

Atmospheric data for ornithology: an introduction

Atsuyoshi MANDA¹, Noriyuki M. YAMAGUCHI^{2,#}, Elham NOURANI² and Yuzo ARISAWA³

¹ Graduate School of Bioresources, Mie University, Kurimamachiya-cho 1577, Tsu-shi, Mie 514–8507, Japan

² Graduate School of Fisheries and Environmental Sciences, Nagasaki University, Bunkyo-machi 1-14, Nagasaki-shi, Nagasaki 852–8521, Japan

³ Weather Information and Communications Service LTD., Minami-ikebukuro 2–8–5, Toshima-ku, Tokyo 171–0022, Japan

ORNITHOLOGICAL SCIENCE

© The Ornithological Society
of Japan 2017

Abstract This paper introduces user-friendly atmospheric data, so-called objective analysis data, for ornithological studies. The data have been interpolated onto grid points distributed at a regular interval in space and time, and are suitable for analyzing using computer data analysis software. Data assimilation techniques, which are basically the application of optimization and control theories, are utilized for producing objective analysis data in order to reduce errors as much as possible and obtain the most reliable dataset. Some examples of objective analysis data are shown and their features are described. Some cautionary notes are also given in order to avoid misinterpretation of the data.

Key words Atmospheric data, Objective analysis

This review paper provides a brief introduction to the atmospheric data used in modern meteorology and which is applicable to ornithology, in particular avian movement ecology. Meteorological data are commonly used in various aspects of ornithological research. Variables such as temperature, precipitation, and solar radiation, are frequently used when building ecological niche models to investigate habitat suitability and predict the ranges of bird species (e.g., Tingley et al. 2009; Gschweng et al. 2012; Limiñana et al. 2015). Moreover, variables related to air movement, i.e. horizontal and vertical wind, are particularly essential in research on migratory birds, as they affect the duration and energetic cost of migration, general routes, and decisions regarding departure points and stop-over sites (Liechti 2006; Shamoun-Baranes et al. 2006; Nourani and Yamaguchi 2017).

Rather than a comprehensive review, this paper is a primer for ornithologists who wish to experiment with these data. The main targets of this review are researchers interested in the relationship between avian movements and atmospheric conditions at the regional or global spatial scales (from ca. 10 km to global). In meteorological terms, “regional” indicates

a specific geographical region, in contrast “global” indicates the entire globe. Numerous kinds of atmospheric data are now available thanks to a rapid increase in the range of observation systems, which now include satellite and ground-based remote sensing technologies as well as sophisticated data analysis systems for synthesizing such observational data. In general, the process of handling observational datasets is difficult, requiring specialist skills. Users may also need to undertake quality control and interpolation of data. Users sometimes require a background in dynamical meteorology in order to obtain physically reasonable data, in addition to computing skills that are rarely used in ornithology. We have chosen to focus on relatively user-friendly datasets (as described later), instead of giving a comprehensive description of various kinds of atmospheric data, as our aim is to support beginners in atmospheric data analysis. The remainder of this paper is organized as follows: first we introduce “objective analysis data” (the most convenient and user-friendly dataset for studying atmospheric fields); then, we provide an example of atmospheric data analysis in an ornithological study, and finally we give some cautionary notes concerning the use of objective analysis data, since the data are not perfect real-life data.

(Received 16 June 2016; Accepted 22 September 2016)

Corresponding author, E-mail: noriyuki@nagasaki-u.ac.jp

Objective analysis data

Modern meteorology strongly relies on so-called “objective analysis data,” which is in short produced by synthesizing outputs of a numerical simulation and various kinds of observational data. Most objective analysis data are “gridded” (i.e., data have been interpolated onto prescribed locations at regular spatial and temporal intervals) in contrast to the majority of observational data that are distributed at locations and times with irregular intervals. Objective analysis data are thus suitable for use with various types of data analysis software. Various kinds of objective analysis data have become available in recent years. This paper introduces some of the most representative and popular datasets, for example for near surface wind and air temperature using objective analysis data (Fig. 1). For more information of the objective analysis data, readers are referred to the website at <https://climatedataguide.ucar.edu/>. Here we have used the National Centers for Environmental Prediction (NCEP) Department of Energy (DOE) Atmospheric Model Intercomparison Project (AMIP-II) reanalysis (R-2) dataset (Kanamitsu et al. 2002), which is one of the most popular objective analysis datasets used in meteorology. We henceforth refer to this dataset as NCEP–DOE reanalysis 2. The NCEP–DOE reanalysis

