Creating country scale predictive pollen and pesticide maps using bee hives as monitoring tools

A modelling exercise in the context of the INSIGNIA EU-pilot project

Bas Buddendorf, Hans Baveco, Ivo Roessink

Wageningen Environmental Research
Wageningen, month 2021

Reviewed by:
Jozef van der Steen (Alveus AB Consultancy)

Buddendorf, W.B., Baveco, J., Roessink, I. 2021. *Creating country scale predictive pollen and pesticide maps using bee hives as monitoring tools. A modelling exercise in the context of the INSIGNIA EU-pilot project.* Wageningen, Wageningen Environmental Research, Report. 27 pp.; 10 fig.; 10 ref.

The widespread decline of pollinators in the agricultural landscape has led to serious concerns over the continued viability of honeybee colonies and provision of critically important pollination services they provide. Honeybees exploit mass-flowering crops and can cover long distances to meet their energy requirements. As a result, the specific pollen found in a colony reflect the foraging landscape. Similarly, the pollutants found in the colony, which are involuntarily collected during foraging reflect the environmental pollution status. As part of the EU INSIGNIA project one of the goals was to predict the pollen diversity in the landscape and potential exposure of honeybees to contaminants. Pollen and pesticides entering the hive that were sampled during the INSIGNIA monitoring campaigns of 2019 and 2020 were used to construct pollen and pesticide classification models. For 2019 predictions were made for Austria and Denmark; for 2020 predictions were made for Austria, Belgium, Denmark, France, Greece, Ireland, Italy, Latvia, and the United Kingdom. Models rely heavily on the input data. As such, high or low predicted numbers of either pollen or pesticides need to be considered in the context of data quality and of the apiary locations that were used to collect the data. With the relatively limited spatial coverage within the current pilot project, linking landscape characteristics and climate to the pollen/pesticide predictions is difficult. Nevertheless, large-scale patterns related to land-use, climate, and floral diversity may be inferred. Model outputs can inform stakeholders, for example beekeepers, and help to make trade-offs between the expected availability and diversity of floral resources and the expected exposure to pesticides within a country and during the growing season.

In 2003 Wageningen Environmental Research implemented the ISO 9001 certified quality management system.
Since 2006 Wageningen Environmental Research has been working with the ISO 14001 certified environmental care system.
By implementing the ISO 26000 guideline, Wageningen Environmental Research can manage and deliver its social responsibility.

Wageningen Environmental Research report | ISSN 1566-7197

Contents

Summary

The widespread decline of pollinators in the agricultural landscape, resulting from agricultural intensification and consequent increase in stress from pesticides and lack of floral resources, has led to serious concerns. Honeybees exploit mass-flowering crops and can cover long distances to meet their energy requirements. Consequently, they are exposed to contaminants present in their local environment and honeybees involuntarily collect these contaminants during foraging. Thus the specific pollen found in a colony reflect the foraging landscape and the pollutants found in the colony reflect the environmental pollution status.

The INSIGNIA project aims to develop a best practice protocol for Apicultural citizen science and to conduct "*A pilot study on the best practices for a European wide monitoring program with honey bee colonies in an Apicultural citizen science (CS) setting to study pesticide exposure of honey bees and investigation of pollen sources*". One of the ambitious goals within the project was to predict the potential exposure of honeybees to contaminants in the landscape. In order to predict exposure at large spatial scales like the country scale, collected pesticide data from honeybee colonies was linked to land-use characteristics. Similar links were established to estimate the diversity of the foraging landscape.

Pollen and pesticides entering the hive that were sampled during the INSIGNIA monitoring campaign were used to construct a 'pollen and pesticide classification model'. For this two main types of input geo-data were collected and prepared: the monitoring data for pollen and pesticides collected in the INSIGNIA study; and two sources of spatial and temporal geodata, i.e. the CORINE Land Use and Land Cover data and a pan-European weather data set. A machine learning approach was used, a so-called random forest. Like real forests, a random forest is made up of individual trees, or classification trees. Observations (from pollen and APIStrip samples) are passed along the branches of a tree and placed in an outcome category (i.e., pollen family or pesticide type). By combining the results of all trees in the random forest model, an ensemble (combined) output is created giving probabilities of finding certain pollen families or pesticides. This makes it a powerful model for the INSIGNIA data. The random forest model can be used as a predictive model that improves as more data are added.

Predictions were made for data collected in 2019 and in 2020. In 2019 this was done for Austria and Denmark, whereas in 2020 predictions were made for Austria, Belgium, Denmark, France, Greece, Ireland, Italy, Latvia, and the United Kingdom. In 2019 pollen diversity in Austria decreased from May to September, from the North-East towards South-West. Over the year pollen diversity was lowest in the North and North-East of the country while in the East, West, and South-West pollen diversity remained high throughout the year. The number of pesticides was consistently low in the East of Austria. Higher pesticide numbers in June-August appeared to coincide with high pollen diversity in May and June but from July onwards patter was seemingly reversed and higher pesticide numbers appeared to be in areas with a lower pollen diversity. Pollen diversity in Denmark was predicted to follow a relatively uniform seasonal pattern with an increase between May and July, and a decrease towards September. Areas to the West and North were generally predicted to be less diverse than the East, with large areas showing a strongly uniform distribution in the predicted number of pollen families. The use of pesticides was at a higher level early in the year, but increased towards the West from May to July The number of pesticides showed more spatial variability in August and September.

The inclusion of additional countries in 2020 allowed for the generation of Europe scale maps. This increase makes more evident that early in the growing season the higher latitude countries like Denmark and Latvia show a reduced pollen diversity compared to for example Austria, Belgium. Conversely, Italy and Greece showed a decreasing pollen diversity compared to other countries as the season progresses. This is possibly due to the fact that the major flowering period for (mass) flowering crops is earlier in the year in the warmer Mediterranean climates of Greece and Italy.

Model outputs can inform stakeholders on the expected availability and diversity of floral resources and on expected exposure to pesticides within a country and during the growing season. This information is

valuable as it allows, for example, beekeepers to make informed decisions on the placement of hives and make trade-offs between forage availability and exposure risk. Importantly, both models rely heavily on the input data. As such, high or low predicted numbers of either pollen or pesticides need to be considered in the context of the apiary locations that were used to collect the so-called model training data and also the quality of the collected data. With the relatively limited spatial coverage within the current pilot project, linking landscape characteristics and climate to the pollen/pesticide predictions is difficult. Nevertheless, large-scale patterns related to land-use, climate, and floral diversity may be inferred. Improvements of the model performance can be achieved through A) a larger set of apiary locations, ideally distributed in a way that limits location bias by ensuring a spatially balanced sample whereby the sample frame (i.e., where and under what conditions is the sample to be taken) is known to be representative or assumed to be; and B) harmonization and standardization of data collection.

1 Introduction

There is serious concern about the widespread decline of pollinators in the agricultural landscape (Potts et al., 2010; Vanbergen and The Insect Pollinators Initiative, 2013). Agricultural intensification leading to increased stress from pesticides and lack of floral resources, is probably a major cause (Goulson et al., 2015). Honeybees exploit mass-flowering crops and can cover long distances to meet their energy requirements. By doing so, honeybees are exposed to contaminants present in the environment and honeybees involuntarily collect these contaminants during foraging. It should be noted that the contaminants collected are not restricted to pollutants in nectar, pollen, water and propolis taken from the soil by the plants themselves, but also include pollutants deposited on the flowers, leaves, buds, and surface water by airborne processes, soil dusting, dust deposition and by human activities such as the application of plant protection products. As a result, the honeybee colony reflects the environmental pollution status of the foraging area (Bromenshenk et al., 1985; Conti and Botrè, 2001; Chauzat et al., 2006; Van der Steen, 2016).

The call on the "*Environmental monitoring of pesticides use through honey bees*" from the European Commission precisely focussed on this principle to facilitate performing a health check of the environment. Within this call, the INSIGNIA project was aiming to develop a best practice protocol for Apicultural citizen science and to conduct "*A pilot study on the best practices for a European wide monitoring program with honey bee colonies in an Apicultural citizen science (CS) setting to study pesticide exposure of honey bees and investigation of pollen sources*". One of the ambitious goals within the project was to predict the potential exposure of honeybees to contaminants in the landscape based on land-use. However, estimating the exposure of honeybees to pesticides on a landscape scale requires the linking of collected pesticide data with land-use characteristics. Similar links are required to estimate the diversity of the foraging landscape.

In order to visualize the occurrence of pollen and pesticides in the landscape, pollen and pesticides entering the hive that were sampled during the INSIGNIA monitoring campaign were used to construct a 'pollen and pesticide classification model'.

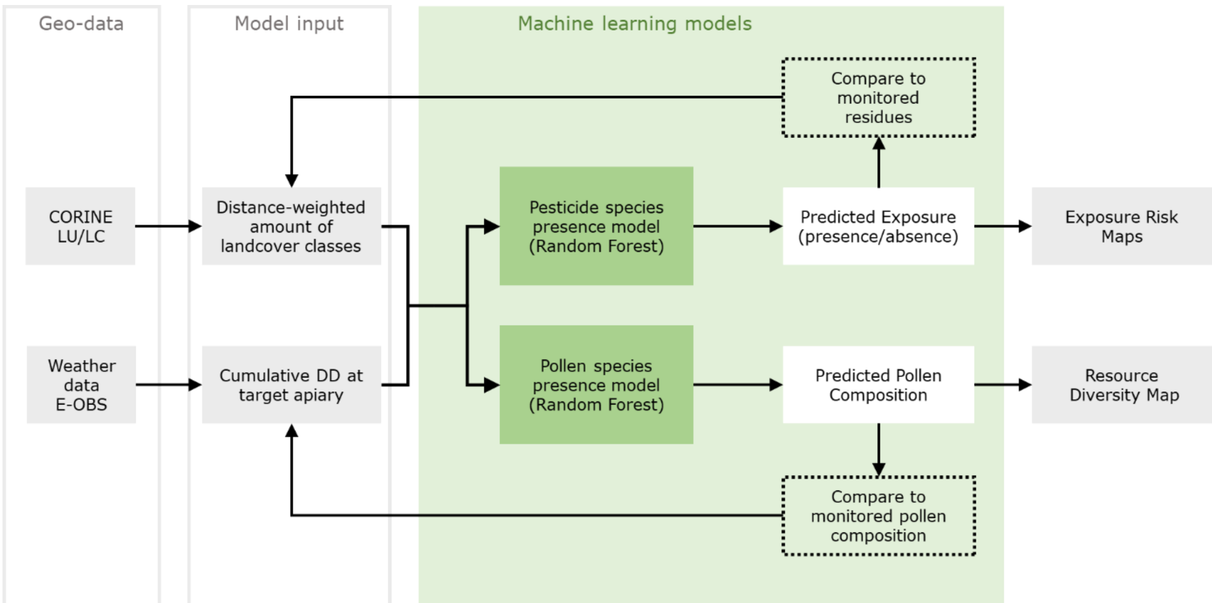


Figure 1. Overview of the machine-learning based modelling of the pesticide and pollen species expected in pollen.

To construct the model, two main types of input geo-data need to be collected and prepared (see Figure 1): monitoring data for pollen and pesticides collected in the INSIGNIA study (green boxes); and two sources of spatial and temporal geodata, i.e. the 'CORINE Land Use and Land Cover' data and the 'pan-European weather' data set. They form the heart of the statistical models, the technical details of which are described in RAPPORT CORINE/GUIDELINE and briefly described here. Before describing these models, first the underlying geo-data sets are discussed as well as the so-called kernel model used to extract relevant landscape information on land-use and weather data for the hive locations. For full details on the underlying data and kernel model, please see RAPPORT CORINE/GUIDELINE.

2 Methods

2.1 Spatial and temporal geodata

2.1.1 CORINE Land Use and Land Cover data

The geo-data used here describe the landscape around the locations of hives and it is used to form a link between the physical landscape and the presence of pollen and pesticides. The geo-data set used is the CORINE Land Cover (CLC) dataset, which is based on satellite imagery. It was first developed in the 1980s and contains information on the land use and land cover over the whole of Europe at a relatively coarse 100m resolution. In other words, the whole of Europe is divided into 1 hectare grids, and each grid cell is placed into a category describing the land-use. In total 44 different land-use classes are recognized. They are sub-divided into five main categories: artificial surfaces (e.g., roads, built up areas), agricultural areas (e.g., pastures, arable fields), forests and semi-natural areas (e.g., rough pastures), wetlands (e.g., bogs, marshes), and waterbodies (e.g., rivers and lakes). Obviously, there is no need to use the entire data set for the whole of Europe for each hive. Therefore, in order to extract the relevant information from the CORINE database (i.e., within the vicinity of a hive) so-called clipped maps are used that contain information on the type of CORINE CLCs in the local area (see box "Geo-data" in Figure 1). In the INSIGNIA study, two spatial scales are used to extract CLC data (see box "Model input" in Figure 1). Firstly, bee foraging is assumed to be limited to the landscape within a 3.25 km radius of the hive. With this distance in mind, for pollen resources a hive location is defined by a clip of 65 by 65 cells with the hive coordinates in the center. In other words, the hive is at the center of a 'patch' with an area of 6.5 by 6.5km. Secondly, the foraging landscape that is used by bees can be influenced by pesticides coming from within this area but also from outside this area by, for example, drift deposition. Therefore, for the landscape that determines the potential sources of exposure to pesticides, a larger clip area is used, i.e., 513 by 513 cells which equates to an area of 51.3 by 51.3km.

2.1.2 Pan-European weather data

The phenology (growth cycle) of plant species determines when that species starts flowering and thus when pollen are available. It is, however, impossible to obtain information on plant phenology for individual plant species from geo-data directly. Yet, the flowering period for plants is largely determined by season and thus strongly driven by temperature across Europe. Therefore, the second source of input geo-data required by the model is on temperature patterns (see Weather data E-OBS in the box "Geo-data" in Figure 1). The number of cumulative degree days (CDD10) above a threshold value (10 °C in this report) is generally considered to be an important species-specific requirement. See Figure 2 for an example of accumulated CDD10s for INSIGNIA apiaries in 2019.

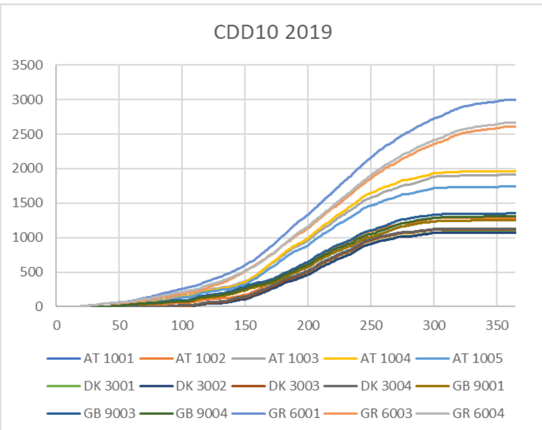


Figure 2. Cumulative degree days with base temperature 10 °C for the INSIGNIA apiaries in 2019

For any location in Europe, and for a large number of years (from 1950 onwards), time-series of daily temperatures were obtained from the daily gridded observational dataset for precipitation, temperature and sea level pressure in Europe (E-OBS), provided by the Copernicus Institute, on a 0.1 degree regular grid (https://www.ecad.eu/download/ensembles/download.php. For the calculations, version 21e was used. An example of the contents of such an E-OBS dataset, for 31 December 2019, is shown with the apiaries in Denmark, Austria, UK and Greece, in Figure 3. With the temperature data pre-processed, the required input geo-data are complete and can be used as the first set of input data (see Figure 1).

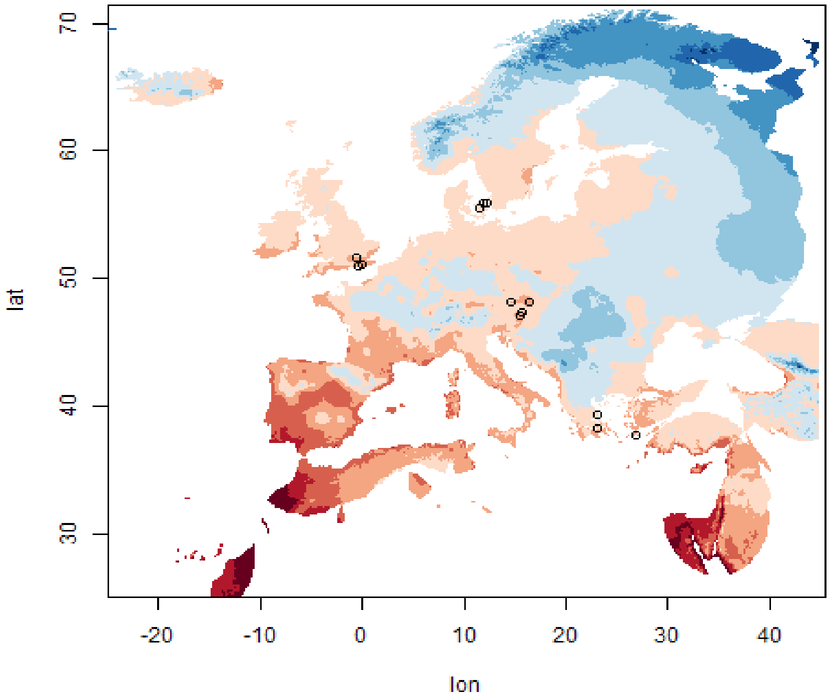


Figure 3. Location of the apiaries plotted on the map with temperatures on 31 December 2019. Indicative colours range from warm (dark red) to cold (dark blue).

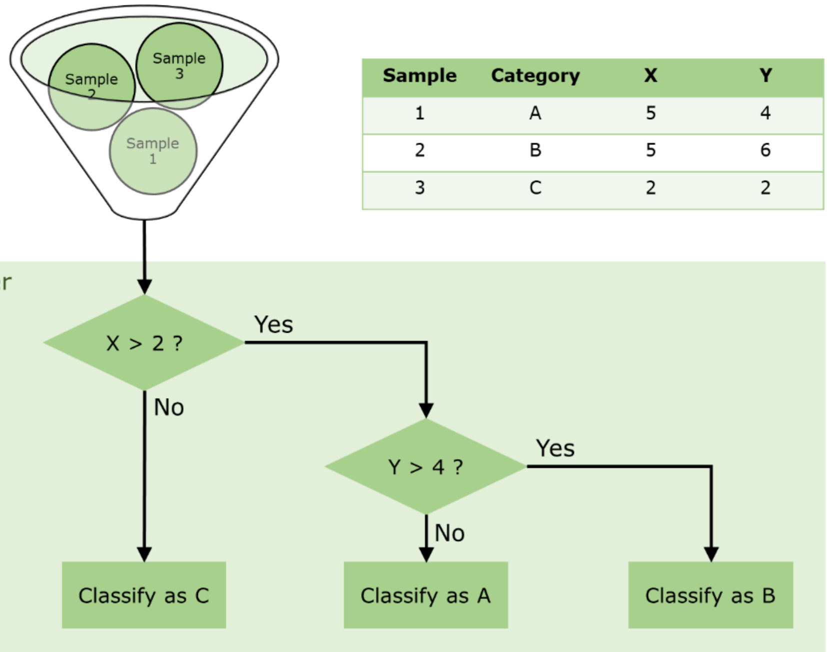
2.2 Kernel model

For a given hive location, the processed CORINE CLC data represents the landscape around the hive. It is assumed that within this local landscape, areas closer to the hive will have a stronger influence on pollen and pesticides than areas that are further away. To account for this, a so-called kernel model is used to analyze the landscape around a hive location. The model calculates the impact for each grid cell around a hive, where grid cells closer to the hive "weigh heavier" than grid cells further away. This way, the further away a cell is, the smaller its influence will be. At this stage the simplifying assumption is made that the influence of a land cover class at the hive location is representative of the experienced foraging landscape used by bees belonging to the hive in the center. In other words, bees are assumed to use all of their surroundings at the same distance equally. Next, for each location all calculated values are summed for each of the CORINE CLCs which gives a total "impact" of a certain land-use category at the location of the hive.

2.3 Pollen and pesticide classification model to create predictive maps

2.3.1 Random forests for classification

The machine learning approach used here is called a random forest. Like real forests, a random forest is made up of individual trees, or so-called classification trees. And like a real forest, each classification tree is slightly different. Figure 4 shows a simplified classification tree with so-called splitting variables X and Y. The tree is a basic classifier that places observations (from pollen and APIStrip samples in this case) in an outcome category based on a set of splitting decisions (Figure 4). By combining the results of all trees in the random forest model, an ensemble (combined) output is created. This makes it a powerful and suitable model for the INSIGNIA data. Namely, without knowing or assuming anything about the precise underlying links between the monitored pollen and pesticides and the surrounding landscape from which they come, the random forest model can be used as a predictive model that improves as more data are added (see boxes around Exposure and Resources in Figure 1). Below follows more background information on random forests in the context of the INSIGNIA pollen and pesticide data analysis; for the random forest models the monitoring data needs to be pre-processed and this is described in the section "Fitting models to INSIGNIA pollen and pesticide data".



Sample	Category	X	Y
1	A	5	4
2	B	5	6
3	C	2	2

Figure 4. Example of a simple classification tree. Samples are associated with explanatory splitting variables X and Y, at each split in the tree samples are divided into groups based on a splitting variable and subsequently placed in a category.

To predict the probability of a pollen family or a pesticide being present in a location, random forest classification models were fitted. For each tree in the forest the training data (i.e., the collected INSIGNIA data) were bootstrapped (sampled), sampling 50% (in-bag fraction) of all data for a country to grow each tree in the forest. This approach limits between tree correlation and so improves classification performance of the random forest. The remaining 50% of data (out-of-bag fraction) are used to test the predictive performance. To further reduce the effect of between tree correlation, at each split in a tree a best splitting variable is selected from a randomly selected subset of the predictor variables (Breiman, 2001). Each time an observation in the training data is passed through the classifier, it is placed into a certain outcome category (predicted outcome); each tree can lead to a different outcome for a sample. Next, from the ensemble of classifiers (i.e., the collection of all trees) a probability estimate is given for the membership of a sample to a certain category based on the proportion of all the trees that predict that category for the sample.

2.3.2 Fitting models to INSIGNIA pollen and pesticide data

Pollen and APIStrip samples were split to individual observations. Distance weighted CORINE CLC data and CDD10 data (see Figure 1) were collected and linked to individual observations, so that each observation is associated with the landscape around the respective hives, and cumulative temperature and month information for pollen and APIStrip samples, respectively. Using the kernel model, CORINE CLC impact was calculated for each apiary (see section "Kernel model"). Note that no assumptions on the likely sources of pollen and/or pesticides are made with respect to the different landcover classes, and therefore here all 44 CORINE CLCs are used. CDD10 information was extracted for the observation's location and date (see section "Pan-European weather data") for pollen and month information was assigned as temporal information for pesticides.

Owing to the fact that most pollen and APIStrip samples had more than one pollen family or pesticide present, the classifier for individual samples is not very accurate (i.e., the out-of-bag error is large). Instead, using the probability estimates of class membership for each sample and applying a cut-off value, we obtain a most likely set of pollen or pesticides in the same sample. If the probability of class membership is equal or larger than the cut-off value, it is accepted in the set of most likely pollen or pesticides. Lower cut-off values thus result in a less strict classifier. Separate models were created for each country, for both pollen and pesticides.

From all random forest models created, the performance was checked against 30 random subsets of 20% of the available pollen or pesticide data. For each sample, all classes with a probability above the cut-off value were accepted as likely to be present and were stored. Each predicted sample set was compared against the observed pollen or pesticides in the sample (see Figure 1). Next, two ratios were calculated: 1) the ratio of accepted classes present in the sample versus actual classes in the sample (true positives); 2) the ratio of accepted classes *not* present in the sample versus the actual classes in the sample (false positives).

Mean true positive and false positive ratios (n = 30) were computed and used to select a cut-off value that gives a good trade-off between the true positives and the false positives. Note here that the implication of having false positives is not the same for the pollen and pesticide models. For the pollen, a false positive would falsely suggest a more diverse foraging landscape. For the pesticides, a false positive would falsely suggest a landscape with a higher presence of pesticides. However, from an environmental risk assessment point of view, which aims to be conservative, this means a false positive for pollen is less conservative, whereas a false positive for pesticides would be more conservative. Hence, for the pollen model, we may accept a slightly lower ratio of true positives if this results in a substantially lower ratio of false positives. Conversely, for the pesticide model, we may accept a slightly higher ratio of false positives if this comes with a substantially higher ratio of true positives.

2.3.3 Country-scale prediction of pollen and pesticides to produce maps

The objective of the final model objects is to make pesticide predictions for five months of the year (i.e. May – September) for the whole country. For pollen, the same was done but using CDD10 values, calculated for the middle of the month for each date/location combination. For this the required data (see Figure 1) was collected based on a 10km spacing of points for which CORINE CLC kernel impacts were calculated as well the cumulative temperature (CDD10).

For every 10km grid cell the model predicts the pollen families and pesticide types present. For simplicity and visualization the results are reported as the number of predicted pollen and pesticides, though for each prediction the actual accepted pollen families and pesticides can be extracted as well. Bear in mind that with the relatively coarse scale of 10km grid cells some cells have their centers in a water body and these are excluded (e.g., coastal areas); and that secondly, the predictions are based on a limited number of spatial and temporal samples, which are extrapolated to the whole country. This means that in practice for any location the real pollen and pesticide numbers might be higher or lower than what the models predict.

In 2019 data were collected from Austria, Denmark, Greece, and Great Britain. Following clean-up and preparation of the raw data only the data for Austria and Denmark could be used in further modelling. Hence, for 2019 the predictions were only made for Austria and Denmark. For the data collected in 2020, which saw an increase in the number of participating countries, predictions were made for Austria, Belgium, Denmark, France, Greece, Ireland, Italy, Latvia, and the United Kingdom.

3 Results and discussion

When interpreting the results below, it is important to keep in mind that pollen and pesticide diversity is largely determined by the bee's behaviour and the placement of apiaries. Practically, this means that if, based on the collected INSIGNIA pollen data, the model predicts a low diversity of pollen this does not necessarily mean that real pollen diversity in the landscape is also low. It could reflect the simple fact that at that time, bees were foraging on a limited number plant species. For the pesticide data, if the model predicts a low number of pesticides this can result from either a real low presence or it can stem from a situation where bees simply don't forage on plants that are treated crops or their foraging sites are not near treated crops at that time. Extrapolation to a whole country using a limited set of apiary locations could also lead to over- or underestimation of the real number of pollen and pesticides.

3.1 2019 Pollen and pesticide predictions

3.1.1 Austria

Pollen diversity in Austria decreases from May to September, from the North-East towards South-West (Figure 3a-e). Over the year pollen diversity is lowest in the North and North-East of the country while in the East, West, and South-West pollen diversity remains high throughout the year. The number of pesticides is consistently low in the East of Austria (Figure 3f-j). Higher pesticide numbers in June-August appear to coincide with high pollen diversity in May and June (Figure 3a-b and 3f-g), but from July onwards patter is seemingly reversed and higher pesticide numbers appear to be in areas with a lower pollen diversity (Figure 5c-e and 5h-j).

3.1.2 Denmark

Pollen diversity in Denmark is predicted to follow a relatively uniform seasonal pattern with an increase between May and July, and a decrease towards September (Figure 6a-e). Areas to the West and North are generally predicted to be less diverse than the East, with Jutland (the largest island) showing a strongly uniform distribution in the predicted number of pollen families in May, August, and September. The use of pesticides is at a higher level early in the year, but increases towards the West from May to July (Figure 6f-h). The number of pesticides shows more spatial variability in August and September (Figure 6i-j).

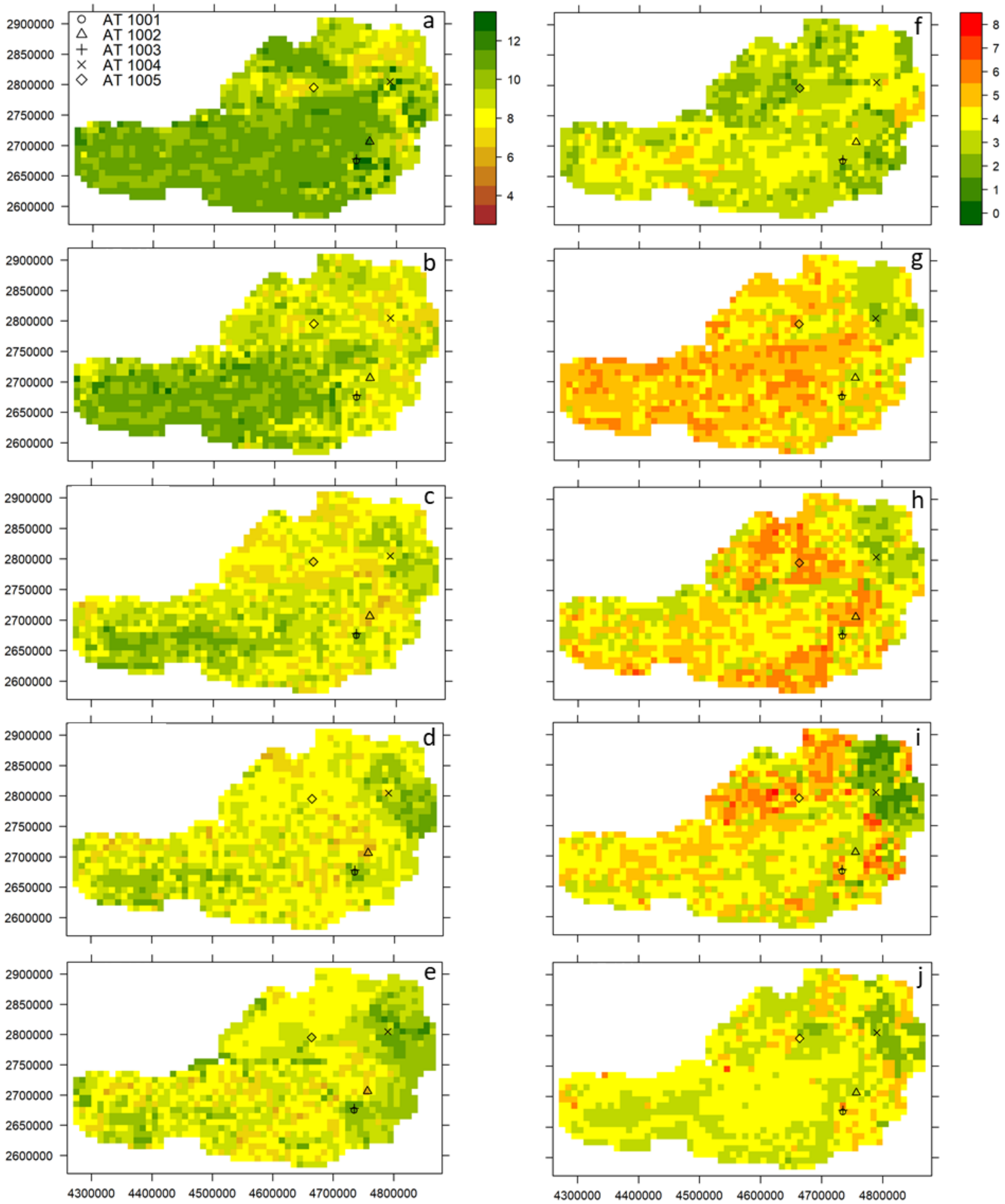


Figure 5: Predicted number of pollen (a-e) and pesticides (f-j) in Austria in May (a and f), June (b and g), July (c and h), August (d and i), and September (e and j). Note that pollen and pesticides have different colour scales. Symbols indicate apiary locations (see legend in a for site IDs).

