

Figure 4. A random point (red point) is chosen within the circle (gray) defining the occurrence. The occurrence is then assigned to the cell the point belongs to (red square). Left: an occurrence (<https://www.gbif.org/occurrence/2235280677>) is totally contained in one of the cells of the reference grid. Center: An occurrence (<https://www.gbif.org/occurrence/1569856810>) spreads over multiple cells. Right: an occurrence (<https://www.gbif.org/occurrence/2235279067>) with small uncertainty spreads over two cells.

The probability distribution is by definition uniform all over the circle so the probability that the random point falls in a specific cell is equal to the proportion of the circle's area covered by the cell. Geometrically it can be demonstrated that no cell has a higher probability to get the occurrence assigned than that one containing the center. However, this doesn't exclude the possibility that the occurrence would be assigned to another cell as shown in Figure. 4.

Step 4. Aggregate occurrences

Aggregating occurrences means *counting* how many occurrences of a specific taxon are in a specific cell and in a specific time interval. Using our example with occurrences of *Reynoutria*, where we decided to produce an occurrence cube at species and year level using a reference grid at 1km resolution, we have to count how many occurrences of *Reynoutria* are there within each year, cell and species. As the occurrence cubes can be used as input for modelling and risk assessment, we store the smallest geographic coordinate uncertainty of the occurrences assigned to a certain cell as value as well. Using a tabular structure (typical of R *data.frames* or pandas *DataFrames*), an occurrence cube would look like a table with as many columns as the sum of the number of dimensions (three) and the number of values (two). In Table 2 we show an excerpt from the example occurrence cube.

As mentioned in Step 1, defining the taxonomic granularity of the occurrence cube implies that occurrences linked to a taxon can come from multiple taxa such as synonyms or taxa with lower rank. For this reason, it can be informative to provide a taxonomic compendium of the occurrence cube as shown in Table. 3. The full occurrence cube and the taxonomic compendium are available on GitHub:

<https://github.com/trias-project/occurrence-cube-paper/tree/9426a29dc6f080920509aa295bd49dad0ea10d26/data/processed>.

Table 2. Tabular representation of the occurrence cube of *Reynoutria* in Belgium from 2000 to 2018. The first three columns represent the temporal, spatial and taxonomic dimensions respectively. Column *year* contains the year the occurrences took place, *eea_cell_code* the cell code from the EEA reference grid at 1km scale, *speciesKey* the GBIF identifier of the species (2889173: *Reynoutria japonica*, 4038485: *Reynoutria bohemica*, 2889088: *Reynoutria sachalinensis*). Taxonomic-spatial-temporal triplets with no occurrences are omitted.

year	eea_cell_code	speciesKey	n	min_coord_uncertainty
2000	1kmE3809N3113	2889173	1	700
2000	1kmE3809N3135	2889173	1	700
...
2006	1kmE3936N3071	2889173	1	49
2006	1kmE3947N3132	2889088	1	700
...
2010	1kmE3883N3121	4038485	1	700
2010	1kmE3884N3121	2889173	1	10
...
2014	1kmE3886N3121	2889173	51	10
2014	1kmE3886N3122	2889173	109	10
...
2018	1kmE4047N3067	2889173	1	2828

