Autoregressive models for time series of random sums of positive variables:

application to tree growth as a function of climate and insect outbreak

Zinsou Max Debaly

ENSAI, Department of Statistics

With **Philippe Marchand** and **Miguel Montoro Girona**

INSTITUT DE RECHERCHE SUR LES FORÊTS, Canada.

Contact Information:
Email: zinsou-max.debaly@ensai.fr

Abstract

We present a broad class of semi-parametric models for time series of random sums of positive variables. Our methodology allows the number of terms inside the sum to be time-varying and is therefore well suited to many examples encountered in the natural sciences. We study the stability properties of the models and provide a valid statistical inference procedure to estimate the model parameters. It is shown that the proposed quasi-maximum likelihood estimator is consistent and asymptotically Gaussian distributed. This work is complemented by simulation results and applied to time series representing growth rates of white spruce (*Picea glauca*) trees from a few dozen sites in Québec (Canada). This time series spans 41 years, including one major spruce budworm (*Choristoneura fumiferana*) outbreak between 1968 and 1991.

1 Introduction

Many ecological studies require measuring the positive dependent variables of random numbers of statistical individuals sampled over time. This approach is often necessary, as 1) researchers cannot observe the entire population, and 2) the individuals observed by researchers depend on time-varying resources. Applications of this statistical approach include studies of species behaviour and ecological services. In this paper, we evaluate the impact of climate change and insect outbreak on tree growth as recorded by growth rings. Spruce budworm (*Choristoneura fumiferana*; SBW) is the most important defoliator of conifer trees in the eastern North American boreal forest. To do so, we present a class of semi-parametric autoregressive models and use them to investigate the relationships between climate, SBW outbreak, and the growth of white spruce.

2 Models and stability results

We denote by $Y_{k,t}, t \in \mathbb{Z}, k = 1, \dots, K$ the time series of the total basal area increment related to the $k-$th observational site, i.e., the sum of increases in the trunk cross-sectional area for the $n_{k,t}$ trees sampled for site k in year t. We aim to model the dynamics of this process both in terms of its own past and in the presence of m additional covariates $X_{k,t} \in \mathbb{R}^m$. In the empirical application presented in empirical section, the covariate process encompasses climate variables, including temperature and precipitation, and the level of SBW-related defoliation of the previous years.

Our model is given by

$$Y_{k,t} = \sum_{l=1}^{n_{k,t}} \zeta_{l,k,t}, \qquad (1)$$

where conditionally on $n_{k,t}, X_{k,t}, n_{k,t}^- = (n_{k,t-s}, s \geq 1)$ and $Y_{k,t}^- = (Y_{k,t-s}, s \geq 1)$, the variables $\zeta_{l,k,t},\ 1 \leq l \leq n_{k,t}$, which represent the basal area increments of individual sampled trees, are distributed identically as a random variable $\zeta_{k,t}$ of mean $\lambda_{k,t}$. Moreover, $(n_{k,t})_{t\in\mathbb{Z}}$ is a sequence of *i.i.d* random variables where, conditionally on $n_{k,t}^-$, the variable $n_{k,t}$ is independent from $X_{k,t}$ and $Y_{k,t}^-$. The mean process is given by

$$\varphi_\delta(\lambda_{k,t}) =: \eta_{k,t} = \omega_k + \sum_{j=1}^{p} \alpha_j \frac{Y_{k,t-j}}{n_{k,t-j}} + \beta^\top X_{k,t}, \qquad k = 1, \dots, K \text{ and } t = 1, \dots, T, \qquad (2)$$

such that $\omega_k \in \mathbb{R}, \alpha_j \in \mathbb{R}, \beta = (\beta_1, \dots, \beta_m) \in \mathbb{R}^m$, and φ_δ is a real-valued function defined on $\mathbb{R}_+$ that can depend on a parameter δ. We choose here $\varphi_\delta(x) = \log(\exp x - 1 - \delta)$. It is worth mentioning, without loss of generality, that the covariate process considered at time t is included in the specification of $\lambda_{k,t}$ because multiple lags of a given set of variables can be included by simply stacking them into a vector. An example is the case of defoliation levels, as shown in our application, as growth can be affected by defoliation occurring up to five years before the present (from $t-5$ to $t-1$).

Theorem 1. *Under some mild assumptions and if $\sum_{j=1}^p |\alpha_j| < 1$, there exists a unique set of K stationary, ergodic sequences $(Y_{k,t}, n_{k,t}, X_{k,t}), k = 1, \dots, K$ that are the solution of equations (1) and (2) with $\mathbb{E}|\eta_{k,0}| < \infty, k = 1, \dots, K$.*

3 Estimation and asymptotic properties

For our application, the K time series are observed between the time points 1 and T. We provide an asymptotic theory for the estimated parameters and present the results of a small simulation study investigating the finite-sample properties of the estimator. In the following section, we make $\lambda_{k,t}$ dependent on the parameter $\theta (\in \Theta$ a compact set); that is

$$\log(\exp \circ \lambda_{k,t}(\theta) - 1 - \delta) = \omega_k + \sum_{j=1}^{p} \alpha_j \frac{Y_{k,t-j}}{n_{k,t-j}} + \beta^\top X_{k,t} =: \eta_{k,t}(\theta), \qquad k = 1, \dots, K \text{ and } t = 1, \dots, T,$$

where $\delta \geq \delta_- > 0$. Let us denote the true, data-generating parameter value by θ_0.

