# NEW METHODS AND APPLICATIONS FOR CONTEXT AWARE MOVEMENT ANALYSIS (CAMA) 

Vanessa da Silva Brum Bastos

A Thesis Submitted for the Degree of PhD at the University of St Andrews


2019

Full metadata for this item is available in St Andrews Research Repository at:
http://research-repository.st-andrews.ac.uk/

Identifiers to use to cite or link to this thesis:
DOI: https://doi.org/10.17630/10023-16812 http://hdl.handle.net/10023/16812

This item is protected by original copyright

# New Methods and Applications for Context Aware Movement Analysis (CAMA) 

Vanessa da Silva Brum Bastos

This thesis is submitted in partial
fulfillment for the degree of PhD at the University of St Andrews
December 3, 2018


## Abstract

Recent years have seen a rapid growth in movement research owing to new technologies contributing to the miniaturization and reduced costs of tracking devices. Similar trends have occurred in how environmental data are being collected (e.g., through satellites, unmanned aerial vehicles, and sensor networks). However, the development of analytical techniques for movement research has failed to keep pace with the data collection advances. There is a need for new methods capable of integrating increasingly detailed movement data with a myriad of contextual data - termed context aware movement analysis (CAMA). CAMA investigates more than movement geometry, by including biological and environmental conditions that may influence movement. However, there is a shortage of methods relating movement patterns to contextual factors, which is still limiting our ability to extract meaningful information from movement data. This thesis contributes to this methodological research gap by assessing the state-of-the art for CAMA within movement ecology and human mobility research, developing innovative methods to consider the spatio-temporal differences between movement data and contextual data and exploring computational methods that allow identification of patterns in contextualized movement data. We developed new methods and demonstrated how they facilitated and improved the integration between high frequency tracking data and temporally dynamic environmental variables. One of the methods, multi-channel sequence analysis, is then used to discover varying human behaviour relative to weather conditions in a large human GPS tracking dataset from Scotland. The second method is developed for combing multi-sensor satellite imagery (i.e., image fusion) of differing spatial and temporal resolutions. This method is applied to a GPS tracking data on maned wolves in Brazil to understand finescale movement behaviours related to vegetation changes across seasons. In summary, this
thesis provides a significant development in terms of new ideas and techniques for performing CAMA for human and wildlife movement studies.

## Acknowledgements

First, I would like to thank my supervisor Dr Urška Demšar for the continuous support throughout my PhD and related research. Your friendship, motivation, kindness, knowledge and guidance helped me immeasurably, especially in the tough times. I also would like to thank my second supervisor Dr Jed Long, one of the brightest researchers I have had the opportunity to work with, for all the brainstorm sessions, insightful witty comments, team spirit, humour and encouragement. I could not have imagined having better mentors for my PhD studies and I am extremely grateful for the years we have worked together.

Great thanks also goes to my friends: Anouk Borst, Bożena Wielgoszewska, Carina Gärtner, Elena Lo Giudice Cappelli, Katarzyna Sila-Nowicka, Paula Duffy and Will Hilles, for all the emotional support, discussions, laughs, encouragement, fun trips and motivation during the course of these four years.

Anouk, Elena and Sila, you have been my heroes in these three last months of insanity!!! I definitely could not have done it without you guys. $๑ \bigcirc \bigcirc$ LOVE YOU LOADS GIRLS $\triangle \subseteq \bigcirc$

Special thanks also goes to Sebastijan Sekulić for lending me a hand when all my coding efforts seemed hopeless and for patiently listening to my grumpy moaning at times.

I would like to extend my gratitude to all my fellow PhD students and Post-doc researchers at the School of Geography and Sustainable Development and at the School of Geosciences, for all the discussions, helpful advice, giggles at lunch time and shared frustration at times. You all have made this journey more interesting than my words could ever possibly describe!

I would also like to thank my undergraduate research supervisor Dr Nelson Jesus Ferreira for persuading me to do a PhD abroad in the first place.
iv

Thanks are also due to Helen Olaez and Maaret Karjalainen for all the administrative help, and Colin Cameron for the IT help during the course of my PhD.

I would also like to thank my husband Rory and my mother-in-law Helena, your support made this journey lighter, successful and possible. $\triangle$ D® I dearly love you $\triangle \bigcirc \bigcirc$

Finally, to my mum Vanda: Obrigada por todo suporte, ensinamentos e afeto durante todos esses anos. Você é, e sempre será a inspiração para que eu me torne um ser humano melhor a cada dia $\wp$ !

## Funding

The research leading to these results has received funding from the Science Without Borders Programme (CAPES BEX 3438/13-1) in the form of the author's PhD scholarship.

## Candidate's declaration

I, Vanessa da Silva Brum Bastos, do hereby certify that this thesis, submitted for the degree of PhD, which is approximately 45,000 words in length, has been written by me, and that it is the record of work carried out by me, or principally by myself in collaboration with others as acknowledged, and that it has not been submitted in any previous application for any degree.

I was admitted as a research student at the University of St Andrews in September 2014.
I received funding from an organisation or institution and have acknowledged the funder(s) in the full text of my thesis.

Date $05 / 12 / 18$
Signature of candidate

## Supervisor's declaration

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of PhD in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date 511212 el8 Signature of supervisor

## Permission for publication

In submitting this thesis to the University of St Andrews we understand that we are giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. We also understand, unless exempt by an award of an embargo as requested below, that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that this thesis will be electronically accessible for personal or research use and that the library has the right to migrate this thesis into new electronic forms as required to ensure continued access to the thesis.

I, Vanessa da Silva Brum Bastos, have obtained, or am in the process of obtaining, thirdparty copyright permissions that are required or have requested the appropriate embargo below.

The following is an agreed request by candidate and supervisor regarding the publication of this thesis:

## Printed copy

Embargo on all of print copy for a period of 2 years on the following ground(s):

- Publication would preclude future publication


## Supporting statement for printed embargo request

Journals will not accept it for publication If it is already published .

## Electronic copy

Embargo on all of electronic copy for a period of 2 years on the following ground(s):

- Publication would preclude future publication


## Supporting statement for electronic embargo request

Publication would preclude further publication

## Title and Abstract

- I agree to the title and abstract being published.


## Underpinning Research Data or Digital Outputs

## Candidate's declaration

I, Vanessa da Silva Brum Bastos, hereby certify that no requirements to deposit original research data or digital outputs apply to this thesis and that, where appropriate, secondary data used have been referenced in the full text of my thesis.

Date $05 / 42 / 18$
Signature of candidate

## List of symbols

- cm - centimetre
- h - hour
- km - kilometre
- m - metre
- mm - millimetre
- min - minute
- s-second
- $W / m^{2} s r$ - watt per square metre per steradian
- ${ }^{\circ} \mathbf{C}$ - degrees Celsius


## List of acronyms

- AM - Arithmetic Mean
- ANOVA - Analysis Of Variance
- ARGOS - Advanced Research and Global Observation Satellite
- ASTER - Advanced Spaceborne Thermal Emission and Reflection Radiometer
- AT - Apparent Temperature
- AVHRR - Advanced Very High Resolution Radiometer
- AVTS - Average Time Spent
- B - Blue
- BIRCH - Balanced Iterative Reducing and Clustering using Hierarchies
- CAMA - Context-Aware Movement Analysis
- CASA - Context-Aware Similarity Analysis
- CBERS - China-Brazil Earth Resources Satellite
- CDTW - Context-based Dynamic Time Warping
- CENAP - National Carnivorous Mammals Research and Conservation Center
- CF - Cluster Feature
- CHI - Calinski-Harabaz Index
- CNP - Canastra National Park
- csv - comma-separated values
- DL - Day Light
- DN - Digital Number
- DTA - Dynamic Trajectory Annotation
- DTA:CS - Dynamic Trajectory Annotation: Cubic Spline
xiv
- DTA:L - Dynamic Trajectory Annotation: Linear
- DTW - Dynamic Time Warping
- EDD - Empirical Density Distribution
- EI - Entropy Index
- ENDVI - Enhanced Normalised Difference Vegetation Index
- Env-DATA - Environmental-data automated track annotation
- ET - Evening Twilight
- ETM+ - Enhanced Thematic Mapper plus
- EURODEER - European Roe Deer Information System
- FI - Far Infra-Red
- G - Green
- GAR - Global Atmospheric Reanalysis
- GPS - Global Positioning System
- HNH - High NDVI home range
- HNH-F - High NDVI home range - Flattened
- HNH-P - High NDVI home range - Peaked
- HR - Home Range
- HRCC - High Resolution CCD Camera
- ICMBio - Chico Mendes Institute for Biodiversity Conservation
- IUCN - International Union for Conservation of Nature
- IDW - Inverse Distance Weighting
- LBSN - Location-Based Social Networks
- LMM - Linear Mixing Model
- LNH - Low NDVI home range
- LNH-F - Low NDVI home range - Flattened
- LNH-P -Low NDVI home range - Peaked
- MCSA - Multi-Channel Sequence Analysis
- MET - Meteorological station
- Met Office - UK Meteorological Office
- MIDAS - Met Office Integrated Data Archive System
- MISH - Meteorological Interpolation based on Surface Homogenized Data Basis
- MSA - Multi-channel Sequence Analysis
- MODIS - Moderate Resolution Imaging Spectroradiometer
- MT - Morning Twilight
- MW - Mother Wavelet
- NA - Neighbour After
- NB - Neighbour Before
- NDBI - Normalised Difference Built-up Index
- NDVI - Normalised Difference Vegetation Index
- NDWI - Normalised Difference Water Index
- NI - Night
- NIMROD - Now casting and Initialization for Modelling Using Regional Observation Data System
- NIR - Near Infra-Red
- NN - Nearest Neighbour
- NNH -Normal NDVI home range
- NNH-F - Normal NDVI home range - Flattened
- NNH-P - Normal NDVI home range - Peaked
- NPP - Net Primary Productivity
- OLI - Operational Land Imager
- OM - Optimal Match
- OBIS-SEAMAP - Spatial Ecological Analysis of Megavertebrate Animal Populations
- PC - Principal Component
- PCA - Principal Components Analysis
- PDF - Probability Density Function
- POI - Place Of Interest
- PVE - Percentage of Variance Explained
xvi
- PRISM - Parameter Regression on Independent Slopes Model
- R - Red
- RS - Remote Sensing
- SPOT - Satellite Pour l'Observation de la Terre
- STAT - Satellite Tracking and Analysis Tool
- SWIR - Short Wavelength Infra-Red
- TA - Trajectory Annotation
- TIR - Thermal Infra-Red
- TM - Thematic Mapper
- UD - Utilisation Distribution
- UK - United Kingdom
- VGI - Volunteered Geographic Information
- VNIR - Visible (RGB) and Near Infra-Red
- WT - Wavelet Transform


## List of Figures

1.1 Structure of the thesis . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 8
2.1 Proposed taxonomy for movement context with non-exhaustive examples of variables and how it relates to the classifications proposed by Nathan et al. (2008) (in orange at the top), Dodge et al. (2008) (in green at the bottom) and Sharif and Alesheikh (2017a) (in purple at the bottom). The red rectangle indicates the type of context this thesis focuses on.
2.2 Spectral reflectance curve of vegetation, soil and water from $0.4 \mu \mathrm{~m}$ to $2.5 \mu \mathrm{~m}$ wavelength. The dominant factors controlling leaf reflectance are the pigments in the palisade mesophyll (chlorophyll content), the scattering in the spongy mesophyll (leaf structure) and the amount of water in the plant. Adapted from SEOS (2018)
2.3 Difference between spatial resolution and pixel size. The first column shows an image grid with an specific resolution overlaying the roof of a house, the other three elements in the row show how the data registered at that resolution will be displayed in an image generated with a pixel size equal, smaller and bigger than the resolution. The red line show the equivalent size of the pixels on the ground.
2.4 Mixture effect illustrated for images with 90 m and 180 m spatial resolution over an urban area. It is clear that there is a higher diversity of objects within a 180 m nominal pixel than within a 90 m one.
2.5 Incompatibilities in spatial and temporal resolutions between movement data and contextual data after Neumann et al. (2015); Shen and Stopher (2014) and Meekan et al. (2017). Ellipses show the spatio-temporal range of movement data for different applications, black dots show the same for free sources of contextual data. The low-left quarter represents the high temporal and spatial resolution (HH) domain. Clockwise from there, we have the quarter with low temporal resolution and high spatial resolution (LH), the quarter with low temporal and spatial resolution (LL), and the quarter with high temporal resolution and low spatial resolution (HL).
2.6 Interpolation in space and time from Dodge et al. (2013). a) The variable data for track-point $P_{i}$ is first interpolated in space based on the data from the available pixels in the contextual dataset native grid around $\mathrm{P}_{\mathrm{i}} \cdot \mathrm{b}$ ) Similar spatial interpolations are conducted at the two nearest available points in time, the nearest before and nearest after the time-stamp of the fix $P_{i}$. Then, the two interpolated spatial values are interpolated in time to the time-stamp of $P_{i}$. . .
3.1 Two main types of mismatches between movement trajectories and environmental data represented in a space-time cube. Fixes are represented by yellow circles and trajectories by black lines. a) Trajectory's fixes are sampled more frequently than environmental data. b) Trajectory's fixes are sampled less frequently than
environmental data.
3.2 TA for a trajectory falling within a single pixel of a stack of spatial layers, the layer is omitted and only the pixel is illustrated for a clearer view. The trajectory is shown in space-time cubes with geographical space at the bottom and hypothetical environmental data displayed at corresponding times. a) Trajectory annotation: the value assigned to $j$ is equal to $v_{t 1}$, the value of the nearest neighbour in time; b) Dynamic trajectory annotation: the value assigned to j is an interpolated value between $\mathrm{v}_{\mathrm{t} 1}$ and $\mathrm{v}_{\mathrm{t} 2}$
3.3 Interpolation methods applied to a simulated trajectory. Trajectories are shown in space-time cubes with geographical space at the bottom and hypothetical environmental data displayed at corresponding times. The colour scale refers to the intensity of the hypothetical environmental value for a fix and/or layer. . .
3.4 Visual summary of datasets used for trajectory annotation. a) Location of Met Office stations and NIMROD coverage with respective resolutions; b) Illustration of 100 seconds segments of simulated trajectories for three Met Office stations.
3.5 Visual explanation of the trapezoidal rule applied to our variables.
3.6 Five interpolation methods with corresponding rainfall rates and precipitation (accumulated rainfall) curves for a segment of the stationary trajectory simulated at Sutton Bonington (Nottinghamshire) weather station. The segment of trajectory is from the $31^{\text {st }}$ of October 2013 , it is shown in space-time cubes with the rainfall rates displayed by different colouring of the fixes. The precipitation is displayed by the red line on the graph under each space-time cube.
3.7 Difference in mm of rainfall for all methods by interpolation method: pink dots show outliers; green line represents the median.
4.1 Comparison between different approaches for obtaining contextual data on one variable. Black lines represent trajectories and yellow circles represent fixes in the trajectory. Horizontal parallelograms with the contextual data and its pixels with different values. A) Contextual data from only one source. B) Contextual data from multiple sources, the time interval between datasets is lower. C) Contextual data from multiple sources and disaggregated. The time interval between datasets is lower and the data sets with finer spatial resolution are used to add details to the coarser datasets. The green grid represents the new pixels of the disaggregated products.
4.2 A) The range of the maned wolf (Chrysocyon brachyurus) in South America. B) Borders of Serra da Canastra National Park (CNP) in Minas Gerais state in Brazil, home of the wolves whose tracking data are used in this study. C) Lobinha "Pup", a female maned wolf of approximately two years old, wearing a GPS tracking collar. Lobinha was rescued from a sugar cane crop by the Chico Mendes Institute for Biodiversity Conservation - Brazilian Ministry of the Environment (ICMBio) and taken to an enclosure, where this picture was taken, for re-adaptation prior to reintroduction to the wilderness.
4.3 The overview of our framework for producing a daily detailed NDVI time series (ENDVI) and using it as contextual data to perform CAMA (Context-Aware Movement Analysis) on maned wolves' trajectories to identify seasonal patterns in their diet. Blue ellipses show inputs, grey rectangles show processing steps, yellow rectangles are secondary outputs and green rectangles show primary outputs, i.e., the final products.
4.4 Home ranges (95\% utilization distribution) for each individual and Canastra National Park (CNP) limits overlayed on top of land use classes. Home ranges of the two individuals in each couple (Table 4.1) intersect to a large extent. The land use map was produced by de Paula (2016) based on automatic and supervised multi-temporal classification (2009-2011) of RapidEye images with 5 m spatial resolution.
4.5 The trade-off between the spatial and the temporal resolutions. Panel A) shows a time-line covering the GPS tracking period, where horizontal dashes indicate availability of satellite images after removing the ones covered by clouds. The type of satellite image is specified by the colour of the bar shown in the legend. MODIS images are plotted on the right time-line, whilst the others are plotted on the left time-line. Panel B) shows scaled pixel sizes overlaying an image from a portion of the study area, highlighting how the heterogeneity of environmental conditions might be camouflaged by the spatial resolution of MODIS. A MODIS pixel covers $250000 \mathrm{~m}^{2}$, a Landsat pixel covers $900 \mathrm{~m}^{2}$, a CBERS pixel covers $400 \mathrm{~m}^{2}$ and an ASTER pixel covers $225 \mathrm{~m}^{2}$.
4.6 The filtered NDVI growth rate extracted from each pixel from MODIS data is combined with land cover data from the finer spatial resolution images to produce a disaggregated NDVI (ENDVI), which has the temporal resolution of MODIS and the spatial details from the finer images. The land cover is used to find the NDVI growth rate for each pixel at the finer images. Growth rates are then applied to the available finer NDVI images to create the temporal series of ENDVI.
4.7 The spatially neighbouring pixels of a target MODIS pixel being downscaled and land cover composition in a MODIS pixel classified from available LANDSAT, ASTER or CBERS images (adapted from Rao et al. (2015)). The neighbourhood shape and size varies for each MODIS pixel being downscaled because it is determined by the number of land cover classes within the targeted MODIS pixel. 87
4.8 Changes in the shape of the probability density function (PDF) according to different ranges of skew and kurtosis. All distributions are plotted within the same x and y limits. The reference range for normality is between -2 and 2 (George and Mallery, 2010).
4.9 Types of probability density function (PDF) according to different ranges of skew and kurtosis. The first character in the name of the distribution refers to the skewness, it can be "L" for low (skew $>2$ ), "H" for high (skew $<-2$ ) or "N" for "normal" $(-2 \leq$ skew $\geq 2)$. The character after the dash refers to the kurtosis of the distribution, it can be "P" for "peaked" (kurtosis $>2$ ), " F " for "flattened" (kurtosis $<2$ ) or absent for when it is similar to the normal distribution $(-2 \leq$ kurtosis $\geq 2$ ). The second and third character are always "NH" and stand for "NDVI in the home range".
4.10 Transformation from $B$ to $B^{\prime}$. The plot on the left, $B(x, y)$, corresponds to the wolves behaviour over the course of 365 days for four states. The plot on the right, $B^{\prime}(x, y)$, represents the same data in the form of a binary matrix of 365 days by 1460 (which is 365 multiplied by the four possible states $(n)$.
4.11 Average ENDVI time series from three pixels from different coffee crops, the time axis covers the same period for which we had a verified land cover map. The crop seasons are shown by background colours explained in the legend.
4.12 Frequency of different values of skew and kurtosis for the ENDVI distributions within the annotated home ranges. The total number of ENDVI distributions is 67631
4.13 Types of probability density function (PDF) according to different ranges of skew and kurtosis that were actually observed in the annotated home ranges. The first character in the name of the distribution refers to the skewness, it can be "L" for low (skew > 2), "H" for high (skew $<-2$ ) or " N " for "normal" ( $-2 \leq$ skew $\geq 2$ ). The character after the dash refers to the kurtosis of the distribution, it can be "P" for "peaked" (kurtosis $>2$ ), " F " for "flattened" (kurtosis $<2$ ) or absent for when it is similar to the normal distribution ( $-2 \leq$ kurtosis $\geq 2$ ). The second and third character are always "NH" and stand for "NDVI in the home range"
4.14 The top panel shows the temporal series of Z-scores of the locations used by Bolt in relation to MODIS NDVI distributions within the home range ( $\mathrm{n}=$ 2569 pixels of 250 m ). The colours in the Z-score plots indicate the type of NDVI or ENDVI distribution (Figure 4.9) for the day. The middle panel shows the temporal series of Z-scores of the locations used by Bolt in relation to the ENDVI distributions within the home range ( $\mathrm{n}=599292$ pixels of 15 m ). Red dashed lines indicate the $10 \%$ confidence interval, points above that show that the animal was selecting areas amongst the $10 \%$ more vegetated in the home range. The bottom panel shows the number of fixes collected each day in the period and the average number of MODIS images available at each day within a 5 days moving window. The beige background indicates when the data quality criteria were fully met, i.e., more than 10 fixes and more than 0.2 MODIS images in a 5 days window.
4.15 The top panel shows the temporal series of Z-scores of the locations used by Jurema in relation to MODIS NDVI distributions within the home range ( $\mathrm{n}=$ 881 pixels of 250 m ). The middle panel shows the temporal series of Z-scores of the locations used by Jurema in relation to the ENDVI distributions within the home range ( $\mathrm{n}=204670$ pixels of 15 m ). Red dashed lines indicate the $10 \%$ confidence interval, points above that show that the animal was selecting areas amongst the $10 \%$ more vegetated in the home range. The bottom panel shows the number of fixes collected each day in the period and the average number of MODIS images available at each day within a 5 days moving window. The beige background indicates when the data quality criteria were fully met, i.e., more than 10 fixes and more than 0.2 MODIS images in a 5 days window
4.16 The top panel shows the temporal series of Z-scores of the locations used by Lais in relation to MODIS NDVI distributions within the home range ( $\mathrm{n}=1926$ pixels of 250 m ). The colours in the Z-score plots indicate the type of NDVI or ENDVI distribution (Figure 4.9) for the day. The middle panel shows the temporal series of Z-scores of the locations used by Lais in relation to the ENDVI distributions within the home range ( $\mathrm{n}=450967$ pixels of 15 m ). Red dashed lines indicate the $10 \%$ confidence interval, points above that show that the animals was selecting areas amongst the $10 \%$ more vegetated in the home range. The bottom panel shows the number of fixes collected each day in the period and the average number of MODIS images available at each day within a 5 days moving window. The beige background indicates when the data quality criteria were fully met, i.e., more than 10 fixes and more than 0.2 MODIS images in a 5 days window.
4.17 The top panel shows the temporal series of Z-scores of the locations used by Loba in relation to MODIS NDVI distributions within the home range ( $\mathrm{n}=$ 1732 pixels of 250 m ). The colours in the Z-score plots indicate the type of NDVI or ENDVI distribution (Figure 4.9) for the day. The middle panel shows the temporal series of Z-scores of the locations used by Loba in relation to the ENDVI distributions within the home range ( $\mathrm{n}=404485$ pixels of 15 m ). Red dashed lines indicate the $10 \%$ confidence interval, points above that show that the animals was selecting areas amongst the $10 \%$ more vegetated in the home range. The bottom panel shows the number of fixes collected each day in the period and the average number of MODIS images available at each day within a 5 days moving window. The beige background indicates when the data quality criteria were fully met, i.e., more than 10 fixes and more than 0.2 MODIS images in a 5 days window.
4.18 The top panel shows the temporal series of Z-scores of the locations used by Luna in relation to MODIS NDVI distributions within the home range ( $\mathrm{n}=$ 550 pixels of 250 m ). The colours in the Z-score plots indicate the type of NDVI or ENDVI distribution (Figure 4.9) for the day. The middle panel shows the temporal series of Z-scores of the locations used by Luna in relation to the ENDVI distributions within the home range ( $\mathrm{n}=128873$ pixels of 15 m ). Red dashed lines indicate the $10 \%$ confidence interval, points above that show that the animals was selecting areas amongst the $10 \%$ more vegetated in the home range. The bottom panel shows the number of fixes collected each day in the period and the average number of MODIS images available at each day within a 5 days moving window. The beige background indicates when the data quality criteria were fully met, i.e., more than 10 fixes and more than 0.2 MODIS images in a 5 days window.
4.19 The top panel shows the temporal series of Z-scores of the locations used by Rose in relation to MODIS NDVI distributions within the home range ( $\mathrm{n}=$ 1655 pixels of 250 m ). The colours in the Z-score plots indicate the type of NDVI or ENDVI distribution (Figure 4.9) for the day. The middle panel shows the temporal series of Z-scores of the locations used by Rose in relation to the ENDVI distributions within the home range ( $\mathrm{n}=387349$ pixels of 15 m ). Red dashed lines indicate the $10 \%$ confidence interval, points above that show that the animals was selecting areas amongst the $10 \%$ more vegetated in the home range. The bottom panel shows the number of fixes collected each day in the period and the average number of MODIS images available at each day within a 5 days moving window. The beige background indicates when the data quality criteria were fully met, i.e., more than 10 fixes and more than 0.2 MODIS images in a 5 days window.
4.20 The top panel shows the temporal series of Z-scores of the locations used by Tay in relation to MODIS NDVI distributions within the home range ( $\mathrm{n}=584$ pixels of 250 m ). The colours in the Z-score plots indicate the type of NDVI or ENDVI distribution (Figure 4.9) for the day. The middle panel shows the temporal series of Z-scores of the locations used by Tay in relation to the ENDVI distributions within the home range ( $\mathrm{n}=138140$ pixels of 15 m ). Red dashed lines indicate the $10 \%$ confidence interval, points above that show that the animals was selecting areas amongst the $10 \%$ more vegetated in the home range. The bottom panel shows the number of fixes collected each day in the period and the average number of MODIS images available at each day within a 5 days moving window. The beige background indicates when the data quality criteria were fully met, i.e., more than 10 fixes and more than 0.2 MODIS images in a 5 days window.
4.21 The real tracking period for each wolf is shown in the top panel and the translation of these data into wolf years is shown in the bottom panel. Each wolf is represented by a different colour and white gaps represent missing data that was accounted for as no data.
4.22 Eigenvalues ranks and percentage of variance explained by each. The total number of ranks was 1460 , however the graph plateaus after the $6^{\text {th }}$ rank
4.23 The top six eigenbehaviours for all the wolves, each row represents one of the states defined previously based on the ENDVI Z-scores (Figure 4.10). The row for the no data $(\mathrm{N})$ state is omitted because it does not provide any information on the behaviour or on relevant data quality for this part of the analysis. The letters in the horizontal axis indicate the seasons, D for dry and W for Wet. The season axis is magnified at the bottom of the picture where it is matched by information on the key biological periods for the species.
4.24 Sequences of states summarising the five first eigenbehaviours. Sequences were generated by selecting the state with the highest eigenvector coefficient at each time during the wolf year for each eigenvector. The letters in the horizontal axis indicate the seasons, D for dry and W for Wet.
5.1 The overview of our framework for identification of groups of similar movement behaviour under specific weather conditions. The framework has two analyses running in parallel: analysis of places and analysis of travel modes. Blue shapes marks travel mode, green shapes marks places, white ellipses represent dataset's sources, rectangles represent variables, beige arrows represent processing steps and hexagons derived results in each step.
5.2 a) East coast of Scotland where the GPS data were collected, and trajectories represented by light blue lines. The green and the red ellipses represent the locations of the Tay and Forth road bridges respectively. b) Sample of two movement travel modes overlaid by the Thiessen polygons used to interpolate MIDAS (Met Office Integrated Data Archive System) data and one pixel of a NIMROD (Met Office's nowcasting system) product for comparison. The frame in the right upper corner illustrates a trajectory sample classified into movement modes and displayed in a space-time cube with rainfall data for a one-hour period. 124
5.3 A multi-channel sequence for a participant over a five-minute period, each channel relates to one of the meteorological variables and movement modes for that minute of the day.
5.4 Sequences of movement modes for each day of the week by participant (vertical axis) and by hour of the day (horizontal axis), missing data are reported in white. For each day, the sequences are ordered by length, but orderings are different for each day (that is, the sequence number on the y chart does not always indicate the same individual).
5.5 Entropy index (EI) by day of the week and by hour of the day (horizontal axis), an EI close to zero indicates that most people are in the same movement mode at that time (lower diversity), while an EI close to one indicates similar proportion of people on the different movement modes (higher diversity).
5.6 Average time expenditure and trend over the week (dashed lines) by movement mode, day of the week and gender. The first row refers to females $(22 \leq \mathrm{n} \leq$ $33)$ and the second row to males $(54 \leq \mathrm{n} \leq 70)$.
5.7 Calinski-Harabaz index (solid line) and optimal number of clusters (dashed line) for MCSA performed on weather variables for travel modes (green) and places (red) on weekdays and on weekend. The reported CHI is divide by ten and refers to the number of clusters used to split the sequences.
5.8 Clusters for MCSA on wind speed, wind direction and travel modes on weekdays (A) and weekends (B). The four top panels describe the wind conditions within each cluster (Types) and the respective proportions of GPS active time spent under the classes of wind speed and direction. The two panels at the bottom show box-plots with the distribution of proportional GPS active time spent on each travel mode by the wind types described on the panels above. The dashed line on box-plots show the average and the continuous line the median. L reports significance from Levene's test and K from Kruskal-Wallis' test.
5.9 Clusters for MCSA on wind speed, wind direction and places on weekdays (A) and weekends (B). The four top panels describe wind conditions within each cluster (Types) and respective proportions of GPS active time spent under the classes of wind speed and direction. The two panels at the bottom show boxplots with the distribution of proportional GPS active time spent on places by the wind types described on the panels above. The dashed line on box-plots show the average and the continuous line the median. L reports significance from Levene's test and K from Kruskal-Wallis' test. . . . . . . . . . . . . . . . .
5.10 Clusters for MCSA on rain and travel modes on weekdays (A) and weekends (B). The two top panels describe the rain conditions within each cluster (Types) and the respective proportions of GPS active time spent under the rainfall classes. The two panels at the bottom show box-plots with the distribution of proportional GPS active time spent on each travel mode by the rain type described on the panel above. The dashed line on box-plots show the average and the continuous line the median. L reports significance from Levene's test and K from Kruskal-Wallis' test.
5.11 Clusters for MCSA on rain and places on weekdays (A) and weekends (B). The two top panels describe the rain conditions within each cluster (Types) and the respective proportions of GPS active time spent under the rainfall classes. The two panels at the bottom show box-plots with the distribution of proportional GPS active time spent on places by the rain type described on the panel above. The dashed line on box-plots show the average and the continuous line the median. L reports significance from Levene's test and K from Kruskal-Wallis'
test.
5.12 Clusters for MCSA on daylight and travel modes on weekdays (A) and weekends (B). The two top panels describe daylight conditions within each cluster (Types) and respective proportions of GPS active time spent under the classes of daylight. The two panels at the bottom show box-plots with the distribution of proportional GPS active time spent on each travel mode by the daylight type described above. The dashed line on box-plots show the average and the continuous line the median. L reports significance from Levene's test and K from Kruskal-Wallis' test.
5.13 Clusters for MCSA on daylight and places on weekdays (A) and weekends (B). The two top panels describe daylight conditions within each cluster (Types) and respective proportions of GPS active time spent under daylight classes. The two panels at the bottom show box-plots with the distribution of proportional GPS active time spent on places by the daylight type described above. The dashed line on box-plots show the average and the continuous line the median. L reports significance from Levene's test and K from Kruskal-Wallis' test.
5.14 Clusters for MCSA run on comfort and travel modes on weekdays (A) and weekends (B). The two top panels describe comfort conditions within each cluster (Types) and respective proportions of GPS active time spent under comfort classes. The two panels at the bottom show the distribution of proportional GPS active time spent on each travel mode by the comfort type described above. The dashed line on box-plots show the average and the continuous line the median. L reports significance from Levene's test and K from Kruskal-Wallis' test.
5.15 Clusters for MCSA run on comfort and places on weekdays (A) and weekends (B). The two top panels describe comfort conditions within each cluster (Types) and respective proportions of GPS active time spent under comfort classes. The two panels at the bottom show box-plots with the distribution of proportional GPS active time spent on places by comfort type described above. The dashed line on box-plots show the average and the continuous line the median. L reports significance from Levene's test and K from Kruskal-Wallis' test.
5.16 Clusters for MCSA run on relative humidity and travel modes on weekdays (A) and weekends (B). The two top panels describe relative humidity conditions within each cluster (Types) and respective proportions of GPS active time spent under relative humidity classes. The two panels at the bottom show box-plots with the distribution of proportional GPS active time spent on each travel mode by relative humidity type described above. The dashed line on box-plots show the average and the continuous line the median. L reports significance from Levene's test and K from Kruskal-Wallis' test
5.17 Clusters for MCSA run on relative humidity and places on weekdays (A) and weekends (B). The two top panels describe relative humidity conditions within each cluster (Types) and respective proportions of GPS active time spent under relative humidity classes. The two panels at the bottom show box-plots with the distribution of proportional GPS active time spent on places by relative humidity type described above. The dashed line on box-plots show the average and the continuous line the median. L reports significance from Levene's test and K from Kruskal-Wallis' test.
5.18 Clusters for MCSA run on temperature and travel modes on weekdays (A) and weekends (B). The two top panels describe temperature conditions within each cluster (Types) and respective proportions of GPS active time spent under temperature classes. The two panels at the bottom show box-plots with the distribution of proportional GPS active time spent on each travel mode by temperature type described above. The dashed line on box-plots show the average and the continuous line the median. L reports significance from Levene's test and K from Kruskal-Wallis' test.
5.19 Clusters for MCSA run on temperature and travel modes on weekdays (A) and weekends (B). The two top panels describe temperature conditions within each cluster (Types) and respective proportions of GPS active time spent under temperature classes. The two panels at the bottom show box-plots with the distribution of proportional GPS active time spent on places by temperature type described above. The dashed line on box-plots show the average and the continuous line the median. L reports significance from Levene's test and K from Kruskal-Wallis' test.
B. 1 The top panel shows the temporal series of Z-scores of the locations used by Amadeo in relation to MODIS NDVI distributions within the home range ( $\mathrm{n}=$ 2162 pixels of 250 m ). The colours in the Z- score plots indicate the type of NDVI or ENDVI distribution (Figure 4.9) for the day. The middle panel shows the temporal series of Z-scores of the locations used by Amadeo in relation to the ENDVI distributions within the home range ( $\mathrm{n}=506822$ pixels of 15 m ). Red dashed lines indicate the $10 \%$ confidence interval, points above that show that the animals was selecting areas amongst the $10 \%$ more vegetated in the home range. The bottom panel shows the number of fixes collected each day in the period and the average number of MODIS images available at each day within a 5 days moving window. The beige background indicates when the data quality criteria were fully met, i.e., more than 10 fixes and more than 0.2 MODIS images in a 5 days window.
B. 2 The top panel shows the temporal series of Z-scores of the locations used by Gamba in relation to MODIS NDVI distributions within the home range ( $\mathrm{n}=$ 850 pixels of 250 m ). The colours in the Z- score plots indicate the type of NDVI or ENDVI distribution (Figure 4.9) for the day. The middle panel shows the temporal series of Z-scores of the locations used by Gamba in relation to the ENDVI distributions within the home range ( $\mathrm{n}=199398$ pixels of 15 m ). Red dashed lines indicate the $10 \%$ confidence interval, points above that show that the animals was selecting areas amongst the $10 \%$ more vegetated in the home range. The bottom panel shows the number of fixes collected each day in the period and the average number of MODIS images available at each day within a 5 days moving window. The beige background indicates when the data quality criteria were fully met, i.e., more than 10 fixes and more than 0.2 MODIS images in a 5 days window.
B. 3 The top panel shows the temporal series of Z-scores of the locations used by Henry in relation to MODIS NDVI distributions within the home range ( $\mathrm{n}=$ 765 pixels of 250 m ). The colours in the Z- score plots indicate the type of NDVI or ENDVI distribution (Figure 4.9) for the day. The middle panel shows the temporal series of Z-scores of the locations used by Henry in relation to the ENDVI distributions within the home range ( $\mathrm{n}=177976$ pixels of 15 m ). Red dashed lines indicate the $10 \%$ confidence interval, points above that show that the animals was selecting areas amongst the $10 \%$ more vegetated in the home range. The bottom panel shows the number of fixes collected each day in the period and the average number of MODIS images available at each day within a 5 days moving window. The beige background indicates when the data quality criteria were fully met, i.e., more than 10 fixes and more than 0.2 MODIS images in a 5 days window.
B. 4 The top panel shows the temporal series of Z-scores of the locations used by Miro in relation to MODIS NDVI distributions within the home range ( $\mathrm{n}=$ 764 pixels of 250 m ). The colours in the Z- score plots indicate the type of NDVI or ENDVI distribution (Figure 4.9) for the day. The middle panel shows the temporal series of Z-scores of the locations used by Miro in relation to the ENDVI distributions within the home range ( $\mathrm{n}=179254$ pixels of 15 m ). Red dashed lines indicate the $10 \%$ confidence interval, points above that show that the animals was selecting areas amongst the $10 \%$ more vegetated in the home range. The bottom panel shows the number of fixes collected each day in the period and the average number of MODIS images available at each day within a 5 days moving window. The beige background indicates when the data quality criteria were fully met, i.e., more than 10 fixes and more than 0.2 MODIS images in a 5 days window.
B. 5 The top panel shows the temporal series of Z-scores of the locations used by Nilde in relation to MODIS NDVI distributions within the home range ( $\mathrm{n}=$ 764 pixels of 250 m ). The colours in the Z- score plots indicate the type of NDVI or ENDVI distribution (Figure 4.9) for the day. The middle panel shows the temporal series of Z-scores of the locations used by Nilde in relation to the ENDVI distributions within the home range ( $\mathrm{n}=174347$ pixels of 15 m ). Red dashed lines indicate the $10 \%$ confidence interval, points above that show that the animals was selecting areas amongst the $10 \%$ more vegetated in the home range. The bottom panel shows the number of fixes collected each day in the period and the average number of MODIS images available at each day within a 5 days moving window. The beige background indicates when the data quality criteria were fully met, i.e., more than 10 fixes and more than 0.2 MODIS images in a 5 days window.
B. 6 The top panel shows the temporal series of Z-scores of the locations used by Samurai in relation to MODIS NDVI distributions within the home range ( $\mathrm{n}=$ 761 pixels of 250 m ). The colours in the Z- score plots indicate the type of NDVI or ENDVI distribution (Figure 4.9) for the day. The middle panel shows the temporal series of Z-scores of the locations used by Samurai in relation to the ENDVI distributions within the home range ( $\mathrm{n}=177753$ pixels of 15 m ). Red dashed lines indicate the $10 \%$ confidence interval, points above that show that the animals was selecting areas amongst the $10 \%$ more vegetated in the home range. The bottom panel shows the number of fixes collected each day in the period and the average number of MODIS images available at each day within a 5 days moving window. The beige background indicates when the data quality criteria were fully met, i.e., more than 10 fixes and more than 0.2 MODIS images in a 5 days window.
C. 1 Distance covered by day vs five primary eigenbehaviours for wolves where there was a match between the decrease in distance covered per day and a higher contribution of the low ENDVI state. Red circles indicate where the match happens. Green rectangle indicates the only couple for which we had simultaneous data, and in which the red circle highlights the increased activity of the male. .

## List of Tables

2.1 Synonymous terminologies used in movement research, building on previous work from Long and Nelson (2013) and Sila-Nowicka (2016)

2.2 A non-exhaustive list of studies using environmental contextual data in CAMA
with reference and specific variable used. ..... 22
2.3 Samples of movement research papers using contextual data with reference, de- scription of movement data and contextual data. MET stands for meteorological stations; RS for remote sensing and GAR for global atmospheric reanalysis mod- els. Dashes indicate that the information was not available on the manuscript. The spatial resolution was converted to km at $0^{\circ}$ latitude where it was originally in degrees to facilitate comparisons. Spatial and temporal resolutions listed are always the best ones reported for irregularly sampled data. ..... 33
3.1 Equations and abbreviations for the five Interpolation methods tested. ..... 47
3.2 Summary of characteristics of datasets used for trajectory annotation. ..... 50
3.3 Descriptive statistics and p-values for Kolmogorov-Smirnov normality test for all methods. All p-values were smaller than 0.01 . ..... 55
3.4 Z-scores reported by two tailed Wilcoxon rank-sum test between interpolationmethods. The multiple testing problem was addressed using Bonferroni correc-tion where $\alpha^{\prime}=\alpha / 4$. Significant values are indicated by a star, where p -value $<\alpha^{\prime}$( $\alpha=0.05$ ).56
$3.5 \chi$ values for non-parametric ANOVA tests. All p-values were smaller than 0.01. ..... 563.6 W statistic for Wilcoxon rank-sum test performed between annotated NIMRODdata from different spatial resolution, the value indicate that the difference in-creases for all methods with degradation of the spatial resolution of the anno-tated data. All p-values were smaller than 0.01 .
3.7 Computational metrics for our algorithms. We present the estimated computational cost in hours for DTA:CS (based on records updated $/ \mathrm{min}$ ) for our data set (140.000.000 GPS points) and real computational cost in hours for other methods. We further note the computational complexity of each algorithm, where u is the number of trajectories and n is the number of points in the longest trajectory, and description of the computational resources used.
4.1 Description of tracked individuals, respective sampling rate, duration, start and end of monitoring period. Names with the same colour indicates a couple (maned wolves take only one mate for life). White indicates the absence of a mate in the tracking dataset. Source: Adapted from de Paula (2016).
4.2 Description of satellite images used to produce NDVI. The letters along with spa-
tial resolutions are specifying the spectral bands for sensors with multiples spa-
tial resolutions, sensors without letters have a uniform spatial resolution along
the spectral bands. R stands for red, G for green, B for blue, NIR for near infra-
red, VNIR for visible (RGB) and near infra-red, SWIR for short wavelength
infra-red, FI for far infra-red and TIR for thermal infra-red. . . . . . . . . . 73
5.1 Summary of contextual datasets with respective sources and specifications. . . 125
5.2 Alphabets for meteorological variables used as contextual data with respective ranges and description. Letters in each alphabet are defined based on standard meteorological classifications (see text for more details).

## Contents

Abstract ..... ii
Acknowledgments ..... v
Funding ..... v
Declaration ..... vii
Permission for electronic publication ..... ix
List of Figures ..... xxvii
List of Tables ..... xxx
1 Introduction ..... 1
1.1 Motivation ..... 1
1.2 Research questions ..... 3
1.3 Structure of the thesis ..... 4
2 Context-Aware Movement Analysis ..... 9
2.1 Introduction ..... 9
2.2 The terminology of movement research ..... 10
2.2.1 A taxonomy for context ..... 13
2.3 The role played by context ..... 18
2.3.1 Movement data types ..... 19
2.3.2 Contextual data types ..... 20
2.3.2.1 Contextual data from remote sensing ..... 21
2.3.2.2 Contextual data from meteorological stations ..... 26
2.4 Integration of context ..... 30
2.5 Current challenges ..... 32
2.5.1 The source of the challenges: mismatch in data resolutions ..... 32
2.5.2 CAMA challenges ..... 35
2.5.2.1 Interpolation within TA and the temporal profile of contextual data ..... 36
2.5.2.2 Spatial structure of contextual data ..... 37
2.5.2.3 How to analyse semantic trajectories ..... 37
2.5.3 Moving forward ..... 38
3 Comparing trajectory annotation methods ..... 41
3.1 Introduction ..... 41
3.2 Methods ..... 44
3.2.1 Dynamic Trajectory Annotation (DTA) ..... 44
3.2.2 Most commonly used trajectory annotation methods ..... 47
3.2.3 Data description ..... 48
3.2.4 Data preprocessing ..... 49
3.2.5 Generating and comparing rainfall rates and cumulative curves ..... 51
3.2.6 Validation against ground truth and statistical comparison of interpola- tion methods ..... 51
3.3 Results ..... 52
3.3.1 Generating and comparing rainfall rates and cumulative curves ..... 52
3.3.2 Validation against ground truth and statistical comparison of TA methods ..... 53
3.3.3 Computational cost of trajectory annotation methods ..... 56
3.4 Discussion ..... 58
4 Seasonality in the diet of maned wolves ..... 63
4.1 Introduction ..... 63
4.1.1 Feeding ecology of maned wolves ..... 66
4.2 Methodology ..... 70
4.2.1 Movement data and study area ..... 71
4.2.2 Contextual data ..... 72
4.2.2.1 Absolute calibration of satellite data ..... 74
4.2.2.2 Land cover classification and calculation of land cover fractions ..... 75
4.2.3 MODIS NDVI growth rate extraction and filtering ..... 76
4.2.4 Calculating NDVI growth rates for land cover fractions to generate the finer NDVI time-series (ENDVI) ..... 78
4.2.5 Context integration and seasonal plots ..... 81
4.2.6 Eigenbehaviour analysis: Identifying seasonal structure in habitat use ..... 90
4.3 Results ..... 93
4.3.1 Calculating NDVI growth rates for land cover fractions to generate finer NDVI time-series (ENDVI) ..... 93
4.3.2 Context integration and seasonal plots ..... 94
4.3.3 Eigenbehaviours: Identifying seasonal structure in habitat use ..... 106
4.4 Discussion ..... 112
5 Weather effects on human mobility ..... 117
5.1 Introduction ..... 117
5.2 Methodology ..... 121
5.2.1 Movement data ..... 122
5.2.2 Contextual data and context integration ..... 124
5.2.3 Trajectory sequencing ..... 126
5.2.4 Context-Aware Similarity Analysis (CASA) ..... 130
5.2.4.1 Multi-channel Sequence Analysis (MCSA) ..... 130
5.2.4.2 Cluster analysis and typology ..... 132
5.3 Results ..... 133
5.3.1 Trajectory sequencing ..... 133
5.3.2 Context-Aware Similarity Analysis (CASA) ..... 138
5.3.2.1 Cluster analysis and typology ..... 138
5.3.2.2 Wind ..... 138
5.3.2.3 Rain ..... 141
5.3.2.4 Daylight ..... 144
5.3.2.5 Comfort ..... 146
5.3.2.6 Humidity ..... 148
5.3.2.7 Temperature ..... 152
5.4 Discussion ..... 152
6 Conclusions ..... 159
6.1 Key findings ..... 160
6.2 Limitations ..... 165
6.3 Future research ..... 166
A Appendices list ..... 169
B Results for the remaining wolves - Chapter 4 ..... 171
C Distance covered by wolf vs eigenbehaviour ..... 181

## Chapter 1

## Introduction

### 1.1 Motivation

Increased availability of remotely sensed environmental data products and the rapid miniaturization and reduced costs of tracking devices have led to exponential growth in movement research (Cagnacci et al., 2010; Kays et al., 2015; Long and Nelson, 2013; Wikelski et al., 2007). Similar trends have occurred in how environmental data are being collected, but the development of analytical techniques has been outpaced by technological advancements (Gudmundsson and Wolle, 2014; Purves et al., 2014). For this reason, new methods capable of performing context aware movement analysis (CAMA) are required to integrate increasingly detailed movement data with a myriad of contextual environmental data.

Context aware movement analysis incorporates not only movement geometry, but also biological and environmental conditions that might be affecting movement (Ahearn et al., 2016; Andrienko et al., 2011; Das and Winter, 2016; Demšar et al., 2015; Dodge et al., 2013). In most studies, individual movement is represented in the form of a trajectory, i.e., the path taken by a single object moving in space over time. Trajectories are captured as movement data by a series of chronologically ordered fixes, often GPS (Global Positioning System) points, which are linked spatial coordinates and time-stamps (Hornsby and Egenhofer, 2002; Lee and Krumm, 2011). Contextual analysis then refers to methods used to integrate each trajectory with context, i.e., with any other type of data that can help to characterize the situation a
moving entity is responding to (Abowd et al., 1999). In movement ecology and human mobility studies, such data often include environmental variables that are retrieved from meteorological stations, radars and biologgers (Demšar et al., 2015).

