

INTERPRETING THE EBP VIEWER MAPS AND GRAPHS

The following instructions have been written in order to help adequately interpret the different maps and graphs shown in the EBP viewer.

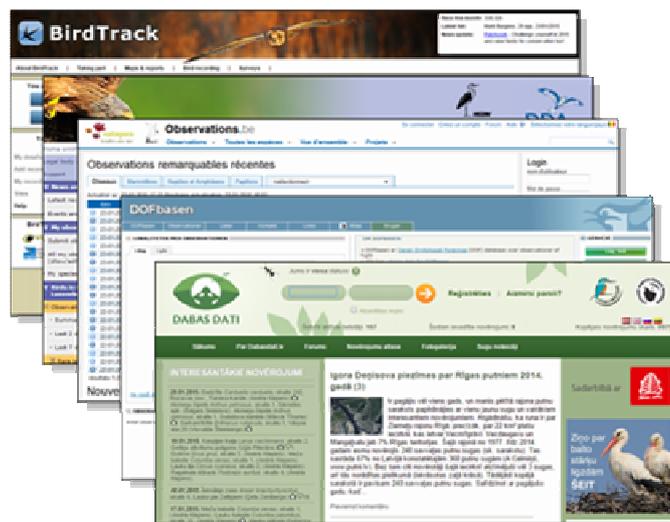
We start by describing the kind of data that is collected by the local online portals contributing to the EBP and the way this data is checked and validated. Then we describe the different maps and graphs depicted in the EBP viewer and give some overall guidelines to help interpret them correctly.

Table of content

| | |
|--|----|
| The data behind the maps..... | 1 |
| Quality checks and overall reliability of the data | 2 |
| Overall considerations on data accuracy and interpretation | 5 |
| The EBP viewer maps and graphs..... | 5 |
| Understanding presence/absence..... | 9 |
| Interpreting occurrence patterns | 10 |
| Some considerations on "all years" maps | 11 |
| Interpreting corrected regional occurrence (CRO) maps..... | 13 |
| Cautionary note about LIVE maps | 15 |

The data behind the maps

The EBP project combines the data from nearly 40 national/regional online portals operating in Europe and, therefore, to understand the maps and graphs shown in the EBP viewer it is very important to know, in the first place, which is the kind of data that is collected by these portals.



The EBP project combines data from the different national/regional online portals operating in Europe

Unlike more traditional monitoring projects, which focus on structured data collection based on standardised recording protocols at a sample of study sites, the online bird data gathering portals aim mainly to obtain data from the relatively unstructured but intensive and widespread activities of birdwatchers.

Online bird recording portals collect data on all bird species year-round thanks to the collaboration of large numbers of volunteer birdwatchers. Data collection protocols are kept reasonably simple while the feedback given back to the participants in the form of maps, lists and tables is, in general, very comprehensive. This attracts maximum participation and ensures that observers use the portal as a kind of online notebook of all their birdwatching activities, promoting widespread data gathering throughout the year.

The basic data collected includes species, date, location, and whether all species detected were reported (i.e. do the data comprise a complete list). Casual records are those that do not form part of complete lists and only provide presence information. In contrast, complete lists contain much more valuable presence/absence data (strictly detection/non detection data). Many observations also include the number of individuals of each species observed (counts) and, in the case of complete lists, some additional contextual data about the whole observational event (e.g. start and end time, distance travelled/area covered). Despite being a very simple protocol, therefore, complete lists not only provide presence/absence data but usually also information about recording effort. Most portals also allow observers to provide additional detailed information on a given observation (e.g. age and sex, breeding evidence, comments) but always in an optional manner. Several portals also include data from standard monitoring schemes (e.g. common breeding bird census, winter waterbird counts), but in general, this is only a small proportion of the whole dataset.

Quality checks and overall reliability of the data

All online portals have a data verification process in place. Usually, a combination of automated data filters and a network of local expert reviewers is used. Data filters are most often used to flag the records which are unusual because of the species (e.g. a vagrant is reported), timing (e.g. a summer visitor observed in mid-winter) or location (e.g. a bird observed well outside its known range for a given time period), but some portals can also flag records when observed in a unusual altitude or because the number of birds reported is too high. These filters are mostly set up once, but can be updated from time to time using the new data collected by the system or through additional expert advice.

To further ensure their effectiveness, filters are often fully embedded in the data entry process, warning the observer about potential errors as they are about to be entered (e.g. showing a warning message or preventing data from being submitted before it is double checked). Most errors are due to typing errors (e.g. selecting the wrong species

from a dropdown menu or typing the wrong number when indicating the count or the date), therefore, if the observer is conveniently warned many of them are already corrected before being submitted.

