International Environmental Modelling and Software Society (iEMSs)
2012 International Congress on Environmental Modelling and Software
Managing Resources of a Limited Planet, Sixth Biennial Meeting, Leipzig, Germany
R. Seppelt, A.A. Voinov, S. Lange, D. Bankamp (Eds.)
http://www.iemss.org/society/index.php/iemss-2012-proceedings

The Application of the Virtual Ecologist Approach to Evaluating the Effects of Uncertainty in Plot Based Monitoring Schemes due to Landscape Spatial and Temporal Heterogeneity

Alex Lechner[1], Daniel Clarke[1], Dion Weatherley[2], Andrew Fletcher[1], Peter Erskine[1], Alexis Comber[3]

[1]University of Queensland, Centre for Mined Land Rehabilitation, Sustainable Minerals Institute, a.lechner,d.clarke,andrew.fletcher,p.erskine(@uq.edu.au)
[2]University of Queensland, WH Bryan Mining & Geology Research Centre, Sustainable Minerals Institute, d.weatherley@uq.edu.au
[3]University of Leicester, Department of Geography, ajc36@leicester.ac.uk

Abstract: Monitoring programs that can detect changes in ecosystem condition are critical to assessing the success of rehabilitation and detecting negative anthropogenic impacts from activities such as mining. However, vegetation communities vary considerably and may exhibit short-term condition changes that mask long-term trends. This paper describes the development of a virtual ecologist (VE) landscape and observation simulation model for time-series data (VELOS_t). VELOS_t can be used to quantify the relationship between vegetation temporal and spatial variability, measurement uncertainty and sampling design to evaluate the robustness of a particular monitoring strategy. The model has four components: i) a landscape model that uses synthetic data to describe vegetation condition; ii) a natural variation model; iii) an environmental impact model and iv) a sampling model to describe plot-based monitoring schemes. The VE model allows users to estimate the expected performance of a range of sampling designs *a priori* and thus estimate detection sensitivity. Using simulated vegetation data, the model assesses whether statistical analyses can distinguish patterns of vegetation abundance from the effects of the observation of these patterns. Furthermore, the VE approach is a useful in testing the uncertainty sources such as imprecise measurement of vegetation cover are easily modelled using the VE approach in comparison to analytical approaches. This paper introduces the virtual ecologist model and provides a simple example of its use to assess the robustness of monitoring scheme design for long-term trend analysis.

***Keywords**:* virtual ecologist; monitoring; simulation modelling; synthetic landscapes; uncertainty

1 INTRODUCTION

1.1 Overview

A representative monitoring program that can detect changes in ecosystem condition is critical to assess the success of rehabilitation and detect negative impacts such as deforestation. However, vegetation communities vary both spatially and temporally and may exhibit superficial change in their condition that masks long-term trends. Measurement accuracy, sampling frequency and intensity will determine whether real changes in vegetation condition can be identified from natural variation. The early identification of trends in long-term condition can inform targeted management activities and has the potential to address negative environmental impacts.

Ensuring that the sampling design is capable of detecting biologically meaningful changes and correctly discern trends in data is fundamental to any meaningful monitoring scheme (Fairweather 1991, Legg and Nagy 2006, Field et al. 2007). The probability of successfully discerning trends in data with statistical methods depends on a number of factors including: i) statistical significance, ii) the effect size and iii) sample size. Power analysis is one method that is often advocated at the design or planning phase of any monitoring scheme to estimate the appropriate sample size required to detect an impact of a specific effect size (Steidl et al. 1997). Analytical power analysis methods can describe the relationship between the basic elements of sampling design such as sample size, however, these methods do not take into account spatial autocorrelation of the sampling points (Legendre 1993) and measurement uncertainty such as observation error (Gorrod and Keith 2009) and precision.

