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 [a,*], Philippe Marchand [b], Miguel Montoro Girona [b,c,d]

[a] *CREST-ENSAI, UMR CNRS 9194, 51 rue Blaise Pascal, BP 37203, 35172 Bruz cedex, France*
[b] *Institut De Recherche Sur Les Forêts, Université Rouyn-Noranda QC, J9X 5E4, Canada*
[c] *Groupe De Recherche En Ecologie De La MRC-Abitibi (GREMA), Canada*
[d] *Restoration Ecology Research Group, Department of Wildlife, Fish and Environmental Studies, Swedish University of Agricultural Sciences, Sweden*

ARTICLE INFO

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. We found significant growth reductions related to budworm-induced defoliation up to two years post-outbreak. Our results also revealed the positive effects of maximum summer temperature, precipitation, and the climate moisture index on white spruce growth. We also identified the negative effects of the climate moisture index in the spring and the maximum temperature of the previous summer. However, the model's performance on this data set was not improved when the interactions between climate and defoliation on growth were considered. This study represents a major advance in our understanding of budworm–climate–tree interactions and provides a useful tool to project the combined effects of climate and insect defoliation on tree growth in a context of greater frequency and severity of outbreaks coupled with the anticipated increases in temperature.

1. Introduction

Many ecological studies require measuring the positive dependent variables of random numbers of statistical individuals sampled over time (Girona et al., 2019). 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 forestry, for example, we can be interested in time series that represent the mass or size of a given tree species. We then randomly sample individual trees each year and observe the corresponding mass or volume, e.g., see Vourlitis et al. (2022). This approach is also applied to evaluate the area occupied by a species in relation to the available resources over time (Labrecque-Foy et al., 2020). In fisheries, we can use this approach to track temporal changes in the weight of fish caught, e.g., Chan et al. (2020).

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 (Girona et al., 2018). In the province of Québec (Canada), the forest area affected by this species of Lepidoptera over the last century covers more twice the size of the Ukraine (Navarro et al., 2018). At the epidemic stage, massive populations of larvae cause widespread damage to tree foliage (Lavoie et al., 2019). SBW affects the main conifer boreal species in Canada, including balsam fir (*Abies balsamea*), white spruce (*Picea glauca*), and black spruce (*Picea mariana*), resulting in a major impact on boreal forest regeneration and dynamics (Martin et al., 2020). Moreover, SBW outbreaks produce important economic consequences through the loss of forest productivity.

Previous works have studied the changes of forest composition following insect outbreaks, e.g., Morin et al. (2021), the response of SBW outbreaks to climate change, e.g, Fleming and Volney (1995) and Berguet et al. (2021), and demography, i.e., the rate of mortality

---

of spruce during outbreaks (Gauthier et al., 2015). However, despite the major implications of future climate change, we continue to have a limited understanding of the combined effects of SBW outbreak and climate change on tree growth. Given that temperature and its variations as well as the timing and amount of precipitation affect organisms' survival, reproduction cycles, and spatial dispersion (Aber et al., 2001), it is critical to understand the links between SBW outbreaks, climate, and tree growth to improve our understanding of the impacts of future climate change on forest productivity (Klapwijk et al., 2013). This concern is amplified by the expected increase in SBW outbreak severity and frequency under future climate scenarios (Navarro et al., 2018; Seidl et al., 2017).

In this paper, we contribute to filling this gap by proposing a broad class of semi-parametric models for positive-valued time series. Time-series data are common in forestry and the standard statistical approaches include descriptive exploratory techniques and linear mixed-effect models with time-varying variables on transformed data, e.g., Girona et al. (2016) and Boulanger and Arseneault (2004), and correlated error terms (Girardin et al., 2016). However, these approaches suffer from several drawbacks. Descriptive exploratory techniques do not allow the drawing of inferences from the data, and linear mixed-effect models, as shown in several papers, e.g., Chou et al. (2015), have demonstrated that specifying a linear model on transformed data often leads to a poor predictive performance. Applying a log transformation, for instance, can make the positive-valued data more normal; nonetheless, the obtained predicted value underestimates the expected value because of Jensen's inequality. Furthermore, models having autocorrelated error terms do not account for the complex, dependent structure of tree-ring growth. To provide a more robust and reliable approach, 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. We also discuss the advantages of applying a repeated-measures design.

Many previous studies have focused on modelling non-Gaussian time series, such as positive-valued processes. Gaussian processes can be represented as linear models, whereas time series of count or binary data are modelled by non-linear dynamics, e.g., Sim (1990), Weiß (2018), and Davis et al. (2016) and references therein. For positive-valued time series data, Engle and Russell (1998) proposed the range volatility model as an alternative to *GARCH* models in finance, and its use has rapidly expanded because of its diverse applicability. We refer the interested reader to the review by Chou et al. (2015). Recently, Aknouche and Francq (2020) considered a positive-valued time series whose conditional distribution had a time-varying mean dependent on exogenous variables. Our approach here differs slightly from theirs, as the positive process under consideration is itself the sum of a random number of other positive variables.

Our approach is driven strongly by the available data at hand, which consist of multiple time series collected from several sites, where the number of sampled individuals varies over time and between sites. Hence, considering an aggregate value such as the sum or the mean of growth rings can produce a loss of variability linked to the sampling scheme. Moreover, in fields such as finance, some modelling relies on considering empirical quantities such as realized volatility; historical returns of investment products within a defined period are then analysed, e.g., Allen et al. (2010). Unlike our framework, which is typical for ecological studies, all transactions on investment products are recorded, i.e., the entire statistical population is observed.

Our paper is organized as follows. In Section 2, we define the model used throughout this paper and discuss our modelling choice. The time-series properties of the models are also assessed in this section. We then present the maximum-likelihood based inference and its asymptotic properties in Section 3. Section 4 contains a small simulation study and an application to empirical data related to the growth of white spruce. All auxiliary lemmas and mathematical proofs are presented in Appendix.

2. Models and stability results

We present a generalized linear dynamic model for the time series of random sums of positive variables. This model relies on an empirical application where we analyse the annual growth of spruce trees subject to climate variation and SBW oubreak. We measured growth using tree cores collected at 1.30 m height from the trunks of individual trees in a stand (Girona et al., 2017). Sample preparation, measurement, and analysis conformed to standard dendroecological protocols (Krause and Morin, 1995). Cores were air-dried, mounted on wood boards, and then sanded. Tree-ring width was measured using WinDendro (Guay et al., 1992) or a manual Henson micrometer having an accuracy of 0.01 mm. The tree-ring series measurements covered the last 41 years and were cross-dated using TSAP-Wi (Rinntech, Heidelberg, Germany).

We denote by $Y_{k,t}, t \in \mathbb{Z}, k = 1, \ldots, K$ 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 Section 4, 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

$$Y_{k,t} = \sum_{l=1}^{n_{k,t}} \zeta_{l,k,t}, \tag{1}$$

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

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

$$t = 1, \ldots, T, \tag{2}$$

such that $\omega_k \in \mathbb{R}, \alpha_j \in \mathbb{R}, \beta = (\beta_1, \ldots, \beta_m) \in \mathbb{R}^m$, and φ_δ is a real-valued function defined on $\mathbb{R}_+$ that can depend on a parameter δ. 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 $t - 1$).

The variables $\zeta_{k,t}$ will be referred to as the unity random variables. We do not make any assumptions about the distribution of the variables $\zeta_{k,t}$. Any distribution on $(0, +\infty)$ can be chosen; examples include the exponential distribution having the parameter $1/\lambda_{k,t}$, the log-normal distribution having parameters $\log \lambda_{k,t} - \sigma^2/2$ and σ, and a Gamma distribution having the parameters $\alpha\lambda_{k,t}$ and α.

Regardless of the distribution of unity random variables, the conditional expectation of $Y_{k,t}$ is $n_{k,t}\lambda_{k,t}$. However under the assumption of the independence of $\zeta_{l,k,t}, 1 \leq l \leq n_{k,t}$, if they are exponentially distributed, the conditional variance is $n_{k,t}\lambda_{k,t}^2$, i.e., a quadratic function of $\lambda_{k,t}$. In our example of Gamma-distributed unity random variables, conditional variance is $n_{k,t}\lambda_{k,t}/\alpha$, i.e., a linear function of $\lambda_{k,t}$. In the case of a log-normal distribution, however, the conditional variance is $n_{k,t}\lambda_{k,t}^4(\exp\sigma^2 - 1)$. With our semi-parametric framework, we will only focus on the estimation of regression parameters $\theta = (\delta, \omega_1, \ldots, \omega_K, \alpha_1, \ldots, \alpha_p, \beta^\top)^\top$ without the need to perform any distributional goodness of fit tests.