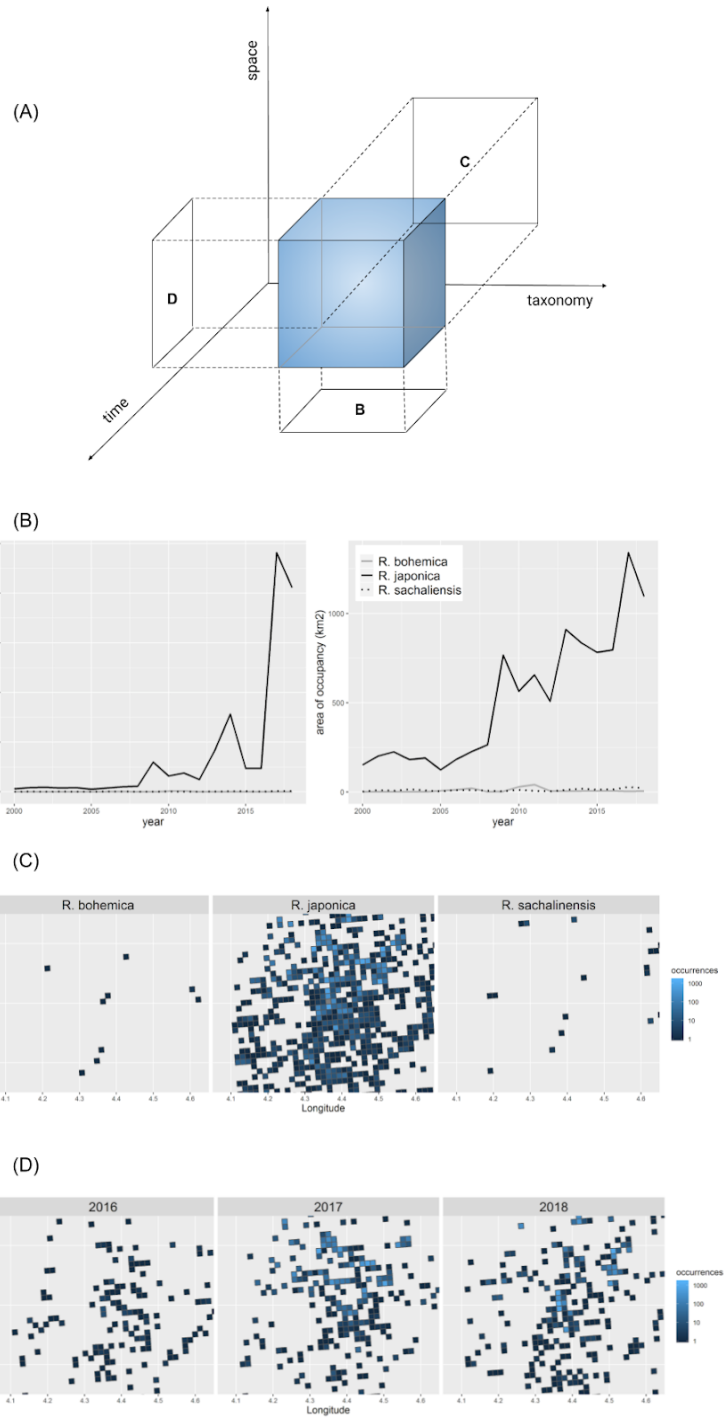
Table 3. Taxonomic compendium of the occurrence cube from GBIF occurrences of genus *Reynoutria*, in Belgium from 2000 to 2018. As shown in column *includes*, occurrences of a species can come from synonyms or infraspecific taxa, described by their GBIF taxon keys and scientific names.

speciesKey	species	includes
2889088	<i>Reynoutria sachalinensis</i>	5334293: <i>Fallopia sachalinensis</i> (Friedrich Schmidt Petrop.) Ronse Decraene 2889088: <i>Reynoutria sachalinensis</i> Nakai
2889173	<i>Reynoutria japonica</i>	5334357: <i>Fallopia japonica</i> (Houtt.) Ronse Decraene 2889173: <i>Reynoutria japonica</i> Houtt. 8361333: <i>Fallopia compacta</i> (Hook.fil.) G.H.Loos & P.Keil 7128523: <i>Fallopia japonica</i> var. <i>japonica</i>
4038485	<i>Reynoutria bohemica</i>	5652296: <i>Fallopia bohemica</i> (Chrték & Chrtková) J.P.Bailey 4038485: <i>Reynoutria bohemica</i> Chrték & Chrtková

The occurrence cube

The resulting occurrence cube can be projected on an orthogonal plane by aggregating along one of the three dimensions, as shown in Figure 5a.