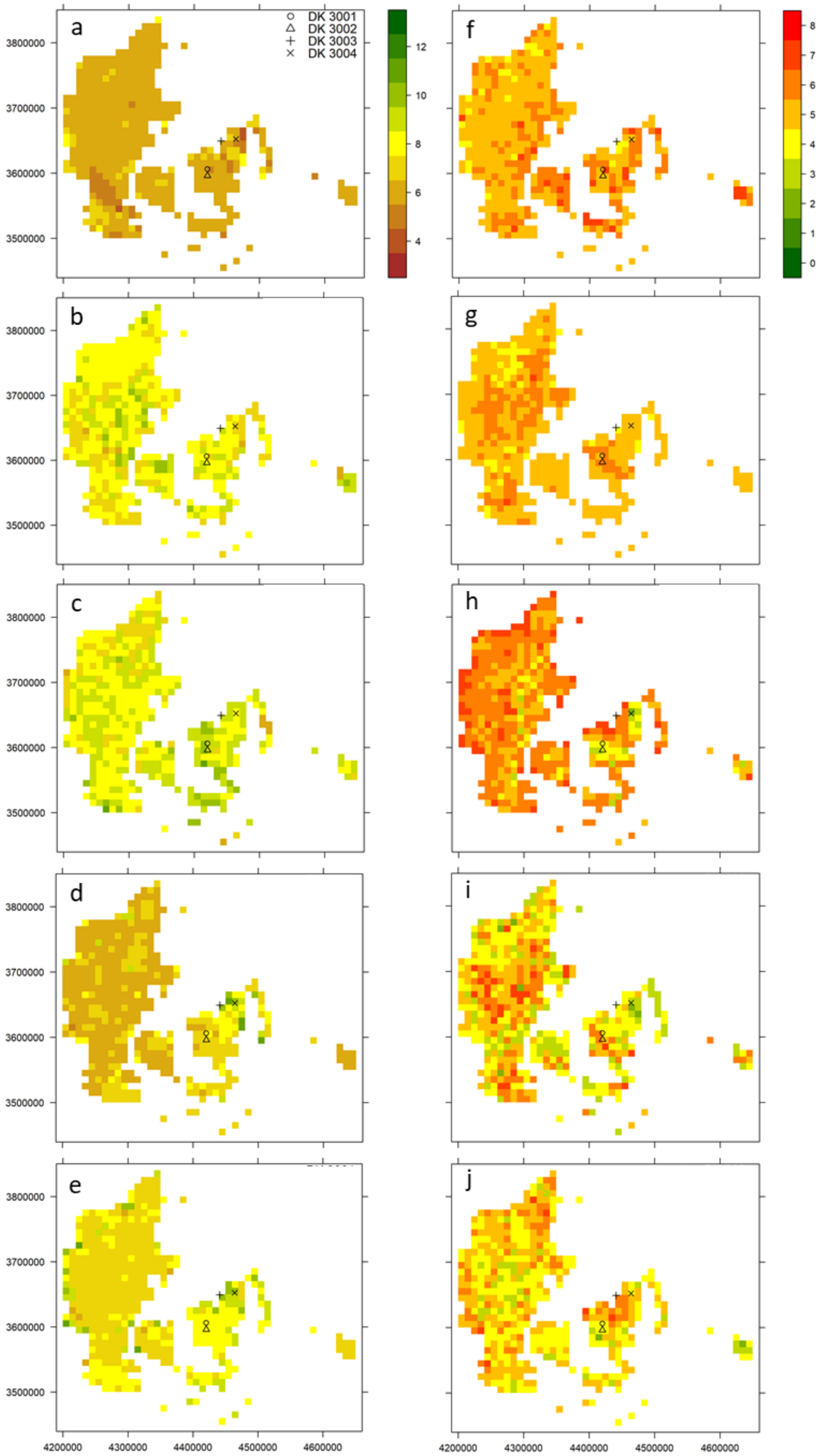


Figure 6: Predicted number of pollen (a-e) and pesticides (f-j) in Denmark in May (a and f), June (b and g), July (c and h), August (d and i), and September (e and j). Note that pollen and pesticides have different colour scales. Symbols indicate apiary locations (see legend in a for site IDs).

3.2 2020 Pollen and pesticide predictions

The inclusion of additional countries in 2020 allows for the generation of Europe scale maps, rather than side-by-side views per country. Full size maps per country are available in the Appendix.

3.2.1 Pollen

In both Austria and Denmark the spatial patterns over the months are similar to patterns predicted in 2019. Owing to the increased number of countries, what is more evident in 2020 is that the higher latitude countries like Denmark and Latvia show a reduced pollen diversity compared to for example Austria, Belgium (Figure 7). Italy and Greece are spatially variable, with a decreasing pollen diversity as the season progresses (Figure 7a-e). This is possibly due to the fact that the major flowering period for (mass) flowering crops is earlier in the year in the warmer Mediterranean climates of Greece and Italy. The picture for France appears to be a mixed mosaic with no major seasonal differences. A possible reason these spatial changes are not predicted for France (which has similar climatic conditions to Greece and Italy in the south, and similar to Belgium in the North) is that the apiary locations in France are clustered in the centre and thus may not capture the same flowering dynamics (see Figure 7). In France the months June and July seem to be the months with the lowest pollen diversity (i.e., more brown grid cells), this could be explained by the mass flowering of plants in the *Asteraceae* (e.g., dandelions), *Fabaceae* (e.g., peas), and *Poaceae* (e.g., maize) families that make up relatively large proportions of the bee's diet (Figure 8) and thus could reflect a less diverse diet of bees during that period. The apparent "browning" of Greece and Italy in August and September could have a similar reason, although different to France in Greece the major pollen family in August and September is from the *Asteraceae* (Figure 8e) and in Italy the major families are *Araliaceae* (e.g., hedera) and *Arecaceae* (i.e., Mediterranean dwarf palm). For the UK, the model predicts that further north (i.e., towards Scotland) the spatial diversity decreases. This could be a consequence of a less diverse landscape. However, it is also possible that the limited spatial coverage of apiaries in the UK, which are centred in the South of the UK, don't cover CORINE CLC landscapes that are commonly found in Scotland. As for Ireland, compared to other months the lower predicted diversity in July stands out (Figure 7b vs 7a,c-e). In Ireland pollen from the *Rosaceae* family make up over a quarter of the returned pollen, which could explain a lower predicted diversity as the diet of bees is less diverse (Figure 8f).

3.2.2 Pesticides

For pesticides, the predicted number and spatial patterns in 2020 are again similar to those predicted in 2019 (Figure 5,6 vs Figure 8). In Latvia the highest numbers of pesticides are predicted towards the end of the season (Figure 8e). In Italy the number of predicted pesticides varies over the year. For example, in August higher numbers are predicted along the Apennines whereas in September higher numbers are predicted along the coastal areas and on the Po plane (Figure 9d vs Figure 9e). There is an obvious difference in the predicted number of pesticides between Belgium, France, Greece, Ireland, and the UK on the one hand and Austria, Denmark, Italy, and Latvia on the other (Figure 9). Greece and Denmark were both present in 2019 and 2020. Considering the collected data for Greece, in 2019 45 pesticides were found (after clean-up and removal of beekeeper substances, Figure 10a) compared to only three substances in 2020 (Figure 10b). Yet, in 2019 only two apiaries with three hives had APIStrips whereas in 2020 there were nine apiaries with two hives that used APIStrips. This could be a real difference in pesticide use between the two years, yet it could also stem from different placement of apiaries, from different feeding behaviour of bees, or from differences in chemical analysis. Ultimately, it is unclear what the root cause of this difference is. Indeed, this example clearly demonstrates the importance of harmonised and standardised data collection protocols and highlights some of the challenges with the monitoring approach, after all: beekeepers have no control over the bee's behaviour and with many people involved there is an increased risk of sample handling error.

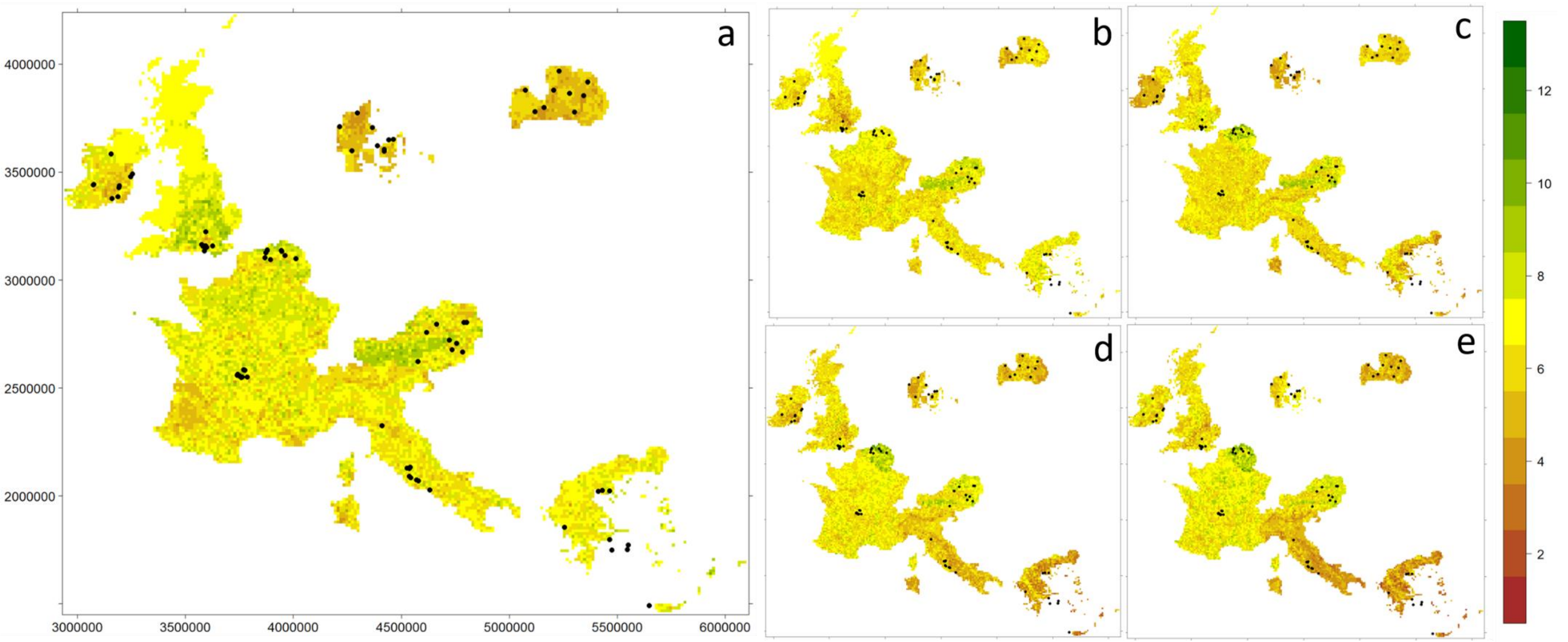


Figure 7: Predicted number of pollen in the 2020 INSIGNIA countries for the month May (a), June (b), July (c), August (d), and September (e). Black dots indicate apiary locations within a country.

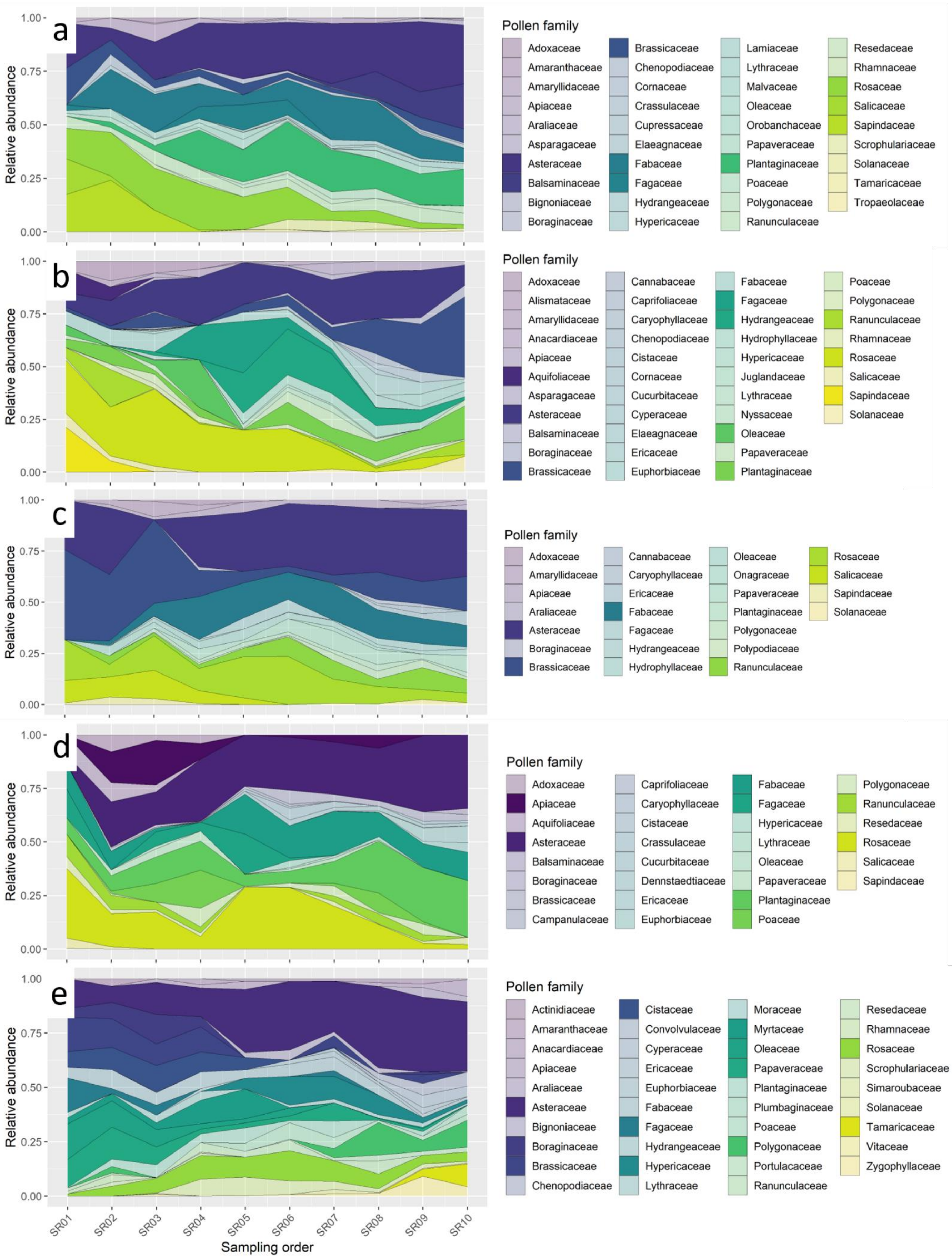


Figure 8: Summed mean relative abundance across apiaries per sampling time in Austria (a), Belgium (b), Denmark (c), France (d), and Greece (e). Plant families forming more than 10% of total abundance at any time in the year are highlighted; minor sources of pollen are transparent. SR1 – SR3 = May; SR4 – SR5 = June; SR6 – SR7 = July; SR8-SR9 = August; SR10 = September.

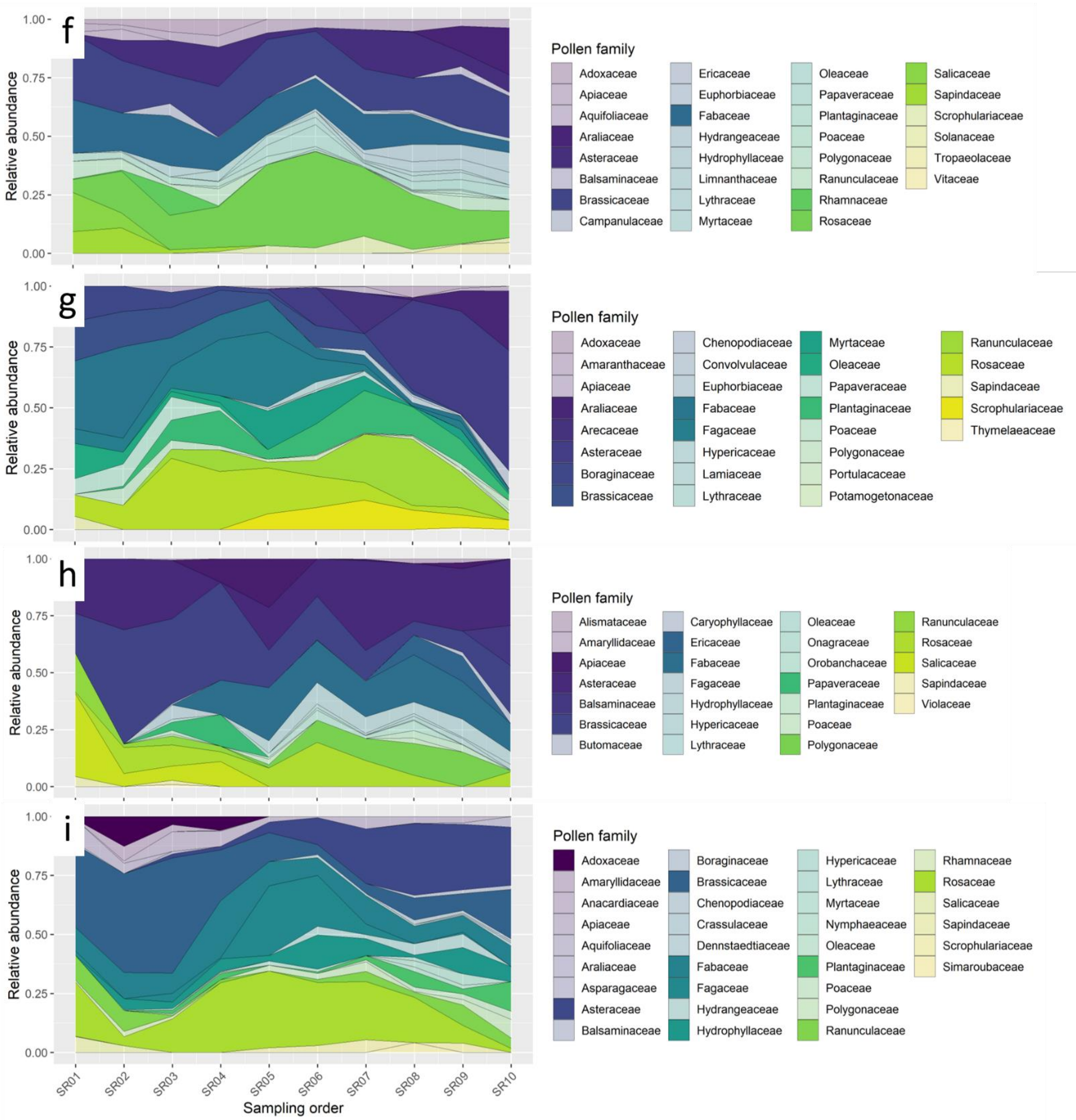


Figure 8 (continued): Ireland (f), Italy (g), Latvia (h), and United Kingdom (i).

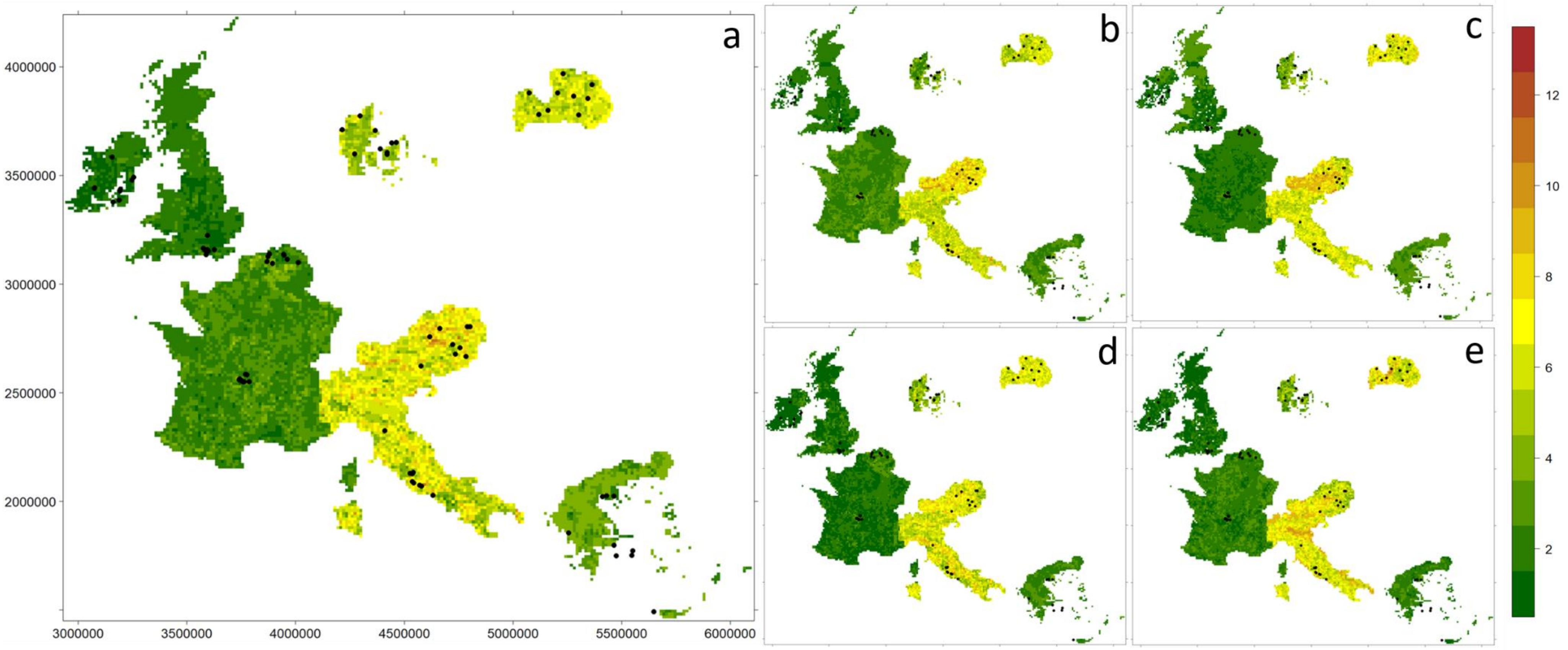


Figure 9: Predicted number of pesticides in the 2020 INSIGNIA countries for the month May (a), June (b), July (c), August (d), and September (e). Black dots indicate apiary locations within a country.

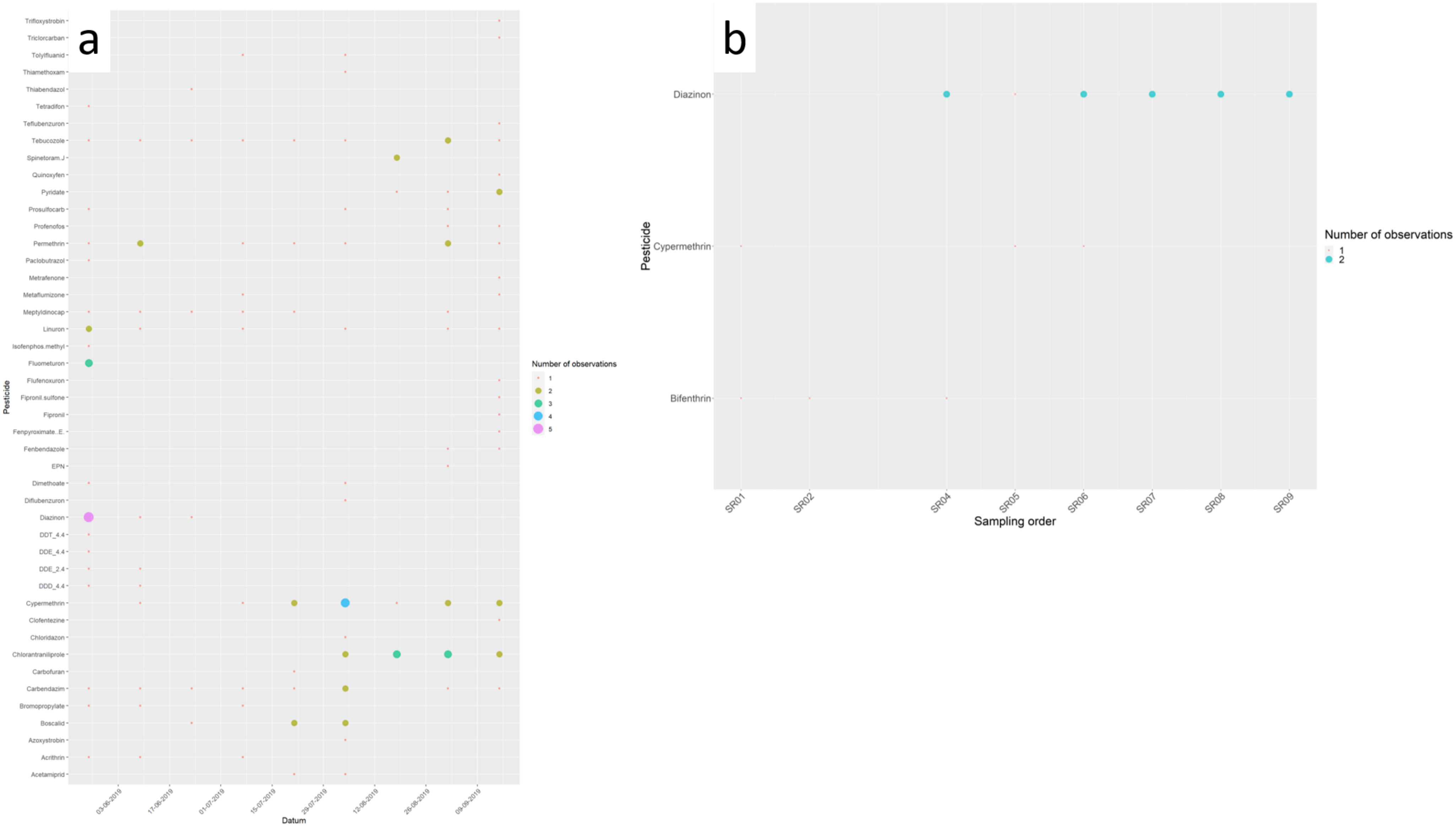


Figure 10: Number of observations for pesticides in Greece, in 2019 (a) and in 2020 (b). Colour and size of dots represent the total number of times a pesticide was found across all apiaries; this gives for 2019 (a) two apiaries with total of six hives; and for 2020 (b): nine apiaries with total of 18 hives.

4 Conclusions

- Using freely available software R (R Core Team, 2020), freely available CORINE CLC and E-OBS geo-data sets and data collected by citizen scientists, predictive models for pollen and pesticide numbers were constructed.
- The models can be used to predict the number and types of pollen and pesticides at large spatial scales for the months May-September (i.e., the main growing season across Europe).
- Model outputs can inform stakeholders on the expected availability of floral resources and on expected exposure to pesticides within a country and during the growing season. This information is valuable as it allows, for example, beekeepers to make informed decisions on the placement of their hives and make trade-offs between forage availability and exposure risk.
- Both models rely heavily on the input data. As such, high or low predicted numbers of either pollen or pesticides need to be considered in the context of the apiary locations that were used to collect the so-called model training data and also the quality of the collected data.
- With the relatively limited spatial coverage within the current pilot project, linking landscape characteristics and climate to the pollen/pesticide predictions is difficult. Nevertheless, large-scale patterns related to land-use, climate, and floral diversity may be inferred.
- Improvements of the model performance can be achieved through:
    - a larger set of apiary locations, ideally distributed in a way that limits location bias by ensuring a spatially balanced sample whereby the sample frame (i.e., where and under what conditions is the sample to be taken) is known to be representative or assumed to be so (Smith et al., 2017).
    - harmonization and standardization of data collection

5 Acknowledgements

We acknowledge the E-OBS dataset from the EU-FP6 project UERRA (http://www.uerra.eu) and the Copernicus Climate Change Service, and the data providers in the ECA&D project (https://www.ecad.eu). We also acknowledge the CORINE CLC dataset provided by the Copernicus Land Monitoring Service (https://land.copernicus.eu/pan-european/corine-land-cover).

6 References

Breiman L, 2001. Random Forests. *Machine Learning, 45, 5-32*.

Bromenshenk JJ, Carlson SR, Simpson JC and Thomas JM, 1985. Pollution Monitoring of Puget Sound with Honey Bees. *Science, 227, 632-634*.

Chauzat M-P, Faucon J-P, Martel A-C, Lachaize J, Cougoule N and Aubert M, 2006. A Survey of Pesticide Residues in Pollen Loads Collected by Honey Bees in France. *Journal of Economic Entomology, 99, 253-262*.

Conti ME and Botrè F, 2001. Honeybees and Their Products as Potential Bioindicators of Heavy Metals Contamination. *Environmental Monitoring and Assessment, 69, 267-282*.

Goulson D, Nicholls E, Botías C and Rotheray EL, 2015. Bee declines driven by combined stress from parasites, pesticides, and lack of flowers. *Science, 347, 1255957*.

Potts SG, Biesmeijer JC, Kremen C, Neumann P, Schweiger O and Kunin WE, 2010. Global pollinator declines: trends, impacts and drivers. *Trends in Ecology & Evolution, 25, 345-353*.