2 provides us with various kinds of atmospheric variables such as wind velocity, air temperature, relative humidity, with a resolution of 2.5 degrees latitude and longitude and at 18 vertical levels (surface, and 1000, 925, 850, 700, 600, 500, 400, 300, 250, 200, 150, 100, 70, 50, 30, 20 and 10 hPa pressure), at six-hourly intervals.

Meteorologists synthesize data using sophisticated “data assimilation” techniques, which are essentially an application of control theory (e.g., Gelb et al. 1974). Control theory is used in many scientific and engineering fields in order to improve the behavior of dynamic systems using the output from those systems. This is not the place to go into the details of data assimilation techniques as that requires a solid background in both optimization and control theories. Interested readers are referred to introductory textbooks in meteorology such as those by Kalney (2002), and Holton and Hakim (2004). Objective analysis data are now indispensable for studying the atmosphere, as obtaining a comprehensive picture of the state of the atmosphere is essentially impossible without them. Despite recent significant improvements to atmospheric observation systems with advanced ground based instruments and satellite-mounted remote-sensing devices, many

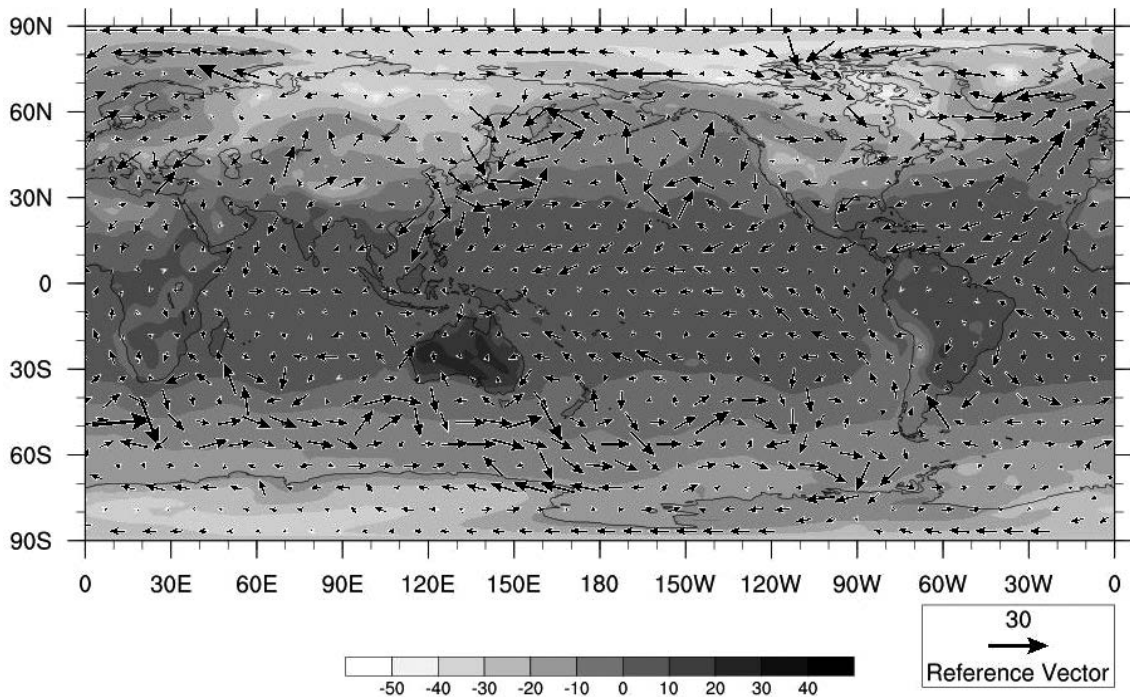


Fig. 1. Horizontal distribution of air temperature at 2 m height (shaded) and horizontal wind velocity at 10 m height (vector) of NCEP–DOE reanalysis 2 at 0000 UTC on 1 January 2015. Reference vector is in m s^{-1} .

