FULL TITLE: rWind: Download, edit and include wind data in ecological and evolutionary analysis.

AUTHOR LIST

Javier Fernández-López, Real Jardín Botánico (CSIC), Madrid, Spain.

ORCiD ID: https://orcid.org/0000-0003-4352-0252

Klaus Schliep, University of Massachusetts Boston, Boston MA, USA

ORCiD ID: https://orcid.org/0000-0003-2941-0161

Corresponding author: Javier Fernández-López E-mail: javfer05@ucm.es

ABSTRACT

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 langauge 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

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) version 3 or greater.

Further examples can be found on the blog of the first author:

http://allthiswasfield.blogspot.com/

DECLARATIONS

Acknowledgments – 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 University 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. Dr Brody Sandel and another anonymous reviewer contributed to improve this manuscript.

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

Author Contributions – J.F.L. conceived the ideas and wrote the first version of rWind R package; K.S. improved rWind functions and wrote new code to increase the package performance; J.F.L. and K.S. 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.

FULL TITLE: rWind: Download, edit and include wind data in ecological and evolutionary analysis.

ABSTRACT

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 langauge 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, Fitch and Brothers 1844, Freeman 1945, Winkworth et al. 2002, Sanmartín, Wanntorp and Winkworth 2007). For example, wind currents can play a decisive role in driving patterns in bird migration (Felicísimo, Muñoz and González-Solis 2008, Vansteelant, Kekkonen and Byholm 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 (J J 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, Bouten and van Loon 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 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 (Etten 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, personal communication). 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 2008), 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 funcionality 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 meters 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 May 6th 2011 and is updated daily. In rWind, these data are obtained via queries to The Pacific Islands Ocean Observing System, coordinated by the University 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.

[table 1 about here]

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 the algorithm proposed by Muñoz et al. (2004) and their variation in Felicísimo, Muñoz and González-Solis (2008) (Equations adapted from Felicísimo, Muñoz and González-Solis 2008, González-Solís et al. 2009, Muñoz et al., 2004),

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

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

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

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 Horizontal Relative Moving Angle 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, Horizontal Factor is equal to two times Horizontal Relative Moving Angle (HRMA) (see equation 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.

To compute "passive" movement cost, that is avoiding movement against wind, we use a variation of equation 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 Horizontal Factor is set to ∞ for all cases in which the Horizontal Relative Moving Angle is more than 90 degrees (Muñoz et al. 2004).

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"transitionLayer" 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 from 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 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 the Supporting Information 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, 2015 February 2nd at 12:00PM.)
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, "active", "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(sl, col = terrain.colors(10), zlim = c(0,7),
 xlab = "longitude", ylab = "latitude")
lines(getMap(resolution = "low"), lwd = 4)
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")
```

[Fig. 1 about here]

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 Supporting Information, we show with this example how rWind can be used to export .png images prepared to be converted in a GIF animation (see Supporting Information 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 downlad 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 = "Lattitude", 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)
```

[Fig. 2 about here]

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

In this example, we focus on *Periploca laevigata* (Aiton, 1789), a Mediterranenan wind-dispersed shrub (Zito, Dötterl and Sajeva 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 (García-Verdugo et al. 2017, figure 2-A-B-C). 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 Supporting Information 1. Fig. 3 shows a wind-connectivity graph from the mainland Africa locations (AGA, TAN, WSAH_A, WSAH_B, see Apendix 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 the with the *P. laevigata* genetic structure measured in García-Verdugo et al. (2017), figure 2-A-B-C, 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.

[Fig. 3 about here]

CONCLUSIONS

Wind is known to be a key factor underlying many ecological, evolutionary and, particularly, biogeographical processes and patterns. 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 of 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.

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. DOI: https://doi.org/10.1038/nature02143

Dijkstra, E. W. 1959. A note on two problems in connexion with graphs. – Numer. Math. 1: 269–271. DOI: 10.1007/BF01386390

Ebach, M. C. and Williams, D. M. 2010. Systematics and Biogeography: Cladistics and Vicariance. – Syst. Biol. 59: 612–614. https://doi.org/10.1093/sysbio/syq050

Etten, J. van. 2017. R Package gdistance: distances and routes on geographical grids. – J. Stat. Softw. 76: 1–21. DOI: 10.18637/jss.v076.i13

Felicísimo, Á. M., Muñoz, J. and González-Solis, J. 2008. Ocean surface winds drive dynamics of transoceanic aerial movements. – PLoS One 3: 1–7. DOI: https://doi.org/10.1371/journal.pone.0002928

Freeman, J. A. 1945. Studies in the distribution of insects by aerial currents. – J. Anim. Ecol. 14: 128–154. DOI: 10.1111/j.1365-2311.1986.tb00304.x

García-Verdugo, C., Mairal, M., Monroy, P., Sajeva, M. and Caujapé-Castells, J. 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. DOI: https://doi.org/10.1111/jbi.13050

González-Solís, J., Felicísimo, A., Fox, J. W., Afanasyev, V., Kolbeinsson, Y. and Muñoz, J. 2009. Influence of sea surface winds on shearwater migration detours. – Mar. Ecol. Prog. Ser. 391: 221–230. DOI: https://doi.org/10.3354/meps08128

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.

Hooker, J. D., Fitch, W. H. and Brothers, R. 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. – London :Reeve Brothers,1844-1860. DOI: https://doi.org/10.5962/bhl.title.16029

Morrone, J. J. and Crisci, J. V. 1995. Historical Biogeography: Introduction to Methods. – Annu. Rev. Ecol. Syst. 26: 373–401. DOI: https://doi.org/10.1146/annurev.es.26.110195.002105

Juan, C., Emerson, B. C., Oromí, P. and Hewitt, G. M. 2000. Colonization and diversification: towards a phylogeographic synthesis for the Canary Islands. – Trends Ecol. Evol. 15: 104–109. DOI: https://doi.org/10.1016/S0169-5347(99)01776-0

Kemp, M. U., Shamoun-Baranes, J., Van Gasteren, H., Bouten, W. and Loon, E.E. 2010. Can wind help explain seasonal differences in avian migration speed? – J. Avian Biol. 41: 672–677. DOI: https://doi.org/10.1111/j.1600-048X.2010.05053.x

Kemp, M. U., Loon, E. E., Shamoun-Baranes, J. and Bouten, W. 2012a. RNCEP: global weather and climate data at your fingertips. – Methods Ecol. Evol. 3: 65–70. DOI: https://doi.org/10.1111/j.2041-210X.2011.00138.x