Records that end up being submitted and which are flagged are subsequently reviewed by a network of local experts who will be responsible for making a final decision regarding the record's validity. The network of local experts is usually formed by a combination of experts on certain regions or areas and experts on certain groups of species. These experts not only validate the data flagged automatically by the internal filters, they also routinely check the entire database using the different tools and outputs provided by each system (tables, maps, graphs) to flag and review additional dubious records. Such networks of experts are usually big and active enough to detect most of the obvious errors in a matter of hours. In the Ornitho system, for instance, it is estimated that 80% of the errors are detected in less than 72 hours. The reviewers usually have special login credentials that allow them not only to see all flagged records but also to communicate with the observers about their records and to check or intervene in other validation processes.

In the case of national or regional rarities, most portals work closely with the corresponding Regional or National Rarities Committees to ensure that the corresponding records also go through the usual validation process for rarities. In fact, some Regional or National Rarities Committees use the online portals as the sole framework where the whole process related to rarities (submission and verification) is done.

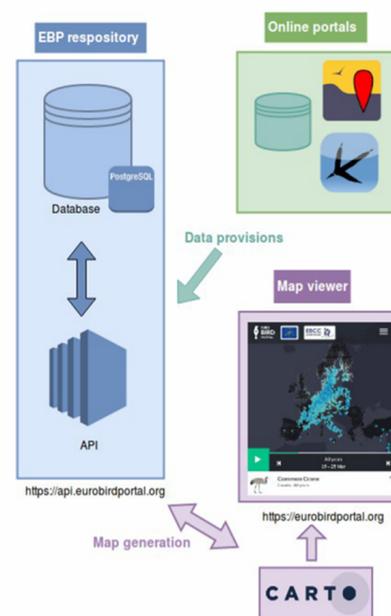
Once a final decision is taken by the corresponding committee or team of reviewers, records in process of validation are either unflagged or marked as unconfirmed (or not-accepted). Unconfirmed records are retained (e.g. for the own use and interest of the individual observer and to be available if future reassessment is required) but, like flagged ones, not shown in public outputs (e.g. listings, graphs, etc.) nor used in any kind of analysis.

Data combination and validations at the EBP level

The EBP project combines the data collected by the local online bird portals using a common data sharing standard and, subsequently, this data is aggregated by week and 30x30 km square (based on the European Environment Agency reference grid ETRS89-LAEA) to be shown in the EBP viewer. For each week and 30x30km square, the data summarizes information on the number of observations of each species and the maximum number of counted birds.

Currently, most of this data handling process is done automatically. The central EBP database repository receives more than 99% of the whole data automatically and on a daily basis. Then, also on a daily basis, the repository generates the tables used by the data visualization platform CARTO to render the different map and graphic visualizations in the EBP viewer.

The local portals only submit data that has not been flagged as potentially erroneous (see above). However, since all the EBP content is update on daily basis and with data up to the previous day some erroneous records or counts may not have been detected yet. Once corrected, the system automatically deals with these errors in subsequent data updates.



For the 105 species shown in the EBP viewer, additional data quality checks are conducted using a specific online tool that allows partners to delete any wrong record that still may be apparent in the viewer maps. This tool works in real-time. Therefore, anytime that a wrong record is deleted all affected maps are automatically corrected.

Note that records of rare species which don't seem unreliable but which are still under formal revision by the corresponding Regional or National Rarities Committees are also included in the maps. Certainly, some of these records could end up being not accepted by the respective committee. However, the reviewing protocols differ among committees and partners and in some cases the process can take several months and, therefore, not including of these observations could artificially alter the seasonal patterns of distribution of these species. In any case, these considerations should always be taken into account when interpreting the maps of these particular species.

Overall considerations on data accuracy and interpretation

As already detailed above, each online bird recording scheme submitting data to the project has its own data validation protocols to deal with potentially erroneous observations. Moreover, additional data quality checks are conducted at the EBP level. However, overall, the EBP demo viewer synthesises more than 320 million bird records in about 55,000 weekly maps. These are too large numbers to completely discard the possibility that some erroneous observations could still be present in some maps. This can be particularly the case in LIVE maps since the data shown can be so recent (e.g. from the previous day) that some erroneous records or counts may not have been detected yet. Once corrected, the system automatically deals with these errors in subsequent data updates. However, the possibility that some erroneous observations could remain unnoticed cannot be completely discarded.

The possible existence of some erroneous records or counts, however, shouldn't distort the overall large-scale patterns of bird distribution shown in the viewer, which is the ultimate interest of the EBP project.

Rather than the possible presence of a few erroneous observations, it is the interpretation of the maps itself that requires an important cautionary note. Online bird recording portals mainly obtain data from the relatively unstructured but intensive and widespread activities of birdwatchers. This means that surveying effort can differ largely from one area to another (even very close ones) or from one week to the next. It is even more relevant that the willingness to report a given species can vary significantly among sites, weeks and the rarity of the species (observers tend to report more often a given species in areas or periods of the year where it is less usual or common). Accordingly, a lack of records of a given species in a given surveyed square and week does not automatically mean that the species was not observed there (it may have been observed but not reported).

