Report No. 7 **MODELLING**

Lifemapper
Comments and Ideas

Arthur D. Chapman
Last modified 23 July 2003

Aim:
To comment on the Lifemapper (University of Kansas 2003) software and its implementation on the Internet.

Constraints:
This paper looks at application of the Lifemapper screen-saver software and its implementation on the Internet. It does not attempt to examine in detail the algorithms used, but identifies a number of issues I have observed from a user point of view.

The comments herein are not meant as a criticism in any way of Lifemapper, but are made in the spirit of cooperation and collaboration in order to improve what is the basis of a very good system.

Background:
Lifemapper was developed in the late 1990s to test the applicability of modelling on the internet, and the use of distributed processing for carrying out such modelling. It began from modest beginnings.

The first modelling of species on the internet was by the Environmental Resources Information Network (ERIN) in Australia as early as 1994. From that early internet modelling arose GARP (Genetic Algorithm for Rule-set Production) (Stockwell and Noble 1992, Payne and Stockwell n.dat.) a program developed by David Stockwell, now at the San Diego Supercomputer Center in the USA. More recently, a desktop version of GARP has been written (Scachetti-Pereira 2002), and that forms the basis of the modelling behind Lifemapper.

From GARP, and the Species Analyst Program (Vieglas *et al.* 1998, Vieglas 1999) arose Lifemapper (Beach *et al.* 2002, University of Kansas 2003). Lifemapper is a truly innovative program, and is an early example of distributed processing to museum biodiversity informatics.

Data:
Species Data
Most species' distribution models use either presence-only data (including records from herbaria or museums and observation data) or presence-absence data from systematic surveys. Most of these data are point-based, although some models also include area-based or grid-based data. All species' collection data are samples of geographic space and inevitably incorporate some degree of spatial bias (Williams *et al.* 2002). Sample sites are generally subsets of sites where the species actually occur and there are few, if any records of where a species may have been looked for, but not found (Margules and Austin 1994).

Plant and animal specimen data held in museums and herbaria provide a vast information resource, providing not only present day information on the locations of these entities, but also historic information going back several hundred years (Chapman and Busby 1994).

These data do, however, have some major drawbacks when it comes to use for modelling distributions. Many of the records carry little geographic information other than a general description of the location where they were collected (Chapman and Milne 1998), the geocoding (latitude and longitude) where given is often not that accurate, especially with historic data (Chapman 1999) and is often added at a later date by those other than the collector (Chapman 1992). It can only supply information on the presence of the entity at a particular time and says nothing about absences in any other place or time (Peterson *et al.* 1998), and it is usually collected opportunistically rather than systematically (Chapman 1999, Williams *et al.* 2002) resulting in large biases – for example, collections that are highly correlated with road networks (Chapman 1999, Peterson and Stockwell 2002).

We cannot neglect this data, however, as it constitutes the largest database of biological information we are ever likely to have. It is estimated that there over three billion records (OECD 1999) worth around $135-150 billion held in this form around the world (Armstrong 1992). The cost of replacing these data with new surveys would be prohibitive. It is not unusual for a single survey to exceed $1 million to conduct (Burbidge 1991). They are an essential resource in any effort to conserve the environment, as they provide the only fully documented record of the occurrence of species in areas that may have undergone habitat change due to clearing for agriculture, urbanization, climate change, or been modified in some other way (Chapman 1999).

Fig. 1. *An example of a Lifemapper application being run as a Screensaver or desktop application (University of Kansas 2003b).*

The Lifemapper project relies on presence-only data, and almost exclusively on specimen data extracted from museums and herbaria. There is a large bias in the data as, for some groups at least, only a few institutions are at present participating by supplying data. For example if one selects "Plants" as one's preferred option, one invariable only receives data from the University of Kansas collections, mainly collected around the US State of Kansas (see, for example, fig. 1). Widespread species such as Taraxicum officinale, Cannabis sativa, etc. provide only records from a small part of the US, and thus the data, by being highly biased, leads to a heavily skewed model. Perhaps a note to this effect may be worth adding as part of the caveat.

