Canadian Journal of Forest Research
Revue canadienne de recherche forestière

More than Moran: Coupling statistical and simulation models to understand how defoliation spread and weather variation drive insect outbreak dynamics

SCHOLARONE™
Manuscripts

1 **Title:** More than Moran: Coupling statistical and simulation models to understand how defoli-

2 ation spread and weather variation drive insect outbreak dynamics

3

4 **Authors:** Hedvig K Nenzén[1,2,*], Pedro Peres-Neto[1], Dominique Gravel[2]

5

6 1: Canada Research Chair in Spatial Modelling and Biodiversity, Département des sciences bi-

7 ologiques, Université du Quebec à Montréal, CP 8888, Succ. Centre Ville, Montréal, Quebec, H3C

8 3P8, Canada.* peres-neto.pedro@uqam.ca

9

10 2: Canada Research Chair on Terrestrial Ecosystems. Département de biologie, chimie et géo-

11 graphique, Université du Quebec à Rimouski, 300 Allée des Ursulines, Rimouski, Quebec, G5L 3A1,

12 Canada ** dominique.gravel@usherbrooke.ca

13

14 * Corresponding author. hedvig.nenzen@gmail.com. Tel: 1-514-4987-3000 ext. 5239. Fax 1-514-

15 987-4647.

16

17 Current Addresses: * Department of Biology, Concordia University, Loyola Campus, 7141 Sher-

18 brooke Street West, Montreal, Quebec H4B 1R6, Canada. ** Département de biologie, Faculté des

19 Sciences, Université de Sherbrooke, 2500, boulevard de l'Université, Sherbrooke (Quebec) J1K 2R1

20 Canada.

Abstract

Understanding the processes that underlie species fluctuations is crucial to the development of efficient management strategies for outbreaks of destructive forest pests. Yet, the role of biotic and abiotic factors as well as their interactions in synchronizing outbreaks is not understood, despite many empirical and theoretical studies of species fluctuations. Here, we use a combined statistical-simulation model to investigate how defoliation spread and autocorrelated weather affect outbreaks of a major defoliator of North American boreal forest, the spruce budworm. We modelled the regional dynamics of spruce budworm and based the model on data collected from spatiotemporal aerial surveys of defoliation from 1968-2015 in Quebec, Canada. The effects of weather on local forest stand defoliation and mortality transitions, along with defoliation spread probability and distance, were estimated statistically. Simulations were run with these estimates to identify the effects of spatiotemporal weather autocorrelation on synchronicity of outbreaks. Defoliation spread together with weather variables was found to best fit the observed outbreak size. Simulation models suggest that positive temporal autocorrelation in weather promotes outbreaks, indicating that a series of suitable years could encourage outbreaks.

Keywords: landscape ecology, model, outbreaks, environmental autocorrelation, defoliation spread

Introduction

Understanding the mechanisms underlying the spatiotemporal population dynamics of species is an ongoing ecological challenge (Liebhold et al., 2004). Outbreaks are high population densities of defoliating insects covering large areas and causing extensive tree mortality, making this phenomena of great concern for management decisions (Boyd et al., 2013). For example, spruce budworm (*Choristoneura fumiferana*) outbreaks profoundly structure the dynamics of North American boreal forests (Fleming, 2000; Pureswaran et al., 2016). Defoliation is so synchronized that during the last outbreak, 50 million hectares of Canada's forests were affected at the same time (Sturtevant et al., 2015), destroying 200 million m^3 of wood and causing large economic and ecological impacts in the region (Morin et al., 2007). The multiple mechanisms hypothesized to synchronize budworm outbreaks are often studied separately, so their relative roles in driving outbreaks are not well understood which could limit efficient management actions. Here, we present a novel method to couple statistical and simulation models to tease apart the drivers of synchronous outbreaks of forest defoliating insects.

Hypotheses about what causes synchronous outbreaks are multiple and connected, but can be broadly divided into abiotic (i.e. environment) or intrinsic (i.e. defoliation spread) drivers (Liebhold et al., 2004). According to the abiotic hypothesis, spatially autocorrelated environmental variation affecting population dynamics synchronizes the abundance of isolated populations (the Moran effect, Moran, 1953). Alternatively, the intrinsic hypothesis suggests that synchrony occurs when individuals from a single population disperse rapidly throughout the landscape (Williams and Liebhold, 2000). In nature, multiple hypotheses may interact to produce outbreaks (Royama, 1984; Cooke et al., 2007). Previous empirical studies concentrated on isolated local populations (Tian et al., 2011), thus not considering the effect of possible dispersal from adjacent populations in driving population changes. Similarly, theoretical models may have underestimated the effects of dispersal as they also only used local population density data (Kendall et al., 1999). Such local population-level ap-

proaches provided only a partial understanding of the mechanisms underlying outbreaks that can be driven by both regional environmental fluctuations and dispersal (Gilbert and O'Connor, 2013). As such, it is necessary to use landscape-level analysis of species dynamics to investigate the effects of both environmental autocorrelation and dispersal.

In order to represent the dynamics of insect outbreaks on a landscape-level, an alternative approach to population-level models, is to consider their population state as in either low (endemic) or high (epidemic) abundance. During endemic periods, budworm could be considered as 'absent' as they cause no detectable forest defoliation (Régnière and Nealis, 2007). During epidemic periods budworm densities are several orders of magnitude higher (Régnière and Nealis, 2007) and the could be considered as 'present' because they cause visible forest defoliation. Therefore, instead of modelling the presence of defoliators we model the presence of defoliation. Large-scale dynamics thus emerge from the local defoliation (from absent to present) and mortality (from present to absent) transitions. Fortunately, large-scale, multi-year empirical defoliation data (e.g., aerial surveys) is readily available, yet rarely used to identify parameters and simulate models (Zhou and Liebhold, 1995). We use defoliation data as a proxy for high insect densities. Defoliating insect species such as the budworm show extreme density fluctuations, which leads us to propose that their dynamics can be sufficiently expressed by presences/absences instead of abundances.

The spatial dynamics of defoliation easily map onto metapopulation and epidemiological models. Metapopulation theory (Levins, 1969) represents transitions between empty and occupied patches driven by colonization and extinction. Similarly, epidemiological models (Kermack and McKendrick, 1927; Keeling and Rohani, 2008) track either negligible or disease-causing densities of the infectious agent among individual hosts. The two methods are mathematically equivalent in some instances (Grenfell and Harwood, 1997) and are supported by extensive theoretical development (Hanski, 1998). For example, spatially-explicit vaccination of the most vulnerable locations can reduce the

87 size of epidemics of infectious diseases (Keeling et al., 2003). Insights from epidemiological models

88 could be used to develop more effective spatial management strategies of insect outbreaks.

89   Transition from endemic to epidemic states is a binomial stochastic process whose drivers can be

90 evaluated from empirical data. Stochastic cell occupancy, state-and-transition, and occupancy mod-

91 els, are all essentially parameterized metapopulation models of species dynamics in space (Moila-

92 nen, 2004; Buckland and Elston, 1993; Bestelmeyer et al., 2004). Statistical models have been used to

93 determine how local stand transitions of various insect species depend on environment and potential

94 defoliation spread from already defoliated neighboring cells (Bouchard and Auger, 2014; Kärvemo

95 et al., 2016). The transitions observed in aerial surveys could be used to parameterize a model of

96 insect defoliations according to simple rules of defoliation spread and environmental effects. Ac-

97 cording to metapopulation and epidemiological theory, the transition from endemic to epidemic is

98 expected to be conditional on both defoliation spread and abiotic conditions. Here, we propose that

99 such an approach of insect outbreaks can be used to investigate the effects of intrinsic and extrinsic

100 drivers of outbreaks.

101   The environment is usually autocorrelated in space and time (Koenig, 2002). Positive autocorre-

102 lation, when environmental values are more similar among adjacent temporal or spatial units than

103 expected by chance, may affect species fluctuations and defoliation spread (Ruokolainen et al., 2009).

104 Synchronous outbreaks may be produced by suitable environments in many locations (Moran effect,

105 spatial autocorrelation). Recent theoretical and experimental studies suggest that suitable environ-

106 mental conditions over consecutive periods further synchronize populations (temporal autocorrela-

107 tion, Gonzalez and Holt, 2002; Massie et al., 2015), but this mechanism has never been demonstrated

108 in nature. We propose that a simulation model parameterized from empirical transitions can assist

109 in disentangling the relative importance and the interactions between drivers of outbreak dynamics.

110   The goal of our study is threefold: 1) determine how weather and defoliation spread drive ob-

111 served budworm outbreak and mortality transitions; 2) quantify the relative importance of these

112 two variables; and 3) determine if spatial or temporal weather autocorrelation synchronize out-

113 breaks. We parameterized a lattice-based metapopulation model from observed outbreaks in Que-

114 bec, Canada. We predict that if the model reproduces the observed outbreaks using the observed

115 autocorrelated weather variables, it suggests that the Moran effect drives budworm outbreaks. Al-

116 ternatively, if the model reproduces the observed outbreaks with non auto-correlated weather, then

117 it would suggest that defoliation spread is the most important driver. As this study only considers

118 defoliation spread and weather, we cannot deduce whether other hypotheses not considered here

119 drive outbreaks. We used budworm as an example species but our coupled statistical-simulation

120 metapopulation model could be applicable to any species that displays large-scale synchrony.

Fig. 1

Materials and methods

Coupled statistical and simulation model

123 We considered a metapopulation approach in which a single spatial unit is a cell, and multiple

124 cells together constitute a landscape (Fig. 1). Each cell can be in two different states: forest F with

125 endemic budworm densities that do not cause defoliation, or defoliated forest I with epidemic bud-

126 worm densities that show signs of defoliation (or infection, hence I). As such, these two cell states

127 correspond to observed endemic and epidemic population densities that vary over orders of mag-

128 nitude (Cappuccino et al., 1998). We studied the transition from forest to defoliation (defoliation

129 transition $F \to I$ or $F \to I$, colonization) and the transition from defoliation back to forest (mortality

130 transition $I \to F$, extinction, Fig. 2). Stochastic state transitions ($F \to I$ and $I \to F$) across all cells

131 (landscape) generate a fluctuating proportion of F and I cells over time.