Note 1 (*Copies of Unity Variables*). In our general setup, the copies $\zeta_{l,k,t}$, $1 \le l \le n_{k,t}$ of the unity random variables $\zeta_{k,t}$ are not required to be independent. In practical applications, where $\zeta_{l,k,t}$ represents, for example, the measure of annual growth for a sampled tree, the general assumption of being identically distributed can be thought as a local stationary condition inside site k at time t.

Note 2 (*Marginal Stationary Distributions*). Note from Eqs. (1) and (2) that $Y_{k,t} = f_{\theta_k}(X_{k,t-s}, n_{k,t-s}, \zeta_{\ell,k,t-s}, s \ge 0, \ell \ge 1)$ for $\theta_k = (\delta, \omega_k, \alpha_1, \ldots, \alpha_p, \beta^\top)^\top$. Then for $1 \le k_1 \ne k_2 \le K$, the distributions of $(Y_{k_1,0}, n_{k_1,0}, X_{k_1,0})$ and $(Y_{k_2,0}, n_{k_1,0}, X_{k_2,0})$ are not equal unless $\omega_{k_1} = \omega_{k_2}$ and $(X_{k_1,0}, \zeta_{\ell,k_1,0}, \ell \ge 1)$ is equal in distribution to $(X_{k_2,0}, \zeta_{\ell,k_2,0}, \ell \ge 1)$. In Section 4, we will investigate the consequences of the latter conditions on the proposed estimation procedure.

Note 3 (*Regression Function* (2)). Note that $\eta_{k,t}$ in (2) does not depend linearly on $Y_{k,t-i}, i = 1, \ldots, p$ but on $Y_{k,t-i}/n_{k,t-i}, i = 1, \ldots, p$. Through (2), we can link the underlying mean process and the empirical estimate of the past mean process. Even for a constant-size process, i.e., $n_{k,t} = n_k, \forall t$ because the regression parameter $\alpha_i, i = 1, \ldots, p$ is free of k, we still cannot yet express $\eta_{k,t}$ as a linear combination of $Y_{k,t-i}, i = 1, \ldots, p$. Moreover, one can expect $Y_{k,t-i}/n_{k,t-i} - \lambda_{k,t-i}, i = 1, \ldots, p$ or more generally $h(Y_{k,t-i}/n_{k,t-i}, \lambda_{k,t-i}), i = 1, \ldots, p$ for some mapping h such that $\mathbb{E}h(Y_{k,t-i}/n_{k,t-i}, \lambda_{k,t-i}) = 0$ in (2) at the place of $Y_{k,t-i}/n_{k,t-i}, i = 1, \ldots, p$. With the latter two mentioned specifications, the Model (1)–(2) defines the *GLARMA* model (see Weiß, 2018 for more details). In its present form, the Model (1)–(2) shares similarities with the well known *ARCH* (Bollerslev, 1986) and INARCH (Weiß, 2018) models. We leave the topic of *GLARMA* specification for a future work.

Note 4 (*Contrast with the Non-Linear Mixed Model*). The Model (1)–(2) has some similarities to well-known mixed models. Indeed, as for mixed models, ω_k represents the site fixed effect, where the random effect is embedded in the distribution of unity variables. The simple example of $\zeta_{l,k,t} = \lambda_{k,t}\epsilon_{l,k,t}$, where $(\epsilon_{l,k,t})_{l \ge 1}$ is a sequence of identically distributed random variables of mean 1, fit with the class of multiplicative random effect models (Cameron and Trivedi, 2013). Nonetheless more complex random effects can be handled. However, the Model (1)–(2) is more general because it allows the individuals sampled over time to change. Indeed, as we will see in Section 3, the individual measurements are no longer needed when the sequence $(Y_{k,t}, n_{k,t})$ is available. Moreover, in terms of the application to resource management, it is often of interest to model and predict a population quantity such as the sum of basal area growth in a forest.

Choice of the link function φ

The logarithmic link function is often applied and coincides with the well-known log-linear model, see for example Cameron and Trivedi (2013) for models of count data. This link function assumes a linear relationship between the logarithm of the mean process and the covariates. However, other link functions can preserve the linear correlation at least on the positive part of $\mathbb{R}$. Consider, for example, the threshold mapping $x \mapsto \max(x, 0)$. This mapping is not smooth and, most of the time, one makes some restrictions on model parameters to obtain the positiveness of the mean directly. Here we will apply the inverse of the softplus function as a link function. Indeed, the softplus function – see Glorot et al. (2011) – is interesting for two reasons. First, in regard to modelling, it preserves the linearity on the positive part of real line. As shown by Weiß et al. (2022) for count time-series modelling, the models defined with the softplus link function are quite close to the truly linear model. This is also relevant for our biological application, as we expect a linear effect of covariates on growth above a certain threshold representing the minimal favourable conditions for growth. The minimum growth expected may not be exactly zero, which is why
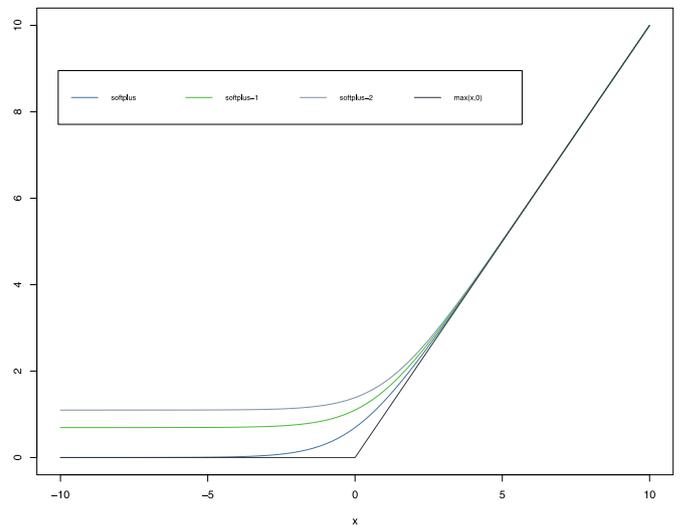

Fig. 1. Comparison between softplus and max(x,0).

we will later consider a slightly different version of softplus, referred to as $\mathrm{softplus}_\delta$ for $\delta > 0$, defined as $\mathrm{softplus}_\delta(x) = \log(1 + \delta + \exp(x))$. Second, and also a technical advantage, is that the mapping $\mathrm{softplus}_\delta$ is infinitely differentiable. Fig. 1 in the Appendix illustrates the difference between the $\mathrm{softplus}_\delta$ link function and $\max(x, 0)$, where $\mathrm{softplus}$ represents $\mathrm{softplus}_0$. One can note that $\mathrm{softplus}_\delta$ is lower bounded by $\log(1 + \delta)$. It is also worth noting that our generalization of the softplus function differs from that of Mei and Eisner (2017) and Weiß et al. (2022) (Eq. 3.2). As noted by the latter, it is possible to mimic the behaviour of the Tobit model with the softplus generalization of Mei and Eisner (2017). In contrast, here we aim to lower bound the softplus link function by a non-zero constant because we require a minimum basal area increment at any time.

Note 5 (*Model Interpretation*). Obviously with the $\mathrm{softplus}_\delta$ link function, the mean process increases with the j−th covariate process if $\beta_j > 0$ and decreases when $\beta_j < 0$. Because $\mathrm{softplus}_\delta(x) \sim_\infty x$, the mean process can be approximated by identity mapping. Therefore all other things remaining equal, the regression function is similar to $\beta_j X_{j,t}$ for large values of $X_{j,t}$ and $\beta_j > 0$ and then increases by $\beta_j \alpha$ for increasing values α of $X_{j,t}$. Let us denote by $\mathrm{RG}(x, y)$ the relative rate of growth of the mean process between x and y, i.e., $\mathrm{RG}_\delta(x, y) = \gamma_\delta(x)/\gamma_\delta(y)$, where γ_δ is the derivative function of $\mathrm{softplus}_\delta$. For $\beta < 0$, $\lim_{x \to \infty} \mathrm{RG}_\delta(\beta(x + \alpha), \beta x) = e^{\beta\alpha}$. Therefore, the rate towards $\log(1 + \delta)$ driven by $X_{j,t}$ is given by $e^{\beta_j \alpha}$ when $\beta_j < 0$. Moreover, when $\delta \sim 0$, by l'Hôpital's rule $\lim_{x \to \infty} \mathrm{RG}_0(x, y) = \lim_{x \to \infty} \mathrm{softplus}_0(\beta(x + \alpha))/\mathrm{softplus}_0(\beta x) = e^{\beta\alpha}$. Therefore, all other things remaining equal, the mean process will be divided by $e^{-\beta\alpha}$ when $X_{j,t}$ increases by α for large values of $X_{j,t}$ and $\beta_j < 0$.

Theorem 1 provides some stability conditions of Model (1)–(2) with the inverse of the softplus function as the link, whereas Lemma 1 in the Appendix represents a general result for φ.