R Core Team, 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

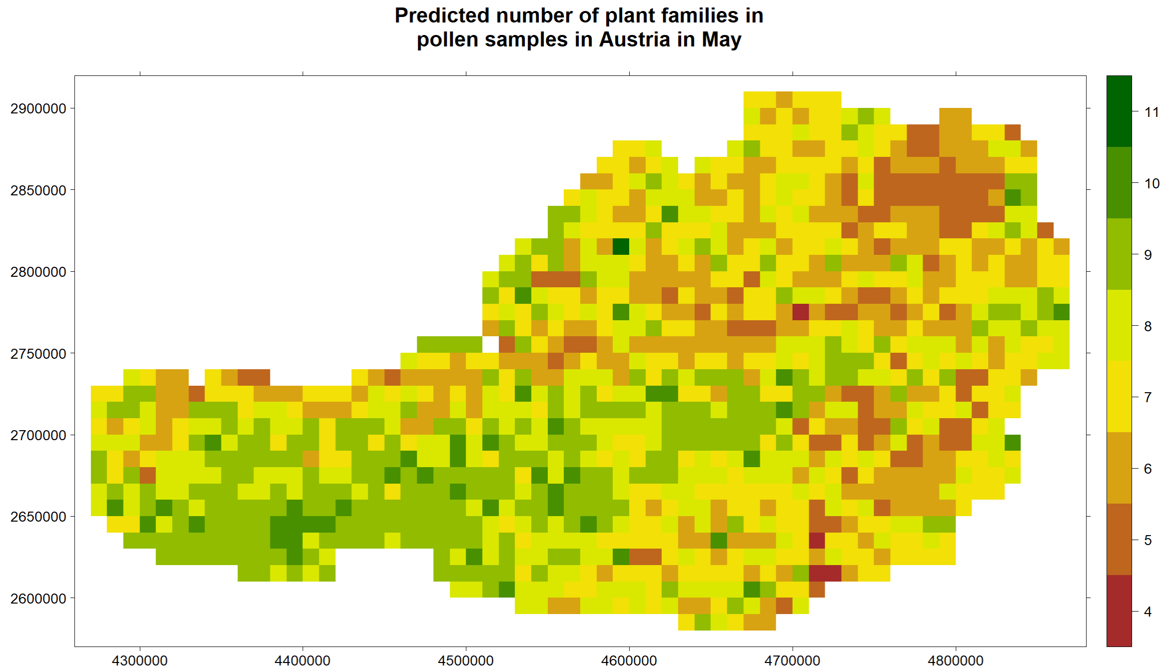
Smith ANH, Anderson MJ and Pawley MDM, 2017. Could ecologists be more random? Straightforward alternatives to haphazard spatial sampling. *Ecography, 40, 1251-1255*.

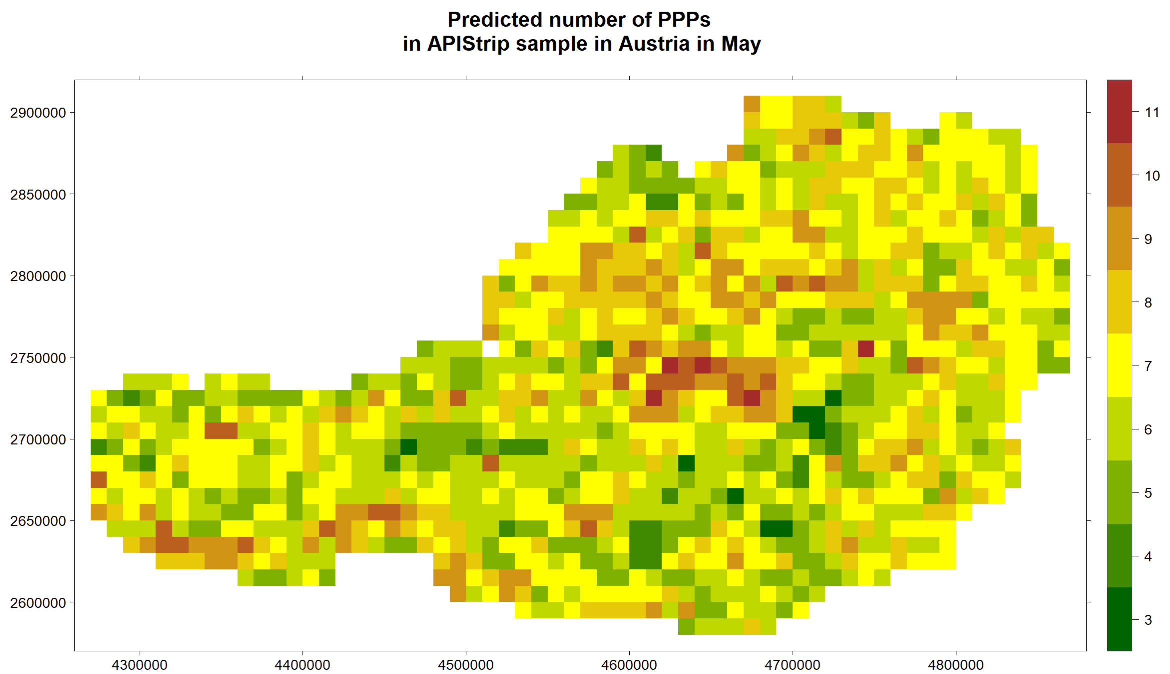
Van der Steen J, 2016. Beehold : the colony of the honeybee (Apis mellifera L) as a bio-sampler for pollutants and plant pathogens.

Vanbergen AJ and The Insect Pollinators Initiative, 2013. Threats to an ecosystem service: pressures on pollinators. *Frontiers in Ecology and the Environment, 11, 251-259*.

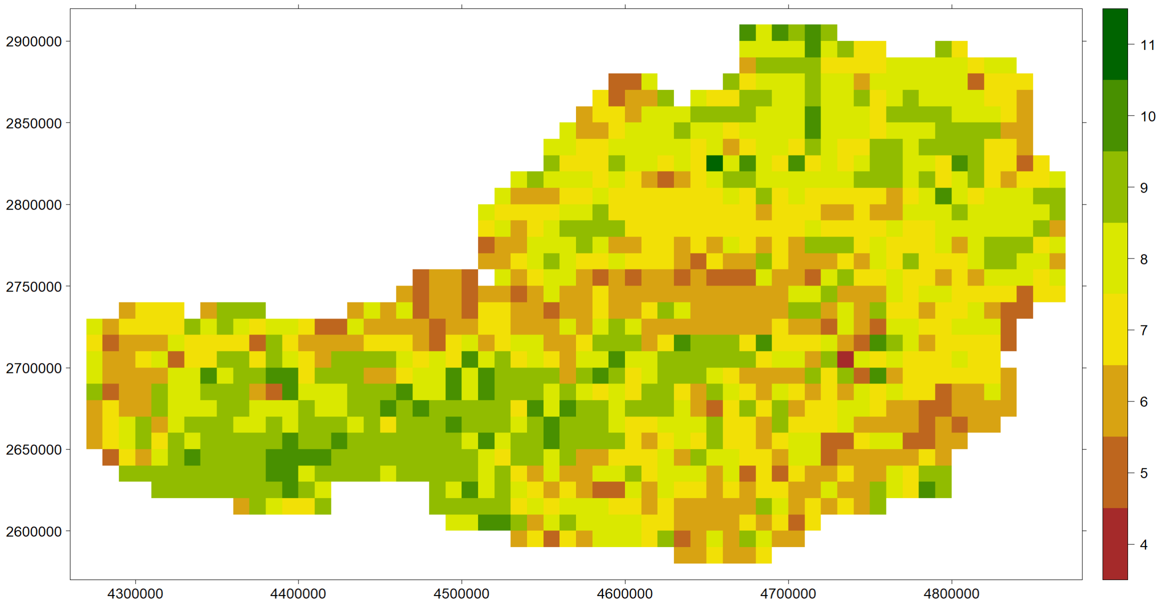
7 Appendix

Below are outputs of the pollen and pesticide classification models for individual countries. Results are presented per country as pairs and showing, for the same month, first the pollen prediction and then the pesticide prediction. Note that the colour scales are different between plots and between pollen and pesticides.

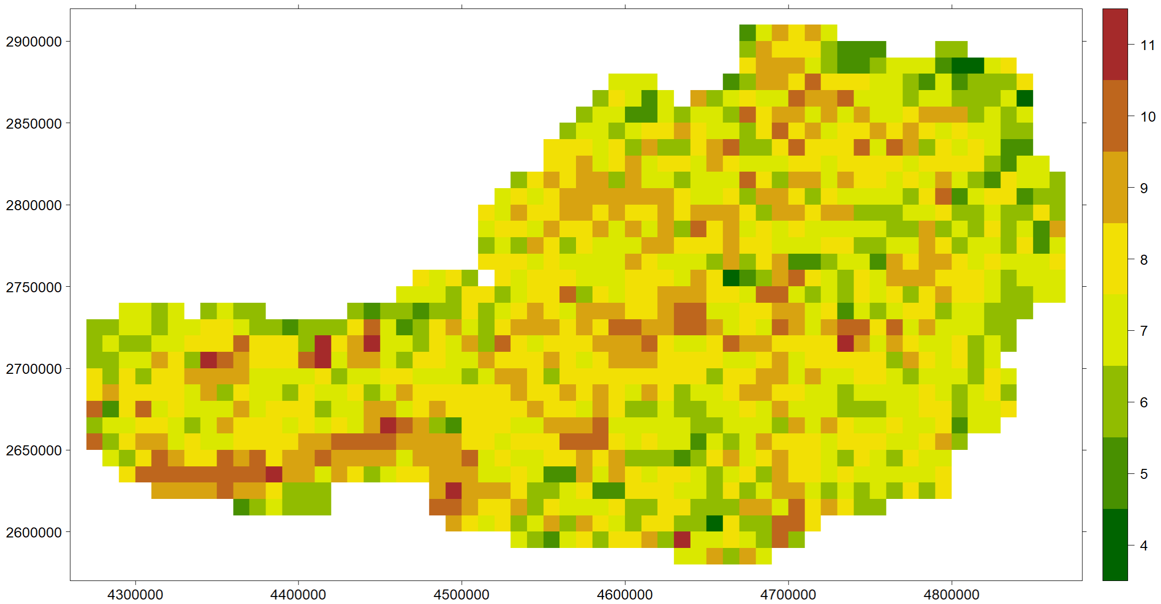




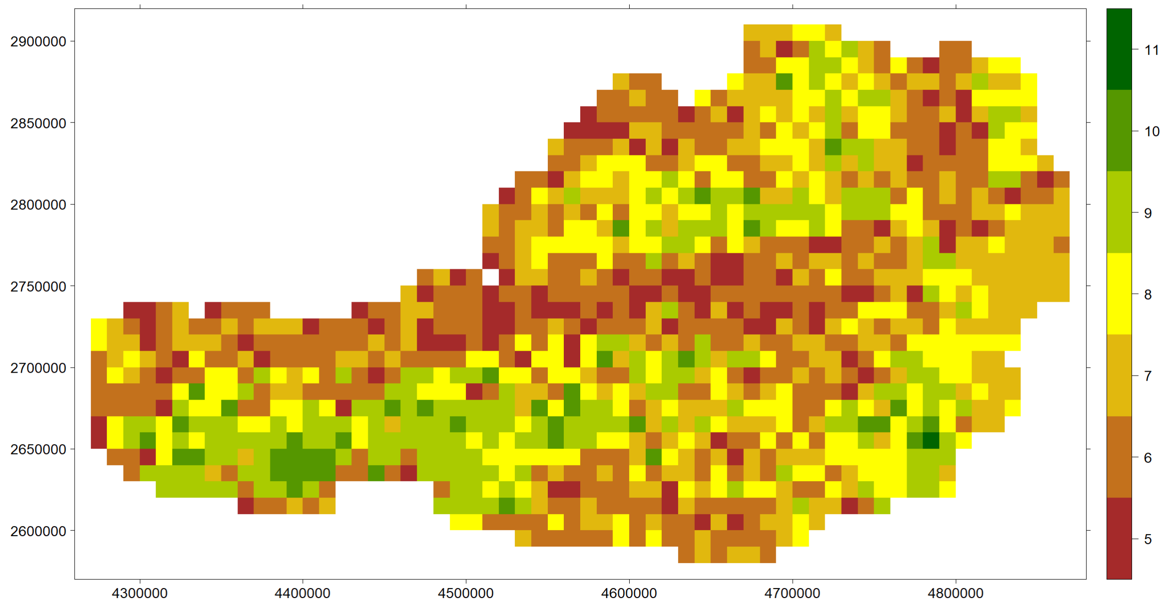
Predicted number of plant families in pollen samples in Austria in June



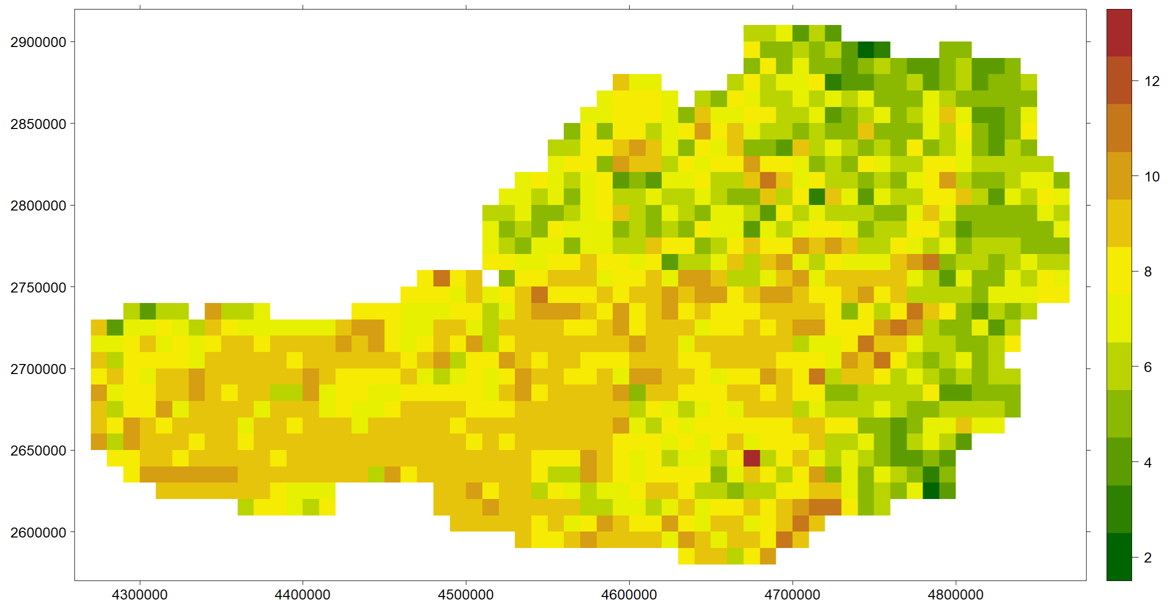
Predicted number of PPPs in APIStrip sample in Austria in June



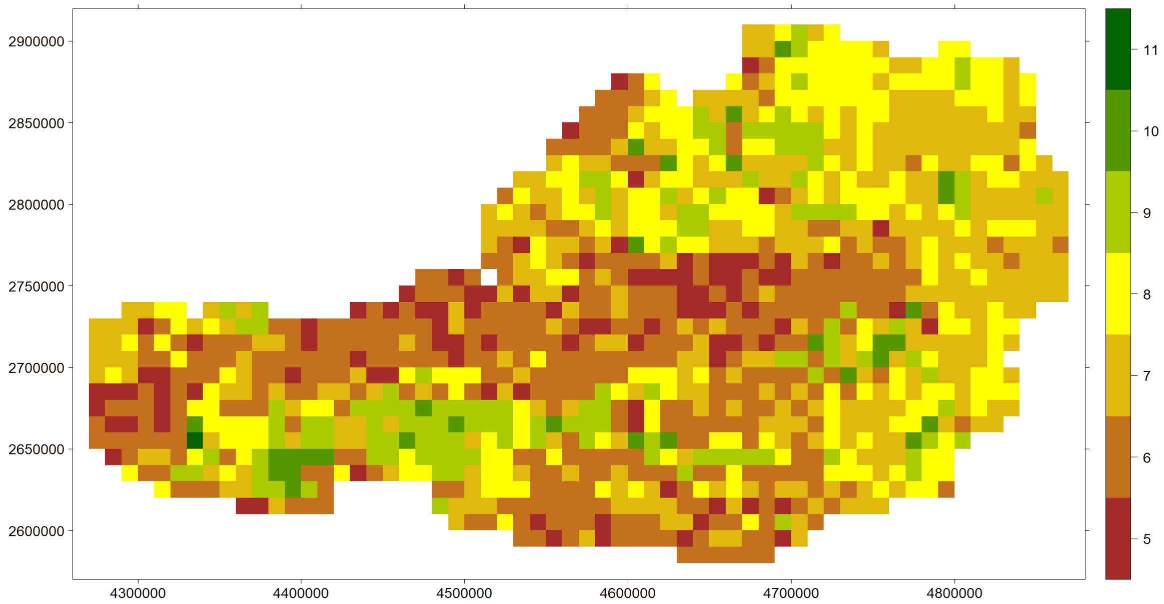
Predicted number of plant families in pollen samples in Austria in July



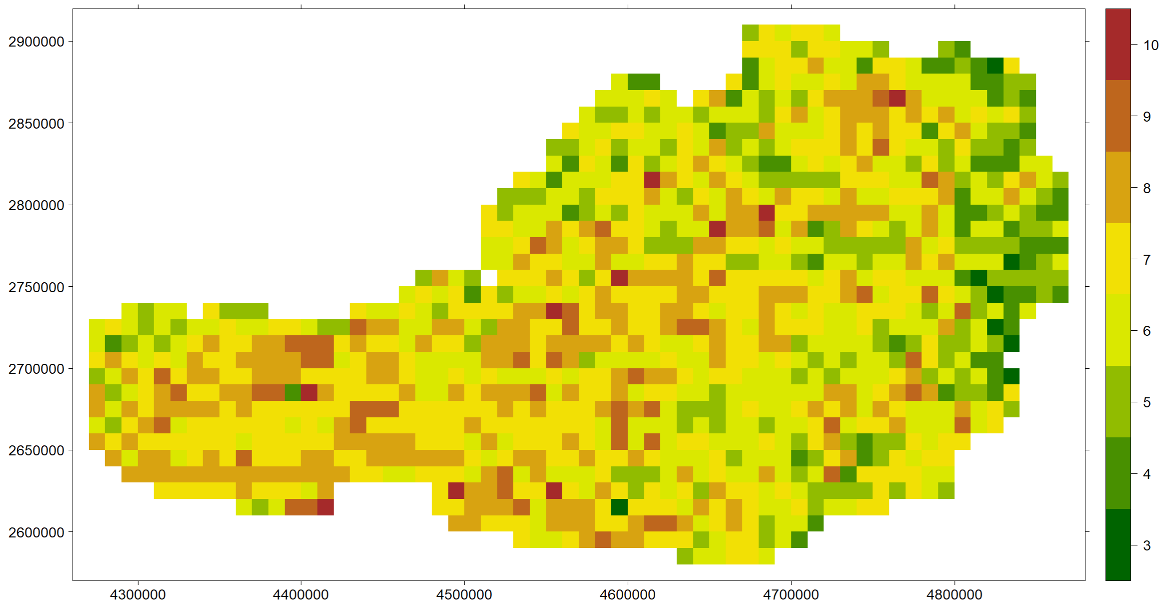
Predicted number of PPPs in APIStrip sample in Austria in July



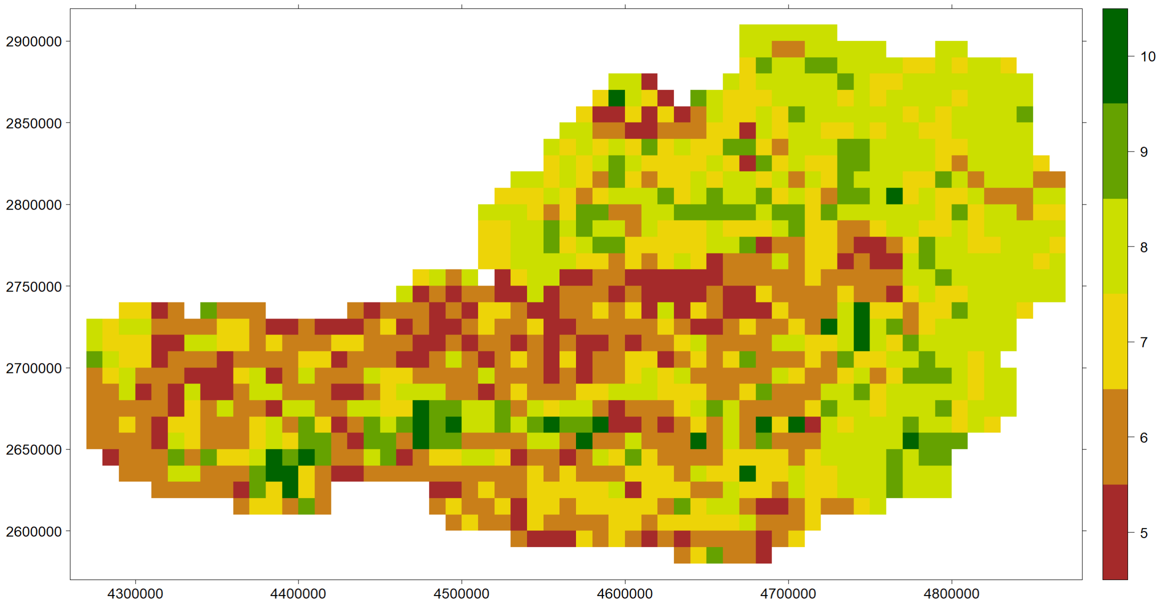
Predicted number of plant families in pollen samples in Austria in August



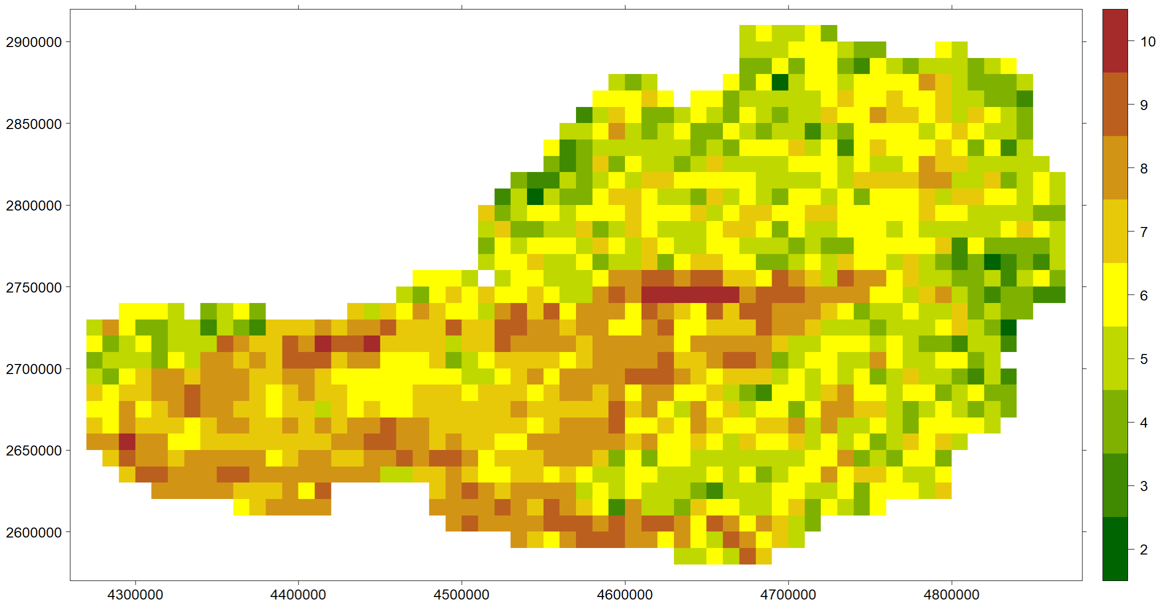
Predicted number of PPPs in APIStrip sample in Austria in August



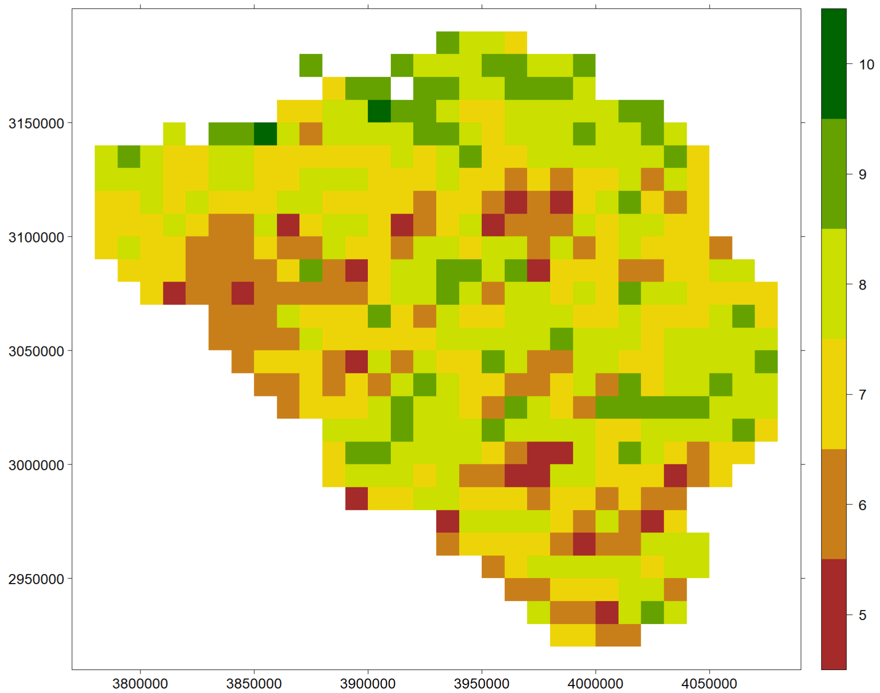
Predicted number of plant families in pollen samples in Austria in September



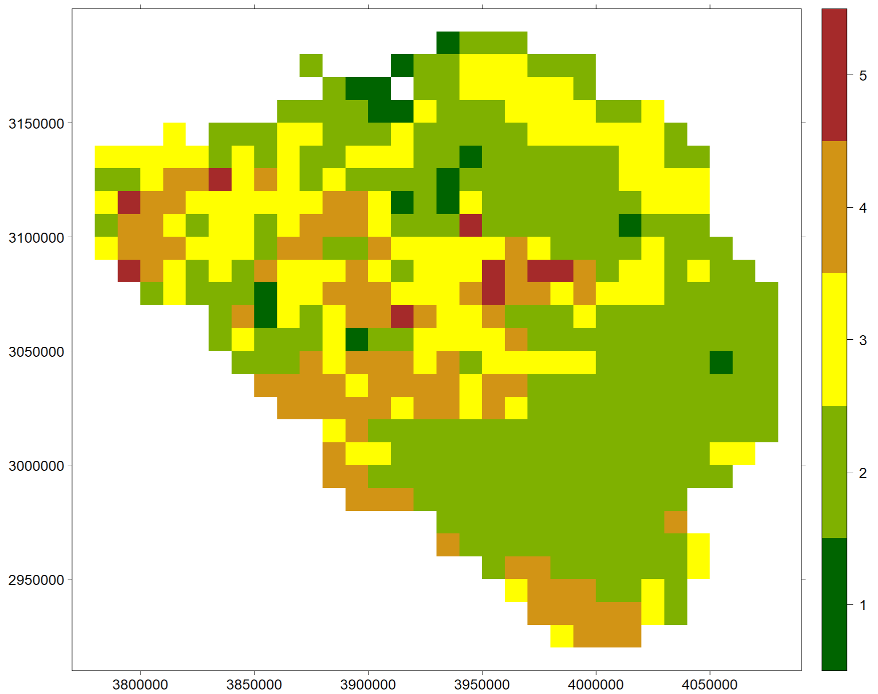
Predicted number of PPPs in APIStrip sample in Austria in September



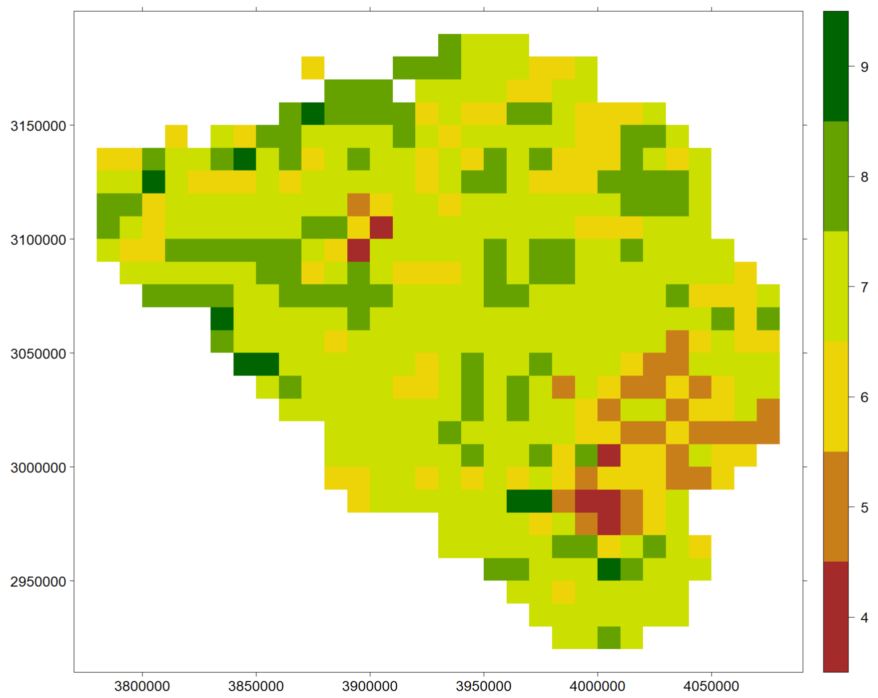
Predicted number of plant families in pollen samples in Belgium in May



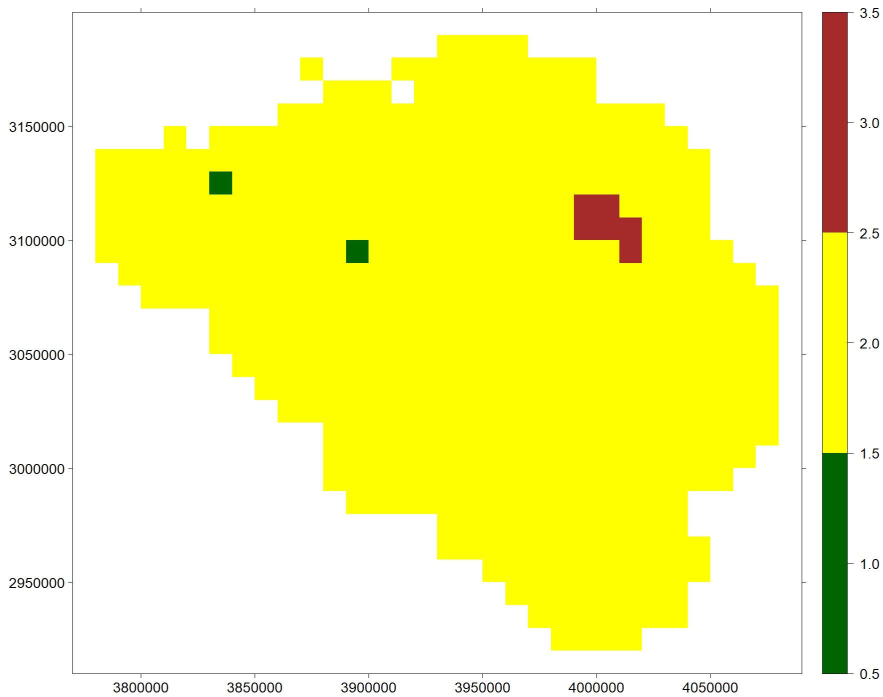
Predicted number of PPPs in APIStrip sample in Belgium in May