132 Our model has two parts. We start by statistically estimating the transition probabilities of defo-

liation and mortality given defoliation spread constraints and climatic conditions, and then we use

the transition probabilities as parameters in a simulation model to test the influence of autocorre-

lation on outbreaks. Defoliation of a single forest cell resulted from two processes corresponding

to the following two hypotheses, respectively: 1) by increasing the insect population density due to

favourable weather; 2) by dispersing individuals arriving from neighboring defoliated cells (Fig. 2

arrows). A candidate set of statistical submodels were chosen *a priori* based on defoliation spread

and weather hypotheses to assess the relative contribution of these two processes (i.e., weather and

defoliation spread) in driving large-scale outbreaks. The probability of defoliation in cell *i* was mod-

eled as a log-linear model of weather and defoliation spread:

$$Probability\{F \to I\}_i = \beta_0 + \beta_n E_i + \alpha_n K_i \tag{1}$$

where β_0 is the intercept, E_i is the set of weather variable, β_n is the set of parameters representing the

effect of weather variables on the transition probability, K_i is the proportion of defoliated neighbors

around cell *i* and α_n is the parameter representing this effect. We tested up to third degree polyno-

mial, i.e. *n* up to 3 (eqn 1), for both weather and defoliation spread variables. This accounted for

complex responses while obtaining parametric relationships for the simulation model. The transi-

tion probability of mortality from defoliation to forest is only dependent on weather. The mortality

transition is set by parameter γ:

$$Probability\{I \to F\}_i = \gamma E_i \tag{2}$$

First, we performed model comparison to select the best-fitting statistical model representing

the observed transitions. Second, the statistical model of transition probabilities was used to carry

out simulation 'experiments' to investigate the effect of spatiotemporal weather autocorrelation on

outbreak dynamics. Our statistical models can only describe the effect of weather for single time

153 steps and cells, whereas simulations integrate dynamics across the entire landscape and over time.

154 We measured the total proportion of defoliated cells in the landscape through time generated by the

155 simulation model and assessed if they were similar to observed outbreaks. We tested the influence

156 of weather autocorrelation on outbreaks by randomizing weather variables E_i, which maintained

157 climatic values but removed autocorrelation structure.

Spatiotemporal data

Spruce budworm defoliation distribution

160 The Ministère des Forêts, Faune et Parc du Quebec (MFFP, 2014) have tracked the temporal (an-

161 nual) and spatial distribution of budworm outbreaks based on aerial surveys and recorded mapped

162 polygons indicating areas of defoliation in Quebec province, Eastern Canada. The data span 1968

163 through 2015, which include one full outbreak (1968-1995), along with the beginning of an ongoing

164 outbreak which started around 2010 (Fig. 1).

165 We converted these polygons to a lattice of 10km x 10km geographical cells (patches, forest stands,

166 from here on referred to as cells arranged in a continuous lattice, Fig. 1), which formed the basis for

167 our coupled statistical-simulation model. Because recent polygons were mapped to a smaller scale

168 than older data, data across all years were aggregated to the same cell size. We overlaid 1km x

169 1km spaced points on the original defoliated polygons, and then resampled this to a regular grid of

170 10km x 10km cells so if at least one of the points within the cell was defoliated, the cell was classified

171 as defoliated (following Bouchard and Auger, 2014). This aggregation also minimized small-scale

172 landscape variations. Any level of defoliation (light, moderate or serious defoliation) within poly-

173 gons were registered as I. We assumed that non-defoliated areas were F and that mortality occurred

174 when defoliation was no longer observed, even though it is unknown if the forest dies or survives

175 defoliation. In total, the study area had 6 446 cells after removing cells that were never defoliated.

176 Each cell was mapped for 46 years, yielding 283 272 transition events (Tables **??**, **??**).

177 After mortality from defoliation, forest regrowth usually takes around 40 years (Burns and Honkala,

178 1990), but occasionally a cell was observed to become re-defoliated before full regrowth. These fast

179 transitions may have arisen because of detection errors and as such were not considered here as

180 representative of the naturally observed defoliation and mortality processes we aim to model. To

181 account for this issue, if a cell became defoliated, not defoliated and then re-defoliated within five

182 years, we converted the intermediate years to the defoliated state.

Fig. 2

Table

Weather variables

184 Historical weather variables covering the same spatiotemporal extent as the defoliation data were

185 obtained from publicly available data. In this data set, weather station data were interpolated to con-

186 tinuous rasters using the ANUSPLIN method (McKenney et al., 2011). Multiple weather variables,

187 in particular extreme temperatures, influence outbreak severity and occurrence (Candau and Flem-

188 ing, 2005). We consequently selected eight potential annual and seasonal weather variables (and

189 elevation, Table 1) for the analysis. Weather one, two and three years before each year t (E_{t-1}, E_{t-2}

190 and E_{t-3}) were also included as weather variables may have a delayed effect on insect population

191 growth and mortality (Aukema et al., 2008), and because there is a lag between population growth

192 and crossing the endemic-epidemic threshold where defoliation is detectable. All weather variables

193 were standardized within the range $0 - 1$ (max-obs)/(max-min) to facilitate comparison between

194 parameters.

Defoliation spread variables

196 The expansion of budworm defoliation to an endemic-density cell is more likely to occur if there

197 are defoliated cells in the vicinity (Bouchard and Auger, 2014). Note that defoliation spread indi-

198 rectly represented the successful dispersal of insect individuals. We considered defoliation spread

199 kernels K_i to represent propagule pressure around the neighbourhood of cell i. We compared K_{it-1}

200 and K_{it-2} computed one and two years prior to the transition t and $t+1$ given that egg-laying adults

201 produced a population increase the following year when larvae emerge. We used two methods and a

202 range of neighborhood sizes in each method to evaluate defoliation spread kernels that characterize

203 defoliation spread:

204 The autologistic kernel represented the proportion of defoliated cells within a certain distance

205 (Yackulic et al., 2015). All cells within the neighborhood were given equal weight. K_i of each potential

206 donor cell i was the number of defoliated cells j within distance d. K_i is standardized by the total

207 number of neighbors so that K is one if all neighbours are defoliated. We calculated K_i for $d = 2 - 30$

208 cells in increments of 1 because the maximum radius of 30 cells covered the majority of the landscape

209 (89 x 176 cells).

The negative exponential kernel included all cells in the landscape but gave a higher weight
to cells close to the focal cell, allowing long-distance defoliation spread (i.e. beyond the threshold
distance in the autologistic kernel). The negative exponential kernel is the sum of the weighted
distances between all cells (Moilanen, 1999):

$$K_i = \sum_{j \neq i} exp(-\delta d_{ij})$$

210 where δ is the average defoliation spread distance and d_{ij} is the distance in number of cells between

211 cell i and j (in this equation α traditionally represents the effect of distance, but we use δ to avoid con-

212 fusion with α probability of defoliation spread). We calculated defoliation spread variables varying

213 $\delta = 2 - 30$ in increments of 2 because this range captured extremes of both short and long-distance

214 defoliation spread (Table ??). The values of negative exponential defoliation spread variable K_i were

215 standardized to $0 - 1$ (max-obs)/(max-min) to be on the same scale as the autologistic kernel.

Statistical models

We determined how weather and defoliation spread influenced transition probabilities by tracking the state transitions of each cell in the observed defoliation data, and estimating how the probabilities were affected by weather and defoliation spread. We constructed separate statistical models for the defoliation $F \rightarrow I$ and mortality $I \rightarrow F$ state transition probabilities. Transition was modelled as a binomial process using logistic regression. The selected weather and defoliation spread variables described above were used as predictors of forest transitions (equation 1 - 2).

Models were evaluated in two ways: model fit (accuracy) and predictive performance on independent data. Goodness-of-fit metrics to assess model fit were ΔAIC and pseudo r^2 for logistic regressions, $r^2 = 1 - (residual deviance / null deviance)$. Cross-validation to assess predictive performance was performed on a random subset of the data and on the second outbreak. Random cross-validation was performed by splitting the data in half, fitting the model on the subset and then estimating model performance on the remaining validation data. This was repeated 10 times and the average error calculated. Temporal cross-validation was performed by fitting the model to the first outbreak (1968-1992) and predicting the model on the second outbreak (1993-2015). We did not estimate temporal cross-validation in $I \rightarrow F$ transitions because there were little data from the second outbreak.

Model comparison

We tested which potential defoliation spread kernel K_i and kernel parameters d and δ best fitted the data. Given that the defoliation spread variables were highly correlated (first axis of a Principal Components Analysis on the correlation matrix explained 84% of the variation, Fig. **??**), we used forward stepwise model selection to rank models with each defoliation spread kernel separately. Only the best-supported kernel was retained for all subsequent analyses (according to AIC, Figs.

239 **??**).

240 The full regression model had one defoliation spread and all eight weather variables with three

241 time delays as predictors (Table 2). We used the first and three best weather variables and interac-

242 tions between them to construct submodels. To determine which candidate weather variable best

243 explained the observed budworm defoliation and mortality transitions, we used backward stepwise

244 model selection because weather variables were not collinear (the first PCA axis on the correlation

245 matrix explained 50% of the variation, Figs. **??**). We estimated a geographical null model to assess if

246 outbreaks were independent of weather and directional in space (anisotropic). Latitude and longi-

247 tude were used as predictors for the geographical null model. We also estimated a defoliation spread

248 null model in which defoliation spread was not limited by distance (global defoliation spread). Pro-

249 portion defoliated area in the previous year was the global defoliation spread predictor, and was

250 compared to the best-fitting distance-based defoliation spread kernel. To assess the effect of host

251 forest species we also tested the effect of dominant host species (boreal, deciduous or mixed) but

252 this variable was not statistically important (Table **??**).

