



# Grid origin affects scaling of species across spatial scales

Jan-Philip M. Witte<sup>1,2\*</sup>, Fangliang He<sup>3</sup> and Cees L.G. Groen<sup>4</sup>

<sup>1</sup>Kiwa Water Research, P.O. Box 1072, 3430 BB Nieuwegein, The Netherlands,

<sup>2</sup>Vrije Universiteit, Institute of Ecological Science, De Boelelaan 1085, 1081 HV Amsterdam, The Netherlands,

<sup>3</sup>Department of Renewable Resources, University of Alberta, 751 General Services Building, Edmonton, Alberta, T6G 2H1

Canada, <sup>4</sup>FLORON Foundation, P.O. Box 9514, 2300 RA Leiden, The Netherlands

## ABSTRACT

**Aim** Distribution maps of species based on a grid are useful for investigating relationships between scale and the number or area of occupied grid cells. A species is scaled up simply by merging occupied grid cells on the observation grid to successively coarser cells. Scale–occupancy relationships (SORs) obtained in this way can be used to extrapolate species down, in other words to compute occupancies at finer scales than the observation scale. In this paper we demonstrate that the SOR is not unique but depends on where one positions the origin of the grid map.

**Innovation** The effect of grid origin on SORs was explored with the aid of the Dutch national data base FLORBASE, which contains the observation records of all 1410 wild vascular plants in the Netherlands on a 1-km square basis. For each species, we generated 2500 unique SORs by scaling up from 1 km, in steps of 1 km, to squares of 50 km. We computed the sensitivity of the SOR to the grid origin for each species, and subsequently analysed the factors that determined this sensitivity. The effect of grid origin on downscaling was demonstrated by means of a simple power function that we used to extrapolate down from both a 2-km and a 5-km grid, to the original 1-km grid.

It appeared that the position of grid origin could have a substantial effect on SORs. The sensitivity of SORs to the position of the grid origin depended on three characteristics of a species' spatial distribution: rarity, degree of spatial clustering and the position of the distribution relative to the border of the investigated area. Rare species with a clustered distribution near the border were particularly highly sensitive. The dependence of SOR on grid origin caused unpredictable and non-random errors in downscaled occupancies.

**Main conclusions** In future, the whole bandwidth of scaled occupancies should be considered when testing and interpreting mathematical relationships between scale and occupancy. Moreover, downscaled occupancies should be interpreted cautiously.

## Keywords

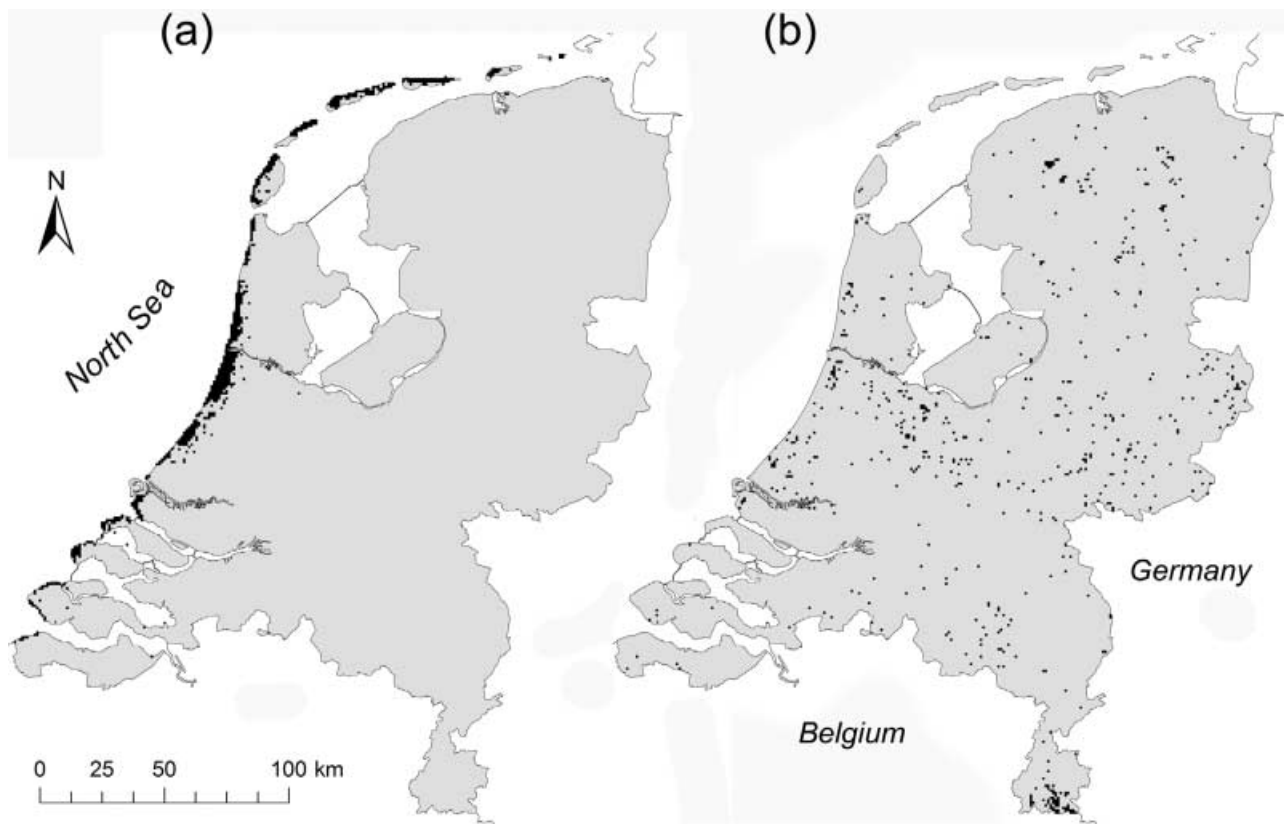
Distribution map, downscaling, FLORBASE, fractal dimension, grid origin, rarity, spatial clustering, spatial scale, upscaling.

\*Correspondence: Jan-Philip Witte, Kiwa Water Research, P.O. Box 1072, 3430 BB Nieuwegein, The Netherlands.  
E-mail: flip.witte@kiwa.nl

## INTRODUCTION

The distribution of species is usually mapped on a grid basis. Grid cells in which a species is observed at least once within a certain period are classified as 'present' and indicated by means of a black dot on such a grid map (Fig. 1). Grid maps have often been used to analyse spatial patterns of occurrence, assemblage composition and species richness, as well as to plan conservation

strategies (see Araújo *et al.*, 2005, and references therein). They have also been employed to determine the degree to which species are rare within a particular country or other human-made administrative region, by totalling the number of occupied cells. Such rarity numbers may be used to monitor species, compile Red Lists and quantify species conservation values (Margules & Usher, 1981; Dony & Denholm, 1985; Wheeler, 1988; Witte, 1998; De Jongh *et al.*, 2003).