Multiple forms of uncertainty within monitoring schemes can be readily investigated using the virtual ecologist (VE) approach (Zurell et al. 2010) that simulates the observation of ecological systems, the ecological system itself and the statistical analyses. This paper describes a virtual ecologist landscape and observation simulation model for time-series data (VELOS_t) and provides a simple example of its use to assess the robustness of monitoring scheme design. We use the VELOS_t model to evaluate the performance of different field based sampling methods in conjunction with a linear trend analyses. It uses simulated vegetation data to describe patterns of vegetation abundance both temporally and spatially and assesses whether statistical analyses can distinguish these patterns from the effects of the observation of these patterns. The observation of landscapes are simulated using fixed plot based field monitoring approaches and the analyses of these plots using trend analysis (Gerrodette 1987). The VELOS_t model simulates spatially autocorrelated changes in the landscape such as a linear decrease in vegetation abundance within a certain part of the landscape. We conclude with a discussion of future research that can use the VELOS_t model described this study.

2 VELOS_t model

The VELOS_t model was developed for high performance computers using the Python's Numpy library with some functions written in C to allow massively parallel processing to overcome computational limitations of desktop PCs. The model, however, can be run on multiple CPUs on a desktop PC with each iteration running in parallel on a separate CPU. Parallel processing is key to processing large volumes of data at high speed to test numerous explanatory variables with high replication.

2.1 VE Model Design

There are four distinct steps in the VELOS_t model that can be summarised as follows (Fig 1). Step 1: Landscape generator and temporal simulation generates random spatially autocorrelated synthetic landscapes that change over time. Step 2: Field observation simulator simulates plot based field observation of these synthetic landscapes. Step 3: Statistical model simulator that analyses the observational data within a trend analysis – linear regression. Step 4: In the final step, the parameters for the landscape generator, temporal simulator and the observation model are stochastically generated to test differences between the trends derived in relation to multiple parameter combinations. Multiple iterations of the VELOS_t model can be ran adjusting the input parameters in steps 1 to 3.

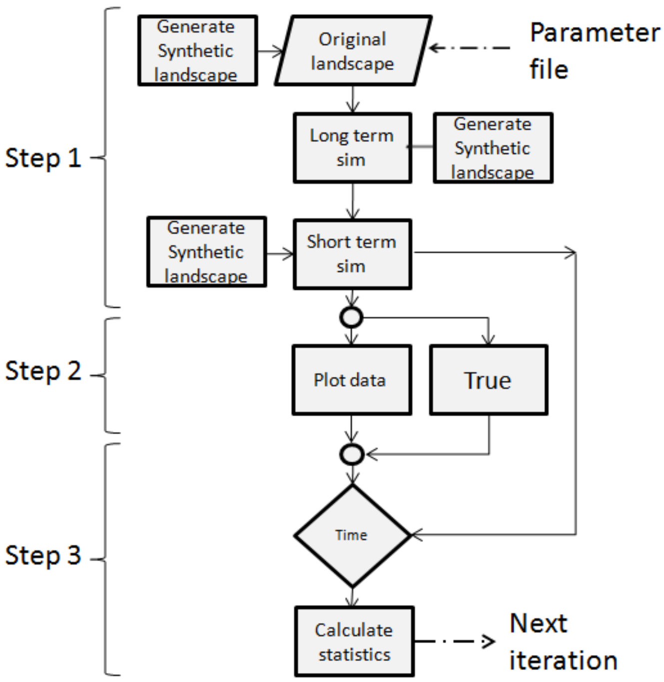


Fig 1. Flow chart of describe the VELOS_t model processing steps.

2.2 Step 1: Landscape generator and temporal change simulator

Vegetation cover was simulated using synthetic gray scale images generated using Saupe's (1988) fractional Brownian motion with midpoint displacement (midpointfM2D). The same algorithm is used in many common synthetic landscape generation programs such as QRULE (Gardner and Urban 2007) – see Fig 2. The midpointfM2D algorithm can randomly generate synthetic multi-fractal landscapes using a variety of fractal dimensions determined by the parameter H, ranging from 0 to 1. In map terms the fractal dimension equates to landscape patterns with different levels of spatial autocorrelation:

H = 0 negative spatial autocorrelation
H = 0.5 no spatial autocorrelation
H =1 positive spatial autocorrelated.