Simulation model

254 We simulated outbreaks using the estimated defoliation and mortality transition probabilities

255 from the full statistical model to investigate the effect of spatiotemporal autocorrelation of weather

256 variables on the outbreak dynamics. We ran stochastic simulations with the statistical transition

257 probabilities and the observed climatic fluctuations, so the simulations occurred in the same land-

258 scape and over the same years as the observed outbreaks. Uncertainty can propagate up and yield

259 different large-scale dynamics even if the model successfully represents the local dynamics (here

260 transitions, Peters et al., 2004).

261 Simulations were initialized with the 1970 weather and run for 45 years. Depending on the state

262 of chosen cell i (F or I), transition probabilities $Probability\{F \to I\}_i$ and $Probability\{I \to F\}_i$ were

263 applied (as explained above, equations 1 and 2). The defoliation spread kernel K_i was calculated

264 in each stand and year from the cells defoliated in the previous year. We then added weather and

265 defoliation spread parameters exactly as in the statistical model to obtain spatiotemporal transition

266 probabilities that ranged $0 - 1$ depending on stand weather and neighbourhood defoliation (equa-

267 tion 1). The final probabilities were then compared to uniformly distributed random values $0 - 1$ to

268 determine if a transition occurs, so transitions were stochastic but influenced by the probabilities.

269 We used asynchronous updating of the landscape by randomly picking and updating 6 446 cells per

270 time step in order to approximate continuous spatial dynamics (Durrett and Levin, 1994).

271   We carried out simulations with and without autocorrelated weather to understand if structured

272 weather variation drives outbreaks (Figure 1, panels in Fig. 4). We randomized weather variables,

273 thus maintaining the frequency distribution of values but removing the autocorrelation structure.

274 First, we randomized weather conditions across time-steps, within each cell, to remove temporal

275 autocorrelation. Then we randomized weather conditions between cells, within each time-step, to

276 remove spatial autocorrelation. We randomized weather in both space and time to remove both

277 temporal and spatial autocorrelation. Simulations with average weather values from all years and

278 locations were carried out to test how weather variation affects outbreak dynamics. Finally, to in-

279 vestigate if the estimated defoliation spread kernel was important in producing outbreaks, we also

280 implemented global defoliation spread with the original weather variables.

281   We characterized simulated outbreaks (30 repetitions per spatiotemporal autocorrelation) in time

282 and space to compare them to observed outbreaks. Simulated defoliations were stochastic and so

283 cannot be compared cell-by-cell to observed defoliations. Instead, to examine if simulated outbreaks

284 were as large, i.e. as spatially synchronous as observed outbreaks, we recorded the proportion of the

285 landscape in state I per year. We also calculated the maximum, standard deviation, and temporal

286 autocorrelation of proportion *I*. Real outbreaks are known to be spatially clustered (i.e. defoliated

287 cells were often connected by their 8 closest neighbors when considering a rasterized landscape,

288 Fig. 1 b), and so we determined if the simulation model was able to reproduce this property. We

289 calculated the maximum cluster size and number of outbreak clusters per simulation to compare the

290 spatial characteristics of observed and simulated landscapes. Fig. 3

Results

292 Overall, the statistical model identified that budworm defoliation spread most often occurs over

293 short distances. We found that the negative exponential kernel best fit the observed defoliation tran-

294 sitions (Fig. **??**, Fig. 3 for example). The defoliation spread probabilities were generally below 0.5,

295 indicating that defoliation spread was episodic and did not always occur. We identified that lagged

296 spring and summer temperatures best explained the defoliation and mortality transitions. The three

297 strongest climatic drivers of the $F \rightarrow I$ transition were in order of importance: maximum summer

298 temperature three years before, summer minimum temperature three years before and number of

299 growing degree days 2 years before (Fig. **??**, Table **??**). Climatic transition probabilities were low,

300 around 0.1 for the most suitable conditions (Fig. **??**, **??**). Minimum spring temperature three years

301 before, precipitation three years before, and minimum temperature of the coldest period two years

302 before were the strongest predictors of mortality transitions (Fig. **??**, Table **??**).

303 The model comparison indicated that both weather and defoliation spread drivers affected out-

304 breaks, but their relative importance depended on how models were evaluated. Models with only

305 weather variables were more accurate (according to AIC), while models with only defoliation spread

306 predicted the second outbreak better (according to temporal cross-validation, Table 2, Table **??**). The

307 full weather model, i.e. with all climatic variables and third degree response, was the most accurate

308 model when predicting on randomly selected data from both outbreaks (cross-validation according

309 to AIC, Table 2). However, the full model was less successful at predicting the second outbreak (ac-

310 cording to temporal cross-validation, Table 2). The model with the best predictive performance was

311 a submodel with three weather variables, defoliation spread and interactions between them (accord-

312 ing to temporal cross-validation, Table 2). Defoliation spread was the best unique variable in terms of

313 both accuracy and predictive performance (according to AIC and cross-validation, Table 2). Adding

314 interactions between weather and defoliation spread increased model fit, though it decreased pre-

315 dictive power (according to AIC and temporal cross-validation, Table 2). Models with interaction

316 so closely fit to the first outbreak are not general enough to predict the second outbreak (overfit).

317 The geographical null model did not fit outbreaks well, indicating that weather was important in

318 synchronizing outbreaks.

319   Simulations with the parameterized metapopulation model revealed that successive periods of

320 temporally autocorrelated weather synchronized outbreaks. Retaining only the temporal weather

321 autocorrelation produced outbreaks that covered 70% of the landscape, compared to the 90% ob-

322 served (example simulations Fig. 4, Fig. 5 a). Simulations with the original autocorrelated weather

323 also produced outbreaks that were as large as the observed outbreaks (Fig. 5, a, b). When retaining

324 only spatial autocorrelation (i.e., removing temporal autocorrelation), outbreaks did not occur dur-

325 ing the same years as observed outbreaks and were on average smaller (40-60%, Fig. 5 a), indicating

326 that the model was less sensitive to spatial autocorrelation. Simulations with constant weather had

327 10% landscape defoliation throughout the simulations with no variability.

328   Temporal dynamics were in general well captured, and spatial dynamics less so. Randomizing

329 weather in time produced outbreaks with lower next-year temporal autocorrelation (correlations

330 $0.7 - 0.9$ with 1 year time lag, Fig. 5 c), with long-range positive temporal autocorrelation (6-16

331 years, Fig. 5 d). Even spatially randomized climatic conditions occasionally resulted in outbreaks

332 that were as temporally autocorrelated as observed outbreaks. The difference between spatially au-

tocorrelated (random time) and completely random weather conditions was small, again indicating that spatial autocorrelation has a weak effect on outbreak dynamics. Simulated outbreaks were more fragmented in space than observed ones. Simulated outbreak clusters were more numerous and smaller in size than the observed ones (Fig. 1, 5 e, f). Simulations with global defoliation spread and observed spatiotemporal autocorrelation yielded landscapes with high proportions of defoliated cells but reflected neither the temporal nor the spatial characteristics of observed outbreaks (Fig. 5).

Table

Fig. 4

Fig. 5

Discussion

We performed a coupled statistical and simulation analysis of 48 years of spruce budworm outbreak observations to understand how defoliation spread and weather drive the dynamics. The estimated parameters from the best statistical model were incorporated into a simulation model to investigate the role of spatiotemporal autocorrelation on the initial development of outbreaks and their spatial structure. The same method can be applied to outbreaks of other insect species that are mapped by defoliation.

We found that defoliation spread between forest cells occurred occasionally over long distances, however short-distance defoliation spread was more common. Individual-based dispersal studies show that adult moths can efficiently travel 150 km (Sturtevant et al., 2013), and the maximum recorded distance is 450 km (Greenbank, 1980). The greater importance of the negative-exponential kernel over the autologistic version indicates that defoliation spread was essentially driven by strong short-distance defoliation spread events and occasional long-distance defoliation spread events. For instance, the probability of defoliation spread was low though a cell was surrounded by defoliated cells in a suitable weather (Fig. **??**). The budworm population is genetically well-mixed during outbreaks (James et al., 2015), supporting effective long-distance dispersal. However, defoliation spread was still limited by distance as the null model with global defoliation spread badly fit the data (Table

356　2). Metapopulation models are often estimated from static (snapshot) distributions and individual

357　dispersal observations (Moilanen, 1999). In contrast to these static methods and due to the detailed

358　spatiotemporal nature of our data, we characterized effective defoliation spread. The strength and

359　shape of dispersal is often the unknown factor in predictive models, and thus in understanding

360　outbreaks (Myers and Cory, 2013). Although dispersal may not be completely represented by defo-

361　liation data, our method describes an effective method to quantify defoliation spread directly from

362　spatiotemporal occupancy data.

363　　　We attempted to disentangle the effects of weather and defoliation spread with an innovative

364　coupling of statistical analysis and simulation modelling. In the end, our results indicate that both

365　weather and defoliation spread interact to drive mortality. Previous studies may have over-estimated

366　the role of the Moran effect in driving population fluctuations because they did not investigate spatial

367　data and thus discounted dispersal. Purely temporal analyses pool data across large regions (Tian

368　et al., 2011), thus possibly confounding local defoliation spread events with climatic effects. Spa-

369　tial analyses suggest that weather drives dynamics if the spatial autocorrelation range of weather is

370　longer than the spatial autocorrelation range of outbreaks (autocorrelation range distance measured

371　in kilometres, Peltonen et al., 2002). Models that exclude dispersal cannot characterize how weather

372　drives epidemic transition events separately from dispersal. Recent process-based statistical studies

373　testing both weather and dispersal (intrinsic and extrinsic) hypotheses have shown that dispersal

374　is more important than environment in synchronizing outbreaks (Preisler et al., 2012; Veran et al.,

375　2014). Similarly, we found that defoliation spread was an important factor in synchronizing defoli-

376　ations in space, although it required an autocorrelated weather trigger (simulations with constant

377　weather vs. simulations with weather, Figure 4).