In many cases, the advice of an expert and further analyses may be required to draw sound conclusions from these maps. The EBP, therefore, can't be made responsible of the inadequate use of the information contained in the viewer or its interpretation by third parties.

The EBP viewer maps and graphs

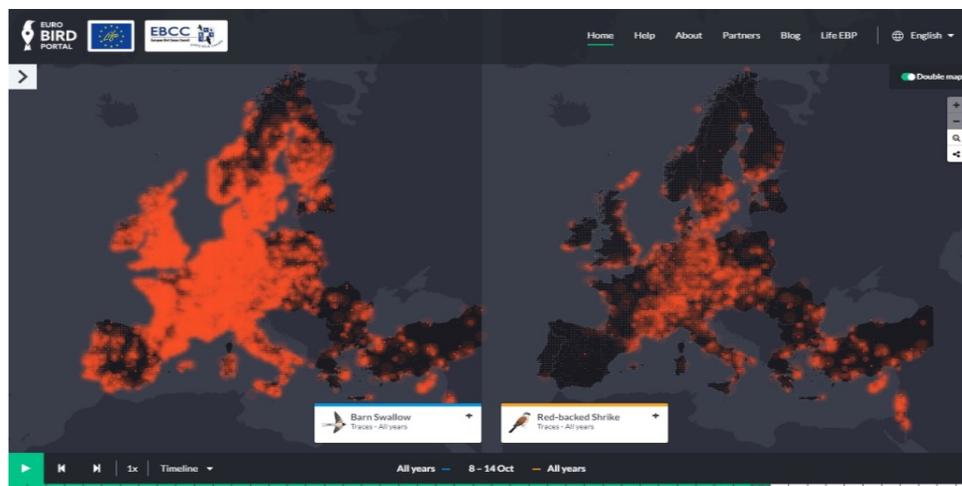
The EBP viewer shows animated maps depicting the week-by-week distributional patterns of 105 species from 2010 up to now at a resolution of 30x30 km. Each animated map depicts 12 continuous months of data (52 weeks) to better visualise the overall year-round seasonal patterns of bird distribution. The user can select LIVE maps (to see that last 52 weeks of data), an entire natural year (e.g. 2018) or an interannual cycle (e.g. July 2017 to June 2018). LIVE maps are the best to see the most recent bird movements (data up to the previous day is provided), natural years are better suited

to assess the overall seasonal movements of long distance migrants (i.e. European summer visitors) and interannual cycles are better for partial migrants. An additional option, "All years", allows the visualization of the overall seasonal patterns of bird distribution across the different years or interannual cycles by combining all the data available since 2010.

There are a total of five types of distributional maps available for each species (see below) and, moreover, four climatic variables can also be selected under the section "Contextual data".

Two visualization modes are available: one depicting a single zoomable map and the other one showing a double map (default option). The later allows two animated maps to be shown simultaneously for direct comparison which, given the number of species, types of maps and years available, means that more than 50 million different map combinations are available to choose from.

To easily learn how to use the EBP viewer and its different functionalities check this [users video guide](#).



The EBP viewer allows easy comparison of the week-by-week distributional patterns of different species or different years of the same species.

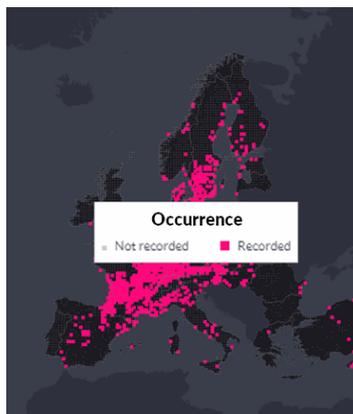
The species maps shown in the EBP demo viewer are based on 320 million bird records collected since 2010 by the on-line bird recording portals run by the project partners. The EBP central database repository aggregates these data by week and 30x30 km square (based on the European Environment Agency reference grid ETRS89-LAEA) summarizing information on the number of observations of each species, the number of counted birds and the recording effort (number of complete lists and total number of records and observers). Four of the species maps (occurrence, traces, counts and

phenology) reflect, in different ways, the raw information contained in the aggregated data, while the fifth one (corrected regional occurrence) uses various analytical procedures to account for heterogeneity in observational effort and species reporting rates.

Thanks to the implementation of an automatic data flow system, most EBP partners submit their data to the EBP central database repository on a daily basis (99%) and, therefore, all maps shown in the viewer are also updated on a daily basis using the most up to date information.