gaps remain in observational data making it difficult to fully describe the mechanics of the atmosphere. Simply put, the data assimilation system interpolates atmospheric fields both in space and time and helps reduce the errors of interpolation as much as possible. The resultant atmospheric fields are considered to be statistically the most reliable under certain *a priori* assumptions (e.g., Kalney 2002).

Various other global objective analysis datasets are widely used by many meteorologists. These include the Japanese 55-year reanalysis (JRA-55) produced by the Japan Meteorological Agency (Kobayashi et al. 2015) and ERA-Interim produced by the Centre for Medium-Range Weather Forecasts (ECMWF)

(Dee et al. 2011). Readers are referred to Dee et al. (2011) and Kobayashi et al. (2015) for detailed specifications of these datasets since the specifications are rather lengthy.

On 1 January 2015, a cyclone was located at around 40°N, 145°E and blew strongly in a northerly direction from the Russian and Chinese coast towards the Japanese archipelago. Although the overall features of the wind fields are similar in the NCEP–DOE reanalysis 2, JRA-55, and ERA-Interim at 0000 UTC datasets, comparison reveals that the details are different (see Fig. 2). For example, the wind field shown by NCEP–DOE reanalysis 2 is rather smooth, compared with the other datasets. The field shows strong

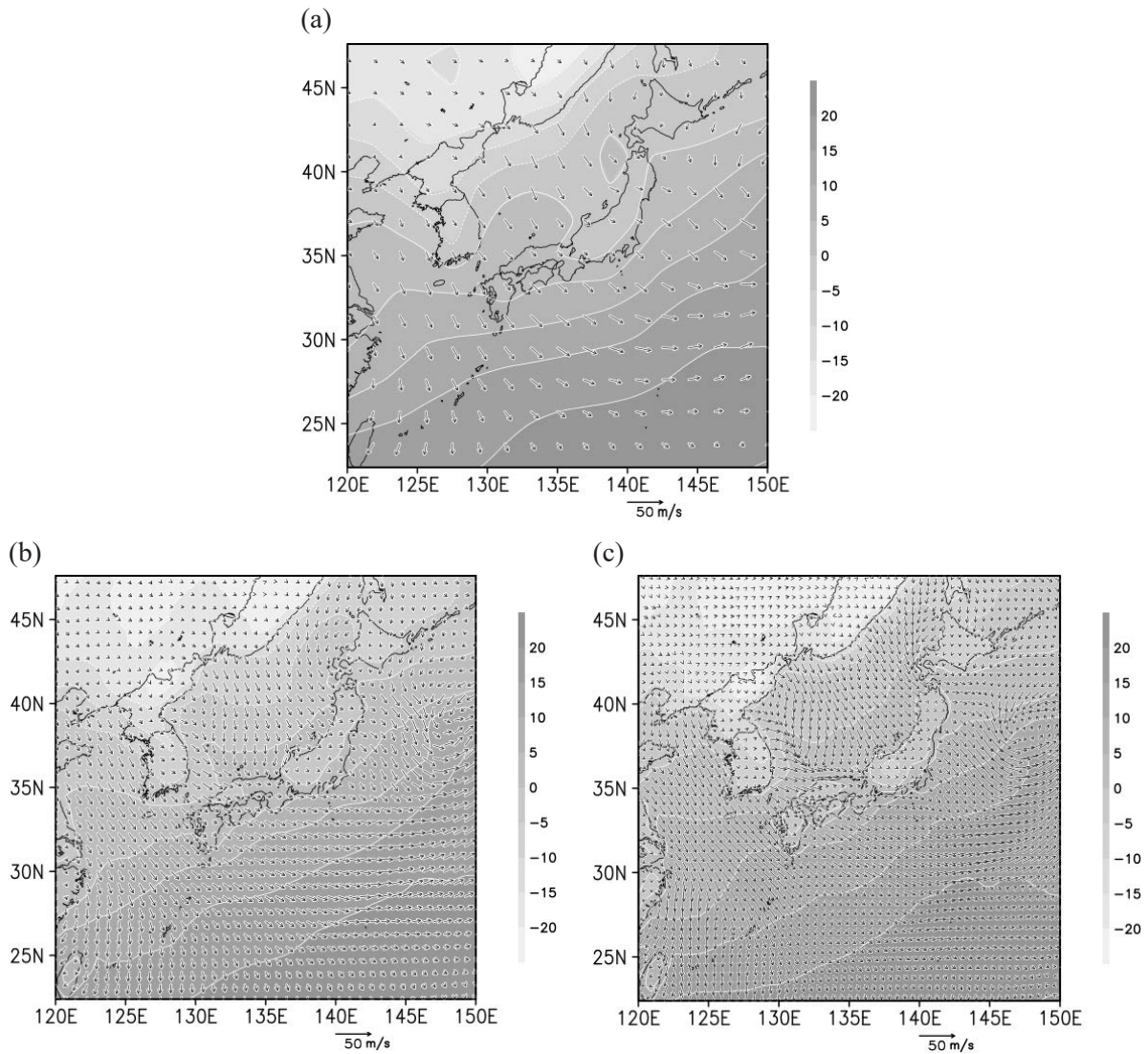


Fig. 2. Horizontal distributions of air temperature at 2 m height (shaded) and horizontal components of wind velocity at 10 m height (vector) at 0000 UTC on 1 January 2015. (a) NCEP–DOE reanalysis 2, (b) ERA-Interim, and (c) JRA-55.

spatial variability near the coast in ERA-Interim and JRA-55, mainly due to the difference in grid spacing in these datasets. The horizontal grid spacing of NCEP–DOE reanalysis 2, ERA-I, and JRA-55 used here are 2.5° , 0.75° , and $\sim 0.55^\circ$, respectively. Please note that some datasets have been re-gridded onto a coarser grid (typically in the order of 100 kilometers) in order to save disk space and some other reasons. It is not