Kemp, M. U., Shamoun-Baranes, J., Loon, E. E., McLaren, J. D., Dokter, A. M. and Bouten, W. 2012b. Quantifying flow-assistance and implications for movement research. – J. Theor. Biol. 308: 56–67. DOI: 10.1016/j.jtbi.2012.05.026

Leonard, P. B., Duffy, E. B., Baldwin, R. F., McRae, B. H., Shah, V. B. and Mohapatra, T. K. 2017. gflow: software for modelling circuit theory-based connectivity at any scale. – Methods Ecol. Evol. 8: 519–526. DOI: https://doi.org/10.1111/2041-210X.12689

Mantel, N. 1967 The detection of disease clustering and a generalized regression approach. – Cancer Res. 27: 209–220. DOI: 10.1158/0008-5472

McRae, B. H. 2006. Isolation by resistance. – Evolution 60: 1551–1561. DOI: https://doi.org/10.1111/j.0014-3820.2006.tb00500.x

McRae, B. H., Dickson, B. G., Keitt, T. H. and Shah, V. B. 2008. Using circuit theory to model connectivity in ecology, evolution, and conservation. – Ecology 89: 2712–2724. DOI:

https://doi.org/10.1890/07-1861.1

Muñoz, J., Felicísimo, A. M., Cabezas, F., Burgaz, A. R. and Martínez, I. 2004. Wind as a long-distance dispersal vehicle in the Southern Hemisphere. – Science 304: 1144–1147. DOI: 10.1126/science.1095210

Murray, D., McWhirter, J., Wier, S. and Emmerson, S. 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, Amer. Meteor. Soc.

Nathan, R. 2006. Long-distance dispersal of plants. – Science 313: 786–788. DOI: 10.1126/science.1124975

Queiroz, A. 2005. The resurrection of oceanic dispersal in historical biogeography. – Trends Ecol. Evol. 20: 68–73. DOI: https://doi.org/10.1016/j.tree.2004.11.006

R Development Core Team. 2008. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Richardson, W. 1990. Wind and orientation of migrating birds: a review. – Experientia 46: 416–425. DOI: 10.1007/BF01952175

Sanmartín, I., Wanntorp, L. and Winkworth, R. C. 2007. West Wind Drift revisited: testing for directional dispersal in the Southern Hemisphere using event-based tree fitting. – J Biogeogr 34: 398–416. DOI: https://doi.org/10.1111/j.1365-2699.2006.01655.x

Shamoun-Baranes, J., Bouten, W. and Loon, E. E. 2010. Integrating Meteorology into research on migration. – Integr. Comp. Biol. 50: 280–292. DOI: 10.1093/icb/icq011

Tøttrup, A. P., Pedersen, L., Onrubia, A., Klaassen, R. H. G. and Thorup, K. 2017. Migration of red-backed shrikes from the Iberian Peninsula: optimal or sub-optimal detour? – J. Avian Biol., 48: 149–154. DOI: 10.1111/jav.01352

Vansteelant, W. M. G., Kekkonen, J. and Byholm, P. 2017. Wind conditions and geography shape the first outbound migration of juvenile honey buzzards and their distribution across sub-Saharan Africa. – Proc Biol Sci. 284: 20170387 DOI: 10.1098/rspb.2017.0387

Winkworth, R. C., Wagstaff, S. J., Glenny, D. and Lockhart, P. J. 2002. Plant dispersal N.E.W.S from New Zealand. – Trends Ecol. Evol. 17: 514–520. DOI: https://doi.org/10.1016/S0169-5347(02)02590-9

Zito, P., Dötterl, S. and Sajeva, M. 2015. Floral volatiles in a *Sapromyiophilous* plant and their importance in attracting house fly pollinators. – J. Chem. Ecol. 41: 340–349. DOI: 10.1007/s10886-015-0568-8

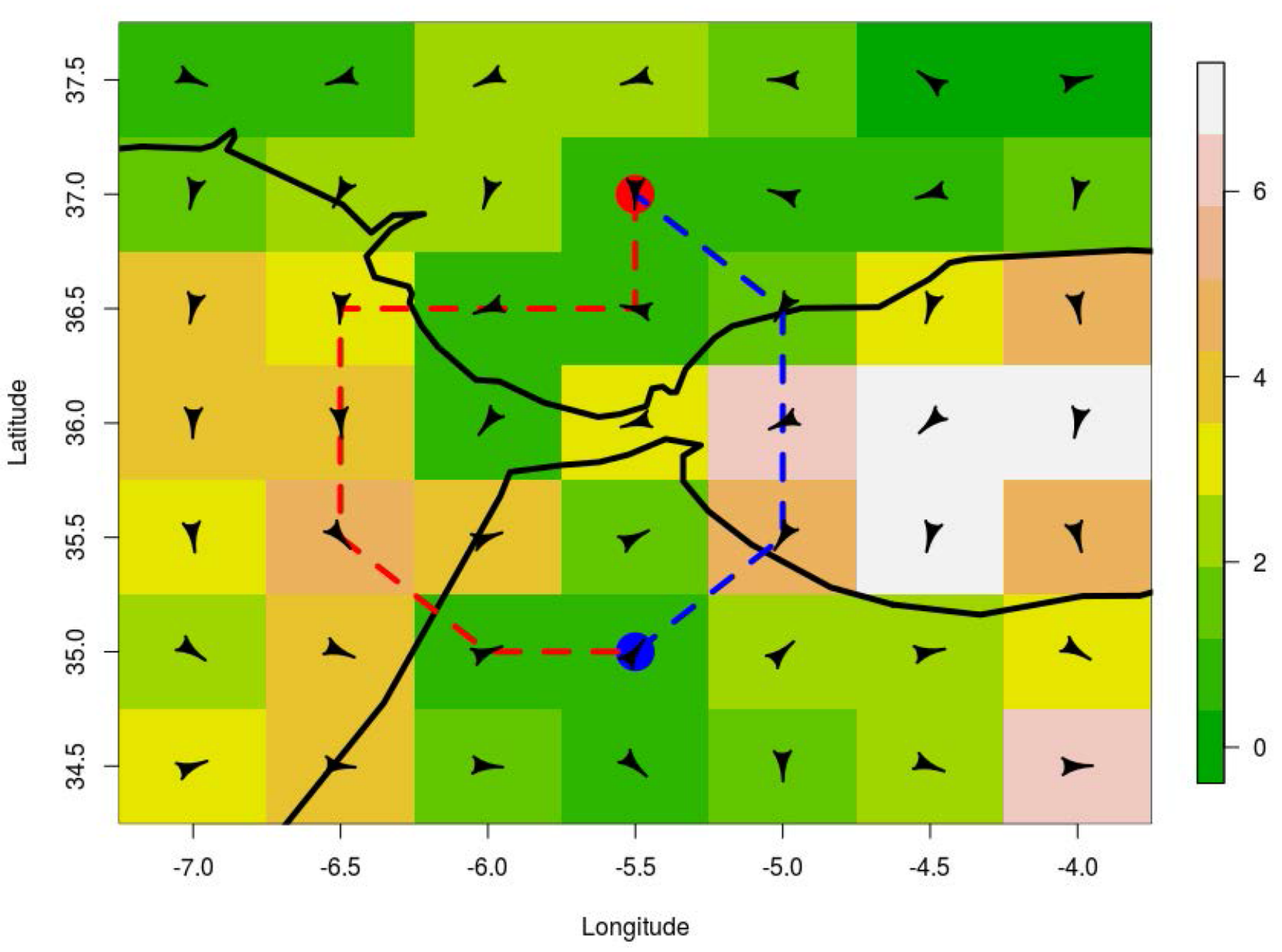
Table 1: Brief description of the main functions included in the R library rWind.