Maps were generated for multiple H values to systematically test for the effect of a range of spatial autocorrelations and fragmentation. Maps with lower H values appear more fragmented than those with higher H values.

Temporal and long-term variation was simulated with a range of spatial autocorrelations, by modifying the original synthetic landscapes with the spatially autocorrelated values in other synthetic landscapes. The landscape simulation model consisted of an initial original landscape modified at each time step with i) a single synthetic landscape to simulate long-term change that is spatially autocorrelated and ii) a new landscape at each time step to simulate natural variability. Change is simulated using simple raster addition thus $\text{Landscape}_t = \text{Landscape}_{t-1}$ + Longterm Landscape + Short-term landscape. Long-term change results in a directional change in value of vegetation cover each time step either positive or negative. Short term change results from random positive and negative changes to the landscape.

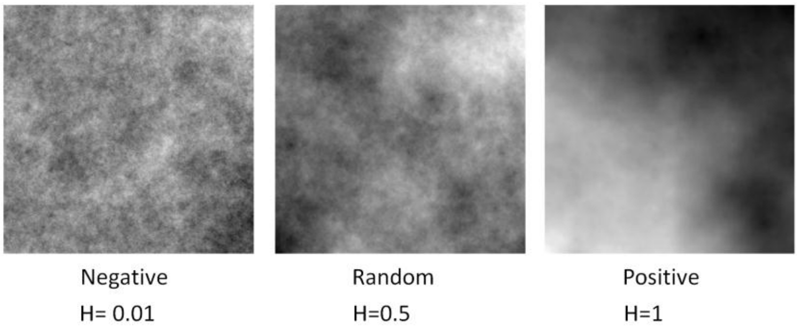


Fig 2. Synthetic landscapes with negative spatial autocorrelation (H= 0.1), random (H=0.5) and positive spatial autocorrelation (H = 1.0). Where black equals vegetation cover of 0 and white equal vegetation cover of 1.

As well as the linear change simulated in the examples, the VELOS_t model can simulate a range of long-term or short-term changes that can be produced by any function (e.g. logarithmic, sine) including random change. The amount of change per step from the long-term and short-term change can also be set or randomised with a range e.g. 10-30% random variability each time step.

In the example described in this paper the range of vegetation cover values over time and spatial variability were parameterised using real data from monitoring sites. Long-term change was simulated for 10 years with 4 seasons each year giving a model with 40 time steps. We simulated a linear long-term change of 40% with short-term random variability.

2.3 Step 2: Field observation simulator

The observation of vegetation cover using plot based field methods was simulated by sampling a subset of the synthetic landscape data using randomly placed non-overlapping circular plots (Fig 3). The model simulated a range of sample sizes, plot areas and sample temporal frequency (i.e. once every year, twice per year). The field observation simulator can also simulate uncertainty in terms of measurement error and uncertainty that results from field data measured categorically such as with the commonly used Braun-Blanquet vegetation cover abundance scale (Wikum and Shanholtzer 1978) (e.g. Category 7:100%- 75 %, Category 6: 75 - 50% etc.).

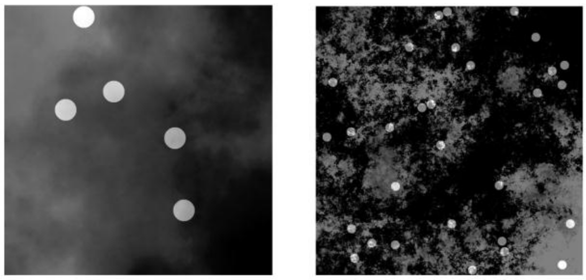


Fig 3. The Field observation simulator tested a range of sampling schemes. left) Five sample plots with a 3% cover, right) 20 sample plots with 3% cover.

