Modelling the potential distribution of weeds

Proceedings of a seminar presented by the Weed Society of Victoria Inc. and held at the Department of Primary Industries, Attwood, Victoria on Thursday 16 April 2009.

Modelling 101 – some modelling basics for non-modellers

Kim Lowell, Department of Primary Industries, 32 Lincoln Square North, Parkville, Victoria 3052, Australia and the Cooperative Research Centre for Spatial Information, University of Melbourne, 723 Swanston Street, Carlton, Victoria 3053, Australia.

Introduction

One of the domains that has been radically altered by the ubiquity of digital technology is algorithmic and quantitative modelling. Over the last five to 10 years, a variety of software modelling platforms have been developed that allow users to create models using diagrammatic interfaces. This has largely dispensed with the need for individuals interested in creating models to be experts in computer science and programming. Instead, domain-specific experts can concentrate on structuring their expansive knowledge in a way that accurately describes known processes and associated interactions, and that can be readily used in a digital environment.

In addition to having computer experts design less technical software modelling support platforms, the widespread use of digital technology has increased the confidence of even technophobic users to employ computers in increasingly sophisticated ways. Hence individuals from non-quantitative disciplines such as natural, biological, and life sciences have an enhanced belief in their own ability to learn and appropriately use sophisticated modelling software.

Hence conceptually computer experts and biological science practitioners have moved closer to a digital centre. The latter are nonetheless generally not completely at ease with quantitative modelling. The goal of this paper is to highlight and explain some basics about modelling in such a way that non-modellers' understanding of modelling is enhanced. The core concepts are presented as questions focused on three areas: model creation, data issues, and model structure.

Question 1 (Model Creation): What is the model for?

Models are essentially a means of simplifying reality. And in a modelling context, reality is simplified for two different reasons.

A. Process or system understanding

First, one may be seeking a means by which one can better understand a process. Someone might first collect data about the presence or concentration of a particular insect pest over a large area. Such data can be converted into useful information a number of ways. For example, one might display the data spatially to understand where pest presence is most prominent. Such a display would inevitably lead to observations or suppositions about the concentration of the pest relative to other factors such as land-use or distance from roads. Models are a way to summarize such interconnections in order to understand, for example, if the concentration of the insect pest is related to rainfall. One might add additional explanatory/related factors to understand if the relation between the concentration of the insect and rainfall is impacted by temporal seasonality.

Models developed to understand a process or a system are generally created using statistical techniques that fit equations to the data collected. Regression analysis, for example, produces an equation that represents the best line through a set of points with the value of a particular coefficient indicating the relationship between a predictor variable – e.g., rainfall – and the output variable of interest – e.g., concentration of insects. Techniques such as regression also indicate the statistical strength of the relationship between two variables.

Hence, this is a data-driven modelling approach that produces one or more equations that describe the strength and nature of relations inherent in a set of data.

B. Prediction or scenario evaluation

Models developed for process or system understanding are often of limited use for prediction. For example, they may be constrained to a particular climate range or insect pest. In contrast, models developed for prediction are generally more widely applicable at a 'cost' of being less accurate for a particular (data-rich) situation or geographical area. Having said that, predictive models can be improved for specific situations by calibrating a generic model structure for a particular area or pest – provided sufficient data are available to do so, of course.

Predictive models are often systems-based and are concerned primarily with developing generalized holistic models and therefore place emphasis on understanding the linkages among different components. In the context of developing a predictive model for a particular pest, a modeller might be concerned with the generalities of how the insect develops in relation to rainfall and spring temperatures, and also how winter temperature impacts the spring population. This is in contrast to examining purpose-collected data using data mining techniques that indicate only what the data set shows.

Such models are developed using structured modelling platforms based on diagrammatic interfaces that rapidly allow connections in a system to be designed and created. Predictive models also are dependent on the strength of the body of science available to describe the dynamics of the phenomenon being modelled. For example, a model that identifies only that a relationship does exist between an insect and rainfall is of limited use; the specific nature of the relationship must be well understood.

The result of a predictive modelling approach is a model structured to describe pest behaviour that can be calibrated and will permit a range of scenarios to be evaluated. Predictive models embody generalized knowledge about a phenomenon rather than being confined to a particular instance of a phenomenon. As such, they provide a means to query what will occur to pest numbers if spring temperature changes or winter survival decreases, for example.

C. Making a link between process understanding and predictive models

Clearly predictive models rely on fundamental research for their basic knowledge, and knowledge developed to understand processes has a wider impact if incorporated into a systems-based predictive model. However, data requirements subtly affect the suitability of process understanding models to be used in predictive models. Suppose that collected data suggest that a particular insect that overwinters on the ground requires a certain soil type as classified based on field observation. To create a regional model for the insect that includes 'capacity for overwintering', it would seem relatively simple to use information from soils maps to determine if the required soil type is present or not at a given location. However, the information associated with the soils map is based on different data collection techniques, may be based on a different taxonomy than was used on the ground, and may not assess soil characteristics in the same way as was done on the ground. In short, the data used to create the understanding of the process is not available as an input into the predictive model.