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

3. Estimation and asymptotic properties

This section is devoted to the estimation of the conditional mean parameters by the quasi-maximum likelihood estimator (QMLE) based on a member of the exponential family. We consider the exponential

QMLE (EQMLE) because this estimator coincides with the maximum likelihood estimator (MLE) when the unity random variables follow the exponential $\Gamma(1, \lambda_{k,t}^{-1})$ distribution, and the copies $\zeta_{l,k,t}$, $1 \le l \le n_{k,t}$ are independent.

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

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

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

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

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

and

$$\hat{\theta}_T = \underset{\theta \in \Theta}{\arg\min} \ r_T(\theta). \tag{4}$$

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

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

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 Taniguchi and Kakizawa (2002) (Thm 3.2.23), which was extended in Klimko and Nelson (1978). 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$_\delta$ 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. *Suppose that assumptions (C.1)–(C.4) in the Appendix are met. Then, almost surely,*

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

If (AN.1)–(AN.3) hold true and θ_0 is 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. Application

4.1. Simulation

We examined the finite-sample performance of the QMLE presented in the previous section through a small simulation study. We present the results for QMLE under two different data-generating processes, here referred to as *Scenario 1* and *Scenario 2*, with $m = 10$ co-variates. First, $X_{k,t}$ does not depend on k and is a sequence of *i.i.d*

random variables distributed as exponential random variables with means $\lambda_1, \dots, \lambda_m$. In the second, for a fixed k, $X_{k,t}$ is sampled independently from exponential distributions of mean $0.4k\lambda_1, \dots, 0.4k\lambda_m$. For the two data-generating processes, for a fixed k, the process $(n_{k,t})_{t \ge 1}$ is independently sampled from a Poisson distribution of mean τ_k as follows: for a fixed $K, \tau_1, \dots, \tau_K$ are independent and distributed as an exponential random variable of mean K. Moreover, we take $p = 1$ and $\delta = 0.5, \beta = (0, 1, -1, 0.5, -0.5, -1.5, 1.5, -2, 2, 0), \alpha_1 = 0.6$, and $\omega_1, \dots, \omega_K$ independently and uniformly sampling in the range $(-0.5K, 0.5K)$ for a fixed K. We sequentially choose $K = 5, 10, 15, 20$ and $T = 50, 100$. The samples are nested, i.e., the sample for the first scenario and $K = 5, T = 50$ is a subset of that of $K = 5, T = 100$. Indeed, our aim here is to evaluate the consequences of increasing K and T on the performance of our estimator. For each sample, we compute the estimator (4) and the corresponding theoretical standard errors (TSE) given by the Gaussian limit distribution. We replicate $B = 100$ times the experiment. Table 1 presents our simulation results. The line EQML refers to the average estimated value of the parameters, and TSE refers to the average value of estimated theoretical standard errors:

$$\text{EQML} = B^{-1} \sum_{b=1}^{B} \hat{\theta}_T^{(b)}, \text{ and } \text{TSE} = B^{-1} \sum_{b=1}^{B} \text{diag}\left\{ \hat{J}^{-1(b)} \hat{V}^{(b)} \hat{J}^{-\top(b)} \right\}^{1/2},$$

where the superscript b represents the index of replication, and $\text{diag} M$ for a matrix M is the diagonal elements of M. It appears that the model parameters are well estimated, except for $\omega_k, k = 1, \dots, K$ when K is very small relative to T, which coincides here with $K = 5, T = 50, 100$. We leave deep simulation studies for a future study.

4.2. Application to white spruce growth series

Dendrochronology, i.e., the study of the time series of tree rings, is a powerful tool for reconstructing past natural and anthropic disturbances (Girona et al., 2016; Boulanger and Arseneault, 2004; Labrecque-Foy et al., 2020). Tree rings represent natural hard disks that record environmental changes and thereby offer the potential to understand the evolution of complex natural phenomena over time, such as disturbances. Dendrochronological data have provided a better understanding of insect outbreak dynamics (Navarro et al., 2018; Camarero et al., 2003; Speer and Kulakowski, 2017).

Here we used the dendroecological series from Jardon et al. (2003), which includes annual tree-ring width measurements for 631 white spruce (*Picea glauca*) trees distributed across 45 sites in southwestern Quebec, Canada, with 1 to 23 trees per site. These time series comprise between 63 and 247 rings. We converted the ring-width increments to basal-area increments (BAI) using the full series; however, because of covariate availability, we limited our analysis to the AD 1955–1995 period (41 years) to study only a single insect outbreak event (see Fig. 2 in the Appendix).

We interpolated climate variables at the study sites for these 41 years using BioSIM (Régnière et al., 2014), a software package that interpolates daily climate station data on the basis of latitudinal and elevational climate gradients and the spatial correlations estimated from 30-year climate normals. We computed the following climate summaries from daily data for the spring (April–June) and summer (July–September) seasons separately: mean of daily maximum temperatures, total precipitation, and the climate moisture index (CMI) equal to the difference between precipitation and potential evapotranspiration (PET). Daily PET values were estimated using the Penman–Monteith equation as implemented in the SPEI package (Beguería and Vicente-Serrano, 2017) in R, on the basis of BioSIM-interpolated values of the minimum and maximum temperature, wind speed at 2 m, solar radiation, dew point temperature, and atmospheric pressure, using the "tall" crop model in SPEI.

One major SBW outbreak occurred in Quebec during the study period, spanning from 1967 to 1991. We obtained annual estimates of the severity of the SBW outbreak at the location of each study

Table 1
Estimation results for the quasi-maximum likelihood estimation.

| K	T	Scenario		α_1	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	β_{10}
				0.6	0	1	−1	0.5	−0.5	−1.5	1.5	−2	2	0
5	50	1	EQMLE	0.567	−0.018	0.985	−1.051	0.631	−0.247	−1.565	1.445	−2.165	1.673	0.011
			TSE	0.060	0.086	0.185	0.286	0.178	0.309	0.332	0.219	0.298	0.211	0.104
		2	EQMLE	0.517	0.019	0.862	−0.772	0.580	−0.493	−1.345	1.138	−1.709	1.908	−0.015
			TSE	0.053	0.097	0.176	0.217	0.193	0.252	0.313	0.214	0.280	0.211	0.104
	100	1	EQMLE	0.362	0.035	0.743	−0.632	0.457	−0.324	−1.219	1.099	−1.655	1.521	−0.080
			TSE	0.065	0.101	0.169	0.303	0.218	0.379	0.401	0.219	0.333	0.232	0.093
		2	EQMLE	0.361	−0.027	0.656	−0.678	0.294	−0.411	−0.956	1.148	−1.237	1.397	0.097
			TSE	0.085	0.114	0.215	0.362	0.232	0.388	0.450	0.268	0.376	0.271	0.121
10	50	1	EQMLE	0.304	−0.048	0.600	−0.487	0.476	−0.460	−0.844	1.012	−1.048	1.311	0.023
			TSE	0.043	0.063	0.122	0.177	0.134	0.199	0.229	0.148	0.146	0.147	0.069
		2	EQMLE	0.311	0.098	0.550	−0.547	0.243	−0.486	−0.762	1.015	−1.233	1.061	0.061
			TSE	0.046	0.084	0.153	0.220	0.175	0.258	0.256	0.183	0.235	0.174	0.079
	100	1	EQMLE	0.328	0.001	0.686	−0.507	0.222	−0.293	−0.838	0.903	−1.091	1.173	0.050
			TSE	0.035	0.056	0.112	0.153	0.107	0.184	0.164	0.108	0.133	0.119	0.059
		2	EQMLE	0.285	0.033	0.676	−0.651	0.258	−0.072	−0.603	0.826	−1.199	1.101	0.042
			TSE	0.039	0.063	0.142	0.175	0.119	0.249	0.205	0.158	0.160	0.160	0.076
15	50	1	EQMLE	0.546	−0.004	0.865	−0.734	0.462	−0.467	−1.418	1.280	−1.854	1.874	0.010
			TSE	0.041	0.069	0.124	0.186	0.125	0.220	0.219	0.140	0.222	0.156	0.062
		2	EQMLE	0.531	0.002	0.894	−0.985	0.405	−0.614	−1.342	1.228	−1.845	1.692	0.0163
			TSE	0.039	0.060	0.119	0.165	0.120	0.190	0.195	0.138	0.180	0.144	0.057
	100	1	EQMLE	0.384	−0.014	0.816	−0.549	0.160	−0.546	−1.226	0.987	−1.314	1.447	0.044
			TSE	0.028	0.040	0.075	0.105	0.080	0.147	0.134	0.096	0.115	0.088	0.042
		2	EQMLE	0.387	0.003	0.740	−0.675	0.471	−0.356	−0.751	1.104	−1.540	1.394	0.058
			TSE	0.053	0.077	0.160	0.230	0.151	0.242	0.273	0.166	0.241	0.168	0.078
20	50	1	EQMLE	0.370	0.018	0.613	−0.612	0.316	−0.337	−0.787	0.915	−1.223	1.277	0.003
			TSE	0.031	0.048	0.092	0.118	0.094	0.151	0.145	0.102	0.115	0.097	0.046
		2	EQMLE	0.369	−0.004	0.616	−0.745	0.350	−0.265	−1.022	0.949	−1.266	1.282	−0.010
			TSE	0.033	0.063	0.116	0.174	0.137	0.228	0.205	0.147	0.186	0.142	0.052
	100	1	EQMLE	0.339	0.002	0.534	−0.491	0.282	−0.393	−0.802	0.862	−1.183	1.199	0.021
			TSE	0.024	0.042	0.076	0.116	0.085	0.138	0.130	0.083	0.098	0.085	0.044
		2	EQMLE	0.311	−0.017	0.636	−0.524	0.294	−0.173	−0.996	1.016	−1.144	1.192	0.039
			TSE	0.025	0.052	0.104	0.128	0.109	0.190	0.151	0.131	0.130	0.122	0.055
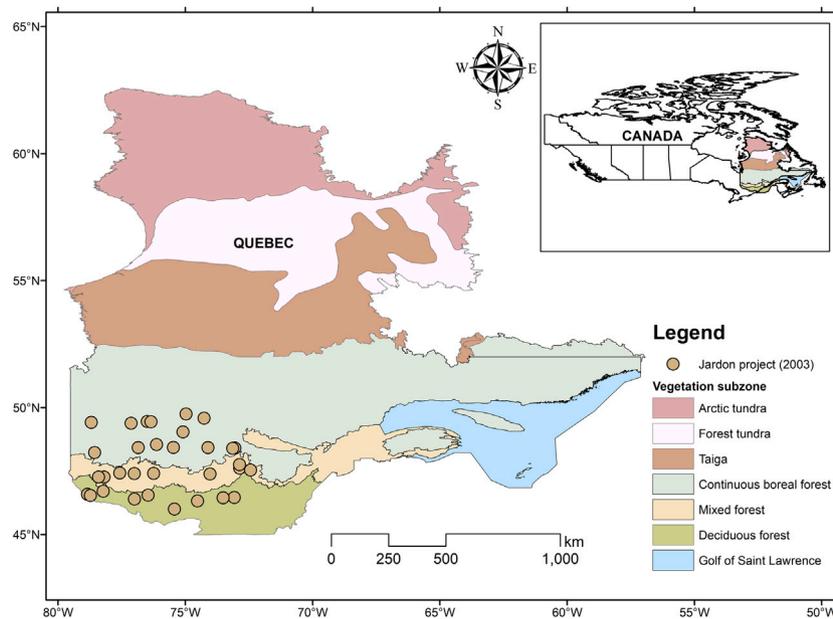
Fig. 2. Location of study sites from Jardon et al. (2003) in Québec (Canada) and the vegetation subzones within the province.

