

Nonparametric upscaling of stochastic simulation models using transition matrices

Pablo A. Cipriotti^{1*}, Thorsten Wiegand², Sandro Pütz^{2,3}, Norberto J. Bartoloni¹ and José M. Paruelo^{1,4}

¹Depto. de Métodos Cuantitativos y Sistemas de Información, IFEVA – Facultad de Agronomía, Universidad de Buenos Aires/ CONICET, Av. San Martín 4453, C1417DSE Ciudad de Buenos Aires, Argentina; ²Helmholtz Centre for Environmental Research - UFZ, Permoserstr. 15, 04318 Leipzig, Germany; ³Department of Bioenergy, Helmholtz Centre for Environmental Research - UFZ, PO Box 500 136, 04301 Leipzig, Germany; and ⁴Laboratorio de Análisis Regional y Teledetección, IFEVA, Facultad de Agronomía, Universidad de Buenos Aires/CONICET, Ciudad de Buenos Aires, Argentina

Summary

1. The problem of scaling up from tractable, small-scale observations and experiments to prediction of large-scale patterns is at the core of ecological theory and application, and one of the central problems in ecology.
2. We present and test a general nonparametric framework to upscale spatially explicit and stochastic simulation models. The idea is to design a state space, defined by the important state variables of the small-scale model, and to divide it into a finite number of discrete states. Transition probabilities are then tallied by monitoring extensive simulation runs of the small-scale model, covering the entire range of initial conditions, states and external drivers that may occur for the desired application. We exemplify our approach by upscaling an individual-based model that simulates the spatiotemporal dynamics of *Festuca pallescens* steppes under sheep grazing in Western Patagonia, Argentina, with a spatial resolution of 0.3 m × 0.3 m and a 0.15-ha extent. The upscaled model simulates a 2500-ha paddock with 0.15-ha resolution and is enriched with additional rules that describe heterogeneity in the local stocking rate at the paddock scale.
3. We obtained 24 transition matrices that governed the upscaled model for different combinations of stocking rates and annual precipitation. The upscaled model produced excellent predictions for the long-term dynamics, but as expected, it did not fully capture the interannual dynamics of the original model. Rules for heterogeneity in the local stocking rate allowed for emergence of realistic vegetation patterns as commonly observed for water points in arid rangelands.
4. Our general nonparametric upscaling approach can be applied to a wide range of stochastic simulation models in which the dynamics can be approximated by a set of states, transitions and external drivers. Because estimation of the transition probabilities can be done parallel, our approach can be applied to a wide range of models of intermediate complexity. Our approach closes a gap in our ability to scale up from small scales, where the biological knowledge is available, to larger scales that are relevant for management.

Key-words: agent-based models, complex systems, graph theory, Markov chains, meta-models, rangelands, spatially explicit models, state-transitions models, succession

Introduction

Many of the phenomena of interest in ecology and conservation appear at larger spatial scales, but field data are typically collected over relatively small areas during observational or experimental studies of short duration (May 1999; Urban 2005). However, to understand large-scale phenomena, broad-scale patterns and processes must be related to those at small scales with which ecologists are most familiar and have data on hand (Wu 1999; Moorcroft, Hurtt & Pacala 2001). As a consequence, the problem of how to transfer information from smaller to larger scales, a process generally known as upscaling, is at the core of ecological theory and application, and is one of

the central problems in ecology (Levin 1992; Wiens *et al.* 1993; Urban 2005; Wu *et al.* 2006; Denny & Benedetti-Cecchi 2012).

The scale of individuals is of fundamental importance in ecology (Wiens *et al.* 1993) because the behaviour and interactions of individuals lead to the emergence of larger scale patterns which are usually of interest in ecology (Grimm *et al.* 2005). Additionally, information on the short-term behaviour of individuals and their interactions is often easy to observe (Pacala *et al.* 1996; Wiegand & Milton 1996). Individual-based models are especially suitable to describe the dynamics at the population or patch scale (Grimm & Railsback 2005, 2012; Grimm *et al.* 2005). Population dynamics emerges in individual-based models as a consequence of the behaviour and interactions of individuals with immediate ecological conditions (Grimm & Railsback 2005). However, what is challenging is

*Correspondence author. E-mail: cipriotti@agro.uba.ar

how to scale across several organizational levels from the individual to the local population scale, and to the landscape scale (Urban 2005; Boulain, Simoni & Gignoux 2007). Several modelling approaches have been developed to accomplish upscaling over more than one level of organization. For example, a recent review by Denny & Benedetti-Cecchi (2012) deals with mechanistic response functions to describe how phenomena interact across scales, and Wu *et al.* (2006) reviewed dynamic model-based upscaling as well as similarity-based upscaling methods that rely on power-law scaling, fractals, etc. (e.g. Kéfi *et al.* 2007).

Upscaling of dynamic models can be done in several ways. First, without changing the original model, one may (i) assume that it is representative. In this case, one may extrapolate the original model directly to the larger scale by running it with the mean values of parameters and inputs averaged across the entire landscape (i.e. 'lumping') or using 'effective' parameters instead of averages (Wu *et al.* 2006). Secondly, without changing the original model, one may (ii) scale up with 'brute force' and implement a copy of the model in each patch (i.e. 'direct extrapolation'). In this approach, the extent of the original model (i.e. a patch) is usually a cell of a grid that is superimposed to the landscape. This allows consideration of geospatially driven parameters and inputs (Rastetter *et al.* 1992; Peters *et al.* 2004; Van de Koppel, Gupta & Vuik 2011), possibly using remote sensed data to define the landscape. This approach has been extensively used in forest gap models (Shugart 1998), but leads only to an adequate upscaling if interactions among the patches can be neglected (Peters *et al.* 2004; Melbourne & Chesson 2006). However, interactions among patches (e.g. due to seed dispersal among patches) need to be considered in many relevant cases (Peters *et al.* 2004). Thus, a hierarchical perspective (Wu & Loucks 1995; Meyer, Wiegand & Ward 2009) can be adopted that (iii) basically uses the original model, but explicitly embeds it into a larger scale model that contains additional rules governing the interactions among patches. While this approach accomplishes an upscaling across three spatial scales (individual to patch, and patch to landscape scale), running in each patch a full copy of the detailed model may lead to excessive computational demands, rendering this approach similarly to (ii) unsuitable for many practical applications.

To escape the twin dilemma of computational intensiveness and larger scale interactions, we need a way to predict the landscape scale dynamics based on individual-level information, possibly supplemented by rules on cross-patch interactions, but without simulating the computationally demanding interactions among individuals. One approach for doing this is (iv) use of analytical approximations based on so-called macroscopic equations (Moorcroft, Hurtt & Pacala 2001; Strigul *et al.* 2008) that can directly predict variables of landscape scale dynamics based on the parameters of the detailed individual-based model. An alternative approach is to (v) fit a larger scale model to the output of a detailed small-scale model (e.g. Acevedo, Urban & Abla 1995; Rastetter *et al.* 2003; Tietjen & Huth 2006). Finally, Seidl *et al.* (2012) (vi) took advantage of the repetition of similar competition situations of individual

trees and replaced its detailed calculation with use of a previously generated library. Thus, once the library is generated by running simulations of the detailed model at smaller spatial scales, the landscape level model achieves reasonable computational speed without losing the information on the detailed behaviour of individuals.