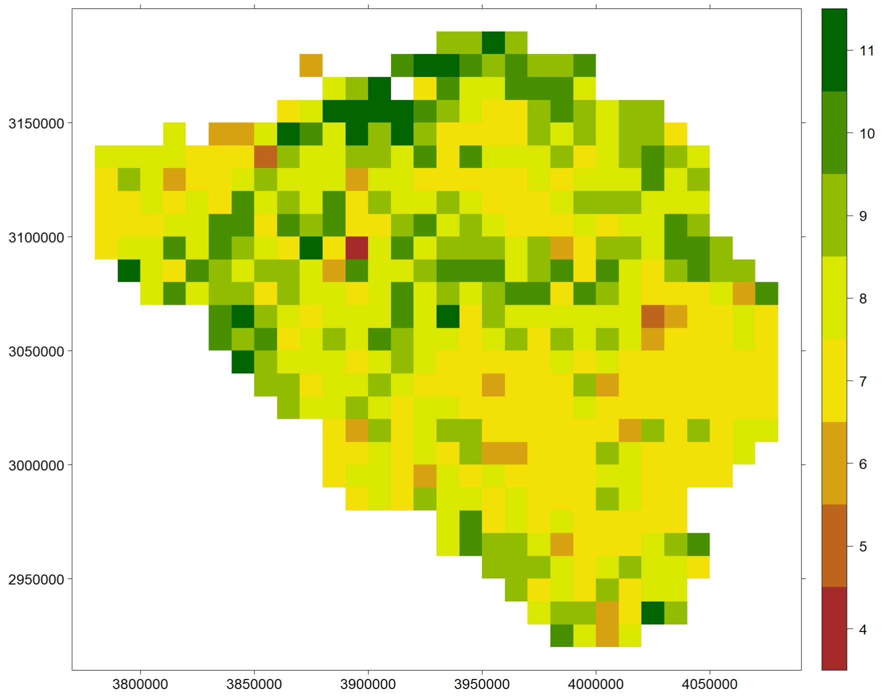
Predicted number of plant families in pollen samples in Belgium in June



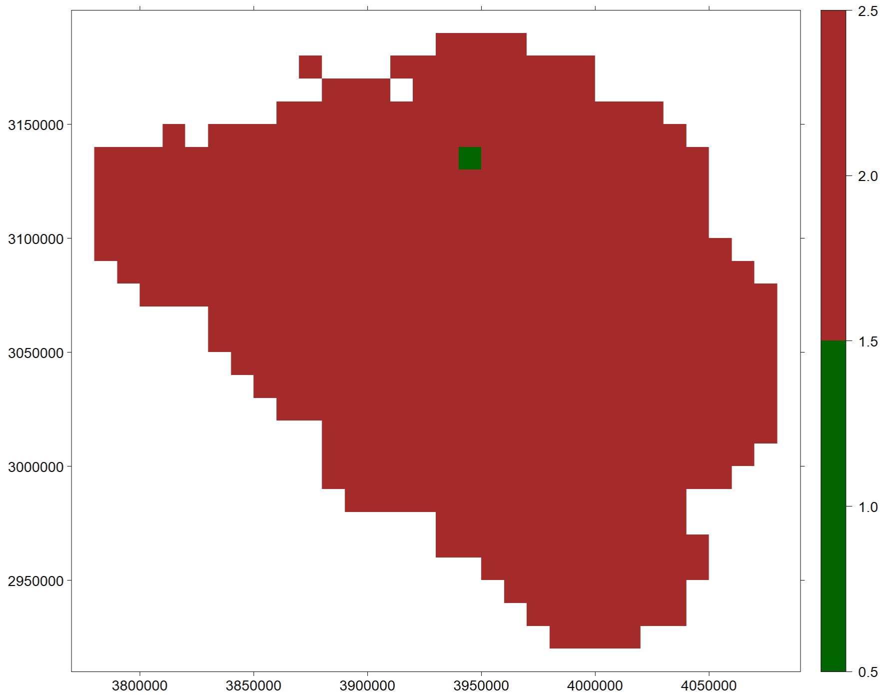
Predicted number of PPPs in APIStrip sample in Belgium in June

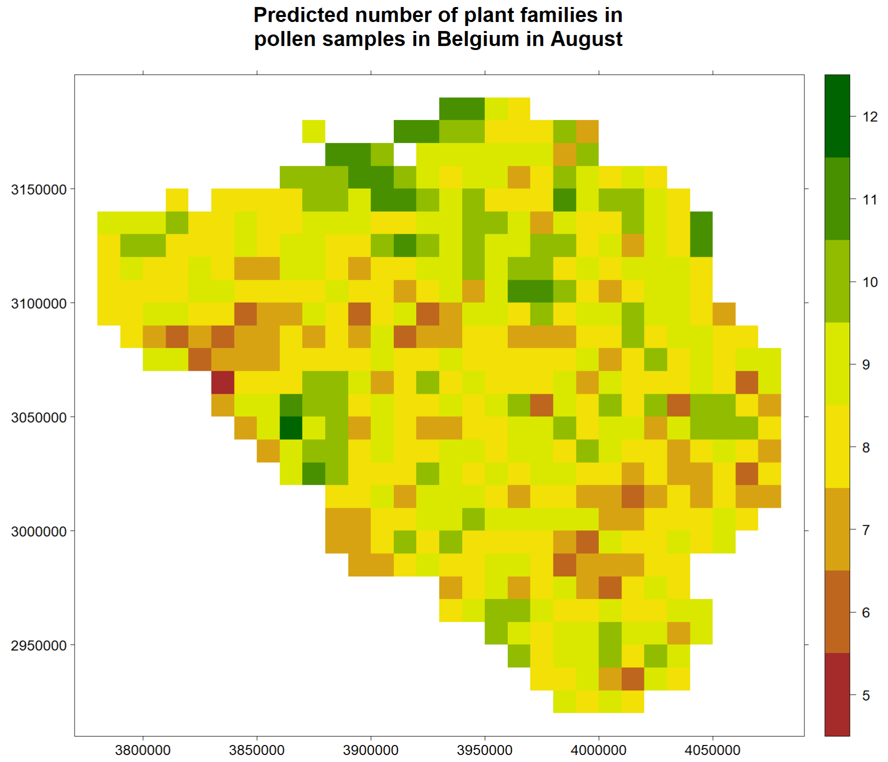


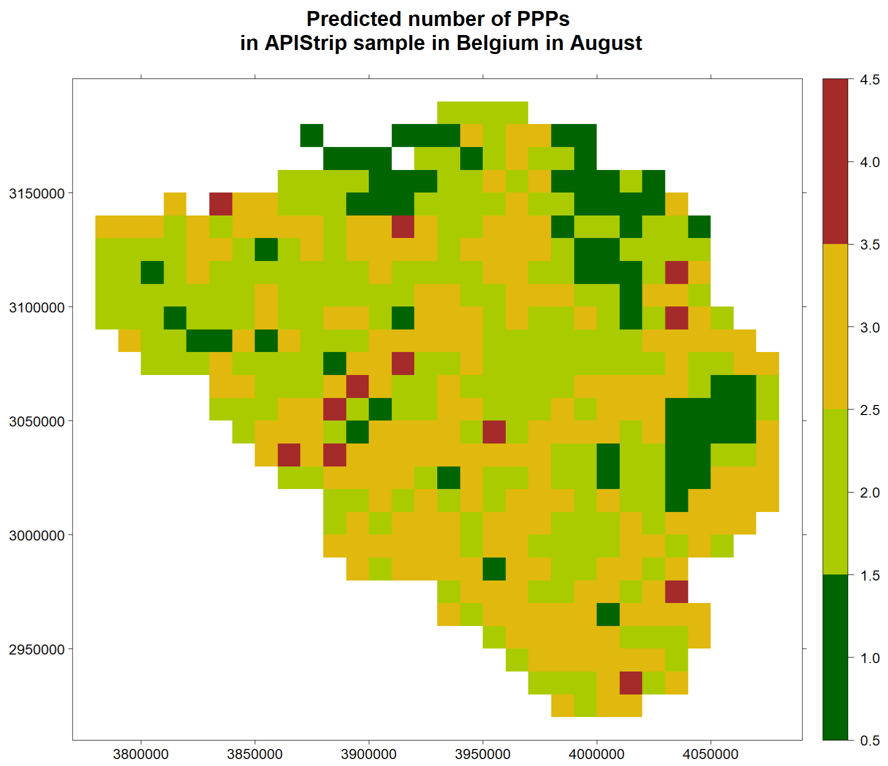
Predicted number of plant families in pollen samples in Belgium in July



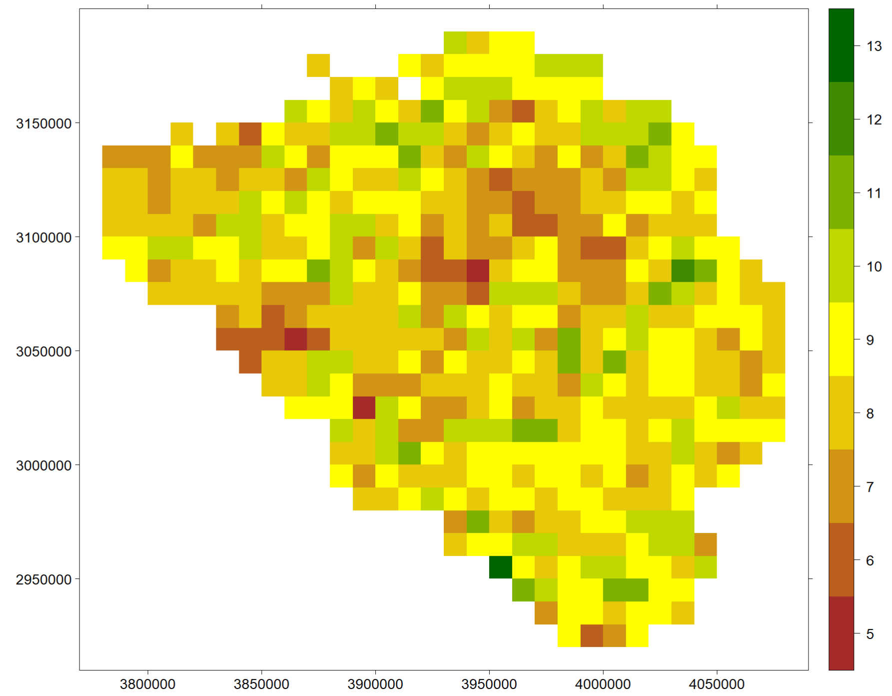
Predicted number of PPPs in APIStrip sample in Belgium in July



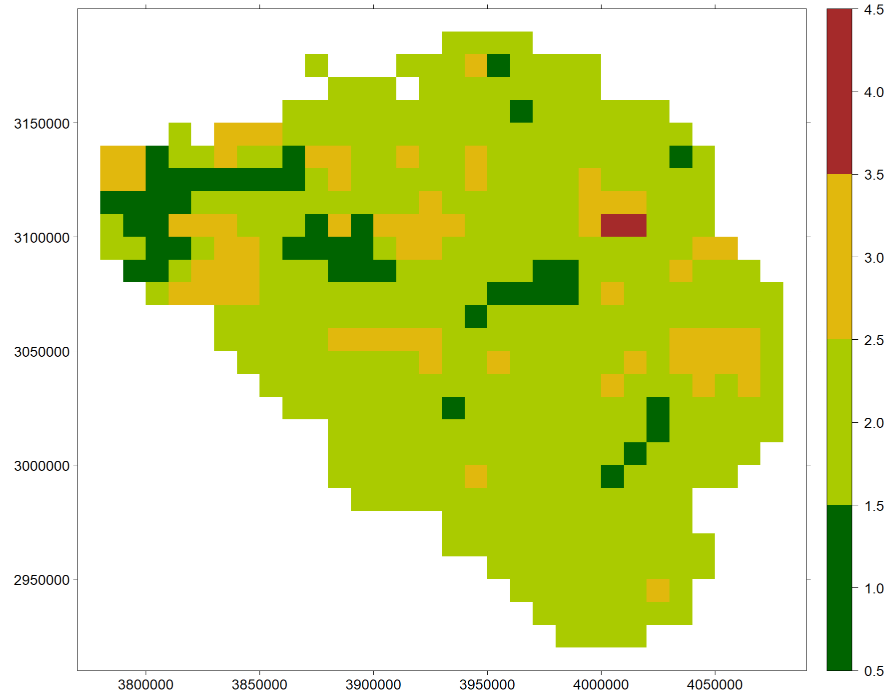

Predicted number of plant families in
pollen samples in Belgium in August
3150000
3100000
3050000
3000000
2950000
3800000
3850000
3900000
3950000
4000000
4050000
12
11
10
9
8
7
6
5


Predicted number of PPPs
in APIStrip sample in Belgium in August
3150000
3100000
3050000
3000000
2950000
3800000
3850000
3900000
3950000
4000000
4050000
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5

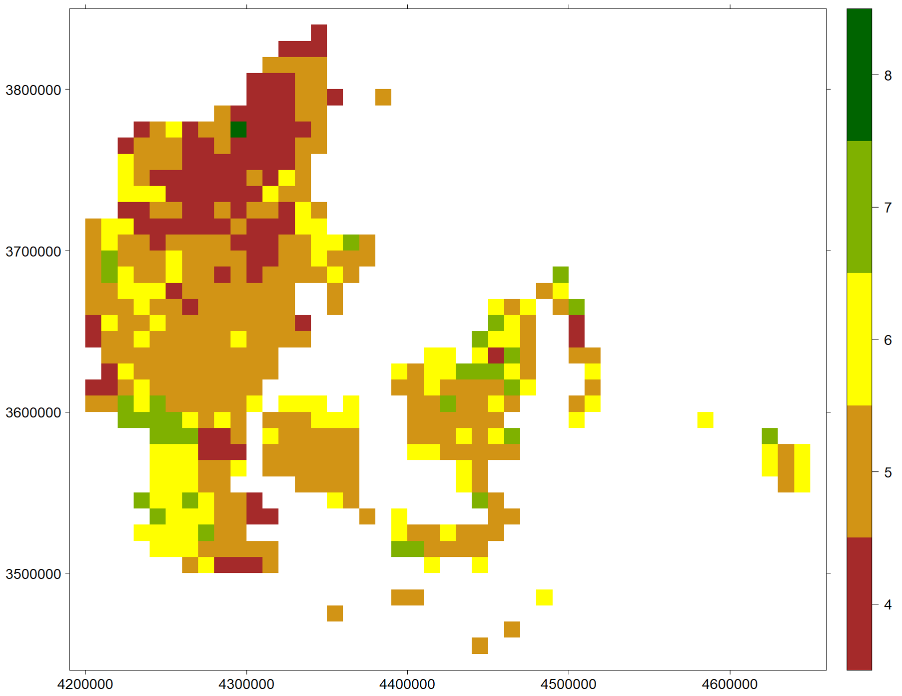
Predicted number of plant families in pollen samples in Belgium in September



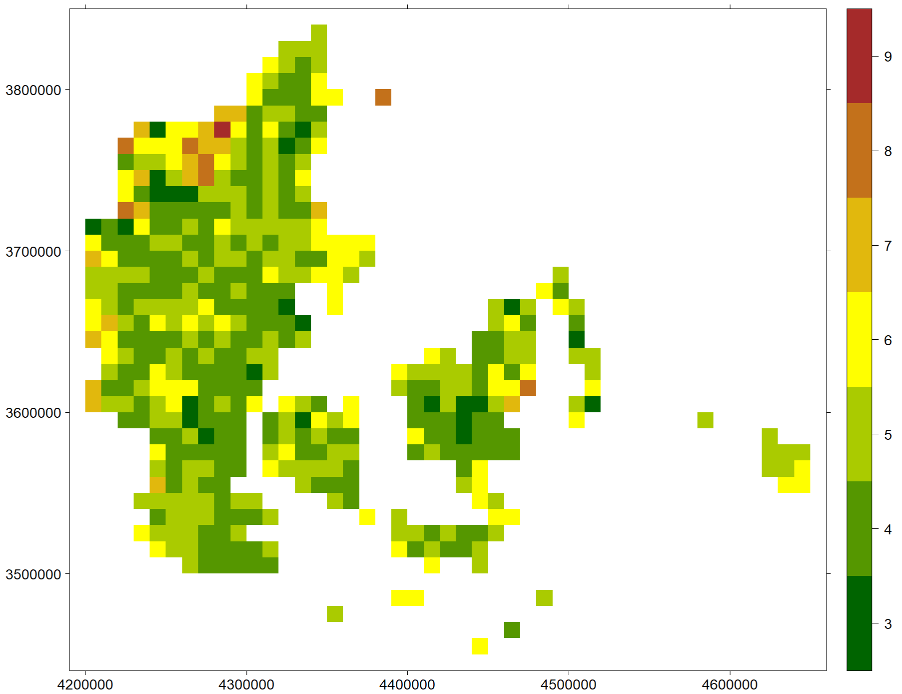
Predicted number of PPPs in APIStrip sample in Belgium in September

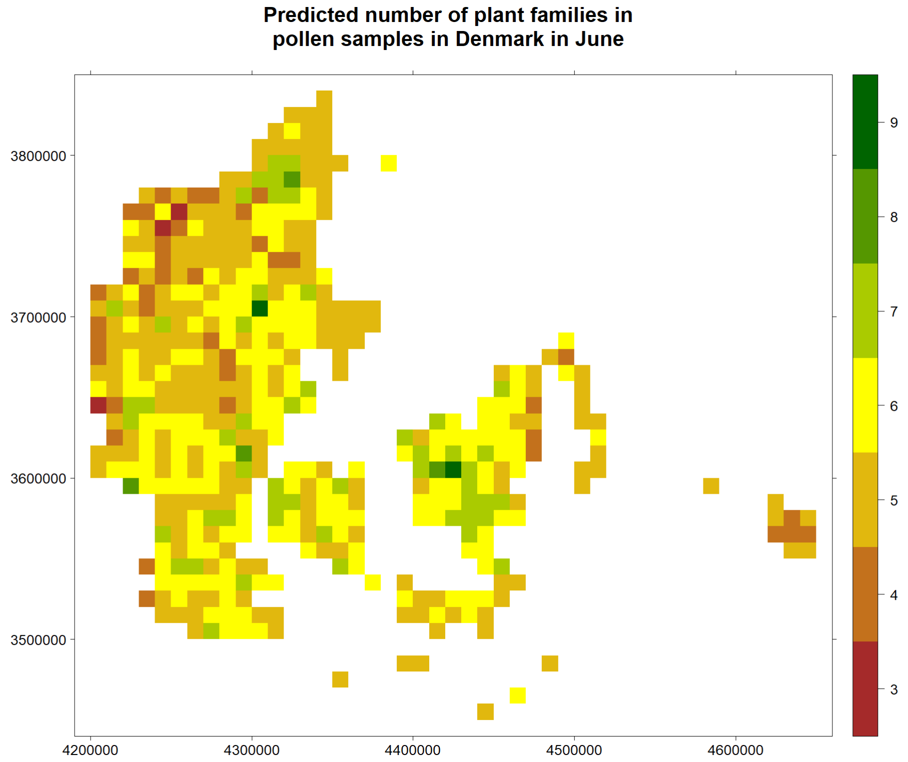


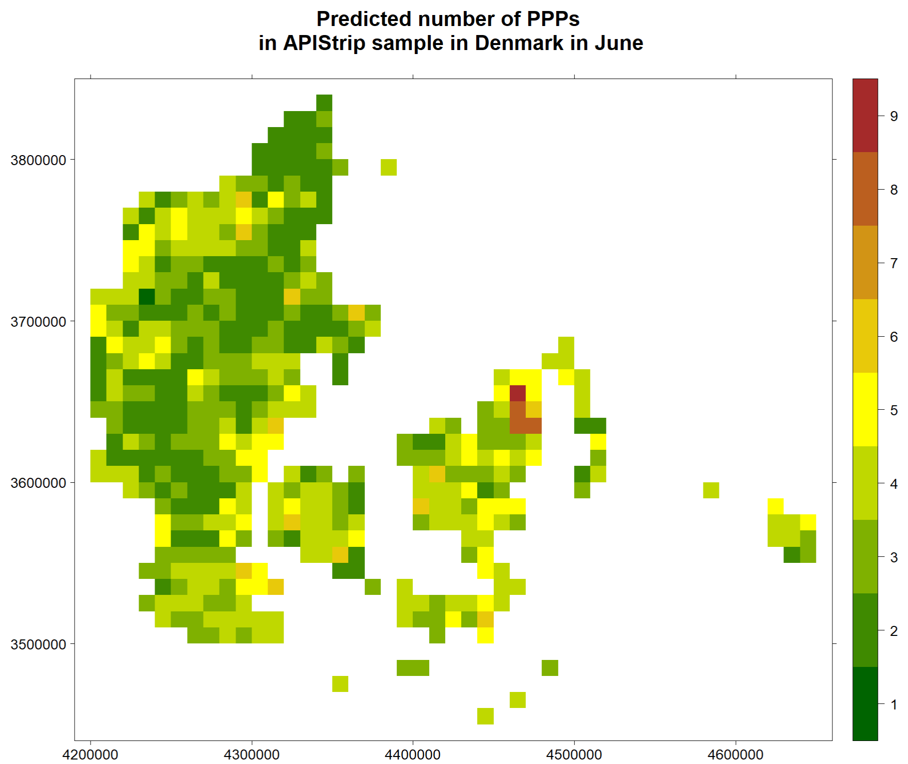
Predicted number of plant families in pollen samples in Denmark in May

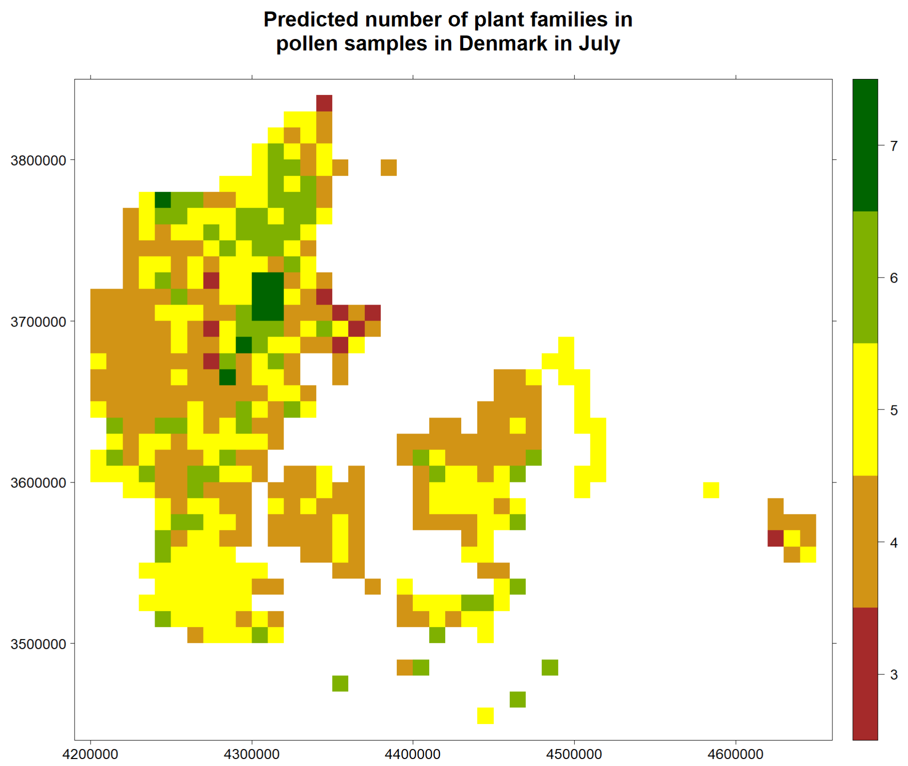


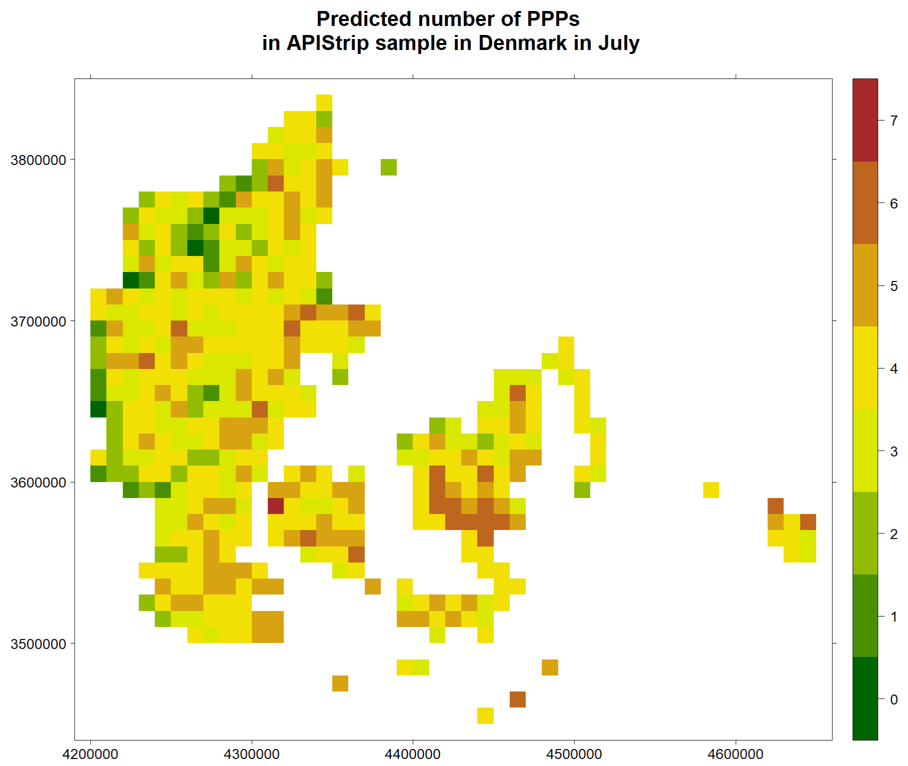
Predicted number of PPPs in APIStrip sample in Denmark in May

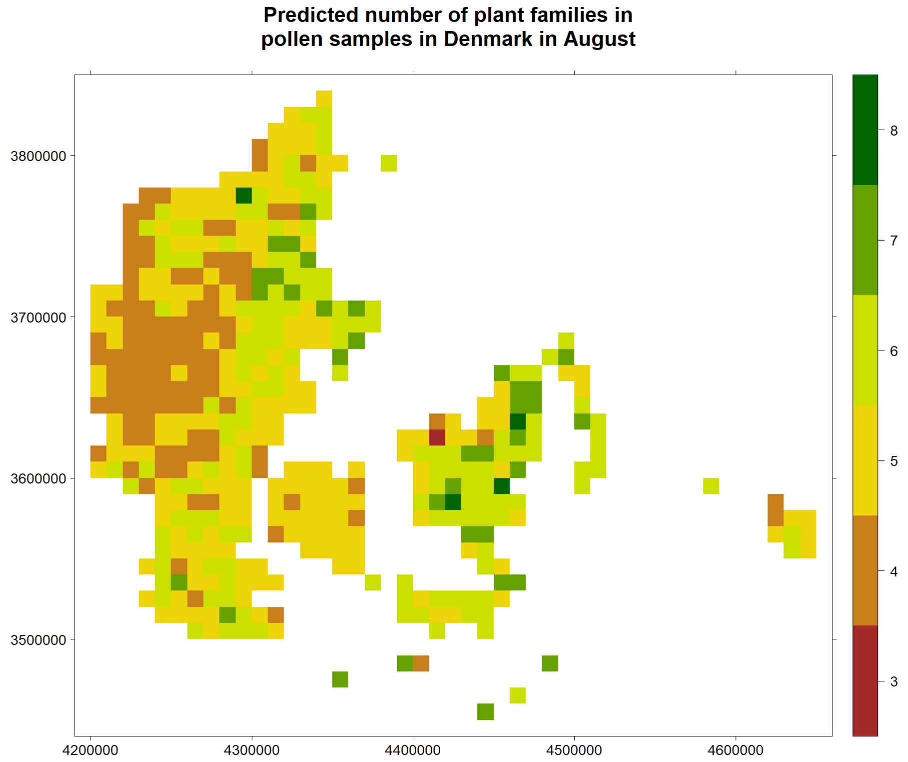


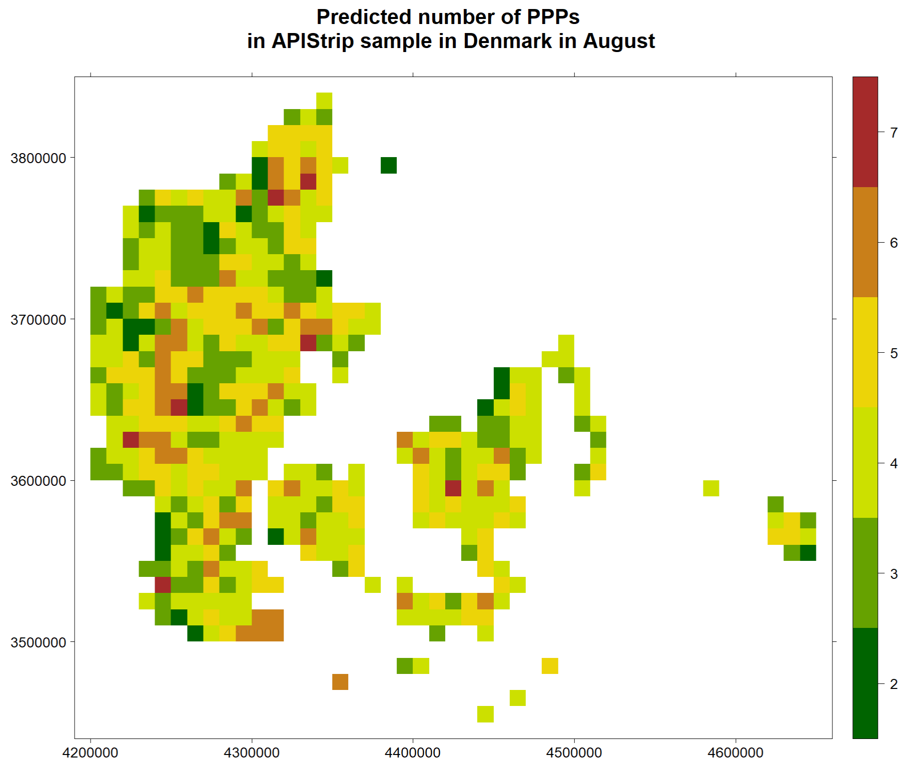

Predicted number of plant families in
pollen samples in Denmark in June
3800000
3700000
3600000
3500000
4200000
4300000
4400000
4500000
4600000
9
8
7
6
5
4
3