site through defoliation maps produced by the Quebec Ministry of Forests, Wildlife and Parks (MFFP, 2019). These maps are digitized versions of hand-drawn outlines of defoliated areas produced by aerial surveys of the affected regions. The defoliation level for each area is classified on a scale of 1 to 3 corresponding to a low (approx. 1%–35%), moderate (36%–70%) or high (71%–100%) fraction of the year's foliage defoliated by SBW. We note that these defoliation levels mainly reflect the status of balsam fir (*Abies balsamea*), which is the main

SBW host and is generally more severely affected than white spruce. Therefore, these defoliation levels are a proxy for outbreak severity, i.e., the potential herbivory pressure exerted by budworm on white spruce at the site.

Because tree growth and its vulnerability to both climate and defoliation depend on tree age, we split the data set and fit our models separately for the five age classes of ≤75, 75–100, 100–125, 125–150, and ≥150 years. We included as covariates the mean daily maximum of temperature, the total precipitation, and the mean CMI for the current and previous spring and summer. Only one of either precipitation and CMI appeared in a given model version because of the correlation between these two variables. We also included as covariates the defoliation levels for the five preceding years, a delay that estimates the time needed to fully regrow the lost foliage after an outbreak. Note that we do not expect defoliation to have a marked effect on the same year's growth ring (Krause et al., 2003). Finally, we considered models having interaction effects of the preceding year's defoliation level and climate variables, representing the possibility that climate conditions can increase or decrease a tree's sensitivity to SBW outbreaks.

Data processing and analyses were performed in R (R Core Team, 2021) with the package dplR (Bunn, 2008) used to process tree-ring data. We minimized the criterion (3) with the R command *nlm* (Dennis and Schnabel, 1983). All developed software is available under the Creative Commons (CC) license (see data availability statements). We used the QAIC criterion for selecting the model. The primary analysis based on partial autocorrelation plots led us to select $p = 1$.

According to the QAIC, the best models were those lacking an interaction between climate and defoliation. Our model results (Figs. 3 and 4) revealed that 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). Moreover, 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.

For the climate variables, 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. However, 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.

It is worth mentioning that D'Orangeville et al. (2018) reported a positive effect of temperature on tree-ring growth and highlighted its transitory nature. D'Orangeville et al. (2018) found high temperatures to be unfavourable for ring growth. Our findings therefore differ slightly, as we observed a negative long-term effect of temperature rather than a reduced ring growth because of extreme temperature values. Moreover, Walker et al. (2015) reported a negative radial growth response to temperature and a positive response to precipitation and CMI. In constrast to these studies, our approach here permitted a much closer inspection of the effects, and we found a more complex relationship between climate and tree-ring growth.

5. 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.

Appendix

A.1. Proofs for the main results

Throughout this section, we will denote by $\zeta_{k,t}^{\infty} = (\zeta_{k,t,l})_{l \geq 1}$, the sequence of copies of the unity random variables $\zeta_{k,t}$. Moreover $\zeta_{k,t}$ can be decomposed into two components: its mean $\lambda_{k,t}$ function of $X_{k,t}$ and a free random variable ζ_t. For example, $\zeta_{k,t} = \lambda_{k,t}\zeta_t$ for a positive random variable ζ_t of mean 1. We will write $\zeta_{k,t} := \zeta_{k,t}(\lambda_{k,t}, \zeta_t)$ to denote the relationship between $\zeta_{k,t}$ and $\lambda_{k,t}$ and ζ_t. Accordingly, $\zeta_{k,t,l} = \lambda_{k,t}\zeta_{t,l}$ with $\zeta_{t,l}, l \geq 1$ *i.i.d* with a mean of 1 or in general $\zeta_{k,t,l} := \zeta_{k,t,l}(\lambda_{k,t}, \zeta_{t,l})$ with $\mathbb{E}\zeta_{k,t,l} = \lambda_{k,t}$. Let $\mathcal{F}_{k,t}$ denote the σ-algebra generated by $\zeta_s, X_{k,s+1}, s \leq t$ and $\mathcal{F}_{k,t,n}$ generated by $n_{k,s}, \zeta_s, X_{k,s+1}, s \leq t$. Finally, we will denote by ϕ_{δ} the inverse of φ_{δ} : $\phi_{\delta}(x) = \varphi_{\delta}^{-1}(x)$. For stability, we will consider the following set of assumptions:

(A.1) The function ϕ_{δ} is υ–Lipschitz, and $\upsilon \sum_{i=1}^{p} |\alpha_i| < 1$.

(ST.1) For $k = 1, \ldots, K, (n_{k,t-1}, \zeta_{k,t-1}^{\infty}, X_{k,t})_{t \in \mathbb{Z}}$ is stationary, ergodic, $(n_{k,t}, \zeta_{k,t}^{\infty})$ is independent from $\mathcal{F}_{k,t-1,n}$, and $\mathbb{E}|X_{k,0}|_1 < \infty$.

(ST.2) For $k = 1, \ldots, K$,

$$E(|\zeta_{k,t}(\lambda_{k,t}, \zeta_t) - \zeta_{k,t}(\overline{\lambda}_{k,t}, \zeta_t)||\mathcal{F}_{k,t-1,n}) \leq |\lambda_{k,t} - \overline{\lambda}_{k,t}|.$$

It is worth noting that the example $\zeta_{k,t} = \lambda_{k,t}\zeta_t$ for a positive random variable ζ_k of mean 1 verifies condition (ST.2).

