Ecological Monographs, 74(4), 2004, pp. 685–701

A NEW TECHNIQUE FOR MAXIMUM-LIKELIHOOD CANONICAL GAUSSIAN ORDINATION

THOMAS W. YEE[1]

Department of Statistics, University of Auckland, Private Bag 92019, Auckland, New Zealand, and
Department of Statistics and Applied Probability, 6 Science Drive 2, National University of Singapore, Singapore 117546

Abstract. Canonical correspondence analysis (CCA) is probably the most popular ordination method in community ecology. However, it is only a heuristic approximation to maximum-likelihood estimated canonical Gaussian ordination (CGO), which is the ''ideal'' method. When proposed in the mid-1980s, CCA held two advantages over CGO: it was computationally cheaper, and its algorithm was not complex. However, an exponential increase in computing speed over the last two decades has meant that computation cost is no longer such a compelling advantage. The computational complexity of CGO has always been its major difficulty, even though it is statistically more sound and simpler to understand than CCA. For these reasons, no general computational framework or software has appeared until now.

This article proposes a new class of statistical regression models called quadratic reduced-rank vector generalized linear models (QRR-VGLMs) for maximum-likelihood estimated CGO. This is achieved by extending a recently developed class of statistical models called RR-VGLMs. The extension is named QRR-VGLMs because of the addition of a quadratic form to each linear predictor, with the consequence that bell-shaped responses can be modeled as functions of latent environmental variables or gradients. QRR-VGLMs have several major positive features; for example, their framework is unifying and broad, so that canonical Gaussian ordination can potentially be performed on a wide range of data types. The two most important special cases of CGO in ecology, multispecies presence/absence and Poisson abundance data, are considered in this article. The methodology is illustrated with a real data set using a software implementation written by the author in the S statistical language. The code, called the VGAM package in R, is object-oriented and free, and it allows QRR-VGLMs to be fitted to moderate-sized data sets conforming reasonably closely to the Gaussian model.

Key words: canonical correspondence analysis; canonical Gaussian ordination; direct gradient analysis; iteratively reweighted least squares; latent variables; maximum-likelihood estimation; ordination; quadratic reduced-rank vector generalized linear models; regression: logistic, Poisson, and reduced-rank; tolerance; unimodal response curve; vector generalized linear models.

INTRODUCTION

Canonical correspondence analysis (CCA) is probably the most popular dimension-reducing/ordination technique among ecologists today. Proposed by ter Braak (1986), CCA is a technique for multivariate direct gradient analysis (ter Braak and Prentice 1988) whereby species abundances **y** are regressed against optimal linear combinations of environmental variables **x** in order to ''explain'' the data as much as possible. Many biologists fit CCA models using the package CANOCO (ter Braak and Šmilauer 1998), which runs under the Windows operating system and has a user-friendly graphical user interface that has contributed to its wide popularity. Unfortunately, practitioners of CCA who are unfamiliar with the literature (e.g., ter Braak 1986, ter Braak and Prentice 1988) are unaware that it is only an approximate solution to canonical Gaussian ordination (CGO) fitted by maximum likelihood estimation. That is, Gaussian ordination with linear external constraints estimated by maximizing the likelihood function can be considered the ''ideal'' methodology because it is statistically sound, whereas CCA is a heuristic approximation to this. For the last two decades, this important gap between ideal and heuristics has not been bridged.

Like all statistical methods, CCA possesses both strengths and weaknesses. One weakness is that CCA makes assumptions relating to optima, tolerances, and abundances that cannot simultaneously hold (ter Braak 1987*b*, ter Braak and Prentice 1988). Another is that it assumes linearity of the **x** and the chi-square-transformed **y** (although Makarenkov and Legendre [2002] is an attempt to offer some flexibility over this). Another deficiency is that CCA models relative abundance rather than absolute abundance. The strengths of CCA include its low computational expense and robustness to the four underlying assumptions (e.g., ter Braak 1987*a*, Palmer 1993). Johnson and Altman (1999) showed that CCA provides a reasonable approximation

Manuscript received 3 February 2003; revised 18 October 2003; accepted 9 November 2003; final version received 15 January 2004. Corresponding Editor: N. C. Kenkel.

[1] E-mail: t.yee@auckland.ac.nz

to CGO when species' tolerances do not differ too much (within a factor of three), are independent of their optimum environment, and species' maxima are within a factor of 10 from each other.

The reasons put forth at the time for using CCA instead of CGO were mainly because of the latter's computational complexity and expense. However, since the mid-1980s when CCA was proposed, computing power has increased exponentially so that computational expense can no longer be considered such a major advantage over CGO. Despite the simplicity of canonical Gaussian ordination (e.g., Eqs. 2–4), a major reason why CGO has not been used is that no software has been written that is easy to use and accessible. Furthermore, no algorithm has been developed to handle responses coming from, more generally, the exponential family (such as GLMs [Nelder and Wedderburn 1972], which handle normal, binomial, and Poisson responses).

To this end, the purpose of this paper is to propose a new class of statistical regression models called quadratic reduced-rank vector generalized linear models (QRR-VGLMs), which allow the maximum likelihood estimation of CGO to be performed on a wide range of statistical models and data types. QRR-VGLMs extend a class of models called RR-VGLMs by allowing for the inclusion of a quadratic form to each linear predictor, so that a bell-shaped response can be modeled for each species. The class of RR-VGLMs was proposed by Yee and Hastie (2003), and they enable the potential benefits of reduced-rank regression to be conveyed to a very wide class of models. RR-VGLMs themselves are a variant of the class of VGLMs, which were introduced by Yee and Wild (1996). VGLMs are a very large class that encompasses a wide range of multivariate response types and models including univariate and multivariate distributions, categorical data analysis, time series, survival analysis, generalized estimating equations, correlated binary data, bioassay data, and nonlinear least-squares problems. VGLMs are expounded with a biological focus in Yee and Mackenzie (2002). QRR-VGLMs inherit many properties of VGLMs; therefore they can be used to fit CGO to a wide range of data types. With the unifying theoretical framework provided by QRR-VGLMs, it is also simpler to write software implementing the method. This article describes a software implementation called VGAM and illustrates its use on some real data.

Ordinary generalized linear models (GLMs; see McCullagh and Nelder 1989) are a special case of the VGLM class. In an article comparing GLMs with CCA, Guisan et al. (1999) highlight one strength that CCA has over GLMs: for community delineation and interpretation CCA is clearly more valuable because it deals with the co-occurrence of species whereas GLMs are a distinct regression. To overcome this limitation of GLMs, they mention the idea of designing a system of simultaneous regression equations (simultaneous GLMs) to integrate species co-occurrence/exclusion information. QRR-VGLMs are a direct answer to this need: Eq. 12 links M species to a common set of latent variables $\mathbf{v}$. Actually, CCA is also a separate regression, but it then projects the fitted values onto an ordination diagram (typically in two dimensions) and gives an impression of community analysis.

Notation and outline of article

In this article, the data consists of matrices $\mathbf{Y}$ and $\mathbf{X}$, which are $n \times M$ and $n \times p$, respectively. There are M species, n sites, and p environmental variables (the first environmental variable is an intercept term.) We write $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n)^T$ as the response matrix (e.g., containing counts or species abundances), and $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)^T$ is the matrix of environmental variables at each site. Elementwise, $\mathbf{Y} = [(y_{ij})]$ and $\mathbf{X} = [(x_{ik})]$, and i will index sites, j will index species, and k will index environmental variables. We write $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})^T$ and $\mathbf{x} = (x_1, \ldots, x_p)^T$, where $x_1 = 1$ is an intercept term, i.e., the first column of $\mathbf{X}$ is a vector of ones, $\mathbf{1}$. All logarithms are to base e. We let $\|\mathbf{x} - \mathbf{y}\|$ be the Euclidean distance between two vectors $\mathbf{x}$ and $\mathbf{y}$, i.e., $\sqrt{(\mathbf{x} - \mathbf{y})^T(\mathbf{x} - \mathbf{y})}$. Intercept terms need special care, therefore we use the vector $\mathbf{x}_{(-1)i}$ to denote the vector $\mathbf{x}_i$ with the first element removed, and $\mathbf{B}_{(-1)}$ denotes a matrix $\mathbf{B}$ with the first row deleted.

While QRR-VGLMs belong to a much larger family of statistical models, we only present the simplest types of QRR-VGLMs (those based on the GLM class) in this article, for brevity. Recall for GLMs that

$$g(\mu_i) = \eta_i = \mathbf{x}_i^T\boldsymbol{\beta} \qquad \text{Var}(Y) = \phi V(\mu) \tag{1}$$

where $\mu = E(Y)$, g is the known link function, η is the linear predictor, ϕ is the dispersion parameter which may be known or unknown, and V is the known variance function.

Canonical Gaussian ordination

One of the simplest and important example of CGO is Poisson data with M species. The rank-1 model (i.e., one latent environmental variable $\nu = \mathbf{c}^T\mathbf{x}$) for species j is the Poisson regression

$$\log \mu_j(\nu) = \eta_j(\nu) \qquad j = 1, \ldots, M \tag{2}$$

where $\mu_j = E(Y_j)$ is the mean abundance, and

$$\eta_j(\nu) = \beta_{(j)1} + \beta_{(j)2}\nu + \beta_{(j)3}\nu^2 \tag{3}$$

$$= \alpha_j - \frac{1}{2}\left(\frac{\nu - u_j}{t_j}\right)^2 \tag{4}$$

is the linear predictor. Eqs. 2–4 are GLMs in the latent variable ν.

The justification for the squared term ν^2 is provided by a large literature dealing with unimodal response curves to environmental variables. This spans much of ecological theory, e.g., the ideas of Shelford, Igoshina, Ellenberg, Hesse, Gause, etc. (see, e.g., Gauch and

Whittaker 1972*b*, Gauch 1982, Austin 1985, ter Braak and Prentice 1988). The ν^2 accommodates symmetric bell-shaped curves on the original **Y** scale.

Eq. 4 is similar to that of the normal or Gaussian distribution, hence the adjective ''Gaussian'' is used. In the second formulation (Eq. 4), u_j is often called species j's optimum and t_j (>0) its tolerance. This representation is often used instead of Eq. 3 because of the ecological interpretations that can be ascribed to the parameters. The tolerance measures how wide the response curve is, i.e., how much deviation the species can tolerate from its optimal environment so that t_j is large for stenoecous species and small for eurycous ones. It is a measure of niche width. The parameter u_j is the value of the gradient (in the ecological rather than the mathematical sense) in which $\mu_j(\nu)$ is a maximum; therefore it is the optimal environment for species j. Since $\mu_j(u_j) = e^{\alpha_j}$, the parameter α_j is directly related to the prevalence/abundance of the species at its optimum. We call $\mu_j(u_j)$ the maximum of species j.

The coefficient $\beta_{(j)3}$ in Eq. 3 is important because a negative value implies the response curve is unimodal about the optimum u_j. If $\beta_{(j)3} = 0$, then the curve is sigmoid, and if $\beta_{(j)3} > 0$, then the curve is ''u'' shaped. Interesting submodels of the above based on ecological theories include testing H_{01}: $\eta_1 = \ldots = \eta_M$ (all species' response surfaces are equal), H_{02}: $t_1 = \cdots = t_M$ (all species' tolerances are equal), and H_{03}: $u_1 = \cdots = u_M$ (all species' optima are equal). All these are considered more generally in *Simplification hypotheses*.

