ESTIMATING BEET YELLOWS SEVERITY AT PLOT RESOLUTION WITH SATELLITE OBSERVATIONS

Samuel SOUBEYRAND [1], François JOUDELAT [2],
Joseph RYNKIEWICZ [3], Edith GABRIEL [1]

[1] INRAE, Avignon, FRANCE
[2] ITB, Paris, FRANCE
[3] Université Paris 1 Panthéon-Sorbonne, Paris, FRANCE

ICPP Satellite Event « How to combine remote sensing with epidemiological modelling to improve plant disease management? »

August 19-20, 2023, Lyon, France

SEPIM and BEYOND Projects

Financé par

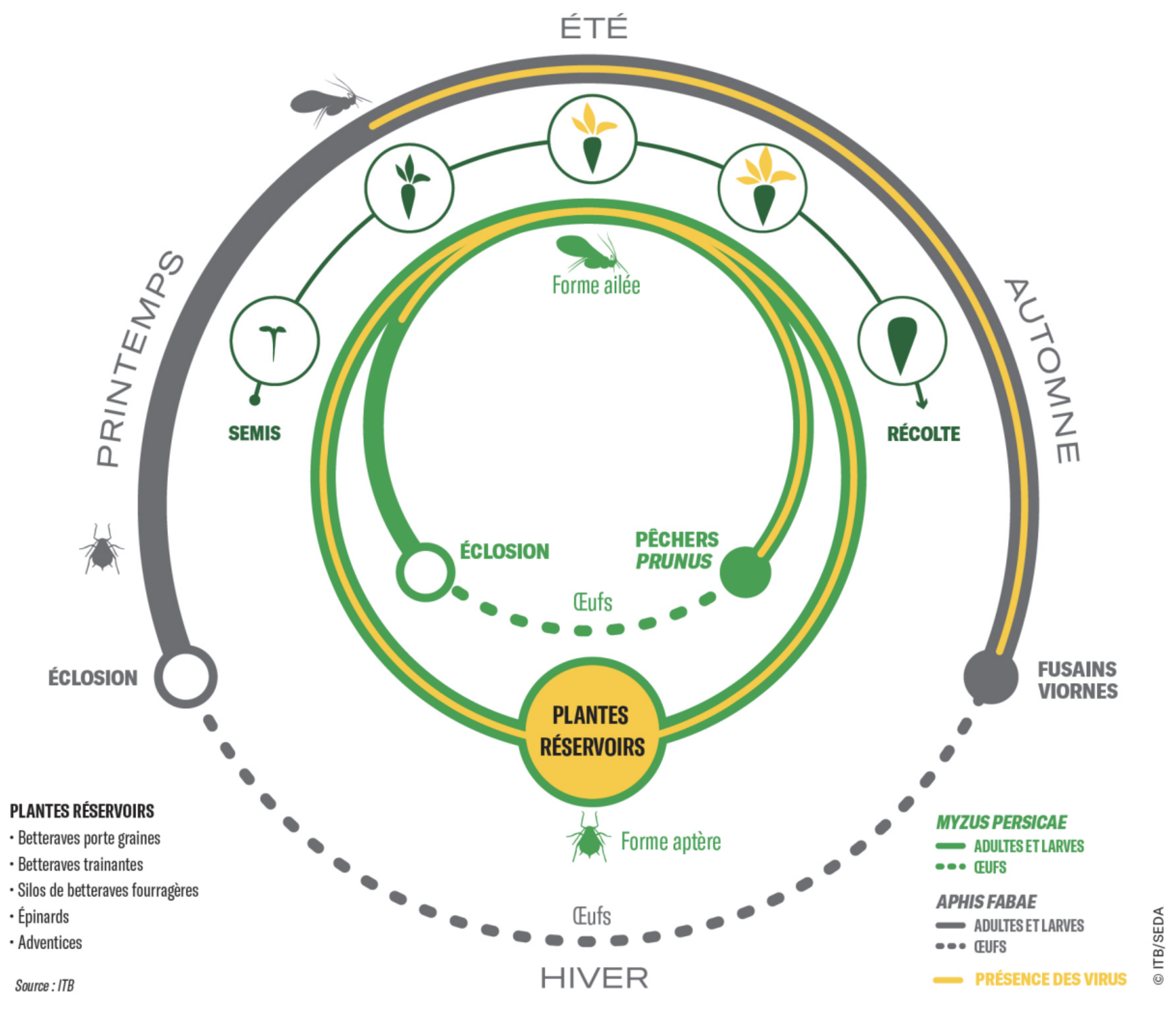
Context

Beet yellows now perceived as a major issue in plant health in France
-> consequence of the modification of regulation concerning the use of phytosanitary products (neonicotinoids) having negative side-effects
-> beet production thus exposed to more hazards

An agroecological approach to this issue requires actions at multiple levels
-> prophylaxis, tactical treatments, insurance systems…

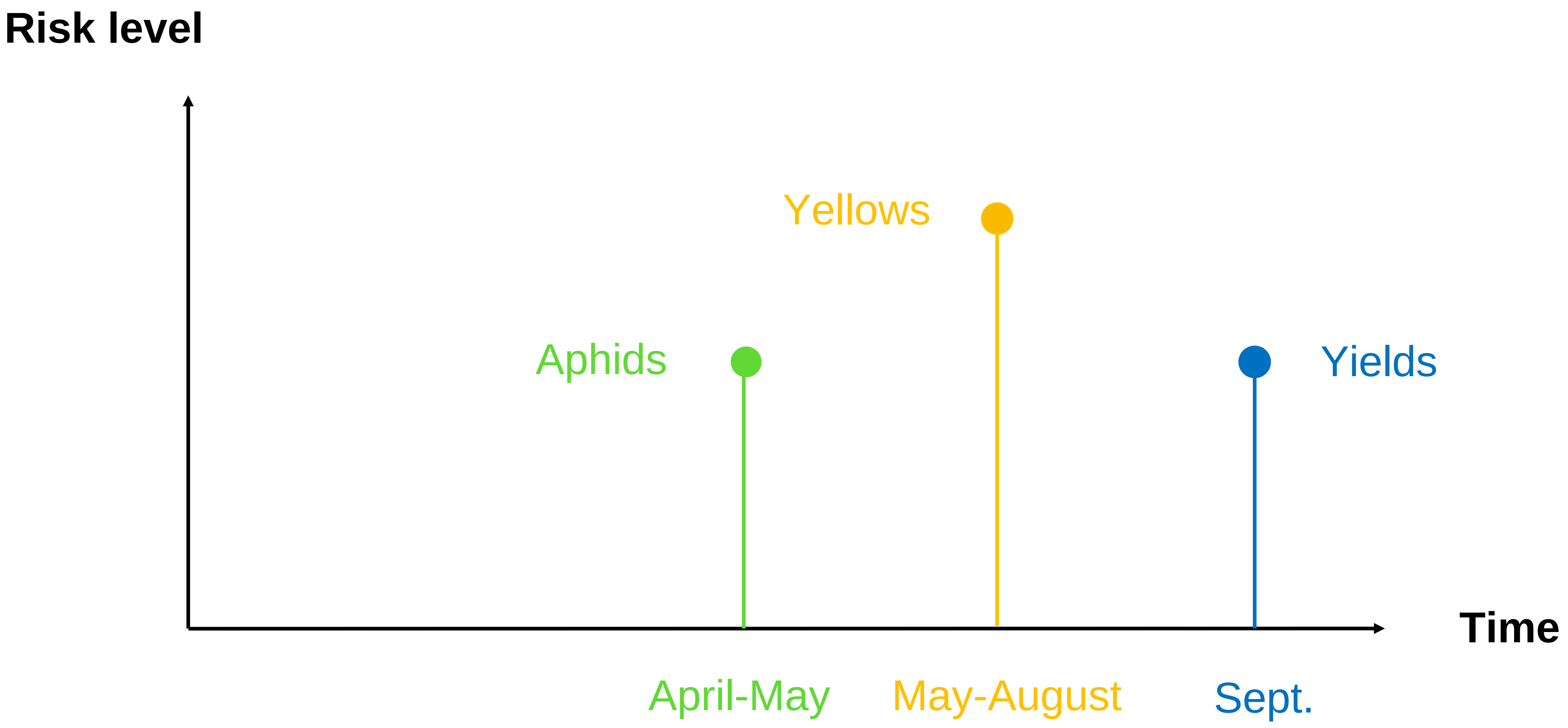
Need for an increased level of information to implement these actions effectively

Need for an increased level of information about different parts of the epidemiological cycle