In this study, we present a general framework for upscaling stochastic, spatially explicit and dynamic simulation models of intermediate complexity from the patch scale to the landscape scale. Our approach is inspired by Seidl *et al.* (2012) and Acevedo, Urban & Abla (1995) and takes advantage of the repetitiveness of the entire dynamics at the patch scale. The idea is to design a state space, defined by the important state variables of the original model, and to divide it into a finite number of discrete states. Libraries of transition probabilities are then tallied by monitoring simulation runs of the detailed model, covering the entire range of initial conditions, parameterizations, states and external drivers relevant for the desired application. Each combination of drivers requires one transition matrix. Finally, larger scale rules can be added to describe interactions among patches. We therefore do not fit the transitions of a predefined Markov chain model as done in earlier approaches (e.g. Acevedo, Urban & Abla 1995), but adopt a more flexible non-parametric approach in which the model itself decides which of the generalized states are the most important ones.

The overall aim of our paper was to present our nonparametric upscaling framework. We exemplify our framework with the individual-based simulation model COIRON that describes the spatial dynamics of the Patagonian *Festuca pallescens* grass steppes in Argentina (Paruelo *et al.* 2008). To this end, we test the upscaled model in its ability to recover the dynamics of COIRON, and we present an application of the upscaled model where consideration of interactions among patches is required to describe the effect of spatial gradients in the stocking rates due to water points.

Materials and methods

SMALL-SCALE MODEL COIRON

The aim of COIRON was to understand mechanisms of degradation in the Patagonian *F. pallescens* grass steppes as driven by stochastic rainfall and sheep grazing (Paruelo *et al.* 2008). This semi-arid ecosystem is dominated by the tussock grass species *F. pallescens* which occurs in areas receiving 300–600 mm of mean annual precipitation concentrated in winter. Additionally, the local stocking rates vary strongly over the paddock scale (e.g. in dependence on the distance to water points), which creates environmental gradients that feedback with local vegetation conditions (Pickup & Chewings 1994).

COIRON describes the dynamics of homogeneous *F. pallescens* stands (patches) with a 38.4 m × 38.4 m extent by considering small-scale processes at the tussock level (with resolution of 0.3 m × 0.3 m; the typical size of one tussock) and external controls such as selective grazing and stochastic precipitation (Paruelo *et al.* 2008). *F. pallescens* shows a negative feedback between the amount of senescent material present in a tussock and sheep grazing behaviour: the lower the amount of senescent material the higher the probability that green biomass is defoliated. This behaviour cannot be adequately described without

considering grazing behaviour at the individual tussock level. COIRON includes for this a detailed grazing model (see Appendix S1, Supporting information). However, the management relevant scale is the paddock or ranch scale which extends to 1000–10 000 ha. Thus, COIRON is a typical example where upscaling is needed.

COIRON is a stochastic grid-based and individual-based simulation model with a time step of 1 year and a spatial resolution (cell) representing a fully grown individual tussock ($0.3 \text{ m} \times 0.3 \text{ m}$). This is the spatial scale at which most of the plant processes and livestock forage decisions take place. The extent comprises 128×128 cells, covering an area of approximately 1500 m^2 . A cell can be occupied by a 'live tussock', a 'dead tussock', be 'empty' or part of a 'large bare patch' (basically a cluster of more than 10 empty cells). The latter are important because large bare patches suffer additional soil water loss due to higher evaporation, which have negative effects on seedling survival. A cell occupied by a live tussock is characterized by the state variables green biomass, dead biomass (which accumulates during previous growing seasons) and potential productivity (which summarizes the state of the tussock based on past precipitation and defoliation). The main input variables of COIRON are the annual rainfall, the local stocking rate and the initial vegetation condition, while the main output variables are the 'vegetation cover' and 'above-ground net primary production' (ANPP) of *F. pallescens*, and the 'cover of large bare patches'.

Changes in the state variables are driven by grazing (defoliation) and water availability resulting from local redistribution of annual precipitation due to root absorption. The dynamics of *F. pallescens* is governed by rules on water dynamics (rule 1), biomass production (rule 2), colonization (rule 3), grazing-induced changes in potential

productivity (rule 4), senescence and littering (rule 5), mortality (rule 6), local grazing pressure (rule 7) and defoliation (rule 8). Several of these rules (i.e. rules 3, 6, 8) are stochastic. The model summarizes abundant field data in this ecosystem collected over more than 40 years (Paruelo *et al.* 2008). The COIRON model is described, parameterized and tested in detail in Paruelo *et al.* (2008); a detailed model description is presented in Appendix S1 (Supporting information).

GENERAL UPSCALING FRAMEWORK

The main idea is to divide the state space, defined by the important state variables of the original model, into discrete states. Transition probabilities are then tallied by monitoring simulation runs of the original model that cover the entire range of initial conditions, parameterization, states and external drivers relevant for the desired application at the broader spatial scale (Fig. 1). For each combination of external drivers, we obtain one transition probability matrix (TPM) (Fig. 1c). The TPMs can be studied by Markov chain or graph theory analysis to gain insights about the system structure and dynamics (Balzer 2000; Strogatz 2001; Daehyun & Phillips 2013; see below 'Model analysis'). Finally, the landscape of the upscaled model is composed of a grid of cells where each cell hosts a copy of the transition probability matrices that replace the detailed small-scale model (Fig. 1e). Thus, the upscaled model can have a substantially larger extent than the original model, but it loses spatial resolution. However, all small-scale rules of the original model (e.g. the negative feedback in COIRON described above; Fig. 1a) are incorporated in the transition matrices. Finally, the upscaled model may contain rules describing large-scale (inter patch)

