Spatial upscaling of process-based vegetation models: An overview of common methods and a case-study for the U.K.

Marcel van Oijen[1], Amanda Thomson[1] & Frank Ewert[2]

[1] Centre for Ecology and Hydrology (CEH-Edinburgh), U.K.
mvano@ceh.ac.uk, amath@ceh.ac.uk

[2] University of Bonn, Germany
frank.ewert@uni-bonn.de

Abstract. Many different process-based models of vegetations are in use today. The majority of these models are parameter-rich, deterministic dynamic models, which require considerable input information and computation time. These characteristics, combined with the fact that the models tend to be parameterised at the point-support spatial scale, have made their use for larger regions problematic. Numerous examples of regional model application do exist, but how upscaling from point to region affects model output uncertainty is generally not considered. We begin by proposing a classification of upscaling methods for process-based models. Seven different methods of spatial upscaling are identified, most of which have been used in practice. We then present the example of the application of forest models to the U.K. at a 20 x 20 km grid. A discussion on upscaling uncertainty, mostly from a Bayesian perspective, concludes the paper.

1 Introduction

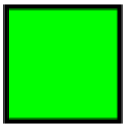
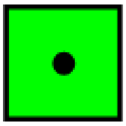
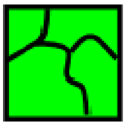
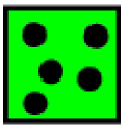
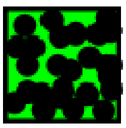
Process-based models (PBMs) have become a standard tool for explaining and predicting changes in ecosystems [11]. This paper addresses PBMs that simulate, deterministically, the cycling of carbon, water and nutrients within vegetation and across the boundaries with soil and atmosphere. The models tend to be complex and highly nonlinear. They are written as sets of differential equations which are solved numerically. In contrast to hydrological models, vegetation models focus on vertical transport, i.e. geographically they operate at point-support. The models require two types of input: (1) time series of local environmental drivers like weather conditions, (2) constant parameters that determine sensitivity of processes to both vegetation state and the environment. The primary outputs of the models are time series of the state variables and process rates, from which summary outputs are often calculated. The perception that PBMs can address the impact of a sheer unlimited range of environmental factors and their interactions, through simulation of processes occurring in the real system, has contributed to their increasing use. For example, for forestry alone, the Register of Ecological Models (http://ecobas.org/www-server/index.html) lists 78 different published models.

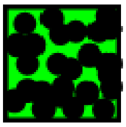
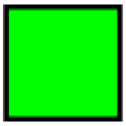
Computationally, the models tend to be demanding. Statistically, the models tend to be characterised by uncertainty in both inputs and model structure. These characteristics hamper model upscaling from point to region, and in particular the quantification of model output uncertainty at the regional scale [5]. The present paper proposes a complete classification of PBM upscaling methods in seven categories, giving examples from the literature where available. The upscaling method currently being considered for the United Kingdom's Greenhouse Gas Inventory is presented in more detail – as a focal point for discussion of the strengths and weaknesses of all seven approaches.

2 AN OVERVIEW OF METHODS FOR UPSCALING PROCESS-BASED MODELS

The problem facing a 'regional' user of point-support PBMs is that the models tend to be too slow to be run exhaustively across the region, even if environmental information happens to be available for every point in the region. This problem can only be solved by working with approximations, of either region or model. This implies using some form of sampling and/or model simplification. We can distinguish seven methods, each described briefly in Table 1.

Table 1: Methods for regional application of point-support process-based models. Each algorithm description ends with a line that shows how to calculate regional totals of model outputs. "y" is model output (stocks, fluxes, etc.) per unit area, "area" is the area of the region, "Total" is the integral of y across the regional area.