possible to assess which dataset is the best when only referring to Fig. 2. The scale of atmospheric data used in any ornithological study should be chosen based on the spatial scale of the bird behavior/characteristics being studied and also on the spatial accuracy of the bird data, such as the tracking points being used.

Although NCEP–DOE reanalysis 2, ERA-Interim, and JRA-55 cover the entire globe, the horizontal resolution of these datasets may be too coarse for those who are interested in regional studies. Some datasets exist with a finer resolution (on the order of ten kilometers or smaller); for example, the mesoscale analysis produced by the Japan Meteorological Agency (Ishikawa & Koizumi 2002; Fig. 3). The term “mesoscale” is used in meteorology in order to describe atmospheric phenomena ranging from an order of kilometers to a thousand kilometers (Markowski & Richardson 2010). The mesoscale analysis has a horizontal resolution of 5 km near the surface, which

is much finer than the global datasets, although it covers only the East Asian region around Japan (see Fig. 3 for a comparison of mesoscale analysis with JRA-55). The mesoscale analysis can reproduce many small-scale atmospheric phenomena owing to its finer model resolution, and both temperature and wind fields at the mesoscale exhibit much higher spatial variabilities compared with JRA-55.

Example of data processing

The volume of objective analysis data is considerable, thus file sizes are large. Therefore, in order to reduce file sizes and increase the speed of data access, meteorologists use special file formats, most typically Unidata netCDF format (<http://www.unidata.ucar.edu/software/netcdf/>). However, as most ornithologists are not familiar with the NetCDF format, we provide an example of data processing using netCDF with R (R Core Team 2015). Additional software packages, such as “RNetCDF” (Michna P with contributions from Milton Woods 2016) and “ncdf4” (Pierce 2015), are required to handle netCDF data in R (see Appendix). These packages can be installed simply by using the “install.packages” function in R.