2.4 Step 3: Statistical model simulator

The VELOS_t model can statistically analyse any form of statistical analyses conducted that can be done programmatically. In the example described in this paper a trend analysis was conducted on the sample data using simple linear regression with the R package. In ecology linear regression can be used with time series data to identify a linear change in vegetation cover over time. The accuracy of the sampling design was evaluated by comparing the difference between the actual linear long-term trend created in the synthetic data with the linear relationship statistically derived from the sampled data. Comparisons between the long-term trend line and the fitted line can be measured using root mean squared error (RMSE).

2.5 Step 4: Stochastic parameter generator

The above 3 steps describe a single iteration of the VE model which can be run an infinite number of times. For each iteration, the VE model stochastically generates parameter values for the landscape generator, temporal simulation and sampling design scenarios from a range of values set with the parameter file. Before the VELOS_t model runs, ranges for all parameters are populated (Table 1) based on a subset of possible values. Commonly this parameterisation would be based on empirical data or expert opinion. In many cases field based data does not provide the necessary information for all model parameters. This is the case for describing landscape spatial autocorrelation, where H values of the long-term and short simulation could not be estimated from plot based data. Table 1 describes a range of possible parameter values for the VELOS_t model and for the examples described in this paper. The sample size and total sample size area and short-term variability were the result of data analysis from a real monitoring program.

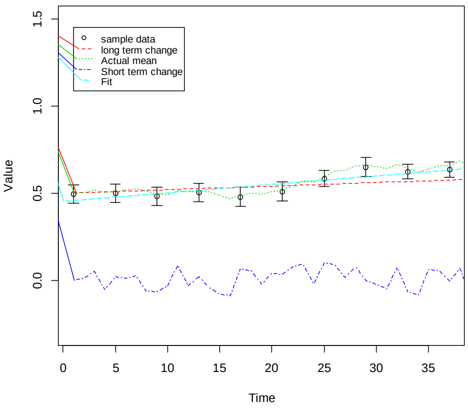
Table 1. Parameters used in VELOS_t model, possible parametisation values and values used in the example.

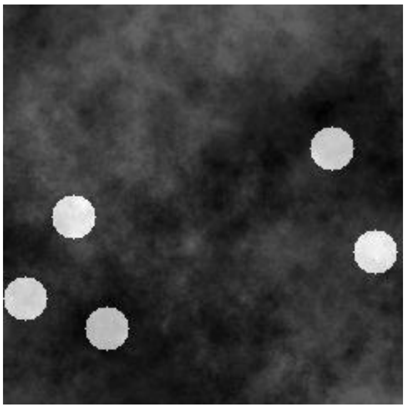
Parameter	Possible values	Example
Time		
Years	Unlimited	10 years
Time steps per year	Unlimited	4 time steps per year
Landscape generator and temporal simulation		
Original landscape – spatial autocorrelation	H values >0 to 1.0	0.5
Long-term modification – spatial autocorrelation	H values >0 to 1.0	0.5
Short-term modification – spatial autocorrelation	H values >0 to 1.0	0.5
Original landscape – Range of vegetation cover values	0 to 1.0	0 to 1
Long-term change – Range of vegetation cover values	0 to 1.0	0 to 0.5
Short-term change – Range of vegetation cover values	0 to 1.0	-0.1 to +1.0
Long-term change function	Unlimited i.e. linear, log etc.	Linear
Short-term change function	Unlimited i.e. sine, random	Random
Field observation simulator		
Number of samples	Unlimited though area dependent (no overlap)	4 – 6
Percentage of area sampled	>0	2.5 – 5%
Measurement error	0-100%	0%
Sampling frequency	Infinite	Twice a year
Stochastic parameter generator		
Number of iterations	Infinite	30