Category	Upscaling method		Algorithm
(1)-(3): Methods that use the original model		(1) Reinterpret the point-support model as being a regional one	1. Re-calibrate model for regional use 2. Get average inputs for region 3. Run model for average conditions 4. Total = y * area
		(2) Select representative point	1. Select a representative point 2. Get inputs for point 3. Run model for point 4. Total = y * area
		(3) Stratify into homogeneous subregions	1. Stratify the region 2. Get average inputs per stratum 3. Run model for all strata 4. Total = $\Sigma(y_i * area_i)$
(4): Method in which the original model is extended		(4) Run for selected points & interpolate	1. Select control points 2. Get inputs for control points 3. Run model for control points 4. Derive geostatistical interpolation model f 5. Total = $\int y(s)$ where $y(s) = f(y_{1..n}, s_{1..n})$
(5)-(7): Methods that rely on making a new model		(5) Create deterministic metamodel & apply exhaustively across the region	1. Create fast metamodel using training set of multiple point-scale I-O 2. Get inputs across whole region 3. Run model for all points 4. Total = mean(y) * area

	(6) Create stochastic emulator & apply exhaustively across the region	1. Create fast emulator using training set of multiple point-scale I-O 2-4. As Method (5)
	(7) Summarize model behaviour & embed in regional model with wider scope than point-support model	1. Summarise model behaviour in the form of a regional summary model 2. Embed summary in regional model 3. Run the regional model using regional-scale inputs 4. Total = y * area

The first three methods calculate regional values without any change to model equations or computer code. The most commonly used method for upscaling PBMs probably is the first one, where the model is simply assumed to be valid at the regional scale, such that it can be supplied with regionally averaged environmental drivers. Because PBMs are nonlinear, this direct regionalisation requires recalibration of all parameters, but this is often not done. An example is the work on the U.K.'s greenhouse gas balance discussed below (§ 3).

The second method consists of identifying a single representative point in the region, and running the PBM only for that point. The point is often chosen to be a weather station [3] but the problem is that we need a point for which model output is representative, not model input.

A refinement is method (3), where the region is stratified into geographic or virtual (not necessarily contiguous) sub-regions. Examples include all model applications where the region of interest is subdivided in small grid cells.

The fourth method involves geostatistics. Whereas the previous two methods are referred to as 'design-based approaches to spatial aggregation' [7], geostatistics is 'model-based'. This is because the original PBM is extended with a stochastic interpolation model that takes the PBM's output for selected control points across the region, and calculates regional values under given assumptions of spatial correlation. This procedure introduces new parameters for the spatial variogram and any regression on covariates. As far as we are aware, geostatistical techniques have not yet been used to scale-up vegetation PBMs.

The last three methods do not use the original PBM code directly, but use summary forms of it instead, to increase computational speed. In method (5) the PBM is summarized as a deterministic metamodel that replaces the dynamic simulation with an approximate analytical solution. The metamodel may then be applied densely across the region, obviating the need for any further upscaling.

The sixth method is more often applied in climate science than in vegetation science. It consists of creating an emulator of the PBM, which predicts the output of the PBM stochastically, e.g. using a Gaussian Process approach. The output of the emulator is a multivariate probability density function (PDF) rather than a vector estimate. Goldstein and Rougier [6] discuss stochastic emulators of environmental models.

The final method consists of using an existing regional-support model, possibly including hydrology or multiple vegetation types, and incorporating in it a summary form of the PBM. Ewert et al. [4] discuss examples of such nested simulation.

3 EXAMPLE: THE UK GREENHOUSE GAS BALANCE AND FOREST MODELS

The annual contribution of afforestation to the greenhouse gas balance of the U.K. is calculated using CFLOW, a simple PBM for stocks and fluxes of carbon in vegetation and soil. Results are reported to government and UNFCCC [15]. Only minor stratification is used, into the constituting countries England, Scotland, Wales and Northern Ireland. Currently we are investigating the scope for finer spatial resolution and the use of a more detailed PBM, 'BASFOR' [17]. Figure 1 shows some of BASFOR's input, output and output uncertainty for a grid of 20x20 km (Northern Ireland not yet included).