In spite of the claims that we (the users) are helping by modelling thousands of species, are models as biased as some of these, worth doing at all? Perhaps a choice on the member profile could be "only species with data from more than 5 institutions" or similar could be added. This would give priority to the modelling of species for which there is an adequate representation and which that may prove of some value, rather than models that may have little no real value.

Environmental data

Just as important as the species data, is the environmental layers used to model the species against. The theory of most environmental models is that species have certain habitat preferences that have an environmental basis. Many models use climatological information such as temperature, rainfall, radiation, evaporation, soil moisture etc. as the basis on which to broadly define the habitat or ecological niche. Other models use vegetation characteristics such as vegetation classes, detailed habitat information, correlated species, etc.

One of the important considerations in choosing appropriate environmental layers is that of scale. Too fine a scale will lead to errors due to mismatching with the biological data being modelled against it. Too coarse a scale may not adequately delineate the appropriate environmental niches. Too often modellers give little consideration to scale in their selection of environmental layers. This may, of course be due to availability as they may only be able to use those layers that are available at the time. An obvious consideration with Lifemapper is the size of the data sets, and the time required to model a species. Layers that are "too fine" will result in models taking hours to produce, and this may be counterproductive by discouraging potential users of the system (however this need not be the case with a better implimentation of the modelling software – see below). On the other hand, layers that are too coarse have very little meaning. For example, slope or aspect in a half-degree grid (ca. 60 km) is a meaningless concept. A scale around 3-arc minutes (ca. 5-6 km) is about ideal for modelling species distributions at a continental scale. Most of the species data can be regarded as being any better than this on average, so anything finer does not make a lot of sense other than for modelling at a regional scale. Anything courser than about 10 minutes does not adequately delineate the environmental characteristics, and even this is too course in most instances. Environmental layers do exist, or are being developed, for most of the world (with the exception of South America) at 3-arc minute grid resolution. A separate report to CRIA (Chapman 2003a) recommends the development of consistent 3-arc minute climate grids for South America as a matter of priority.

The present implementation of Desktop GARP (Scachetti-Pereira 2002) as used in Lifemapper is, as I understand it, inefficient in that it imports the total environmental grid space and maps each cell in the analysis, rather than just importing those parts being used during the analysis and only using the whole grid at the mapping and visualisation stage as is done in other implementations of GARP (Payne and Stockwell n.dat.). This means that using the methodology in Desktop GARP, a fine scale environmental grid will cause a slowing down in the processing of the model. This version of GARP perhaps needs modification to allow for use of much finer scale environmental layers than are used at present. Lifemapper will also need to address this issue.

The choice of layers that have some meaning from an environmental point of view is also important. I believe the selection of just one or two months as examples of temperature or rainfall makes very little sense biologically as I will explain more fully later.

Data Accuracy
Errors in data are common and to be expected. Errors in species' data are particularly common and need to be catered for. Errors in spatial position (geocoding) and in taxonomic circumscription are two of the major causes of error in modelling. Assessment of the accuracy of input data is essential otherwise the results of any modelling will be meaningless. Correcting errors in data and weeding out the bad records is a time consuming and tedious process (Williams *et al.* 1994). A detailed discussion of data error is beyond the scope of this paper and is the subject of separate reports to CRIA (Chapman 2003c, 2003d).

Although one can produce a perfectly good looking model from poor data; a model that may even be regarded as being "better than random", it does not mean that the model has any meaning in reality when predicting the occurrence or non-occurrence of a particular species. The better and more accurate the input data, the better the resultant model is likely to be.

Lifemapper:
Below I have expanded on the Layers used in Lifemapper and provided some comments.

