

ECOGRAPHY

Software note

rWind: download, edit and include wind data in ecological and evolutionary analysis

Javier Fernández-López and Klaus Schliep

J. Fernández-López (<http://orcid.org/0000-0003-4352-0252>) (javfer05@ucm.es), Real Jardín Botánico (CSIC), Madrid, Spain. – K. Schliep (<http://orcid.org/0000-0003-2941-0161>), Univ. of Massachusetts Boston, Boston MA, USA.

Ecography

42: 804–810, 2019

doi: 10.1111/ecog.03730

Subject Editor: Brody Sandel
Editor-in-Chief: Miguel Araújo
Accepted 30 October 2018

1) Wind connectivity has been identified as a key factor driving many biological processes.

2) Existing software available for managing wind data are often overly complex for studying many ecological processes and cannot be incorporated into a broad framework.

3) Here we present rWind, an R language package to download and manage surface wind data from the Global Forecasting System and to compute wind connectivity between locations.

4) Data obtained with rWind can be used in a general framework for analysis of biological processes to develop hypotheses about the role of wind in driving ecological and evolutionary patterns.

Keywords: R, wind connectivity, landscape genetics



Introduction

The bulk movement of air across the surface of the Earth, that is, wind, has been broadly recognized as an important influence in biological processes related to species distribution and biogeography (Hooker et al. 1844, Freeman 1945, Winkworth et al. 2002, Sanmartín et al. 2007). For example, wind currents can play a decisive role in driving patterns in bird migration (Felicísimo et al. 2008, Vansteelant et al. 2017), island colonization (Harvey 1994, Juan et al. 2000), gene flow between populations (Calsbeek and Smith 2003), and dispersal ecology (Muñoz et al. 2004, Nathan 2006). Models of wind-mediated processes are frequently criticized due to the lack of available empirical wind data and their inherent infalsifiability (Morrone and Crisci 1995, Ebach and Williams 2010). However, in recent decades the development of modern monitoring systems for atmospheric conditions and the public availability of these data from these systems (Shamoun-Baranes et al. 2010) have promoted the incorporation of quantitative wind data into research (Kemp et al. 2010, Tøttrup et al. 2017). The development of tools to access and manage these data has also increased, and several models and R packages have been created in order to study the effect



www.ecography.org

of wind on specific biological processes (e.g. bird migration, Kemp et al. 2012a, b). However, these models are often very specialized and usually require input data such as radio-tracked locations or bird flying altitudes, which are not always available. In addition, it is usually quite difficult to adapt these models into a more general framework in order to analyze the role of wind connectivity in evolutionary processes such as alternative population genetics models (e.g. landscape ecology), or species dispersal versus vicariance models in biogeography (e.g. oceanic island colonizations, Queiroz 2005).

In this context, a simpler approach is useful in order to formulate hypotheses about connectivity between individuals, populations or communities to be later tested with any source of information, from simple presence records, to genetic data (microsatellite data, NGS data, etc.). Presently, available software to compute connectivity in ecology, including Circuitscape (McRae 2006, McRae et al. 2008), *gdistance* and *raster* (Etten 2017, Hijmans 2017) or *GLFOW* (Leonard et al. 2017), are based primarily on the inclusion of friction layers: maps with habitat suitability or any other geographical/ecological characteristic that may influence dispersal or movement ability. Least cost path or connectivity are then computed taking into account these layers via multiple algorithms (e.g. Dijkstra algorithm, Dijkstra 1959). However, wind data have several peculiarities that make them not particularly adaptable to these models (Etten pers. comm.). First, wind connectivity depends on two factors obtained from the same data: wind speed and direction. Second, wind based connectivity is directional and place dependent, i.e. it not only depends on the wind speed and direction at source cell, but also the location of the target cell.