Figure 5. (A) The occurrence cube and its projections on the temporal/taxonomic plane (B), the taxonomic/spatial plane (C) and the temporal/spatial plane (D). (B) Number of occurrences (left) and number of 1x1km cells or *area of occupancy* (right) of *Reynoutria bohemica*, *R. japonica* and *R. sachalinensis* per year. Both indicators can be seen as ways of projecting the occurrence cube on the temporal/taxonomic dimensions. (C) Projecting the occurrence cube along the taxonomic/spatial plane, thus getting a heatmap of the number of occurrences for each of the *Reynoutria sp.* in Belgium. The maps are zoomed for better readability. (D) Projecting the occurrence cube along the temporal/spatial plane, thus getting a heatmap of the number of occurrences of genus *Reynoutria* in Belgium for each year. The maps are zoomed for better readability.



Aggregating along the spatial dimensions means projecting the cube on the taxonomic and temporal dimensions. Counting the number of occurrences we get the abundance, counting the number of occupied cells we get the area of occupancy as shown in Figure 5b. Aggregating along the temporal dimension means projecting the cube on the spatial and taxonomic dimensions. Based on our example, it means counting how

many occurrences of each species of *Reynoutria* are in each cell during the entire period 2000 - 2018, as shown in Figure 5c. Similarly, aggregating along the taxonomic dimension, we project the cube on the spatial and temporal dimensions. We are then counting the number of occurrences of genus *Reynoutria* per cell and year as shown in Figure 5d.

We applied this methodology to larger taxonomic, spatial and temporal constraints as well. We created and published occurrence cubes at species level for Belgium and Italy (Oldoni et al., 2020a) and the occurrence cubes for non-native taxa in Belgium and Europe (Oldoni et al., 2020b). All these occurrence cubes are at year level and are based on EEA reference grids at 1km scale.

Notes

Data, scripts and figures are open and available on GitHub:

<https://github.com/trias-project/occurrence-cube-paper>

Acknowledgements

This work has been funded under the Belgian Science Policies Brain program (BelSPO BR/165/A1/TrIAS).

References

Blagoderov, V., Kitching, I.J., Livermore, L., Simonsen, T.J., & Smith, V.S. (2012). No specimen left behind: industrial scale digitization of natural history collections. *ZooKeys*, (209), 133.

<https://doi.org/10.3897/zookeys.209.3178>

Chandler, M., See, L., Copas, K., Bonde, A.M., López, B.C., Danielsen, F., Legind, J.K., Masinde, S., Miller-Rushing, A.J., Newman, G., Rosemartin, A., & Turak, E. (2017). Contribution of citizen science towards international biodiversity monitoring. *Biological Conservation*, 213, 280-294.

<https://doi.org/10.1016/j.biocon.2016.09.004>

Hardisty, A.R., Belbin, L., Hobern, D., McGeoch, M.A., Pirzl, R., Williams, K.J., & Kissling, W.D.D. (2018). Research infrastructure challenges in preparing essential biodiversity variables data products for alien invasive species. *Environmental Research Letters*. <https://doi.org/10.1088/1748-9326/aaf5db>

Hardisty, A.R., Michener, W.K., Agosti, D., García, E.A., Bastin, L., Belbin, L., Bowser, A., Buttigieg, P.L., Canhos, D.A.L., Egloff, W., De Giovanni, R., Figueira, R., Groom, Q., Guralnick, R.P., Hobern, D., Hugo, W., Koureas, D., Ji, L., Los W., Manuel, J., Manset, D., Poelen, J., Saarenmaa, H., Schigel, D., Uhler, P.F., & Kissling, W.D. (2019). The Bari Manifesto: An interoperability framework for essential biodiversity variables. *Ecological informatics*, 49, 22-31. <https://doi.org/10.1016/j.ecoinf.2018.11.003>