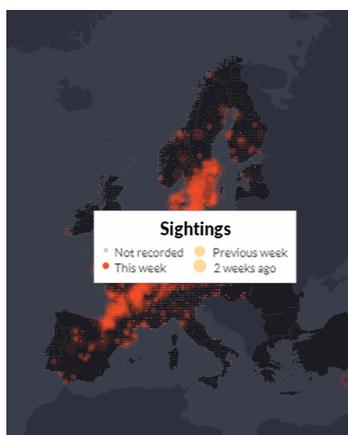
The maps depicting climatic variables have been prepared using the E-OBS high-resolution gridded dataset v10.0. Like species maps, the original E-OBS dataset is aggregated by week and 30x30 km square.

In all cases, only those 30x30 km squares including terrestrial areas are shown in the maps.



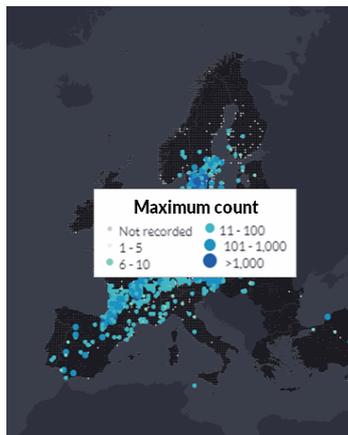
Occurrence map

These maps show the 30x30 km squares where the selected species was recorded or not each week. Note that not all species detected end up being reported and, thus, the absence of records in a given square does not automatically mean that the species was not observed there.



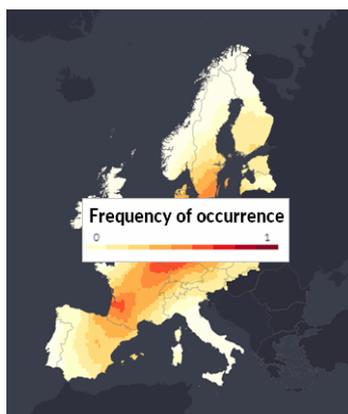
Traces map

These maps show the 30x30 km squares where the selected species has been recorded during each week but also during the two previous ones, enhancing the visualization of rapid temporal changes in distribution and their “traces” over time. The orange symbol indicating that the species was recorded in a given week increases in size and transparency the older the record (i.e. up to 2 weeks ago). Note that not all species detected end up being reported and, thus, the absence of records in a given square does not automatically mean that the species was not observed there.



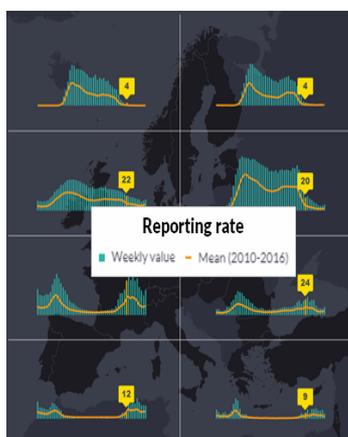
Counts map

These maps show the maximum count recorded for the selected species in each 30x30 km square and week (in "All years" maps, the average of the maximum count of each year is shown). Note that the original counts used to obtain these figures mostly refer to casual counts or rough estimates of the number of birds detected in a given site and date. Only rarely they refer to formal censuses or exhaustive counts. These maps, therefore, only show a rough approximation of the real variability in bird numbers across space and time. Also take into account that not all species detected end up being reported and not all detected species are counted; thus the absence of records in a given square does not automatically mean that the species was not observed there.



Corrected regional occurrence (CRO) map

These maps make use of a complex set of spatial and temporal aggregation and smoothing procedures to account for differences in observational effort and reporting activity of the observers (for more details see below). They are specifically intended to help visualize large-scale temporal changes in bird distributions and thus they should not be interpreted at a local scale. For each 30x30 km square and week these maps depict the estimated regional frequency of occurrence of the selected species. Modelling the spatial and temporal dynamics of bird distributions is one of the main but challenging objectives of the EBP project. This work is underway and, therefore, it has to be stressed that the maps shown here are very preliminary.

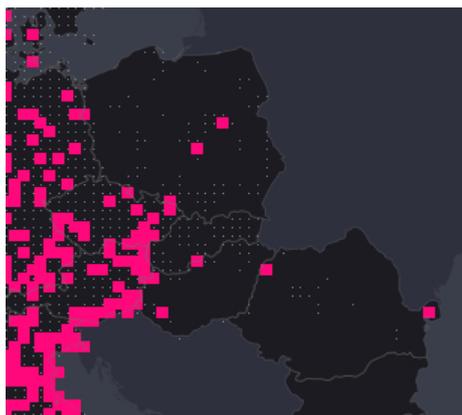


Phenological map

These maps depict the phenology of the selected species in eight different geographical sectors according to the percentage of 30x30 km squares where the species was recorded in each sector and week in relation to the total number of squares known to have been surveyed in each given sector and week. The weekly value of the selected year ("Weekly value") and the mean for all years available are shown.

