As in the second set of analyses, a MSO was computed from a partial RDA that controlled for the seven pre-selected MEM eigenfunctions. The results of this fourth set of analyses were almost identical to those shown in Fig. 14.13b: the variogram of the explained plus residual variation was now entirely within the confidence envelope of the variogram of $\mathbf{Y}_{\text{res}|\mathbf{W}}$ and the separate variograms of explained and residual variation were flat, showing that the MEM eigenfunctions successfully controlled for the spatial correlation of the detrended mite data that was not well explained by the detrended environmental variables.

The results of this analysis could have differed from Fig. 14.13b if the mite data had been structured by a broad-scale spatial trend running in a different direction from that of the environmental data. Had that been the case, the trend in the undetrended response data $\mathbf{Y}$ shown in Fig. 14.13b would not be modelled by the undetrended environmental variables $\mathbf{X}$. However, the fourth set of analyses described in the present paragraph, with $\mathbf{Y}$ and $\mathbf{X}$ having been both detrended, would not have been impaired by these differing trends.

The R code to run the MSO analyses reported in the present ecological application is found in Section 7.5.3 of Borcard *et al*. (2011).

14.5 Other eigenfunction-based methods of spatial analysis

This section describes additional statistical methods based on spatial eigenfunctions that were not covered in the previous sections.

1 — Space-time interaction

A commonly used approach to test hypotheses about natural or man-made environmental changes, including climate change, is to sample portions of ecosystems repeatedly over time. This type of sampling is usually done without replication of sites; in this way, the sampling effort can be spent on maximizing the expanse of space covered by the study. If the sampling sites and times are represented by dummy variables or Helmert contrasts, as in paragraphs 3 and 4 of Subsection 11.1.10, one can use canonical analysis to study the effect of the sites on species composition while controlling for the effect of time, and vice versa. An important limit of this approach is that the interaction between space and time cannot be estimated for lack of replicates. Assessing that interaction is, however, of great interest in such studies because a significant interaction would indicate that the spatial structure of the univariate or multivariate response data has changed through time, and conversely that the temporal variations differed significantly among the sites, thus indicating, for example, the signature of climate change on ecosystems.

Legendre *et al*. (2010) described a statistical method to *analyse the interaction* between the space (S) and time (T) factors *in space-time studies without replication*; the acronym of the method is STI (for *space-time interaction*). The method can be applied to multivariate response data, e.g. ecological community composition, through partial RDA. The method consists in representing the space and/or time factors by spatial and/or temporal eigenfunctions (MEM, Sections 14.1 and 14.2, or AEM,

STI

Section 14.3). It is not necessary to represent both space and time by eigenfunctions: for example, if there are many sites and only a few sampling times, e.g. 2 or 3, spatial relationships may be coded using spatial eigenfunctions and temporal relationships using dummy variables or Helmert contrasts. Coding the space and/or time factors by spatial and/or temporal eigenfunctions requires fewer coding variables than dummy variables or Helmert contrasts. The interaction can be represented by variables obtained by computing the Hadamard product of each eigenfunction that codes for space with each eigenfunction that codes for time. Enough degrees of freedom are saved to correctly estimate the residual fraction of variation and test the significance of the interaction term.

The above paper gives details about the computation method. The R package STI is available to carry out the calculations (Section 14.7). The paper also contains two applications to real species assemblage data: an analysis of Trichoptera (insects, 56 species) emerging from a stream and captured in 22 emergence traps during 100 days, grouped into 10 consecutive 10-day periods, and a study of four surveys conducted between 1982 and 1995 in the Barro Colorado Island permanent forest plot (315 species of trees). Another application is found in Laliberté *et al.* (2009) where tree seedling abundances at 40 sites along a transect in a temperate forest understory, monitored during a 9-year period, were analysed for space-time interaction. The analysis of spatio-temporal data is also discussed in Cressie & Wikle (2011).

2 — *Multiscale codependence analysis*

A causal relationship between an explanatory (**x**) and a response variable (**y**) across space implies that the two variables are correlated. When the correlation between **x** and **y** is not significant, the causal hypothesis must be abandoned. Conversely, a significant correlation can be interpreted as support of the causal hypothesis that **x** may have an effect on **y**. Given the multiscale nature of ecological processes, one may wonder at which scales **x** is an important predictor of **y**. The same question can be asked about pairs of variables forming a bivariate time series; for simplicity, the presentation here will focus on space.