Here we introduce *rWind*, a package in the R language for statistical computing and graphics (R Development Core Team), designed specifically to download and process wind data from the Global Forecasting System. From these data, users can obtain wind speed and direction layers in order to compute connectivity values between locations. *rWind* fills the gap between wind data accessibility and their inclusion in a general framework to be applied broadly in ecological or evolutionary studies.

In the following section, we describe the data used by *rWind*, provide a brief description of the functions in the library, and detail the algorithm used to compute connectivity values. Finally, we provide three examples to illustrate the general functionality of the package.

Description

The Global Forecasting System wind data

The Global Forecasting System (GFS) atmospheric model is a dataset from the National Oceanic and Atmospheric Administration (NOAA) and National Centers for Environmental Prediction (NCEP). In this database, wind

is stored as velocity vector components (U: eastward_wind and V: northward_wind) at 10 m above the Earth's surface. The resolution of the data is 0.5 degrees, approximately 50 km. Wind velocities have been registered six times per day (00:00 – 03:00 – 06:00 – 09:00 – 12:00 – 15:00 – 18:00 – 21:00 (UTC)), since 6th May 2011 and is updated daily. In *rWind*, these data are obtained via queries to The Pacific Islands Ocean Observing System, coordinated by the Univ. of Hawaii School of Ocean and Earth Science and Technology (SOEST). A raw plain text file with gridded data is obtained for each dataset requested by the user, with the date and time of the data, the location (longitude and latitude coordinates) the wind vectors (U and V components) and wind speed and direction. These data can either be exported in a .csv file or stored internally as an 'rWind data frame' in R. In Table 1 we present the functions contained in *rWind* package, with a brief description of each.

Cost/connectivity computation

One of the most important functions of the *rWind* package is the computation of a cost matrix between selected locations based on wind data ('flow.dispersion' function).

To calculate the movement cost from any starting cell to one of its 8 adjacent cells (Moore neighborhood), we take three parameters: wind speed at starting cell, wind direction at the starting cell (azimuth), and the position of the target cell.

To compute this cost, we implemented by default the algorithm proposed by Muñoz et al. (2004) and their variation in Felicísimo et al. (2008) (equations adapted from Muñoz et al. 2004, Felicísimo et al. 2008, González-Solís et al. 2009),

$$Cost = \frac{HF}{S} \quad (1)$$

where HF is the horizontal factor and S the wind speed at the starting cell. Equation 2 shows how the horizontal factor is obtained:

$$HF = \begin{cases} 0.1 & \text{if } HRMA=0 \\ 2 \times HRMA & \text{if } HRMA \neq 0 \end{cases} \quad (2)$$

where HRMA (Horizontal Relative Moving Angle) is the angle in degrees between the azimuth and the direction of the movement trajectory to the target cell. This difference is used to penalize the connectivity (increasing the cost) between both cells when deviations from the exact wind vector azimuth increases. If the HRMA is zero (i.e. movement is in the exactly same direction as azimuth), the parameter called Horizontal Factor (HF) is set to 0.1. Otherwise, HF is equal to two times HRMA (Eq. 2). This algorithm is used to compute 'active' movement costs. In other words, it allows the organism to move against wind directions as birds do during migration.

Table 1. Brief description of the main functions included in the R library rWind.

Function	Description
wind.dl wind.dl_2	Downloads wind data from the Global Forecast System (GFS) of the USA's National Weather Service (NWS) (<www.ncdc.noaa.gov/data-access/model-data/model-datasets/global-forecast-system-gfs>) and returns either a .csv file or a data.frame.
wind.mean	Takes a list of wind data downloaded with wind.dl_2 and returns the mean (average) of the time series in a data.frame.
tidy	Takes an 'rWind_series' object from wind.dl_2 and joint and tidy up wind data in a single data.frame.
wind2raster	wind2raster creates a raster stack file (gridded) from wind data downloaded, with two raster layers: wind direction and wind speed.
flow.dispersion	It takes input from raster stack with two raster layers: direction and speed. flow.dispersion computes movement conductance through a flow either, sea or wind currents following any computation defined by the user and provided as a function. It returns either a sparse cost matrix or a conductance TransitionLayer object.