The loss function from the exponential quasi-maximum likelihood is given by

$$r_T(\theta) = \sum_{k=1}^{K} T^{-1} \sum_{t=1}^{T} \left(\frac{Y_{k,t}}{\lambda_{k,t}(\theta)} + n_{k,t} \log \circ \lambda_{k,t}(\theta) \right) =: \sum_{k=1}^{K} T^{-1} \sum_{t=1}^{T} \ell_{k,t}(\theta) =: \sum_{k=1}^{K} \ell_k(\theta) \qquad (3)$$

and

$$\hat{\theta}_T = \underset{\theta \in \Theta}{\operatorname{argmin}}\ r_T(\theta). \qquad (4)$$

The derivative of $\lambda_{k,t}(\theta)$ with respect to θ is given by

$$\begin{aligned} \frac{\partial \lambda_{k,t}(\theta)}{\partial \theta} &=: \dot{\lambda}_{k,t}(\theta) \\ &= \left(\frac{1}{1+\delta+e^{\eta_{k,t}(\theta)}}, \frac{e^{\eta_{k,t}(\theta)}}{1+\delta+e^{\eta_{k,t}(\theta)}} \left(\iota_k, \frac{Y_{k,t-1}}{n_{k,t-1}} \dots \frac{Y_{k,t-p}}{n_{k,t-p}}, X_{k,t}^\top \right) \right)^\top. \end{aligned}$$

where ι_k is a vector of size K with 1 at the $k-$th position and 0 elsewhere. We will denote by $\dot{\lambda}_{k,t}$ (resp. $\lambda_{k,t}$), the vector $\dot{\lambda}_{k,t}(\theta)$ (resp. $\lambda_{k,t}(\theta)$), evaluated at the point $\theta = \theta_0$.

We will study the asymptotic properties of the QMLE estimator (4). To do so, we employ [5] (Thm 3.2.23), which was extended in [4]. The lemmas in our Appendix produce the general result for the asymptotic properties of QMLE (4). The following theorem represents the consistency and the asymptotic normality of (4) for the softplus_δ link function. Let us set

$$V_k = \mathbb{E}\left[\frac{1}{\lambda_{k,0}^2} \left(n_{k,0} - \frac{Y_{k,0}}{\lambda_{k,0}} \right)^2 \dot{\lambda}_{k,0} \dot{\lambda}_{k,0}^\top \right], and J_k = \mathbb{E}\left[n_{k,0} \frac{1}{\lambda_{k,0}^2} \dot{\lambda}_{k,0} \dot{\lambda}_{k,0}^\top \right].$$

Theorem 2. *Under mild regularity assumptions, almost surely,*

$$\lim_{T\to\infty} \hat{\theta}_T = \theta_0.$$

Coupled with some additional mild assumptions and θ_0 being located in the interior of Θ,

$$\lim_{T\to\infty} \sqrt{T}(\hat{\theta}_T - \theta_0) = \mathcal{N}(0, J^{-1} V J^{-\top}),$$

where $J = \sum_{k=1}^K J_k$ and $V = \sum_{k=1}^K V_k$.

4 Results

1. Higher defoliation levels led to reduced tree-ring growth, but this effect vanished after two years; however, note that while the direct effect vanished, expected growth remained lower in the successive years because of the large estimated first-order autocorrelation coefficient (0.8–0.9, depending on age class) ;
2. There was no significant effect of defoliation on the following year's growth for the youngest and oldest trees although it produced an effect two years following the defoliation. The results differed markedly for middle-aged trees, which were significantly affected one year after defoliation but not in the second year ;
3. High maximum temperatures in the summer increased growth, with up to a 5.6 cm² increase in basal area from a 10 °C increase in summer maximum temperature. However, the previous summer's temperature had a negative effect on growth. Finally, the spring CMI was negatively correlated with tree-ring growth, whereas the summer CMI had a positive effect ;
4. Both the CMI and precipitation in the previous spring increased the tree-ring growth of the current year: 100 mm greater precipitation led to at least a 6.8 cm² increase in basal area growth.

Conclusions

Here we developed a new time-series model to handle data having a time-varying number of sampled individuals. We provided a valid statistical inference procedure and applied the model to assessing the combined effect of climate and SBW outbreak on white spruce tree-ring growth in several sites in eastern Canada. We assumed a fixed number of ecological sites K. For future work, we plan to investigate the case of diverging K and the length n of observed series. Because many other ecological studies rely on binary variable or count data, it may be useful to extend the framework of this paper to these data types.

References

[1] Zinsou Max Debaly, Philippe Marchand, and Miguel Montoro Girona. Autoregressive models for time series of random sums of positive variables: application to tree growth as a function of climate and insect outbreaks, 2022, to appear in Ecological Modelling.

[2] Zinsou Max Debaly and Lionel Truquet. Iterations of dependent random maps and exogeneity in nonlinear dynamics. *Econometric Theory*, page 1–38, 2021.

[3] Zinsou Max Debaly and Lionel Truquet. Multivariate time series models for mixed data. *to appear in Bernoulli*, 202x.

[4] Lawrence A Klimko and Paul I Nelson. On conditional least squares estimation for stochastic processes. *The Annals of statistics*, pages 629–642, 1978.

[5] Masanobu Taniguchi and Yoshihide Kakizawa. *Asymptotic theory of statistical inference for time series*. Springer Science & Business Media, 2002.