**Figure 1** Distribution of two species in the Netherlands according to FLORBASE: (a) dune pansy (*Viola curtisii*) and (b) woodruff (*Galium odoratum*). Both species have a comparable rarity ( $O_{1\text{ km}} = 662$  and  $576 \text{ km}^2$ , respectively), but their distribution differs considerably with respect to the degree of spatial clustering.

Grid maps of different territories often vary considerably within the width of the grid squares. The national grid of the Netherlands, for instance, has a resolution of 1 km, Great Britain's grid is based on 10-km squares, while the grid cells of the *Atlas Florae Europaeae* approximately measure 50 km. Moreover, grid maps of the same country may vary over time (Van der Meijden *et al.*, 2000). To be able to compare countries or inventory periods with different grid resolutions, distribution data need to be upscaled or downscaled. Downscaling may also be advisable if one is interested in using a finer-scale distribution to better represent the 'actual' rarity of species.

Upscaling of grid information is considered straightforward; a cell of a coarser grid is classified as 'present' whenever this cell comprises one or more black dots of the original finer observation grid. Downscaling distribution, by contrast, is much more challenging. This involves inferring fine-grid distribution from coarse-grid information. Two species that display the same coarse-grid information can have very different fine-scale distributions (He & Condit, 2007). Recently, much attention has been paid to developing downscaling methods and to applying these methods for conservation purposes (Kunin, 1998; He & Gaston, 2000; Kunin *et al.*, 2000; Warren *et al.*, 2003; Witte & Torfs, 2003; Hartley *et al.*, 2004; Tosh *et al.*, 2004; Wilson *et al.*, 2004; Fagan *et al.*, 2005; He & Reed, 2006). Downscaling is performed by

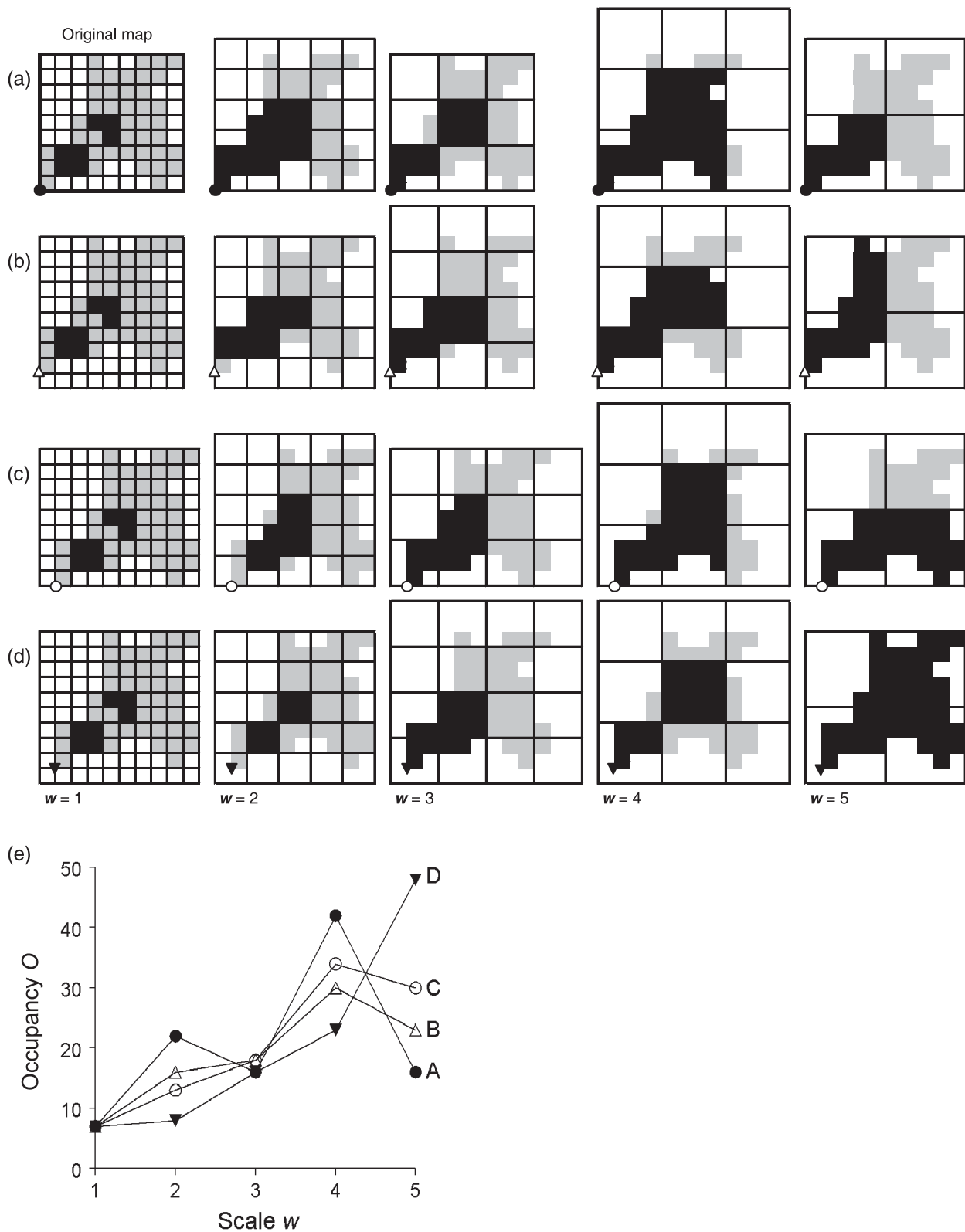
extrapolating down from relationships between scale (grid resolution) and the number or area of occupied cells. Such scale-occupancy relationships (hereafter SORs) can be generated from the original observation data by merging cells to successively coarser scales. All the previous studies assume that each species has a unique SOR for a given map.

In this paper, we demonstrate that the SOR is not unique but depends on the origin of the grid map. We will explore which kinds of species are especially sensitive to grid origin and demonstrate how this dependence influences downscaling occupancies from coarse-scale data.

