A method to predict connectivity for partially migratory waterbird species from tracking data - Supplementary Information

Contents

1 Hidden Markov modelling to detect movement classes for ibis

We used the hidden Markov model (HMM) to classify states of behaviour using the ibis telemetry data. The intention was to discriminate between hypothesised behaviour states that *a priori* include waterbird roosting, foraging, shifting and long-distance travel that includes migratory movements.

We focus on the method of Patterson et al. [2009], which is implemented in the `moveHMM` package [Michelot et al., 2016] in the R programming language. This package was chosen for its balance of statistical rigour, and computationally efficient implementation, which is very important for data of this size, ease of use along with the accompanying tools for evaluating model fit and visualisation.

1.1 Methods

The primary assumptions for use of movement data in HMMs are that measurement error in positions is negligible and that there is a regular sampling unit e.g., one positional observation per hour. We consider the measurement error to be negligible relative to the distance travelled post quality control of the points.

For each of the 48 birds, the telemetry points were processed for position (longitude, latitude) and date and closest hour of telemetry reading. The telemetry devices were programmed to read at hourly intervals between 7:00 and 19:00 and a fix at 24:00. To accommodate the regular sampling unit requirement of the HMM and `moveHMM` we imputed non-measured hourly values with missing values, which can be handled by the `moveHMM` package. The `prepData` function was used to calculate step distance and angle between points (see Michelot et al. [2019] for definitions of these quantities). Known states can be passed to the algorithm so we allocated midnight fixes as a state to represent the roosting state as it is known that ibis roost at night.

The `moveHMM` HMM software requires an initialisation of the parameters of the step length distribution, which is modelled using a gamma distribution, and the parameters of the turning angle distribution, which is modelled using a von Mises distribution. Furthermore, the number of behaviour states needs to be specified for each run of the algorithm i.e., the number of states is not estimated from the data. For the step distribution, the mean and variance of each of the states are estimated and require initialisation along with the mean and concentration of each of the states for the von Mises distribution. For each of the states, steps that are exactly equal to zero are modelled using a zero-inflated form of the gamma distribution. A starting value for the proportion of steps that are exactly zero is also required to run the algorithm.

We hypothesised that three or four behaviour states would best model the data for these species based on current best ecological knowledge. However, to investigate this we ran for each bird six runs of the `fitHMM` function using new random starting parameters that were drawn from the observed distributions of step length and angle. Six runs were chosen to ensure that the space of reasonable starting values was explored and that the numerical maximisation routine of the `fitHMM` function might converge to the best approximation to the global maximum of the likelihood function. Percentiles of the observed distributions were used as proxies for states in the data and random values generated from distributions that best summarised the distribution of the starting parameter e.g., chi-squared distribution for the standard deviation of the step distance. Therefore, for each bird six runs for each of the two, three and four state models were run with random starting values.

We ran the `moveHMM` software using all individual ibis in each run of the six runs with different starting parameters for each of the three movement states (2, 3, or 4), which is a total of 18 model runs. The HMM analysis included 250,533 telemetry points for 48 Straw-necked ibis.

1.2 Results

The number of movement states was chosen by comparing the output from the 18 runs for each bird and comparing the Akaike information criterion (AIC) of each model fit. We found the HMM fitted with four movement states was both stable across the six model runs and returned the lowest AIC (Table 1).

Species	No. states	AIC	No. models
SNI	2	398603	3
SNI	2	416524	3
SNI	3	386972	5
SNI	3	436124	1
SNI	4	383173	5
SNI	4	417685	1

Table 1 Summary of HMM model AIC used for model selection. The HMM was run six times using varying starting parameters assuming 2, 3, or 4 hidden movement states. The method should converge to a similar or exactly the same log-likelihood and subsequent AIC for different assumed states, but can converge to local maxima depending on initial starting parameters. The table summarises the number of the six models run that converged to the same AIC. It further shows which model gave the lowest AIC, which was taken as the best fitting model. For example, for SNI of the six models run assuming four movement classes, 5 converged to an AIC of 383173 and 1 to an AIC of 417685.

Table 2 shows the HMM model parameters for the lowest AIC model. Table 2 shows how the HMM model classification has differentiated the hourly distances travelled and how movement state four captures the substantial movements.

Component	Parameters	HMM state 1	2	3	4
Distance	mean	0.1	0.56	2.53	27.52
	sd	0.09	0.44	2.57	17.11
	zero-mass	0.07	0.00	0.00	0.00
Angle	mean	3.13	-3.12	-0.05	0.00
	concentration	0.28	0.31	0.09	1.79

Table 2 Summary of HMM model parameters selected from lowest AIC model of four movement states. The mean (km) and standard deviation (sd in km) refer to the zero-inflated gamma distribution with zero-mass reflecting the proportion of 0 km movements, and the parameters of the angle parameters (in radians) are those of the von Mises distribution. For reference, if the von Mises angle is zero with a concentration value greater than 1 then the angle between successive points is small and variation off the continuous path of the individual is small. Mean values close to π with a small concentration value indicate a lot of variability in the angle of movement and a less directed path.

2 Distance modelling

In this section we describe the approach used to predict the distance travelled by ibis using telemetry data and environmental covariates. Separate generalised additive models (GAMs) for models were fit for telemetry data classified into HMM states 3 and 4 respectively.

The section first describes the collation of several derived environmental covariates (wind, local and landscape water, and near real-time weather). It then describes how data were summarised to the daily scale, before outlining the statistical GAM models used to determine the best fit relationship between distance travelled and the environmental predictors.

2.1 Deriving additional environmental covariates

A list of potentially useful environmental covariates was generated based on our knowledge of ibis life history. The following set of physical variables were hypothesised to be of high interest

- Inundation states and landscape water availability (wet/dry, extent, duration, depth)
- Normalized difference vegetation index (NDVI)
- Rainfall
- Temperature
- Wind direction and speed
- Barometric pressure

Further important species-specific variables included:

- Timing of breeding: Adults moving to breeding sites (spring, summer); juveniles fledging and leaving natal breeding sites (summer); adults leaving breeding sites (summer)

- Season/calendar month: e.g. no long-distance movements in winter; most long-distance movements in spring, late summer, early autumn
- Daily behaviour: Ibis depart on long-distance trips between 9 – 11 a.m. and stop at the end of the day before dusk (4 – 6 p.m.). Ibis always roost/sleep at night.

Some covariates that were identified in this process (e.g. NDVI, rainfall and temperature) were readily available from existing data products. However, often the available products lack real-time information that is scalable to the spatial and temporal resolution of the telemetry data (e.g. landscape water availability, wind speed and direction). In the following subsections we describe how we supplemented existing information with more detailed (i) wind, (ii) inundation and (iii) weather data.

2.1.1 Wind as a driver of movement

We hypothesised that wind is a key covariate driving movements for waterbirds and we sought data that aligned with the waterbird telemetry data. We used the R package `rWind` [Fernández-López and Schliep, 2018] to align the wind speed and direction for each of the waterbird telemetry points. The package `rWind` is designed specifically to download and process wind data from the Global Forecasting System, which is an atmospheric model from the National Oceanic and Atmospheric Administration (NOAA) and National Centers for Environmental Prediction (NCEP). The `rWind` package requests data from the Pacific Islands Ocean Observing System, coordinated by the University of Hawaii School of Ocean and Earth Science and Technology (SOEST).

For each telemetry point, we aligned wind stored as velocity vector components (U: eastward wind and V: northward wind) at 10 m above the Earth's surface at a spatial resolution of 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)) and the value for the closest hour taken. For each bird's telemetry points a raw plain text file 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 was generated and aligned to each bird's telemetry point information and covariates.

To align the wind speed and direction with the telemetry points we first took all available telemetry points and used functions in the `adehabitatLT` R package [Calenge, 2006] to compute the distances between telemetry points and the angle between each move and the x-axis (anti-clockwise). This is in contrast to the angle computed by `moveHMM`, which is the angle between successive moves. The wind direction angle is reported as clockwise from the y-axis. The bird and wind angle were both converted to the angle anti-clockwise from the x-axis and in degrees. With the bird and wind angle aligned we calculated the absolute difference between the two angles named θ. We put this on a $0-180$-degree scale to represent the full benefit of the wind i.e., when $\theta = 0$ the bird is directly aligned with the wind and when $\theta = 180$ the bird is flying directly into a headwind. Whether this angle is to the left or right we don't discriminate and hence the $0-180$-degree scale.

We compute a composite measure of wind velocity aligned with the bird's direction and termed *wind benefit*. We hypothesise that small distances travelled are unaffected by wind and thus the direction and speed of the surrounding wind has little influence. When birds are performing longer distance movements having a strong wind velocity aligned with the direction of movement could be beneficial to energy conservation and distance travelled. Wind benefit is the component of the wind velocity that is aligned with the bird's velocity i.e., $WB = W \times \cos(\theta)$. Given the wind speed and direction from `rWind` and the absolute angle between the bird and the wind's direction, we computed WB for each telemetry point. Figure 1 shows a conceptualisation of this calculation. In the computation, the angle is in radians.

Figure 2 shows that there is evidence that wind benefit is a strong driver of longer-distance moves for straw-necked ibis.

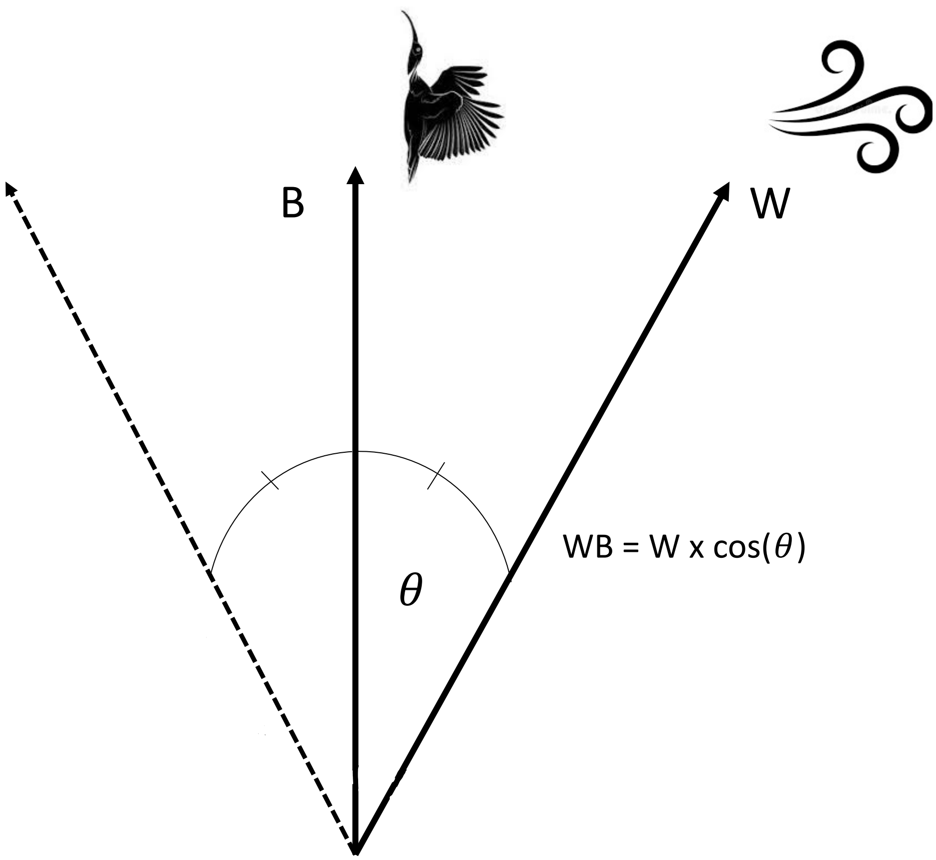

Fig. 1 Representation of the computation of wind benefit (WB). B represents the bird velocity (speed, direction) and W the wind velocity. The wind benefit is simply the component of the wind velocity that is aligned with the direction of the bird. For example, when $\theta = 0$ the bird receives the full benefit of the wind and when $\theta = 180$ the bird is flying into a direct headwind and receives the full negative velocity of the wind.

