



**HAL**  
open science

# Predicting the seasonal flight activity of *Myzus persicae* , the main aphid vector of Virus Yellows in sugar beet

Martin Luquet, Sylvain Poggi, Christelle Buchard, Manuel Plantegenest, Yann  
Tricault

## ► To cite this version:

Martin Luquet, Sylvain Poggi, Christelle Buchard, Manuel Plantegenest, Yann Tricault. Predicting the seasonal flight activity of *Myzus persicae*, the main aphid vector of Virus Yellows in sugar beet. *Pest Management Science*, 2023, 10.1002/ps.7653. hal-04161101

**HAL Id: hal-04161101**

**<https://hal.science/hal-04161101>**

Submitted on 3 Aug 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

LUQUET Martin (Orcid ID: 0000-0002-4656-4923)  
Tricault Yann (Orcid ID: 0000-0002-7778-5163)

## **Predicting the seasonal flight activity of *Myzus persicae*, the main aphid vector of Virus Yellows in sugar beet**

Forecast of seasonal aphid flight activity in a sugar beet production area

LUQUET Martin<sup>1\*</sup>, POGGI Sylvain<sup>2</sup>, BUCHARD Christelle<sup>2</sup>, PLANTEGENEST Manuel<sup>3</sup>, TRICAULT Yann<sup>1</sup>

<sup>1</sup> IGEPP, INRAE, Institut Agro, Univ Rennes, 49045 Angers, France

<sup>2</sup> IGEPP, INRAE, Institut Agro, Univ Rennes, 35653 Le Rheu, France

<sup>3</sup> IGEPP, INRAE, Institut Agro, Univ Rennes, 35000 Rennes, France

\* Corresponding author: Martin Luquet, INRAE – UMR 1349 IGEPP, L'Institut Agro Rennes-Angers, 2  
rue Le Nôtre, F-49045 Angers, +33 2 41 22 54 73, martin.luquet.pro@gmail.com

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the [Version of Record](#). Please cite this article as doi: [10.1002/ps.7653](https://doi.org/10.1002/ps.7653)

This article is protected by copyright. All rights reserved.

## ABSTRACT

### BACKGROUND

Virus Yellow (VY), a disease caused by several aphid-borne viruses, is a major threat to the global sugar beet production. Following the ban of neonicotinoid-based seed treatments against aphids in Europe, increased efforts are needed to monitor and forecast aphid population spread during the sugar beet growing season. In particular, predicting aphid flight seasonal activity could allow anticipation of the timing and intensity of crop colonisation and contribute to the proper implementation of management methods. Forecasts should be made early enough to assess risk, but can be updated as the season progresses to refine management. Based on a long-term suction-trap dataset gathered between 1978 to 2014, we built and evaluated a set of models to predict the flight activity features of the main VY vector, *Myzus persicae*, at any location in the French sugar beet production area (circa  $4.10^5$  ha). Flight onset dates, length of flight period and cumulative abundance of flying aphids were predicted using climatic and land-use predictors as well as geographical position.

### RESULTS

Our predictions outperformed current models published in the literature. The importance of the predictor variables varied according to the predicted flight feature but winter and early spring temperature always played a major role. Forecasts based on temperature were made more accurate by adding predictors related to aphid winter reservoirs. In addition, updating the model parameters to take advantage of new weather data acquired during the season improved the flight forecast.

### CONCLUSION

Our models can be used as a tool for the mitigation in sugar beet crops.

## KEYWORDS

Epidemiological surveillance, Forecasting, *Myzus persicae*, Regression model, Sugar beet, Virus Yellow

Accepted Article

## 1. INTRODUCTION

Virus Yellows (VY) is one of the major economic threats to sugar beet production worldwide, and can lead to substantial yield losses<sup>1,2</sup>. The disease is caused by aphid-borne viruses, with the green peach aphid *Myzus persicae* being the most efficient vector<sup>1,3</sup>. Disease management has mainly focused on vector control with aphicides. In Europe, over the last three decades, neonicotinoids (NNIs) and in particular their use in seed coatings have been very effective in controlling aphids and preventing VY outbreaks<sup>3</sup>. However, the adverse effects of these insecticides on non-target organisms have been well documented for bees<sup>4</sup> and other invertebrates and vertebrates<sup>4,5</sup>. Their deleterious effects on pollinators led the European Union to ban the use of NNIs on outdoor crops, including sugar beet, in 2018. Following this ban, a resurgence of the disease is underway in various European countries<sup>6,7</sup>. This is the case in France, the world's second largest sugar beet producer<sup>8</sup>, as shown by the aphid and yellows observation data collected on untreated sugar beet crops over the last 3 years<sup>9–11</sup>). To counter the risk of collapse of the French sugar beet industry, the use of NNIs-coated seeds has been reauthorized on a derogation basis for the 2021 and 2022 growing seasons<sup>12</sup>. Following a decision of the European Court of Justice<sup>13</sup>, these emergency exceptions can no longer be implemented. There is therefore an urgent need to develop effective new tools, such as risk prediction models, to monitor and control the disease and its main vector.

The susceptibility of sugar beets to VY-carrying aphids decreases as the plant develops<sup>14</sup> with early and heavy infestations of this spring crop resulting in higher viral infection rates and yield losses<sup>15,16</sup>. Thus, the timing and intensity of crop colonisation by winged aphids are key parameters for disease development. Accurate forecasting of seasonal aphid flight activity is thus a key step in anticipating VY risks. Depending on populations, *M. persicae* overwinter on peach trees as eggs or on winter crops or wild plants as parthenogenetic females<sup>17</sup>. Winged migrants originating from these reservoirs colonise crops (sugar beet but also many other crops<sup>18</sup>) in spring. During spring and summer, aphid populations reproduce continuously and spread in and among crops. Finally, aphids migrate back to their winter

Accepted Article

hosts in autumn<sup>17</sup>. The phenology of aphid flights thus follows seasonal patterns that can be described by their onset dates, duration and cumulative abundance of the aerial aphid pool<sup>19,20</sup>. Predicting these flight features allows summarising aphid flight activity in order to forecast VY prevalence: indeed, it has been shown that *M. persicae* flight onset dates and abundance are tightly linked to VY incidence at a regional scale in the UK<sup>21,22</sup>.

In order to deploy disease mitigation strategies, predictions of aphid flight activity should be made sufficiently early, using relevant predictors. The role of winter temperature is known to modulate flight onset dates, duration and abundance<sup>20,22-24</sup>, likely through its effects on the survival and development rates of source populations<sup>24</sup>. Land use may also significantly determine aphid flight activity<sup>25</sup>. The number of flying migratory *M. persicae* is, for instance, positively correlated with the surrounding area planted with rapeseed, which is thought to be a major winter reservoir for this pest<sup>23</sup>. Aphid flight phenology is also related to geographic location<sup>20,23,24</sup>. Thus, different sources of information must be combined to accurately anticipate seasonal aphid activity at a time when measures can still be taken to reduce aphid infestation or VY transmission to beets.

Ideally, the accuracy of the forecasting system should improve as the season progresses (e.g. integrating new weather data) to support farmers' operational choices in real time. Early information could guide the choice of variety to be sown, as resistant or tolerant varieties should soon be available<sup>2</sup>.

In the French beet production area (Fig. 1), this decision is usually made by mid-February. Later, the modulation of the sowing date and the choice of installing a companion plant (made in early March in France), two methods likely to limit colonisation by aphids, can be implemented<sup>26,27</sup>. Finally, aphid flight forecasts could help in deciding on the use of curative measures (e.g. insecticides, biocontrol products)<sup>27,28</sup> after beets have emerged and during their period of highest susceptibility to VY. In France, this usually occurs in mid-April in the main beet production area.

In our study, we aimed at predicting the seasonal migration of *M. persicae* at any location in the French sugar beet production area. We predicted three different features summarizing seasonal aphid flight activity (flight onset date, cumulative abundance, flight duration), and for three key periods where

Accepted Article

strategic decisions can be made by beet growers to mitigate the effects of VY (choice of the variety, sowing date and implementation of curative treatments). Model development and parameter estimation were based on aerial aphid activity data collected by a French network of suction traps<sup>29</sup> that operated for more than 30 years. We built and compared the performance of several regression models including different sources of information (climate data collected over various time windows, land-use data and geographical position). We assessed the relative importance of each factor in predicting aphid flights and identified the best models. We verified the improved predictive power of our models compared to the regression model described by Qi et al. (2004)<sup>22</sup>, updated in Dewar & Qi (2021)<sup>6</sup>, adapted here to French data and based on winter temperature only. With this analysis, we contribute to the development of forecasting tools for the epidemiological surveillance of virus yellows at different times of the beet growing season.

## 2. MATERIAL AND METHODS

### 2.1 Study area

The study focused on the French sugar beet production area, which covers approximately 400 000 ha cultivated annually<sup>8</sup> in five administrative regions concentrated in the North and East of France (Fig. 1). The landscape in this area is dominated by arable crops, although permanent grasslands are well represented in the west (Normandy region). The climate, temperate oceanic according to Köppen classification (*Cfb* class<sup>30</sup>, average annual temperature:  $10.6 \pm 0.8^\circ\text{C}$ , average annual rainfall:  $738 \pm 110$  mm), varies both longitudinally and latitudinally and is characterised by rather mild winters compared to the rest of France (Fig. 1). During winter (November to March), the most contrasting sites differ by a maximum of  $3.1^\circ\text{C}$  for the average temperature and 76.8 mm for the average cumulative rainfall.

< Figure 1 >

## 2.2 Aphid data

### 2.2.1 The Agraphid network

Aphid data used are from the Agraphid trap network<sup>29</sup> that was part of the former European networks Euraphid<sup>31</sup> and EXAMINE<sup>32</sup>. This historical network of 20 suction traps located in France, Belgium and Switzerland (Fig. 1) was launched in 1977 and has been capturing aphids and other insects since then. However, the number of traps in operation has varied over time and only one trap is still in permanent operation in 2022, located in western France (Fig. 1), outside the sugar beet production area. These 12.2m suction traps<sup>33</sup> were designed to catch representative samples of the aerial aphid population from a wide area (about 80 to 100 km around<sup>33,34</sup>). Individuals of more than 200 aphid species were counted and identified daily during the period of operation of each trap.

Of the 20 traps included in the network, 8 were located in the French sugar beet production area. We also considered the two Belgian traps to belong to this zone because of their proximity. These 10 traps operated for varying lengths of time over a 36-year period (1978 to 2014, Fig. 1), for a total of 156 trap-years. These data were used for model fitting and validation (Fig. 1, see §2.4.2). Aphid flight data from the 10 remaining traps outside the region of interest (181 trap-years) were also used to test model performances in environmental conditions that are different from those in which they were trained (Fig. 1, see §2.4.2). This was done as a way to evaluate their potential transferability outside of the area of interest.

### 2.2.2 Features of seasonal flight activity

For each trap-year, we extracted daily counts of winged *M. persicae* from the Agraphid data, and calculated the features summarising seasonal flight activity in spring and summer as follows:

- The onset of the flight period was estimated by the date of first capture (*D1c*) or the date of fifth capture (*D5c*). While *D1c* has been used in similar studies (see *e.g.*<sup>6,22</sup>), *D5c* is more



robust to winter by-catch<sup>25</sup>. Both variables were expressed as the year number of days since January 1 (excluding February 29<sup>th</sup> for leap years, as the database is coded as such).

- The cumulative abundance of flying aphids (*Ab*) was calculated as the sum of trap catches between January 1 and June 30. This period has been shown to be optimal for predicting beet yellows in England<sup>21</sup> and has been retained here for comparison with previous studies<sup>6,22</sup>. While flights can occur up to the end of August in the French sugar beet production area<sup>17</sup>, extending the period over which daily abundances are summed did not change the results significantly (results not shown here).
- The total duration of spring and summer flights (*Fd*) was estimated following Bell et al. (2015)<sup>20</sup>, by computing the number of days between the 5th and 95th percentiles of total aphid captures between weeks 10 (early March; no aphid was captured before this week in the area of interest) to 35 (late August; latest potential end of summer flight<sup>17</sup>).