## INNOVATION

### The problem of grid origin

Figure 2 illustrates the problem of grid origin with a hypothetical distribution on an observation grid measuring  $9 \times 9$  square cells. Cells occupied by a certain species are in black, grey cells reflect empty cells and white cells are outside the investigated area (e.g. foreign countries or sea). The top-left map in Fig. 2 contains the original data from the observation grid. All the other maps are derived from this basic map. From left to right, the grid square width  $w$  increases from 1 to 2, 3, 4 and 5, respectively. With that,



**Figure 2** The effect of grid origin on occupancy scaling, as demonstrated by a hypothetical observation grid of  $9 \times 9$  square cells. From left to right: scaling up from a grid with a square width of  $w_0 = 1$  to grids with  $w = 2, 3, 4$  and  $5$ . From top to bottom: (a) the original origin, (b) origin moved one cell up, (c) origin moved one cell to the right, (d) origin moved one cell up and one to the right. On maps (a)–(d) the original origin is indicated with different symbols ( $a = \bullet$ ,  $b = \triangle$ ,  $c = \circ$ ,  $d = \blacktriangledown$ ). Black = occupied area, grey = empty area, white = outside the study area (e.g. foreign countries or sea). (e) Scale–occupancy relationships corresponding to the above four different grid origins, illustrating the effect of changing the grid origin.

the total area of occupied grid cells, the occupancy  $O$ , changes as well. Following Witte & Torfs (2003), we exclude the area outside the area of investigation from the computation of  $O$  (thus, it is only the black part in Fig. 2 that counts). Moreover, for each value of  $w$  we adjust the number of rows and the number of columns of the grid in such a way that dividing these numbers by  $w$  yields integers. Since  $O$  is calculated as the area occupied within the investigated area, adjusting the grid size does not influence the result: any size will do, as long as the grid comprises the country. For instance, in Fig. 2(a) the grid measures  $9 \times 9$  for  $w = 1$  and  $w = 3$ ,  $10 \times 10$  for  $w = 2$  and  $w = 5$ , and  $12 \times 12$  for  $w = 4$ .

The consequence of changing the grid origin becomes clear when we compare the maps from top to bottom (Fig. 2). Compared with the maps in row (a), the origin of the maps in row (b) is one grid cell lower; in the maps in row (c) it is one grid cell to the left; and it is one left and one down in row (d). Note in each map the location of the origin of the original map (a) is indicated by a small symbol. In Fig. 2 it is clearly visible that the area of occupied black cells not only changes with  $w$  (from left to right) but with grid origin as well (from top to bottom). The consequence of changing  $w$  and origin on the SORs is summarized in Fig. 2(e). Large differences in  $O$  are visible, especially at  $w = 5$ , where  $O_a = 16$  and  $O_d = 48$  grid cells. Moreover, the SORs of (a), (b) and (c) have a wavy pattern, all three with a local maximum at  $w = 4$ , and (a) also at  $w = 2$ . Scaling up is apparently already not straightforward given the example of the four SORs in Fig. 2.

Given that one could use any cell as the grid origin for scaling occupancy, this becomes still more problematic. However, there is a maximum to the number of potentially different values for  $O$  at each scale  $w$ . In Fig. 2 we showed four grid origins that generated unique solutions for  $w = 2$ . So if, for example, we move the original origin of Fig. 2(a) down by three cells, we obtain the same patterns as in row (b); two down and one to the left produces the same patterns as row (c), and so on. In general, at a scale of  $w$  there are  $(w/w_0)^2$  different ways to scale up. Since we scaled up to  $w = 5$ , there are  $(5/1)^2 = 25$  possible SORs for the hypothetical example in Fig. 2.

It should be noted that most studies on SORs apply grids with a fixed number of rows and columns, which limits the number of  $w$  values that can be explored. For instance, with a fixed map of  $12 \times 12$  cells, it is only possible to investigate  $w = 2, 3, 4$  and  $6$  (because dividing 12 by other values does not yield integers). Because of this approach, the wavy nature of many SORs is easily obscured.

The effect of grid aggregation and grid origin on upscaling can be further explored by means of Appendix S1 in Supplementary Material, in which one can define any country and species distribution within a  $9 \times 9$  grid, and immediately see the corresponding SORs.

## Results from real observation data

We analysed the seriousness of the problem with the aid of the Dutch national data base known as FLORBASE (Van der Meijden *et al.*, 1996), which contains the observation records

collected in the period 1975–98 of all 1410 Dutch wild vascular plants. The data in FLORBASE are stored in a grid consisting of  $1 \times 1$  km cells. Figure 1 illustrates the FLORBASE distribution of two of those species. We used a map of the Netherlands with a resolution of 1 km to calculate  $O_w$ , which is the land-covered area of a species at the scale of  $w$  km. At the scale of  $1 \times 1$  km, the total land area of the Netherlands measures  $35,812 \text{ km}^2$ . As in Fig. 2, only the land-covered part of the grid squares was involved in the computations of  $O_w$ , hence this entity can never exceed  $35,812 \text{ km}^2$ . Each species was scaled up from  $w_0 = 1$  km on the base map to  $w = 50$  km, in steps of 1 km. For each value of  $w$ , we adjusted the number of rows and the number of columns of the national grid, just as we did in Fig. 2.