This point is fundamental to the importance of understanding whether a model is being created to understand a process or system, or if it is being developed for predictive purposes. Though it is difficult for scientists to accept, if prediction is the goal of a particular model, then the only variables/components that can be included in the model are those for which input data will be available. This means that a number of variables that might be included in a process/system understanding model must be left out of a predictive model even if those variables are understood to be highly related to the behaviour of the phenomenon/system of interest

Question 2 (Data Issues): What if my data do not support my modelling needs?

The final point of the previous section presents considerable difficulty for someone who has gained enhanced understanding of a process through the creation of a model, but finds that the model has little predictive use because it is dependent on data that are available for the situation or area of interest. In most cases, collecting the appropriate data for the entire situation/area of interest is not a viable solution. In the soil example used earlier, it is simply not economically feasible to visit every location in a region to create a soils map using the ground-based techniques that were used to collect the original data.

Two alternatives are possible. First, one can evaluate the suitability of available data to serve as a surrogate for the desired data. For example, one may have evaluated soil type on the ground by assessing the percent of sand in the top 10 cm of the soil profile. The soils maps may record the soil structure for the top 10 cm, or the percent of sand in the surface horizon regardless of depth. The relationship of the ground-based measurements can be compared against both of these by overlaying the locations of the field-based observations on the soils map, extracting the soils map information for each, and analyzing the data. Note that in doing this, it is unnecessary that the two sets of data give equivalent estimates – only that the relationship between the two is statistically strong. If so, the reality that soil map information consistently overestimates sand content by 10%, for example, is unimportant; for predictive purposes sand content from the soils map need merely be reduced by 10% to have useful data.

Second, one can seek surrogate information obtained from an understanding of relations among ancillary factors. For example, it may be known that the amount of sand is related to the underlying geology and the steepness of the terrain around a given location as determined by a digital elevation model (DEM). Using the field-based data for calibration, it may be discovered that these factors are related to weathering and erosional processes in some poorly understood yet statistically significant way. If so, one needs only have a map of underlying geology, a DEM, and the mathematical expression of the relationship, and one can produce the information required for predictive modelling.

Question 3 (Model Structure): Does my model need to be very complex?

The second approach described in the previous section adds a level of complexity to a model. Instead of having a simple input – percent sand in the top 10 cm of a soil profile – one has a model that produces that input. The model used for percent sand may in turn use other models to describe its inputs. And if one considers a multi-factor model and all of the potential inputs and interactions among the various components, models can rapidly become quite complex.

Such complexity may be highly undesirable. Generally speaking, as model complexity increases, general usability decreases. This latter is a consideration particularly when a model is being developed with the specific goal of being widely distributed – particularly to non-experts.

Of equal importance is model performance, and this impacts the necessary and appropriate level of complexity in two ways. First, as model complexity increases, model precision generally decreases. Because relationships in natural systems cannot be modelled with perfect precision, increasingly complex models connect one imprecise relationship with another, and another, and… This also causes the imprecision in individual model components to compound. Second, additional complexity may have little impact on model improvement. For example, one might never consider developing an insect pest model without temperature – something that is known to be affected by elevation as well as general weather patterns. However, the inclusion of elevation may only improve the model slightly. In model parlance, this means that the phenomenon being modelled has low sensitivity to elevation even though it has high sensitivity to temperature which is in turn (paradoxically) related to elevation. If a phenomenon is not sensitive to a particular factor – even if that factor on its own is known to be highly related to the behaviour of the phenomenon being modelled – there is little point in including that factor in the model.

Increasing model complexity may also impact its accuracy. The nature of ground-based studies is that they focus on a limited number of factors that can be studied through direct data collection. This means that the impact(s) on a phenomenon of interest of interactions among factors are often not understood. Creating complex models sometimes has the effect of describing the impact of factor interaction on a modelled phenomenon when no empirical knowledge about the true impact exists. If this occurs, the accuracy of model outputs will be unknown. The author is aware of a number of hydrological models that estimate streamflow, recharge, and other factors based in part on vegetative evapotranspiration. Vegetative evapotranspiration is in turn estimated by crop and forest growth models. The net effect is that the models estimate different amounts of vegetative water consumption depending on what crop is being grown or tree species is present in a forest even though these have not been explicitly studied.

Conclusions

Advances in computer technology and in the comfort with which experts in biological sciences use computers are facilitating an increasing capability for the creation of quantitative models developed for a variety of purposes. The first step in model development is determining if a given model is being used to better understand a particular process or system, or if it is to be used for prediction. The goal of the model will impact the data that can be used for model development and possibly the complexity of the resulting model. However, practitioners are well advised that increased model complexity will not necessarily lead to better or more accurate models.