It is not always necessary to directly manipulate NetCDF files, because some platforms convert this format into more user-friendly raster formats. For scientists working with surface-level atmospheric

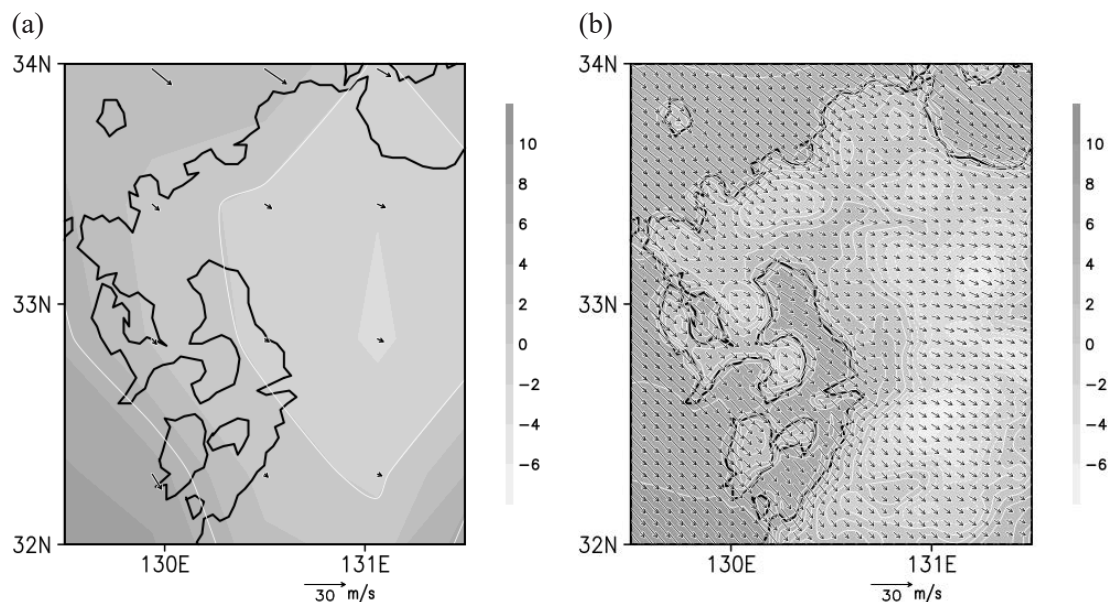


Fig. 3. Close-up of horizontal distributions of air temperature at 2 m height (shaded) and horizontal components of wind velocity at 10 m height (vector) around Kyushu Island, Japan at 0000 UTC on 1 January 2015. (a) JRA-55 and (b) mesoscale analysis.

data, WorldClim is very popular, whereas migration researchers favor Movebank (<https://www.movebank.org/>) and Uva-bits (<http://www.uva-bits.nl/>). Movebank and Uva-bits automatically annotate telemetry data with climate and weather variables and present them in more user-friendly formats than NetCDF. For most ornithologists, who are not familiar with netCDF format, the RNCEP package (Kemp et al. 2012) in R is very convenient. This package allows for downloading and analyzing NCEP/NCAR data in R, with no need to manipulate the original netCDF format.

Various other alternative data formats are also used in meteorology, such as Gridded Binary (GRIB). The R package, “rgdal” is capable of handling GRIB files. GRIB files can also be converted to netCDF using the climate data operators (cdo) utility developed by the Max-Planck Institute for Meteorology (<https://en.wikipedia.org/wiki/GRIB>). Handling datasets in formats other than those in netCDF is somewhat difficult and requires additional computational skills, which are beyond the scope of this introductory review.

Cautionary notes on Objective Analysis Data

It must be noted that objective analysis data are not real-life data; they should be considered “best guess” data using available observation data and simulation models. These data do not contain information on atmospheric phenomena smaller than their grid spacing. Moreover, the objective analysis data are not error-free since the simulation models used in the data assimilation techniques cannot represent all physical phenomena perfectly, furthermore observational data contain some errors. It is essential therefore that we take care when using objective analysis data.