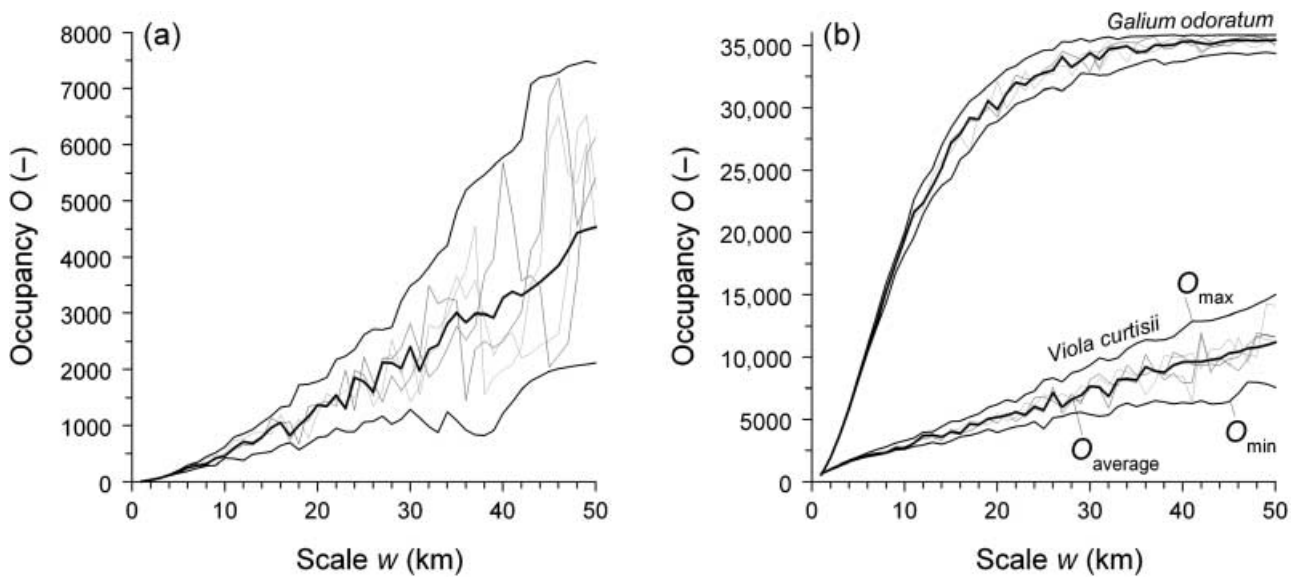
By way of example, four arbitrarily chosen SOR curves out of the 2500 are presented in Fig. 3 for three species, a rare one ( $O_{1 \text{ km}} = 9 \text{ km}^2$ ) and the two fairly common species from Fig. 1 ( $O_{1 \text{ km}} = 662$  and  $576 \text{ km}^2$ ). For these three species, we also drew the minimum ( $O_{\min}$ ), the maximum ( $O_{\max}$ ) and the average ( $O_{\text{average}}$ ) occupancies (bold lines) calculated at each scale  $w$  (at, for instance,  $w = 10$  km, these three entities are based on 100  $O_{10 \text{ km}}$  values).

Figure 3 shows that the rare species is more susceptible to grid origin than the two common ones; the relative bandwidth of possible SORs is larger in Fig. 3(a) than in Fig. 3(b) (note that the  $O$ -scales of (a) and (b) are different). This is no coincidence, as shown by the results in Fig. 4 where, for all species, the sensitivity  $R$  is plotted against the occupancy at the observation scale of  $w_0 = 1$  km, for  $w = 2, 5, 25$  and  $50$  km. Here,  $R$  is defined as

$$R = \frac{O_{\max} - O_{\min}}{O_{\text{average}}} \quad (1)$$

The shape of the data points in Fig. 4 can be explained by starting from a hypothetical species, occurring in every 1-km square of the country (i.e.  $O_{1 \text{ km}} = 35,812 \text{ km}^2$ ). It is self-evident that the  $O_w$  of this species always equals  $O_{1 \text{ km}}$  (and, thus,  $R = 0$ ), no matter what scale the  $w$  and no matter what position of the grid origin. If we proceed from this extremely common species to a progressively rarer species, the sheer probability increases that a merged square consisting of occupied 1-km squares splits into two to four different squares when we take another grid origin. Vice versa, occupied neighbouring squares might ‘accidentally’ merge into one bigger square at another grid origin.

From the description above, it becomes clear that susceptibility to grid origin not only relates to rarity but also to the extent to which species are clustered. To explore the relationship between  $R$  and clustering, we express the degree of clustering in the spatial autocorrelation coefficient  $r_a$ . For each species, we calculated this correlation based on the four adjacent squares of each 1-km square (i.e. ‘at a distance of 1 km’). Let us assume that the grid is parametrized by  $(i, j)$ , where  $i$  counts the cells in an east–west direction and  $j$  does so in a north–south direction. Let us assume that  $X(i, j)$  is any quantity defined on this grid with mean  $\bar{X}$  and standard deviation  $S$ , and  $N$  is the number of grid squares. The Pearson product–moment correlation in the north–south direction then becomes:  $r_{\text{NS}} = (1/(N \times S^2)) \sum_{i,j} (X(i,j) - \bar{X})(X(i,j+1) - \bar{X})$ .



**Figure 3** Examples of scale–occupancy relationships (SORs) for three species: (a) man orchid (*Aceras anthropophorum*), a rare one with  $O_{1\text{ km}^2} = 9\text{ km}^2$ ; (b) dune pansy (*Viola curtisii*) and woodruff (*Galium odoratum*), which have a comparable rarity (Fig. 1) but a very different degree of spatial clustering ( $r_a = 0.71$  vs.  $0.13$ ). For all three species, four arbitrarily chosen SOR curves out of the 2500 are presented (thin lines) as well as the minimum, average and maximum values of  $O$  (bold lines).

In a similar way, we can compute the correlation in the east–west direction,  $r_{EW}$ . Here we used the average of  $r_{NS}$  and  $r_{EW}$  as the overall spatial correlation between neighbouring cells:

$$r_a = \frac{1}{2N \times S^2} \sum_{i,j} [(X(i,j) - \bar{X})(X(i+1,j) - \bar{X}) + (X(i,j) - \bar{X})(X(i,j+1) - \bar{X})] \quad (2)$$