378　　　Among the abiotic drivers, a suite of weather predictors best described the observed transitions.

379　We found that delayed spring and summer temperatures influenced both defoliation and mortality

transitions, suggesting that weather during the reproductive season causes populations to reach epidemic insect densities years later, similar to what Aukema et al. (2008) observed. Multiple weather variables are necessary to reproduce outbreak dynamics because these may affect insect development periods and life histories differently. In general, winter temperatures affect overwintering survival, maximum spring temperature affects larval emergence, and summer temperatures affect feeding rates (Régnière et al., 2012; Gray, 2013). The full model with all weather variables was more accurate, suggesting that a specific combination of weather characteristics was important ('a perfect storm', Wilmers et al., 2007). Moreover, interactions between predictors increased fit, and may indicate that weather and defoliation spread interact (e.g., defoliation spread is greater in warmer temperatures).

The difficulty in identifying a single weather driver may be explained statistically. First, it is difficult to fit autocorrelated predictor variables (Kissling and Carl, 2008). Spatiotemporal autoregressive models that account for spatiotemporal autocorrelation are not generalized to logistic regression with complex response functions (Cressie and Wikle, 2011, but see Griffith and Peres-Neto, 2006). Moreover, our goal is not to remove autocorrelation, but to understand how it affects species dynamics. Cross-validation suggested that the full model overfit the data and included the maximum number of variables to produce a spurious relationship. Simpler models with fewer variables are generally preferred for prediction purposes (Cuddington et al., 2013).

The simulation model suggests that non-random weather variation produced outbreaks even though weather itself was a weak statistical predictor (as in the Moran effect). Simulation models can draw such conclusions because they simulate several years, not just one year ahead. The model reproduced the emergent temporal characteristics of budworm outbreaks; 0.9 temporal autocorrelation with a one year lag, and 10 year range in positive autocorrelation (Fig. 5 c, d). We identified that temporal weather autocorrelation drives large-scale outbreaks, because retaining only temporal

autocorrelation produced outbreaks that were similar to both spatial and temporal outbreak characteristics. Non-random temporal environmental variation has been shown to inflate the Moran effect in microcosm experiments (Gonzalez and Holt, 2002; Massie et al., 2015) and theoretical population models (Vasseur, 2007). For the budworm, several consecutive dry summers have been associated with defoliations (Greenbank, 1956). This simulation model allowed us to identify the importance of autocorrelated temporal weather variation at the landscape scale. However, because weather has a high spatial autocorrelation (analysis not shown), randomizing weather in space does not change dynamics and may have led us to underestimate the role of spatial autocorrelation.

To estimate the effect of weather and defoliation spread, we used to most complete set of spatio-temporal data available for the spruce budworm, but this presence-absence dataset still presents some limitations. Only the data for Québec province is publicly available, so our study is bounded by areas with outbreaks that we did not track but which could have nevertheless affected outbreaks. For example, defoliation could have spread into Québec from the provinces of Ontario and New Brunswick. This edge-effect could have led us to underestimate the role of dispersal in our study. Further, aerial survey data may omit low-density population dynamics that are under the defoliation threshold. This cross-scale assumption is partially accounted for by the statistical significance of the weather variables three years previously to the defoliation transition. A large-scale study may find different climate-responses than a population-level study. For example, Robert et al. (2012) showed that tree-ring patterns indicate that forest management determined spruce budworm outbreaks. Finally, as we use observed data we cannot identify causes, only correlations between outbreaks and climate. However, a simulation model was employed to improve the plausibility of the results, and can be considered as the only type of 'experiment' possible to investigate landscape-scale outbreaks. Moreover, theoretical experiments such as ours can suggest future avenues for investigation.

Our simulation model showed that spatiotemporal autocorrelation affected outbreaks, but it did

not completely reproduce the observed dynamics. Observed defoliation spread events are episodic in time and clustered in space, yielding a small number of large outbreak clusters (maximum 30 clusters and a single cluster covering 85% of the study area, Fig. 5 e, f). Defoliation spread is difficult to model because it is inherently stochastic and may be driven by unpredictable atmospheric conditions such as wind direction (Sturtevant et al., 2013). Because our model already contained all possible observed climatic spatial autocorrelation, it is possible that additional hypotheses are required to completely predict outbreak patterns (Liebhold et al., 2004). For instance, the Allee effect decreases mating success at endemic population densities (Régnière et al., 2013) and may alter dispersal and establishment, and possibly landscape-scale outbreak dynamics. Spatially correlated distributions of natural enemies such as parasitoids may synchronize insect densities (Roland and Taylor, 1997; Pureswaran et al., 2016), but there is little available empirical data. The effect of local environmental conditions such as forest composition may further mediate natural enemy interactions (Bouchard and Auger, 2014). However, Gray (2013) found that average climate variables were a better predictor of large-scale outbreak duration and severity than forest composition. Detailed data on natural enemies and forest composition could improve predictions.

Conclusions

Dramatic spruce budworm outbreaks have inspired the development of many models, including statistical and simulation models. Here we reinterpreted outbreaks as a metapopulation process so that forest defoliation data could be employed as an indicator of low and high insect densities. It is beneficial to estimate parameters on the same scale at which outbreaks are observed, because emergent outbreak characteristics may not be predicted from local mechanisms (Fleming et al., 2002). Simple metapopulation models are as useful as complex local population models, especially when the ecological mechanisms are hard to parameterize or simply unknown (Harrison

et al., 2011). Simple models are also more applicable because of their ease of use and generality, and can easily be extended to other species. The relative importance of defoliation spread is one of the main hurdles for understanding outbreaks (Myers and Cory, 2013) and here we present a method to quantify defoliation spread kernels from aerial survey data. The landscape-scale of our process-based metapopulation model makes it relevant for the development of large-scale forestry and budworm population management practices. Pest management considers pests as immobile, possibly preventing effective management (Tscharntke et al., 2007). Because metapopulation models can be analogous to epidemiological models (Grenfell and Harwood, 1997), widely tested epidemiological techniques to minimize transmission can be considered. Our metapopulation model could be used to generate management strategies that minimize defoliation spread and thus outbreaks, such as spatially-explicit harvesting or spraying programmes. Moreover, to our knowledge for the first time in nature, results suggest for that spatiotemporal autocorrelation in weather values may generate insect outbreaks.

Acknowledgements

References

Aukema, B.H., Carroll, A.L., Zheng, Y., Zhu, J., Raffa, K.F., Dan Moore, R., Stahl, K., and Taylor, S.W. 2008. Movement of outbreak populations of mountain pine beetle: Influences of spatiotemporal patterns and climate. Ecography **31**: 348–358. doi:10.1111/j.0906-7590.2007.05453.x.

Bestelmeyer, B.T., Herrick, J.E., Brown, J.R., Trujillo, D.A., and Havstad, K.M. 2004. Land management in the American southwest: a state-and-transition approach to ecosystem complexity. Environ. Manage. **34**: 38–51. doi:10.1007/s00267-004-0047-4.

Bouchard, M. and Auger, I. 2014. Influence of environmental factors and spatio-temporal covariates during the initial development of a spruce budworm outbreak. Landsc. Ecol. **29**(1): 111–126. doi:10.1007/s10980-013-9966-x.

Boyd, I.L., Freer-Smith, P.H., Gilligan, C.A., and Godfray, H.C.J. 2013. The consequence of tree pests and diseases for ecosystem services. Science **342**(2013): 1235773. doi:10.1126/science.1235773.

Buckland, S.T. and Elston, D.A. 1993. Empirical Models for the Spatial Distribution of Wildlife. J. Appl. Ecol. **30**(3): 478. doi:10.2307/2404188.

Burns, R.M. and Honkala, B.H. 1990. Silvics of North America: 1. Conifers; 2. Hardwoods. vol.2. Agriculture Handbook 654. U.S. Department of Agriculture, Forest Service, Washington, DC.

Candau, J.N. and Fleming, R.A. 2005. Landscape-scale spatial distribution of spruce budworm defoliation in relation to bioclimatic conditions. Can. J. For. Res. **35**: 2218–2232. doi:10.1139/X05-078.

Cappuccino, N., Lavertu, D., Bergeron, Y., and Régnière, J. 1998. Spruce budworm impact, abundance and parasitism rate in a patchy landscape. Oecologia **114**(2): 236–242. doi:10.1007/s004420050441.

Cooke, B.J., Nealis, V.G., and Régnière, J. 2007. Insect defoliators as periodic disturbances in northern forest ecosystems. In Plant Disturb. Ecol. Process Response, edited by E. Johnson and K. Miyanishi, Elsevier, Burlington, Massachusetts, USA., pp. 487–525.

Cressie, N. and Wikle, C. 2011. Statistics for Spatio-Temporal Data. Wiley.

Cuddington, K.M., Fortin, M.j., Gerber, L.R., Hastings, A., Liebhold, A.M., O'Connor, M., and Ray, C. 2013. Process-based models are required to manage ecological systems in a changing world. Ecosphere **4**(2): 1–12. doi:10.1890/ES12-00178.1.

Durrett, R. and Levin, S.A. 1994. The importance of being discrete (and spatial). Theor. Popul. Biol. **46**(3): 363–394. doi:10.1006/tpbi.1994.1032.

Fleming, R.A. 2000. Climate change and insect disturbance regimes in Canada's boreal forests. World Resour. Rev. **12**(3): 521–555.

Fleming, R.A., Barclay, H.J., and Candau, J.N. 2002. Scaling-up an autoregressive time-series model (of spruce budworm population dynamics) changes its qualitative behavior. Ecol. Modell. **149**(1-2): 127–142. doi:10.1016/S0304-3800(01)00519-1.

Gilbert, B. and O'Connor, M.I. 2013. Climate change and species interactions: Beyond local communities. Ann. N. Y. Acad. Sci. **1297**: 98–111. doi:10.1111/nyas.12149.

Gonzalez, A. and Holt, R.D. 2002. The inflationary effects of environmental fluctuations in source-sink systems. Proc. Natl. Acad. Sci. **99**(23): 14872–14877.

Gray, D.R. 2013. The influence of forest composition and climate on outbreak characteristics of the spruce budworm in eastern Canada. Can. J. For. Res. **1195**: 1181–1195.