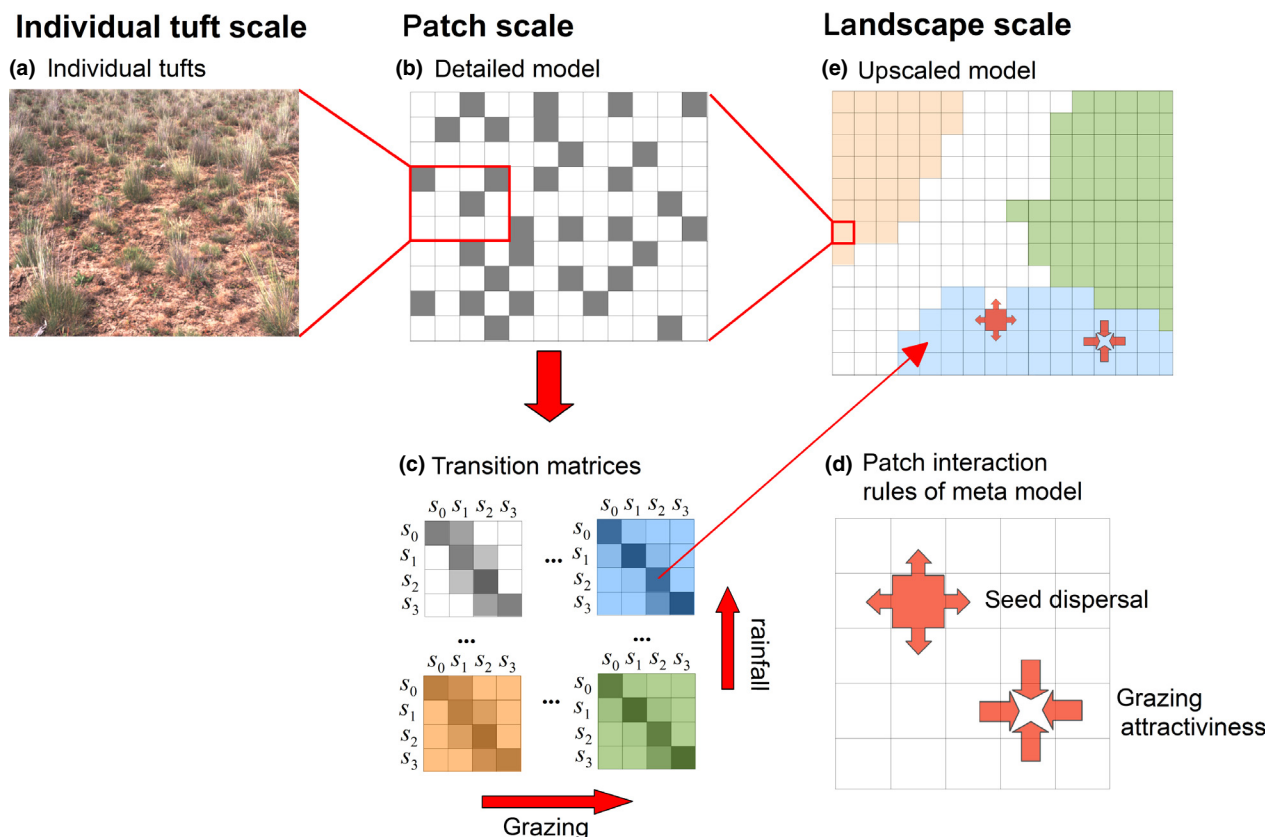


Fig. 1. General scheme of the upscaling approach indicating scale transitions and additional rules of the upscaled meta-model. The colours of the transition matrices and cells in the upscaled model indicate different environmental conditions.

Table 1. Criteria to the definition of the 192 states (0–191) from the small-scale model (COIRON) based on different levels from three main output variables: *F. pallescens* cover (8 levels), above-ground net primary production (6 levels) and cover of large bare gaps (4 levels)

Cover (%)	ANPP (kg DM m ⁻² yr ⁻¹)	Cover of large bare gaps (%)
0–10	0–175	0–25
11–20	176–350	26–50
21–30	351–750	51–75
31–40	751–1100	75–100
41–50	1101–2200	
51–60	>2200	
61–80		
80–100		

interactions such as gradients in local grazing pressure due to distance to water points (Fig. 1d).

Our upscaling framework therefore consists of four steps and is supplemented by two additional steps of model analysis: (i) definition of the states of the original model and the external drivers, (ii) definition of simulation scenarios that cover the full range of situations expected to occur at the large scale (e.g. initial conditions, climate time series, local stocking rates) and account for parameter uncertainty, (iii) running of the original model for all scenarios determined in (ii) to sample the libraries of transition probabilities, (iv) embed the transition probabilities into the large-scale model that describe larger scale processes and environmental drivers, and the possibly spatial interactions among patches.

DEFINITION OF STATES AND EXTERNAL DRIVERS FOR THE SMALL-SCALE MODEL

Our framework does not rely on *a priori* defined states; instead, states result from systematic combination of the state variables of the original model. Thus, we may possibly obtain a large number of states but are not constrained by *a priori* model structures. We define for each state variable a reasonable number of levels which are then combined with that of the other state variables to yield the generalized states of the original model.

The important state variables of COIRON are vegetation cover, ANPP of *F. pallescens*, and cover of large bare gaps. Based on information on the range of conditions found in Patagonian *F. pallescens* steppes (Paruelo *et al.* 1993; Bertiller & Bisigato 1998), we defined 192 possible vegetation states (0–191) that resulted from eight levels of *F. pallescens* cover, six levels of ANPP and four levels of cover of large bare gaps (Table 1). For easier representation of the results, we ordered the states with increasing forage quality from state 0 (<10% of vegetation cover, <175 g m⁻² yr⁻¹ of ANPP and 76–100% of cover with large bare gaps) to the best vegetation condition at state 191 (>80% of vegetation cover, >2200 g m⁻² yr⁻¹ of ANPP and <25% of cover of large bare gaps).

The important external drivers of the year-to-year dynamics of COIRON are the annual precipitation and the local stocking rate. We selected 24 environmental states that resulted from three levels of annual precipitation (dry years: < 315 mm yr⁻¹; typical years: 315–425 mm yr⁻¹; and wet years: > 425 mm yr⁻¹) and eight local stocking rates ranging from zero to 4 sheep per ha (i.e. 0, 0.25, 0.5, 0.75, 1, 1.25, 2 and 4 sheep per ha). Thus, our library of TPMs consists of 24 matrices that govern all possible year-to-year transitions among the 192 possible vegetation states.

DEFINITION OF SIMULATION SCENARIOS

To obtain complete TPMs the original model must be executed under a wide range of typical situations, defined by initial conditions, climatic scenarios and local stocking rates. Additionally, we can account for parameter uncertainty by repeating the procedure for different model parameterizations. We repeated the COIRON simulations for 30 different initial conditions, nine precipitation time series, eight stocking rates and eight alternative model parameterizations, yielding a total of 17 280. Details can be found in Appendix S2, Supporting information (*Definition of simulation scenarios*). Each simulation lasted 200 years. Thus, we monitored a total of $17\,280 \times 200 = 3\,456\,000$ individual transitions.

SAMPLING OF THE TRANSITION PROBABILITIES

For each simulation year t , we determined the initial state s_0 (at year $t-1$), the new state s_n (at year t), the annual precipitation class i and the stocking rate j (which was fixed for a given simulation). This information was used to tally transition probability matrices $\mathbf{P}_{ij}$ that counted the number of transitions from the old state s_0 to the new state s_n , conditionally on an annual precipitation class i and stocking rate j . The final matrices with transition probabilities $\mathbf{P}_{ij}(s_0, s_n)$ were calculated as

$$\mathbf{P}_{ij}(s_0, s_n) = \frac{n_{ij}(s_0, s_n)}{\sum_{s=0}^{191} n_{ij}(s_0, s)} \quad (1)$$

where the numerator n_{ij} is the number of transitions from the old state s_0 to the new state s_n , and the denominator is the sum across all transitions from the old state s_0 to whatever state s . However, we applied equation 1 only if the sum in the denominator was larger than zero, and defined $\mathbf{P}_{ij}(s_0, s_n) = 0$ otherwise.