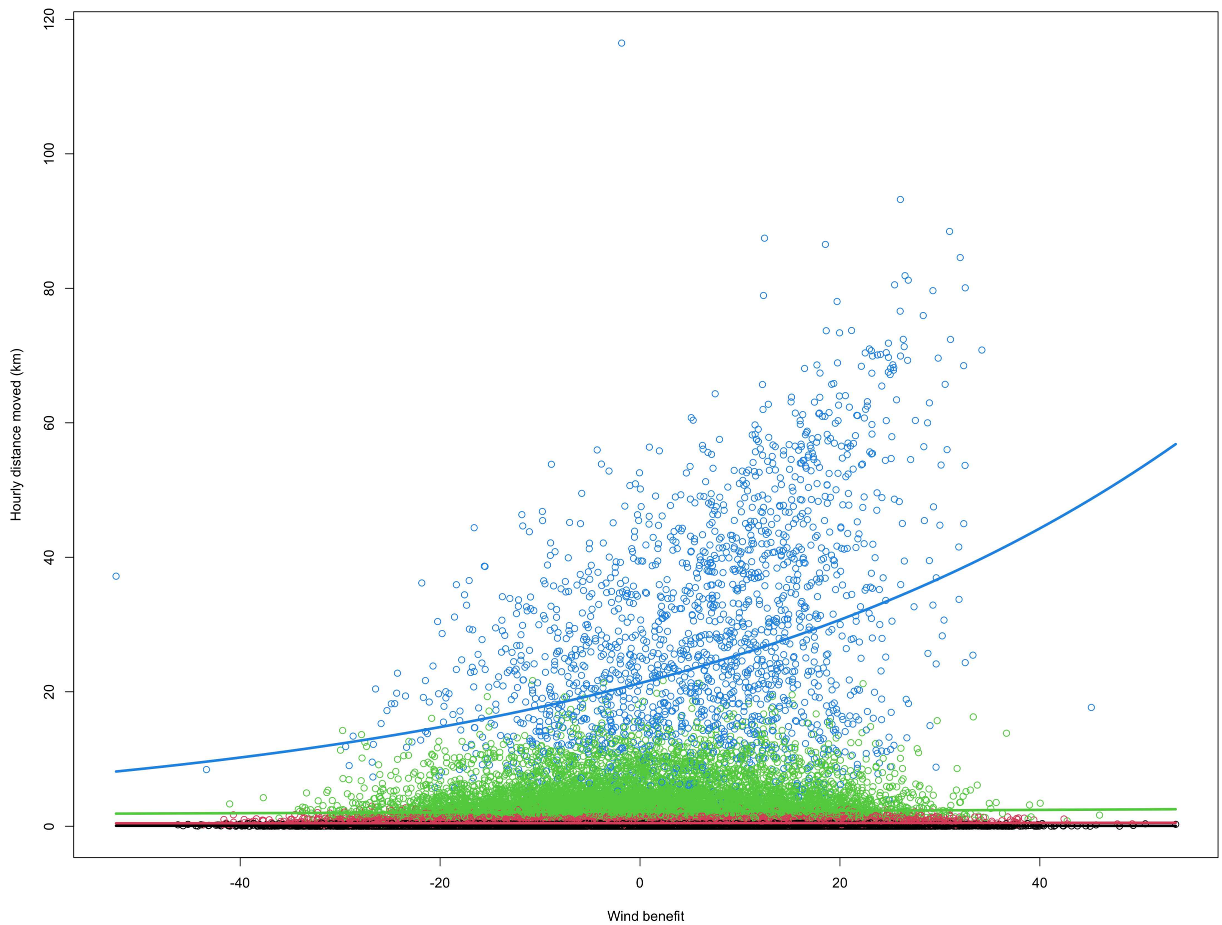

Fig. 2 Relationship between wind benefit and hourly distance moved for straw-necked ibis. Points represent individual telemetry points and point colours represent HMM movement state with black being state 1, red state 2, green state 3 and blue state 4. The lines are from a fitted log-linear model with an interaction effect for state and movement state. The figure shows the distinct correlation between long distance moves and wind benefit for Straw-necked ibis.

2.1.2 Measure of local and landscape inundation

To overcome some of the temporal resolution limitations of the Water Observed from Space (WOfS) inundation product, which is time averaged over 30 years, we investigated the Copernicus Sentinel-2 remote sensing data. The Sentinel-2 data have been shown to work well at classifying the presence or absence of water and have the advantage of being measured on a finer time scale than WOfS, with pairs of images for every location every two weeks available at a higher pixel resolution (10 m x 10 m). The intention was to investigate the capacity of the Sentinel-2 water observations to be a near real-time measure of inundation local to the bird's position. The pre-processed Sentinel data were downloaded from the National Computing Infrastructure (NCI) at `http://dapds00.nci.org.au/thredds/catalogs/if87/catalog.html`.

The primary difficulty with matching remote sensing data with telemetry points is that the points and spatial data must be matched in space and in time. The tiles produced from the Sentinel program do not cover the whole of eastern Australia at each time snapshot but rather cover it over a period of 14 days in a series of snapshots. Therefore, it is frequently the case that a bird may be in a position that does not have a Sentinel measure and then move before an image of that position is taken. To overcome this we undertook a process of averaging over many before and after images of each position and favouring Sentinel images that were close in time to the bird's presence and had low cloud cover.

Initially, we intersected all telemetry points with the small-memory boundaries files provided by the NCI for each Sentinel tile at each time point. For each telemetry point, all Sentinel tile boundaries from the prior 20 and post 20 days following the date of observation were gathered and then intersected to subset the files to those that overlapped with the position of the observed telemetry points. This typically left $2 - 8$ tiles for each telemetry point that both overlapped and were observed within 20 days on either side of the bird's observation date.

For many of the telemetry points, the same Sentinel tile captures multiple bird movements. Therefore, we subsetted the total list of Sentinel tiles that overlapped to the unique

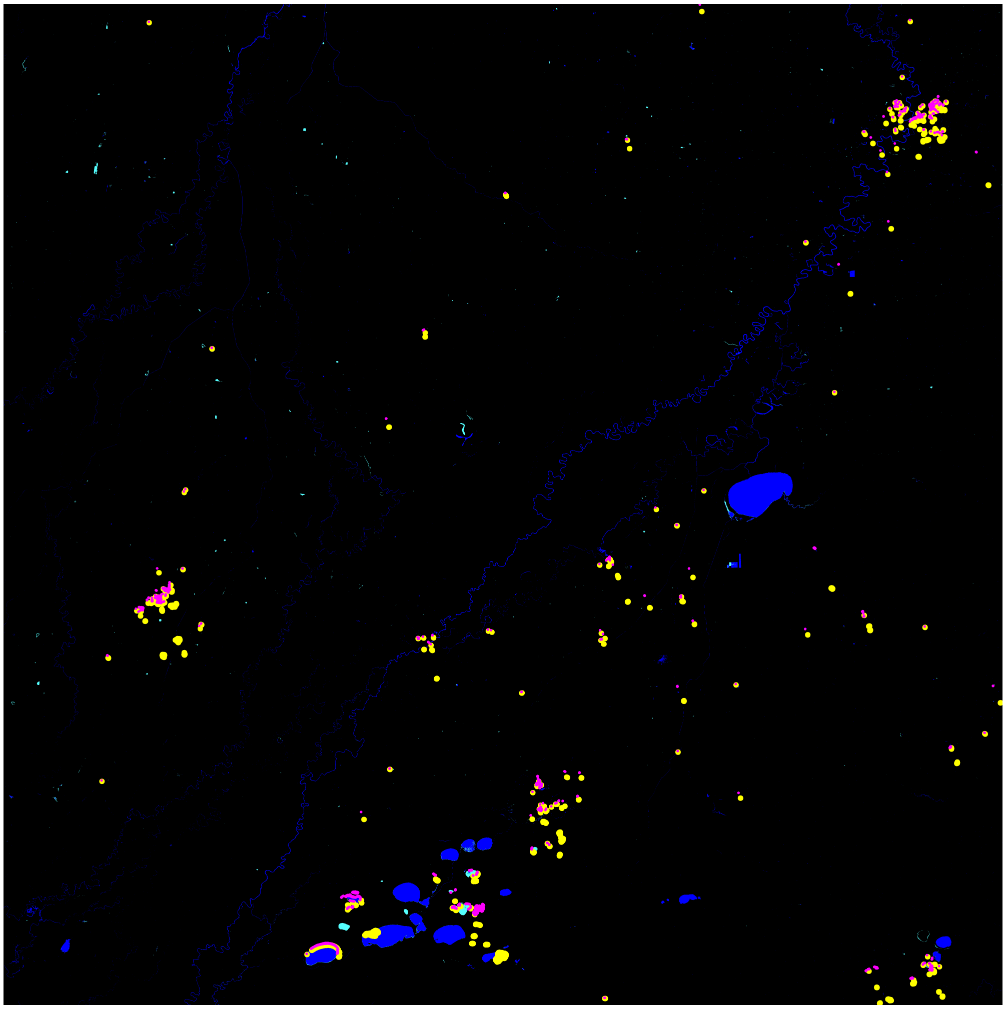

Fig. 3 Example of a Sentinel FMASK tile. Colours correspond to the FMASK algorithm classification of each pixel with black - valid, cloud - pink, cloud shadow - yellow, snow - cyan, and water - blue. Tile was taken from 2018-01-16 and is 50 km by 50 km in size.

set that covered all bird telemetry observations, which was 3,514 Sentinel tiles. Each Sentinel tile is approximately 50 km by 50 km but they do vary in size. For each tile, we downloaded the FMASK file, which is the version of the file that has been classified for cloud and water. The FMASK tile was generated using the pythonfmask module (http://pythonfmask.org), which classifies each 10 m by 10 m pixel as 0 - Null/Fill Value, 1 - Valid, 2 - Cloud, 3 - Cloud Shadow, 4 - Snow, and 5 - Water. Figure 3 shows an example of the Sentinel FMASK tiles used in the analysis.

To derive proxy measures of inundation for each waterbird telemetry point we read the set of Sentinel tiles corresponding to each telemetry point and computed the proportion of pixels in the tile classified as water along with the proportion of pixels in the tile as cloud or cloud shadow. Furthermore, we drew a 10 km radius circle around each telemetry point and again computed the water and cloud proportions for this local section of the tile. Therefore, for each telemetry point, a landscape and a local proxy measure of water on the landscape was available from 2 – 8 tiles as well as the proportion of cloud in each tile. To collapse these multiple measures we calculated a weighted average of the time-point measures. The weights correspond to 1) the inverse of how many days were between the observed telemetry point and the Sentinel tile measurement with measurement on the same day given a weight of two, and 2) the weights for cloud were inversely proportional to the proportion of cloud cover up to a maximum value of 100. For example, if a tile was observed two days away from the telemetry point and the tile was 10% cloud then the weight for that tile would be $1/2 \times 1/0.1$. The broad concept is that the weighted average preferences tiles that are measured close in time and with the lowest cloud proportion. Therefore, for each bird's telemetry point set, we have a measure of the proportion of local and landscape water present that is time-averaged to preference the closest water observation with the lowest cloud.