Lemma 1. *Under the assumptions (A.1)–(ST.2), there exists a unique set of K stationary, ergodic sequences $(Y_{k,t}, n_{k,t}, X_{k,t}), k = 1, \ldots, K$ that are a solution of Eqs. (1) and (2) with $\mathbb{E}|\eta_{k,0}| < \infty, k = 1, \ldots, K$.*
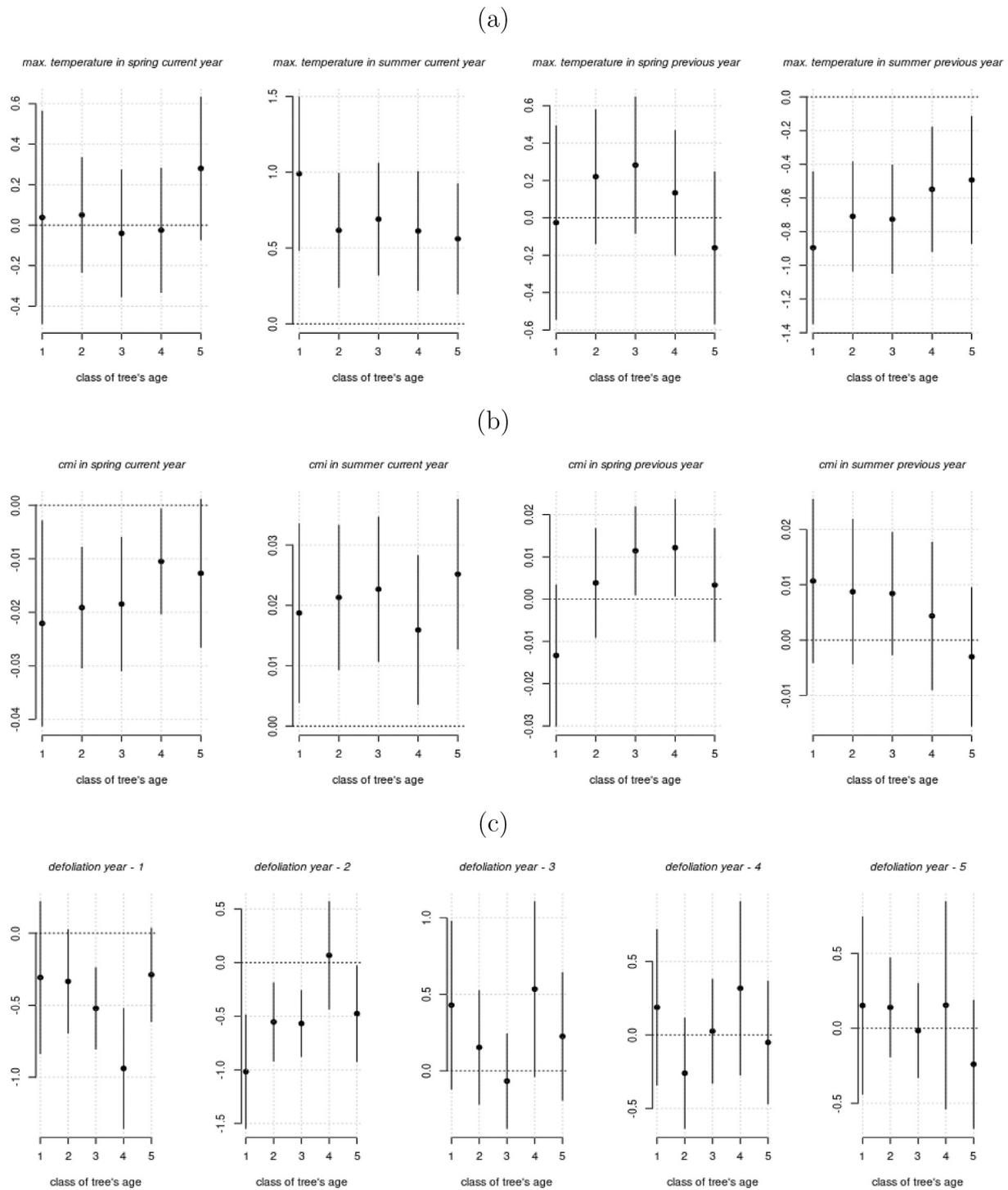
$$(a)$$

$$(b)$$

$$(c)$$

Fig. 3. Model with Temperature + CMI + Defoliation. Age classes: 1, ≤75; 2, 75–100; 3, 100–125; 4, 125–150, and 5, ≥150 years; (a) effects of maximum temperature in spring and summer on current and preceding years; (b) effects of the CMI index in spring and summer on the current and preceding years; and (c) the delayed effect of defoliation level. The dashed horizontal line corresponds to zero.

The proof of Lemma 1 uses the techniques of iterated random maps. We refer interested readers to Debaly and Truquet (2021a) theorems 2 and 4, which investigated the problem of solving recursive stochastic equations with covariates or Debaly and Truquet (2021b) in the case where no covariates are included in the dynamic.

Proof of Lemma 1. From (2),

$$\eta_{k,t} = \omega_k + \sum_{j=1}^{p} \alpha_j \frac{1}{n_{k,t-j}} \sum_{\ell=1}^{n_{k,t-j}} \zeta_{k,t-j,\ell}(\phi_\delta(\eta_{k,t-j}), \zeta_{t-j,\ell}) + \beta^\top X_{k,t}.$$

Then under condition (ST.1), the processes $(\eta_{k,t} = \varphi_\delta(\lambda_{k,t})_{t\in\mathbb{Z}}, k = 1, \ldots, K$ obey some recursive stochastic equations,

$$\eta_{k,t} = f(\eta_{k,t-1}, \ldots, \eta_{k,t-p}; n_{k,t-1}, \ldots, n_{k,t-p}, \zeta_{k,t-1}^\infty, \ldots, \zeta_{k,t-p}^\infty, X_{k,t}).$$

And with (A.1), for $k = 1, \ldots, K, (x, y) \in \mathbb{R}^{2p}$,

$$E(|f(x; n_{k,t-1}, \ldots, n_{k,t-p}, \zeta_{k,t-1}^\infty, \ldots, \zeta_{k,t-p}^\infty, X_{k,t})$$
$$- f(y; n_{k,t-1}, \ldots, n_{k,t-p}, \zeta_{k,t-1}^\infty, \ldots, \zeta_{k,t-p}^\infty, X_{k,t})||\mathcal{F}_{k,t-1,n}) \leq v\alpha^\top|x - y|$$

(a)

(b)

(c)

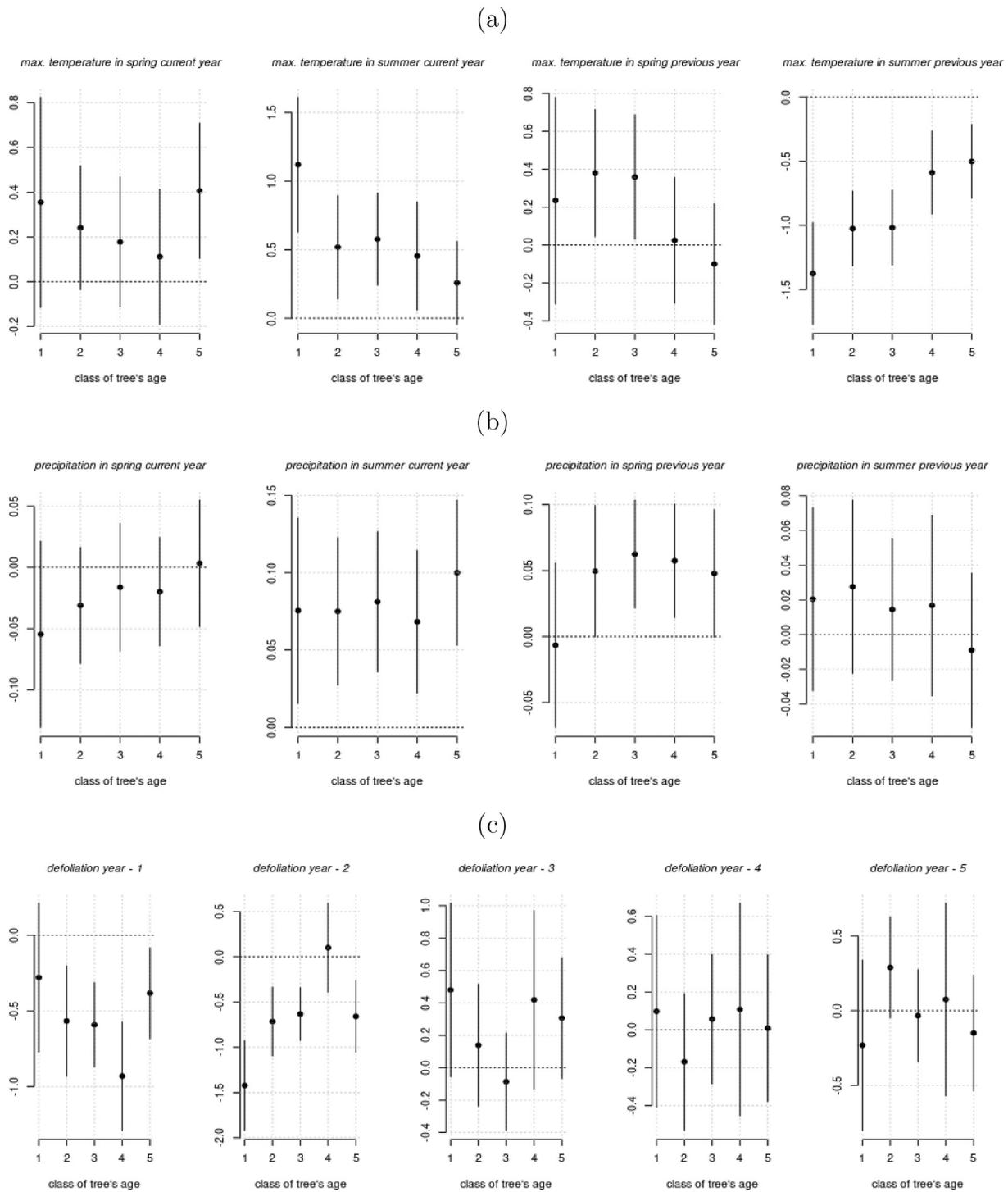**Fig. 4.** Model with Temperature + Precipitation + Defoliation. Age classes: 1, ≤75; 2, 75–100; 3, 100–125; 4, 125–150, and 5, ≥150 years; (a) effects of maximum temperature in spring and summer on current and preceding years; (b) effects of the precipitation index in spring and summer on the current and preceding years; and (c) delayed effect of defoliation level. The dashed horizontal line corresponds to zero.