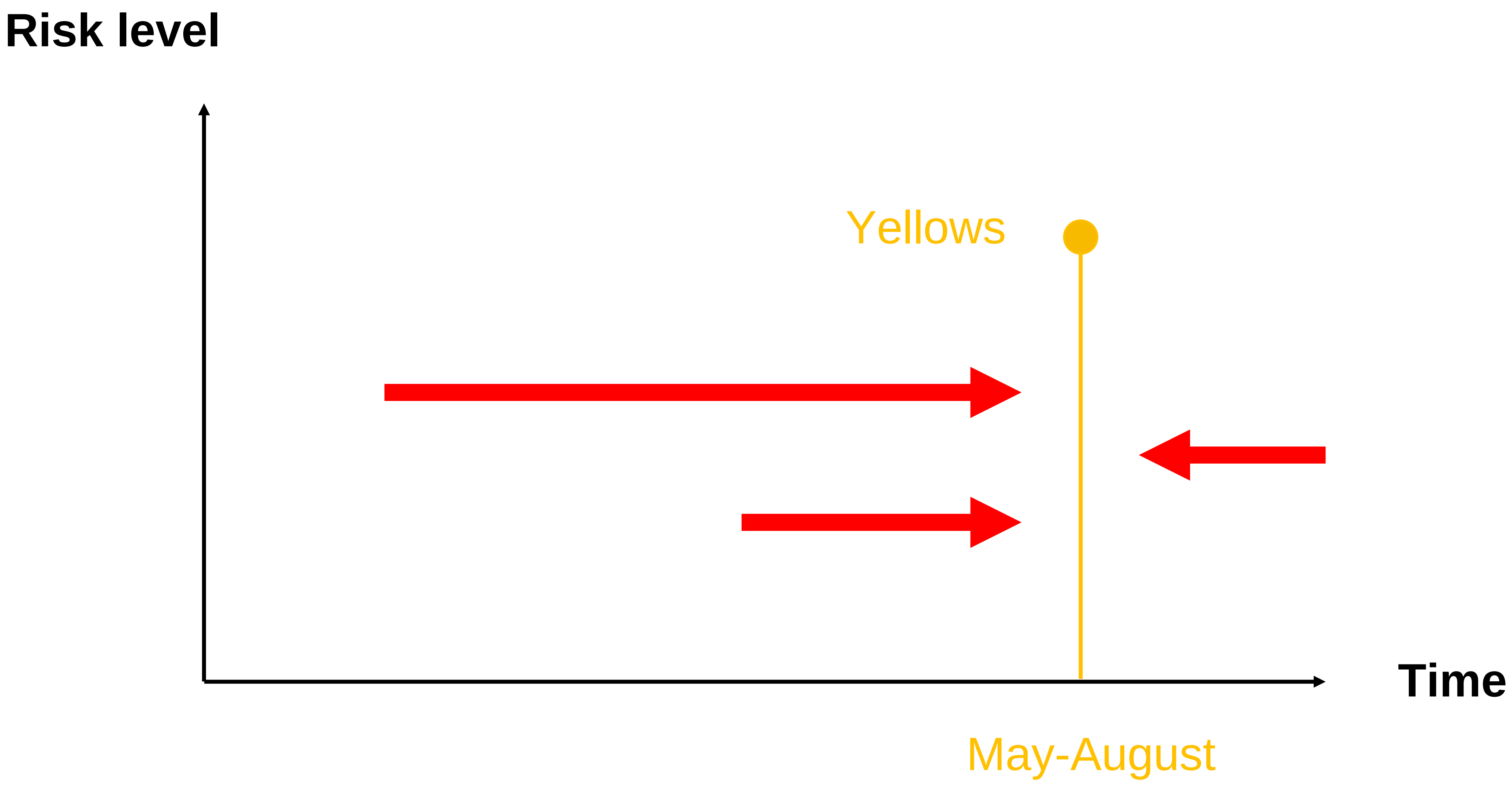


https://www.itbfr.org/fileadmin/user_upload/PDF/Fiches_Bioagresseurs/Gestion_integree_-_jaunisse_2020_web_01.pdf

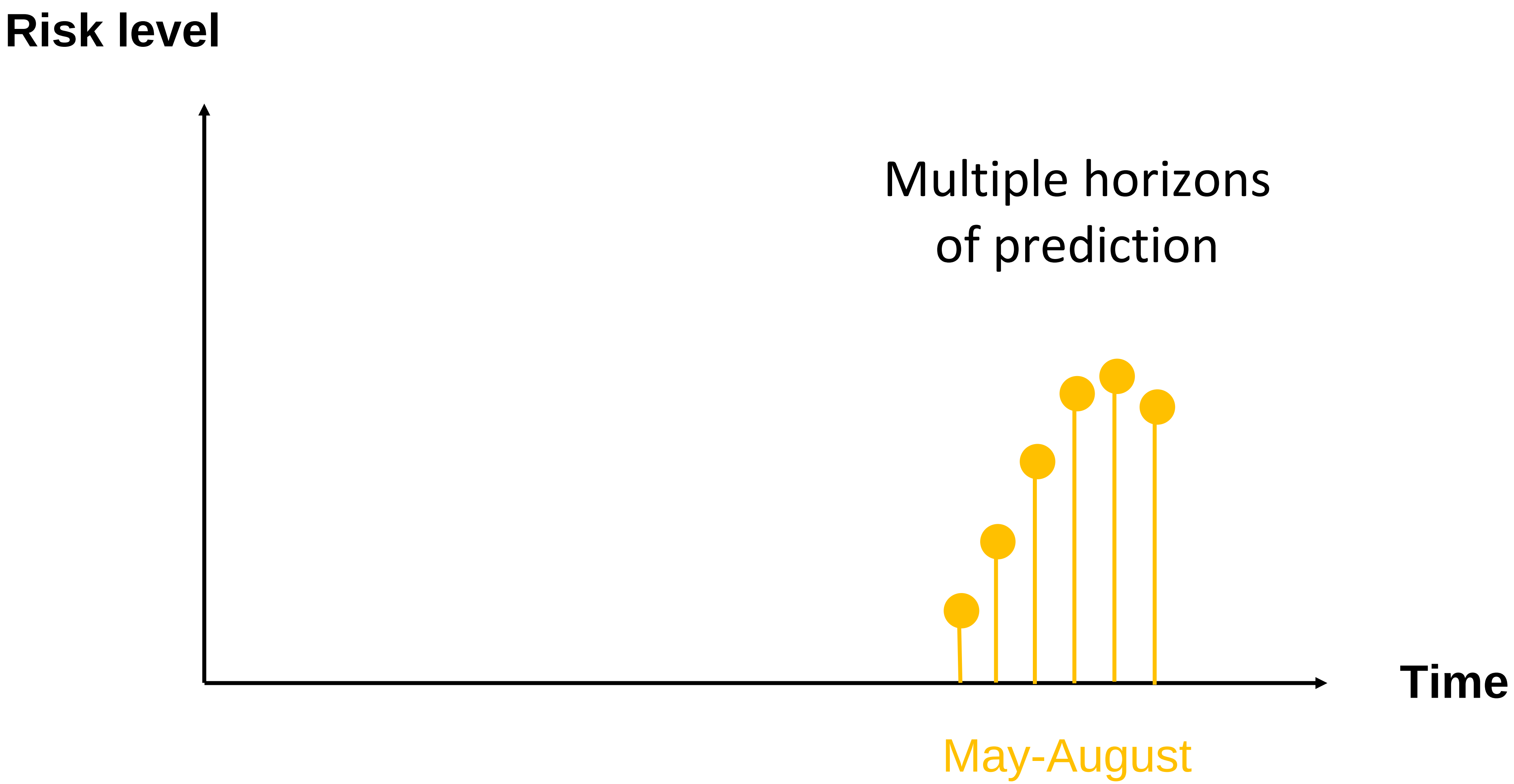
Need for an increased level of information concerning different variables relevant at different times



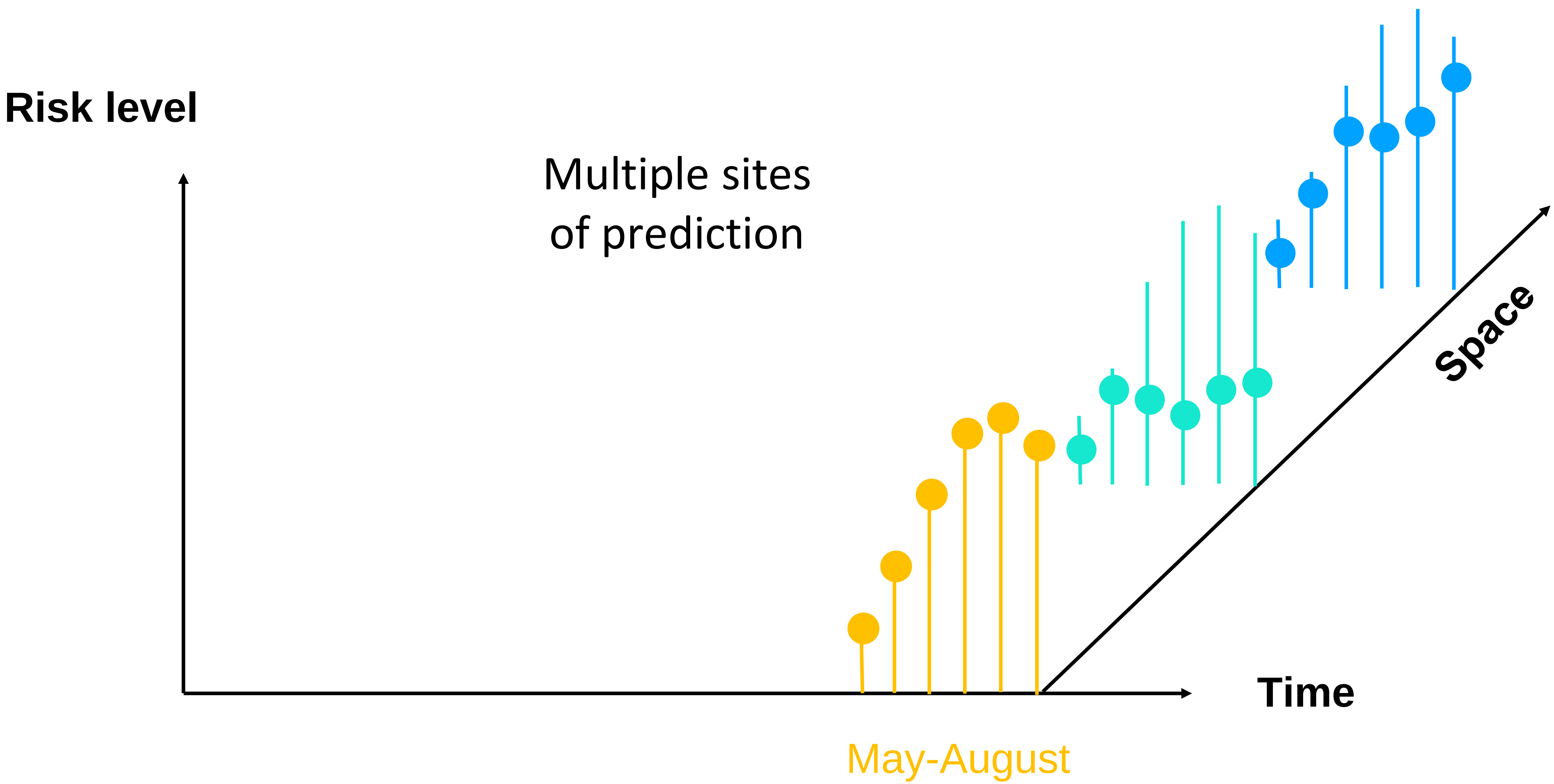
Need for an increased level of information at different times from the prediction horizon