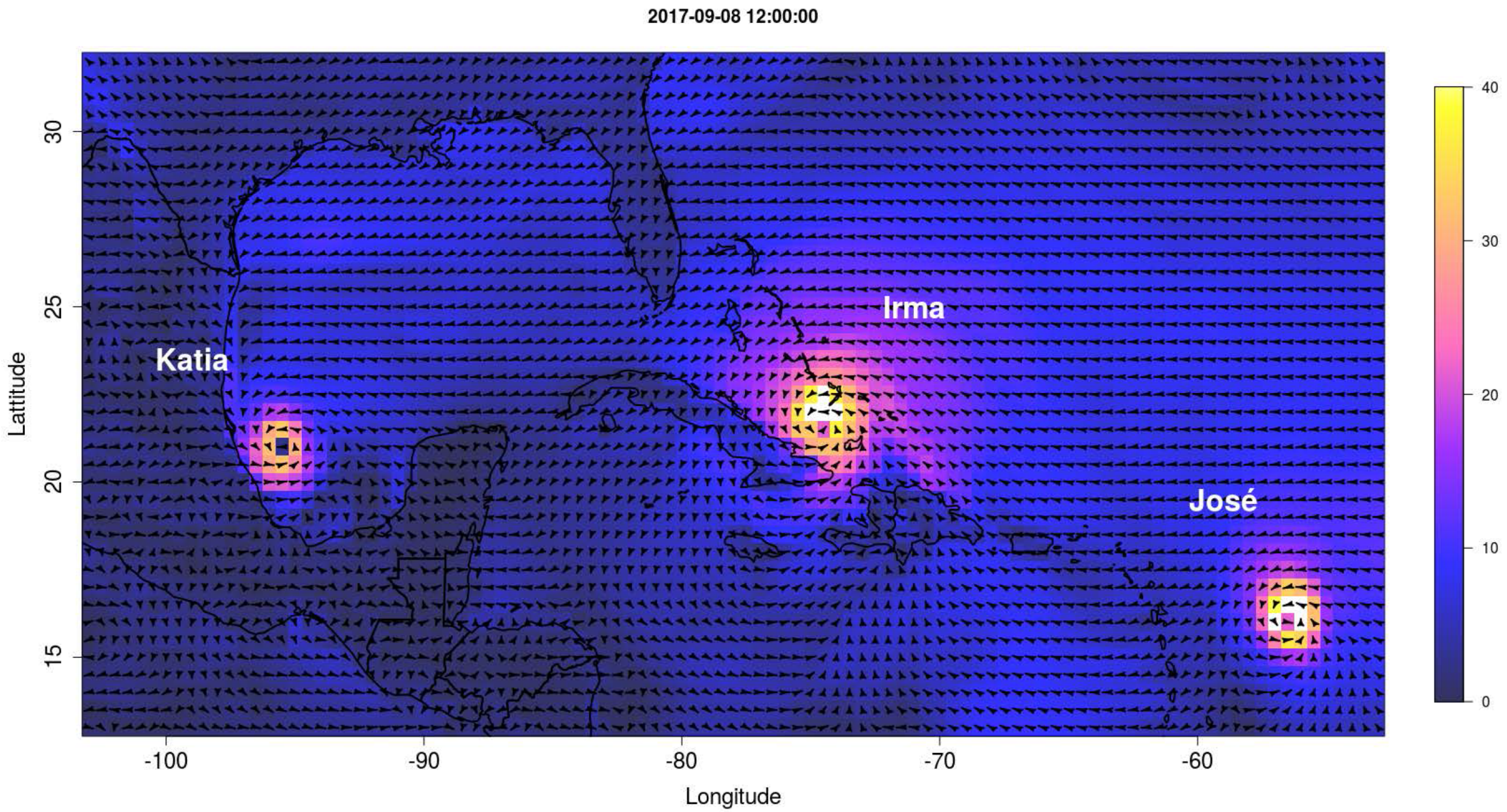
function	description
wind.dl wind.dl_2	downloads wind data from the Global Forecast System (GFS) of the USA's National Weather Service (NWS) (https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/global-forcast-system-gfs) and returns either a .csv file or a data.frame.
wind.mean	Takes a list of wind data downloaded with wind.dl_2 and retunrs 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 crates 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. It returns either, a sparse cost matrix or a conductance TransitionLayer object.