We kept only reliable data and discarded all those that seemed doubtful (e.g. due to a mechanical problem with the trapping system - see Appendix S1 for details about data exclusion). Where necessary, we imputed missing daily abundance values by rolling averages over the previous 7 days and the following 7 days. Because of some variation in the timing of trap malfunctioning events, this led to slight differences in the number of available trap-years among the variables calculated, but these were always greater than 292 and 126 for all traps and for those located in the sugar beet area, respectively (Table S1.4 in Appendix S1).

## 2.3 Environmental data

### 2.3.1 Meteorological data

The SAFRAN analysis system<sup>35,36</sup> developed by the French national meteorological service Météo-France provides time series of different near-surface meteorological variables on an 8km grid across

France and part of the neighbouring countries (including Belgium) by spatial interpolation. Daily mean temperature and cumulative rainfall were extracted from SAFRAN for each 8km x 8km cell containing a suction trap for the whole period of interest (1978-2014). These data were used to compute various meteorological variables, detailed in section 2.4.2.

### 2.3.2 Land-use data

Data on land areas covered by forest, grassland and crops come from the French<sup>37-39</sup> and Belgian<sup>40</sup> national censuses. They were extracted at the NUTS-3 scale for France (“departments”) and at the NUTS-2 scale for Belgium (“provinces”). These administrative areas are of similar size, ranging from 2284 to 8169 km<sup>2</sup> in our study area. From these data, we calculated several variables identified as potential predictors of *M. persicae* flight activity in literature<sup>23,24</sup>): the semi-natural land area (*SN\_area*), which is the sum of the areas covered by forest or permanent grassland, the total crop area (*Crop\_area*), which is the sum of the areas covered by arable land (including temporary crops, temporary meadows and fallows) or perennial crops, and the area occupied by rapeseed crops (*RSC\_area*), a suspected winter reservoir for *M. persicae*<sup>23</sup>. It should be noted that rapeseed crops were also included in *Crop\_area* where they accounted for  $4\% \pm 0.03$  (SD) of the total area during the study period, but with weak correlation between rapeseed crop area and total crop area ( $r = 0.17$ ). For some years in France, where rapeseed area was missing in the national Agreste<sup>37-39</sup> census, data provided by the French Institute of Oilseeds and Protein Crops, Terres Inovia, were used instead. To homogenise the estimates between traps, the land-use variables (semi-natural land, crops and rapeseed) were calculated within an 80 km radius of the trap, by averaging the values for each department/province, weighted by its relative area in the buffer zone. This radius value was set to the size of the area for which aphids captured by a suction trap are considered representative<sup>33</sup>. In the absence of information on land use (Germany, Luxembourg, Spain), the missing part was excluded and the land-use variables were calculated on the remaining area. This exclusion procedure concerned 5 traps; in the worst case, 35% of the circle area was excluded (westernmost trap).

## 2.4 Modelling

### 2.4.1 Overview

Models were fitted to predict flight features (response variables) at each trap-year on trap data pertaining to the sugar beet production area (Fig. 1, Table S1.3). For each response variable, 3 alternative models were fitted (detailed in the next section), which combined different sets of local predictors including meteorological, land-use and geographical variables. Each model was then fitted in 3 different versions, in which the meteorological predictors were computed over optimal time windows determined over 3 different periods. These periods ended on average dates on which VY disease management decisions can be made. Although these strategic deadlines can vary from one region to another, they are relatively homogeneous in the French sugar beet production area (*French Sugar Beet Technical Institute, ITB, pers. comm.*). The first deadline was mid-February (set to the 14<sup>th</sup>), the deadline for farmers to choose their beet varieties. The second one was early March (set to the 5<sup>th</sup>), the deadline for farmers to adjust the sowing dates and/or decide to sow a companion crop. The third deadline was mid-April (set to the 15<sup>th</sup>), when the first leaves appear in the field. This is when sugar beets become susceptible to yellows viruses and farmers may decide to apply curative treatments.

### 2.4.2 Model fitting and model performance

We fitted separate linear models to predict the start of the flight period ( $D1c$  or  $D5c$ ), its duration ( $Fd$ ) and the log-transformed cumulative abundance of flying aphids ( $\log + 0.01$  – the transformed variable was named  $\log Ab$ ). For each flight feature (response variable) and each prediction date (February 14, March 5 and April 15), the following models were fitted (Table 1):

*Model M<sub>1</sub>: model by Qi et al. (2004) applied to the French context*

$M_1$  was based on the approach developed by Qi et al.<sup>22</sup> for England and re-used by Dewar and Qi (2021)<sup>6</sup>. Following these authors, the flight features obtained from the Agraphid dataset were predicted from the daily mean temperature averaged over the period January 1 to February 14.

*Model M<sub>2</sub>: prediction from thermal accumulation over a critical period (optimized winter time window)*

$M_2$  is similar to the  $M_1$  model, but considers all possible time windows over which to calculate the degree-day accumulation. For each response variable, we fitted all possible regressions with the sum of degree-days above 0°C calculated over any time window between 1 November and each prediction date (deadline) detailed above. Time windows of all possible lengths (including one day length) were considered, resulting in more than 13000 models for each response variable. Additionally, we also tested a 4°C threshold as an alternative to the 0°C threshold, as it is the developmental threshold for *M. persicae* estimated by Whalon & Smilowitz (1977)<sup>41</sup>. However, using this threshold reduced model predictive accuracy (results not shown). Model performance was assessed using cross-validation (see below) and the time window that minimized the root mean squared error of prediction (RMSE<sub>p</sub>, see below) was selected, hereafter named “critical period”<sup>42</sup>.

*Model M<sub>3</sub>: prediction from meteorological variables, land-use variables and geographical position*

The  $M_3$  model included meteorological variables (temperature and rainfall), land-use variables and geographical coordinates as predictors (Table 1), selected following a review of literature<sup>20,22,23,25,43-45</sup>. Again, the models were fitted with all possible time windows to calculate meteorological variables. However, for the sake of computational time, the same window was used for temperature (cumulative degree-days) and mean rainfall. For each response and prediction date, the best model was selected according to the RMSE<sub>p</sub> criterion following cross-validation.

For model selection, a 10-fold cross-validation procedure, repeated 3 times<sup>46</sup> was applied. Two metrics were calculated on held-out samples: the root mean squared error of prediction, RMSE<sub>p</sub>, and the normalized root mean squared error of prediction, NRMSE<sub>p</sub>. NRMSE<sub>p</sub> is a standardized metrics that

allows for comparison of model prediction performance for different response variables independently of their scale. It was calculated by dividing  $RMSE_p$  by its interquartile range (i.e. the difference between 25<sup>th</sup> and 75<sup>th</sup> percentile<sup>47</sup>). The models exhibiting the minimal  $RMSE_p$  were selected and fitted on the full dataset and the RMSE of the final models was calculated. In models  $M_3$ , the relative importance of predictors was assessed using model coefficients standardised by partial standard deviation<sup>48,49</sup>. In addition, to examine the spatial variation in performance, we calculated the RMSE for each trap used to fit the models. To assess model performance when increasing distance from the French sugar beet production area, we also predicted flight features for traps located outside the sugar beet production area (Table S1.3c) and computed the  $RMSE_p$  for each of these traps.

All analyses were carried out using the R version 4.2.1 programming environment<sup>50</sup>, using the tidyverse package version 1.3.2<sup>51</sup> for data curation, the caret package version 6.0-93<sup>52</sup> for cross-validations and the MuMIn package version 1.47.1<sup>53</sup> for the calculation of partial standardized coefficients. The management and analysis of spatial data were conducted the sf package (version 1.0-8)<sup>54</sup> and the rmapshaper package (version 0.4.6)<sup>55</sup>. Data visualisation was done thanks to ggplot2 (3.4.2)<sup>56</sup>, cowplot (1.1.1)<sup>57</sup>, ggpubr (0.4.0)<sup>58</sup>, RColorBrewer (1.1-3)<sup>59</sup>, ggh4x (0.2.2)<sup>60</sup>, scales (1.2.1)<sup>61</sup>, ggson (0.5.0)<sup>62</sup> and ggnewscale (0.4.8)<sup>63</sup> packages.

## 3. RESULTS

### 3.1 Model comparison

Table 2 displays the root mean squared error of prediction ( $RMSE_p$ ) values obtained for the 3 alternative models ( $M_1$  to  $M_3$ ), and the 4 flight features (the two flight onset dates proxies  $D1c$  and  $D5c$ , flight duration  $Fd$  and log-cumulative abundance  $log\_Ab$ ), at the 3 selected prediction dates.  $RMSE_p$  values showed that  $M_2$  and  $M_3$  models almost always outperformed the  $M_1$  model (Table 2), except for the log-cumulative abundance prediction on the first prediction date. Models containing

meteorological, land-use and geographical covariates ( $M_3$ ) always outperformed the models including temperature predictors only ( $M_1$  and  $M_2$ , Table 2).

Predictions also tended to improve with delaying prediction dates for all response variables (Table 2), although the performance improvement was inferior to the one brought by the choice of a better model (e.g.  $M_3$  over  $M_2$ , Table 2). In the most accurate models ( $M_3$ ), using the later prediction date allowed reducing  $RMSE_p$  by 1.7 days for  $D5c$ , 1.8 days for  $Fd$  and 0.2 for  $logAb$  (corresponding to an approximate 20% reduction in abundance estimation error).

< Table 2 >

Examination of the normalized root mean squared error of prediction ( $NRMSE_p$ ) criterion (Table 3) showed that the logarithm of the cumulative abundance ( $logAb$ ) was the flight feature for which predictions were the least accurate, followed by flight duration ( $Fd$ ) and flight onset proxies ( $D1c$  and  $D5c$ ). Prediction accuracy for the date of 5<sup>th</sup> capture ( $D5c$ ) was substantially higher than for the date of 1<sup>st</sup> capture ( $D1c$ ), decreasing the  $NRMSE_p$  by 31-39% across prediction dates.

< Table 3 >

In subsequent analyses,  $D5c$  was retained as the sole proxy of flight onset. The relationships between observed  $D5c$ ,  $logAb$  and  $Fd$  and their predictions using the best models are shown in Fig. 2 (a, b, c). Predictions were overall linearly related to observations and matched their range for all response variables, with homogeneous accuracy across traps. However, a few log-cumulative abundance predictions departed from the expected  $y=x$  relationship for observations corresponding to no aphids being captured on a given trap-year (resulting in  $logAb = \log(0 + 0.01) = -4.6$ , Fig. 2c). Note that the linear relationship between predicted and observed abundance on the logarithmic scale implied that when predicted  $logAb$  were back-transformed to actual abundances, errors tended to increase for higher abundance values (Fig. 2d).

Accepted Article

For  $D5c$  and across all prediction dates, prediction errors (expressed as the difference between the observed and predicted values) ranged from -45 (45 days too early) to 40 (40 days too late, Fig. 2a). On average, respectively for the best model at each prediction date, predictions deviated from observed  $D5c$  by approximately 11, 12, and 13 days (*Mean Absolute Error* or MAE) in the Agraphid dataset. Note that these MAE values are lower than RMSE values displayed in Fig. 2a because the latter gives more weight to larger errors. For flight duration ( $Fd$ ), predictions ranged from 52 days too short to 76 days too long (Fig. 2b). On average, they deviated from the observations by approximately 15 to 16 days. Back-transformed  $\log Ab$  predictions matched the range of observed cumulative abundance ( $Ab$ ) values (from 0 to more than 2000 aphids annually caught in a trap, Fig. 2d). On average, these predictions mis-estimated the actual values by a factor of 3.2 (indicating that, on average, the estimated annual aphid abundance was 3.2 times either too low or too high) to 3.7, depending on the prediction date. However, misestimation of aphid abundance could be as high as a factor of 100. In some cases, these misestimations resulted in large discrepancies in terms of absolute numbers. For example, in one instance, the back-transformed cumulative abundance of aphids was predicted as 2126 instead of 286, and in another case, the predicted abundance was 2226 instead of 219.