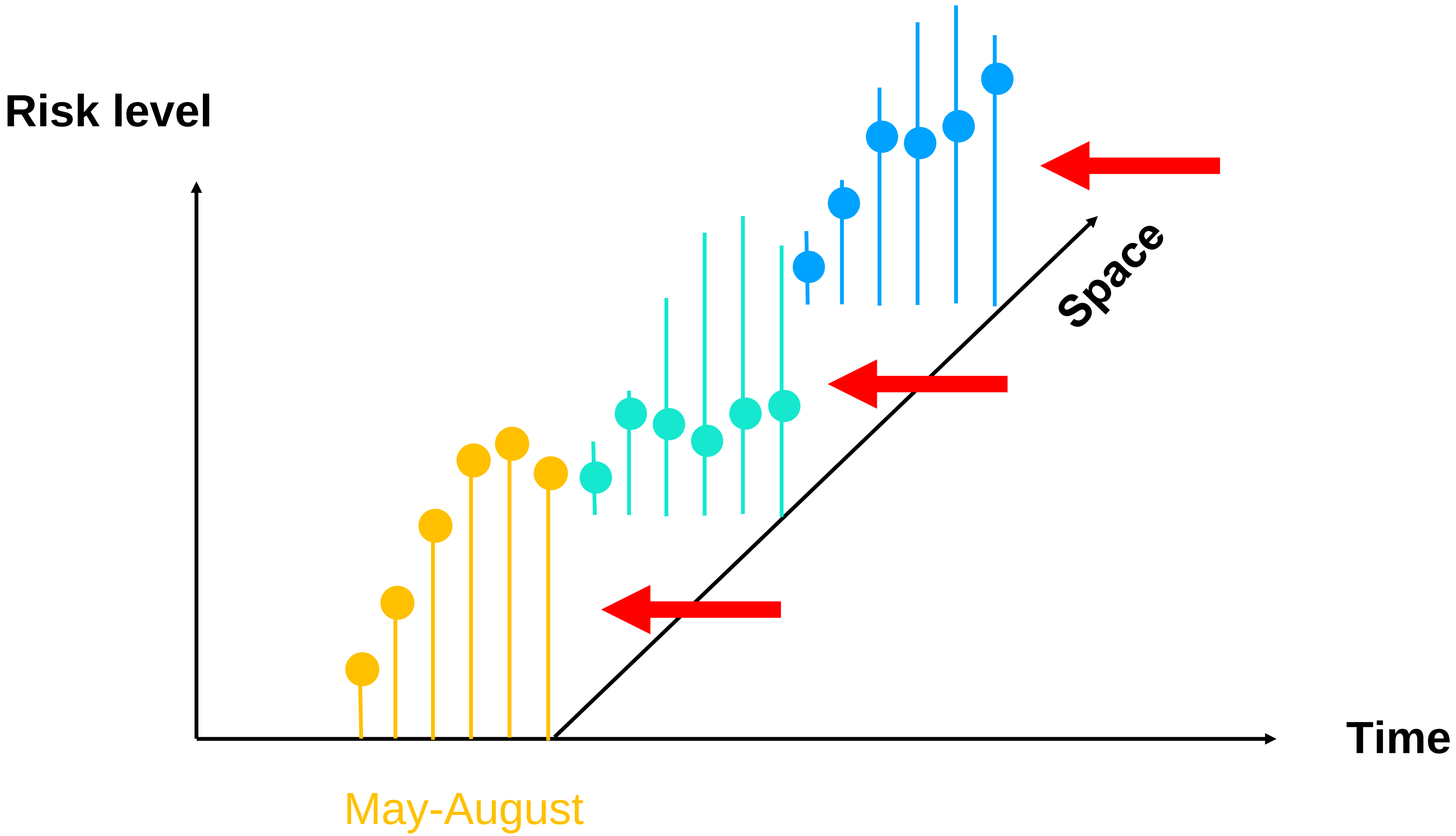
Need for an increased level of information at different times from the prediction horizon



Need for an increased level of information at different times from the prediction horizon



Our focus here: To infer disease progress in the past over large territories



Final objective: To inform crop insurance solutions and claims and/or governmental compensations (how much disease dynamics explain yield losses?)

Our focus here: To infer disease progress in the past over large territories

Specific goal: To estimate the temporal evolution of beet yellows severity at the plot resolution for a large number of plots

Different types of observations may be mobilized:

- Human vision
- Genetic-based diagnostic (plants, insect vectors…)
- Smartphone camera
- Drone camera
- Satellite images

Proposal: an approach combining

- field observations based on human vision (a priori precise but partial)
- satellite observations (a priori less precise but with a high spatial coverage rate)

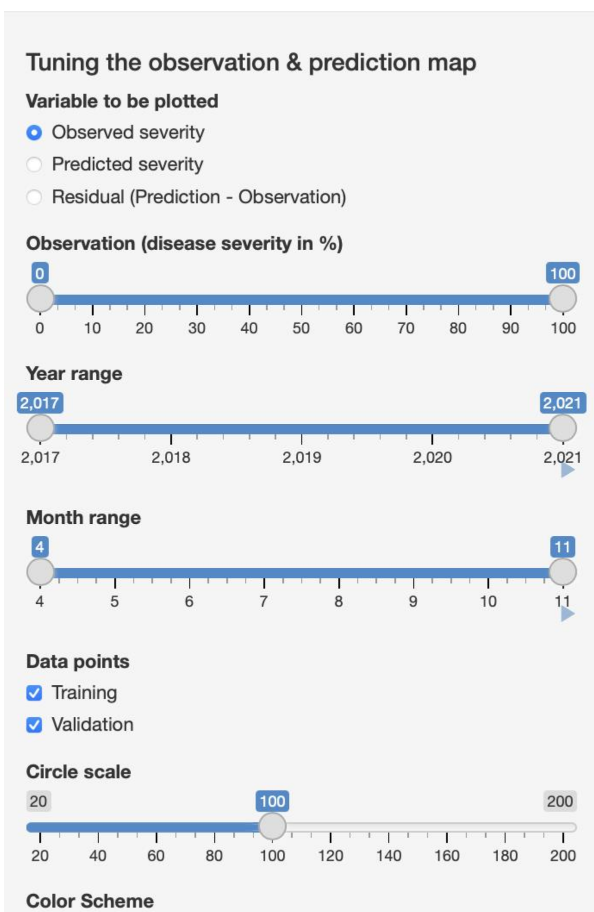
Field observations

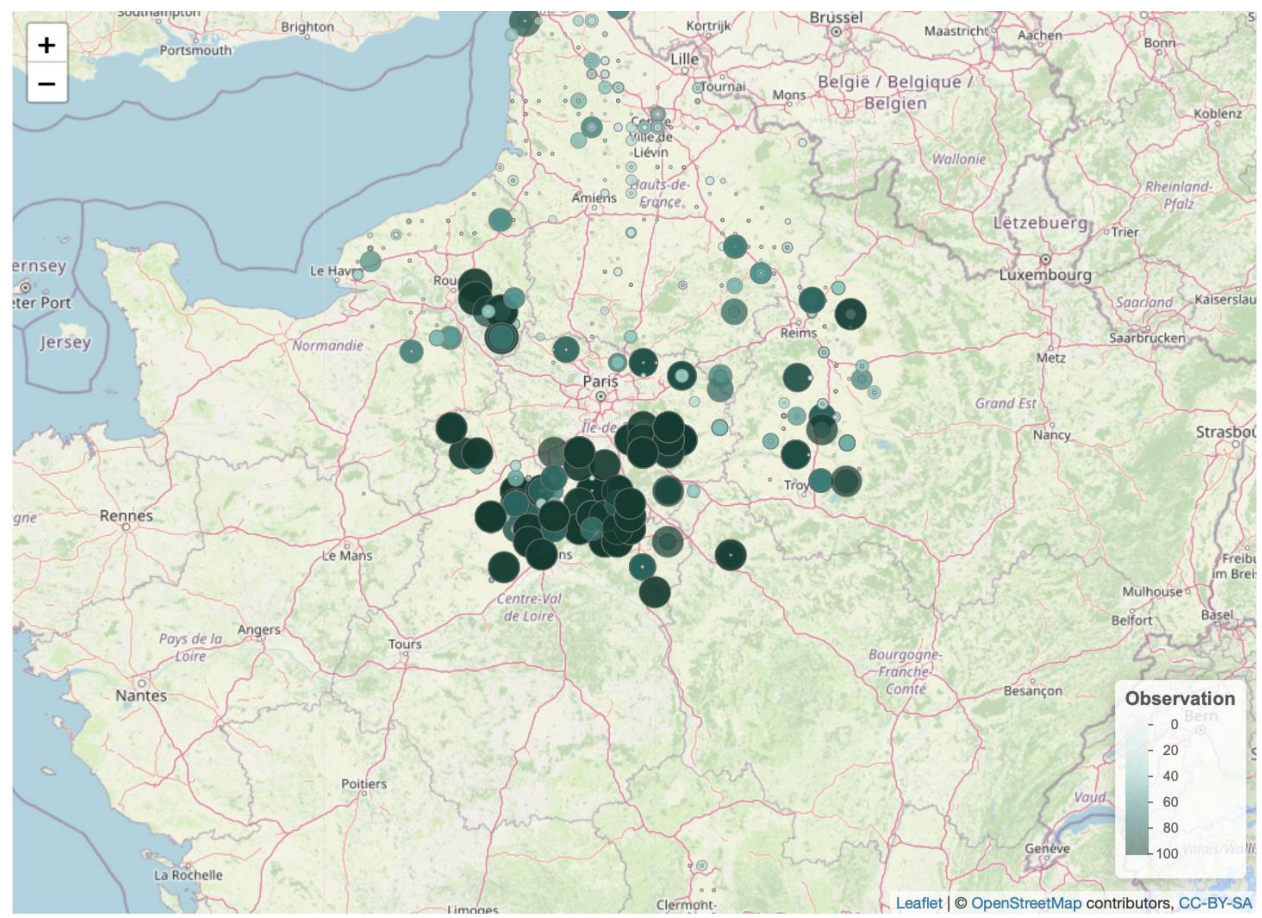
Visual assessment of the severity of beet yellows in the plot (percentage of vegetation cover with symptoms)

- Five years of data (1528 observations for 621 plots)

2017	2018	2019	2020	2021
63	35	256	930	244

- For each plot, from 1 to 14 observations in the same year (average = 2.5)
- A few additional variables: observation date, coordinates, sowing date, beet variety