Greenbank, D. 1956. The role of climate and dispersal in the inititiation of outbreaks of the spruce budworm in New Brunswick. 1. The role of climate. Can. J. Zool. **34**: 453–476.

516 Greenbank, D. 1980. Spruce budworm (Lepidoptera: Tortricidae) moth flight and dispersal: new

517 understanding from canopy observations, radar, and aircraft. Mem. Entomol. Soc. Canada **112**:

518 1–49. doi:10.4039/entm112110fv.

519 Grenfell, B.T. and Harwood, J. 1997. (Meta)population dynamics of infectious diseases. Trends Ecol.

520 & Evol. **12**: 395–399.

521 Griffith, D.A. and Peres-Neto, P.R. 2006. Spatial modeling in Ecology: the flexibility of the eigenfunc-

522 tion spatial analyses. Ecology **87**(10): 2603–2613. doi:10.1890/0012-9658(2006)87[2603:SMIETF]2.

523 0.CO;2.

524 Hanski, I. 1998. Metapopulation Dynamics. Nature **396**(6706): 41–49. doi:10.1038/23876.

525 Harrison, P.J., Hanski, I., and Ovaskainen, O. 2011. Bayesian state-space modeling of metapopulation

526 dynamics in the Glanville fritillary butterfly. Ecol. Monogr. **81**(4): 581–598. doi:10.1890/11-0192.1.

527 James, P.M.A., Cooke, B.J., Brunet, B.M.T., Lumley, L.M., and Felix, A. 2015. Life-stage differences

528 in spatial genetic structure in an irruptive forest insect : implications for dispersal and spatial

529 synchrony. Mol. Ecol. **24**: 296–309. doi:10.1111/mec.13025.

530 Kärvemo, S., Johansson, V., Schroeder, M., and Ranius, T. 2016. Local colonization-extinction dy-

531 namics of a tree-killing bark beetle during a large-scale outbreak. Ecosphere **7**(3): 1–14. doi:

532 10.1002/ecs2.1257.

533 Keeling, M.J. and Rohani, P. 2008. Modeling infectious diseases in humans and animals. Princeton

534 University Press, Princeton, NJ.

535 Keeling, M.J., Woolhouse, M.E.J., May, R.M., Davies, G., and Grenfell, B.T. 2003. Modelling vac-

536 cination strategies against foot-and-mouth disease. Nature **421**(6919): 136–42. doi:10.1038/

537 nature01343.

538 Kendall, B.E., Briggs, C.J., Murdoch, W.W., Turchin, P., Ellner, S.P., McCauley, E., Nisbet, R.M., and

539 Wood, S.N. 1999. Why do populations cycle? A synthesis of statistical and mechanistic modeling

540 approaches. Ecology **80**: 1789–1805. doi:10.1890/0012-9658(1999)080[1789:WDPCAS]2.0.CO;2.

541 Kermack, W.O. and McKendrick, A.G. 1927. A Contribution to the Mathematical Theory of Epi-

542 demics. Proc. R. Soc. A Math. Phys. Eng. Sci. **115**(772): 700–721. doi:10.1098/rspa.1927.0118.

543 Kissling, W.D. and Carl, G. 2008. Spatial autocorrelation and the selection of simultaneous autore-

544 gressive models. Glob. Ecol. Biogeogr. **17**: 59–71. doi:10.1111/j.1466-8238.2007.00334.x.

545 Koenig, W.D. 2002. Global patterns of environmental synchrony and the Moran effect. Ecography

546 **25**(3): 283–288.

547 Levins, R. 1969. Some Demographic and Genetic Consequences of Environmental Heterogeneity for

548 Biological Control. Bull. Ecol. Soc. Am. **15**(3): 237–240.

549 Liebhold, A.M., Koenig, W.D., and Bjørnstad, O.N. 2004. Spatial Synchrony in Population Dynamics.

550 Annu. Rev. Ecol. Evol. Syst. **35**(1): 467–490. doi:10.1146/annurev.ecolsys.34.011802.132516.

551 Massie, T.M., Weithoff, G., Kuckländer, N., Gaedke, U., and Blasius, B. 2015. Enhanced Moran

552 effect by spatial variation in environmental autocorrelation. Nat. Commun. **6**: 5993. doi:

553 10.1038/ncomms6993.

554 McKenney, D.W., Hutchinson, M.F., Papadopol, P., Lawrence, K., Pedlar, J., Campbell, K., Milewska,

555 E., Hopkinson, R.F., Price, D.T., and Owen, T. 2011. Customized Spatial Climate Models for North

556 America. Publicly available at gmaps.nrcan.gc.ca/cl_p/climatepoints.php. Bull. Am. Meteorol.

557 Soc. **92**(12): 1611–1622. doi:10.1175/2011BAMS3132.1.

558 MFFP 2014. Ministère des Forêts de la Faune et des Parcs Direction de la protection des forêts. Service

de la gestion des Ravageurs. Donnees sur les perturbations naturelles: Tordeuse des bourgeons de l'épinette. Données ouvertes .

Moilanen, A. 1999. Patch occupancy models of metapopulation dynamics: efficient parameter estimation using implicit statistical inference. Ecology **80**(3): 1031–1043.

Moilanen, A. 2004. SPOMSIM: Software for stochastic patch occupancy models of metapopulation dynamics. Ecol. Modell. **179**: 533–550. doi:10.1016/j.ecolmodel.2004.04.019.

Moran, P.A.P. 1953. The statistical analysis of the Canadian lynx cycle. I. Structure and prediction. Aust. J. Zool. **1**(2): 291–298.

Morin, H., Jardon, Y., and Gagnon, R. 2007. Relationship Between Spruce Budworm Outbreaks and Forest Dynamics in Eastern North America. In Plant Disturb. Ecol. Process Response, edited by E. Johnson and K. Miyanishi, Elsevier, New York, pp. 555–577.

Myers, J.H. and Cory, J.S. 2013. Population Cycles in Forest Lepidoptera Revisited. Annu. Rev. Ecol. Evol. Syst. **44**(1): 565–592. doi:10.1146/annurev-ecolsys-110512-135858.

Peltonen, M., Liebhold, A.M., Bjørnstad, O.N., and Williams, D.W. 2002. Spatial synchrony in forest insect outbreaks: roles of regional stochasticity and dispersal. Ecology **83**: 3120–3129. doi:10. 1890/0012-9658(2002)083[3120:SSIFIO]2.0.CO;2.

Peters, D.P.C., Herrick, J.E., Urban, D.L., Gardner, R.H., and Breshears, D.D. 2004. Strategies for ecological extrapolation. Oikos **106**(3): 627–636. doi:10.1111/j.0030-1299.2004.12869.x.

Preisler, H.K., Hicke, J.A., Ager, A.A., and Hayes, J.L. 2012. Climate and weather influences on spatial temporal patterns of mountain pine beetle populations in Washington and Oregon. Ecology **93**(11): 2421–2434. doi:10.1890/11-1412.1.

580   Pureswaran, D.S., Johns, R., Heard, S.B., and Quiring, D. 2016. Paradigms in Eastern Spruce Bud-

581   worm (Lepidoptera: Tortricidae) Population Ecology: A Century of Debate. Environ. Entomol. p.

582   nvw103. doi:10.1093/ee/nvw103.

583   Régnière, J., Delisle, J., Pureswaran, D.S., and Trudel, R. 2013. Mate-finding Allee effect in spruce

584   budworm population dynamics. Entomol. Exp. Appl. **146**: 112–122. doi:10.1111/eea.12019.

585   Régnière, J. and Nealis, V.G. 2007. Ecological mechanisms of population change during outbreaks

586   of the spruce budworm. Ecol. Entomol. **32**(5): 461–477. doi:10.1111/j.1365-2311.2007.00888.x.

587   Régnière, J., Powell, J., Bentz, B., and Nealis, V.G. 2012. Effects of temperature on development,

588   survival and reproduction of insects: experimental design, data analysis and modeling. J. Insect

589   Physiol. **58**(5): 634–47. doi:10.1016/j.jinsphys.2012.01.010.

590   Robert, L.e., Kneeshaw, D., and Sturtevant, B.R. 2012. Effects of forest management legacies on

591   spruce budworm (Choristoneura fumiferana) outbreaks. Can. J. For. Res. **475**: 463–475. doi:

592   10.1139/X2012-005.

593   Roland, J. and Taylor, P. 1997. Insect parasitoid species respond to forest structure at different spatial

594   scales. Nature **386**(6626): 710–713. doi:10.1038/386710a0.

595   Royama, T. 1984. Population Dynamics of the Spruce Budworm Choristoneura fumiferana. Ecol.

596   Monogr. **54**: 429–462.

597   Ruokolainen, L., Lindén, A., Kaitala, V., and Fowler, M.S. 2009. Ecological and evolutionary dy-

598   namics under coloured environmental variation. Trends Ecol. & Evol. **24**(10): 555–63. doi:

599   10.1016/j.tree.2009.04.009.

600   Sturtevant, B.R., Achtemeier, G.L., Charney, J.J., Anderson, D.P., Cooke, B.J., and Townsend, P.A.

601   2013. Long-distance dispersal of spruce budworm (Choristoneura fumiferana Clemens) in Min-

602 nesota (USA) and Ontario (Canada) via the atmospheric pathway. Agric. For. Meteorol. **168**:

603 186–200.

604 Sturtevant, B.R., Cooke, B.J., Kneeshaw, D., and Maclean, J.E. 2015. Modeling Insect Disturbance

605 Across Forested Landscapes: Insights from the Spruce Budworm. In Simul. Model. For. Landsc.

606 Disturbances, edited by A.H. Perera, B.R. Sturtevant, and L.J. Buse, Springer, pp. 1–321. doi:10.

607 1007/978-3-319-19809-5.

608 Tian, H., Stige, L.C., Cazelles, B., Kausrud, K.L., Svarverud, R., Stenseth, N.C., and Zhang, Z. 2011.

609 Reconstruction of a 1,910-y-long locust series reveals consistent associations with climate fluctua-

610 tions in China. Proc. Natl. Acad. Sci. **108**(6): 14521–14526. doi:10.1073/pnas.1100189108.