LARGE-SCALE MODEL

Space is represented in the upscaled model as a grid of 135×210 cells (4180 ha) where each cell represents a $38.4 \text{ m} \times 38.4 \text{ m}$ patch with the extent of the original model. Thus, we change both aspects of scale, extent and grain. The upscaled model has a substantially larger extent than the original model (4180 ha vs. 0.15 ha), but a coarser grain (1500 m^2 vs. 0.09 m^2) (Fig. 1). After defining an initial state for each cell, the transition matrices are applied to update the state of each cell depending on its stocking rate and climate condition. We can also include spatially heterogeneous environmental conditions by applying different transition matrices for cells with different environmental conditions.

However, new processes may appear at the larger scales that are not included in the detailed smaller scale model. Here, we consider spatial heterogeneity in the local stocking rate as example. In semi-arid rangelands such as the *F. pallescens* steppe, the local stocking rate emerges from a number of factors, including (i) the global stocking rate in the paddock, (ii) the state of the focal patch, (iii) the state of the immediate neighbourhood of the focal patch and (iv) the distance of the focal patch to the nearest water point or other sites able to concentrate livestock in the paddock (Pickup & Chewings 1994). To account for these large-scale effects, we estimated for each patch in the paddock an index of ‘relative grazing attractiveness’ that allows for estimation of the local stocking rate in each patch (see Appendix S2, Supporting information ‘Relative grazing attractiveness’ for details).

Model analysis

Graph theory analysis of transition matrices. We can analyse the transition matrices with tools of graph theory as the vegetation states can be interpreted as vertices and the transitions as edges. This yields directed graphs because each edge has a direction and a value (i.e. the transition probability). To assess whether and under what conditions multiple pathways of degradation or recovery exist, we determined cohesive blocks of states (i.e. groups of vertices which are internally more connected than vertices belonging to other blocks; White & Harary 2001) and articulation vertices (which link different otherwise unconnected parts of the graph). We used the *igraph* package (Csardi & Nepusz 2006) in the R environment (R Core Team 2013) for graphs analysis.

Verification of the up-scaled model. To assess how well the upscaled model captured the detailed dynamics of COIRON, we compared the predictions of the upscaled model for different environmental situations (i.e. three rainfall scenarios and three stocking rates) with the outputs from COIRON. To this end, we simulated with the upscaled model a rectangular area composed of 100×100 cells (1474 ha). We started from a homogeneous initial condition where all cells were in good condition (i.e. state 187) and simulated three different stocking rates and three different climate series. See details in Appendix S2, Supporting information (*Verification of the up-scaled model*).

Application of the upscaled model to a water point. We implemented a water point in the centre of the paddock and simulated spatially heterogeneous stocking rates (see Appendix S2, Supporting information 'Relative grazing attractiveness') to assess its consequences for the *F. pallescens* dynamics at the landscape scale. As initial condition, we assigned to all 135×210 cells of the paddock a good vegetation condition (ID = 187) and simulated the upscaled model for 100 years. For comparison, we simulated a no grazing treatment and a treatment with high stocking rate but no water point. During the simulations, we recorded the vegetation state of each cell at steps of 25 years. To determine the average vegetation states at increasing distances from the water point, we applied a moving window on the resulting map of vegetation states.

Results

ANALYSIS OF THE TPMS OF THE UP-SCALED MODEL

Only between 28 and 45 of the 192 possible states were realized during the 17 280 COIRON simulations and only between 0.38% and 0.5% of the 192^2 theoretically possible transitions were realized (Fig. 2). The 147 states that were not realized in any of the 24 TPMS were unlikely biologically. However, in other cases some states did occur only under certain precipitation scenarios and/or stocking rates (Fig. 2a–c). Generally, matrices corresponding to high precipitation and intermediate stocking rates showed the highest number of states.

Most of the observed transitions occurred near to the main diagonal of the TPMS (Fig. 2a–c) and the highest transition probabilities occurred on the main diagonal (i.e. stasis; Fig. 2a–c). This was expected because the *F. pallescens* steppes do not show dramatic shifts in vegetation condition during

1 year, even under the dry climate and overgrazing conditions here studied. Matrices for dry years and overgrazing (i.e. >1 sheep per ha) yielded a lower number of transitions ($\approx 0.37\%$) compared with wet years and moderate or zero stocking rates (0.41–0.50%). Graph representations of TPMS show that the networks became simpler with increasing stocking rate (i.e. fewer states and transitions; Fig. 2d–f). We also found that degradation or recovery trajectories under a given environmental condition pass mostly through articulation vertices (Fig. 2d–f). Thus, multiple pathways under the same environmental conditions were unlikely.

TESTING THE UPSCALED MODEL

Generally, the upscaled model captured the long-term trends of the detailed COIRON model well (Fig. 3). The *F. pallescens* cover predicted by the upscaled model was almost in all years inside the 95% simulation envelopes of the detailed simulations (Fig. 3a–d), and that for ANPP (Fig. 3e–h). However, the upscaled model was unable to capture the interannual variability (especially in ANPP) and aspects of the short-term dynamics (e.g. 10–20 year cycles) of the original model. This was expected because the definition of the states did not contain variables to capture a 'memory' introduced by the typical life span of *F. pallescens*. However, what is more important is that such smaller errors in describing the short-term fluctuations in population dynamics evened out and had no consequences for the ability of the upscaled model to capture the dynamics of the steppe on the longer term.

APPLICATION OF THE UPSCALED MODEL TO A WATER POINT

Spatial heterogeneity in the local stocking rate induced by a water point in the centre of the paddock produced patterns typically observed in large paddocks (>2000 ha; e.g. Pickup, Bastin & Chewings 1998) (Fig. 4c,f). As expected, no spatial structure emerged without grazing or with spatially homogeneous grazing (Figs. 4a,b,d,e). After 25 years, the locally elevated stocking rate close to the water point resulted in an aggregation of low relative forage values within 1000 m from the water point, but after 100 years the forage value of the entire paddock declined to levels similar to that of homogeneous grazing (cf. Figs. 4c,f).