Understanding presence/absence

It is very important to understand how presence/absence information is depicted in the EBP maps. All these maps only depict 30x30 km squares that have been surveyed in a given week (i.e. those where at least some species were recorded). The larger symbols (pink, orange or blue depending on the map type) indicate the existence of records of the selected species in a surveyed square and week while the small grey symbols indicate the lack of them. Therefore, areas not covered at all (no species recorded) in any given week can easily be identified since they appear black in the maps (no symbol is shown). However, note that though, in general, the absence of records of a given species in a surveyed square and week does denote a lack of observations of the species it cannot be ruled out that the species was observed but not reported (remember that, as indicated above, most of the data are collected as casual records and, therefore, not all species observed are necessarily reported, unlike in complete lists).



2015



2016

Two close-up snapshots of the EBP viewer showing the occurrence of the Garganey in early March (5-11 March) in two different years. The symbols identify the 30x30 surveyed squares (i.e. those where at least some species were recorded in the given week), the large pink square signals the ones where the selected species (the Garganey in this case) was reported and the small grey square the ones where it was not. The areas shown in black (no symbol depicted) signal the 30x30 km squares where no single observation (of any species) was reported. Note how the level of coverage in that particular week has changed from one year to the next, particularly in Poland.

Interpreting occurrence patterns

There are several important issues to take into account when interpreting the phenological graphs and all maps based on occurrence shown in the EBP viewer.

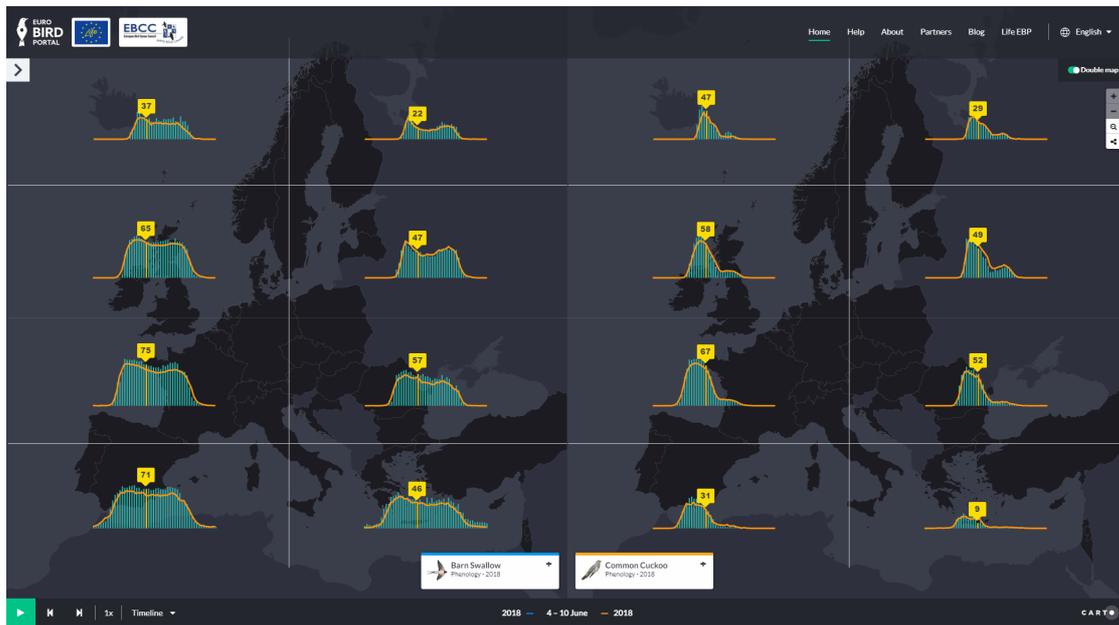
First of all, we have to take into account that a large part of the EBP data is still collected as casual records and, therefore:

- 1) We can not be completely sure about the real number of squares that have been surveyed. The problem is that in the case of casual records we only know that a given square has been surveyed if some of the 105 EBP target species were reported.
- 2) In general, the absence of records of a given species in a surveyed square and week does denote a lack of observations of the species but cannot be ruled out that the species was observed but not reported. Again, as most of the data are collected as casual records not all species observed are necessarily reported.

Since the relative occurrence values used in the EBP viewer graphs are based on a rather coarse spatiotemporal unit (30x30km square and week), we consider that these two objections probably have a rather overall limited effect, nonetheless, it is important to take always them into account when interpreting the results.

There is, however, one issue that is of paramount relevance when interpreting occurrence patterns irrespective of whether data is originating from casual records or complete lists: **detectability**.

Even where data are in the form of complete lists, or can be coerced into a similar format, occurrence and detection probability are confounded. An obvious example is that migratory songbirds sing very actively when they arrive on the breeding grounds. However as the summer progresses the proportion of complete lists recording such species declines, not through any reduction in abundance but simple because males are more engaged in other aspects of breeding and hence are less conspicuous. This drop off in detection is not seen in species such as Barn Swallows that are mainly detected by sight. Dispersal of waterfowl or waders across the breeding grounds could result in similar effects in some areas. On arrival many birds may be concentrated on water bodies where they are easily observed and that are well covered by observers. As the breeding season progresses and birds spread out onto breeding territories they may become less easy to detect.