The study of movement behaviour has enhanced our knowledge of individual and population dynamics in ecology (Schick et al., 2008) and provided insights into the complex humanenvironment interactions associated with, for example, commuting (Beecham et al., 2014; Gong et al., 2012), tourist behaviour (Meijles et al., 2014; Versichele et al., 2012), and retail choice decisions (Thakuriah et al., 2016). Nonetheless, simultaneously analysing individual movement alongside external and internal contextual factors is expected to lead to new discoveries about how individuals move and interact with the environment. For example, changes in environmental conditions, such as wind, temperature and precipitation may trigger different movement behaviours, which are reflected as movement patterns.

Despite the potential of CAMA for advancing the understanding of movement behaviour, the overwhelming majority of existing movement research over the past two decades have been geometry-focused (Laube and Purves, 2011). As a consequence, the developments in geometrybased movement analysis are much more advanced than current methods for performing CAMA (Laube et al., 2007; Long and Nelson, 2013; Purves et al., 2014). Another issue is that there have been very few studies that have identified the key challenges in conducting CAMA, such as the large and complex datasets, the differences in temporal and spatial resolutions of movement data and contextual data, how to operationally link these datasets, and finally what methods are most appropriate for extracting meaningful inferences on movement behaviour from linked data.

The shortage of methods relating movement patterns to environmental contextual factors is still in many ways limiting our ability to extract meaningful information from movement data. Further, given the rapid increase in availability of remotely sensed data products describing environmental conditions, there are significant opportunities to enhance our analysis of individual movement by linking to these data. However, these new technologies come with an increasing need for tools capable of looking simultaneously at movement data and data on the conditions surrounding movement, i.e., geographical data that describe environmental
conditions that may be influencing individual movement decisions. Therefore, there is a need to invent and design novel and informative methods that facilitate CAMA, and identify how those methods can be translated into new insights into movement behaviour across the wide range of disciplines interested in movement. This thesis contributes to this methodological research gap.

### 1.2 Research questions

The goal of this thesis is to contribute towards advancing the field of CAMA by addressing the following three research questions

1. How can CAMA methods properly account for the temporal dynamics of contextual data (e.g., contextual factors that change over time)?

In terms of temporal dynamics, wind and temperature are examples of contextual variables with varied temporal cycles, yet they are often processed using the same interpolation methods, i.e., the interpolation methods are chosen disregarding the characteristics and scale of each variable. Therefore, in order to better deal with the temporal incompatibilities in trajectory annotation (TA), there is a need to compare the current interpolation methods and their implications, and also search for methods that take into account the temporal progression of contextual variables in a way more suited to natural progression of contextual phenomena.
2. How can CAMA methods better address challenges associated with data structures of contextual data (e.g., issues posed by different spatial and/or temporal resolutions, data representations, etc.)?

In addition to being collected for different purposes and by diverse means, movement data are collected point-wise whilst context is collected in a variety of forms, from raster to point and area data. Current methods disregard the structure of contextual data and often interpolate at very coarse level in an attempt to deal with the spatial incompatibilities between movement data and contextual data. There is a need to find new and
better ways to deal with the spatial mismatch between the resolutions of movement data and contextual data.
3. How can we make meaningful inferences about behaviour from contextualized movement data using modern computational methods?

In order to identify movement patterns that can be linked to behaviour, it is essential to examine similarities in the data from the contextual perspective. While spatio-temporal similarity has been thoroughly explored, contextual similarity has remained largely neglected. Context awareness is still a new trend and there are only a few approaches of using analytical methods to better understand movement behaviour from semantic trajectories, i.e., movement trajectories that have been linked to contextual data. Therefore, there is a need to find new ways to make meaningful inferences about behaviour from contextualized movement data within CAMA.

Given the current lack of analytical tools and theoretical background on CAMA, one of the primary goals of this thesis is to investigate the challenges posed by CAMA and to develop methods capable of overcoming these challenges. Three research objectives are used to separate this task into key contributions: 1) assess the state-of-the art for CAMA within movement ecology and human mobility research; 2) develop innovative methods to take into account the spatio-temporal differences between movement data and contextual data; and 3) explore computational methods that allow identification of patterns in contextualized movement data.

### 1.3 Structure of the thesis

This thesis contains six chapters (including this one) which attempt to assess and understand the spatial behaviour of people and wildlife using GPS movement data and environmental data through the development of frameworks for performing CAMA. The core of the thesis consists of three empirical chapters (Chapters 3, 4 and 5) where movement datasets and contextual data are analysed through aiming to better understand the movement behaviour of both people and wildlife. The structure of the thesis is summarised in Figure 1.1.

## Chapter 2: Context-Aware Movement Analysis

In this chapter we examine the state-of-the-art in the field of CAMA for human mobility and movement ecology studies. We introduce the terminology of movement research, analyse previous taxonomies for movement context and develop our own. We also discuss the role played by context in movement, as well as the data types used to represent movement and context. We focus particularly on movement data from GPS devices and contextual data from remote sensing and meteorological stations. In addition, we look at how contextual datasets and movement datasets are currently integrated and what are the challenges to overcome in this area. We conclude by indicating how this thesis will try to address these challenges.

## Chapter 3: Comparing trajectory annotation methods

In this chapter we look at how CAMA considers the context within which movement occurs by using trajectory annotation (TA) to associate environmental and other contextual data with trajectories. TA depends on spatial and temporal interpolation methods to estimate nonexistent data values. The diversity of temporal and spatial scales and resolutions of environmental data result in uncertainties within interpolation methods and must be assessed before meaningful information can be extracted. We formalize a TA method, the dynamic trajectory annotation (DTA), which aims to address the case of temporal mismatch where trajectory fixes are collected more frequently than environmental data. We compare the performance of DTA with the following commonly used interpolation methods: NN (nearest neighbour), NB (neighbour before), NA (neighbour after) and AM (arithmetic mean). The results indicate that DTA is likely to outperform the NA and NN method, and that the DTA method produces smoother transitions, being more suitable for visualization purposes, with the burden of added computational time. Future directions will include expanding DTA for trajectories with coarser sampling intervals relative to environmental data.

Chapter 4: Seasonal response in the diet of maned wolves: A study using multi-sensor image fusion and a sequence based behavioural analysis

In this chapter we make use of geospatial data on movement and context to design new methods to investigate the interactions between maned wolves Chrysocyon brachyurus and seasonal availability of vegetation. Understanding the influence of vegetation on maned wolves is
of interest for the preservation of this near-threatened species. The effect of vegetation seasonality on movement behaviour can be explored through Context-Aware Movement Analysis (CAMA), which integrates movement geometry with its context. More specifically, we propose a new method that uses multi-source remote sensing data to overcome spatio-temporal incompatibilities between movement data and contextual data. This is the first time that CAMA for a particular variable has been attempted by combining data from several satellites. In the second part of the analysis we represent a wolf's movement as a temporal sequence of states that describe the conditions of the vegetation in the point where a GPS fix was recorded in relation to the vegetation of the entire home range for that wolf. This moves our data representation from trajectories to sequences. Movement patterns can then be identified by aligning these sequences using a new sequence analysis method, the so-called eigenbehaviours, to identify recurrent behaviour related to vegetation availability.

## Chapter 5: Weather effects on human mobility: A study using multi-channel sequence analysis

In this chapter we propose new methods to investigate the effects of weather conditions on human movement patterns. Understanding the influence of weather on human behaviour is of interest for diverse applications, such as urban planning and traffic engineering. The effect of weather on movement behaviour can be explored through Context-Aware Movement Analysis (CAMA), which integrates movement geometry with its context. More specifically, we use multi-channel sequence analysis (MSA) to represent a person's movement as a multidimensional sequence of states, describing either the type of movement or the state of the environment throughout time. In contrast with the previous chapter where we used the sequence analysis for one variable only, the MSA allows us to investigate the effects of several meteorological variables at the same time. Similar movement patterns can be identified by comparing and aligning multi-channel sequences using clustering and other data mining methods. We used trajectories from a GPS tracking study of commuting in the Scottish town of Dunfermline and linked them to weather data on wind, temperature and rainfall. We then converted these contextualised trajectories into multi-channel sequences which were clustered into groups of similar behaviours under specific weather typologies. Our findings show that
the CAMA + MSA method can successfully identify the response of commuters to variations in environmental conditions. We also discuss our findings on how travel modes and time spent at different places are affected by meteorological conditions, mainly wind, but also rainfall, daylight duration, temperature, comfort and relative humidity.

## Chapter 6: Conclusions

In the final chapter we revisit the research questions and objectives to present the main conclusions from each of the empirical chapters. We also discuss the implications of this work for future and current research in the area of CAMA, particularly regarding study designs when collecting movement data in both human and wildlife research. Finally, we present the contributions of this thesis to developing a new understanding of the movement behaviour of humans and animals based on GPS data. We look at the limitations of this empirical work, and outline a number of directions for the future research.



## Chapter 2

## Context-Aware Movement Analysis

### 2.1 Introduction

Movement is a fundamental attribute of our world, it encompasses animate and inanimate entities in varied ways and plays a key role in many ecological and evolutionary processes (Nathan and Giuggioli, 2013). Living organisms in particular, show exceedingly frequent and diversified motion driven by a set of environmental and biological factors interacting across multiple spatial and temporal scales (Nathan et al., 2008; Dodge et al., 2013).

In the past two decades, the advances in location based technologies have resulted in greatly improved quantity and quality of movement data (Demšar et al., 2015). At the same time, there has been a significant increase on the availability of other sources of data, such as meteorological stations, satellites, radars and biologging sensors. This simultaneity brought new opportunities and challenges to movement research: we can now not only look at movement itself but also gather insights into the environment and conditions under which movement happened. In order to gain insights from the now widely available contextual data collected alongside movement data, we need first to work on the development of methodologies that accommodate the simultaneous analysis of the moving entity and its environment (Laube and Purves, 2011; Purves et al., 2014). These methodologies are termed Context-Aware Movement Analysis (CAMA), where the Context-Awareness represents the simultaneous consideration of the environmental conditions within which movement occurs (Sharif and Alesheikh, 2017a).

This chapter summarises the current developments in CAMA in the fields of human mobility and movement ecology. In Section 2.2 we define the key terms in movement research and propose a taxonomy for the different types of context. In Section 2.3 we explain the main types of contextual data and explore the importance of simultaneously looking at context and movement. In section 2.4 we explain how movement data and contextual data are currently linked. In section 2.5 we use previous studies to showcase the current challenges and potential of CAMA. We conclude this chapter by indicating how these challenges will be tackled within the scope of this thesis.

### 2.2 The terminology of movement research

Movement research is an extremely multidisciplinary field with diverse and perhaps redundant terminology (Long and Nelson, 2013). This section introduces the terminology relevant to this thesis with definitions and lists synonymous terms with references in Table 2.1.

- Movement - a continuous process defined by the change of spatial location over time (Hornsby and Egenhofer, 2002; Long and Nelson, 2013).
- Movement data - a discrete collection of spatial and temporal coordinates describing samples of the movement of one or more objects, possibly including attributes (Long and Nelson, 2013).
- Movement database - a database of trajectories of one or more moving entities possibly with attributes for all records (Long and Nelson, 2013).
- Trajectory - a collection of chronologically ordered spatio-temporal coordinates describing samples of the movement of one object in geographical space and time (Hu et al., 2013) within diverse contexts (Gao et al., 2013).
- Point - one spatio-temporal record in a trajectory (Jeung et al., 2011), representing a single sample of movement.
- Context - a set of factors, situations, events or information that may be of interest for an entity because it has an influence upon its behaviour (Bolchini et al., 2009), which makes that context core to representing and reasoning its movement (Orellana and Renso, 2011). These are factors in the surroundings of a location point that might be affecting the shape of the trajectory (Demšar et al., 2015). In Section 2.2.1 we will explore this term and its subdivisions in more detail.
- Contextual data - a collection of chronologically ordered datasets describing one or more contexts under which an entity was moving.
- Contextual layer - a single record of spatio-temporal data describing a context in which an entity was moving at a certain time.
- Semantic trajectory - a trajectory annotated with contextual data relevant to the movement analysis (Alvares et al., 2007).
- Trajectory annotation (TA) - the procedure by which spatial and temporal coordinates are used to link contextual layers to location points within a trajectory, resulting in a semantic trajectory (Dodge et al., 2013).

Table 2.1: Synonymous terminologies used in movement research, building on previous work from Long and Nelson (2013) and Sila-Nowicka (2016).

| Adopted term | Synonymous | Reference |
| :---: | :---: | :---: |
| Point | Movement fix | Long and Nelson (2013) |
|  | trajectory fix | Dodge et al. (2009) |
|  | location | Brillinger et al. (2004) |
|  | observation | Harris (2001) |
|  | record | Sila-Nowicka (2016) |
|  | fixation | Dodge et al. (2008) |
|  | anchor | Hägerstrand (1970) |
| Trajectory | Movement path | Long and Nelson (2013) |
|  | space-time path | Hägerstrand (1970) |
|  | trip-chain | Kondo and Kitamura (1987) |
|  | geospatial lifeline | Laube et al. (2007) |
|  | trace | Jiang et al. (2015) |
|  | track | Brillinger et al. (2004) |
| Movement | Mobility data | Sila-Nowicka (2016) |
| database | GPS data | Kays et al. (2015) |
| Context | Context | Sharif and Alesheikh (2017a) |
|  | influencing factors | Dodge et al. (2008) |
|  | motivation context | Bogorny et al. (2014) |
|  | movement context | Sharif and Alesheikh (2017a) |
|  | modality context | Nathan et al. (2008) |
|  | geographic context | Buchin et al. (2014) |
|  | environmental context | Demšar et al. (2015) |
|  | semantic geographic information | Alvares et al. (2007) |
|  | biological context | Martin et al. (2013) |
|  | geographical embedding | Laube (2014) |
|  | spatio-temporal context | Andrienko et al. (2011) |

### 2.2.1 A taxonomy for context

Defining context is an important step towards the development of context-aware methods for movement analysis, because it allows a better understanding of the state-of-the-art in CAMA and underpins future research steps. Perhaps due to its novelty, current literature shows considerable variety in definitions of context (Sharif and Alesheikh, 2017a), as well as many subdivisions.

In the current literature, the following definitions have been used for "context":

- "That part of a situation or data that influences movement or is influenced by movement" (Sharif and Alesheikh, 2017a, p.19).
- What enables or limits movement (Purves et al., 2014).
- "[..] the locational circumstances of a moving agent, the external factors connected to the underlying landscape [...] or the surrounding environment [...] in which movement takes place" (Buchin et al., 2014, p.102).
- "[...] the complex and heterogeneous physical space, in which characteristics vary from place to place and change over time; complex and heterogeneous physical time, in which day differs from night, summer from winter, and so on; static and dynamic objects existing in space, as well as events occurring over time" (Andrienko et al., 2011, p.1347).
- Various influencing factors that impact and constraint movement (Dodge et al., 2008).

We also found the following subdivisions of context:

- internal and external, " $[$ the former is any factor that is related to the MPO [moving-point objects], characteristics, states, and conditions, such as the intention, location, direction, and speed, while the latter is dedicated to the geographical and environmental conditions during the move[...]" (Sharif and Alesheikh, 2017a, p.427).
- the ones "[...] com[ing] from additional data collected with trajectories[,] describ[ing] the space within which movement can occur [and the ones] added by detailed knowledge, and initial hypotheses, about the process under investigation" Purves et al. (2014, p.4).
- intrinsic properties of the moving object, spatial constraints, environment against which the movement takes place and influences of other agents (Dodge et al., 2008).

Sharif and Alesheikh (2017a) was the first study to actually focus on defining and classifying context by developing a 4M taxonomy. They first divide context into internal and external, but rename the latter as milieu context and subdivide the former into motivation context, movement context and modality context, which leads to the 4 M 's defined below.

- Motivation context: " [...] any propellant, driver, or reason [...] or navigation capacity" (Sharif and Alesheikh, 2017a, p.6).
- Movement context: " [...] the entities' quantitative parameters (e.g. speed and acceleration)" (Sharif and Alesheikh, 2017a, p.9).
- Modality context: " [...] a fundamental property of mobility that relates to the entity's condition mode in which something exists or is experienced or expressed [...]" (Sharif and Alesheikh, 2017a, p.9).
- Milieu context: " [...] pertains to any external factor" (Sharif and Alesheikh, 2017a, p.9).

Despite the progress made by Sharif and Alesheikh (2017a), their definitions are still vague, confusing and data-centred rather than focused on the reality of movement. Also, there is a huge overlap between the subdivisions and the meaning of the names are not readily apparent. For the purpose of this thesis we define movement context as follows: one or a set of variables that can be linked to an entity and might be influencing its behaviour, therefore becoming useful for understanding, modelling and predicting movement. These variables can be dynamic or static, depending on the temporal scale of the study and of the variable.

Based on the current literature and in our experience with CAMA, we suggest that movement context can be divided into the following taxonomy, illustrated and related to the work by Nathan et al. (2008), Dodge et al. (2008) and Sharif and Alesheikh (2017a) in Figure 2.1.

- Physiological context: variables inherent to the organism's regular functioning that can trigger a specific movement pattern. The hunger hormone ghrelin, for example, has been found to affect stopover decisions of migratory birds (Goymann et al., 2017). In humans, sex hormones seem to be related to differences in physical activity levels between genders (Bowen et al., 2011).
- Environmental context: variables inherent to the movement's surrounding location, i.e., inherent to the place where movement is happening. The term surrounding here is purposely vague, because the distance threshold to which environment influences movement is moving entity dependant. For example, in an animal ecology context, eagles can see clearly as far as three thousand metres whilst rhinos cannot distinguish between a human and a tree at four metres distance. Thus, the surrounding threshold for an environmental context that is visually perceived would be fundamentally different between these two species.

Next we identify different sub-types of environmental variables, which may differ depending on the application.

- Natural context: variables inherent to the movement's surrounding location that describe the landscape and are not man-made, such as wind fields (Safi et al., 2013), temperature (Edwards et al., 2015), vegetation coverage and state (Pettorelli et al., 2005).
- Circumstantial context: variables inherent to the movement's surrounding location that are not cyclical or permanent at the temporal scale in which movement is being analysed, such as tsunamis, floods (Pregnolato et al., 2017), earthquakes (Gething and Tatem, 2011) and road closures (Cole et al., 1997).
- Anthropogenic context: variables inherent to the movement's surrounding location that are created by the human presence, such as pollution (Dewulf et al., 2016), traffic (Javid and Javid, 2018), transportation mode (Dabiri and Heaslip, 2018), retail and activity options (Siła-Nowicka et al., 2016; Sila-Nowicka, 2018).
- Demographical context: variables inherent to the organism that allows a direct comparison to meaningful groupings of the population and also might explain divergences in movement patterns amongst those groups.

Next we identify different sub-types of demographical variables, which may differ depending on the application.

- Biological context: variables inherent to the organisms that do not depend on any other individual, for example age (Vazquez-Prokopec et al., 2013) and sex (SiłaNowicka et al., 2016).
- Sociological context: variables that describe the organisms relation to others or group belonging, it is more common in human mobility analyses but the same would apply in terms of ethological variables for animals. Some examples are, community and group belonging (Bode et al., 2015; Shi et al., 2015; Toole et al., 2015; Yang et al., 2018), economic status (Kimijima and Nagai, 2017), social interaction (Mollgaard et al., 2017).

In the recent work by Sharif and Alesheikh (2017a), one controversial issue has been the use of the term movement context. On the one hand, Sharif and Alesheikh (2017a) argue that speed, distance, turning angle and acceleration are movement context. On the other hand, Dodge et al. (2008) contend the same attributes as primary and secondary parameters of movement. We agree with Dodge et al. (2008) that those are movement parameters and we use movement context as a general term encompassing all the subdivisions in our taxonomy, as it can be seen in Figure 2.1. Though we concede that movement parameters can help modelling movement, we still insist that movement parameters are not context because they do not full-fill the definition of it. Firstly, movement parameters do not need to be linked to the moving entity, as they are derived from movement samples. Secondly, these parameters do not influence behaviour but rather are an integrated and resultant part of it. In fact, movement parameters are good descriptors but do not fully explain the reasoning and motivation for movement behaviour, which is a main trait of contextual data.

We are aware that the boundaries between classes of movement context are, at times, imprecise but to the best of our knowledge our proposed taxonomy reflects the division between internal and external factors by Nathan et al. (2008), which is a consensus in the movement ecology literature. This consensus is also illustrated in Figure 2.1, where we show the relationship between our taxonomy and other divisions of movement context. Different divisions are shown in distinct colours and dashed lines identify how they relate to each other.


Figure 2.1: Proposed taxonomy for movement context with non-exhaustive examples of variables and how it relates to the classifications proposed by Nathan et al. (2008) (in orange at the top), Dodge et al. (2008) (in green at the bottom) and Sharif and Alesheikh (2017a) (in purple at the bottom). The red rectangle indicates the type of context this thesis focuses on.

In Figure 2.1, at the top in orange, we have the division by Nathan et al. (2008), which comes from ecology and it is the starting point for the other three classifications of context. In the second horizontal block of Figure 2.1 we illustrated our taxonomy, in which we classify physiological and demographical-biological context as internal factors, and demographical-sociological and environmental as external factors. In the third row of Figure 2.1 in green we have the division by Dodge et al. (2008), in which the external factors by Nathan et al. (2008) are divided into other agents, equivalent to our demographical-sociological context, and spatial-constraints, equivalent to our environmental context. In the division by Dodge et al. (2008), intrinsic con-
text is equivalent to the internal division defined by Nathan et al. (2008). The last horizontal block of Figure 2.1 shows the classification by Sharif and Alesheikh (2017a), in which as opposed to Dodge et al. (2008), the authors rename external context as milieu and create three types of internal context. However, as said before, their definitions are fundamentally unclear and for this reason we do not try to relate them to our taxonomy. This thesis focuses on our definition of environmental context surrounding movement (see red rectangle in Figure 2.1).

### 2.3 The role played by context

Movement is triggered by a combination of internal and external factors interacting at different temporal and spatial scales, i.e., it is a product of physiological and environmental variability over a scale ranging from seconds to years (Nathan et al., 2008). Reproductive calendar, nourishment, circadian cycle and age are a few examples of internal factors that affect the movement of living organisms, whilst temperature, wind, land cover and the presence of predators are examples of external factors (Martin et al., 2013). During summer, for example, people are more likely to feel hot (a physiological change) because of the higher temperatures (an environmental change), which might trigger more frequent visits to the ice cream shop, a trip to the beach, a change of the transportation mode or simply an overall increase, or decrease depending on the typical weather, on time spent outdoors. Similarly, some birds are known to migrate in response to the changes in intensity of the sunshine and duration of daylight, which also produce physiological changes, such as the decrease in size of feeding related organs (Nebel, 2010).

The comprehension of how a specific context triggers different movement patterns is of vital importance for understanding behaviour, which is of particular interest for biodiversity conservation (Sutherland et al., 2013), wildlife management (Urbano et al., 2010), epidemiology (Holden, 2006), human mobility (Gao, 2014) and urban planning (Willis et al., 2004). Unveiling behavioural mechanisms is the ultimate goal of many movement analyses because the modelling of behaviour will also allow for its prediction. Solely analysing trajectories enables us to identify movements, but fail to reveal the reasoning and motivation behind those. In contrast, taking
context into consideration can lead to inferences about the drivers of movement, contributing towards the understanding and prediction of behaviour. Yet, there are still few studies taking context into account, perhaps due to the technical challenges posed by integrating movement data and contextual data (Dodge et al., 2008; Laube, 2014). Many of these challenges starts with the different nature of these data sets, which we explore in the next sections.

### 2.3.1 Movement data types

The core of any movement data is the collection of positional information through time (Laube, 2014) which has historically been performed with a variety of techniques (Long and Nelson, 2013). At its onset, movement data were collected via rudimentary techniques such as banding/ringing (Knox, 1983), scale clipping (Blanchard and Finster, 1933), manual travel surveys and travel diaries (Palmer et al., 2013). Currently, most of the data acquisition is performed remotely (Long and Nelson, 2013) via radio telemetry (Salvatori et al., 1999; Fryxell et al., 2008; Cagnacci et al., 2010), surveillance radar (Gauthreaux and Belser, 2005), mobile phone records (Horanont et al., 2013; Louail et al., 2014; Isaacman et al., 2011; Phithakkitnukoon, Smoreda and Olivier, 2012), bluetooth data (Versichele et al., 2012), satellite tracking systems such as the ARGOS (Advanced Research and Global Observation Satellite) data collection and location system (Coyne and Godley, 2005; Prosser et al., 2015), light-level geolocators (Lisovski et al., 2018) and GPS trackers (Urbano et al., 2010; Cagnacci et al., 2010; Siła-Nowicka et al., 2016; De Groeve et al., 2016). Of these, data collection via GPS tracking has been increasingly prevalent in movement studies, both for humans and animals (David et al., 2014; Demšar et al., 2015; Wikelski et al., 2007; Tomkiewicz et al., 2010) and it is the type on which we will focus from now onwards.

In contrast to the early methods, trajectory sampling is automated in GPS tracking. However, there is a trade-off between sampling intervals and duration, which are defined based on the research goals and device limitations (Johnson and Ganskopp, 2008). The sampling rate determines the granularity (Hornsby and Egenhofer, 2002) or temporal resolution and volume of movement data (Cagnacci et al., 2010; Urbano and Cagnacci, 2014). Finer temporal resolutions are associated with higher sampling rates, bigger volume and provide more certain
trajectories, while coarser temporal resolutions are related to lower sampling rates, smaller volume and provide more uncertain trajectories (Long and Nelson, 2013). Sampling frequencies can vary widely (Urbano and Cagnacci, 2014): from milliseconds (Shamoun-Baranes et al., 2006), to seconds (Meijles et al., 2014; Elliott et al., 2014; Xiao et al., 2015; Siła-Nowicka et al., 2016; Roeleke et al., 2018), minutes (Safi et al., 2013; Wang et al., 2015; Shaffer et al., 2017), hours (Dodge et al., 2013; Davies et al., 2013; De Groeve et al., 2016; Fullman et al., 2017) and even days (Cagnacci et al., 2010; Urbano and Cagnacci, 2014; Parlin et al., 2018). Ideally the sampling rate must be higher than the frequency of the behaviour under investigation, but that is not always possible due to battery constraints (David et al., 2014).

The constraints related to battery life can be reduced by setting the GPS device, so that it is switched-off when no movement is detected (David et al., 2014). This results in samples being taken at regular and/or irregular intervals (Calenge et al., 2009), in the first case location points are spread equally in time producing regular trajectories while in the second case location points are spread unevenly in time producing irregular trajectories. This strategy is also reasonable in regards to the biology of organisms, as there is no pattern to be detected when people are sleeping or during the day for nocturnal animals. Yet, it adds to the already high computational complexity of trajectories.

Processing trajectories involves many challenges (Laube and Purves, 2011; Laube, 2014; Shamoun-Baranes et al., 2012), from dealing with very large volumes of data from trajectories with high temporal resolution to dealing with the uncertainty from trajectories with low temporal resolution. These challenges, which we will explore more about in Section 2.5, are further enhanced when contextual data are included in the analysis (Demšar et al., 2015; Cagnacci et al., 2010; Neumann et al., 2015).

### 2.3.2 Contextual data types

Context is represented by a variety of data types in the computational environment (Laube, 2014; Dodge et al., 2013), which can be acquired from users, inference systems or sensors (Shaffer et al., 2017). User generated contextual data are primarily employed in human movement research and consists of interviews, questionnaires, Location-Based Social Networks (LBSN) and
volunteered geographic information (VGI) (Sun et al., 2013), as for example Flickr (Barchiesi et al., 2015), FourSquare (Agryzkov et al., 2017; Krueger et al., 2014), Facebook (Turk, 2013) and Twitter (Blanford et al., 2015; Jurdak et al., 2015) data.

Inferred contextual data, such as movement modes (Siła-Nowicka et al., 2016; Dabiri and Heaslip, 2018; Bolbol et al., 2012), activities (Sila-Nowicka, 2018; Wan and Lin, 2016), places of interest (POI's) (Ashbrook and Starner, 2003; Thierry et al., 2013; Jiang et al., 2015; SiłaNowicka et al., 2016; Sila-Nowicka, 2018) and density of predators (Block et al., 2011) are generated by inference algorithms, segmentation methods, classifiers and forecasting models (Schick et al., 2013).

Sensed contextual data are often used to represent physiological context and/or environmental context. The first is exclusively measured by in-situ co-located sensors (Jeltsch et al., 2013; Demšar et al., 2015) that simultaneously register movement and physiological variables, such as heart rate (Louzao et al., 2014; Richardson et al., 2018). The second can be measured in-situ or remotely and it is the type on which this thesis focuses. More specifically, sensed environmental context has been acquired $i n$-situ by meteorological stations, particularly for assessing the effect of wind on avian flight and for evaluating human response to weather (Table 2.2 ), and remotely by satellites, drones, airplanes and camera traps to obtain data on diverse variables (Neumann et al., 2015), of which we list a few examples in Table 2.2.

### 2.3.2.1 Contextual data from remote sensing

Remote sensing uses sophisticated sensors to measure the spectral electromagnetic energy, i.e. the radiance $\left(W / m^{2} s r\right)$, emitted/reflected by an object or geographic area without coming into direct contact with it, to then extract information on the biophysical properties of those objects (Fussell and Rundquist, 1986; Jensen, 2006). The main advantage of remote sensing comes from the capability of performing systematic data collection over large areas at a relatively low cost, which is useful for characterizing the environment in order to understand the mechanisms behind movement (Pettorelli et al., 2014; Neumann et al., 2015). This is particularly important for large-scale movements such as animal migration, where environmental conditions cannot be measured locally because they are needed over very large areas (Dodge et al., 2008).

Table 2.2: A non-exhaustive list of studies using environmental contextual data in CAMA with reference and specific variable used.

| Variable | Reference |
| :---: | :---: |
| Weather | Horanont et al. (2013) |
|  | Phithakkitnukoon, Leong, Smoreda and Olivier (2012) |
|  | Vanky et al. (2017) |
|  | Sila-Nowicka (2018) |
| Brum-Bastos et al. (2018) |  |
| Wind | Safi et al. (2013) |
|  | Richardson et al. (2018) |
|  | Shamoun-Baranes et al. (2010) |
| Yoda et al. (2012) |  |
| Rainfall | Gutierrez Illan et al. (2017) |
|  | Kleyheeg et al. (2017) |
| Surface temperature | Henry et al. (2015) |
|  | Kappes et al. (2015) |
|  | Henry et al. (2015) |
| Snow coverage | Howey et al. (2017) |
|  | Fullman et al. (2017) |
|  | Cagnacci et al. (2011) |
| Topography | Getz and Saltz (2008) |
|  | Katzner et al. (2012) |
|  | Widmann et al. (2015) |
|  | Kittle et al. (2015) |
|  | Fullman et al. (2017) |
|  | Kleyheeg et al. (2017) |
| Musiani et al. (2010) |  |
|  | Bartlam-Brooks et al. (2013) |
|  | Bohrer et al. (2014) |
|  | Buchin et al. (2015) |
|  | Henry et al. (2015) |
|  | Thorup et al. (2017) |
|  |  |

Most remote sensing products are delivered as digital numbers (DN), i.e. integers assigned to pixels, which allow the visualisation of images by applying different shades of grey but do not provide any biophysical information on the objects. Passive remote sensing systems are further constantly calibrated by also measuring the radiance coming from the Sun, which allows the calculation of the proportion of the received electromagnetic energy that bounces back from an object, or the so-called spectral reflectance (Rees, 2001). While DN are affected by diverse factors, reflectance is a property inherent to objects so that different objects
have specific spectral reflectance curves. Often called spectral signatures, these curves are well documented and describe how much energy a target reflects in each portion of the electromagnetic spectrum, which allows us to distinguish between objects in satellite images but also to extract biophysical information on the targets. The use of reflectance, which can be obtained by applying sensor-specific equations to the DN , instead of DN is particularly important in multi-temporal and multi-sensor studies, because of recurrent variations in the illumination of scenes that make DN incomparable between images of different dates and/or sensors. A forest patch, for example, may not have the same DN in images from different satellites but it will have very similar, if not the same, reflectance values in both. One of the reasons for that are differences between the four resolutions intrinsic to remote sensing.

Remote sensing data are characterised by radiometric resolution, spectral resolution, temporal resolution and spatial resolution (Jensen, 2006; Novo, 2010; Rees, 2001). Radiometric resolution is measured in bits and defines the levels of quantisation captured by an image, i.e., it defines how capable a sensor is of discerning between close objects. It also determines the maximum DN in an image, which is equal to $2^{n}$, where $n$ is the radiometric resolution expressed in number of bits. Common radiometric resolutions are 8 bits (Landsat 4 and 5), 11 bits (QuickBird and IKONOS) and 16 bits (Landsat 8), which yields respectively 256, 2048 and 65536 possible DN. Therefore, the same forest patch on the same day and time will have a different DN in different remote sensing systems.

Spectral resolution refers to the number and width of bands, i.e. intervals in the electromagnetic spectrum, within which reflectance is measured by a remote sensing instrument. It can vary widely, from one large band in WorldView-I (DigitalGlobe, 2016) to 36 narrower bands in MODIS (Moderate Resolution Imaging Spectroradiometer) (Toller and Isaacman, 2006), as the bands are defined based on the spectral signature of the target they are being designed for monitoring. If a satellite system is designed to monitor vegetation, for example, it will have bands in the red and near infra-red intervals of the electromagnetic spectrum, because those intervals are known to capture features that are singular to vegetation. The low reflectance in the red interval, because of chlorophyll absorbency, combined with the high reflectance in the near-infrared caused by multiple reflections within the leaves (more specifically within the
spongy mesophyll) (Figure 2.2) are well known features that allow identifying, assessing and distinguishing vegetation from other targets, such as water and soil (Figure 2.2) (Ponzoni and Shimabukuro, 2010). There are trade-offs between the width of the bands, the spatial resolution which and the temporal resolution.


Figure 2.2: Spectral reflectance curve of vegetation, soil and water from $0.4 \mu \mathrm{~m}$ to $2.5 \mu \mathrm{~m}$ wavelength. The dominant factors controlling leaf reflectance are the pigments in the palisade mesophyll (chlorophyll content), the scattering in the spongy mesophyll (leaf structure) and the amount of water in the plant. Adapted from SEOS (2018)

Temporal resolution refers to the interval of time it will take for a satellite to orbit again above the same point on Earth, i.e., how long it takes until the next image of the same area is acquired (Jensen, 2006). Revisiting periods vary from minutes to days, a longer revisit period is usually related to a higher spatial resolution and fewer bands. On the contrary, shorter revisiting periods are often related to lower spatial resolution but more bands. For example, MODIS has 36 multi-spectral bands with spatial resolution varying from 250 m to 1 km and daily revisit (Toller and Isaacman, 2006), whilst Landsat OLI (Operational Land Imager) has eight bands with 30 m spatial resolution and 16 days revisit period (USGS, 2016). Higher temporal resolution is achieved by projecting systems with bigger scene sizes (swatch), which in turn requires the detectors to be projected over a larger area coarsening the spatial resolution.

Spatial resolution is often erroneously defined as the pixel size of an image generated by a remote sensing system. The resolution defines the size of the smallest object that a sensor is capable of discerning as a single unit, while the pixel size defines the smallest unit with which an image is being displayed (Figure 2.3). Spatial resolution is unchangeable and inherent to the system that generated the image (Jensen, 2006; Novo, 2010; Rees, 2001), but the pixel size can be changed. For example, if a sensor has a spatial resolution of 10 metres and an image from that sensor is displayed at full resolution, each pixel represents an area of $10 \mathrm{~m} \times 10 \mathrm{~m}$ on the ground, in which case the pixel size and resolution are the same (see column one and two of Figure 2.3). However, it is possible to display the same image with a bigger or smaller pixel size different than the resolution (see column one, three and four of Figure 2.3). Increasing the pixel size will degrade the discernibility of objects, but decreasing the pixel size will not do the opposite, because a pixel can not show more than what the spatial resolution is capable of capturing.

The signal registered at a full resolution pixel is the result of the integration of an weighted average of the energy reflected/emitted by all the objects within the area delimited by the spatial resolution (Zhan et al., 2013), this is known as mixture effect (Choodarathnakara et al., 2012). Coarser spatial resolutions result in higher diversity of signals being registered at the same pixel (Figure 2.4) and therefore a more intense mixture effect, which becomes problematic when trying to identify targets and patterns from remote sensing data (Choodarathnakara et al., 2012; Zhan et al., 2013).

The mixture effect is also increased by resampling remote sensing data to a pixel size different from the spatial resolution because it interpolates multiples pixels, i.e., multiple mixtures of objects, to estimate a new value at a certain location. In that sense, the value of the pixel which a location falls within is the best estimate for it and no interpolation will improve the accuracy beyond it. On the other hand, sub-pixel modelling and multi-sensor approaches have been used in the last years to tackle this problem (Zhan et al., 2013).


Figure 2.3: Difference between spatial resolution and pixel size. The first column shows an image grid with an specific resolution overlaying the roof of a house, the other three elements in the row show how the data registered at that resolution will be displayed in an image generated with a pixel size equal, smaller and bigger than the resolution. The red line show the equivalent size of the pixels on the ground.

### 2.3.2.2 Contextual data from meteorological stations

Each country has established its network of meteorological stations with the purpose of predicting the weather (Day, 1966), a collective phenomena of the atmosphere, that can be monitored by observing its meteorological elements (Gole, 1970). These elements, namely atmospheric temperature, atmospheric pressure, relative humidity, precipitation, cloudiness, visibility and


Figure 2.4: Mixture effect illustrated for images with 90 m and 180 m spatial resolution over an urban area. It is clear that there is a higher diversity of objects within a 180 m nominal pixel than within a 90 m one.
wind, are measured at least four times a day at the same Greenwich time by weather stations everywhere in the world (Day, 1966; Gole, 1970; Met Office, 2016).

Weather stations are organised in a network collecting point-wise information at multiple locations, which ideally have level ground and no trees, buildings or hills around that could introduce bias (Met Office, 2015). The measurements of the meteorological elements can be done automatically or manually, depending on the type of station (Met Office, 2016). Manual stations provide data four times a day, while automatic ones usually provide data at hourly intervals data and sometimes even sub-hourly intervals (Day, 1966). In the United Kingdom, for example, there are 200 automatic meteorological stations, which are set up approximately 40 km apart (Met Office, 2016) and collect data at one minute intervals. There are no data collected in-between stations, which means that the values of any variable measured at a meteorological station need to be inferred at those locations.

The extrapolation of point-wise meteorological data to other locations is done by applying spatial interpolation methods, which can be deterministic, probabilistic or mixed (Sluiter, 2008). Deterministic methods use only the geometric properties of the point observations to create a continuous surface. Probabilistic methods allow the inclusion of the variance in the interpolation process and also assess the statistical significance of the predictions. Mixed methods are the ones specially developed for meteorological purposes by combining deterministic and probabilistic methods (Tveito, 2010; Beek, 1991).

The Meteorological Interpolation based on Surface Homogenized Data Basis (MISH) (Szentimrey et al., 2010) and the Parameter Regression on Independent Slopes Model (PRISM) (Daly et al., 1997) are examples of mixed methods specially developed for meteorology. These methods increase accuracy by using information from the climate data series (30 years or more) to estimate the spatial trend differences and the covariances of the surface being interpolated. However, they are still not fully understood (Tveito, 2010) and require massive amounts of data and computational resources.

Kriging is an example of one of the most used probabilistic interpolation methods for weather variables. It incorporates the concept of randomness by accepting that a continuous attribute varies too irregularly to be modelled by a simple function, so that its variation is better described by a stochastic surface with an attribute that respects the concepts of spatial dependence and spatial autocorrelation (Chatterjee and Chowdhury, 2017). Kriging methods produce good results for not too sparse data (Sluiter, 2008), but they require prior estimation and validation of the spatial auto-covariance structure, which can be difficult (Hutchinson, 1995).

Amongst the deterministic methods, Thiessen or Voronoi polygons are the simplest and perform well for dense measurement networks (Sluiter, 2008). This method draws distancebased boundaries between data points by creating one polygon centred at each station, i.e., it predicts the attributes of unsampled points based on those of the nearest sampled point (Hartkamp et al., 1999), which results in a collection of polygons delineating zones that are linked to temporal data-series on the variables measured at the meteorological station. The Inverse Distance Weighting (IDW), another deterministic method widely used in meteorology, includes multiple observations by using the distance from each point data as a weight to estimate the value at a certain location (Tveito, 2010). Other less frequently used deterministic methods are splines, linear regressions and artificial neural networks.

The choice of the interpolation method should take into account the variability of the meteorological variable, the density of the meteorological network and the variations in altitude across the study site (Hartkamp et al., 1999; Tveito, 2010). Temperature in Europe, for example, rarely varies much over regions smaller than $50 \times 50 \mathrm{~km}^{2}$ (Beek, 1991) and decreases about
$0.6^{\circ} \mathrm{C}$ per 100 metres of altitude (Hough and Jones, 1997). Therefore, with stations at intervals equal or smaller to 50 km and relatively constant altitude, it is not necessary to apply more sophisticated interpolation methods and the use of deterministic interpolation, such as Thiessen polygons, is appropriate. Other similar examples are wind and humidity, which generally show little variance in a 50 to 150 km radius, except in coastal areas where the wind varies more and humidity is higher. In contrast, precipitation shows large spatio-temporal variability, not only in intensity but also in form: rain, hail, sleet or snow (Beek, 1991). Therefore, deterministic methods do not perform well with this variable, so that it usually requires more sophisticated methods from probabilistic theory, such as kriging and regression.

The large spatio-temporal variability is one of the reasons why rainfall is frequently monitored also via meteorological radars, which are more capable of retrieving data at sub-hourly interval for monitoring larger areas. Radar measurements are taken off-nadir which results on varying spatial resolution across different ranges, i.e., the further away a point is from the radar, the coarser is the spatial resolution at that point (Jensen, 2006). The accuracy and suitability of the data for different applications is determined by its spatial resolution, which is also linked to the interpolation method applied to the raw radar data.

Regardless of the interpolation method the final data set will have zones linked to the value of the meteorological element and timestamps indicating when they were collected. These zones and the timestamps from the data-series will determine which values of the weather variable are attributed to the points in a trajectory in CAMA. The zones are used for spatial intersection, i.e., trajectory points within a certain Thiessen polygon will receive the values registered at the meteorological station at the centre of that polygon, and the timestamps are used for temporal intersection, i.e., ideally trajectory points within a certain Thiessen polygon or grid cell would receive the values registered at the meteorological station at the centre of that polygon at the time when the location point was recorded. However, there is a mismatch between movement data and meteorological data temporal resolutions, whilst the first are commonly sampled at sub-hourly intervals, the second are sampled at hourly intervals. This asynchrony highlights the need for methods to deal with the temporal incompatibilities between movement data and contextual data, which is one of the challenges of performing CAMA (Dodge et al., 2013;

Demšar et al., 2015).
There are many other characteristics and challenges to remote sensing and meteorological data, but the aforementioned ones are the most relevant for integrating movement data and contextual data. The complexity of processing trajectories is increased by adding contextual information (Demšar et al., 2015; Shamoun-Baranes et al., 2012), the challenge of combining contextual and movement datasets require an intermediate step to integrate all these diverse data sources which have different sampling scales in both space and time (Shamoun-Baranes et al., 2012; Dodge et al., 2013; Coyne and Godley, 2005). The integration is often done via trajectory annotation (Mandel et al., 2011; Dodge et al., 2013), which we explain in Section 2.4 .

### 2.4 Integration of context

The integration of contextual variables and movement paths is done via trajectory annotation (TA) (Mandel et al., 2011; Dodge et al., 2013), a process by which contextual datasets are associated with movement trajectories by performing spatio-temporal intersections to add the value of the contextual variable as an attribute of the points in the trajectory (Safi et al., 2013; Urbano et al., 2010). Trajectory annotation comes from computer science, more specifically from web-browsing, an area in which it is used to add data on important variables encountered through a particular path to the object whose path was recorded (Mandel et al., 2011). Similarly in movement research, an annotated path integrates the values of contextual variables co-located in time and space with the moving organism (Mandel et al., 2011). The annotation of trajectories results in the so-called semantically enriched trajectories (Parent et al., 2013; Safi et al., 2013; Yan et al., 2013), which support the understanding of how context affects movement behaviour.

The extremely large amount of data involved in the process requires TA to be as automated as possible, which led to the development of systems to support the task (Dodge et al., 2008). Most of these systems come from movement ecology, where there is a tradition to make them accessible for other researchers, while systems for TA of human tracking are much more
individualised and closed, so that the methods are often only accessible to authors themselves.
The Satellite Tracking and Analysis Tool (STAT) (Coyne and Godley, 2005), currently Wild Life Tracking, was the first freely available web system to provide tools for management, analysis and integration of environmental data. STAT is primarily focused on data from Argos (a worldwide satellite based tracking and environmental monitoring system for data collection and transmission) and provides oceanographic contextual information on bathymetry, sea surface temperature, chlorophyll, sea surface height and currents (Coyne and Godley, 2005). The Spatial Ecological Analysis of Megavertebrate Animal Populations (OBIS-SEAMAP) (Halpin et al., 2006, 2009) provides a very similar array of contextual variables and despite being a multi-species platform, it still exclusively focused on the maritime organisms.

The first platform to focus on terrestrial animals, still single-species and very specialised, was the EUropean ROe DEER Information System (EURODEER). It was a web-based collaborative platform designed to store shared movement data and a wide set of contextual variables, such as snow coverage and topography (Cagnacci et al., 2011). Another system and the most popular, the environmental-data automated track annotation (Env-DATA) system is "capable of managing and analysing movement trajectories linked to large remote sensing, climatic, and land use datasets will greatly facilitate the next generation of research into movement ecology" (Dodge et al., 2013, p.13). Very similar in architecture but more recent ZoaTrack (Dwyer et al., 2015) also provides tools for integrating contextual data into movement analysis but it is geographically focused on Australasia.

The Env-DATA expanded the capabilities of Movebank (Kranstauber et al., 2011), a free on-line multi-species database of animal tracking data, to assimilate context in the analysis of animal movement data (Dodge et al., 2008). The system offers more than twenty contextual variables derived from remote sensing data and global reanalysis models, as well as three different options of spatio-temporal interpolation methods to match contextual data and movement samples. In addition, DynamoVis (Xavier and Dodge, 2014; Xavier et al., 2018), a visualization tool has been recently added. DynamoVis uses combinations of visual variables to animate multivariate representations of movement, so that it can be visualised within context (Xavier and Dodge, 2014; Xavier et al., 2018). The combination of Env-DATA, MoveBank
and DynamoVis systems are up-to-date undoubtedly the most advanced in terms of facilitating CAMA, particularly the use of remote sensing products as contextual data (Dodge et al., 2013). Yet, CAMA is still a challenging task and it is unclear whether trajectory annotation is the best approach to perform it because of the differences in spatial and temporal resolutions of contextual data and trajectories (Demšar et al., 2015). These incompatibilities and the main challenges to be faced when performing CAMA are explored in Section 2.5.

### 2.5 Current challenges

### 2.5.1 The source of the challenges: mismatch in data resolutions

Movement data can be collected with temporal resolutions ranging from miliseconds to days and at times with sub meter accuracy. Contextual data, on the other hand, might only be available at best hourly, if from meteorological stations, or half-daily, if from orbital remote sensing, and with spatial resolution varying from several tens to hundreds of metres (Dodge et al., 2008; Demšar et al., 2015). The differences in spatial and temporal resolutions of movement data and contextual data are amongst the most pressing issues in CAMA. Table 2.3 lists a few examples of recent studies where this spatio-temporal mismatch is found, there are other many examples in the literature but we do not intend here to create an exhaustive list.

Generally, there are two possible types of temporal incompatibility when performing TA, in the first one movement data are sampled more frequently than contextual data, while in the second one movement data are sampled less frequently than contextual data. The first case can be represented by HH quarter in Figure 2.5, in which there are different phenomena being tracked at fine scale for humans and wildlife, but there are only coarser sources of contextual data. The second case can be illustrated by the LL quarter in Figure 2.5, in which there are different phenomena being tracked at coarser scale for humans and wildlife, but there are only finer sources of contextual data.