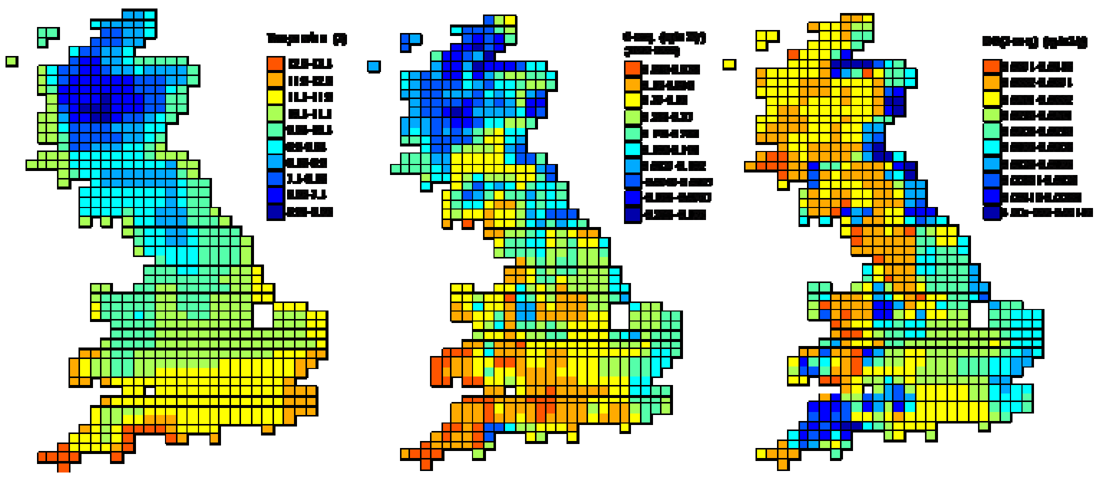


Figure 1: Process based modeling of C-sequestration in the UK, 1920-2000. Cells are 20x20 km. Left: mean temperature (source: UKCIP). Mid: average annual C-sequestration simulated using forest model BASFOR. Right: uncertainty (S.D.) of outputs shown in mid panel.

BASFOR is a 40-parameter point-support model. Upscaling to the national scale obviously used method (3), stratification, by way of the 655 grid cells. However, it is clear that the strata are not homogeneous, as soil type, elevation and distance from the sea – and thus weather – change over much shorter length scales than 20 km. The BASFOR study thus also used method (1) to scale from point to cell. However, interpreting BASFOR as a cell-scale model was dubious because parameterisation had been carried out at the point-scale. This was done by means of Bayesian calibration using Markov Chain Monte Carlo exploration of the 40-dimensional parameter space, with growth data from small forest stands [16, 17]. The uncertainty shown on the right in Fig. 1 thus only represents point-scale parameter uncertainty (the posterior PDF from the Bayesian calibration), and does not include the additional uncertainty from the upscaling.

4 QUANTIFICATION OF UNCERTAINTY ASSOCIATED WITH UPSCALING

The example of the previous section showed the most common approach to PBM upscaling: combining methods (1) and (3). Method (1) strictly requires reparameterisation of the PBM, using regional I-O data, but this is rarely done. An exception is the work of Patenaude et al. [12], who calibrated the 3PG model using remotely sensed data from a small forested region. Recent years have seen an increasing abundance of data on structural characteristics of vegetation, measured by remote sensing, and on gas-fluxes, measured by eddy-covariance towers. This makes direct regional parameterisation increasingly possible. If the parameterisation is carried out using probabilistic techniques like Bayesian calibration [12], uncertainty quantification will be included.

The sampling-based methods (2)-(4) require uncertainty quantification to account for assumptions of representativeness of strata or control points. Recent developments in geostatistics may provide methodology. Bayesian kriging, for example, affords a means to quantify uncertainties comprehensively, including the uncertainty of the spatial interpolation parameters [1].

Methods (5)-(7) use new models that approximate the original PBM. Probabilistic frameworks for dealing with such 'models of models' are still subject of intense research [6, 8], and there is not yet a generally accepted method. The debate is mainly about how to account for the discrepancy between models and reality.

5 DISCUSSION

Classification systems for methods of model upscaling have been presented before (e.g. [4]). Our approach was intended to be more wide-ranging in that we include both methods from spatial statistics and methods that have been used in practice for PBMs.

Upscaling involves approximation and therefore introduces uncertainty about the quality of that approximation. We found no studies that quantified spatial upscaling uncertainty for PBMs, except for limited comparisons of sampling schemes (see also [13, 14]). For temporal upscaling there is the work of Nonhebel [9] who showed that a PBM for wheat overestimated yields by 5-15% when weather data were supplied as 10-day averages rather than daily values. This was because weather averaging removed extreme conditions. The same is likely to apply with spatial averaging.