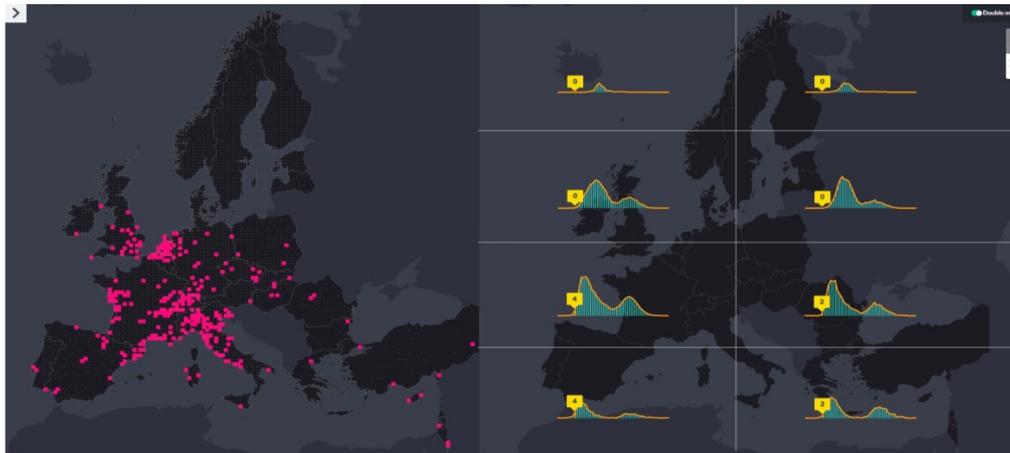
Common Cuckoo (right) detectability drops markedly once the singing activity finishes while Barn Swallows (left), which are mainly detected by sight, are easily observed from arrival to departure from Europe. It's very important to note that the patterns shown in the EBP phenological graphs and occurrence maps are the result of a combination of occurrence and detection probabilities.

The most important point is that when interpreting EBP phenology graphs and occurrence maps we should be always very aware that they represent some unquantified combination of occurrence and detection probabilities.

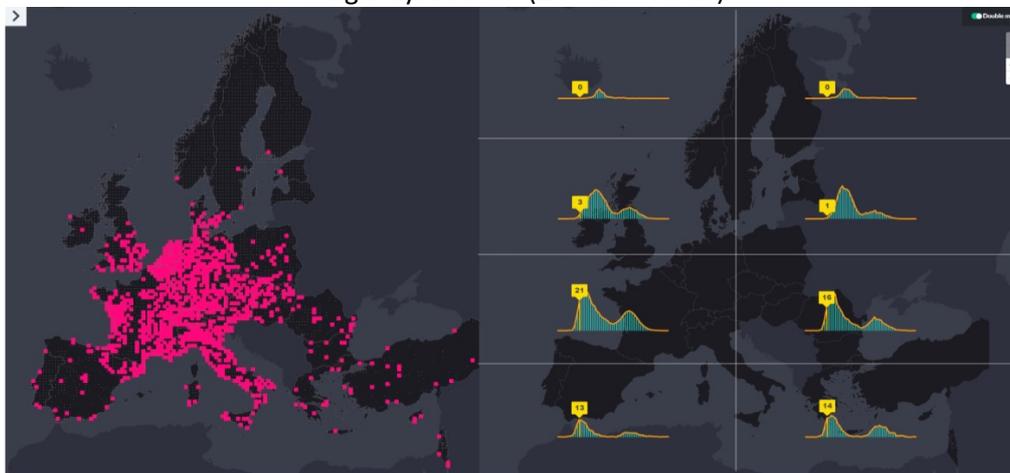
Some considerations on "all years" maps

The all years maps have the advantage of combining all the available information since 2010, which is particularly useful for areas where data or coverage is sparse. However, the combination of several years of data in occurrence maps can sometimes be misleading, for example suggesting that a given species is more common in a given area and time period than it is in reality. Imagine ten records in a given week widespread in different 30x30 km squares of the same country. These can be quite conspicuous in the maps, but taking into account that 7 years of data have been combined, the real frequency of the species in the area and time period would be, really, very low, close to zero. This, therefore, has to be taken into account when interpreting these maps.

Count maps, particularly in some gregarious species such as wildfowl, can be used to overcome this problem since quantity is complementing the coarse occurrence information. Another option is to use the phenology maps, which give an indication of the relative frequency of the occurrence of the species in different sectors of Europe.