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

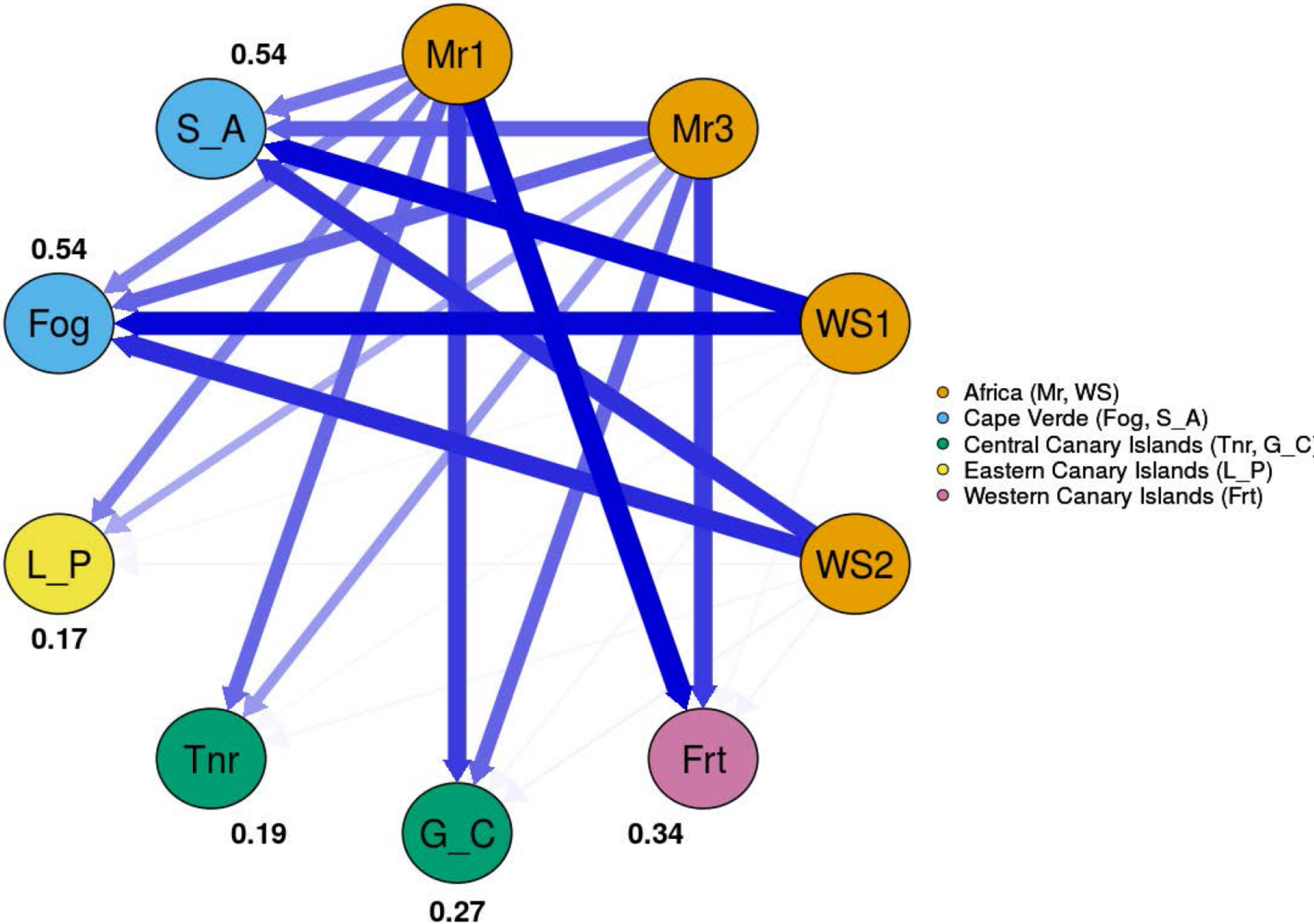
Figure 2: Snapshot of the wind speed and direction from September 8th 2017 showing hurricanes Irma, José, and Katia. Wind speed is given in meters per second and a cut off at 40 m/s was added.

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 × 100).


37.5
37.0
36.5
36.0
35.5
35.0
34.5
Latitude
-7.0
-6.5
-6.0
-5.5
-5.0
-4.5
-4.0
Longitude
6
4
2
0



2017-09-08 12:00:00
Katia
Irma
José
Lattitude
Longitude
30
25
20
15
-100
-90
-80
-70
-60
40
30
20
10
0


Mr1
Mr3
S_A
Fog
WS1
WS2
L_P
Tnr
G_C
Frt
0.54
0.54
0.17
0.19
0.27
0.34
Africa (Mr, WS)
Cape Verde (Fog, S_A)
Central Canary Islands (Tnr, G_C)
Eastern Canary Islands (L_P)
Western Canary Islands (Frt)

rWind vignette

LOADING PACKAGES

First, we load the main packages we will use in this vignette. This vignette was written under the rWind version 1.0.0

```
# use install.packages() if some is not installed
# and you can install the latest development version using the command
# devtools::install_github("jabiologo/rWind")
library(rWind)
library(raster)
library(gdistance)
```

EXAMPLE 1: Anisotropic shortest paths across Strait of Gibraltar

In this simple example, we introduce the most basic functionality of rWind, to get the shortest paths between two points across Strait of Gibraltar. Notice that, as wind connectivity is anisotropic (direction dependent), shortest path from A to B usually does not match with shortest path from B to A.

First, we will download wind data of a selected date (e.g. 2015 February 12th) and we will fix and transform them into two raster layers, with values of wind direction and wind speed.

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

```
## [1] "2015-02-12"
## [1] "2015-02-12 12:00:00 downloading..."
```

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

Then, we will use `flow.dispersion` function to obtain a transitionLayer object with conductance values, which will be used later to obtain the shortest paths.

```
Conductance<-flow.dispersion(wind_layer,"active", "transitionLayer")
```

Now, we will use `shortestPath` function from `gdistance` package to compute shortest path from our `Conductance` object between the two selected points.