To compute 'passive' movement cost, that is avoiding movement against wind, we use a variation of Eq. 2

$$HF = \begin{cases} 0.1 & \text{if } HRMA=0 \\ \infty & \text{if } > 90 \\ 2 \times HRMA & \text{if } 0 < HRMA \leq 90 \end{cases} \quad (3)$$

where HF is set to ∞ for all cases in which the HRMA is more than 90 degrees (Muñoz et al. 2004). Though this algorithm is provided by default in 'flow.dispersion' as a straightforward way to compute movement cost based in wind data, custom functions can be defined by the users to be used with 'flow.dispersion'.

Two outputs are possible from the 'flow.dispersion' function. First, the 'raw' mode creates a sparse matrix (class 'dgCMatrix') from the Matrix R package with transition costs between all cells at the study area. Second, the 'transition-Layer' mode creates a TransitionLayer object with conductance values (1/cost) between cells which can be used with the 'gdistance' R package (Etten 2017) to compute the shortest path or movement cost between two locations.

Examples

To illustrate some functionalities of rWind, we have designed three brief, fully reproducible examples. In the first, we show the very basic functionality of rWind to download and manage wind data and to compute the shortest paths between two points with the help of gdistance package. In the second, we use rWind to download and plot wind data during hurricanes that occurred in the Caribbean during the month of September, 2017. Finally, in the third example, we show how rWind can be used to obtain wind connectivity between mainland and islands to test hypotheses about evolutionary processes in wind-dispersed plants.

Example 1: getting shortest wind paths across Strait of Gibraltar

The Strait of Gibraltar is an important geographical connection point between Europe and Africa. Many birds and other organisms use this point to complete their migratory

routes between both continents, since the minimum distance between both coasts is around 14 km (Bernis and Tellería 1981). For this reason, the study of wind patterns in this region is relevant to understanding how they affect animal migratory behavior or other ecological processes (Richardson 1990). In this simple example, we introduce the most basic functionality of rWind, to obtain the anisotropic (direction-dependent) shortest paths between two points across the Strait of Gibraltar. The following code produces Fig. 1, for an extended example see Supplementary material Appendix 1–1.

First, we load the packages that we will use

```
library(rWind)
library(raster)
library(rworldmap)
library(gdistance)
library(fields)
library(lubridate)
library(shape)
```

Now, we download wind data for the Strait of Gibraltar at the selected date and time (in our example, 12 February 2015 at 12:00 pm)

```
w <- wind.dl(2015, 2, 12, 12, -7, -4,
            34.5, 37.5)
```

Next, we create a raster stack with wind direction and speed.

```
wind_layer <- wind2raster(w)
```

With this raster stack, we can compute conductance values (1/cost) to be used later to get the shortest paths between the two points using gdistance package.

```
Conductance <- flow.dispersion(wind_
  layer, output="transitionLayer")
```

```
AtoB <- shortestPath(Conductance,
  c(-5.5, 37), c(-5.5, 35),
  output="SpatialLines")
```

```
BtoA <- shortestPath(Conductance,
  c(-5.5, 35), c(-5.5, 37),
  output="SpatialLines")
```

Finally, we can plot the wind data with the shortest paths.

```
image.plot(s1, col=terrain.colors(10),
  zlim=c(0,7),
  xlab="longitude", ylab="latitude")
lines(getMap(resolution="low"), lwd=4)
```