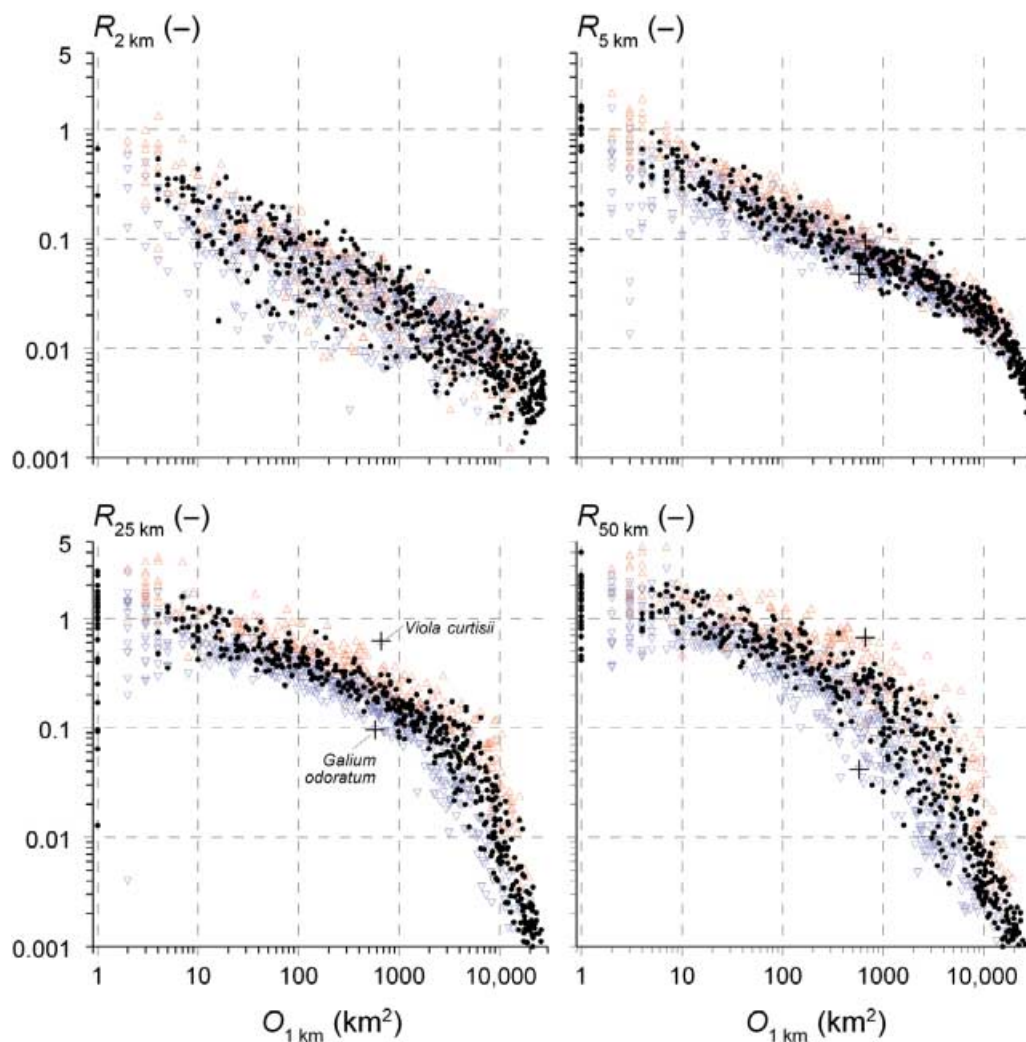
In our calculation,  $X$  could assume the value of 0 (absent) or 1 (present).

In Fig. 5,  $r_a$  is plotted against  $O_{1\text{ km}^2}$ . This figure shows there is a strong correlation between rarity ( $O_{1\text{ km}^2}$ ) and the extent to which species are clustered ( $r_a$ ): clustering increases with commonness. Because of the dependency between clustering and rarity, we need to judge  $r_a$  in relation to  $O_{1\text{ km}^2}$ . To do this, we distinguished three classes of relative clustering: ‘high’, ‘average’ and ‘low’ clustering, depicted in Fig. 4 by red, black and blue dots, respectively. Species that exceed the regression line ( $\hat{r}_a = 0.041 + 0.043 \ln O_{1\text{ km}^2}$ ) by 0.05 point ( $r_a > \hat{r}_a + 0.05$ ) were classified as ‘high’, species with  $r_a < \hat{r}_a - 0.05$  as ‘low’ and the rest of the species (those occurring near the regression line) as ‘average’ (Fig. 5).

From Fig. 4 we can conclude that, for a given rarity, it is the relatively clustered species that are more sensitive and that less aggregated species are less sensitive. By way of example, Fig. 3(b) shows the SORs of the two species in Fig. 1, *Viola curtisii* ( $O_{1\text{ km}^2} = 576\text{ km}^2$ ,  $r_a = 0.71$ ,  $R_{25\text{ km}^2} = 0.62$ ) and *Galium odoratum* ( $O_{1\text{ km}^2} = 662\text{ km}^2$ ,  $r_a = 0.13$ ,  $R_{25\text{ km}^2} = 0.10$ ). These species have similar rarity on the base scale but are very different in their sensitivity, as indicated by the black crosses in Fig. 4. Clustered species are usually confined to habitats with a distinct geographical area. In the case of *V. curtisii*, this area is the coastal dunes in the west of the country (Fig. 1a), the only place in the Netherlands with a suitable habitat for this species: dry, nutrient-poor and calcareous

grasslands. Widespread species are often generalists in environmental terms, and thus common. But this is not always the case, some species can occur in small but distinct sites all over the country, for instance in narrow strips along ditches or on stony walls. The resolution of 1 km is too coarse to detect such habitats. *Galium odoratum* is a typical species in moist and alkaline deciduous forests, but also in the old orchards of country houses and estates that exist all over the country (Fig. 1b).

A final factor determining sensitivity is the position of the species’ distribution relative to the border of the investigated area. This factor affects the very rare species in particular, and is responsible for the increased spread of the data points in Fig. 4 from  $O_{1\text{ km}^2} \approx 10\text{ km}^2$  down to  $O_{1\text{ km}^2} = 1\text{ km}^2$ . Take, for example, a very rare species, occurring in a single 1-km square. When this square occurs somewhere in the middle of the country, the species will be completely immune to grid origin:  $O_w = w^2\text{ km}^2$ , thus  $R = 0$ , as long as the country’s border does not intersect the occupied merged square. A few species behave like this at a scale of 2, 5 and 25 km, but because of the logarithmic  $y$ -axis they are not visible in Fig. 4. On the other hand, when such a rare species (with  $O_{1\text{ km}^2} = 1\text{ km}^2$ ) occurs precisely at the border of the country,  $O_w$  can theoretically vary between  $1\text{ km}^2$  (when an area of  $w^2 - 1\text{ km}^2$  of the merged cell lies abroad) and  $w^2\text{ km}^2$  (when the merged cell completely falls within the country). In that case,  $R$  takes the maximum value of  $w^2$  at scale  $w$ . In a similar way, we can argue that the sensitivity of rare species occurring in more than one 1-km square also has a theoretical maximum of  $R = w^2$  at scale  $w$ , as long as none of the occupied 1-km squares merge into one bigger square at scale  $w$ . In practice, the theoretical maximum is far from reached within our data set (Fig. 4). The maximum  $R$  observed for  $w = 2, 5, 25$  and  $50\text{ km}$  is 1.3, 2.2, 3.6



**Figure 4** Relationship between rarity (defined as the occupancy at the observation scale,  $O_{1\text{km}}$ ) and sensitivity  $R$  to the position of the grid origin (defined by equation 1) at a merging scale of, from top-left to bottom-right: square width  $w = 2, 5, 25$  and  $50$  km. Each dot represents one plant species. Colours and symbols depict different degrees of relative spatial clustering:  $\Delta$  = high,  $\bullet$  = average,  $\nabla$  = low.

and 4.6, respectively (where values of 4, 25, 625 and 2500 are theoretically possible).