Another simple example of CGO is with the presence/absence data of M species at a site. Then Eq. 2 becomes

$$\text{logit } \mu_j(\nu) = \eta_j(\nu) \qquad j = 1, \ldots, M \tag{5}$$

the so-called ''Gaussian logit model'' for species j (ter Braak and Looman 1986) applied to the hypothetical environmental gradient ν. Here, the logit link function is used, but other link functions include the probit and complementary log-log links; VGAM easily allows the user to choose among these.

Vector Generalized Linear Models

VGLMs, which are described in detail in Yee and Hastie (2003), involve a set of M linear predictors η_j:

$$\eta_j(\mathbf{x}) = \beta_j^T\mathbf{x} = \beta_{(j)1}x_1 + \cdots + \beta_{(j)p}x_p \qquad j = 1, \ldots, M. \tag{6}$$

The VGLM class encompasses all known GLMs (Eq. 1), and are deliberately general so that it includes as many distributions and models as possible. We attempt to be limited only by the assumption that the regression coefficients enter through a set of linear predictors. Eq. 6 can be written as follows:

$$\boldsymbol{\eta} = \boldsymbol{\eta}(\mathbf{x}_i) = \begin{bmatrix} \eta_1(\mathbf{x}_i) \\ \vdots \\ \eta_M(\mathbf{x}_i) \end{bmatrix} = \boldsymbol{\eta}_0 + [\mathbf{B}_{(-1)}]^\top\mathbf{x}_{(-1)i} = \boldsymbol{\eta}_0 + \begin{bmatrix} \boldsymbol{\beta}_{(-1)1}^\top\mathbf{x}_{(-1)i} \\ \vdots \\ \boldsymbol{\beta}_{(-1)M}^\top\mathbf{x}_{(-1)i} \end{bmatrix} \tag{7}$$

where $\boldsymbol{\eta}_0 = (\beta_{(1)1}, \ldots, \beta_{(M)1})^\top$ is a vector of intercepts, and $\mathbf{B} = (\boldsymbol{\beta}_1\boldsymbol{\beta}_2 \cdots \boldsymbol{\beta}_M)$ is $p \times M$.

Most VGLMs have a log-likelihood function $\ell = \sum_{i=1}^n \ell_i(\eta_1, \ldots, \eta_M)$, or at least a log-quasilikelihood function (McCullagh and Nelder 1989). Maximum likelihood estimation is performed by iteratively reweighted least squares (IRLS; see, e.g., Green 1984) using either Newton-Raphson or Fisher scoring.

With multispecies Poisson data, Poisson regression fits within the VGLM framework (Eq. 6) because

$$\boldsymbol{\eta}_i = \log \boldsymbol{\mu}_i = (\log \mu_{i1}, \ldots, \log \mu_{iM})^\top. \tag{8}$$

Here, Y_{ij} has a Poisson distribution with mean $\mu_{ij} = E(Y_{ij})$, and is the mean of the ith row of **Y**. Thus $\log \mu_i = \eta_i = \mathbf{B}^\top\mathbf{x}_i$ allows M simultaneous Poisson regressions on the environmental variables $\mathbf{x}_i$—one for each species.

In a similar manner, multispecies binary data fits within the VGLM framework by letting $Y_{ij} = 1$ or 0 for presence and absence, respectively, of species j at site i. Then instead of Eq. 8, one simply has

$$\boldsymbol{\eta}_i = \text{logit } \boldsymbol{\mu}_i = \{\log[\mu_{i1}/(1 - \mu_{i1})], \ldots, \log[\mu_{iM}/(1 - \mu_{iM})]\}^\top. \tag{9}$$

The two important topics of constraints on the functions and the scaling parameter for VGLMs are briefly summarized in Appendix B.

Reduced-rank VGLMs

Motivation for the class of reduced-rank VGLMs (RR-VGLMs) can be obtained from the observation that, for various VGLM models, the matrix of regression coefficients $\mathbf{B}_{(-1)}$ in Eq. 7 may be very large for the data at hand. For example, for $M \geq 5$ species and $p \geq 9$ environmental variables, there are at least 40 regression coefficients. Consequently, it is clearly advantageous to reduce the number of parameters. Reduced-rank regression can achieve this by constraining $\mathbf{B}_{(-1)}$ to be of a lower rank, i.e., $\mathbf{B}_{(-1)}$ can be replaced by the approximation

$$\mathbf{B}_{(-1)} = \mathbf{C}\mathbf{A}^\top \tag{10}$$

where $\mathbf{C} = (\mathbf{c}_{(1)}\mathbf{c}_{(2)} \cdots \mathbf{c}_{(R)})$ is $(p - 1) \times R$, $\mathbf{A} = (\mathbf{a}_{(1)}\mathbf{a}_{(2)} \cdots \mathbf{a}_{(R)}) = (\mathbf{a}_1 \cdots \mathbf{a}_M)^\top$ is $M \times R$, and R ($\leq \min(M, p - 1)$) is the rank (**A** and **C** are both of full-column rank). Eq. 10 shows that $\mathbf{B}_{(-1)}$ is approximated by the (outer) product of two ''thin'' R-column matrices, and when applied to the VGLM class, this gives rise to the

class of RR-VGLMs. A more parsimonious model results if $R \ll p - 1$ because the resulting number of parameters is often much less than the full model (the difference is $(M - R)(p - 1 - R)$).

The reduced-rank idea is also important for a number of reasons:

1) The concept is readily interpretable because $[\boldsymbol{c}_{(1)}^{\mathsf{T}}\boldsymbol{x}_{(-1)}, \ldots, \boldsymbol{c}_{(R)}^{\mathsf{T}}\boldsymbol{x}_{(-1)}]^{\mathsf{T}}$ $[= \mathbf{C}^{\mathsf{T}}\boldsymbol{x}_{(-1)} = \boldsymbol{\nu}$, say] can be interpreted as a vector of latent variables or hypothetical environmental variables—linear combinations of the original predictor variables that give more explanatory power. They often can be thought of as a proxy for some underlying variable behind the mechanism of the process generating the data. In plant ecology, the idea is an important one, for example, in direct and indirect gradient analysis (see, e.g., ter Braak and Prentice 1988). The role of $\mathbf{C}$ can be considered as choosing the ''best'' regressors from a linear combination of the original regressors, and $\mathbf{A}$ as regression coefficients of these new regressors.

2) The reduced-rank approximation provides a vehicle for a low-dimensional view of the data via an ordination diagram.

3) It allows for flexible generalizations based on smoothing.

4) Reduced-rank regression/redundancy analysis (RDA; ter Braak and Prentice 1988, ter Braak and Looman 1994) for quantitative linear data can be replaced by RR-VGLMs for binary data and counts. Features such as biplots are available for RR-VGLMs.

Embedding the reduced-rank constraint within a large class of models such as VGLMs has the additional advantage that many models become just special cases, and software is more easily written and used.

Partial RR-VGLMs

More flexibility is gained by partitioning $\mathbf{x}$ into $(\mathbf{x}_1^{\mathsf{T}}, \mathbf{x}_2^{\mathsf{T}})^{\mathsf{T}}$, and $\mathbf{B} = (\mathbf{B}_1^{\mathsf{T}}\ \mathbf{B}_2^{\mathsf{T}})^{\mathsf{T}}$ accordingly. Yee and Hastie (2003) propose the class of partial RR-VGLMs, which is given by

$$\boldsymbol{\eta} = \mathbf{B}_1^{\mathsf{T}}\mathbf{x}_1 + \mathbf{A}\mathbf{C}^{\mathsf{T}}\mathbf{x}_2 = \mathbf{B}_1^{\mathsf{T}}\mathbf{x}_1 + \mathbf{A}\boldsymbol{\nu} \tag{11}$$

where $\mathbf{B}_1$ is of full rank (or rather, is subject to known constraint matrices) and $\mathbf{B}_2 = \mathbf{C}\mathbf{A}^{\mathsf{T}}$. Partial RR-VGLMs are RR-VGLMs but with the reduced-rank component only applied to a subset of the regressors. We let $p_j = \dim(\mathbf{x}_j)$.

Eq. 11 can be interpreted as a reduced-rank regression of $\mathbf{x}_2$ after adjusting for covariates in $\mathbf{x}_1$. It follows the same idea as partial CCA (ter Braak 1988), where the effects of a set of covariables $\mathbf{x}_1$ have been partialed out to allow the relationship between a set of response variables and a set of variables of interest to be seen. Eq. 11 restricts a species' response to be a linear function of $\boldsymbol{\nu}$, i.e., sigmoid, but in this paper we extend Eq. 11 to give the class of quadratic (partial) RR-VGLMs. This enables (partial) CGO to be performed. For QRR-VGLMs, $\mathbf{x}_1$ often contains the intercept term only and $\mathbf{x}_2$ are the ''real'' environmental variables.

QUADRATIC RR-VGLMS

Quadratic (partial) RR-VGLMs extend Eq. 11 by adding quadratic terms in $\boldsymbol{\nu}$ to give

$$\begin{aligned}\boldsymbol{\eta} &= \mathbf{B}_1^{\mathsf{T}}\mathbf{x}_1 + \mathbf{A}\boldsymbol{\nu} + \begin{pmatrix}\boldsymbol{\nu}^{\mathsf{T}}\mathbf{D}_1\boldsymbol{\nu}\\ \vdots\\ \boldsymbol{\nu}^{\mathsf{T}}\mathbf{D}_M\boldsymbol{\nu}\end{pmatrix}\\ &= \boldsymbol{\alpha} - \frac{1}{2}\begin{pmatrix}(\boldsymbol{\nu} - \mathbf{u}_1)^{\mathsf{T}}\mathbf{T}_1^{-1}(\boldsymbol{\nu} - \mathbf{u}_1)\\ \vdots\\ (\boldsymbol{\nu} - \mathbf{u}_M)^{\mathsf{T}}\mathbf{T}_M^{-1}(\boldsymbol{\nu} - \mathbf{u}_M)\end{pmatrix}\end{aligned} \tag{12}$$

where $\mathbf{D}_j$ are $R \times R$ symmetric matrices, and $\boldsymbol{\alpha}$ is some vector depending on $\mathbf{x}_1$, $\mathbf{u}_j$, and the tolerance matrices $\mathbf{T}_j$. The definition (Eq. 12) is also more general than the ''partial canonical Gaussian ordination'' model (ter Braak 1988) because, for example, the tolerance matrices can differ. The justification for adding the quadratic forms in Eq. 12 is to accommodate the unimodal response tenet mentioned above.

The jth response surface in Eq. 12 is bell shaped in the latent variables $\boldsymbol{\nu}$ if and only if $\mathbf{D}_j$ is negative-definite, i.e., if and only if $\mathbf{T}_j = -\frac{1}{2}\mathbf{D}_j^{-1}$ is positive-definite. The matrices $\mathbf{D}_j$ and $\mathbf{T}_j$ control the (ellipsoidal) contours of the bell-shaped surface in the R-dimensional latent variable space, for example, if they are diagonal, then the axes of the ellipsoids are parallel to the ν_r $(r = 1, \ldots, R)$ (ordination) axes. The contours represent points that have the same fitted value (e.g., abundance or probability). The fitted $\hat{\mathbf{C}}$ are called (estimated) canonical coefficients, the jth column of $\hat{\mathbf{C}}$ are called the jth canonical coefficients (which are interpreted as weights), and

$$\mathbf{u}_j = \mathbf{T}_j\mathbf{a}_j \tag{13}$$

is the optimum of species j.