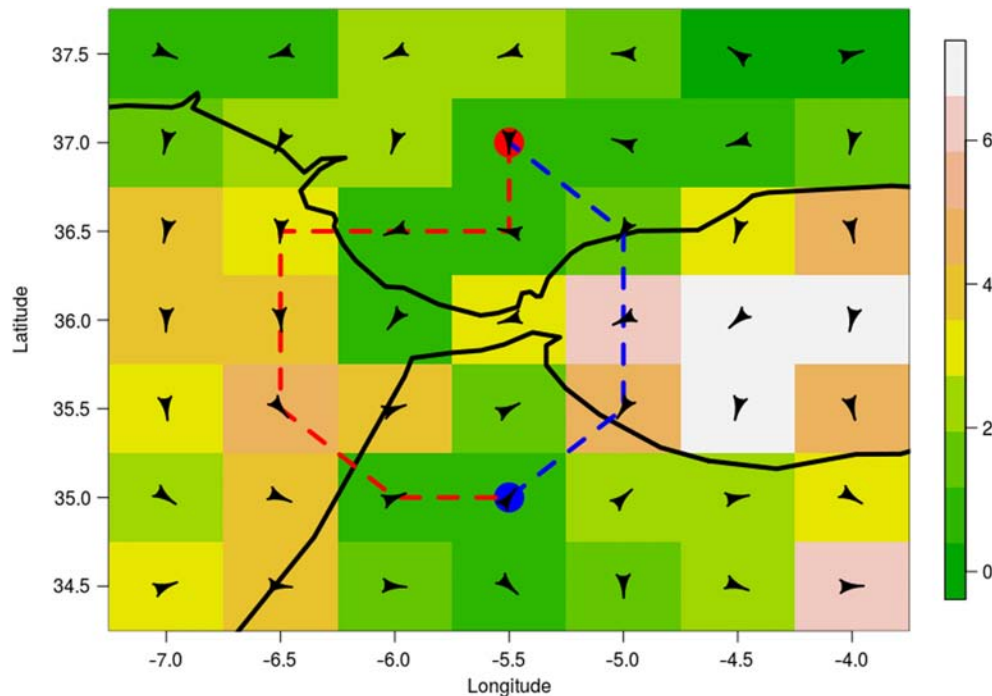


Figure 1. Shortest paths following wind speed and direction of 2015 12 February at 12:00 (UTM) between two points across Strait of Gibraltar.

```
points(-5.5, 37, pch=19, cex=3.4,
  col="red")
points(-5.5, 35, pch=19, cex=3.4,
  col="blue")
lines(AtoB, col="red", lwd=4, lty=2)
lines(BtoA, col="blue", lwd=4, lty=2)
Arrowhead(w$lon, w$lat, angle=
  arrowDir(w), arr.length=0.4, arr.
  type="curved")
```

Example 2: monitoring and plotting the Caribbean hurricanes Irma, José, and Katia (September 2017)

During the first days of September 2017, three hurricanes (named Irma, José, and Katia) hit the Caribbean at the same time. In this brief example we use `rWind` to display hurricanes in a straightforward way (Fig. 2). In the Supplementary material Appendix 1–2, we show with this example how `rWind` can be used to export .png images prepared to be converted in a GIF animation (Supplementary material Appendix 1 and 2).

```
First, use lubridate R package to create a sequence of dates.
dt <- seq(ymd_hms(paste(2017, 9, 3,
  00, 00, 00, sep="-")),
  ymd_hms(paste(2017, 9, 11, 21, 00, 00,
  sep="-")), by="3 hours")
```

Then, we download the data using `wind.dl_2` and the sequence of dates, and we create the raster stacks for each date and time

```
wind_series <- wind.dl_2(dt, -103,
  -53, 13, 32)
wind_series_layer
  <- wind2raster(wind_series)
```

Finally, plot the hurricanes.

```
image.plot(wind_series_
  layer[[45]]$wind.speed,
  col=bpy.colors(1000, alpha=0.8),
  zlim=c(0, 40),
  main=wind_series[[45]]$time[1],
  xlab="Longitude",
  ylab="Latitude", cex.lab=1.5, cex.
  axis=1.5)
alpha <- arrowDir(wind_series[[45]])
Arrowhead(wind_series[[45]]$lon, wind_
  series[[45]]$lat, angle=alpha,
  arr.length=0.12, arr.type="curved")
lines(getMap(resolution="low"), lwd=2)
text(-99, 23.5, labels="Katia", cex=2,
  col="white", font=2)
text(-71, 25, labels="Irma", cex=2,
  col="white", font=2)
text(-59, 19.5, labels="José", cex=2,
  col="white", font=2)
```