611 Tscharntke, T., Bommarco, R., Clough, Y., Crist, T.O., Kleijn, D., Rand, T.A., Tylianakis, J.M., Van

612 Nouhuys, S., and Vidal, S. 2007. Conservation biological control and enemy diversity on a land-

613 scape scale. Biol. Control **43**: 294–309. doi:10.1016/j.biocontrol.2007.08.006.

614 Vasseur, D.A. 2007. Environmental colour intensifies the Moran effect when population dynamics

615 are spatially heterogeneous. Oikos **116**(10): 1726–1736. doi:10.1111/j.2007.0030-1299.16101.x.

616 Veran, S., Simpson, S.J., Sword, G.A., Deveson, E., Piry, S., Hines, J.E., and Berthier, K. 2014. Modeling

617 spatiotemporal dynamics of outbreaking species: influence of environment and migration in a

618 locust. Ecology **96**(3): 737–748. doi:10.1890/14-0183.1.

619 Williams, D.W. and Liebhold, A.M. 2000. Spatial synchrony of spruce budworm outbreaks in eastern

620 North America. Ecology **81**(10): 2753–2766.

621 Wilmers, C.C., Post, E., and Hastings, A. 2007. A perfect storm: the combined effects on population

622 fluctuations of autocorrelated environmental noise, age structure, and density dependence. Am.

623 Nat. **169**(5): 673–83. doi:10.1086/513484.

624   Yackulic, C.B., Nichols, J.D., Reid, J., and Der, R. 2015. To predict the niche, model colonization and

625   extinction. Ecology **96**(1): 16–23. doi:10.1890/14-1361.1.

626   Zhou, G. and Liebhold, A.M. 1995. Forecasting the spatial dynamics of gypsy moth outbreaks using

627   cellular transition models. Landsc. Ecol. **10**(3): 177–189. doi:10.1007/BF00133030.

Tables

Table 1: The potential weather variables used in the analysis and their time lags tested. The interpolated values covered each cell in the study area (Fig. 1 b) from 1965- 2013. From McKenney et al., 2011. Bottom five lines, the types of weather randomizations.

weather Variable	Time lag
Max/min temperature of warmest/coldest period	1 year before
Max/min temperature of coldest month	2 years before
Max/min temperature of spring (April-May)	3 years before
Max/min temperature of summer (June-July)	
Length of growing season (Growing Degree Days)	
Total Precipitation	
Elevation (constant in time)	
	Original time, original space
	Random time, original space
	Original time, random space
	Random time, random space
	Average time and space (constant)

Table 2: Model selection results on defoliation FI transitions. The three best climate variables were summer maximum temperature $t - 3$, summer minimum temperature $t - 3$ and number of growing degree days $t - 2$. The variables had third-degree response terms. K = the number of variables.

	K	AIC	deltaAIC	pseudo-r^2	random residuals	temporal residuals
all climate + dispersal	79	19475	0	0.425	2439	1375
all climate	76	24374	4899	0.300	2842	1582
3 climate + dispersal	13	27599	8124	0.214	3390	1387
1 climate x dispersal	16	28084	8609	0.186	3610	1395
1 climate + dispersal	7	28860	9386	0.175	3558	1384
dispersal	4	28889	9414	0.168	3610	1378
3 climate (with interactions)	64	29708	10234	0.156	3329	1575
global dispersal	4	30548	11073	0.120	3688	1490
3 climate	10	32076	12602	0.066	3721	1581
1 climate	4	33678	14203	0.015	3868	1572
latitude x longitude	16	34596	15121	0.013	3751	1562

Figure Captions

Fig. 1. Left, the proportion of outbreaks over time, expressed as the proportion of cells in the defoliated state. Right, the geographic extent of the study area in Quebec province, Eastern Canada. Each square is a 'cell' (6 446 in total) and the colour indicates its state: Dark grey (red online) is defoliated/defoliated, grey (green online) is forest non-defoliated cell in 2015. Grey areas were never defoliated and excluded from the analysis.

Fig. 2. Schematic representation of a metapopulation approach to forest insect outbreaks, indicating the forest F and defoliated I states. The arrows indicate the parameters that set the state transitions defoliation $F \rightarrow I$ and mortality $I \rightarrow F$.

Fig. 3. Map of αK calculated from the cells defoliated in 2013 (in dark grey, red online). The defoliation spread kernel K was calculated using the negative exponential kernel with $\delta = 10$.

Fig. 4. The simulated proportion of defoliated cells with different types of spatiotemporally auto-correlated weather. Dotted lines indicates the observed outbreaks, and each thin line indicates the proportion of defoliated landscape during one simulated outbreak.

Fig. 5. Summary results from simulation model. Horizontal dotted lines indicate the values in the observed outbreaks. Each plot shows the values from 100 simulations. Boxes encompass the 25% - 75% quartiles of the data, with the median indicated by the thick line through the centre of each box. Whiskers extending from the box encompass the 95% quartiles, and extreme observations are shown as circles. a, Maximum proportion of defoliated area. b, Standard deviation in the simulated proportion of defoliated area I. c, Temporal autocorrelation with a 1 year time lag. d, Range of positive temporal autocorrelation (in years). e, Maximum number of contiguous outbreak clusters per simulation. f, Largest contiguous outbreak cluster size per simulation (number of cells).
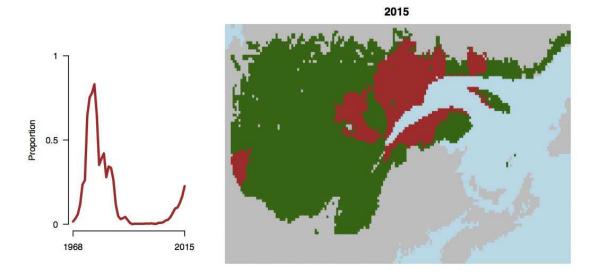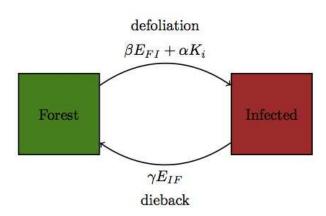
Figures

Fig. 1: Left, the proportion of outbreaks over time, expressed as the proportion of cells in the defoliated state. Right, the geographic extent of the study area in Quebec province, Eastern Canada. Each square is a 'cell' (6 446 in total) and the colour indicates its state: Dark grey (red online) is defoliated/defoliated, grey (green online) is forest non-defoliated cell in 2015. Grey areas were never defoliated and excluded from the analysis.

defoliation

$$\beta E_{FI} + \alpha K_i$$

$$\beta = \text{probability of spontaneous infection}$$
$$\alpha = \text{probability of infection spread}$$
$$\gamma = \text{probability of a dieback to forest}$$
$$E_{FI} = \text{predicted climate suitability of defoliation}$$
$$E_{IF} = \text{predicted climate suitability of dieback}$$
$$K_i = \text{dispersal kernel}$$

Forest   Infected

$$\gamma E_{IF}$$

dieback

Fig. 2: Schematic representation of a metapopulation approach to forest insect outbreaks, indicating the forest F and defoliated I states. The arrows indicate the parameters that set the state transitions defoliation $F \rightarrow I$ and mortality $I \rightarrow F$.

Probability of dispersal α K

Predicted probability
of dispersal

☐	1
☐	0.75
☐	0.5
☐	0.25
■	0

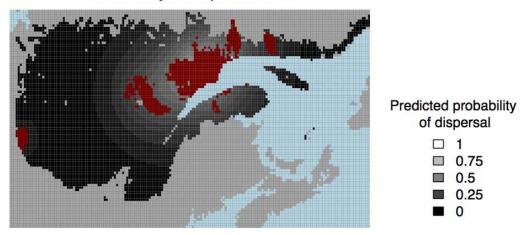Fig. 3: Map of αK calculated from the cells defoliated in 2013 (in dark grey, red online). The defoliation spread kernel K was calculated using the negative exponential kernel with $\delta = 10$.
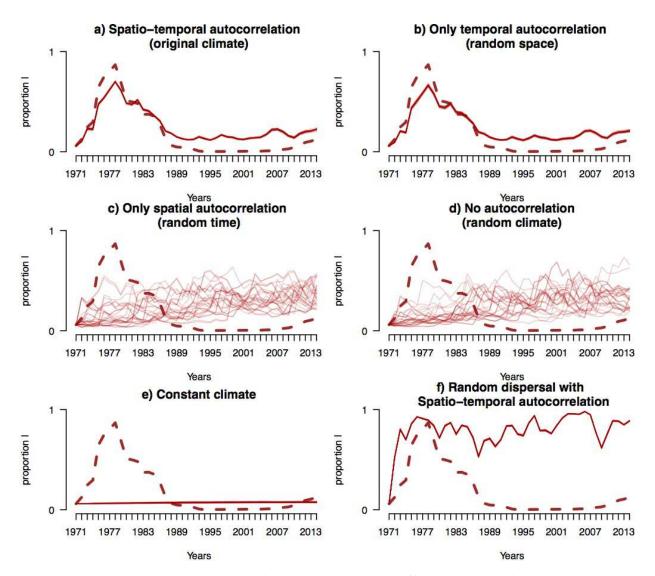
Fig. 4: The simulated proportion of defoliated cells with different types of spatiotemporally auto-correlated weather. Dotted lines indicates the observed outbreaks, and each thin line indicates the proportion of defoliated landscape during one simulated outbreak.
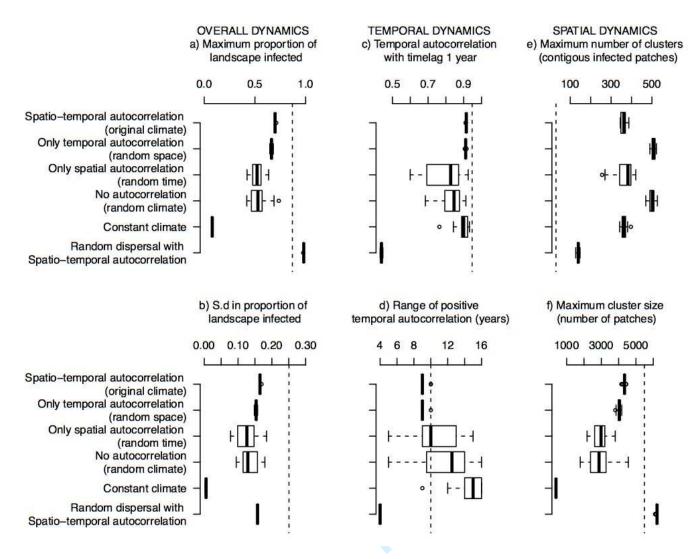
Fig. 5: Summary results from simulation model. Horizontal dotted lines indicate the values in the observed outbreaks. Each plot shows the values from 100 simulations. Boxes encompass the 25% - 75% quartiles of the data, with the median indicated by the thick line through the centre of each box. Whiskers extending from the box encompass the 95% quartiles, and extreme observations are shown as circles. a, Maximum proportion of defoliated area. b, Standard deviation in the simulated proportion of defoliated area I. c, Temporal autocorrelation with a 1 year time lag. d, Range of positive temporal autocorrelation (in years). e, Maximum number of contiguous outbreak clusters per simulation. f, Largest contiguous outbreak cluster size per simulation (number of cells).