Discussion

The problem of scaling up information from small scales to understand and predict large-scale patterns is at the core of ecological theory and application. This issue arises because most of the biological information is available at small scales, but larger scale patterns are usually relevant for management and policy (Levin 1992; Urban 2005). We presented here a general approach to upscale detailed dynamic simulation models operating at one scale (in our example the patch scale) to the next higher scale (in our example the landscape scale). Our approach allows for the development of larger scale simulation

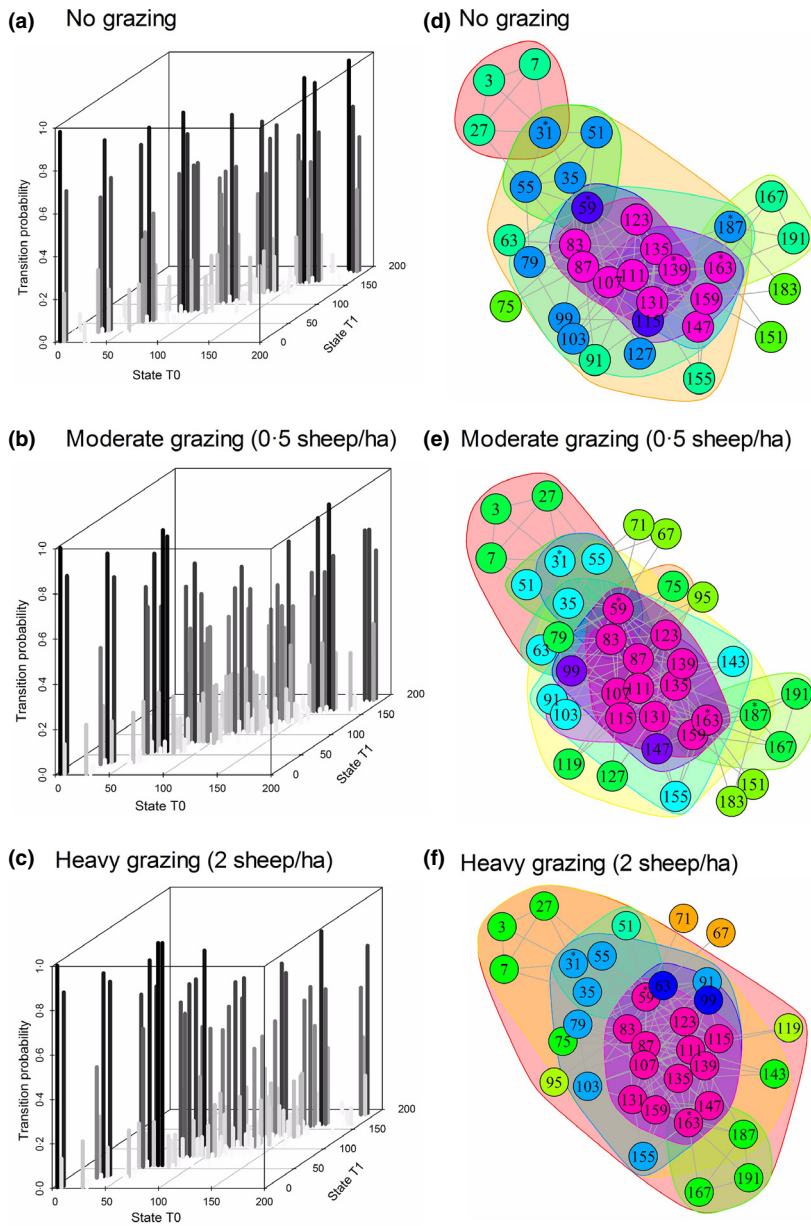


Fig. 2. Analysis of transitions probability matrices (TPMs). (a–c) TPMs as 3D plots (black bars: transition rates close to 1, light grey bars: rates close to 0) for a typical climate series ($\text{MAP} = 375 \text{ mm yr}^{-1}$; $\text{CV} = 20\%$) and for three stocking rates: 0 (a), 0.5 (b) and 2 (c) sheep per ha. (d–f) Cohesive blocks for the same matrices interpreted as undirected graphs. States in a cohesive block (encircled by polygons with the same colour) are internally more connected than with states belonging to other blocks. The colour of the state (i.e. circles) indicates the number of connections (e.g. low: ochre, medium: light blue, and high: magenta). Asterisks indicate the articulation vertices.

models (meta-models) which nevertheless retains the essential details and biological processes from the local scale model.

ANALYSIS OF TRANSITION MATRICES

If a detailed model of vegetation dynamics is available, our approach solves long-standing problems in rangeland ecology related with the state-and-transition concept: the definition of (vegetation) states, the estimation of transition probabilities (e.g. Westoby, Walker & Noy-Meir 1989; Briske, Fuhlendorf & Smeins 2005) and the inclusion of spatial effects (Bestelmeyer, Goolsby & Archer 2011). Analysis of the transition matrices using Markov chain tools can reveal critical states under specific environmental conditions such as absorbing states and states participating in the steady state. Transition probabilities are in general difficult to determine with field studies because dynamic aspects such as transitions among

states occur usually at a much larger time-scale than duration of most field studies (Wiegand, Milton & Wissel 1995; Groffman *et al.* 2006). Application of graph theory allowed us to reveal the detailed properties of the system dynamics with direct implications for management, such as identification of trajectories of degradation or recovery or occurrence of 'critical' states which need to be crossed during degradation (i.e. articulation vertices; Daehyun & Phillips 2013). Graph theory can also capture resilient aspects of networks (Strogatz 2001).

ASSUMPTIONS OF OUR APPROACH

Clearly, upscaling comes with the cost of losing some of the detailed information of the original small-scale model, but in many cases such detail is not really relevant for questions asked at larger scales. The 'art' is therefore to maintain what is essential and to discard what is not essential. A potential weakness