Satellite observations

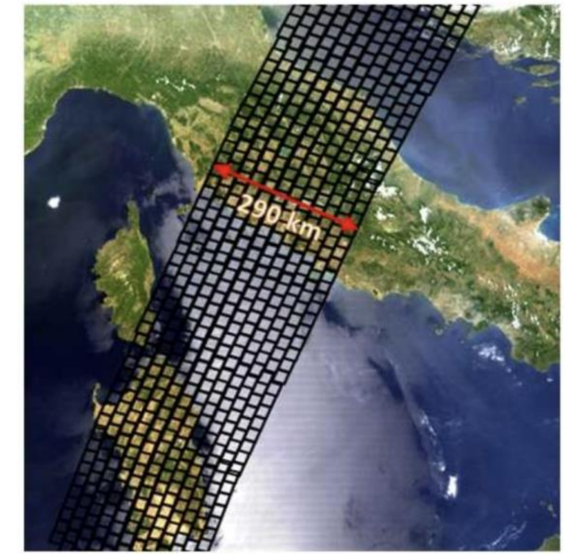


Figure 13: Typical orbit showing layout of Level-1B product granules

Use of Sentinel-2 images

- Free and open data
- Theoretical passage over a covered point: every 2-3 days at mid-latitudes
  In practice (in our study): time lag from 0 to 14 days (average = 6.4 days)
- 13 spectral bands:
  - 4 bands at 10 m: 490 nm (B2), 560 nm (B3), 665 nm (B4), 842 nm (B8)
  - 6 bands at 20 m: 705 nm (B5), 740 nm (B6), 783 nm (B7), 865 nm (B8a), 1 610 nm (B11), 2 190 nm (B12)
  - 3 bands at 60 m: 443 nm (B1), 945 nm (B9) and 1 375 nm (B10)

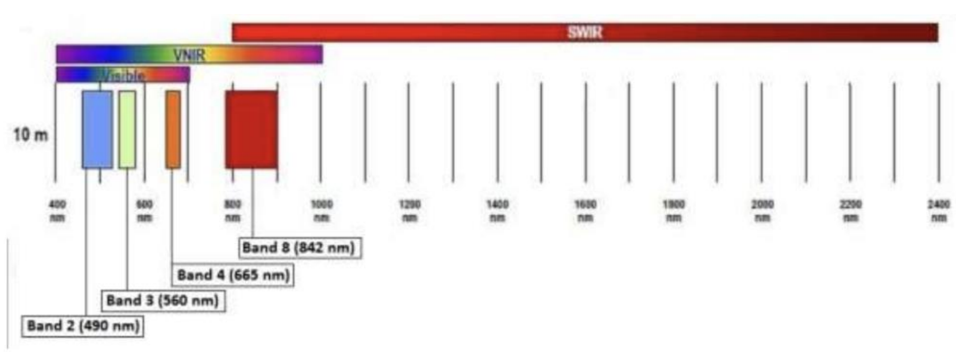


Figure 18: Sentinel-2 10 m Spatial Resolution Bands: B2 (490 nm), B3 (560 nm), B4 (665 nm) and B8 (842 nm)

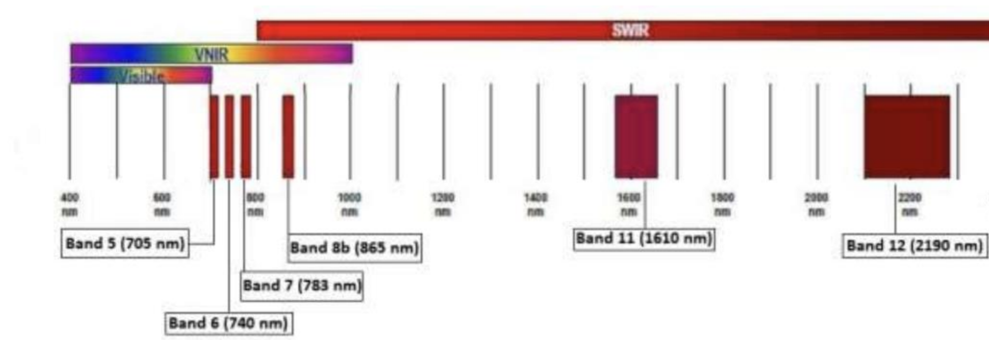
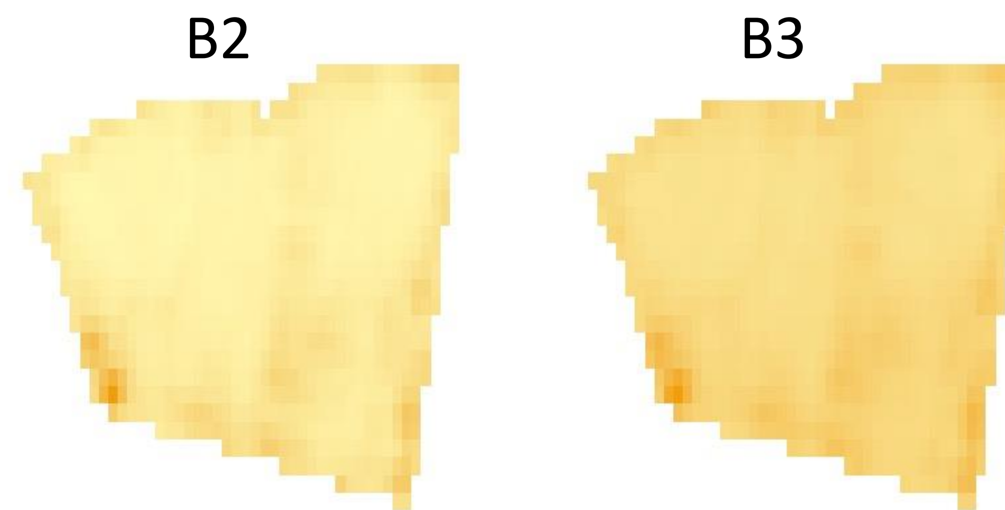


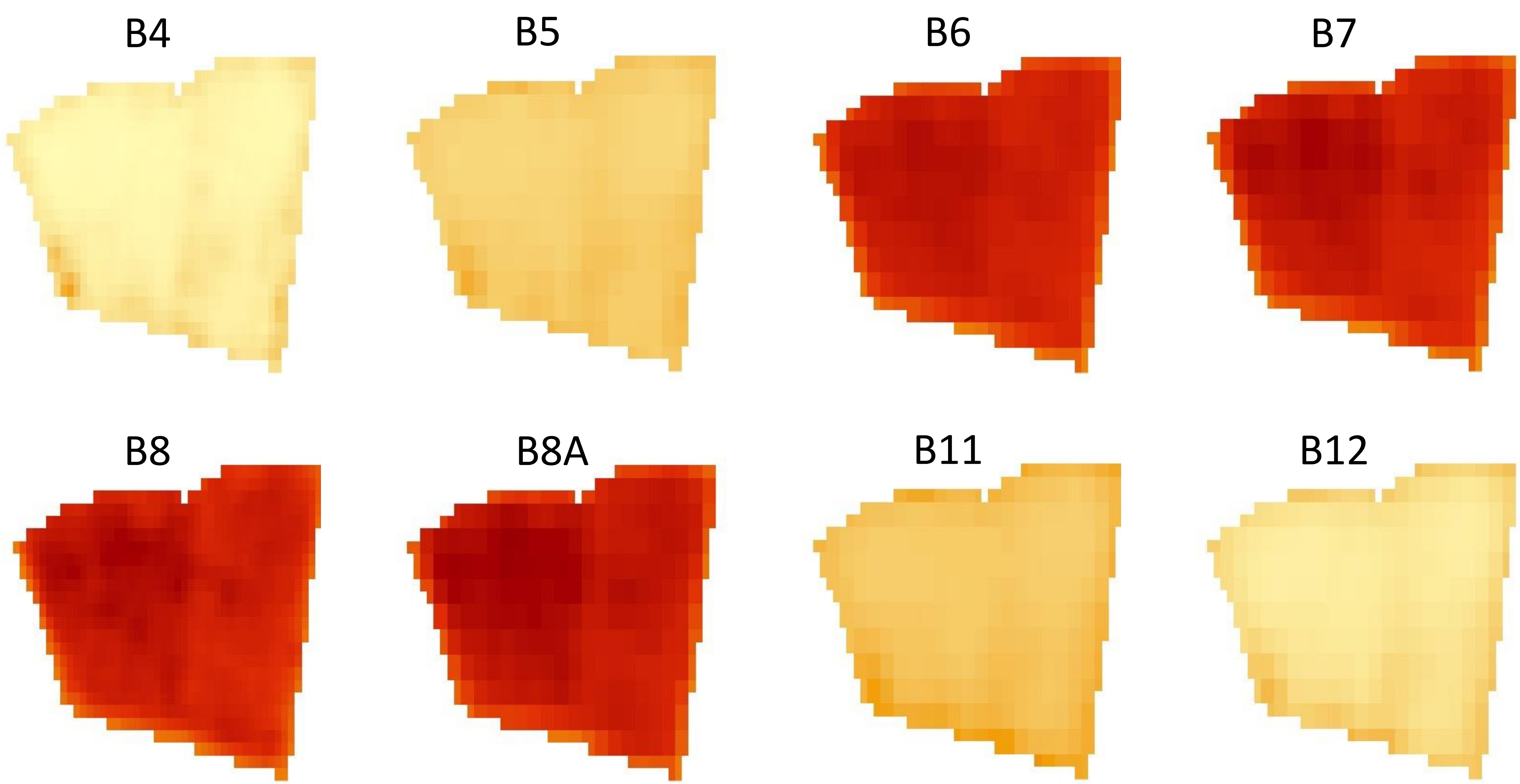
Figure 19: Sentinel-2 20 m Spatial Resolution Bands: B5 (705 nm), B6 (740 nm), B7 (783 nm), B8b (865 nm), B11 (1610 nm) and B12 (2190 nm)

