



UvA-DARE (Digital Academic Repository)

How birds weather the weather: avian migration in the mid-latitudes

Kemp, M.U.

Publication date
2012

[Link to publication](#)

Citation for published version (APA):

Kemp, M. U. (2012). *How birds weather the weather: avian migration in the mid-latitudes*. [Thesis, fully internal, Universiteit van Amsterdam].

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

RNCEP: global weather and climate data at your fingertips

Kemp, M.U., van Loon, E.E., Shamoun-Baranes, J., & Bouten, W. (2012) *Methods in Ecology and Evolution* **3**:65-70

2.1 Summary

1. Atmospheric conditions strongly influence ecological systems, and tools that simplify the access and processing of atmospheric data can greatly facilitate ecological research.
2. We have developed RNCEP, a package of functions in the open-source R language, to access, organize and visualize freely available atmospheric data from two long-term high-quality data sets with global coverage.
3. These functions retrieve data, via the Internet, for either a desired spatiotemporal extent or interpolated to a point in space and time. The package also contains functions to temporally aggregate data, producing user-defined variables, and to visualize these data on a map.
4. By making access to atmospheric data and integration with biological data easier and more flexible, we hope to facilitate and encourage the exploration of relationships between biological systems and atmospheric conditions.

2.2 Introduction

Atmospheric conditions at different scales in space and time strongly influence a wide range of biological processes: from short-term atmospheric conditions, especially in the lowest troposphere (Stull, 1988), described collectively as weather, to longer-term average conditions (i.e. climate). For instance, the sex of some reptilian species' offspring is temperature-dependent (Janzen, 1994); humidity influences the rate of dehydration and therefore the level of activity in most amphibians (Shoemaker and Nagy, 1977); low temperatures and storms can influence avian reproduction (Wingfield, 1984); and atmospheric conditions are drivers of migration and long-distance transport in numerous taxa (Drake and Farrow, 1988; Dingle, 1996; Nathan, 2005; Newton, 2008). Recent trends in atmospheric conditions, for example rising temperatures and oscillations in pressure systems (e.g. North Atlantic Oscillation), have profound effects on marine and terrestrial ecosystems (e.g. Hughes, 2000; Ottersen et al., 2001; Walther et al., 2002; Parmesan, 2006). The connection between climate and biology cannot be overstated, as a species' ability to inhabit a particular geographic range is largely determined by the area's climate (MacArthur, 1972; Guisan and Zimmermann, 2000; Tarroso and Rebelo, 2010). Within the past century, climate change has influenced the geographical ranges and abundance of numerous species (Parmesan and Yohe, 2003; Root et al., 2003), in some cases leading to population and species extinctions (Pounds et al., 1999; Parmesan, 2006).

Clearly, this intimate relationship between short- and long-term atmospheric conditions and biological systems demonstrates the need to incorporate atmospheric data in ecological studies. A large amount of atmospheric data, stored in diverse formats, can be accessed through various Internet portals (Shamoun-Baranes et al., 2010); however, not all data are freely available. The National Centers for Environmental Prediction (NCEP)/National Center for Atmospheric Research (NCAR) Reanalysis data set (Kalnay et al., 1996), hereafter called R-1, and NCEP/Department of Energy (DOE) Reanalysis II data set (Kanamitsu et al., 2002), hereafter called R-2, are high-quality, well-documented, freely-available data sets with global coverage of numerous atmospheric variables. In recent years, these NCEP data sets have been increasingly used in ecological research including studies of phenology (e.g. Chmielewski and Rötzer, 2002; Cook et al., 2005), land cover and atmospheric interaction (e.g. Lawton et al., 2001; Marengo et al., 2008), and bird migration (e.g. Shamoun-Baranes et al., 2010, and references therein). NCEP data can be accessed via ftp where users can download a full year of data, a web service where users can select a temporally and spatially continuous subset of data, or queried directly from the Internet using the Open-source Project for a

network Data Access Protocol (OPenDAP Sgouros, 2004). In all cases, NCEP data are stored in the netCDF data format (a well-documented binary format for storing array-oriented scientific data; Rew et al., 2010) and the user needs to extract the data using specially developed software tools.