1. Scale

> "They were originally on a scale of 0.5 degree cells, and were later generalized to 1 km cells to match the scale of terrain and other data sets." (University of Kansas 2003)

Climate data is available for the whole of the earth's surface at 10-minute grid resolution (Hijmans 1999, New *et al.* 2002). If one is going to generalize to a 1 km grid, why not use the 10-min data rather than the 0.5 degree data (a improvement in resolution of 9 times). I have seldom seen data as mixed in scale as the Lifemapper layers are, prove very successful in modelling. The use of 1km would, to me seem to be too fine for most of the biological data. I would be tempted to generalize the 10-minute climate data to about 3-minute resolution and use either Cubic Convolution or Bilinear Interpolation algorithms to reclassify the terrain data to the same scale. However, because of the scale issue with slope and aspect mentioned above, using terrain at 1 km and the climate at 10 minute may be a reasonable compromise. They should, however, be brought to the same coordinate system, one being area based and the other geographic.

2. Climate layers

> "The main climate parameters used by Lifemapper are:
>
> - Cloud cover
> - Diurnal temperature range
> - Ground-frost frequency
> - Maximum temperature
> - Mean temperature
> - Minimum temperature
> - Precipitation
> - Solar radiation
> - Vapor pressure
> - Wet-day frequency
> - Winds

> All those parameters are used as year averages from 1961 to 1990, and also the averages for the months of January and July for the same period." (University of Kansas 2003)

The use of just two months (***January and July***) makes very little sense to me for a number of reasons. January and July are not necessarily the most relevant or important months in many parts of the world. This has been recognised in a number of modelling programs, for example FloraMap (Jones and Gladkov 2001) and ANUCLIM (Houlder *et al*. 2000). Both treat the problem in different ways, however, both have recognised it as an issue.

One needs to think about what the biological entities (plants or animals) are responding to. For the example here, I will use plants. One of the most important factors on determining where a plant may grow is the relationship between rainfall and temperature. Agronomists have relied on this knowledge for thousands of years, and hence summer and winter plantings of different crops. Some species respond to rainfall at certain times of the year (Spring or Autumn are common), others at other times. Rainfall in the middle of winter has little or no effect on a species that is dormant during that period. In more tropical areas, however, this may be the key growing period, as many plants shut down to 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 a much larger detrimental effect on a plant than the same long dry period during the middle of winter.

For these reasons, and as a result of many years of modelling experience, the developers of ANUCLIM incorporated a number of environmental layers that have proved (over 30 years of modelling) to be key driving forces for a large majority of species' distributions. These layers are:

- Rainfall of the hottest quarter
- Rainfall of the coolest quarter
- Mean temperature of the wettest quarter
- Mean temperature of the driest quarter
- Rainfall of the wettest quarter
- Rainfall of the driest quarter
- Mean temperature of the warmest quarter
- Mean temperature of the coolest quarter

If one wished to reduce the number of layers, then I would suggest just using the first four.

It has also been found that using the three-monthly (quartile) periods has more meaning than using just single months.

Using this concept, I recently created layers for South America at 10-minute resolution and mapped the results (fig.2). You begin to see distinct summer/winter rainfall patterns emerging that are not obvious from using just the January and July monthly figures.

Figure 2 also shows some areas where the driest three months occur in the same period as the wettest. This may appear an anomaly, but the three wettest months could be January, February, March, and the three driest March, April and May. This does occur in some areas with very low rainfall - for example, in some of the desert areas in north-eastern Brazil.