https://sentinel.esa.int/web/sentinel/missions/sentinel-2

Satellite observations

Examples of images for a given plot and a given passage date





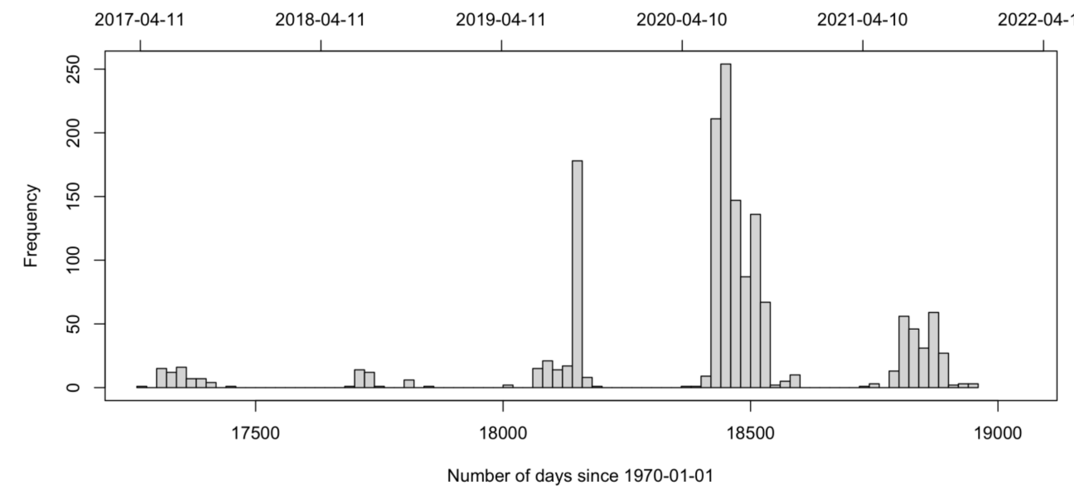
Satellite observations

43 indicators built from the images and additional variables

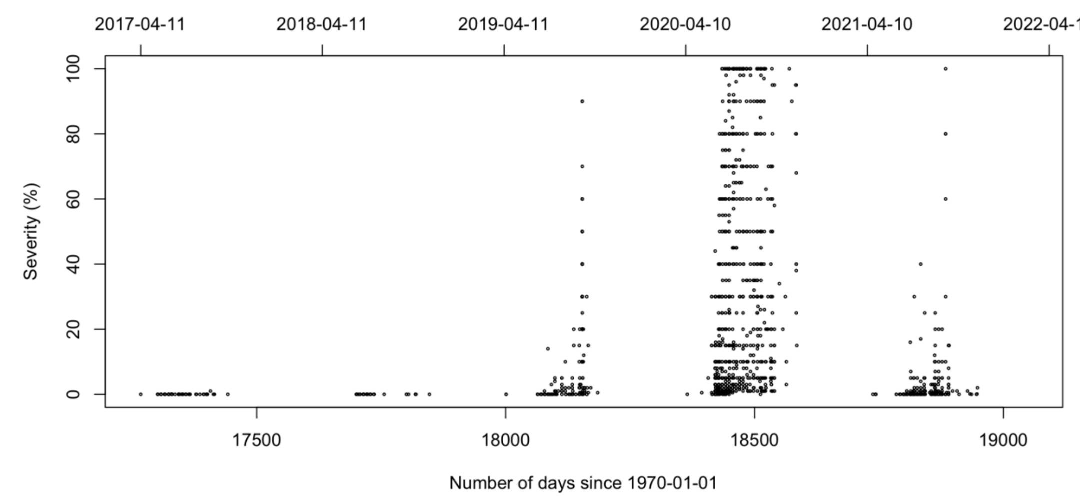
```
 [1] "NDVI"              "VARI"                 "EVI"              "NDRE"
 [5] "mNDblue"           "MTCI"                 "TCARI_OSAVI"      "TCARI_OSAVI_705_750"
 [9] "MCARI_OSAVI"       "MCARI_OSAVI_705_750"  "CIgreen"          "CIrededge"
[13] "NDWI1"             "NDWI2"                "SR_SWIR"          "MCARI"
[17] "MCARI2"            "GNDVI"                "MSAVI"            "CVI"
[21] "inv.SIPI3"         "SR_740_705"           "REPLI"            "TCI"
[25] "ARI"               "inv.CARI"             "LCI"              "MTVI1"
[29] "MTVI2"             "FRE_B2"               "FRE_B3"           "FRE_B4"
[33] "FRE_B8"            "FRE_B5"               "FRE_B6"           "FRE_B7"
[37] "FRE_B8A"           "FRE_B11"              "FRE_B12"          "angle1.11"
[41] "angle2.11"         "timelag"              "j_satellite"
```

Resulting data set

2046 observations plot x satellite x date (with a time lag in general between the field observation and the satellite observation)



Five years of data with heterogeneous
- quantities of data
- temporal patterns of observation
- frequencies of observation in plots
- levels of disease severity…

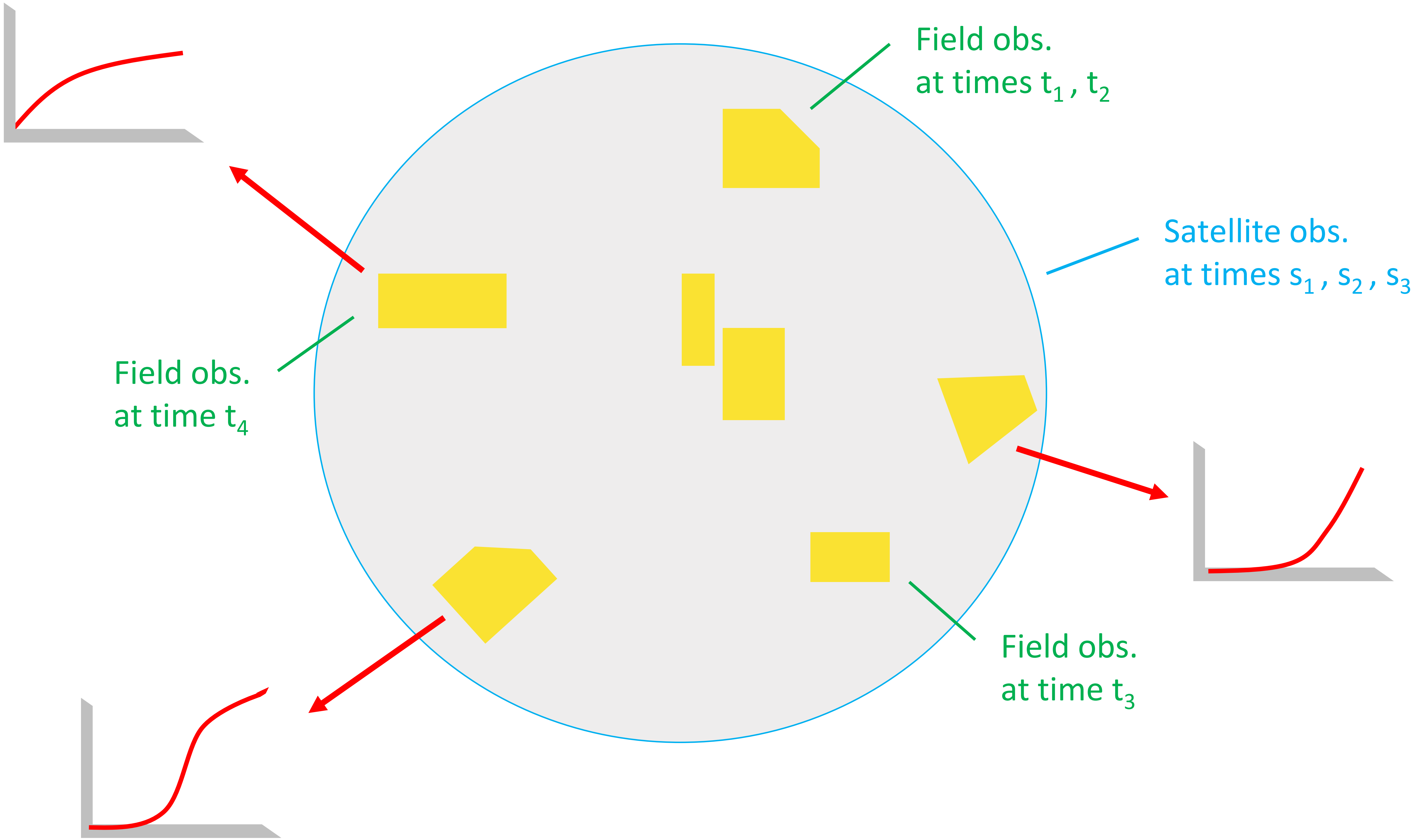


Variables:
- Response variable: disease severity (field observation)
- Explanatory variables: raw satellite images or 48 indicators characterizing the images and additional variables
- Reconstruction of missing values (<1%; 12% for variety, 10% for sowing date) and standardization of variables

```
 [1] "variete"              "SAFRAN_X"             "SAFRAN_Y"              "NDVI"
 [5] "VARI"                 "EVI"                  "NDRE"                  "mNDblue"
 [9] "MTCI"                 "TCARI_OSAVI"          "TCARI_OSAVI_705_750"   "MCARI_OSAVI"
[13] "MCARI_OSAVI_705_750"  "CIgreen"              "CIrededge"             "NDWI1"
[17] "NDWI2"                "SR_SWIR"              "MCARI"                 "MCARI2"
[21] "GNDVI"                "MSAVI"                "CVI"                   "inv.SIPI3"
[25] "SR_740_705"           "REPLI"                "TCI"                   "ARI"
[29] "inv.CARI"             "LCI"                  "MTVI1"                 "MTVI2"
[33] "FRE_B2"               "FRE_B3"               "FRE_B4"                "FRE_B8"
[37] "FRE_B5"               "FRE_B6"               "FRE_B7"                "FRE_B8A"
[41] "FRE_B11"              "FRE_B12"              "angle1.11"             "angle2.11"
[45] "timelag"              "j_notation"           "j_semis"               "j_satellite"
```

Methodology of analysis

Coupling partial field observations and high-coverage satellite observations to infer disease progress curves for every plots

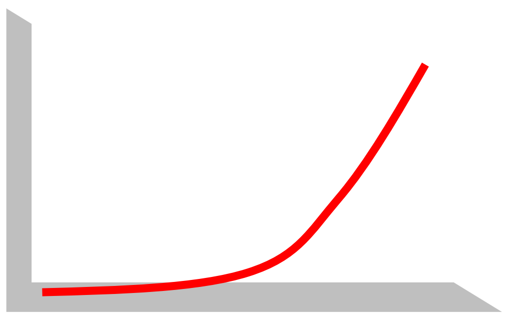


Methodology of analysis

Tools to link disease severity with explanatory variables

- Convolutional neural network (VGG16) applied to images
- Regression models applied to indicators and additional variables
  - Neural network
  - Random forest
  - 0&1-inflated Beta GLM with AIC-based stepwise selection of explanatory variables
  - ...
- Post hoc model refinement:
  - Recycling smoothed versions of preliminary predictors (with different smoothing bandwidths) as explanatory variables to use information from neighborhoods