One feature about fitting (partial) CGO models using maximum likelihood estimation is that the rank-R solution cannot be obtained from a higher rank solution. This is unlike CCA, where the canonical coefficients are the same no matter what the dimension, i.e., the CCA solution is nested. Users of CGO need to specify a value of R.

Simplification hypotheses

In Eq. 12, the $\mathbf{T}_j$ are general symmetric matrices. Consequently, the total number of parameters grows very quickly with R. For many applications it is not practical to have so many parameters, and some simplification is strongly recommended. The following hypothesis tests give some simplications, and some of these can be handled with the constraints-on-the-functions framework.

H_{00}: $\mathbf{D}_j = \mathbf{O}$, $j = 1, \ldots, M$. This states that all the species' responses are sigmoid and not quadratic, so

that the QRR-VGLM becomes a RR-VGLM. This hypothesis will be more likely to be accepted if there is little environmental variability in the sample.

H_{01}: $\eta_1 = \cdots = \eta_M$. This is a very strong hypothesis. It states that the response surfaces of all the species are identical.

H_{02}: $\mathbf{T}_1 = \cdots = \mathbf{T}_M$. That is, all species have the same tolerance matrix. VGAM has an argument `EqualTolerances` for this. If true, this hypothesis simplifies the interpretation of the ordination diagram and decreases the computational cost of estimating the model. Ter Braak and Prentice (1988) note that when the tolerance of the species are allowed to differ, the likelihood function typically contains many local maxima.

H_{03}: $\mathbf{u}_1 = \cdots = \mathbf{u}_M$. That is, all species have the same optimum. Fitting this model does not appear possible using the constraints-on-the-functions framework.

H_{04}: $\mathbf{T}_j = t_j^2 \mathbf{I}_R$. We call this the ''spherical'' assumption because the tolerance with respect to each canonical axis is the same. For rank $R = 2$, the elliptical contours are circular. The spherical hypothesis was assumed in ter Braak (1987*b*:113). The hypothesis H_{04} must be made in reference with a normalization, e.g., Eq. 14.

H_{02} is particularly important to the user because, if it is true, then the latent variable space can be rotated so that all the $\mathbf{T}_j$ become diagonal. The canonical axes can then be stretched/shrunken so that all the $\mathbf{T}_j$ become $\mathbf{I}_R$. At this stage, the ordination diagram becomes particularly easy to interpret.

Theoretically, CGO with Poisson and binary responses can be fitted to any data set satisfying *nM* greater than or equal to the number of parameters. The benefits of having any of the above hypotheses holding is that it reduces the number of parameters to be estimated. Consequently, there is a gain in speed and there may be a greater chance of successful convergence.

Ordination diagrams

A major by-product of CGO is an ordination diagram (also called a latent variable plot in this article), which is practical for ranks $R = 1$ and 2. They enable one to visualize relationships among the site scores $\hat{\boldsymbol{\nu}}_i$, the environmental variables $\mathbf{x}_2$, and the optima $\hat{\mathbf{u}}_j$. Relative abundances can also be read off CGO diagrams.

A CGO latent variable plot for rank $R = 1$ is straightforward: the x-axis is $\hat{\nu}$ and the y-axis is $\hat{\mu}$ or $\hat{\eta}$. For a CGO diagram for a rank $R = 2$ model, the x-axis is $\hat{\nu}_1$ and the y-axis is $\hat{\nu}_2$. However, ideally, one would want directions and Euclidean distances to have a natural meaning, as well as latent variables that are uncorrelated. All these ideals can be met provided an equal-tolerances assumption ($\mathbf{T}_j = \mathbf{T}$, say, for all j) is made.

Distances between points on an ordination diagram must be viewed in terms of the quantity $(\boldsymbol{\nu} - \mathbf{u}_j)^T \mathbf{T}_j^{-1} (\boldsymbol{\nu} - \mathbf{u}_j)$ in Eq. 12. This term is like a squared Mahalanobis distance, therefore proximities must be viewed with respect to the contours associated with the bell-shaped response surfaces defined by Eq. 12. This is because, for example, the abundance or probability of occurrence of a species decreases with distance from its optimum. We want the Mahalanobis distance and Euclidean distance to coincide by having $\mathbf{T}_j = \mathbf{I}_R$.

However, currently VGAM uses the normalization Eq. 14 during the estimation process, which leads to site scores which are uncorrelated, but in general, the $\hat{\mathbf{T}}_j$ will be non-diagonal. Since the elliptical contours of the $\hat{\mathbf{T}}_j$ are needed to judge distances correctly, it is a very good idea if a transformation can be done to circumvent the need for contours to be displayed on the ordination diagram. Fortunately, this is possible by transforming the $\hat{\mathbf{T}}_j$ so that they are $\mathbf{I}_R$. This results in contours that are spherical, therefore distances between two points naturally measure the dissimilarity between the two points. That is, the ordination diagram becomes a ''distance plot''—the ''distance rule,'' whereby a site that is close to an optimum is more likely to contain the species than a site that is far from the optimum, holds. For example, in Fig. 2, sites 12 and 11 are the most abundant sites of *Pardosa monticola*.

By default, VGAM's functions `Coef()` and `lvplot()` (for ''latent variable plot'') transform the $\mathbf{T}_j$ to $\mathbf{I}_R$ if the $\hat{\mathbf{T}}_j$ are constrained to be equal and are positive-definite. However, in order to retain the uncorrelatedness of the latent variables, the site scores are first rotated so that the $\hat{\mathbf{T}}_j$ are diagonal and then the $\hat{\mathbf{T}}_j$ are fully transformed into $\mathbf{I}_R$. Consequently, Eq. 14 no longer holds, though $\text{Var}(\hat{\boldsymbol{\nu}}_i)$ remains diagonal.

Such CGO diagrams need to be scaled so that the (circular) contours appear circular, i.e., a correct aspect ratio. When this is done, the circular contours can be omitted from the plot. Also, CGO diagrams will generally result in the range of the site scores along the second latent axis being smaller than the first canonical axis. One explanation for this is that the first canonical axis will usually do better at separating out the species so that there is more overlap between species on the second axis. This feature can be seen in Figs. 2 and 3.

How can a rank-2 CGO diagram be constructed if the $\mathbf{T}_j$ are not all equal? To do this, one can use Eq. 14 and add the contours of the response surfaces in order to supply a metric for judging distances and directions. Without the contours, the ordination diagram will be deceptive and misleading. When the $\mathbf{T}_j$ are very different, the interpretation is unfortunately complicated.

Regardless of the normalization of the latent variable plot, for $R = 2$, one has the $\hat{\boldsymbol{\nu}}$ as the canonical axes, with the optima $\hat{\mathbf{u}}_j$ superimposed, and the rows of $\hat{\mathbf{C}}$ can be plotted as arrows emanating from the origin. By projecting the arrows onto the canonical axes ν_r, the

way in which each latent variable depends on each environmental variable can be readily seen. The n individual site scores $\hat{\boldsymbol{\nu}}_i$ can also be added to the plot. It is sometimes useful to add the convex hull of the site scores to the rank-2 latent variable plot to give an idea of the sampled range of environments. Any optimum lying outside this region should be viewed with greater uncertainty.

It can be noted that CGO diagrams have several features not available to CCA. One is that contours of the bell-shaped surfaces can be added to the plot. VGAM allows the contour levels to be either relative to the species' maximum (e.g., 95%) or an absolute value, or simply be a specified number of tolerance values away from the optimum. Another feature is that optima are not plotted unless the fitted response surface is actually bell shaped; CCA plots them (weighted averages) regardless. Finally, note that if $\mathbf{x}_2$ has mean $\mathbf{0}$ then $E(\boldsymbol{\nu}) = E(\mathbf{C}^T\mathbf{x}_2) = \mathbf{0}$; this implies that if the covariates $\mathbf{x}_2$ have been centered, then the sample mean (centroid) of the site scores $\hat{\boldsymbol{\nu}}_i$ over the entire data set will be located at the origin.

Residuals

With QRR-VGLMs, there are several types of residuals. These include ''response'' residuals $y_{ij} - \mu_{ij}$, Pearson residuals, deviance residuals (applicable for binary and Poisson data), and working residuals based on the adjusted dependent vectors $\mathbf{z}_i$ from the Newton-Raphson/Fisher-scoring algorithm. Residual analysis is potentially very useful to check the adequacy of fitted models and is highly recommended. In VGAM, the generic function `plot()` is used to produce residual plots, whereas `lvplot()` produces a latent variable plot or ordination diagram.

Some Technical Details

Normalizations

The factorization Eq. 10 is not unique because $\mathbf{AC}^T = \mathbf{AMM}^{-1}\mathbf{C}^T$ for any nonsingular matrix $\mathbf{M}$. There are a number of possible uniqueness constraints, e.g., Yee and Hastie (2003) use corner constraints. For QRR-VGLMs, it is convenient to consider two normalizations. The first is to restrict $\mathbf{C}$ so that

$$\text{Var}(\boldsymbol{\nu}) = \text{Var}(\mathbf{C}^T\mathbf{x}_2) = \mathbf{I}_R. \tag{14}$$

This normalization means the site scores (latent variables) are uncorrelated and have unit variance. It is always possible to use this normalization, therefore this is currently used by VGAM during the fitting stage.

Using Eq. 14 has several advantages. As well as being always possible, the tolerance of each species can also be compared to the total variation of the latent variables, i.e., unity. Another advantage is that it is quite numerically stable.

The following second normalization is only practical when the tolerance matrices $\mathbf{T}_j$ are equal and correspond to bell-shaped response surfaces, i.e., $\mathbf{T}_j$ is positive-definite under H_{02}. Then we choose $\mathbf{M}$ so that

$$\mathbf{T}_j = \mathbf{I}_R \qquad j = 1, \ldots, M \tag{15}$$

i.e., the contours of the response surface are spherical (circular when the rank is two). The normalization Eq. 15, which gives an ecological scale to the canonical axes because the individual tolerances are all unity, is similar in aim to ''Hill's scaling.'' Also, Eq. 15 has the important property that in a rank-2 latent variable plot, distances in the ordination diagram between optima $\hat{\mathbf{u}}_j$ and site scores $\hat{\boldsymbol{\nu}}_j$ measure dissimilarities between these quantities—the smaller the distance the more similar.

In general, both normalizations cannot hold simultaneously. When the second normalization is applicable (it is to be preferred in latent variable plots), we apply a rotation followed by a scaling operation. This results in a relaxing of the first normalization (Eq. 14) so that $\text{Var}(\boldsymbol{\nu})$ is merely diagonal. The rotation does not affect Eq. 14 because the $\text{Var}(\mathbf{R}\boldsymbol{\nu}) = \mathbf{I}_R$ for all orthogonal $\mathbf{R}$. Consequently, Eq. 14 is not unique.