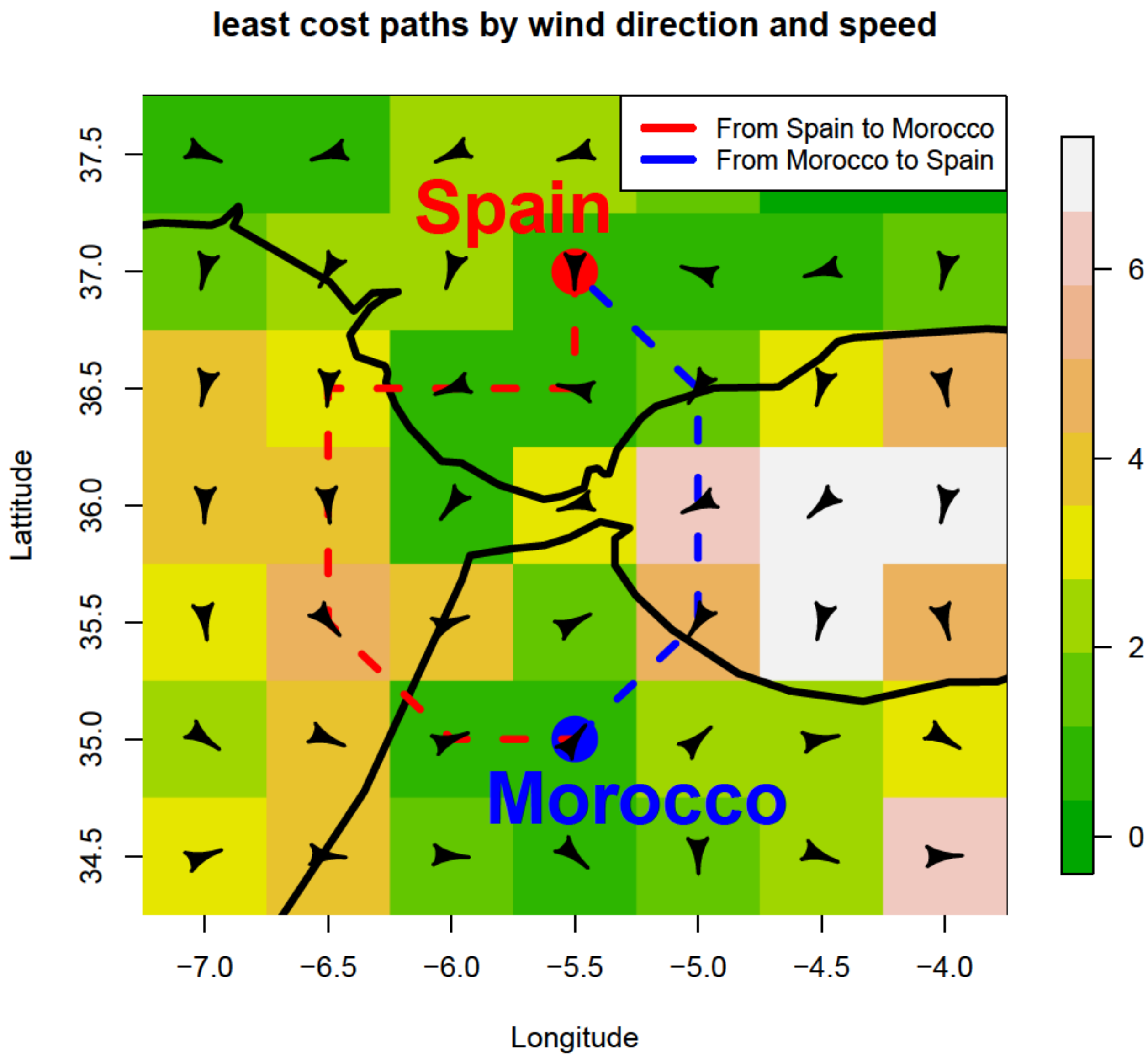
```
AtoB<- shortestPath(Conductance,
                    c(-5.5, 37), c(-5.5, 35), output="SpatialLines")
BtoA<- shortestPath(Conductance,
                    c(-5.5, 35), c(-5.5, 37), output="SpatialLines")
```

Finally, we plot the map and we will add the shortest paths as lines and some other features.

```
library(fields)
library(shape)
library(rworldmap)
```

```
image.plot(wind_layer$wind.speed, main="least cost paths by wind direction and speed",
     col=terrain.colors(10), xlab="Longitude", ylab="Lattitude", zlim=c(0,7))

lines(getMap(resolution = "low"), lwd=4)

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)

alpha <- arrowDir(w)
Arrowhead(w$lon, w$lat, angle=alpha, arr.length = 0.4, arr.type="curved")

text(-5.75, 37.25,labels="Spain", cex= 2.5, col="red", font=2)
text(-5.25, 34.75,labels="Morocco", cex= 2.5, col="blue", font=2)
legend("topright", legend = c("From Spain to Morocco", "From Morocco to Spain"),
    lwd=4 ,lty = 1, col=c("red","blue"), cex=0.9, bg="white")
```



EXAMPLE 2: Hurricanes visualization

This example shows how to download a time series data with rWind and edit them to obtain a gif map with wind speed and directions. For this example, we will use hurricanes wind data ocurred during the first days of September 2017.

First, we will download wind data from 3rd to 11th of September 2017 by each three hours using wind.dl_2. To do that, we will use **lubridate** package to create a list of dates/times to be used by wind.dl_2. It could take a while…

```
library(lubridate)
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")
wind_series <- wind.dl_2(dt,-103,-53,13,32)
```

In a second step, we will obtain raster layers for wind speed and direction of each date-time downloaded. wind2raster can take a list as argument.

```
wind_series_layer <- wind2raster(wind_series)
```

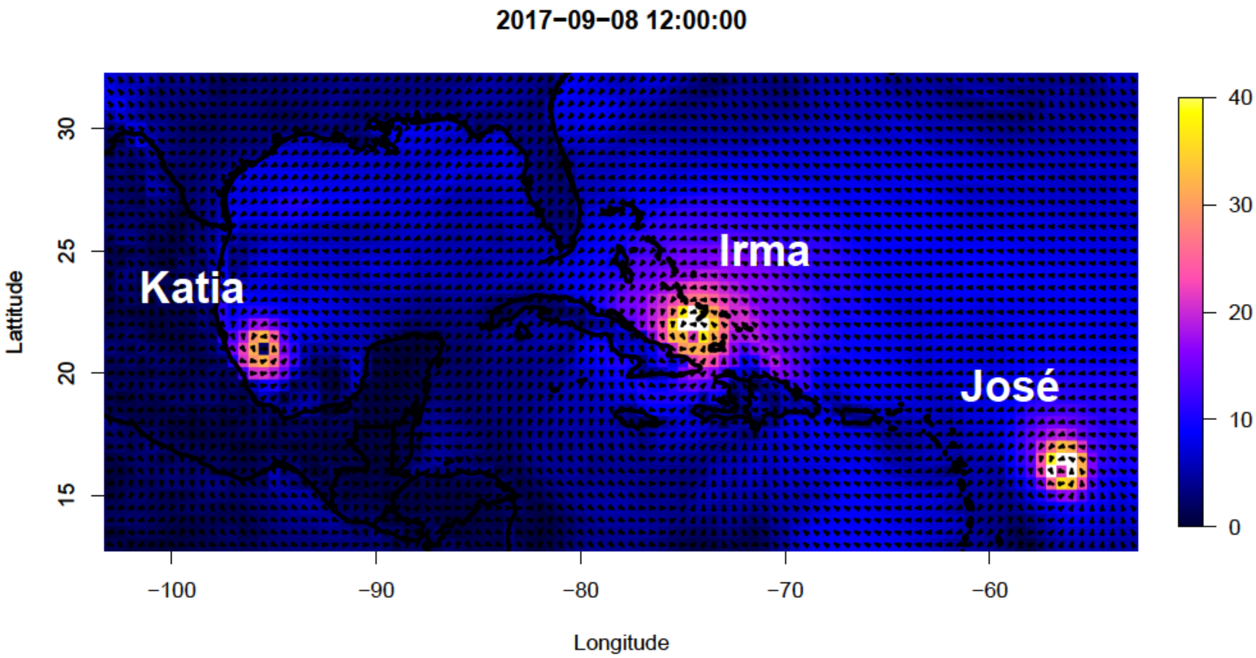
Finally, we will export the entire time series as PNG format. You should check your current work directory (`getwd`) to know where the PNG files will be stored.

```
id<-0
for (i in 1:72) {
  id <- sprintf("%03d", i)
  png(paste("hurricane",id,".png", sep=""), width=1100, height=570, units="px",
      pointsize=18)
  image.plot(wind_series_layer[[i]]$wind.speed, col=bpy.colors(1000),
             zlim=c(0,40), main =wind_series[[i]]$time[1])
  lines(getMap(resolution = "low"), lwd=3)
  dev.off()
}
```