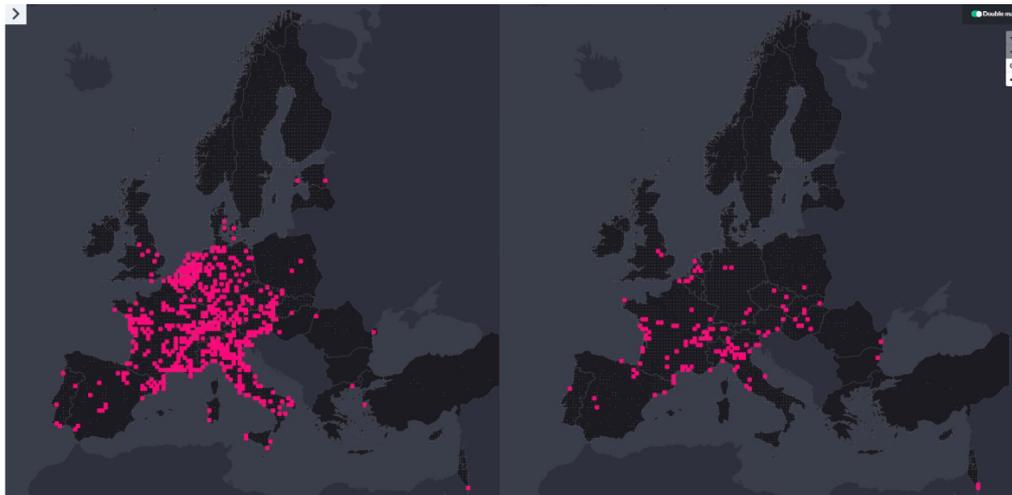
Garganey. Week 9 (26 Feb – 4 Mar)



Garganey. Week 11 (12 – 18 Mar)

The occurrence map of the Garganey shows quite a lot of occurrence dots already in late February-beginning of March in the British Isles and around the Netherlands (week 95) which may suggest that the arrival takes place in these areas more or less at the same time that in more southern European regions. However the phenological graph clearly indicates that the passage takes place earlier the further to the south (note how the peak of frequency changes from south to north).

Checking maps of different years can also help to understand the nature of a given occurrence pattern. For example, to see if the pattern is consistent enough across years, or mostly due to records collected in a very unusual year, or the consequence of the combination of small numbers of records collected across many years.



Snapshot of the occurrence map of the Garganey on week 10 (5 – 11 March) in 2015 (left) and 2016 (right). Note that in 2015 migration was much more advanced than in 2016.

Interpreting corrected regional occurrence (CRO) maps

The observational effort and the willingness to report a given species are unevenly distributed in space and time. Hence, it is important to produce maps that can account for such sources of variation and uncertainty. As a first step to produce these maps, we used a set of spatial and temporal aggregation and smoothing procedures. The objective was to create maps intended to help visualize large-scale spatiotemporal changes in bird distribution rather than details on bird distribution dynamics at fine spatial and temporal scales. First of all, data was aggregated using spatial and temporal rules to increase the sampling effort and the amount of bird occurrence information available when/where the local sampling effort was insufficient. In a second step, a spatial smoothing approach was applied to calculate the expected frequency of occurrence of each species over a predefined area in the neighbourhood.

1) Estimation of an adequate sampling effort and calculation of reporting rates based on a spatial and temporal aggregation procedure

The total number of observations reported for all the species during a given week and in a given 30x30 km square has been used as an indirect estimate of sampling effort. For each species, reporting rates were calculated in each square and week as the percentage of observations of the species relative to the total number of observations. To calculate consistent reporting rates, however, a minimum adequate sampling effort had to be used. To estimate such adequate sampling effort, reporting rates calculated in all 30x30 km squares and during the different weeks along the year were plotted against the total number of observations in those squares and those weeks. As expected, the variability in estimated reporting rates decreased with increasing sampling effort, i.e. with an increasing number of observations used to estimate it. The adequate sampling effort was defined as the total number of observations above

which the reporting rate became stable. Reporting rates were then estimated at different spatial and temporal resolutions depending on the sampling effort in each square and each week.

If the sampling effort was sufficient, reporting rates were calculated at fine spatial (30x30 km) and temporal (1 week) resolutions. If the total number of observations in a focal 30x30 km square and week was below the adequate sampling effort, bird observations reported during the previous week and/or in squares around the focal square were also added. Such inclusion of additional bird observations increases the sampling effort but decreases the spatial and/or temporal resolution at which reporting rates are estimated. In order to find the best compromise between sampling effort and resolution, this aggregation procedure was carried out in an iterative way by using a gradually increasing number of squares in the neighbourhood until the adequate sampling effort was reached. The maximum number of squares used to increase the sample of bird observations from the neighbourhood was limited to a radius distance of 120 km. In poorly sampled regions or during certain periods of the year, the adequate sampling effort was not reached using this maximum radius distance and the reporting rates were impossible to estimate.