To investigate the concept that within a landscape there may be different types of water reservoirs e.g., lakes/dams versus rivers, we used the `raster` package to convert the water in each of the Sentinel tiles to polygon objects. We then characterised all the water reservoirs in each tile greater than 0.5 square km in size. For each telemetry point, we computed using the same weighted average method above the number of large water reservoirs in the landscape, the sum of the area of the largest reservoirs, the average and maximum size of all polygons as well as similar measures at the local scale. If a polygon intersected the 10km radius circle then the whole polygon was included in the area calculations.

2.1.3 Near real-time weather variables

Weather variables are hypothesised to be key drivers of waterbird movement and decision making. To incorporate weather into our analyses we sought data from the European Centre for Medium-Range Weather Forecasts 5th Generation Reanalysis data product (ERA5) [Hersbach et al., 2020]. ERA5 generation used the Integrated Forecasting System (IFS) Cy41r2 and benefits from improvements in model physics, core dynamics and data assimilation. Additionally, ERA5 has a much higher spatial resolution of 31 km over the ERA-Interim, and has hourly output. ERA5 is a comprehensive reanalysis, from year 1979 to near real-time and assimilates many observations in the upper air and near surface. The ERA5 atmospheric model is coupled with a land surface model and a wave model to produce the output.

The ERA5 release has a vast number of weather variables but we focussed on the surface and single-level instantaneous parameters. We acquired data on temperature, pressure and rain. These three variables are hypothesised to be important in triggering movement and potentially in course changes during flight or en route to a common destination. ERA5 data is replicated through the Copernicus Climate Data Store (CDS) and hosted for use by the Australian research community at the National Computing Infrastructure (NCI) (details at https://opus.nci.org.au/x/R4DYB). Data access is granted by user request and was accessed from /g/data/rt52/era5 on the Gadi supercomputer filesystem.

For each year ERA5 stores monthly global `.nc` multidimensional scientific data format files for each variable. Using the `ncdf4` R package we read the slice that includes the Australian continent for the month that includes each bird's telemetry point (e.g., see Figure 4). The closest day, hour and (latitude, longitude) point were oriented to each telemetry point and the ERA variable value corresponding to the closest in time and space taken.

For temperature, we used the two-metre temperature variable. This corresponds to the temperature of the air at 2m above the surface of land, sea or inland waters. The two-metre temperature variable is calculated by interpolating between the lowest model level and the

Earth's surface, taking account of the atmospheric conditions. The value is reported in Kelvin and was then converted to Celsius.

For pressure, we took the mean sea level pressure value. This parameter is the pressure of the atmosphere and is a measure of the weight that all the air in a column vertically above the area of Earth's surface would have at that point if the point were located at the mean sea level. Mean sea level pressure maps can be used to identify the locations of low and high-pressure systems, often referred to as cyclones and anticyclones. Contours of mean sea level pressure also indicate the strength of the wind. The unit for pressure is Pascals (Pa).

For precipitation, we used the large-scale rain rate variable. This parameter measures the rainfall intensity, at the specified time, generated by the cloud scheme in the ECMWF Integrated Forecasting System (IFS). The cloud scheme represents the formation and dissipation of clouds and large-scale precipitation due to changes in atmospheric quantities (such as pressure, temperature and moisture) predicted directly by the IFS at spatial scales of a grid box or larger. The units for LSRR are equivalent to mm per second time-averaged and are recommended to be interpreted as local to the reported point.

Following ERA5 data collation, the waterbird telemetry points were matched to the ERA5 variable results for modelling.

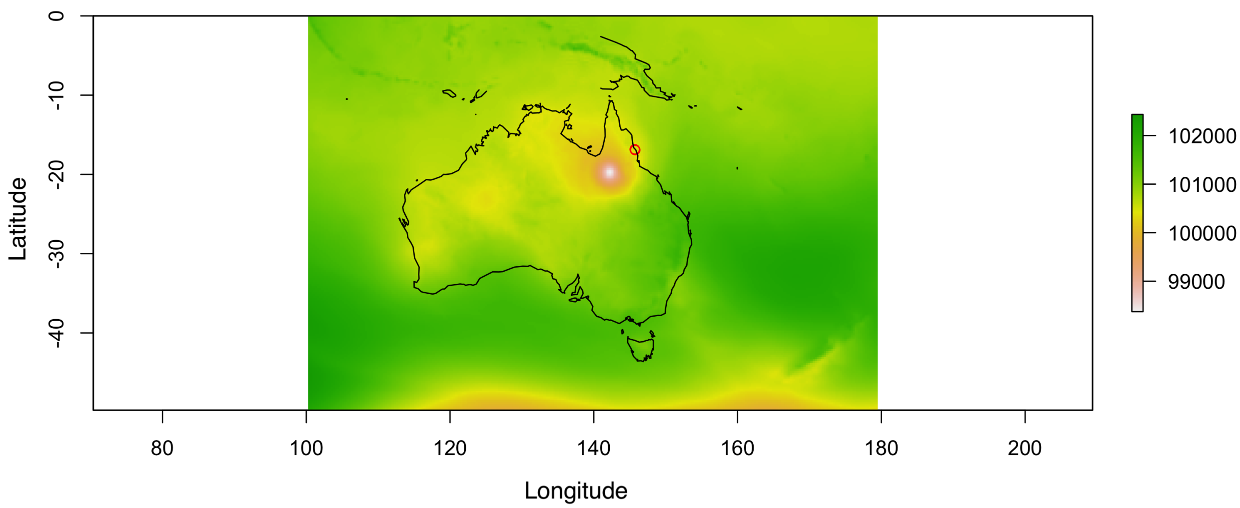

Fig. 4 Example of spatial extent and resolution for ERA5 weather variables. Displayed in the mean sea level pressure gradient for the hours covering the days in February, 2011 which includes the cyclone Yasi weather event. The red circle corresponds to Cairns city, which was the closest populous city and the colours represent mean sea-level pressure in Pascals.

2.2 Summarising data to the daily scale

For the straw-necked ibis, movements can be expected to be partitioned into daily units. For example, straw-necked ibis are known to roost at night and rarely travel at that time. Initial hourly modelling of distances travelled was not representative of a core movement unit. The modelling of hourly distances is too detailed; it is hypothesised that decisions to move or stay are made at a daily time scale and thus hourly changes in distance travelled are a partition of the decision time frame that the birds are hypothesised to make decisions on long-distance movement. Furthermore, environmental variables at the hourly scale are highly correlated, which limits the amount of independent information. Also, the number of points at the hourly time scale increases the computational burden.

We summarised the hourly straw-necked ibis telemetry data to a daily scale. Of particular importance is that straw-necked ibis are known on average to perform their largest flights during the day between the hours of 10 a.m. and 4 p.m. This allowed for a natural partition of the daily movements and the environmental covariates into a departure and destination set. For example, if an individual did perform a long-distance movement between the hours of 10 a.m. and 4 p.m. then the environmental variables at the morning and afternoon locations have a natural departure-destination interpretation. Environmental variables were summarised to the morning and afternoon components, cumulative values, and daily averages, where it made sense. Only days that included greater than or equal to 9 telemetry points were summarised. The value of 9 was chosen to balance the inclusion of as many days as possible but such that a representative sample of observations in the morning and afternoon periods were maintained.

The following is a detailed description of how each variable was summarised to the daily scale for the straw-necked ibis data set.

- For the Australian National Aquatic Ecosystem classification we summarised the class that the bird was most present in during the morning and afternoon observation period. We further summarised the ANAE class that the bird was most present in for each day.

- Sentinel local and landscape water variables were averaged for the morning and afternoon periods and a difference variable was constructed using the difference in these averages. These variables are intended to capture whether a bird moved from a dry to wet location or vice versa.

- Wind was summarised using the wind benefit metric. Wind benefit was averaged over the day to form the mean variable and the maximum wind benefit was also calculated.

- Temperature in degrees Celsius was averaged for the morning and afternoon periods and a difference variable was computed. A daily mean was also calculated.

- Mean sea-level pressure in Pascals was averaged for the morning and afternoon periods and a difference variable was computed. A daily mean was also calculated.

- Rainfall intensity was averaged for the morning and afternoon periods and a difference variable was computed. A daily mean was also calculated.

- Date was coded using numeric monthly values (i.e. 1-12) then fitted in the GAM using a cyclic cubic spline. This approach avoids any discontinuities in movements that occur between January and December.

- HMM state variables are intended to capture the movement state in which each bird was most in for that day. We summarised the state in which the bird was most in over the day i.e., if the bird was in movement state four for 8 of the 14 telemetry points between 7 a.m. and 7 p.m. then the bird's day would be assigned a 4. In some instances a bird maybe only in movement state four for four hours of the day. If we only relied on the max value then the state of this day could be assigned a 1. To capture the daily movement we computed the average state that the bird was in, which allowed for a continuous measure of movement state. We also summarise the number of instances during the day that each individual was in movement state 4.

- We computed the daily distance travelled in kms by summing the hourly distances.

Post summary, 9,522 days were available across 45 straw-necked ibis individuals with 26 derived environmental variables for modelling. Table 3 summarises the variables available for modelling.

Variable	Description
BandID	Band identifier for each bird
Date	Date summarised
DayDist	Total daily distance travelled summed over 7 a.m. to 7 p.m. fixes
AnaeMorn	ANAE category most present in before and including 10 a.m. on date
AnaeAft	ANAE category most present in after and including 4 p.m. on date
AnaeTop	ANAE category most present in over day.
TempMorn	Average temperature before and including 10 a.m.
TempAft	Average temperature after and including 4 p.m.
TempDiff	Difference between morning and afternoon temperature averages
TempMean	Mean temperature over all fixes between 7 a.m. to 7 p.m.
PressMorn	Average mean sea-level pressure before and including 10 a.m.
PressAft	Average temperature after and including 4 p.m.
PressDiff	Difference between morning and afternoon pressure averages
PressMean	Mean pressure over all fixes between 7 a.m. to 7 p.m.
RainMorn	Average rainfall intensity before and including 10 a.m.
RainAft	Average rainfall intensity after and including 4 p.m.
RainDiff	Difference between morning and afternoon rainfall intensity averages
RainMean	Mean rainfall intensity over all fixes between 7 a.m. to 7 p.m.
MaxState	The HMM movement state most present in during fixes between 7 a.m. to 7 p.m.
MaxStateWtAvg	The HMM movement state average over fixes between 7 a.m. to 7 p.m.
No4State	The number of fixes in HMM movement state four (longest distance)
Sent10kmMorn	Average proportion of local 10km area with water present over fixes before and including 10 a.m.
Sent10kmAft	Average proportion of local 10km area with water present over fixes after and including 4 p.m.
Sent10kmDiff	Difference between morning and afternoon Average proportion of local 10km area with water present
SentLndspMorn	Average proportion of local 10km area with water present over fixes before and including 10 a.m.
SentLndspAft	Average proportion of local 10km area with water present over fixes after and including 4 p.m.
SentLndspDiff	Difference between morning and afternoon Average proportion of local 10km area with water present
WindBenefitMean	Average wind benefit over all fixes between 7 a.m. to 7 p.m.
WindBenefitMax	Maximum wind benefit over all fixes between 7 a.m. to 7 p.m.
Sp	Species of waterbird for which individual and day is assigned
Sex	Sex of waterbird for which individual and day is assigned
AgeCat	Age (adult or juvenile) category of waterbird for which individual and day is assigned

Table 3 Summary of variables available for GAM modelling.