The exported PNG files can be converted into a GIF format using several softwares. You can use imagmagick `convert -delay 10 *.png hurricane.gif` which will result in an animation like the Supporting Information 2

The following code is used to create figure 2 in the manuscript.

```
image.plot(wind_series_layer[[45]]$wind.speed, col=bpy.colors(1000), zlim=c(0,40),
           main =wind_series[[45]]$time[1], xlab="Longitude",
           ylab="Lattitude")
alpha <- arrowDir(wind_series[[45]])
Arrowhead(wind_series[[45]]$lon, wind_series[[45]]$lat, angle=alpha, arr.length = 0.07,
          arr.type="curved")
lines(getMap(resolution = "low"), lwd=3)
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)
```



In addition, we can compute some statistics from the wind data downloaded.

First, we can use **wind.mean** to compute speed and direction averages for each location in our study area:

```
mean_wind <- wind.mean(wind_series)
head(mean_wind)
```

```
##         time lat   lon    ugrd10m    vgrd10m      dir    speed
## 1 2017-09-03  13 257.0  0.6798113 -2.2501722 163.1896 2.350621
## 2 2017-09-03  13 257.5 -2.1322760  0.7382131 289.0963 2.256448
## 3 2017-09-03  13 258.0 -1.9516941  1.1400587 300.2908 2.260275
## 4 2017-09-03  13 258.5 -1.9550353  1.1605026 300.6932 2.273528
## 5 2017-09-03  13 259.0 -2.0296657  1.1105695 298.6861 2.313635
## 6 2017-09-03  13 259.5 -2.1109363  1.0253549 295.9075 2.346786
```

We can also use **tidy** function to put all the data in a single data.frame and use **dplyr** package to compute maximum speed, for example:

```
t_wind_series <- tidy(wind_series)

library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:igraph':
##
##     as_data_frame, groups, union

## The following objects are masked from 'package:raster':
##
##     intersect, select, union

## The following objects are masked from 'package:stats':
##
##     filter, lag
```

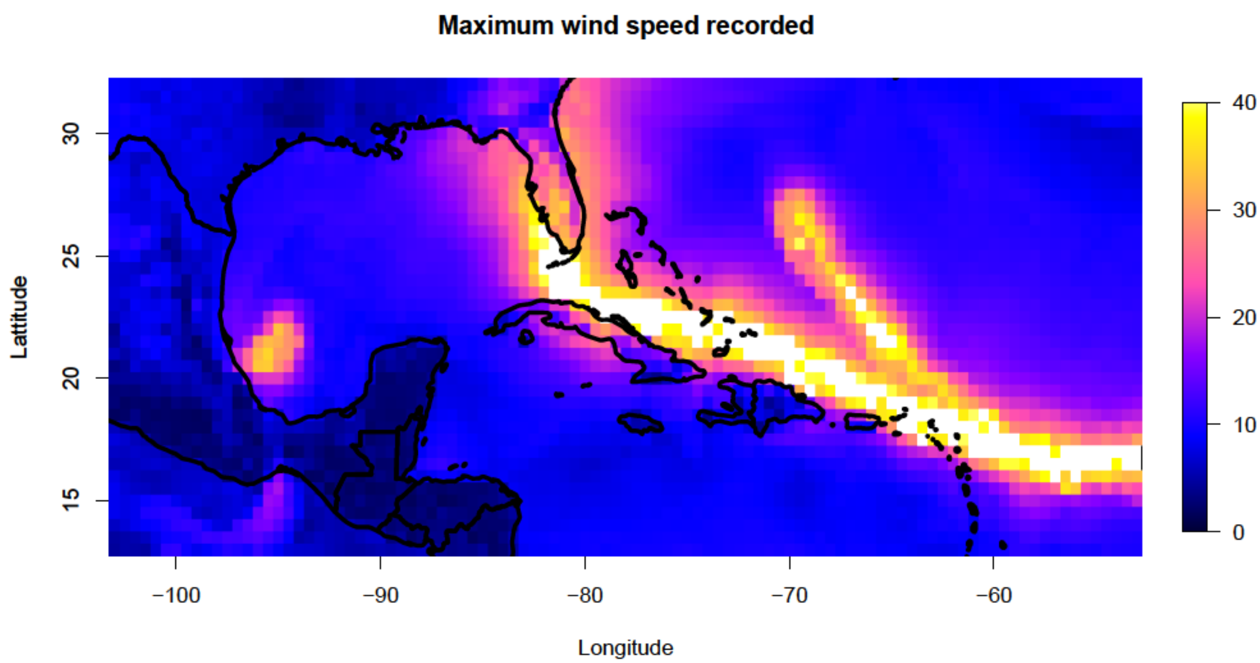
```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
g_wind_series <- t_wind_series %>% group_by(lat, lon)

max_ww <- g_wind_series %>% summarise(speed = max(speed))

maxw <- cbind(max_ww$lon, max_ww$lat, max_ww$speed)

head(maxw)
```

```
##         [,1] [,2]      [,3]
## [1,] -103.0   13  4.509488
## [2,] -102.5   13  5.205593
## [3,] -102.0   13  5.026556
## [4,] -101.5   13  4.746515
## [5,] -101.0   13  7.129943
## [6,] -100.5   13 11.128306
```

