# SCIENTIFIC DATA (110110)

### **DATA DESCRIPTOR**

Received: 28 March 2019 Accepted: 16 August 2019

Published online: 08 October 2019

## **OPEN** ClimCKmap, a spatially, temporally and climatically explicit distribution database for the Italian fauna

Silvio Marta 101,2, Michele Brunetti, Gentile Francesco Ficetola 1, Fabio Stoch, Giovanni Amori<sup>5</sup>, Donatella Cesaroni<sup>6</sup>, Valerio Sbordoni<sup>6</sup> & Antonello Provenzale<sup>7</sup>

Understanding and counteracting biodiversity losses requires quantitative knowledge on species distribution and abundance across space and time, as well as integrated and interoperable information on climate conditions and climatic changes. In this paper we developed a new biodiversity-climate database for Italy, ClimCKmap, based on the critical analysis, quality estimation and subsequent integration of the CKmap database with several high-resolution climate datasets. The original database was quality-checked for errors in toponym, species name and dating; the retained records were georeferenced and their distribution polygonised via Voronoi tessellation. We then integrated the species distribution information with several high-resolution climatic datasets: average monthly minimum and maximum temperature and total monthly precipitation were reconstructed for each Voronoi cell and year. The resulting database contains 268,977 occurrence records from 8,445 binomials and 16,332 localities, dating between 1680 and 2006 CE. This dataset, fully available at https://doi. org/10.6084/m9.figshare.7906739.v4 and http://hdl.handle.net/21.11125/a91f85cb-befd-4e14-8e83-24f17c4a0491, represents the largest, fully quality-checked, spatially, temporally and climatically explicit distribution database ever assembled for the Italian fauna, now ready for scientific exploitation.

#### **Background & Summary**

Biological diversity, along with the associated ecosystem services and genetic resources, is essential to support the wealth of human society. In the last decades, international efforts have been made to collect, standardise and share a huge amount of distribution data, e.g. via the Global Biodiversity Information Facility (GBIF; https:// www.gbif.org). Despite the efforts to collect as much information as possible and compile comprehensive databases, several limitations are still present<sup>1</sup> and biodiversity inventories suffer from various types of bias<sup>2</sup>. Sources of bias include the unequal sampling effort due to site accessibility and attractiveness<sup>3,4</sup>, the introduction of data from non-systematic (i.e. opportunistic or occasional) surveys<sup>5</sup>, the inconsistency of sampling methodologies across space and time<sup>6</sup> and the taxonomical attractiveness (i.e. species or groups receiving particular attention for being rare, beautiful or somehow charismatic)7. Additionally, distribution data from biodiversity inventories are often affected by errors in spatial coordinates<sup>8</sup>, taxonomic misidentification or changes in nomenclature<sup>9</sup> or need retrospective georeferencing prior to use<sup>10,11</sup>. Even considering all these potential limitations, biodiversity inventories still remain a powerful tool for a deeper understanding of ecosystem complexity and for tracking long-term biodiversity trajectories. In this frame, historical data (e.g. properly labelled specimens from private collections, museums or herbaria) represent a crucial source of information on past species distribution and phenology<sup>12,13</sup>. For example, comparing present-day information with data spanning the period of accelerated anthropogenic habitat destruction and climate change (e.g. the nineteenth and twentieth centuries) allowed to detect local extinctions and altitudinal shifts following human disturbance or recent climate change 14-16.

<sup>1</sup>Department of Environmental Science and Policy, University of Milan, Via G. Celoria 26, 20133, Milan, Italy. <sup>2</sup>Institute of Ecosystem Studies, ISE-CNR c/o Department of Biology, University of Rome 'Tor Vergata', Via della Ricerca Scientifica, 00133, Rome, Italy. <sup>3</sup>Institute of Atmospheric Sciences and Climate, ISAC-CNR, 40129, Bologna, Italy. <sup>4</sup>Evolutionary Biology & Ecology - C.P. 160/12, Université Libre de Bruxelles, Avenue F.D. Roosevelt 50, B-1050, Brussels, Belgium. <sup>5</sup>Institute of Research on Terrestrial Ecosystems, IRET-CNR, Viale dell'Università 32, 00185, Rome, Italy. <sup>6</sup>Department of Biology, University of Rome 'Tor Vergata', Via della Ricerca Scientifica, 00133, Rome, Italy. <sup>7</sup>Institute of Geosciences and Earth Resources, IGG-CNR, National Research Council, 56124, Pisa, Italy. Correspondence and requests for materials should be addressed to S.M. (email: silvio.marta@hotmail.it)