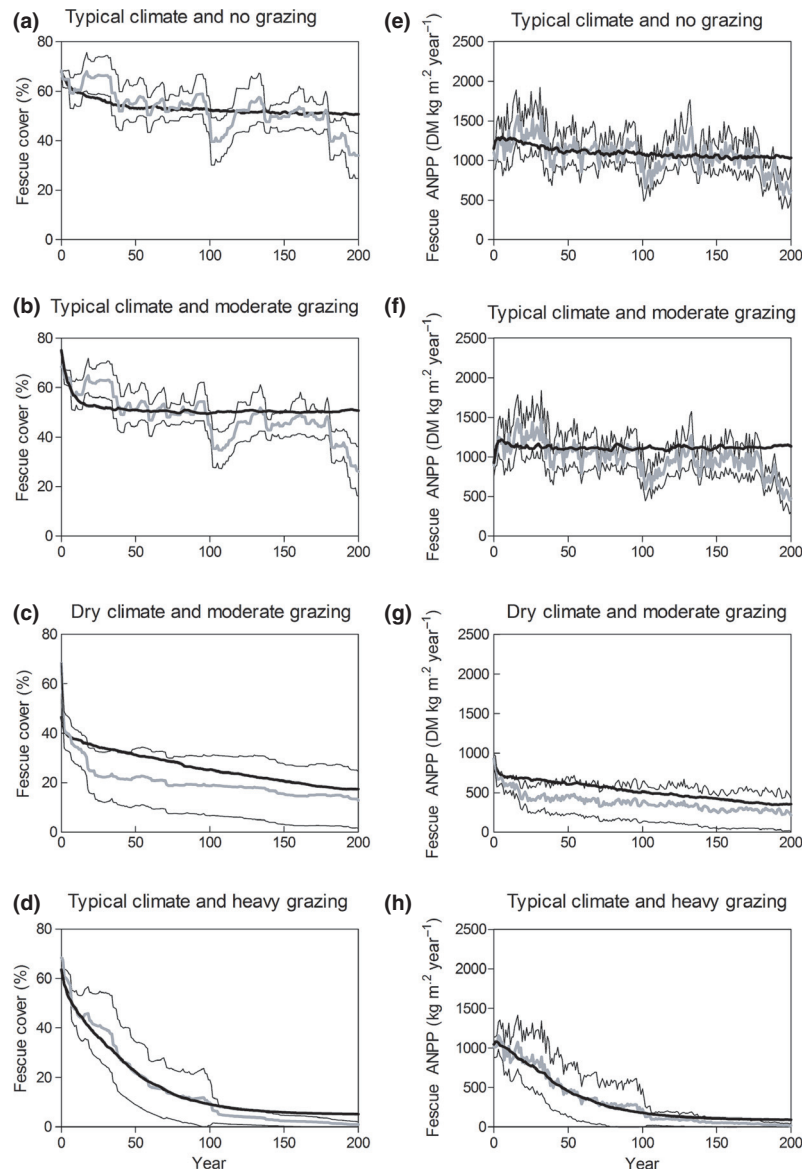


Fig. 3. Test of the upscaled model. Time series of *F. pallescens* cover (a–d) and aerial net primary productivity (e–h) predicted by the upscaled model (black bold line) and mean (bold grey line) and 95% simulation envelopes (thin grey lines) based on 8 parameterizations with the small-scale model. (a, e): typical climate series (MAP = 375 mm·yr⁻¹; CV = 20%) and no grazing, (b, f): typical climate series and 0.5 sheep per ha, (c, g): dry climate series (MAP = 308 mm yr⁻¹; CV = 21%) and 0.5 sheep per ha, (d, h): typical climate series and 2 sheep per ha.

of our approach is that the number of states may become very high if the upscaled model should capture more detail (e.g. also keep track of the short-term dynamics; Fig. 3). However, in general this will not be a severe problem because most of the theoretically possible states will not be realized (in our case <0.5% were realized). If strong correlations between state variables are discovered during the simulations, it is also always possible to reduce the number of state variables that define the states *a posteriori*.

Our approach is nonparametric and does therefore not depend on a particular structure of the large-scale model, but it keeps track of the observed transitions in an abstract state space. Therefore, the most important assumption of our approach is that changes in the states of a given patch depend for a set of environmental condition only on the previous state (s_t) (i.e. a first-order Markov dependence property), and that these changes are invariants through time (Balzter 2000). However, complex systems such as *F. pallescens* steppes may show memory effects that can arise in our example if large cohorts of

F. pallescens establish and die together. Appropriate definition of the states can accommodate memory effects in a way that the Markov property still holds (see also Acevedo, Urban & Abla 1995). For example, if the transition probabilities depend on the 'age' of the state (which may be relevant if cohorts of *F. pallescens* die together), we can use an additional state variable that counts the number of years the patch persisted in a given state before changing. Our approach can also accommodate situations where different vegetation types (e.g. driven by different soil types) co-occur in one landscape by upscaling for each vegetation type a detailed model, and integrating the resulting transition matrices into the framework of the upscaled meta-model.

In our example, we assumed that the change in state of a given patch does not depend on the state of neighbored patches (i.e. spatial independence property). This is a good approximation for our model, given that a patch simulated by COIRON covers an area of approximately 1500 m² and may host in the best state (ID = 191) roughly 12 000–16 000 tufts.

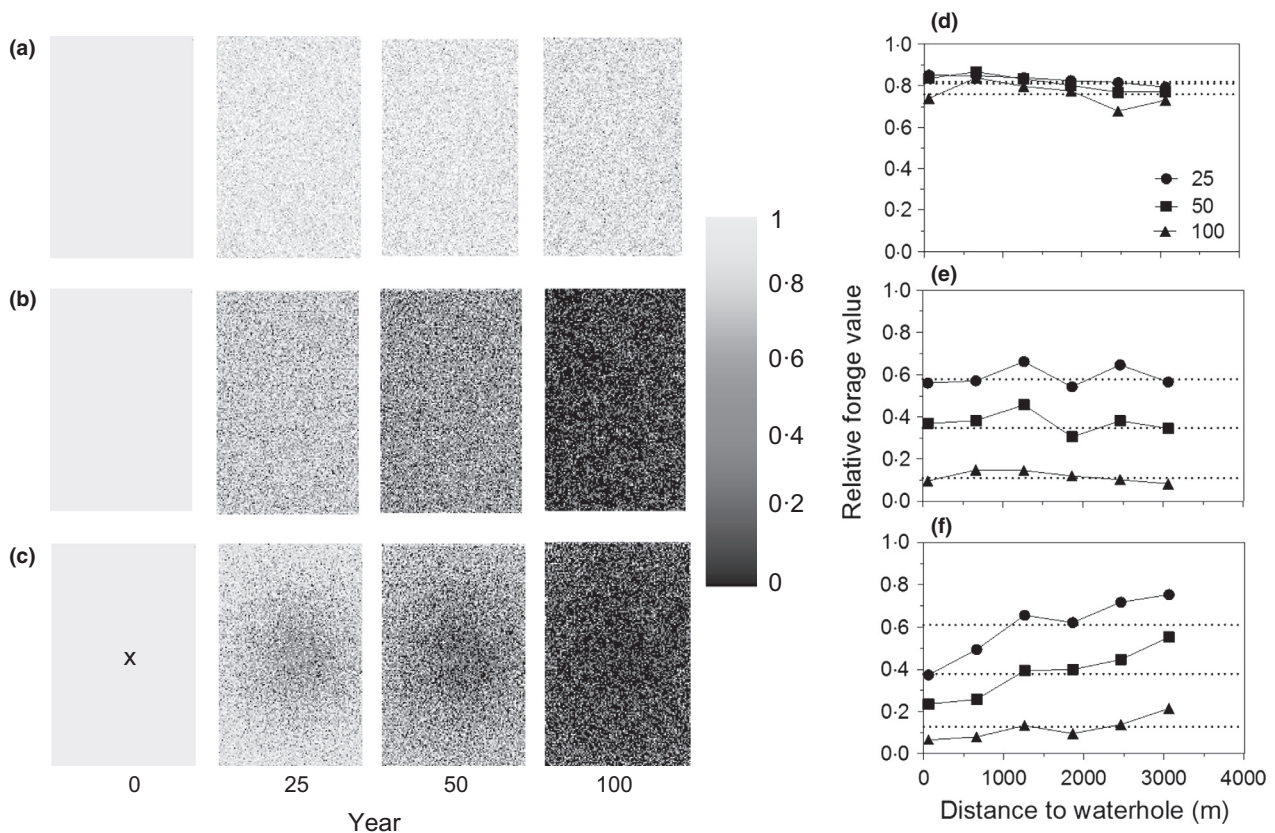


Fig. 4. Application of the upscaled model to a water point. (a–c) Time series of the relative forage value of the vegetation states (= ID/191) of a 135×210 cell paddock at years 0, 25, 50 and 100. All simulations started with a homogeneous paddock in good condition (ID = 187). Light grey: highest relative forage value (1), black: lowest forage value (0). (d–f) Relationships between the average relative forage value at different distance from the water point for years 25, 50 and 100. The panels show results for the following grazing conditions: (a, d) no grazing; (b, e) high stocking rates and homogeneous grazing; (c, f) high stocking rate and a water point at the centre of paddock (marked with 'x') that generates a spatially heterogeneous grazing pressure.