### The effect on downscaling

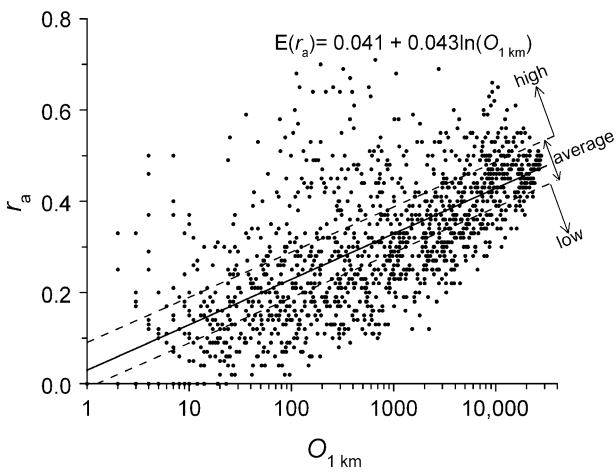
SORs have been widely used to describe the spatial distribution of species, to extrapolate fine-scale occupancies from those of coarse scales, and to determine species rarity (Gaston, 1994; Kunin, 1998; He & Gaston, 2000; Witte & Torfs, 2003; Wilson *et al.*, 2004; Fagan *et al.*, 2005; He & Condit, 2007). The dependence of SORs on map origin causes serious problems for the utility of the relationships. Here we demonstrate how this can influence the outcomes of occupancy extrapolations to finer scales. We limit our investigation to the 530 rarest species in FLORBASE, occurring in fewer than 200 1-km squares. This selection was made because the SORs of rare species are more susceptible to grid origin and are less affected by 'grid saturation'. Grid saturation (clearly visible in the SORs of *G. odoratum*; Fig. 3b) is the pheno-

menon whereby SORs level off at very coarse scales because the maximum area occupied is limited to the total investigated area.

There is some discussion in the literature about the best method to scale down, and researchers are still exploring various methods (He & Gaston, 2000; Kunin *et al.*, 2000; He & Condit, 2007). Since we merely want to expose the possible effects of grid origin, we will not become involved in this discussion. Instead, we apply the most commonly used method. We extrapolate down using a simple SOR power function (Hastings & Sugihara, 1993; Gaston, 1994; Kunin, 1998):

$$\hat{O}(w) = \alpha \times w^\beta \quad (3)$$

The parameters  $\alpha$  and  $\beta$  were found by log-log linear regression:  $\log \hat{O}(w) = \log \alpha + \beta \log w$ . For each of the 530 rare species, we first computed an artificial 'observation occupancy' at a base grid of  $w_0 = 2$  km. We then aggregated up this base distribution to produce maps at  $w = 4, 6$  and  $8$  km grids. Scaling up from 2 to



**Figure 5** Relationship between rarity  $O_{1\text{ km}}$  and spatial autocorrelation  $r_a$ . The regression line  $\hat{r}_a = 0.041 + 0.043 \ln O_{1\text{ km}}$  is used to classify species into three classes of relative clustering: ‘high’ (more than 0.05 higher than the regression line), ‘low’ (more than 0.05 lower than the line) and ‘average’ (near the regression line).

8 km resulted in  $(8/2)^2 = 16$  different SORs. Each SOR was fitted by model (3), resulting in 16 models. Regression was done by ordinary least squares (OLS) and resulted in very good fits in nearly all of the cases ( $r > 0.95$ ). By substituting the designated  $w = 1$  km into these models, we then arrived at 16 possible outcomes for the occupancy at the 1-km scale,  $\hat{O}_{1\text{ km}}$ . In the same way, we scaled down to the 1 km grid from a base grid with a resolution of  $w_0 = 5$  km that we scaled up to data points for  $w = 10, 15$  and  $20$  km.

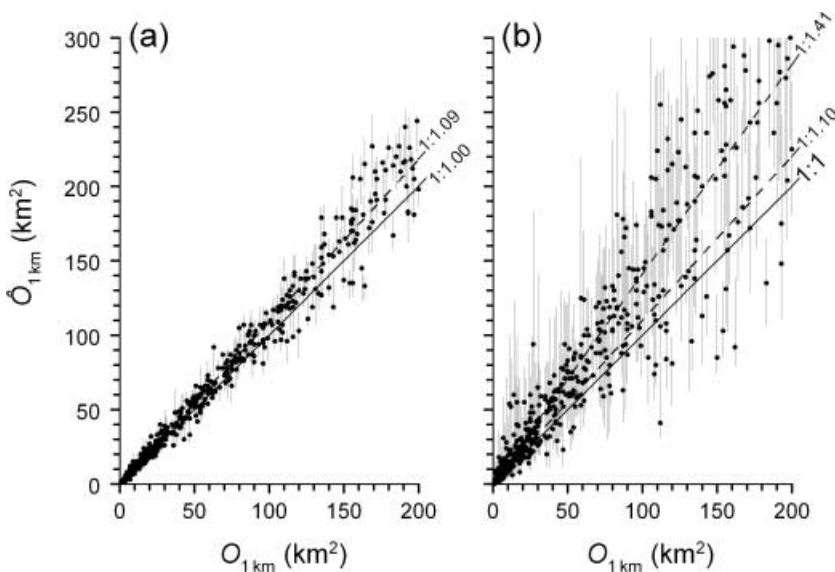
In Fig. 6, the predicted values are plotted against the observed ones. The predicted average, minimum and maximum occupancy at  $w = 1$  km are presented. Figure 6 shows how sensitive the predicted occupancy can be to grid origin, especially when scaling down from a very coarse scale, such as  $w_0 = 5$  km to  $w = 1$  km

(Fig. 6b). Figure 6 also shows the 1:1 line (solid lines), as well as the regression through both the average and the minimum expected occupancy ( $\hat{O}_{1\text{ km, average}}$  and  $\hat{O}_{1\text{ km, min}}$ ) (dashed lines). The predicted values based on  $\hat{O}_{1\text{ km, average}}$  (the dots in Fig. 6) appear to be systematically higher than the observed values. Using a base grid of  $w_0 = 2$  km, the 1-km predictions are 9% too high ( $\hat{O}_{1\text{ km, average}} = 1.09 O_{1\text{ km}}$ ), whereas a 5-km grid results in an overestimate of 41% ( $\hat{O}_{1\text{ km, average}} = 1.41 O_{1\text{ km}}$ ). Although frequently used, equation (3) often overestimates occupancy at finer scales, as has already been observed (Witte & Torfs, 2003; He & Condit, 2007). If we regress on the minimal occupancies,  $\hat{O}_{1\text{ km, min}}$ , the predictions correspond better, on average, to the observations: scaling down from a 2-km grid leads to predicted values that on average equal the observations ( $\hat{O}_{1\text{ km, min}} = 1.00 O_{1\text{ km}}$ ), whereas a 5-km grid leads to predictions that are 10% too high ( $\hat{O}_{1\text{ km, min}} = 1.10 O_{1\text{ km}}$ ).