Distribution data from collections frequently exhibit high resolution at both the spatial (collection locality) and temporal (complete date or year) scales; such biodiversity inventories are intrinsically spatially-accurate multi-temporal datasets, spanning over wide areas and temporal intervals. Comparable multi-temporal fine-grained abiotic data are thus required to shed light on the patterns of biodiversity response to environmental changes, and to identify the processes underlying such responses<sup>17</sup>. During the past decades, a substantial set of climatologies has been made available to researchers dealing with drivers of species distribution, but each of the most popular climate datasets clearly has some limitations. WorldClim<sup>18,19</sup> has very broad spatial extent (global) and high resolution (30 arcsec), but lacks of temporal extent (multi-year climatologies: 1950-2000 and 1970-2000, respectively). On the contrary, datasets such as CRUT4<sup>20</sup>, CRU TS 4.02<sup>21</sup>, GISSTEMP<sup>22</sup>, NOAA<sup>23</sup> and BEST<sup>24</sup> are global, providing gridded anomalies across the World at monthly resolution with a high temporal extent (first year ranging from mid-18th century for BEST to 1901 for CRU TS 4.02), but with quite coarse spatial resolution (ranging from 5° to 0.5°). Finally, the recently published CHELSA<sup>25</sup> combines the high spatial extent and resolution of WorldClim (worldwide at 30 arcsec) with the high temporal resolution of CRU TS (monthly means), but is temporally limited to 1979-2013. A major issue of all these datasets is that they are originated from the broad-scale interpolation of data from weather stations, and can be locally inaccurate. Finally, Deblauwe et al.<sup>26</sup> showed that such global datasets have fluctuating performances across the globe, and proposed the use of remotely sensed data to obtain more accurate climatologies, particularly in areas with a limited number of ground stations. Unfortunately, both the Deblauwe dataset and remotely-sensed data are temporally limited to the last forty years.

Our aim here was to produce and share a climatically explicit database on the distribution of Italian fauna with high spatial and temporal resolution. Based on GBIF data, Italy stands out among western countries for the low completeness of its species inventory<sup>27</sup>, even if it is one of the most biodiverse countries in Europe. Italy is formally neither a voting nor an associated participant of GBIF (https://www.gbif.org/the-gbif-network); consequently, there was no massive data contribution to the GBIF database from Italian institutions to date<sup>28</sup>. To fill this gap, in this work we updated, cleaned and georeferenced with greater precision raw distribution data from the extensive database for the Italian fauna (CKmap 5.3.8; ref.<sup>29</sup>), and integrated it with a high-resolution reconstruction of monthly temperature and precipitation in the location and at the time specimens were collected. The dataset compiled here represents the most spatially and temporally comprehensive and accurate faunal distribution database available for the Italian fauna, as well as one of the first attempts to finely yet massively reconstruct climate conditions under which biodiversity data were collected over more than two centuries.

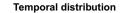
#### Methods

CKmap 5.3.8 (ref.<sup>29</sup>; English edition – version 5.4) reports the taxonomical data and the occurrence in Italy of over 10,000 terrestrial and freshwater animal species<sup>30</sup>. Species distribution was mapped attributing each record (i.e. the occurrence of a species in a single location) to the  $10 \times 10 \,\mathrm{km}$  UTM cell (ED50 datum, MGRS system) on the basis of a gazetteer. The gazetteer stored in the database (available for download at <a href="http://www.faunaitalia.it/documents/TCI.zip">http://www.faunaitalia.it/documents/TCI.zip</a>) included 46,961 toponyms taken from the "Touring Club Italiano" (TCI) atlas, accurately georeferenced using topographic maps of Italy at the scale 1:25,000<sup>30</sup>.

The geographic cleaning phase included quality checks for both CKmap and the reference gazetteer. All toponyms were checked for double spacing and/or additional spacing at the beginning and at the end of the string; all the records with an empty locality field were discarded. In the gazetteer, the precision of coordinates was indicated with the letters A, F and G³0. All G records were excluded, given they represent the centroid of broad spatial polygons (i.e., mountain ranges or rivers). A one-to-many join then allowed linking each of the remaining toponyms (A and F; i.e. individual locations) to the corresponding distribution records. An R code allowing partial matches (i.e. allowing up to 5% of letter deletion, insertion or substitution) was run on non-matching records and putative matches were carefully verified by hand prior to assignment. The distribution of the retained records was polygonised via a Voronoi tessellation using the deldir R package³¹ and the resulting Voronoi diagram was clipped using coastline and administrative boundaries. We used this approach because database compilers described each collection locality using the closer toponym³0. For each Voronoi cell, spatial precision was calculated as the distance between the sample point (toponym) and the farthest vertex (i.e. the positive pole, for bounded cells). The dimension of each Voronoi cell, and hence the record precision, obviously depends on the local toponym density.