< Figure 2 >

### 3.2 Relative importance of predictors

The relative importance of the predictor variables and the direction of the association with the response variables (i.e. flight features) were estimated using their partial standardised coefficients for each of the best models selected and each prediction date (Fig. 3). Cumulative temperature during the critical period ( $T_{crit}$ ) had the strongest influence on all flight features. Mild winter temperatures consistently led to early flight onset dates (lower  $D5c$ ), longer flight durations (higher  $Fd$ ) and more abundant aerial aphid populations (higher  $\log Ab$ ). Conversely, increased rainfall ( $R_{crit}$ ) was

associated to later, shorter and less abundant flights in most models, but the effects of rainfall were overall weaker than those of temperature. Land use predictors measured at 80km around the traps also influenced flight features. Increasing the area of semi-natural land (*SN\_area*) was associated with earlier flights, while increasing the area sown to rapeseed crops (*RSC\_area*) led to earlier, more abundant and moderately shorter flights. Finally, geographical position influenced response variables. Aphid numbers decreased from south to north (increased *Latitude*) and from west to east (increased *Longitude*). Flights also tended to be later and longer in the East (increased *Longitude* and *LonSq*). The product of latitude and longitude (*LatLon*) was retained in most models and was overall positively associated with log-cumulative abundance (*logAb*), negatively with flight duration (*Fd*) and had contrasting effects on flight onset date (*D5c*) depending on the prediction date.

< Figure 3 >

### 3.3 Meteorological critical windows

The critical periods over which temperature accumulation and mean rainfall were calculated in the best models were overall consistent across flight features (Fig. 4; see Fig. S2.1 in Appendix S2 for critical periods for all models). For the two first prediction dates (February 14 and March 5), the critical period started in early January (*D5c, Fd*) or late December (*logAb*) and ended on the prediction date or a few days earlier. When predictions were made on the last prediction date (April 15), the critical period started in mid-November (*D5c, Fd*) or late December (*logAb*) and ended two to three weeks before the April 15 deadline. In other words, weather conditions between late March and mid-April did not appear to influence the flight features of *Myzus persicae* observed later in the beet area.

< Figure 4 >



### 3.4 Spatial variation in prediction performance and assessment of model performance outside the sugar beet production area

For all response variables and prediction dates, the prediction errors (RMSE) were rather homogenous among the traps located in the sugar beet production area. In contrast, prediction accuracy tended to decrease for traps located outside of the sugar beet production area, which were not used to build models (Fig. 5). Moreover, prediction errors increased with distance to the South or West. The higher  $RMSE_p$  for the traps located outside the production area reached 5, 2.3 and 6.3 times the mean RMSE inside the production area, respectively for the flight onset date ( $D5c$ , most westerly trap), flight duration ( $Fd$ , most southern trap) and log-cumulative abundance ( $logAb$ , most south-westerly trap, Fig. 5). Using less performant models,  $M_1$  and  $M_2$ , allowed decreasing error in log-cumulative abundance estimation for traps not pertaining to the production area, with RMSE values similar to the ones obtained in the traps used to build the models (Fig. S3.1 in Appendix S3). This was not the case for  $D5c$  and  $Fd$ , however (Fig. S3.1).

< Figure 5 >

## 4. DISCUSSION

Using the French historical database of suction trap catches, we were able to accurately predict the seasonal flight activity features of *M. persicae* (flight onset date, duration, and log-cumulative abundance) from simple linear models based on one or a few predictors. Meteorological and land-use predictors, as well as geographical position, allowed capturing a significant part of the variation of the three response variables. There are few references on the performance of aphid flight prediction models but ours reached levels comparable to those proposed by Bourhis et al. (2021)<sup>64</sup>. Their neural network approach resulted in a  $RMSE_p$  of 20.0 days for the prediction of aphid flight onset dates (various species including *M. persicae*), while we obtained values ranging from 14.3 to 23.2 days,

depending on the proxy used for flight onset date, the prediction date and the predictors selected. Comparison of models including various response variables, predictors and prediction dates allowed us to identify possible sources of improvement in aphid flight modelling and its magnitude.

First, we showed that for a given set of predictors, prediction accuracy varied with the flight feature considered. Using meteorological and land-use predictors, flight onset variables were more accurately predicted than flight duration or log-cumulative abundance. Abundance was the least accurately predicted feature of aphid flight, sometimes leading to notable discrepancies between the observed and predicted values. This seems to confirm the conclusions of Bell et al. (2023)<sup>65</sup>, who recently found a low to moderate intrinsic predictability of *M. persicae* aerial numbers using suction-trap data in England. They suggested that the difficulty in predicting yearly aerial aphid numbers could be due to the multiplicity of factors acting on flying individuals, rendering annual time series highly stochastic<sup>65</sup>. Besides, we showed that for a given flight feature, the choice of the proxy influences the quality of the prediction. Indeed, the date of 5<sup>th</sup> capture (*D5c*) was systematically better predicted than the date of 1<sup>st</sup> capture (*D1c*). This could be due to incidental captures of single aphids before the formation of the migratory pool, leading to *D5c* being a more robust indicator of the flight onset than *D1c*<sup>25</sup>. This should be particularly relevant in warmer regions where flying aphids are occasionally observed in winter. In our analysis, predicting *D5c* rather than *D1c* led to a reduction in prediction errors by about 5 days, which shows the operational importance of the choice of the response variables used as proxies.

Although response variables differed in their predictability, refining the predictor set consistently improved their prediction. Depending on response variables, optimising the time window over which cumulative degree-days were computed, compared to the use of a fixed time window, reduced prediction error up to 20%. The gain in model performance obtained when optimising the time window increased when predicting later in the season, suggesting that newly acquired temperature data in late winter and early spring allowed bringing additional information on aerial aphid populations. In most models, the selected time window at least included the period from early-January to mid-February,

confirming the strong role of winter temperature in determining spring flight patterns<sup>23,24</sup>. This is likely due to the influence of temperature at this period of the year on the survival and development rates of overwintering parthenogenetic females<sup>24</sup>, and possibly on the hatching dates of overwintering eggs<sup>66</sup>. In later-predicting models, selected time windows consistently spanned until early April, probably reflecting the influence of early spring temperature on the production of winged individuals forming the aerial pool<sup>24</sup>. It is noteworthy that screening later than April 1<sup>st</sup> did not improve models, suggesting that temperature conditions are no longer an influential factor at this point.

Added to winter and early spring temperature, other environmental predictors significantly improved the accuracy of the predictions. Accounting for land-use, rainfall and geographical position improved predictions, but their relative importance depended on the considered response variables. Land use mainly affected the abundance of the *M. persicae* aphid pool, as well as the timing of the flight onset. The area of rapeseed was positively associated with the numbers of *M. persicae* trapped, as previously found<sup>23</sup>. It also had a positive influence on flight earliness. Rapeseed, is an overwintering host for *M. persicae* parthenogenetic populations in winter<sup>67</sup> and has been increasingly grown in France and other European countries since the end of the twentieth century<sup>8</sup>. Our results suggest that rapeseed crops are a source of spring migration, allowing aphid populations to quickly produce large numbers of flying aphids in early spring. However, there has been no evidence of rapeseed being a host plant of beet yellows viruses until now. The role of this aphid reservoir in the epidemiology of the disease therefore remains uncertain. Similarly, the positive influence of semi-natural land area on flight earliness could reflect the role of forests or grasslands as winter reservoirs for aphids. However, the host range of *M. persicae* is very broad, including many wild plants<sup>55</sup>, and it is difficult to assess the refuge habitat role of uncultivated areas without knowing their composition. Rainfall also influenced aphid flight features, as in a previous study<sup>23</sup>, with more rain leading to later, shorter and less abundant flights. This could reflect direct negative effects of precipitation on the survival and development of initial populations<sup>68</sup>, or on the take-off of winged aphids<sup>69</sup>, but also indirect effects through the modulation of crop growth or planting dates<sup>70</sup>. To limit the computation time, we imposed the same critical period for rainfall and

temperature. Future research is needed to disentangle the effects on aphids of these two climatic factors, which may also act through complex interactions<sup>70</sup>. Finally, the influence of geographical position variables confirmed that aphid flight features have a strong spatial structure<sup>45,71,72</sup>. This persists even when considering weather and land use predictors and could be due either to the effects of other spatially structured predictors not taken into account here (e.g. climatic factors at higher temporal scales than yearly weather), or to biological processes operating on a large spatial scale such as dispersal<sup>71,72</sup>. The relationship between these geographical predictors and response variables are purely correlative, which likely explain why models performed poorly when predicting outside of the spatial area for which they were developed. However, prediction performance was homogeneous across the French sugar beet production area, preventing the need to build several regional models. Overall, this suggests that a similar approach could be used in other regions; however, the poor performance of our models when increasing the distance indicates that they should be fitted using local data.

Finally, model performances improved over the season, encouraging a continuous in-season updating of meteorological information to predict aphid flight features. However, predictions made at the first deadline (February 14), corresponding to the time when farmers can choose the beet variety to sow, were almost as good as the predictions made at the last deadline (April 15). This suggests that accurate predictions of *M. persicae* flight activity can be produced as early as late winter to inform the Virus Yellows mitigation strategy, when many management actions can still be implemented<sup>26,27,73</sup>. In the near future, such predictions could be made even earlier by using seasonal weather forecasts, which already demonstrated their usefulness as a tool to predict pest population dynamics in other systems<sup>74</sup>. However, this strategy of risk prediction and mitigation assumes a strong correlation between aphid flight features and disease expression in the field because it ignores the viruliferous status of aphids<sup>21</sup>. Further research is needed to better understand the spatiotemporal dynamics of the different viruses involved in beet yellows and to integrate this knowledge into prediction models.

Despite the rather satisfactory accuracy of our models' predictions, from an academic point of view, several considerations have to be made regarding their real-life application. First, errors are not negligible from an operational point of view: for instance, the error in flight onset prediction ( $D5c$ ) was on average two weeks, which impedes the forecasting of a precise colonisation date at a particular geographical location. The same can be said for errors on flight abundance (on average a factor of  $e^{1.7} = 5.5$ ) and flight duration (20 days on average), which are even more complicated to directly translate in terms of economic risk (see also the last paragraph). However, our models could help anticipating whether flights will be particularly early, abundant or long in a given year: for instance, a two-week error is much lower than the range of possible variation from one year to another on a given site, which can be up to two months in our dataset. This is of great importance for risk assessment, as the susceptibility to Virus Yellows decreases during beet development<sup>14</sup>. In the same manner, our models can help to identify the patterns of geographical variation in aphid flight features across sugar beet production regions, as prediction errors were rather homogeneous across traps.

One other practical limitation of our modelling approach in an applied context is that it requires data that may not be available at the time prediction is needed. In particular, land-use data needed to run  $M_3$  model becomes available in national censuses only on  $n+1$  year (<https://agreste.agriculture.gouv.fr/agreste-web/methodon/S-SAA/methodon/>), and may be difficult to estimate from satellite data – in particular rapeseed crop area, usually detected from its flowers<sup>75</sup> that start blooming in April in Western Europe<sup>76</sup>. If practitioners such as the industry or technical institutes aimed at using such a tool for prediction, they would have to obtain it another way, e.g. through surveys filled by farmers (in our region of interest, it is often the same farmers who grow rapeseed and beet, and there are enough of them to be representative of the whole region). If land-use data is still missing, practitioners could try imputing them using data from previous years: for semi-natural and crop land area, this should usually not vary a lot from one year to another (never more than 8% in our data), but this cannot be ensured for oilseed rape crop area depending on the economic

context (up to 96% inter-year variation in our data). Otherwise the second most performant model, only relying on meteorological predictors, should be used instead.