Table 2.3: Samples of movement research papers using contextual data with reference, description of movement data and contextual data. MET stands for meteorological stations; RS for remote sensing and GAR for global atmospheric reanalysis models. Dashes indicate that the information was not available on the manuscript. The spatial resolution was converted to km at $0^{\circ}$ latitude where it was originally in degrees to facilitate comparisons. Spatial and temporal resolutions listed are always the best ones reported for irregularly sampled data.

| $\begin{array}{c}\text { Moving } \\ \text { entity }\end{array}$ | $\begin{array}{c}\text { Movement data } \\ \text { temporal resolution }\end{array}$ | $\begin{array}{c}\text { Contextual } \\ \text { variable }\end{array}$ | $\begin{array}{c}\text { Contextual data resolution } \\ \text { temporal }\end{array}$ |  | Refatial | Reference |
| :---: | :---: | :---: | :--- | :---: | ---: | :---: | \(\left.\begin{array}{c}Data <br>

source\end{array}\right)\)


Figure 2.5: Incompatibilities in spatial and temporal resolutions between movement data and contextual data after Neumann et al. (2015); Shen and Stopher (2014) and Meekan et al. (2017). Ellipses show the spatio-temporal range of movement data for different applications, black dots show the same for free sources of contextual data. The low-left quarter represents the high temporal and spatial resolution (HH) domain. Clockwise from there, we have the quarter with low temporal resolution and high spatial resolution (LH), the quarter with low temporal and spatial resolution (LL), and the quarter with high temporal resolution and low spatial resolution (HL).

Either mismatch types make the use of pre-processing measures, such as interpolation or aggregation, necessary. However, while finer contextual data can be generalised or aggregated without many issues, pre-processing coarser contextual data requires more caution. The reason for this is that finer contextual data allows to extract trends and cycles at coarser scales than the contextual data were collected at because the information is already there, while coarser contextual data requires the use of inference or auxiliary data sets to extract trends at a finer scale, simply because the information is not in the original data. For example, if we were
to describe the weather hourly and we know that it was raining at 13:00, 14:00, 15:00, 16:00 and 17:00, it is possible to generalise this information to "it was raining between 13:00 and 17:00". However, if we only knew that it was raining at 13:00 and at 17:00, it is less possible to know what happened between those points in time without looking for auxiliary data in the forecast or inferring from previous experiences. Movement studies have been using interpolation methods to infer contextual data at unsampled spatio-temporal coordinates and overcome the mismatch between resolutions before performing TA, which is one of the challenges involved in performing CAMA.

### 2.5.2 CAMA challenges

The annotation process starts by finding at least four pixels in the grid of the contextual data spatially adjacent to the point being annotated ( $\mathrm{P}_{\mathrm{i}}$ in Figure 2.6 a ), at the two timestamps before ( t ) and after ( t ') the fix to be annotated (Figure 2.6 b ). Then, the values of the four neighbour pixels $\mathrm{v}_{1}, \mathrm{v}_{2}, \mathrm{v}_{3}, \mathrm{v}_{4}$ at t and $\mathrm{v}^{\prime}{ }_{1}, \mathrm{v}^{\prime}{ }_{2}, \mathrm{v}^{\prime}{ }_{3}, \mathrm{v}^{\prime}{ }_{4}$ at t ' are interpolated in space to compute the value of the contextual variable at $t$ and $t$ ', which are subsequently interpolated in time to compute the contextual value at time $t_{i}$ when the fix $\mathrm{P}_{\mathrm{i}}$ was collected (Figure 2.6 b ). Current annotation systems, such as STAT and ZoaTrack, offer at least one method for temporal interpolation and multiple ones for spatial interpolation. Env-DATA, the most popular annotation system, offers the nearest neighbour (NN), bilinear, and inverse distance weighted (IDW) methods for spatial interpolation, and the NN and IDW for temporal interpolation (Dodge et al., 2013). For details on these methods see Chapter 3.

In Env-Data, for example, the finest contextual dataset has a spatial resolution of 1 km (See table 1 in Dodge et al. (2013, p.6)), which means that, in the best case scenario, it would interpolate from an area of at least $4 \mathrm{~km}^{2}$ to estimate the value of a contextual variable at point $P_{i}$ at $t$ and from another $4 \mathrm{~km}^{2}$ to estimate it at t ', and then interpolate the values at t and t' to estimate the context at $\mathrm{t}_{\mathrm{i}}$. Following, resulting semantic trajectories are delivered in commaseparated values (csv) format that can be visualised in DynamoViz (Dodge et al., 2008; Xavier et al., 2018). However, none of these systems include any statistical and analytical methods for knowledge discovery and data mining of the annotated trajectories to perform context-aware
similarity analysis (CASA).


Figure 2.6: Interpolation in space and time from Dodge et al. (2013). a) The variable data for track-point $P_{i}$ is first interpolated in space based on the data from the available pixels in the contextual dataset native grid around $\left.P_{i} . b\right)$ Similar spatial interpolations are conducted at the two nearest available points in time, the nearest before and nearest after the time-stamp of the fix $P_{i}$. Then, the two interpolated spatial values are interpolated in time to the time-stamp of $P_{i}$.

There are three main problems to be considered in this process :

1. The temporal profile of the context being annotated, i.e., how that contextual variable progresses or evolves in time and/or space.
2. The spatial structure of contextual data, i.e., issues related to the data collection, such as the mixture effect in coarse pixels.
3. What to do with semantic trajectories, i.e., once trajectories have been annotated with contextual data, how can we identify patterns in these trajectories that can be linked with movement behaviour? This can be for example by looking at movement and contextual similarity.

### 2.5.2 1 Interpolation within TA and the temporal profile of contextual data

In terms of temporal profile, rainfall and vegetation phenology, are examples of contextual variables with different temporal cycles, yet they are often processed using the same interpolation
methods, i.e., the interpolation methods are chosen disregarding the specificity of each variable and scale. Perhaps because the implications of using different methods for temporal interpolation of contextual data, prior TA, have not yet been tested nor compared. Therefore, in order to better deal with the temporal incompatibilities in TA, there is a the need to compare the current interpolation methods and its implications, but also search for methods that take into account the temporal progression of contextual variables.

### 2.5.2.2 Spatial structure of contextual data

Regarding the spatial structure of the data, considering the mixture effect inherent to remote sensing data (See section 2.3.2.1), which is specially prominent in data with coarse spatial resolution, and taking into account that "everything is related to everything else, but near things are more related than distant things" (Tobler, 1970) as stated by the first law of geography, it is naive to assume that the spatial interpolation of such coarse pixels will, in fact, improve the accuracy of the contextual data at point $\mathrm{P}_{\mathrm{i}}$. On the contrary, it is only mixing more signals and adding more noise to the data, since the other four pixels being interpolated are further away from $P_{i}$ than the pixel in which $P_{i}$ falls within, which is already an average o signals from a $1 \mathrm{~km}{ }^{2}$ area as discussed in Section 2.3.2.1.

Therefore, there is also the need to find new and better ways to deal with the spatial mismatch between the resolutions of movement data and contextual data, particularly for when contextual data are coarser than movement data, which is the most common case as movement data are collected point-wise whilst context is mostly collected by areas.

### 2.5.2.3 How to analyse semantic trajectories

In order to identify patterns in movement that can be linked to movement behaviours, "it is essential to not only explore the spatial and temporal dimensions of the moving objects but also examine how identical they are from contextual perspective" (Sharif and Alesheikh, 2017b, p.428). Yet, while spatio-temporal similarity has been thoroughly explored (Ranacher and Tzavella, 2014), contextual similarity and a combined contextual-spatio-temporal one, which
is relevant for CAMA, have remained largely neglected (Sharif and Alesheikh, 2017b), perhaps because of the complexity of semantic trajectories that require further analytical approaches. Buchin et al. (2014) presented a method for extending geometric similarity measures to perform CAMA. De Groeve et al. (2016) used sequence alignment techniques to cluster animals according to their types of habitat. Sharif and Alesheikh (2017b) generalised the dynamic time warping method to a context-based dynamic time warping, in which elements that have similar contexts can be matched even if they are asynchronous. These methods were able to distinguish between moving entities that have a similar spatio-temporal track but within different contexts. However, these are just a few first approaches of using analytical methods to better understand movement behaviour from semantic trajectories. Context awareness is still a new trend (Sharif and Alesheikh, 2017a) and there is the need to keep exploring more options for performing CAMA.

Therefore, there is also a need to develop innovative methods to perform CAMA, as for example methods that can handle more than one contextual variable at time and that also provide a way to visualise context-aware similarity groups.

### 2.5.3 Moving forward

In Chapter 3 of this thesis we tackle the temporal incompatibilities of CAMA in by exploring the current interpolation methods used in TA and comparing those with a dynamic trajectory annotation method (DTA), which we introduce in the same chapter. In Chapter 4 we use multi-source multiple resolution contextual information from remote sensing, a novel approach in movement research, to overcome spatial and temporal incompatibilities between movement data and contextual data. More specifically, we propose that instead of using interpolation, it is possible to improve contextual data by using multiple sources for one contextual variable instead of only one source so that the best spatial and temporal resolution are preserved. Further in the same chapter, we introduce a new approach for identification of prevalent behaviours from semantic trajectories, the use of Eigen decomposition to look at individuals with similar behaviour in different seasons. In Chapter 5 we deal with temporal incompatibilities to contextualise human movement with weather context from meteorological stations and remote
sensing radar. Further in the same chapter we identify behaviour patterns from semantic trajectories annotated with multiple contextual variables using another new method in movement research, the Multi-Channel Sequence Alignment analysis.

## Chapter 3

## Comparing trajectory annotation methods

### 3.1 Introduction

The process by which environmental data are associated with a movement trajectory is termed trajectory annotation (TA) (Mandel et al., 2011) and results in a so-called semantically enriched trajectory (Parent et al., 2013; Safi et al., 2013). TA is limited by the mismatch between temporal and spatial resolutions of environmental data and trajectory data. For example, weather radar data are typically collected at five minute intervals, which is infrequent compared to the tracking resolution of GPS data for humans and birds whose movement is potentially affected by precipitation (e.g., 5 seconds, humans, Siła-Nowicka et al. (2016); 125 milliseconds, gulls, buzzards and swifts, Shamoun-Baranes et al. (2006); 3 minutes, gulls, Stienen et al. (2016). Many other examples exist (Coyne and Godley, 2005; Dodge et al., 2013, 2014; Palm et al., 2015; Robinson et al., 2010; Safi et al., 2013; Shamoun-Baranes et al., 2006; Therrien et al., 2015). The opposite problem of movement being sampled less frequently than environmental layers is common in movement ecology, where there is a trade-off between the life of the battery on the GPS tracker and the sampling frequency. In particular, larger mammals are often tracked at low frequencies (e.g., 3-4 GPS points per day; (Cagnacci et al., 2011), which is less frequent than available data on weather conditions which can affect wildlife movement patterns.

In terms of spatial resolution, each location recorded in a trajectory data set corresponds to an exact point, while environmental data are typically represented in the form of a regular grid (i.e., raster data), with a spatial resolution ranging from centimetres (e.g., with modern drone systems) to hundreds of kilometres. To summarize, based on differences in temporal resolution between trajectory and environmental data we have three possible situations: 1) trajectory data sampled at a finer temporal resolution than environmental data (Figure 3.1 a ), 2) trajectory data sampled at a coarser temporal resolution than environmental data (Figure 3.1 b ), and 3 ) when trajectory data and environmental data are sampled at corresponding temporal resolutions, which does not pose a temporal mismatch problem. This chapter addresses the first issue, i.e., temporal interpolation for when trajectory data are sampled at a finer sampling rate than environmental data.


Figure 3.1: Two main types of mismatches between movement trajectories and environmental data represented in a space-time cube. Fixes are represented by yellow circles and trajectories by black lines. a) Trajectory's fixes are sampled more frequently than environmental data. b) Trajectory's fixes are sampled less frequently than environmental data.

We can conceptualise the mismatch between the temporal resolution of a trajectory and
environmental data as a case of missing data. If trajectories are sampled more frequently than environmental data values, the unobserved values need to be estimated at the time $t_{n}$ when the fix was recorded. The unobserved values can be estimated via interpolation methods (Tang et al., 2016, 2017), such as the nearest neighbour (NN) which is commonly chosen for TA for being straightforward to compute. NN assigns the nearest environmental record in space and/or time to the fix (Coyne and Godley, 2005; Dodge et al., 2013). Other common methods are: the neighbour before (NB), which assigns the data value immediately before to the fix; the neighbour after (NA), which assigns the data value immediately after to the fix; and the arithmetic mean (AM), which assigns the sum divided by two of the data values immediately before and after to the fix.

These methods disregard the continuity inherent to most natural phenomena often considered in CAMA, such as precipitation, temperature, humidity, wind direction and wind speed. This means that the environmental variable is not being modelled realistically, creating bias and possibly introducing spurious relationships between environmental conditions and movement patterns. In contrast, Dynamic Trajectory Annotation (DTA) can be used to better approximate the modelled environmental variable as a continuously varying natural phenomenon and compare its performance against other methods using simulated movement data and real environmental data. Instead of assigning the closest value in time and space to each fix, DTA interpolates the environmental variable at the unknown time by estimating intermediate values between two given environmental layers at two consecutive times. The DTA method aims to address a common discrepancy between temporal collection scales of trajectories and environmental data, the case where trajectory fixes are collected more frequently than environmental data.

The novelty of our work resides first in the comparison of these methods for temporal interpolation for TA and second in the creation of an open Python package (VANJU) with all TA methods to perform CAMA. Similar TA methods are available in Movebank.org through Env-DATA (Dodge et al., 2013), but they are limited to the datasets already in MoveBank and require the movement data to be uploaded to the web. Env-DATA also performs spatial interpolation before temporal interpolation, main point in which our approaches differ. We see
spatial interpolation as a pre-processing step, in which the aim is to standardize the different resolutions from multiple contextual data sources rather than match them to the points in a trajectory. For this reason, our method focuses on temporal interpolation for TA.

The primary objective of this chapter is to measure and compare the performance of interpolation methods most commonly used for TA and DTA. The rest of the chapter is structured as follows: section 3.2.1 describes and formalizes DTA, section 3.2.2 introduces the interpolation methods most commonly used for TA, namely: NB (Neighbour Before), NA (Neighbour After), NN (Nearest Neighbour) and AM (Arithmetic Mean); section 3.2.3 describes the data sets and how we built our simulated trajectories for which we had real accumulated rainfall data to be used as ground truth; section 3.2.4 describes how we annotated trajectories with rainfall rates from radar data using five different temporal interpolation methods to calculate the accumulated rainfall to be compared to the ground truth data (section 3.2.5). In section 3.2.6 we evaluate the performance of each method by analysing the difference between the real accumulated rainfall and the accumulated rainfall estimate. The results are discussed in section 3.3 and we conclude with some considerations about the potential of the different annotation methods and some ideas for further use of the DTA method with other environmental variables for performing CAMA (section 3.4). Parts of this chapter are published as the conference abstracts: Brum-Bastos et al. (2016) and Brum-Bastos et al. (2015).

### 3.2 Methods

### 3.2.1 Dynamic Trajectory Annotation (DTA)

This section describes and formalizes the DTA method, which takes continuity into account and is applicable when environmental data have coarser temporal resolution than the trajectory data. Consider two layers of the same environmental variable at times $t_{1}$ and $t_{2}$, with values $\mathrm{v}_{\mathrm{t} 1}$ and $\mathrm{v}_{\mathrm{t} 2}$, and a trajectory fix j at $\mathrm{t}_{\mathrm{n}}$, where $\mathrm{t}_{1}<\mathrm{t}_{\mathrm{n}}<\mathrm{t}_{2}$ and $\mathrm{t}_{\mathrm{n}}-\mathrm{t}_{1}<\mathrm{t}_{2}-\mathrm{t}_{\mathrm{n}}$. TA using the NN method assigns $v_{t 1}$ to the fix $j$, as illustrated in Figure $3.2 a$, because $v_{t 1}$ at $t_{1}$ is the nearest neighbour of the fix j in time. This process can be misleading because environmental conditions may have changed considerably between $t_{1}$ and $t_{n}$. With DTA the environmental
values are interpolated as a continuous function between $t_{1}$ and $t_{2}$. The simplest interpolation possible is a linear function, as illustrated in Figure 3.2 b, with more complex functions also possible, e.g. a cubic spline. We call the linear case DTA:L and the spline version DTA:CS.


Figure 3.2: TA for a trajectory falling within a single pixel of a stack of spatial layers, the layer is omitted and only the pixel is illustrated for a clearer view. The trajectory is shown in space-time cubes with geographical space at the bottom and hypothetical environmental data displayed at corresponding times. a) Trajectory annotation: the value assigned to $j$ is equal to $\mathrm{v}_{\mathrm{t} 1}$, the value of the nearest neighbour in time; b) Dynamic trajectory annotation: the value assigned to j is an interpolated value between $\mathrm{v}_{\mathrm{t} 1}$ and $\mathrm{v}_{\mathrm{t} 2}$.

In the simplest approach, DTA:L, the value $\mathrm{v}_{\mathrm{tn}}$ is calculated assuming that the change rate between $t_{1}$ and $t_{2}$ is linear; for this, a linear function is derived between each pair of subsequent values of the environmental variable (Equation 3.1).

$$
\begin{equation*}
v_{t_{n}}=\left(\frac{v_{t_{2}}-v_{t_{1}}}{t_{2}-t_{1}}\right) t_{n}+\left[v_{t-1}-\left(\frac{v_{t_{2}}-v_{t_{1}}}{t_{2}-t_{1}}\right) t_{1}\right] \tag{3.1}
\end{equation*}
$$

Most environmental datasets are collected at a regular temporal resolution $\Delta t$; therefore we can say that $\mathrm{t}_{2}-\mathrm{t}_{1}=\Delta \mathrm{t}$ and $\mathrm{t}_{1}=0$ for any pair of environmental variables so that Equation 3.1 is simplified into Equation 3.2, as follows:

$$
\begin{equation*}
v_{t_{n}}=\left(\frac{v_{t_{2}}-v_{t_{1}}}{\Delta t}\right) t_{n}+v_{t_{1}} \tag{3.2}
\end{equation*}
$$

In the second and more complex approach, the DTA:CS, the value $\mathrm{v}_{\mathrm{tn}}$ is calculated by fitting
piece-wise cubic polynomials which pass through four control points, i.e., four environmental values: $\mathrm{v}_{\mathrm{t} 0}, \mathrm{v}_{\mathrm{t} 1}, \mathrm{v}_{\mathrm{t} 2}$ and $\mathrm{v}_{\mathrm{t} 3}$. The second derivative of each polynomial is set to zero at the endpoints $\left(\mathrm{t}_{0}, \mathrm{v}_{\mathrm{t} 0}\right)$ and $\left(\mathrm{t}_{3}, \mathrm{v}_{\mathrm{t} 3}\right)$ to provide the boundary conditions for the system of two equations. For the same fix $j$, we calculate $v_{t n}$ using the second piece of the spline, i.e. $v_{t n}=$ $\mathrm{S}_{2}\left(\mathrm{t}_{\mathrm{n}}\right)$. Each $\mathrm{i}^{\text {th }}$ piece of the cubic spline can then be represented as follows in Equation 3.3.

$$
\begin{equation*}
S_{i}\left(t_{n}\right)=a_{i}+b_{i}\left(t_{n}-t_{i}\right)+c_{i}\left(t_{n}-t_{i}\right)^{2}+d_{i}\left(t_{n}-t_{i}\right)^{3} \tag{3.3}
\end{equation*}
$$

Given the set of fixes and environmental values $\left(\mathrm{t}_{0}, \mathrm{v}_{\mathrm{t} 0}\right),\left(\mathrm{t}_{1}, \mathrm{v}_{\mathrm{t} 1}\right),\left(\mathrm{t}_{2}, \mathrm{v}_{\mathrm{t} 2}\right)$ and $\left(\mathrm{t}_{3}, \mathrm{v}_{\mathrm{t} 3}\right)$ we need to find the set of three splines $\mathrm{v}_{\mathrm{i}}(\mathrm{t})$ for $\mathrm{i}=0,1,2$ and 3 . These splines must satisfy Equations 3.4, 3.5, 3.6 and 3.7 as follows:

$$
\begin{equation*}
S_{i}\left(t_{i}\right)=V_{t_{i}}=S_{i-1}\left(t_{i}\right) \quad \text { for } \quad i=1,2 \quad \text { and } \quad 3 \tag{3.4}
\end{equation*}
$$

$$
\begin{equation*}
S_{i}^{\prime}\left(t_{i}\right)=S_{i-1}^{\prime}\left(t_{i}\right) \quad \text { for } \quad i=1,2 \quad \text { and } \quad 3 \tag{3.5}
\end{equation*}
$$

$$
\begin{equation*}
S_{i}^{\prime \prime}\left(t_{i}\right)=S_{i-1}^{\prime \prime}\left(t_{i}\right) \quad \text { for } \quad i=1,2 \quad \text { and } \quad 3 \tag{3.6}
\end{equation*}
$$

$$
\begin{equation*}
S_{0}^{\prime \prime}\left(t_{0}\right)=S_{3}^{\prime \prime}\left(t_{3}\right)=0 \tag{3.7}
\end{equation*}
$$

Equations 3.4, 3.5, 3.6 and 3.7 can be rearranged into a symmetric tridiagonal system to find $a_{i}, b_{i}, c_{i}$ and $d_{i}$ (Bartels et al., 1998). Then $a_{2}, b_{2}, c_{2}$ and $d_{2}$ are replaced in Equation 3.3 to estimate $\mathrm{V}_{\mathrm{tn}}$.

Note however that while we theoretically introduced the DTA:CS method, the computation
cost for fitting the cubic spline to our data set proved to be too expensive for practical purposes (see section 3.3 for details). We therefore note the mathematical possibility of using this method, but it is omitted in our comparative analysis. From here on we also refer to DTA:L method as simply DTA.

### 3.2.2 Most commonly used trajectory annotation methods

The equations for each of the five TA methods used in the comparative analysis are described in Table 3.1. Figure 3.3 illustrates the differences between methods for three hypothetical trajectories. Hypothetical environmental layers and annotated trajectories are shown in spacetime cubes, that is, volumes where the two bottom dimensions represent the geographic plane and the third dimension represents time. Trajectories are shown as a polyline in each cube and a hypothetical environmental variable is shown as horizontal layers at $t_{1}$ and $t_{2}$ times. Different TA methods generated distinct annotated trajectories: for NB, NA and AM the changes in the environmental variable rate occur where a fix intersects the layer in time, while for NN changes occur mainly half way between $t_{1}$ and $t_{2}$, and for DTA changes are smoother and not restricted to the intersections.

Table 3.1: Equations and abbreviations for the five Interpolation methods tested.

| Interpolation method | Abbreviation |  | Equation |
| :---: | :---: | :---: | :---: |
| Neighbour before | NB |  | $\mathrm{v}_{\mathrm{t}_{\mathrm{n}}}=\mathrm{v}_{\mathrm{t}_{1}}$ |
| Neighbour after | NA |  | $\mathrm{v}_{\mathrm{t}_{\mathrm{n}}}=\mathrm{v}_{\mathrm{t}_{2}}$ |
|  |  | If | $t_{\mathrm{n}}-t_{1}>t_{2}-t_{\mathrm{n}}:$ |
| Nearest neighbour | NN | Else : | $v_{\mathrm{t}_{\mathrm{n}}}=v_{\mathrm{t}_{1}}$ |
|  |  |  |  |
|  | $v_{\mathrm{t}_{\mathrm{n}}}=v_{\mathrm{t}_{2}}$ |  |  |
| Arithmetic mean | AM | $v_{t_{n}}=\frac{v_{t_{1}}+v_{t_{2}}}{2}$ |  |
|  |  |  |  |
| Dynamic trajectory annotation | DTA | $v_{t_{n}}=\left(\frac{v_{t_{2}}-v_{t_{1}}}{\Delta t}\right) t_{n}+v_{t_{1}}$ |  |



Figure 3.3: Interpolation methods applied to a simulated trajectory. Trajectories are shown in space-time cubes with geographical space at the bottom and hypothetical environmental data displayed at corresponding times. The colour scale refers to the intensity of the hypothetical environmental value for a fix and/or layer.

### 3.2.3 Data description

The five interpolation methods were evaluated against each other using the following data sets covering the period from $28^{\text {th }}$ of September 2013 to $10^{\text {th }}$ of January 2014. Spatial and temporal resolutions for all data are given in Table 3.2.

- Observed hourly accumulated precipitation from UK Met Office meteorological stations:

Defined as the accumulated precipitation captured during each hour interval (e.g., 06:00 - 07:00), collected from 78 Met Office meteorological stations as ground truth data. The accumulated precipitation is registered by rain-gauges that store rainfall only in liquid form during a one hour period; rain-gauges consist of a circular collector, delineating a $750 \mathrm{~cm}^{2}$ sampling area, and a funnel that conducts the collected rain into an automatic
measuring mechanism or into a reservoir where it may be measured by a human at a later time (Met Office, 2015). Figure 3.4a shows their location.

- Rainfall rates from the UK Met Office "Now casting and Initialization for Modelling Using Regional Observation Data System" (NIMROD) radars: Radar measurements are taken off-nadir which results on varying spatial resolution across different ranges, i.e., the further away a point is from the radar, the coarser is the spatial resolution at that point (Jensen, 2006). The spatial resolution, which here varies between $1 \mathrm{~km}, 2 \mathrm{~km}$ and 5 km , determines the accuracy and suitability of the data for different applications, for the NIMROD radar it works as follows. 5 km coverage provides useful qualitative data and a good overall picture of the extent of precipitation at a regional scale. 2 km coverage provides good quantitative data and shows more detailed distribution of precipitation intensities. It is suitable for more demanding rainfall monitoring and hydrological applications. 1 km coverage provides the most detailed quantitative information, down to the scale of individual clouds. It is designed to assist real-time monitoring of small urban catchments and sewer systems (Met Office, 2007). Each station was identified with its respective NIMROD coverage to assess how the degradation of the spatial resolution influences the accuracy of all interpolation methods (Figure 3.4).
- Simulated trajectories: For each accumulated precipitation interval we created a simulated trajectory at each station that would enable us to test interpolation methods at all stations. These simulated trajectories are stationary in space and fixes are sampled at five seconds (Figure 3.4 b ), following a typical sampling rate used in movement research on human mobility (Siła-Nowicka et al., 2016). The spatial stationarity of simulated trajectories is important to evaluate the interpolation methods, it allows comparison between ground truth data (actual accumulated rainfall at each station) and annotated data.


### 3.2.4 Data preprocessing

The rainfall rates from the Met Office NIMROD radar were used as contextual data to annotate simulated trajectories. This dataset was chosen because it is possible to derive precipitation

Table 3.2: Summary of characteristics of datasets used for trajectory annotation.

| Data set | Type | $\Delta \mathrm{t}$ | Spatial resolution | Source |
| :---: | :---: | :---: | :---: | :---: |
| Trajectories | Point | 5 s | - | Simulated |
| Rainfall | Raster | 5 min | $1-5 \mathrm{~km}$ | NIMROD radars |
| Precipitation | Point | 1 h | - | Met Office stations |

accumulation from the radar values, which allows a direct comparison to the ground truth accumulation values from meteorological stations.


Figure 3.4: Visual summary of datasets used for trajectory annotation. a) Location of Met Office stations and NIMROD coverage with respective resolutions; b) Illustration of 100 seconds segments of simulated trajectories for three Met Office stations.

Simulated trajectories and the NIMROD data were stored in a PostGIS spatial data base and a Python script, using Psycopg library, was developed to manipulate the datasets. First we intersected the NIMROD rainfall data in space and time to annotate each fix with the rainfall value from the neighbour raster before ( $\mathrm{NB}-\mathrm{t}_{1}$ ) and from the neighbour raster after ( $\mathrm{NA}-\mathrm{t}_{2}$ ) in time, i.e., for a fix at $20: 53, t_{1}=20: 50$ and $t_{2}=20: 55$. Values for the other interpolation methods were calculated according to equations in section 3.2.2 using the value from the NB
as $\mathrm{v}_{\mathrm{t}_{1}}$ and the value from the NA as $\mathrm{v}_{\mathrm{t}_{2}}$.
The intervals in which the accumulated precipitation was zero and those where there was a failure in the acquisition of NIMROD data were excluded from the analysis. This procedure guaranteed that for every station used in testing, there were twelve NIMROD rainfall layers for each hourly accumulation interval, i.e., one layer at each five minutes; which excluded potential influences of data failures on the accuracy of annotated values. For each precipitation accumulation interval we created attributes identifying the meteorological station, start time, end time and date.

### 3.2.5 Generating and comparing rainfall rates and cumulative curves

A mass curve of rainfall is a plot of cumulative rainfall against time, from which the total accumulation and intensity of rainfall at any instant of time can be found. It is always a rising smooth curve and may have horizontal sections which indicate periods of no rainfall (Raghunath, 2006, p.37). To analyse the performance of each interpolation for visualization and assess which method had an accumulation curve that was the closest to physical reality, we calculated the accumulated precipitation (mm) for each accumulation interval for each interpolation method at each station by applying the trapezoidal rule to integrate annotated rainfall rates ( $\mathrm{mm} / \mathrm{h}$ ) in time. The trapezoidal rule is a technique that calculates a definite integral by approximating the region under a curve as a trapezoid and calculating its area; we applied it for each pair of consecutive annotated rainfall rates within a precipitation accumulation interval and summed these up to estimate the accumulated precipitation for that hour (Figure 3.5).

### 3.2.6 Validation against ground truth and statistical comparison of interpolation methods

The comparison amongst interpolation methods was based on the difference of the estimated accumulation from the real accumulated precipitation obtained from the meteorological data. The difference was calculated by subtracting the accumulated value estimated by the interpolation method from the value observed at the respective station and time. We computed quantiles, average, standard deviation and skewness of the difference for each method. We


Area $(A) \square$ rainfall $(\mathrm{mm})=$ rate $B(\mathrm{~mm} / \mathrm{h}) \times \Delta t(\mathrm{~h})+\frac{[\text { rate } A(\mathrm{~mm} / \mathrm{h})-\text { rate } B(\mathrm{~mm} / \mathrm{h})] \times \Delta t}{2}$
Rainfall $(\mathrm{mm})$ for Accumulation interval $=$ Area $(A)+$ Area(B) $\ldots+$ Area $(\mathrm{n})$
Figure 3.5: Visual explanation of the trapezoidal rule applied to our variables.
tested the distribution of differences of each method for normality and applied the Wilcoxon rank-sum test to evaluate if the differences from one method were more likely to be larger than the differences from another method. We also conducted two non-parametric ANOVA tests on the methods: the first one to analyse if there were differences between groups of stations under different NIMROD coverages, as coarser spatial resolution could imply poorer performance; the second one to analyse if there were differences between individual stations, as the accuracy of measurements in the station gauges can affect the results. We also used the Wilcoxon ranksum test to qualify the effect of the spatial resolution of annotated data on the accuracy of the methods, i.e., if accuracy increases or decreases with the degradation of spatial resolution.

### 3.3 Results

### 3.3.1 Generating and comparing rainfall rates and cumulative curves

To illustrate how each method models the accumulated precipitation and the instantaneous rainfall, we selected a trajectory segment with the maximum rainfall rate registered in our
database (Figure 3.6). The record is from the $31^{\text {st }}$ of October 2013 at Sutton Bonington (Nottinghamshire) weather station between 17:55 and 18:15. Figure 3.6 shows how this trajectory was annotated with rainfall rates and accumulated rainfall curves for each interpolation method. Fixes are coloured accordingly to their rainfall rate and shown in space-time cubes. The respective accumulated precipitation curve is displayed as the red line on the graphs under each space-time cube.

The different interpolation methods generated visually distinct annotated trajectories. The 3D graphs in Figure 3.6 show that the changes in rainfall rates were abrupt in all methods except DTA. This means that the rainfall rate in the annotated trajectories in NB, NA and NN increased from zero to a $335 \mathrm{~mm} / \mathrm{h}$ plateau in five seconds and stayed at that level for five minutes, followed by dropping back to zero. AM resulted in abrupt transitions as well, but the increment was half that of the first three methods and covering a longer interval of 10 minutes. The DTA resulted in smoother transitions, while still preserving the $335 \mathrm{~mm} / \mathrm{h}$ peak and the $0 \mathrm{~mm} / \mathrm{h}$ borders - a pattern that corresponds much better to the real rainfall progression as registered by the NIMROD system.

It is also possible to identify differences between the accumulation curves; the DTA curve (Figure 3.6) is the closest to a continuous smooth growth curve, which is what would be expected from a real mass curve of rainfall (Raghunath, 2006, p.37). Additionally, the accumulation curve for DTA can be split into at least three sections with different inclinations, which correspond to moving through precipitation of different strengths and accumulating rainfall at three different but consistent rates. These sections are not identifiable on the accumulation charts of the other methods, all which have similar shapes differing only by the starting and ending time of the accumulation.

### 3.3.2 Validation against ground truth and statistical comparison of TA methods

The distribution of differences between estimated and real precipitation accumulation was similar for all methods (Figure 3.7). Accuracy ranged from -10 mm to 10 mm of precipitation with a high precision reported by the mean close to 0 and the standard deviation of approxi-


Figure 3.6: Five interpolation methods with corresponding rainfall rates and precipitation (accumulated rainfall) curves for a segment of the stationary trajectory simulated at Sutton Bonington (Nottinghamshire) weather station. The segment of trajectory is from the 31 st of October 2013, it is shown in space-time cubes with the rainfall rates displayed by different colouring of the fixes. The precipitation is displayed by the red line on the graph under each space-time cube.
mately 0.8 for all methods (Table 3.3). All methods failed the Kolmogorov Smirnov normality test ( $\alpha=0.05$ ) and the skewness values (See Table 3.3) indicate a slightly positive asymmetrical distribution, which means that all methods are likely to underestimate the accumulated
rainfall.


Figure 3.7: Difference in mm of rainfall for all methods by interpolation method: pink dots show outliers; green line represents the median.

Table 3.3: Descriptive statistics and p-values for Kolmogorov-Smirnov normality test for all methods. All p-values were smaller than 0.01 .

| Algorithm | Skewness | Q1 | Q2 | Q3 | Q44 | Average | StD |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| NB | 0.380 | -0.277 | 0.084 | 0.308 | 8.683 | 0.048 | 0.870 |
| NA | 0.374 | -0.279 | 0.103 | 0.339 | 8.674 | 0.059 | 0.895 |
| NN | 0.373 | -0.281 | 0.094 | 0.324 | 8.678 | 0.053 | 0.882 |
| AM | 0.391 | -0.281 | 0.093 | 0.323 | 8.679 | 0.053 | 0.880 |
| DTA | 0.380 | -0.282 | 0.094 | 0.324 | 8.678 | 0.053 | 0.881 |

To check whether the differences from one method were likely to be larger than other methods we converted the differences to absolute differences, as we wanted to analyse only the distance to the ground truth value and not if there was an overestimate or underestimate. For each method, we applied four Wilcoxon rank-sum two-tailed tests and ANOVAs on absolute differences, which allowed us to do a pair-wise comparison of methods (see Table 3.4 for Z-scores and significance).

The results indicate that the differences between the NB and NA method were significant and NB outperformed NA in this case; however the significance can be related to the large

Table 3.4: Z-scores reported by two tailed Wilcoxon rank-sum test between interpolation methods. The multiple testing problem was addressed using Bonferroni correction where $\alpha^{\prime}=\alpha / 4$. Significant values are indicated by a star, where p-value $<\alpha^{\prime}(\alpha=0.05)$.

|  | NA | NN | AM | DTA |
| :---: | :---: | :---: | :---: | :---: |
| NB | $-3.18^{*}$ | -1.51 | -1.30 | -1.43 |
| NA |  | 1.67 | 1.88 | 1.75 |
| NN |  |  | 0.21 | 0.07 |
| AM |  |  |  | -0.13 |

volume of data used in the test. There were no significant differences between any of the tests for the NN, AM and DTA methods in terms of absolute differences. The non-parametric ANOVA tests reported p-values close to zero (See Table 3.5) for all methods, which shows they all have their accuracy affected by the resolution of the annotated rainfall data ( $1 \mathrm{~km}, 2 \mathrm{~km}$ or 5 km , See Figure 3.4) and also by factors related to the individual stations, such as the rain gauge precision.

Table 3.5: $\chi$ values for non-parametric ANOVA tests. All p-values were smaller than 0.01.

|  | by resolution | by station |
| :---: | :---: | :---: |
| NB | 19.02 | 1101.55 |
| NA | 18.58 | 999.96 |
| NN | 20.03 | 1053.83 |
| AM | 19.36 | 1059.11 |
| DTA | 19.79 | 1056.69 |

The Wilcoxon rank-sum test between differences under different NIMROD coverages indicated that a coarser spatial resolution negatively influenced accuracy in all cases. The negative W values with significant p -values indicate that the differences tended to be larger in the second group than in the first group (See Table 3.6), i.e. the accuracy of all methods was decreased by the coarsening of the spatial resolution of the annotated data.

### 3.3.3 Computational cost of trajectory annotation methods

The time spent to perform TA increased with the use of more complex interpolation methods, Table 3.7 shows the computational complexity and how long the TA process took for each interpolation method for a database with approximately $140,000,00$ fixes. The computational

Table 3.6: W statistic for Wilcoxon rank-sum test performed between annotated NIMROD data from different spatial resolution, the value indicate that the difference increases for all methods with degradation of the spatial resolution of the annotated data. All p-values were smaller than 0.01 .

| Resolutions (km) | NB | NA | NN | AM | DTA |
| :---: | :---: | :---: | :---: | :---: | :---: |
| $1-2$ | -2.81 | -2.80 | -2.86 | -2.84 | -2.85 |
| $2-5$ | -2.40 | -2.35 | -2.49 | -2.43 | -2.48 |
| $1-5$ | -4.26 | -4.21 | -4.37 | -4.29 | -4.34 |

complexity is a measure of how many steps the algorithm will perform in the worst case for an input of a certain size. The number of steps is measured as a function of that size (Blakey, 2010). The complexities in our table are calculated for annotation and interpolation algorithms only and do not include pre-processing or intersecting raster and fixes.

Table 3.7: Computational metrics for our algorithms. We present the estimated computational cost in hours for DTA:CS (based on records updated/min) for our data set (140.000.000 GPS points) and real computational cost in hours for other methods. We further note the computational complexity of each algorithm, where $u$ is the number of trajectories and $n$ is the number of points in the longest trajectory, and description of the computational resources used.

| Algorithm | Computational <br> cost (hours) | Computational <br> complexity | Computational system |
| :---: | :---: | :---: | :---: |
| NB | 5 | $\mathrm{O}(\mathrm{un})$ |  |
| NA | 5 | $\mathrm{O}(\mathrm{un})$ | Intel® Xeon® CPU X5660 |
| NN | 24 | $\mathrm{O}(\mathrm{un})$ | $2.80 \mathrm{GHz}(2$ processors) |
| AM | 24 | $\mathrm{O}(\mathrm{un})$ | 192 GB RAM 12 cores |
| DTA:L | 720 | $\mathrm{O}(\mathrm{un})$ |  |
| DTA:CS | 5000 | $\mathrm{O}\left(\mathrm{un}^{3}\right)$ | Intel Xeon® E5-2686 v4 2.30 GHz <br> (Broadwell) processors |
|  |  |  | 160 GB RAM 40 cores |

The NB and NA method were the less computationally costly, which was expected due to the simplicity of the calculations behind them. The NN and AM method were approximately five times costlier then NB and NA whilst the DTA-L was 144 times costlier. The DTA-CS method was more than 6 times costlier than the DTA-L method, because it not only took more time but also it was calculated in parallel using forty cores, whilst the DTA-L was performed using twelve cores. Note also that the estimated running time to calculate the DTA-CS, even when parallelised onto forty cores, was 5000 hours which is approximately 7 months - too long
for any practical analysis. While introducing the possibility to use the spline interpolation in DTA theoretically, based on this running time we decided not to include it in our comparative analysis. The computational costs indicate whether one method is more suitable than other according to the size of the movement data, the infrastructure available to process the data and the time to be spent on the task.

### 3.4 Discussion

The aim of this chapter was to introduce and evaluate Dynamic Trajectory Annotation (DTA) methods that enable trajectories and environmental data to be combined to perform CAMA. We compared the DTA methods with commonly used interpolation methods (NB, NA, NN and AM) by annotating simulated trajectories with rainfall rates from meteorological radar data, computing accumulated rainfall values based on the annotated rates and comparing these values to accumulated precipitation from meteorological stations, which we used as ground truth.

The DTA generates cumulative curves more similar to the expected continuous smooth growth curve produced by a real mass curve of rainfall (Raghunath, 2006), which indicates that the DTA is a better model for representing how this rainfall is accumulated along a trajectory. Further, in terms of visual analysis the changes on rainfall rates were abruptly represented for all methods but the DTA method. Abrupt transitions and omission of peaks and valleys, which is present in the other four interpolation methods, may mask relationships between movement modes and rainfall, which might make the recognition of certain movement patterns harder. For example, if we were analysing movement that is heavily influenced by rainfall, such as people exercising outside, discovering the rainfall thresholds that drive the decision of going out or staying at home would be much harder with the annotated trajectories from NB, NA, NN and AM. The reason for this is that the other methods do not capture gradual increases and instead represent rainfall as a binary $0 / 1$ phenomenon at a 5 s rate. This may occasionally happen, however, from meteorological perspective, it is not typically the way that rain accumulates (Dirk, 2004).

Our visual analysis show that the cumulative curves generated by the NB, NA, NN and

AM methods are not in accordance with a continuous smooth growth curve, which is expected from an empirical real mass curve of rainfall derived from in-situ measurements and for which there is no pre-defined equation (Raghunath, 2006). The DTA curve is closer to a real mass curve of rainfall, which indicates that the DTA is a better model for representing how this rainfall is accumulated along a trajectory. Further, in terms of visual analysis the changes on rainfall rates were abruptly represented for all methods but the DTA method. The DTA method resulted in smoother transitions preserving peaks, borders, and showing the gradual increase of the rainfall, which is consistent with meteorological reality. The DTA:CS is likely to produce an even more realistic curve (Hutchinson, 1995), but with additional computational burden.

Our quantitative analyses show that all methods are likely to slightly underestimate the accumulated rainfall. We also found that NN, AM and DTA methods have higher accuracy than the other methods, but do not differ amongst themselves significantly. Further, all methods have their accuracy reduced by the coarsening of the spatial resolution of the environmental data used for annotation. Data with the best possible spatial resolution should therefore be used for performing CAMA and the accuracy could be further improved by using multiple sources of data.

There is no significant difference in accuracy between NN, AM and DTA methods but they perform better than NB and NA when calculating accumulated values, thus the use of one of the NN, AM and DTA is recommended when performing CAMA. The NN and AM methods have equally low computational cost and are less time consuming when compared to the DTA method. However, if a quicker and low computational cost method is the requirement we would not encourage the use of the AM method because it arbitrarily creates values that were not in the original data, without considering the behaviour of the variable in real life. It would be better to stick with the original data values in this case and use the NN method.

The DTA method was superior for modelling the rainfall mass curve and it was as accurate as the NN method, for this reason it is recommended for CAMA in scenarios where representing continuously varying phenomenon (like rainfall accumulation) is of high importance. Its capability of capturing gradual changes and preserving peaks and valleys from the original data
makes the DTA method a better choice when attempting to elicit the relationships between the environmental variable and fine-scale movement patterns. This is not restricted to the accumulation of rainfall, and the recommendation can be extended to any environmental variable whose behaviour between two temporal points can be approximated as a linear function over time. Other environmental variables may change differently over time and for these, the DTA can easily be extended from linear interpolation into more complex forms, for example $2^{\text {nd }}$ or $3^{\text {rd }}$ order polynomials, using an appropriate temporal function.

The application of DTA is particularly interesting for movement research within human mobility and wildlife ecology, where movement behaviour may be contextualised by other dynamic environmental variables such as air temperature, vegetation indices, humidity, wind speed, air pollution and snow coverage (Cagnacci et al., 2011; Horanont et al., 2013; Howarth and Hoffman, 1984; Phithakkitnukoon, Smoreda and Olivier, 2012; Siła-Nowicka et al., 2016). These variables are often obtained from remotely sensed data sources for which the temporal resolution can vary from five minutes to days and in most cases does not match the temporal resolution of movement data. In addition, it is common in movement research to simultaneously consider more than one environmental variable, which makes the choice of the interpolation method for trajectory annotation even more relevant. In such cases, CAMA would have to go beyond just integrating trajectories with one environmental variable: it would require the integration of several environmental variables amongst themselves to model how their interaction is influencing movement. We address this problem in the following two chapters.

In summary, the DTA method is suitable for the annotation of trajectories with continuous numeric environmental variables. The method is available as an open source Python script on GitHub and can be accessed through https://github.com/vsbrumb/DTA. We used DTA in a study of human behaviour (Chapter 5), to link GPS trajectories of commuters with weather data to explore how temperature, wind speed, humidity and 'feels like' temperature (i.e. the equivalent temperature perceived by humans, when the combined effect of air temperature, relative humidity and wind speed is taken into account) might influence commuting patterns. Future work could also include further methodological developments, where we extend the DTA for the case where trajectory fixes are sampled at a lower rate than environmental data, a case
that is common in movement ecology (Cagnacci et al., 2011; De Groeve et al., 2016; Dodge et al., 2014; Gaston et al., 2016).

## Chapter 4

## Seasonal response in the diet of

## maned wolves: A study using

multi-sensor image fusion and a sequence based behavioural analysis

### 4.1 Introduction

Combining contextual data from remote sensing imagery and GPS tracking data is possible through context-aware movement analysis (CAMA), a method for performing movement analysis which had a dramatic rise in the recent years. In the last ten years many studies have attested the potential of remotely sensed indicators for ecological research (Pettorelli et al., 2011; Neumann et al., 2015; Kerr and Ostrovsky, 2003; Turner et al., 2003), and in particular the use of NDVI (Normalized Difference Vegetation Index) (Pettorelli et al., 2005). NDVI (Rouse et al., 1973) is the most widely applied and calibrated vegetation index (Xue and Su , 2017). The widespread use of NDVI relates to its high correlation with vegetation productivity and capacity to detect seasonal dynamics (Hurley et al., 2014), sensitivity to canopy structure and photosynthetic activity (Xue and $\mathrm{Su}, 2017$ ). The index also correlates with forage quantity
and quality, which has made it extremely valuable for assessing habitat quality (Hurley et al., 2014).

The most commonly used NDVI datasets are obtained from SPOT (Satellite Pour l'Observation de la Terre), MODIS (Moderate Resolution Imaging Spectroradiometer) and AVHRR (Advanced Very High Resolution Radiometer). They are usually available at a temporal resolution of $10-15$ days and spatial resolution between 250 m and 8 km , which adds a lot of uncertainty to estimate changes on vegetation coverage. There are other satellites that provide higher spatial resolution data, however they have a low temporal resolution (Pettorelli et al., 2005) and usually do not offer pre-processed NDVI.

Up to now, the main strategies used to combine data from remote sensing and movement data were developed by Coyne and Godley (2005) and Dodge et al. (2013). These methods try to deal with spatial and temporal incompatibilities between contextual data and movement data by using a set of interpolation methods that are applied in space and time to estimate the contextual variables at unknown times (temporal resolution) and at the exact point where the GPS fix was collected (spatial resolution).