with $\alpha = (\alpha_1, \ldots, \alpha_p)$. Moreover,

$E|f(x; n_{k,-1}, \ldots, n_{k,-p}, \zeta_{k,-1}^\infty, \ldots, \zeta_{k,-p}^\infty, X_{k,0})| < \infty$. Then, from Debaly and Truquet (2021a) Theorem 4, we obtain the stationary and ergodic solution with $\mathbb{E}|\eta_{k,0}| < \infty, k = 1, \ldots, K.\square$

Theorem 1 is a straight consequence of Lemma 1 and follows the Lipschitz property of $x \mapsto \log(\exp(x) + 1 + \delta)$ for any $\delta > 0$. For the asymptotic results for $\hat{\theta}_T$, the following assumptions are necessary:

(A.2) The conditions (A.1) and (ST.1) are met, and θ_0 verifies: $\upsilon \sum_{i=1}^p |\alpha_{i,0}| < 1$.

(A.3) For $k = 1, \ldots, K, \mathbb{E}n_{k,0} < \infty$ and

$$\mathbb{E}\sup_\theta \left(\frac{\phi_\delta(\eta_{k,0}(\theta_0))}{\phi_\delta(\eta_{k,0}(\theta))} + |\log \circ \phi_\delta(\eta_{k,0}(\theta))| \right) < \infty.$$

(A.4) For $(\delta, \bar{\delta}) \in [\delta_-, \infty)^2, (\eta, \bar{\eta}) \in \mathbb{R}^2,$

$$\phi_\delta(\eta) = \phi_{\bar{\delta}}(\bar{\eta}) \Rightarrow (\delta = \bar{\delta}, \eta = \bar{\eta}).$$

(C.1) The conditions (ST.1) and (ST.2) are met, and θ_0 verifies: $\sum_{i=1}^{p} |\alpha_{i,0}| < 1$.

(C.2) For $k = 1, \ldots, K$, $\mathbb{E} n_{k,0} < \infty$.

(C.3) For $k = 1, \ldots, K$, conditionally on $X_{k,0}$, the distribution of $\left(\frac{Y_{k,-1}}{n_{k-1}}, \ldots, \frac{Y_{k,-p}}{n_{k-p}} \right)$ is not supported by an hyperplan of $\mathbb{R}^p$.

(C.4) For $k = 1, \ldots, K$, the distribution of $X_{k,0}$ is not degenerate.

Lemma 2. *Let us suppose that the assumptions (A.2)–(A.4) and (C.3)–(C.4) are met. Then, almost surely,*

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

We do not prove Lemma 2. Similar results for time-series models can be found in Diop and Kengne (2021), Aknouche and Francq (2020), and Debaly and Truquet (2022) among others.

Proof of consistency part of Theorem 2. We will check (A.2) to (A.4).

- (A.2) comes from (C.1).
- One can note that here $\phi_\delta(x) = \log(1 + \delta + \exp(x))$, and $\phi_\delta(x) \geq \log(1 + \delta)$, $\phi_\delta(x) \leq \kappa_1(\theta)(1 + |x|)$, and $|\log \circ \phi_\delta(x)| \leq \kappa_2(\theta)(1 + |x|) + \kappa_3(\theta)$, where $\kappa_i, i = 1, 2, 3$ are continuous functions of θ. Then (A.3) holds because $\mathbb{E} \sup_\theta |\eta_{k,0}(\theta)| < \infty$. Indeed $\mathbb{E} Y_{k,0}/n_{k,0} = \phi_\delta(\eta_{k,0}) < \infty$ because $\mathbb{E} |\eta_{k,0}| < \infty$.
- For (A.4), we note that

$$\phi_\delta(\eta) = \phi_{\bar{\delta}}(\bar{\eta}) \Rightarrow \delta - \bar{\delta} = \exp \bar{\eta} - \exp \eta,$$

and $0 = \lim_{\eta \to -\infty, \bar{\eta} \to -\infty} \exp \bar{\eta} - \exp \eta = \delta - \bar{\delta}$. Then $\delta = \bar{\delta}$, and $\eta = \bar{\eta}$. $\square$

Let us set $\sigma_{k,0}^2 = \text{Var} \left(\frac{Y_{k,0}}{\lambda_{k,0}} | \mathcal{F}_{k,-1,n} \vee n_{k,0} \right)$, $\partial_\delta \phi_\delta$ the derivative of ϕ_δ with respect to δ, and $\theta_{-\delta}$ the vector of parameters without δ. We will consider the following assumptions for the asymptotic distribution of $\hat{\theta}_T$.

(A.1) The function ϕ_δ is twice continuously differentiable and for $k = 1, \ldots, K$,

$$\mathbb{E} \frac{\sigma_{k,0}^2}{\phi_\delta^2(\eta_{k,0})} \left[\partial_\delta \phi_\delta(\eta_{k,0}(\theta_0))^2 + \phi_\delta'(\eta_{k,0}(\theta_0))^2 \| \nabla_{\theta_{-\delta}} \eta_{k,0}(\theta_0) \|_2^2 \right] < \infty, \text{ and}$$

$$\mathbb{E} \frac{1}{\phi_\delta^2(\eta_{k,0})} \left[\partial_\delta \phi_\delta(\eta_{k,0}(\theta_0))^2 + \phi_\delta'(\eta_{k,0}(\theta_0))^2 \| \nabla_{\theta_{-\delta}} \eta_{k,0}(\theta_0) \|_2^2 \right] < \infty.$$

(A.2) For $k = 1, \ldots, K$, the distribution of $(\partial_\delta \phi_\delta(\eta_{k,0}(\theta_0)), \phi_\delta'(\eta_{k,0}(\theta_0)) \nabla_{\theta_{-\delta}} \eta_{k,0}(\theta_0))$ is not degenarate.

(A.3) For $k = 1, \ldots, K$, $\mathbb{E} \sup_\theta |W_{k,0}^{i,j}(\theta)| < \infty$, where $W_{k,0}^{i,j}(\theta)$ is one of the following quantities for all pairs i, j.

$$\frac{1}{\phi_\delta^2(\eta_{k,0}(\theta))} \left(\frac{\phi_\delta(\eta_{k,0}(\theta_0))}{\phi_\delta(\eta_{k,0}(\theta))} + 1 \right) \frac{\partial \phi_\delta(\eta_{k,0}(\theta))}{\partial \theta_i} \frac{\partial \phi_\delta(\eta_{k,0}(\theta))}{\partial \theta_j},$$

$$\frac{1}{\phi_\delta^2(\eta_{k,0}(\theta))} \frac{\phi_\delta(\eta_{k,0}(\theta_0))}{\phi_\delta(\eta_{k,0}(\theta))} \frac{\partial \phi_\delta(\eta_{k,0}(\theta))}{\partial \theta_i} \frac{\partial \phi_\delta(\eta_{k,0}(\theta))}{\partial \theta_j},$$

$$\frac{1}{\phi_\delta(\eta_{k,0}(\theta))} \left(\frac{\phi_\delta(\eta_{k,0}(\theta_0))}{\phi_\delta(\eta_{k,0}(\theta))} + 1 \right) \frac{\partial^2 \phi_\delta(\eta_{k,0}(\theta))}{\partial \theta_i \partial \theta_j}.$$

(AN.1) The K stationary sequences solution of (1)–(2) are independent of each other.