3 RESULTS

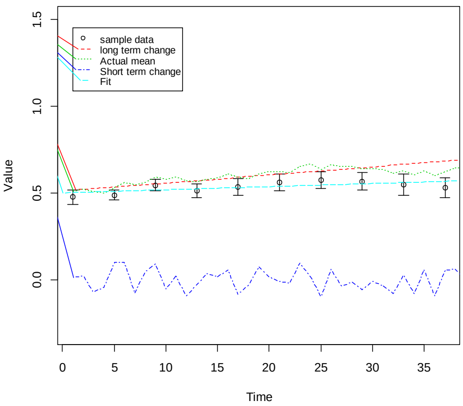
To demonstrate the potential of the VELOS_t model we simulated 30 replications and from those replications we chose two examples: where a monitoring scheme and statistical test failed and where they were successful. In the first example, the combination of sampling point number, size and location along with landscape spatial heterogeneity, and spatial heterogeneity in changes over time resulted in the fitted line closely representing the long-term change (Fig 4a). In this case the fitted line had a high r^2 value of 0.743 that was significant (P value = 0.001) and was very close to the long-term trend line with a RMSE of 0.034. In the second example the fitted value was still significant (p=0.03) with a lower but still relatively high r^2 of 0.46, but the RMSE was higher at 0.33. The trend of the fitted line was very different compared to the true long-term trend. The key to a successful monitoring program is its ability to accurately identify a change in the long-term trend. These simple examples show how significant relationships can be derived with a statistical test, while in one case the fitted curve resembles the long-term relationship and thus correctly identifies a long-term trend. The other case results in the slope of the fitted line approximating 0 indicating no trend which would suggest no change in the monitored vegetation abundance.

a)





b)



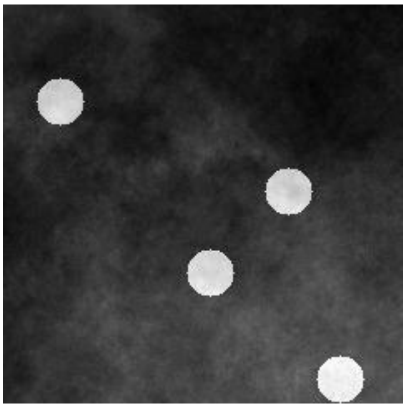


Fig 4. Example of two iteration of the model with the following parameters:, sampling intensity = 2 times a year, total time 10 years (i.e. 40 time steps x-axis), short-term variability = -/+ 10% and original landscape H=0.49. The y-axis describes vegetation cover values. a) Sample size 5 with 4.7% of total area. RMSE = 0.034 and r^2=0.743 (P = 0.001). b) Sample size 4 with 4.2% of total area. RMSE = 0.073 and r^2=0.464 (P =0.030).

There are a range of factors affecting monitoring design success or failure. In example 2, two factors are interacting to result in the inability of the statistical test to retrieve the true consistent change. Firstly random variability has masked the long-term trend (as seen by a drop in mean landscape vegetation cover from time 30 onwards) and secondly the plots are not found in areas representing the total condition of the landscape (Fig 4b). However, using only the significance value as a measure of the accuracy of the test one may conclude the statistical test was successful in retrieving the true ecological pattern. Only in a completely measured, and therefore virtual world can we know that the true ecological pattern is very different from the one fitted to the sample data.

The successful retrieval of trends in long-term change was predominantly the result of random differences in spatial pattern and random variability in the total differences per time step for short-term variability. Such large variability in the outcome resulted from very few of the parameters being changed from the range of possibilities (Table 1). Table 2 shows the average values for the fitted curve for all 30 iterations. As can be seen from this table the fit was usually significant and the differences between the true long-term change and the fitted line were moderate with an average 0.054 RMSE. Only 77% of the results showed the correct positive trend in the long term change.

Table 2 Data Summary (n=30). Positive slope % indicate the percentage of runs in which the slope was greater than 0. Sig p (0.05) % indicate the number of times the fitted curve had a p value less than 0.05.