Despite representing progress in relation to the previous approaches, these methods disregard the structure of contextual data when performing spatial interpolation. There is a common misconception between pixel size and spatial resolution when it comes to raster data produced by remote sensing satellites. The pixel size changes when spatial interpolation is performed, but the spatial resolution cannot be changed as it is inherent and particular to each imaging system (See Figure 2.3 in Chapter 2). That said and knowing that the level of detail in a dataset is determined by its spatial resolution and not by its pixel size, interpolating in space to find a value at a very specific trajectory point brings no improvement, it actually adds noise from neighbouring pixels that are less related to the one in analysis. This is especially problematic when working with coarse resolution data, such a as MODIS dataset ( $0.5-1 \mathrm{~km}$ ). In this case it is definitely more appropriate to use the value of the pixel in which the point fell within, because that value is an integrated average of that area, which is more likely to be highly correlated with the point we need to measure than neighbouring pixels.

To overcome these spatio-temporal incompatibilities we propose a multi-source disaggre-
gation approach to produce better quality and more accurate contextual datasets for CAMA. That is, we obtain NDVI data from several satellites with varying spatial and temporal resolution and design a new methodology to create a fused NDVI data set with increased temporal resolution and level of detail for CAMA. This approach has not been attempted before in movement research and the difference between the usual single source approach and ours is shown in Figure 4.1. Our approach consists of using daily MODIS images with a 250 m spatial resolution and a finer image (with 15 to 30 m spatial resolution) collected fortnightly to derive daily images with the same spatial resolution as the finer source images. The purpose of this approach is to create a new disaggregated NDVI data set, which can then be linked to movement trajectories to detect finer scale movement patterns related to contextual changes. We have tested this methodology in a case study of maned wolves (Chrysocyon brachyurus), the largest south American canids which live in the Brazilian Cerrado. Our new methodology can help explore how change in vegetation leads to dietary changes in manned wolves diet and how these changes are reflected as different movement patterns that vary with seasons and availability of vegetation. Maned wolves are omnivorous, their diet is mostly guided by the availability of prey and vegetation, therefore we expect that changes in NDVI, i.e., changes in food availability from vegetation, may appear as changes in movement patterns that we can detect with our proposed methodology.

Our new methodology is a three-step process. In Step 1, we use daily MODIS coarse NDVI and higher resolution satellite images from seven other sensors to produce daily NDVI data at higher level of detail and 15 m pixel size. We do this by adapting a downscaling method proposed by Rao et al. (2015), in which higher resolution NDVI and land cover classification are used to obtain a temporal NDVI series with higher spatial detail, i.e., at MODIS subpixel level. In Step 2, we use the daily finer detailed NDVI time series to annotate maned wolves' trajectories. We first explore the resulting semantic trajectories by looking at the distribution of NDVI within home ranges compared to the distribution of NDVI values for the GPS locations. In Step 3, we use a sequence-based method, the so-called Eigenbehaviours (Eagle and Pentland, 2009) to identify seasonal patterns in the feeding habits of maned wolves during the study period. We compare these behaviour with already known habitat preferences


Figure 4.1: Comparison between different approaches for obtaining contextual data on one variable. Black lines represent trajectories and yellow circles represent fixes in the trajectory. Horizontal parallelograms with the contextual data and its pixels with different values. A) Contextual data from only one source. B) Contextual data from multiple sources, the time interval between datasets is lower. C) Contextual data from multiple sources and disaggregated. The time interval between datasets is lower and the data sets with finer spatial resolution are used to add details to the coarser datasets. The green grid represents the new pixels of the disaggregated products.
and known biology of the species.

### 4.1.1 Feeding ecology of maned wolves

Maned wolves, the largest South American canids (de Paula and Desbiez, 2014), are savannahadapted omnivores found south of the Amazon Forest. More specifically, their range extends from Bolivia into eastern Brazil, through northern Argentina and Uruguay, to central Paraguay (Deem and Emmons, 2005) (Figure 4.2A). Considered "vulnerable" until 1996 by IUCN (International Union for Conservation of Nature), the species is currently classified as "near threatened" by IUCN (de Paula and DeMatteo, 2015) and it is still considered "vulnerable" by the Brazilian environmental authorities (ICMBio, 2016).

The main threat to the species comes from the continuous large scale habitat losses (Noss and Lima, 2007), which are especially significant in Brazil because of the extensive conversion of Cerrado (Brazilian savannah) into farmland (Fonseca et al., 1994). Only $20 \%$ of Cerrado


Figure 4.2: A) The range of the maned wolf (Chrysocyon brachyurus) in South America. B) Borders of Serra da Canastra National Park (CNP) in Minas Gerais state in Brazil, home of the wolves whose tracking data are used in this study. C) Lobinha "Pup", a female maned wolf of approximately two years old, wearing a GPS tracking collar. Lobinha was rescued from a sugar cane crop by the Chico Mendes Institute for Biodiversity Conservation - Brazilian Ministry of the Environment (ICMBio) and taken to an enclosure, where this picture was taken, for re-adaptation prior to reintroduction to the wilderness.
is still covered by native vegetation (Myers et al., 2000) and less than $2.5 \%$ of it is protected by law (Klink et al., 2005). One of the protected areas, the Serra da Canastra National Park (CNP) (Figure 4.2B) has been key to the preservation of maned wolves.

The CNP is the headquarters of the project "Behavioural biology and conservation of the maned wolf (Chrysocyon brachyurus) in the Cerrado of Minas Gerais" founded in 2004 by ten research institutes as a joint effort to protect the species. The extensive conversion of the park's surroundings into farmland has exposed the wolves to many anthropogenic threats, such as road traffic, culling and disease contamination by domestic animals (Deem and Emmons, 2005), all of which can result in large fluctuations in population size, eventually leading to extinction (de Paula and Desbiez, 2014).

Maned wolves are the main actors in dispersal of various fruits, including the ones endemic
to Cerrado. They provide population and disease control by preying on insects and rats that are vector of diseases, such as hantavirus infection and leptospirosis (Consorte-McCrea and Santos, 2013). They are a keystone and umbrella species (de Paula et al., 2013), i.e., they have a disproportionate effect on their environment and surrounding organisms relative to their population size (Mills and Doak, 1993) and play a crucial role in maintaining the structure of the ecological community (Roberge and Angelstam, 2004). The protection of a umbrella species indirectly protects many other species in the ecological community (Ray, 2005). For this reason maned wolves are often the targeted species for conservation-related decisions. In addition, they are an important part of the local cultural heritage and can, with adequate planning, become a tourist attraction (Consorte-McCrea and Santos, 2013). This may boost the local economy and generate more motivation to protect the ecosystem.

The indifference, sometimes even antipathy (Consorte-McCrea and Santos, 2013), towards the species is erroneously founded on the idea that their diet mostly consists of poultry or livestock (Motta-Junior et al., 2014). Regardless of maned wolf's preference for wolf's fruit (Solanum lycocarpum), miscellaneous fruits and rodents (Queirolo and Motta-Junior, 2007), the pressures of habitat impoverishment caused by anthropogenic presence might encourage individuals to venture into human-occupied areas, which leads to conflict and higher mortality (Motta-Junior et al., 2014). Despite the existence of a few studies on the species, the general knowledge about its ecology and population dynamics is still insufficient to secure preservation, which could be done through conservation of habitat, especially in extremely altered landscapes (de Paula et al., 2008) such as Cerrado.

The cornerstone of any preservation effort is knowing what must be preserved (Rodrigues et al., 2014). Therefore, understanding feeding ecology and related spatial patterns is of prime importance, in particular for a species with a seasonal variation of dietary composition (MottaJunior et al., 2014). Most studies on feeding habits of maned wolves were performed in captive population. Only a couple have been done in the wilderness and those are mostly based on faecal and gastric analysis (Motta-Junior et al., 2014). These studies describe the diet of maned wolves as largely based on fruits and small mammals, the proportions of which vary according to seasonal fluctuations in food availability (Motta-Junior et al., 2014).

Wolf's fruit and small mammals are more abundant in the dry season, whilst miscellaneous fruits and arthropods are plenty in the wet season (Motta-Junior et al., 2014; Bueno and Motta-Junior, 2009). In the CNP, our study area, a previous study by Queirolo and MottaJunior (2007) on 399 faeces samples indicates the following dietary breakdown: $29.72 \%$ from miscellaneous fruits, $22.02 \%$ from rodents, $12.61 \%$ from birds, $9.89 \%$ from wolf's fruit, $9.83 \%$ from grasses and fruit, $7.40 \%$ from lizards, $4.32 \%$ from insects and arthropods, $1.78 \%$ from snakes, $1.42 \%$ from opossums, $0.89 \%$ from armadillos and $0.28 \%$ from other medium and small mammals.

Overall, little is known about maned wolves diet in the wild and research on the impact of environmental changes is essential to identify how fitness and reproduction are affected by diet composition (Bueno and Motta-Junior, 2009; de Paula et al., 2013). In particular, the diet variability has been previously studied in terms of the percentage of food types, but its spatial and temporal patterns have not been explored. The seasonal pattern found in faecal samples collected by a few studies seems to indicate a diet with a temporal trophic opportunistic pattern (Motta-Junior et al., 2014).

We propose that the existence of a temporal trophic pattern can be investigated by looking at the NDVI (Normalized Difference Vegetation Index) utilisation distribution within the home ranges versus the surrounding areas. NDVI is a normalised ratio between the spectral reflectance in the red and near infra-red portions of the electromagnetic spectrum, which is a proxy for the content and state of live green vegetation. It is an indicator of vegetation vigour, and is often used as a surrogate estimator of net primary productivity (NPP) (Xu et al., 2012). NPP is the difference between the rate at which plants in an ecosystem assimilate carbon to produce energy and the rate at which they use part of this energy. It determines the amount of energy available to be transferred from the plants to other trophic levels in the ecosystem (Chapin et al., 2011). In other words, it directly determines the amount of food available for the consumption of herbivores and omnivores, and consequently indirectly to the carnivores. Thus, as maned wolves are omnivorous, NDVI can be used as an indicator of the resources available for their maintenance in a given environment (Bueno and Motta-Junior, 2009). In addition, when combined with GPS tracking data NDVI can provide insights into the seasonal
changes of dietary composition and if those are reflected as specific spatial patterns.

### 4.2 Methodology

To study the influence of vegetation state and content on seasonal patterns of maned wolves' diet we used a three-step process (Figure 4.3). In Step 1, we used daily MODIS coarse NDVI and higher resolution satellite images from seven other sensors to produce daily NDVI data at a higher level of detail and 15 m pixel size. We did this by adapting a downscaling method proposed by Rao et al. (2015), in which higher resolution NDVI and land cover classification are used to obtain a NDVI temporal series with a higher spatial detail, i.e., at MODIS sub-pixel level. First we performed the absolute calibration of satellite data to guarantee consistency amongst the measurements from different satellites. Then we extracted NDVI growth rates from MODIS and performed land cover classification on the finer resolution data, so that we could compute land cover fractions within a MODIS pixel. The land cover fraction and the daily NDVI growth rates from MODIS were used to calculate the NDVI growth rate for each land cover fraction, which was then applied to the finer NDVI data to generate a time series of daily NDVI with higher level of spatial detail. In Step 2, we use the daily finer detailed NDVI time series to annotate maned wolves' trajectories. We explored these annotated semantic trajectories by looking at the distribution of NDVI within home ranges compared to the distribution of NDVI values for the GPS locations. In step 3 we transformed trajectories into temporal sequences that described how each individual is making use of the vegetation within their home range. These sequences are then analysed using the so-called eigenbehaviours (Eagle and Pentland, 2009), a method that finds typical movement patterns in sequential data. This allowed us to identify seasonal structures in the spatial patterns in the feeding habits of maned wolves.

### 4.2.1 Movement data and study area

We analysed a wildlife movement dataset generated by GPS collars attached to 13 maned wolves (Table 4.1 and Figure 4.4), more specifically the models Pinnacle Lite G5C 275D produced by Sirtrac, 3300S and Iridium Track 1D produced by Lotek Wireless Inc (de Paula, 2016). The data were collected by our collaborators in the National Carnivorous Mammals Research and Conservation Center (CENAP) at the Chico mendes Institute for Biodiversity Conservation(ICMBio). The dataset, collected between March 2007 and July 2015, has the highest number of total GPS locations (54.196) recorded for this species in the wilderness (de Paula, 2016). The tracking data included seven females and six males, with the tracking period varying from 59 to 841 days (Table 4.1).

Table 4.1: Description of tracked individuals, respective sampling rate, duration, start and end of monitoring period. Names with the same colour indicates a couple (maned wolves take only one mate for life). White indicates the absence of a mate in the tracking dataset. Source: Adapted from de Paula (2016).

| Name | Sex | Weight <br> $(\mathrm{kg})$ | Age at ${ }^{\text {st }}$ <br> capture | Sampling <br> rate (h) | Monitored <br> days | GPS <br> points | Start <br> date | End <br> date |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Amadeu | M | 29 | 6 | $1-6$ | 461 | 3514 | $20 / 03 / 2007$ | $20 / 10 / 2008$ |
| Lais | F | 31 | 6 | $0.4-6$ | 433 | 3726 | $03 / 05 / 2007$ | $28 / 07 / 2008$ |
| Bolt | M | 34 | 6 | 1 | 547 | 12917 | $01 / 10 / 2013$ | $31 / 03 / 2015$ |
| Rose | F | 27 | 4 | 1 | 356 | 8527 | $20 / 07 / 2014$ | $10 / 07 / 2015$ |
| Gamba | M | 30 | 6 | 4 | 98 | 562 | $30 / 04 / 2009$ | $05 / 08 / 2009$ |
| Tay | F | 32 | 7 | 2 | 841 | 8760 | $14 / 03 / 2007$ | $05 / 12 / 2009$ |
| Miro | M | 30 | 7 | $2-5$ | 382 | 2687 | $25 / 03 / 2011$ | $29 / 02 / 2012$ |
| Luna | F | 30 | 7 | 1 | 295 | 5775 | $28 / 02 / 2013$ | $19 / 11 / 2013$ |
| Samurai | M | 32 | 4 | 4 | 86 | 496 | $26 / 08 / 2009$ | $19 / 11 / 2009$ |
| Jurema | F | 28 | 3 | 4 | 510 | 2846 | $03 / 09 / 2009$ | $25 / 01 / 2011$ |
| Henry | M | 27 | 3 | 4 | 250 | 1451 | $12 / 05 / 2010$ | $22 / 01 / 2011$ |
| Loba | F | 27 | 3 | $3-11$ | 563 | 2317 | $15 / 06 / 2012$ | $29 / 12 / 2013$ |
| Nilde | F | 27 | 7 | 2 | 59 | 618 | $30 / 03 / 2011$ | $27 / 05 / 2011$ |

Tracking data are part of the project "Behavioural biology and conservation of the maned wolf (Chrysocyon brachyurus)" and were collected in the Canastra National Park (CNP) and its surroundings. Figure 4.4 shows the study area, land uses, the boundary of the CNP and the individual home ranges derived from tracking data. Home ranges (HR) are the areas used by an individual during its normal activities for foraging, mating and rearing (Burt, 1943). We defined them by delineating the $95 \%$ utilisation distribution (UD) for each individual (Hayne,

1949; Mohr, 1947) (Figure 4.4), where UD was calculated as a kernel density surface. The overlay of HRs and land uses show that only two individuals stay completely within the CNP, five transit between the CNP and its surrounding areas, and the remaining six are based outside the park in landscapes extremely altered by humans, mainly farmlands (Figure 4.4).

### 4.2.2 Contextual data

NDVI is a proxy for the content and state of the live green vegetation and its computation requires information on the spectral reflectance in the red and near infra-red portions of the electromagnetic spectrum, as shown in Equation 4.1 from Rouse et al. (1973).

$$
\begin{equation*}
N D V I=\frac{N I R_{\rho}-\operatorname{Red}_{\rho}}{N I R_{\rho}+\operatorname{Red}_{\rho}} \tag{4.1}
\end{equation*}
$$

Here $N I R_{\rho}$ is the reflectance in near infra-red interval $(800-1000 \mathrm{~nm})^{1}$ and $\operatorname{Red}_{\rho}$ is the reflectance in the red interval ( $650-700 \mathrm{~nm})^{1}$.

NDVI values range from -1 to 1 . Values smaller than 0.1 are usually related to bare rocks, sand, or snow; values around 0.2 to 0.5 are related to sparse vegetation such as shrub, grasslands or senescence crops; values between 0.6 and 1.0 correspond to dense vegetation, such as tropical forests or crops at their peak growth stage (Rouse et al., 1973; Hurley et al., 2014; Xue and Su, 2017). Information on NDVI can be obtained through ready-to-use satellite products, such as the daily NDVI data from MODIS (Moderate Resolution Imaging Spectroradiometer), or by processing raw satellite images to obtain spectral reflectance bands in the red and NIR ranges and then compute the NDVI.

Most ecological studies have used daily pre-processed NDVI (Pettorelli et al., 2011) from MODIS, which has a coarse spatial resolution (Table 4.2 and Figure 4.5). In our approach we integrate MODIS NDVI and higher spatial resolution satellite data in order to minimise information loss and maximise data availability for performing CAMA. We used the daily reflectance (MOD09) and cloud masks (MOD33) from Terra - MODIS and raw images from

[^0]Terra - ASTER (Advanced Spaceborne Thermal Emission and Reflection Radiometer), Landsat 4-5-7-8, CBERS 2 (China-Brazil Earth Resources Satellite) and CBERS 2B. Satellite data are described in detail in Table 4.2, Figure 4.5 shows the distribution of images over the tracking period and illustrates the differences in spatial and temporal resolutions (Figure 4.5).

Table 4.2: Description of satellite images used to produce NDVI. The letters along with spatial resolutions are specifying the spectral bands for sensors with multiples spatial resolutions, sensors without letters have a uniform spatial resolution along the spectral bands. R stands for red, G for green, B for blue, NIR for near infra-red, VNIR for visible (RGB) and near infra-red, SWIR for short wavelength infra-red, FI for far infra-red and TIR for thermal infra-red.

| Satellite | Sensor | Temporal resolution (days) | Spatial <br> resolution (m) |
| :---: | :---: | :---: | :---: |
| CBERS 2-2B | High Resolution |  |  |
|  | CCD Camera (HRCC) | 26 | 20 |
| Landsat 4-5 | Thematic Mapper (TM) | 16 | 30 |
| Landsat 7 | Enhanced Thematic | 16 | 30 |
|  | Mapper plus (ETM+) |  |  |
| Landsat 8 | Operational Land Imager (OLI) | 16 | 30 |
|  | Moderate Resolution |  | 250 (R/NIR) |
| Terra | Imaging | 1-2 | 500 (B/G/SWIR) |
|  | Spectroradiometer (MODIS) |  | 1000 (VNIR/FI) |
|  | Advanced Spaceborne |  | 15(VNIR) |
| Terra | Thermal Emission and | 16 | 30 (SWIR) |
|  | Reflection Radiometer (ASTER) |  | 90 (TIR) |

### 4.2.2.1 Absolute calibration of satellite data

Satellite images are most often provided in digital numbers (DN), which represent the intensity of the pixel at each scanned location (Jensen, 2006). These values are related to the spectral solar radiance reflected by the surface and measured by the sensor (Rees, 2001; Novo, 2010). As opposed to DNs, the spectral solar radiance is a physical quantity that can provide valuable information on the state of the imaged surface (Jensen, 2006) and can also be used for intercalibrating temporal multi-source data series. Inter-calibration is done by converting spectral radiance to spectral reflectance, which decreases the bias introduced by different atmospheric conditions and illumination (Rees, 2001) and produces more accurate results when applying change detection algorithms to temporal series (Ponzoni and Shimabukuro, 2010). In other words, a DN of five, for example, does not hold the same meaning for images from different satellites, not even for bands of the same satellite and often not even for images and bands of the same satellite on different dates. Spectral reflectance, however, is a physical property inherent to the object being measured, holding the same meaning across satellites and is therefore the appropriate quantity to use for keeping the consistency and accuracy of measurements in a multi-source methodology.

The daily spectral reflectance dataset from MODIS (MOD09) has already gone through absolute calibration prior to being released as a product. Thus, we only had to calibrate the 35 images from the other seven satellites, which was done using sensor-specific scaling parameters and equations that are provided in the meta data for each image and user's handbook of each sensor. Following this step, we resampled all the finer resolution images, i.e. all except MODIS, to 15 m in order to match the finest spatial resolution in our datasets and keep the maximum information on the land cover. We used the nearest neighbour method to perform the resampling, which does not create new values that were not in the original data (Meneses and Almeida, 2012). This is important when working with biophysical parameters to preserve the relationship between what was measured on the ground by the satellite and the biophysical variable being analysed. After absolute calibration the calibrated spectral reflectance bands were used to computed the fine resolution NDVI, NDWI (Normalized Difference Water Index)
(Gao, 1996) and NDBI (Normalized Difference Built-Up Index)(Zha et al., 2003).

### 4.2.2.2 Land cover classification and calculation of land cover fractions

The land cover classification on the finer scale must be determined in advance so that the land cover fractions and the daily NDVI growth rate by land cover can be calculated (Figure 4.3). Previous studies have used high spatial-resolution land cover maps to obtain the fractions, however, land cover maps are not always available. In such cases, the land cover map can be obtained by unsupervised classification of all finer resolution images available. Unsupervised classification is recommended because it achieves relatively accurate classification results and keeps the method as automated as possible (Rao et al., 2015). in addition, the use of multiple land cover maps is important to account for changes in land cover during the study period.

Land cover classification was performed on the resampled spectral reflectance, NDWI, NDVI and NDBI bands from the finer resolution sensors, i.e., all except MODIS. The inclusion of radiometric indices often improves the accuracy of the classification, because these indices highlight the differences between land covers (Jensen, 2006; Ponzoni and Shimabukuro, 2010; Novo, 2010). We performed automatic non-supervised classification by using the BIRCH (Balanced Iterative Reducing and Clustering using Hierarchies) (Zhang et al., 1996) clustering algorithm, which is a memory efficient method optimised for large datasets (Xu and Wunsch, 2009).

BIRCH uses the summary statistics of the original dataset instead of the entire data to identify the clusters, greatly reducing the need for storage and memory ( Xu and Wunsch, 2009). The basic representation unit for a cluster in BIRCH is the cluster feature (CF), a tuple of three elements: the number of data points in the cluster, the linear sum and the square sum of the data points (Zhang et al., 1996). The CFs are stored in a a tree, the CF tree, which is built dynamically during the clustering process ( Xu and Wunsch, 2009). The algorithm requires three input parameters: 1) a threshold for the maximum allowed radius for the cluster resulting from the grouping of a sub cluster and the closest sample, the cluster is partitioned if the radius is bigger than the chosen threshold; a branching factor that determines the maximum number of sub clusters in each node, if a new sample is added such that the number of sub clusters exceeds the branching factor the node is split into two; and the number
of clusters after the final clustering step, which treats the sub clusters from the leaves as new samples.

We used a branching factor of 100 for all images, varied the threshold from 0.2 to 0.7 by 0.02 for each image and set the number of clusters to None, which means keeping all sub clusters, i.e., not pruning the tree. We chose to keep all sub clusters and a higher branching factor because having different land covers mixed in the same cluster would introduce errors in our analysis. The threshold range and increment was heuristically chosen after performing a few tests on one of the images: for each threshold we computed the Calinski-Harabaz Index (CHI) (Calinski and Harabasz, 1974) to measure the intra and inter quality of our partitions, i.e., land cover classes, and for each image we kept the threshold that achieved the best performance, i.e., lowest within-cluster dispersion and highest between-cluster dispersion. Finally, for each of the 35 land cover classifications ( 15 m ) we calculated the percentage of each land cover within the correspondent MODIS pixel ( 250 m ) for each pixel, which produced 35 maps of land cover fractions for our study period. The use of multiple land cover fractions maps is important to account for possible changes in the land cover within a MODIS pixel between acquisition of the finer resolution data.

### 4.2.3 MODIS NDVI growth rate extraction and filtering

NDVI growth rate is a value that represents the change in vegetation for each pixel in a image between two images captured at different days. The extraction of daily NDVI growth rates from MODIS images allows the creation of a temporal series describing the vegetation dynamics within each coarse pixel. The information on the daily dynamics of vegetation can then be broken-down according to the distribution of different land covers within that pixel, which comes from the classification of the finer resolution images. Prior to using MODIS NDVI data however, these need to be filtered. The filtering is important to reduce the impact of outliers cause by the presence of clouds missed during the masking of images, occasional saturation of detectors leading to overestimation of NDVI and other factors that can bias the growth rates.

MODIS data have been previously used for extracting growth rates and understanding vegetation dynamics (Lu et al., 2015; Rao et al., 2015; Eckert et al., 2015). Their high temporal
resolution allows daily monitoring of changes in vegetation and NDVI growth rates were found to be well represented by a linear function when constrained to such a short time-spam (Rao et al., 2015). However, the extraction of accurate NDVI growth rates requires a rigorous filtering process to de-noise the data series, i.e., to reduce the known interference of clouds, atmosphere dynamics, variability on the detectors that register reflectance and other factors. Wavelet transform (WT) is specially efficient in identifying and reducing noise while preserving useful information in time-series (Lu et al., 2007) and it has been widely used in the extraction of vegetation patterns via radiometric indices (Sakamoto et al., 2005; Priyadarshi et al., 2017; Leguizamon et al., 2000; Epinat et al., 2001; Dutta, 2012).

The basic idea behind WT is that a chosen finite function, the mother wavelet (MW), will be translated and dilated as a base for expanding other functions (Grossmann and Morlet, 1984), i.e., estimating the closest possible function based on the signal samples. A wavelet is a zero mean finite rapidly decaying wave like oscillation able to capture acute changes in signal. There are wavelets of various shapes and the choice of a MW must be done in accordance with the application. WT filters signals in the time-frequency domain by decomposing the noisy signal, deriving the wavelet coefficients and using the inverse WT to obtain the modified coefficients and deliver the de-noised signal (Priyadarshi et al., 2017).

There are two main coefficients in WT, scaling and shifting, the first refers to stretching or shrinking the MW in time to fit the signal samples, while the second refers to delaying or advancing the MW to align with the signal samples. Stretching helps capturing the slowly varying changes and compression captures rapid changes. WT allows keeping both levels of detail by using multiple scale decomposition of the noisy signal.

We used the 3152 MODIS NDVI images to create a temporal profile of NDVI for each one of the pixels in our study area. For this we first used the cloud mask product (MOD35) (Strabala, 2018) corresponding to each MODIS NDVI image to remove pixels contaminated by clouds. In the next step we converted the images into time-series of NDVI values, one timeseries for each pixel, and applied two consecutive WT using the Daubechies 4 (Daubechies, 1990) MW. This MW has been extensively used for de-noising (Leguizamon et al., 2000) and it is commonly used for NDVI (Kaddar et al., 2017). We performed a four level soft threshold

WT of the NDVI temporal series for each pixel (3152 samples/pixel) in the MODIS data, then we reconstructed the series, then repeated the procedure once again to obtain the final filtered NDVI pixel series. The NDVI daily growth rates were then computed for each pair of MODIS images by calculating the first derivative for each filtered NDVI pixel time series.

### 4.2.4 Calculating NDVI growth rates for land cover fractions to generate the finer NDVI time-series (ENDVI)

The filtered NDVI daily growth rates extracted from MODIS reflect the average of the vegetation dynamics within the $62500 \mathrm{~m}^{2}$ covered by each pixel. In order to obtain more detailed information on the vegetation dynamics within each pixel, we calculated the contribution of each land cover fraction to the growth rates (Figure 4.6). This is possible by using the 35 land cover classification based on the finer images. The contribution of each land cover fraction is then applied to the finer NDVI data to generate daily ENDVI images (Figure 4.6). In the next step, we plotted the ENDVI values for pixels of coffee crops, which have well defined seasons, to evaluate if the ENDVI was reporting the vegetation changes within growing and harvesting times, which in the absence of ground truth data, gave us a proxy of the quality of our ENDVI estimates.

This section is an adapted summary of the methods for producing finer spatial-resolution NDVI time series presented by Rao et al. (2015). Despite the complex vegetation dynamics driving NDVI changes, short-term changes in NDVI values can be interpreted as linear. Between two timestamps this can be expressed as Equation 4.2.

$$
\begin{equation*}
N D V I_{t_{n+1}}=N D V I_{t_{n}}+k\left(t_{n+1}-t_{n}\right) \tag{4.2}
\end{equation*}
$$

Here $N D V I_{t_{n}}$ is the NDVI value at the date $t_{n}, N D V I_{t_{n+1}}$ is the NDVI value at the date $t_{n+1}$, and $k$ is the corresponding growth rate between $t_{n+1}$ and $t_{n}$. This equation allows the prediction of NDVI on a given date by using the growth rate in that particular period and one NDVI observation as baseline. The NDVI growth rate between any two dates can be calculated using MODIS NDVI time-series data, derived for example from the MODIS daily reflectance
product (MOD09). However, estimating NDVI growth rates for finer scale pixels from the growth rate at the MODIS pixel is a more complex issue that must be solved in order to obtain a finer image-like NDVI prediction. Following Rao et al. (2015) we employed a linear mixing model (LMM) to solve the issue (Equation 4.3).

$$
\begin{equation*}
N D V I(x, y, t)=\sum_{c=1}^{n} f_{c}(x, y, t) * N D V I_{c}(x, y, t)+\varepsilon(x, y, t) \tag{4.3}
\end{equation*}
$$

Here $N D V I(x, y, t)$ is the NDVI value of a MODIS pixel $(x, y)$ at time $t, f_{c}(x, y, t)$ and $N D V I_{c}(x, y, t)$ are respectively the fractional coverage and the NDVI value of land cover class $c$ within the MODIS pixel $(x, y)$ at time $t, n$ is the total number of land cover classes within the MODIS pixel $(x, y)$ and $\varepsilon(x, y, t)$ is the error introduced by LMM. Considering Equation 4.2, Equation 4.3 and under the assumption that no land cover changes occur between $t_{1}$ and $t_{2}$, we can write the relationship between the growth rate of a MODIS pixel and the corresponding finer pixels as Equation 4.4.

$$
\begin{equation*}
k^{M O D I S}\left(x, y, t_{1} \rightarrow t_{2}\right)=\sum_{c=1}^{n} f_{c}\left(x, y, t_{1}\right) * k_{c}^{F I N E R}\left(x, y, t_{1} \rightarrow t_{2}\right) \tag{4.4}
\end{equation*}
$$

Here $k^{\operatorname{MODIS}}\left(x, y, t_{1} \rightarrow t_{2}\right)$ is the growth rate of a MODIS pixel $(x, y)$ from $t_{1}$ to $t_{2}$, $k_{c}^{F I N E R}\left(x, y, t_{1} \rightarrow t_{2}\right)$ is the growth rate of land cover $c$ on the finer pixel scale within the corresponding MODIS pixel from $t_{1}$ to $t_{2}$.

There are several land cover classes within a MODIS pixel (Figure 4.7), which means that the estimate of the unknown parameters, and therefore the computation of the finer scale NDVI growth rates, requires at least $n$ MODIS pixels (Rao et al., 2015) (Figure 4.7). These MODIS pixels are arranged in a system of equations, so that there are at least $n$ equations to find the unknown parameters and the growth rate for each land cover type.

The estimate growth rates by land cover fractions are then computed by solving a linear system of $n+1$ equations, where each line express the combination of land covers and respective
coarse NDVI growth rate for a MODIS pixel neighbouring the target MODIS pixel being downscaled (Figure 4.7). The neighbours for each downscale were selected by their centroid to centroid distance to the target pixel (Figure 4.7) and the system was solved using a constrained least square method with upper and lower boundaries, as shown in Equation 4.5.

$$
\begin{equation*}
k_{\min }^{M O D I S}-S t D\left(k^{M O D I S}\right) \leq k_{c}^{F I N E R} \leq k_{\max }^{M O D I S}+S t D\left(k^{M O D I S}\right) \tag{4.5}
\end{equation*}
$$

Here $S t D\left(k^{M O D I S}\right), k_{\min }^{M O D I S}$ and $k_{\max }^{M O D I S}$ are the standard deviation, the minimum and the maximum values of the MODIS growth rate for the entire study area at the given date. The constraints are used to avoid unreasonable rates retrieved by Equation 4.5 that might be caused by classification errors and noise in the time-series (Rao et al., 2015). Once the LMM is solved and the estimate growth rates of all land cover classes are calculated at the finer spatial scale, Equation 4.2 can be used to predict the finer NDVI value at $t_{2}$ (Equation 4.6).

$$
\begin{equation*}
E N D V I_{t_{n+1}}=N D V I_{t_{n}}^{F I N E R}+k_{n}\left(t_{n+1}-t_{n}\right) \tag{4.6}
\end{equation*}
$$

Here $E N D V I_{t_{n+1}}$ is the disaggregated NDVI value at the date $t_{n}$ at the finer resolution, $N D V I_{t_{n+1}}^{F I N E R}$ is the NDVI value at the date $t_{n+1}$ at the finer resolution image, and $k_{n}$ is the corresponding growth rate between $t_{n+1}$ and $t_{n}$ for the land cover class $n$.

We evaluated the ENDVI time-series by selecting points in the study area, classifying them in accordance to the detailed land cover map (Figure 4.4), extracting their ENDVI temporal profile and comparing with what would be expected from that specific land cover type. In a crop of coffee, for example, NDVI values are expected to fall in September when coffee beans are planted and increase after that when the growing season starts.

### 4.2.5 Context integration and seasonal plots

Movement data, MODIS NDVI and the ENDVI contextual data were integrated using trajectory annotation. This method links GPS fixes and contextual information according to their temporal and spatial coordinates. We applied the nearest neighbour trajectory annotation which intersects in space the contextual layer which is the closest in time for each fix to find the value of the contextual variable at the time when the fix was collected (Brum-Bastos et al., 2016). We also annotated individual home ranges, derived from the GPS points collected during the entire period for each wolf (Figure 4.4), with the daily distribution of MODIS NDVI and ENDVI, i.e., we assigned all the NDVI and ENDVI values within the home range to this particular home range. This allowed us to compare the vegetation available within the home range to the vegetation visited by the wolf.

We computed daily skew and kurtosis for the MODIS NDVI and ENDVI distributions within each home range, which we combined to generate a typology of NDVI distributions to characterize the state and content of vegetation within the home ranges, so that we could infer the likelihood of occurrence for higher and lower NDVI values from their empirical density distributions (EDD). The skewness is a measure of asymmetry relative to the mean, a distribution can be right-skewed (skew $>2$ ) or left-skewed (skew $<-2$ ). A positive skew indicates that the right tail of the distribution is longer and the mass of the distribution is concentrated to the left of the mean (Figure 4.8 A ). A negative skew indicates that the left tail of the distribution is longer and the mass of the distribution is concentrated to the right of the mean (Figure 4.8 C). The kurtosis is a measure of the weight of the tail of the distribution relative to a normal distribution, a distribution can be light-tailed/peaked (kurtosis $>2$ ) or heavily-tailed/flat (kurtosis $<-2$ ). A positive kurtosis indicates that the probability is low at the tails, i.e., less frequent outliers, and the mass of the distribution is concentrated around the mean (Figure 4.8 D). A negative kurtosis indicates that the probability is higher at the tails, i.e., more frequent outliers, and the mass of the distribution is more dispersed around the mean (Figure 4.8 F). Distributions with skew and kurtosis between -2 and 2 are considered within the range of normality (George and Mallery, 2010) (Figure 4.8 B and E).

In comparison to a normal distribution of vegetation state and content, a positive skew indicates a home range where the occurrence of dense and healthy vegetation (higher NDVI values) is less likely and a negative skew indicates a home range within which the occurrence of sparse vegetation and other land covers (lower NDVI values) is more likely. Similarly, in comparison to a normal distribution a positive kurtosis indicates a home range within which the land cover is more homogeneous (NDVI values closer to the mean) and a negative kurtosis indicates a home range within which the land cover is more heterogeneous (NDVI values more dispersed).

Based on the ranges of skew and kurtosis (Figure 4.8) we defined nine types of NDVI distributions to characterize the state and content of vegetation within the home ranges, which we named according to the following convention: the first character in the name of the distribution refers to the skewness, it can be "L" for low (skew > 2), "H" for high (skew <-2) or "N" for "normal" $(-2 \leq$ skew $\geq 2)$. The character after the dash refers to the kurtosis of the distribution, it can be "P" for "peaked" (kurtosis > 2), "F" for "flattened" (kurtosis < 2) or absent for when it is similar to the normal distribution ( $-2 \leq$ kurtosis $\geq 2$ ). The second and third character are always "NH" and stand for "NDVI in the Home range". The nine types of NDVI distribution are illustrated in Figure 4.9 and described here in terms of vegetation state and content:

- LNH-P (Low NDVI home range - Peaked): Homogeneous home range with predominantly lower NDVI values, i.e., predominance of barren rocks, shrub, grasslands and presence of very few patches of dense vegetation.
- LNH-F (Low NDVI home range - Flattened): Heterogeneous home range with predominantly lower NDVI values, i.e., predominance of barren rocks, shrub and grasslands and more noticeable presence of dense vegetation
- LNH (Low NDVI home range - Normal): Homogeneous home range with predominantly lower NDVI values, i.e., predominance of barren rocks, shrub, grasslands and presence of almost no dense vegetation
- HNH-P (High NDVI home range - Peaked): Homogeneous home range with predominantly higher NDVI values, i.e., predominance of densely vegetated areas such as crops and forests, with very few patches of barren rocks and grasslands.
- HNH-F (High NDVI home range - Flattened): Heterogeneous home range with predominantly higher NDVI values, i.e., predominance of densely vegetated areas such as crops and forests, with more noticeable presence of barren rocks and grasslands.
- HNH (High NDVI home range): Homogeneous home range with predominantly higher NDVI values, i.e., predominance of densely vegetated areas such as crops and forests, with almost no barren rocks and grasslands.
- NNH-P (Normal NDVI home range - Peaked): Homogeneous home range with predominantly average NDVI values, i.e., grasslands, with almost no barren rocks or dense vegetation.
- NNH-F (Normal NDVI home range - Flattened): Heterogeneous home range with predominantly average NDVI values, i.e., grasslands, some barren rocks and dense vegetation.
- NNH (Normal NDVI home range): Homogeneous home range with predominantly average NDVI values, i.e., grasslands, with very few barren rocks and dense vegetation.

For each day we further identified which of the nine types was the best fit of NDVI distribution within each home range. In the next step we used the distribution to compute Z-scores for each ENDVI and MODIS NDVI values where GPS points were registered on that day, i.e., locations the maned wolf visited on that day. Visited locations with Z-scores higher than 1.28 indicate preference for areas with higher NDVI, i.e. the wolf is choosing areas with higher NPP. Conversely, locations with Z-scores lower than - 1.28 indicate preference for areas with lower NDVI, i.e., the wolf is choosing areas with lower NPP. We plotted the Z-scores for each wolf along with the availability of MODIS NDVI and the number of fixes to evaluate the representativeness of the obtained Z-scores (See Figure 4.14 for example).


Figure 4.3: The overview of our framework for producing a daily detailed NDVI time series (ENDVI) and using it as contextual data to perform CAMA (Context-Aware Movement Analysis) on maned wolves' trajectories to identify seasonal patterns in their diet. Blue ellipses show inputs, grey rectangles show processing steps, yellow rectangles are secondary outputs and green rectangles show primary outputs, i.e., the final products.


Figure 4.4: Home ranges ( $95 \%$ utilization distribution) for each individual and Canastra National Park (CNP) limits overlayed on top of land use classes. Home ranges of the two individuals in each couple (Table 4.1) intersect to a large extent. The land use map was produced by de Paula (2016) based on automatic and supervised multi-temporal classification (2009-2011) of RapidEye images with 5 m spatial resolution.


Figure 4.5: The trade-off between the spatial and the temporal resolutions. Panel A) shows a time-line covering the GPS tracking period, where horizontal dashes indicate availability of satellite images after removing the ones covered by clouds. The type of satellite image is specified by the colour of the bar shown in the legend. MODIS images are plotted on the right time-line, whilst the others are plotted on the left time-line. Panel B) shows scaled pixel sizes overlaying an image from a portion of the study area, highlighting how the heterogeneity of environmental conditions might be camouflaged by the spatial resolution of MODIS. A MODIS pixel covers $250000 \mathrm{~m}^{2}$, a Landsat pixel covers $900 \mathrm{~m}^{2}$, a CBERS pixel covers $400 \mathrm{~m}^{2}$ and an ASTER pixel covers $225 \mathrm{~m}^{2}$.


Figure 4.6: The filtered NDVI growth rate extracted from each pixel from MODIS data is combined with land cover data from the finer spatial resolution images to produce a disaggregated NDVI (ENDVI), which has the temporal resolution of MODIS and the spatial details from the finer images. The land cover is used to find the NDVI growth rate for each pixel at the finer images. Growth rates are then applied to the available finer NDVI images to create the temporal series of ENDVI.


Figure 4.7: The spatially neighbouring pixels of a target MODIS pixel being downscaled and land cover composition in a MODIS pixel classified from available LANDSAT, ASTER or CBERS images (adapted from Rao et al. (2015)). The neighbourhood shape and size varies for each MODIS pixel being downscaled because it is determined by the number of land cover classes within the targeted MODIS pixel.


Figure 4.8: Changes in the shape of the probability density function (PDF) according to different ranges of skew and kurtosis. All distributions are plotted within the same x and y limits. The reference range for normality is between -2 and 2 (George and Mallery, 2010).








Figure 4.9: Types of probability density function (PDF) according to different ranges of skew and kurtosis. The first character in the name of the distribution refers to the skewness, it can be "L" for low (skew $>2$ ), " H " for high (skew $<-2$ ) or "N" for "normal" $(-2 \leq$ skew $\geq 2)$. The character after the dash refers to the kurtosis of the distribution, it can be "P" for "peaked" (kurtosis $>2)$, " $F$ " for "flattened" (kurtosis $<2$ ) or absent for when it is similar to the normal distribution $(-2 \leq$ kurtosis $\geq 2)$. The second and third character are always "NH" and stand for "NDVI in the home range".

### 4.2.6 Eigenbehaviour analysis: Identifying seasonal structure in habitat use

Contextualised movement data sets are highly dimensional and as such can be difficult to visualize and interpret. However, often not all of the variables are necessary to understand the movement patterns linked to behaviour because there often exists a smaller intrinsic dimensionality in the data set that explains most of the variance (Demšar et al., 2013). Therefore, it is often of interest to reduce the dimensionality of the data in order to understand it better, which can be done via principal components analysis (PCA) techniques such as Eigen decomposition.

Eigendecomposition has been used in many applications in computer science (Eagle and Pentland, 2009), such as object and face recognition (Turk and Pentland, 1991), shape and movement description (Pentland and Sclaroff, 1991), data interpolation (Pentland, 1992) and computer animation (Pentland and Williams, 1989). The repeating structures behind animal behaviour can also be retrieved with this method: commonly repeated behavioural patterns can be found by identifying eigenbehaviours, i.e., the principal components of an individual's behavioural dataset (Eagle and Pentland, 2009).

The Eigenbehaviours are a set of vectors that characterize the variation in the behaviour of an entity during a time period. These vectors are eigenvectors of the covariance matrix, or Principal Components (PC's) of behaviour data. Eigenvectors with the highest eigenvalues usually represent a repeated behaviour, such as using greener areas during dawn for forage. A linear combination of an individual's eigenvectors can precisely reconstruct the behaviour from each day in the data, also making possible to accurately predict an individual's subsequent behaviour based on its eigenvectors (Eagle and Pentland, 2009; Hurley et al., 2014). For more details on the mathematics and calculation of eigenvectors/PCs please refer to Jolliffe (2002).

In the context of movement, eigenbehaviour analysis has been used on sequential data representing people's daily behaviours (Eagle and Pentland, 2009). We adapt this method to investigate repeated behaviours of maned wolves. For this, their semantic trajectories first need to be converted into sequences, because eigenbehaviour analysis requires regularly sampled categorical behavioural data indicating the possible states for an individual (Eagle and Pentland, 2009).

As most trajectories were sampled more than once a day but with different temporal resolutions, we computed daily average Z-scores for each wolf to make our data regularly sampled. We then classified the average Z-scores into high ENDVI (H), low ENDVI (L) and average ENDVI (A). So that, for each day of each wolf we had a description of their choice of vegetated area in relation to the vegetation available in their entire home range. This produced codified sequences at daily temporal resolution describing the habitat use preference for each wolf during the study period. These sequences were then evaluated in terms of contextual data quality and representativeness of the Z-score. Sequences with long data gaps in which there were less than $1 / 5$ of MODIS data were excluded of further analysis, as well as the ones with less than 10 fixes recorded per day.

The remaining sequences were ordered according to what we called a wolf year, starting on the $1^{\text {st }}$ of July and ending on the $30^{\text {th }}$ of June. This year corresponds to a cycle in wolf ecology and starts at the time when the whelping rate peaks in the year cycle. This cycle consists of the following parts:

1. Whelping : between June and September (dry season)
2. Non-reproductive: between October and February (wet season)
3. Breeding : between March and June (dry season)

Each wolf year was treated as a different sequence, even when there were multiple sequences/years linked to the same wolf. Then, we characterized all the wolves by $B(x, y)$, a two-dimensional W by 365 array, where W is the total number of wolf years in the study and 365 is the number of days within a year, leap years had the extra day disregarded. B contains $n$ classes of ENDVI Z-score, one for each day of the year, which in our case corresponded to the four wolf behaviours relative to the vegetation: high ENDVI (H), low ENDVI (L) and average ENDVI (A) and a no data class (N). We converted the $B$ into $B^{\prime}$ a W by $365 X n$ array of binary values (Figure 4.10).

In order to obtain the repetitive behaviours of wolves during the study period, we applied eigendecomposition to the $B^{\prime}$ matrix of states. This resulted in a $1460 X 1460$ matrix of


Figure 4.10: Transformation from $B$ to $B^{\prime}$. The plot on the left, $B(x, y)$, corresponds to the wolves behaviour over the course of 365 days for four states. The plot on the right, $B^{\prime}(x, y)$, represents the same data in the form of a binary matrix of 365 days by 1460 (which is 365 multiplied by the four possible states $(n)$.
eigenvectors that can are ranked according to their eigenvalue. The vectors with the highest eigenvalues are considered an individual's primary eigenbehaviours (Eagle and Pentland, 2009). We calculated the percentage of variance explained (PVE) by each eigenvector in a scree plot and used it as our criteria for dimensionality reduction. We plotted the PVE and used the point where the graph levelled off as the threshold for selecting the eigenbehaviours to be kept and interpreted (Jolliffe, 2002). This is important as it is not feasible to interpret 1460 different eigenbehaviours and the components with low eigenvalues reflect individual behaviour, which is not our focus. The eigenvectors with higher eigenvalues reflect behaviours that are common to most wolves in the study, i.e., population behaviour and are the ones we are interested in keeping.

### 4.3 Results

### 4.3.1 Calculating NDVI growth rates for land cover fractions to generate finer NDVI time-series (ENDVI)

As we did not have access to temporal ground-truthing data, we plotted the ENDVI time series to visualise if the ENDVI was reporting the seasonal changes on vegetation in a reasonable manner. For this we selected pixels from areas where the multi-temporal land cover classification by de Paula (2016) indicated a specific crop type instead of a generic "farmland" area. The reason for this choice is that the knowledge of the crop type allows the comparison between the agricultural calendar (harvest, plating, growing) for that specific crop in that region and the ENDVI values. The only specified crop in the area was coffee fields, which have well defined seasons.

The average ENDVI time series from three pixels from different coffee crops is shown in Figure 4.11 , the time axis covers the same period for which we had a verified land cover map (Figure 4.4). The knowledge of the land cover allowed us to compare the ENDVI curve to the expected NDVI pattern for that specific land cover, taking into account dry and rainy seasons or harvesting and planting seasons. In addition to being the only specified farmland, coffee crops have a well known NDVI signature which has been used for mapping coffee fields in Brazil (Alves et al., 2016; Bernardes et al., 2012) and allowed us to verify if our ENDVI product was in accordance with the land cover dynamics.

