

# Generic Model to Predict the Outbreak of Insects in European Forests <sup>†</sup>

Dorian Collot <sup>1,\*</sup> and Christelle Robinet <sup>2</sup>

<sup>1</sup> INRAE Val de Loire, site d'Orléans, Unité de Recherche de Zoologie Forestière (URZF) 2163 Avenue de la pomme de pin - CS 40001 – Ardon 45075 Orléans cedex 2, France ; dorian.collot@inrae.fr

<sup>2</sup> INRAE Val de Loire, site d'Orléans, Unité de Recherche de Zoologie Forestière (URZF) 2163 Avenue de la pomme de pin - CS 40001 – Ardon 45075 Orléans cedex 2, France ; christelle.robinet@inrae.fr

\* Correspondence: dorian.collot@inrae.fr

† Presented at the 1st International Electronic Conference on Entomology (IECE 2021), 1–15 July 2021; Available online: <https://iece.sciforum.net/>.

**Abstract:** Insect pests are one of the major threats to forests. Although invasive species cause more and more impacts, native species could also become real pests. The population dynamics of insects relies on several factors, going from weather to stand conditions. Due to global change, insects could face conditions they have never encountered, leading to unusual population outbreaks. Forest managers need to consider these possible emergent pests. However, the biology of these new pests is generally poorly described and predicting insect outbreaks is thus very challenging. In this context, we have developed a generic model of emergence to describe local outbreaks. This model describes the probability of occurrence of an outbreak at a given time and at a given area, based on several conditions (34 variables). It has been built and parametrized on different orders of European forest pests. This parametrization allows obtaining species profiles that can be used as a baseline to make predictions even if poor data are available on the pest, to ensure the genericity of the model. This is to our knowledge the very first generic outbreak model that has been developed so far. This model was coded in R and a user-friendly version using a shiny app was developed. In this work, we are going to present the model and its validation.

**Keywords:** Generic model; pest; probability of outbreak; Europe; Climate

**Citation:** Collot, D.; Robinet, C. Generic Model to Predict the Outbreak of Insects in European Forests. *Proceedings* **2021**, *68*, x. <https://doi.org/10.3390/xxxxx>

Published: 30 June 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Pest outbreaks are major threats for forests [1]. Global changes can affect the occurrence of outbreak: their frequency could be modified [2], and species could make outbreak in new places [3].

Predicting the occurrence of such events would help to design more efficient ways to manage pests and thus to protect the forest. By anticipating the year an outbreak is expected, or the place where it could occur, the surveillance could be optimized accordingly. In the same way, monitoring only a few species, the more likely to outbreak could contribute to the management optimization.

Outbreaks rely on several environmental factors [4], and modelling can be used to sum up the biological knowledge and to make predictions. Various models have been developed to tackle this question (for instance [5][6][7]). These models can successfully predict the occurrence of outbreak but they only focus on some variables. Capitalizing on all these studies could allow to have a more general understanding of potential outbreak drivers, and predict them even in new conditions.

The model should also be designed to be easily applied to less studied species. In fact, some species are particularly well described such as *Ips typographus* [6]–[8], because they cause damages for a long time. However, due to climate changes, other species could

make more damages, for which less data are available. A generic model could therefore help to make prediction for this species, and help to manage them before they become a threat.

In this work, we have built a model to compute a probability of outbreak occurring based on environmental factors reported to be involved in forest insect outbreaks. The parametrization of the model was made for four orders of forest insect pests, focusing on European forests.

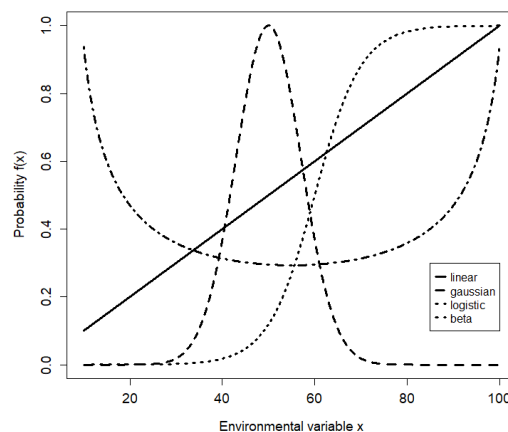
## 2. Methods

### 2.1. Mathematical model

The model is inspired by data fusion methods used in imagery [8]. For each environmental variable  $x_i$ , a probability that an outbreak occurs is calculated. A total probability  $P(outbreak | (x_1, \dots, x_n))$  is then calculated by averaging these probabilities, leading to:

$$P(outbreak | (x_1, \dots, x_n)) = g(f_1(x_1), \dots, f_n(x_n)), \tag{1}$$

The functions  $f_i$  have to provide a value between 0 and 1. In this work, four type of functions, described on Figure 1 have been used. These functions encompass a wide range of responses to environment.



**Figure 1.** The four type of functions  $f_i$  used in the model.

The function  $g$  is defined as the geometric means of the different  $f_i$ . This model has been implemented in R [10].

### 2.2. Variables of the model

To develop a generic model, we selected 34 input variables. These variables have been selected based on their accessibility (i.e., datasets easy to retrieve), and their expected effect on outbreak.

Climatic data have been retrieved from the dataset CRU-TS 4.03 [11] downsampled with WorldClim 2.1 [12]. This dataset contains the monthly sum of precipitations and average monthly minimal and maximal temperatures from 1960 to 2018 throughout Europe. A total of 13 climatic variables have been defined: the mean temperature on each trimester, the sum of precipitation on each trimester, the de Martonne aridity index for each trimester, and the lowest annual temperature. The mean temperature was computed using the average monthly minimal and maximal temperature and then averaged by trimester. The aridity index  $I_{dM}$  is defined as [13]:

$$I_{dM} = (\text{sum of precipitation}) / (\text{mean temperature} + 10), \tag{1}$$

Each climatic variable is considered for the year in which we aim to estimate the probability of outbreak but also the previous year, leading to 26 climatic variables.

A variable described the richness of the soil. The data was retrieved from the FAO/UNESCO dataset [14]. Three variables described the topography: the altitude, the slope and its orientation. The altitude was based on the SRTM (Shuttle Radar Topography Mission) data, found on the WorldClim website. The slope (in %) and its direction were computed using the altitude. Forest covering is described by three variables: the presence of conifer, the presence of deciduous tree, and the density of the host of the pest. This data has been retrieved from the EFI (European Forest Institute) [15]. Therefore, a total of 7 variables describes the stand conditions.

The 34<sup>th</sup> variable is the occurrence of an outbreak on the site the previous year.

### 2.3. Parameters inference and validation of the model

Recorded outbreaks have been extracted from DSF database (French Forest Health Department, depending on the Ministry of Agriculture). In order to have enough outbreak data to estimate the parameters correctly, the species have been grouped by orders. Four orders have been studied in this work: Lepidoptera, Coleoptera, Hemiptera and Hymenoptera.

ABC-SMC methods [16] were used to infer the parameters values for the different order of pests.

The prediction of the model has then been compared to outbreak recorded in the literature. The Hymenoptera profile has been compared to data of *Diprion pini* [6], the Lepidoptera profile to data of *Dendrolimus pini* [18], the Hemiptera to *Elatobium abientinum* [19] and the Coleoptera profile to data of *Ips typographus*. For each profile, the Area Under the ROC Curve (AUC) has been calculated.

## 3. Results

The model predicts the probability for an outbreak to occur based on several environmental variables. The parameters estimation has been made for four pest orders (Lepidoptera, Coleoptera, Hymenoptera and Hemiptera), leading to four generic profile.

The different profiles have then been tested on the data found in the literature. The data used for the validation are different from the ones used for estimating the parameters to ensure an independent validation. The ROC curves and their AUC are represented in Figure 2. The AUC is between 0.57 and 0.75. This score is quite good, especially for a generic model. The generic outbreak model fits the best for the Lepidoptera profile and the lowest for the Hemiptera profile.