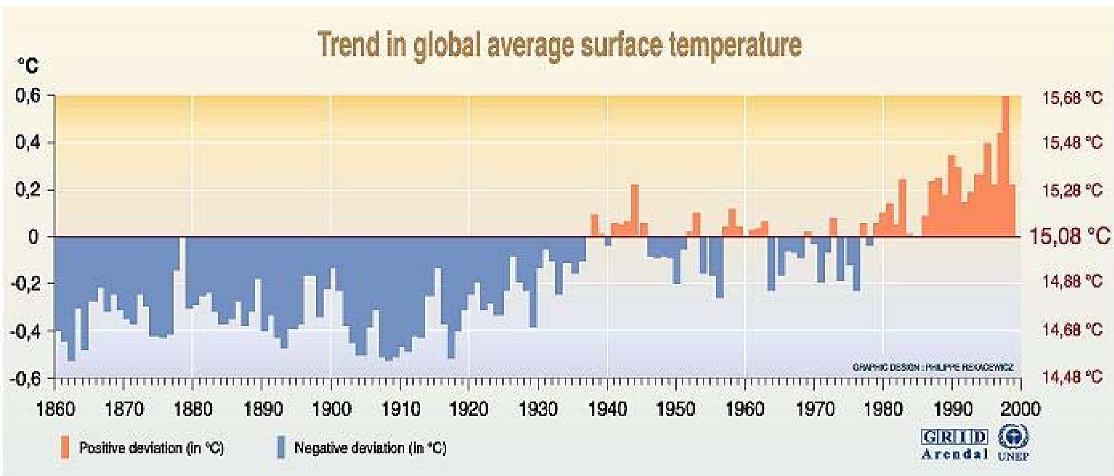
By using just January and July layers you also tend to negate the possibility of modelling between the northern and southern hemispheres. If you have a species with most records in the northern hemisphere, for example, and develop your rules (i.e. it likes a hot July), then applying those rules to the southern hemisphere won't work, where July is the cool month.


A


B

Fig.2. *Maps of South America showing when the driest (**A**) and wettest (B) three months occur – red = summer (three months beginning Nov, Dec, Jan), brown = autumn (three months beginning Feb, Mar, Apr), blue = winter (three months beginning May, Jun, Jul) and green = spring (three months beginning Aug, Sep, Oct).*

Wind is a layer that has proved of very little value in modelling in most parts of the world. In the USA, for example there are many meteorological stations that record wind (speed, direction, etc.). In much of the rest of the world, however, such stations are very sparse. To develop surfaces for wind, such criteria as distance from coast have been tried, along with slope and aspect, however these prove quite unreliable in developing robust climate surfaces, especially once the distance from the coast is greater than ca. 250 km. Also, wind has very little impact on species occurrences except in a very few areas (southern Argentina, high mountains, coastal areas and islands). The climate in these areas is generally delineated by the other layers being used, so I see no good reason for including it. There are other more important layers that could be included.



Source: School of environmental sciences, climatic research unit, university of East Anglia, Norwich, United Kingdom, 1999.

Fig. 3. Combined land-surface and sea-surface temperatures (degrees Centigrade) 1861-1998, relative to the temperature between 1961 and 1990 (from UNEP 2003).

One layer not used in Lifemapper, that often proves of great value is the Coefficient of Variation of Rainfall. It is easily calculated from the raw data.

The use of a climate period of 1961-1990. This period covers the beginning of a climate change response to greenhouse (see fig. 3) and is not truly representative of the longer-term climate for many species. It may be perfectly satisfactory for short-lived species (annuals, short-lived perennials, animals, etc.), however there may be problems using such layers for longer-lived tree species. Just because a tree lives in a particular location in today's climate, does not mean that that is its preferred climate. For example, the Huon Pine in Tasmania is a tree that lives for 2000 years. The climate satisfactory for it to thrive and reproduce may be quite different from today's climate. It may still be growing at point 'X', but is it reproducing, or just surviving? By modelling the species using the climate of the last thirty years, we are making some massive assumptions. Similarly, many wrong assumptions are made about climate-change and its affects on species' distributions. Often I hear statements such as 'this species will move …', or 'the rainforest will move …', but what is more likely is that the rainforest may well survive for many hundreds of years in its existing location, even though the climate may not be 'satisfactory', but it will gradually decline as the composite species fail to reproduce. Some seedlings may thrive on the edges in new areas, but reproduction will not keep up with current rates of climate change and whole vegetation types will not just 'move' from one place to another. This concept also applies to past changes in climate.