### DISCUSSION

The very first decision in mapping species distribution is perhaps to determine where to position the origin of the grid map. By demonstrating how different grid origins result in different SORs we have shown that this decision can have a substantial effect on distribution aggregation. The sensitivity of SORs to grid origin appears to depend on three characteristics of the spatial distribution of a species: rarity, the degree of spatial clustering and the position of the distribution relative to the border of the investigated area.

Despite its importance, the problem of grid origin has not hitherto been investigated in the literature. If this problem is not properly recognized, there could be several consequences. First of all, SOR is a widespread tool for describing the spatial distribution of species. The dependence of the SOR on grid origin, as shown in this study, means that one will get very different SORs using different grid origins. There is no ‘correct’ or ‘incorrect’ SOR. All SORs constructed from different grid origins are



**Figure 6** Observed versus predicted occupancy on the 1-km grid ( $O_{1\text{ km}}$  vs.  $\hat{O}_{1\text{ km}}$ ). Predicted values based on a 2 km grid (a) as well as a 5 km grid (b). For each species, 16 predictions are available. The predicted average occupancy is indicated by a dot, the predicted minimum and maximum by a vertical line through this dot. The solid line is the 1:1 line, the dotted lines represent the regressions both on the average occupancy [(a)  $\hat{O}_{1\text{ km, average}} = 1.09 O_{1\text{ km}}$ ,  $r^2 = 0.98$ ; (b)  $\hat{O}_{1\text{ km, average}} = 1.41 O_{1\text{ km}}$ ,  $r^2 = 0.86$ ] and the minimum predicted occupancy [(a)  $\hat{O}_{1\text{ km, min}} = 1.00 O_{1\text{ km}}$ ,  $r^2 = 0.98$ ; (b)  $\hat{O}_{1\text{ km, min}} = 1.10 O_{1\text{ km}}$ ,  $r^2 = 0.83$ ]. Note:  $\hat{O}_{1\text{ km}} > 300 \text{ km}^2$  are not visible in (b).

technically correct, yet they only represent the spatial feature when scaled from particular origins. We lack criteria to judge which SOR best describes the overall feature of the spatial distribution of the species. This problem limits the utility of SORs for describing the spatial distribution of species.

Second, the primary use of SORs is to downscale occupancies from coarse scales and thus to determine species rarity. The lack of unique SORs makes the downscaled results highly variable and biased. Before new methods are available to deal with this problem, precautions must be taken to interpret any downscaled occupancies (at fine scales). As a consequence, the rarity status of species defined from such fine-scale occupancies should not be taken for granted. We advise evaluating and, if considered necessary, correcting rarity figures based on expert judgement or independent information, such as species lists from vegetation samples (relevés) (Van der Meijden *et al.*, 2000; De Iongh *et al.*, 2003).

Third, the different SORs from different grid origins will thwart scaling occupancies across spatial scales. However, one is only able to interpret and test mathematical relationships between scale and occupancy if the whole bandwidth of scaled occupancies is considered. This requires computing SORs based on all possible different grid origins. This, unfortunately, may demand considerable computation time. On a standard PC (1.73 GHz, 504 Mb RAM) it took our computer program, which was efficiently written in a direct language (FORTRAN-77), 5 min to scale all 1410 vascular species of the Dutch flora up from 1 to 4 km in steps of 1 km (16 SORs per species), but scaling up from 1 to 50 km (2500 SORs) required 50 h. Naturally, the development of faster processors may solve this 'problem' in the course of time.

The variation of SORs with grid origin raises a serious methodological challenge for scaling distribution. The scaling methods so far available only model smooth SOR curves (Gaston, 1994; Kunin, 1998; He & Gaston, 2000; Kunin *et al.*, 2000; Warren *et al.*, 2003; He & Reed, 2006). The SORs resulting from different grid origins are wavy; furthermore, the wavy variation is not a random effect resulting from sampling error. This fact will make modelling SORs solely on the basis of distribution data very difficult, if not impossible. The use of the power-law model in this study resulted in the systematic over-prediction of occupancies. As an *ad hoc* solution, we suggested a simple method to minimize this problem. This involved taking the lowest downscaled occupancy from all SORs corresponding to different grid origins. We have no explanation for this improvement and do not know whether downscaling this way in general is a good solution. However, whatever current downscaling method one chooses, the data points will deviate from the regression line due to the arbitrarily chosen position of the grid origin.

The big challenge now is to find a mathematical solution that not only has a theoretical basis but also takes the problem of grid origin into account. One such solution may be to incorporate ecological knowledge and more geographical information into the scaling method, with the aim of limiting the upscaled area of the species to the area where species can potentially occur. For instance, many rare species in the Netherlands are restricted to nature reserves, which are all known in GIS. Just as in this paper

we only considered the investigated area, and omitted foreign territory and sea, for these rare species we could also omit the area outside nature reserves. Furthermore, we think the variation of SORs will decrease and the reliability of the downscaled occupancies will increase if we only consider the area of potential sites within the observed grid square at the base scale. In addition, geographical information about the habitat of the species could be used. For instance, we know that the distribution of *V. curtisii* is restricted to calcareous dunes (Fig. 1a), so we should avoid extrapolating the occupied area of this species outside these dunes when scaling up. The ecological requirements of all plant species in the Netherlands are well known in terms of the habitat factors 'salinity', 'moisture regime', 'nutrient availability' and 'acidity' (Runhaar *et al.*, 2004; Witte *et al.*, 2007). Moreover, based on a soil map and a groundwater depth map, these habitat factors are mapped nationwide on a scale of 1:50,000 (Klijn *et al.*, 1996). Therefore, for the Netherlands at least, it would be possible to limit the upscaled area of species to the habitats where they can potentially occur, and thus minimize the variation of SORs and improve the reliability of downscaled occupancies.