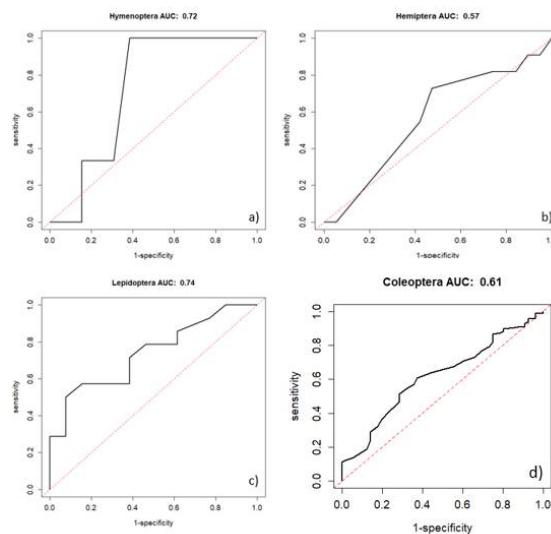


Figure 2. ROC curves and AUC for each pest profiles.

#### 4. Discussion

In this work, we build a generic model to predict the outbreak of the forest pest and parametrize it for different pest orders. The results are quite good, considering that the model is generic. The accuracy is limited by the need to consider any species of the same order. The model is either able to determine years unsuitable for an outbreak, as for Hymenoptera, or to predict “suitable” year, as for Lepidoptera. This difference could be explained by the effect of the variable considered. If a variable is a driver, and trigger the outbreak, the model will have a higher sensitivity, but if the variable is an “inhibitor” and prevent the outbreak, the model will have a higher specificity.

Two directions should be considered to improve this generic outbreak model.

First the profile could represent smaller clades. Orders cover a huge diversity of species that could behave differently to environmental variables. Families or genus gather less species, but with more similar life-history. Considering such smaller clades would allow refining the parametrization. However, the model being too specific in that case would be less generic and therefore less useful when considering any kind of forest insect pests as we have not enough data to do this for all possible small clades. The optimal way to group and describe species outbreak is a tricky question.

Secondly, the variables could be selected differently. Here the variable has been selected for their easy accessibility. As a consequence, these variables are average or proxy or other variables. For instance, the evapotranspiration is often relevant, but is difficult to measure. This variable is linked to the temperature and the precipitation, which are considered in the model. Considering the real driver rather than proxy could increase the model accuracy. This problem is easy to fix in this model, given that the dataset is available. Since the variables are independent, a variable can be added (or removed) without reperforming the estimation of the parameters. This is one of the strengths of this generic model. This generic outbreak model can therefore also be used for well-studied species with a refined approach. The user can use the profile as a baseline, and add variables (and their parameters) in order to increase the accuracy of the model.

This model could be useful for management of forest stand. The model could help to identify a group of species to control. It could also help to discard some years or some areas unsuitable for outbreaks: monitoring efforts could be reduced to save time and resources. In contrast, the model could be used to manage a stand, and could provide, for given year, the most likely pest order to outbreak. The management strategy could then be adapted depending on the type of pest to reallocate monitoring efforts.

#### 5. Conclusions

We built a generic model to predict the probability of outbreak of European forest pests and we parametrized it for different pest orders. This model allows to make predictions and could help to manage forest stands. This work was done in the frame of the European project HOMED and will feed a toolbox for the assessment and management of emerging or invasive pests of European forests.

#### References

1. T. D. Ramsfield, B. J. Bentz, M. Faccoli, H. Jactel, and E. G. Brockerhoff, “Forest health in a changing world: Effects of globalization and climate change on forest insect and pathogen impacts,” *Forestry*, vol. 89, no. 3, pp. 245–252, 2016, doi: 10.1093/forestry/cpw018.
2. K. J. Haynes, A. J. Allstadt, and D. Klimetzek, “Forest defoliator outbreaks under climate change: Effects on the frequency and severity of outbreaks of five pine insect pests,” *Glob. Chang. Biol.*, vol. 20, no. 6, pp. 2004–2018, 2014, doi: 10.1111/gcb.12506.
3. M. Matek and M. Pernek, “First record of *Dendrolimus pini* outbreak on aleppo pine in Croatia and severe case of population collapse caused by entomopathogen *Beauveria bassiana*,” *South-East Eur. For.*, vol. 9, no. 2, pp. 91–96, 2018, doi: 10.15177/see-for.18-17.
4. L. Rasche and R. A. J. Taylor, “A generic pest submodel for use in integrated assessment models,” *Trans. ASABE*, vol. 60, no. 1, pp. 147–158, 2017, doi: 10.13031/trans.11931.
5. P. Baier, J. Pennerstorfer, and A. Schopf, “PHENIPS-A comprehensive phenology model of *Ips typographus* (L.) (Col.,