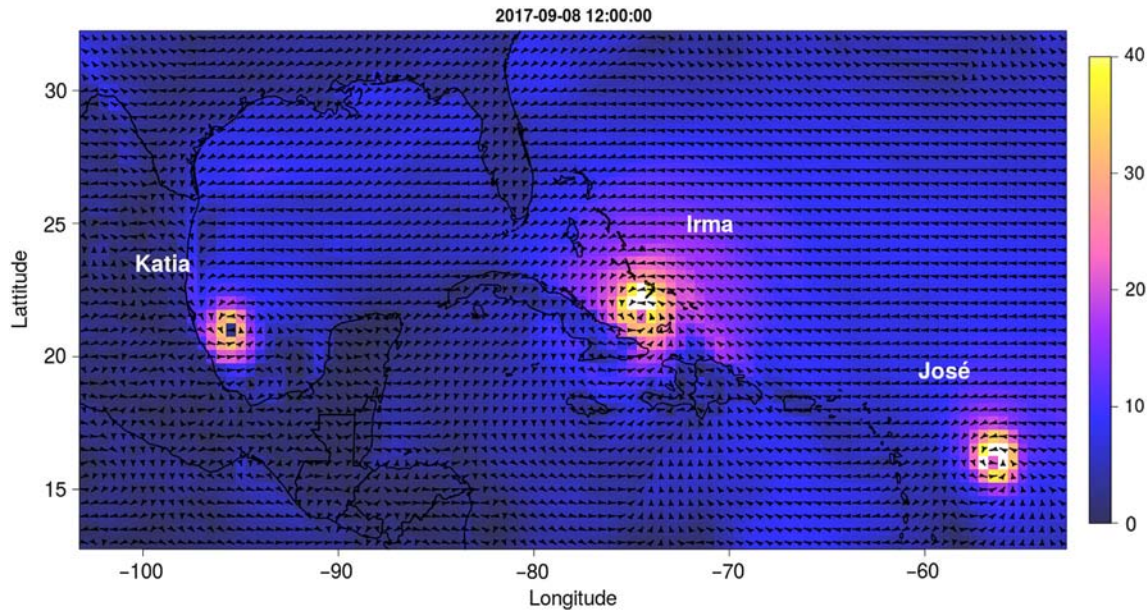


Figure 2. Snapshot of the wind speed and direction from 8 September 2017 showing hurricanes Irma, José, and Katia. Wind speed is given in meters per second and a cut off at 40 m s^{-1} was added.

Example 3: measuring wind connectivity between northwestern Africa and southern Macaronesian islands (Canary Islands and Cape Verde)

In this example, we focus on *Periploca laevigata*, a Mediterranean wind-dispersed shrub (Zito et al. 2015) found in the southern Mediterranean and west African regions, and on the Macaronesian Islands. Specifically, we compare the genetic structure of northwestern African and Macaronesian populations of *P. laevigata* obtained by García-Verdugo et al. (2017) with wind connectivity between those areas computed with rWind.

In their research, García-Verdugo et al. detected a close genetic relation between northwestern African populations, eastern Canary Islands populations and Cape Verde populations (Fig. 2A, B, C in García-Verdugo et al. 2017). In this example we compute wind connectivity from locations sampled on the African mainland by García-Verdugo et al. (2017) with their sampled islands of Fuerteventura (eastern Canary Islands), Gran Canaria and Tenerife (central Canary Islands), La Palma (western Canary Islands), and Santo Antão and Fogo (Cape Verde). A complete script with analyses and plots created for this example is included in the Supplementary material Appendix 1–3. Figure 3 shows a wind-connectivity graph from the mainland Africa locations (AGA, TAN, WSAH_A, WSAH_B, see Appendix S1 in García-Verdugo et al. (2017)) to all the island locations. Our analyses showed that wind connectivity observed between mainland Africa locations and Cape Verde islands is higher than those between Africa and Canary Islands. Western/central Canary Islands showed the lowest values of wind connectivity, while the eastern Canary Islands were connected only with Moroccan mainland. These results are in agreement with the *P. laevigata* genetic structure measured in Fig. 2A,