However, expansion of large bare patches or seed input from neighbored patches may influence the transition probabilities. Our approach can accommodate such 'internal' neighbourhood effects by expanding the state space by adding a state variable such as the difference in *F. pallelescens* cover between the focal patch and the neighbouring patches. In this case, we need to expand the extent of the original model to 3×3 patches. We did not consider here memory effects, but except of losing some of the detailed short-term fluctuations, the upscaled model predicted the long-term trends very well (Fig. 3). Clearly, the level of detail in the states depends on the level of detail required for the objectives of the study. Testing the upscaling (e.g. Fig. 3) will show whether the states are detailed enough.

COMPARISON WITH OTHER APPROACHES

Our approach has a number of advantages compared to alternative upscaling methods. For example, model lumping or use of effective parameters does not allow for representation of spatial heterogeneity, and scaling up by brute force will not be feasible for the models of intermediate complexity targeted by our approach. Fitting a larger scale model to the output of a detailed smaller scale model (e.g. Tietjen & Huth 2006)

should require similar simulation effort compared with estimation of transition matrices as all relevant circumstances should be covered. Macroscopic equations (e.g. Strigul *et al.* 2008) are fascinating, but may work only for certain systems where key assumptions such as the perfect plasticity approximation hold (e.g. Strigul *et al.* 2008). When considering detailed states, our approach approximates that embed the original model into a larger scale model (e.g. Meyer, Wiegand & Ward 2009), but again, this may be computationally unfeasible. However, if the scientific question asked requires maintenance of more detail the best one can do to increase the computational speed is to systematically use libraries that approximate parts of the model which are heavily repeated (e.g. Seidl *et al.* 2012).

FEASIBILITY OF OUR APPROACH

In principle, our approach can be applied to any dynamic models for which definition of states and transitions makes sense. However, the computational effort to estimate transitions matrices is the limiting factor. It depends proportionally on (i) the time needed for one simulation run of the original model, (ii) the number of initial conditions, (iii) the number of parameterizations and (iv) the dimensionality of

environmental conditions. For example, if one simulation of the original model takes one minute, our 17 280 simulations runs would need 12 days. However, tallying the transition matrices can easily be parallelized and, for example, 200 cores would reduce the simulation time to 1.5 h. Thus, our upscaling approach is feasible for models of intermediate complexity where one run requires few minutes and where the dimensionality of the environmental drivers is not too high (we used $9 \times 8 = 72$ different environmental conditions).

Conclusions

Upscaling represents a difficult methodological task (Wu & David 2002; Urban 2005), but is needed in a wide range of scientific disciplines to extrapolate knowledge from small scales (small extent and/or fine grains) and low hierarchical levels to larger scales and upper hierarchical levels. Our nonparametric upscaling approach applies in a way *brute force* to derive transition probabilities for generalized states. Not relying on a predefined model structure provides it with flexibility that allows application for a great variety of models and situations. It is clear that the transition probabilities must be enriched in most cases with additional rules that describe phenomena occurring at the larger scale. While our approach can describe the influence of the state of a neighboured patch by appropriate definition of the state variables and simulating a larger area of the original model, inclusion of higher scale processes can be accomplished by adopting a meta-model approach (Urban 2005). In this approach, rules are added to the transition probabilities that describe spatial heterogeneity at the large scale as well as spatial interactions among patches. We expect that our nonparametric upscaling approach can be applied to a wide range of simulation models of intermediate complexity and closes a gap in our ability to scale up from small scales where biological knowledge is available to the larger, management relevant scales.

Acknowledgements

We thank Andreas Huth, the Associate Editor and two anonymous reviewers for comments. The studies reported in this manuscript comply with the ethics guidelines and current laws of the Republic of Argentina. This work was supported by grants from DAAD and Consejo Nacional de Investigaciones Científicas y Tecnológicas (PIP0034). TW was supported by the European Research Council (ERC) advanced grant 233066. PAC was supported by a post-doc Georg Forster fellowship from the Alexander von Humboldt Foundation.

Data accessibility

This manuscript does not use field data, because it was based on simulations with a published model for Patagonian grass steppes (Paruelo *et al.* 2008).

References

Acevedo, M.F., Urban, D.L. & Ablan, M. (1995) Transition and gap models of forest dynamics. *Ecological Applications*, **5**, 1040–1055.
Balzter, H. (2000) Markov chain models for vegetation dynamics. *Ecological Modelling*, **126**, 139–154.