2.3 Generalised additive modelling of daily distances travelled

We modelled the daily distances moved using the environmental variables detailed and summarised in Table 3. For many of these variables, a linear relationship between the variation in the environmental variable and the average daily distance moved may not be the best fit. However, a priori hypotheses of the relationship for each variable is challenging. Generalised additive models (GAMs) are an extension of generalized linear model in which the response variable (daily distance travelled) depends linearly on unknown smooth functions of some predictor variables.

2.3.1 Methods

We use GAMs implemented in the `mgcv` R package [Wood, 2004] to model daily distance moved as a function of smooth functions of the derived environmental variables.

As an initial step we imputed variables with missing data with the mean value over all individuals. Specifically, temperature, pressure and rain variables along with Sentinel water variables contained missing values. The primary reason for missing data is that all morning or afternoon fixes were missing and thus no average could be computed. Approximately 1% of the covariate data were missing.

The daily distance variable is positive and does not contain zeroes. Therefore, we fitted a GAM using a natural logarithm transformation of the daily distances and a Gaussian family with an identity link. Residual plot investigations showed that this was sufficient to provide an adequate model fit. Formally, we model

$$y_i = \beta_0 + f_1(x_{1i}) + f_2(x_{2i}) + \dots f_p(x_{pi}) + \sum_{j=1}^{m} Z_j b_j + \varepsilon_i.$$

where $b_j \sim N(0, \sigma_b^2)$ and $\varepsilon_i \sim N(0, \sigma^2)$. Above the x_is contain the environmental variables and the Zs $n \times m$ model matrix encoding the random intercept for each of $m = 45$ straw-necked ibis. The elements of y_i correspond to the natural log transformed distances and β_0 controls the global intercept. In the above formulation we do not specify the size of p as this is to be determined via investigation and interrogation of results.

In the `mgcv` package random effects are specified as special cases of 'smooths' using terms like `s(z, bs='re')` in the model formula arguments of functions `gam`. We use the REML smoothing parameter estimation method when fitting the GAM model.

2.3.2 GAM model using maximum daily movement state

As an initial model, we fitted the following variables using all the straw-necked ibis data: max HMM state, a smooth term for month using cyclic cubic regression splines and smooth terms using thin-plate splines for Sentinel water differences, wind benefit average, rainfall intensity difference, temperature difference and pressure difference and the random intercept term for bird. We did not have any a priori notion of whether each of the variables would have a different relationship with distance moved given which movement state the bird was in. Therefore, as an initial pass we fitted each of the covariates using a factor-smooth interaction where the factor was the HMM movement state that the bird was maximally in for each day. These initial models include all variables of interest fitted as a smooth factor by movement state variables. The model fit statistics are included in Supp Tables S1 and S2.

One difficulty of the fully interacting model (by movement state) is that it makes the inclusion and exclusion of a variable difficult as you must then assess that variable as a fully interacting term. Furthermore, the data are not balanced with respect to movement states because there are far fewer points in the larger movement classes. The total variation in daily distances may be substantially increased with the inclusion of all data and may mask other processes that are acting in these larger movement classes but cannot be discriminated due to their small sample size relative to the total variation. We therefore chose to split the data by movement model classes, which makes inclusion and exclusion of variables simpler, allows for sex interaction terms and for simpler interpretation. This further reduced the computational burden of exploring a large set of models.

2.3.3 Customised GAM model for HMM state 4

If we only take days in which the movement state four is the maximum state then only 100 data points are available for GAM modelling. We hypothesised that birds may only be in movement state four for a few hours of the day and thus we included days in which there was at least one fix that was classified as being in movement state four. For these days there are 527 days available for analysis.

Given that we observed differences between adult and juvenile split sample analyses we use these state-4 data to run an analysis including month and all difference variables fitted by age class. Table S3 shows the results, where we see that the differentiation by age class leads to differentiated associations. However, if we fit the model without age differentiation we observe that the 'no age differentiation' model has a substantially lower AIC (975 drops to 957) (Table S4). This suggests that the relationships are similar enough between the age classes to not warrant different smooths to describe them.

By definition weather variables are highly correlated with season. To remove this effect we fit all morning and afternoon weather variables by month. To compare models we remove weather variables via significance of terms in the models and AIC comparison.

We converged on two models that included month as a measure of season, wind benefit, morning and afternoon temperatures as well as the random effect for bird. The model `mgcv` output is summarised in Tables S5 and S6. The model that only includes seasonal terms, wind benefit and inter-bird variation is the most parsimonious model although temperature appears to have some evidence for playing a role in September and October.

Wind was investigated by month and relationships were significant and similar in smooth shape by month. One concern is that wind could be directional and stronger in the months that birds travel. If birds travel along common routes when doing long-distance movements then wind could be correlated with time of year and times when birds are travelling long distances. Although it can be challenging to discriminate whether the birds are not using wind but simply moving when the wind is strong and directional, we do not believe that month, rather than wind, is the key driver of distance travelled. This is because the observed patterns of long-distance movements are not regular over the seasons, i.e. ibis have been observed making long-distance movements in most months, and they do not follow consistent directions that would indicate that they were waiting for particular seasonal effects such as prevailing monthly winds or responding to other monthly-scale seasonal cues such as temperature. Instead, the highly variable directions and timing of movements support the idea that birds are strategically using wind benefits on a daily rather than monthly time scale. As a result, there is less evidence that wind is an intermediate correlate to distance travelled. There is some evidence for its effect for movement state three (see next subsection), which are shorter movements. These smaller movements would likely be less correlated with time of year than longer-distance, more 'migratory-like' movements and perhaps further evidence that wind is assisting the very long ibis movements.

2.3.4 Customised GAM model for HMM movement state three

To focus on movement state three, we split the data and used those days where the major proportion of fixes was in movement state three. This left 275 days for modelling across 44 individuals. We followed a similar model selection procedure as that for the state four models and arrived at a final model that included season, sentinel water differences, and inter-birds differences as random effects. The largest model fitted as well as the reduced model are summarised in Tables S8 and Table S9. See Figures 7, 8, and 9 for the most influential covariate relationships. Seasonal and Sentinel water differences were the variables that were most important for explaining variation in movement state three. The relationship

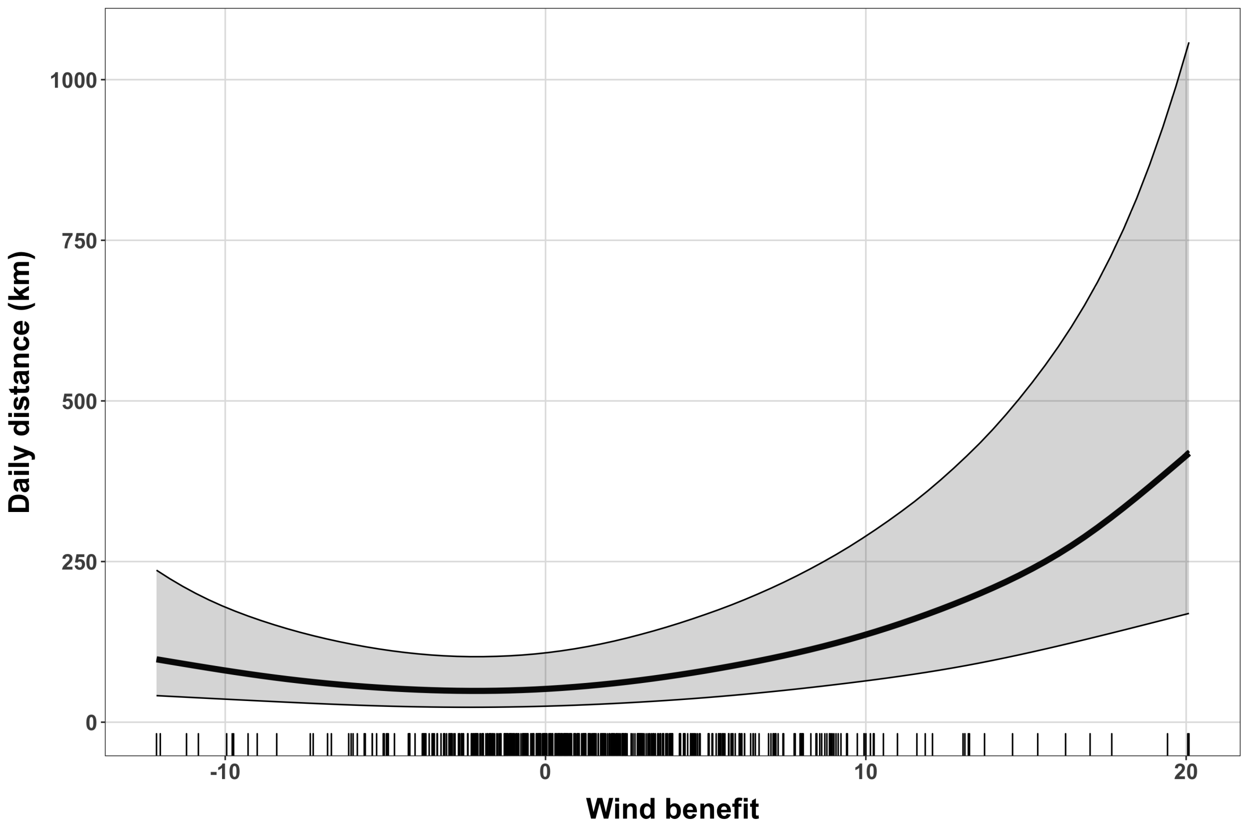

Fig. 5 Wind benefit relationship in movement state four for straw-necked ibis from GAM model with lowest AIC. Solid black line shows expected daily distance and shaded area the 95% prediction interval for daily distance with a other terms in the model held constant. Rug ticks show observed wind benefit measurements. The wind benefit for the day is averaged over each days measurements (avg. of 14 per day).

with season was similar to that observed in the HMM state 4 modelling. The relationships with Sentinel water differences (Figures 8 and 9) both suggested that ibis move further when they travel between dry and wet landscapes. However it should be noted that comparatively few data points support the edges of Sentinel water at the landscape variation (50km).