Finally, we can transform this data into a raster file to be ploted and "track" the maximum speed recorded in our study area for our time lapse.

```r
rmax <- rasterFromXYZ(maxw)

image.plot(rmax, col=bpy.colors(1000), zlim=c(0,40),
           main = "Maximum wind speed recorded", xlab="Longitude",
           ylab="Lattitude")
lines(getMap(resolution = "low"), lwd=3)
```



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

Wind currents are often thought to play a key role in species dispersal processes, particulary in oceanic island colonization. In this example, we will compare genetic structure of the wind-dispersed shrub *Periploca laevigata* obtained by García-Verdugo et al. (2017) with wind connectivity between northwestern Africa and Canary and Cape Verde archipelagos using rWind package. 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, etc.), this preliminary analysis shows how rWind can be useful in the formulation of new hypotheses in biogeographical studies.

First, we set the locations used by García-Verdugo et al. (2017). Since the rWind data resolution is about 50 km, some close sampled locations were merged.

```
loc <- matrix(c(-9.4729, -11.0422, -14.0443, -13.0395, -14.0000, -15.6000,
                -16.6000, -17.8600, -24.3800, -25.1800, 30.3331,  28.3376,
                26.2368, 26.3028, 28.4000, 28.0000, 28.2700, 28.7300, 14.9300,
                17.0700), 10, 2)
colnames(loc) <- c("lon", "lat")
rownames(loc) <- c("Morocco1", "Morocco3", "WSahara1", "WSahara2",
                   "Fuerteventura", "Gran_Canaria", "Tenerife", "La_Palma",
                   "Fogo", "San_Antonio")
```

Second, we define the temporal scale of wind data. Since *P. laevigata* fruits are available from spring to summer, we select May, June, July and August winds from 2012 to 2017. We sample wind once per day each 5 days. Wind sampling could be more intense, but we reduce it for computation reasons. The spatial window is set between -27 and -7 longitudinal degrees and 14 and 31 latitudinal degrees. We will get a total of 150 wind data for this region.

```
dt <- c(seq(ymd_hms(paste(2012,5,3,00,00,00, sep="-")),
            ymd_hms(paste(2012,8,31,12,00,00, sep="-")),by="5 days"),
        seq(ymd_hms(paste(2013,5,3,00,00,00, sep="-")),
            ymd_hms(paste(2013,8,31,12,00,00, sep="-")),by="5 days"),
        seq(ymd_hms(paste(2014,5,3,00,00,00, sep="-")),
            ymd_hms(paste(2014,8,31,12,00,00, sep="-")),by="5 days"),
        seq(ymd_hms(paste(2015,5,3,00,00,00, sep="-")),
            ymd_hms(paste(2015,8,31,12,00,00, sep="-")),by="5 days"),
        seq(ymd_hms(paste(2016,5,3,00,00,00, sep="-")),
            ymd_hms(paste(2016,8,31,12,00,00, sep="-")),by="5 days"),
        seq(ymd_hms(paste(2017,5,3,00,00,00, sep="-")),
            ymd_hms(paste(2017,8,31,12,00,00, sep="-")),by="5 days"))
```

Next, we create two objects to store costs obtained for each wind data between all locations and path lines between two selected locations as an example (Western Sahara and Santo Antao, Cape Verd)

```
paths <- list(1:150)
cost_list <- array(NA_real_, dim=c(10,10,150))
```

Now, we will execute the next actions:

1. Download wind data (`wind.dl_2()`).
2. Transform data in raster layers (speed and direction, `wind2raster()`).
3. Get `Conductance` matrices for each wind data downloaded.
4. Compute `Costs` between locations and store them in a `cost_list` object.
5. If the `Cost` is not `Inf`, get the shortest path between selected locations, and store it into `paths` object.

WARNING: Notice that you will download and managge 150 wind datasets, so it could take a while... you can reduce wind size sample (e.g. sampling each 10 days), but it could affect to connectivity values. We advice to use as much data as possible in order to obtain accurate results

```
w <- wind.dl_2(dt,-27,-7,14,31)

path_layers <- wind2raster(w)

Conductance <- flow.dispersion(path_layers,"passive", output="transitionLayer")

for (i in 1:150){
  cost_list[,,i] <- costDistance(Conductance[[i]],loc)
  if (costDistance(Conductance[[i]],loc[3,] , loc[9,]) != Inf){
    paths[[i]] <- shortestPath(Conductance[[i]], loc[3,], loc[9,],
                               output="SpatialLines")
  }
}

connectivity <- 1/cost_list
```

Now, we obtain the connectivity average between locations and rename columns and rows.

```
conn_avg <- apply(connectivity, c(1, 2), mean, na.rm = TRUE)
rownames(conn_avg) <- rownames(loc)
colnames(conn_avg) <- rownames(loc)
```

Since in this example we are specially interested in wind connectivity from Africa mainland to islands, we select just this part of the matrix. Then, we build a new matrix whit those values in the low triangle, ready to be plotted with **qgraph** R package

```
mat <- matrix(0,10,10)
mat[5:10,1:4] <- t(conn_avg[1:4,5:10])
colnames(mat) <- c(colnames(t(conn_avg[1:4,5:10])),
                   rownames(t(conn_avg[1:4,5:10])))
rownames(mat) <- c(colnames(t(conn_avg[1:4,5:10])),
                   rownames(t(conn_avg[1:4,5:10])))
mat[5:10,1:4]
```