In this paper we have shown that the choice of the grid origin can substantially affect both upscaling and downscaling of grid information. This may not be a problem as long as one is interested in the general mathematical shape of the relationship between scale and occupancy, but for testing and interpreting SORs we advise the consideration of all possible grid origins. We believe the latter may even be a prerequisite if one wants to use occupancy figures to compile a Red List, to monitor species, or to quantify species conservation values. Improvement of scaling techniques will not only be a matter of finding a better mathematical relationship between scale and occupancy, but presumably also will require extra geographical information, especially about species habitats.

## ACKNOWLEDGEMENTS

This study was preformed in the framework of the joint research programme of the Dutch Water Utility sector. The FLORON Foundation provided the FLORBASE distribution data of plant species in the Netherlands. We thank Professor Diniz-Filho and three anonymous referees for their valuable comments on the manuscript.

## REFERENCES

- Araújo, M.B., Thuiller, W., Williams, P.H. & Reginster, I. (2005) Downscaling European species atlas distributions to a finer resolution: implications for conservation planning. *Global Ecology and Biogeography*, **14**, 17–30.
- De Iongh, H.H., Bánki, O.S., Bergmans, W. & Van der Werf, M.J. (eds) (2003) *The harmonization of Red Lists for threatened species in Europe*. Commission for International Nature Protection, Leiden, The Netherlands.
- Dony, J.G. & Denholm, I. (1985) Some quantitative methods of assessing the conservation value of ecologically similar sites. *Journal of Applied Ecology*, **22**, 229–238.



- Fagan, W.F., Aumann, C., Kennedy, C.M. & Unmack, P.J. (2005) Rarity, fragmentation, and the scale dependence of extinction risk in desert fishes. *Ecology*, **86**, 34–41.
- Gaston, K.J. (1994) *Rarity*. Chapman & Hall, London.
- Hartley, S., Kunin, W.E., Lennon, J.J. & Pockock, M.J. (2004) Coherence and discontinuity in the scaling of species' distribution patterns. *Proceedings of the Royal Society B: Biological Sciences*, **271**, 81–88.
- Hastings, H.M. & Sugihara, G. (1993) *Fractals; a user's guide for the natural sciences*. Oxford University Press, Oxford.
- He, F. & Condit, R. (2007) The distribution of species: occupancy, scale, and rarity. *Scaling biodiversity* (ed. by D. Storch, P.L. Marquet and J.H. Brown), pp. 32–50. Cambridge University Press, Cambridge.
- He, F. & Gaston, K.J. (2000) Estimating species abundance from occurrence. *The American Naturalist*, **156**, 553–559.
- He, F. & Reed, W. (2006) Downscaling abundance from the distribution of species: occupancy theory and applications. *Scaling and uncertainty analysis in ecology: methods and applications* (ed. by J. Wu, B. Jones, H. Li and O. Loucks), pp. 89–108. Springer, Dordrecht, the Netherlands.
- Klijn, F., Groen, C.L.G. & Witte, J.P.M. (1996) Ecoseries for potential site mapping, an example from the Netherlands. *Landscape and Urban Planning*, **35**, 53–70.
- Kunin, W.E. (1998) Extrapolating species abundance across spatial scales. *Science*, **281**, 1513–1515.
- Kunin, W.E., Hartley, S. & Lennon, J.J. (2000) Scaling down: on the challenge of estimating abundance from occurrence patterns. *The American Naturalist*, **156**, 560–566.
- Margules, C. & Usher, M.B. (1981) Criteria used in assessing wildlife conservation potential: a review. *Biological Conservation*, **21**, 79–109.
- Runhaar, J., Van Landuyt, W., Groen, C.L.G., Weeda, E.J. & Verloove, F. (2004) Herziening van de indeling in ecologische soortengroepen voor Nederland en Vlaanderen. *Gorteria*, **30**, 12–26.
- Tosh, C.A., Reyers, B. & Jaarsveld, A.S. (2004) Estimating the abundances of large herbivores in the Kruger National Park using presence–absence data. *Animal Conservation*, **7**, 55–61.
- Van der Meijden, R., Groen, C.L.G., Vermeulen, J.J., Peterbroers, T., Van 't Zelfde, M. & Witte, J.P.M. (1996) *Eindrapport FLORBASE-1*. Rijksherbarium, Leiden, the Netherlands.
- Van der Meijden, R., Odé, B., Groen, C.L.G. & Witte, J.P.M. (2000) Bedreigde en kwetsbare vaatplanten in Nederland. *Gorteria*, **26**, 85–208.
- Warren, M., McGeoch, M.A. & Chown, S.L. (2003) Predicting abundance from occupancy: a test for an aggregated assemblage. *Journal of Animal Ecology*, **72**, 468–477.
- Wheeler, B.D. (1988) Species richness, species rarity and conservation evaluation of rich-fen vegetation in lowland England and Wales. *Journal of Applied Ecology*, **25**, 331–353.
- Wilson, R.J., Thomas, C.D., Fox, R., Roy, D.B. & Kunin, W.E. (2004) Spatial patterns in species distributions reveal biodiversity change. *Nature*, **432**, 393–396.
- Witte, J.P.M. (1998) *National water management and the value of nature*. PhD thesis, Agricultural University of Wageningen, the Netherlands.
- Witte, J.P.M. & Torfs, P.J.J.F. (2003) Scale dependency and fractal dimension of rarity. *Ecography*, **26**, 60–68.
- Witte, J.P.M., Wójcik, R., Torfs, P.J.J.F., De Haan, M.W.H. & Hennekens, S. (2007) Bayesian classification of vegetation types with Gaussian mixture density fitting to indicator values. *Journal of Vegetation Science*, **18**, 605–612.

## SUPPLEMENTARY MATERIAL

The following supplementary material is available for this article:

**Appendix S1** Effect of grid aggregation and grid origin on upscaling.

This material is available as part of the online article from: <http://www.blackwell-synergy.com/doi/abs/10.1111/j.1466-8238.2008.00387.x>

(This link will take you to the article abstract).

Please note: Blackwell Publishing is not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

## BIOSKETCHES

**Jan-Philip Witte** specializes in the relationship between groundwater and vegetation biodiversity, with an emphasis on modelling the ecological effects of water management.

**Fangliang He** is interested in community ecology, species diversity and conservation, ecological methodologies and modelling, and spatial statistics.

**Cees Groen** is a botanist, specializing in setting up and analysing data banks for the benefit of nature conservation.

Editor: José Alexandre F. Diniz-Filho