- Scolytinae) as a tool for hazard rating of bark beetle infestation," *For. Ecol. Manage.*, vol. 249, no. 3, pp. 171–186, 2007, doi: 10.1016/j.foreco.2007.05.020.
6. K. Möller, R. Hentschel, A. Wenning, and J. Schröder, "Improved outbreak prediction for common pine sawfly (*Diprion pini* L.) by analyzing floating 'climatic windows' as keys for changes in voltinism," *Forests*, vol. 8, no. 9, 2017, doi: 10.3390/f8090319.
  7. M. Kosunen, T. Kantola, M. Starr, M. Blomqvist, M. Talvitie, and P. Lyytikäinen-Saarenmaa, "Influence of soil and topography on defoliation intensity during an extended outbreak of the common pine sawfly (*Diprion pini* L.)," *IForest*, vol. 10, no. 1, pp. 164–171, 2017, doi: 10.3832/ifer2069-009.
  8. Bloch I. and H. Maitre, "Fusion de données en traitement d'images: modèles d'information et de décision," *Trait. Du Signal*, vol. 11, no. 6, pp. 435–446, 1994.
  9. I. Bloch, "Information Combination Operators for Data Fusion: A Comparative Review with Classification," *Trans. Syst. MAN, Cybern.*, vol. 26, no. 1, pp. 52–67, 1996.
  10. R Core Team, "R: A Language and Environment for Statistical Computing." 2018.
  11. I. Harris, P. D. Jones, T. J. Osborn, and D. H. Lister, "Updated high-resolution grids of monthly climatic observations - the CRU TS3.10 Dataset," *Int. J. Climatol.*, vol. 34, no. 3, pp. 623–642, 2014, doi: 10.1002/joc.3711.
  12. S. E. Fick and R. J. Hijmans, "WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas," *Int. J. Climatol.*, vol. 37, no. 12, pp. 4302–4315, 2017, doi: 10.1002/joc.5086.
  13. G. Pellicone, T. Caloiero, and I. Guagliardi, "The De Martonne aridity index in Calabria (Southern Italy)," *J. Maps*, vol. 15, no. 2, pp. 788–796, 2019, doi: 10.1080/17445647.2019.1673840.
  14. P. A. Sanchez, W. Couto, and S. W. Buol, "The fertility capability soil classification system: Interpretation, applicability and modification," *Geoderma*, vol. 27, no. 4, pp. 283–309, 1982, doi: 10.1016/0016-7061(82)90019-2.
  15. D. J. Brus *et al.*, "Statistical mapping of tree species over Europe," *Eur. J. For. Res.*, vol. 131, no. 1, pp. 145–157, 2012, doi: 10.1007/s10342-011-0513-5.
  16. T. Toni, D. Welch, N. Strelkowa, A. Ipsen, and M. P. H. Stumpf, "Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems," *J. R. Soc. Interface*, vol. 6, pp. 187–202, 2009, doi: 10.1098/rsif.2008.0172.
  17. M. a Beaumont, W. Zhang, and D. J. Balding, "Approximate Bayesian computation in population genetics.," *Genetics*, vol. 162, no. 4, pp. 2025–35, Dec. 2002.
  18. A. Lesniak, "CLIMATIC AND METEOROLOGICAL CONDITIONS OF THE PINE MOTH ( *DENDROLIMUS PINI* L .) OUTBREAKS," *Ekol. Pol.*, vol. 24, no. 4, pp. 515–547, 1976.
  19. S. A. Estay, M. Lima, F. A. Labra, and R. Harrington, "Increased outbreak frequency associated with changes in the dynamic behaviour of populations of two aphid species," *Oikos*, vol. 121, no. 4, pp. 614–622, 2012, doi: 10.1111/j.1600-0706.2011.19525.x.