The database cleaning consisted in the deletion of inaccurate records. All the records reporting an inaccurate taxonomical classification (e.g. using open nomenclature<sup>32</sup> instead of the Linnean binomials) were discarded. The remaining binomials were cleaned by removing subspecific classification and checking for extra spacing and typos using the above-mentioned approach. Furthermore, all records without collection date were discarded and the remaining records were checked for agreement with publication year. All the records with publication year (if any) later than collection year were retained; we also retained all the records without publication year, which represented unpublished collection specimens. We finally removed duplicate records (i.e. occurrence records of the same taxon from the same location and year), except when they were from different sources or were collected at different elevations.

For each retained record, we reconstructed average monthly minimum and maximum temperature and total monthly precipitation exploiting the huge amount of meteorological observations available for Italy since the 18<sup>th</sup> century<sup>33</sup>. We used the anomaly method for climate reconstruction<sup>34</sup>, which is based on the independent reconstruction of the climatologies (i.e. the climate normal over the standard reference period) and the deviations from them (i.e. the anomalies with respect to the same baseline period). Climatologies are characterized by strong spatial gradients and a large number of weather stations is necessary to capture them, even if available for a short period. Consequently, an interpolation technique that exploits the dependence of climate normals on orography and other geographic parameters is necessary to reconstruct such climatologies<sup>35–37</sup>. Anomalies, on the other hand, are linked to climate change and variability; for this reason, they are usually characterized by higher



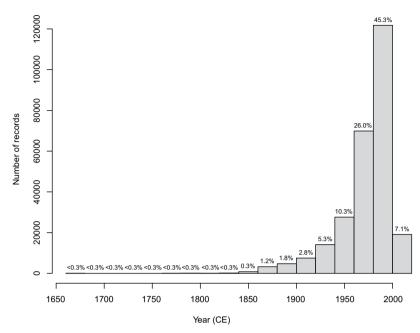


Fig. 1 Temporal distribution of the records from ClimCKmap. Total number of records: 268,977 from 8,445 binomials and 16,332 localities. Temporal coverage: 1680–2006 CE. The numerals above each bar indicate the percentage of records falling in that time period.

spatial coherence<sup>34</sup>. A limited number of weather stations can therefore be sufficient to capture spatial patterns through simpler interpolation methods, but it is pivotal to use long-term time series to have a satisfactory temporal coverage over the past. These series must in turn be corrected for errors deriving from the history of the stations (changes of station and instrument location, instrument replacements, changes in observation protocols, etc.) which could interfere with the actual climate signal<sup>33,38,39</sup>. For each record, we reconstructed the climate information of the Voronoi cell centre using average cell elevation following Brunetti *et al.*<sup>40</sup>. For each centre we first estimated the climate normals calculating a weighted linear regression of the data from nearby stations as a function of elevation. We assigned greater weights to the stations with elevation and topographic position similar to that of the location of interest, as derived from a 30 arcsec resolution digital elevation model<sup>36,37</sup>. Anomalies were interpolated at the same locations exploiting an improved version of the dataset presented in Brunetti *et al.*<sup>33</sup> and then, by combining anomalies and climatologies, temporal series of temperature and precipitation in absolute values were obtained for each record location. Bioclimatic variables were finally calculated on monthly temperature and precipitation using the dismo R package<sup>41</sup>. All the above-mentioned analyses were run using custom R and Fortran codes.

#### **Data Records**

The final dataset includes 268,977 occurrence records deriving from 8,445 binomials and 16,332 localities, dating between 1680 and 2006 CE (Fig. 1). A complete description of the database structure is reported in the Online-only Table 1.

The full database is made accessible at https://doi.org/10.6084/m9.figshare.7906739.v4 42 and http://hdl.handle.net/21.11125/a91f85cb-befd-4e14-8e83-24f17c4a0491.