The reporting rates calculated across a range of spatial and temporal resolutions were then allocated to each 30x30 km square and to each week (i.e. the basic spatial and temporal resolutions of the EBP demo viewer) irrespective of the spatial and/or temporal level of aggregation used for the calculation.

Average level of the spatial and temporal aggregation used for each 30x30 km square and year (the lower, the better the resolution).

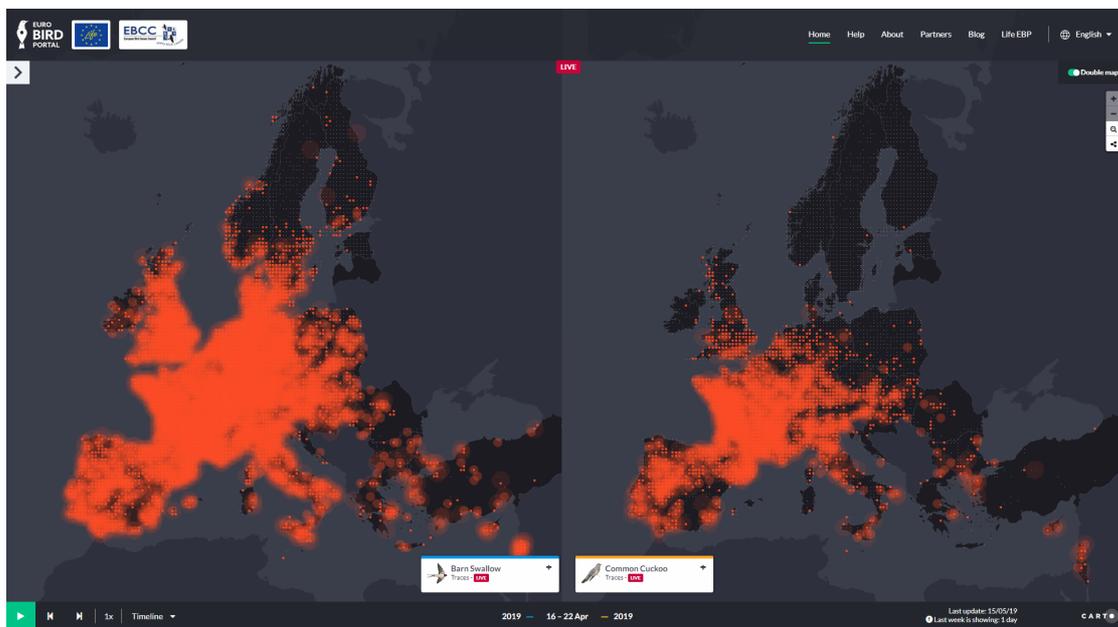
2) Spatial smoothing procedure

For each 30x30 km square and each week, the reporting rate estimated above was converted into a binary species presence/absence variable: 1 when reporting rate was higher than 0 (presence) and 0 when reporting rate was equal to 0 (absence). It is important to note that this conversion is based on reporting rates that were previously estimated for each square and each week in a consistent way according to a spatial and temporal aggregation procedure so that sampling effort was considered as adequate to estimate species presences and absences. Based on this binary information, a “corrected regional species frequency of occurrence” value was calculated for each 30x30 km square and each week as the proportion of squares where the species was estimated to be present within a 250 km radius distance (100 km for terns, shorebirds, ducks, geese and swans). The value was only calculated when there were at least 50% of squares with estimated reporting rates within the 250 km radius distance. Note that in the viewer maps areas where estimated occurrences are more unreliable due to data scarcity (Low Resolution areas) can be shaded off by turning on the corresponding layer.

To summarise, corrected regional occurrence (CRO) maps make use of a complex set of spatial and temporal aggregation and smoothing procedures to account for differences in observational effort and reporting activity of the observers. They are specifically intended to help visualize large-scale temporal changes in bird distribution and thus they should not be interpreted at a local scale. It is very much a work in progress and, therefore, it has to be stressed that the maps are still very preliminary.

Cautionary note about LIVE maps

LIVE maps show data from the previous day back to one year ago. As in all other EBP maps data is grouped in 52 weeks. To be able to visualise distributional patterns in near real-time is one of the most exciting developments of the EBP project, however, it is important to take into account a few specific questions when interpreting these maps.



LIVE maps show data collected up to the previous day and, thus, their interpretation require some caution.

Despite most EBP data (99%) is updated daily some online portals still submit their data at longer intervals and, therefore, the areas covered by these portals can appear blank in the most recent weeks.

Note that the most recent week depicted in these maps, usually the current one, may not have finished yet and, thus, special attention must be paid when interpreting distributional patterns of such incomplete weeks.

Note also that, despite the different data validation protocols put in place, LIVE maps show so recent data that some erroneous records or counts may not have been detected yet. Once corrected, the system automatically deals with these errors in



subsequent data updates. However, the possibility that some erroneous observations could remain unnoticed cannot be completely discarded (see also above).