To facilitate the extraction, organization, aggregation and visualization of NCEP R-1 and R-2 data, we have developed the RNCEP package of functions in the R language for statistical computing and graphics (R Development Core Team, 2010). R is a freely available open-source computing environment that is highly extendable and can be run on multiple platforms. R is extensively used in ecological research with tools tailored to the ecological community (e.g. Calenge, 2006; Kneib and Petzoldt, 2007). This chapter describes the functionality of the RNCEP package.

2.3 The NCEP data sets

The NCEP/NCAR R-1 and NCEP/DOE R-2 are freely available state-of-the-art gridded reanalysis data sets with global coverage of many relevant atmospheric variables spanning 1957 to present and 1979 to present, respectively. Data for many variables are available at 17 pressure levels ranging from 1000 to 10 mb. Other variables describe conditions either at or near the surface. These data have a spatial resolution of $2.5^\circ \times 2.5^\circ$ and a temporal resolution of 6 h (00, 06, 12, 18 h UTC). Still other variables in these data sets are given on a global T62 Gaussian grid with 192 equally spaced longitudinal and 94 variably spaced latitudinal grid points, also in 6 h intervals.

2.4 The R scripts

Source code and binary distributions of the RNCEP package, with associated help files, are available via the CRAN repository (<http://www.cran.r-project.org/>) and can be installed on most systems by entering `install.packages('RNCEP', dependencies=TRUE)` into an R command prompt. Installed in this way, help files may be called using standard R syntax (e.g. `?RNCEP`), and dependent packages are installed automatically. RNCEP depends on the R packages `abind` (Plate and Heiberger, 2004), `fields` (Furrer et al., 2010), `fossil` (Vavrek, 2011), `maps` (Becker et al., 2010), `tcltk` and `tgpp` (Gramacy and Taddy, 2010); each is freely available via the CRAN repository.

We have designed several functions that can be used to facilitate common tasks when working with atmospheric data. These tasks are described in more detail in the following text and presented in schematic workflows (Figures 2.1 and 2.2).