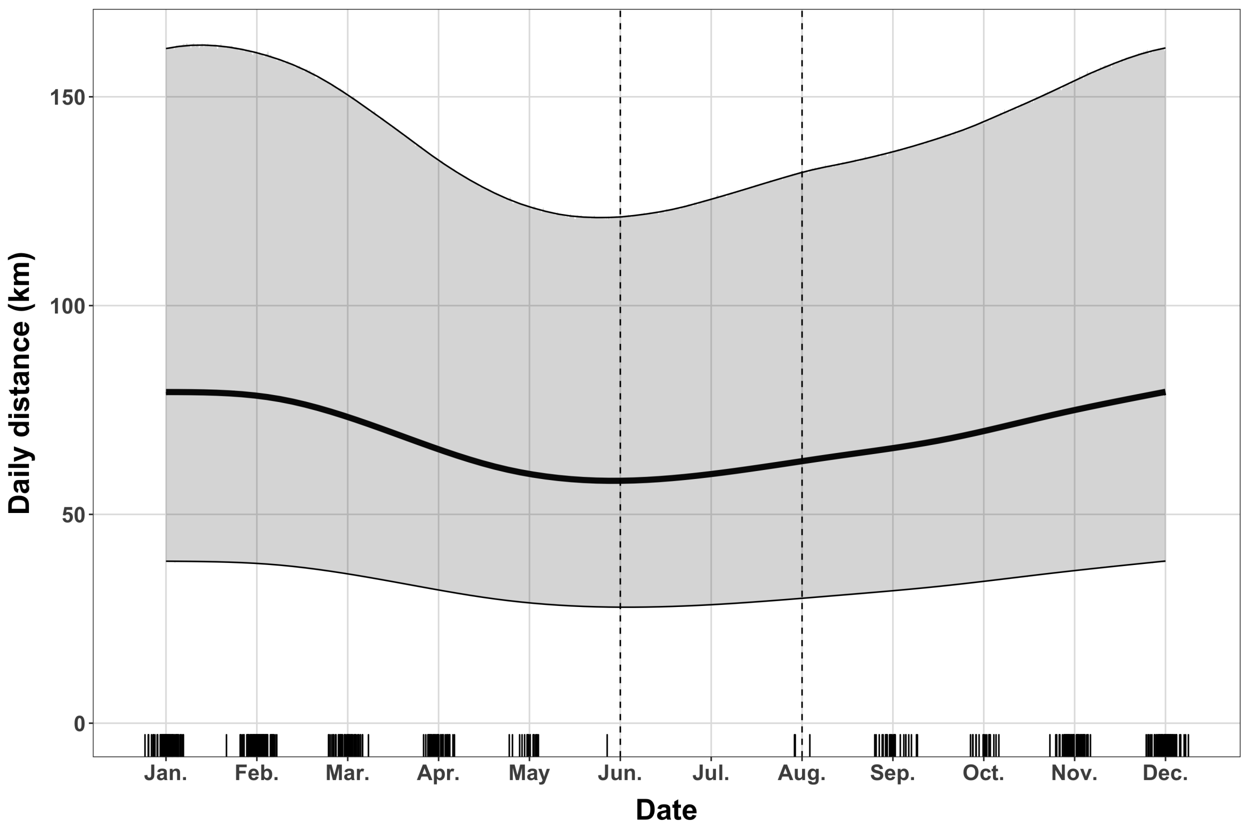

Fig. 6 Seasonal relationship in movement state four for straw-necked ibis from GAM model with lowest AIC.. The solid black line shows expected daily distance and shaded area the 95% prediction interval for daily distance with a other terms in the model held constant. Rug ticks show observed wind benefit measurements. No movement state four days were observed in July and only three in Jun and August combined. Individual SNI move very little in the winter months and this part of the curve should be interpreted with caution.

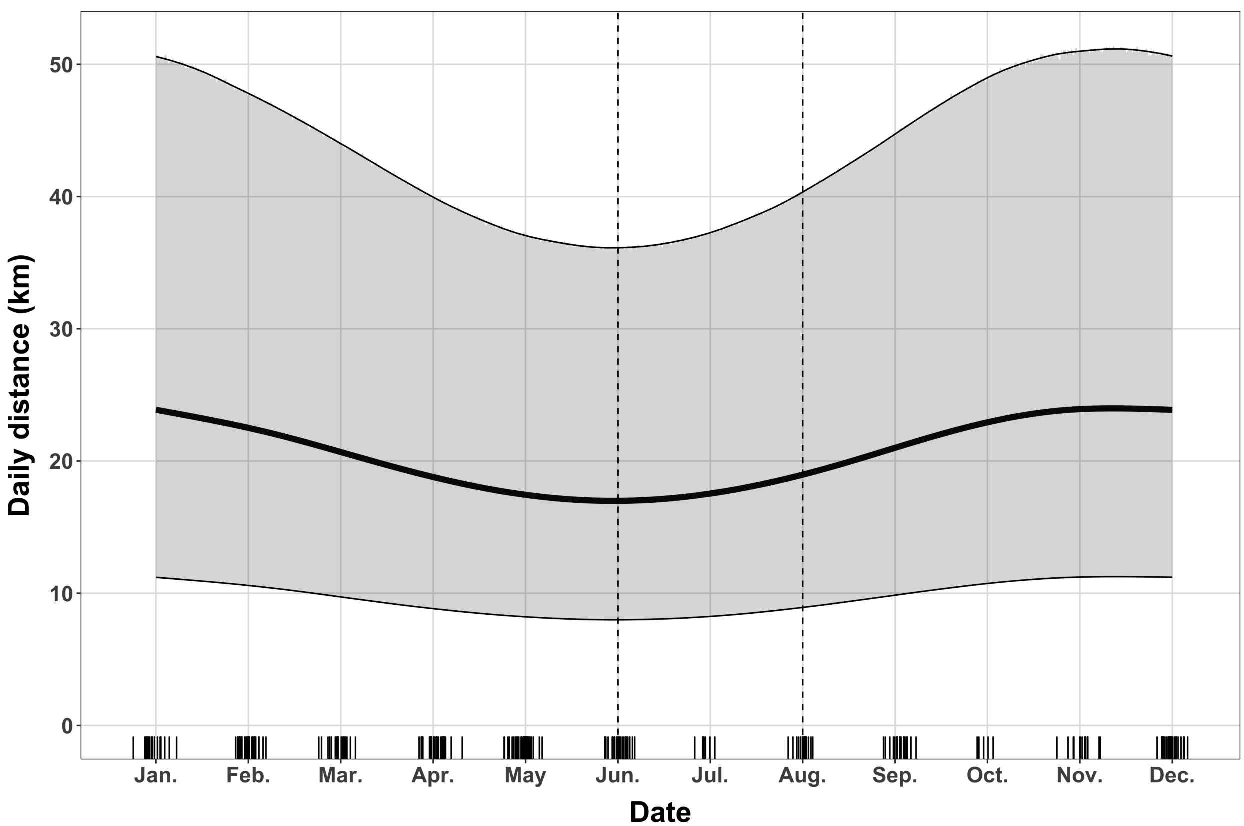

Fig. 7 **GAM smooth for season from distance model fitted to movement state three and straw-necked ibis.**

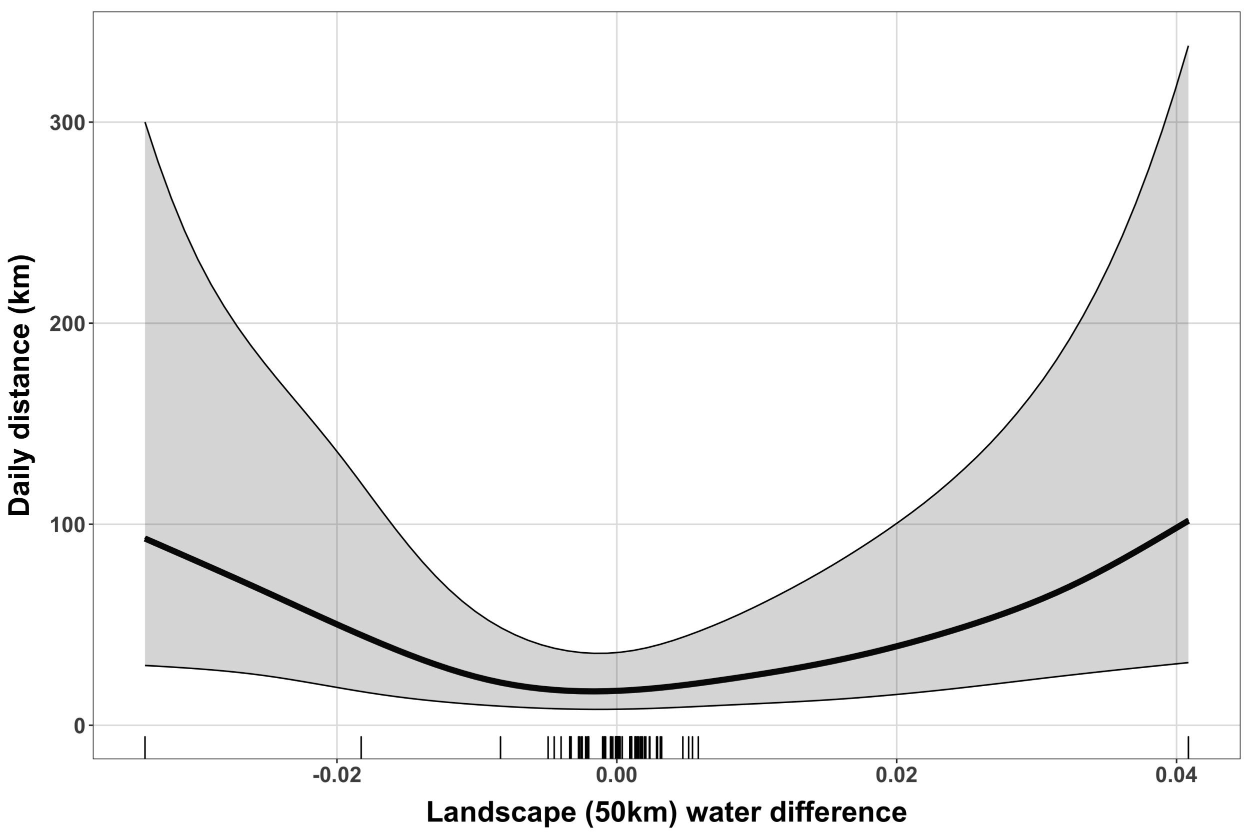

Fig. 8 **GAM smooth for Sentinel water differences at landscape scale movement state three and straw-necked ibis.**

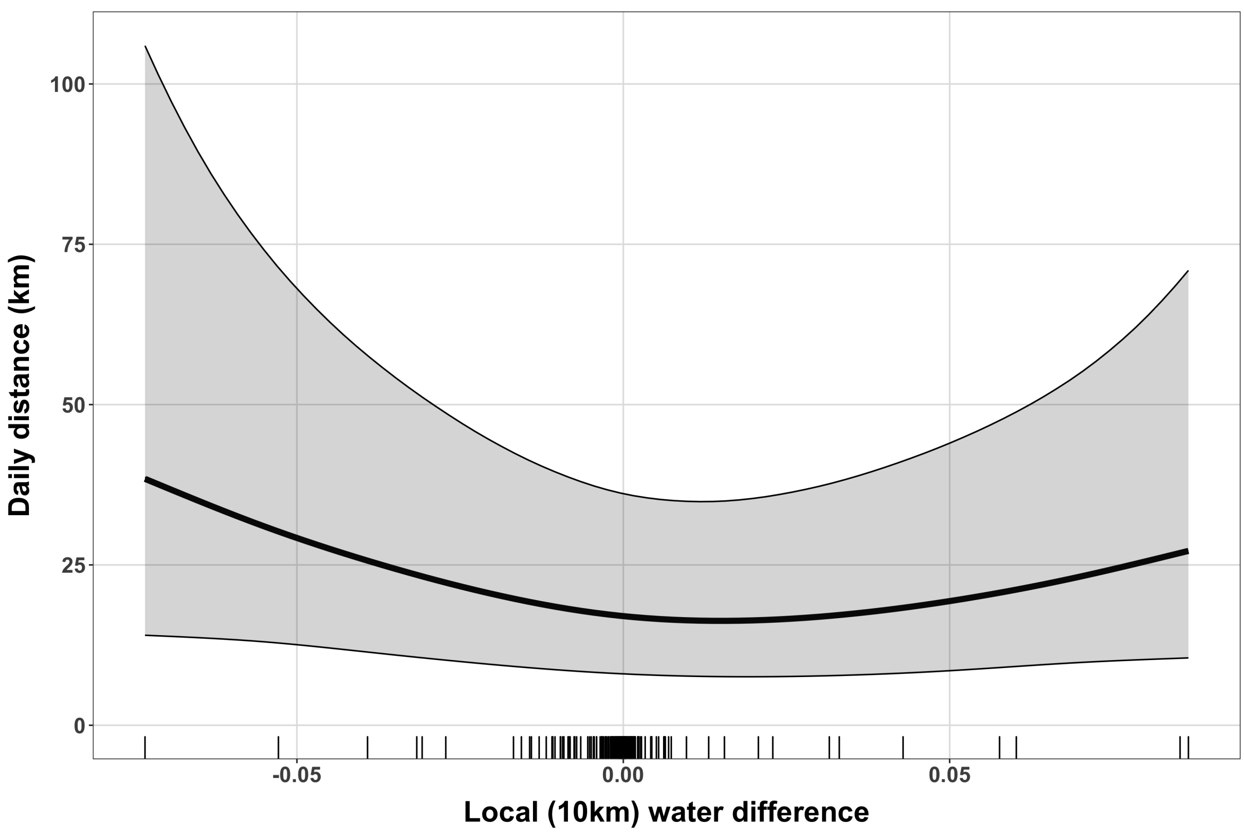

Fig. 9 **GAM smooth for Sentinel water differences at local scale (10 km) for movement state three and straw-necked ibis.**

3 Connectivity modelling

3.1 Computing prediction intervals from Generalised Additive Models

In order to be able to simulate from the GAM, we need to define the prediction interval for a given set of environmental covariates. More specifically, we need to define the prediction interval for some fixed values of the covariates fitted from a linear mixed model of log-transformed distances travelled on environmental covariates (fitted as fixed effects) and bird fitted as a random effect. The prediction interval summarises our uncertainty about the distribution of future values conditional on the model, the uncertainty in the parameters and the error of the prediction.