Modelling of fish distributions using atmospheric climate layers. Another concern I have is the use of these layers for modelling fish distributions. I'd be interested to hear what fish experts say about the results (although from the scale of the modelling it may be difficult to say a lot). Atmospheric climate often has little to do with the location of fish species, whereas criteria such as water temperature, oxygen content, pH, stream flow, etc., etc. have a far greater influence. Water temperature may have little relationship to atmospheric temperature in the area. It may be more related to the temperature hundreds of kilometres upstream – in the mountains, for example. Also flow-rate may be totally unrelated to the rainfall of the area, but directly related to rainfall hundreds of kilometres upstream. Attempts to model fish species in Australia, using climate layers such as these, has proved totally unreliable, and of little value.

3. Terrain-related layers

> "The terrain related layers were obtained from the United States Geological Survey - USGS on a scale of 1:250.000. Those layers were later processed to grids with 1 km cells.

- **Aspect**: describes the direction of maximum rate of change in the elevations between each cell and its eight neighbors. It can essentially be thought of as the slope direction;
- **Flow directions**: defines the direction of flow from each cell in the Digital Elevation Model to its steepest down-slope neighbor;
- **Flow accumulation**: defines the amount of upstream area draining into each cell. It is essentially a measure of the upstream catchment area. The flow direction layer is used to define which cells flow into the target cell;
- **Slope**: describes the maximum change in the elevations between each cell and its eight neighbors. The slope is expressed in integer degrees of slope between 0 and 90;
- **Compound Topographic Index** or **CTI**: it is commonly referred to as the **Wetness Index**, is a function of the upstream contributing area and the slope of the landscape. The CTI is calculated using the flow accumulation (FA) layer along with the slope. In areas of no slope, a CTI value is obtained by substituting a slope of 0.001. This value is smaller

than the smallest slope obtainable from a 1000 m data set with a 1m vertical resolution." (University of Kansas 2003).

I have no problem with these layers. The Wetness Index, linked to soil texture can produce another valuable layer in Soil water-holding Capacity. It may be worth looking at.

4. Vegetation and land-use

"Vegetation and land use data was obtained from **University of Maryland** and it comprises the layers:

- Percentage of tree cover
- Land use land cover type" (University of Kansas 2003).

My main comment here is related to what you are trying to model. Is it historic distribution or present day distribution? If you want to show the species range prior to clearing, one needs to use a "pre-European" or "pre agricultural revolution" land cover layer. If, on the other hand, you want to use a current day distribution, then it will depend on the organism being modelled. If it is a forest tree species, then it won't occur where the land has been thoroughly cleared (for wheat cropping, urbanisation, etc.), in which case you want that layer to be over-riding and thus best used as a GIS layer at the end to remove cleared areas from the modelled distribution. In other cases, the species may thrive perfectly well in cleared areas (weeds, grasses along roadsides and fence lines, various insect species, etc.).

I have always found these layers better used in a GIS as a clip layer after modelling. I have not found them very successful when used within the model themselves. More stable layers – soil texture, surficial geology, etc. may be better layers to include, although these may be difficult to obtain as consistent layers for the whole world.

5. GARP

There are a number of issues I have raised with respect to the GARP method of modelling in a separate report to CRIA (Chapman 2003b). In that report I suggested a number of ways that I believe that GARP could be improved as a modelling tool. These include the development of probability surfaces rather than straight presence/absence output, use of different methods for resampling environmental layers, use of different environmental layers as discussed above, incorporation of non-climate layers such as ndvi, soil texture, soil moisture, surficial geology, etc., incorporation of 'what-if' scenarios and an improved visual output.

