Recreational and tourism value of healthy rivers

Final Report

Water & Environment Research Program Theme 4:

Research Question 12.2

Deliverable 12.2.4

Report prepared by

James C.R. Smart[1,2]

Jeremy Harte[1]

[1]Australian Rivers Institute, Griffith University, Brisbane

[2]School of Environment and Science, Griffith University, Brisbane

For:

Murray-Darling Water and Environment Research Program (MD-WERP)

August 2024

AUSTRALIAN RIVERS INSTITUTE

Griffith UNIVERSITY

An Australian Government Initiative

Funded through the Murray–Darling Water and Environment Research Program

Version	Issue Date
v0.5	13th August 2024
V1.0	23rd August 2024
V1.1	6th September 2024

Approvals

Author signature: J. Smart .... Date 6/09/2024

Author Name: Associate Professor James C. R. Smart

Author signature: .... Date 6/09/2024

Author Name: Mr Jeremy Harte

Author contact details

☏ 07 3735 5290

✉ j.smart@griffith.edu.au

https://www.griffith.edu.au/australian-rivers-institute

Report citation

Smart, J. C. R. and Harte, J. 2024. Murray Darling Basin Authority Water & Environment Research Program Theme 4: Research Question 12.2, Deliverable 12.2.4 *Recreational and tourism value of healthy rivers – Final Report*. Australian Rivers Institute, Griffith University, Brisbane.

1. Contents

List of Figures

List of Tables

Abbreviations

ABS	Australian Bureau of Statistics
BuD	'Birder user Day' derived from geo-located, time-stamped bird species lists from the eBird.org citizen science website.
eBird	A citizen science website on which bird watchers can log bird species checklists. Each checklist is time-stamped and geo-located. (eBird, Cornell Lab of Ornithology, Ithaca, New York: www.ebird.org)
CC	Australian Hydrological Geo-spatial Fabric contracting catchments (CCs), grouped into River Regions within the Murray Darling Basin.
i.i.d	Independent and identically distributed (used when considering whether regression residuals conform with regression assumption)
MDB	Murray Darling Basin
MDBA	Murray Darling Basin Authority
MNDVi	Modified Normalised Difference Water Index used to report average percentage inundation area per sub-catchment at monthly resolution. Derived using imagery from the green and Shortwave-Infrared (SWIR) bands from the Sentinel-2 satellite.
NB	Negative binomial distribution and the negative binomial count data regression model which uses a negative binomial distribution for its count data dependent variable.
NDVi	Normalised Difference Vegetation Index used to report average landscape 'greenness' per sub-catchment at monthly resolution. Derived from Advanced Very High Resolution Radiometer instruments onboard the NOAA series of satellites operated by the US.
PuD	'Photo user Day' derived from geo-located, time-stamped photo posts from the Flickr.com photo posting website.
RE	Random effect: a sub-catchment-specific constant term in a count data regression model. The random effect terms are assumed to be drawn from a normal distribution. The count data regression model estimates the variance of this distribution as part of the modelling process.
VIF	Variance inflation factor: used to assess whether the extent of collinearity between dependent variables is likely to be problematic for regression estimation.
VGI	Volunteered geographic information
ZIP	Zero-inflated Poisson count data regression model
ZINB	Zero-inflated Negative Binomial count data regression model

Acknowledgements

The authors acknowledge and thank the Murray Darling Basin Authority's Water and Environment Research Program (WERP) for providing funding for this research.

eBird citizen science data were obtained for research use, with authorization (see Appendix 1), from the Cornell Lab of Ornithology, 159 Sapsucker Woods Road, Ithaca, NY, 14850, USA; eBird@cornell.edu (Sullivan et al., 2014). The authors particularly thank Jenna Curtis (eBird Project Leader) for helpful comments and suggestions.

2. Executive Summary

This research study from WERP Theme 4 RQ12.2 seeks to determine whether improvements in riverine ecosystem health increase tourism and recreation visitation rates at locations across the Basin.

Metrics of monthly visitation count at sub-catchment scale are produced as Birder user Days (BuDs) from geo-located, time stamped bird species sighting lists posted to the citizen science website eBird, and Photo user Days (PuDs) derived from geo-located, time stamped photo posts to the Flickr photo posting web site. Monthly BuD and PuD counts for 3170 sub-catchments, grouped into 28 River Regions, are collated over 73 months from March 2013 to March 2019. This produces a data set of more than 230,000 data points.

Count data regressions are used to quantify potential linkages between monthly BuD or PuD counts for sub-catchments within a River Region and a suite of sub-catchment resolution spatio-temporal, spatial and temporal variables. Remotely sensed monthly average normalized difference vegetation index (NDVI) 'greenness' and percentage inundation area per sub-catchment (detected using modified normalised difference water index (MNDWI)) are used as alternative proxies for environmental condition. These environmental condition proxies together with monthly total rainfall, monthly maximum temperature, spatial variables such as distance from a sub-catchment's centroid to the nearest local population centre and travel time through the road network to the nearest major city, and temporal factors such as month of the year and year are used as potential drivers of BuD or PuD counts in the regressions.