Predicted number of PPPs
in APIStrip sample in Denmark in June
3800000
3700000
3600000
3500000
4200000
4300000
4400000
4500000
4600000
9
8
7
6
5
4
3
2
1

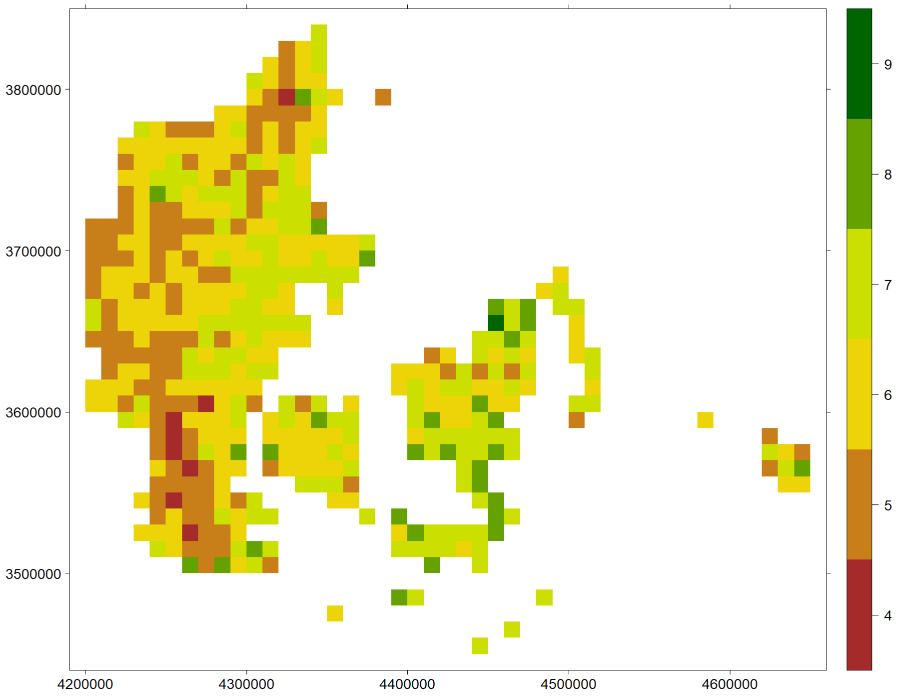

Predicted number of plant families in
pollen samples in Denmark in July
3800000
3700000
3600000
3500000
4200000
4300000
4400000
4500000
4600000
7
6
5
4
3


Predicted number of PPPs
in APIStrip sample in Denmark in July
3800000
3700000
3600000
3500000
4200000
4300000
4400000
4500000
4600000
7
6
5
4
3
2
1
0

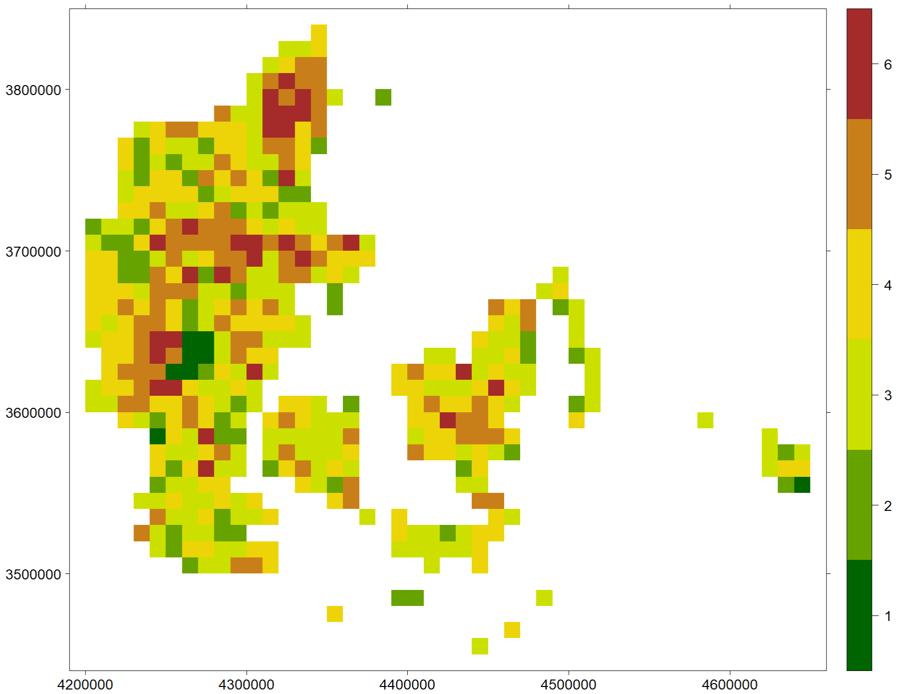

Predicted number of plant families in
pollen samples in Denmark in August
3800000
3700000
3600000
3500000
4200000
4300000
4400000
4500000
4600000
8
7
6
5
4
3

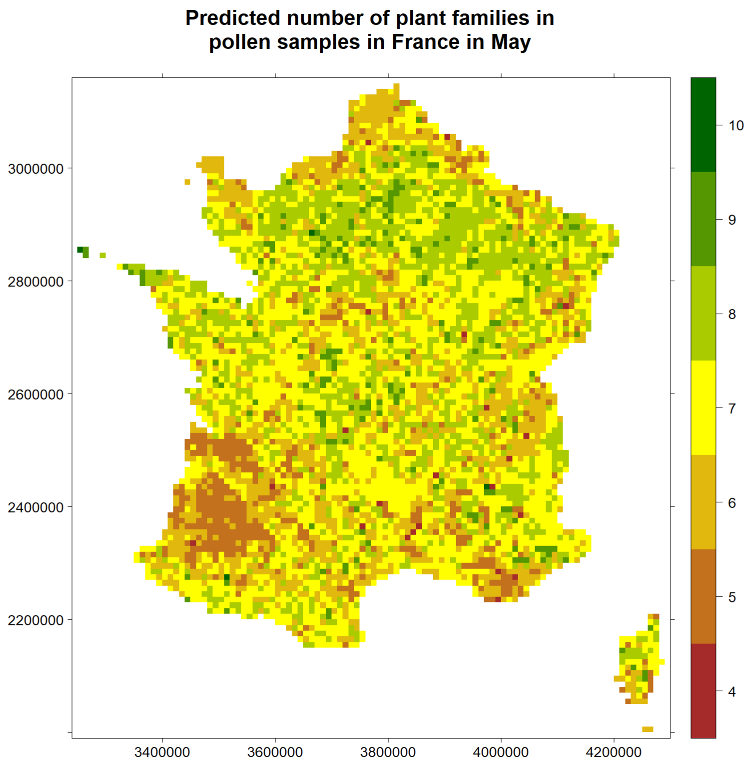

Predicted number of PPPs
in APIStrip sample in Denmark in August
3800000
3700000
3600000
3500000
4200000
4300000
4400000
4500000
4600000
7
6
5
4
3
2

Predicted number of plant families in pollen samples in Denmark in September

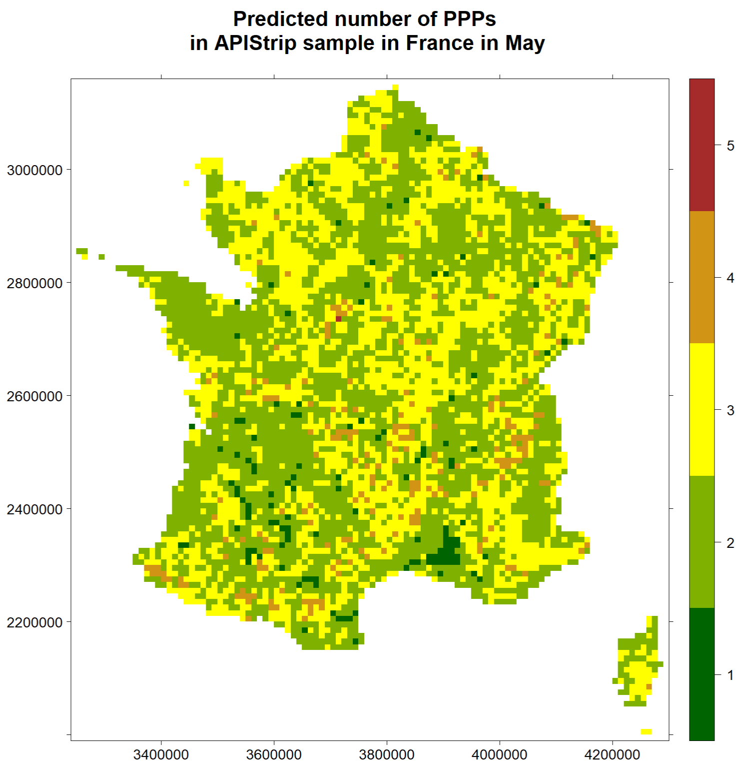


Predicted number of PPPs in APIStrip sample in Denmark in September



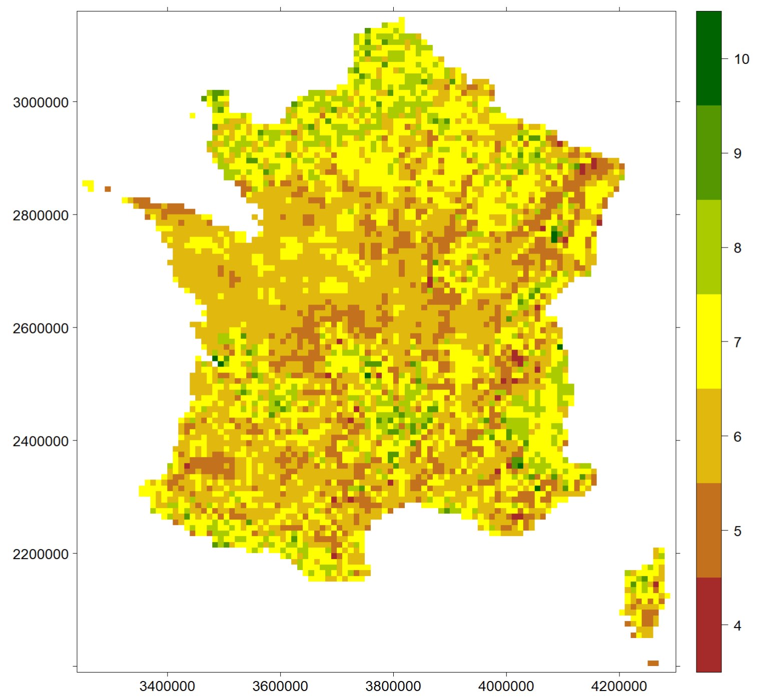


Predicted number of plant families in
pollen samples in France in May
3000000
2800000
2600000
2400000
2200000
3400000
3600000
3800000
4000000
4200000
10
9
8
7
6
5
4

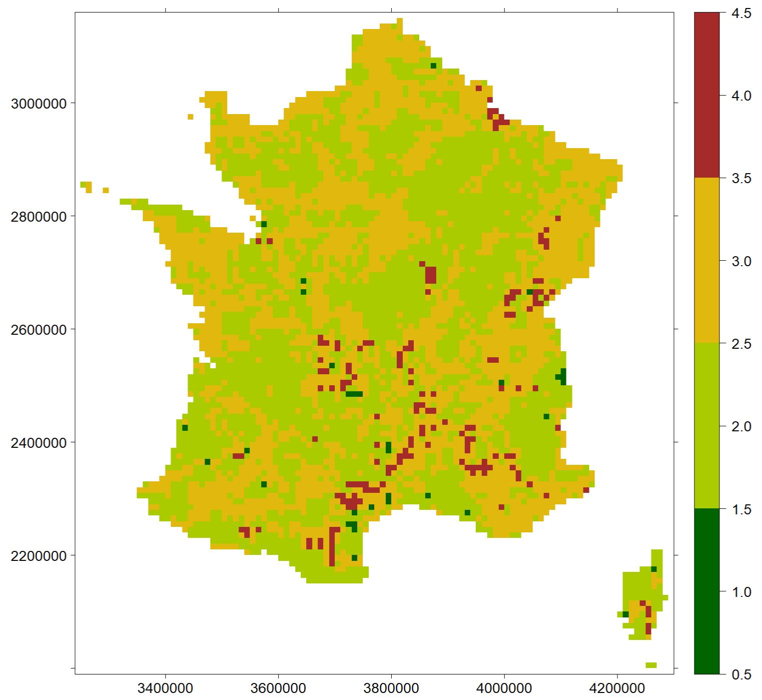


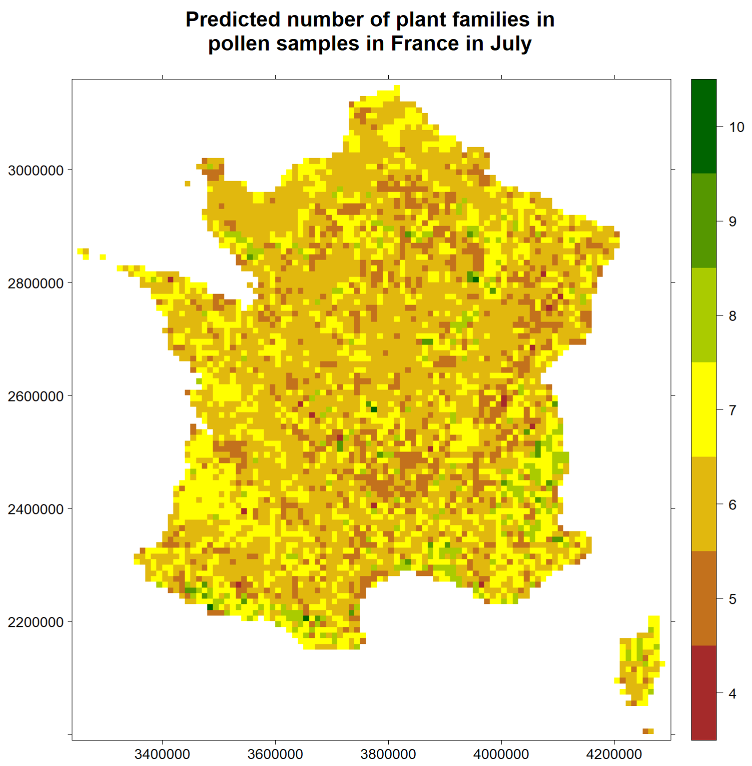
Predicted number of PPPs
in APIStrip sample in France in May
3000000
2800000
2600000
2400000
2200000
3400000
3600000
3800000
4000000
4200000
5
4
3
2
1

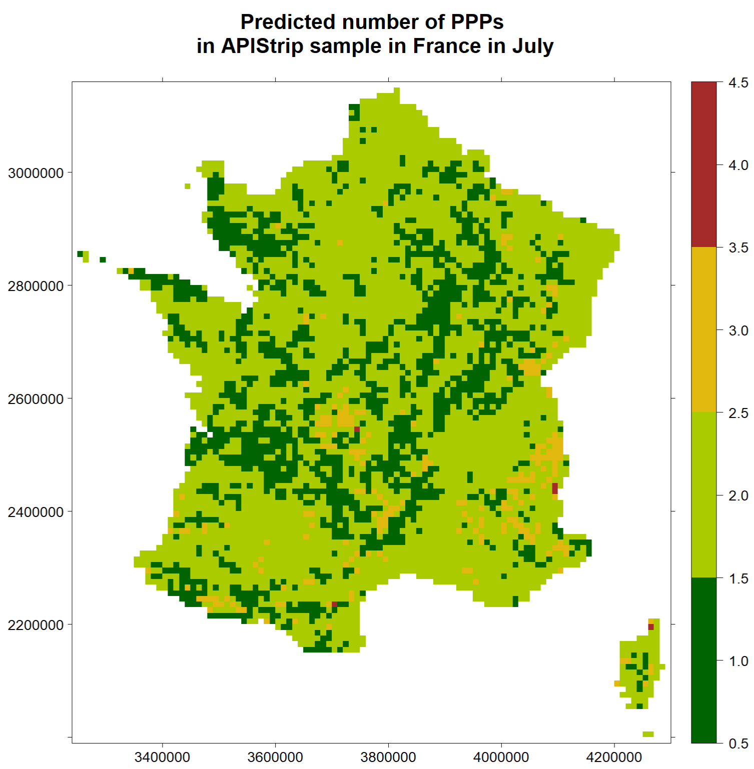
Predicted number of plant families in pollen samples in France in June



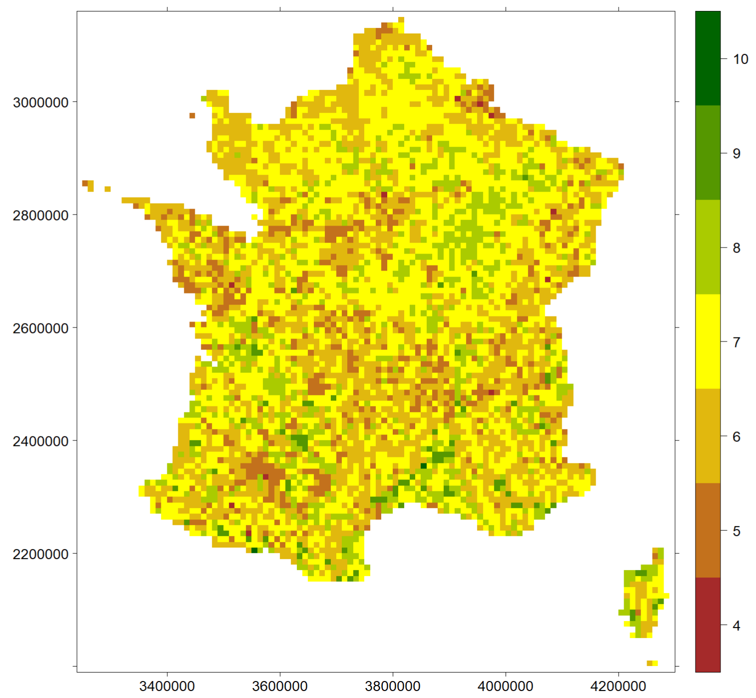
Predicted number of PPPs in APIStrip sample in France in June



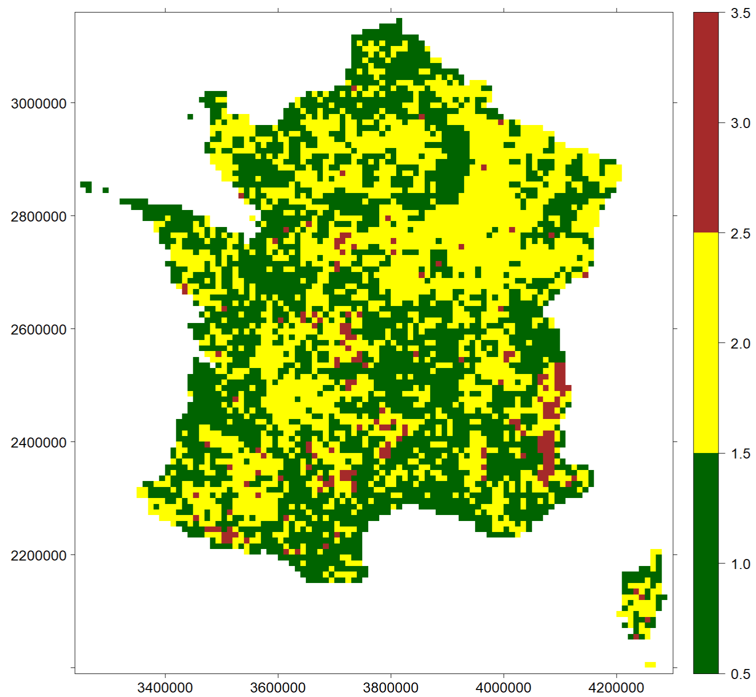

Predicted number of plant families in
pollen samples in France in July
3000000
2800000
2600000
2400000
2200000
3400000
3600000
3800000
4000000
4200000
10
9
8
7
6
5
4


Predicted number of PPPs
in APIStrip sample in France in July
3000000
2800000
2600000
2400000
2200000
3400000
3600000
3800000
4000000
4200000
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5

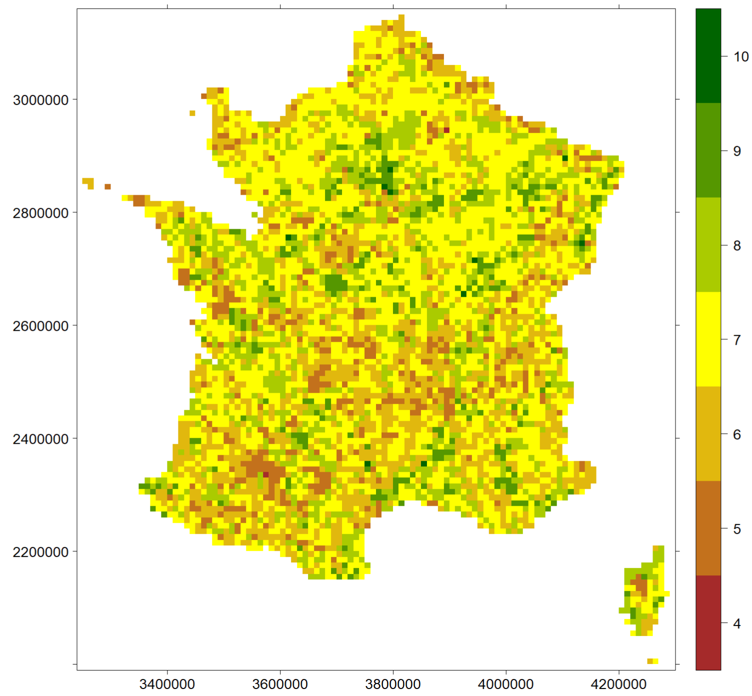
Predicted number of plant families in pollen samples in France in August



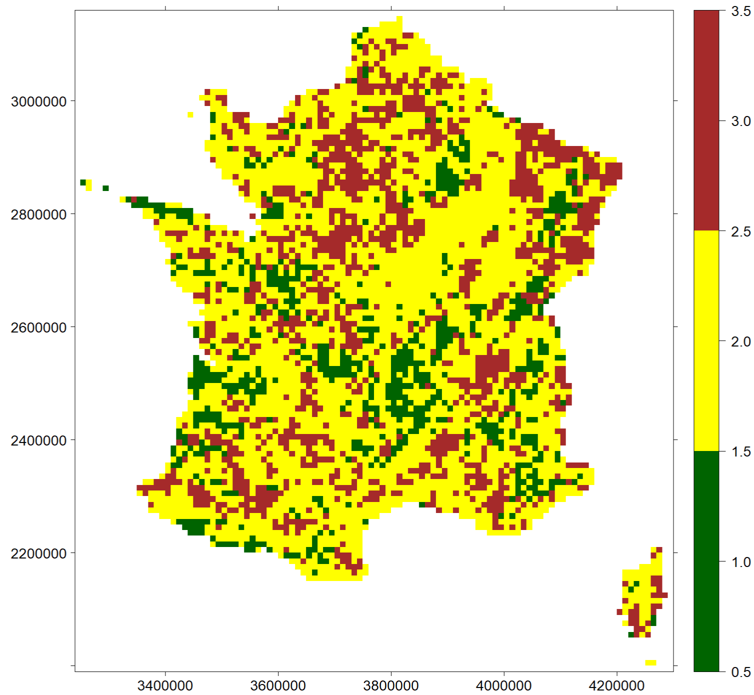
Predicted number of PPPs in APIStrip sample in France in August

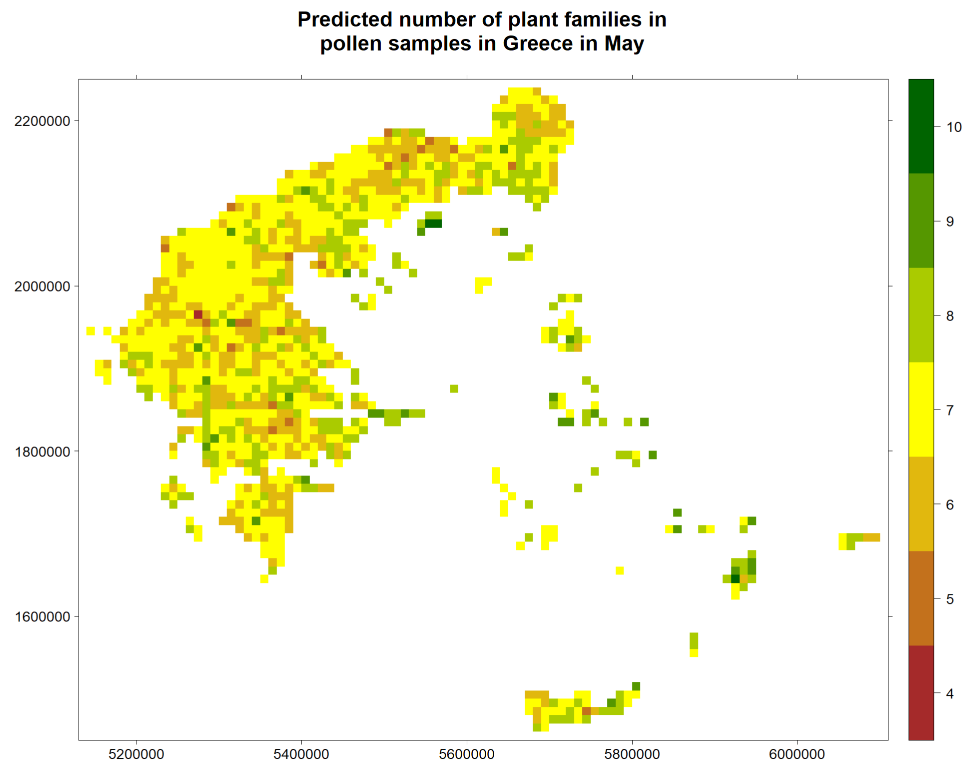


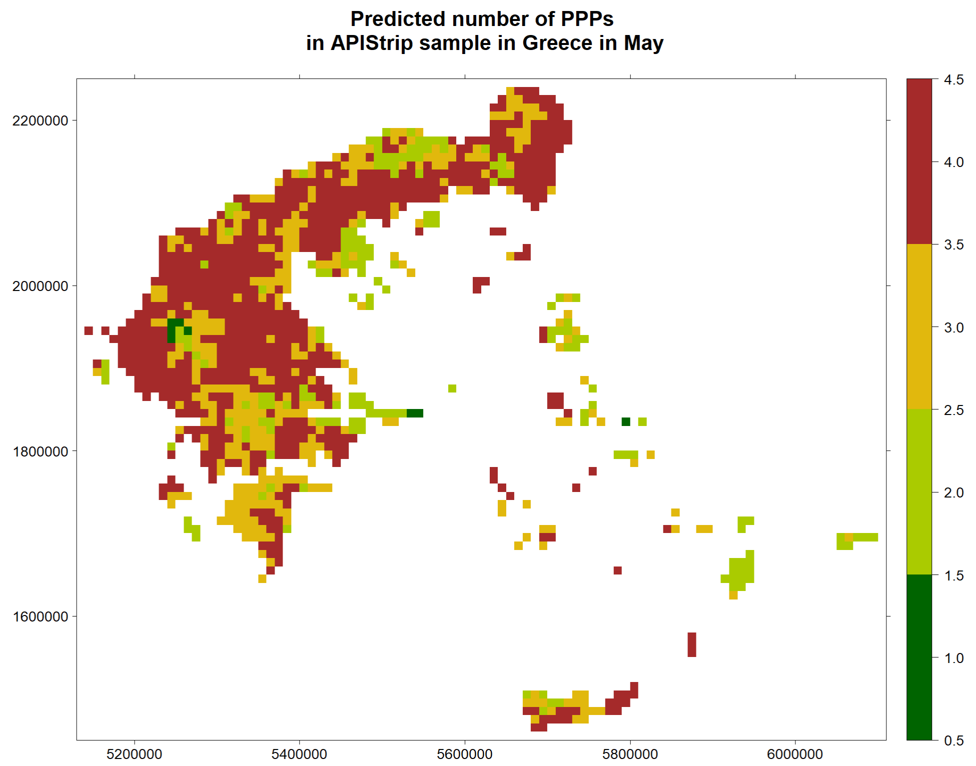
Predicted number of plant families in pollen samples in France in September



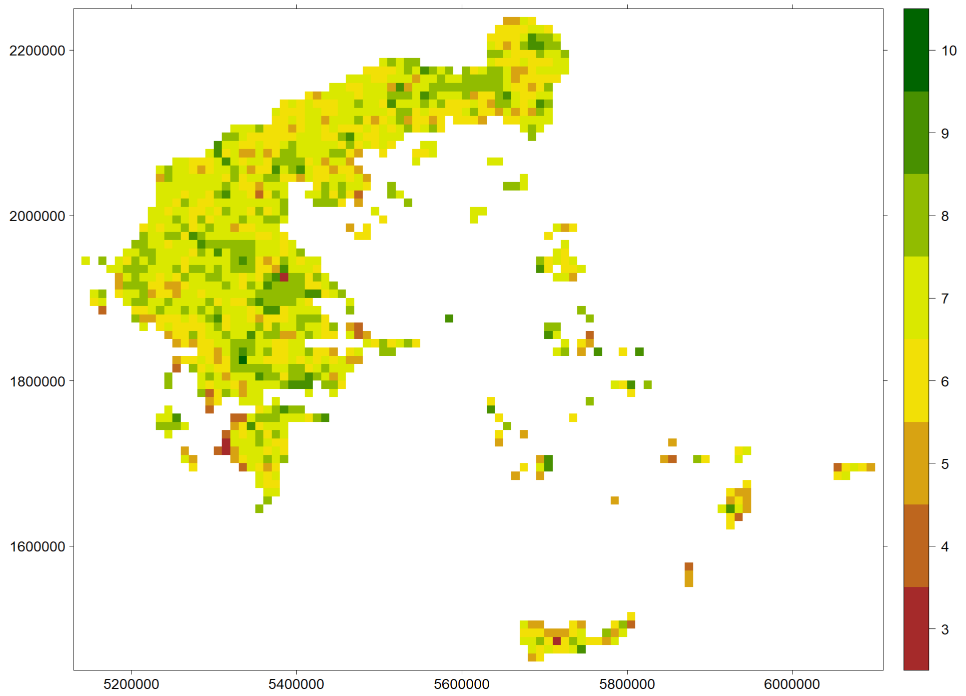
Predicted number of PPPs in APIStrip sample in France in September