#### **Technical Validation**

The starting dataset (CKmap 5.3.8; ref.<sup>29</sup>) included 548,868 occurrence records from 10,132 taxa. After taxonomic simplification and cleaning, 544,764 records from 10,064 binomials (i.e. species) were retained. The subsequent geographic cleaning and georeferencing phase allowed assigning coordinates to 470,232 records from 19,574 localities. The removal of non- or dubiously-dated records restricted the size of the resulting dataset to 277,310 records. The removal of duplicates led to a final dataset including 268,977 distribution records deriving from 8,445 binomials and 16,332 localities, dating between 1680 and 2006 CE (Fig. 1). The georeferencing methodology returned satisfactory results in term of precision (Fig. 2; median accuracy = 2480 m; for 95% of records accuracy was between 1251 and 4474 m). Climate reconstructions did not allow assigning climate data to only a small fraction of the total dataset (894 records dating between 1680 and 1921 CE; 0.3%) due to the lack of meteorological data (pre-instrumental period) or the incompleteness of the local time series. Monthly means of minimum and maximum temperature and total precipitation were thus reconstructed for the remaining part of the records (268,083 records dating between 1790 e 2006 CE).

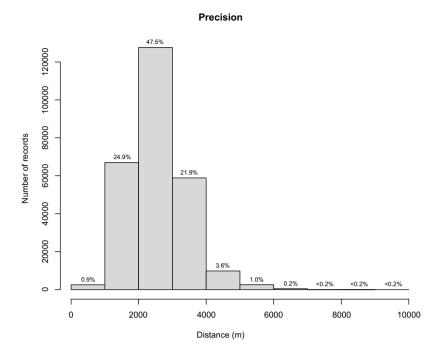


Fig. 2 Distance between sample point and the farthest vertex of the Voronoi cell (precision). Median = 2480 m; 95% confidence interval = 1251–4474 m. The numerals above each bar indicate the percentage of records falling in that distance class.

#### **Usage Notes**

In the last two centuries, biodiversity has been facing abrupt climate and environmental changes globally, mainly due to human activities. Huge efforts have been spent to expand our knowledge of species distribution worldwide, but the picture is still far from complete. In fact, apart from rare exceptions<sup>43</sup>, low taxonomic coverage and temporal gaps still limit our understanding of ecosystem functioning under climate change and biodiversity loss. The dataset we compiled and presented here aimed at moving a step forward to fill this gap, by providing spatially, temporally and climatically accurate distribution information for the fauna of one of the most biodiverse countries in Europe. ClimCKmap is currently the largest spatially, temporally and climatically explicit distribution database ever assembled for the Italian fauna. Future research should consider exploiting the database, for example deriving empirical relationships between climate change and biodiversity loss and using them to evaluate impacts of future climate change scenarios, as well as to identify climate-biodiversity change hotspots. Proper use of the data will also allow for developing knowledge-based biodiversity conservation measures.

This database has of course some potential issues that have to be taken into account when working with it. First, we opted for retaining the original classification and taxonomy, i.e. a revised form of the checklist of the Italian fauna<sup>44</sup>. Additionally, some authors reported biases due to spatially unequal sampling effort<sup>45,46</sup>, as well as temporal and taxon-specific biases in the completeness of the database<sup>46</sup>. Researchers interested in exploiting the database should thus evaluate case by case if a taxonomic updating for the group under analysis is needed, and be aware of the potentially (spatially and/or temporally) biased nature of the database itself.

#### **Code Availability**

Custom R and Fortran codes used to run the analyses, together with an example dataset are available at https://doi.org/10.6084/m9.figshare.7906739.v4 (ref. 42).

#### References

- 1. Hortal, J. et al. Seven shortfalls that beset large-scale knowledge of biodiversity. Annu. Rev. Ecol. Evol. S. 46, 523-549 (2015).
- Yang, W., Ma, K. & Kreft, H. Geographical sampling bias in a large distributional database and its effects on species richness-environment models. J. Biogeogr. 40, 1415–1426 (2013).
- Reddy, S. & Dávalos, L. M. Geographical sampling bias and its implications for conservation priorities in Africa. J. Biogeogr. 30, 1719–1727 (2003).
- 4. Ficetola, G. F., Bonardi, A., Sindaco, R. & Padoa-Schioppa, E. Estimating patterns of reptile biodiversity in remote regions. *J. Biogeogr.* 40, 1202–1211 (2013).
- 5. Franklin, J., Serra-Diaz, J. M., Syphard, A. D. & Regan, H. M. Big data for forecasting the impacts of global change on plant communities. *Global Ecol. Biogeogr.* **26**, 6–17 (2017).
- Colwell, R. K. et al. Models and estimators linking individual-based and sample-based rarefaction, extrapolation and comparison of assemblages. J. Plant Ecol. 5, 3–21 (2012).
- 7. Engemann, K. *et al.* Limited sampling hampers "big data" estimation of species richness in a tropical biodiversity hotspot. *Ecol. Evol.* 5, 807–820 (2015).
- 8. Belbin, L., Daly, J., Hirsch, T., Hobern, D. & La Salle, J. A specialist's audit of aggregated occurrence records: An 'aggregator's' perspective. *ZooKeys* 305, 67–76 (2013).