The GAM model is complex and there is no established theory to directly compute the prediction interval for new data given the GAM model. Therefore, we simulate from the posterior predictive distribution model using the following theory and procedure. The large-sample posterior distribution for the regression spline coefficients of a GAM is

$$\boldsymbol{\beta}|\mathbf{y}, \boldsymbol{\lambda}, \phi \sim N(\widehat{\boldsymbol{\beta}}, \mathbf{V}_{\boldsymbol{\beta}})$$

where $\widehat{\boldsymbol{\beta}}$ is the maximum penalized likelihood estimate of $\boldsymbol{\beta}$, which has the form $(\mathbf{X}^\top\mathbf{W}\mathbf{X}+\mathbf{S})^{-1}\mathbf{X}^\top\mathbf{W}\mathbf{z}$, $\mathbf{V}_{\boldsymbol{\beta}} = (\mathbf{X}^\top\mathbf{W}\mathbf{X}+\mathbf{S})^{-1}\phi$, $\mathbf{W}$ is the diagonal weight matrix at convergence of the GAM optimisation algorithm and $\mathbf{S} = \sum_j \lambda_j \mathbf{S}_j$. The entities $\mathbf{W}$ and ϕ are equal to the identity matrix and σ^2 when a normal response and identity link are used. These results are used to simulate from the posterior prediction distribution of bird flight distances given the fitted GAM model for each of the movement state classes.

The simulation is easier, if we compute the prediction matrix $\mathbf{X}_p$, mapping the model coefficients to the linear predictor, for any desired set of predictor variable values. The `mgcv` packages has the `predict.gam` function to compute the `lpmatrix`. We first extract the parameter estimates and their covariance matrix

```
beta <- coef(gam.mod)
V    <- vcov(gam.mod)
```

We then simulate a number of parameter vectors. This is done by assuming approximate normality and using the Cholesky trick.

```
num_beta_vecs <- nsim
Cv            <- chol(V)
nus           <- rnorm(num_beta_vecs * length(beta))
nu_mat        <- matrix(nus,  nrow = length(beta), ncol = num_beta_vecs)
beta_sims     <- beta + t(Cv) \%*\% nu_mat
```

We then calculate the linear predictors

```
covar_sim <- predict(gam.mod, newdata = data.new, type = "lpmatrix")
linpred_sim <- covar_sim \%*\% beta_sims
```

As this was a Normal model we use the multivariate normal to simulate for the posterior prediction distribution.

```
  y_sim <- matrix(rnorm(n    = prod(dim(linpred_sim)),
                        mean = linpred_sim,
                        sd   = summary(gam.mod)$scale),
                        nrow = nrow(linpred_sim),
                        ncol = ncol(linpred_sim))
```

and exponentiate these values to place them back on the positive distance scale.

3.2 A habitat preference model for ibis movements in the Basin

Birds do not fly in a straight line between origin and destination points. Instead, we hypothesized that they prefer to move through some landscapes over others. The relative ease or difficulty of moving through particular landscapes is quantified using a resistance surface. The resistance surface can then be used to generate least cost paths that represent the preferred routes between two points. We used an existing habitat preference model to generate the resistance surface for ibis in the Basin.

The observed telemetry points for the 44 SNI individuals were matched to random points in a spatial buffer for which availability points were randomly sampled. For the presence-availability points remote sensing data on a set of environmental covariates were obtained. An infinitely weighted logistic regression approach [Fithian and Hastie, 2013] is applied in a linear mixed modelling framework for the pooled set of observations to model individual variability and estimate the primary environmental variables that drive the distribution of these bird species at the scale of the population.

To create the availability points, a buffer was created around each telemetry point, where the buffer distance was the distance travelled along that line from the previous point. Within the buffer area for the entire 24 hours, 14 random points were generated (the same as the maximum number of tracking fixes per day). Using this method, long distance movements were therefore allocated larger buffer areas (and choice of habitats) than were short distance movements. These random available points and the presence points were then used to sample remote sensing environmental layers.

The context variables available used in the analysis include Multi-resolution Valley Bottom Flatness (MRVBF) [Gallant, John, Dowling, Trevor & Austin, Jenet, 2016], National Vegetation Information System (NVIS) [NVIS Technical Working Group, 2013], Catchment scale land use of Australia (CLUM) [ABARES, 2021], Australian National Aquatic Ecosystem (ANAE) [Brooks S. and J., 2013] and Water Observations from Space (WOfS) [Australia, 2015].

Presence and available points were combined into a single covariate matrix with each row representing an observed telemetry point with each column containing covariate information. Rows were differentiated as presence or availability points using an indicator variable with presence encoded as 1 and availability as 0. Some measures had substantial missing values and thus to make downstream analyses simpler and complete, missing values were imputed with the mean value or the missing variable for each of the factor variables. For example, NVIS has a factor 99 variables that represents unknown/no data. ANAE also had substantial missing values as it only classifies pixels that are wetlands and thus a new level of the ANAE factor variable was created to represent *not an ANAE category* and missing values imputed with this dummy value.

The total data after initial quality control and subsetting to complete cases for habitat selection analyses included 281,894 presence-availability points for 285 variables when all levels of factor variables were converted to 0,1 variables for each level using the model.matrix() function in R.

The habitat preference model fit an infinitely weighted logistic regression mixed model (IWGLMM) with $W = 10^5$ using the `glmmTMB` software [Brooks et al., 2017]. Initial analyses showed poor convergence when fitting all covariates due to complete or near-complete separation of factor levels or highly correlated variables. To overcome this numerical instability, levels of factor variables were removed if the level contained no information, or if at least one cell count of the two-way table of the factor level versus presence-availability status was less than 10. To reduce the dimension of the IWGLMM and remove highly correlated variables we applied IWLR Lasso [Fithian and Hastie, 2013] using the `glmnet()` software in R. Ten-fold cross-validation was used to choose the best lasso penalty parameter under an AUC loss measure within the `cv.glmnet` function in R [Friedman et al., 2010]. The lasso analysis chooses a best subset of non-zero variables, which are then used to fit the IWGLMM analysis. The results of the IWLR Lasso complement the IWGLMM analyses but are viewed in this analysis as an initial covariate screening method as manual screening of highly correlated continuous and factor variables to a set that led to IWGLMM convergence was practically infeasible.

Post quality control and dimension reduction using the IWLR Lasso a dataset (size $\sum_n I_n \times p$) of size $281,894 \times 148$ was used to run the IWGLMM. All variables chosen from the IWLR Lasso were fitted in the IWGLMM with the inclusion of quadratic terms for WOfS. Auto-correlation error structures were investigated but the fitting of them imposed substantial computational complexity and standard error and p-values results appeared marginally altered.

To compute the cost surface a spatial points object was created for the whole Murray-Darling basin at a 500 metre resolution, which was approximately 4 million points. Variables were matched to each of the points for MRVBF, NVIS, CLUM, ANAE and WOfS. A model matrix was constructed from these data, which contained 324 variables. The variables set was matched to the 147 variables for which coefficients were available from the IWGLMM analysis. The linear predictor was computed through matrix multiplication of the model matrix with the estimated coefficients. The linear predictor is summarised as the cost C of moving through a grid cell:

$$C = Intercept + ANAE + MRVBF + NVIS + CLUM + WOFS + (WOFS)^2 \qquad (1)$$

The linear predictor was exponentiated to be on the natural scale after the link function transformation and then scaled by the maximum value to be a proportional value as it is a relative measure of density. The surface used and the coefficients are available in the BitBucket repository associated with the manuscript.

3.3 Using the posterior predictive distribution of distance to investigate connectivity

To make use of the predicted distance model (section 3.1) we took the main breeding habitats of the SNI and derived the least cost path between these habitats based on habitat selection models generated in other work (section 3.2). Given the least-cost path distance, we simulate the connectivity between these points by asking how many instances of a five-day flight in movement state four or state-three exceeds the least-cost path distance. The proportion of flights exceeding the distance is the probability of moving between these two breeding sites under the model and covariates stipulated for the GAM simulation.

To create outputs that varied by month, we derived seasonal wind maps for the Murray-Darling Basin using wind data derived from the Bureau of Meteorology Atmospheric high-resolution Regional Reanalysis for Australia (BARRA). These data have wind data in 'u' and 'v' vector components at the height of 10 metres at an hourly time scale. The wind data also cover the dates of the bird telemetry recordings. We downloaded all wind data for the years 2015–2019 at a 12 km resolution for the whole basin. The data were downloaded from the National Computing Infrastructure at:

`https://dapds00.nci.org.au/thredds/catalog/cj37/BARRA`.

For each day of the year, we averaged the wind vector components between 9 and 12 a.m. because this is the most likely time that individuals departed for long-distance movements in the telemetry data set. Over the five years, the wind vector components were pooled by month and then averaged. The final product was a monthly spatial data set that showed wind speed and direction.

To make the connectivity maps more realistic we took each of the wetland habitat connections and their least-cost paths and computed the average wind benefit along the path. To do this, each least-cost path was decomposed into a set of points and the angle computed between each set of successive points. The wind benefit was computed using the path angle between points and the wind speed and direction for all points along all paths for each season. These average wind benefits are then used as inputs to the simulation function as the wind benefit for each season and each combination of least-cost path origin and destination.

References

ABARES. Catchment scale land use of australia – update december 2020. 10.25814/aqjw-rq15, 2021. Data collection.

Geoscience Australia. Australian water observations from space (wofs) - water summary, filtered. bioregional assessment source dataset. http://data.bioregionalassessments.gov.au/dataset/719a5433-2af0-4601-8036-a03f77199442 2015. Data collection.

Mollie E Brooks, Kasper Kristensen, Koen J van Benthem, Arni Magnusson, Casper W Berg, Anders Nielsen, Hans J Skaug, Martin Machler, and Benjamin M Bolker. glmmtmb balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R journal*, 9(2):378–400, 2017.

Butcher R. Brooks S., Cottingham P. and Hale J. Murray-darling basin aquatic ecosystem classification: Stage 2 report. Technical report, Murray-Darling Basin Authority, Canberra, Australia, 2013.

C. Calenge. The package adehabitat for the r software: tool for the analysis of space and habitat use by animals. *Ecological Modelling*, 197:1035, 2006.

Javier Fernández-López and Klaus Schliep. rwind: download, edit and include wind data in ecological and evolutionary analysis. *Ecography*, 42(4), 2018.