The very fact that PBMs tend to be too slow, computationally, to be run exhaustively across heterogeneous regions – necessitating the upscaling techniques we discussed in this paper – makes it hard to quantify upscaling uncertainty. There is a need for a practical probabilistic framework that produces PDFs for the regional model output, conditional on input distribution (environment and parameters), upscaling assumptions and upscaling method. Possibly the recent flourishing, in ecological research, of Bayesian hierarchical models [2, 10, 11] will lead to such a framework.

6 REFERENCES

[1] Banerjee, S., B.P. Carlin and A.E. Gelfand (2004). "Hierarchical Modeling and Analysis for Spatial Data." Chapman Hall, Boca Raton: 472 pp.

[2] Cressie, N., C.A. Calder, J.S. Clark, J.M. Ver Hoef and C.K. Wikle (2009). "Accounting for uncertainty in ecological analysis: the strengths and limitations of hierarchical statistical modeling." Ecological Applications 19: 553-570.

[3] De Koning, G.H.J. and C.A. Van Diepen (1993). "Crop production potential of rural areas within the European Communities IV". WRR, The Hague: 84 pp.

[4] Ewert, F., H. Van Keulen, M. van Ittersum, K. Giller, P. Leffelaar and R. Roetter (2006). "Multi-scale analysis and modelling of natural resource management options." iEMSs Third biennial Meeting, Burlington, USA, July 2006. International Environmental Modelling and Software Society: 6 pp.

[5] Goidts, E. (2009). Soil organic carbon evolution at the regional scale: overcoming uncertainties & quantifying driving forces. PhD-thesis, Univ. cath. Louvain: 241 pp.

[6] Goldstein, M. and J. Rougier (2009). "Reified Bayesian modelling and inference for physical systems." Journal of Statistical Planning and Inference 139: 1221-1239.

[7] Heuvelink, G.B.M. and E.J. Pebesma (1999). "Spatial aggregation and soil process modelling." Geoderma 89: 47-65.

[8] Kennedy, M. C. and A. O'Hagan (2001). "Bayesian calibration of computer models." Journal of the Royal Statistical Society Series B-Statistical Methodology 63: 425-450.

[9] Nonhebel, S. (1994). "The effects of use of average instead of daily weather data in crop growth simulation models." Agricultural Systems 44: 377-396.

[10] Ogle, K. (2009). "Hierarchical Bayesian statistics: merging experimental and modelling approaches in ecology. " Ecological Applications 19: 577-581.

[11] Ogle, K. and J.J. Barber (2008) "Bayesian data-model integration in plant physiological and ecosystem ecology." Progress in Botany 69:281-311

[12] Patenaude, G., R. Milne, M. Van Oijen, C.S. Rowland and R.A. Hill (2008). "Integrating remote sensing datasets into ecological modelling: a Bayesian approach." International Journal of Remote Sensing 29: 1295 - 1315.

[13] Rastetter, E.B., A.W. King, B.J. Cosby, G.M. Hornberger, R.V. O'Neill and J.E. Hobbie (1992). "Aggregating fine-scale ecological knowledge to model coarser-scale attributes of ecosystems." Ecological Applications 2: 55-70.

[14] Russell, G. & Van Gardingen, P.R. (1997). "Problems with using models to predict regional crop production." In: Van Gardingen, P.R., G.M. Foody and P.J. Curran (Eds), "Scaling-up: From Cell to landscape". CUP, Cambridge: 273-294.

[15] Thomson, A.M. and M. Van Oijen (Eds) (2007). "Inventory and Projections of UK Emissions by Sources and Removals by Sinks due to Land Use, Land Use Change and Forestry." Centre for Ecology and Hydrology / DEFRA, London: 197 pp.

[16] Van Oijen, M., J. Rougier and R. Smith (2005). "Bayesian calibration of process-based forest models: bridging the gap between models and data." Tree Physiology 25: 915-927.

[17] Van Oijen, M. and A. Thomson (subm.). "Towards Bayesian uncertainty quantification for forestry models used in the U.K. GHG Inventory for LULUCF.“ Climatic Change.