Besides, despite our effort to rigorously measure model prediction performance, one should remain cautious about the accuracy of model predictions in a context of climate change. Indeed, our data was available only until 2014, but temperature and precipitation patterns are rapidly changing and will continue doing so in the next decades, leading to novel weather conditions that are not represented in our data. Insect populations exhibit complex and non-linear responses to climate changes, making them prone to shifts in their functioning that are very difficult to predict<sup>77,78</sup>. This additional layer of uncertainty should be kept in mind when predicting in new situations.

Finally, we need to improve our knowledge of the relationship between aphid flight features and colonisation rates in sugar beet fields. Indeed, it remains challenging to interpret predictions of aphid flight duration and abundance, along with their associated errors, from an operational standpoint. It is unclear how these predictions translate into agronomically relevant variables, such as aphid numbers within sugar beet fields, levels of Virus Yellows infestation, or the potential effects on yield. In a study on barley yellow dwarf disease, Fabre et al. (2010)<sup>79</sup> estimated an aphid deposition rate from the comparisons between ground observations and counts of winged individuals in suction traps. This approach could be applied to Virus Yellows. Predictions from our regression models could then be used as inputs in process-based models such as Werker et al. (1998)<sup>21</sup> to attain Virus Yellows level prediction in French sugar beet crops.

## ACKNOWLEDGEMENTS

This research was supported by the French SEPIM project as part of the French National Plan for Research and Innovation (PNRI). We are very grateful Fabienne Maupas as well as the French Technical Institute for Sugar Beet, ITB (“Institut Technique de la Betterave”) for their valuable advice when

designing the analysis. We thank Terres Inovia for sending us additional data on rapeseed crop surfaces, as well as Corentin Barbu and Baptiste Rayon from INRAE for their help in acquiring and handling AGRESTE data. We acknowledge the GenOuest bioinformatics core facility (<https://www.genouest.org>) for providing the computing infrastructure.

## CONFLICT OF INTEREST DECLARATION

The authors declare no conflict of interest.

## 5. REFERENCES

- 1 Stevens M, Liu H-Y, and Lemaire O, Virus Diseases, Sugar Beet, John Wiley & Sons, Ltd, pp. 256–285 (2006).
- 2 Hossain R, Menzel W, Lachmann C, and Varrelmann M, New insights into virus yellows distribution in Europe and effects of beet yellows virus, beet mild yellowing virus, and beet chlorosis virus on sugar beet yield following field inoculation, *Plant Pathol* **70**:584–593 (2021).
- 3 Hauer M, Hansen AL, Manderyck B, Olsson Å, Raaijmakers E, Hanse B, *et al.*, Neonicotinoids in sugar beet cultivation in Central and Northern Europe: Efficacy and environmental impact of neonicotinoid seed treatments and alternative measures, *Crop Prot* **93**:132–142 (2017).
- 4 Wood TJ and Goulson D, The environmental risks of neonicotinoid pesticides: a review of the evidence post 2013, *Environ Sci Pollut Res* **24**:17285–17325 (2017).
- 5 Gibbons D, Morrissey C, and Mineau P, A review of the direct and indirect effects of neonicotinoids and fipronil on vertebrate wildlife, *Environ Sci Pollut Res* **22**:103–118 (2015).
- 6 Dewar AM and Qi A, The virus yellows epidemic in sugar beet in the UK in 2020 and the adverse effect of the EU ban on neonicotinoids on sugar beet production, *Outlooks Pest Manag* **32**:53–59 (2021).
- 7 Mahillon M, Groux R, Bussereau F, Brodard J, Debonneville C, Demal S, *et al.*, Virus yellows and syndrome “basses richesses” in western Switzerland: a dramatic 2020 season calls for urgent control measures, *Pathogens* **11**:885, Multidisciplinary Digital Publishing Institute (2022).

- 8 Food and Agriculture Organization of the United Nations (FAO), FAOSTAT statistical database, FAO., [Rome] (2021).
- 9 Institut Technique de la Betterave (ITB), Le Cahier Technique, *Better Fr* **1115**:13–16 (2020).
- 10 Institut Technique de la Betterave (ITB), Le Cahier Technique, *Better Fr* **1134**:17–20 (2021).
- 11 Institut Technique de la Betterave (ITB), Le Cahier Technique, *Better Fr* **1150**:11–14 (2022).
- 12 European Food Safety Authority (EFSA), Evaluation of the emergency authorisations granted by Member State France for plant protection products containing imidacloprid or thiamethoxam, *EFSA Support Publ* **18**:6968E (2021).
- 13 CJEU (Court of Justice of the European Union), CJEU (Court of Justice of the European Union). Judgment of 19 January 2023, Pesticide Action Network Europe and Others, not yet published, *ECLI:EU:C:2023:30 C-162/21* (2023).
- 14 Schop S, Kloth KJ, Raaijmakers E, and van der Vlugt RAA, The effect of mature plant resistance in sugar beet (*Beta vulgaris* spp. *vulgaris*) on survival, fecundity and behaviour of green peach aphids (*Myzus persicae*), *Bull Entomol Res*:1–8 (2022).
- 15 Smith HG and Hallsworth PB, The effects of yellowing viruses on yield of sugar beet in field trials, 1985 and 1987, *Ann Appl Biol* **116**:503–511 (1990).
- 16 van der Werf W, Westerman PR, Verweij R, and Peters D, The influence of primary infection date and establishment of vector populations on the spread of yellowing viruses in sugar beet, *Ann Appl Biol* **121**:57–74 (1992).
- 17 Guillemaud T, Mieuze L, and Simon J-C, Spatial and temporal genetic variability in French populations of the peach–potato aphid, *Myzus persicae*, *Heredity* **91**:143–152, Nature Publishing Group (2003).
- 18 Blackman RL and Eastop VF, Aphids on the world’s crops: an identification and information guide, 2nd ed., Chichester (UK) Wiley (2000).
- 19 Hullé M, Coquio S, and Laperche V, Patterns in flight phenology of a migrant cereal aphid species, *J Appl Ecol* **31**:49–58, [British Ecological Society, Wiley] (1994).
- 20 Bell JR, Alderson L, Izera D, Kruger T, Parker S, Pickup J, *et al.*, Long-term phenological trends, species accumulation rates, aphid traits and climate: five decades of change in migrating aphids, ed. by Wilson K, *J Anim Ecol* **84**:21–34 (2015).
- 21 Werker AR, Dewar AM, and Harrington R, Modelling the incidence of virus yellows in sugar beet in the UK in relation to numbers of migrating *Myzus persicae*, *J Appl Ecol* **35**:811–818 (1998).
- 22 Qi A, Dewar AM, and Harrington R, Decision making in controlling virus yellows of sugar beet in the UK, *Pest Manag Sci* **60**:727–732 (2004).
- 23 Cocu N, Harrington R, Rounsevell I M D A, Worner SP, Hullé M, and the EXAMINE project participants, Geographical location, climate and land use influences on the phenology and numbers of the aphid, *Myzus persicae*, in Europe: Environmental influences on aphid distribution, *J Biogeogr* **32**:615–632 (2005).
- 24 Harrington R, Clark SJ, Welham SJ, Verrier PJ, Denholm CH, Hullé M, *et al.*, Environmental change and the phenology of European aphids, *Glob Change Biol* **13**:1550–1564 (2007).



- 25 Luquet M, Hullé M, Simon J-C, Parisey N, Buchard C, and Jaloux B, Relative importance of long-term changes in climate and land-use on the phenology and abundance of legume crop specialist and generalist aphids, *Insect Sci* **26**:881–896 (2019).
- 26 Heathcote GD, Effect of plant spacing and time of sowing of sugar beet on aphid infestation and spread of virus yellows, *Plant Pathol* **19**:32–39 (1970).
- 27 Verheggen F, Barrès B, Bonafos R, Desneux N, Escobar-Gutiérrez AJ, Gachet E, *et al.*, Producing sugar beets without neonicotinoids: An evaluation of alternatives for the management of viruses-transmitting aphids, *Entomol Gen* **42**:491–498 (2022).
- 28 Laurent A, Favrot A, Maupas F, Royer C, and Makowski D, Assessment of non-neonicotinoid treatments against aphids on sugar beets, *Crop Prot* **164**:106140 (2023).
- 29 Hullé M, Agraphid, un réseau de surveillance des populations de pucerons : base de données associée et domaines d'application, presented at the ANPP - Journées sur les bases de données Européennes en Protection des Plantes - 14 et 15 Octobre 1991, 1991, Strasbourg.
- 30 Beck HE, Zimmermann NE, McVicar TR, Vergopolan N, Berg A, and Wood EF, Present and future Köppen-Geiger climate classification maps at 1-km resolution, *Sci Data* **5**:180214, Nature Publishing Group (2018).
- 31 Cavalloro R, “Euraphid” Network: trapping and aphid prognosis : Proceedings of a meeting of the EC-Experts Group, Catania, Italy, 7 to 9 November 1988, Commission of the European Communities, Directorate-General Telecommunications, Information Industries and Innovation (1990).
- 32 Harrington R, Verrier P, Denholm C, Hullé M, Maurice D, Bell N, *et al.*, “EXAMINE” (EXploitation of Aphid Monitoring in Europe): An EU Thematic Network for the study of global change impacts on aphids, 2004.
- 33 Taylor LR, 202–239 (1974).
- 34 Hullé M and Gamon A, Relations entre les captures des pièges à succion du réseau Agraphid., ed. by Cavalloro R, “Euraphid” Network: Trapping and Aphid Prognosis., CEC Commission of the European Communities, Bruxelles [BEL], pp. 165–177 (1990).
- 35 Durand Y, Brun E, Merindol L, Guyomarc’h G, Lesaffre B, and Martin E, A meteorological estimation of relevant parameters for snow models, *Ann Glaciol* **18**:65–71, Cambridge University Press (1993).
- 36 Vidal J-P, Martin E, Franchistéguy L, Baillon M, and Soubeyroux J-M, A 50-year high-resolution atmospheric reanalysis over France with the Safran system, *Int J Climatol* **30**:1627–1644 (2010).
- 37 Agreste, Enquête TERUTI-LUCAS. Occupation du territoire - Séries longues. Service de la Statistique et de la Prospective (SSP) du Ministère de l’Agriculture, de l’Alimentation, de la Pêche, de la Ruralité et de l’Aménagement du Territoire., 2021. <https://agreste.agriculture.gouv.fr/agreste-web/disaron/W0021/detail/> [accessed 20 May 2022].
- 38 Agreste, Statistique Agricole Annuelle. Cultures développées (hors fourrage, prairies, fruits, fleurs et vigne). Service de la Statistique et de la Prospective (SSP) du Ministère de l’Agriculture, de l’Alimentation, de la Pêche, de la Ruralité et de l’Aménagement du Territoire., 2021.

[https://agreste.agriculture.gouv.fr/agreste-web/disaron/SAANR\\_DEVELOPPE\\_2/detail/](https://agreste.agriculture.gouv.fr/agreste-web/disaron/SAANR_DEVELOPPE_2/detail/) [accessed 22 April 2022].