Even if all the $\mathbf{T}_j$ are not all equal, it is possible to choose one of the species and rotate its (assumed bell-shaped) response surface so that its elliptical contours have either minor-axis or major-axis parallel to the first ordination axis. This is equivalent to transforming $\mathbf{T}_j$ to be diagonal for that species. It is more convenient to choose the minor axis rather than the major axis because the former, corresponding to a lower tolerance, often is associated with a greater separation between all the species. When the ordination diagram is rotated as such, the interpretation that can be ascribed is that the latent variables are uncorrelated, and that the two latent variables do not interact for this species. In practice, the species best chosen for this rotation might be a dominant species whose full environmental range is reflected in the data set.

Note that Eqs. 14 and 15 differ from the common normalizations of CCA, viz. Hill's scaling and biplot scaling. Consequently, the ordination diagrams presented here cannot be compared ''exactly'' to those of CCA. Currently, by default, VGAM's `Coef()` and `lvplot()` for CGO objects choose Eq. 15 if `EqualTolerances = TRUE`; otherwise Eq. 14 is used.

Interpretation of the estimated canonical coefficients $\hat{\mathbf{C}}$ should be made only when the site scores $\hat{\boldsymbol{\nu}}_i$ are uncorrelated, cf. principal components analysis.

Inference and identification of the number of latent variables

Up to this point, we have assumed the rank R to be known whereas in practice it is rarely known. Its determination is an important subproblem. Unfortunately, the likelihood ratio test is not valid due to technical reasons (see, e.g., the discussion in Anderson [1984]). Yee and Hastie (2003) suggest the use of information theoretic-quantities such as minimizing the AIC (Akaike 1974) to determine R.

VGAM has a methods function `Coef.qrrvglm()` that shows which of the fitted $\eta_j(\boldsymbol{\nu})$ are bell-shaped. One ad hoc method is to increase R until many species' response curves are no longer bell-shaped. Ideally, they are bell shaped for $R = 1$ and 2. As R increases, it is less likely for species' response surfaces to be bell-shaped, therefore it is sometimes necessary to assume an equal-tolerances assumption $\mathbf{T}_j = \mathbf{T}$. In practice, it pays to keep R low if possible (e.g., between 1 and 3, inclusive) because the computational cost increases very rapidly with R. This compares with CCA, where fits are often rank-2 because they are plotted on computer screen and/or paper (although the analysis itself is of full rank) and CANOCO as currently programmed extracts only the first four axes.

Empirically, if a higher-rank model is fitted to lower-rank data (rank R, say), then only the ''first'' R tolerances will be finite under Eq. 14. That is, the contours of the ellipsoids will be elongated along the $R + 1$, $R + 2$, . . . canonical axes. Consequently, there is much ''overlap'' between the species in these higher dimensions. The practical lesson from this is that, provided the data and the statistical model agree, very wide tolerances relative to the site scores among all the species can be indicative of a model with too high rank. This is coupled with the problem that too high rank models become difficult to fit in the first place.

Estimation

Successful application of QRR-VGLMs requires a fast computer, experimentation, patience, and at least a rudimentary understanding of how they are estimated. Such an account is given in Appendix A because of its higher technical level. In this section, we consider important practical issues, which are based on Appendix A.

Initial values

Canonical Gaussian ordination estimated by maximum likelihood will converge faster if good initial values are provided. As Koojiman (1977*b*) notes (for non-canonical Gaussian ordination), good initial values can be crucial, and the quality of these become more important with increasing numbers of parameters.

One idea to obtain initial estimates is by fitting the smallest rank-R QRR-VGLM that is possible. For example, in Poisson or binary Gaussian ordination, this means only fitting R species. Given the $\hat{\mathbf{C}}$ from the initial model, this can be fed into the full QRR-VGLM as initial values. Practical experience with this ''trick'' has shown it can dramatically decrease the total computing time. A related idea is to fit a CCA model and use the solution as initial values for the QRR-VGLM—note however, that software such as CANOCO return standardized coefficients and use a different normalization.

Given the solution of a rank-R QRR-VGLM, initial values for the rank-$(R + 1)$ model could be $\mathbf{C}^0_{R+1} = (\hat{\mathbf{C}}_R, \boldsymbol{\varepsilon})$ where $\boldsymbol{\varepsilon}$ is a vector of p_2 random normal variates with zero means.

In the absence of inputted initial values by the user, VGAM currently chooses each element of $\mathbf{C}$ to be independent normal random variates with zero means.

It should be noted that it is a good idea for each x_k to be checked for outliers, high amounts of skew, etc., and appropriate action such as transformations be applied. For example, Palmer (1993) advocated using the logarithm transformation for soil chemical data. Ideally, $\nu_r = \mathbf{c}_{(r)}^T\mathbf{x}_2$ needs to be a plausible underlying gradient with bell-shaped responses for each species. If the individual variables in $\mathbf{x}_2$ are ''well behaved,'' then linear combinations of such variables should also be well behaved.

Dispersion parameters

Up until now, it has been assumed that the dispersion parameters ϕ_j are known, so that QRR-VGLMs are maximum likelihood estimates. When the ϕ_j are unknown, they may be estimated by the method-of-moments (McCullagh and Nelder 1989), and although the final model is no longer fully estimated by maximum likelihood, we loosely still say that QRR-VGLMs give maximum likelihood estimates because the ϕ_j are usually treated as nuisance parameters.

Approximate standard errors

Standard errors for parameter estimates are generally available from an IRLS algorithm. For RR-VGLMs, there are complications due to the fact that $\mathbf{A}$ and $\mathbf{C}$ are not simultaneously estimated under a single regression; however, Yee and Hastie (2003) show how to compute the complete variance-covariance matrix for the estimated elements of $\mathbf{A}$, $\mathbf{B}_1$, and $\mathbf{C}$ by fitting two overlapping models and then combining their estimated variance-covariance matrices. For QRR-VGLMs, one cannot use this method, and furthermore, the second derivative matrix with respect to the elements of $\mathbf{C}$ appear intractable.

The following is a crude approximation. If we treat $\hat{\mathbf{C}}$ as known, and then produce standard errors for the estimated elements of $\mathbf{A}$, $\mathbf{B}_1$, and $\mathbf{D}_j$, then these standard errors will be generally too small, nevertheless, they can still be useful for approximate inference. Standard errors for $\hat{\mathbf{C}}$, therefore, are currently unavailable.

Example

In this section, VGAM is used to fit QRR-VGLMs to a Dutch hunting spider data set (ter Braak 1986). VGAM is based on Version 4 of the S language (Chambers 1998) and uses the modeling ideas of Chambers and Hastie (1993), viz. formula, data frames, and generic functions. Two texts on S modeling are Venables and Ripley (2002) and Dalgaard (2002). All timings presented here were calculated using R 1.8.0 on a 2.4-GHz Pentium 4 machine running Linux. VGAM is under continual

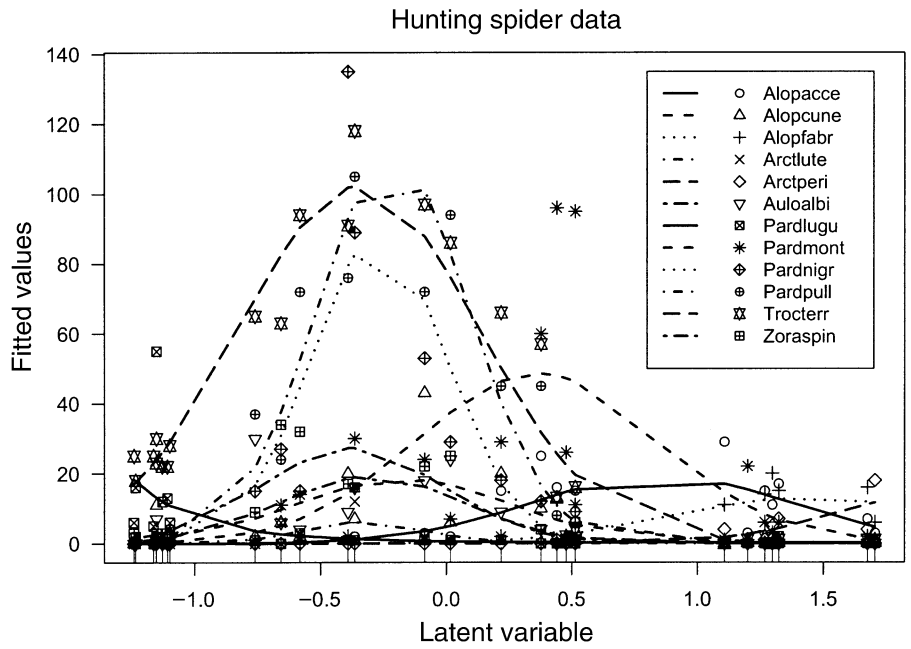

FIG. 1. Output from the generic function `lvplot()` when applied to a rank-1 Poisson QRR-VGLM with unequal tolerances. The fitted values are joined by lines; hence the lack of smoothness in the curves. The rug plot at the bottom shows jittered (slightly to separate ties) values of the site scores $\hat{\nu}_i = \hat{\mathbf{c}}^T \mathbf{x}_{2i}$, $i = 1, \ldots, n$. The species are: Alopacce, *Alopecosa accentuata*; Alopcune, *Alopecosa cuneata*; Alopfabr, *Alopecosa fabrilis*; Arctlute, *Arctosa lutetiana*; Arctperi, *Arctosa perita*; Auloalbi, *Aulonia albimana*; Pardlugu, *Pardosa lugubris*; Pardmont, *Pardosa monticola*; Pardnigr, *Pardosa nigriceps*; Pardpull, *Pardosa pullata*; Trocterr, *Trochosa terricola*; and Zoraspin, *Zora spinimana*.

development. Therefore future changes to its usage cannot be ruled out. Software implementing VGLMs, RR-VGLMs, and vector generalized additive models (VGAMs) are packaged in the VGAM package for S-PLUS and R.[2]

Hunting spider data

Ter Braak (1986) applied CCA to $M = 12$ species (see Fig. 1 caption) of hunting spiders in a Dutch dune area, and a similar analysis is presented here based on Poisson CGO via QRR-VGLMs (Eqs. 8 and 12). The data, which originally came from Van der Aart and Smeek-Enserink (1975), consists of abundances (numbers trapped over a 60-wk period) and six environmental variables. There were $n = 28$ sites. In the CCA analysis, a square-root transformation was applied to the abundances to reduce the effect of large values; we defer such an analysis to later. For comparison with the CCA analysis of ter Braak (1986), each of the six environmental variables was standardized to zero mean and unit variance.