(AN.2) For $k = 1, \ldots, K$, $\mathbb{E} n_{k,0}^2 < \infty$,

$$\mathbb{E} \sigma_{k,0}^4 < \infty.$$

(AN.3) For $k = 1, \ldots, K$,

$$\mathbb{E} |X_{k,0}|_1^4 < \infty, \text{ and} \mathbb{E} Y_{k,0}^4 < \infty.$$

Lemma 3. *Under the assumptions of Lemma 1, and if (A.1)–(A.3) and (AN.1) hold, then*

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

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

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

$$\dot{\lambda}_{k,0} = (\partial_\delta \phi_\delta(\eta_{k,0}(\theta_0)), \phi_\delta'(\eta_{k,0}(\theta_0)) \nabla_{\theta_{-\delta}} \eta_{k,0}(\theta_0))^\top.$$

As for Lemma 2, we do not prove Lemma 3. We refer the interested reader to Diop and Kengne (2021), Aknouche and Francq (2020), and Debaly and Truquet (2022), among others.

Proof of asymptotic normality part of Theorem 2. For the proof of asymptotic normality part of Theorem 2, one can note that in the single framework ($k = 1$), assumptions (AN.2) yield the asymptotic normality of $\sqrt{T} \nabla \ell_k(\theta_0)$ using the central limit theorem for martingale difference. Next,

$$\frac{\partial^2 \ell_{k,t}(\theta)}{\partial \theta_i \partial \theta_j} = \frac{1}{\lambda_{k,t}^2(\theta)} \left(\frac{Y_{k,t}}{\lambda_{k,t}(\theta)} - n_{k,t} \right) \frac{\partial \lambda_{k,t}(\theta)}{\partial \theta_i} \frac{\partial \lambda_{k,t}(\theta)}{\partial \theta_j}$$
$$+ \frac{1}{\lambda_{k,t}^2(\theta)} \frac{Y_{k,t}}{\lambda_{k,t}(\theta)} \frac{\partial \lambda_{k,t}(\theta)}{\partial \theta_i} \frac{\partial \lambda_{k,t}(\theta)}{\partial \theta_j}$$
$$- \frac{1}{\lambda_{k,t}(\theta)} \left(\frac{Y_{k,t}}{\lambda_{k,t}(\theta)} - n_{k,t} \right) \frac{\partial^2 \lambda_{k,t}(\theta)}{\partial \theta_i \partial \theta_j} =: I_{k,t}(\theta) + II_{k,t}(\theta) + III_{k,t}(\theta).$$

For the first term,

$$\sup_\theta |I_{k,t}(\theta)| \leq n_{i,t} \left(\frac{\lambda_{i,t}}{\log(1 + \delta_-)} + 1 \right) \sup_\theta \frac{1}{\lambda_{k,t}^2(\theta)} \frac{\partial \lambda_{k,t}(\theta)}{\partial \theta_i} \frac{\partial \lambda_{k,t}(\theta)}{\partial \theta_j},$$

and

$$\frac{1}{\lambda_{k,t}(\theta)} \dot{\lambda}_{k,t}(\theta) \preccurlyeq \kappa_{\delta_-} \left(1, \iota_k, \frac{Y_{k,t-1}}{n_{k,t-1}} \ldots \frac{Y_{k,t-p}}{n_{k,t-p}}, X_{k,t}^\top \right)^\top,$$

where for $x = (x_1, \ldots, x_d), y = (y_1, \ldots, y_d), x \preccurlyeq y$ means $x_i \leq y_i, i = 1, \ldots, d$, and κ_{δ_-} is a function of δ_-. Then, $\mathbb{E} \sup_\theta |I_{k,t}(\theta)| < \infty$ under assumption (AN.3). It can be shown similarly that $\mathbb{E} \sup_\theta |II_{k,t}(\theta)| < \infty$ and $\mathbb{E} \sup_\theta |III_{k,t}(\theta)| < \infty$. By the Taylor expansion of $r_T(\cdot)$ between $\hat{\theta}_T$ and θ,

$$0 = \sqrt{T} \nabla r_T(\hat{\theta}_T) = \sum_{k=1}^{K} \sqrt{T} \nabla \ell_k(\hat{\theta}_T)$$
$$= \left(\sum_{k=1}^{K} \sqrt{T} \nabla \ell_k(\theta_0) \right) + \left(\sum_{k=1}^{K} \nabla^2 \ell_k(\theta_0) \right) \sqrt{T}(\hat{\theta}_T - \theta_0) + o_{\mathbb{P}}(1).$$

The independence condition for path (AN.1), assumption (AN.2), and the central limit theorem for martingale difference allows us to conclude $\sum_{k=1}^{K} \sqrt{T} \nabla \ell_k(\theta_0)$ converges in distribution to a central Gaussian vector of variance V as T tends to infinity. The assumption (AN.3) and ergodic theorem entail that $\sum_{k=1}^{K} \nabla^2 \ell_k(\theta_0)$ converges to J. Moreover, conditions (AN.1), (C.3), and (C.4) ensure that the matrix J is invertible. $\square$

References

Aber, John, Neilson, Ronald P., McNulty, Steve, Lenihan, James M., Bachelet, Dominique, Drapek, Raymond J., 2001. Forest processes and global environmental change: predicting the effects of individual and multiple stressors: we review the effects of several rapidly changing environmental drivers on ecosystem function, discuss interactions among them, and summarize predicted changes in productivity, carbon storage, and water balance. BioScience 51 (9), 735–751.

Aknouche, Abdelhakim, Francq, Christian, 2020. Count and duration time series with equal conditional stochastic and mean orders. Econom. Theory 1–33. http://dx.doi.org/10.1017/S0266466620000134.

Allen, D.E., McAleer, M., Scharth, M., 2010. Realized Volatility Risk. Working Paper, University of Canterbury. Department of Economics, and Finance, Department of Economics and Finance, College of Business and Economics, University of Canterbury, URL https://books.google.fr/books?id=YD-RtgEACAAJ.

Beguería, Santiago, Vicente-Serrano, Sergio M., 2017. SPEI: Calculation of the standardised precipitation-evapotranspiration index. URL https://CRAN.R-project.org/package=SPEI. R package version 1.7.

Berguet, Cassy, Martin, Maxence, Arseneault, Dominique, Morin, Hubert, 2021. Spatiotemporal dynamics of 20th-century spruce budworm outbreaks in eastern canada: Three distinct patterns of outbreak severity. Front. Ecol. Evol. (ISSN: 2296-701X) 8, http://dx.doi.org/10.3389/fevo.2020.544088, URL https://www.frontiersin.org/article/10.3389/fevo.2020.544088.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. J. Econometrics 31, 307–327.

Boulanger, Yan, Arseneault, Dominique, 2004. Spruce budworm outbreaks in eastern quebec over the last 450 years. Can. J. Forest Res. 34 (5), 1035–1043.

Bunn, Andrew G., 2008. A dendrochronology program library in r (dplr). Dendrochronologia 26 (2), 115–124.

Camarero, J.J., Martín, E., Gil-Pelegrín, E., 2003. The impact of a needleminer (epinotia subsequana) outbreak on radial growth of silver fir (abies alba) in the aragón pyrenees: a dendrochronological assessment. Dendrochronologia 21 (1), 3–12.

Cameron, A. Colin, Trivedi, Pravin K., 2013. Regression Analysis of Count Data. Vol. 53. Cambridge University Press.

Chan, Bunyeth, Ngor, Peng Bun, Hogan, Zeb S., So, Nam, Brosse, Sébastien, Lek, Sovan, 2020. Temporal dynamics of fish assemblages as a reflection of policy shift from fishing concession to co-management in one of the world's largest tropical flood pulse fisheries. Water 12 (11), http://dx.doi.org/10.3390/w12112974, URL https://www.mdpi.com/2073-4441/12/11/2974.

Chou, Ray Yeutien, Chou, Hengchih, Liu, Nathan, 2015. Range volatility: a review of models and empirical studies. In: Handbook of Financial Econometrics and Statistics. pp. 2029–2050.

Davis, Richard A., Holan, Scott H., Lund, Robert, Ravishanker, Nalini, 2016. Handbook of Discrete-Valued Time Series. CRC Press.

Debaly, Zinsou Max, Truquet, Lionel, 2021a. Iterations of dependent random maps and exogeneity in nonlinear dynamics. Econom. Theory 1–38. http://dx.doi.org/10.1017/S0266466620000559.

Debaly, Zinsou Max, Truquet, Lionel, 2021b. A note on the stability of multivariate non-linear time series with an application to time series of counts. Statist. Probab. Lett. 179, 109196.