39 Agreste, Statistique Agricole Annuelle. Utilisation du territoire. Service de la Statistique et de la Prospective (SSP) du Ministère de l'Agriculture, de l'Alimentation, de la Pêche, de la Ruralité et de l'Aménagement du Territoire., 2021. [https://agreste.agriculture.gouv.fr/agreste-web/disaron/SAANR\\_DEVELOPPE\\_2/detail/](https://agreste.agriculture.gouv.fr/agreste-web/disaron/SAANR_DEVELOPPE_2/detail/) [accessed 22 April 2022].

40 Statbel, Exploitations agricoles et horticoles. Direction générale Statistique - Statistics Belgium., 2022. <https://statbel.fgov.be/fr/themes/agriculture-peche/exploitations-agricoles-et-horticoles/plus> [accessed 31 May 2022].

41 Whalon ME and Smilowitz Z, Determination of consultant temperature developmental thresholds for *Myzus persicae* (Sulz.), *J N Y Entomol Soc* **85** (1977).

42 Pierre J-S, Hullé M, Gauthier J-P, and Rispe C, Critical windows: A method for detecting lagged variables in ecological time series, *Ecol Inform* **61**:101178 (2021).

43 Bell JR, Botham MS, Henrys PA, Leech DI, Pearce-Higgins JW, Shortall CR, *et al.*, Spatial and habitat variation in aphid, butterfly, moth and bird phenologies over the last half century, *Glob Change Biol* **25**:1982–1994 (2019).

44 Harrington R, Hullé M, Pickup J, and Bale JS, Forecasting the need for early season aphid control: geographical variation in the relationship between winter temperature and early season flight activity of *Myzus persicae*, 1992, Wellesbourne, Warwick : Association of Applied Biologists (AAB).

45 Sheppard LW, Bell JR, Harrington R, and Reuman DC, Changes in large-scale climate alter spatial synchrony of aphid pests, *Nat Clim Change* **6**:610–613 (2016).

46 Hastie T, Tibshirani R, and Friedman J, *The Elements of Statistical Learning*, Springer New York, New York, NY (2009).

47 Otto SA, How to normalize the RMSE [Blog post]., <https://www.marinedatascience.co/blog/2019/01/07/normalizing--Rmse>, 2019. <https://www.marinedatascience.co/blog/2019/01/07/normalizing-the-rmse/> [accessed 7 March 2023].

48 Bring J, How to standardize regression coefficients, *Am Stat* **48**:209–213, Taylor & Francis (1994).

49 Cade BS, Model averaging and muddled multimodel inferences, *Ecology* **96**:2370–2382 (2015).

50 R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria (2022).

51 Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, *et al.*, Welcome to the Tidyverse, *J Open Source Softw* **4**:1686 (2019).

52 Kuhn M, *caret: Classification and Regression Training* (2022).

53 Bartoń K, *MuMIn: Multi-Model Inference* (2022).

- 54 Pebesma E, Simple Features for R: Standardized Support for Spatial Vector Data, *R J* **10**:439–446 (2018).
- 55 Teucher A and Russell K, rmapshaper: Client for “mapshaper” for “Geospatial” Operations. R package version 0.4.6. <https://CRAN.R-project.org/package=rmapshaper> (2022).
- 56 Wickham H, ggplot2: Elegant Graphics for Data Analysis, 2nd ed. 2016 edition, Springer, New York, NY (2016).
- 57 Wilke CO, cowplot: Streamlined Plot Theme and Plot Annotations for “ggplot2”. R package version 1.1.1, <<https://CRAN.R-project.org/package=cowplot>>. (2020).
- 58 Kassambara A, ggpubr: “ggplot2” Based Publication Ready Plots. R package version 0.4.0, <<https://CRAN.R-project.org/package=ggpubr>>. (2023).
- 59 Neuwirth E, RColorBrewer: ColorBrewer Palettes. R package version 1.1-3, <<https://CRAN.R-project.org/package=RColorBrewer>>. (2022).
- 60 Brand T van den, ggh4x: Hacks for “ggplot2”. R package version 0.2.2, <<https://CRAN.R-project.org/package=ggh4x>>. (2023).
- 61 Wickham H, Seidel D, and RStudio, scales: Scale Functions for Visualization. R package version 1.2.1, <<https://CRAN.R-project.org/package=scales>>. (2022).
- 62 Baquero OS, ggsn: North Symbols and Scale Bars for Maps Created with “ggplot2” or “ggmap”. R package version 0.5.0, <<https://CRAN.R-project.org/package=ggsn>>. (2019).
- 63 Campitelli E, ggnewscale: Multiple Fill and Colour Scales in “ggplot2”. R package version 0.4.8, <<https://CRAN.R-project.org/package=ggnewscale>>. (2022).
- 64 Bourhis Y, Bell JR, van den Bosch F, and Milne AE, Artificial neural networks for monitoring network optimisation—a practical example using a national insect survey, *Environ Model Softw* **135**:104925 (2021).
- 65 Bell JR, Clark SJ, Stevens M, and Mead A, Quantifying inherent predictability and spatial synchrony in the aphid vector *Myzus persicae*: field-scale patterns of abundance and regional forecasting error in the UK, *Pest Manag Sci* **79**:1331–1341 (2023).
- 66 Bale JS, Masters GJ, Hodkinson ID, Awmack C, Bezemer TM, Brown VK, *et al.*, Herbivory in global climate change research: direct effects of rising temperature on insect herbivores, *Glob Change Biol* **8**:1–16 (2002).
- 67 Smith HG and Hinckes JA, Studies on beet western yellows virus in oilseed rape (*Brassica napus* ssp. *oleifera*) and sugar beet (*Beta vulgaris*), *Ann Appl Biol* **107**:473–484 (1985).
- 68 Chen C, Harvey JA, Biere A, and Gols R, Rain downpours affect survival and development of insect herbivores: the specter of climate change?, *Ecology* **100**:e02819 (2019).
- 69 Rohitha BH and Penman DR, Flight of the bluegreen aphid, *Acyrthosiphon kondoi* Shinji (Homoptera: Aphididae). II. The effect of weather: Multiple regression and flight threshold analyses, *N Z J Zool* **13**:209–214 (1986).
- 70 Crossley MS, Lagos-Kutz D, Davis TS, Eigenbrode SD, Hartman GL, Voegtlin DJ, *et al.*, Precipitation change accentuates or reverses temperature effects on aphid dispersal, *Ecol Appl* **32**:e2593 (2022).

- 71 Cocu N, Harrington R, Hullé M, and Rounsevell MD, Spatial autocorrelation as a tool for identifying the geographical patterns of aphid annual abundance, *Agric For Entomol* **7**:31–43 (2005).
- 72 Estay SA, Lima M, and Harrington R, Climate mediated exogenous forcing and synchrony in populations of the oak aphid in the UK, *Oikos* **118**:175–182 (2009).
- 73 Francis F, Then C, Francis A, Gbangbo YAC, Iannello L, and Ben Fekih I, Complementary strategies for biological control of aphids and related virus transmission in sugar beet to replace neonicotinoids, *Agriculture* **12**:1663, Multidisciplinary Digital Publishing Institute (2022).
- 74 Neta A, Levi Y, Morin E, and Morin S, Seasonal forecasting of pest population dynamics based on downscaled SEAS5 forecasts, *Ecol Model* **480**:110326 (2023).
- 75 d’Andrimont R, Taymans M, Lemoine G, Ceglar A, Yordanov M, and van der Velde M, Detecting flowering phenology in oil seed rape parcels with Sentinel-1 and -2 time series, *Remote Sens Environ* **239**:111660 (2020).
- 76 Habekotté B, A model of the phenological development of winter oilseed rape (*Brassica napus* L.), *Field Crops Res* **54**:127–136 (1997).
- 77 Kiritani K, Different effects of climate change on the population dynamics of insects, *Appl Entomol Zool* **48**:97–104 (2013).
- 78 Scott ER, Wei J-P, Li X, Han W-Y, and Orians CM, Differing non-linear, lagged effects of temperature and precipitation on an insect herbivore and its host plant, *Ecol Entomol* **46**:866–876 (2021).
- 79 Fabre F, Dedryver C-A, Plantegenest M, Hullé M, and Rivot E, Hierarchical Bayesian Modelling of plant colonisation by winged aphids: Inferring dispersal processes by linking aerial and field count data, *Ecol Model* **221**:1770–1778 (2010).

## SUPPORTING INFORMATION

**Appendix S1** Criteria applied for data inclusion.

**Appendix S2** Critical periods for the calculation of meteorological windows for all selected models.

**Appendix S3** Spatial variation in predictions from all models.

## FIGURE LEGENDS

**Figure 1** French sugar beet production area (in blue) and the Agraphid network. Circles correspond to traps considered as belonging to the French sugar beet production area and were used for both model fitting and model validation. Triangles correspond to traps used only for testing models in conditions different from those in the French sugar beet production area. Colour of a given trap and associated label corresponds to the average temperature ( $^{\circ}\text{C}$ ) from November 1 to March 30 at that location, between 1978 and 2022. The horizontal bar represents the years of operation of the trap (in black) between 1978 and 2022.

**Figure 2** Relationship between the flight features observed in each trap (x-axis) and their prediction from the best model predicting each of them (y-axis) at each prediction date (February 14, March 5, April 15). The flight features represented are, **a** the date of 5<sup>th</sup> capture ( $D5c$ , expressed as the day number of the year), **b** the number of days between the 5<sup>th</sup> and 95<sup>th</sup> percentiles of captures in spring and summer ( $Fd$ ), **c** the log-cumulative abundance (+0.01) between January 1 and June 30 ( $\log Ab$ ), and **d** the cumulative abundance ( $Ab$ ) back-transformed from  $\log Ab$ . Observations are representative of aphid aerial samples on an 80km area. The colours of the dots correspond to the 10 traps used for model fitting. In each facet, the dashed red line corresponds to the identity line  $y = x$ . Root mean squared error of prediction ( $\text{RMSE}_p$ ) is measured *via* cross-validation, while root mean squared error ( $\text{RMSE}$ ) is obtained after fitting the model on the full dataset. For  $Ab$  such metrics are not displayed as errors should be interpreted multiplicatively (see main text).

**Figure 3** Partial standardised coefficients associated with predictors for the best models at each prediction date and for each aphid flight feature:  $D5c$  (date of 5<sup>th</sup> capture),  $Fd$  (number of days between

the 5<sup>th</sup> and 95<sup>th</sup> percentile flight),  $\log Ab$  (log-cumulative abundance + 0.01) between January 1 and June 30. Partial standardised coefficient can be interpreted as the change in  $y$  following a 1 standard deviation change in  $x$ , other predictors being held constant (Bring 1994, Cade 2015). The definition of predictors is reported in Table 2.

**Figure 4** Periods on which meteorological variables (sum of degree-days  $>0^{\circ}\text{C}$  and mean rainfall) were calculated in best performing models ( $M_3$ ) at each prediction date and for each aphid flight feature:  $D5c$  = Date of 5<sup>th</sup> capture,  $Fd$  = number of days between the 5<sup>th</sup> and 95<sup>th</sup> percentile flight,  $\log Ab$  = log-cumulative abundance (+0.01) between January 1 and June 30. Dashed lines correspond to the 3 prediction dates: February 14 (choice of a sugar beet variety, in blue), March 5 (decisions on sugar beet sowing dates and sowing a companion plant, in green) and April 15 (curative measures, in red).

**Figure 5** Residual Mean Squared Error (RMSE) or Residual Mean Squared Error of prediction (RMSE<sub>p</sub>) obtained when predicting each trap from the Agraphid trapping network, respectively for the traps located inside (triangles) or outside (circles) the sugar beet production area.  $D5c$  = Date of 5<sup>th</sup> capture,  $Fd$  = number of days between the 5<sup>th</sup> and 95<sup>th</sup> percentile flight,  $\log Ab$  = log-cumulative abundance (+0.01) between January 1 and June 30. To allow meaningful comparison across response variables, the intensity of the colour gradient depends on how the RMSE<sub>p</sub> on the independent dataset departs from the RMSE on the training dataset.