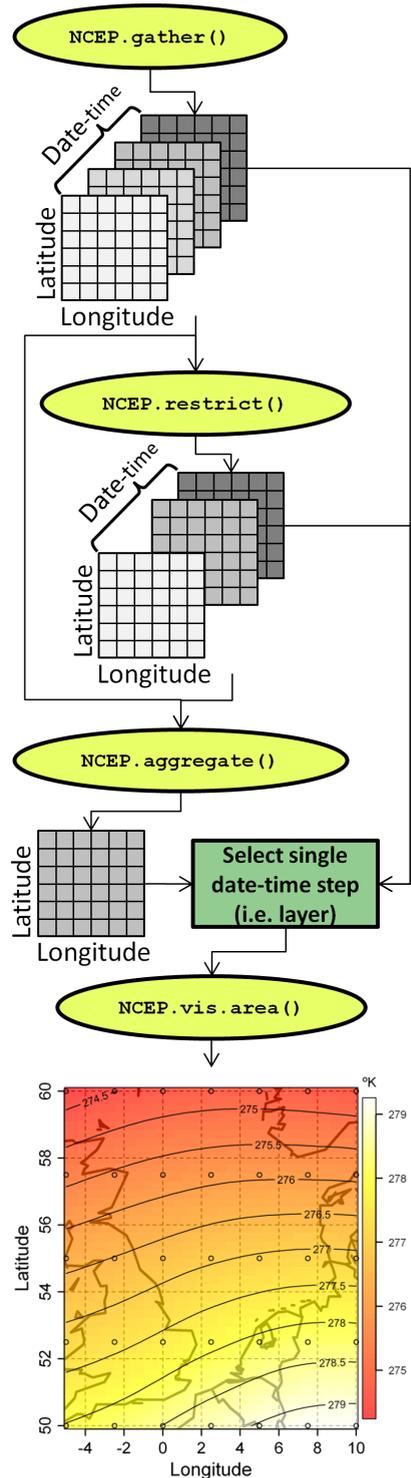


Figure 2.1: This flowchart illustrates a workflow to retrieve and organize data from the R-1 or R-2 dataset for a specified spatiotemporal extent using the RNCEP package. In the workflow, yellow ovals indicate functions in the RNCEP package. `NCEP.gather()` obtains data for a desired spatiotemporal extent. These data are arranged in a three-dimensional array comprising latitude, longitude, and date-time. Any undesired date-times can be removed from the array using `NCEP.restrict()`. `NCEP.aggregate()` temporally aggregates the data array to calculate user-defined summary statistics. `NCEP.vis.area()` visualizes a single date-time (or aggregated date-time) at any point in this process. We show a filled contour plot of mean temperature (K) at 1800 UTC from the 850 mb pressure level from 1 October to 15 October 1970-2000. At any stage, these data may be exported in various formats including those readable by third-party GIS software.

Utilizing R's extensive facilities, one can perform GIS and statistical operations on any retrieved atmospheric data from within R or data may be exported in a variety of formats, ranging from R's internal format to formats that can be imported into third-party GIS software (see the RNCEP help files).

2.5 R functions to retrieve and organize reanalysis data from a specified spatiotemporal extent

2.5.1 Task 1 – Gather NCEP data

The `NCEP.gather()` function retrieves data from the R-1 or R-2 data set, utilizing the OPenDAP method of data access, for a given spatiotemporal range. Users must specify (i) the variable of interest, (ii) the level (i.e. pressure level, surface or T62 Gaussian grid) from which the variable is to be obtained, (iii) the desired spatial range, (iv) the desired month and year ranges and (v) whether data should come from the R-1 or R-2 data set. With these parameters specified, `NCEP.gather()` returns to the R environment a three-dimensional array containing the specified variable over the desired spatiotemporal range. For example, the user can extract air temperature at the 850 mb pressure level from the R-1 data set for October, 1970-2000, 50°N-60°N, 5°W-10°E (Figure 2.1).

2.5.2 Task 2 – Restrict NCEP data

The structure of the NCEP data sets makes it difficult to download an interrupted time series of data per year. For example, one cannot easily obtain data for only 1200 UTC every day for a specific month. The function `NCEP.restrict()` can be used to remove unwanted temporal intervals from the data imported into R in Task 1 using `NCEP.gather()`. Using `NCEP.restrict()`, one can remove data for a specified year, month, day, hour or any combination of the four. For the example in Figure 2.1, this function is used to restrict the temperature data to 1800 UTC and 1-15 October.

2.5.3 Task 3 – Aggregate NCEP data

Once an array of atmospheric data has been obtained, and the time series restricted, derived variables may be calculated by temporally aggregating or summarizing the array, for example by calculating a mean, percentage of occurrence or an accumulation. For instance, one could calculate mean maximum temperature to explain Malaria epidemics (Githeko and Ndegwa, 2001), the

frequency of tailwind assistance for migrating birds to explain the timing of spring arrival (Sinelschikova et al., 2007), or seasonal temperature accumulation (i.e. degree days) to explain variability in plant phenology Wang (1960). The `NCEP.aggregate()` function is applied to perform these temporal aggregations. It summarizes data at each grid point and returns a new array with the same spatial dimensions as the input array. The user specifies the function to apply (either an internal R-function such as ‘mean’, ‘max’ or ‘sum’ or a function created by the user) and whether or not to aggregate each temporal component: year, month, day and hour. While data returned by `NCEP.gather()` and `NCEP.restrict()` are technically weather data as they describe atmospheric conditions at relatively short temporal intervals, `NCEP.aggregate()` can be used to derive climate data by averaging the atmospheric data over a sufficiently long period. In the example in Figure 2.1, mean temperature at 1800 UTC from 1 to 15 October 1997-2000 is calculated.