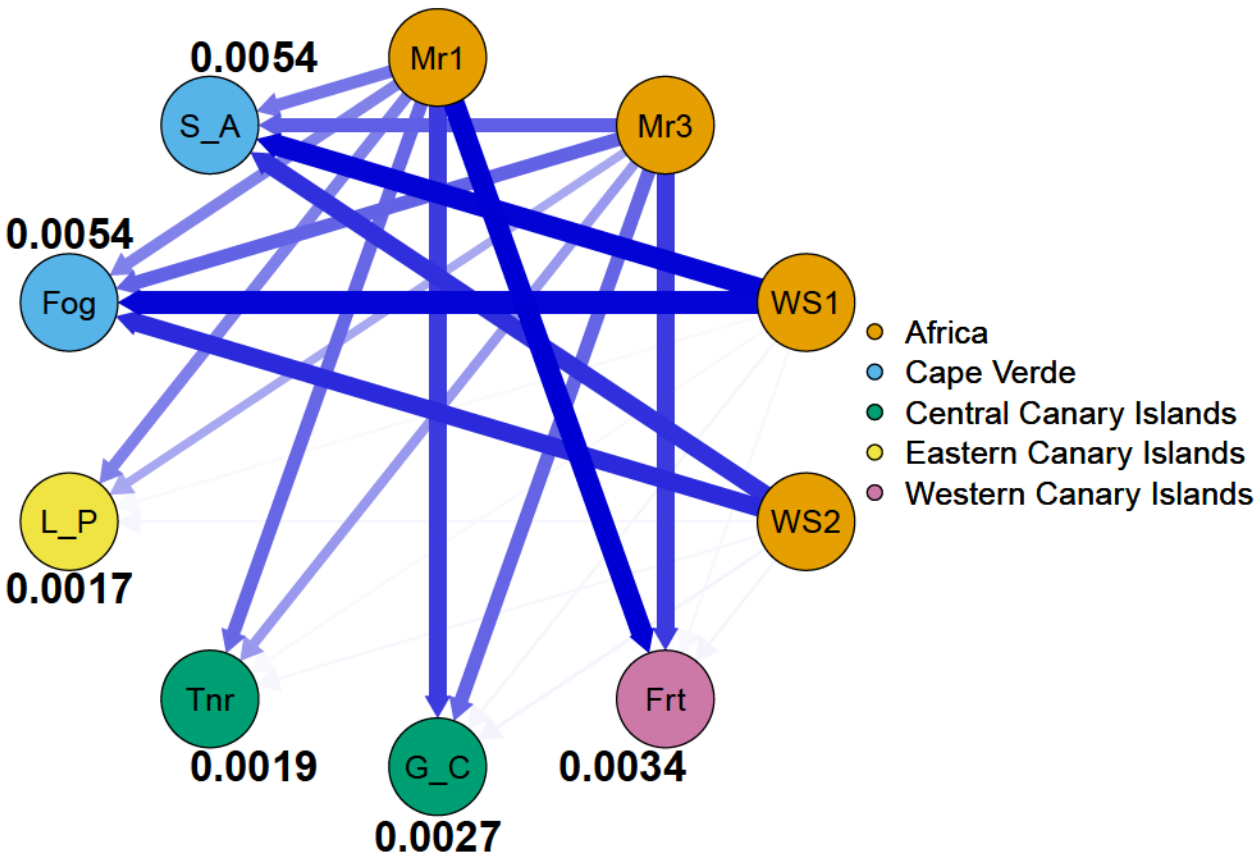
```
##                 Morocco1    Morocco3     WSahara1     WSahara2
## Fuerteventura 0.007319083 0.005658942 0.0002379433 0.0003069165
## Gran_Canaria  0.005625220 0.004453675 0.0002783037 0.0003479783
## Tenerife      0.004350546 0.002944165 0.0002090746 0.0002624865
## La_Palma      0.003670016 0.002491726 0.0001969059 0.0002472273
## Fogo          0.003617335 0.004497506 0.0072149708 0.0061264604
## San_Antonio   0.003958045 0.004596927 0.0072439731 0.0059671579
```

Now, we can use **qgraph** R package to plot a graph with wind connectivity from locations in Africa mainland to Macaronesian islands and compare this connectivity with another source of data, as genetic structure of *P. laevigata* from García-Verdugo et al. (2017). The following code is used to create figure 2 in the manuscript.

```
library(qgraph)
```

```
gr <- as.factor(c("Africa","Africa","Africa","Africa","Western Canary Islands",
                  "Central Canary Islands","Central Canary Islands",
                  "Eastern Canary Islands","Cape Verde","Cape Verde"))
qgraph(t(mat), layout= "circle", groups=gr, theme="colorblind", vsize=8,
       edge.width=1.7)
```

```
text( 0.5,-1, round(mean(mat[5,1:4]), 4), cex=1.5, font=2)
text( 0,-1.2, round(mean(mat[6,1:4]), 4), cex=1.5, font=2)
text( -0.5,-1, round(mean(mat[7,1:4]), 4), cex=1.5, font=2)
text( -1,-0.5, round(mean(mat[8,1:4]), 4), cex=1.5, font=2)
text( -1,0.5, round(mean(mat[9,1:4]), 4), cex=1.5, font=2)
```



Finally, we can use the shortest paths stored between the two selected locations (Western Sahara and Santo Antao (Cape Verd)) to plot them as a lines kernels over a map.

First, we should remove the NULL entries from the **path** object (they were created if **Inf** cost was obtained). Next, we merge all the path lines in the **paths_merged** object

```
paths_clean <- paths[!sapply(paths, is.null)]
paths_merged <- paths_clean[[1]]

for (h in 2:length(paths_clean)) {
  paths_merged <- rbind(paths_merged,paths_clean[[h]])
}
```

Now, we use R package **spatstat** to create a kernel distribution of lines with the function **density** and then transform this object into a raster layer. We can apply here transformation or thresholds for better representation. We remove all kernel density under the 10% of the maximum kernel value.

```
library(spatstat)
```

```
paths_psp <- as(paths_merged, "psp")
lines_kernel <- density(paths_psp, sigma=0.4, dimyx=c(350,410))
kernel <- raster(lines_kernel)
```

```
kernel[kernel<(maxValue(kernel)*0.1)]<-NA
```

Finally, we use **ggmap** package to plot our paths kernel density over a map

```
library(ggmap)
polyg <- rasterToPolygons(kernel)
polyg@data$id <- 1:nrow(polyg@data)

polygFort <- fortify(polyg, data = polyg@data)
polygFortMer <- merge(polygFort, polyg@data, by.x = 'id', by.y = 'id')

study_area <- ggmap(get_map(location = c(-29,10,-5,36), maptype = "hybrid"))
```

```
study_area +
geom_polygon(data = polygFortMer,
             aes(x = long, y = lat, group = group, fill = layer),
             alpha = 0.7, size = 0) +
scale_fill_gradientn(colours = bpy.colors(255))
```