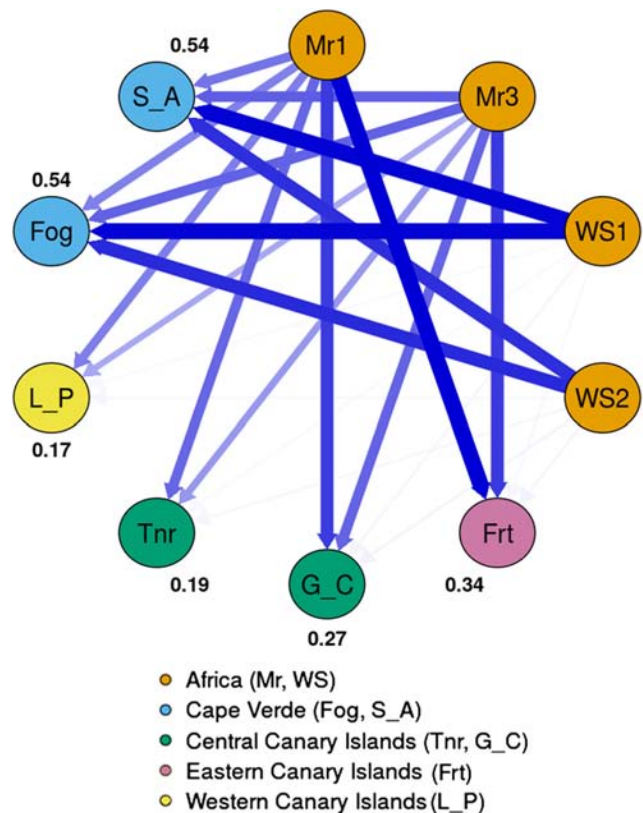


Figure 3. Wind connectivity graph from northwestern Africa mainland and southern Macaronesian Islands for the period of April–August of 2012–2017. This graph shows a higher connectivity of north western Africa mainland (Mr1, Mr3, WS1, WS2) with Cape Verde islands (Fog, S_A) than with Canary Islands (Frt, G_C, Tnr, L_P), even when the distance between African and Canary Island locations is smaller. Mainland connectivity average is showed over each island locality (connectivity is measured in arbitrary units $\times 100$).

B, C in García-Verdugo et al. (2017), suggesting wind connection may play a role in genetic structuring. Although in this simple example several important issues are not taken into account, such as spatio-temporal scales and the lack of a specific statistical framework (e.g. Mantel test, Mantel 1967), this preliminary analysis shows how rWind can be useful in the formulation of new hypotheses in biogeographical studies.

Conclusions

Wind is known to be a key factor underlying many ecological, evolutionary and, particularly, biogeographical processes. Therefore it is important to include wind data in analyses of evolutionary history, dispersal, and phylogeography to help understand and test the role that wind plays in shaping biological patterns. rWind provides new tools to include wind data in ecological, evolutionary, and biogeographic studies, computing connectivity matrices that can be easily applied to many existing analyses, from landscape ecology to bird migration models. Although other software exists to manage atmospheric data (RNCEP (Kemp et al. 2012a), IDV (Murray et al. 2003)), rWind uses a simpler model to compute wind mediated connectivity which does not require additional data. Moreover, rWind is specifically designed to interact with other R packages such as raster and gdistance (Etten 2017), which allows it to take advantage of the diverse functionality of these libraries, and to easily export wind data in a raster format to be used in a Geographic Information System (GIS) environment. In addition, it also provides the option to export data as plain text files, and therefore to be ported into any other software. We plan to extend functionalities of rWind to model connectivity from other sources such as sea currents and fluvial networks.