MCA   Guénard *et al.* (2010) developed *multiscale codependence analysis* (MCA) to address the above question and test the significance of the correlations between two variables at different spatial scales. The method is based on spatial eigenfunctions, MEM or AEM, which correspond to different and identifiable spatial scales: indeed, a Moran's *I* statistic (eq. 13.1) can be computed for each eigenfunction. If the sampling is regular along a transect, eq. 14.1 can be used to determine the wavelegths of the k eigenfunctions, which are assembled in a matrix called **W**, of size $n \times k$. Correlation coefficients are computed between **y** and each of the k eigenfunctions, and written in a vector $\mathbf{r_{yW}}$ of length k. Similarly, correlation coefficients are computed between **x** and each of the k eigenfunctions, and written in a vector $\mathbf{r_{xW}}$. The Hadamard product of the two vectors, $\mathbf{r_{yW}}$ and $\mathbf{r_{xW}}$, is the vector of *codependence coefficients*, which reflect the strength of the **x**-**y** correlations at the different scales represented by the eigenfunctions in matrix **W**. Each codependence coefficient can be tested for

significance using a τ (tau) statistic obtained by computing the product of the *t*-statistics associated with the two correlation coefficients. The testing procedure is described in the paper. An R package is available for the calculations (Section 14.7).

In the above paper, the method was applied to model the river habitat of juvenile Atlantic salmon (parr). MCA showed that variables describing substrate composition of the river bed were the most influential predictors of parr abundance at the 0.4 – 4.1 km scales whereas mean channel depth was more influential at the 200 – 300 m scales. This example shows that when properly assessed, the multiscale structuring observed in nature may be used to refine our understanding of natural processes.

3 — *Estimating and controlling for spatial structure in modelling*

The examples and applications reported in Sections 14.1 to 14.3 show that spatial eigenfunctions can efficiently model all kinds of spatial structures in data. Can they be used to find a solution to the problem described in Subsection 1.1.2, that spatial correlation inflates the level of type I error in tests of species-environment relationships in regression and canonical analysis?

A species-environment relationship after controlling for spatial structure can be represented by fraction [a] in a Venn diagram (e.g. Figs. 10.10) showing the partitioning of the variation of response data, univariate $\mathbf{y}$ or multivariate $\mathbf{Y}$, with respect to environmental (left circle) and spatial variables (right circle). A real example is shown in Fig. 14.7. Using numerical simulations, Peres-Neto & Legendre (2010) showed that spatial eigenfunctions provided an effective answer to the problem. Firstly, one must determine if the spatial component of $\mathbf{y}$ or $\mathbf{Y}$ is significant. This can be done by regression of $\mathbf{y}$, or canonical analysis of $\mathbf{Y}$, against all MEM spatial predictors, or by univariate (for $\mathbf{y}$) or multivariate (for $\mathbf{Y}$) variogram analysis. Secondly, if the spatial component is significant, one can select a subset of spatial predictors, and use the environmental ($\mathbf{X}$) and the selected spatial predictors (covariables $\mathbf{W}$) in a partial regression (for $\mathbf{y}$, Subsection 10.3.5) or partial canonical analysis (for $\mathbf{Y}$, Subsection 11.1.6).

For the analysis of community composition data, the authors found that a species-by-species forward selection procedure, described in their paper, was to be preferred to a global, community-based selection. In this method, eigenfunctions are selected for each species independently, and the union of the selected sets is used as the matrix of MEM covariables in canonical analysis. This provides an effective method of control for type I error in the assessment of species-environment relationships. The paper also showed that polynomial regressors (Subsection 13.2.1) did not produce tests of significance with correct levels of type I error.

The Peres-Neto & Legendre (2010) paper provides theoretical support to the effect observed in Ecological application 14.4, that MEM used as covariables in canonical analysis effectively controlled for the spatial correlation observed in the species-environment relationship in the first part of the analysis of the mite data.