_ramp_dem_v2.gd.html.
[27] http://www.nrm.qld.gov.au/pests/psas/pdfs/Pricklyacacia.pdf.

[28] http://www.ngdc.noaa.gov/seg/fliers/globedem.shtml.
[29] http://www.mpm.edu/cr/insight/Papers/Noonan GIS review.pdf.
[30] http://gis.esri.com/library/userconf/proc97/proc97/to400/pap382/p382.htm.

11 spacecraft. Int. J. Rem. Sens. 16:1931-1942.

Rao, C. R. N., and J. Chen. 1996. Post-launch Calibration of the Visible and Near-infrared Channels of the Advanced Very High Resolution Radiometer on NOAA-14 Spacecraft. [cited 26 February 2004].[31].

Rao, K. S., and Y. S. Rao. 1999. Seminar on Recent Developments in Differential SAR interferometry and its applications. Lecture Notes, IIT, Bombay 1-11.

Reynolds, J. F., P. R. Kemp, and J. D. Tenhunen. 2000. Effects of long-term rainfall variability on evapotranspiration and soil water distribution in the Chihuahuan Desert: a modeling analysis. Plant Ecol. 150:145-159. [cited 18 April 2005][32].

Sattler, P., and R. Williams (eds). 1999. The Conservation of Queensland's Bioregional Ecosystems. Environment Protection Authority, Brisbane, Qld.

Sobel, D. 1995. Longitude: The True Story of a Lone Genius Who Solved the Greatest Scientific Problem of His Time. Walker Publishing Co.

Soberón, J., J. Golubov, and J. Sarakhán. 2000. Predicting the effects of *Cactoblastis cactorum* Berg on the *Platyopuntia* of Mexico: A model on the route of invasion. Pp. 95-97 *in* Assessment and Management of Alien Species that Threaten Ecosystems, Habitats and Species. CBD Technical Series No. 1. Montreal, Canada: Convention on Biological Diversity. [cited 10 February 2004].[33].

Thackway, R., and I. Cresswell. (eds) 1995. An Interim Biogeographic Regionalisation for Australia: A Framework for Setting Priorities in the National Reserves System Cooperative Program. (Version 4.0). Australian Nature Conservation Agency, Canberra [cited 10 February 2004].[34].

USGS. 2003. GTOPO30 Documentation. USGS-NASA. [cited 16 October 2003].[35].

Wang, Q. J., F. L. N. McConachy, F. H. S. Chiew, R. James, G. C. de Hoedt, and W. J. Wright. 2004. Climate Atlas of Australia Maps of Evaporation. Bureau of Meterology, Canberra. [cited 6 August 2004].[36].

Wein, J. 2002. Predicting species occurrences: Progress, problems, and prospects. Pp. 739-749 *in* M. J. Scott, et al. (eds). Predicting Species Occurrences: Issues of Accuracy and Scale. Island Press, Washington, D.C.

Wieczorek, J. 2001. MaNIS: Georeferencing Guidelines. University of California, Berkeley –

MaNIS. [cited 26 January 2004].[37].

Wikipedia. 2004. Wikipedia the free encyclopedia. [cited 26 May 2004].[38].

Williams, P. H., C. R. Margules, and D. W. Hilbert. 2002. Data requirements and data sources for biodiversity priority area selection. J. Biosc. 27:327-338.

Zheng X., and R. E. Basher. 1996. Spatial modelling of New Zealand temperature normals. Int. J. Climatol. 16:307-319.

---

[31] http://orbit-net.nesdis.noaa.gov/ora/text/nrao02.txt.
[32] http://www.ees.nmt.edu/vivoni/ecohydro/pdf/jornadarainfallvariabiliity.pdf.
[33] http://www.biodiv.org/doc/publications/cbd-ts-01.pdf.
[34] http://www.deh.gov.au/parks/nrs/ibra/version4-0/framework/index.html.
[35] http://edcdaac.usgs.gov/gtopo30/README.html.
[36] http://www.bom.gov.au/climate/averages/climatology/evapotrans/text/et-txt.shtml.

[37] http://dlp.cs.Berkeley.edu/manis/GeorefGuide.html.
[38] http://en.wikipedia.org.