Software availability

The stable version of rWind is released regularly on the Comprehensive R Archive Network (CRAN): <<https://CRAN.R-project.org/package=rWind>> and can be installed in R by typing the following command: `install.packages("rWind")`.

The development version of rWind is hosted on github: <<https://github.com/jabiologo/rWind>>, rWind is distributed under GNU Public Licence (GPL) ver. 3 or greater.

Further examples can be found on the blog of the first author: <<http://allthiswasfield.blogspot.com/>>.

To cite rWind or acknowledge its use, cite this Software note as follows, substituting the version of the application that you used for 'version 0':

Fernández-López, J. and Schliep, K. 2019. rWind: download, edit and include wind data in ecological and evolutionary analysis. – *Ecography* 42: 000–000 (ver. 0).

Acknowledgements – Wind data in rWind are kindly provided by the Pacific Islands Ocean Observing System (PacIOOS), funded through the National Oceanic and Atmospheric Administration

(NOAA) as a Regional Association within the U.S. Integrated Ocean Observing System (IOOS). PacIOOS is coordinated by the Univ. of Hawaii School of Ocean and Earth Science and Technology (SOEST). We thank Conor Delaney and Jacob van Etten for help and advices with weather data and the raster and gdistance packages. We also thanks Kristin M. Winchell and Liam Revell for their comments and suggestions. Brody Sandel and another anonymous reviewer contributed to improve this manuscript.

Funding – JF-L was supported by Predoctoral Grants from the Spanish Ministry of Economy and Competitiveness (CGL2012-35559; CGL2015-67459-P; BES-2013-066429). This work was partially made during a JF-L short stay (EEBB-I-17-12634), hosted in Revell Lab at Univ. of Massachusetts, Boston. This work was supported in part by a grant from the National Science Foundation (DEB 1350474 to Liam Revell).

Author contributions – JF-L conceived the ideas and wrote the first version of rWind R package; KS improved rWind functions and wrote new code to increase the package performance; JF-L and KS led the writing of the manuscript. Both authors contributed critically to the drafts and gave final approval for publication.

Conflicts of interest – Nothing to disclosure.

References

- Bernis, F. and Tellería, J. L. 1981. La migración de las aves en el Estrecho de Gibraltar: (época posnupcial): migración de cigüeñas y falconiformes por Gibraltar v. 2. La migración de las aves no planeadoras por Gibraltar. – *Cátedra de Zoología de Vertebrados, Facultad de Biología, Univ. Complutense*.
- Calsbeek, R. and Smith, T. B. 2003. Ocean currents mediate evolution in island lizards. – *Nature* 426: 552–555.
- Dijkstra, E. W. 1959. A note on two problems in connexion with graphs. – *Numer. Math.* 1: 269–271.
- Ebach, M. C. and Williams, D. M. 2010. Systematics and biogeography: cladistics and vicariance. – *Syst. Biol.* 59: 612–614.
- Etten, J. van. 2017. R Package gdistance: distances and routes on geographical grids. – *J. Stat. Softw.* 76: 1–21.
- Felicísimo, Á. M. et al. 2008. Ocean surface winds drive dynamics of transoceanic aerial movements. – *PLoS One* 3: e2928.
- Freeman, J. A. 1945. Studies in the distribution of insects by aerial currents. – *J. Anim. Ecol.* 14: 128–154.
- García-Verdugo, C. et al. 2017. The loss of dispersal on islands hypothesis revisited: implementing phylogeography to investigate evolution of dispersal traits in *Periploca* (Apocynaceae). – *J. Biogeogr.* 44: 2595–2606.
- González-Solís, J. et al. 2009. Influence of sea surface winds on shearwater migration detours. – *Mar. Ecol. Prog. Ser.* 391: 221–230.
- Harvey, L. E. 1994. Spatial patterns of inter-island plant and bird species movements in the Galápagos Islands. – *J. R. Soc. N. Z.* 24: 45–63.
- Hijmans, R. J. 2017. raster: geographic data analysis and modeling. – R package ver. 2.6-7, <<https://CRAN.R-project.org/package=raster>>.
- Hooker, J. D. et al. 1844. The botany of the Antarctic voyage of H.M. discovery ships Erebus and Terror in the years 1839–1843: under the command of Captain Sir James Clark Ross. – Reeve Brothers, London.
- Juan, C. et al. 2000. Colonization and diversification: towards a phylogeographic synthesis for the Canary Islands. – *Trends Ecol. Evol.* 15: 104–109.