William Fithian and Trevor Hastie. Finite-sample equivalence in statistical models for presence-only data. *The annals of applied statistics*, 7(4):1917, 2013.

Jerome Friedman, Trevor Hastie, and Rob Tibshirani. Regularization paths for generalized linear models via coordinate descent. *Journal of statistical software*, 33(1):1, 2010.

Gallant, John, Dowling, Trevor & Austin, Jenet. Multi-resolution valley bottom flatness (mrvbf). version v3. https://doi.org/10.4225/08/5701C885AB4FE, 2016. Data collection.

Hans Hersbach, Bill Bell, Paul Berrisford, Shoji Hirahara, András Horányi, Joaquín Muñoz-Sabater, Julien Nicolas, Carole Peubey, Raluca Radu, Dinand Schepers, et al. The era5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730):1999–2049, 2020.

Theo Michelot, Roland Langrock, and Toby A Patterson. movehmm: an r package for the statistical modelling of animal movement data using hidden markov models. *Methods in Ecology and Evolution*, 7(11):1308–1315, 2016.

Théo Michelot, Roland Langrock, and Toby Patterson. movehmm: An r package for the analysis of animal movement data, 2019.

NVIS Technical Working Group. *Australian Vegetation Attribute Manual: National Vegetation Information System, version 7.0.* Department of the Environment and Energy, Canberra, ACT, Australia, 2013.

Toby A Patterson, Marinelle Basson, Mark V Bravington, and John S Gunn. Classifying movement behaviour in relation to environmental conditions using hidden markov models. *Journal of Animal Ecology*, 78(6):1113–1123, 2009.

S. N. Wood. Stable and efficient multiple smoothing parameter estimation for generalized additive models. *Journal of the American Statistical Association*, 99(467):673–686, 2004.

4 Supplementary Tables

```
Family: gaussian
Link function: identity

Formula:
LogDayDist ~ MaxState + s(Month, by = MaxState, bs = "cc", k = 5) +
    s(Sent10kmDiff, by = MaxState, bs = "tp") + s(SentLndspDiff,
    by = MaxState, bs = "tp") + s(WindBenefitMean, by = MaxState,
    bs = "tp") + s(RainDiff, by = MaxState, bs = "tp") + s(TempDiff,
    by = MaxState, bs = "tp") + s(PressDiff, by = MaxState, bs = "tp") +
    s(BandID, bs = "re")

Parametric coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.91094    0.12439   7.323 2.63e-13 ***
MaxState2    0.52084    0.03638  14.316  < 2e-16 ***
MaxState3    1.71151    0.06386  26.801  < 2e-16 ***
MaxState4    3.74382    0.12652  29.590  < 2e-16 ***

Approximate significance of smooth terms:
                                edf Ref.df       F  p-value
s(Month):MaxState1           2.962049  3.000 544.147  < 2e-16 ***
s(Month):MaxState2           0.002085  3.000   0.000  0.57605
s(Month):MaxState3           1.477615  3.000   1.814  0.03521 *
s(Month):MaxState4           0.002823  3.000   0.001  0.41873
s(Sent10kmDiff):MaxState1    8.909968  8.997  86.288  < 2e-16 ***
s(Sent10kmDiff):MaxState2    5.466755  6.243   6.171 1.71e-06 ***
s(Sent10kmDiff):MaxState3    1.623977  2.035   0.578  0.56658
s(Sent10kmDiff):MaxState4    1.002679  1.005   0.583  0.44702
s(SentLndspDiff):MaxState1   8.863317  8.982  63.934  < 2e-16 ***
s(SentLndspDiff):MaxState2   3.651690  4.254   8.164 4.78e-07 ***
s(SentLndspDiff):MaxState3   2.223976  2.678   1.920  0.10974
s(SentLndspDiff):MaxState4   1.001401  1.003   0.094  0.76097
s(WindBenefitMean):MaxState1 5.175962  6.373  10.386  < 2e-16 ***
s(WindBenefitMean):MaxState2 1.293031  1.544   1.934  0.21917
s(WindBenefitMean):MaxState3 1.001030  1.002   0.140  0.70945
s(WindBenefitMean):MaxState4 1.596406  1.995   2.798  0.06551 .
s(RainDiff):MaxState1        1.009651  1.019   0.000  0.99698
s(RainDiff):MaxState2        1.000917  1.002   0.232  0.63074
s(RainDiff):MaxState3        1.000974  1.002   0.013  0.91225
s(RainDiff):MaxState4        1.002752  1.005   0.932  0.33413
s(TempDiff):MaxState1        3.716896  4.689   4.103  0.00178 **
s(TempDiff):MaxState2        1.000637  1.001   2.411  0.12043
s(TempDiff):MaxState3        1.000978  1.002   2.369  0.12370
s(TempDiff):MaxState4        1.001049  1.002   1.109  0.29201
s(PressDiff):MaxState1       4.087137  5.120   3.381  0.00440 **
s(PressDiff):MaxState2       1.001163  1.002   0.000  0.99987
s(PressDiff):MaxState3       1.001151  1.002   0.021  0.88813
s(PressDiff):MaxState4       1.001489  1.003   0.071  0.79310
s(BandID)                   41.295283 43.000  41.679  < 2e-16 ***
R-sq.(adj) =  0.482   Deviance explained = 48.8%
-REML =  13042  Scale est. = 0.91977   n = 9297
```

Table S1 Straw-necked ibis all individuals GAM model output.

```
Family: gaussian
Link function: identity

Formula:
LogDayDist ~ MaxState + s(Month, by = MaxState, bs = "cc", k = 5) +
    s(Sent10kmDiff, by = MaxState, bs = "tp") + s(SentLndspDiff,
    by = MaxState, bs = "tp") + s(WindBenefitMean, by = MaxState,
    bs = "tp") + s(RainDiff, by = MaxState, bs = "tp") + s(TempDiff,
    by = MaxState, bs = "tp") + s(PressDiff, by = MaxState, bs = "tp") +
    s(BandID, bs = "re")

Parametric coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.47857    0.05531   26.73   <2e-16 ***
MaxState2    0.46901    0.04502   10.42   <2e-16 ***
MaxState3    1.66851    0.07782   21.44   <2e-16 ***
MaxState4    3.54505    0.14565   24.34   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Approximate significance of smooth terms:
                                edf Ref.df      F  p-value
s(Month):MaxState1          2.948e+00  3.000 47.290  < 2e-16 ***
s(Month):MaxState2          1.241e+00  3.000  1.110 0.088958 .
s(Month):MaxState3          1.681e+00  3.000  2.560 0.008365 **
s(Month):MaxState4          3.487e-04  3.000  0.000 0.972735
s(Sent10kmDiff):MaxState1    8.802e+00  8.986 46.736  < 2e-16 ***
s(Sent10kmDiff):MaxState2    4.865e+00  5.780  4.261 0.000625 ***
s(Sent10kmDiff):MaxState3    1.524e+00  1.880  0.370 0.666887
s(Sent10kmDiff):MaxState4    1.001e+00  1.001  0.131 0.717629
s(SentLndspDiff):MaxState1   8.345e+00  8.758 48.678  < 2e-16 ***
s(SentLndspDiff):MaxState2   3.791e+00  4.467  4.025 0.002341 **
s(SentLndspDiff):MaxState3   1.159e+00  1.295  1.719 0.228491
s(SentLndspDiff):MaxState4   1.001e+00  1.002  0.039 0.845347
s(WindBenefitMean):MaxState1 4.873e+00  6.070  8.550  < 2e-16 ***
s(WindBenefitMean):MaxState2 1.001e+00  1.003  1.294 0.255631
s(WindBenefitMean):MaxState3 1.001e+00  1.001  0.461 0.497210
s(WindBenefitMean):MaxState4 1.001e+00  1.001  7.576 0.005923 **
s(RainDiff):MaxState1       1.167e+00  1.317  0.271 0.600645
s(RainDiff):MaxState2       1.000e+00  1.001  0.212 0.645573
s(RainDiff):MaxState3       1.000e+00  1.000  0.622 0.430540
s(RainDiff):MaxState4       1.000e+00  1.000  0.076 0.783618
s(TempDiff):MaxState1       3.059e+00  3.889  3.614 0.006972 **
s(TempDiff):MaxState2       1.001e+00  1.002  5.638 0.017589 *
s(TempDiff):MaxState3       1.001e+00  1.001  1.215 0.270311
s(TempDiff):MaxState4       1.000e+00  1.001  1.243 0.264952
s(PressDiff):MaxState1      3.583e+00  4.546  2.077 0.071783 .
s(PressDiff):MaxState2      1.000e+00  1.001  0.991 0.319647
s(PressDiff):MaxState3      1.001e+00  1.002  0.021 0.887820
s(PressDiff):MaxState4      1.001e+00  1.002  0.134 0.714787
s(BandID)                   1.422e+01 17.000 11.274  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

R-sq.(adj) =  0.434   Deviance explained = 44.3%
-REML = 6390.6  Scale est. = 0.8196    n = 4741
```


Table S2 Straw-necked ibis all adult individuals GAM model output.

```
Family: gaussian
Link function: identity

Formula:
LogDayDist ~ s(Month, by = AgeCat, bs = "cc") + s(Sent10kmDiff,
    by = AgeCat, bs = "tp") + s(SentLndspDiff, by = AgeCat, bs = "tp") +
    s(WindBenefitMean, by = AgeCat, bs = "tp") + s(RainDiff,
    by = AgeCat, bs = "tp") + s(TempDiff, by = AgeCat, bs = "tp") +
    s(PressDiff, by = AgeCat, bs = "tp") + s(BandID, bs = "re")

Parametric coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.32744    0.04814   89.89   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Approximate significance of smooth terms:
                                    edf Ref.df      F  p-value
s(Month):AgeCatAdult              2.916  8.000  3.076 0.000674 ***
s(Month):AgeCatJuvenile           1.607  8.000  0.720 0.052930 .
s(Sent10kmDiff):AgeCatAdult       1.000  1.000  0.089 0.766098
s(Sent10kmDiff):AgeCatJuvenile    1.000  1.000  0.117 0.732632
s(SentLndspDiff):AgeCatAdult      2.762  3.469  0.856 0.427833
s(SentLndspDiff):AgeCatJuvenile   1.000  1.000  7.951 0.005001 **
s(WindBenefitMean):AgeCatAdult    3.955  4.929 22.475  < 2e-16 ***
s(WindBenefitMean):AgeCatJuvenile 3.974  4.919 12.955  < 2e-16 ***
s(RainDiff):AgeCatAdult           1.000  1.000  0.670 0.413329
s(RainDiff):AgeCatJuvenile        1.000  1.000  0.081 0.775880
s(TempDiff):AgeCatAdult           1.001  1.002  7.101 0.007952 **
s(TempDiff):AgeCatJuvenile        1.000  1.000  0.000 0.998339
s(PressDiff):AgeCatAdult          1.000  1.000  4.264 0.039454 *
s(PressDiff):AgeCatJuvenile       1.267  1.481  0.858 0.316997
s(BandID)                        15.143 38.000  1.058 1.97e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

R-sq.(adj) =  0.355   Deviance explained = 40.4%
-REML = 504.69  Scale est. = 0.33143   n = 526
AIC =  975.1654
```