Some precautions should be considered when using these data. First of all, it is important to pay attention to the grid spacing of the objective analysis data. All of the physical quantities in objective analysis data, such as temperature, and wind velocity, represent special averages. For example, consider an objective analysis dataset with a horizontal grid spacing of 5 km and a vertical grid spacing of 1 km; all of the values in this dataset are averages in a cuboid with a length and width of 5 km and a height of 1 km. We cannot say anything about atmospheric phenomena whose scale is smaller than this cuboid. To be more precise, a simulation model needs several grids within atmospheric phenomena in order to resolve them. For example, consider an atmospheric wave whose wave-

length is 10 km. A simulation model needs at least two grid points to represent a crest and trough of the wave, that is, the horizontal grid spacing must be 5 km or smaller. Please note that two grid points are the minimum requirement for representing the existence of a wave, while more grids are required in order to represent the features of the wave.

Let us consider two specific examples. If we want to resolve a mid-latitude cyclone with a horizontal scale on the order of 1,000 km, the grid spacing should be on the order of 100 km or smaller. This is one of the reasons why typical objective analyses have a grid spacing on the order of 100 km. If one wants to reproduce the kind of mesoscale convective systems that are responsible for many extreme weather events such as torrential rainfalls (e.g., Markowski & Richardson 2010), the grid spacing should be on the order of 10 km or smaller in order to resolve systems whose horizontal scale is on the order of 100 km, which is why mesoscale analysis has a horizontal grid spacing of 10 km. We need to examine the scale of the atmospheric phenomena relevant to our research interests before conducting detailed data analysis, in order to avoid misinterpretation of data.

Second, we must be aware of the discrepancies among different objective analysis data sets. Although all the wind fields are shown at the same time and height in the three panels of Fig. 2, they are not identical because they were produced using different data assimilation systems. They use different simulation models, observational data, and data assimilation techniques. Users should therefore examine the differences among the datasets looking for inconsistencies before interpreting the data. If a portion of one objective analysis dataset is not consistent with that of other datasets, then that portion may be an artifact.

As meteorological observation systems have improved, and as the quantity of observational data has increased, the accuracy of objective analysis has improved, thus in general the quality of recent state-of-the-art objective analysis data is better than that of older data sets (Wu et al. 2015). Improvements to the data assimilation systems have also helped to improve accuracy.

Concluding remarks

Objective analysis data give us a comprehensive view of the state of the atmosphere, i.e., the data are distributed at a regular interval in both space and time without missing values, something that is

unavoidable when using only conventional observational data. Objective analysis data have become tractable, even for non-specialists in meteorology, thanks to the development of useful computer tools. In the hope of expanding the use of these techniques in the field of ornithology, we have introduced objective analysis data as user-friendly atmospheric data that can be combined with R packages and a sample code that helps to manipulate the netCDF format. The use of such data will help researchers achieve the essential understanding of the atmospheric environment that affects, perhaps even drives, aspects of bird movements such as migration.

Although the overall structure of the data format and the software described in this paper have not changed substantially over the last decade, it should be noted that, as in any field, continued refinements and improvements are being made, due to rapid progress in the development of observation and analysis systems.

ACKNOWLEDGMENTS

NCEP–DOE_Reanalysis 2 data were provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their Web site at <http://www.esrl.noaa.gov/psd/>. JRA-55 data were provided by Japan Meteorological Agency from their website at <http://jra.kishou.go.jp/>. ERA-Interim data were provided by ECMWF from their website at <http://apps.ecmwf.int/datasets/>. The mesoscale analysis data were provided by the JMA and distributed by the Research Institute for Sustainable Humanosphere, Kyoto University (<http://database.rish.kyoto-u.ac.jp/arch/jmadata/>). This work was supported in part by JSPS KAKENHI Grant Number 26440245 to A. Manda and N. M. Yamaguchi.