Suppose the data is stored in a data frame called `hspider`. A rank-1 model with unequal tolerances can be fitted using the following:

```
p1 = cgo(cbind(Alopacce, Alopcune, Alopfabr,
               Arctlute, Arctperi, Auloalbi,
               Pardlugu, Pardmont, Pardnigr,
               Pardpull, Trocterr, Zoraspin)
    ~ WaterCon + BareSand + FallTwig
      + CoveMoss + CoveHerb + ReflLux,
      family = quasipoissonff,
      data = hspider,
      Crow1positive = FALSE,
      EqualTolerances = FALSE).
```

The VGAM family function `quasipoissonff` is used so that a dispersion parameter ϕ_j may be estimated for each species. The environmental variable names correspond, in order, to those in Table 1. Residual plots of `p1` (not given here) did not reveal any gross outliers. The ordination diagram

```
lvplot(p1, 1col = 1:12, llwd = 2,
       llty = 1:12, y = TRUE,
       pch = 1:12, pcol = 1:12,
       las = 1, main = "Hunting spider data")
```

gives Fig. 1. Most of the arguments refer to graphical parameters such as line width, colors, etc. and are optional. With a dozen species, the plot is quite cluttered, especially with the abundances overlaid as points (different line and symbol for each species). The plot suggests the model provides a reasonable fit to these data.

[2] ⟨http://www.r-project.org/⟩

TABLE 1. Estimated rank-1 Poisson QRR-VGLMs applied to the standardized hunting spider data.

	$\hat{\mathbf{C}}$ (count data)		$\hat{\mathbf{C}}$ (square-root count data)			
Variable	Unequal	Equal	Unequal	Equal	CCA1	CCA2
Water content	−0.119	−0.150, −0.356	−0.215	−0.127, −0.224	−0.51	−0.41
Bare sand	0.261	0.234, 0.554	0.227	0.244, 0.432	0.33	−0.10
Fallen twigs	−0.306	−0.387, −0.918	−0.302	−0.421, −0.745	−0.14	0.37
Cover moss	0.107	0.134, 0.318	0.204	0.192, 0.339	0.05	−0.27
Cover herbs	−0.172	−0.128, −0.304	−0.089	−0.064, −0.113	−0.28	−0.15
Light refl.	0.406	0.297, 0.703	0.253	0.219, 0.388	0.27	−0.03
Deviance	1176.00	1585.11	167.22	252.29		

Notes: ''Unequal'' and ''equal'' refer to the tolerances. For the equal-tolerances column, the LHS estimate corresponds to the normalization Eq. 14, whereas the RHS estimate corresponds to $\hat{\mathbf{T}}_j = \mathbf{I}_2$. The unequal-tolerances columns correspond to Eq. 14. The first two canonical coefficients of CCA (ter Braak 1986) are included (CCA1 and CCA2) for ''comparison.'' The log-transformed environmental variables are water content (percentage of soil dry mass), bare sand (percent cover of bare sand), fallen twigs (percent cover of fallen leaves and twigs), cover moss (percent cover of the moss layer), cover herbs (percent cover of the herb layer), and light refl. (reflection of the soil surface with cloudless sky).

Coefficients and other useful quantities can be obtained using the generic function `Coef()`, for example, typing `Coef(p1)` produces

C matrix:

	lv
WaterCon	−0.119
BareSand	0.261
FallTwig	−0.306
CoveMoss	0.107
CoveHerb	−0.172
ReflLux	0.406

	Optimum	Maximum	Tolerance
Alopacce	0.854	19.29	0.508
Alopcune	−0.169	18.38	0.431
Alopfabr	1.445	13.03	0.532
Arctlute	−0.327	6.17	0.241
Arctperi	1.993	14.59	0.430
Auloalbi	−0.298	19.24	0.365
Pardlugu	NA	NA	NaN
Pardmont	0.363	48.60	0.481
Pardnigr	−0.270	87.90	0.269
Pardpull	−0.213	110.39	0.303
Trocterr	−0.347	102.29	0.474
Zoraspin	−0.377	27.25	0.360.

It appears that all the environmental variables are important for the latent variable, and that the strongest weighting is for light reflection. In Table 1, the first canonical coefficients of CCA ''agree'' with the QRR-VGLM coefficients in that the signs match.

The output reveals that all species except *Pardosa lugubris* had a fitted quadratic that was bell shaped. In general, the estimated response curve of a species has a higher chance of being bell shaped the larger the environmental range of the sample data. At their optimal environment, *Pardosa pullata* and *Trochosa terricola* are the most abundant species. The location of the optima, when overlaid on Fig. 1, show seven species with their estimated optimum $\hat{u}_j$ clustered between −0.4 and −0.1 on the latent variable $\hat{\nu}$ scale.

The estimates of the dispersion parameters were:

Alopacce	Alopcune	Alopfabr	Arctlute	Arctperi
2.802	7.526	2.136	1.139	0.848

Pardmont	Pardnigr	Pardpull	Trocterr	Zoraspin
12.358	5.575	5.394	14.480	3.115

Auloalbi	Pardlugu
5.730	8.849.

Not surprisingly, there is a tendency for species to exhibit overdispersion relative to a Poisson distribution. If one was willing to assume that each species' dispersion parameter was equal (not realistic here), then the estimate of this common value would be $\hat{\phi} = 4.47$ for the unequal-tolerances model—a significant amount of overdispersion.

To test whether the species have equal tolerances, the above analysis was repeated but with the argument `EqualTolerances = TRUE`. This gave a deviance of 1585.11 whereas deviance was 1176.00 for the unequal-tolerance model. Then a likelihood ratio test gives P value $P[\chi^2_{11} > 409.11/4.46] \approx 0$, which gives extremely strong evidence against H_{02}: $t_1^2 = \cdots = t_{12}^2$. This is not surprising, since the species' tolerances differ by a factor of 2. The tolerance for *Pardosa lugubris* is undefined because its response curve is not bell-shaped, and with so little data from this species, it is not surprising that its fitted curve is ''U''-shaped. Table 1 also gives the estimated **C** from the equal-tolerances model, and these appear to be a little different from the unequal-tolerances model.

Using a number of different sequences of random initial **C** values, estimation typically took between 20 s

and 80 s (median, 30 s) to converge for the unequal-tolerances model. The equal-tolerances model typically took between 10 s and 30 s.

Note that, if a few species are not bell shaped, fitting a model with equal tolerances may force all species to become bell shaped because the common tolerance pools information from all species to give an overall estimate. For these data, $\hat{t} = 0.422$ under normalization (Eq. 14).

Practical experience with this data showed that convergence to the global maximum likelihood solution was generally the case for the unequal-tolerances model. With the equal-tolerance model, convergence to its global solution was less likely.

Square root of the count data

The analysis described in the last section is dominated by the abundant species. For a closer comparison with ter Braak (1986), we repeated the analysis with the square root of these data. Doing this can result in a distribution that can never be exactly Poisson but can nevertheless be closer to a Poisson distribution than the original data.

The estimated regression coefficients are given in Table 1. It can be seen that the signs of the canonical coefficients $\hat{\mathbf{C}}$ match in both cases; this is a comforting property. The latent variable plot (not given) is quite similar to Fig. 1, and is arguably superior in that the residual plots are better and a small number of species no longer dominate the analysis. The estimated dispersion parameters for the unequal variances model are:

Alopacce	Alopcune	Alopfabr	Arctlute	Arctperi
0.523	0.711	0.771	0.279	0.115

Pardmont	Pardnigr	Pardpull	Trocterr	Zoraspin
0.970	0.599	0.654	0.564	0.549

Auloalbi	Pardlugu
0.836	1.319

which indicate most species are now under-dispersed relative to a Poisson dispersion. This seems to suggest the square-root transformation is too severe, and raising the counts to an intermediate power such as 3/4 might be better.

Rank-2 model

An equal-tolerances rank-2 model was fitted to 10 species, resulting in a fit with a deviance of 856.5. Compared to the rank-1 model, it was necessary to assume equal tolerances to get any useful results (it is usually difficult for individual species to be bell shaped in two dimensions; pooling their tolerances can stabilize the results.) Additionally, it appeared necessary to omit two species from the analysis because of numerical reasons.

The code

```
r2 = cgo(cbind(Alopacce, Alopcune, Alopfabr,
               Arctlute, Arctperi, Auloalbi,
               Pardmont, Pardnigr, Pardpull,
               Trocterr)
         ~ WaterCon + BareSand + FallTwig
           + CoveMoss + CoveHerb + ReflLux,
             family = quasipoissonff,
             data = hspider,
             Rank = 2,
             EqualTolerances = TRUE.
             Crow1positive = c(FALSE, FALSE))
```

typically took between 35 and 60 s to compute. Then $\hat{\mathbf{C}}^{\mathsf{T}}$ is

	WaterCon	BareSand	FallTwig	CoveMoss
lv1	−0.284	0.807	−0.86677	0.277
lv2	−0.574	0.180	0.00410	−0.849

	CoveHerb	ReflLux
lv1	−0.271	0.4822
lv2	0.392	−0.0249

which uses $\hat{\mathbf{T}}_j = \mathbf{I}_R$ scaling, and is similar to the rank-1 results. The second latent variable can be interpreted as a contrast between `CoveHerb` and the two variables `WaterCon` and `CoveMoss`. Compared to the CCA model of ter Braak (1986),the first column of $\hat{\mathbf{C}}$ largely agrees but the second column differs somewhat (see Table 1) especially in the variable `CoveHerb`.

The code

```
clr = (1:(10 + 1))[−7] # Omit yellow color
adj = c(−0.1, −0.1, −0.1, 1.1, 1.1, 1.1,
         −0.1, −0.1, −0.1, 1.1)
lvplot(r2, label = TRUE, xlim = c(−2.7,5.1),
       ellipse = FALSE, C = TRUE,
       Cadj = c(1.1. −0.1, 1.2, 1.1, 1.1,
                −0.1),
       adj = adj, las = 1, chull = TRUE,
       pch = "+", pcol = clr, sites = TRUE)
```

produces Fig. 2. Both ordination axes are plotted on a common scale. The arrows display $\hat{\mathbf{C}}$ and show how much the environmental variables make up the two latent variables—by projecting them onto the canonical axes. Ellipsoid contours at 95% of the maximum values are plotted later to avoid clutter. The plot here preserves the rank-1 feature (Fig. 1) that there is a cluster of species optima along the first ordination axis.

There are many other interesting features of the plot. The distribution of the site scores $\hat{\boldsymbol{\nu}}_i$ are spread over an elliptical region (delineated by the convex hull) and contrasts sharply with CCA where the site scores fall on the path of an arch—no doubt a spurious artifact. Additionally, the ''order'' and the length of the arrows representing the canonical coefficients $\hat{\mathbf{C}}$ differ between CGO and CCA substantially. To compare Fig. 2 and the CCA biplot (Fig. 1 of ter Braak 1986) more rigorously, a Procrustes analysis would be suitable.