Debaly, Zinsou Max, Truquet, Lionel, 2022. Multivariate time series models for mixed data. (in press) in Bernoulli, 202x.

Dennis, J.E., Schnabel, R.B., 1983. Numerical Methods for Unconstrained Optimization and Nonlinear Equations. Prentice-Hall. Inc. Englewood Clis.

Diop, Mamadou Lamine, Kengne, William, 2021. Inference and model selection in general causal time series with exogenous covariates. arXiv preprint arXiv:2102.02870.

D'Orangeville, Loïc, Houle, Daniel, Duchesne, Louis, Phillips, Richard P., Bergeron, Yves, Kneeshaw, Daniel, 2018. Beneficial effects of climate warming on boreal tree growth may be transitory. Nature Commun. 9 (1), 1–10.

Engle, Robert F., Russell, Jeffrey R., 1998. Autoregressive conditional duration: A new model for irregularly spaced transaction data. Econometrica 66 (5), 1127–1162, ISSN 00129682, 14680262. URL http://www.jstor.org/stable/2999632.

Fleming, Richard A., Volney, W. Jan A., 1995. Effects of climate change on insect defoliator population processes in canada's boreal forest: some plausible scenarios. Water Air Soil Pollut. 82 (1), 445–454.

Gauthier, Sylvie, Bernier, Patrick, Kuuluvainen, T., Shvidenko, A.Z., Schepaschenko, D.G., 2015. Boreal forest health and global change. Science 349 (6250), 819–822.

Girardin, Martin P., Bouriaud, Olivier, Hogg, Edward H., Kurz, Werner, Zimmermann, Niklaus E., Metsaranta, Juha M., Jong, Rogier de, Frank, David C., Esper, Jan, Büntgen, Ulf, Guo, Xiao Jing, Bhatti, Jagtar, 2016. No growth stimulation of canada's boreal forest under half-century of combined warming and co2 fertilization. Proc. Natl. Acad. Sci. (ISSN: 0027-8424) 113 (52), E8406–E8414. http://dx.doi.org/10.1073/pnas.1610156113, URL https://www.pnas.org/content/113/52/E8406.

Girona, Miguel Montoro, Morin, Hubert, Lussier, Jean-Martin, Ruel, Jean-Claude, 2019. Post-cutting mortality following experimental silvicultural treatments in unmanaged boreal forest stands. Front. Forests Glob. Change 4.

Girona, Miguel Montoro, Morin, Hubert, Lussier, Jean-Martin, Walsh, Denis, 2016. Radial growth response of black spruce stands ten years after experimental shelterwoods and seed-tree cuttings in boreal forest. Forests 7 (10), 240.

Girona, Miguel Montoro, Navarro, Lionel, Morin, Hubert, 2018. A secret hidden in the sediments: Lepidoptera scales. Front. Ecol. Evol. 6, 2.

Girona, Miguel Montoro, Rossi, Sergio, Lussier, Jean-Martin, Walsh, Denis, Morin, Hubert, 2017. Understanding tree growth responses after partial cuttings: A new approach. PLoS One 12 (2), e0172653.

Glorot, Xavier, Bordes, Antoine, Bengio, Yoshua, 2011. Deep sparse rectifier neural networks. In: Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics. pp. 315–323, JMLR Workshop and Conference Proceedings.

Guay, Régent, Gagnon, Réjean, Morin, Hubert, 1992. A new automatic and interactive tree ring measurement system based on a line scan camera. For. Chron. 68 (1), 138–141.

Jardon, Yves, Morin, Hubert, Dutilleul, Pierre, 2003. Périodicité et synchronisme des épidémies de la tordeuse des bourgeons de l'épinette au québec. Can. J. Forest Res. 33 (10), 1947–1961.

Klapwijk, Maartje J., Csóka, György, Hirka, Anikó, Björkman, Christer, 2013. Forest insects and climate change: Long-term trends in herbivore damage. Ecol. Evol. 3 (12), 4183–4196.

Klimko, Lawrence A., Nelson, Paul I., 1978. On conditional least squares estimation for stochastic processes. Ann. Statist. 629–642.

Krause, Cornelia, Gionest, F., Morin, Hubert, MacLean, David A., 2003. Temporal relations between defoliation caused by spruce budworm (choristoneura fumiferana clem.) and growth of balsam fir (abies balsamea (l.) mill). Dendrochronologia 21 (1), 23–31.

Krause, Cornelia, Morin, Hubert, 1995. Changes in radial increment in stems and roots of balsam fir [abies balsamea (l.) mill.] after defoliation spruce budworm. For. Chron. 71 (6), 747–754.

Labrecque-Foy, Julie-Pascale, Morin, Hubert, Girona, Miguel Montoro, 2020. Dynamics of territorial occupation by north american beavers in canadian boreal forests: A novel dendroecological approach. Forests 11 (2), 221.

Lavoie, Janie, Girona, Miguel Montoro, Morin, Hubert, 2019. Vulnerability of conifer regeneration to spruce budworm outbreaks in the eastern canadian boreal forest. Forests 10 (10), 850.

Martin, Maxence, Girona, Miguel Montoro, Morin, Hubert, 2020. Driving factors of conifer regeneration dynamics in eastern canadian boreal old-growth forests. PLoS One 15 (7), e0230221.

Mei, Hongyuan, Eisner, Jason M., 2017. The neural hawkes process: A neurally self-modulating multivariate point process. Adv. Neural Inf. Process. Syst. 30.

MFFP, 2019. Données sur les perturbations naturelles - insecte : Tordeuse des bourgeons de l'épinette. URL https://www.donneesquebec.ca/recherche/fr/dataset/donnees-sur-les-perturbations-naturelles-insecte-tordeuse-des-bourgeons-de-lepinette. (Accessed 19 May 2019).

Morin, Hubert, Gagnon, Réjean, Lemay, Audrey, Navarro, Lionel, 2021. Chapter thirteen - revisiting the relationship between spruce budworm outbreaks and forest dynamics over the holocene in eastern north america based on novel proxies. In: Johnson, Edward A., Miyanishi, Kiyoko (Eds.), Plant Disturbance Ecology, second ed. Academic Press, San Diego, ISBN: 978-0-12-818813-2, pp. 463–487. http://dx.doi.org/10.1016/B978-0-12-818813-2.00013-7, URL https://www.sciencedirect.com/science/article/pii/B9780128188132000137.

Navarro, Lionel, Morin, Hubert, Bergeron, Yves, Girona, Miguel Montoro, 2018. Changes in spatiotemporal patterns of 20th century spruce budworm outbreaks in eastern canadian boreal forests. Front. Plant Sci. 9, 1905.

R Core Team, 2021. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, URL https://www.R-project.org/.

Régnière, Jacques, Saint-Amant, Rémi, Béchard, Ariane, Moutaoufik, Ahmed, 2014. BioSIM 10: User's Manual. Laurentian Forestry Centre Québec, QC, Canada.

Seidl, Rupert, Thom, Dominik, Kautz, Markus, Martin-Benito, Dario, Peltoniemi, Mikko, Vacchiano, Giorgio, Wild, Jan, Ascoli, Davide, Petr, Michal, Honkaniemi, Juha, et al., 2017. Forest disturbances under climate change. Nature Clim. Change 7 (6), 395–402.

Sim, Chiaw-Hock, 1990. First-order autoregressive models for gamma and exponential processes. J. Appl. Probab. 27 (2), 325–332.

Speer, James H., Kulakowski, Dominik, 2017. Creating a buzz: insect outbreaks and disturbance interactions. In: Dendroecology. Springer, pp. 231–255.

Taniguchi, Masanobu, Kakizawa, Yoshihide, 2002. Asymptotic Theory of Statistical Inference for Time Series. Springer Science & Business Media.

Vourlitis, George L., Pinto, Osvaldo Borges, Dalmagro, Higo J., de Arruda, Paulo Enrique Zanella, de Almeida Lobo, Francisco, de Souza Nogueira, José, 2022. Tree growth responses to climate variation in upland and seasonally flooded forests and woodlands of the cerrado-pantanal transition of Brazil. Forest Ecol. Manag. (ISSN: 0378-1127) 505, 119917. http://dx.doi.org/10.1016/j.foreco.2021.119917, URL https://www.sciencedirect.com/science/article/pii/S0378112721010082.

Walker, Xanthe J., Mack, Michelle C., Johnstone, Jill F., 2015. Stable carbon isotope analysis reveals widespread drought stress in boreal black spruce forests. Global Change Biol. 21 (8), 3102–3113.

Weiß, Christian H., 2018. An Introduction to Discrete-Valued Time Series, Chapter 4. John Wiley & Sons.

Weiß, Christian H, Zhu, Fukang, Hoshiyar, Aisouda, 2022. Softplus ingarch models. Stat. Sinica 32 (3).