REFERENCES

- Dee DP, Uppala SM, Simmons AJ, Berrisford P, Poli P, Kobayashi S et al. (2011) The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quart J R Meteorol Soc* 137: 553–597.
- Gelb A, Kasper JF, Nash RA, Price CF & Sutherland AA (1974) *Applied Optimal Estimation*. MIT press, Cambridge.
- Gschweng M, Kalko EK, Berthold P, Fieldler W & Fahr J (2012) Multi-temporal distribution modelling with satellite tracking data: predicting responses of a long-distance migrant to changing environmental conditions. *J Appl Ecol* 49: 803–813.
- Holton JR & Hakim GJ (2004) *An introduction to dynamic meteorology*. Academic Press, Massachusetts.
- Ishikawa Y & Koizumi K (2002) One month cycle experiments of the JMA mesoscale 4-dimensional variational data assimilation (4D-Var) system. *CAS/JSC WGN Res. Activ Atmos Oceanic Modell* 32: 0126–0127.
- Kalney E (2002) *Atmospheric modeling, data assimilation and predictability*. Cambridge University Press, Cambridge.
- Kanamitsu M, Ebisuzaki W, Woollen J, Yang S-K, Hnilo JJ, Fiorino M et al. (2002) NCEP-DOE AMIP-II Reanalysis (R-2). *Bull Am Meteorol Soc* 83: 1631–1643.
- Kemp MU, van Loon EE, Shamoun-Baranes J & Bouten W (2012) RNCEP: global weather and climate data at your fingertips. *Meth Ecol Evol* 3: 65–70.
- Kobayashi S, Ota Y, Harada Y, Ebata A, Moriya M, Onoda H et al. (2015) The JRA-55 Reanalysis: General specifications and basic characteristics. *J Meteor Soc Japan* 93: 5–48.
- Liechti F (2006) Birds: blowin' by the wind? *J Ornithol* 147: 202–221.
- Limiñana R, Arroyo B, Terraube J, McGrady M & Mougeot F (2015) Using satellite telemetry and environment niche modelling to inform conservation targets for a long-distance migratory raptor in its wintering grounds. *Oryx* 49: 329–337.
- Markowski P & Richardson Y (2010) *Mesoscale meteorology in midlatitudes*. Royal Meteorological Society, Wiley, West Sussex.
- Michna P & Woods M (2016) RNetCDF: Interface to NetCDF Datasets. R package version 1.8-2. Available at <http://CRAN.R-project.org/package=RNetCDF> (accessed on 10 June 2016).
- Nourani E & Yamaguchi NM (2017) The influence of atmospheric conditions on migratory behavior of soaring birds: a review. *Ornithol Sci* 16: 5–15.
- Pierce D (2015) ncd4: Interface to Unidata netCDF (Version 4 or Earlier) Format Data Files. R Package version 1.15. Available at <http://CRAN.R-project.org/package=ncd4> (accessed on 10 June 2016).
- R Core Team (2015) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna. Available at: <https://www.R-project.org/> (accessed on 8 June 2016).
- Shamoun-Baranes J, van Loon E, Alon D, Alpert P, Yom-Tov Y & Leshem Y (2006) Is there a connection between weather at departure sites, onset of migration and timing of a soaring-bird autumn migration in Israel? *Global Ecol Biogeogr* 15: 541–552.
- Tingley MW, Monahan WB, Beissinger SR & Moritz C

Atmospheric data for ornithology

(2009) Birds track their Grinnellian niche through a century of climate change. *Proc Nat Acad Sci USA* 106: 19637–19643.

Wu R, Yan S, Shan Y, Dang Q & Shun G (2015) Deep image: Scaling up image recognition. *arXiv preprint arXiv: 1501.02876*.

Appendix. A sample R script as an example of converting a netCDF format file directly downloaded from ECMWF to a dataframe and saved it as a csv file. This script was created by Elham Nourani (email: mahle68@gmail.com), Laboratory of Animal Ecology, Nagasaki University, Japan in June 2016.