Sub-catchment BuD and PuD counts vary considerably across space and through time. Many sub-catchments record no BuD or PuD counts for many, or all, months between March 2013 and March 2019, while other sub-catchments are visitation hotspots. This makes it difficult for the regressions to predict BuD or PuD counts accurately month by month, so complex models have to be used. Regression models were developed initially for sub-catchments in the Lachlan River Region, then equivalent models were used to predict BuD counts for sub-catchments in 19 River Regions across the Basin. The main objective was to determine which of the potential driving factors were convincingly associated with BuD counts in the various River Regions.

Results showed that inaccessibility (proxied by distance from the nearest local town) and sub-catchment land area were convincingly associated with BuD counts. Association was negative for distance from the nearest town, and positive for catchment land area. Of the two candidate environmental condition variables, percentage inundation area in a sub-catchment was identified as being positively associated with BuD counts in several River Regions, when some forms of regression model were used. Whilst these associations look promising, further investigation will be required to determine whether they are an artefact of the complex modelling, or a real effect for regions such as the Macquarie-Bogan, Lachlan, Murrumbidgee and Murray Riverina in which large flows into lower catchment wetlands could plausibly attract high numbers and variety of birds which then attract visits from birdwatchers.

This research has shown that eBird-derived BuDs show real promise as a metric of visitation rate that is available across the whole of the Basin. Cross calibration with visitation by the broader outdoor recreational community should be pursued as strong progress in this direction is already evident in the literature (Cameron and Kolstoe, 2022b; Langemeyer et al., 2023). Whilst birdwatching (proxied by BuDs) may plausibly increase as riverine ecosystem health improves, other forms of outdoor recreation and ecotourism may also respond to changes in riverine ecosystem health. It is important to keep in mind that findings relating to associations between environmental condition and birdwatching visitation are only part of the broader picture regarding outdoor recreation and ecotourism visitation.

A major challenge remains in identifying better metrics of riparian environmental condition that can be applied across the whole Basin. Remotely sensed ecosystem condition metrics developed for use in SEEA Ecosystem Accounts could be worth exploring (Harwood et al., 2023).

RQ12.2 Extension B (‘eBird’) research will conclude by using the regression methodologies developed in this study to explore whether environmental condition proxied by surface water area (from Digital Earth Australia’s Water Observations from Space) or site-specific NDVI greenness at individual birdwatching hotspots across the Basin affect birder visitation rate (as proxied by BuDs) at those sites.

3. Research Objective

Project Objective:

Theme 4 RQ12.2 set out to address the central question: do improvements in riverine ecosystem health increase tourism and recreation visitation rates to Basin locations?

If so, can we quantify:

(i) the increase in visitation rates to particular locations and
(ii) the increase in tourism/recreation expenditure in local economies that follow from these higher visitation rates?

The objective of this project is to establish and, if possible, quantify the link between riverine ecosystem health and tourism/recreation visitation rates for main catchments across the Basin. We do this by analysing location-specific proxies for visitation rate derived from social media data (photo posts to www.flickr.com) or citizen science data (bird species viewing lists posted to www.eBird.org), together with location-specific proxies of environmental condition (environmental greenness via remotely-sensed NDVi or remotely sensed percentage inundation area per sub-catchment) and other relevant location-specific factors that could affect visitation rate across the Basin (e.g., travel time through the road network to major cities, 'crow flies' distances to nearest local population centre, nearest point on the road network, distance to nearest wetland, distance to nearest Ramsar site, distance to nearest Important Bird Area). Counts of Flickr-derived Photo User Days (PuDs) or eBird-derived Birder User Days (BuDs) are summed per month over a 6-year period (March 2013 – March 2019), i.e., 73 months in total. Analyses are undertaken at sub-catchment resolution separately for main catchments across the Basin

This Report describes findings on the central research question regarding the potential link between visitation rate and environmental condition. The Final Report on RQ12.2 Extension B (Deliverable RQ12.2.8) will report on approaches for using LGA-resolution data from Tourism Research Australia to estimate the increases in tourism/recreation expenditure that could follow from improvements in environmental condition for indicative main catchments across the Basin.