- 9. Wiser, S. K. Achievements and challenges in the integration, reuse and synthesis of vegetation plot data. *J. Veg. Sci.* 27, 868–879 (2016).
- 10. Guralnick, R. P., Wieczorek, J., Beaman, R. & Hijmans, R. J. & BioGeomancer Working Group. BioGeomancer: automated georeferencing to map the world's biodiversity data. *PLoS Biol.* 4, e381 (2006).
- 11. Gratton, P. et al. A world of sequences: can we use georeferenced nucleotide databases for a robust automated phylogeography? J.Biogeogr. 44, 475–486 (2017).
- 12. Lavoie, C. Biological collections in an ever changing world: Herbaria as tools for biogeographical and environmental studies. *Perspect Plant Ecol.* **15**, 68–76 (2013).
- Lister, A. M., Climate Change Research Group. Natural history collections as sources of long-term datasets. Trends Ecol. Evol. 26, 153–154 (2011).
- 14. Shaffer, H. B., Fisher, R. N. & Davidson, C. The role of natural history collections in documenting species declines. *Trends Ecol. Evol.* 13, 27–30 (1998).
- 15. Lenoir, J., Gégout, J. C., Marquet, P. A., De Ruffray, P. & Brisse, H. A significant upward shift in plant species optimum elevation during the 20th century. *Science* 320, 1768–1771 (2008).
- 16. Tingley, M. W. & Beissinger, S. R. Detecting range shifts from historical species occurrences: new perspectives on old data. *Trends Ecol. Evol.* **24**, 625–633 (2009).
- 17. Parmesan, C. Ecological and evolutionary responses to recent climate change. Annu. Rev. Ecol. Evol. S. 37, 637-669 (2006).
- 18. Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G. & Jarvis, A. Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* 25, 1965–1978 (2005).
- 19. Fick, S. E. & Hijmans, R. J. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* 37, 4302–4315 (2017).
- 20. Jones, P. D. *et al.* Hemispheric and large-scale land-surface air temperature variations: An extensive revision and an update to 2010. *J. Geophys. Res.-Atmos.* **117**(D5), 1–29 (2012).
- 21. Harris, Í., Jones, P. D., Osborn, T. J. & Lister, D. H. Updated high-resolution grids of monthly climatic observations—the CRU TS3. 10 Dataset. *Int. J. Climatol.* 34, 623–642 (2014).
- 22. Hansen, J., Ruedy, R., Sato, M. & Lo, K. Global surface temperature change. Rev. Geophys. 48, 1-29 (2010).
- 23. Smith, T. M., Reynolds, R. W., Peterson, T. C. & Lawrimore, J. Improvements to NOAA's historical merged land-ocean surface temperature analysis (1880–2006). *J. Climate* 21, 2283–2296 (2008).
- 24. Rohde, R. et al. A new estimate of the average Earth surface land temperature spanning 1753 to 2011. Geoinfor. Geostat.: An Overview 1, 1-7 (2013).
- 25. Karger, D. N. et al. Climatologies at high resolution for the earth's land surface areas. Sci. Data 4, 170122 (2017).
- 26. Deblauwe, V. et al. Remotely sensed temperature and precipitation data improve species distribution modelling in the tropics. Global Ecol. Biogeogr. 25, 443–454 (2016).
- 27. Meyer, C., Kreft, H., Guralnick, R. & Jetz, W. Global priorities for an effective information basis of biodiversity distributions. *Nat. Commun.* **6**, 8221 (2015).
- 28. Cesaroni, D. et al. DNA Barcodes of the animal species occurring in Italy under the European "Habitats Directive" (92/43/EEC): a reference library for the Italian National Biodiversity Network. Biogeographia 32, 5–23 (2017).
- 29. Stoch, F. *CKmap* 5.3.8, http://www.faunaitalia.it/documents/CKmap\_54.zip (Ministero dell'Ambiente e della Tutela del Territorio, Direzione Protezione della Natura, 2000).
- 30. Ruffo, S. & Stoch, F. Checklist and Distribution of the Italian Fauna. 10,000 Terrestrial and Inland Waters Species. (Memorie del Museo Civico di Storia Naturale di Verona, 2.serie, Sezione Scienze della Vita, 2006).