More than Moran: Coupling statistical and simulation models to understand how defoliation spread and weather variation drive outbreak dynamics

March 7, 2017

Parameter results

The weather prediction E_{FI} was low even in the most suitable weather ($\beta = 0.008$, Figure S7).

The mean weather prediction of infections E_{FI} was higher along the coast and in the south. In models with only the best weather variable, the probability of spontaneous outbreaks was highest at moderate maximum summer temperatures, two years previously (Figure S7c).

The mean defoliation mortality to forest E_{IF} was more probable on the geographic range edges, meaning that the outbreaks die back faster in unsuitable weather closer to the range edges (γ, Figure S7b). In unsuitable weather, (i.e. in extreme spring maximum temperatures) the probability of an defoliated forest becoming forest in the next year was close to one (Figure S7 d).

Dispersal probabilities were highest at around $K_i = 0.5$ (Figure S7 e). Dispersal probabilities were also affected by weather, so in suitable weather the probability of defoliation spread can reach 1 (thin lines in Figure S7 e).

Multiple weather variables influence the transition of outbreaks. Summer and spring temperatures generally had a large effect on infection probabilities. Temperature during the growth period determines the insect-temperature response curves in the laboratory (Régnière 1987). In contrast to a similar smaller-scale study with linear weather relationships, precipitation and elevation variables were not selected (Bouchard and Auger 2014). Temperature variables were selected here because our study spanned a wide environmental gradient, to which nonlinear temperature responses fit better. Non-parametric models have suggested that temperature determines cumulative budworm damage (Candau and Fleming 2011; Gray 2013), and here we show that temperature also affects infection transitions.

Analysis of spatial autocorrelation in the statistical model

It was necessary to separate the relative contribution of defoliation spread and (multiple) weather variables to outbreaks, so predictions were divided up to defoliation spread- and weather-dependent partial predictions. Weather predictions E_{FI} represented the contribution of multiple weather variables to the probability of transition. The total probability of transition comes from adding the weather predictions E_{FI} and neighbourhood infection K:

$$Probability\{F \rightarrow I\}_i = E_{FI} + \alpha_1 K + \alpha_2 K^2 + \alpha_3 K^3$$

We analyzed if a higher autocorrelation in E_{FI} was associated with more successful infection transitions. We examined the effect of autocorrelation in E_{FI} in time. We assessed if a series of years with suitable weather for outbreaks occurred during large-scale outbreaks, i.e. in years with temporally autocorrelated predictions. To calculate the temporal autocorrelation of predictions, we first calculated the annual mean of infection E_{FI} and mortality E_{IF} predictions. We then estimated during which years the predictions were asynchronous (phase coherence in wavelet cross-correlation, R-package biwavelet, Gouhier and Grinsted 2012).

The effect of temporal autocorrelation in weather predictions E on infections shows that the outbreak in 1971-1980 occurred at the same time as a high weather predictions E_{FI}, and low average weather predictions E_{FI}. This combination led to many new infection transitions that stayed defoliated longer (Figure S9 a). Infection and mortality transition probabilities were asynchronous for several years at the beginning of the outbreak (not significant, Figure S9 b). However, the probabilities were also asynchronous between outbreaks.
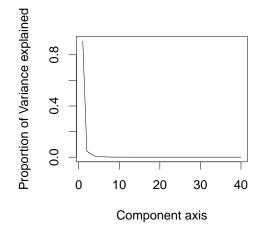
Figure S1: Proportion of variance explained by the first 25 axes of a PCA on the defoliation spread variables K.
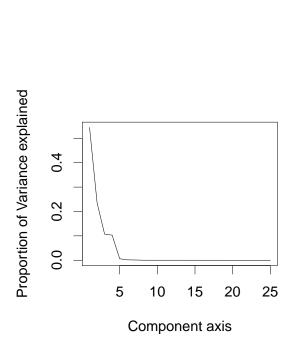
Figure S2: Proportion of variance explained by the first 25 axes of a PCA on the weather variables.

$\delta = 7$ $\delta = 12$

$\delta = 17$
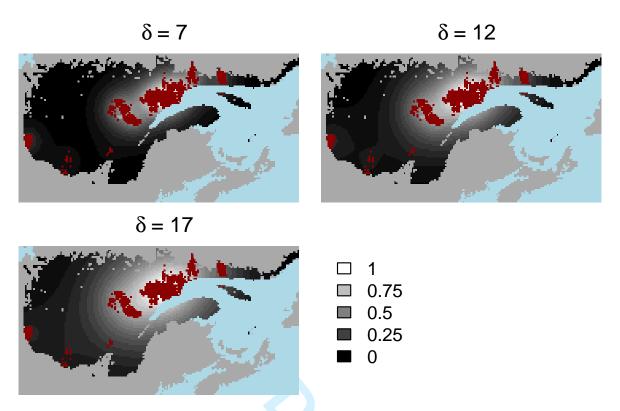
Figure S3: Examples of K with different values for δ, the effect of distance to defoliated cells.
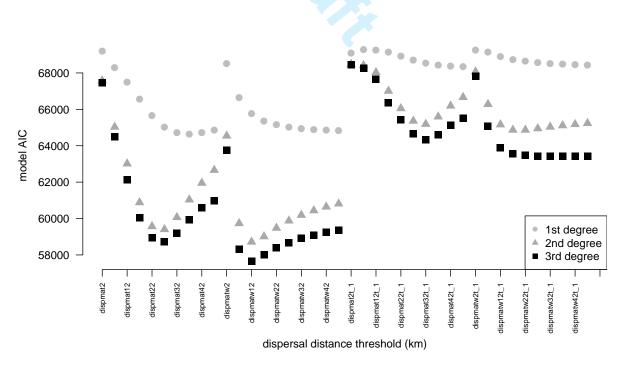
Figure S4: The likelihood profile from selection of the best defoliation spread kernel K_i with first, second and third degree response functions. The estimated AIC with defoliation spread kernels calculated with different δ values, from a forwards stepwise logistic regression. The autologistic and negative exponential kernels are both tested with infection distributions from one and two years previously.
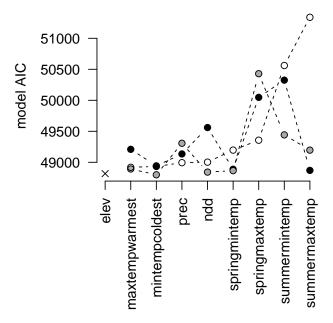
Figure S5: Selection of best weather variable. AIC results of backward model selection results on FI infection transitions, third-degree response terms.
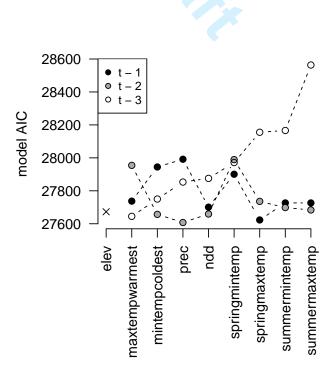
Figure S6: AIC results, backward selection. Model selection results on IM transitions, response terms with 3 degrees and estimated on all data.

Table S1: AIC results, backward selection. Model selection results on FI transitions, with 1,2 and 3-degree response terms and estimated on all data. Each AIC is obtained when carrying out a glm with all weather variables except that one. The variable when removed results in the highest AIC is the most important variable.

	AIC, degree 1	deltaAIC, degree 1	AIC, degree 2	deltaAIC, degree 2	AIC, degree 3	deltaAIC, degree 3
summermaxtempt_3	61162	0	54267	0	51339	0
summermintempt_3	59273	1889	53039	1228	50561	778
nddt_2	58068	3094	52490	1777	50429	910
summermintempt_1	59091	2071	53315	952	50325	1014
nddt_1	58654	2508	52693	1574	50048	1291
springmaxtempt_1	58012	3150	51784	2483	49561	1778
summermintempt_2	57908	3254	51704	2563	49443	1896
nddt_3	57819	3342	51791	2476	49357	1982
maxtempwarmestt_2	58884	2278	52058	2209	49307	2032
springmintempt_1	57835	3327	51746	2521	49209	2130
prect_3	57774	3387	51603	2664	49197	2142
summermaxtempt_2	58039	3122	51702	2565	49196	2143
maxtempwarmestt_1	58861	2301	52130	2138	49135	2204
springmaxtempt_3	57807	3355	51391	2876	49003	2336
maxtempwarmestt_3	59459	1703	52166	2101	48994	2345
mintempcoldestt_1	57891	3271	51691	2577	48943	2396
mintempcoldestt_3	57793	3369	51554	2713	48933	2406
springmintempt_3	57767	3395	51815	2452	48919	2419
springmintempt_2	57840	3321	51557	2711	48892	2447
prect_1	57775	3387	51463	2804	48888	2451
summermaxtempt_1	57816	3345	51634	2633	48870	2469
prect_2	57872	3290	51524	2743	48865	2474
springmaxtempt_2	57932	3230	51451	2817	48842	2497
elev	57768	3394	51509	2758	48818	2521
mintempcoldestt_2	57929	3233	51536	2731	48799	2540

5

Table S2: AIC results, backward selection. Model selection results on IF transitions, with 1,2 and 3-degree response terms and estimated on all data. Each AIC is obtained when carrying out a glm with all weather variables except that one. The variable when removed results in the highest AIC is the most important variable.