The ENDVI was able to capture the seasonal variations in the phenology of the coffee crops. As expected the lower values are found around the planting season, which is coherent with the field being bare soil during that period and therefore having the lowest ENDVI responses. The highest ENDVIs are found immediately before the harvest season, reflecting the productive peak of the plant. There is a spike in ENDVI values after October 2011, which is a result of the lack of higher resolution images between August 2011 and July 2012, a result of malfunctioning of the Landsat 5 and 7 satellites during that period. This indicates a high sensitivity of the model to the absence of higher resolution images. Therefore a series of more higher resolution NDVI images spread evenly across the study period is preferred and more likely to produce


Figure 4.11: Average ENDVI time series from three pixels from different coffee crops, the time axis covers the same period for which we had a verified land cover map. The crop seasons are shown by background colours explained in the legend.
more accurate ENDVI.
Despite this sensitivity to the lack of high resolution images, the ENDVI preserved the seasonal trends, i.e., seasonal changes were still preserved even where absolute values were not as accurate. Since the changes and trends are more relevant for our study than the absolute values of ENDVI, as they give us information on the dynamics of the landscape in which maned wolves were moving, this did not pose a particular problem.

### 4.3.2 Context integration and seasonal plots

This section shows the results for the seasonal plots derived from the semantic trajectories and annotated home ranges. Figure 4.12 shows the frequencies for different values of skew and kurtosis computed for the daily ENDVI distributions within the home ranges. The distributions are predominantly positively skewed (Figure 4.12 A ) and show only positive kurtosis (Figure 4.12 B), which means that only six out of the nine distributions (Figure 4.9) characterising the vegetation within the home ranges were possible. The absence of negative kurtosis indicates that the home ranges are predominantly homogeneous in terms of vegetation, and the predominance of positive skewness indicates that home ranges have mostly lower NDVI, therefore predominance of grasslands and sparse vegetation, which is in agreement with the phenology of the Brazilian Savannah.


Figure 4.12: Frequency of different values of skew and kurtosis for the ENDVI distributions within the annotated home ranges. The total number of ENDVI distributions is 67631 .

The home ranges showed ENDVI distributions of type NNH, NNH-P, LNH-P and rarely HNH-P (Figure 4.13). HNH-P distributions were observed between November and December of 2009 for the couple Samurai (Figure B.6) and Jurema (Figure 4.15). Generally the Z-scores of the areas used by the wolves showed almost a cyclic rhythm in terms of transitioning from high Z-scores to medium and the low Z-scores. Overall, it was also possible to see that the use of the ENDVI gives a more complete picture of the landscape dynamics, as the distributions are changed from NDVI to ENDVI and the latter ones are more compatible with what would be expected from that landscape. In addition, the ENDVI product was able to retrieve data where there was cloud coverage and no MODIS NDVI image available.


Figure 4.13: Types of probability density function (PDF) according to different ranges of skew and kurtosis that were actually observed in the annotated home ranges. The first character in the name of the distribution refers to the skewness, it can be "L" for low (skew $>2$ ), "H" for high (skew $<-2$ ) or " N " for "normal" ( $-2 \leq$ skew $\geq 2$ ). The character after the dash refers to the kurtosis of the distribution, it can be "P" for "peaked" (kurtosis > 2), "F" for "flattened" (kurtosis $<2$ ) or absent for when it is similar to the normal distribution ( $-2 \leq$ kurtosis $\geq 2$ ). The second and third character are always "NH" and stand for "NDVI in the home range".

We selected the seven wolves with the best data quality to present here, their results are shown in Figures 4.14 to 4.20. The results for the remaining wolves can be found in Appendix B. The figures show the type of distribution for the MODIS NDVI versus the Z-score for the wolf's locations (top panel), the type of distribution for the ENDVI versus the Z-score for the wolf's locations (middle panel), and the data quality in terms of number of fixes registered on
that day and the average number of MODIS images in a five-day window (bottom panel). In the next paragraphs, we analyse these graphs for the selected seven wolves in more detail.

The wolf Bolt established his home range inside the CNP area. His home range is covered mostly by heath with no more than three small patches of young forest at the South-West border (Figure 4.4). This wolf was tracked between the end of 2013 and 2015, the period in within which the study area was going trough an intense drought His results are shown in Figure 4.14. The ENDVI time series shows a clear pattern of landscape transitioning from a higher net primary productivity (NNH), to a medium net primary productivity (NNH-P) and finally to a low net primary productivity (LNH-P). However, the proportion of time in which the landscape is HNH is lower compared with that of a year without drought. As the landscape transitions into a state of lower NPP there is a change in the periodicity of Bolt's visits to locations with higher Z-scores of ENDVI, i.e., locations with more food availability. This periodicity seems to change from daily/every other day visits, in the NNH period, to five to seven day intervals between visits in the LNH-P period, reaching a maximum of 15 days interval at the end of the dry season in March 2015.

The wolf Jurema established her home range outside the CNP. Her home range is covered mostly by pasture and heath, a big patch of farmland and very few patches of mature forest (Figure 4.4). This wolf was tracked between 2009 and 2011. Her results are shown in Figure 4.15. The ENDVI time series show a clearer pattern in terms of landscape transitioning from a higher NPP (NNH) in the dry season, to a medium NPP (NNH-P) and finally to a low NPP(LNH-P) int he wet season. As the landscape transitions into a state of lower NPP there is a change in the periodicity of visits to locations with higher Z-scores of ENDVI, i.e., location with more food availability. It seems to transition from daily/every other day visits, in the NNH home range, to five to seven days intervals in the LNH-P home range. However, the quality of contextual data is rather unstable which can affect the results in the graph.

The wolf Lais, Amadeo's partner, established her home range in the CNP border, her home range is covered mostly by heath, some pasture and few patches of young forest (Figure 4.4). This wolf was tracked between 2007 and 2008, period in which the state of Minas Gerais was going through an intense drought. Her results are shown in Figure 4.16. The ENDVI time


Figure 4.14: The top panel shows the temporal series of Z-scores of the locations used by Bolt in relation to MODIS NDVI distributions within the home range ( $\mathrm{n}=2569$ pixels of 250 m ). The colours in the Z-score plots indicate the type of NDVI or ENDVI distribution (Figure 4.9) for the day. The middle panel shows the temporal series of Z-scores of the locations used by Bolt in relation to the ENDVI distributions within the home range ( $\mathrm{n}=599292$ pixels of 15 $\mathrm{m})$. Red dashed lines indicate the $10 \%$ confidence interval, points above that show that the animal was selecting areas amongst the $10 \%$ more vegetated in the home range. The bottom panel shows the number of fixes collected each day in the period and the average number of MODIS images available at each day within a 5 days moving window. The beige background indicates when the data quality criteria were fully met, i.e., more than 10 fixes and more than 0.2 MODIS images in a 5 days window.


Figure 4.15: The top panel shows the temporal series of Z-scores of the locations used by Jurema in relation to MODIS NDVI distributions within the home range ( $\mathrm{n}=881$ pixels of 250 m ). The middle panel shows the temporal series of Z-scores of the locations used by Jurema in relation to the ENDVI distributions within the home range ( $\mathrm{n}=204670$ pixels of 15 m ). Red dashed lines indicate the $10 \%$ confidence interval, points above that show that the animal was selecting areas amongst the $10 \%$ more vegetated in the home range. The bottom panel shows the number of fixes collected each day in the period and the average number of MODIS images available at each day within a 5 days moving window. The beige background indicates when the data quality criteria were fully met, i.e., more than 10 fixes and more than 0.2 MODIS images in a 5 days window.
series show a clearer pattern in terms of landscape transitioning from a higher NPP (NNH), to a medium NPP (NNH-P) and finally to a low NPP (LNH-P). As the landscape transitions into a
state of lower NPP there is a change in the periodicity of visits to locations with higher Z-scores of ENDVI, i.e., location with more food availability. It seems to transition from daily/every other day visits, in the NNH home range, to five to seven days intervals in the LNH-P home range. However, the data quality was low in the middle months of the tracking period.

The wolf Loba, established her home range within the CNP area. Her home range is covered mostly by heath, very few pasture and young forest patches and two patches of mature forest (Figure 4.4). This wolf was tracked between 2012 and 2014, period in which the state of MG was going trough an intense drought in 2014. Her results are shown in Figure 4.17. The ENDVI time series show a clearer pattern in terms of landscape transitioning from a higher NPP (NNH), to a medium NPP (NNH-P) and finally to a low NPP (LNH-P). As the landscape transitions into a state of lower NPP there is a change in the periodicity of visits to locations with higher Z-scores of ENDVI, i.e., location with more food availability. It seems to transition from daily/every other day visits, on the NNH home range, to five to seven days intervals on the LNH-P home range. However, the data quality was very poor during the tracking period.

The wolf Luna, established her home range outside the CNP area, her home range is covered mostly by heath and pasture, with very few patches of shrubland and coffee crops (Figure 4.4). This wolf was tracked during 2013, a typical year in terms of climate in the region. He results are shown in Figure 4.18. The ENDVI time series show a very clear pattern where the dry season has a LNH-P distribution of ENDVI and the beginning of the wet season has a NNH distribution, i.e., higher NPP during the wet season. As the landscape transitions into a state of lower NPP there is a change in the periodicity of visits to locations with higher Z-scores of ENDVI, i.e., location with more food availability. It seems to transition from daily/every other day visits, in the NNH home range, to five to seven days intervals in the LNH-P home range. However, the data quality was very poor during the tracking period.

The wolf Rose, Bolt's partner, established her home range inside the CNP area, her home range is covered mostly by heath with no more than three small patches of young forest at the South-West border (Figure 4.4). This wolf was tracked between 2014 and 2015. Her results are shown in Figure 4.19.The ENDVI time series show a very clear pattern where the wet season has a NNH and NNH-P distribution of ENDVI and the dry season has a LNH-P distribution,


Figure 4.16: The top panel shows the temporal series of Z-scores of the locations used by Lais in relation to MODIS NDVI distributions within the home range ( $\mathrm{n}=1926$ pixels of 250 m).The colours in the Z-score plots indicate the type of NDVI or ENDVI distribution (Figure 4.9) for the day. The middle panel shows the temporal series of Z-scores of the locations used by Lais in relation to the ENDVI distributions within the home range ( $\mathrm{n}=450967$ pixels of 15 m ). Red dashed lines indicate the $10 \%$ confidence interval, points above that show that the animals was selecting areas amongst the $10 \%$ more vegetated in the home range. The bottom panel shows the number of fixes collected each day in the period and the average number of MODIS images available at each day within a 5 days moving window. The beige background indicates when the data quality criteria were fully met, i.e., more than 10 fixes and more than 0.2 MODIS images in a 5 days window.


Figure 4.17: The top panel shows the temporal series of Z-scores of the locations used by Loba in relation to MODIS NDVI distributions within the home range ( $\mathrm{n}=1732$ pixels of 250 m).The colours in the Z-score plots indicate the type of NDVI or ENDVI distribution (Figure 4.9) for the day. The middle panel shows the temporal series of Z-scores of the locations used by Loba in relation to the ENDVI distributions within the home range ( $\mathrm{n}=404485$ pixels of 15 m ). Red dashed lines indicate the $10 \%$ confidence interval, points above that show that the animals was selecting areas amongst the $10 \%$ more vegetated in the home range. The bottom panel shows the number of fixes collected each day in the period and the average number of MODIS images available at each day within a 5 days moving window. The beige background indicates when the data quality criteria were fully met, i.e., more than 10 fixes and more than 0.2 MODIS images in a 5 days window.


Figure 4.18: The top panel shows the temporal series of Z-scores of the locations used by Luna in relation to MODIS NDVI distributions within the home range ( $\mathrm{n}=550$ pixels of 250 m ). The colours in the Z-score plots indicate the type of NDVI or ENDVI distribution (Figure 4.9) for the day. The middle panel shows the temporal series of Z-scores of the locations used by Luna in relation to the ENDVI distributions within the home range ( $\mathrm{n}=128873$ pixels of 15 $\mathrm{m})$. Red dashed lines indicate the $10 \%$ confidence interval, points above that show that the animals was selecting areas amongst the $10 \%$ more vegetated in the home range. The bottom panel shows the number of fixes collected each day in the period and the average number of MODIS images available at each day within a 5 days moving window. The beige background indicates when the data quality criteria were fully met, i.e., more than 10 fixes and more than 0.2 MODIS images in a 5 days window.
i.e., higher NPP during the dry season. As the landscape transitions into a state of lower NPP there is a change in the periodicity of visits to locations with higher Z-scores of ENDVI, i.e., location with more food availability. It seems to transition from daily/every other day visits, on the NNH home range, to five to seven days intervals on the LNH-P home range.

The wolf Tay, Gamba's partner, established her home range in the CNP border. Her home range is covered mostly by heath and pasture, with scattered islands of shrubland and some farmland (Figure 4.4). This she wolf was tracked between 2007 and 2010, period in which the study area was going trough an intense drought in 2008. Her results are shown in Figure 4.20. The ENDVI time series show a very clear pattern where the wet season has a NNH and NNH-P distribution of ENDVI and the dry season has a LNH-P distribution, i.e., higher NPP during the dry season. As the landscape transitions into a state of lower NPP there is a change in the periodicity of visits to locations with higher Z-scores of ENDVI, i.e., location with more food availability. It seems to transition from daily/every other day visits, in the NNH home range, to five to seven days intervals in the LNH-P home range.


Figure 4.19: The top panel shows the temporal series of Z-scores of the locations used by Rose in relation to MODIS NDVI distributions within the home range ( $\mathrm{n}=1655$ pixels of 250 m ). The colours in the Z-score plots indicate the type of NDVI or ENDVI distribution (Figure 4.9) for the day. The middle panel shows the temporal series of Z-scores of the locations used by Rose in relation to the ENDVI distributions within the home range ( $\mathrm{n}=387349$ pixels of 15 $\mathrm{m})$. Red dashed lines indicate the $10 \%$ confidence interval, points above that show that the animals was selecting areas amongst the $10 \%$ more vegetated in the home range. The bottom panel shows the number of fixes collected each day in the period and the average number of MODIS images available at each day within a 5 days moving window. The beige background indicates when the data quality criteria were fully met, i.e., more than 10 fixes and more than 0.2 MODIS images in a 5 days window.


Figure 4.20: The top panel shows the temporal series of Z-scores of the locations used by Tay in relation to MODIS NDVI distributions within the home range ( $\mathrm{n}=584$ pixels of 250 m ). The colours in the Z-score plots indicate the type of NDVI or ENDVI distribution (Figure 4.9) for the day. The middle panel shows the temporal series of Z-scores of the locations used by Tay in relation to the ENDVI distributions within the home range ( $\mathrm{n}=138140$ pixels of 15 $\mathrm{m})$. Red dashed lines indicate the $10 \%$ confidence interval, points above that show that the animals was selecting areas amongst the $10 \%$ more vegetated in the home range. The bottom panel shows the number of fixes collected each day in the period and the average number of MODIS images available at each day within a 5 days moving window. The beige background indicates when the data quality criteria were fully met, i.e., more than 10 fixes and more than 0.2 MODIS images in a 5 days window.

### 4.3.3 Eigenbehaviours: Identifying seasonal structure in habitat use

This section describes the results obtained by applying eigen decomposition to the sequences of maned wolf behaviour that were generated from the semantic trajectories. Figure 4.21 shows
all the annotated trajectories sequenced in different wolf years, which resulted in a total of 28 sequences of wolf years for which there was information on the choice of vegetated areas by the wolves. Twenty two of those were selected for the analysis of eigenbehaviours, because six of the sequences were either too short or with too low quality contextual data, or both.


Figure 4.21: The real tracking period for each wolf is shown in the top panel and the translation of these data into wolf years is shown in the bottom panel. Each wolf is represented by a different colour and white gaps represent missing data that was accounted for as no data.

We kept the eigenbehaviours that explained most of the variance in the dataset, which we evaluated based on the percentage of variance explained by the eigenvalues. For this we created a scree plot (Figure 4.22) which is a plot of the number of eigenvalues $v s$ the percentage of variance each vector contributes. A heuristic way of performing dimensionality reduction is to select the eigenvectors until the point where the scree plot plateaus. The first eigenbehaviour explained $45 \%$ of the variance in the data, the second explained $31.6 \%$, the third explained $5.08 \%$, the fourth explained $4.08 \%$ and the fifth explained $2.57 \%$. The percentage of variance explained plateaus from the $6^{\text {th }}$ eigenbehaviour forwards around $1 \%$, for that reason we kept only the first five eigenbehaviours that accounted in total for $88.33 \%$ of the variance in the whole dataset.


Figure 4.22: Eigenvalues ranks and percentage of variance explained by each. The total number of ranks was 1460 , however the graph plateaus after the $6^{\text {th }}$ rank.

The eigenvectors with the highest eigenvalues represent behaviours that are common to most wolves in the study, i.e., the primary eigenbehaviours. Figure 4.23 shows the five first eigenbehaviours with respective absolute eigenvector values. Higher eigenvector values (red) show a higher contribution of the state at that point in time and lower eigenvector values (white) show lower contribution of the behavioural state at that point in time (panels H, L and A in the $B^{\prime}$ matrix in Figure 4.10). The first eigenbehaviour corresponds to years where wolves stay in areas of average ENDVI before and during the wet season and start to choose areas with higher NPP at the end of the wet season onwards. The second eigenbehaviour show the opposite trend, where wolves stay in areas of average ENDVI after and during the dry season and preferentially select areas of high ENDVI at the middle of the wet season. The third eigenbehaviour corresponds to years where wolves stay in areas of average ENDVI intermittently, before and during the second half of the wet season and start to choose areas of higher ENDVI at the first half of the wet season and at the beginning of the dry season.

The fourth eigenbehaviour corresponds to years where wolves stay in areas of average ENDVI for the first half of the year and particularly before the wet season, and choose areas of higher ENDVI at the middle of the first dry season as well as during the entirety of the second dry season. The fifth eigenbehaviour corresponds to years where wolves choose areas of higher ENDVI the entire year and then stay within areas of average ENDVI between the end of wet season and the beginning of the dry season.

The second row shows that there seems to be a persistent trend of choosing areas of low ENDVI at very specific times of the year, this can be seen in all eigenbehaviours.







We combined the computed eigenbehaviours into sequences of three behavioural states by taking the state with the highest eigenvector coefficients at any given time. This resulted in the sequences shown in Figure 4.24. The first eigenbehaviour shows a preferential selection of high ENDVI areas after the wet season, the second one shows less preferential selection of high ENDVI areas and only in the first half of the wet season, the third one shows preferential selection of high ENDVI in the first half of wet season and less so at the beginning of the dry season. The fourth eigenbehaviour shows preferential selection of high ENDVI areas from the middle of the wet season onwards and the fifth shows preferential selection of high ENDVI areas during the entire year. The second, third and fourth eigenbehaviours show a pattern of preferential selection of low ENDVI areas in the second third of the wet season and at the end of the dry season.


Figure 4.24: Sequences of states summarising the five first eigenbehaviours. Sequences were generated by selecting the state with the highest eigenvector coefficient at each time during the wolf year for each eigenvector. The letters in the horizontal axis indicate the seasons, D for dry and W for Wet.

### 4.4 Discussion

The recent widespread availability and quality of geospatial data on movement and context presents new challenges and opportunities for developing innovative methods to understand the interactions between wildlife movement and environment. We were interested in developing a multi-source disaggregation approach to produce contextual datasets with higher temporal resolution and level of detail for CAMA. We did that by using daily MODIS images with 250 meters spatial resolution and finer images ( 15 to 30 m spatial resolution) collected fortnightly to derive daily images with the same spatial resolution as the finer source images. This methodology allowed us to produce not only a series of contextual data with better temporal coverage than the original MODIS and finer resolution data, but also with a higher level of spatial detail on the contextual variable.

Whilst some contextual variables, such as built-up area, can be considered static over a large period of time, many contextual variables relevant for CAMA, such as vegetation phenology, have an inherent temporal variability (Urbano and Cagnacci, 2014). This temporal variability is best represented by dynamic contextual information that corresponds as close as possible to the conditions encountered by an individual moving across the landscape (Moorcroft, 2012) and the use of static data, in this case, not only introduces bias but also limits statical inference (Basille et al., 2013). High temporal resolution series of remote sensing data can provide dynamic contextual data of medium to low spatial resolution and have been extensively used for CAMA (Pettorelli et al., 2006; Remelgado et al., 2018; Pettorelli et al., 2011; Dodge et al., 2014; Urbano and Cagnacci, 2014; Neumann et al., 2015). Traditionally, these spatially coarse datasets have been used when temporal variability is a key component of the study, because of their high temporal resolution (Basille et al., 2013). Yet, the spatial resolution of these datasets ( 250 m for MODIS, 1 km for SPOT, 8 km for AVHRR) does not match the current average error of less than 20 m (Frair et al., 2010), in GPS tracking, which leads to a mismatch between tracking data and contextual layers (Urbano and Cagnacci, 2014).

Remote sensing data have become a standard source of contextual data, however the combination of multiple sensors is still not a routine. This disregards opportunities to capitalise
on the different characteristics of varied satellites (Bühne and Pettorelli, 2017), such as their high temporal resolution or high spatial resolution. The approach we proposed in this thesis capitalises on the strengths of multiple satellites by using data fusion to produce a series of contextual layers for CAMA. This is the first time that contextual data from multiple satellites are used to annotate trajectories and perform CAMA. Usually, studies are restricted to a single source of contextual data, which is commonly a pre-processed contextual variable (Urbano and Cagnacci, 2014; Bühne and Pettorelli, 2017). The higher level of spatial detail alongside the better coverage produced by our method enabled us to capture the seasonal fluctuations of context within the home ranges, which was particularly interesting for analysing the values of the contextual variable that were used in comparison to the ones available in the area.

Analysing contextualised trajectories is a cumbersome task and most algorithms disregard contextual data in the process, which is one of the current drawbacks of methods for movement analysis (Buchin et al., 2012). The most common approaches for exploring semantic trajectories are the use of map animation or space-time cubes, which are both limited in the number of trajectories it can show, the time period it can cover, and the ability to represent contextual variables (Andrienko et al., 2011). This chapter suggests the use of eigendecomposition as an alternative approach to perform CAMA. This method does not have a limitation on the number of trajectories or the time period to be covered in the analysis. In addition, this method is less sensitive to gaps in the trajectories, as the PC's are calculated at each time and by behavioural state. This means that the "no data" sections are shown as separate dimensions and can be excluded without hindering the identification of the relevant behavioural patterns.

We used the seasonality in the diet of maned wolves to test our hypothesis that our approach would enable the detection of finer scale movement patterns linked to contextual changes. More specifically, we wanted see if we can identify the diet with a temporal trophic opportunistic pattern indicated by previous studies. Our results seem to support this hypothesis, as all wolves showed higher interval between visits to high-NDVI locations when the home range distribution showed lower food availability, and more frequent visits when food availability was higher.

We also found that most wolves will choose greener areas during the dry season, which
agrees with the current literature in which wolves have special preference for wolf's fruit in the dry season. These fruit grow on a flowering shrub with height up to 5 m and large leaves, which has a higher NDVI response than most vegetation in the Cerrado, such as the heath and grasslands. In addition, even though the dietary habits of maned wolf are unknown in wild, most wolves double their food intake during breeding season as nourishment is a key factor for successful reproduction (Sillero-Zubiri and Gottelli, 1995). This seems to match with our results that show that wolves are choosing areas with higher ENDVI during the breeding season, i.e., they are actively choosing greener areas in a period of low food availability and when they require more energy to ensure successful reproduction.

It is possible that visits to areas of low Z-scores may be linked to denning, as typically wolves choose rocky areas as for this purpose and those have lower NDVI. More specifically, we believe that the ones happening during the dry season are related to whelping, mainly the ones for which we found female wolves decreasing the distance they cover in a day at that exact time of the year (See C. 1 in Appendix C). On the other hand, visits to areas with higher NDVIs, i.e., higher Z-scores may be linked to feeding and foraging since these animals eat not only fruits but also small mammals that are often found in vegetated areas. In addition, considering that vegetation phenology is mostly driven by precipitation and that in the CNP there are two very well-defined dry and wet seasons, it makes sense that most of the ENDVI plots show the existence of two, sometimes three types of distribution within the home ranges.

As the landscape transitions into a state of lower NPP there is a change in the periodicity of wolves visits to locations with higher Z-scores of ENDVI, i.e., locations with more food availability. The frequency seems to transition from daily/every other day visits, when food availability is higher, to 5 days or even 15 days, when food availability is lower. This exact feeding pattern (feast-famine) has been reported in the literature for other species of wolf (Stahler et al., 2006).

Some wolves like Jurema and Samurai, showed preference for higher NDVI areas during unexpected periods, considering the natural phenology of their landscape. However, these wolves were mostly in highly anthropic areas, and particularly had their home ranges in areas with large patches of farmland. The second eigenbehaviour seem to match the harvest and
planting season for the crops that are found in that region (soy, sugar-cane, coffee and beans), which may mean that the presence of farmland could be interfering with the natural cycle of the species and consequently with the reproductive success. The avoidance periods showed by this eigenbehaviour seem to be related to whelping. The period of the year when this eigenbehaviour occurs and the decreased distance covered by female wolves point to the same possibility.

To summarise, the multi-source image fusion allows taking advantage of complementary information produced by different satellites (Bühne and Pettorelli, 2017). It is a new analysis tool for CAMA where contextual data is required at higher temporal resolution and level of detail than readily available. The main advantage of this approach is that it is general and can be applied for other species and other contextual variables derived from remote sensing data, such as marine net primary productivity, land surface temperature, humidity, air pollution or snow coverage. It is common in movement research to simultaneously need daily, even hourly contextual data, but also high spatial resolution, particularly for studies in areas with heterogeneous environments. The use of eigenbehaviours showed potential for studying contextualised trajectories, however there is the need for more studies where the patterns found can be compared to observational data on behaviour.

## Chapter 5

## Weather effects on human mobility: A study using multi-channel sequence analysis

### 5.1 Introduction

The spread of geolocated smartphones and the decreasing price of GPS devices have contributed towards the production of large amounts of data on human movement of unprecedented spatiotemporal quality (Meekan et al., 2017; Stanley et al., 2018). New human mobility studies attempt to link such movement data with contextual information (such as points of interest) to gather insights into, for example, commuting behaviour (Beecham et al., 2014; Gong et al., 2012), tourist behaviour (Meijles et al., 2014; Versichele et al., 2012), or retail choice decisions and human activities (Siła-Nowicka et al., 2016). However, integrating high resolution GPS trajectories and dynamic spatio-temporal contextual information remains an underexplored approach for studying the effects of weather on human movement, despite its relevance for urban planning (Givoni, 1974; Ng, 2012), traffic engineering (Dunne and Ghosh, 2013), retail planning (Thakuriah et al., 2016), tourism (de Freitas, 2003), health (Tucker and Gilliland, 2007), psychology (Nerlich and Jaspal, 2014) and epidemiology (Horowitz, 2002).

Specific weather conditions often trigger changes in human behaviour, for example, higher temperatures increase aggressiveness (Anderson, 2001; Carlsmith and Anderson, 1979) and lower temperatures contribute to irritability and combativeness (Schneider et al., 1980; Worfolk, 1997). Different components of weather have different magnitudes of importance, for example, air temperature, direct solar radiation and wind speed have a more significant influence on human behaviour than humidity (de Montigny et al., 2012). However, it is challenging to understand how weather influences human behaviour because the responses are partially a result of individual preferences (de Freitas, 2015). Some individuals are more responsive to the thermal component of weather, i.e. the combined effects of air temperature, humidity and solar radiation, while some are more receptive to physical components like rain, and others are more greatly affected by the aesthetic components, such as cloud coverage and sunshine. Yet, most individuals do respond to the combination of all three of these components (de Freitas, 1990).

Traditionally, these interactions have been explored through questionnaires and multidimensional scaling methods within the field of human biometeorology (Cabanac, 1971; de Freitas, 1990; Manu et al., 2016; Stanley et al., 2018). With the increased availability of tracking and environmental data we however propose that the effect of weather on movement behaviour can be explored through Context-Aware Movement Analysis (CAMA), which integrates movement geometry with its context, i.e. with the surrounding biological and environmental conditions that might be affecting movement (Andrienko et al., 2011; Demšar et al., 2015; Dodge et al., 2013). More specifically we use multi-channel sequence analysis (MCSA) to represent a person's movement as a sequence of states, describing either the type of movement or the state of the environment throughout time. Similar movement patterns can then be identified (termed context aware similarity analysis) by comparing and aligning mobility sequences.

Similarity analysis is one of the most common tasks in movement analytics and consists of using distance measures and grouping methods to split trajectories (Demšar et al., 2015) into groups of elements more similar amongst them than to other groups (Jain et al., 1999), which followed by clustering allows the identification of spatio-temporal movement patterns that might be linked to behaviour (Dodge et al., 2014). Similarity is often established based on
geometry or physical attributes; geometrical similarity solely relies on measures of spatial and temporal distances, and physical similarity relies on movement attributes such as speed, turning angle, acceleration and direction (Demšar et al., 2015). Context-aware similarity is based on multiple attributes (Andrienko et al., 2011; Demšar et al., 2015; Sharif and Alesheikh, 2017a) describing the conditions within which the movement took place.

Context-awareness is a recent trend (Sharif and Alesheikh, 2017b), as a result there are few context-aware methods for assessing similarity between trajectories. Sharif and Alesheikh (2017a) generalized the dynamic time warping (DTW) to develop a context-based dynamic time warping (CDTW) method, which matches trajectories with contextual similarity even if they are not concurrent. This method is highly dependent on arbitrary weights for the contextual variables, restricted to numeric context and disregards changes of context between two points in time. i.e., same contexts are considered similar even when they are not concurrent. De Groeve et al. (2016) uses single channel sequence alignments and hamming distance to understand the temporal variation of habitat use by roe deer; the similarity is measured by the cost to transform a sequence of habitat use into another. This method is able to handle only one contextual variable at time, therefore it is not able to handle the interactive effects of multiple contextual variables on movement. Buchin et al. (2014) modified existing similarity measures to make them context-aware, more specifically they defined the distance between two points as the sum of their contextual and spatial distances. The transition costs between contexts are defined by the user and the method is restricted to contextual data in the form of polygonal divisions.

We propose the use of multi-channel sequence analysis (MCSA) to perform context-aware similarity analysis (CASA) and cluster trajectories into groups of similar behaviour. MCSA is a new analysis tool for movement data where contextual information can now be readily combined with detailed tracking datasets. The main advantage of this approach is that it is also possible to consider as many channels (contextual variables) as desired at once. It is common in movement research to simultaneously consider multiple environmental variables, which makes MCSA particularly relevant for studying human mobility, traffic, transportation and wildlife ecology; areas in which movement behaviour may be contextualised by other
dynamic environmental variables such as air temperature, vegetation indices, humidity, wind speed, air pollution and snow coverage. Single channel analysis has been used before to explore spatio-temporal patterns on the activity of visitors in Akko's Old city - Israel (Shoval and Isaacson, 2007) and to analyse sequential habitat use by roe deer in North-East Italy (De Groeve et al., 2016). Shoval and Isaacson (2007) focused on sequences of locations, i.e. the movement itself, while De Groeve et al. (2016) emphasized sequences of habitat use classes, i.e. the context surrounding movement. Horanont et al. (2013) looked at GPS traces from mobile phone users, coarse scale movement data, hourly temperature, rainfall and wind speed to explore the independent effects of each variable on people's activity patterns. We innovate by applying MCSA, for the very first time, to perform CAMA of fine scale human movement data to simultaneously consider movement and context by looking at the combined and single effects of six meteorological variables.

Despite the novelty of MCSA in movement research, sequence analysis has been consistently used in medical and social sciences, particularly within bioinformatics and life courses research (Idury and Waterman, 1995; Abbott, 1995; Abbott and Tsay, 2000). In bioinformatics, a sequence represents the DNA molecule as a string of characters (which stand for specific nucleotides), between a precise start and end point; the comparison of similarities and differences between those strings allows the identification of nucleotide sequences related to genetic diseases and traits. We propose that the same principle can be applied to movement trajectories for identifying groups of people with similar movement patterns, i.e., clusters of similar behaviour Billari (2001). Further, we propose to not only represent the trajectories with one sequence only, but to use multi-channel sequence analysis (MCSA), which allows for comparison of sequences consisting of several dimensions (channels) (Gauthier et al., 2010). For this, we link data from a GPS tracking study to weather data and convert the information into multi-channel sequences in a first fully data-driven attempt to explore weather effects on human movement patterns.

The primary objective of this chapter is using multi-channel sequence analysis (MCSA) to perform context-aware similarity analysis (CASA) and cluster trajectories into groups of similar behaviour. The rest of the chapter is structured as follows: section 5.2.1 describes the GPS
tracking data; section 5.2.2 describes the meteorological data used in our analysis, explains how they were combined with the GPS tracking data and finally converted into sequences; sections 5.2.4.1 and 5.2.4.2 describe how multi-channel sequence analysis was applied to identify changes in group movement patterns related to weather. The results are presented in section 5.3 and discussed in section 5.4. We conclude with some considerations about our findings, the potential of this methodology and ideas for future research (section 5.4). Parts of this chapter are published as the journal article: Brum-Bastos et al. (2018).

### 5.2 Methodology

To study the influence of weather on human mobility behaviour we used a five-step process (Figure 5.1). In Step 1, we integrate trajectories with contextual data by using trajectory annotation to link GPS points to weather variables, which resulted in contextualized trajectories. In Step 2, we transform those trajectories into multi-channel sequences by creating alphabets with codes for each weather variable, travel mode and places. In Step 3, we use optimal matching distances (Abbott and Tsay, 2000) to calculate a dissimilarity matrix describing the degree of difference between each pair of multi-channel sequences in our dataset. In Step 4, we use Ward's clustering (Murtagh and Legendre, 2011) algorithm to partition the sequences into similarity based groups, which represent groups of people showing similar movement behaviour under particular weather conditions. In Step 5, we perform statistical tests to validate and understand differences between groups.

Trajectory annotation and sequencing were performed using PostgreSQL 9.4 database manager, VANJU library and its dependencies under Python 2.7, for more details refer to (BrumBastos et al., 2016). The MCSA, including optimal matching distances, Ward's clustering and statistical tests, was performed using TraMineR 1.8-9 and cluster 1.14.4 libraries under R 3.4.1, for more details on the equations used by these libraries please refer to Gabadinho et al. (2009) and Maechler et al. (2018) respectively.


Figure 5.1: The overview of our framework for identification of groups of similar movement behaviour under specific weather conditions. The framework has two analyses running in parallel: analysis of places and analysis of travel modes. Blue shapes marks travel mode, green shapes marks places, white ellipses represent dataset's sources, rectangles represent variables, beige arrows represent processing steps and hexagons derived results in each step.

### 5.2.1 Movement data

We analysed a human movement dataset where GPS devices were carried by volunteers from the Kingdom of Fife - UK (Figure 5.2a) (Siła-Nowicka et al., 2016). The data were collected between the $28^{\text {th }}$ of September 2013 and the $10^{\text {th }}$ of January 2014 as part of the GEOCROWD
project (Siła-Nowicka et al., 2016), in which 6000 individuals were randomly selected by postcode from the voting registry (focusing on the three major towns in Fife) and invited via letter to participate in the study. In total, 206 individuals accepted the invitation and provided usable data whereby they were tracked for two consecutive weeks within the study time spam. GPS devices recorded participant positions every five seconds, representing a very high-resolution trajectory of participant locations. The GPS trackers were coupled with accelerometers, which turned off the GPS when the individual was not moving (Oshan et al., 2014). The aim of the GEOCROWD project was to develop new movement analytics methods that would allow researchers to find out as much as possible from the actual GPS data while participants were asked to do as little as possible (i.e. the only task was to carry a GPS device and mail it back after two weeks). Therefore, very little auxiliary data were collected and beyond gender and age of the participants, which were sourced from the electoral register together with the address of each participant, no other demographic or ground truth data were collected. For more details on data collection refer to Oshan et al. (2014).

In this chapter we re-analyse the GEOCROWD data from the town called Dunfermline (Figure 5.2a), which had the highest number of participants ( $\mathrm{n}=91$ ), of which 23 were female, 41 were male, and 27 did not declare their gender. Looking at the ages of our participants: 10 were between 21 and 34 years old, 46 were between 35 and 60 years old, 8 were between 61 and 65 years old, and 27 did not declare their age. As stated above, apart from their home address, gender, and age, no further information about participants or their activities were available for our secondary data analysis.

The participant trajectories were classified into movement classes (Walk, Train, Bus and Vehicle, Traffic Stop, Bus Stop, Train Stop) and stop classes (Home, Work, Shopping, Unidentified Stop) (Figure 5.2b) (Siła-Nowicka et al., 2016). The classification achieved $85 \%$ accuracy, which was assessed by comparing a 200 m range from the recorded home addresses with the home location found by the classification algorithm (for more details on data segmentation and classification refer to Siła-Nowicka et al. (2016)).


Table 5.1: Summary of contextual datasets with respective sources and specifications.

| Source | Variables | Data type | Geometry | Temporal <br> resolution | Spatial <br> resolution |
| :---: | :--- | :--- | :---: | :---: | :---: |
| Weather Cam <br> (UK Weather, 2013) | Daylight | Categoric | Point | 24 h | - |
| NIMROD <br> (MetOffice, 2003) | Rainfall | Numeric | Raster | 5 m | $1-5 \mathrm{~km}$ |
| MIDAS | Temperature, <br> Relative humidity, <br> Wind speed, | Numeric | Point | 1 h | - |
| Wind direction |  |  |  |  |  |

We associated MIDAS data with trajectory points using Thiessen Polygons around each meteorological station ( $\mathrm{n}=109$, Figure 5.2 b ). From the MIDAS meteorological variables we also derived the apparent temperature (AT) using Equation 5.1, which considers the combined effects of temperature, humidity and wind (Steadman, 1994).

$$
\begin{equation*}
A T=T a+0.33 e-0.70 W s-4 \tag{5.1}
\end{equation*}
$$

Here $T a$ is the air temperature in ${ }^{\circ} \mathrm{C}, e$ is the water vapour pressure in hPa calculated from the relative humidity and temperature, and $W s$ is the wind speed in $\mathrm{m} / \mathrm{s}$.

The Weather Cam data was used to calculate dusk, sunset, sunrise and dawn times (for a central location in the study area) as at this latitude daylight length varies by approximately 4.5 hours from September to January. Daylight categories were annotated to trajectories according to the following rules: Morning Twilight (MT) for GPS points recorded in the period between dawn and sunrise, Day Light (DL) for GPS points recorded between sunrise and sunset, Evening Twilight (ET) for GPS points recorded between sunset and dusk, Night (NI) for GPS points recorded between dusk and dawn.

### 5.2.3 Trajectory sequencing

Sequence analysis requires a finite alphabet, in which each letter originally represents genomic nucleotides (Idury and Waterman, 1995). In single channel sequence analysis, a sequence is a one dimensional ordered list of characters from one alphabet, representing successive states (Abbott and Tsay, 2000). However, most phenomena are multidimensional and require multiple alphabets. This means that each dimension gets its own bespoke alphabet and instead of having the data object represented as one sequence, the object now has as many different sequences as there are dimensions, which are called channels (therefore the name Multi-Channel Sequence Analysis). The alignment, i.e. similarity, then needs to be calculated across all channels (Gauthier et al., 2010). This multi-channel approach is therefore a shift from looking at individual units towards analysing context, connections and events (Abbott, 1995).

We created several bespoke alphabets, one for movement mode (e.g., walking and driving) and one for each weather variable in our data. For this, we had to translate the GPS track of each participant into a multi-channel sequence consisting of time units, to which the characters were assigned (Figure 5.3). Weather conditions were categorized to create weather-based alphabets (Table 5.2). Rainfall was classified based on the UK Met Office scale for rainfall intensity, wind Speed according to an adaptation of the Beaufort scale (Royal Meteorological Society, 2017), wind direction according to the cardinal and collateral points, apparent temperature according to the (VDI, 2008) thermal perception scale, humidity and temperature according to the 1991-2000 seasonal climate normals for Dunfermline from Jenkins et al. (2009). Climate normals are a three-decade average of weather variable commonly used to characterize local climates (Ayoade, 1986).

The multi-channel sequences were then generated for each volunteer and day (illustrated in Figure 5.3) by taking the modal weather condition (for each variable described in Table 5.2) and movement mode for each 1-minute interval for each participant. To each time unit we assigned descriptors for the weather variables and the respective movement mode, which are linked to the descriptor for the following time unit building multiple chronologically arranged strips. These sequences can be analysed alongside strips of contextual variables to understand not only
the responses to specific variables, but also to different combinations of those variables and the identification of patterns relative for specific age groups, gender or other profiling information. The number of channels in a MCSA is defined by the number of variables under consideration, in our case eight variables therefore, eight channels by definition. The use of modal attributes for each 60 second segment (as the data were collected at a 5 second frequency) filtered out possible noise from the raw data and represents an appropriate scale of analysis for studying human movement.


Figure 5.3: A multi-channel sequence for a participant over a five-minute period, each channel relates to one of the meteorological variables and movement modes for that minute of the day.

We calculated the entropy index (EI) for the movement mode channel for all sequences of at each minute (Billari, 2001). The EI is a measure of the complexity induced by the distribution of states in a group of sequences (Gabadinho et al., 2009), which in our case can be used to observe the diversity of places and travel modes across the week and hours of the day. In our analysis, an EI closer to one indicates an even distribution of a contextual variable across movement modes (alphabet states), while an EI closer to zero indicates a high level of association with one mode. We also looked at the average time expenditure at home, socialising, shopping, walk, public transport and vehicle by gender and on each day of the week. The average time
expenditure was calculated by first computing the amount of time spent in each movement mode and dividing it by the total GPS active time for each participant, keeping in mind that each state in our sequences corresponded to one minute. Following this, we calculated the mean for the gender of participants (male, female).
Table 5.2: Alphabets for meteorological variables used as contextual data with respective ranges and description. Letters in each alphabet are defined based on standard meteorological classifications (see text for more details).

| Thermal perception ( ${ }^{\circ} \mathrm{C}$ ) |  |  | Rainfall (mm/h) |  |  | Wind coming from direction ( ${ }^{\circ}$ ) |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Letter | Description | Range | Letter | Description | Range | Letter | Description | Range |
| VC | Very Cold | $<=-39$ | DR | Dry | 0 | N | North | >337.5-22.5 |
| CD | Cold | $>-39-26$ | VS | Very Slight | $>0-0.5$ | NE | North East | $>22.5-67.5$ |
| CL | Cool | $>-26-13$ | SL | Slight | $>0.5-1$ | E | East | $>67.5-112.5$ |
| SC | Slightly Cool | $>-13-0$ | LM | Low Moderate | $>1-2$ | SE | South East | >112.5-157.5 |
| CF | Comfortable | $>0-20$ | MO | Moderate | $>2-4$ | S | South | $>157.5-202.5$ |
| SW | Slightly Warm | $>20-26$ | HV | Heavy | $>4-10$ | SW | South West | $>202.5-247.5$ |
| W | Warm | >26-32 | VH | Very Heavy | $>10-50$ | W | West | $>247.5-292.5$ |
| H | Hot | $>32-38$ | VI | Violent | >50 | NW | North West | >292.5-337.5 |
| VH | Very Hot | >38 |  |  |  |  |  |  |
|  | Humidity(\%) |  |  | Temperature ( ${ }^{\circ} \mathrm{C}$ |  |  | Wind spee | (m/s) |
| Letter | Description | Range | Letter | Description | Range | Letter | Description | Range |
| EH | Extremely High | $>90$ | EL | Extremely Low | <=5 | CM | Calm | <= 3 |
| AA | Above Average | >85-90 | AN | Average minimum | >5-7 | BR | Breeze | >3-14 |
| AV | Average | >80-85 | AV | Average Average | $>7-10$ | GA | Gale | $>14-24$ |
| BA | Below Average | $>75-80$ | AX | Average Maximum | >10-13 | ST | Storm | $>24$ |
| LW | Low | $>70-75$ | EH | Extremely High | $>13$ |  |  |  |

### 5.2.4 Context-Aware Similarity Analysis (CASA)

### 5.2.4.1 Multi-channel Sequence Analysis (MCSA)

We divided our analysis into two streams, by separately analysing travel modes (walk, public transport and vehicle) and places (home, social places and shopping), since the choice of travel mode and of staying in a place are not necessarily affected in the same way by weather (Derrick Sewell et al., 1968). When the destination is obligatory, such as work, people are more likely to change their travel mode, for example, driving to work instead of walking under heavy rain; however, if the destination is linked to leisure, such as shopping, people might simply postpone the task instead of changing the travel mode to get there (Connolly, 2008; Zivin, 2014). We further split the analysis into weekdays and weekends to reflect different movement motivations (for example, travel to work during workdays is usually obligatory regardless of weather conditions while people have more voluntary choices about their mobility during weekends).

Sequence analysis requires cost matrices, which were computed separately for travel modes and places and for weekends and weekdays. We used the optimal matching (OM) distance to compute similarity between sequences as this method has shown potential for identifying groups with matching movement behaviour De Groeve et al. (2016). The distance between two sequences is assessed by quantifying their differences based on a matrix with the costs for substituting, deleting or inserting letters to transform one sequence into the other. The substitution costs are given by symmetrical matrices that represent the costs of transitioning between each pair of states in the alphabet (Gabadinho et al., 2009). In our case, the costs for transitions between the states of travel modes, places, wind speed and wind direction were computed using transition rates calculated from the sequences for computing the cost matrices, as shown in Equation 5.2.

$$
\begin{equation*}
F(i, j)=1-P(i, j)-P(j, i) \tag{5.2}
\end{equation*}
$$

Here $F(i, j)$ is the substitution cost and $P(i, j)$ is the transition rate from state $i$ to $j$.
The costs for transitions between the states of thermal comfort, temperature, humidity,
daylight and rainfall were defined by ordering the classes of each variable (alphabets) by their intensity and calculating the cost to replace one class by another with Equation 5.3.

$$
\begin{equation*}
F\left(i_{n}, j_{n+1}\right)=\frac{|n-(n+1)|}{(z-1)} \tag{5.3}
\end{equation*}
$$

Here $F\left(i_{n}, j_{n+1}\right)$ is the cost between the classes $i$ and $j$ with intensity order n and $\mathrm{n}+1$, and z is the number of classes for that variable (size of the alphabet). The cost for replacing null values by any other class (insertion) was zero and likewise to substitute any other class by null (deletion), because for our study they are related to periods for which we had no information on the participant's movement. This procedure resulted in ten cost matrices, one for each weather variable, two for travel modes and two for places (weekdays and weekend). The cost matrices are then used to calculate the optimal match (OM) score, for example, given an alphabet $A$ with size $Z$, pick sequences $I$ and $J$ based on alphabet $A$. The sequences are aligned in time and the OM cost is calculated by summing up the costs of substitutions ( $C_{S_{i} S_{j}}$ ), deletions and insertions (d) needed to modify the sub sequences of $J$, so that it turns into $I$. The OM is the less costly and is computed using Equation 5.4, in which each line defines a possible OM score for two sub sequences, depending on which of the procedures, insertion, deletion or substitution, is cheaper (Gauthier et al., 2010).

$$
F(i, j)=\min \left\{\begin{array}{l}
F(i-1, j-1)+C_{S_{i} S_{j}}  \tag{5.4}\\
F(i-1, j)+d \\
F(i, j-1)+d
\end{array}\right.
$$

Here $F(i-1, j-1)$ represents the OM score of a subsequence containing the 1 to $i-1$ characters of sequence $I$ against a subsequence containing 1 to $j-1$ in sequence $J$ (Gabadinho et al., 2009; Gauthier et al., 2010). The OM cost is computed for each channel between all multi-channel sequences and the cost between two multi-channel sequences is the summed costs between their channels. We calculated the OM distances simultaneously considering three channels for wind: movement mode, wind speed and wind direction; and two channels for the remaining
weather conditions, where the places or the travel modes were always the first channel and the variables were considered in turns as the second channel. A $k$ by $k$ dissimilarity matrix, where $k$ is the number of sequences, represents the level of alignment between each two multi-channel sequences, i.e., a similarity measure between two moving people.