Tools to model disease progress maps

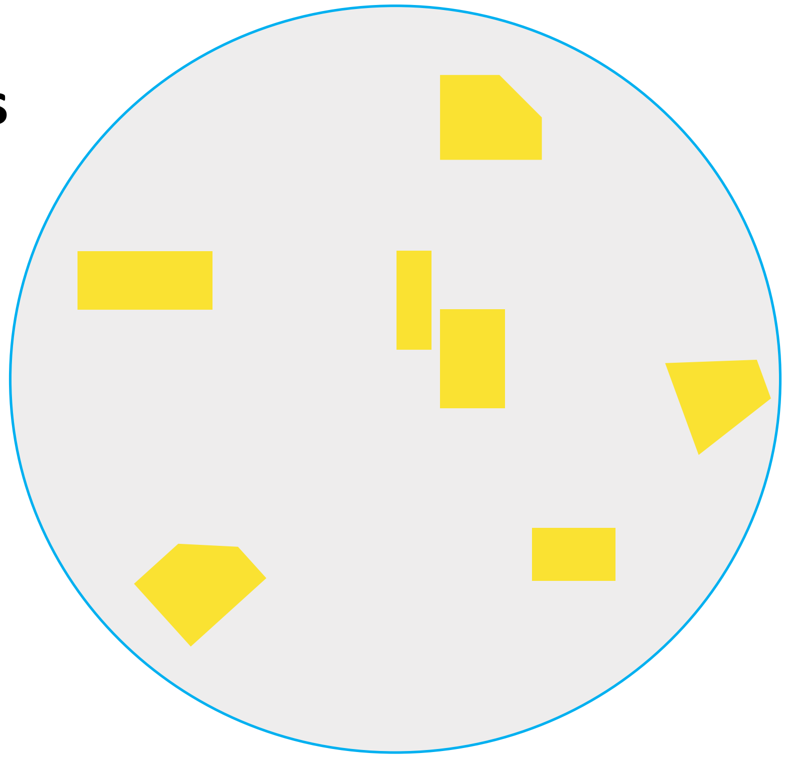


- Logistic model fitted to
  - Observed severities for each plot (with at least 3 observation times)
  - Predicted severities based on satellite data computed at the observation times
- Spatial smoothing of disease progress curves with a bandwidth selected by cross-validation

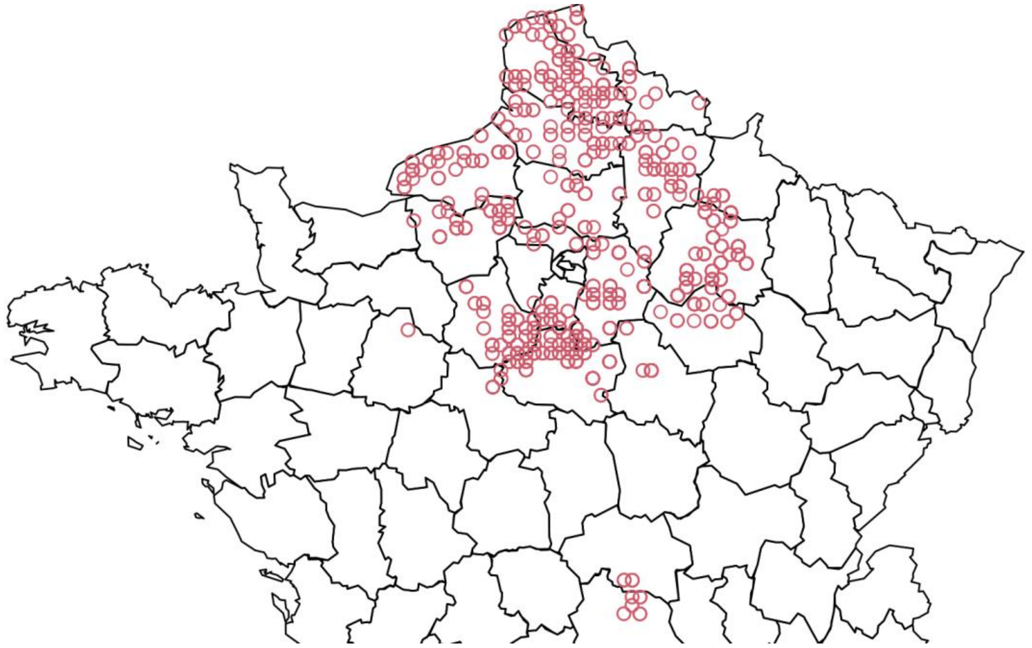
Methodology of analysis

Training and validation sets, and exploration of sampling strategies

Forming the training set to mimic different sampling strategies



1) **Random**: Completely spatially and temporally random sampling
2) **Stratified**: Spatially stratified sampling by administrative division and temporally stratified sampling by year
3) **Stratified wrt plots**: Strategy (2) applied to plots instead of observations (i.e., all observation times for a given plot are either included in or excluded from the training set)
4) **Exhaustive in 2017-2020, Stratified in 2021**: Exhaustive sampling for all the years except the last one, and spatially stratified sampling by administrative division for the last year

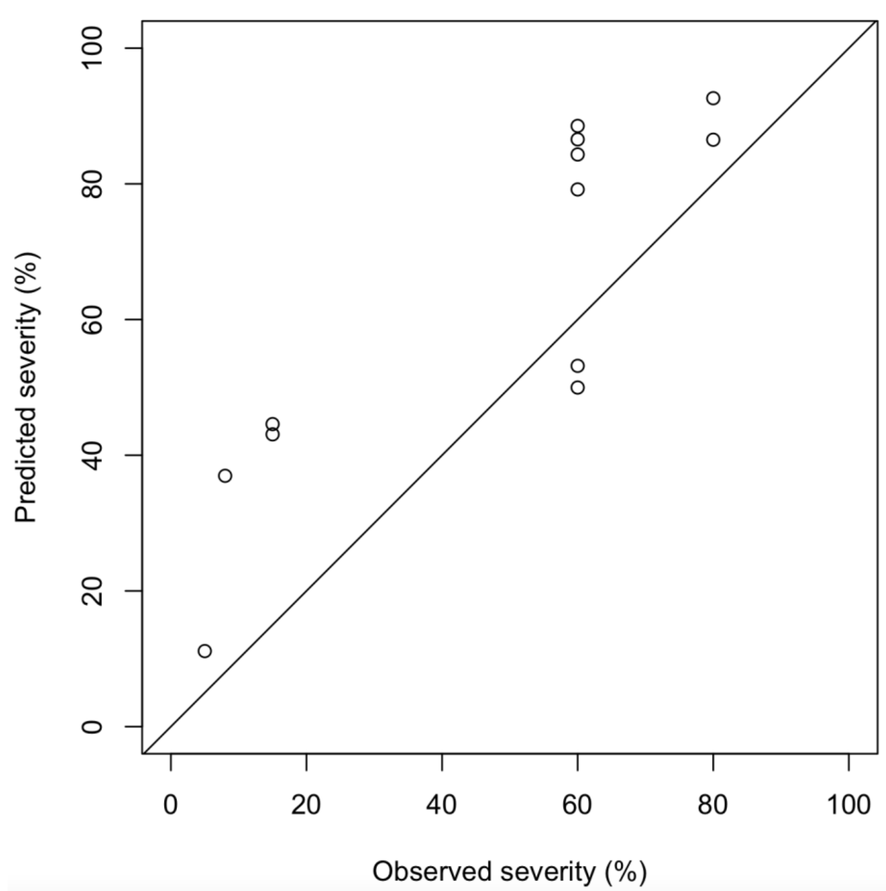


Methodology of analysis

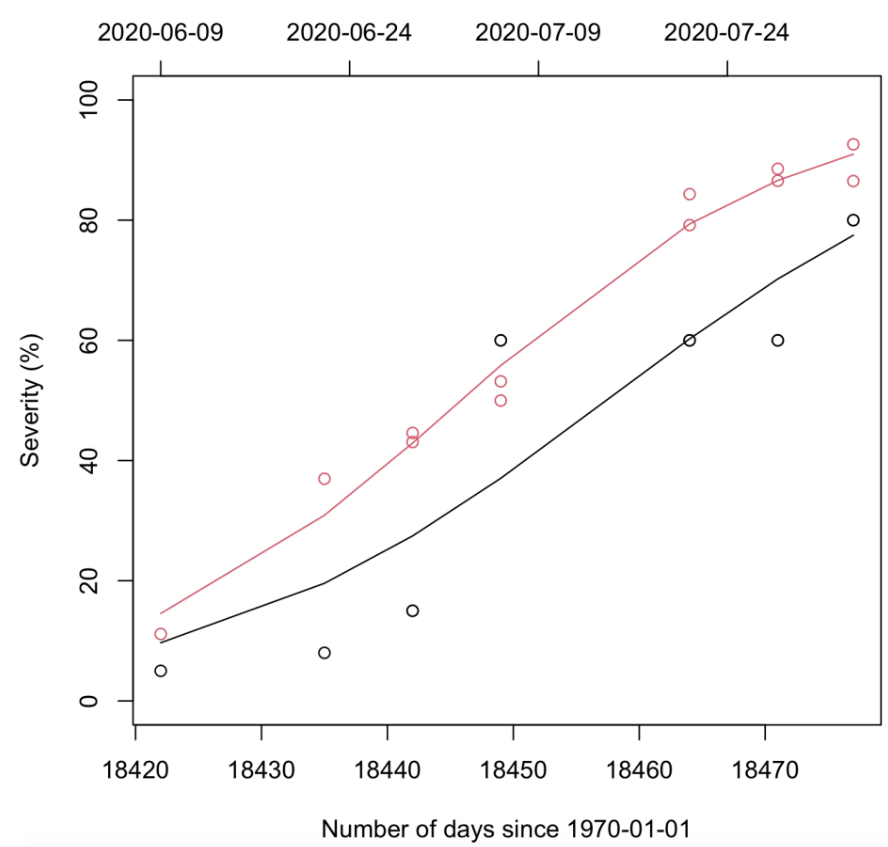
Validation criteria

- Root mean square error (RMSE)
- Coefficient of determination (R2)
- Two scales of validation:

  At plot scale: Comparison between observed severities and predicted severities

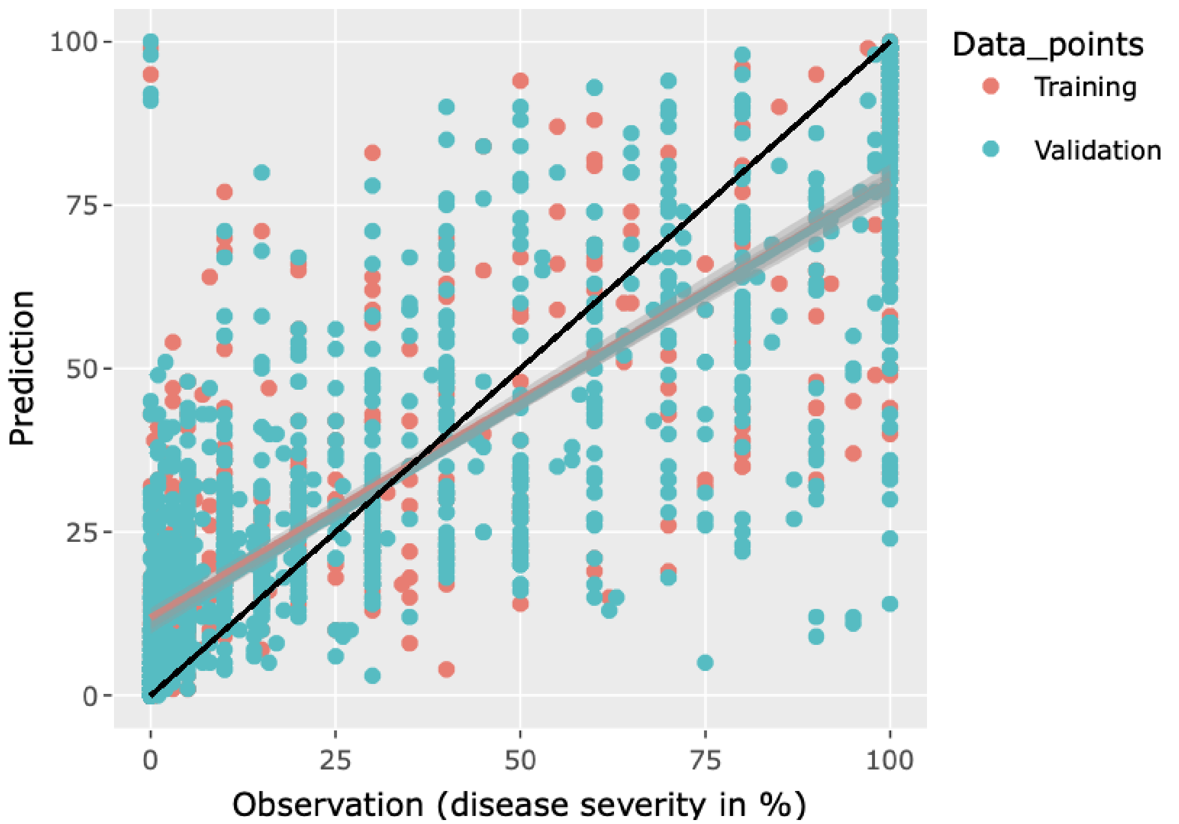


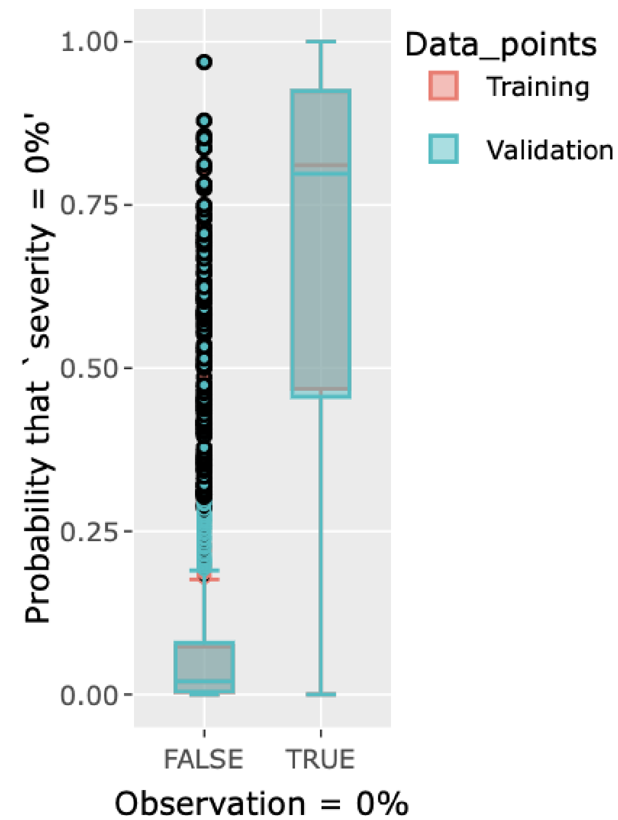
At regional scale: Comparison between smoothed disease progress maps obtained from either observed severities or predicted severities

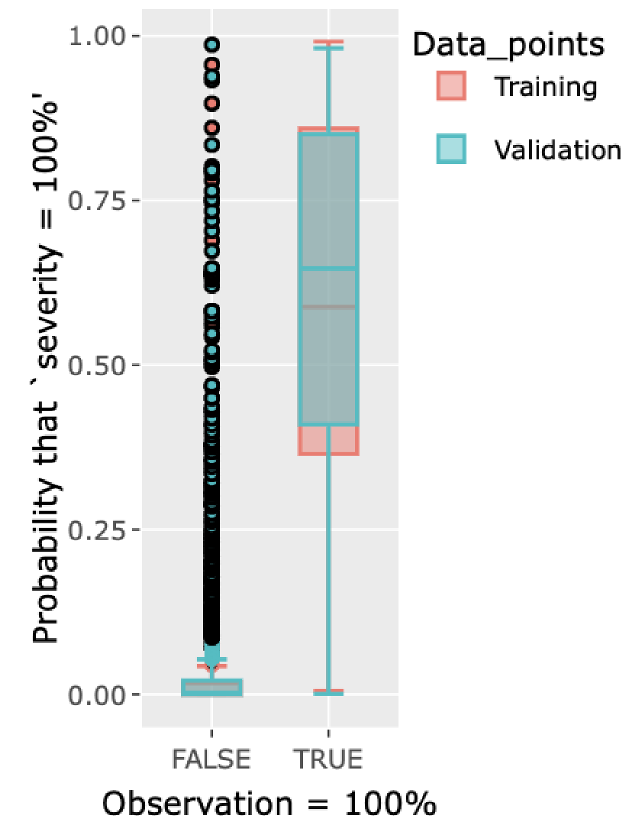


Results

Predicted versus observed severities at plot scale







- Quite large uncertainty in prediction
- However,
  - Rather consistent trend in predictions
  - Rather consistent probabilities for the severity to be equal to 0% or 100%
  - Consistency of performance in training and validation sets

Predictions versus observations

Table of prediction performance

Data points	RMSE	R2
All	0.195	0.702
Training	0.192	0.707
Validation	0.200	0.693

Results

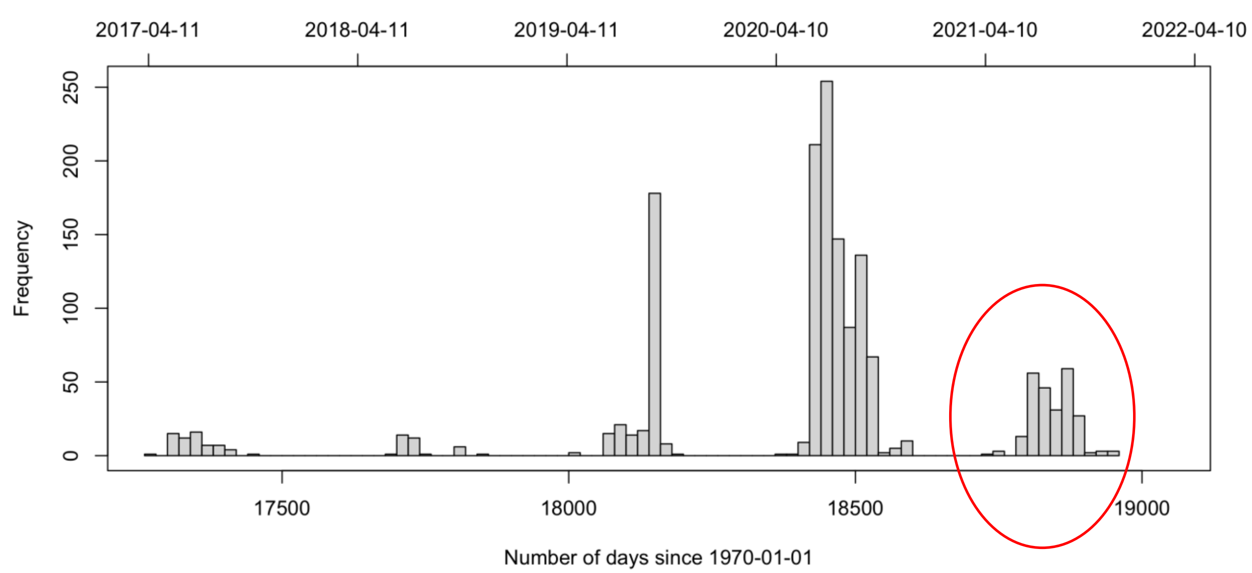
Performance on validation data at plot scale: Random versus stratified sampling strategies