2.6 R functions to interpolate reanalysis data to specified points in space and time

The function `NCEP.interp()` interpolates variables from the R-1 or R-2 data set to specified locations in space and time (see Figure 2.2 for a workflow of this procedure). The user must specify the atmospheric variable, level (again, pressure level, surface, or T62 Gaussian grid) from which the variable should be obtained, and spatial and temporal location to which the variable should be interpolated. Further optional arguments include parameters to control interpolation. `NCEP.interp()` will accept vectors as arguments and can, therefore, easily be applied to all of the points in an ecological data set with a single command. Thus, users could calculate the temperature and wind conditions at each location along the entire migratory route of an individually tracked animal (e.g. Shamoun-Baranes et al., 2003a, 2010, and references therein).

`NCEP.interp()` queries the NCEP data base, utilizing OPenDAP via the Internet, obtaining data from the eight grid points surrounding the desired location in space and time. If the method of interpolation is given as ‘linear’, the function performs trilinear interpolation in latitude, longitude and time. Alternatively, if the method of interpolation is ‘IDW’, the function performs inverse distance weighting (Shepard, 1968) in space followed by linear interpolation in time. The user can turn off interpolation in space or time or both, in which case `NCEP.interp()` performs ‘nearest-neighbor’ interpolation returning the value of the closest grid point in space or time or both, respectively. Spatial interpolation is always performed assuming a spherical grid rather than a planar surface. To indicate the precision of an interpolated

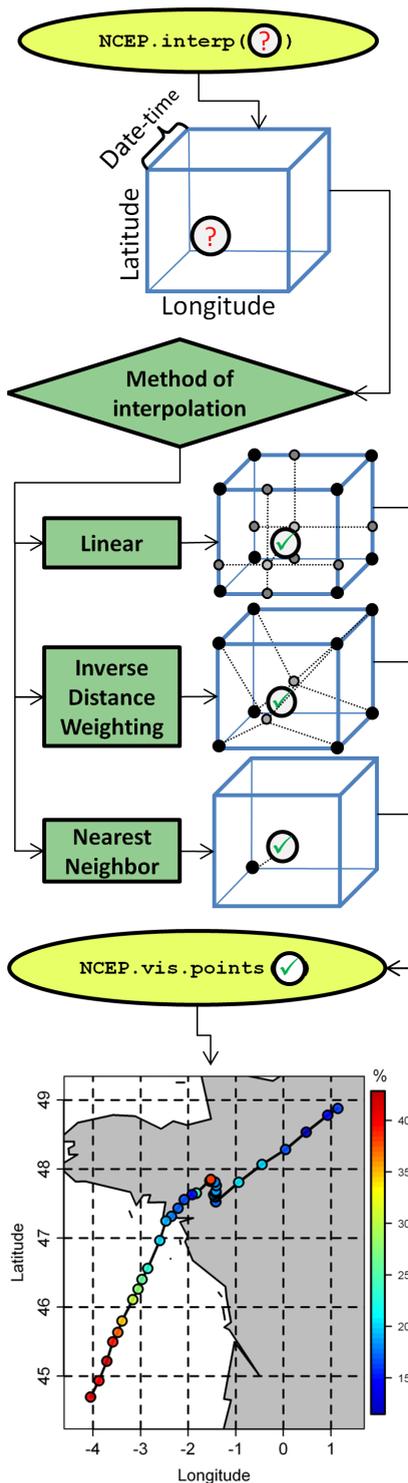


Figure 2.2: This flowchart illustrates a workflow to interpolate weather data from the R-1 or R-2 dataset to specified points in space and time using the RNCEP package. In the workflow, yellow ovals indicate functions in the RNCEP package. `NCEP.interp()` retrieves the eight data points in the R-1 or R-2 dataset surrounding the desired point in space and time. In the flowchart, a question-mark indicates a point in space and time to which interpolation will be performed. The cube indicates the eight grid points in the R-1 or R-2 dataset surrounding the desired point in space and time. The function interpolates the data from the surrounding grid points using the specified method of interpolation. The question-mark becomes a check-mark after interpolation: the value is now known. `NCEP.vis.points()` visualizes interpolated data in their proper location on a map using color to indicate their value. Here, we see a migrating lesser black-backed gull (*Larus fuscus* Linnaeus, 1758) traveling from northeast to southwest. Its location, recorded by a mounted GPS tracking device, is shown between 17 and 19 September 2008 (median measurement interval: 21 min). The color of the points indicates the linearly-interpolated percentage of total cloud cover. The gull delayed crossing the Bay of Biscay for over 24 h, likely due to inclement weather suggested by the increased cloud cover.