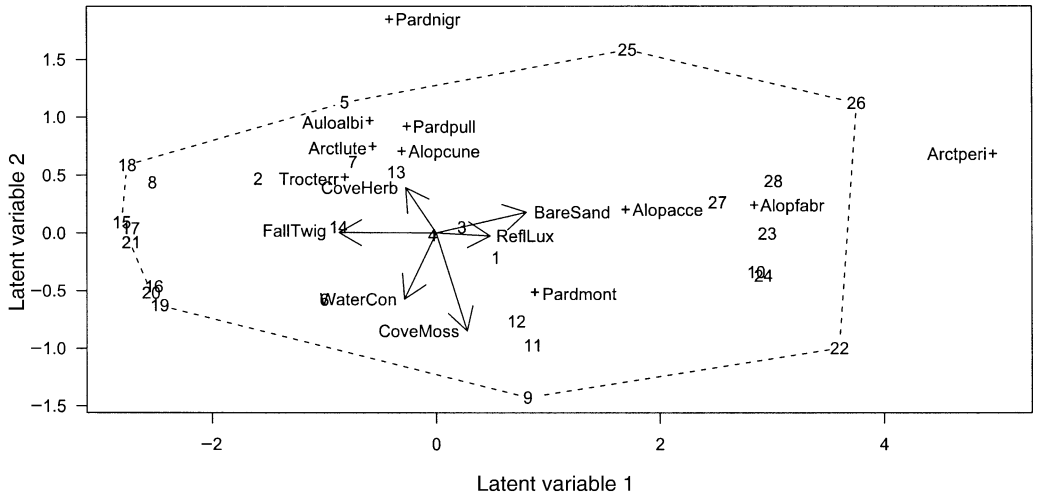

FIG. 2. CGO diagram for the hunting spider data (using $\hat{\mathbf{T}}_j = \mathbf{I}_R$). The dashed line is a convex hull of the labeled site scores (i for $\hat{\nu}_i$). The ''+'' symbols denote the position of the species' optima, and the arrows denote the canonical coefficients $\hat{\mathbf{C}}$. Species abbreviations are in the Fig. 1 legend; abbreviations for environmental variables are defined in Table 1.

Fig. 2 is quite cluttered, so

```
lvplot(r2, label = TRUE, xlim = c(−2.8, 5.0),
       las = 1, chull = TRUE,
       adj = adj, ecol = clr,
       pch = "+", pch = clr,
       sites = TRUE)
```

was used to produce Fig. 3. As an example of interpreting CGO diagrams, the latent variable plot suggests sites 3, 4, and 5 have similar abundances of *Alopecosa cuneata*, but sites 3 and 4 are expected to have more *Pardosa monticola* than site 5. The latent variable plot does not imply that site 5 has the same absolute abundance of *Trochosa terricola* and *Alopecosa cuneata* because the contours are relative to each species' maximum.

Relative abundances of one species at two sites can be readily read off from the CGO plots. For example, let A_9 and A_{23} be the absolute abundance of *Pardosa monticola* for sites 9 and 23, respectively. Then, because $\hat{\mathbf{T}} = \mathbf{I}_2$, we have $A_9 = \exp(\alpha_{\text{Pardmont}} - \frac{1}{2}\|\boldsymbol{\nu}_9 - \mathbf{u}_{\text{Pardmont}}\|^2)$ and $A_{23} = \exp(\alpha_{\text{Pardmont}} - \frac{1}{2}\|\boldsymbol{\nu}_{23} - \mathbf{u}_{\text{Pardmont}}\|^2)$, where the Euclidean distances are measured in latent variable units. In Figs. 2 and 3, it can be seen that $\|\hat{\boldsymbol{\nu}}_9 - \hat{\mathbf{u}}_{\text{Pardmont}}\| \approx 1$ and $\|\hat{\boldsymbol{\nu}}_{23} - \hat{\mathbf{u}}_{\text{Pardmont}}\| \approx 2$; therefore $\hat{A}_9/\hat{A}_{23} \approx \exp\{\frac{1}{2}(2^2 - 1^2)\} \approx 4.5$. That is, site 9 is expected to have ~4.5 times more counts of *Pardosa monticola* than site 23. In fact, the actual ratio is 26/6 $\approx$ 4.3. In general, if site i_2 is c times further from a species optimum $\mathbf{u}$ than site i_1 ($c > 1$), then

$$\frac{A_{i_1}}{A_{i_2}} = \exp\left\{\frac{\|\boldsymbol{\nu}_{i_1} - \mathbf{u}\|^2}{2}(c^2 - 1)\right\}.$$

In the same vein, relative abundances of two species at one site can be readily read off from the CGO diagrams. Let A'_j be the absolute abundance of species j at site i. Then, because

$$A'_j = \exp\left(\alpha_j - \frac{1}{2}\|\boldsymbol{\nu}_i - \mathbf{u}_j\|^2\right)$$

we have

$$\frac{A'_{j_1}}{A'_{j_2}} = \exp(\alpha_{j_1} - \alpha_{j_2}) \times \exp\left\{\frac{1}{2}(\|\boldsymbol{\nu}_i - \mathbf{u}_{j_2}\|^2 - \|\boldsymbol{\nu}_i - \mathbf{u}_{j_1}\|^2)\right\}.$$

When the maxima of the two species j_1 and j_2 are equal (i.e., $\alpha_{j1} = \alpha_{j2}$), then the first exponential can be ignored. To give a few specific examples, we note that, from the output from the rank-1 model, the maxima of the following pairs of species are very similar: *Alopecosa accentuata* and *Alopecosa cuneata*; *Alopecosa fabrilis* and *Arctosa perita*; *Pardosa pullata* and *Trochosa terricola*. Now $\|\hat{\boldsymbol{\nu}}_{26} - \hat{\mathbf{u}}_{\text{Arctperi}}\| \approx \|\hat{\boldsymbol{\nu}}_{26} - \hat{\boldsymbol{u}}_{\text{Alopfabar}}\| \approx 1$ so site 26 is expected to have approximately the same *Arctosa perita* and *Alopecosa fabrilis* counts. Another example is site 6:

$$\|\hat{\boldsymbol{\nu}}_6 - \hat{\mathbf{u}}_{\text{Trocterr}}\| \approx 1$$

$$\|\hat{\boldsymbol{\nu}}_6 - \hat{\mathbf{u}}_{\text{Trocterr}}\| \approx 1$$

$$\|\hat{\boldsymbol{\nu}}_6 - \hat{\mathbf{u}}_{\text{Pardpull}}\| \approx 1.5.$$

Therefore site 6 is expected to have approximately $\exp\{\frac{1}{2}(1.5^2 - 1)\} \approx 2$ times more *Trochosa terricola* counts than *Pardosa pullata* counts. In fact, the actual ratio is 63/24 $\approx$ 2.6.

Note that when $\hat{\mathbf{T}} = \mathbf{I}_2$, the absolute abundance of a species at a location that is a distance k away (in latent variable units) from the species' optimum is $\exp(-\frac{1}{2}k_2)$

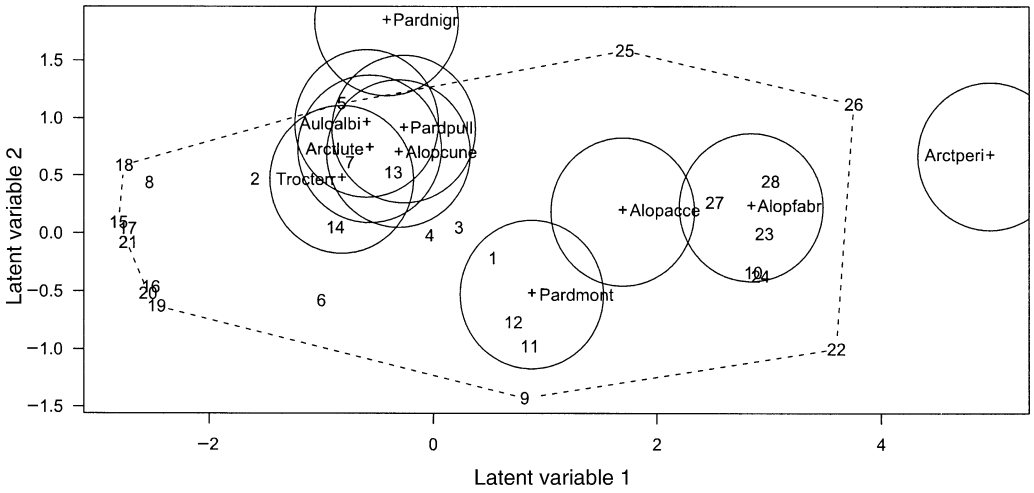

FIG. 3. Fig. 2 with contours included and arrows representing $\hat{\mathbf{C}}$ omitted. The circular contours denote 95% of each species' maximum (these are called relative contours). Species abbreviations are defined in the Fig. 1 legend.

multiplied by the species' maximum. For example, $\|\hat{\boldsymbol{\nu}}_6 - \hat{\mathbf{u}}_{\text{Troterr}}\| \approx 1$ implies the absolute abundance of *Trochosa terricola* at site 6 has been reduced by a factor of $1 - e^{-1/2} \approx 40\%$. Site 12 is approximately a distance two away from $\hat{\mathbf{u}}_{\text{Troterr}}$, therefore the absolute abundance of *Trochosa terricola* at site 12 is expected to be $\sim 102.29 \times \exp(-\frac{1}{2}2^2) \approx 13.8$. The actual count is 13.

Note that optima lying outside the convex hull of site scores are subject to more statistical error and therefore must be viewed with less confidence. In Fig. 2, *Arctosa perita* and *Pardosa nigriceps* are two such species.

The above analysis was tried on the square root of the abundance data but the estimated **T** was not positive-definite. Evidently, it appears that a square root transformation is too severe. Using a power of 0.75 worked (but with site 22 deleted because it is an influential point), and the resulting canonical coefficients are

	WaterCon	BareSand	FallTwig	CoveMoss
lv1	−0.376	0.601	−0.870	0.131
lv2	−0.356	0.165	0.024	−0.531

	CoveHerb	ReflLux
lv1	−0.317	0.420
lv2	0.697	−0.191

and the latent variable plot is given in Fig. 4. One can see that is roughly similar to Fig. 3. As with the raw counts, compared to the CCA model of ter Braak (1986), the $\hat{\mathbf{C}}$ here is roughly similar except the second canonical coefficient's `CoveHerb` is substantially different. When compared to CCA (see Fig. 1 of ter Braak 1986), Fig. 4 shows a few similarities,

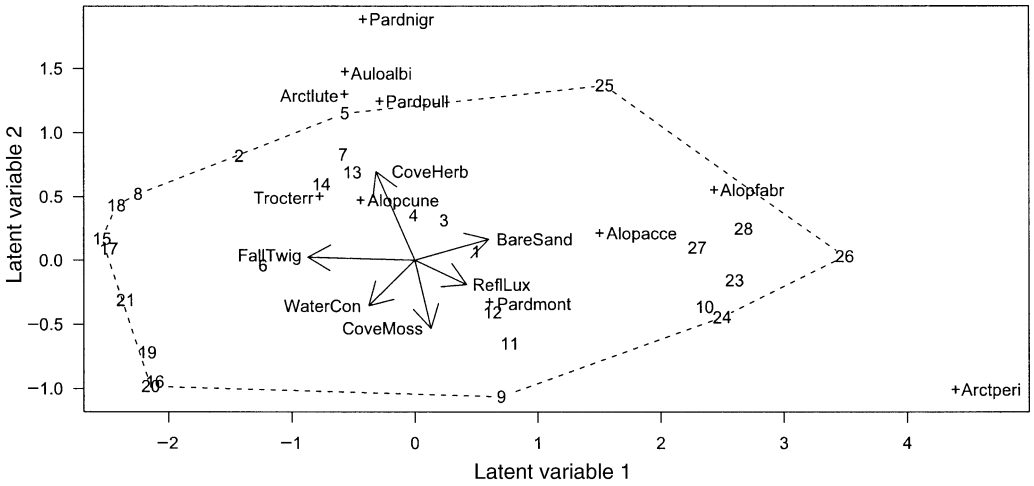

FIG. 4. The same type of analysis as shown in Fig. 2, but applied to transformed abundance data $y_{ij}^{0.75}$ instead of the raw counts y_{ij}. Site 22 has been omitted from the ordination. Species abbreviations are in the Fig. 1 legend; abbreviations for environmental variables are defined in Table 1.