Mean RMSE	Positive slope %	Mean r^2 value	Mean p value	Sig p (0.05) %
0.054	0.777	0.56	0.0754	0.73

4 DISCUSSION AND CONCLUSION

The VE (Zurell et al 2010) approach to simulating ecological data and observers and subjecting them to statistical analyses allows for some elements of the complexity of real world phenomena to be studied in a systematic way. The simulation model represents an idealized scenario where there are no ‘unknown’ sources of uncertainty present in the data or in the ecological model. It allows us to evaluate the performance of ecological models by using a known truth described by synthetic ecological data. If theoretical ecological relationships cannot be derived correctly using simulation models, there can be little confidence in recovering these from more complex ‘real’ data (Austin et al. 2006). When using real data ecologists can only make inference about causal processes based on empirical data. The virtual ecologist approach allows for the comparison of what contributes to the patterns that emerge from a statistical analysis: the data or uncertainties that result from the observation of the data. Thus, using this approach we can test whether patterns derived from ecological models are independent of the observation method.

This paper uses a small dataset to introduce the VELOS_t model and illustrate how it works. Future work will focus creating generalisations to guide ecological research. However, there are a few key problems that need to be addressed in order to make generalisations. Firstly, the potential output from the VELOS_t model is multi-dimensional with massive datasets that in some cases are correlated, thus requiring sophisticated data mining techniques to derive patterns from the data. Secondly, the patterns derived from VELOS_t are dependent on the parameterisation. For example, a highly spatially autocorrelated landscape may be a key driver of the probability of a statistical test not working, but the probability of certain H values existing in reality needs to be empirically determined. Thus realistic settings for the model need to be derived from real data.

The VE approach used in VELOS_t can be considered as a ‘sandbox’ for testing whether an ecological model is robust to uncertainty and thus identify what aspects of a model need improvement. A modelling framework, as developed in this study, has the potential to assess uncertainty in any model that may be called in Python. In the future there are plans to make the code open source and allow for the participation of the ecological community in its development.

ACKNOWLEDGMENTS

The project was funded by the SMI Excellence award and University of Queensland Early career research grant. I would like to acknowledge the help of David Unwin and Enzhen Luo.

REFERENCES

Austin, M. P., L. Belbin, J. A. Meyers, M. D. Doherty, and M. Luoto. 2006. Evaluation of statistical models used for predicting plant species distributions: Role of artificial data and theory. Ecological Modelling **199**:197-216.

Fairweather, P. G. 1991. Statistical power and design requirements for environmental monitoring. Australian Journal of Marine and Freshwater Research **42**:555-567.

Field, S. A., P. J. O'Connor, A. J. Tyre, and H. P. Possingham. 2007. Making monitoring meaningful. Austral Ecology **32**:485-491.

Gardner, R. H. and D. L. Urban. 2007. Neutral models for testing landscape hypotheses. Landscape Ecology **22**:15-29.

Gerrodette, T. 1987. A Power Analysis for Detecting Trends. Ecology **68**:1364-1372.

Gorrod, E. J. and D. A. Keith. 2009. Observer variation in field assessments of vegetation condition: Implications for biodiversity conservation. Ecological Management & Restoration **10**:31-40.

Legendre, P. 1993. Spatial Autocorrelation: Trouble or New Paradigm? Ecology **74**:1659-1673.

Legg, C. J. and L. Nagy. 2006. Why most conservation monitoring is, but need not be, a waste of time. Journal of Environmental Management **78**:194-199.

Saupe, D. 1988. Algorithms for random fractals. Pages 71-113 *in* H. O. Peitgen and D. Saupe, editors. The science of fractal images. Springer-Verlag, New York.

Steidl, R. J., J. P. Hayes, and E. Schauber. 1997. Statistical Power Analysis in Wildlife Research. The Journal of Wildlife Management **61**:270-279.

Wikum, D. A. and G. F. Shanholtzer. 1978. Application of the Braun-Blanquet cover-abundance scale for vegetation analysis in land development studies. Environmental Management **2**:323-329.

Zurell, D., U. Berger, J. S. Cabral, F. Jeltsch, C. N. Meynard, T. Munkemuller, N. Nehrbass, J. Pagel, B. Reineking, B. Schroder, and V. Grimm. 2010. The virtual ecologist approach: simulating data and observers. Oikos **119**:622-635.