Kissling, W.D., Ahumada, J.A., Bowser, A., Fernandez, M., Fernández, N., García, E.A., Guralnick, R.P., Isaac, N.J.B., Kelling, S., Los, W., McRae, L., Mihoub, J.-B., Obst, M., Santamaria, M., Skidmore, A.K., Williams, K.J., Agosti, D., Amariles, D., Arvanitidis, C., Bastin, L., De Leo, F., Egloff, W., Elith, J., Hobern, D., Martin, D., Pereira, H.M., Pesole, G., Peterseil, J., Saarenmaa, H., Schigel, D., Schmeller, D.S., Segata, N., Turak, E., Uhlir, P.F., Wee, B., & Hardisty, A.R. (2018). Building essential biodiversity variables (EBVs) of species distribution and abundance at a global scale. *Biol Rev*, 93: 600-625. <https://doi.org/10.1111/brv.12359>

Oldoni, D., Groom, Q., Adriaens, T., Davis, A.J.S., Reyserhove, L., Strubbe, D., Vanderhoeven, S., & Desmet, P. (2020a). Occurrence cubes at species level for European countries (Version 20200205) [Data set]. Zenodo. <http://doi.org/10.5281/zenodo.3637911>

Oldoni, D., Groom, Q., Adriaens, T., Davis, A.J.S., Reyserhove, L., Strubbe, D., Vanderhoeven, S., & Desmet, P. (2020b). Occurrence cubes for non-native taxa in Belgium and Europe (Version 20200204) [Data set]. Zenodo. <http://doi.org/10.5281/zenodo.3635510>

Pereira, H.M., Ferrier, S., Walters, M., Geller, G.N., Jongman, R.H.G., Scholes, R.J., Bruford, M.W., Brummitt, N., Butchart, S.H.M., Cardoso, A.C., Coops, N.C., Dulloo, E., Faith, D.P., Freyhof, J., Gregory, R.D., Heip, C., Höft, R., Hurr, G., Jetz, W., Karp, D.S., McGeoch, M.A., Obura, D., Onoda, Y., Pettorelli, N., Reyers, B., Sayre, R., Scharlemann, J.P.W., Stuart, S.N., Turak, E., Walpole, M., & Wegmann, M. (2013). Essential biodiversity variables. *Science*, 339(6117), 277-278. <https://doi.org/10.1126/science.1229931>

Pocock, M.J., Roy, H.E., August, T., Kuria, A., Barasa, F., Bett, J., Githiru, M., Kairo, J., Kimani, J., Kinuthia, W., Kissui, B., Madindou, I., Mbogo, K., Mirembe, J., Mugo, P. Muniale, F.M., Njoroge, P., Njuguna, E.G., Izava, Olendo, M.I., Opige, M., Otieno, T.O., Ng'weno, C.C., Pallangyo, E., Thenya, T., Wanjiru, A., & Trevelyan, R. (2019). Developing the global potential of citizen science: Assessing opportunities that benefit people, society and the environment in East Africa. *Journal of applied ecology*, 56(2), 274-281. <https://doi.org/10.1111/1365-2664.13279>

Vanderhoeven, S., Adriaens, T., Desmet, P., Strubbe, D., Backeljau, T., Barbier, Y., Brosens, D., Cigar, J., Coupremagne, M., De Troch, R., Eggermont, H., Heughebaert, A., Hostens, K., Huybrechts, P., Jacquemart, A., Lens, L., Monty, A., Paquet, J., Prévot, C., Robertson, T., Termonia, P., Van De Kerchove, R., Van Hoey, G., Van Schaeybroeck, B., Vercayie, D., Verleye, T., Welby, S., & Groom, Q. (2017). Tracking Invasive Alien Species (TriAS): Building a data-driven framework to inform policy. *Research Ideas and Outcomes*, 3: e13414. <https://doi.org/10.3897/rio.3.e13414>