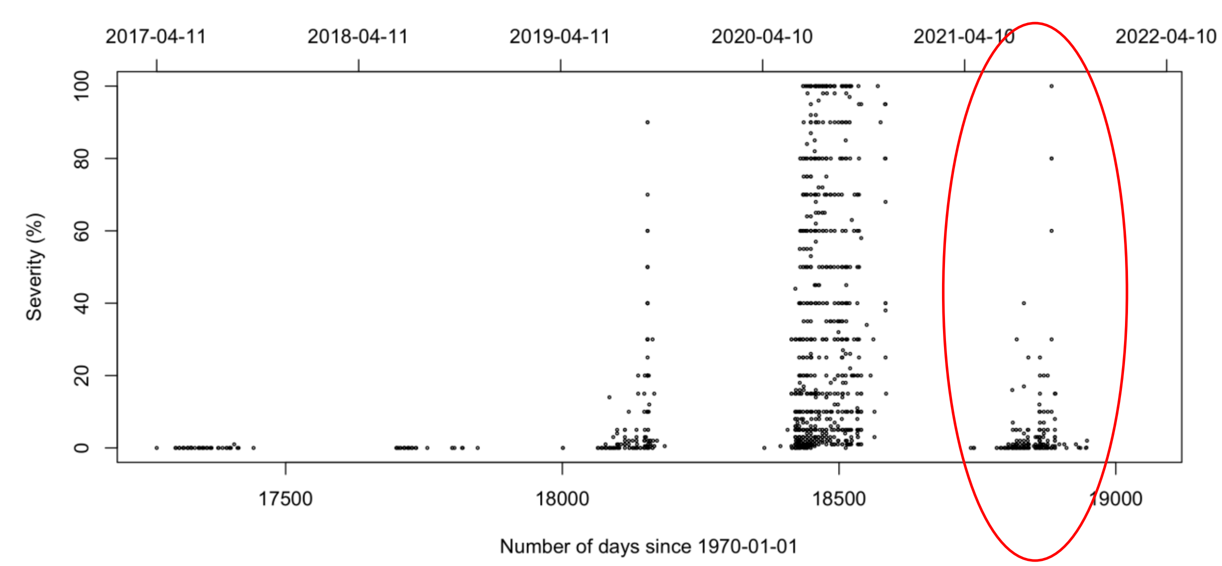
Sampling strategy	Training proportion	RMSE	R2
(1) Random	0.5	0.20	0.68
(1) Random	0.7	0.19	0.71
(2) Stratified	0.5	0.20	0.70
(2) Stratified	0.7	0.20	0.69
(3) Stratified wrt plots	0.5	0.23	0.61
(3) Stratified wrt plots	0.7	0.22	0.63

Results

Performance on validation data at plot scale: Prediction for a specific year

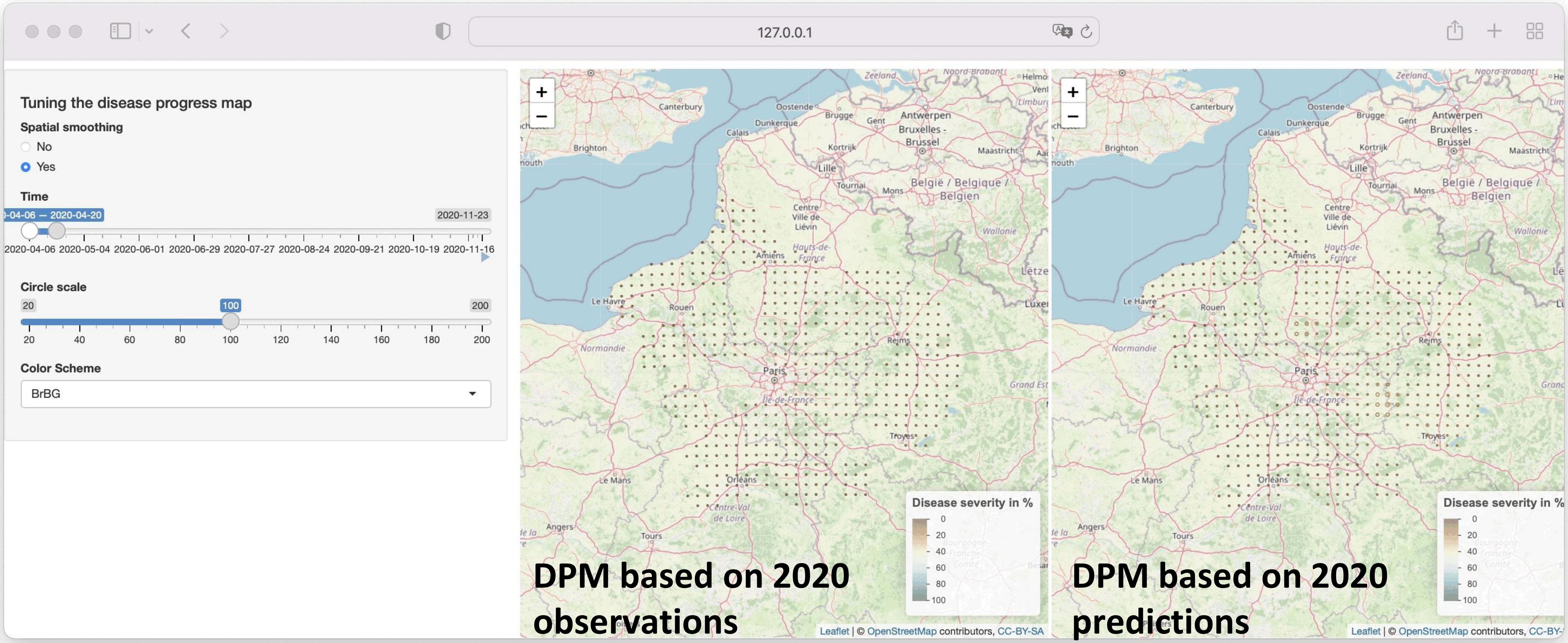
Sampling strategy	Training proportion	RMSE(2021)	R2(2021)
(2) Stratified	0.5	0.12	0.24
(2) Stratified	0.7	0.10	0.24
(4) Exhaustive in 2017-2020 No data in 2021	0.0	0.16	0.03
(4) Exhaustive in 2017-2020 Stratified in 2021	0.2	0.14	0.14
(4) Exhaustive in 2017-2020 Stratified in 2021	0.5	0.14	0.23





Results

Disease progress map (DPM) at regional scale



Results

Performance at regional scale: Random versus stratified sampling strategies

Sampling strategy	Training prop.	RMSE-2020	R2-2020	RMSE-2021	R2-2021
(1) Random	0.5	0.11	0.90	0.16	0.59
(1) Random	0.7	0.11	0.91	0.15	0.48
(2) Stratified	0.5	0.11	0.92	0.15	0.42
(2) Stratified	0.7	0.11	0.91	0.15	0.51
(3) Stratified wrt plots	0.5	0.11	0.91	0.11	0.65
(3) Stratified wrt plots	0.7	0.11	0.92	0.14	0.64

Results

Performance at regional scale: Prediction for a specific year

Sampling strategy	Training proportion	RMSE(2021)	R2(2021)
(2) Stratified	0.5	0.15	0.42
(2) Stratified	0.7	0.15	0.51
(4) Exhaustive in 2017-2020 No data in 2021	0.0	0.19	0.48
(4) Exhaustive in 2017-2020 Stratified in 2021	0.2	0.17	0.61
(4) Exhaustive in 2017-2020 Stratified in 2021	0.5	0.16	0.56

Conclusions and perspectives

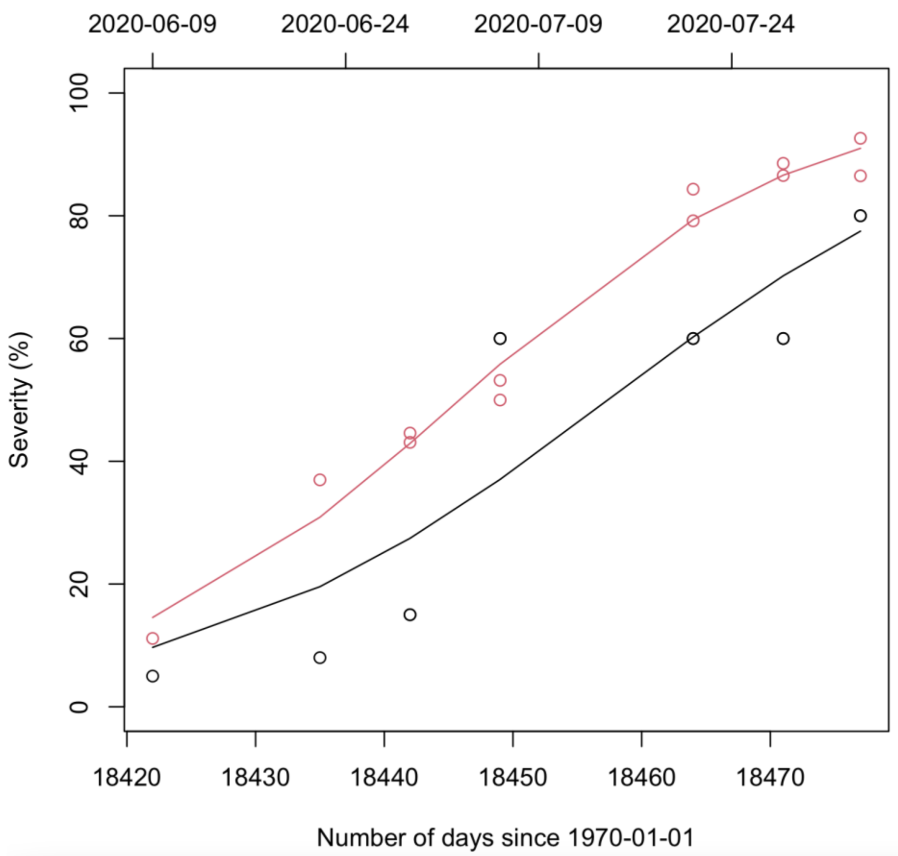
- Random and stratified sampling strategies approximately equally perform

- Based on Sentinel-2 data and the considered model, using no field observation for the year of interest and betting only on satellite observations may lead to poor performance

- The post-hoc model refinement typically allows a 30%-increase of R2: The preliminary smoothed predictors used as complementary explanatory variables in this post-hoc approach are surrogates for coupled effects “year x area” at diverse spatial scales
  -> Applying this refinement to other models than the 0&1-inflated Beta GLM

- Including the 0&1-inflation in other models

- Encouraging results but not completely satisfactory because of the relatively high prediction uncertainty
  -> Using satellites with higher-spatial resolution
  -> Annotating the higher-spatial resolution images and using a model adapted to annotated images
  -> Using drone-based photographs to make field observations more reliable



- In the insurance and compensation perspective, including yield data (generally at low spatial resolution) in the analysis
  -> Deeper integration of heterogeneous data

Conclusions and perspectives

Challenges identified in the BEYOND project concerning the use of satellite-based information in the context of plant health surveillance:

- Exploiting satellite remote sensing to contribute to the surveillance of plant diseases or syndromes in a multi-layer surveillance strategy

- Developing consistent integration methods of in-field data and remote sensing data in the inference of unknowns (parameters and latent processes) of spatio-temporal models

- Using satellite remote sensing to refine knowledge about the spatial distribution of crops and reservoirs at a finer resolution (in terms of crop / reservoir categories) than the resolution of typical land-use databases