- Kemp, M. U. et al. 2010. Can wind help explain seasonal differences in avian migration speed? – *J. Avian Biol.* 41: 672–677.
- Kemp, M. U. et al. 2012a. RNCEP: global weather and climate data at your fingertips. – *Methods Ecol. Evol.* 3: 65–70.
- Kemp, M. U. et al. 2012b. Quantifying flow-assistance and implications for movement research. – *J. Theor. Biol.* 308: 56–67.
- Leonard, P. B. et al. 2017. gflow: software for modelling circuit theory-based connectivity at any scale. – *Methods Ecol. Evol.* 8: 519–526.
- Mantel, N. 1967 The detection of disease clustering and a generalized regression approach. – *Cancer Res.* 27: 209–220.
- McRae, B. H. 2006. Isolation by resistance. – *Evolution* 60: 1551–1561.
- McRae, B. H. et al. 2008. Using circuit theory to model connectivity in ecology, evolution, and conservation. – *Ecology* 89: 2712–2724.
- Morrone, J. J. and Crisci, J. V. 1995. Historical biogeography: introduction to methods. – *Annu. Rev. Ecol. Syst.* 26: 373–401.
- Muñoz, J. et al. 2004. Wind as a long-distance dispersal vehicle in the Southern Hemisphere. – *Science* 304: 1144–1147.
- Murray, D. et al. 2003. The integrated data viewer: a web-enabled application for scientific analysis and visualization. – 19th Int. Conf. on Interactive Information and Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology, Long Beach, CA, Am. Meteorol. Soc.
- Nathan, R. 2006. Long-distance dispersal of plants. – *Science* 313: 786–788.
- Queiroz, A. 2005. The resurrection of oceanic dispersal in historical biogeography. – *Trends Ecol. Evol.* 20: 68–73.
- Richardson, W. 1990. Wind and orientation of migrating birds: a review. – *Experientia* 46: 416–425.
- Sanmartín, I. et al. 2007. West wind drift revisited: testing for directional dispersal in the Southern Hemisphere using event-based tree fitting. – *J. Biogeogr.* 34: 398–416.
- Shamoun-Baranes, J. et al. 2010. Integrating meteorology into research on migration. – *Integr. Comp. Biol.* 50: 280–292.
- Tøttrup, A. P. et al. 2017. Migration of red-backed shrikes from the Iberian Peninsula: optimal or sub-optimal detour? – *J. Avian Biol.* 48: 149–154.
- Vansteelant, W. M. G. et al. 2017. Wind conditions and geography shape the first outbound migration of juvenile honey buzzards and their distribution across sub-Saharan Africa. – *Proc. R. Soc. B* 284: 20170387.
- Winkworth, R. C. et al. 2002. Plant dispersal N.E.W.S from New Zealand. – *Trends Ecol. Evol.* 17: 514–520.
- Zito, P. et al. 2015. Floral volatiles in a *Sapromyiophilous* plant and their importance in attracting house fly pollinators. – *J. Chem. Ecol.* 41: 340–349.

Supplementary material (Appendix ECOG-03730 at <www.ecogeography.org/appendix/ecog-03730>). Appendix 1–2.