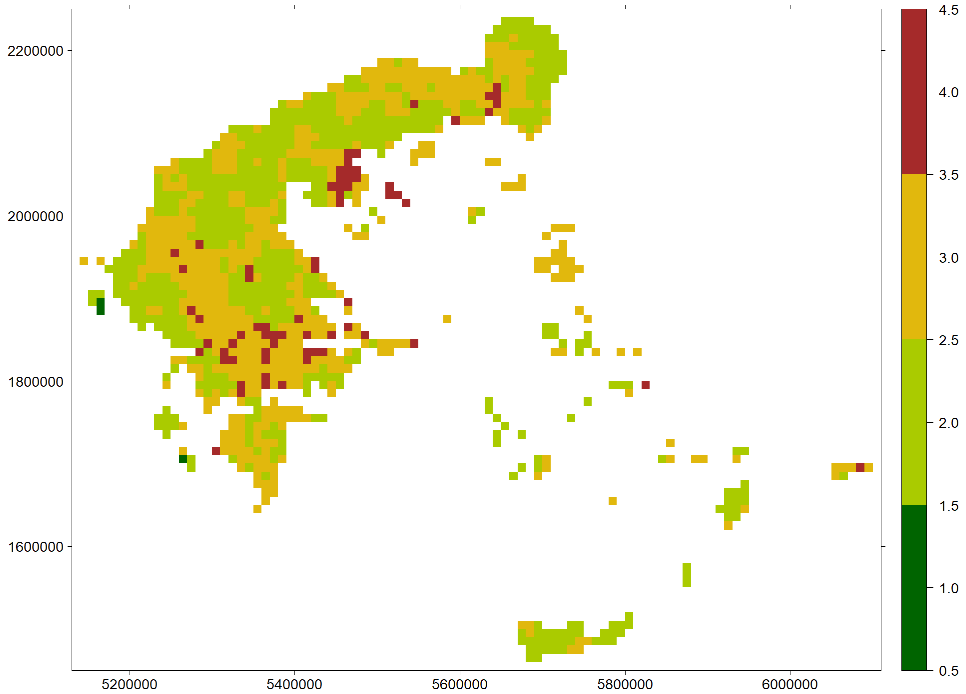

Predicted number of plant families in
pollen samples in Greece in May
2200000
2000000
1800000
1600000
5200000
5400000
5600000
5800000
6000000
10
9
8
7
6
5
4


Predicted number of PPPs
in APIStrip sample in Greece in May
2200000
2000000
1800000
1600000
5200000
5400000
5600000
5800000
6000000
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5

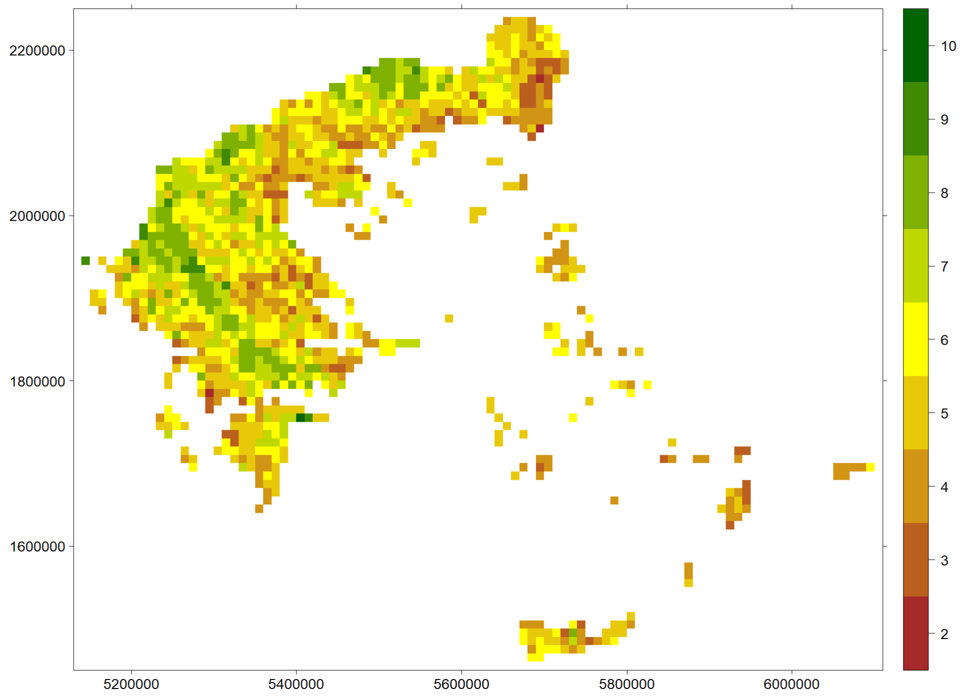
Predicted number of plant families in pollen samples in Greece in June



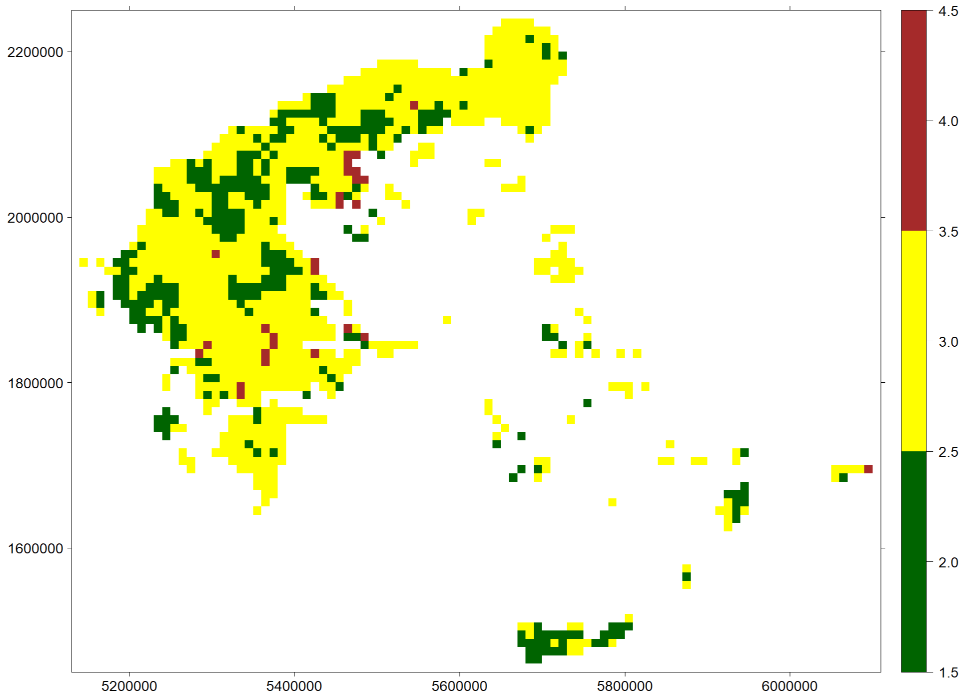
Predicted number of PPPs in APIStrip sample in Greece in June



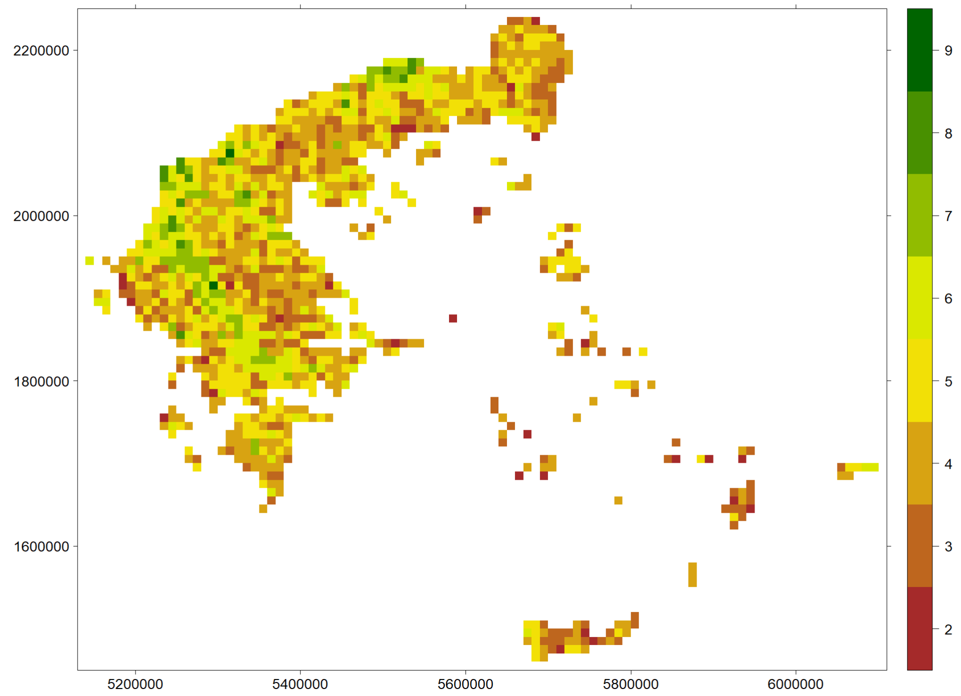
Predicted number of plant families in pollen samples in Greece in July



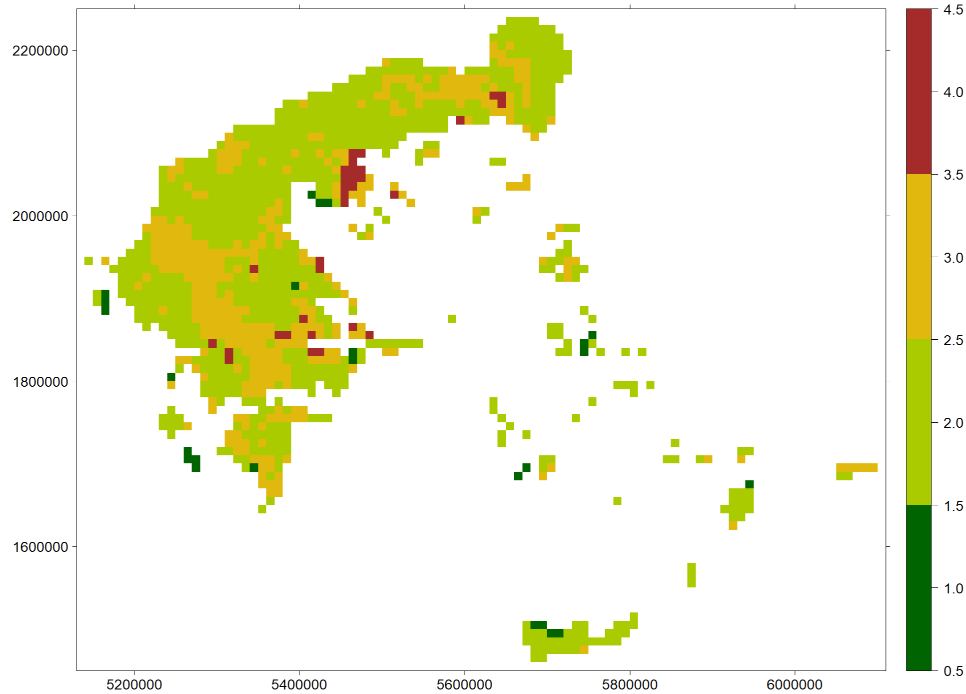
Predicted number of PPPs in APIStrip sample in Greece in July



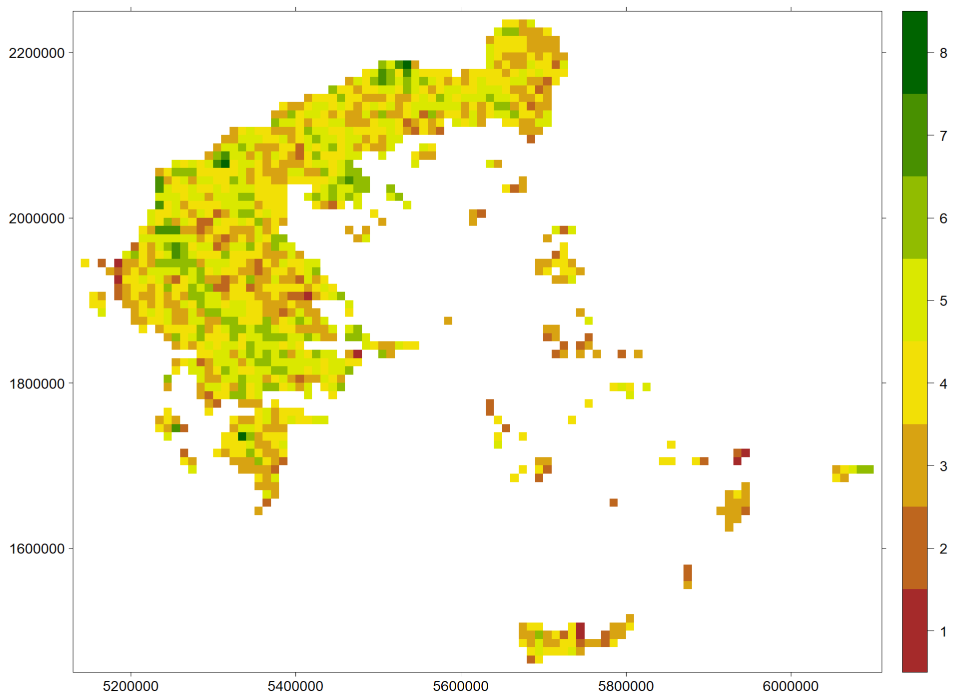
Predicted number of plant families in
pollen samples in Greece in August



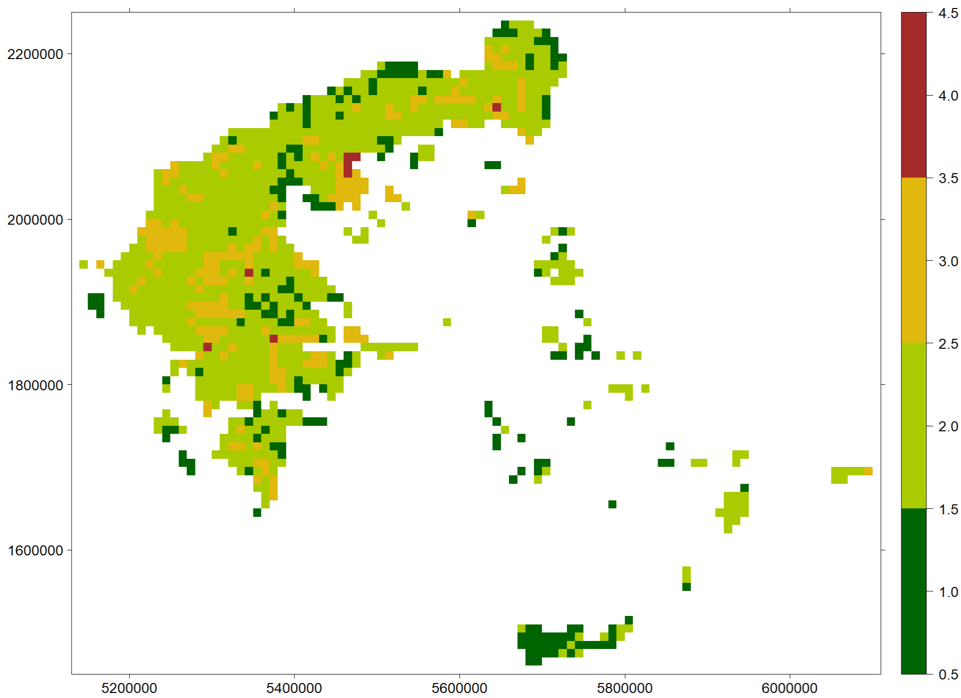
Predicted number of PPPs
in APIStrip sample in Greece in August



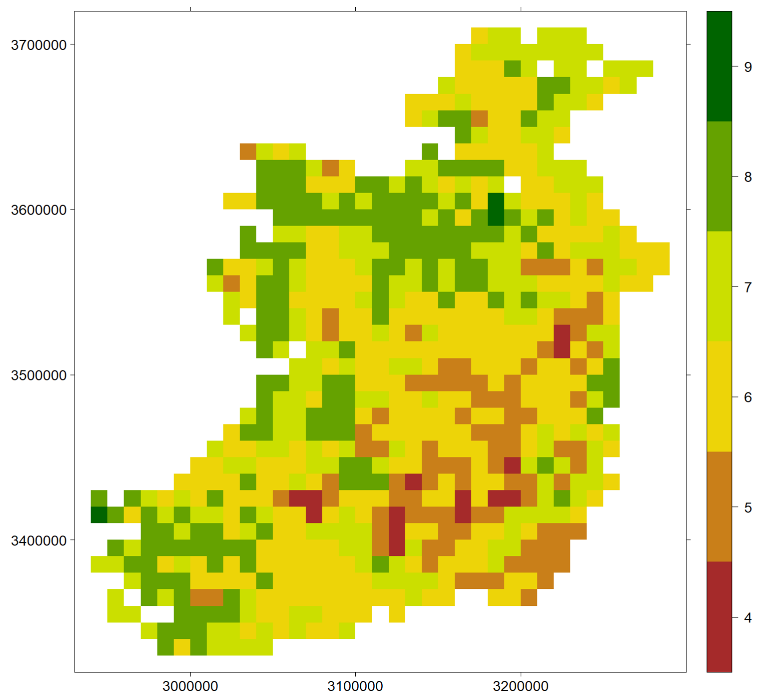
Predicted number of plant families in
pollen samples in Greece in September



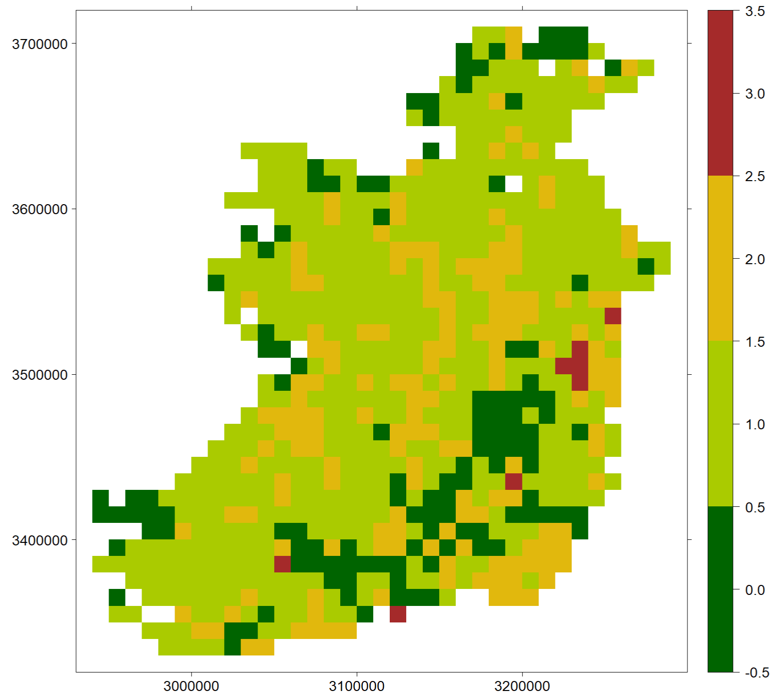
Predicted number of PPPs
in APIStrip sample in Greece in September



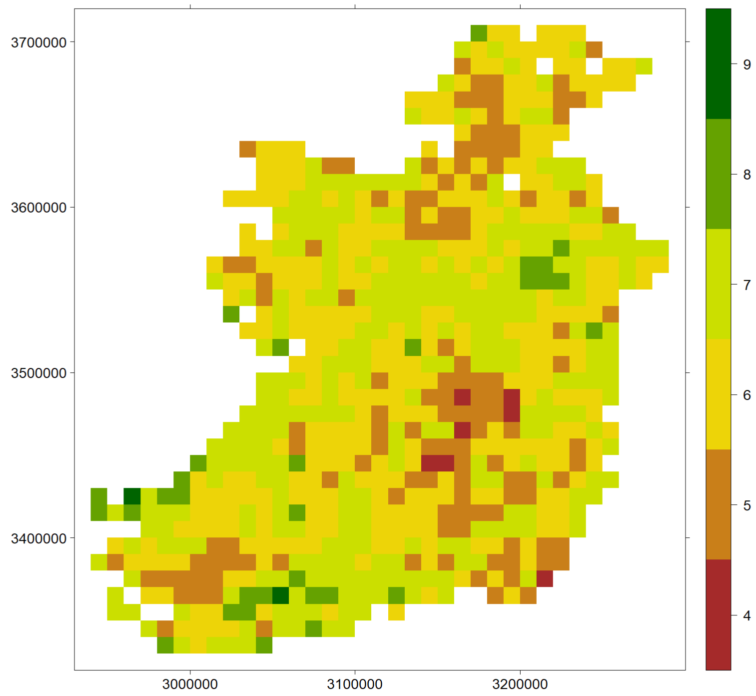
Predicted number of plant families in
pollen samples in Ireland in May



Predicted number of PPPs
in APIStrip sample in Ireland in May

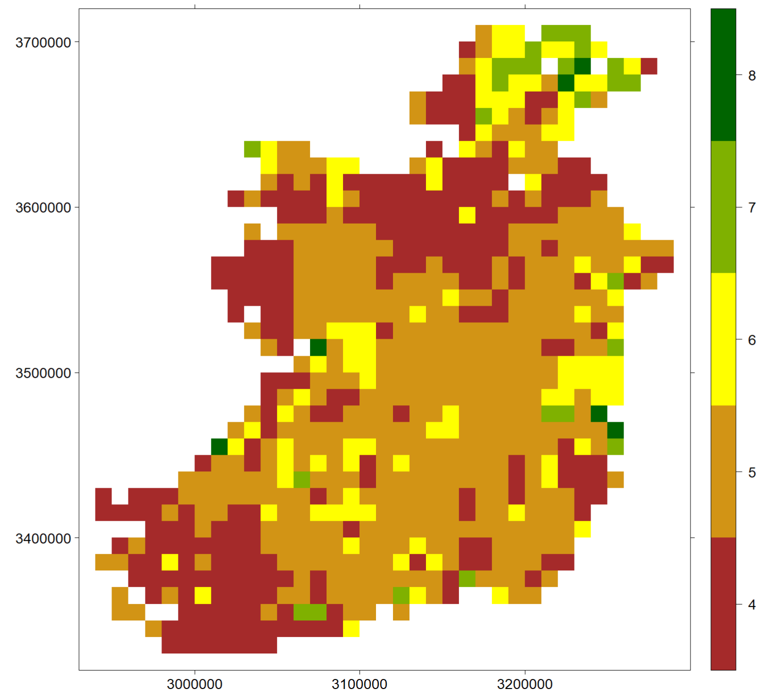


Predicted number of plant families in
pollen samples in Ireland in June



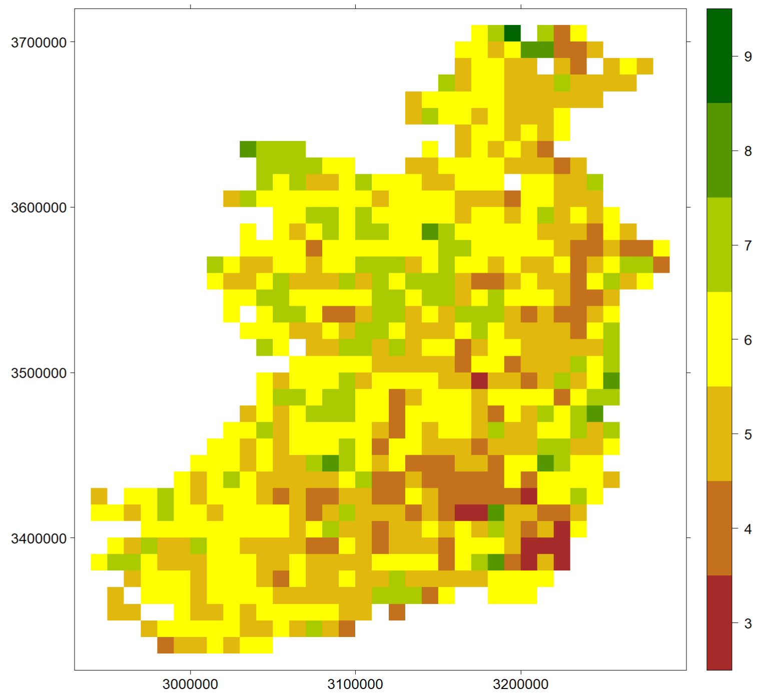
Predicted number of PPPs
in APIStrip sample in Ireland in June

Predicted number of plant families in pollen samples in Ireland in July

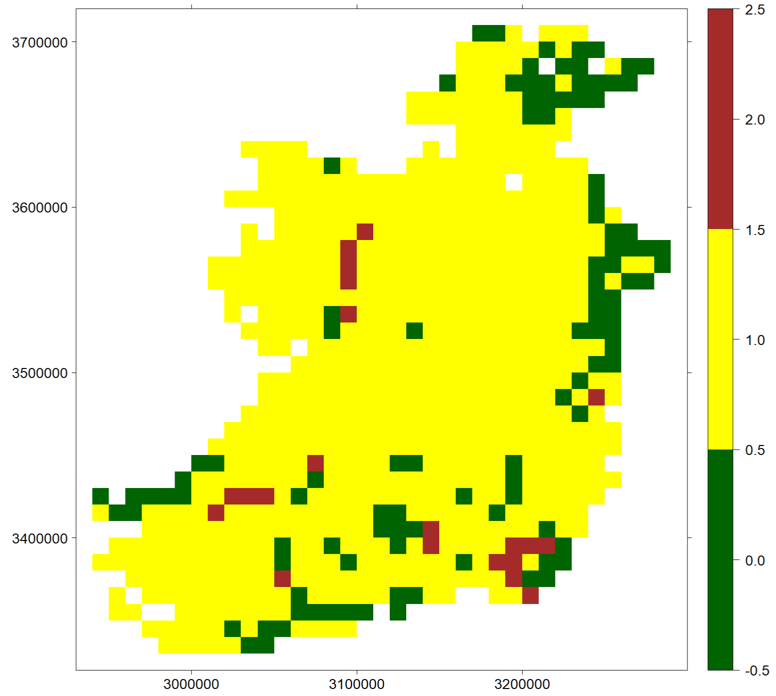


Predicted number of PPPs in APIStrip sample in Ireland in July

Predicted number of plant families in
pollen samples in Ireland in August

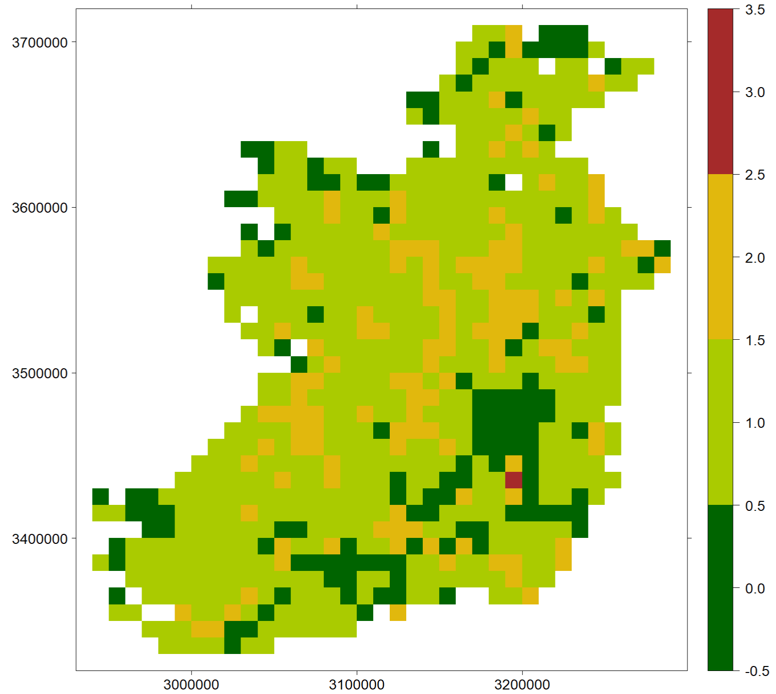


Predicted number of PPPs
in APIStrip sample in Ireland in August

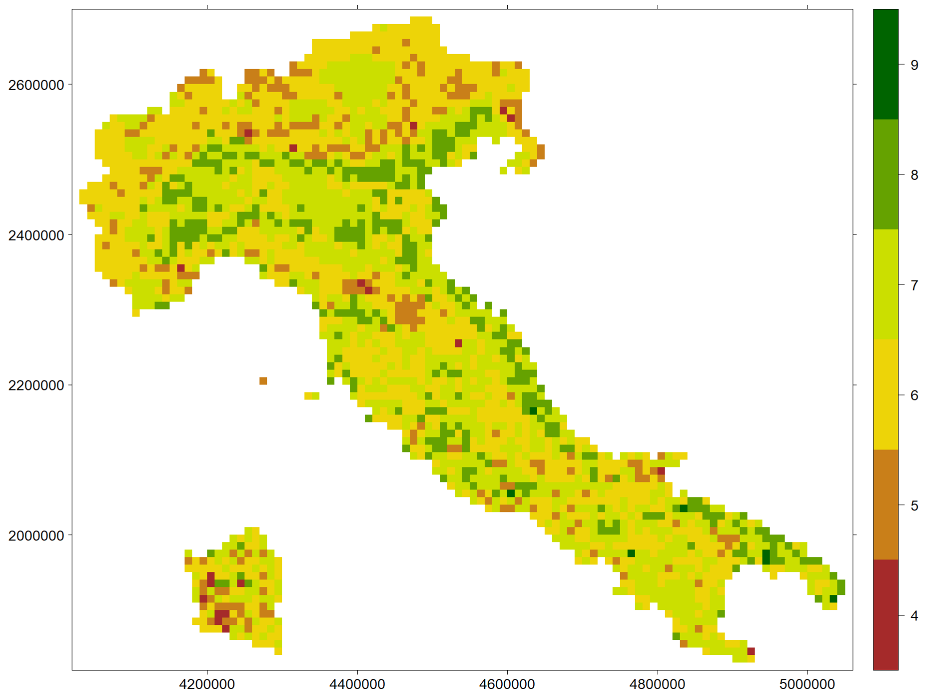


Predicted number of plant families in
pollen samples in Ireland in September

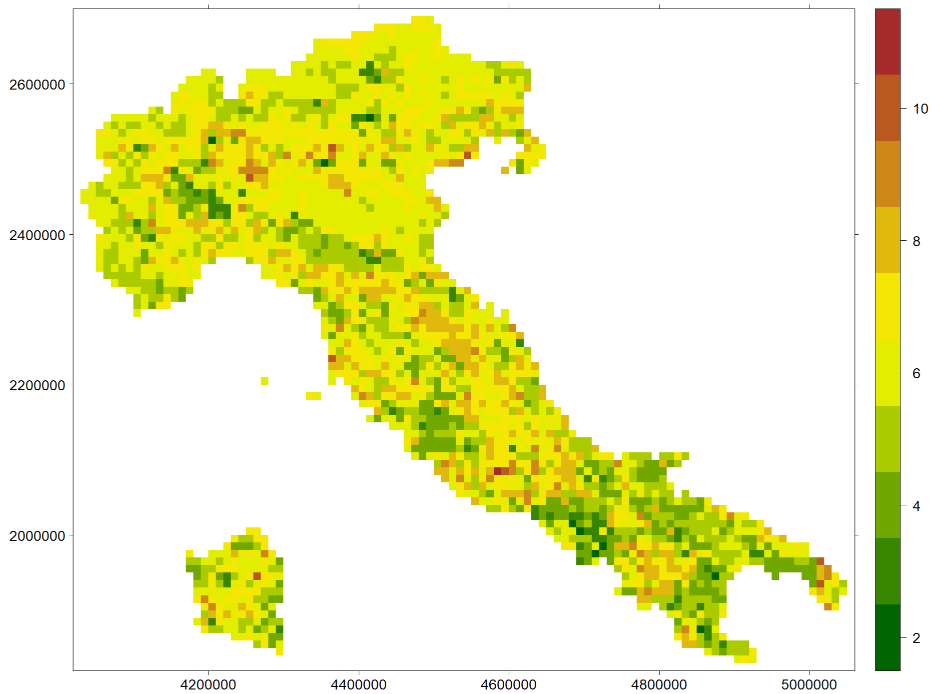
Predicted number of PPPs
in APIStrip sample in Ireland in September