Modelling resources

Books

Goodenough, K. and McKinion, J. (eds) (1992). Basics of insect modelling. American Society of Agricultural Engineers. ASAE Monograph Number 10.

Grace, J. (2006). Structural equation modelling and natural systems. (Cambridge Press).

Papjorgji, P. and Pardalos, P. (eds) (2009). Advances in modelling agricultural systems. (Springer).

Peart, R. and Shoup, W. (eds) (1998). Agricultural systems modelling and simulation. (CRC Press).

Vohnout, K. (2003). Mathematical modelling for systems analysis in agricultural research. (Elsevier).

Journals

Computers and electronics in agriculture. http://www.elsevier.com/wps/find/journaldescription.cws_home/503304/description#description. (Elsevier).

Ecological modelling. http://www.elsevier.com/wps/find/journaldescription.cws_home/503306/description. (Elsevier).

Environmental modelling and software. http://www.elsevier.com/wps/find/journaldescription.cws_home/422921/description#description. (Elsevier).

Journal of biological dynamics. http://www.tandf.co.uk/journals/titles/17513758.asp. (Taylor and Francis).

Conferences

IASTED International Conference on Modelling and Simulation. Annual. MS 2009: http://www.iasted.org/CONFERENCES/home-670.html.

International Conference on User Modelling, Adaptation, and Personalization. Biennial in odd-numbered years. UMAP 2009: http://umap09.fbk.eu/.

MODSIM – International Conference on Modelling and Simulation. Biennial in odd-numbered years. MODSIM 09: http://www.mssanz.org.au/modsim09/.

Weeds in a warmer world: predicting the impact of climate change on Australia's alien plant species using MaxEnt

Peter D. Wilson[A], **Paul O. Downey**[B], **Michelle Leishman**[A], **Rachael Gallagher**[A], **Lesley Hughes**[A] and **Jessica O'Donnell**[A]

[A] Department of Biological Sciences, Macquarie University, New South Wales 2109, Australia.

[B] Pest Management Unit, Parks and Wildlife, New South Wales Department of Environment and Climate Change, PO Box 1967, Hurstville, New South Wales 1481, Australia.

Summary

Australia is now host to many thousands of introduced plant species, and about 3000 of these have established self-sustaining populations in the wild. Of these, approximately 450 are classified as invasive weeds nationally or regionally, and are being targeted with control measures. Two questions arise in the context of climate change: what changes might occur in the distribution of the 450 species known to be highly invasive, and which species in the pool of 3000 might emerge to become future serious pests. We are applying an advanced modelling tool, MaxEnt, to provide a strategic overview of a large portion of the 450 highly invasive species. Preliminary results suggest differing responses of weed species in northern and southern Australia linked to predicted major shifts in rainfall pattern.

Introduction

Since the earliest European contact with the Australian continent in the late 16th Century, a steady stream of introduced species have been entering Australia's ecological communities with varying impacts on the native biota. Several thousand plant species have been introduced deliberately either as agricultural or horticultural stock, or as ornamental garden plants (Groves *et al.* 2005) or accidentally. Through complex pathways involving genetic change, developmental responses to new environments and chance dispersal events, a portion of these introduced plants have become established as self-sustaining populations in the wild. Current estimates suggest that at least 3000 plant species have become naturalized in Australia (Groves *et al.* 2003) with approximately 450 of these now classified as highly invasive pests. The balance of the 3000 naturalized species represents a pool from which it is likely that new invasive species might emerge, especially given that many of these are garden escapes and still available for sale or grown in gardens (Groves *et al.* 2005).

The mechanisms by which an introduced plant species makes the transition following introduction to establish self-sustaining wild populations are complex and not fully understood. They appear to involve a combination of genetic changes, phenotypic and developmental changes, and ecological interactions (both positive and negative) within the new environment. It is clear that the climate experienced by populations of a species is a powerful driver of ecological and micro-evolutionary processes. Climate directly influences a species' establishment, growth, reproduction, and survival. Climate also has indirect influences on invasive species via its impact on species within the ecological communities of which an invasive species is a part. Climate data is thus an obvious candidate to use as a surrogate for detailed ecological models of how a species responds to its environment, with the added advantage that modelled climate data is available in GIS coverage for current climate conditions. Future climate models also allow us to make predictions of distributions under certain constraints or caveats.

Given the paucity of detailed genetic, physiological or population data for the majority of the 3000 naturalized species, other forms of inference about likely changes in distribution and abundance are required. Species distribution models (SDMs) represent one tool that may assist our management of invasive plants. Not only can SDMs guide our understanding of current distributions and the response of species to the cumulative influences of past conditions, they offer the prospect of some degree of prediction under novel environmental conditions such as climate change.

The motivation for our project is twofold. First, we wish to investigate through experimental methods the way in which key groups of invasive plant species will respond to climate change, particularly to increased CO_2 concentration. This