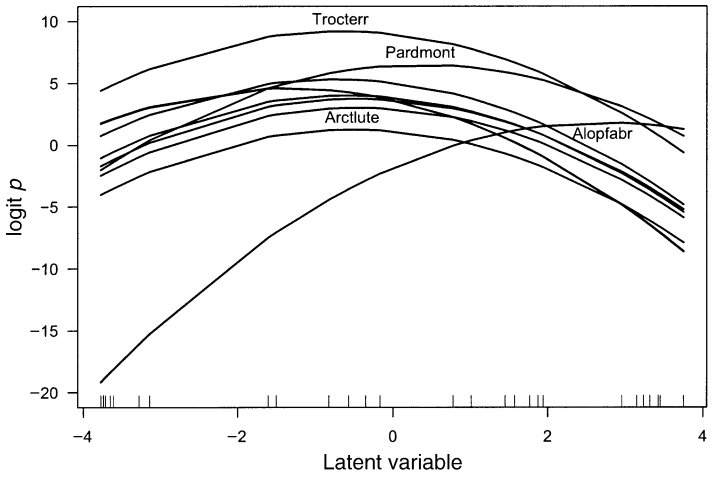

FIG. 5. Ordination diagram for the hunting spider (presence/absence) data. Some selected curves are labeled at the position of their optimas and maxima. This latent variable plot is comparable to Figs. 2–4 because equal tolerances have been assumed; therefore the tolerances have been scaled to unity. Species abbreviations are in the Fig. 1 legend.

but there are many differences. An obvious one is that the optima of *Pardosa nigriceps* and *Arctosa perita* differ greatly—but these optima lie outside the convex hull, therefore are subject to greater uncertainty.

Presence/absence data

For purely illustrative purposes of a canonical Gaussian logit ordination, the hunting spider species data were converted to presence/absence and a rank-1 model was fitted. QRR-VGLMs with binary responses are more prone to numerical instability if the fitted probabilities become very small during iterations. This is because the IRLS algorithm becomes unstable for the inner minimization problem described in Appendix A. The numerical problems are often due to species with small tolerances, and a possible method of circumventing this problem is to assume that all the species have common tolerances, i.e., a common $\mathbf{D}_j$ (an assumption that is made by CCA). Another way numerical problems can occur is, for example, for a rank-1 model, if the data are sorted with respect to the latent variable and the resulting response vector is a sequence of absences followed by a sequence of presences followed by a sequence absences (i.e., $\mathbf{y} = (\mathbf{0}^T, \mathbf{1}^T, \mathbf{0}^T)^T$); then a Gaussian model will result in estimates that go to infinity. Numerical problems such as these occur with ordinary logistic regression (see e.g., Rousseeuw and Christmann 2003). For this reason, fitting QRR-VGLMs with binary responses generally requires n large and $n \gg M$, e.g., $n > 100$ for $M = 2$.

For the hunting spiders data, there are only 28 sites, which would be far too few for QRR-VGLMs to work in general. It was therefore not surprising that numerical difficulties were encountered with an unequal tolerances assumption. With an equal tolerances assumption, it was necessary to omit species 1 and 5 from the analysis:

```
attach(hspider)
 ybin = 0 + (cbind(Alopacce, Alopcune, Alopfabr,
                   Arctlute, Arctperi, Auloalbi,
                   Pardlugu, Pardmont, Pardnigr,
                   Pardpull, Trocterr, Zoraspin)
             > 0) # Matrix of 0's and 1's
detach( )
b1 = cgo(ybin[, -c(1,5)] ~ Watercon
         + BareSand + FallTwig + CoveMoss
         + CoveHerb + ReflLux,
         family = quasibinomialff(mv = TRUE),
         data = hspider,
         EqualTolerances = TRUE,
         Crow1positive = FALSE)
lvplot(b1, type = "predictors", llwd = 2,
       las = 1, ylab = "logit p",
       ylim = c(-20, 11), 1col = 1:10)
c1 = Coef(b1);
cts = c("Trocterr", "Pardmont", "Alopfabr",
        "Arctlute")
text(c1@Optimum[1, cts],
     logit(c1@Maximum[cts]) + 1.0, cts)
```

Estimation typically took between 10 and 30 s, and a deviance of 154.6 was obtained. The latent variable plot for the above model is given in Fig. 5. It is plotted on the η (logit) scale, therefore the fitted curves are quadratics. For this, $\hat{t} \approx 0.36$ is the estimated tolerance under Eq. 14, and

	WaterCon	BareSand	FallTwig	CoveMoss
lv	−0.127	0.185	−0.498	0.228

	CoveHerb	ReflLux
lv	−0.031	0.15

is $\hat{\mathbf{C}}^T$; these coefficients agree with the rank-1 unequal tolerances Poisson model with respect to their signs.

In some applications, given abundance data, canonical Gaussian logit ordination will work better than a Poisson CGO, especially if the abundances are not

TABLE 2. Statistics pertaining to estimated **C** matrices from simulated rank-1 Poisson unequal-tolerances QRR-VGLMs fitted to the hunting spider data.

Variable	$\hat{\mathbf{C}}$ (original model)	Mean	SD
Water content	−0.119	−0.116	0.017
Bare sand	0.261	0.261	0.008
Fallen twigs	−0.306	−0.308	0.018
Cover moss	0.107	0.107	0.011
Cover herbs	−0.172	−0.174	0.013
Light refl	0.406	0.406	0.015

Notes: The sample mean and standard deviation are based on 100 simulations. See Table 1 notes for definitions of variables.

Poisson distributed. However, if the abundances are Poisson distributed, conversion to presence/absence will result in a loss of information and therefore give inferior inferences. In general, Poisson CGO is preferred over canonical Gaussian logit ordination because it is less prone to numerical problems and does not require n to be as large.

A simulation experiment

When fitting a statistical model to real data, often we have no idea how much to believe our model because the ''truth'' is unknown. Furthermore, if there are alternative solutions, it is often not easy to determine which is closer to the ''truth.'' One method to address this problem is to check the assumptions behind the statistical model, e.g., by model diagnostics such residual analyses. Another method is to use simulated data sets with various properties in order to gauge some idea of the likely performance of the model. In this subsection, a small simulation study is reported to check that the proposed methodology is reasonable in this respect. This was done for the rank-1 unequal tolerances Poisson model applied to the hunting spider data. For each site–species combination, a Poisson abundance was randomly generated based on the estimated mean from the original model (Table 1). This was performed 100 times. A QRR-VGLM was fitted to each of these simulated data sets, and Table 2 shows the amount of variation in the estimated **C** matrices. It can be seen that the fitted models conform closely to the ''true'' model. Furthermore, all latent variable plots were very similar to Fig. 1. A closer analysis of the results showed that separate 95% confidence intervals of each coefficient covers the corresponding coefficient from the original model, and that the signs of the regression coefficients matched those from the original model. All these results are encouraging—it gives confidence that QRR-VGLMs can be expected to fit at least some data sets well, provided the data and the fitted model match. More simulation studies, with the help of software tools such as CoenoFlex (*available online*)[3] and COMPAS (Minchin 1987a), are needed to determine the practical limitations and strengths of QRR-VGLMs.

In the simulations, the distribution of execution times, using randomly chosen starting values, had a median of 50 s and had values ranging between 20 s and 95 s.

DISCUSSION

Gaussian response surfaces hold an important role in ecological theory, but despite their simple functional form, the capability to fit them to common forms of biological data in a constrained ordination has only now been achieved. This paper has sought to show that QRR-VGLMs are a suitable vehicle for maximum-likelihood-estimated CGO, and with the software described in this article, practitioners can experiment with the methodology on their own data sets.

It is beneficial to summarize the advantages and disadvantages of the methodology presented in this article. In doing this, one must be careful to distinguish between CGO and QRR-VGLMs because the former represents the ''problem'' and the latter is one particular ''solution.''

Advantages of CGO include:

1) CGO is easy to understand and has clear statistical assumptions. Independence of the n sites and, given the model, M species is assumed. Strictly speaking, there is some violation of this assumption due to spatial correlation, but this will be so with any regression technique such as linear models, GLMs, and so forth applied to these data. Dependence between species could possibly be handled using generalized estimating equations (see e.g., Wild and Yee 1996). The distribution and functional form (Eq. 12) is explicitly stated.

2) CGO provides estimates for the maxima, optima, and tolerances of species, which are highly interpretable. Each $\hat{\mathbf{u}}_j$ can be classified as an optimum, a minimum, or a saddle point. Maxima and contours are not available in CCA. CGO diagrams can be highly intuitive and informative.

3) CGO bypasses the complications, heuristics, and statistical inefficiencies of other methods such as weighted-averaging and detrending. Gaussian ordination is built on a firm statistical bed of theory.

4) The model is arguably more flexible than CCA. For example, CCA makes four assumptions (including equal tolerances, equal maxima, homogeneously distributed optima) that cannot simultaneously hold. CGO is not required to make any of these assumptions, though an equal tolerances assumption makes the ordination diagram very interpretable.

5) Residual diagnostic plots are available.

Disadvantages of canonical Gaussian ordination include:

1) Often the statistical assumptions are too strong and unrealistic in practice. For example, a limitation of Eq. 12 is that skewed and multimodal responses are

[3] ⟨http://labdsv.nr.usu.edu⟩

not handled. Furthermore, full specification of the statistical model may not be possible.

2) CGO is sensitive to departures from the model assumptions, as well as to outliers, sparse data, high leverage points, multicollinearity, etc. For example, when the $\mathbf{x}_2$ includes highly correlated environmental variables, the result is that $\mathbf{C}$ is unstable; however, this problem occurs in other (unpenalized) regression methods.

3) The log-likelihood function often has multiple maxima, and convergence to a local solution can never be ruled out.

4) Good initial values can be difficult to obtain, and this becomes more important as the number of parameters increases.

5) Statistically rigorous determination of the rank R is difficult. See the discussion of Anderson (1984).

6) CGO is numerically very intensive. However, as CPU power continues to increase exponentially according to Moore's law, the size of data sets that can be practically handled will increase consequentially.

Like every statistical methodology, QRR-VGLMs have both strengths and weaknesses. QRR-VGLMs have the following advantages:

1) The framework of QRR-VGLM is large, so that it can potentially perform CGO on many data types. For example, in theory, distributions such as the zero-inflated Poisson and the negative binomial and data types such as compositional data could be handled.

2) The latent variables defined by $\boldsymbol{\nu} = \mathbf{C}^\top \mathbf{x}_2$ can accommodate linear combinations of nonlinear functions of the explanatory variables. For example (for purely illustrative purposes),

```
cgo(ymatrix ~ poly(x1, 2) + x2 + x3
    + I(x2 * x3), . . .)
```

will use $\boldsymbol{\nu}_r = \mathbf{c}_{(r)}^\top (x_1, x_1^2, x_2, x_3, x_2x_3)^\top$. In particular, $\mathbf{x}_2$ can include factors and basis functions such as orthogonal polynomials and B-splines. In contrast to the natural capability of the QRR-VGLM algorithm to handle such extensions, the proposal of Makarenkov and Legendre (2002) for polynomial CCA is ad hoc and empirical.