I see a number of directions that GARP (and hence Lifemapper) may consider for the future. These include:

a. Extension into the aquatic environment

This may be a longer-term research project, but there is a lot of research going on around the world in the fresh-water aquatic environments. It makes no sense to use broad climatic layers in a model for modelling fish species as mentioned above. The water temperature, oxygen content, pH, water flow, etc. may be far more important.

b. Extension into marine areas

Consideration could be given in the longer term to extending GARP (Lifemapper) modelling methodologies into the marine. There are very few, if any, good models available for modelling species in a marine environment. I believe that GARP would be ideal in this scenario. This is a major research project however, and would need funding.

c. Inclusion of a range of climate periods

As mentioned above with long-lived trees, and for purposes of studying climate change, I would see value in GARP in the future including options of using a series of climate layers covering several 30-year periods (1910-1940, 1940-1980, 1970-2000) as well as a 100-year period to capture variations across the century.

d. Inclusion of ndvi as a layer

Normalised Difference Vegetation Index (ndvi) – a derived layer from Remote Sensing that has a strong link to vegetation, could be included as a layer in GARP as suggested above. This could be used (without necessarily knowing what the colours actually mean) to exclude cleared areas, agricultural land, etc. The inclusion of layers from different months would be able to increasing differentiate between different vegetation types. These layers are available for the whole world at 1 km resolution (AVHRR).

Conclusion

One can lie with maps just as easily as one can lie with statistics (Wein 2002), and probably have them believed easier.

It is so easy to produce a pretty map showing a species model, but what does it actually mean to the environment. How does it reflect the environmental variables, and is it of value in understanding and managing the species or the environment. I believe that in many cases, the reason for doing modelling has been lost site of, and that the results have not been looked at critically from an environmental viewpoint. In many cases, environmental layers have been used without critical examination of why those layers are being used, and whether they are the best layers in the circumstances. A model is only one of a range of hypotheses on the possible distribution of a species or its possible range. Too often people see a model as an end product in itself – it isn't – it is merely another tool to help in our understanding of the environment and the species that make up that environment. Modelling should not just be another academic exercise!

I believe GARP has come a long way since its early days as a command driven program back in Australia in the mid 1990s. The development of Desktop GARP has improved its useability and made it available to a much wider audience. It still has some way to go of course!

Lifemapper has shown a good beginning, and statistics are showing an increasing interest in it. I believe it is at a stage, however, where it may require an extensive overview, to cover some of the issues I have mentioned above. I look forward to seeing and using future versions.

References:

Armstrong, J.A. (1992). The funding base for Australian biological collections. *Australian Biologist* **5(1):** 80-88.

Beach, J.H., Stewrat, A.M., Yorontsov, Y., Scachetti Piera, R., Stockwell, D.R.B., Vieglas, D.A. and Downie, S.R. (2002). The Future of Biodiversity: Mapping and Predicting the Distribution of Life with Distributed Computation http://gis.esri.com/library/userconf/proc02/pap0769/p0769.htm.

Burbidge, A.A. (1991). Cost Constraints on Surveys for Nature Conservation **in** Margules, C.R. and Austin, M.P. (eds). *Nature Conservation: Cost Effective Biological Surveys and Data Analysis*. Canberra: CSIRO

Chapman, A.D. (1992). Quality Control and Validation of Environmental Resource Data **in** *Data Quality and Standards: Proceedings of a Seminar Organised by the Commonwealth Land Information Forum, Canberra, 5 December 1991*. Canberra: Commonwealth land Information Forum.

Chapman, A.D. (1999). Quality Control and Validation of Point-Sourced Environmental Resource Data pp. 409-418 **in** Lowell, K. and Jaton, A. eds. *Spatial accuracy assessment: Land information uncertainty in natural resources*. Chelsea, MI: Ann Arbor Press.

Chapman, A.D. (2003a). *The Case for a 3 minute Climate Surface for South America*. Internal report No.3 to CRIA, Draft - May 2003.

Chapman, A.D. (2003b). *Environmental Modelling in CRIA – Issues and Ideas*. Internal report No. 4 to CRIA, Draft - June 2003.

Chapman, A.D. (2003c). *Notes on Environmental Data Quality – a. A Discussion Paper.* Internal report No.5 to CRIA, Draft in Preparation - July 2003.