### 5.2.4.2 Cluster analysis and typology

The dissimilarity matrix can be used to find whether people were showing similar movement behaviour under certain weather conditions. For this we apply a clustering algorithm to the dissimilarity matrix for each weather variable for both travel modes and places. We used Ward's clustering, a hierarchical bottom-up algorithm that computes dissimilarities between two groups as the increase in the error sum of squares after merging those groups. The algorithm starts with each sequence as their own group and successively merges them into clusters based on the minimum increase in the error sum of squares, until it becomes a single cluster (Murtagh and Legendre, 2011). For selecting the optimal number of clusters, we used the CalinskiHarabaz Index (CHI) (Calinski and Harabasz, 1974) that considers the within and between groups dispersion as shown in Equation 5.5.

$$
\begin{equation*}
C H I=\frac{\operatorname{trace}(B)}{\operatorname{trace}(W)} \tag{5.5}
\end{equation*}
$$

Here $W$ and $B$ are the within and between group dispersion matrices, the trace of $W$ is the sum of the within cluster variance and the trace of $B$ is the sum of the between cluster variances; a higher CHI indicates a better data partition (Ahmed, 2012), because it shows that the within group distances are lower and the between groups distances are higher. We varied the number of clusters from the number of sequences (i.e. the maximum possible number of clusters, if every sequence is allocated to its own cluster) to one and used the configuration with highest CHI, except where the maximum CHI resulted in individuals' clusters, to assign the multi-channel sequences into their final clusters. The combination of values of weather and movement modes in each cluster then defined a type of the group. Note that the types are not consistent between variables, i.e., we found different clusters for each weather variable, thus the typology is specific
for each variable.
We then looked at the distribution of the proportion of time spent in different travel modes and places for the weather conditions associated with each cluster. We expected this would give insights into the different behavioural patterns in individuals related to weather (i.e., an overall picture of the effects of the weather conditions within each cluster on movement modes). We tested the significance of the differences using Kruskal-Wallis and Levene's tests and we assumed that a statistically significant difference between medians or variances of each cluster was enough evidence to support the existence of different behavioural groups. We further used discrepancy analysis to verify if and how behavioural groups were related to age and gender. This method evaluates the strength of the association between the groups of sequences and a categorical covariate (Studer et al., 2011) by calculating the share of discrepancy according to Equation 5.6 and looking at its p-value.

$$
\begin{equation*}
S D=\frac{S S_{B}}{S S_{T}} \tag{5.6}
\end{equation*}
$$

Here $S D$ is the share of discrepancy, $S S_{B}$ is the sum of square distances within the age or gender groups, and $S S_{T}$ is the total sum of square distances between all sequences (Batagelj, 1988).

### 5.3 Results

### 5.3.1 Trajectory sequencing

The different movement modes for each participant for each day of the week are shown in Figure 5.4, most sequences start between six and eight in the morning and have a minimum of $58 \%$ and a maximum of $98 \%$ of missing data, i.e., minutes for which the GPS tracker was off and movement modes are unknown (white gaps in Figure 5.4). The entropy index (EI) (5.5) provides some insight into participants daily movement behaviour. The EI increases between 4:00 am and 7:00 am on weekdays, but only rises between 8:00 am and 10:00 am on weekends, indicating higher diversity of movement modes earlier on weekdays. Sunday has the highest

EI and similarly to Saturday, it drops and rise between 3:00 pm and 6:00 pm.
The average time spent (AVTS) walking did not change substantially across weekdays and between genders (Figure 5.6). The AVTS at home varied throughout the week, being the highest on Sunday and lowest on Wednesday for both genders (dashed orange lines on Figure 5.6). The low values on Wednesday might be related to the higher average time spend socializing in comparison to other days of the week (dashed red lines on Figure 5.6). Moreover, women seem to spend more time socializing and to concentrate social activities on Tuesdays, Wednesdays and Saturdays; while men socialise very little on Tuesdays and keep a steady, but lower than women, average from Wednesday to Monday.


Figure 5.4: Sequences of movement modes for each day of the week by participant (vertical axis) and by hour of the day (horizontal axis), missing data are reported in white. For each day, the sequences are ordered by length, but orderings are different for each day (that is, the sequence number on the y chart does not always indicate the same individual).








Figure 5.5: Entropy index (EI) by day of the week and by hour of the day (horizontal axis), an EI close to zero indicates that most
 on the different movement modes (higher diversity).

Figure 5.6: Average time expenditure and trend over the week (dashed lines) by movement mode, day of the week and gender. The first row refers to females $(22 \leq \mathrm{n} \leq 33)$ and the second row to males $(54 \leq \mathrm{n} \leq 70)$

### 5.3.2 Context-Aware Similarity Analysis (CASA)

### 5.3.2.1 Cluster analysis and typology

Overall the CHI was higher on weekdays for all weather variables for travel modes and places (Figure 5.7) indicating the existence of a clear division amongst behavioural groups and a more homogeneous movement behaviour within these groups regarding the weather effects in comparison to the weekend. This could also be reflective of the higher diversity of activities during the weekends shown by the higher EI; more variety might lead to lower separability between groups and higher within group distances making more difficult to identify group's responses to weather. This usually results in a higher optimum number of cluster, as seen on the weekend chart (Figure 5.7 right). On weekdays the CHI followed a similar pattern for travel modes and places for all weather variables but relative humidity, for which the index was about two times higher for travel modes (Figure 5.7 left). This indicates a stronger distinction between the two behavioural groups regarding travel modes and relative humidity on weekdays, which might be related to people using relative humidity as a proxy for rainfall to plan their journeys. Relative humidity, rain and comfort showed the highest discernibility for travel mode differences during the week, while for places the highest discernibility was associated with temperature, rain and comfort during the weekend.

Next, we present our findings while the remaining set for each meteorological variable. The typologies are specific for each variable, i.e., Type 1 for wind is not the same as type 1 for rainfall. The analysis of shared discrepancy did not show significant correlation between the behavioural clusters and gender or age groups, all SD were lower than 0.01 with nonsignificant p-values ( $\alpha=0.1$ ). Significant values for Levene's (L) and Kruskal-Wallis' tests (K) are reported on the heading of each graph on the pictures by the following symbology: ${ }^{* * *}$ for $\alpha=0.001,{ }^{* *}$ for $\alpha=0.01,{ }^{*}$ for $\alpha=0.5$, for $\alpha=0.1$.

### 5.3.2.2 Wind

Figure 5.8 shows the clusters for MCSA on wind on weekdays (Figure 5.8 A ) and weekends (Figure 5.8B). The top box-plot shows the distribution of the GPS active time spent under wind


Figure 5.7: Calinski-Harabaz index (solid line) and optimal number of clusters (dashed line) for MCSA performed on weather variables for travel modes (green) and places (red) on weekdays and on weekend. The reported CHI is divide by ten and refers to the number of clusters used to split the sequences.
blowing from each direction and the middle one shows the distribution of the GPS active time spent under different wind intensities. Both box-plot panels are divided into Type 1 and Type 2, which refer to the two clusters found by the MCSA analysis and for which the distribution of the GPS active time in different travel modes is shown on the box-plot panel at the bottom. This box-plot shows the difference between groups with different distribution of time spent on travel modes, while the remaining panels describe the wind conditions encountered within those groups. There were no significant differences on the average time spent on different travel modes on weekdays under different wind conditions (Figure 5.8A), on the weekend however we found significant differences on the average time expenditure in public transport and vehicle (Figure 5.8B). CASA clustering showed a significantly lower use of public transportation with concurrent increase on the use of vehicles under more windy conditions coming from NorthEast, North-West and South-West (Type 2).

Figure 5.9 shows the clusters for MCSA on wind on weekdays (Figure 5.9A) and weekends


Figure 5．8：Clusters for MCSA on wind speed，wind direction and travel modes on weekdays （A）and weekends（B）．The four top panels describe the wind conditions within each cluster （Types）and the respective proportions of GPS active time spent under the classes of wind speed and direction．The two panels at the bottom show box－plots with the distribution of proportional GPS active time spent on each travel mode by the wind types described on the panels above．The dashed line on box－plots show the average and the continuous line the median．L reports significance from Levene＇s test and K from Kruskal－Wallis＇test．
（Figure 5．9B）．The top boxplot shows the distribution of the GPS active time spent under wind blowing from each direction and the middle one shows the distribution of the GPS active time spent under different wind intensities．Both box－plot panels are divided into five types on weekdays and two types on weekends，which refer to the clusters found by the MCSA analysis and for which the distribution of the GPS active time in different activities is shown on the box－plot panel at the bottom．This boxplot shows the difference between groups with different distribution of time spent on activities，while the remaining panels describe the wind
conditions encountered within those groups. For places, there were significant differences in the average time expenditure at home and shopping during the week, and in socialising on the weekend (Figure 5.9). There were five clusters based on wind during weekdays, but Type 1, Type 3 and Type 4 are very similar in terms of time spent at places and they do not show any pattern in terms of wind direction and strength. We are not able to draw conclusions about Type 5 because of its small number of participants ( $\mathrm{n}=7$ ); Type 2 however, showed a lower proportion of time spent at home with concurrent increase of time spent shopping under more windy conditions. CASA clustering on weekend showed a significant decrease on the proportional time spent socialising under more windy conditions coming from NE, NW and SW (Type 2). Whereas weekend Type 1 does not show any prevailing direction and its strength alternates between calm and gale for around $88 \%$ of the time.

### 5.3.2.3 Rain

Figure 5.10 shows the clusters for MCSA on rainfall on weekdays (Figure 5.10A) and weekends (Figure 5.10B). The top box-plot shows the distribution of the GPS active time spent under different rainfall intensities. The box-plot panel is divided into three types on weekdays and four types on weekends, which refer to the clusters found by the MCSA analysis and for which the distribution of the GPS active time in different travel modes is shown on the box-plot panel at the bottom. This box-plot shows the difference between groups with different distribution of time spent on travel modes, while the remaining panels describe the rainfall conditions encountered within those groups. There were no significant differences in the average time expenditure for different travel modes on weekends under different rain conditions (Figure 5.10B), on weekdays however we found significant differences in the average time spent in public transport (Figure 5.10A). CASA clustering showed that in comparison to more drier conditions (Type 1 and Type 2), public transport is significantly less used under heavy rainfall (Type 3) with a concurrent, but not statistically significant, increase on the use of vehicles and decrease on walking.

Figure 5.11 shows the clusters for MCSA on rainfall on weekdays (Figure 5.11A) and weekends (Figure 5.11B). The top box-plot shows the distribution of the GPS active time spent


Figure 5.9: Clusters for MCSA on wind speed, wind direction and places on weekdays (A) and weekends (B). The four top panels describe wind conditions within each cluster (Types) and respective proportions of GPS active time spent under the classes of wind speed and direction. The two panels at the bottom show box-plots with the distribution of proportional GPS active time spent on places by the wind types described on the panels above. The dashed line on box-plots show the average and the continuous line the median. L reports significance from Levene's test and K from Kruskal-Wallis' test.
under different rainfall intensities. The box-plot panel is divided into three types on weekdays and six types on weekends, which refer to the clusters found by the MCSA analysis and for which the distribution of the GPS active time in different activities is shown on the box-plot panel at the bottom. This box-plot shows the difference between groups with different distribution of time spent on activities, while the remaining panels describe the rainfall conditions encountered within those groups. The only significant difference for places was on the average time expenditure at home on weekends and weekdays under different rain conditions (Figure


Figure 5.10: Clusters for MCSA on rain and travel modes on weekdays (A) and weekends (B). The two top panels describe the rain conditions within each cluster (Types) and the respective proportions of GPS active time spent under the rainfall classes. The two panels at the bottom show box-plots with the distribution of proportional GPS active time spent on each travel mode by the rain type described on the panel above. The dashed line on box-plots show the average and the continuous line the median. L reports significance from Levene's test and K from Kruskal-Wallis' test.
5.11). On weekdays we found one cluster with predominantly dry conditions (Type 1), a second cluster with dry conditions but an even distribution of time amongst the other rain states (Type 2) and a third cluster with violent rain (Type 3). As expected, people spend more time at home under violent rain (Type 3), but surprisingly people also spend more time at home under predominantly dry conditions (Type 1) in comparison when there is a mix of dry and different rain conditions (Type 2).

On weekends, Type 5 and Type 6 have such a sparse number of members that we considered them outliers. Under heavy rainfall (Type 3) there is a higher average time expenditure at home, while less time is spent at home under predominantly dry conditions, even with the


Figure 5.11: Clusters for MCSA on rain and places on weekdays (A) and weekends (B). The two top panels describe the rain conditions within each cluster (Types) and the respective proportions of GPS active time spent under the rainfall classes. The two panels at the bottom show box-plots with the distribution of proportional GPS active time spent on places by the rain type described on the panel above. The dashed line on box-plots show the average and the continuous line the median. L reports significance from Levene's test and K from KruskalWallis' test.
remaining time being almost evenly distributed amongst the other rain conditions (Type 2 and Type 4); the driest conditions (Type 1) showed the higher time expenditure at home.

### 5.3.2.4 Daylight

Figure 5.12 shows the clusters for MCSA on daylight on weekdays (Figure 5.12A) and weekends (Figure 5.12B). The top box-plot shows the distribution of the GPS active time spent under different light conditions. The box-plot panel is divided into three types on weekdays and two types on weekends, which refer to the clusters found by the MCSA analysis and for which the distribution of the GPS active time in different travel modes is shown on the box-plot panel
at the bottom. This box-plot shows the difference between groups with different distribution of time spent on travel modes, while the remaining panels describe the daylight conditions encountered within those groups. There were no significant differences on the average time expenditure for different travel modes on weekends nor on weekdays under different daylight conditions (Figure 5.12).


Figure 5.12: Clusters for MCSA on daylight and travel modes on weekdays (A) and weekends (B). The two top panels describe daylight conditions within each cluster (Types) and respective proportions of GPS active time spent under the classes of daylight. The two panels at the bottom show box-plots with the distribution of proportional GPS active time spent on each travel mode by the daylight type described above. The dashed line on box-plots show the average and the continuous line the median. L reports significance from Levene's test and K from Kruskal-Wallis' test.

Despite not being statistically significant, less daylight time resulted in less time walking (Type 3) compared to more time walking under more daylight hours (Type 1 and Type 2).

A similar decrease is observed on the time expenditure in public transport, with a concurrent increase on time expenditure in vehicles. This trend reverses on weekends, in which walking and public transport are more prominent than the use of vehicle in a group exposed to more night hours (Type 2), while the use of vehicles prevails in a group with more daylight hours (Type 1).

Figure 5.13 shows the clusters for MCSA daylight on weekdays (Figure 5.13A) and weekends (Figure 5.13 B ). The top box-plot shows the distribution of the GPS active time spent under different light conditions. The box-plot panel is divided into three types on weekdays and two types on weekends, which refer to the clusters found by the MCSA analysis and for which the distribution of the GPS active time in different activities is shown on the box-plot panel at the bottom. This box-plot shows the difference between groups with different distribution of time spent on activities, while the remaining panels describe the daylight conditions encountered within those groups. The analysis for daylight and places, was significant for all places both on weekend and weekdays. There are 3 daylight types on weekdays, Type 1 has more daylight, Type 2 is the one with more night time and Type 3 is an almost even mix of day and night (Figure 5.13). Type 1 has less time spent at home than Type 2, however we are unsure why the time expenditure at home is the lowest for Type 3. There is less shopping and socialising in the group with more night hours (Type 2). On weekends (Figure 5.13 B ). more time is spent at home under brighter conditions (Type 2), while under lower light conditions (Type 1) more time is spent shopping and socialising.

### 5.3.2.5 Comfort

Figure 5.14 shows the clusters for MCSA on comfort on weekdays (Figure 5.14 A ) and weekends (Figure 5.14 B ). The top box-plot shows the distribution of the GPS active time spent under different comfort conditions. The box-plot panel is divided into two types, which refer to the clusters found by the MCSA analysis and for which the distribution of the GPS active time in different travel modes is shown on the box-plot panel at the bottom. This box-plot shows the difference between groups with different distribution of time spent on travel modes, while the remaining panels describe the thermal comfort conditions encountered within those


Figure 5.13: Clusters for MCSA on daylight and places on weekdays (A) and weekends (B). The two top panels describe daylight conditions within each cluster (Types) and respective proportions of GPS active time spent under daylight classes. The two panels at the bottom show box-plots with the distribution of proportional GPS active time spent on places by the daylight type described above. The dashed line on box-plots show the average and the continuous line the median. L reports significance from Levene's test and K from Kruskal-Wallis' test.
groups. The only significant difference for comfort and travel modes happened on weekdays for the average time spent in vehicles under different comfort conditions (Figure 5.14). Slightly uncomfortable conditions (Type 2) were associated with significant higher use of vehicles, less walking and less use of public transport. For the weekend participants were split into one large group and an individual, therefore limiting interpretation. There were no significant differences or meaningful visual patterns from the average time expenditure on different places both on weekdays and weekends under different comfort levels (Figure 5.15).


Figure 5．14：Clusters for MCSA run on comfort and travel modes on weekdays（A）and week－ ends（B）．The two top panels describe comfort conditions within each cluster（Types）and respective proportions of GPS active time spent under comfort classes．The two panels at the bottom show the distribution of proportional GPS active time spent on each travel mode by the comfort type described above．The dashed line on box－plots show the average and the con－ tinuous line the median．L reports significance from Levene＇s test and K from Kruskal－Wallis＇ test．

## 5．3．2．6 Humidity

Figure 5.16 shows the clusters for MCSA on relative humidity on weekdays（Figure 5．16A） and weekends（Figure 5．16B）．The top box－plot shows the distribution of the GPS active time spent under different relative humidity conditions．The box－plot panel is divided into types， which refer to the clusters found by the MCSA analysis and for which the distribution of the GPS active time in different travel modes is shown on the box－plot panel at the bottom．This box－plot shows the difference between groups with different distribution of time spent on travel


Figure 5．15：Clusters for MCSA run on comfort and places on weekdays（A）and weekends （B）．The two top panels describe comfort conditions within each cluster（Types）and respective proportions of GPS active time spent under comfort classes．The two panels at the bottom show box－plots with the distribution of proportional GPS active time spent on places by comfort type described above．The dashed line on box－plots show the average and the continuous line the median．L reports significance from Levene＇s test and K from Kruskal－Wallis＇test．
modes，while the remaining panels describe the meteorological conditions encountered within those groups．The only significant difference for relative humidity and travel modes happened on weekdays on the average time spent walking under different relative humidity（Figure 5．16）． It seems that more humid conditions（Type 1）were associated with a significantly higher time expenditure walking．On weekends the time spent in public transport is visually higher when humidity is lower（Type 2）．

Figure 5.17 shows the clusters for MCSA relative humidity on weekdays（Figure 5．17A） and weekends（Figure 5．17B）．The top box－plot shows the distribution of the GPS active time


Figure 5.16: Clusters for MCSA run on relative humidity and travel modes on weekdays (A) and weekends (B). The two top panels describe relative humidity conditions within each cluster (Types) and respective proportions of GPS active time spent under relative humidity classes. The two panels at the bottom show box-plots with the distribution of proportional GPS active time spent on each travel mode by relative humidity type described above. The dashed line on box-plots show the average and the continuous line the median. L reports significance from Levene's test and K from Kruskal-Wallis' test.
spent under different relative humidity conditions. The box-plot panel is divided into two types, which refer to the two clusters found by the MCSA analysis and for which the distribution of the GPS active time in different activities is shown on the box-plot panel at the bottom. This box-plot shows the difference between groups with different distribution of time spent on activities, while the remaining panels describe the meteorological conditions encountered within those groups. For the analysis on places (Figure 5.17), there were significant differences on time spent at home and socialising on the weekend. Also, when relative humidity is lower
(Type 2), more time is spent at home and less time is spent socialising, while it goes the other way around for Type 1, for which the relative humidity is higher. We believe that these patterns are more related to rain than to relative humidity, as higher humidity is closely related to probability of rain.


Figure 5.17: Clusters for MCSA run on relative humidity and places on weekdays (A) and weekends (B). The two top panels describe relative humidity conditions within each cluster (Types) and respective proportions of GPS active time spent under relative humidity classes. The two panels at the bottom show box-plots with the distribution of proportional GPS active time spent on places by relative humidity type described above. The dashed line on box-plots show the average and the continuous line the median. L reports significance from Levene's test and K from Kruskal-Wallis' test.

### 5.3.2.7 Temperature

Figure 5.18 shows the clusters for MCSA on temperature on weekdays (Figure 5.18A) and weekends (Figure 5.18B). The top box-plot shows the distribution of the GPS active time spent under different temperature conditions. The box-plot panel is divided into two types, which refer to the clusters found by the MCSA analysis and for which the distribution of the GPS active time in different travel modes is shown on the box-plot panel at the bottom. This box-plot shows the difference between groups with different distribution of time spent on travel modes, while the remaining panels describe the meteorological conditions encountered within those groups. For temperature and travel modes the only significant differences happened on weekdays for the average time spent walking and on weekends on public transport (Figure 5.19). Higher temperatures on weekdays (Type 1) led to a significant higher time expenditure walking, and slightly less time spent on public transport under temperatures close to the average historical maximum (Type 2). On the other hand, extremely elevated temperatures (Type 1) show more time spent on public transport. Types 3,4 and 5 had a low number of trajectories $(\mathrm{n}<9)$ assigned to them and were considered outliers on which we cannot not draw conclusions.

### 5.4 Discussion

The recent widespread availability and quality of geospatial data on movement and context presents opportunities for developing new methods to understand the interactions between movement behaviour and environment. We were interested on how weather affects human movement, in particular the choice of travel mode and time spent on activities. Our methodology was efficient in identifying groups of specific behaviour under certain weather conditions, and it can be expanded to other types of movement and contextual data. We investigated the impact of wind (strength and direction), rainfall, daylight, comfort, relative humidity and temperature, on the proportion of GPS active time spent on travel modes (walk, public transport, vehicles) and places (home, shopping, socialising). Differences were observed between the time expenditure on different travel modes and places across the day, week and between genders.


Figure 5.18: Clusters for MCSA run on temperature and travel modes on weekdays (A) and weekends (B). The two top panels describe temperature conditions within each cluster (Types) and respective proportions of GPS active time spent under temperature classes. The two panels at the bottom show box-plots with the distribution of proportional GPS active time spent on each travel mode by temperature type described above. The dashed line on box-plots show the average and the continuous line the median. L reports significance from Levene's test and K from Kruskal-Wallis' test.

The analysis of the entropy index (EI) showed a high diversity of movement modes in the early morning on weekdays and weekends, with a positive shift of three hours on weekends. Horanont et al. (2013) found the same entropy pattern, despite analysing weekdays and weekends together, when using GPS traces from mobile phone to explore the effects of weather on daily routine. We found that during weekdays there is a drop with subsequent rise on the EI, which does not exist on weekends because the EI is higher from 10 am throughout the afternoon. Horanont et al. (2013) found a very similar variation for specific extreme weather conditions,


Figure 5.19: Clusters for MCSA run on temperature and travel modes on weekdays (A) and weekends (B). The two top panels describe temperature conditions within each cluster (Types) and respective proportions of GPS active time spent under temperature classes. The two panels at the bottom show box-plots with the distribution of proportional GPS active time spent on places by temperature type described above. The dashed line on box-plots show the average and the continuous line the median. L reports significance from Levene's test and K from Kruskal-Wallis' test.
according to meteorological information provided by the authors, which they attributed to the weather conditions. However, we believe it is related to similar differences to the ones we found between weekdays and weekend, and that it is more likely that the extreme weather events reported by Horanont et al. (2013) took place on a weekend. Similarly to Ryan et al. (2010), we found that people have more varied activities on weekends, which was shown by the highest EI on Sunday and Saturday. This happens because people have more scope for freedom of action on weekends, in contrast to the external controls imposed on weekdays by work and
school (Ryan et al., 2010).
The average time spent (AVTS) walking did not change substantially across weekdays and between genders (Figure 5.6). The AVTS at home varied throughout the week, being the highest on Sunday and lowest on Wednesday for both genders (dashed orange lines on Figure

Similarly to Stover et al. (2012), we found that the wind strength and direction exerted considerable influence on weekends on the use of public transport and vehicles, which is possibly related to traffic restrictions at the Tay and Forth bridges (See dashed ellipses in Figure 5.2A) under high winds, which are more likely to come from NW, SW and NE in the Central Belt of Scotland. There were at least ten occasions during our data collection during which the bridges were either closed or had restrictions on the type of vehicle and speed limits because of high winds (Traffic Scotland (@trafficscotland), 2017). It is possible that these restrictions are reflected in our findings during these windy periods, since the participants in our study were mostly commuters from Fife to Edinburgh or Dundee (Siła-Nowicka et al., 2016) and therefore typically have to cross one of these two bridges daily.

As opposed to what Guo et al. (2007) found in Chicago, rain during the weekend had no key role on travel modes, but heavy rain decreases the use of public transport during the week. This could be explained by the fact that discretionary passengers are more affected by rain than commuters (Changnon, 1996), i.e., people are obliged to go out for their daily duties on weekdays and therefore might adapt their travel modes, while on weekends they can opt to stay at home under heavy rain. In addition, similarly to what Chen et al. (2017) found when studying the impact of rainfall on taxi use, we also found a trend of more vehicular use under rainy weather, and less walking in heavier rain.

We found that daylight length seems to factor into mobility decisions differently on weekends in comparison to weekdays. During the week, less daylight hours were linked to less walking and less public transport use, but more vehicular use; on weekends the same daylight conditions resulted in the opposite pattern It is not clear why this may be. In addition, daylight seems to play a major role on time expenditure at certain places; weekdays with more dark hours are more likely to be spent at home, while more time is spent at home on weekends under more daylight hours. Temperature increase seems to have a positive effect on walking (Cools et al.,

2010; Tucker and Gilliland, 2007), which makes sense in Scotland because the temperate and oceanic climate gives people more opportunities for outdoor activities during summer. It is likely that in places with more tropical climates the temperature effect would be different, in the USA for example, areas with a more tropical weather showed a decline in physical activity on hotter months, while areas with cold weather showed an increase on warmer months (Tucker and Gilliland, 2007).

The application of multi-channel sequence analysis on semantic trajectories was efficient for identification of movement patterns. Even though our method does not use the exact coordinates, the multi-channel sequences keep the spatial component through places and travel modes, which allows us to link movement patterns to environmental conditions and identify responses. Our methodology works both with categorical and numerical contextual data, considers the change of context between two timestamps, is able to handle multiple contextual variables and their interactions at once, and can deal with contextual data in any the form. These capabilities make it more able to deal with complex contextual situations than previous methodologies, such as those established by Sharif and Alesheikh (2017a); De Groeve et al. (2016); Buchin et al. (2014).

MCSA clusters are useful for simplifying the increasingly large and complex tracking datasets, the creation of typologies allows the generalization and reduction of thousands of trajectories to a few representative trajectories. In addition, MCSA can help with the recent increasing demand for Context-Aware methods, as it is able to perform similarity analysis taking context into account and also allows for visualization of movement patterns and contextual variables simultaneously along the time axis. Another advantage here is that the time units are flexible, i.e., the sequences can be arranged at daily, weekly, monthly or hourly scale, which allows for multi-scale detection of movement patterns.

To summarise, multi-channel sequence analysis represents a new analysis tool for movement data where contextual information can now be readily combined with detailed tracking datasets. The main advantage of this approach is that it also is possible to consider as many channels (variables) as desired at once. It is common in movement research to simultaneously consider multiple environmental variables, which makes MCSA particularly relevant for
studying human mobility, traffic, transportation and wildlife ecology; areas in which movement behaviour may be contextualised by other dynamic environmental variables such as air temperature, vegetation indices, humidity, wind speed, air pollution and snow coverage. MCSA can help performing Context-Aware Similarity Analysis (CASA), which improves our understanding of how movement is affected by the combination of multiple contextual variables. In addition, MCSA is a good approach to summarise large movement dataset into clusters expressing specific typologies, i.e., a group of similar movement patterns.

## Chapter 6

## Conclusions

In this thesis we have attempted to contribute towards advancing the field of CAMA by investigating its challenges and developing methods capable of overcoming them. Specifically, we asked the following three questions:

1. How can CAMA methods properly account for the temporal dynamics of contextual data (e.g., contextual factors that change over time)?
2. How can CAMA methods better accommodate issues associated with data structures in contextual data (e.g., challenges posed by different spatial and/or temporal resolutions, data representations, etc.)?
3. How can we make meaningful inferences on behaviour from contextualized movement data using modern computational methods?

Aiming to answer these questions, we addressed three related research objectives: 1) assess the state-of-the art for CAMA within movement ecology and human mobility research; 2) develop innovative methods to take into account the spatio temporal differences between movement data and contextual data; and 3) explore computational methods that allow for meaningful inferences from contextualized movement data.

We tackled these research objectives by combining the concepts covered in Chapters 2 and 3 with methods from other research areas, which we then used to analyse movement data in

Chapters 4 and 5.

### 6.1 Key findings

How can CAMA methods properly account for the temporal dynamics of contextual data?
In order to answer this question, we designed a comparative experiment amongst temporal interpolation methods used for TA within CAMA. First we formalized a dynamic trajectory annotation (DTA), which was then compared to interpolation methods which are most commonly used for TA, namely: NB (Neighbour Before), NA (Neighbour After), NN (Nearest Neighbour) and AM (Arithmetic Mean). The comparison was based on the annotation of simulated trajectories with rainfall data from meteorological radars, for which we had real accumulated rainfall data to be used as ground truth.

The DTA method was superior for modelling rainfall mass curves and it was as accurate as the NN method. We therefore recommend it for CAMA in scenarios where representing continuously varying phenomenon is of high importance. Its capability of capturing gradual changes and preserving peaks and valleys from the original data makes the DTA method a good choice when attempting to elicit the relationships between the environmental variable and fine-scale movement patterns. This is not restricted to the accumulation of rainfall, and the recommendation can be extended to any environmental variable whose behaviour between two temporal points can be approximated as a linear function over time. Other environmental variables may change differently over time and for these, the DTA can be extended from linear interpolation into more complex forms, for example $2^{\text {nd }}$ or $3^{\text {rd }}$ order polynomials, using an appropriate temporal function. However, our research was limited to one contextual variable, rainfall, and the application of DTA to other variables might require further validation.

CAMA methods can properly account for the temporal dynamics of contextual data by using more complex interpolation methods that are capable of modelling the progression of the contextual variable. Simpler methods, such as NN and AM are well suited for situations where absolute values are more relevant than progression, or when the differences between the values of the contextual variable are more pronounced. In addition, the choice of interpolation
method should take into account not only the behaviour of the contextual variable being modelled, but also the structure in which this variable has been recorded and represented, and how this variable will be considered in the movement analysis. These considerations, along with computational costs, should guide the choice of the most appropriate TA method for the analysis.

How can CAMA methods better accommodate issues associated with data structures in contextual data?

In order to answer this question, we developed a multi-source disaggregation approach for remotely sensed data to produce contextual datasets with a higher temporal resolution and level of detail for CAMA. That is, we obtained NDVI data from several satellites with varying spatial and temporal resolution and designed a new methodology to create a fused NDVI data with increased temporal resolution and level of detail for CAMA. This approach has not been attempted before in movement research and consists of using daily MODIS images with a 250 m spatial resolution and a finer image (with a 15 to 30 m spatial resolution) collected fortnightly to derive daily images with the same spatial resolution as the finer source images. The purpose of this approach is to create a new disaggregated NDVI data, which can then be linked to movement trajectories to detect finer scale movement patterns linked to contextual changes.

We have tested this methodology in a case study of maned wolves Chrysocyon brachyurus, the largest south American canids which live in the Brazilian Cerrado. We explored the seasonality in the diet of maned wolves to test the hypothesis that our approach would enable the detection of finer scale movement patterns linked to contextual changes. More specifically, we wanted to see if we could identify from our semantically enriched movement data the diet with a temporal trophic opportunistic pattern indicated by previous studies. Our results seem to support this hypothesis, as all wolves showed higher interval between visits to high-NDVI locations when the home range distribution showed lower food availability, and more frequent visits when food availability was higher.

The approach we proposed in this thesis capitalises on the strengths of multiple satellites by using data fusion to produce a series of contextual layers for CAMA. This is the first time that contextual data from multiple satellites is used to annotate trajectories and perform CAMA.

Usually, studies are restricted to a single source of contextual data, which is commonly a preprocessed contextual variable (Urbano and Cagnacci, 2014; Bühne and Pettorelli, 2017). A higher level of spatial detail alongside a better coverage produced by our method enabled us to capture the seasonal fluctuations of context within the home ranges of individual wolves, which was particularly interesting for analysing the values of the contextual variable that were used in comparison to the ones available in the area.

Analysing contextualised trajectories is a cumbersome task and most algorithms disregard contextual data in the process, which is one of the current drawbacks of methods for movement analysis (Buchin et al., 2012). The most common approaches for analysis of semantic trajectories are the use of map animation or space-time cubes, which are both limited in the number of trajectories it can show, the time period it can cover, and the ability to represent contextual variables (Andrienko et al., 2011). This chapter suggests the use of eigendecomposition as an alternative approach for analysis of contextualised trajectories, a sequence-method that uses a principal components analysis disaggregation to identify the most important sequences in the data (which in movement context correspond to the most frequent behaviour patterns). This method does not have a limitation on the number of trajectories nor on time period to be covered in the analysis. In addition, this method is less sensitive to gaps in the trajectories, as the PC's are calculated at each time and by behavioural state, which means that the "no data" sections will are kept as a separate dimension, without hindering the identification of the relevant behavioural patterns.

Capturing the seasonal fluctuations of context within the home ranges, which was particularly interesting for analysing the values of the contextual variable that were used in comparison to the ones available in the entire area. In that sense, the use of a typology for the distribution of the contextual variable and the transformation of the used values into Z-scores was efficient to assess whether individuals were using certain areas as a choice or only resource. It is also encouraging that the patterns we identified seem to correspond to what is know about maned wolves' diet from ecological studies. However more work is needed to confirm our results, potentially through other CAMA studies or traditional ecological observational research.

Multi-source image fusion is a new analysis tool for CAMA where contextual data is required
at higher temporal resolution and level of detail than readily available. The main advantage of this approach is that it can be applied to other species and other contextual variables, such as marine net primary productivity, land surface temperature, humidity, air pollution and snow coverages, which minimises information loss in that dimensions. It is common in movement research to simultaneously need daily, even hourly contextual data, but also high spatial resolution, particularly for studies in areas with heterogeneous environments.

CAMA methods can better accommodate issues associated with data structures in contextual data by looking beyond the pre-processed available products and exploring other sources of contextual data. We propose the use of multiple sources as the way forward for movement analysis where high spatial and temporal resolution are required. This work represents a novel contribution to movement ecology research by demonstrating the potential of multi-source image fusion for capturing fine scale environmental responses in animal tracking data. These methods are well developed in other research areas (e.g., forest monitoring, climate modelling) but remain limited in their application to movement modelling.

How can we make meaningful inferences on behaviour from contextualized movement data using modern computational methods?

In order to answer this question, we adapted techniques from other research fields and used them to perform CAMA. More specifically, we used a different representation paradigm of movement: instead of traditional representation of movement as trajectories we used the contextual information along with the temporal information from trajectories to create sequences that represented behavioural states of each individual. We then used two sequence analysis methods, the multi-channel sequence analysis (MCSA) to explore the effect of multiple weather variables on human movement, and the eigendecomposition to identify seasonal patterns in the use of vegetated areas and dietary composition of maned wolves. Such sequence methods are commonly used in longitudinal studies in demography (Abbott and Tsay, 2000) and in human behaviour (Eagle and Pentland, 2009), but this thesis is one of the first attempts of their application in the context of movement analysis.

The use of MCSA to explore how weather affects human movement, in particular the choice of travel mode and time spent on activities, was efficient in identifying groups of specific be-
haviour under certain weather conditions, and it can be expanded to other types of movement and contextual data. We investigated the impact of wind (strength and direction), rainfall, daylight, comfort, relative humidity and temperature, on the proportion of GPS active time spent on travel modes (walk, public transport, vehicles) and places (home, shopping, socialising).

The application of MCSA on semantic trajectories was further efficient for identification of movement patterns. Even though our method does not use the exact coordinates, the multi-channel sequences keep the spatial component through places and travel modes, which allows us to link movement patterns to environmental conditions and identify responses. Our methodology works both with categorical and numerical contextual data, considers the change of context between two timestamps, is able to handle multiple contextual variables and their interactions at once, and can deal with contextual data in any the form. These capabilities make it more able to deal with complex contextual situations than previous methodologies.

MCSA clusters are useful for simplifying the increasingly large and complex tracking datasets, the creation of typologies allows the generalization and reduction of thousands of trajectories to a few representative trajectories. In addition, MCSA can help with the recent increasing demand for Context-Aware methods, as it is able to perform similarity analysis taking context into account and also allows for visualization of movement patterns and contextual variables simultaneously along the time axis. Another advantage here is that the time units are flexible, i.e., the sequences can be arranged at daily, weekly, monthly or hourly scale, which allows for multi-scale detection of movement patterns.

The use of eigendecomposition allowed us to find patterns in how wolves use vegetated areas according to different seasons. It also reflected a well-known feeding habit for wolves, the so-called feast-famine regimen. The use of eigenbehaviours showed potential for studying contextualised trajectories, and the patterns we found are supported by current literature on the species. However there is a need for more studies where the patterns found can be compared to observational data on behaviour, so that the methodology can be validated. The use eigenbehaviours is a good approach to contextualise and summarise large movement datasets in order to enhance the understanding of how animals choose to use resources according to their availability.

### 6.2 Limitations

Despite contributing towards advancing the field of CAMA, the investigation conducted in this thesis has four main limitations. First, our studies are limited in number of contextual variables. In developing innovative methods to take into account the spatio temporal differences between movement data and contextual data, we only explored a few contextual variables but we believe that it would be beneficial to test our methods for the variables commonly used in the latest CAMA research, such as snow coverage (Therrien et al., 2015; Leblond et al., 2015), land cover (Ladin et al., 2018), atmospheric pressure (Liechti et al., 2018), wind fields (Safi et al., 2013; Shamoun-Baranes et al., 2004). There are at least 17 contextual variables that we did not cover in this thesis and are routinely available for trajectory annotation in Movebank (Dodge et al., 2008), therefore understanding how these variables behave under our methodology is relevant.

Second, the use of the methods developed in this study might be limited by their complexity. This is both in terms of computational complexity (i.e. time required to run the methods) and perceptual complexity, a factor that often prevents methods to be used by techincally less-savy users. As a an example of the first problem, the more sophisticated annotation methods, DTA:L and DTA:C, introduce extra computational costs, which might be a limitation for users with less powerful hardware and/or huge datasets and time constraints to analyse it. As an example of the latter problem, we are aware that the intricacy of some steps, such as the ones related to image fusion or absolute calibration of remote sensing images, can be off putting. Single source CAMA already requires the choice of appropriate spatio-temporal scales, interpolation techniques, data sources, formats, projection systems and data transformations, which are a challenge that limits many users from performing CAMA (Dodge et al., 2008). The addition of multiple sources of contextual data and more advanced remote sensing techniques tops-up the challenges, yet it is highly beneficial for movement studies (Neumann et al., 2015; Bühne and Pettorelli, 2017; Dodge et al., 2008) as such data and procedures are capable of enhancing the quality of semantic trajectories, which in turn will facilitate the identification of patterns of movement as a response to changes in context.

Third, our inferences on human and animal behaviour were limited by the lack of groundtruthing observational data. In exploring computational methods that allow for meaningful inferences from contextualized movement data, we found movement patterns which we hypothesized to be linked to specific events for humans or wildlife. We used events from the newspapers and behaviours previously reported in the literature to validate our findings. However without the appropriate observational data, we can only create hypotheses based on our findings and literature, which is not particularly helpful for development of new methodologies as their accuracy then cannot be easily assessed.

Fourth, our analysis was limited by the irregularity and uneven coverage of tracking data. For both case studies, humans and maned wolves, the tracking data were not consistent amongst individual and sometimes not even for the same individual. This was not a major issue for the wildlife study case, as we normalised the contextual variables for each day by the occurrence of that context in the home range. For the human mobility study, however, there was not a clear way to normalise the context during different days. We tried using climatological parameters for that, yet it is unclear how effective that was and maybe that is the reason why we had some clusters of outliers. Another issue was the irregularity of GPS trajectories in the wildlife studies, which limited our inferences to a daily scale. As not all individuals had sub-hourly data, we could not explore the within day patterns of how maned wolves use vegetation, only seasonal patterns were analysed.

### 6.3 Future research

Considering the aforementioned limitations, we propose that future research should focus on testing our methodologies for a wider range of contextual variables to assess their suitability and accuracy. Once these tests are performed, it might be possible to start working towards decreasing the complexity of these methods, so that they are accessible to a wider range of users.

The complexity of the methodologies can be tackled by using specialised software for pre processing remote sensing data and developing scripts and libraries to run the data fusion for
movement analysis. There is also a potential of adding the remote sensing disaggregation tools into the already existing systems. For example, movebank.org provides not only a repository for animal tracking data, but also a set of tools for analysis of these, including a tool for trajectory annotation (Env-DATA, (Dodge et al., 2008)). Our proposed multi-source tool could be linked to that. Further, to make sure that these types of new methods are relevant to the actual problems in the application community (e.g. movement ecology), this would require the integration of remote sensing researchers and animal movement ecologists. This would provide a new application area for remote sensing research and it would give access to a whole new toolbox for data processing and information extraction for movement research. It is important to support this integration by not only using remote sensing data, but also publishing in journals that provide such integration, such as the "Remote Sensing in Ecology and Conservation" and developing multidisciplinary projects.

In order to better validate methods for inferences on human and animal behaviour, we need to produce tracking data with appropriate ground-truthing observational data, so that we are able to confirm the patterns found in our analyses. In the future, small pilot studies using GPS data and observational data may be able to bridge this gap. We believe that the use of daily logs for humans and observational notes for animals will be of great help to understand how movement patterns and contextual data are shown in CAMA.

Analysing movement data alongside contextual variables is still a recent trend in movement analysis, which can bring new insights into the behaviour of different species according to variations in the surrounding environment. We believe these tools must be sought in other research areas, in which similar data complexity is recurrent. Remote sensing, in particular, has a lot to contribute in terms of enhancing and capitalising products to obtain the best possible representation of contextual variables. There are plenty of research on how to improve remote sensing data spatio-temporal resolutions for the most varied applications, which could be replicated for movement research. Yet, the complexity of such methods require a higher integration between remote sensing community and movement researchers, so that knowledge can be transferred both ways and a new fruitful partnership between those areas can flourish.

We also believe that the definition of which contextual variables will be considered in the
analysis should happen before the data collection. A lot of GPS data collection is still happening without the appropriate definition of sampling rates, or even a description of the equipment set-up, which adds more challenges on the pile of issues to be dealt with when performing CAMA. More specifically, different sampling rates and equipment set up have a direct effect on the reliability of the inferences performed on a dataset via CAMA. However, contextual variables are still not a main consideration when design GPS data collection, when they should be normal practice.

These challenges can be foreseen and reduced if the research team is multidisciplinary and takes part in all the design of data collection (Sedlmair et al., 2012). We believe that ideally, movement research studies should count with at least an ecologist/social scientist (the expert on the moving entity) and a remote sensing/ data scientist (the expert on the contextual data and movement data), so that the most can be made of the data analysis but also that it can be confirmed or tested in terms of the real movement in analysis.

## Appendix A

## Appendices list

1. Results for the remaining wolves - Chapter 4
2. Distance covered by wolf $v s$ eigenbehaviour - Chapter 4

## Appendix B

## Results for the remaining wolves Chapter 4

This appendix includes the results for the remaining wolves that were not analysed within Chapter 4.

The wolf Amadeo established his home range in the CNP border, his home range is covered mostly by heath, some pasture and few patches of young forest (Figure 4.4). This wolf was tracked between 2007 and 2008, period in which the state of MG was going trough an intense drought. His results are shown in Figure B.1. The ENDVI time series show a clearer pattern in terms of landscape transitioning from a higher NPP (NNH), to a medium NPP (NNH-P) and finally to a low NPP (LNH-P). As the landscape transitions into a state of lower NPP there is a change in the periodicity of visits to locations with higher Z-scores of ENDVI, i.e., location with more food availability. It seems to transition from daily/every other day visits, on the NNH home range, to five to seven days intervals on the LNH-P home range. However, the data quality threshold was met only in the first six months of tracking.


Figure B.1: The top panel shows the temporal series of Z-scores of the locations used by Amadeo in relation to MODIS NDVI distributions within the home range ( $\mathrm{n}=2162$ pixels of $250 \mathrm{~m})$. The colours in the Z- score plots indicate the type of NDVI or ENDVI distribution (Figure 4.9) for the day. The middle panel shows the temporal series of Z-scores of the locations used by Amadeo in relation to the ENDVI distributions within the home range ( $\mathrm{n}=506822$ pixels of 15 m ). Red dashed lines indicate the $10 \%$ confidence interval, points above that show that the animals was selecting areas amongst the $10 \%$ more vegetated in the home range. The bottom panel shows the number of fixes collected each day in the period and the average number of MODIS images available at each day within a 5 days moving window. The beige background indicates when the data quality criteria were fully met, i.e., more than 10 fixes and more than 0.2 MODIS images in a 5 days window.

The wolf Gamba established his home range in the CNP border, his home range is covered
mostly by heath and pasture, with scattered islands of shrubland and some farmland (Figure 4.4). This wolf was tracked during 2009, period in which the state of MG was going trough an intense drought. His results are shown in Figure B.2. The ENDVI time series does'nt show a clear pattern for this home range, because the tracking period is mostly during the dry season. As the landscape transitions into a state of lower NPP there seem to have a change in the periodicity of visits to locations with higher Z-scores of ENDVI, i.e., location with more food availability. It seems to transition from daily/every other day visits,ion the NNH home range, to five to seven days intervals in the LNH-P home range.


Figure B.2: The top panel shows the temporal series of Z-scores of the locations used by Gamba in relation to MODIS NDVI distributions within the home range ( $\mathrm{n}=850$ pixels of 250 m ). The colours in the Z- score plots indicate the type of NDVI or ENDVI distribution (Figure 4.9) for the day. The middle panel shows the temporal series of Z-scores of the locations used by Gamba in relation to the ENDVI distributions within the home range ( $\mathrm{n}=199398$ pixels of $15 \mathrm{~m})$. Red dashed lines indicate the $10 \%$ confidence interval, points above that show that the animals was selecting areas amongst the $10 \%$ more vegetated in the home range. The bottom panel shows the number of fixes collected each day in the period and the average number of MODIS images available at each day within a 5 days moving window. The beige background indicates when the data quality criteria were fully met, i.e., more than 10 fixes and more than 0.2 MODIS images in a 5 days window.

The wolf Henry established his home range outside the CNP, his home range is covered mostly by pasture, some coffee crops and farmland, heath and very few patches of young forest (Figure 4.4). The colours in the Z- score plots indicate the type of NDVI or ENDVI distribution (Figure 4.9) for the day. This wolf was tracked between 2010 and 2011. His results are shown in Figure B.3. The ENDVI time series show a clearer pattern in terms of landscape transitioning from a higher NPP (NNH), to a medium NPP (NNH-P) and finally to a low NPP (LNH-P). As the landscape transitions into a state of lower NPP there is a change in the periodicity of visits to locations with higher Z-scores of ENDVI, i.e., location with more food availability. It seems to transition from daily/every other day visits, on the LNH-P home range, to five to seven days intervals on the HNH home range, which is the opposite behaviour when compared to most wolves in the study.

The wolf Miro, Luna's partner, established his home range outside the CNP area, his home range is covered mostly by heath and pasture, with very few patches of shrubland and coffee crops (Figure 4.4). This wolf was tracked between 2011 and 2012. His results are shown in Figure B.4. The ENDVI time series show a clearer pattern in terms of landscape transitioning from a higher NPP (NNH), to a medium NPP (NNH-P) and finally to a low NPP (LNH-P). As the landscape transitions into a state of lower NPP there is a change in the periodicity of visits to locations with higher Z-scores of ENDVI, i.e., location with more food availability. It seems to transition from daily/every other day visits, in the NNH home range, to five to seven days intervals in the LNH-P home range. However, the data quality was very poor during the tracking period.