result, `NCEP.interp()` calculates the standard deviation of the values used to perform the interpolation. Thus, precision is described in the same units as the interpolated output, with smaller values indicating less variability among the predictor points. Some variables in these data sets (e.g. cloud cover) should not be temporally interpolated as they describe conditions over an interval of time rather than at a specific point. For these variables, `NCEP.interp()` will not perform interpolation in time and instead automatically retrieves data describing the interval within which the specified date-time falls.

2.7 R functions to visualize reanalysis data

Regardless of whether data are obtained over a spatiotemporal extent or interpolated to a point in space and time, it is often desirable to visualize these data on a map.

`NCEP.vis.area()` produces a contour plot of a single date-time (or single aggregated date-time) from a data-array (Figure 2.1). Specifying only the input data-array and date-time to visualize, the user can quickly obtain a representative map. The map's spatial dimensions, for instance, are automatically set according to the spatial range of the input data-array. The user can also manually configure each aspect of the map.

`NCEP.vis.points()` produces a map indicating the value of a variable interpolated to point locations as obtained using `NCEP.interp()` (Figure 2.2). The color of each point indicates the interpolated value. The user only needs to specify the location of each point and the value of the variable at that point to produce a representative map, yet all aspects of the map are configurable.

2.8 Conclusion

RNCEP is intended to streamline access to and organization of atmospheric data from two freely available high-quality gridded data sets with global coverage. Although other tools exist to extract and visualize these data, e.g. GrADS (Doty et al., 1995) and IDV (Murray et al., 2003), the RNCEP package integrates those functions that are particularly useful for an ecologist: data download, format conversion, subset, aggregation, interpolation and visualization. The package enables rapid adjustment of the spatial and temporal ranges obtained, translating logical user input into the necessary commands in OPeNDAP and returns data in an easily interpretable, unpacked format. Further, RNCEP does not require that any data be stored locally, nor must any particular data base connection be maintained. With only an Internet connection, data may be obtained on-demand, explored, manipulated, integrated

into R statistical analyses and either saved in various formats or discarded. In the future, other gridded data sets (e.g. the NCEP Global Ocean Data Assimilation System) that are available online in a stable format may be accessed using the approach of the RNCEP package. In this way, many functions in the RNCEP package may be applied directly to these new data sets.

Ecology is inherently a multidisciplinary field at the interface between various Earth systems. As such, tools to integrate data from the different systems, enabling efficient and versatile research, are an important aspect of an e-science environment (Hey and Trefethen, 2005). We hope RNCEP will facilitate and encourage the incorporation of atmospheric data into ecological research and promote the exploration of relationships between biological systems and atmospheric conditions. To cite RNCEP or acknowledge its use, cite this article as follows, substituting the version of the package that you used for ‘R package version 1.0.1’:

Kemp, M. U., van Loon, E. E., Shamoun-Baranes, J., and Bouten, W. (2011). RNCEP: global weather and climate data at your fingertips. *Methods in Ecology and Evolution*, 3: 65-70. doi: 10.1111/j.2041-210X.2011.00138.x (R package version 1.0.1).

Acknowledgments

NCEP R-1 and NCEP R-2 data were provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their website at <http://www.esrl.noaa.gov/psd>. R-2 data were produced with the support of the U.S. National Weather Service and U.S. Department of Energy. Please acknowledge the use of NCEP data in any documents or publications. Our studies are facilitated by the BiG Grid infrastructure for e-Science (<http://www.biggrid.nl>), Flysafe2 and the Dutch National Authority for Data concerning Nature (GaN; <http://www.gegevensautoriteitnatuur.nl>). The authors thank the associate editor of *Methods in Ecology and Evolution* and two anonymous reviewers for constructive comments on an earlier draft of the manuscript.