Table S3 Straw-necked ibis all individuals age class state four.

```
Family: gaussian
Link function: identity

Formula:
LogDayDist ~ s(Month, bs = "cc") + s(Sent10kmDiff, bs = "tp") +
    s(SentLndspDiff, bs = "tp") + s(WindBenefitMean, bs = "tp") +
    s(RainDiff, bs = "tp") + s(TempDiff, bs = "tp") + s(PressDiff,
    bs = "tp") + s(BandID, bs = "re")

Parametric coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.31654    0.04908   87.95   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                      edf Ref.df      F  p-value
s(Month)            2.741  8.000  3.202  0.00083 ***
s(Sent10kmDiff)     1.000  1.000  0.109  0.74113
s(SentLndspDiff)    2.651  3.379  1.478  0.19282
s(WindBenefitMean)  4.481  5.529 31.024  < 2e-16 ***
s(RainDiff)         1.000  1.000  0.705  0.40165
s(TempDiff)         1.001  1.002  4.815  0.02868 *
s(PressDiff)        1.000  1.000  1.496  0.22181
s(BandID)          16.430 38.000  1.225 4.49e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.347   Deviance explained = 38.5%
-REML = 496.26  Scale est. = 0.33536   n = 526
AIC = 957.1762
```

Table S4 Straw-necked ibis all individuals state four.

```
Family: gaussian
Link function: identity

Formula:
LogDayDist ~ s(Month, bs = "cc", k = 6) + s(WindBenefitMean,
    bs = "tp") + s(TempMorn, by = MonthFac, bs = "tp", k = 5) +
    s(TempAft, by = MonthFac, bs = "tp", k = 5) + s(BandID, bs = "re")

Parametric coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   4.3853     0.0489   89.67   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Approximate significance of smooth terms:
                            edf    Ref.df      F p-value
s(Month)                9.610e-04 4.000e+00  0.000 0.35927
s(WindBenefitMean)      4.516e+00 5.563e+00 32.083 < 2e-16 ***
s(TempMorn):MonthFac1   1.564e+00 1.872e+00  0.454 0.63493
s(TempMorn):MonthFac2   1.919e+00 2.347e+00  0.695 0.66864
s(TempMorn):MonthFac3   1.000e+00 1.000e+00  0.474 0.49158
s(TempMorn):MonthFac4   1.000e+00 1.000e+00  0.002 0.96679
s(TempMorn):MonthFac5   1.000e+00 1.000e+00  1.746 0.18699
s(TempMorn):MonthFac6   1.000e+00 1.000e+00  0.714 0.39839
s(TempMorn):MonthFac8   1.280e+00 1.459e+00  0.151 0.72662
s(TempMorn):MonthFac9   1.645e+00 1.975e+00  4.863 0.00579 **
s(TempMorn):MonthFac10  1.848e+00 2.273e+00  7.357 0.00044 ***
s(TempMorn):MonthFac11  1.372e+00 1.651e+00  0.538 0.63653
s(TempMorn):MonthFac12  1.000e+00 1.000e+00  2.378 0.12369
s(TempAft):MonthFac1    1.591e+00 1.910e+00  1.165 0.24624
s(TempAft):MonthFac2    1.000e+00 1.000e+00  0.985 0.32142
s(TempAft):MonthFac3    1.653e+00 2.001e+00  0.499 0.61298
s(TempAft):MonthFac4    1.000e+00 1.000e+00  2.860 0.09143 .
s(TempAft):MonthFac5    2.708e+00 3.106e+00  3.251 0.01945 *
s(TempAft):MonthFac6    4.121e-17 4.121e-17  0.000 1.00000
s(TempAft):MonthFac8    1.078e+00 1.128e+00  0.130 0.82744
s(TempAft):MonthFac9    1.976e+00 2.346e+00  2.146 0.15786
s(TempAft):MonthFac10   1.000e+00 1.000e+00  8.795 0.00317 **
s(TempAft):MonthFac11   1.000e+00 1.000e+00  0.000 0.99673
s(TempAft):MonthFac12   1.000e+00 1.000e+00  0.779 0.37789
s(BandID)               1.155e+01 3.800e+01  0.558 0.00761 **
AIC = 960.3795
```

Table S5 Straw-necked ibis all individuals state four with second best AIC and temperature by month.

```
Family: gaussian
Link function: identity

Formula:
LogDayDist ~ s(Month, bs = "cc", k = 6) + s(WindBenefitMean,
    bs = "tp") + s(BandID, bs = "re")

Parametric coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.32848    0.04615   93.79   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Approximate significance of smooth terms:
                      edf Ref.df      F  p-value
s(Month)            2.582  4.000  8.806 7.19e-05 ***
s(WindBenefitMean)  4.363  5.392 31.814  < 2e-16 ***
s(BandID)          14.846 38.000  1.037 2.46e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

R-sq.(adj) =  0.335   Deviance explained = 36.2%
-REML = 487.02  Scale est. = 0.34192   n = 526
AIC=956.0898
```

Table S6 Straw-necked ibis all individuals state four with best AIC and no temperature but a seasonal term.

```
Family: gaussian
Link function: identity

Formula:
LogDayDist ~ s(SentLndspDiff, bs = "tp") + s(WindBenefitMean,
    bs = "tp") + s(TempMorn, bs = "tp") + s(BandID, bs = "re")

Parametric coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.30832    0.07185   59.97   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Approximate significance of smooth terms:
                    edf Ref.df      F  p-value
s(SentLndspDiff)   1.476  1.810  7.793  0.00308 **
s(WindBenefitMean) 4.444  5.450 19.709  < 2e-16 ***
s(TempMorn)        2.838  3.565  9.983 2.53e-06 ***
s(BandID)          9.598 22.000  1.445 4.32e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

R-sq.(adj) =  0.521   Deviance explained = 57.5%
-REML = 140.99  Scale est. = 0.24955   n = 166
```

Table S7 Straw-necked ibis juvenile individuals state four with best AIC to highlight landscape water effect.

```
Family: gaussian
Link function: identity

Formula:
LogDayDist ~ s(Month, bs = "cc", k = 10) + s(Sent10kmDiff, bs = "tp") +
    s(SentLndspDiff, bs = "tp") + s(WindBenefitMean, by = MonthFac,
    bs = "tp", k = 5) + s(RainDiff, by = MonthFac, bs = "tp",
    k = 5) + s(TempDiff, by = MonthFac, bs = "tp", k = 5) + s(PressDiff,
    by = MonthFac, bs = "tp", k = 5) + s(BandID, bs = "re")

Parametric coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.15675    0.06562   48.11   <2e-16 ***

Approximate significance of smooth terms:
                                edf Ref.df     F  p-value
s(Month)                      1.317  8.000 0.492  0.02622 *
s(Sent10kmDiff)               2.272  2.852 1.567  0.20380
s(SentLndspDiff)              4.175  4.870 6.803 2.08e-05 ***
s(WindBenefitMean):MonthFac1  1.954  2.332 2.352  0.09344 .
s(WindBenefitMean):MonthFac2  1.000  1.000 0.001  0.97905
s(WindBenefitMean):MonthFac3  1.000  1.000 0.005  0.94371
s(WindBenefitMean):MonthFac4  1.746  2.055 1.395  0.24798
s(WindBenefitMean):MonthFac5  1.467  1.790 0.681  0.38456
s(WindBenefitMean):MonthFac6  1.000  1.000 0.278  0.59832
s(WindBenefitMean):MonthFac7  1.000  1.000 1.300  0.25553
s(WindBenefitMean):MonthFac8  1.000  1.000 0.009  0.92555
s(WindBenefitMean):MonthFac9  1.000  1.000 0.140  0.70917
s(WindBenefitMean):MonthFac10 1.000  1.000 0.073  0.78719
s(WindBenefitMean):MonthFac11 2.836  3.252 2.642  0.04280 *
s(WindBenefitMean):MonthFac12 1.000  1.000 0.005  0.94578
s(RainDiff):MonthFac1         1.640  1.874 0.897  0.41805
s(RainDiff):MonthFac2         1.912  2.268 1.190  0.30907
s(RainDiff):MonthFac3         1.000  1.000 2.391  0.12366
s(RainDiff):MonthFac4         1.000  1.000 0.035  0.85077
s(RainDiff):MonthFac5         1.000  1.000 0.314  0.57557
s(RainDiff):MonthFac6         1.349  1.582 0.532  0.67571
s(RainDiff):MonthFac7         1.000  1.000 0.003  0.95350
s(RainDiff):MonthFac8         1.000  1.000 0.060  0.80680
s(RainDiff):MonthFac9         1.721  1.885 1.704  0.25482
s(RainDiff):MonthFac10        1.000  1.000 2.498  0.11556
s(RainDiff):MonthFac11        1.000  1.000 0.016  0.90092
s(RainDiff):MonthFac12        1.000  1.000 0.167  0.68327
s(TempDiff):MonthFac1         1.000  1.000 0.337  0.56244
s(TempDiff):MonthFac2         1.000  1.000 0.293  0.58868
s(TempDiff):MonthFac3         2.833  3.285 1.764  0.13767
s(TempDiff):MonthFac4         1.000  1.000 0.120  0.72890
s(TempDiff):MonthFac5         1.372  1.617 0.223  0.66187
s(TempDiff):MonthFac6         1.000  1.000 0.498  0.48103
s(TempDiff):MonthFac7         1.000  1.000 0.099  0.75359
s(TempDiff):MonthFac8         1.000  1.000 1.943  0.16495
s(TempDiff):MonthFac9         1.000  1.000 0.051  0.82208
s(TempDiff):MonthFac10        1.000  1.000 0.324  0.56958
s(TempDiff):MonthFac11        1.000  1.000 0.357  0.55103
s(TempDiff):MonthFac12        1.000  1.000 7.014  0.00874 **
s(PressDiff):MonthFac1        1.000  1.000 0.412  0.52180
s(PressDiff):MonthFac2        1.000  1.000 1.490  0.22370
s(PressDiff):MonthFac3        1.451  1.709 0.441  0.49898
s(PressDiff):MonthFac4        1.000  1.000 0.316  0.57464
s(PressDiff):MonthFac5        1.000  1.000 0.341  0.55993
s(PressDiff):MonthFac6        1.000  1.000 0.280  0.59709
s(PressDiff):MonthFac7        1.000  1.000 1.097  0.29616
s(PressDiff):MonthFac8        1.000  1.000 0.014  0.90580
s(PressDiff):MonthFac9        1.000  1.000 0.411  0.52223
s(PressDiff):MonthFac10       1.000  1.000 1.042  0.30856
s(PressDiff):MonthFac11       2.107  2.441 1.981  0.11061
s(PressDiff):MonthFac12       1.000  1.000 0.130  0.71858
s(BandID)                     8.740 27.000 0.568  0.02707 *
R-sq.(adj) =  0.338   Deviance explained = 51.9%
-REML =  225.5  Scale est. = 0.21401   n = 275
AIC= 440.9035
```


Table S8 Straw-necked ibis all individuals and days with max state three with all terms fitted.

```
Family: gaussian
Link function: identity

Formula:
LogDayDist ~ s(Month, bs = "cc", k = 10) + s(SentLndspDiffAbs,
    bs = "tp") + s(Sent10kmDiffAbs, bs = "tp", k = 3) + s(BandID,
    bs = "re")

Parametric coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.20048    0.06277   50.99   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Approximate significance of smooth terms:
                       edf Ref.df      F  p-value
s(Month)             2.604  8.000  6.983  2.9e-06 ***
s(SentLndspDiffAbs)  1.818  2.210 18.197  < 2e-16 ***
s(Sent10kmDiffAbs)   1.000  1.001  7.565 0.006373 **
s(BandID)           13.869 27.000  1.282 0.000447 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

R-sq.(adj) =   0.33   Deviance explained = 37.7%
-REML = 202.23  Scale est. = 0.21648   n = 275
AIC=381.8005
```

Table S9 Straw-necked ibis all individuals and days with max state three and reduced terms with the absolute value on the Sentinel water terms.