- 31. Turner, R. Deldir: Delaunay Triangulation and Dirichlet (Voronoi) Tessellation, https://cran.r-project.org/package=deldir (2015).
- 32. Bengtson, P. Open nomenclature. Palaeontology 31, 223-227 (1988).
- 33. Brunetti, M., Maugeri, M., Monti, F. & Nanni, T. Temperature and precipitation variability in Italy in the last two centuries from homogenised instrumental time series. *Int. J. Climatol.* 26, 345–381 (2006).
- 34. Mitchell, T. D. & Jones, P. D. An improved method of constructing a database of monthly climate observations and associated high-resolution grids. *Int. J. Climatol.* 25, 693–712 (2005).
- 35. Daly, C. et al. Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. Int. J. Climatol. 28, 2031–2064 (2008).
- Brunetti, M., Maugeri, M., Nanni, T., Simolo, C. & Spinoni, J. High-resolution temperature climatology for Italy: interpolation method intercomparison. *Int. J. Climatol.* 34, 1278–1296 (2014).
- Crespi, A., Brunetti, M., Lentini, G. & Maugeri, M. 1961–1990 high-resolution monthly precipitation climatologies for Italy. Int. J. Climatol. 38, 878–895 (2018).
- 38. Aguilar, E., Auer, I., Brunet, M., Peterson, T. C. & Wieringa, J. *Guidelines on Climate Metadata and Homogenization*. World Climate Programme Data and Monitoring WCDMP-No. 53, WMO-TD No. 1186 (World Meteorological Organization, 2003).
- 39. Venema, V. K. C. et al. Benchmarking homogenization algorithms for monthly data. Clim. Past 8, 89-115 (2012).
- 40. Brunetti, M. et al. Projecting North Eastern Italy temperature and precipitation secular records onto a high-resolution grid. Phys. Chem. Earth 40, 9–22 (2012).
- 41. Hijmans, R. J., Phillips, S., Leathwick, J. & Elith, J. Dismo: Species distribution modelling, https://cran.r-project.org/package=dismo (2015).
- 42. Marta, S. et al. 2019\_ClimCKmap: a spatially, temporally and climatically explicit distribution database for the Italian fauna. figshare. https://doi.org/10.6084/m9.figshare.7906739.v4 (2019).
- 43. Oliver, T. H. et al. Declining resilience of ecosystem functions under biodiversity loss. Nat. Commun. 6, 10122 (2015).
- 44. Minelli, A., Ruffo, S. & La Posta, S. Checklist delle Specie della Fauna Italiana (Vol. 107) (Calderini, 1993).
- 45. Barbosa, A. M., Fontaneto, D., Marini, L. & Pautasso, M. Is the human population a large-scale indicator of the species richness of ground beetles? *Anim. Conserv.* 13, 432–441 (2010).
- Girardello, M., Martellos, S., Pardo, A. & Bertolino, S. Gaps in biodiversity occurrence information may hamper the achievement of international biodiversity targets: insights from a cross-taxon analysis. *Environ. Conserv.* 45, 370–377 (2018).

#### **Acknowledgements**

S.M. was supported by the ISE-CNR/Project of Interest 'NextData' - 'Montane butterflies and mammals as ecosystem indicators of climate effects: upgrading the NextData bank', S.M. and G.F.F. are funded by the European Research Council under the European Community's Horizon 2020 Programme, Grant Agreement No. 772284 ('IceCommunities - Reconstructing community dynamics and ecosystem functioning after glacial retreat').

#### **Author Contributions**

S.M., A.P., V.S., D.C. and G.A. conceived the idea. S.M. retrieved, cleaned and georeferenced distribution data, while M.B. associated climatic information to each distribution record. G.F.F., F.S., G.A., D.C., V.S. and A.P. jointly supervised the work. S.M., M.B. and G.F.F. led the writing with substantial contributions from all the other authors.

#### **Additional Information**

**Competing Interests:** The authors declare no competing interests.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

The Creative Commons Public Domain Dedication waiver http://creativecommons.org/publicdomain/zero/1.0/ applies to the metadata files associated with this article.

© The Author(s) 2019