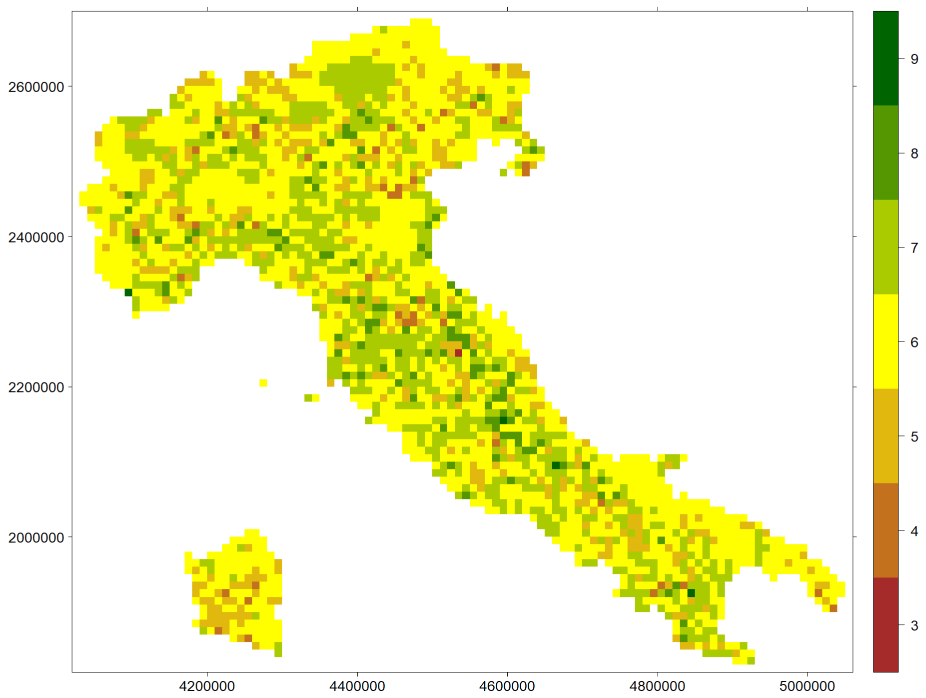
Predicted number of plant families in
pollen samples in Italy in May



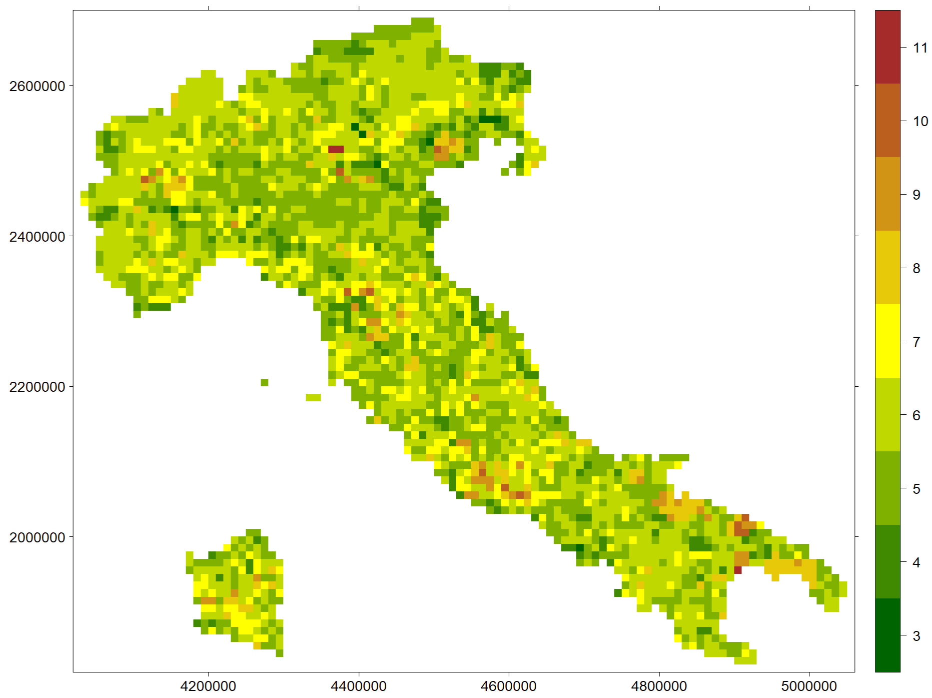
Predicted number of PPPs
in APIStrip sample in Italy in May



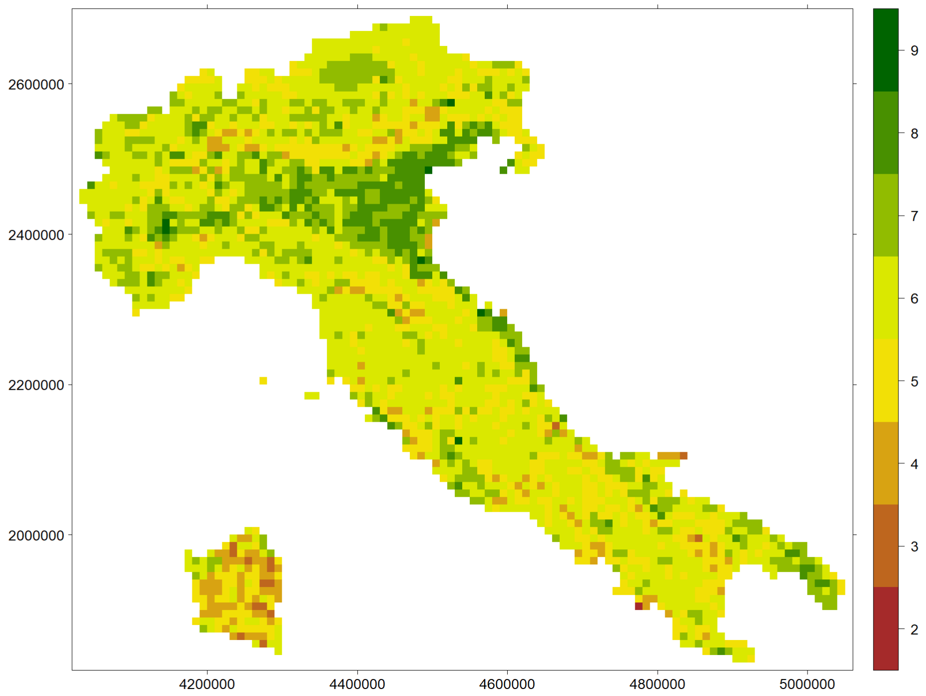
Predicted number of plant families in
pollen samples in Italy in June



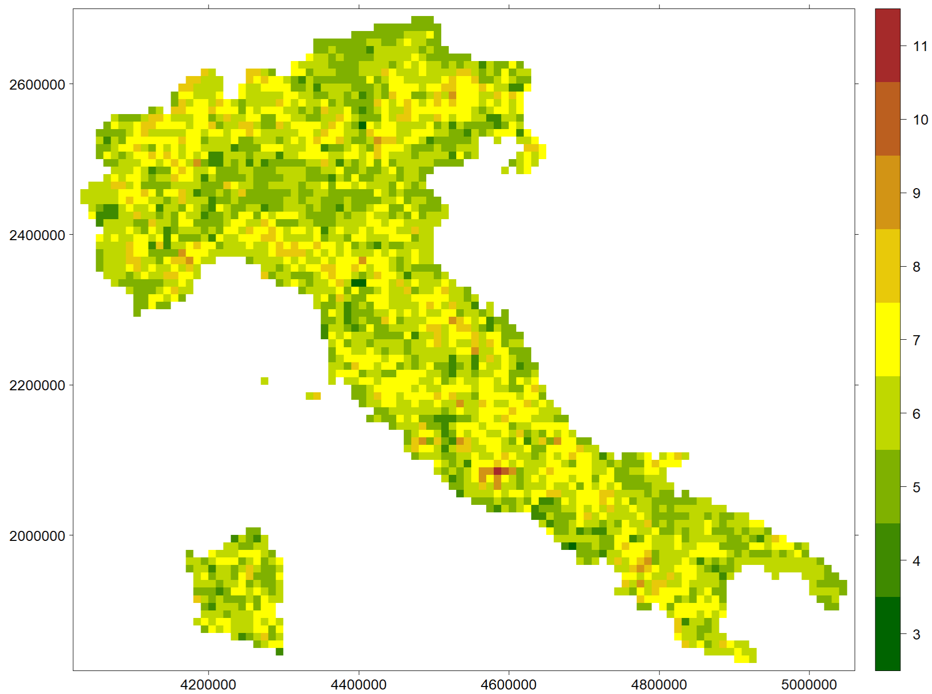
Predicted number of PPPs
in APIStrip sample in Italy in June

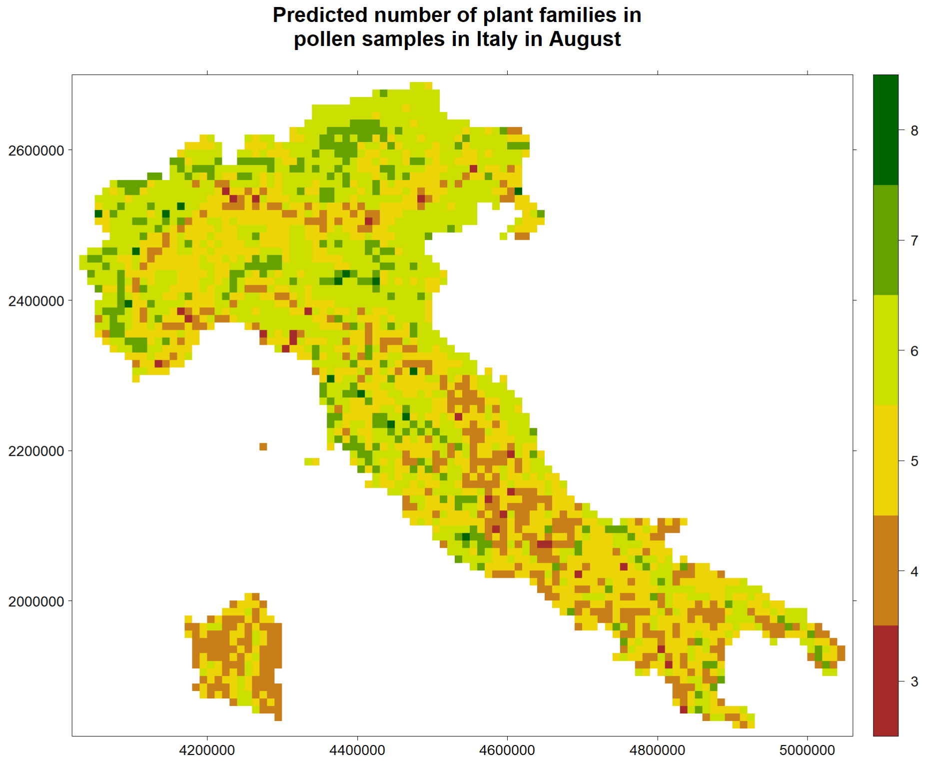


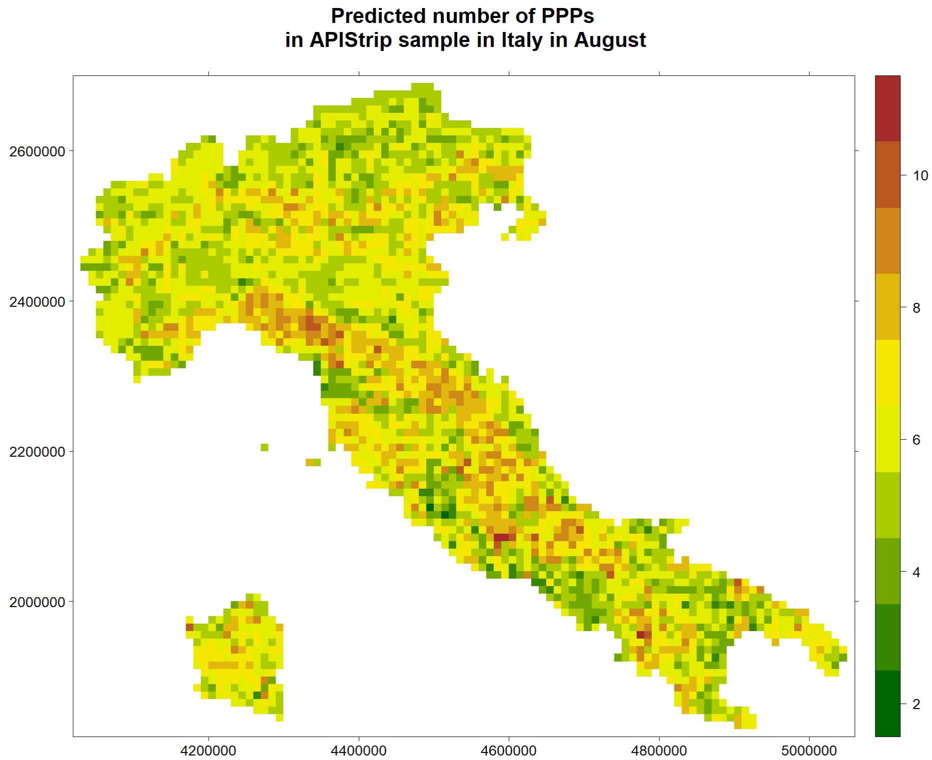
Predicted number of plant families in pollen samples in Italy in July



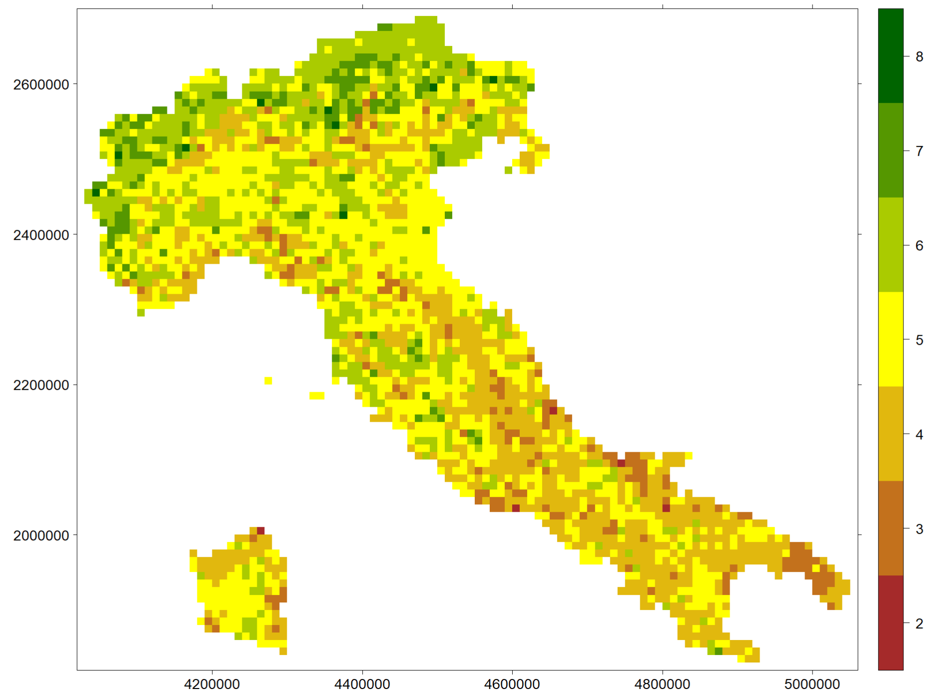
Predicted number of PPPs in APIStrip sample in Italy in July



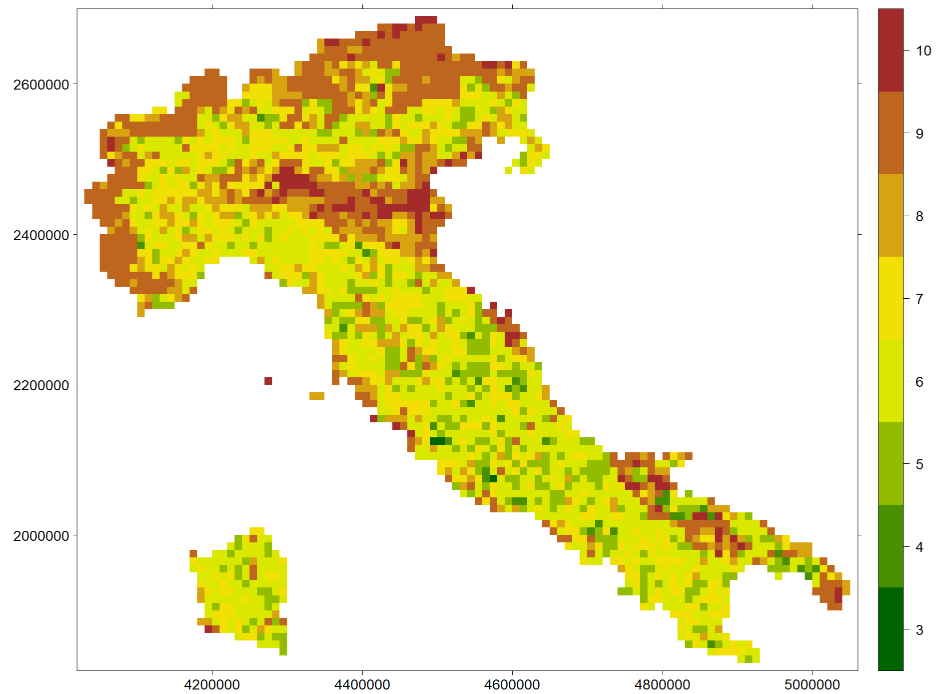

Predicted number of plant families in
pollen samples in Italy in August
2600000
2400000
2200000
2000000
4200000
4400000
4600000
4800000
5000000
8
7
6
5
4
3


Predicted number of PPPs
in APIStrip sample in Italy in August
2600000
2400000
2200000
2000000
4200000
4400000
4600000
4800000
5000000
10
8
6
4
2

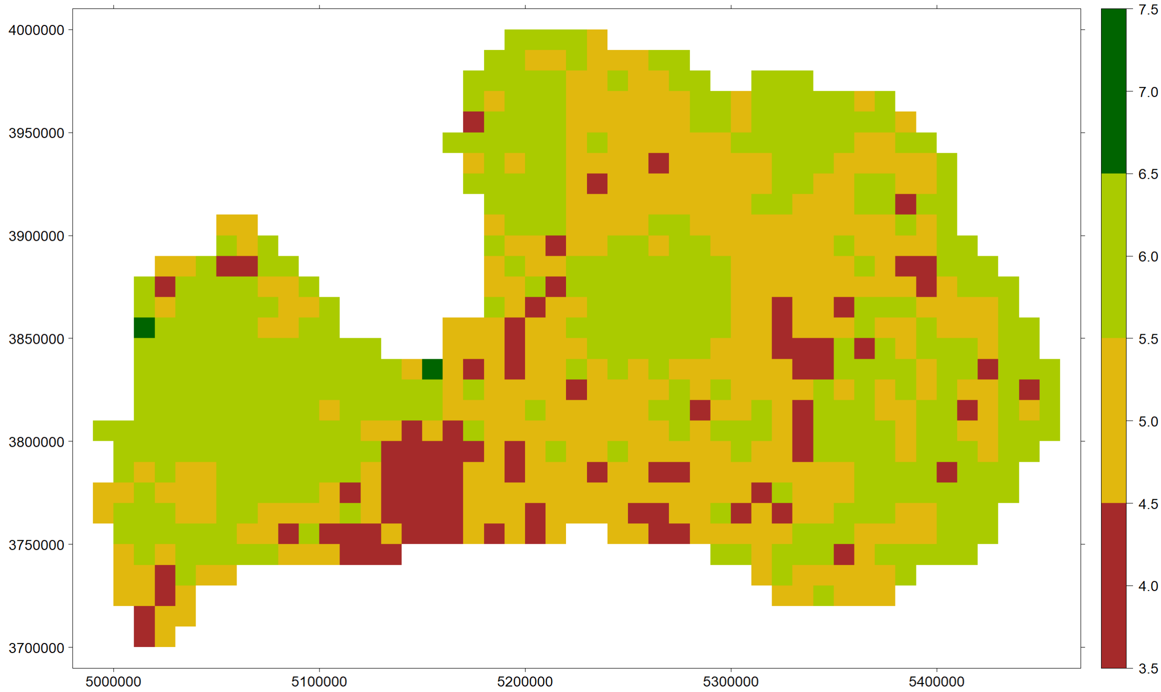
Predicted number of plant families in pollen samples in Italy in September



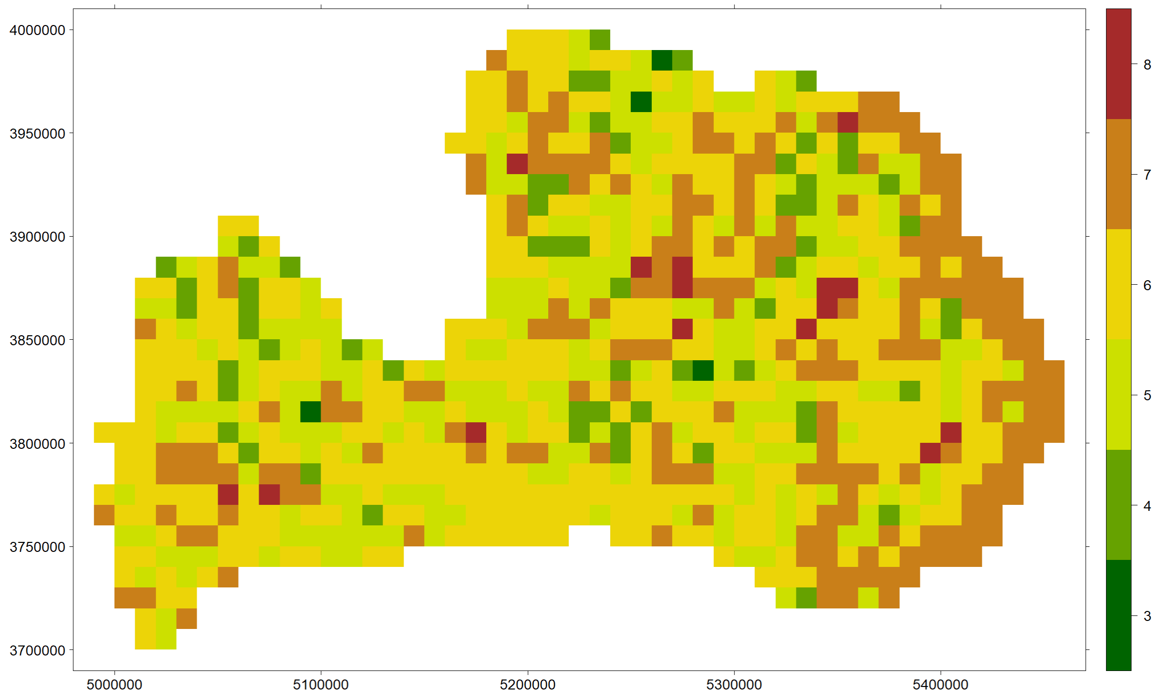
Predicted number of PPPs in APIStrip sample in Italy in September



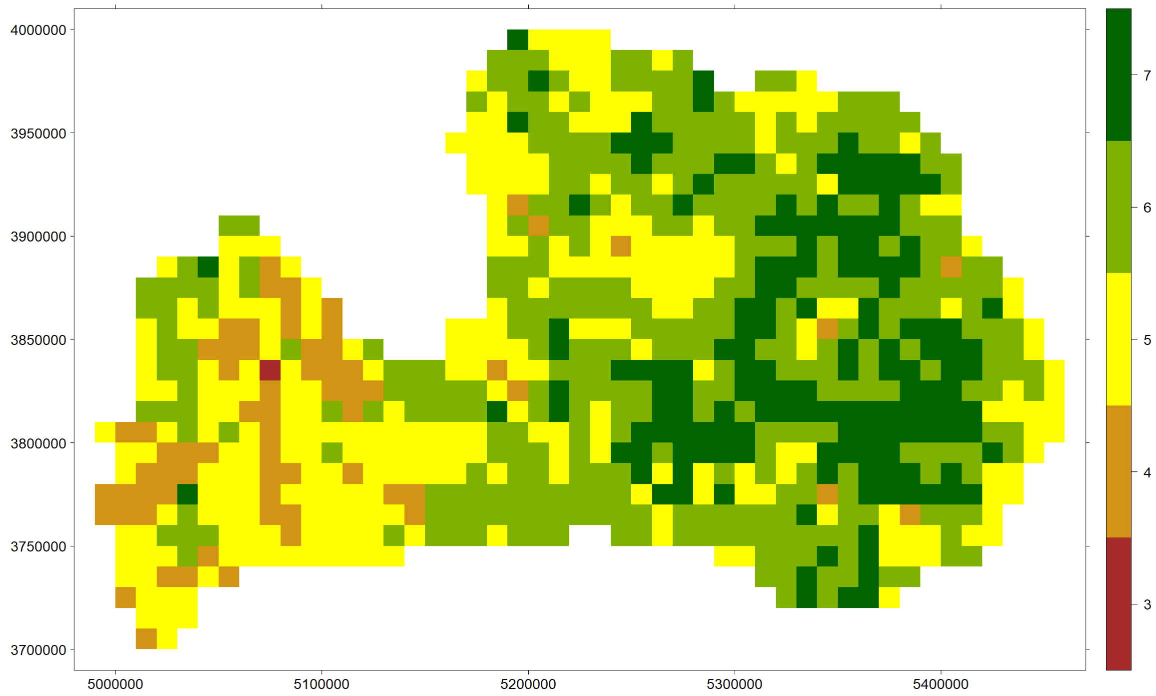
Predicted number of plant families in pollen samples in Latvia in May



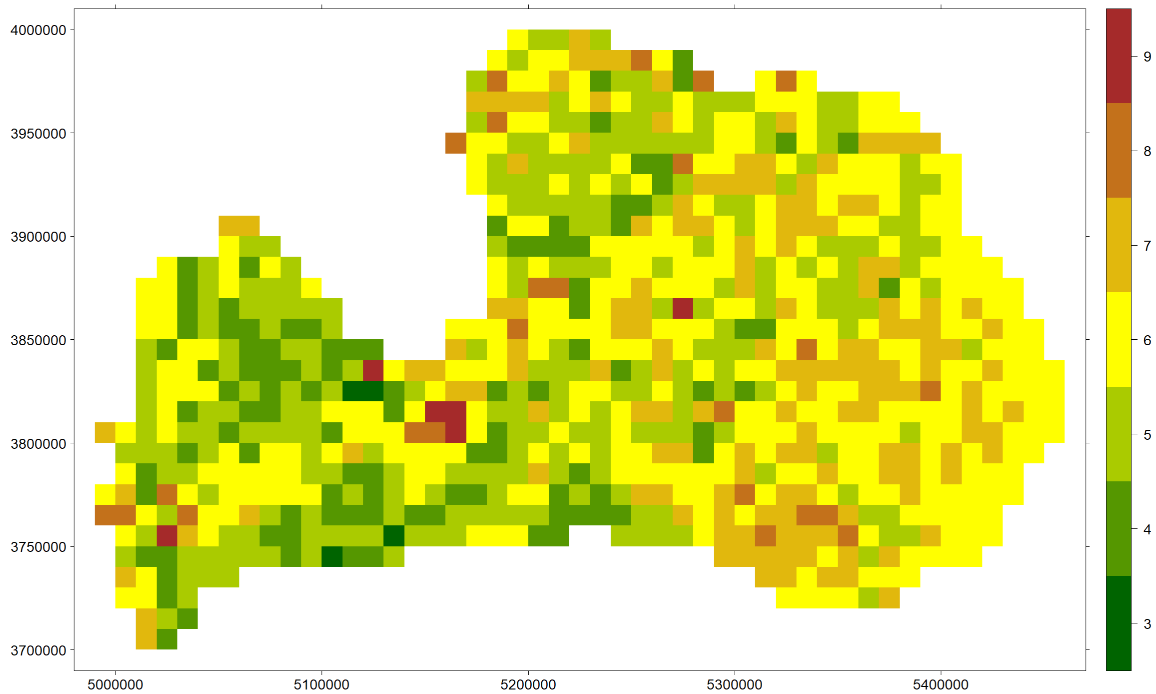
Predicted number of PPPs in APIStrip sample in Latvia in May