Bertiller, M.B. & Bisigato, A. (1998) Vegetation dynamics under grazing disturbance. The state-and-transition model for the Patagonian steppes. *Ecología Austral*, **8**, 191–199.
Bestelmeyer, B.T., Goolsby, D.P. & Archer, S.R. (2011) Spatial perspectives in state-and-transition models: a missing link to land management? *Journal of Applied Ecology*, **48**, 746–757.
Boulain, N., Simoni, G. & Gignoux, J. (2007) Changing scale in ecological modelling: a bottom up approach with an individual based vegetation model. *Ecological Modelling*, **203**, 257–269.
Briske, D.D., Fuhlendorf, S.D. & Smeins, F.E. (2005) State-and-Transition models, thresholds, and rangeland health: a synthesis of ecological concepts and perspectives. *Rangeland Ecology & Management*, **58**, 1–10.
Csardi, G. & Nepusz, T. (2006) The *igraph* software package for complex network research. Inter Journal Complex Systems, 1695. <http://igraph.org>
Daehyun, K. & Phillips, J.D. (2013) Predicting the structure and mode of vegetation dynamics: an application of graph theory to state-and-transition models. *Ecological Modelling*, **265**, 64–73.
Denny, M. & Benedetti-Cecchi, L. (2012) Scaling up in Ecology: mechanistic approaches. *Annual Review of Ecology, Evolution, and Systematics*, **43**, 1–22.
Grimm, V. & Railsback, S.F. (2005) *Individual-Based Modeling and Ecology*. Princeton University Press, Princeton, NJ, USA.
Grimm, V. & Railsback, S.F. (2012) Pattern-oriented modelling: a 'multi-scope' for predictive systems ecology. *Philosophical Transactions of the Royal Society B, Biological Sciences*, **367**, 298–310.
Grimm, V., Revilla, E., Berger, U., Jeltsch, F., Mooij, W., Railsback, S.F. *et al.* (2005) Pattern-oriented modeling of agent-based complex Systems: lessons from ecology. *Science*, **311**, 987–991.
Groffman, P.M., Baron, J.S., Blett, T., Gold, A.J., Goodman, I., Gunderson, L.H. *et al.* (2006) Ecological thresholds: the key to successful environmental management or an important concept with no practical application? *Ecosystems*, **9**, 1–13.
Kéfi, S., Rietkerk, M., Alados, C.L., Pueyo, Y., Papanastasis, V.P., ElAich, A. & de Ruiter, P.C. (2007) Spatial vegetation patterns and imminent desertification in Mediterranean arid ecosystems. *Nature*, **449**, 213–217.
Levin, S.A. (1992) The problem of pattern and scale in ecology: the Robert H. MacArthur Award Lecture. *Ecology*, **73**, 1943–1967.
May, R.M. (1999) Unanswered questions in ecology. *Philosophical Transactions of the Royal Society B, Biological Sciences*, **354**, 1951–1959.
Melbourne, B.A. & Chesson, P. (2006) The scale transition: scaling up population dynamics with field data. *Ecology*, **87**, 1478–1488.
Meyer, K.M., Wiegand, K. & Ward, D. (2009) Patch dynamics integrate mechanisms for savanna tree-grass coexistence. *Basic and Applied Ecology*, **10**, 491–499.
Moorcroft, P.R., Hurtt, G.C. & Pacala, S.W. (2001) A method for scaling vegetation dynamics: the ecosystem demography model (ED). *Ecological Monographs*, **71**, 557–586.
Pacala, S.W., Canham, C.D., Saponara, J., Silander, J.A., Kobe, R.K. Jr & Ribbens, E. (1996) Forest models defined by field measurements: II. Estimation, error analysis and dynamics. *Ecological Monographs*, **66**, 1–43.
Paruelo, J.M., Bertiller, M.B., Schlichter, T. & Coronato, F. (1993) Secuencias de deterioro en distintos ambientes patagónicos: su caracterización mediante el modelo de estados y transiciones. Convenio Argentino-Alemán de cooperación técnica INTA-GTZ. Proyecto Ludepa-SME, San Carlos de Bariloche, p. 110.
Paruelo, J.M., Pütz, S., Weber, G., Bertiller, M.B., Golluscio, R.A., Aguiar, M.R. & Wiegand, T. (2008) Long-term dynamics of a semiarid grass steppe under stochastic climate and different grazing regimes: a simulation analysis. *Journal of Arid Environments*, **72**, 2211–2231.
Peters, D.P.C., Pielke, R.A., Bestelmeyer, B.T., Allen, C.D., Munson-McGee, S. & Havstad, K.M. (2004) Cross-scale interactions, nonlinearities, and forecasting catastrophic events. *Proceedings of the National Academy of Sciences*, **101**, 15130–15135.
Pickup, G., Bastin, G.N. & Chewings, V.H. (1998) Identifying trends in land degradation in non-equilibrium rangelands. *Journal of Applied Ecology*, **35**, 365–377.
Pickup, G. & Chewings, V.H. (1994) A grazing gradient approach to land degradation assessment in arid areas from remotely-sensed data. *International Journal of Remote Sensing*, **15**, 597–617.
R Core Team (2013) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
Rastetter, E.B., King, A.W., Cosby, B.J., Hornberger, G.M., O'Neill, R.V. & Hobbie, J.E. (1992) Aggregating fine-scale ecological knowledge to model coarser-scale attributes of ecosystems. *Ecological Applications*, **2**, 55–70.

- Rastetter, E.B., Aber, J.D., Peters, D.P.C. & Ojima, D.S. (2003) Using mechanistic models to scale ecological processes across space and time. *BioScience*, **53**, 1–9.
- Seidl, R., Rammer, W., Scheller, R.M. & Spies, T.A. (2012) An individual-based process model to simulate landscape-scale forest ecosystem dynamics. *Ecological Modelling*, **231**, 87–100.
- Shugart, H.H. (1998) *Terrestrial Ecosystems in Changing Environments*. Cambridge Studies in Ecology, Cambridge University Press, Cambridge.
- Strigul, N., Pristinski, D., Purves, D., Dushoff, J. & Pacala, S. (2008) Scaling from trees to forests: tractable macroscopic equations for forest dynamics. *Ecological Monographs*, **78**, 523–545.
- Strogatz, S.H. (2001) Exploring complex networks. *Nature*, **410**, 268–276.
- Tietjen, B. & Huth, A. (2006) Modelling dynamics of managed species-rich forests - an aggregated approach. *Ecological Modelling*, **199**, 421–432.
- Urban, D.L. (2005) Modeling ecological processes across scales. *Ecology*, **86**, 1996–2006.
- Van de Koppel, J., Gupta, R. & Vuik, C. (2011) Scaling-up spatially-explicit ecological models using graphics processors. *Ecological Modelling*, **222**, 3011–3019.
- Westoby, M., Walker, B. & Noy-Meir, I. (1989) Opportunistic management for rangelands not at equilibrium. *Journal of Range Management*, **42**, 266–274.
- White, D.R. & Harary, F. (2001) The cohesiveness of blocks in social networks: node connectivity and conditional density. *Sociological Methodology*, **31**, 305–359.
- Wiegand, T. & Milton, S.J. (1996) Vegetation change in semiarid communities: simulating probabilities and time scales. *Vegetation*, **125**, 169–183.
- Wiegand, T., Milton, S.J. & Wissel, C. (1995) A simulation model for a shrub ecosystem in the semiarid Karoo, South Africa. *Ecology*, **76**, 2205–2211.
- Wiens, J.A., Stenseth, N.C., Van Horne, B. & Ims, R.A. (1993) Ecological mechanisms and landscape ecology. *Oikos*, **66**, 369–380.
- Wu, J. (1999) Hierarchy and scaling: extrapolating information along a scaling ladder. *Canadian Journal of Remote Sensing*, **25**, 367–380.
- Wu, J. & David, J.L. (2002) A spatially explicit hierarchical approach to modeling complex ecological systems: theory and applications. *Ecological Modelling*, **153**, 7–26.
- Wu, J. & Loucks, O.L. (1995) From balance of nature to hierarchical patch dynamics: a paradigm shift in Ecology. *Quarterly Review of Biology*, **70**, 439–466.
- Wu, J., Jones, K.B., Li, H. & Loucks, O.L. (2006) *Scaling and Uncertainty in Ecology. Methods and Applications*. p. 353. Springer, the Netherlands.

Received 20 May 2015; accepted 16 August 2015

Handling Editor: Justin Travis

Supporting Information

Additional Supporting Information may be found in the online version of this article.

Appendix S1. COIRON simulation model for semi-arid grasslands.

Appendix S2. Methodological specifications.