**Table 1** Predictive variables included in each model of aphid seasonal flight activity.

Variable name	Variable description	Data source	Models
<i>TJF14</i>	Mean daily temperature from January 1 <sup>st</sup> to February 14 <sup>th</sup>	SAFRAN	$M_1$
<i>T_crit</i>	Cumulative degree-days (base temperature = 0°C, no upper threshold) over the critical time window	SAFRAN	$M_2, M_3, M_4$
<i>R_crit</i>	Mean of daily rainfall over the critical time window	SAFRAN	$M_3, M_4$
<i>Latitude</i>	Trap latitude		$M_3, M_4$
<i>Longitude</i>	Trap longitude		$M_3, M_4$
<i>LatSq</i>	Trap latitude squared		$M_3, M_4$
<i>LonSq</i>	Trap longitude squared		$M_3, M_4$
<i>LatLon</i>	Trap latitude * trap longitude		$M_3, M_4$
<i>RSC_area</i>	Rapeseed crop area (in ha) around a trap (80km radius)	Agreste, Statbel, Terres Inovia	$M_3, M_4$
<i>Crop_area</i>	Crop area (arable land + permanent crops, in ha) around a trap (80km radius)	Agreste, Statbel	$M_3, M_4$



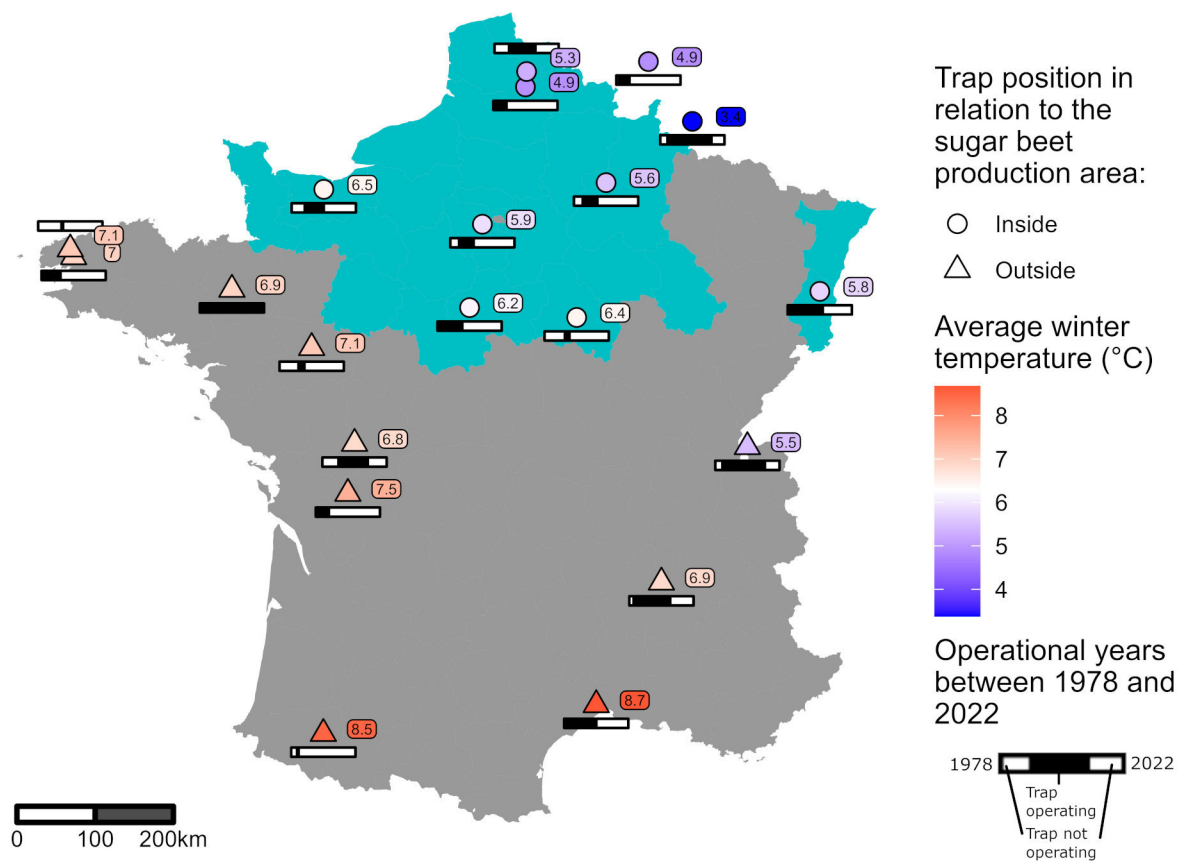
<i>SN_area</i>	Semi-natural land area (forest + permanent grasslands, in ha) around a trap (80km radius)	Agreste, Statbel	$M_3, M_4$
----------------	---	------------------	------------

**Table 2** Root mean squared error of prediction ( $RMSE_p$ ) values (mean  $\pm$  SEM) for each model, each response variable (flight feature) and each prediction date. Values in bold correspond to the model(s) with the best predictive capacities at each prediction date, for each flight feature. *D1c* = Date of first capture, *D5c* = Date of fifth capture, *Fd* = number of days between the 5<sup>th</sup> and 95<sup>th</sup> percentiles of captures of captures in spring and summer, *logAb* =  $\log(\text{cumulative abundance from January 1 to June } 30 + 0.01)$ .  $RMSE_p$  unit is the same as the response (note that *D1c* and *D5c* are expressed as day of the year numbers).

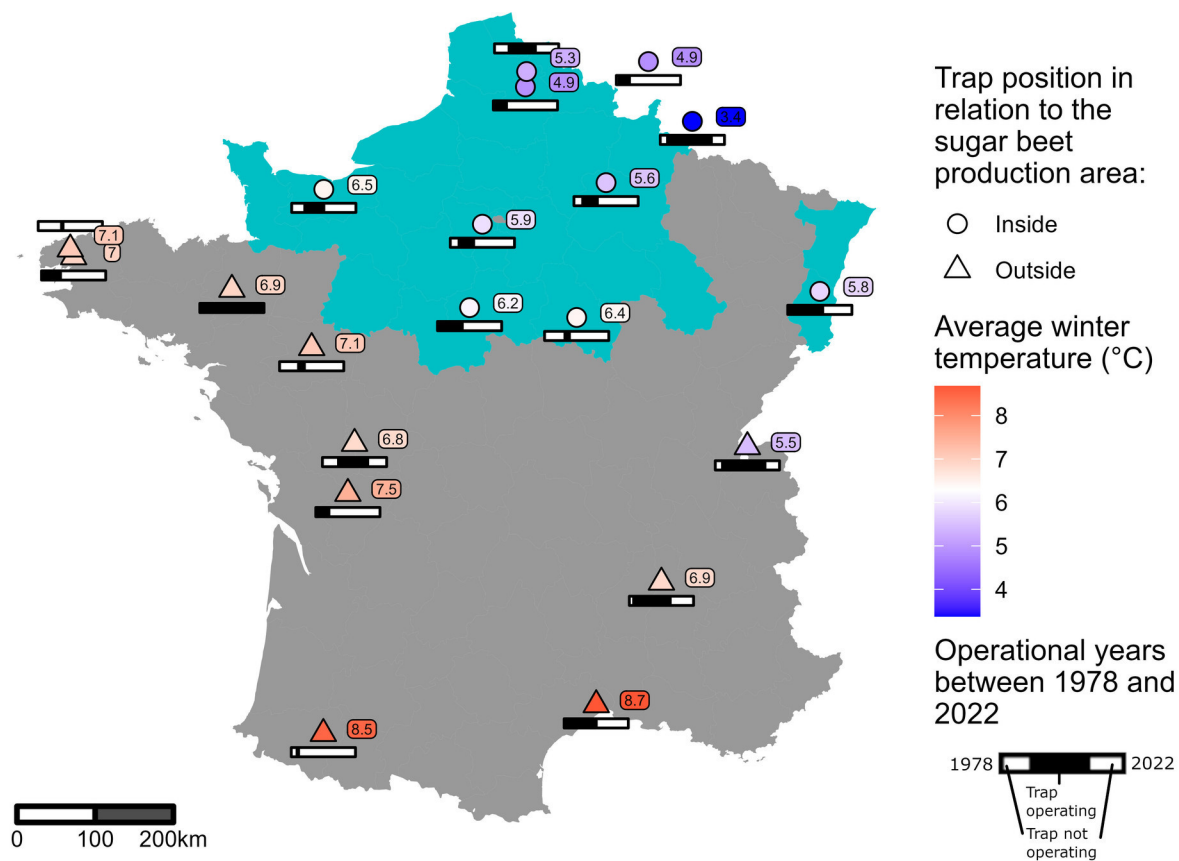
<i>Response variable</i>	<i>Prediction date</i>	<i>M1</i>	<i>M2</i>	<i>M3</i>
<i>D1c</i>	February 14	23.2 ( $\pm$ 0.1)	22.8 ( $\pm$ 0.9)	<b>21.1 (<math>\pm</math> 0.9)</b>
<i>D1c</i>	March 5	23.2 ( $\pm$ 0.1)	21.2 ( $\pm$ 0.7)	<b>19.7 (<math>\pm</math> 0.8)</b>
<i>D1c</i>	April 15	23.2 ( $\pm$ 0.1)	20 ( $\pm$ 0.7)	<b>18.9 (<math>\pm</math> 0.7)</b>
<i>D5c</i>	February 14	19.9 ( $\pm$ 0.1)	19.1 ( $\pm$ 0.7)	<b>16 (<math>\pm</math> 0.6)</b>
<i>D5c</i>	March 5	19.9 ( $\pm$ 0.1)	17.6 ( $\pm$ 0.5)	<b>14.9 (<math>\pm</math> 0.6)</b>
<i>D5c</i>	April 15	19.9 ( $\pm$ 0.1)	16.7 ( $\pm$ 0.5)	<b>14.3 (<math>\pm</math> 0.6)</b>
<i>Fd</i>	February 14	22.8 ( $\pm$ 0.1)	22 ( $\pm$ 0.6)	<b>20.2 (<math>\pm</math> 0.6)</b>
<i>Fd</i>	March 5	22.8 ( $\pm$ 0.1)	21.4 ( $\pm$ 0.6)	<b>19.5 (<math>\pm</math> 0.6)</b>
<i>Fd</i>	April 15	22.8 ( $\pm$ 0.1)	20.8 ( $\pm$ 0.6)	<b>19.4 (<math>\pm</math> 0.6)</b>
<i>logAb</i>	February 14	2.3 ( $\pm$ 0)	2.3 ( $\pm$ 0.1)	<b>1.9 (<math>\pm</math> 0.1)</b>
<i>logAb</i>	March 5	2.3 ( $\pm$ 0)	2.2 ( $\pm$ 0.1)	<b>1.8 (<math>\pm</math> 0.1)</b>
<i>logAb</i>	April 15	2.3 ( $\pm$ 0)	2.1 ( $\pm$ 0.1)	<b>1.7 (<math>\pm</math> 0.1)</b>

**Table 3** Normalized root mean squared error of prediction (NRMSE<sub>p</sub>) values (mean ± SEM) for each model, each response variable (flight feature) and each prediction date. Values in bold correspond to the model(s) with the best predictive capacities at each prediction date, for each flight feature. *D1c* = Date of first capture, *D5c* = Date of fifth capture, *Fd* = number of days between the 5<sup>th</sup> and 95<sup>th</sup> percentiles of captures in spring and summer, *logAb* = log(cumulative abundance from January 1 to June 30 + 0.01).