	AIC, degree 1	deltaAIC, degree 1	AIC, degree 2	deltaAIC, degree 2	AIC, degree 3	deltaAIC, degree 3
springmintempt_3	34641	1618	29896	0	28564	0
prect_3	34744	1515	28992	904	28166	397
mintempcoldestt_3	36259	0	29393	503	28155	408
summermintempt_1	35301	958	29249	647	27992	572
maxtempwarmestt_2	34638	1621	29200	696	27989	574
maxtempwarmestt_3	34804	1455	29060	837	27972	592
summermaxtempt_2	34633	1626	29009	887	27955	609
nddt_1	35141	1118	29059	837	27945	619
maxtempwarmestt_1	34778	1481	29256	641	27901	663
springmaxtempt_3	34659	1600	29015	881	27875	688
summermintempt_3	34649	1610	28955	941	27853	711
nddt_3	34641	1618	29078	818	27750	814
summermaxtempt_1	34718	1542	28970	926	27737	826
mintempcoldestt_2	34813	1446	28938	958	27736	828
prect_1	34679	1581	28837	1059	27726	837
springmintempt_1	34633	1627	28973	923	27726	837
springmaxtempt_1	34635	1624	28780	1116	27701	863
prect_2	34684	1575	28836	1060	27698	865
springmintempt_2	34985	1274	28880	1016	27684	880
elev	34925	1334	28885	1011	27673	891
springmaxtempt_2	34911	1348	28776	1120	27659	904
nddt_2	34662	1598	28768	1128	27657	907
summermaxtempt_3	34661	1598	28782	1115	27645	919
mintempcoldestt_1	34761	1498	28783	1113	27623	941
summermintempt_2	34661	1598	28780	1116	27608	956

Table S3: Model selection results on defoliation FI transitions also testing submodels of the effect of forest type and years since outbreaks. We removed all data points where the number of years since the last outbreak and the forest type was unknown, which resulted in a smaller data set (192 056 instead of 235 744 observations). The three best climate variables were summer maximum temperature $t-3$, summer minimum temperature $t-3$ and number of growing degree days $t-2$. The variables had third-degree response terms. K = the number of variables.

	K	AIC	deltaAIC	pseudo-r^2	random residuals	temporal residuals
all climate + dispersal	79	7232	0	0.300	13	0
all climate	76	8171	939	0.234	0	83
3 climate + dispersal	13	8764	1532	0.140	75	132
3 climate (with interactions)	64	8861	1629	0.145	53	105
dispersal	4	9072	1840	0.091	109	153
1 climate x dispersal	16	9107	1875	0.127	44	136
latitude x longitude	16	9179	1947	0.111	73	158
1 climate + dispersal	7	9183	1951	0.108	68	143
3 climate	10	9469	2237	0.056	100	171
global dispersal	4	9721	2489	0.045	93	172
years since outbreaks	4	10010	2778	0.048	50	176
1 climate	4	10163	2931	0.013	81	185
forest type	4	10238	3006	0.000	87	187
3 climate x dispersal	256	133008	125776	-6.919	935	186

Table S4: Model selection results on mortality IF transitions. The predictors had third-degree response terms. K = the number of variables. The random residuals are estimated on half of the data selected randomly, and repeated 10 times.

	K	AIC	deltaAIC	pseudo-r^2	rand. res.
all climate	76	13700	0	0.337	19
3 climate x dispersal	256	14539	839	0.314	0
latitude x longitude	16	18787	5087	0.087	44
3 climate	10	18230	4530	0.111	48
1 climate	4	18207	4507	0.104	65

	Forest_K1	I	Forest_K0
Infected	7821	47368	93
Forest	161975	6913	65855

Table S5: The total transition events, when t0 is either Infected or Forest with no outbreaks ($K < 0.01$) and with outbreaks ($K > 0.01$).

	Forest_K1	I	Forest_K0
Infected	0.04606	0.87264	0.00141
Forest	0.95394	0.12736	0.99859

Table S6: The total transition probabilities, when t0 is either Infected, Forest with no outbreaks ($K < 0.01$) and with outbreaks ($K > 0.01$).

References

Bouchard, M., and I. Auger. 2014. Influence of environmental factors and spatio-temporal covariates during the initial development of a spruce budworm outbreak. Landsc. Ecol. 29:111–126.

Candau, J.-N., and R. A. Fleming. 2011. Forecasting the response of spruce budworm defoliation to climate change in Ontario.

Gouhier, T. C., and A. Grinsted. 2012. biwavelet: Conduct univariate and bivariate wavelet analyses. R package version 0.12.

Gray, D. R. 2013. The influence of forest composition and climate on outbreak characteristics of the spruce budworm in eastern Canada. Can. J. For. Res. 1195:1181–1195.

Régnière, J. 1987. Temperature-dependent development of eggs and larvae of Choristoneura fumiferana (Clem.) (Lepidoptera: Tortricide) and simulation of its seasonal history. Can. Entomol. 119:717–728.
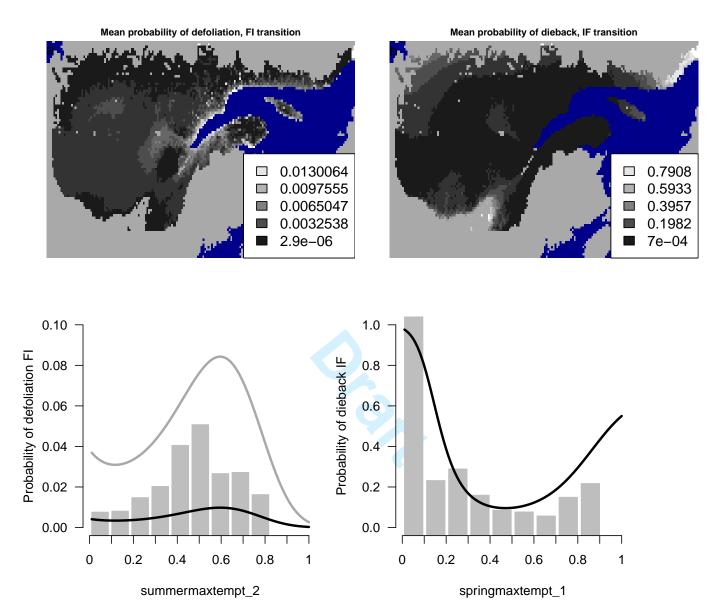
Figure S7: The predicted weather predictions *E* estimated with a logistic regression with third-degree relationships. Top left, E_{FI}, mean probability of forest infection, calculated from all years. White colours indicate a high probability of a forest cell transitioning from the forest state *F* to defoliated *I*, and black colours indicates a low probability of transition (the forest stays as forest). These probabilities are only based on weather and do not include the effect of defoliation spread. Top right, E_{IF}, mean probability of mortality, calculated from all years. White colours indicate a high probability of a forest cell transitioning from the defoliated state *I* to the forest state *F*, and black colours indicates a low probability of transition (the defoliated area stays defoliated). Bottom left, the effect of one weather variable (selected by backward AIC) on the probability of *FI* infection. The black line shows the probability of spontaneous infection ($K_i = 0.0$) and the grey line shows the probability of infection when the defoliation spread kernel $K_i > 0.1$. The grey bars show the observed transitions. Bottom right, the effect of one weather variable (selected by backwards AIC) on the probability of *IF* mortality. The grey bars show the observed transitions.
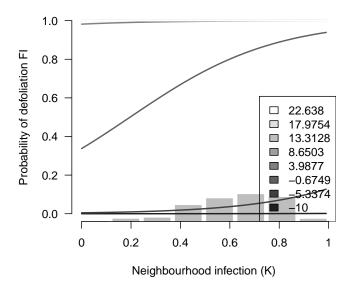
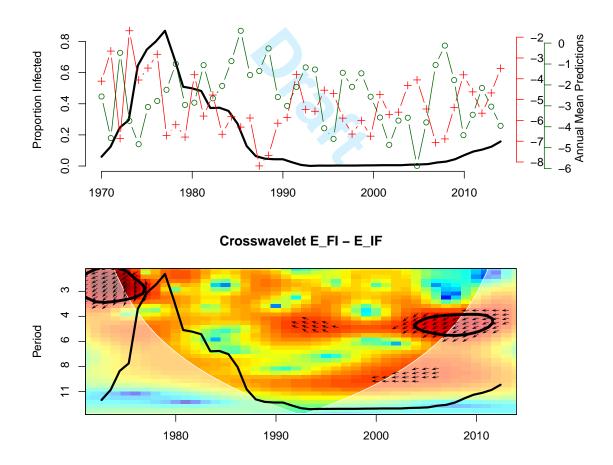Figure S8: Estimated kernel for the defoliation spread, and the lines represent the different weather predictions.

Crosswavelet E_FI – E_IF

Figure S9: The predicted weather predictions E and defoliation spread parameters. Bottom, the mean predicted probabilities of infection E_{FI} (red lines) and mortality E_{IF} (green lines), in all cells. The solid red line is the annual observed total proportion of defoliated forest. Top, cross-correlation wavelet between E_{FI} and E_{IF}. The colours represent the power of the period of the fluctuations, from dark blue (low values), to dark red (high values). The arrows show the phase angles between the fluctuations in the climatic predictions. Arrows pointing left indicate asynchrony in values and arrows pointing down indicate that the E_{FI} fluctuates before the E_{IF} by a quarter period ($90°$).