The wolf Nilde, established her home range outside the CNP area, her home range is covered mostly by heath and pasture, with very few patches of farmland and and shrubland (Figure 4.4). This she-wolf was tracked during 2011. Her results are shown in Figure B.5. The ENDVI time series show a very clear pattern where the wet season has a LNH-P distribution of ENDVI and the dry season has a NNH distribution, i.e., higher NPP during the dry season. As the landscape transitions into a state of lower NPP there is a change in the periodicity of visits to locations with higher Z-scores of ENDVI, i.e., location with more food availability. It seems to transition from daily/every other day visits, in the NNH home range, to five to seven


Figure B.3: The top panel shows the temporal series of Z-scores of the locations used by Henry in relation to MODIS NDVI distributions within the home range ( $\mathrm{n}=765$ pixels of 250 m ). The colours in the Z- score plots indicate the type of NDVI or ENDVI distribution (Figure 4.9) for the day. The middle panel shows the temporal series of Z-scores of the locations used by Henry in relation to the ENDVI distributions within the home range ( $\mathrm{n}=177976$ pixels of 15 m ). Red dashed lines indicate the $10 \%$ confidence interval, points above that show that the animals was selecting areas amongst the $10 \%$ more vegetated in the home range. The bottom panel shows the number of fixes collected each day in the period and the average number of MODIS images available at each day within a 5 days moving window. The beige background indicates when the data quality criteria were fully met, i.e., more than 10 fixes and more than 0.2 MODIS images in a 5 days window.


Figure B.4: The top panel shows the temporal series of Z-scores of the locations used by Miro in relation to MODIS NDVI distributions within the home range ( $\mathrm{n}=764$ pixels of 250 m ). The colours in the Z- score plots indicate the type of NDVI or ENDVI distribution (Figure 4.9) for the day. The middle panel shows the temporal series of Z-scores of the locations used by Miro in relation to the ENDVI distributions within the home range ( $\mathrm{n}=179254$ pixels of 15 m ). Red dashed lines indicate the $10 \%$ confidence interval, points above that show that the animals was selecting areas amongst the $10 \%$ more vegetated in the home range. The bottom panel shows the number of fixes collected each day in the period and the average number of MODIS images available at each day within a 5 days moving window.The beige background indicates when the data quality criteria were fully met, i.e., more than 10 fixes and more than 0.2 MODIS images in a 5 days window.
days intervals in the LNH-P home range.


Figure B.5: The top panel shows the temporal series of Z-scores of the locations used by Nilde in relation to MODIS NDVI distributions within the home range ( $\mathrm{n}=764$ pixels of 250 m ). The colours in the Z- score plots indicate the type of NDVI or ENDVI distribution (Figure 4.9) for the day. The middle panel shows the temporal series of Z-scores of the locations used by Nilde in relation to the ENDVI distributions within the home range ( $\mathrm{n}=174347$ pixels of 15 m ). Red dashed lines indicate the $10 \%$ confidence interval, points above that show that the animals was selecting areas amongst the $10 \%$ more vegetated in the home range. The bottom panel shows the number of fixes collected each day in the period and the average number of MODIS images available at each day within a 5 days moving window. The beige background indicates when the data quality criteria were fully met, i.e., more than 10 fixes and more than 0.2 MODIS images in a 5 days window.

The wolf Samurai, Jurema's partner, established his home range outside the CNP, his home range is covered mostly by pasture and heath, a big patch of farmland and very few patches of mature forest (Figure 4.4). This wolf was tracked during 2009. His results are shown in Figure B.6. The ENDVI time series does not show any clear pattern and the time series is also limited and with a varying quality.


Figure B.6: The top panel shows the temporal series of Z-scores of the locations used by Samurai in relation to MODIS NDVI distributions within the home range ( $\mathrm{n}=761$ pixels of $250 \mathrm{~m})$. The colours in the Z- score plots indicate the type of NDVI or ENDVI distribution (Figure 4.9) for the day. The middle panel shows the temporal series of Z-scores of the locations used by Samurai in relation to the ENDVI distributions within the home range ( $\mathrm{n}=177753$ pixels of 15 m ). Red dashed lines indicate the $10 \%$ confidence interval, points above that show that the animals was selecting areas amongst the $10 \%$ more vegetated in the home range. The bottom panel shows the number of fixes collected each day in the period and the average number of MODIS images available at each day within a 5 days moving window. The beige background indicates when the data quality criteria were fully met, i.e., more than 10 fixes and more than 0.2 MODIS images in a 5 days window.

## Appendix C

## Distance covered by wolf $v s$ <br> eigenbehaviour

This appendix includes the plot of the distance covered vs eigenbehaviour mentioned in the discussion of Chapter 4.

The distance covered by an individual wolf during each day is represented by a black line plot overlaying the five primary eigenbehaviours for low NDVI, i.e., choosing areas with lower ENDVI Chapter. These areas may be linked to denning and more specifically to whelping if the behaviour occurs during the dry season. During whelping and in subsequent days, the distance covered by the female is expected to drop dramatically. The areas indicated by a red circle in Figure C.1, where there are lower distances covered by a female wolf with a simultaneous high contribution of green avoidance, may indicate successful reproduction of that female. It is also interesting to note that at the same time Luna show a decrease in distance covered and simultaneously high contribution of green avoidance, her partner Miro shows an increase in the distance covered in those days. The literature states that male maned wolves are responsible for hunting and providing food for the mother and pups during whelping and in the subsequent days (de Paula et al., 2013), which is what we might be seeing in our results. However, more tracking data and detailed behavioural observations would be needed to confirm this hypothesis.


Figure C.1: Distance covered by day vs five primary eigenbehaviours for wolves where there was a match between the decrease in distance covered per day and a higher contribution of the low ENDVI state. Red circles indicate where the match happens. Green rectangle indicates the only couple for which we had simultaneous data, and in which the red circle highlights the increased activity of the male.

## Bibliography

Abbott, A. (1995), 'Sequence Analysis: New Methods for Old Ideas', Annual Review of Sociology 21(21), 93-113.

URL: http://socsci2.ucsd.edu/ aronatas/project/academic/abbott on sequence analsys.pdf http://www.jstor.org/stable/2083405

Abbott, A. and Tsay, A. (2000), 'Sequence Analysis and Optimal Matching Methods in Sociology', Sociological Methods \& Research 29(1), 3-33.
URL: http://journals.sagepub.com/doi/10.1177/0049124100029001001
Abowd, G. D., Dey, A. K., Brown, P. J., Davies, N., Smith, M. and Steggles, P. (1999), Towards a Better Understanding of Context and Context-Awareness, in 'International Symposium on Handheld and Ubiquitous Computing', Springer, Berlin, Heidelberg, Karlsruhe (Germany), pp. 304-307.

URL: http://link.springer.com/10.1007/3-540-48157-5_29
Agryzkov, T., Martí, P., Tortosa, L. and Vicent, J. F. (2017), 'Measuring urban activities using Foursquare data and network analysis: a case study of Murcia (Spain)', International Journal of Geographical Information Science 31(1), 100-121.

URL: https://www.tandfonline.com/doi/full/10.1080/13658816.2016.1188931

Ahearn, S. C., Dodge, S., Simcharoen, A., Xavier, G. and Smith, J. L. (2016), 'A contextsensitive correlated random walk: a new simulation model for movement', International Journal of Geographical Information Science 8816(September), 1-17.
URL: https://www.tandfonline.com/doi/full/10.1080/13658816.2016.1224887

Ahmed, K. I. (2012), Acoustic data optimisation for seabed mapping with visual and computational data mining, Doctorate thesis, National University of Ireland, Maynooth.

Alvares, L. O., Bogorny, V., Kuijpers, B., de Macedo, J. A. F., Moelans, B. and Vaisman, A. (2007), 'A model for enriching trajectories with semantic geographical information', Proceedings of the 15th Annual ACM International Symposium on Advances in Geographic Information Systems (GIS '07) (i), 22:1-22:8.

URL: http://portal.acm.org/citation.cfm?doid=1341012.1341041

Alves, H. M. R., Volpato, M. M. L., Vieira, T. G. C., Maciel, D. A., Gonçalves, T. G. and Dantas, M. F. (2016), 'Characterization and spectral monitoring of coffee lands in Brazil', ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLI-B8, 801-803.

URL: http://landsat.usgs.gov/landsat8.php http://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XLI-B8/801/2016/isprs-archives-XLI-B8-801-2016.pdf

Anderson, C. A. (2001), 'Heat and violence', Current Directions in Psychological Science 10(1), 33-38.

URL: http://www2.psych.ubc.ca/ schaller/308Readings/Anderson2001.pdf

Andrienko, G., Andrienko, N. and Heurich, M. (2011), 'An event-based conceptual model for context-aware movement analysis', International Journal of Geographical Information Science 25(9), 1347-1370.

Ashbrook, D. and Starner, T. (2003), 'Using GPS to learn significant locations and predict movement across multiple users', Personal and Ubiquitous Computing 7(5), 275-286.

URL: http://link.springer.com/10.1007/s007r9-003-0240-0

Ayoade, J. (1986), Introdução à climatologia para os trópicos, Difel, São Paulo.

Barchiesi, D., Preis, T., Bishop, S. and Moat, H. S. (2015), 'Modelling human mobility patterns using photographic data shared online', Royal Society Open Science 2(8).

URL: http://dx.doi.org/10.1098/rsos.150046orviahttp://rsos.royalsocietypublishing.org.

Bartels, R. H., Beatty, J. C. and Barsky, B. A. (1998), Hermite and Cubic Spline Interpolation, in M. Morgan, ed., 'An Introduction to Splines for Use in Computer Graphics and Geometric Modelling', Morgan Kaufmann, San Francisco, CA, chapter 3, pp. 9-17.

Bartlam-Brooks, H. L., Beck, P. S., Bohrer, G. and Harris, S. (2013), 'In search of greener pastures: Using satellite images to predict the effects of environmental change on zebra migration', Journal of Geophysical Research: Biogeosciences 118(4), 1427-1437.

URL: http://doi.wiley.com/10.1002/jgrg. 20096

Basille, M., Fortin, D., Dussault, C., Ouellet, J. P. and Courtois, R. (2013), 'Ecologically based definition of seasons clarifies predator-prey interactions', Ecography 36(2), 220-229.

URL: http://doi.wiley.com/10.1111/j.1600-0587.2011.07367.x

Batagelj, V. (1988), Classification and Related Methods of Data Analysis, in H. H. Bock, ed., 'n Classification and related methods of data analysis', Amsterdam, chapter Generalize, pp. 67-74.

Beecham, R., Wood, J. and Bowerman, A. (2014), 'Studying commuting behaviours using collaborative visual analytics', Computers, Environment and Urban Systems 47, 5-15.

URL: http://linkinghub.elsevier.com/retrieve/pii/S0198971513001014

Beek, E. G. (1991), Spatial interpolation of daily meteorological data Theoretical evaluation of available techniques, Technical report, DLO The Winand Staring Centre,, Wageningen (The Netherlands).

URL: http://edepot.wur.nl/360405

Bernardes, T., Alves Moreira, M., Adami, M. and Friedrich Theodor Rudorff, B. (2012),
'Monitoring biennial bearing effect on coffee yield using MODIS remote sensing imagery', International Geoscience and Remote Sensing Symposium (IGARSS) (August), 3760-3763.

Billari, F. C. (2001), 'Sequence Analysis in Demographic Research', Canadian Studies in Population 28(2), 439-458.

Blakey, E. (2010), A Model-Independent Theory of Computational Complexity: From Patience to Precision and Beyond, Doctorate thesis, Oxford University.

URL: http://www.cs.ox.ac.uk/people/bob.coecke/Blakey.pdf

Blanchard, F. N. and Finster, E. B. (1933), 'A Method of Marking Living Snakes for Future Recognition, with a Discussion of Some Problems and Results', Ecology 14(4), 334-347.

URL: http://doi.wiley.com/10.2307/1932657

Blanford, J. I., Huang, Z., Savelyev, A. and MacEachren, A. M. (2015), 'Geo-located tweets. Enhancing mobility maps and capturing cross-border movement', PLoS ONE 10(6), 1-16.

Block, B. A., Jonsen, I. D., Jorgensen, S. J., Winship, A. J., Shaffer, S. A., Bograd, S. J., Hazen, E. L., Foley, D. G., Breed, G. A., Harrison, A.-L., Ganong, J. E., Swithenbank, A., Castleton, M., Dewar, H., Mate, B. R., Shillinger, G. L., Schaefer, K. M., Benson, S. R., Weise, M. J., Henry, R. W. and Costa, D. P. (2011), 'Tracking apex marine predator movements in a dynamic ocean', Nature 475(7354), 86-90.

URL: http://www.nature.com/articles/nature10082

Bode, N. W. F., Wood, A. J. and Franks, D. W. (2015), Group movement and animal social networks, in 'Animal Social Networks', Oxford University Press, pp. 74-83.

Bogorny, V., Renso, C., de Aquino, A. R., de Lucca Siqueira, F. and Alvares, L. O. (2014), 'CONSTAnT - A Conceptual Data Model for Semantic Trajectories of Moving Objects', Transactions in GIS $\mathbf{1 8}(1), 66-88$.

URL: http://doi.wiley.com/10.1111/tgis.12011

Bohrer, G., Beck, P. S., Ngene, S. M., Skidmore, A. K. and Douglas-Hamilton, I. (2014), 'Elephant movement closely tracks precipitation-driven vegetation dynamics in a Kenyan forest-savanna landscape', Movement Ecology 2(1), 2.

URL: http://www.movementecologyjournal.com/content/2/1/2

Bolbol, A., Cheng, T., Tsapakis, I. and Haworth, J. (2012), 'Inferring hybrid transportation modes from sparse GPS data using a moving window SVM classification', Computers,

Environment and Urban Systems 36(6), 526-537.
URL: https://www.sciencedirect.com/science/article/pii/S0198971512000543

Bolchini, C., Curino, C. A., Orsi, G., Quintarelli, E., Rossato, R., Schreiber, F. A. and Tanca, L. (2009), 'And what can context do for data?', Communications of the ACM 52(11), 136. URL: http://portal.acm.org/citation.cfm?doid=1592761.1592793

Bowen, R. S., Turner, M. J. and Lightfoot, J. T. (2011), 'Sex Hormone Effects on Physical Activity Levels', Sports Medicine 41(1), 73-86.

Brillinger, D. R., Preisler, H. K., Ager, A. a. and Kie, J. G. (2004), 'An exploratory data analysis (EDA) of the paths of moving animals', Journal of Statistical Planning and Inference 122(12), 43-63.

URL: http://linkinghub.elsevier.com/retrieve/pii/S0378375803002404

Brum-Bastos, V. d. S., Long, J. A. and Demšar, U. (2015), New methodological approaches for cross-scale integration of environmental remotely sensed data with spatio-temporal movement data, in 'Evolving GIScience: A celebration of the life and work of Peter Fisher', Leicester.

Brum-Bastos, V., Long, J. and Demšar, U. (2016), Dynamic trajectory annotation for integrating environmental and movement data, in 'Visually-supported Computational Movement Analysis Workshop - AGILE 2016', Helsinki.

URL: http://viz.icaci.org/vcma2016/wp-content/uploads/2016/07/vcma_brumbastospaper.pdf

Brum-Bastos, V. S., Long, J. A. and Demšar, U. (2018), 'Weather effects on human mobility: A study using multi-channel sequence analysis', Computers, Environment and Urban Systems 70, 1-17.

URL: https://www.sciencedirect.com/science/article/pii/S0198971518300425?via\%3Dihub

Buchin, K., Sijben, S., van Loon, E. E., Sapir, N., Mercier, S., Arseneau, T. J. M. and Willems, E. P. (2015), 'Deriving movement properties and the effect of the environment from the

Brownian bridge movement model in monkeys and birds', Movement Ecology 3(1), 18.
URL: http://www.movementecologyjournal.com/content/3/1/18

Buchin, M., Dodge, S. and Speckmann, B. (2012), Context-Aware Similarity of Trajectories, in N. Xiao, ed., 'Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)', Vol. LNCS7478, pp. 43-56.

URL: https://link.springer.com/content/pdf/10.1007\%2F978-3-642-33024-7_4.pdf http://link.springer.com/10.1007/978-3-642-33024-7-4

Buchin, M., Dodge, S. and Speckmann, B. (2014), 'Similarity of trajectories taking into account geographic context', Journal of Spatial Information Science $9(9), 101-124$.

URL: http://josis.org/index.php/josis/article/view/179

Bueno, A. d. A. and Motta-Junior, J. C. (2009), 'Feeding habits of the maned wolf, Chrysocyon brachyurus (Carnivora: Canidae), in southeast Brazil', Studies on Neotropical Fauna and Environment $\mathbf{4 4}(2), 67-75$.

URL: http://www.tandfonline.com/doi/abs/10.1080/01650520902891413

Bühne, H. and Pettorelli, N. (2017), 'Better together: Integrating and fusing multispectral and radar satellite imagery to inform biodiversity monitoring, ecological research and conservation science', Methods in Ecology and Evolution 38(1), 42-49.

URL: http://doi.wiley.com/10.1111/ijlh.12426\
http://doi.wiley.com/10.1111/2041210X.12942

Burt, W. H. (1943), 'Territoriality and Home Range Concepts as Applied to Mammals', Journal of Mammalogy $\mathbf{2 4}(3), 346$.

URL: https://academic.oup.com/jmammal/article-lookup/doi/10.2307/1374834

Cabanac, M. (1971), 'Physiological role of pleasure', Science $\mathbf{1 7 3}(4002)$, 1103-1107.
URL: http://www.ncbi.nlm.nih.gov/pubmed/5098954

Cagnacci, F., Boitani, L., Powell, R. a. and Boyce, M. S. (2010), 'Animal ecology meets

GPS-based radiotelemetry: a perfect storm of opportunities and challenges.', Philosophical transactions of the Royal Society of London. Series B, Biological sciences 365(1550), 2157-62.

Cagnacci, F., Focardi, S., Heurich, M., Stache, A., Hewison, a. J. M., Morellet, N., Kjellander, P., Linnell, J. D. C., Mysterud, A., Neteler, M., Delucchi, L., Ossi, F. and Urbano, F. (2011), 'Partial migration in roe deer: migratory and resident tactics are end points of a behavioural gradient determined by ecological factors', Oikos 120(12), 1790-1802.

URL: http://doi.wiley.com/10.1111/j.1600-0706.2011.19441.x

Calenge, C., Dray, S. and Royer-Carenzi, M. (2009), 'The concept of animals' trajectories from a data analysis perspective', Ecological Informatics 4(1), 34-41.

URL: https://www.sciencedirect.com/science/article/pii/S1574954108000654

Calinski, T. and Harabasz, J. (1974), 'A dendrite method for cluster analysis', Communications in Statistics - Theory and Methods 3(1), 1-27.

URL: http://www.tandfonline.com/doi/abs/10.1080/03610927408827101

Carlsmith, J. M. and Anderson, C. A. (1979), 'Ambient Temperature and the Occurrence of Collective Violence: ANew Analysis', Journal of Personality and Social Psychology 37(3), 337-344.

URL: https://public.psych.iastate.edu/caa/abstracts/1979-1984/79CA.PDF

Changnon, S. A. (1996), 'Effects of summer precipitation on urban transportation', Climatic Change 32(4), 481-494.

URL: http://link.springer.com/10.1007/BF00140357

Chapin, F. S., Matson, P. A. and Vitousek, P. M. (2011), Trophic Dynamics, in 'Principles of Terrestrial Ecosystem Ecology', Springer New York, New York, NY, chapter 10, pp. 297-320.

URL: http://link.springer.com/10.1007/978-1-4419-9504-9

Chatterjee, T. and Chowdhury, R. (2017), Improved Sparse Approximation Models for Stochastic Computations, in 'Handbook of Neural Computation', Elsevier, pp. 201-223.

Chen, D., Zhang, Y., Gao, L., Geng, N. and Li, X. (2017), 'The impact of rainfall on the temporal and spatial distribution of taxi passengers', PLOS ONE 12(9).

URL: $\quad$ https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5584943/pdf/pone.0183574.pdf http://dx.plos.org/10.1371/journal.pone. 0183574

Choodarathnakara, A. L., Kumar, D. A., Koliwad, D. and Patil, D. (2012), 'Mixed Pixels : A Challenge in Remote Sensing Data Classification for Improving Performance', International Journal of Advanced Research in Computer Engineering \& Technology 1(9), 261-271.

URL: https://pdfs.semanticscholar.org/63b9/7ce83281e012da96d7be39d84a1739fae67c.pdf
Cole, E. K., Pope, M. D. and Anthony, R. G. (1997), 'Effects of road management on movement and survival of Roosevelt elk', The Journal of Wildlife Management 61(4), 1115-1126.

URL: https://www.jstor.org/stable/3802109?origin=crossref
Connolly, M. (2008), 'Here Comes the Rain Again: Weather and the Intertemporal Substitution of Leisure', Source Journal of Labor Economics Journal of Labor Economics 26(1), 73-100. URL: http://www.jstor.org/stable/10.1086/522067 http://about.jstor.org/terms

Consorte-McCrea, A. G. and Santos, E. F. (2013), The relationships between the maned wolf and people, in E. F. Consorte-McCrea, A. Santos, ed., 'Ecology and conservation of the maned wolf : multidisciplinary perspectives', CRC Press, Boca Raton, chapter 3, pp. 35-52. URL: http://create.canterbury.ac.uk/13738/

Cools, M., Moons, E., Creemers, L. and Wets, G. (2010), 'Changes in Travel Behavior in Response to Weather Conditions: Whether Type of Weather and Trip Purpose Matter?', Journal of the Transportation Research Board 2157, 22-28.

Coyne, M. S. and Godley, B. J. (2005), 'Satellite Tracking and Analysis Tool (STAT): an integrated system for archiving, analyzing and mapping animal tracking data', Marine ecology progress series 301, 1-7.

Dabiri, S. and Heaslip, K. (2018), 'Inferring transportation modes from GPS trajectories using a convolutional neural network', Transportation Research Part C: Emerging Technologies

86, 360-371.
URL: https://doi.org/10.1016/j.trc.2017.11.021

Daly, C., Taylor, G. and Gibson, W. (1997), 'The Prism approach to mapping precipitation and emperature', 10th AMS Conference on Applied Climatology (1), 1-4.

Das, R. D. and Winter, S. (2016), 'A context-sensitive conceptual framework for activity modeling', Journal of Spatial Information Science 12(12), 45-85.

URL: https://minerva-access.unimelb.edu.au/bitstream/handle/11343/122943/Das_Contextsensitive.pdf?sequence $=1$

Daubechies, I. (1990), 'The wavelet transform, time-frequency localization and signal analysis', IEEE Transactions on Information Theory 36(5), 961-1005.

URL: http://ieeexplore.ieee.org/document/57199/

David, A., Latham, M. C., Latham, M. C., Anderson, D. P., Cruz, J., Herries, D., Hebblewhite, M., Latham, A. M., Latham, M. C., Anderson, D. P., Cruz, J., Herries, D. and Hebblewhite, M. (2014), 'The GPS craze: six questions to address before deciding to deploy GPS technology on wildlife', New Zealand Journal of Ecology 39(1), 142-152.
URL: http://www.telemetrysolutions.com/http://newzealandecology.org/nzje/3204.pdf

Davies, N., Gramotnev, G., Seabrook, L., Bradley, A., Baxter, G., Rhodes, J., Lunney, D. and McAlpine, C. (2013), 'Movement patterns of an arboreal marsupial at the edge of its range: a case study of the koala', Movement Ecology $\mathbf{1}(1), 8$.

URL: http://movementecologyjournal.biomedcentral.com/articles/10.1186/2051-3933-1-8

Day, J. A. (1966), Weather variables and their measurament, in 'The science of weather', Addison Wesley, Oregon, chapter 2, pp. 33-50.
de Freitas, C. R. (1990), 'Recreation climate assessment', International Journal of Climatology $\mathbf{1 0}(1), 89-103$.
de Freitas, C. R. (2003), 'Tourism climatology: evaluating environmental information for decision making and business planning in the recreation and tourism sector', International

Journal of Biometeorology 48(1), 45-54.
URL: https://link.springer.com/content/pdf/10.1007\%2Fs00484-003-0177-z.pdf
de Freitas, C. R. (2015), 'Weather and place-based human behavior: recreational preferences and sensitivity', International Journal of Biometeorology 59(1), 55-63.

URL: https://link.springer.com/content/pdf/10.1007\%2Fs00484-014-0824-6.pdf
http://link.springer.com/10.1007/s00484-014-0824-6

De Groeve, J., Van de Weghe, N., Ranc, N., Neutens, T., Ometto, L., Rota-Stabelli, O. and Cagnacci, F. (2016), 'Extracting spatio-temporal patterns in animal trajectories: An ecological application of sequence analysis methods', Methods in Ecology and Evolution 7(3), 369-379.
de Montigny, L., Ling, R. and Zacharias, J. (2012), 'The Effects of Weather on Walking Rates in Nine Cities', Environment and Behavior 44(6), 821-840.

URL: http://eab.sagepub.com http://journals.sagepub.com/doi/10.1177/0013916511409033
de Paula, R. (2016), Adequabilidade ambiental dos biomas brasileiros à ocorrência do loboguará (Chrysocyon brachyurus) e efeitos da composição da paisagem em sua ecologia espacial, atividade e movimentação., PhD thesis, University of São Paulo.

URL: http://www.teses.usp.br/teses/disponiveis/11/11150/tde-05072016-114911/pt-br.php
de Paula, R. C. and DeMatteo, K. (2015), Chrysocyon brachyurus, Maned Wolf, in 'The IUCN Red List of Threathened Species', Vol. 8235.
de Paula, R. C., Medici, P. and Gonzalves Morato, R., eds (2008), Plano de Ação para a Conservação do Lobo-Guará, Edições IBAMA, Brasília.

URL: http://www.procarnivoros.org.br/pdfs/plano_de_acao_lobo_guara_2008.pdf
de Paula, R. and Desbiez, A. (2014), Maned wolf population viability, in A. Consorte-McCrea and E. Santos, eds, 'Ecology and conservation of Manned Wolf : Multidisciplinary perspectives', 1 edn, Boca Raton, FL, chapter 2, p. 322.
de Paula, R., Rodrigues, F., Bizerril, M., Songsasen, N., Eizirik, E., Morato, R., Santos, J., Azevedo, F., May, J., Soares, C., Rocha, F. and Arrais, R. (2013), The Maned Wolf Conservation Project, in 'Ecology and Conservation of the Maned Wolf', CRC Press, pp. 177192.

URL: http://www.crcnetbase.com/doi/abs/10.1201/b15607-17

Deem, S. L. and Emmons, L. H. (2005), 'Exposure of free-ranging maned wolves (Chrysocyon brachyurus) to infectious and parasitic disease agents in the Noël Kempff Mercado national park, Bolivia', Journal of Zoo and Wildlife Medicine 36(2), 192-197.

URL: http://www.bioone.org/doi/abs/10.1638/04-076.1

Demšar, U., Buchin, K., Cagnacci, F., Safi, K., Speckmann, B., Van de Weghe, N., Weiskopf, D. and Weibel, R. (2015), 'Analysis and visualisation of movement: an interdisciplinary review', Movement Ecology 3(1), 5.

URL: http://www.movementecologyjournal.com/content/3/1/5
Demšar, U., Harris, P., Brunsdon, C., Fotheringham, A. S. and McLoone, S. (2013), 'Principal Component Analysis on Spatial Data: An Overview', Annals of the Association of American Geographers 103(1), 106-128.

URL: http://www.tandfonline.com/doi/abs/10.1080/00045608.2012.689236

Derrick Sewell, W., Kates, R. W. and Philips, L. (1968), 'Human responde to weather and climate : geographical contributions', The geoographical review LVIII(2), 262-280.

Dewulf, B., Neutens, T., Lefebvre, W., Seynaeve, G., Vanpoucke, C., Beckx, C. and Van de Weghe, N. (2016), 'Dynamic assessment of exposure to air pollution using mobile phone data', International Journal of Health Geographics 15(1), 14.

URL: http://www.ncbi.nlm.nih.gov/pubmed/27097526

DigitalGlobe (2016), WorldView-1 Relative Radiometric Response Curves, Technical report, DigitalGlobe.

URL: https://www.digitalglobe.com/resources/technical-information

Dirk, R. (2004), College on Soil Physics -30th Anniversary (1983-2013) Frequency analysis of rainfall data, Technical report, K.U.Leuven, Leuven, Belgium.

URL: http://indico.ictp.it/event/a12165/session/21/contribution/16/material/0/0.pdf

Dodge, S., Bohrer, G., Bildstein, K., Davidson, S. C., Weinzierl, R., Bechard, M. J., Barber, D., Kays, R., Brandes, D., Han, J. and Wikelski, M. (2014), 'Environmental drivers of variability in the movement ecology of turkey vultures (Cathartes aura) in North and South America.', Philosophical transactions of the Royal Society of London. Series B, Biological sciences $\mathbf{3 6 9}$ (1643), 20130195.

URL: http://www.ncbi.nlm.nih.gov/pubmed/24733950

Dodge, S., Bohrer, G., Weinzierl, R., Davidson, S. C., Kays, R., Douglas, D., Cruz, S., Han, J., Brandes, D. and Wikelski, M. (2013), 'The environmental-data automated track annotation (Env-DATA) system: linking animal tracks with environmental data', Movement Ecology $\mathbf{1}(1), 3$.

URL: http://www.movementecologyjournal.com/content/1/1/3

Dodge, S., Weibel, R. and Laube, P. (2009), 'PhD Showcase : Exploring Movement - Similarity Analysis of Moving Objects', SIGSPATIAL Special 1(3), 11-16.

URL: https://dl.acm.org/citation.cfm?id=1645427

Dodge, S., Weibel, R. and Lautenschütz, A.-K. (2008), 'Towards a taxonomy of movement patterns', Information Visualization $7(3-4), 240-252$.

Dunne, S. and Ghosh, B. (2013), 'Weather Adaptive Traffic Prediction Using Neurowavelet Models', IEEE Transactions on Intelligent Transportation Systems 14(1), 370-379.

URL: http://ieeexplore.ieee.org/document/6414631/

Dutta, R. (2012), Wavelet based pattern analysis to detect changes in vegetation growth, in 'The 33rd Asian Conference on Remote Sensing'.

Dwyer, R. G., Brooking, C., Brimblecombe, W., Campbell, H. a., Hunter, J., Watts, M. and Franklin, C. E. (2015), 'An open Web-based system for the analysis and sharing of animal
tracking data', Animal Biotelemetry 3(1), 1-11.
URL: http://www.animalbiotelemetry.com/content/3/1/1

Eagle, N. and Pentland, A. S. (2009), 'Eigenbehaviors: identifying structure in routine', Behavioral Ecology and Sociobiology 63(7), 1057-1066.

URL: http://realitycommons.media.mit.edu/pdfs/eigenbehaviors.pdf

Eckert, S., Hüsler, F., Liniger, H. and Hodel, E. (2015), 'Trend analysis of MODIS NDVI time series for detecting land degradation and regeneration in Mongolia', Journal of Arid Environments 113, 16-28.

URL: http://www.eic.mn/DLDbase/upload/2014/tadesertbook/3/20141208_2263_3.pdf http://linkinghub.elsevier.com/retrieve/pii/S0140196314001797

Edwards, N. M., Myer, G. D., Kalkwarf, H. J., Woo, J. G., Khoury, P. R., Hewett, T. E. and Daniels, S. R. (2015), 'Outdoor Temperature, Precipitation, and Wind Speed Affect Physical Activity Levels in Children: A Longitudinal Cohort Study', Journal of Physical Activity and Health 12(8), 1074-1081.

URL: http://www.ncbi.nlm.nih.gov/pubmed/25423667

Elliott, K. H., Chivers, L. S., Bessey, L., Gaston, A. J., Hatch, S. A., Kato, A., Osborne, O., Ropert-Coudert, Y., Speakman, J. R. and Hare, J. F. (2014), 'Windscapes shape seabird instantaneous energy costs but adult behavior buffers impact on offspring', Movement Ecology 2(1), 17.

URL: http://www.movementecologyjournal.com/content/2/1/17

Epinat, V., Stein, A., de Jong, S. M. and Bouma, J. (2001), 'A wavelet characterization of high-resolution NDVI patterns for precision agriculture', International Journal of Applied Earth Observation and Geoinformation 3(2), 121-132.

URL: https://www.sciencedirect.com/science/article/pii/S0303243401850030 http://linkinghub.elsevier.com/retrieve/pii/S0303243401850030

Fonseca, G., Rylands, A., Costa, C., Machado, R. and Leite, Y., eds (1994),

Livro vermelho dos mamíferos ameaçadso de extinção, 1 edn, Fundação Biodiversitas, Belo Horizonte, Brazil.

Frair, J. L., Fieberg, J., Hebblewhite, M., Cagnacci, F., DeCesare, N. J. and Pedrotti, L. (2010), 'Resolving issues of imprecise and habitat-biased locations in ecological analyses using GPS telemetry data', Philosophical Transactions of the Royal Society B: Biological Sciences 365(1550), 2187-2200.

Fryxell, J. M., Hazell, M., Borger, L., Dalziel, B. D., Haydon, D. T., Morales, J. M., McIntosh, T. and Rosatte, R. C. (2008), 'Multiple movement modes by large herbivores at multiple spatiotemporal scales', Proceedings of the National Academy of Sciences 105(49), 1911419119.

URL: http://www.ncbi.nlm.nih.gov/pubmed/19060190

Fullman, T. J., Joly, K. and Ackerman, A. (2017), 'Effects of environmental features and sport hunting on caribou migration in northwestern Alaska', Movement Ecology 5(1), 4.

URL: http://movementecologyjournal.biomedcentral.com/articles/10.1186/s40462-017-0095-z

Fussell, J. and Rundquist, D. (1986), 'On defining remote sensing', Photogrammetric engineering and remote sensing $\mathbf{5 2}(9), 1507-1511$.

URL: https://www.asprs.org/wp-content/uploads/pers/1986journal/sep/1986_sep_15071511.pdf

Gabadinho, A., Ritschard, G., Studer, M. and Müller, N. S. (2009), 'Mining Sequence Data in $R$ with the TraMineR package: A user's guide'.

URL: https://cran.r-project.org/web/packages/TraMineR/TraMineR.pdf

Gao, B.-c. (1996), 'NDWI-A normalized difference water index for remote sensing of vegetation liquid water from space', Remote Sensing of Environment 58(3), 257-266.

URL: http://linkinghub.elsevier.com/retrieve/pii/S0034425796000673

Gao, S. (2014), ‘Spatio-Temporal Analytics for Exploring Human Mobility Patterns and Urban

Dynamics in the Mobile Age', Spatial Cognition \& Computation pp. 1-41.
URL: http://www.tandfonline.com/doi/abs/10.1080/13875868.2014.984300

Gao, S., Yang, J.-a., Janowicz, K., Hu, Y. and Yan, B. (2013), TrajAnalyst : Matching Data to Trajectory Analysis Modules via a Conceptual Framework Triple TrajAnalyst Bivariate Single, in 'Eighth International Conference on Geographic Information Science (GIScience)', number L, Vienna, Austria, pp. 2-5.

Gaston, A., Bllazquez-Cabrera, S., Garrote, G., Mateo-Sanchez, M. C., Beier, P., Simon, M. A. and Saura, S. (2016), 'Response to agriculture by a woodland species depends on cover type and behavioural state: insights??from resident and dispersing Iberian lynx', Journal of Applied Ecology 53(3), 814-824.

Gauthier, J.-A., Widmer, E. D., Bucher, P. and Notredame, C. (2010), '1. Multichannel Sequence Analysis Applied to Social Science Data', Sociological Methodology 40(1), 1-38.

URL: http://journals.sagepub.com/doi/10.1111/j.1467-9531.2010.01227.x

Gauthreaux, S. A. J. and Belser, C. G. (2005), Radar ornithology and the conservation of migratory birds, Technical report.

URL: http://gis.fs.fed.us/psw/publications/documents/psw_gtr191/Asilomar/pdfs/871875.pdf

George, G. and Mallery, P. (2010), SPSS for Windows Step by Step, 10th edn.
URL: https://wps.ablongman.com/wps/media/objects/385/394732/george4answers.pdf

Gething, P. W. and Tatem, A. J. (2011), 'Can mobile phone data improve emergency response to natural disasters?', PLoS Medicine 8(8), e1001085.

URL: http://www.ncbi.nlm.nih.gov/pubmed/21918644

Getz, W. M. and Saltz, D. (2008), 'A framework for generating and analyzing movement paths on ecological landscapes', Proceedings of the National Academy of Sciences 105(49), 1906619071.

URL:
http://doi.wiley.com/10.1111/j.1365-2427.2006.01554.x http://www.pnas.org/cgi/doi/10.1073/pnas. 0801732105

Givoni, B. (1974), 'Architectural and urban planning in relation to weather and climate.', $\underline{\text { Progress in biometeorology. Division A: Progress in human biometeorology 1(1A), 183-93, }}$ 632.

URL: http://www.ncbi.nlm.nih.gov/pubmed/4438650

Gole, F. W. (1970), Part I - Introduction to the atmosphere, in 'Introduction to meteorology', John Wiley \& Sons, Inc., New York, chapter 1-5, pp. 3-70.

Gong, H., Chen, C., Bialostozky, E. and Lawson, C. T. (2012), 'A GPS/GIS method for travel mode detection in New York City', Computers, Environment and Urban Systems 36(2), 131139.

URL: http://www.sciencedirect.com/science/article/pii/S0198971511000536

Goymann, W., Lupi, S., Kaiya, H., Cardinale, M. and Fusani, L. (2017), 'Ghrelin affects stopover decisions and food intake in a long-distance migrant', Proceedings of the National Academy of Sciences 114(8), 1946-1951.

URL: http://www.ncbi.nlm.nih.gov/pubmed/28167792

Grossmann, A. and Morlet, J. (1984), 'Decomposition of Hardy Functions into Square Integrable Wavelets of Constant Shape', SIAM Journal on Mathematical Analysis 15(4), 723736.

URL: http://epubs.siam.org/doi/10.1137/0515056

Gudmundsson, J. and Wolle, T. (2014), 'Football analysis using spatio-temporal tools', Computers, Environment and Urban Systems 47, 16-27.

URL: http://linkinghub.elsevier.com/retrieve/pii/S0198971513000847

Guo, Z., Wilson, N. and Rahbee, A. (2007), 'Impact of Weather on Transit Ridership in Chicago, Illinois', Transportation Research Record: Journal of the Transportation Research

Board 2034, 3-10.
URL: http://trrjournalonline.trb.org/doi/10.3141/2034-01

Gutierrez Illan, J., Wang, G., Cunningham, F. L. and King, D. T. (2017), 'Seasonal effects of wind conditions on migration patterns of soaring American white pelican', PLOS ONE $12(10)$, e0186948.

URL: http://dx.plos.org/10.1371/journal.pone. 0186948

Hägerstrand, T. (1970), 'What about people in Regional Science?', Papers of the Regional Science Association .

Halpin, P. N., Read, A. J., Best, B. D., Hyrenbach, K. D., Fujioka, E., Coyne, M. S., Crowder, L. B., Freeman, S. A. and Spoerri, C. (2006), 'OBIS-SEAMAP: Developing a biogeographic research data commons for the ecological studies of marine mammals, seabirds, and sea turtles', Marine Ecology Progress Series 316, 239-246.

Halpin, P., Read, A., Fujioka, E., Best, B., Donnelly, B., Hazen, L., Kot, C., Urian, K., LaBrecque, E., Dimatteo, A., Cleary, J., Good, C., Crowder, L. and Hyrenbach, K. D. (2009), 'OBIS-SEAMAP: The World Data Center for Marine Mammal, Sea Bird, and Sea Turtle Distributions', Oceanography 22(2), 104-115.

URL: http://dx.doi.org/10.5670/oceanog.2011.80. https://tos.org/oceanography/article/obis-seamap-the-world-data-center-for-marine-mammal-sea-bird-and-sea-turtle

Harris, B. (2001), 'Accessibility: Concepts and Applications', Journal of Transportation and Statistics and Statistics $4(1)$.

Hartkamp, A. D., De Beurs, K., Stein, A. and White, J. W. (1999), 'Interpolation Techniques for Climate Variables Interpolation', Soil Science p. 26.

URL: www.cimmyt.cgiar.org

Hayne, D. W. (1949), 'No Title', 30(1).
URL: http://www.jstor.org/stable/1375189 https://academic.oup.com/jmammal/articlelookup/doi/10.2307/1375189

Henry, D. A., Ament, J. M. and Cumming, G. S. (2015), ‘Exploring the environmental drivers of waterfowl movement in arid landscapes using first-passage time analysis', Movement Ecology 4(1), 8.

URL: http://www.movementecologyjournal.com/content/4/1/8

Holden, C. (2006), 'Inching Toward Movement Ecology', Science 313(August), 779 -782.

Horanont, T., Phithakkitnukoon, S., Leong, T. W., Sekimoto, Y., Shibasaki, R. and Lambiotte, R. (2013), 'Weather effects on the patterns of people's everyday activities: A study using GPS traces of mobile phone users', PLoS ONE 8(12), 1-14.

URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3867318/pdf/pone.0081153.pdf
Hornsby, K. and Egenhofer, M. J. (2002), 'Modeling moving objects over multiple granularities', Annals of Mathematics and Artificial Intelligence 36, 177-194.

URL: http://link.springer.com/article/10.1023/A:1015812206586

Horowitz, S. (2002), 'Weather-Health Links: Epidemiologic Research, Clinical Observations, and Patient Perceptions', Alternative \& complementary therapies pp. 34-40.

URL: http://www.sld.cu/galerias/pdf/p34_s.pdf

Hough, M. N. and Jones, R. J. A. (1997), 'The United Kingdom Meteorological Office rainfall and evaporation calculation system: MORECS version 2.0-an overview', Hydrology and Earth System Sciences 1(2), 227-239.

URL: http://www.hydrol-earth-syst-sci.net/1/227/1997/

Howarth, E. and Hoffman, M. S. (1984), 'A multidimensional approach to the relationship between mood and weather', $\underline{\text { British Journal of Psychology 75(1), 15-23. }}$

Howey, L. A., Wetherbee, B. M., Tolentino, E. R. and Shivji, M. S. (2017), 'Biogeophysical and physiological processes drive movement patterns in a marine predator', Movement Ecology 5(1), 16.

URL: http://movementecologyjournal.biomedcentral.com/articles/10.1186/s40462-017-0107-z

Hu, Y., Janowicz, K., Carral, D., Scheider, S., Kuhn, W., Berg-Cross, G., Hitzler, P., Dean, M. and Kolas, D. (2013), A geo-ontology design pattern for semantic trajectories, in T. Tenbrink, J. Stell, A. Galton and Z. Wood, eds, 'Spatial Information Theory : Lecture Notes in Computer Science', Vol. 8116 LNCS, Springer, Scarborough, pp. 438-456.

Hurley, M. A., Hebblewhite, M., Gaillard, J. M., Dray, S., Taylor, K. A., Smith, W. K., Zager, P. and Bonenfant, C. (2014), 'Functional analysis of normalized difference vegetation index curves reveals overwinter mule deer survival is driven by both spring and autumn phenology', Philosophical Transactions of the Royal Society B: Biological Sciences 369(1643), 2013019620130196.

URL: http://www.ncbi.nlm.nih.gov/pubmed/24733951
Hutchinson, M. F. (1995), 'Interpolating mean rainfall using thin plate smoothing splines', International journal of geographical information systems 9(4), 385-403.

URL: http://www.tandfonline.com/doi/abs/10.1080/02693799508902045

ICMBio (2016), Executive Summary Brazil Red Book of Threatened Species of Fauna.
URL: www.icmbio.gov.br

Idury, R. M. and Waterman, M. S. (1995), 'A New Algorithm for DNA Sequence Assembly', Journal of Computational Biology 2(2), 291-306.
URL: http://www.ncbi.nlm.nih.gov/pubmed/7497130

Isaacman, S., Becker, R., Cáceres, R., Kobourov, S., Martonosi, M., Rowland, J. and Varshavsky, A. (2011), Identifying Important Places in People's Lives from Cellular Network Data, in 'Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)', Vol. 6696 LNCS, pp. 133-151.

URL: http://link.springer.com/10.1007/978-3-642-21726-5_9

Jain, A. K., Murty, M. N., Flynn, P. J., Rosenfeld, A., Bowyer, K., Ahuja, N. and Jain, A. (1999), 'Data Clustering: A Review', ACM Computing Surveys 31(3).

Javid, R. J. and Javid, R. J. (2018), 'A framework for travel time variability analysis using
urban traffic incident data', IATSS Research 42(1), 30-38.
URL: https://www.sciencedirect.com/science/article/pii/S0386111216300395 https://linkinghub.elsevier.com/retrieve/pii/S0386111216300395

Jeltsch, F., Bonte, D., Pe'er, G., Reineking, B., Leimgruber, P., Balkenhol, N., Schröder, B., Buchmann, C. M., Mueller, T., Blaum, N., Zurell, D., Böhning-Gaese, K., Wiegand, T., Eccard, J. A., Hofer, H., Reeg, J., Eggers, U. and Bauer, S. (2013), 'Integrating movement ecology with biodiversity research - exploring new avenues to address spatiotemporal biodiversity dynamics', Movement Ecology 1(1), 6.

URL: http://movementecologyjournal.biomedcentral.com/articles/10.1186/2051-3933-1-6

Jenkins, G., Perry, M. and Prior, J. (2009), The climate of the United Kingdom and recent trends, Technical report, UKCP, Met Office Hadley Centre, Exeter, UK.

URL: http://www.ukcip.org.uk/wp-content/PDFs/UKCP09_Trends.pdf

Jensen, J. R. (2006), Active and Passive Microwave Remote Sensing, in K. C. Clarck, ed., 'Remote Sensing of the Environment: An Earth Resource Perspective', $2^{\text {a }}$ edn, Prentice Hall, Englewood Cliffs, p. 592.

Jeung, H., Yiu, M. L. and Jensen, C. S. (2011), Trajectory Pattern Mining, in ‘Computing with Spatial Trajectories', Springer New York, New York, NY, pp. 143-177.

Jiang, S., Alves, A., Rodrigues, F., Ferreira, J. and Pereira, F. C. (2015), 'Mining point-of-interest data from social networks for urban land use classification and disaggregation', Computers, Environment and Urban Systems 53, 36-46.

URL: https://www.sciencedirect.com/science/article/pii/S0198971514001355

Johnson, D. D. and Ganskopp, D. C. (2008), 'GPS collar sampling frequency: Effects on measures of resource use', Rangeland Ecology and Management 61(2), 226-231.

URL: https://www.sciencedirect.com/science/article/pii/S1550742408500261

Jolliffe, I. (2002), Principal Component Analysis, Springer Series in Statistics, 2nd edn,

Springer-Verlag, New York.
URL: http://link.springer.com/10.1007/0-387-22440-8_1

Jurdak, R., Zhao, K., Liu, J., AbouJaoude, M., Cameron, M. and Newth, D. (2015), 'Understanding human mobility from Twitter', PLoS ONE 10(7), e0131469.

Kaddar, B., Fizazi, H., Kaddar, B. and Fizazi, H. (2017), 'Spatiotemporal Analysis for NDVI Time Series Using Local Binary Pattern and Daubechies Wavelet Transform', International Review of Aerospace Engineering (IREASE) $\mathbf{1 0}(2), 96$.

Kappes, M. A., Shaffer, S. A., Tremblay, Y., Foley, D. G., Palacios, D. M., Bograd, S. J. and Costa, D. P. (2015), 'Reproductive constraints influence habitat accessibility, segregation, and preference of sympatric albatross species', Movement Ecology 3(1), 34.