3) In principle, any rank R can be fitted, subject to the model being estimable. This contrasts with previous (non-canonical) solutions that were restricted to $R = 1$ and $R = 2$ only (e.g., Gauch et al. 1974, Kooijiman 1977, Goodall and Johnson 1982, ter Braak 1985).

4) The approximate statistical significance of the elements of $\mathbf{A}$, $\mathbf{B}_1$, and $\mathbf{D}_j$ can be tested with QRR-VGLMs.

5) Statistical tests for certain hypotheses can be conducted.

6) Overdispersion and underdispersion can be detected in the data, a feature unavailable with CCA.

The major disadvantages of QRR-VGLMs include:

1) It requires considerable amounts of memory because IRLS constructs large model matrices to perform the least squares fits. Although direct maximization of the log-likelihood function could avoid this problem, this would probably introduce other difficulties.

2) Specialist software is needed. This contrasts with methods such as CCA, which can, in theory, be fitted using a number of algorithms. Additionally, the overall algorithm is sophisticated so that any implementation would be a major undertaking.

3) Convergence problems may occur. The algorithm is susceptible to numerical problems such as underflow and overflow, and this is aggravated with dirty data.

4) Greater skill and knowledge is required to use QRR-VGLMs successfully. The user needs to know elements of statistical computing and numerical analysis for a complete understanding of QRR-VGLMs.

5) ''Exact'' standard errors for all estimated parameters is currently not possible. This may well be improved in the future. The use of the bootstrap or simulation to provide standard errors is currently not practical, as it may be too computationally demanding—it would require fitting multiple QRR-VGLM models. Permutation testing (ter Braak and Šmilauer 1998) is also a possibility.

Overall, it is recommended that QRR-VGLMs be applied to clean (no outliers, multicollinearity, or influential observations; the variables are not heavily skewed; etc.) data sets of moderate size that conform reasonably closely to the Gaussian model, and that CCA be applied to large and/or noisy/dirty data sets due to its lower computational expense and robustness. To give a rough idea of the practicability, with the computer configuration described above, rank-1 unequal tolerance QRR-VGLMs have been fitted to artificial data sets of size $n = 500$ sites, $M = 5$ species, and $p = 5$ covariates, with median execution time of around 70 s. This takes about the same time as a $n = 60$, $M = 10$, and $p = 10$ configuration. Rank-2 equal tolerance QRR-VGLMs have been fitted to size $n = 100$, $M = 10$, and $p = 5$ data sets with median execution time of around 130 s. The computational cost increases very rapidly in both M and R.

There is considerable room for more work. First, methods of relaxing the strong model assumptions need to be developed. It may be possible to develop methods similar to QRR-VGLMs which give a similar answer for far less computation. Second, although R is relatively fast, improvements in speed would result if all the numerically intensive parts were written in a compiled language such as C or FORTRAN. (However, this may not be necessary because R may become a compiled language.) Third, there is need for more experience with QRR-VGLMs to be gained, e.g., simulation studies on its robustness and sensitivity to departures from the model. This could be along the lines of Minchin (1987*a*) and

Palmer (1993) who used COMPAS Minchin (1987*b*) to generate species abundances (see also Johnson and Altman 1999). Fourth, compositional data has not been considered in this article. For this, its relationship with the stereotype model (Anderson 1984) and RR-VGLMs will hopefully be given elsewhere. More importantly, the relationships between latent variable plots with QRR-VGLMs and biplots of RR-VGLMs and CCA is deferred to elsewhere. Fifth, this paper has been motivated by ecological theory to follow a model-driven approach by fitting Eq. 12; however, one important extension is to allow smoothing to determine the species response to ν_r rather than fitting the a priori quadratic functional form. This data-driven approach is motivated by the observation that there is substantial evidence for species responses being more complex than the Gaussian response, e.g., recent work by Bio et al. (1998) and Ejrnaes (2000). Such a data-driven approach would automatically allow the user to visualize each species' response as a smooth data-driven function of ν_r, cf. generalized additive models (Yee and Mitchell 1991) relative to GLMs. The need for checking the adequacy of the model by visual means has been demonstrated recently by Johnson and Altman (1999), who, for the rank-1 Poisson CCA model, proposed (LOWESS) smoothing y_{ij} vs. ν_{ir} in order to see whether a bell-shaped curve was justified.

Acknowledgments

The author wishes to thank Dr. C. ter Braak for helpful discussions and a copy of his books, and Brian McArdle for comments on the original manuscript. Both referees and the Corresponding Editor made many beneficial and perceptive comments that led to major improvements in the manuscript.

Literature Cited

Akaike, H. 1974. A new look at the statistical model indentification. IEEE Transactions on Automatic Control **19**: 716–723.

Anderson, J. A. 1984. Regression and ordered categorical variables (with discussion). Journal of the Royal Statistical Society, Series B, Methodological **46**:1–30.

Austin, M. P. 1985. Continuum concept, ordination methods, and niche theory. Annual Review of Ecological Systems **16**:39–61.

Bio, A. M. F., R. Alkemade, and A. Barendregt. 1998. Determining alternative models for vegetation response analysis: a non-parametric approach. Journal of Vegetation Science **9**:5–16.

Chambers, J. M. 1998. Programming with data: a guide to the S language. Springer-Verlag, New York, New York, USA.

Chambers, J. M., and T. J. Hastie, editors. 1993. Statistical models in S. Chapman and Hall, New York, New York, USA.

Dalgaard, P. 2002. Introductory statistics with R. Springer-Verlag, New York, New York, USA.

Ejrnaes, R. 2000. Can we trust gradients extracted by detrended correspondence analysis? Journal of Vegetation Science **11**:565–572.

Gauch, H. G. 1982. Multivariate analysis in community ecology. Cambridge University Press, Cambridge, UK.

Gauch, H. G., G. B. Chase, and R. H. Whittaker. 1974. Ordinations of vegetation samples by Gaussian species distributions. Ecology **55**:1382–1390.

Gauch, H. G., and R. H. Whittaker. 1972. Coenocline simulation. Ecology **53**:446–451.

Goodall, D. W., and R. W. Johnson. 1982. Non-linear ordination in several dimensions. Vegetatio **48**:197–208.

Green, P. J. 1984. Iteratively reweighted least squares for maximum likelihood estimation, and some robust and resistant alternatives. Journal of the Royal Statistical Society, Series B, Methodological **46**:149–192.

Guisan, A., S. B. Weiss, and A. D. Weiss. 1999. GLM versus CCA spatial modelling of plant species distribution. Plant Ecology **143**:107–122.

Johnson, K. W., and N. S. Altman. 1999. Canonical correspondence analysis as an approximation to Gaussian ordination. Environmetrics **10**:39–52.

Kooijman, S. A. L. M. 1977. Inference about dispersal patterns. Dissertation. University of Leiden, Leiden, The Netherlands.

Makarenkov, V., and P. Legendre. 2002. Nonlinear redundancy analysis and canonical correspondence analysis based on polynomial regression. Ecology **83**:1146–1161.

McCullagh, P., and J. A. Nelder. 1989. Generalized linear models. Second edition. Chapman and Hall, London, UK.

Minchin, P. R. 1987*a*. An evaluation of the relative robustness of techniques for ecological ordination. Vegetation **69**:89–107.

Minchin, P. R. 1987*b*. Simulation of multidimensional community patterns: towards a comprehensive model. Vegetatio **71**:145–156.

Nelder, J. A., and R. W. M. Wedderburn. 1972. Generalized linear models. Journal of the Royal Statistical Society, Series A, General **135**:370–384.

Palmer, M. 1993. Putting things in even better order: the advantages of canonical correspondence analysis. Ecology **74**:2215–2230.

Rousseeuw, P. J., and A. Christmann. 2003. Robustness against separations and outliers in logistic regression. Computational Statistics and Data Analysis **43**:315–332.

ter Braak, C. J. F. 1985. Correspondence analysis of incidence and abundance data: properties in terms of a unimodal response model. Biometrics **41**:859–873.

ter Braak, C. J. F. 1986. Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. Ecology **67**:1167–1179.

ter Braak, C. J. F. 1987*a*. The analysis of vegetation–environment relationships by canonical correspondence analysis. Vegetatio **69**:69–77.

ter Braak, C. J. F. 1987*b*. Ordination. Pages 99–173 *in* R. H. G. Jongman, C. J. F. ter Braak, and O. F. R. van Tongeren, editors. 1987. Data analysis in community and landscape ecology. Cambridge University Press, Cambridge, UK.

ter Braak, C. J. F. 1988. Partial canonical correspondence analysis. Pages 551–558 *in* H. H. Bock, editor. Classification and related methods of data analysis. North-Holland, Amsterdam, The Netherlands.

ter Braak, C. J. F., and C. W. N. Looman. 1986. Weighted averaging, logistic regression and the Gaussian response model. Vegetatio **65**:3–11.

ter Braak, C. J. F., and C. W. N. Looman. 1994. Biplots in reduced-rank regression. Biometrical Journal **36**:983–1003.

ter Braak, C. J. F., and I. C. Prentice. 1988. A theory of gradient analysis. Pages 271–317 *in* Advances in ecological research. Volume 18. Academic Press, London, UK.

ter Braak, C. J. F., and P. Šmilauer. 1998. CANOCO reference manual and user's guide to CANOCO for Windows. Software for canonical community ordination, version 4. Microcomputer Power, Ithaca, New York, USA.

Van der Aart, P. J. M., and N. Smeek-Enserink. 1975. Correlations between distributions of hunting spiders (Lycosidae, Ctenidae) and environmental characteristics in a dune area. Netherlands Journal of Zoology **25**:1–45.

Venables, W. N., and B. D. Ripley. 2002. Modern applied statistics with S. Fourth edition. Springer-Verlag, New York, New York, USA.

Wild, C. J., and T. W. Yee. 1996. Additive extensions to generalized estimating equation methods. Journal of the Royal Statistical Society, Series B, Methodological **58**: 711–725.

Yee, T. W., and T. J. Hastie. 2003. Reduced-rank vector generalized linear models. Statistical Modelling **3**:15–41.

Yee, T. W., and M. Mackenzie. 2002. Vector generalized additive models in plant ecology. Ecological Modelling **157**:141–156.

Yee, T. W., and N. D. Mitchell. 1991. Generalized additive models in plant ecology. Journal of Vegetation Science **2**: 587–602.

Yee, T. W., and C. J. Wild. 1996. Vector generalized additive models. Journal of the Royal Statistical Society, Series B, Methodological **58**:481–493.

APPENDIX A

Estimation of quadratic reduced-rank vector generalized linear models (QRR-VGLMs) is explained in ESA's Electronic Data Archive: *Ecological Archives* M074-016-A1.

APPENDIX B

A discussion of constraints-on-the-functions and the scaling parameter is presented in ESA's Electronic Data Archive: *Ecological Archives* M074-016-A2.