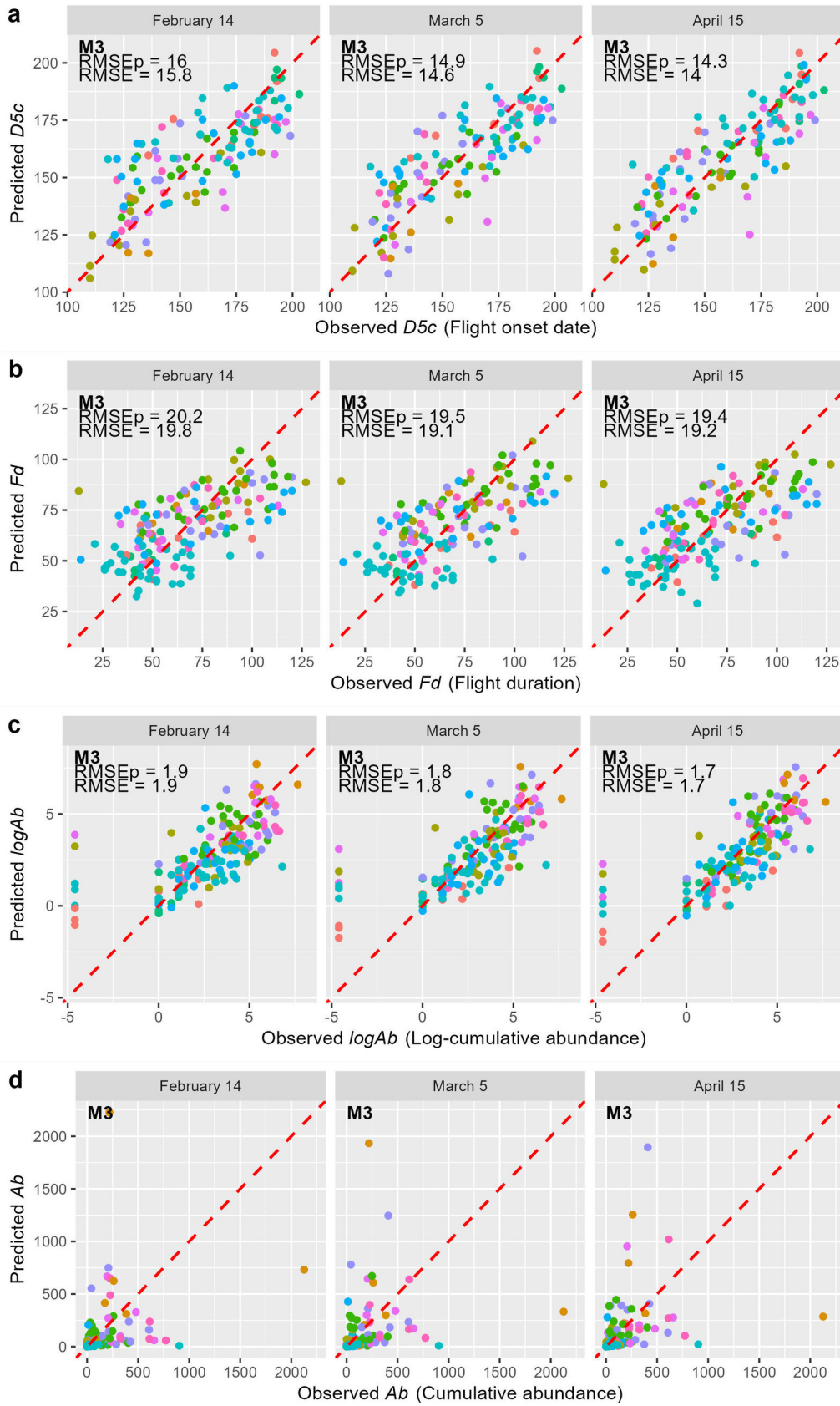
<i>Response variable</i>	<i>Prediction date</i>	<i>M1</i>	<i>M2</i>	<i>M3</i>
<i>D1c</i>	February 14	0.65 (±0.01)	0.64 (±0.04)	<b>0.6 (±0.04)</b>
<i>D1c</i>	March 5	0.65 (±0.01)	0.6 (±0.03)	<b>0.56 (±0.04)</b>
<i>D1c</i>	April 15	0.65 (±0.01)	0.56 (±0.03)	<b>0.53 (±0.03)</b>
<i>D5c</i>	February 14	0.45 (±0)	0.44 (±0.02)	<b>0.37 (±0.02)</b>
<i>D5c</i>	March 5	0.45 (±0)	0.4 (±0.01)	<b>0.34 (±0.01)</b>
<i>D5c</i>	April 15	0.45 (±0)	0.38 (±0.01)	<b>0.33 (±0.01)</b>
<i>Fd</i>	February 14	0.64 (±0)	0.62 (±0.02)	<b>0.57 (±0.02)</b>
<i>Fd</i>	March 5	0.64 (±0)	0.6 (±0.02)	<b>0.55 (±0.02)</b>
<i>Fd</i>	April 15	0.64 (±0)	0.59 (±0.02)	<b>0.55 (±0.02)</b>
<i>logAb</i>	February 14	0.83 (±0.01)	0.83 (±0.05)	<b>0.69 (±0.05)</b>
<i>logAb</i>	March 5	0.83 (±0.01)	0.81 (±0.04)	<b>0.66 (±0.05)</b>
<i>logAb</i>	April 15	0.83 (±0.01)	0.76 (±0.04)	<b>0.61 (±0.04)</b>



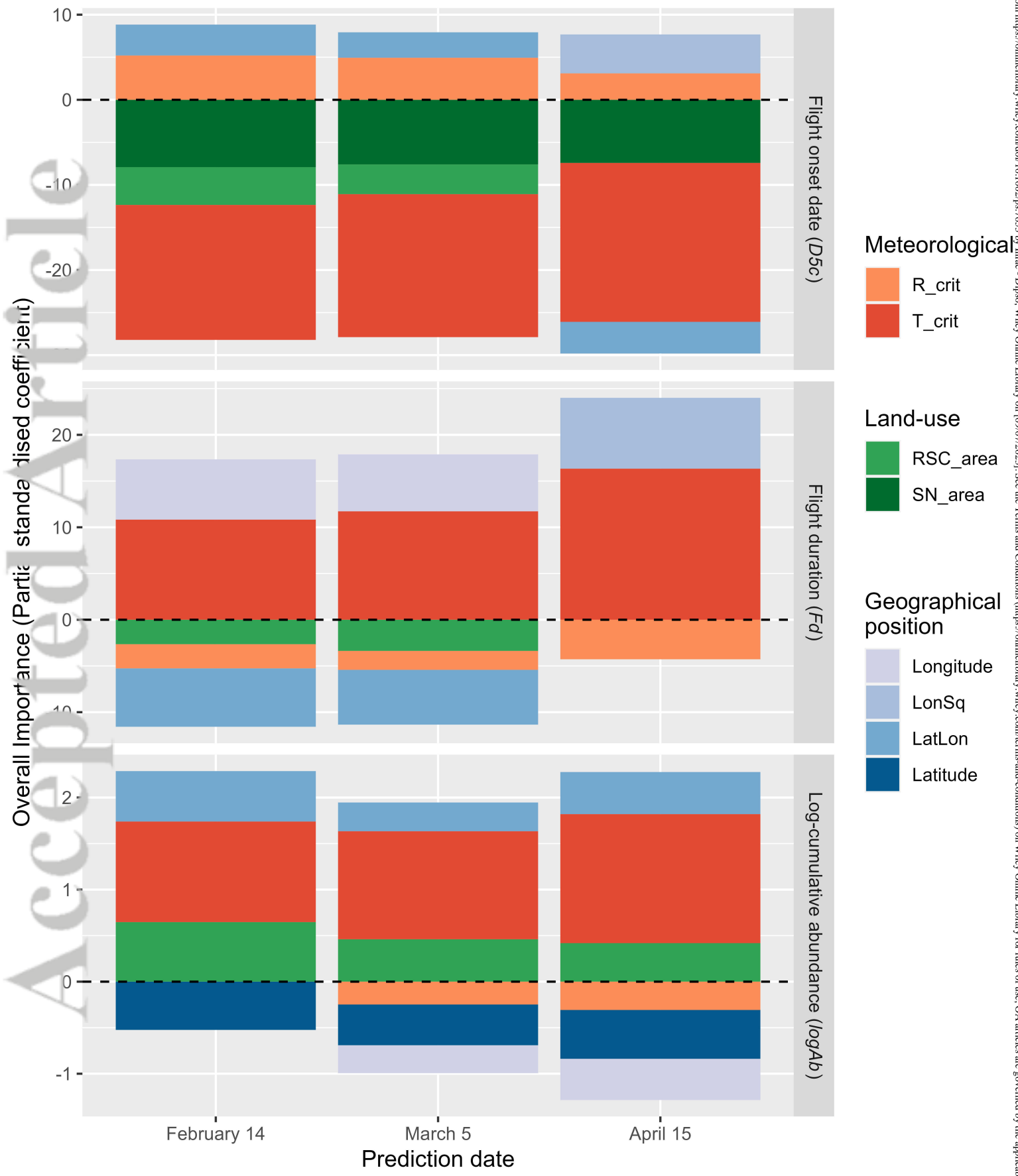
Luquet\_et\_al\_Fig1\_20230407\_Low\_resolution\_figure\_for\_peer\_review.jpg



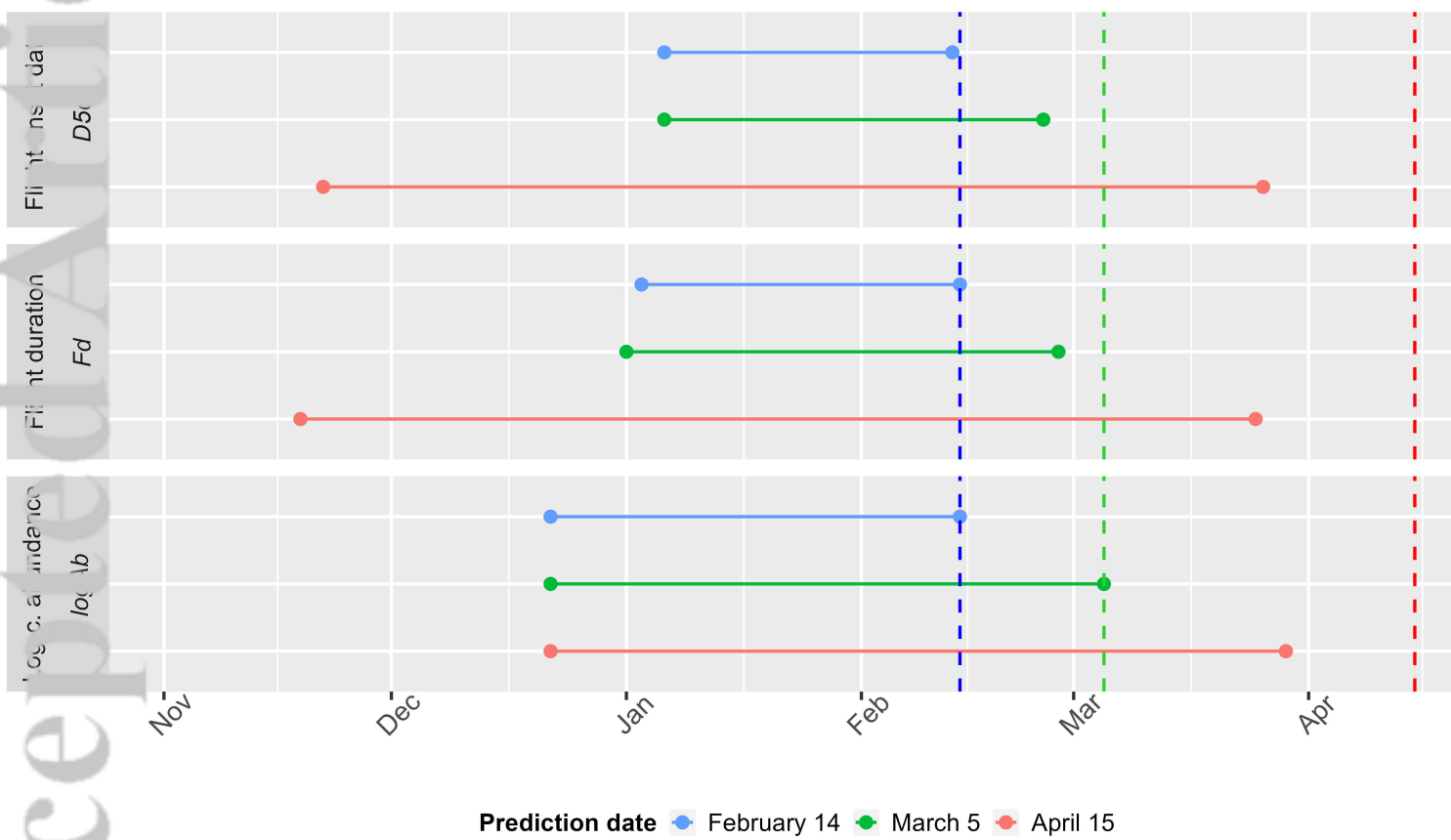
Luquet\_et\_al\_Figure\_1\_revised\_20230602.jpg