URL: http://www.movementecologyjournal.com/content/3/1/34

Katzner, T. E., Brandes, D., Miller, T., Lanzone, M., Maisonneuve, C., Tremblay, J. A., Mulvihill, R. and Merovich, G. T. (2012), 'Topography drives migratory flight altitude of golden eagles: Implications for on-shore wind energy development', Journal of Applied Ecology 49(5), 1178-1186.

URL: http://doi.wiley.com/10.1111/j.1365-2664.2012.02185.x

Kays, R., Crofoot, M. C., Jetz, W. and Wikelski, M. (2015), 'Terrestrial animal tracking as an eye on life and planet', Science $\mathbf{3 4 8 ( 6 2 4 0 )}$, aaa2478-aaa2478.

URL: http://www.sciencemag.org/cgi/doi/10.1126/science.aaa2478

Kerr, J. T. and Ostrovsky, M. (2003), 'From space to species: ecological applications for remote sensing', Trends in Ecology \& Evolution 18(6), 299-305.

URL: https://www.sciencedirect.com/science/article/pii/S0169534703000715 http://linkinghub.elsevier.com/retrieve/pii/S0169534703000715

Kimijima, S. and Nagai, M. (2017), 'Human Mobility Analysis for Extracting Local Interactions under Rapid Socio-Economic Transformation in Dawei, Myanmar', Sustainability $9(9), 1598$.

URL: http://www.mdpi.com/2071-1050/9/9/1598

Kittle, A. M., Bukombe, J. K., Sinclair, A. R., Mduma, S. A. and Fryxell, J. M. (2015), 'Landscape-level movement patterns by lions in western Serengeti: Comparing the influence of inter-specific competitors, habitat attributes and prey availability', Movement Ecology 4(1), 17.

URL: http://movementecologyjournal.biomedcentral.com/articles/10.1186/s40462-016-0082-9

Kleyheeg, E., van Dijk, J. G., Tsopoglou-Gkina, D., Woud, T. Y., Boonstra, D. K., Nolet, B. A. and Soons, M. B. (2017), 'Movement patterns of a keystone waterbird species are highly predictable from landscape configuration', Movement Ecology 5(1), 2.

URL: http://movementecologyjournal.biomedcentral.com/articles/10.1186/s40462-016-0092-7

Klink, C. a., Klink, C. a., Machado, R. B. and Machado, R. B. (2005), 'A conservação do Cerrado brasileiro', Megadiversidade 1(1), 147-155.

Knox, A. (1983), 'The location of the ringing registers of the Aberdeen University BirdMigration Inquiry', 4(3), 148-148.

URL: http://www.tandfonline.com/action/journalInformation?journalCode=tram20

Kondo, K. and Kitamura, R. (1987), 'Time-space constraints and the formation of trip chains', Regional Science and Urban Economics 17(1), 49-65.

URL: https://www.sciencedirect.com/science/article/pii/0166046287900688

Kranstauber, B., Cameron, A., Weinzerl, R., Fountain, T., Tilak, S., Wikelski, M. and Kays, R. (2011), 'The Movebank data model for animal tracking', Environmental Modelling and Software 26(6), 834-835.

Krueger, R., Thom, D. and Ertl, T. (2014), Visual analysis of movement behavior using web data for context enrichment, in ‘IEEE Pacific Visualization Symposium', IEEE, pp. 193-200. URL: http://ieeexplore.ieee.org/document/6787167/

Ladin, Z. S., Van Nieuland, S., Adalsteinsson, S. A., D'Amico, V., Bowman, J. L., Buler, J. J.,

Baetens, J. M., De Baets, B. and Shriver, W. G. (2018), 'Differential post-fledging habitat use of Nearctic-Neotropical migratory birds within an urbanized landscape', Movement Ecology $\mathbf{6}(1), 17$.

Laube, P. (2014), Computational Movement Analysis, Springer Berlin Heidelberg, Berlin, Heidelberg.

URL: http://www.springer.com/series/10028 http://link.springer.com/10.1007/978-3-540-72680-7_22

Laube, P., Dennis, T., Forer, P. and Walker, M. (2007), 'Movement beyond the snapshot - Dynamic analysis of geospatial lifelines', Computers, Environment and Urban Systems 31(5), 481-501.

URL: https://www.sciencedirect.com/science/article/pii/S0198971507000531

Laube, P. and Purves, R. S. (2011), 'How fast is a cow? Cross-Scale Analysis of Movement Data', Transactions in GIS 15(3), 401-418.

URL: http://doi.wiley.com/10.1111/j.1467-9671.2011.01256.x

Leblond, M., St-Laurent, M. H. and Côté, S. D. (2015), 'Caribou, water, and ice - fine-scale movements of a migratory arctic ungulate in the context of climate change', Movement Ecology 4(1), 14.

Lee, W.-c. and Krumm, J. (2011), Trajectory Preprocessing, in Y. Zheng and X. Zhou, eds, ‘Computing with Spatial Trajectories', Springer, New York, pp. 3-33.

Leguizamon, S., Menenti, M. and Roerink, G. (2000), Characterisation of the dynamic response of the Vegetation Cover in South America by Wavelet Multiresolution analysis of NDVI time series, in 'International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives', Vol. 33.

URL: http://www.isprs.org/proceedings/XXXIII/congress/part7/768_XXXIII-part7.pdf

Liechti, F., Bauer, S., Dhanjal-Adams, K. L., Emmenegger, T., Zehtindjiev, P. and Hahn,
S. (2018), 'Miniaturized multi-sensor loggers provide new insight into year-round flight behaviour of small trans-Sahara avian migrants', Movement Ecology 6(1), 1-10.

Lisovski, S., Schmaljohann, H., Bridge, E. S., Bauer, S., Farnsworth, A., Gauthreaux, S. A., Hahn, S., Hallworth, M. T., Hewson, C. M., Kelly, J. F., Liechti, F., Marra, P. P., Rakhimberdiev, E., Ross, J. D., Seavy, N. E., Sumner, M. D., Taylor, C. M., Winkler, D. W., Wotherspoon, S. J. and Wunder, M. B. (2018), 'Inherent limits of light-level geolocation may lead to over-interpretation', Current Biology 28(3), R99-R100.

URL: http://www.ncbi.nlm.nih.gov/pubmed/29408264

Long, J. a. and Nelson, T. a. (2013), 'A review of quantitative methods for movement data', International Journal of Geographical Information Science (July), 1-27.

Louail, T., Lenormand, M., García, O., Picornell, M., Herranz, R., Frías-Martinez, E., Ramasco, J. J. and Barthelemy, M. (2014), 'From mobile phone data to the spatial structure of cities', Scientific reports 4, 5276.

URL: http://arxiv.org/abs/1401.4540

Louzao, M., Wiegand, T., Bartumeus, F. and Weimerskirch, H. (2014), 'Coupling instantaneous energy-budget models and behavioural mode analysis to estimate optimal foraging strategy: An example with wandering albatrosses', Movement Ecology 2(1), 8 .

URL: http://www.movementecologyjournal.com/content/2/1/8

Lu, L., Kuenzer, C., Wang, C., Guo, H. and Li, Q. (2015), 'Evaluation of Three MODIS-Derived Vegetation Index Time Series for Dryland Vegetation Dynamics Monitoring', Remote Sensing 7(6), 7597-7614.

URL: http://www.mdpi.com/2072-4292/7/6/7597

Lu, X., Liu, R., Liu, J. and Liang, S. (2007), 'Removal of Noise by Wavelet Method to Generate High Quality Temporal Data of Terrestrial MODIS Products', Photogrammetric Engineering \& Remote Sensing 73(10), 1129-1139.

URL: https://pdfs.semanticscholar.org/8e98/cfcc3ea68db5faff9852d351763469317207.pdf
http://openurl.ingenta.com/content/xref?genre=articleधissn=0099-
$1112 \mathcal{G}$ volume $=73$ Gissue $=10 \xi$ spage $=1129$

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M. and Hornik, K. (2018), 'cluster: Cluster Analysis Basics and Extensions. R package version 2.0.7.'.

URL: https://cran.r-project.org/web/packages/cluster/index.html

Mandel, J. T., Bohrer, G., Winkler, D. W., Barber, D. R., Houston, C. S. and Bildstein, K. L. (2011), 'Migration path annotation: Cross-continental study of migration-flight response to environmental conditions', Ecological Applications 21(6), 2258-2268.

Manu, S., Shukla, Y., Rawal, R., Thomas, L. E. and de Dear, R. (2016), 'Field studies of thermal comfort across multiple climate zones for the subcontinent: India Model for Adaptive Comfort (IMAC)', Building and Environment 98, 55-70.

URL: http://linkinghub.elsevier.com/retrieve/pii/S0360132315302171

Martin, J., van Moorter, B., Revilla, E., Blanchard, P., Dray, S., Quenette, P. Y., Allainé, D. and Swenson, J. E. (2013), 'Reciprocal modulation of internal and external factors determines individual movements', Journal of Animal Ecology 82(2), 290-300.

URL: http://doi.wiley.com/10.1111/j.1365-2656.2012.02038.x

Meekan, M. G., Duarte, C. M., Fernández-Gracia, J., Thums, M., Sequeira, A. M., Harcourt, R. and Eguíluz, V. M. (2017), 'The Ecology of Human Mobility', Trends in Ecology and Evolution 32(3), 198-210.

URL: http: //dx.doi.org/10.1016/j.tree.2016.12.006

Meijles, E. W., de Bakker, M., Groote, P. D. and Barske, R. (2014), 'Analysing hiker movement patterns using GPS data: Implications for park management', Computers, Environment and Urban Systems 47, 44-57.

URL: https://www.sciencedirect.com/science/article/pii/S0198971513000665

Meneses, P. R. and Almeida, T. D. (2012), Processamento digital de imagens, Brasília.

Met Office (2007), Fact sheet No. 15 - Weather Radar, Technical Report 3.
URL: http://www.metoffice.gov.uk/learning/library/publications/factsheets

Met Office (2012), 'Met Office Integrated Data Archive System (MIDAS) Land and Marine Surface Stations Data (1853-current)'.

URL: http://catalogue.ceda.ac.uk/uuid/220a65615218d5c9cc9e4785a3234bd0

Met Office (2015), 'How we measure rainfall - Met Office'.
URL: http://www.metoffice.gov.uk/guide/weather/observations-guide/how-we-measurerainfall

Met Office (2016), 'Weather stations - Met Office'.
URL: https://www.metoffice.gov.uk/learning/making-a-forecast/first-steps/observations/weather-stations

MetOffice (2003), ' 1 km Resolution UK Composite Rainfall Data from the Met Office Nimrod System'.

URL: http://catalogue.ceda.ac.uk/uuid/27dd6ffba67f667a18c62de5c3456350

Mills, L. S. and Doak, D. F. (1993), 'The Keystone-Species Concept in Ecology and Conservation', BioScience 43(4), 219-224.

URL: https://academic.oup.com/bioscience/article-lookup/doi/10.2307/1312122

Mohr, C. O. (1947), 'Table of Equivalent Populations of North American Small Mammals', The American Midland Naturalist, 37(1), 223-249.

URL: https://www.jstor.org/stable/pdf/2421652.pdf

Mollgaard, A., Lehmann, S. and Mathiesen, J. (2017), 'Correlations between human mobility and social interaction reveal general activity patterns', PLoS ONE $12(12)$, e0188973.

Moorcroft, P. R. (2012), 'Mechanistic approaches to understanding and predicting mammalian space use: recent advances, future directions', Journal of Mammalogy 93(4), 903-916.

URL: http://www.bioone.org/doi/abs/10.1644/11-MAMM-S-254.1

Motta-Junior, J. C., Queirolo, D. and Bueno, A. d. A. (2014), Feeding ecology: A review, in E. F. Consorte-McCrea, A. Santos, ed., 'Ecology and Conservation of the Maned Wolf: Multidisciplinary Perspectives', CRC Press, Boca Raton, chapter 7, pp. 87-98.

Murtagh, F. and Legendre, P. (2011), 'Ward's Hierarchical Clustering Method: Clustering Criterion and Agglomerative Algorithm'.

URL: http://arxiv.org/abs/1111.6285 http://dx.doi.org/10.1007/s00357-014-9161-z

Musiani, M., Morshed Anwar, S., McDermid, G. J., Hebblewhite, M. and Marceau, D. J. (2010), 'How humans shape wolf behavior in Banff and Kootenay National Parks, Canada', Ecological Modelling 221(19), 2374-2387.

URL: https://www.sciencedirect.com/science/article/pii/S0304380010003078

Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B. and Kent, J. (2000),
'Biodiversity hotspots for conservation priorities', Nature 403(6772), 853-858.
URL: http://www.nature.com/articles/35002501

Nathan, R., Getzb, W. M., Revillac, E., Holyoakd, M., Kadmona, R., Saltze, D. and Smousef, P. E. (2008), 'A movement ecology paradigm for unifying organismal movement research', Proceedings of the National Academy of Sciences 1035(49), 19052-19059.

Nathan, R. and Giuggioli, L. (2013), 'A milestone for movement ecology research', Movement Ecology 1(1), 1-3.

URL: http://www.movementecologyjournal.com/content/1/1/1

Nebel, S. (2010), 'Animal Migration', Nature Education Knowledge 3(10), 77.
URL: https://www.nature.com/scitable/knowledge/library/animal-migration-13259533

Nerlich, B. and Jaspal, R. (2014), 'Images of Extreme Weather: Symbolising Human Responses to Climate Change', Science as Culture 23(2), 253-276.

URL: http://www.tandfonline.com/doi/abs/10.1080/09505431.2013.846311

Neumann, W., Martinuzzi, S., Estes, A. B., Pidgeon, A. M., Dettki, H., Ericsson, G. and Radeloff, V. C. (2015), 'Opportunities for the application of advanced remotely-sensed data
in ecological studies of terrestrial animal movement', Movement Ecology 3(1), 8 .
URL: http://www.movementecologyjournal.com/content/3/1/8

Ng, E. (2012), 'Towards planning and practical understanding of the need for meteorological and climatic information in the design of high-density cities: A case-based study of Hong Kong', International Journal of Climatology 32(4), 582-598.

URL: http://doi.wiley.com/10.1002/joc.2292

Noss, A. and Lima, E. D. S. (2007), 'Vertebrate Conservation and Biodiversity', (October 2014).

URL: http://link.springer.com/10.1007/978-1-4020-6320-6

Novo, E. M. L. d. M. (2010), Sensoriamento remoto: princípios e aplicações, $4^{\text {a }}$ edn, Blucher, São Paulo.

Orellana, D. and Renso, C. (2011), Developing an Interactions Ontology for Characterising Pedestrian Movement Behaviour, in 'Movement-Aware Applications for Sustainable Mobility', number January, IGI Global, chapter 5, pp. 62-86.

URL: http://services.igi-global.com/resolvedoi/resolve.aspx? doi=10.4018/978-1-61520-7695.ch005

Oshan, T., Vandrol, J., Sila-nowicka, K., Fotheringham, S. A. and Pozdnoukhov, A. (2014), Spatiotemporal patterns, Technical report, USTAN.

Palm, E. C., Newman, S. H., Prosser, D. J., Xiao, X., Ze, L., Batbayar, N., Balachandran, S. and Takekawa, J. Y. (2015), 'Mapping migratory flyways in Asia using dynamic Brownian bridge movement models', Movement Ecology 3, 1-10.

URL: http://www.movementecologyjournal.com/content/3/1/3

Palmer, J. R., Espenshade, T. J., Bartumeus, F., Chung, C. Y., Ozgencil, N. E. and Li, K. (2013), 'New Approaches to Human Mobility: Using Mobile Phones for Demographic Research', Demography 50(3), 1105-1128.

URL: http://link.springer.com/10.1007/s13524-012-0175-z

Parent, C., Spaccapietra, S., Renso, C., Andrienko, G., Andrienko, N., Bogorny, V., Damiani, M. L., Gkoulalas-Divanis, A., Macedo, J., Pelekis, N., Theodoridis, Y. and Yan, Z. (2013), 'Semantic trajectories modeling and analysis', ACM Computing Surveys 45(4), 42:1-42:32.

URL: http://dl.acm.org/citation.cfm?doid=2501654.2501656
Parlin, A. F., Nardone, J. A., Dougherty, J. K., Rebein, M., Safi, K. and Schaeffer, P. J. (2018), 'Activity and movement of free-living box turtles are largely independent of ambient and thermal conditions', Movement Ecology 6(1), 1-9.

URL: https://movementecologyjournal.biomedcentral.com/articles/10.1186/s40462-018-0130-8

Pentland, A. P. (1992), 'Fast solutions to physical equilibrium and interpolation problems', The Visual Computer 8(5-6), 303-314.

URL: http://link.springer.com/10.1007/BF01897117
Pentland, A. and Sclaroff, S. (1991), 'Closed-form solutions for physically based shape modeling and recognition', IEEE Transactions on Pattern Analysis and Machine Intelligence 13(7), 715-729.

URL: http://ieeexplore.ieee.org/document/85660/
Pentland, A. and Williams, J. (1989), Good vibrations: model dynamics for graphics and animation, in 'Proceedings of the 16th annual conference on Computer graphics and interactive techniques - SIGGRAPH '89', Vol. 23, ACM Press, New York, New York, USA, pp. 215-222.

URL: http://portal.acm.org/citation.cfm? ${ }^{\text {doid }=74333.74355}$

Pettorelli, N., Gaillard, J. M., Mysterud, A., Duncan, P., Stenseth, N. C., Delorme, D., Van Laere, G., Toïgo, C. and Klein, F. (2006), 'Using a proxy of plant productivity (NDVI) to find key periods for animal performance: The case of roe deer', Oikos $\mathbf{1 1 2 ( 3 ) , 5 6 5 - 5 7 2 .}$

URL: http://doi.wiley.com/10.1111/j.0030-1299.2006.14447.x
Pettorelli, N., Ryan, S., Mueller, T., Bunnefeld, N., Jedrzejewska, B., Lima, M. and Kausrud, K. (2011), 'The Normalized Difference Vegetation Index (NDVI): Unforeseen successes in animal ecology', Climate Research 46(1), 15-27.

Pettorelli, N., Safi, K. and Turner, W. (2014), 'Satellite remote sensing, biodiversity research and conservation of the future', Philosophical Transactions of the Royal Society B: Biological Sciences $369(1643), 20130190$.

URL: http://www.ncbi.nlm.nih.gov/pubmed/24733945

Pettorelli, N., Vik, J. O., Mysterud, A., Gaillard, J.-M., Tucker, C. J. and Stenseth, N. C. (2005), 'Using the satellite-derived NDVI to assess ecological responses to environmental change', Trends in Ecology \& Evolution 20(9), 503-510.

Phithakkitnukoon, S., Leong, T. W., Smoreda, Z. and Olivier, P. (2012), 'Weather Effects on Mobile Social Interactions: A Case Study of Mobile Phone Users in Lisbon, Portugal', $\underline{\text { PLoS }}$ ONE 7(10), e45745.

URL: http://dx.plos.org/10.1371/journal.pone. 0045745

Phithakkitnukoon, S., Smoreda, Z. and Olivier, P. (2012), 'Socio-geography of human mobility: A study using longitudinal mobile phone data', PLoS ONE 7(6), 1-9.

URL: http://www.ncbi.nlm.nih.gov/pubmed/22761748

Ponzoni, F. J. and Shimabukuro, Y. E. (2010), Sensoriamento remoto no estudo da vegetação, $10^{\mathrm{a}}$ edn, Parêntese, São José dos Campos.

Pregnolato, M., Ford, A., Wilkinson, S. M. and Dawson, R. J. (2017), 'The impact of flooding on road transport: A depth-disruption function', Transportation Research Part D: Transport and Environment 55, 67-81.

URL: https://www.sciencedirect.com/science/article/pii/S1361920916308367\#f0005

Priyadarshi, N., Chowdary, V. M., Srivastava, Y. K., Das, I. C. and Jha, C. S. (2017), 'Reconstruction of time series MODIS EVI data using de-noising algorithms', Geocarto International pp. 1-19.

URL: https://doi.org/10.1080/10106049.2017.1333535

Prosser, D. J., Palm, E. C., Takekawa, J. Y., Zhao, D., Xiao, X., Li, P., Liu, Y. and Newman, S. H. (2015), 'Movement analysis of free-grazing domestic ducks in Poyang Lake, China: a dis-
ease connection', International Journal of Geographical Information Science 30(5, SI), 869880.

Purves, R. S., Laube, P., Buchin, M. and Speckmann, B. (2014), 'Moving beyond the point: An agenda for research in movement analysis with real data', Computers, Environment and Urban Systems 47, 1-4.

Queirolo, D. and Motta-Junior, J. C. (2007), 'Prey availability and diet of maned wolf in Serra da Canastra National Park, southeastern Brazil', Acta Theriologica 52(4), 391-402.

Raghunath, H. (2006), Hydrology, New Age International (P) Limited, New Delhi.
Ranacher, P. and Tzavella, K. (2014), 'How to compare movement? A review of physical movement similarity measures in geographic information science and beyond', Cartography and Geographic Information Science 41(3), 286-307.

URL: http://www.ncbi.nlm.nih.gov/pubmed/27019646

Rao, Y., Zhu, X., Chen, J. and Wang, J. (2015), 'An Improved Method for Producing High Spatial-Resolution NDVI Time Series Datasets with Multi-Temporal MODIS NDVI Data and Landsat TM/ETM+ Images', Remote Sensing 7(6), 7865-7891.

URL: http://www.mdpi.com/2072-4292/7/6/7865

Ray, J. C. (2005), Large carnivores and the conservation of biodiversity, Island Press.
Rees, W. (2001), Physical Principles of Remote Sensing, $2^{\text {a }}$ edn, Cambrydge University Press, Cambridge.

Remelgado, R., Leutner, B., Safi, K., Sonnenschein, R., Kuebert, C. and Wegmann, M. (2017), 'Linking animal movement and remote sensing - mapping resource suitability from a remote sensing perspective'.

URL: http://doi.wiley.com/10.1002/rse2.70

Remelgado, R., Leutner, B., Safi, K., Sonnenschein, R., Kuebert, C. and Wegmann, M. (2018), 'Linking animal movement and remote sensing - mapping resource suitability from a remote
sensing perspective', Remote Sensing in Ecology and Conservation 4(3), 211-224.
URL: http://doi.wiley.com/10.1002/rse2.70

Richardson, P. L., Wakefield, E. D. and Phillips, R. A. (2018), 'Flight speed and performance of the wandering albatross with respect to wind', Movement Ecology $\mathbf{6}(1), 3$.

URL: https://movementecologyjournal.biomedcentral.com/articles/10.1186/s40462-018-0121-9

Roberge, J. and Angelstam, P. (2004), 'Usefulness of the Umbrella Species Concept as a Conservation Tool', Conservation Biology 18(1), 76-85.

URL: http://doi.wiley.com/10.1111/j.1523-1739.2004.00450.x

Robinson, W. D., Bowlin, M. S., Bisson, I., Shamoun-Baranes, J., Thorup, K., Diehl, R. H., Kunz, T. H., Mabey, S. and Winkler, D. W. (2010), 'Integrating concepts and technologies to advance the study of bird migration', Frontiers in Ecology and the Environment 8(7), 354361.

Rodrigues, F. H., Desbiez, A., Grando, R. L., Lacerda, A. C., Hass, A. and Silva, W. R. (2014), Conservation of a Population of Maned Wolves, Chrysocyon brachyurus, on a Small Reserve in the Cerrado, in E. F. Consorte-McCrea, A. Santos, ed., 'Conservation of a Population of Maned Wolves, Chrysocyon brachyurus, on a Small Reserve in the Cerrado - Ecology and Conservation of the Maned Wolf', CRC Press, Boca Raton, chapter 10, pp. 131-152.

URL: https://www.taylorfrancis.com/books/9781466512603/chapters/10.1201\%2Fb1560717

Roeleke, M., Teige, T., Hoffmeister, U., Klingler, F. and Voigt, C. C. (2018), 'Aerial-hawking bats adjust their use of space to the lunar cycle', Movement Ecology $\mathbf{6}(1), 11$.

URL: https://movementecologyjournal.biomedcentral.com/articles/10.1186/s40462-018-0131-7

Rouse, J. W., Haas, R. H., Schell, J. A. and Deering, D. W. (1973), Monitoring vegetation systems in the Great Plains with ERTS, in S. C. Freden and M. A. Becker, eds, 'Third

ERTS Symposium', Vol. 1 of SP-351, 3010-317, NASA, National Library of Agriculture (NAL), Greenbelt, pp. 309-317.

Royal Meteorological Society (2017), 'Beaufort Scale'.
URL: https://www.rmets.org/weather-and-climate/observing/beaufort-scale

Ryan, R. M., Bernstein, J. H. and Brown, K. W. (2010), 'Weekends, Work, and Well-Being: Psychological Need Satisfactions and Day of the Week Effects on Mood, Vitality, and Physical Symptoms', Journal of Social and Clinical Psychology 29(1), 95-122.

Safi, K., Kranstauber, B., Weinzierl, R., Griffin, L., Rees, E. C., Cabot, D., Cruz, S., Proaño, C., Takekawa, J. Y., Newman, S. H., Waldenström, J., Bengtsson, D., Kays, R., Wikelski, M. and Bohrer, G. (2013), 'Flying with the wind: scale dependency of speed and direction measurements in modelling wind support in avian flight', Movement Ecology 1(1), 1-13.

URL: http://movementecologyjournal.biomedcentral.com/articles/10.1186/2051-3933-1-4 http://www.movementecologyjournal.com/content/1/1/4

Sakamoto, T., Yokozawa, M., Toritani, H., Shibayama, M., ISHITSUKA, N. and OHNO, H. (2005), 'A crop phenology detection method using time-series MODIS data', Remote Sensing of Environment 96(3-4), 366-374.

URL: https://www.sciencedirect.com/science/article/pii/S0034425705001057
http://linkinghub.elsevier.com/retrieve/pii/S0034425705001057

Salvatori, V., Skidmore, A. K., Corsi, F. and Van Der Meer, F. (1999), 'Estimating temporal independence of radio-telemetry data on animal activity', Journal of Theoretical Biology 198(4), 567-574.

URL: http://www.ncbi.nlm.nih.gov/pubmed/10373355

Schick, R. S., Loarie, S. R., Colchero, F., Best, B. D., Boustany, A., Conde, D. a., Halpin, P. N., Joppa, L. N., McClellan, C. M. and Clark, J. S. (2008), 'Understanding movement data and movement processes: current and emerging directions.', Ecology letters 11(12), 1338-50.

URL: http://www.ncbi.nlm.nih.gov/pubmed/19046362

Schick, R. S., Roberts, J. J., Eckert, S. A., Halpin, P. N., Bailey, H., Chai, F., Shi, L. and Clark, J. S. (2013), 'Pelagic movements of pacific leatherback turtles (dermochelys coriacea) highlight the role of prey and ocean currents', Movement Ecology $\mathbf{1}(1), 11$.

URL: http://movementecologyjournal.biomedcentral.com/articles/10.1186/2051-3933-1-11

Schneider, F., Lesko, W. and Garret, W. (1980), 'Helping behavior in hot, comfortable, and cold temperatures: A field study', Environment and Behavior $12(2), 231-241$.

URL: http://journals.sagepub.com/doi/pdf/10.1177/0013916580122007

Sedlmair, M., Meyer, M. and Munzner, T. (2012), 'Design Study Methodolgy: Reflections from the Trenches and the Stacks', IEEE Trans. Visualization and Computer Graphics 1(12), 2431-2440.

SEOS (2018), 'Introduction to Remote Sensing :Spectral Reflectance Properties'.
URL: http://www.seos-project.eu/modules/remotesensing/remotesensing-c01-p05.html

Shaffer, S. A., Cockerham, S., Warzybok, P., Bradley, R. W., Jahncke, J., Clatterbuck, C. A., Lucia, M., Jelincic, J. A., Cassell, A. L., Kelsey, E. C. and Adams, J. (2017), 'Population-level plasticity in foraging behavior of western gulls (Larus occidentalis)', Movement Ecology 5(1), 27.

URL: https://movementecologyjournal.biomedcentral.com/articles/10.1186/s40462-017-0118-9

Shamoun-Baranes, J., Bouten, W., Belle, J. V., Buurma, L. and Gasteren, H. V. (2004), Influence of meteorological conditions on flight altitudes of birds, in ' 16 th Biometeorology and Aerobiology', Vancouver, pp. 96-98.

URL: https://ams.confex.com/ams/AFAPURBBIO/techprogram/paper_79056.htm

Shamoun-Baranes, J., Bouten, W. and van Loon, E. E. (2010), 'Integrating Meteorology into Research on Migration', Integrative and Comparative Biology 50(3), 280-292.

URL: https://academic.oup.com/icb/article-lookup/doi/10.1093/icb/icq011

Shamoun-Baranes, J., van Loon, E. E., Purves, R. S., Speckmann, B., Weiskopf, D. and

Camphuysen, C. J. (2012), ‘Analysis and visualization of animal movement', Biology Letters 8(1), 6-9.

Shamoun-Baranes, J., van Loon, E., van Gastere, H., van Belle, J., Bouten, W. and Buurma, L. (2006), 'A Comparative Analysis of the Influence of Weather on the Flight Altitudes of Birds', Bulletin of the American Meteorological Society 87(1), 47-62.

URL: http://journals.ametsoc.org/doi/abs/10.1175/BAMS-87-1-47 http://journals.ametsoc.org/doi/10.1175/BAMS-87-1-47

Sharif, M. and Alesheikh, A. A. (2017a), 'Context-aware movement analytics: implications, taxonomy, and design framework', Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery p. e1233.
URL: http://doi.wiley.com/10.1002/widm. 1233

Sharif, M. and Alesheikh, A. A. (2017b), 'Context-awareness in similarity measures and pattern discoveries of trajectories: a context-based dynamic time warping method', GIScience and Remote Sensing 54(3), 426-452.

Shen, L. and Stopher, P. R. (2014), 'Review of GPS Travel Survey and GPS Data-Processing Methods', Transport Reviews 34(3), 316-334.

URL: http://dx.doi.org/10.1080/01441647.2014.903530
Shi, L., Chi, G., Liu, X. and Liu, Y. (2015), 'Human mobility patterns in different communities: a mobile phone data-based social network approach', Annals of GIS 21(1), 15-26.

URL: http://www.tandfonline.com/doi/full/10.1080/19475683.2014.992372
Shoval, N. and Isaacson, M. (2007), 'Sequence alignment as a method for human activity analysis in space and time', Annals of the Association of American ... 97(May 2006), 282297.

Sila-Nowicka, K. (2016), Using GPS Trajectories For Further Understanding of Spatial Behaviour, Phd thesis, University of St Andrews.

Sila-Nowicka, K. F. A. (2018), 'Calibrating spatial interaction models from GPS tracking data: An example of retail behaviour (Under review)', Computers, Environment and Urban Systems .

Siła-Nowicka, K., Vandrol, J., Oshan, T., Long, J. A., Demšar, U., Fotheringham, A. S., Sila-Nowicka, K., Vandrol, J., Oshan, T., Long, J. A., Demšar, U. and Fotheringham, A. S. (2016), 'Analysis of human mobility patterns from GPS trajectories and contextual information', International Journal of Geographical Information Science 30(5), 881-906.

URL: http://www.tandfonline.com/doi/abs/10.1080/13658816.2015.1100731 http://www.tandfonline.com/doi/full/10.1080/13658816.2015.1100731

Sillero-Zubiri, C. and Gottelli, D. (1995), 'Diet and Feeding Behavior of Ethiopian Wolves (Canis simensis)', Journal of Mammalogy 76(2), 531-541.

URL: https://academic.oup.com/jmammal/article-lookup/doi/10.2307/1382361

Sluiter, R. (2008), Interpolation methods for climate data - Literature review, Technical report, KNMI.

Stahler, D. R., Smith, D. W. and Guernsey, D. S. (2006), 'Foraging and Feeding Ecology of the Gray Wolf (Canis lupus): Lessons from Yellowstone National Park, Wyoming, USA', The Journal of Nutrition $\mathbf{1 3 6}(7), 1923 \mathrm{~S}-1926 \mathrm{~S}$.

URL: https://academic.oup.com/jn/article/136/7/1923S/4664711

Stanley, K., Yoo, E.-H., Paul, T. and Bell, S. (2018), 'How many days are enough?: capturing routine human mobility', International Journal of Geographical Information Science 32(7), 1485-1504.

URL: https://www.tandfonline.com/doi/full/10.1080/13658816.2018.1434888

Steadman, R. G. (1994), 'Norms of apparent temperature in Australia', Australian Meteorology Magazine 43, 1-16.

URL: http://www.bom.gov.au/amm/docs/1994/steadman.pdf

Stienen, E. W., Desmet, P., Aelterman, B., Courtens, W., Feys, S., Vanermen, N., Verstraete,
H., Van de Walle, M., Deneudt, K., Hernandez, F., Houthoofdt, R., Vanhoorne, B., Bouten, W., Buijs, R.-J., Kavelaars, M. M., Müller, W., Herman, D., Matheve, H., Sotillo, A. and Lens, L. (2016), 'GPS tracking data of Lesser Black-backed Gulls and Herring Gulls breeding at the southern North Sea coast', ZooKeys 555, 115-124.

URL: http://zookeys.pensoft.net/articles.php?id=6173

Stover, V. W., Consulting Associates, N. and Edward McCormack, S. D. (2012), 'The Impact of Weather on Bus Ridership in Pierce County The Impact of Weather on Bus Ridership in Pierce County, Washington', Journal of Public Transportation 15(1), 95-110.

URL: https://www.nctr.usf.edu/wp-content/uploads/2012/04/JPT15.1Stover.pdf
Strabala, K. (2018), 'MODIS Cloud Mask User's Guide'.
URL: http://cimss.ssec.wisc.edu/modis/CMUSERSGUIDE.PDF

Studer, M., Ritschard, G., Gabadinho, A. and Müller, N. S. (2011), 'Discrepancy analysis of state sequences', Sociological Methods and Research 40(3), 471-510.

Sun, Y., Fan, H., Helbich, M. and Zipf, A. (2013), Progress in Location-Based Services, number January 2013 in 'Lecture Notes in Geoinformation and Cartography', Springer Berlin Heidelberg, Berlin, Heidelberg.

URL: http://link.springer.com/10.1007/978-3-642-34203-5

Sutherland, W. J., Freckleton, R. P., Godfray, H. C. J., Beissinger, S. R., Benton, T., Cameron, D. D., Carmel, Y., Coomes, D. A., Coulson, T., Emmerson, M. C., Hails, R. S., Hays, G. C., Hodgson, D. J., Hutchings, M. J., Johnson, D., Jones, J. P. G., Keeling, M. J., Kokko, H., Kunin, W. E., Lambin, X., Lewis, O. T., Malhi, Y., Mieszkowska, N., Milner-Gulland, E. J., Norris, K., Phillimore, A. B., Purves, D. W., Reid, J. M., Reuman, D. C., Thompson, K., Travis, J. M. J., Turnbull, L. A., Wardle, D. A. and Wiegand, T. (2013), 'Identification of 100 fundamental ecological questions', Journal of Ecology 101(1), 58-67.

Szentimrey, T., Bihari, Z. and Szalai, S. (2010), Comparison of Geostatistical and Meteorological Interpolation Methods (What is What?), in 'Spatial Interpolation for Climate Data:

The Use of GIS in Climatology and Meteorology', ISTE, London, UK, pp. 45-56.
URL: http://doi.wiley.com/10.1002/9780470612262.ch4

Tang, M., Agrawal, P., Nie, F., Pongpaichet, S. and Jain, R. (2016), 'A graph based multimodal geospatial interpolation framework', Proceedings - IEEE International Conference on Multimedia and Expo 2016-Augus, 1-6.

Tang, M., Wu, X., Agrawal, P., Pongpaichet, S. and Jain, R. (2017), 'Integration of Diverse Data Sources for Spatial PM2.5 Data Interpolation', IEEE Transactions on Multimedia $19(2), 408-417$.

Thakuriah, P. V., Sila-Nowicka, K. and Paule, J. G. (2016), 'Sensing Spatiotemporal Patterns in Urban Areas : Analytics and Visualizations Using the Integrated Multimedia City Data', Built Environment (October), 415-429.

Therrien, J.-f., Pinaudcebccnrsfr, D. P., Gauthier, G., Lecomte, N. and Bildstein, K. L. (2015), 'Is pre-breeding prospecting behaviour affected by snow cover in the irruptive snowy owl ? A test using state-space modelling and environmental data annotated via Movebank', Movement Ecology pp. 1-8.

Thierry, B., Chaix, B. and Kestens, Y. (2013), 'Detecting activity locations from raw GPS data: A novel kernel-based algorithm', International Journal of Health Geographics 12, 14. URL: http://www.ncbi.nlm.nih.gov/pubmed/23497213

Thorup, K., Tøttrup, A. P., Willemoes, M., Klaassen, R. H. G., Strandberg, R., Vega, M. L., Dasari, H. P., Araújo, M. B., Wikelski, M. and Rahbek, C. (2017), 'Resource tracking within and across continents in long-distance bird migrants', Science Advances 3(1), e1601360.

URL: http://advances.sciencemag.org/content/advances/3/1/e1601360.full.pdf http://advances.sciencemag.org/lookup/doi/10.1126/sciadv. 1601360

Tobler, W. R. (1970), 'A Computer Movie Simulating Urban Growth in the Detroit Region', Economic Geography 46, 234.

URL:
https://dds.cepal.org/infancia/guia-para-estimar-la-pobreza-infantil/bibliografia/capitulo-IV/Tobler Waldo (1970) A computer movie simulation urban growth in the Detroit region.pdf http://www.jstor.org/stable/143141?origin=crossref

Toller, G. and Isaacman, A. (2006), MODIS Level 1B Product User's Guide, Technical report, Goddard Space Flight Center, Greenbelt, MD.

URL: https://mcst.gsfc.nasa.gov/sites/mcst.gsfc/files/file_attachments/M1054.pdf

Tomkiewicz, S. M., Fuller, M. R., Kie, J. G. and Bates, K. K. (2010), 'Global positioning system and associated technologies in animal behaviour and ecological research.', $\underline{\text { Philosophical }}$ transactions of the Royal Society of London. Series B, Biological sciences 365(1550), 21632176.

URL: http://www.ncbi.nlm.nih.gov/pubmed/20566494

Toole, J. L., Herrera-Yaqüe, C., Schneider, C. M. and González, M. C. (2015), 'Coupling human mobility and social ties', Journal of the Royal Society Interface $12(105), 20141128$.

URL: http://www.ncbi.nlm.nih.gov/pubmed/25716185

Traffic Scotland (@trafficscotland) (2017), "Twitter advanced search: "cars only" from:trafficscotland since:2013-09-28 until:2014-01-10'.

URL: https://twitter.com/search?l=छq=\"cars only\%22 from\%3Atrafficscotland since\%3A2013-09-28 until\%3A2014-01-10Esrc=typd

Tucker, P. and Gilliland, J. (2007), 'The effect of season and weather on physical activity: A systematic review', Public Health 121(12), 909-922.

Turk, M. and Pentland, A. (1991), 'Eigenfaces for Recognition', Journal of Cognitive Neuroscience $3(1)$, 71-86.

URL: http://www.ncbi.nlm.nih.gov/pubmed/23964806

Turk, V. (2013), 'Facebook's migration station: what our location updates say about human movement'.

URL: https://www.wired.co.uk/article/facebook-migration-station

Turner, W., Spector, S., Gardiner, N., Fladeland, M., Sterling, E. and Steininger, M. (2003), 'Remote sensing for biodiversity science and conservation', Trends in Ecology \& Evolution 18(6), 306-314.

URL: https://www.sciencedirect.com/science/article/pii/S0169534703000703 http://linkinghub.elsevier.com/retrieve/pii/S0169534703000703

Tveito, O. E. (2010), The Developments in Spatialization of Meteorological and Climatological Elements, in 'Spatial Interpolation for Climate Data', ISTE, London, UK, chapter 6, pp. 7386.

URL: http://doi.wiley.com/10.1002/9780470612262.ch6

UK Weather (2013), 'Dunfermline monthly Sunset Sunrise times UK'.
URL: http://www.ukweathercams.co.uk

Urbano, F. and Cagnacci, F. (2014), Spatial Database for GPS Wildlife Tracking Data, Springer International Publishing, Cham.

URL: http://link.springer.com/10.1007/978-3-319-03743-1

Urbano, F., Cagnacci, F., Calenge, C., Dettki, H., Cameron, A. and Neteler, M. (2010), 'Wildlife tracking data management: a new vision', Philosophical Transactions of the Royal Society B: Biological Sciences 365(1550), 2177-2185.

USGS (2016), Landsat 8 (L8) Data Users Handbook, 2 edn, USGS, Sioux Falls, South Dakota.

Vanky, A. P., Verma, S. K., Courtney, T. K., Santi, P. and Ratti, C. (2017), 'Effect of weather on pedestrian trip count and duration: City-scale evaluations using mobile phone application data', Preventive Medicine Reports 8, 30-37.

URL: https://www.sciencedirect.com/science/article/pii/S2211335517301158

Vazquez-Prokopec, G. M., Bisanzio, D., Stoddard, S. T., Paz-Soldan, V., Morrison, A. C., Elder, J. P., Ramirez-Paredes, J., Halsey, E. S., Kochel, T. J., Scott, T. W. and Kitron, U. (2013), 'Using GPS Technology to Quantify Human Mobility, Dynamic Contacts and Infec-
tious Disease Dynamics in a Resource-Poor Urban Environment', PLoS ONE 8(4), e58802.
URL: http://www.ncbi.nlm.nih.gov/pubmed/23577059

VDI (2008), 'VDI 3787, Part I: Environmental Meteorology, Methods for the human biometeorological evaluation of climate and air quality for the urban and regional planning at regional level. Part I: Climate'.

Versichele, M., Neutens, T., Delafontaine, M. and Van de Weghe, N. (2012), 'The use of Bluetooth for analysing spatiotemporal dynamics of human movement at mass events: A case study of the Ghent Festivities', Applied Geography 32(2), 208-220.

URL: http://linkinghub.elsevier.com/retrieve/pii/S0143622811000944

Wan, N. and Lin, G. (2016), 'Classifying Human Activity Patterns from Smartphone Col-
lected GPS data: A Fuzzy Classification and Aggregation Approach', Transactions in GIS $\mathbf{2 0}(6), 869-886$.

Wang, Y., Nickel, B., Rutishauser, M., Bryce, C. M., Williams, T. M., Elkaim, G. and Wilmers, C. C. (2015), 'Movement, resting, and attack behaviors of wild pumas are revealed by triaxial accelerometer measurements', Movement Ecology 3(1), 2.

URL: http://www.movementecologyjournal.com/content/3/1/2

Widmann, M., Kato, A., Raymond, B., Angelier, F., Arthur, B., Chastel, O., Pellé, M., Raclot, T. and Ropert-Coudert, Y. (2015), 'Habitat use and sex-specific foraging behaviour of Adélie penguins throughout the breeding season in Adélie Land, East Antarctica', Movement Ecology 3(1), 30.

URL: http://www.movementecologyjournal.com/content/3/1/30

Wikelski, M., Kays, R. W., Kasdin, N. J., Thorup, K., Smith, J. A. and Swenson, G. W. (2007), 'Going wild: what a global small-animal tracking system could do for experimental biologists.', The Journal of experimental biology $\mathbf{2 1 0}(2), 181-186$.

URL: http://www.ncbi.nlm.nih.gov/pubmed/17210955

Willis, A., Gjersoe, N., Havard, C., Kerridge, J. and Kukla, R. (2004), 'Human movement
behaviour in urban spaces: implications for the design and modelling of effective pedestrian environments', Environment and Planning B: Planning and Design 31(6), 805-828.

URL: http://www.envplan.com/abstract.cgi?id=b3060

Worfolk, J. B. (1997), 'Keep Frail Elders Warm! The thermal instabilities of the old have not received sufficient at- tention in basic educational programs', Geriatric Nursing 18(1), 7-11. URL: http://www.gnjournal.com/article/S0197-4572(97)90123-3/pdf

Xavier, G. and Dodge, S. (2014), An exploratory visualization tool for mapping the relationships between animal movement and the environment, in 'Proceedings of the 2nd ACM SIGSPATIAL International Workshop on Interacting with Maps - MapInteract '14', ACM Press, New York, New York, USA, pp. 36-42.

URL: http://dx.doi.org/10.1145/2483669.2483682

Xavier, G., Dodge, S. and Wong, W. Y. (2018), ‘DynamoVis - Dynamic Visualization of Animal Movement Data.', Retrieved from the Data Repository for the University of Minnesota .

URL: https://doi.org/10.13020/D6PH49

Xiao, G., Juan, Z. and Zhang, C. (2015), 'Travel mode detection based on GPS track data and Bayesian networks', Computers, Environment and Urban Systems 54, 14-22.

URL: https://www.sciencedirect.com/science/article/pii/S0198971515000587

Xu, C., Li, Y., Hu, J., Yang, X., Sheng, S. and Liu, M. (2012), 'Evaluating the difference between the normalized difference vegetation index and net primary productivity as the indicators of vegetation vigor assessment at landscape scale', Environ Monit Assess 184, 12751286.

Xu, R. and Wunsch, D. C. (2009), Clustering, John Wiley \& Sons, Ltd., New Jersey.

Xue, J. and Su, B. (2017), 'Significant Remote Sensing Vegetation Indices: A Review of Developments and Applications', Journal of Sensors 2017, 1-17.

URL: http://www.agrotic.org/blog/wp-content/uploads/2017/07/Significant-Remote-Sensing-Vegetation-Indices.pdf https://www.hindawi.com/journals/js/2017/1353691/

Yan, Z., Chakraborty, D., Parent, C., Spaccapietra, S. and Aberer, K. (2013), 'Semantic trajectories', ACM Transactions on Intelligent Systems and Technology 4(3), 1.

URL: http://dx.doi.org/10.1145/2483669.2483682
Yang, C., Xiao, M., Ding, X., Tian, W., Zhai, Y., Chen, J., Liu, L. and Ye, X. (2018), ‘Exploring human mobility patterns using geo-tagged social media data at the group level', Journal of Spatial Science pp. 1-18.

URL: https://www.tandfonline.com/doi/full/10.1080/14498596.2017.1421487
Yoda, K., Tajima, T., Sasaki, S., Sato, K. and Niizuma, Y. (2012), 'Influence of local wind conditions on the flight speed of the great cormorant phalacrocorax carbo', International Journal of Zoology 2012, 1-7.
URL: http://www.hindawi.com/journals/ijz/2012/187102/

Zha, Y., Gao, J. and Ni, S. (2003), 'Use of normalized difference built-up index in automatically mapping urban areas from TM imagery', International Journal of Remote Sensing 24(3), 583-594.

URL: http://www.tandfonline.com/doi/abs/10.1080/01431160304987

Zhan, W., Chen, Y., Zhou, J., Wang, J., Liu, W., Voogt, J., Zhu, X., Quan, J. and Li, J. (2013), 'Disaggregation of remotely sensed land surface temperature: Literature survey, taxonomy, issues, and caveats', Remote Sensing of Environment 131(19), 119-139.

URL: http://linkinghub.elsevier.com/retrieve/pii/S0034425712004804
Zhang, T., Ramakrishnan, R. and Livny, M. (1996), BIRCH: An efficient clustering method for very large databases, in 'Proceedings of the 1996 ACM SIGMOD international conference on Management of data - SIGMOD '96', Vol. 25, ACM Press, New York, New York, USA, pp. 103-114.

URL: http://portal.acm.org/citation.cfm?doid=233269.233324
Zivin, J. G. (2014), 'Temperature and the Allocation of Time: Implications for Climate Change', Journal of Labor Economics 32(1).

URL: https://www.journals.uchicago.edu/doi/pdfplus/10.1086/671766


[^0]:    ${ }^{1}$ Theoretical limit according to Jensen (2006), these limits may vary from satellite to satellite but will stay within this range.