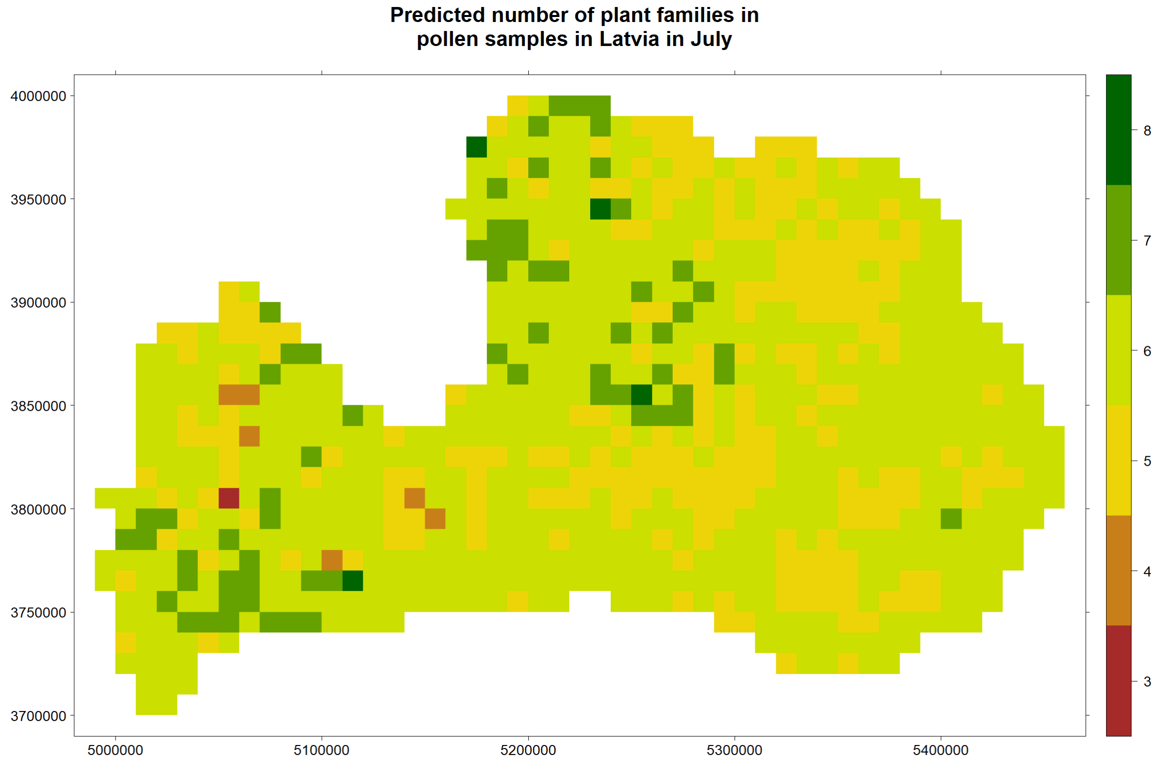


Predicted number of plant families in pollen samples in Latvia in June

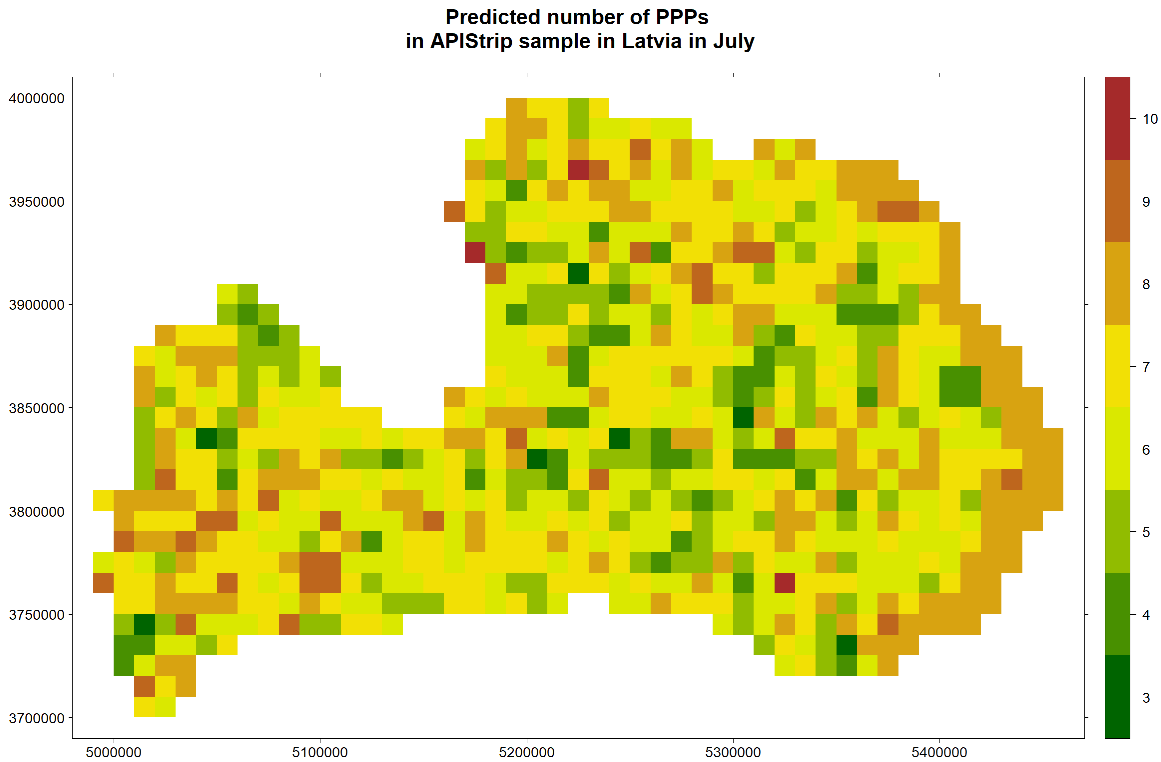


Predicted number of PPPs in APIStrip sample in Latvia in June



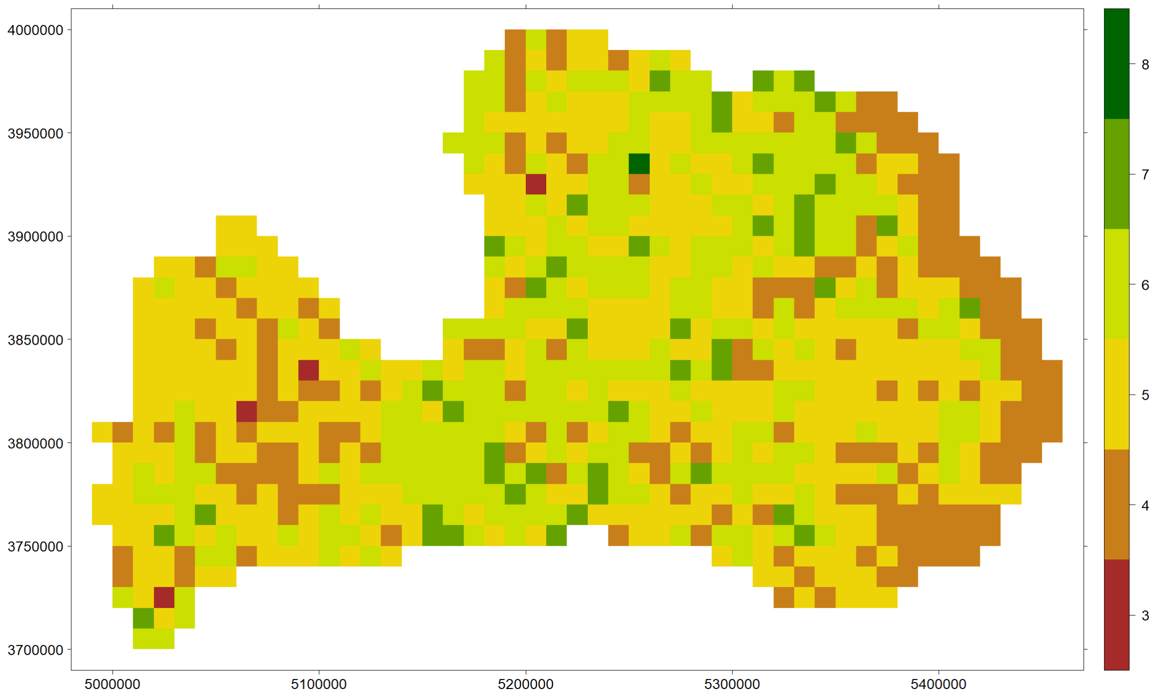


Predicted number of plant families in
pollen samples in Latvia in July
4000000
3950000
3900000
3850000
3800000
3750000
3700000
5000000
5100000
5200000
5300000
5400000
8
7
6
5
4
3

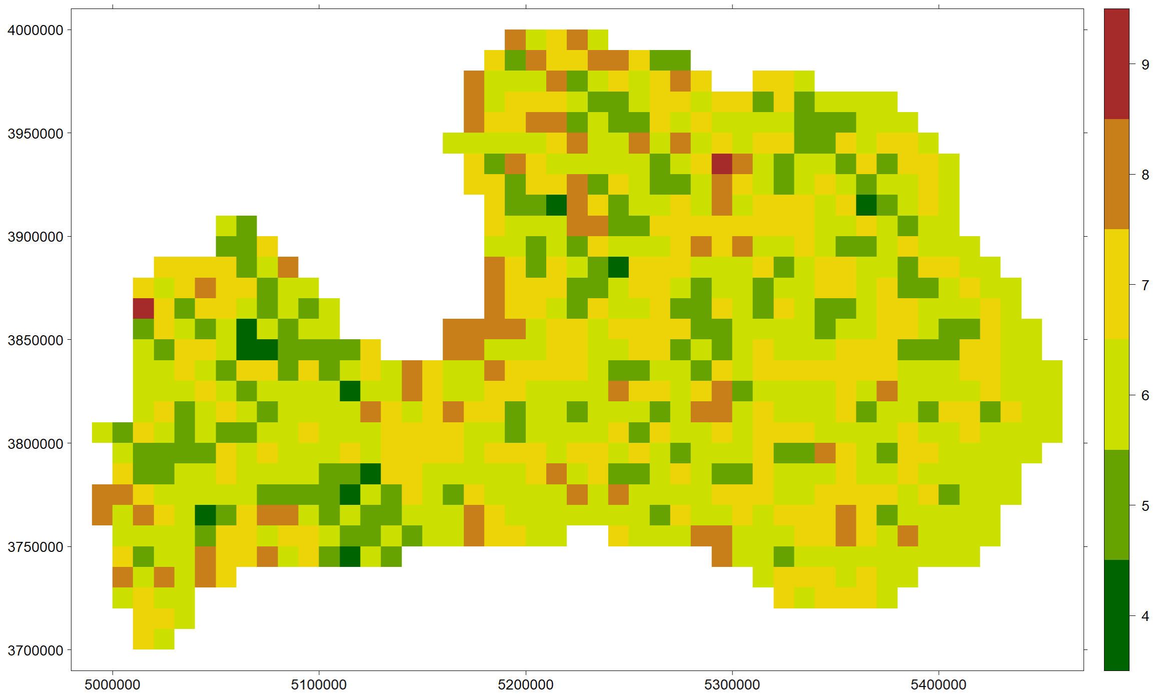


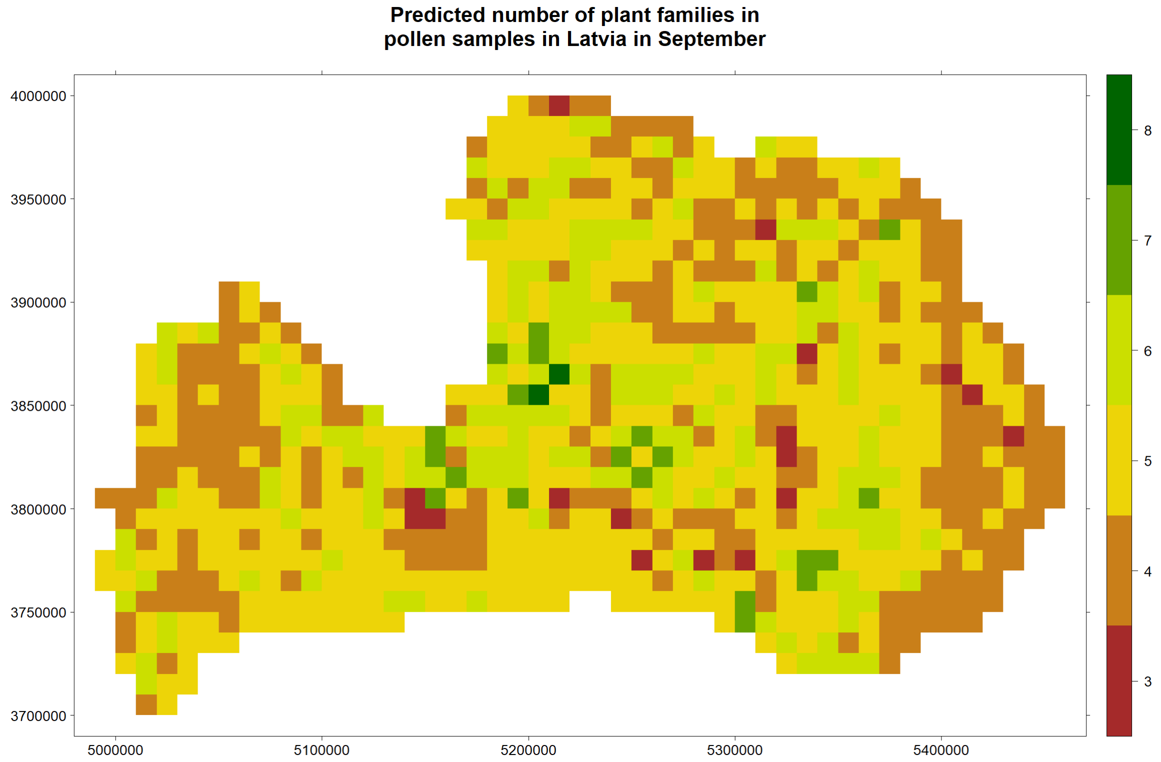
Predicted number of PPPs
in APIStrip sample in Latvia in July
4000000
3950000
3900000
3850000
3800000
3750000
3700000
5000000
5100000
5200000
5300000
5400000
10
9
8
7
6
5
4
3

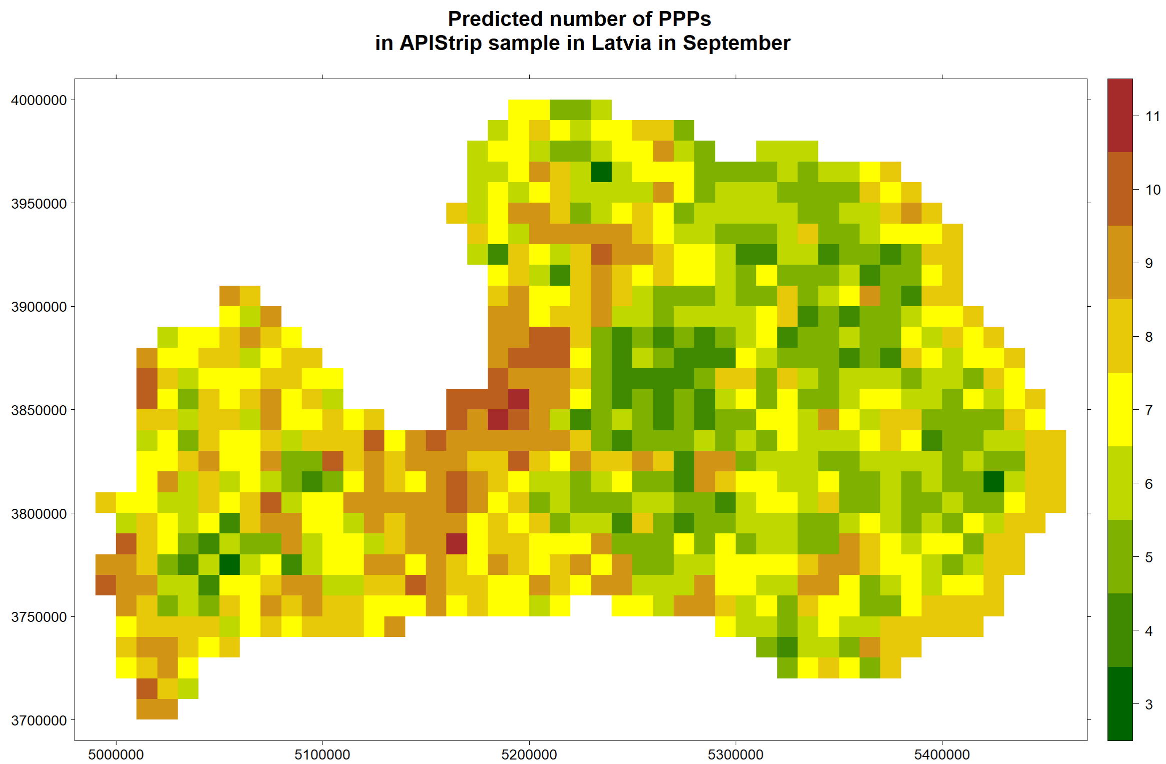
**Predicted number of plant families in
pollen samples in Latvia in August**

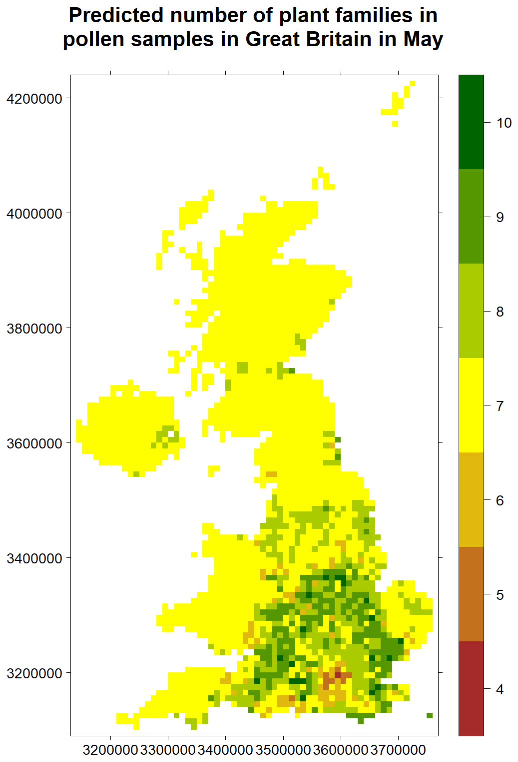


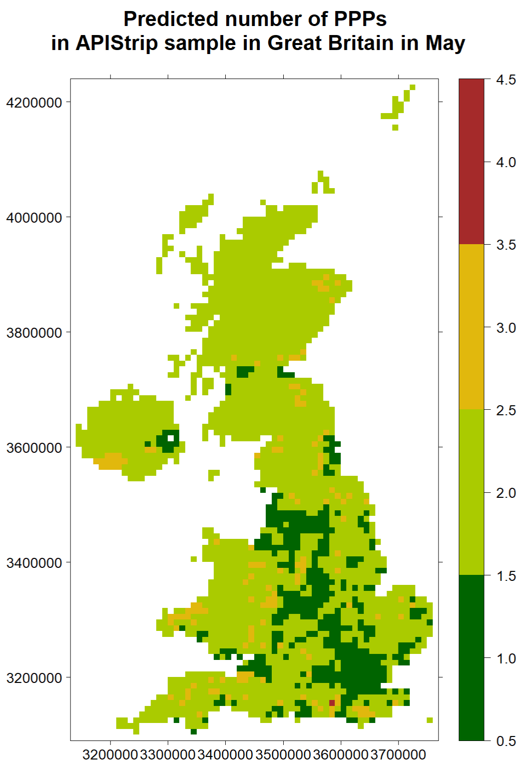
**Predicted number of PPPs
in APIStrip sample in Latvia in August**



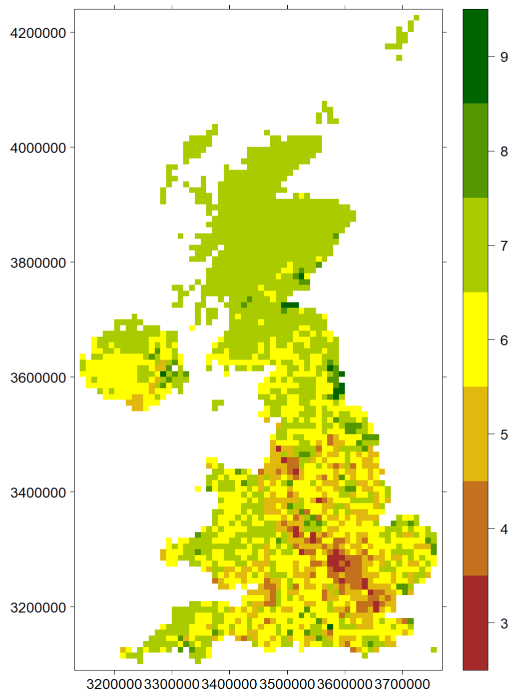

Predicted number of plant families in
pollen samples in Latvia in September
4000000
3950000
3900000
3850000
3800000
3750000
3700000
5000000
5100000
5200000
5300000
5400000
8
7
6
5
4
3


Predicted number of PPPs
in APIStrip sample in Latvia in September
4000000
3950000
3900000
3850000
3800000
3750000
3700000
5000000
5100000
5200000
5300000
5400000
11
10
9
8
7
6
5
4
3

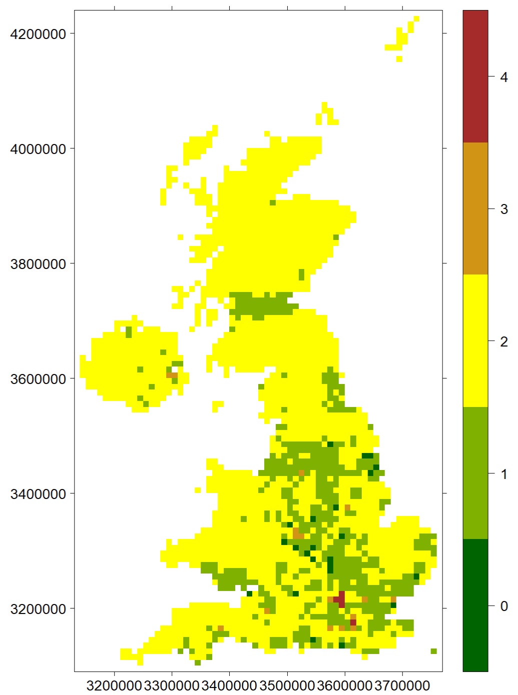

Predicted number of plant families in
pollen samples in Great Britain in May
4200000
4000000
3800000
3600000
3400000
3200000
3200000 3300000 3400000 3500000 3600000 3700000
10
9
8
7
6
5
4


Predicted number of PPPs
in APIStrip sample in Great Britain in May
4200000
4000000
3800000
3600000
3400000
3200000
3200000 3300000 3400000 3500000 3600000 3700000
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5

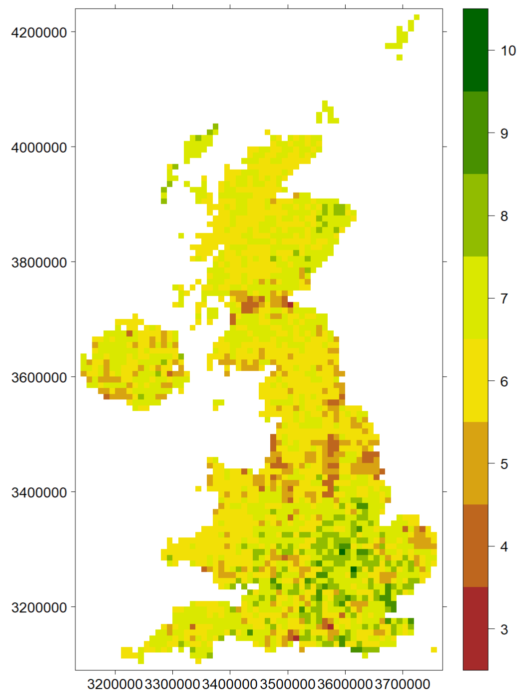
Predicted number of plant families in pollen samples in Great Britain in June



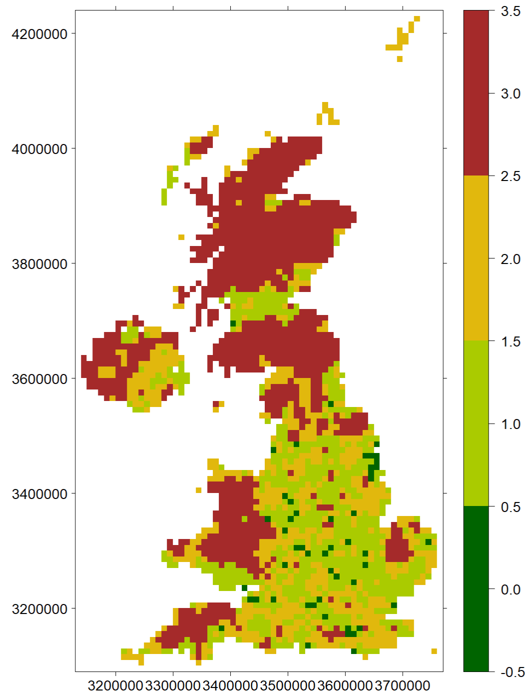
Predicted number of PPPs in APIStrip sample in Great Britain in June



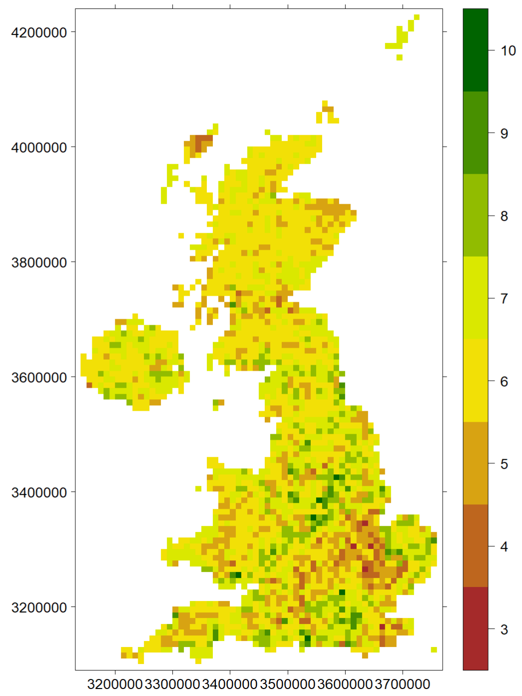
Predicted number of plant families in
pollen samples in Great Britain in July



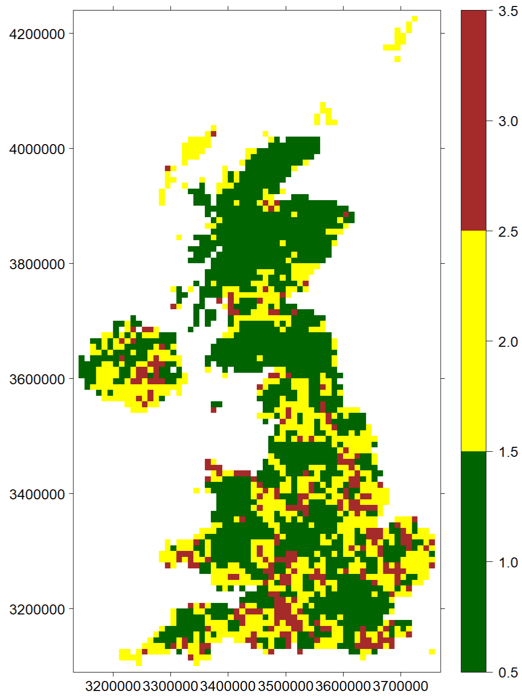
Predicted number of PPPs
in APIStrip sample in Great Britain in July



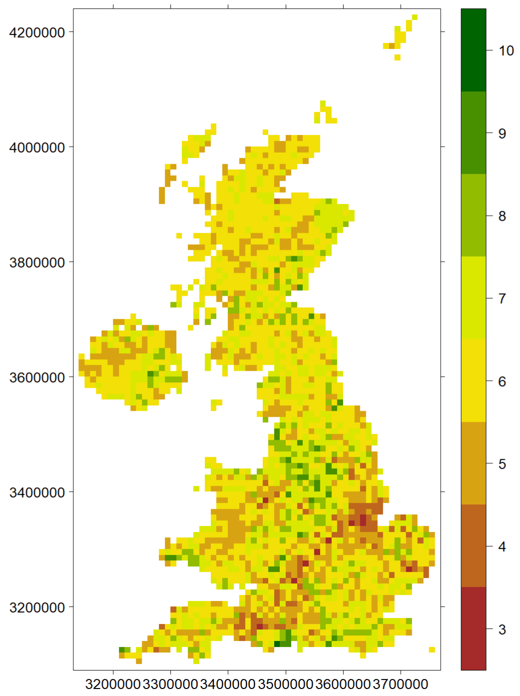
Predicted number of plant families in pollen samples in Great Britain in August



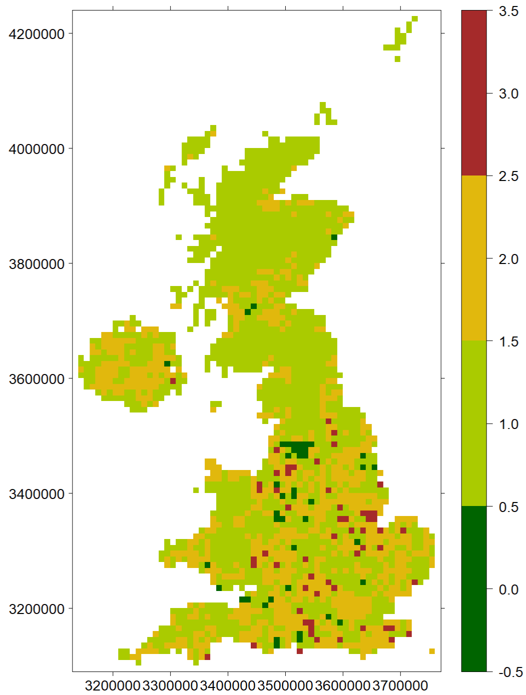
Predicted number of PPPs in APIStrip sample in Great Britain in August



Predicted number of plant families in pollen samples in Great Britain in September



Predicted number of PPPs in APIStrip sample in Great Britain in September



Wageningen Environmental Research
P.O. Box 47
6700 AA Wageningen
The Netherlands
T +31 (0)317 48 07 00
www.wur.nl/environmental-research

Wageningen Environmental Research
Report
ISSN 1566-7197

The mission of Wageningen University & Research is "To explore the potential of nature to improve the quality of life". Under the banner Wageningen University & Research, Wageningen University and the specialised research institutes of the Wageningen Research Foundation have joined forces in contributing to finding solutions to important questions in the domain of healthy food and living environment. With its roughly 30 branches, 6,800 employees (6,000 fte) and 12,900 students, Wageningen University & Research is one of the leading organisations in its domain. The unique Wageningen approach lies in its integrated approach to issues and the collaboration between different disciplines.