Luquet\_et\_al\_Figure\_2\_low\_resolution\_figure\_for\_review.jpg



Luquet\_et\_al\_Fig3\_revised\_20230627.jpg

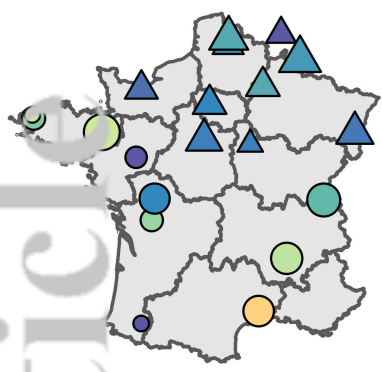


Luquet\_et\_al\_Fig4\_20230320.jpg

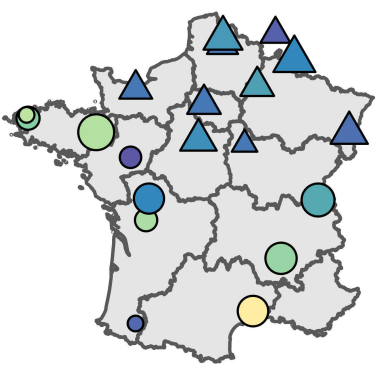


D5c

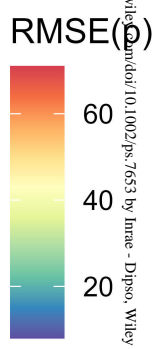
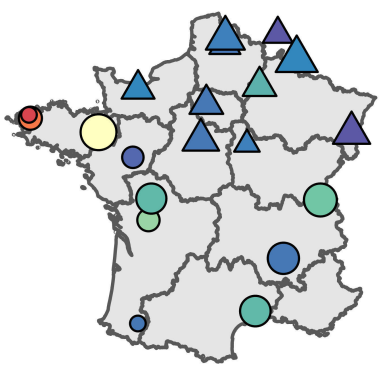
February 14



March 5

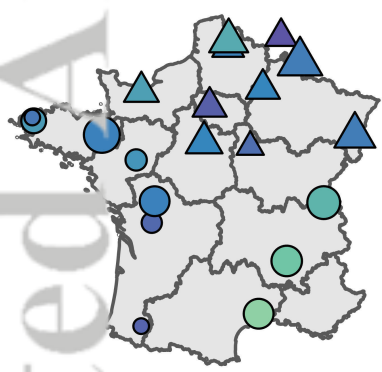


April 15

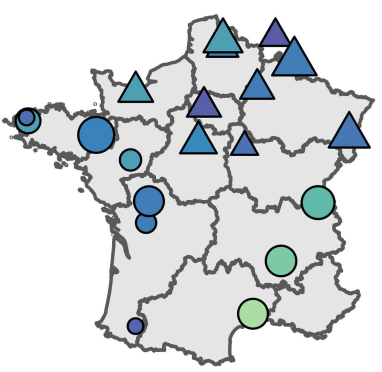


Fa

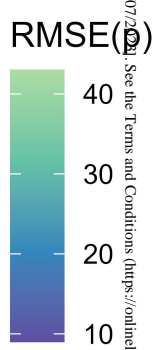
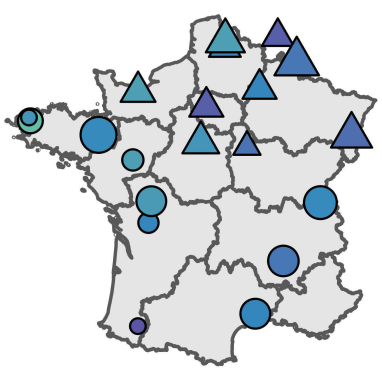
February 14



March 5

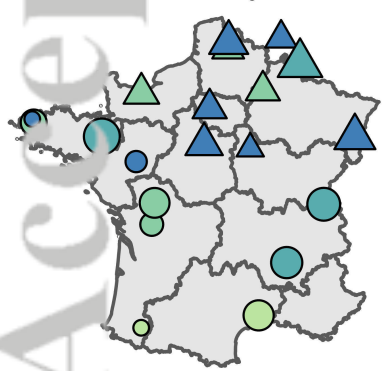


April 15

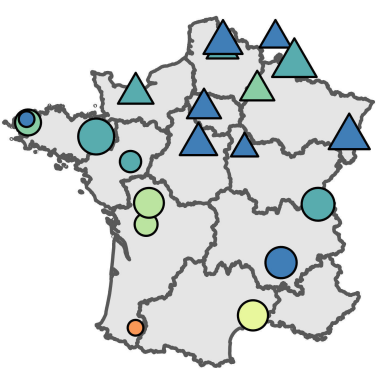


logAu

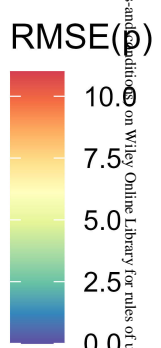
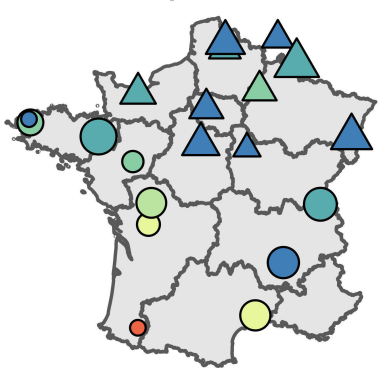
February 14



March 5



April 15



Trap status for model construction

○ Not used    △ Used

Sample size for performance evaluation

● 10    ● 20    ● 30    ● 40

Luquet\_et\_al\_Fig5\_20230320.jpg



Accepted Article

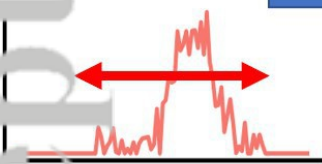


### Seasonal flight activity

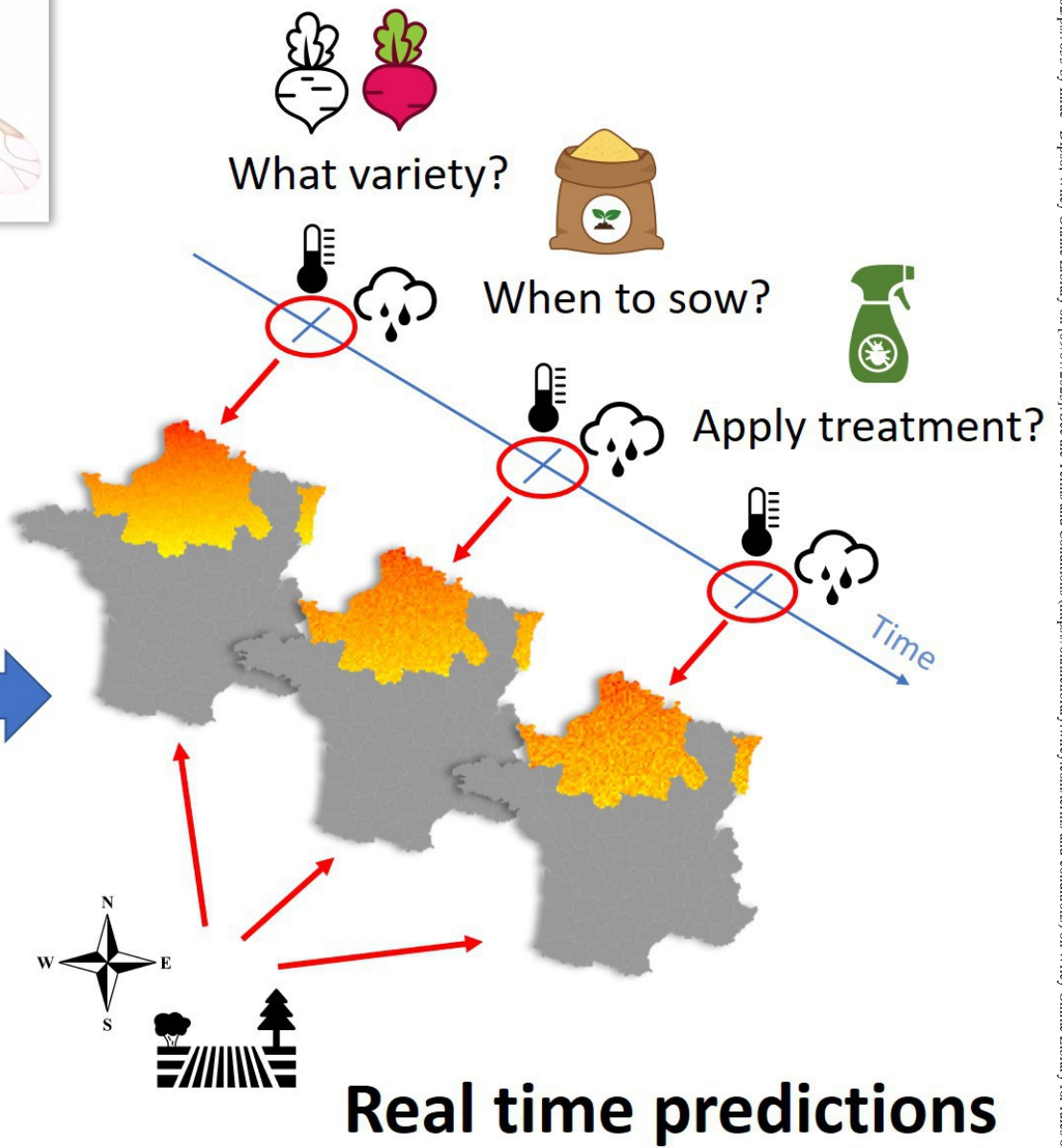
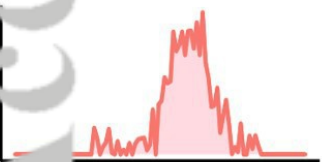
Onset date



Duration



Cumul. abundance



## Real time predictions

Luquet\_et\_al\_Graphical\_abstract\_image\_20230403.jpg

# **Predicting the seasonal flight activity of *Myzus persicae*, the main aphid vector of Virus Yellows in sugar beet**

LUQUET Martin\*, POGGI Sylvain, BUCHARD Christelle, PLANTEGENEST Manuel, TRICAULT Yann

We built regression models to forecast the main characteristics of aphid flight. Predictions are updated at strategic moments for the mitigation of aphid-borne virus yellows in sugar beet crops.

Accepted Article