Chapman, A.D. (2003d). *Notes on Environmental Data Quality – b. Data Cleaning Tools.* Internal report No.6 to CRIA, Draft in Preparation - June 2003.

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

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

Hijmans, R.J. (1999). Global Climate Surfaces at a 10’ Resolution. *Production Systems and Natural Resource Management* Working Paper No. 3. Lima, Peru: International Potato Centre.

Houlder, D. Hutchinson, M.J., Nix, H.A. and McMahaon, J. (2000). ANUCLIM 5.1 Users Guide. Canberra: Cres, ANU. http://cres.anu.edu.au/outputs/anuclim.html

Jones P.G. and Gladkov, A. (2001). *Floramap Version 1.01*. Cali, Colombia: CIAT. http://www.floramap-ciat.org/ing/floramap101.htm.

Margules, C.R. and Austin, M.P. (1994). Biological models for monitoring species decline: the construction and use of data bases. *Philos. Trans. R. Soc., London* **B344:** 69-74.

New, M., Lister, D., Hulme, M and Makin, I. (2002). A high-resolution data set of surface climate over global land areas. *Climate Research* **21**: 1-25.

OECD (1999). *Final Report of the Megascience Forum Working Group on Biological Informatics*. Paris: OECD

Payne, K. and Stockwell, D.R.B. (n.dat.). GARP Modelling System User’s Guide and Technical Reference. San Diego: SDSC. http://biodi.sdsc.edu/Doc/GARP/Manual/manual.html

Peterson, A.T., Navarro-Siguenza, A.G. and Benitez-Diaz, H. (1998). The need for continued scientific collecting: A geographic analysis of Mexican bird specimens. *Ibis* **140:** 288-294. [not seen]

Peterson, A.T., Stockwell, D.R.B. and Kluza, D.A. (2002). Distributional Prediction Based on Ecological Niche Modelling of Primary Occurrence Data pp. 617-623 **in** Scott, M.J. *et al.* eds. *Predicting Species Occurrences. Issues of Accuracy and Scale.* Washington: Island Press.

Scachetti-Pereira, R. (2002). *Desktop Garp*. Lawrence, Kansas: University of Kansas Center for Research. http://www.lifemapper.org/desktopgarp/ [Accessed 30 Jan 2004].

Stockwell, D.R.B. and Peters, D. (1999). The GARP Modeling System: problems and solutions to automated spatial prediction. *International Journal of Geographical Information Science* **13(2):** 143-158.

University of Kansas (2003). LifeMapper, Lawrence, Kansas: University of Kansas - Informatics Biodiversity research Center http://www.lifemapper.org.

UNEP (2003). Vital Climate Graphics. United Nationions Environment Programme - GRID Agenda. http://www.grida.no/climate/vital/.

Vieglais, D.A., Stockwell, D.R.B., Cundari, C.M., Beach, J.H., Peterson, A.T. and Krishtalka, L. (1998). The species analyst: Tools enabling a comprehensive distributed biodiversity network. *Biodiversity, Biotechnology & Biobusiness, 2nd Asia Pacific Conference on Biotechnology, 23-27 November, Perth, Western Australia.*

Vieglas, D.A. (1999). The Species Analyst. Integrating Disparate Biodiversity Resources using Information Retrieval Standards (Z39.50). 1999 TDWG Meeting, Harvard University, Cambridge, USA 29-31 October 1999. http://www.tdwg.org/davetdwg.htm.

Wein, J. (2002). Predicting Species Occurrences: Progress, Problems, and Prospects pp. 739-749 **in** Scott, M.J. et al. eds. *Predicting Species Occurrences. Issues of Accuracy and Scale*. Washington: Island Press.

Williams, P.H., Marguiles, C.R. and Hilbert, D.W. (2002). Data requirements and data sources for biodiversity priority area selection. *J. Biosc.* **27(4):** 327-338.