4. Background

4.1. Tourism and recreation in the Murray Darling Basin

The total gross value added (GVA) across Tourism Research Australia's (TRA) regions that intersect with the MDB is estimated at $6 billion for financial year 2019-20 (Aither, 2022). This is very similar to the $6 billion total gross value of irrigated agricultural production (GVIAP) reported by the same source for irrigated agricultural production in the Basin for the year 2020. Note, however, that GVIAP is defined as the gross value of agricultural commodities that are produced with the assistance of irrigation, where 'the gross value of commodities produced' is the value placed on recorded production at the wholesale prices realised in the marketplace. In contrast, GVA (the statistic quoted for tourism) is defined as the total value of outputs at basic prices *less* the total intermediate consumption at purchasers' prices (Australian Bureau of Statistics, 2022). 'Tourism' is classified in the tourism industry as involving a trip of 30kms or more from home (Rod Hillman, Head of Ecotourism Australia: pers comm.)

Despite tourism's substantial contribution to local Basin economies, it has not yet been possible to quantify the link between riverine ecosystem condition and tourism/recreation visitation numbers. Until such a link can be established convincingly, it is not possible to evidence the additional tourism and recreation expenditures that could be generated by improving riverine ecosystem health. Research Question 12.2 in WERP Theme 4 aims to address this research gap.

In this research we use Birder user Days (BuDs) derived from geo-located, time stamped bird species sighting lists posted to the citizen science website eBird (eBird, 2021; Kolstoe and Cameron, 2017; Sullivan et al., 2014, 2009), and Photo user Days (PuDs) (Ciesielski and Stereńczak, 2021; Ghermandi, 2022a; Langemeyer et al., 2023; Sinclair et al., 2020a) derived from geo-located, time stamped photo posts to the Flickr photo posting web site as proxies for visitation for birdwatching and (primarily) landscape photography purposes. Whilst birdwatching and landscape photography may plausibly increase as riverine ecosystem health improves (Camacho-Valdez et al., 2020; Sinclair et al., 2021, 2019a), other forms of outdoor recreation and ecotourism may also respond to changes in riverine ecosystem health (Andersen et al., 2019; Huang et al., 2020; Stithou et al., 2012). It is important to keep in mind that findings relating to associations between environmental condition and visitation for birdwatching or outdoor photography purposes are only part of the broader picture regarding outdoor recreation and ecotourism visitation.

4.2. Volunteered geographic information as a data source

Traditionally, visitation rate has been estimated directly via surveys (Tourism Research Australia, 2022). However, when adequate sample sizes and representative sampling are required, direct survey approaches are time consuming and costly. Consequently, data volunteered by members of the public via social media (termed Volunteered Geographic Information (VGI) (Cameron and Kolstoe, 2022a; Goodchild, 2007)) are being used increasingly as a data source to quantify overall visitation rate (e.g., (Teles da Mota et al., 2022), spatial and temporal variation in visitation (Hausmann et al., 2019), visitor activities, experiences and sentiment (Bhatt and Pickering, 2022), and for estimating the demand for and valuation of visitor experiences (Sinclair et al., 2019, 2020).

Geo-located, time-stamped photo posts from the photo-sharing sites Flickr (https://www.flickr.com) and Instagram (https://www.instagram.com) [Instagram removed the ability to obtain geo-locations from Instagram posts in 2018 (Toivonen et al., 2019)] have been widely used in studies of visitation rate, spatial and temporal variation in visitation, and to investigate which types of activities visitors undertake at particular locations via automated or manual categorisation of photo content. User-posted walking and cycling routes from activity sites such as Strava (https://www.strava.com) have been used to provide high resolution spatial information on spatial and temporal variation in visitor activity (Norman et al., 2019; Norman and Pickering, 2019). Twitter posts have also been widely used as data sources for sentiment analysis (Ghermandi, 2022b; Teles da Mota et al., 2022; Teles da Mota and Pickering, 2021; Tenkanen et al., 2017).

Geo-located photo posts on Flickr have been particularly widely used to produce proxies for visitation rate. Recent reviews by (Ghermandi, 2022b; Wilkins et al., 2021b) have generally found acceptable levels of correlations between geo-located Flickr photo-posts and separate assessments of visitation rate from, for example, separately administered site-based surveys. A well-established recommendation is to amalgamate posts from the same photo-poster on at a given location on a particular day into a single ‘photo user day’ (PUD) for visitation rate assessment (Ghermandi, 2022b; Tenkanen et al., 2017; Wilkins et al., 2021b). Use of PUDs as a proxy for visitation rate has been found to be reasonably reliable (Ghermandi, 2022b; Sinclair et al., 2020b), although reliability can be affected by factors such as the overall popularity of the location, the age profile of visitors to that location and temporal variation in the popularity of the social media site.

Geo-located, time-stamped bird sighting checklists on the eBird citizen science website (https://ebird.org/home) have also been used as a source of VGI in visitation research (Cameron and Kolstoe, 2022b; Guilfoos et al., 2023; Kolstoe and Cameron, 2017). The eBird site operators (Cornell Lab of Ornithology) facilitate and encourage use of eBird data for scientific research and have granted permission for eBird data to be used in this project (see authorisation email in Appendix 1). As described for Flickr-derived PuDs, bird species listing posts from the same eBird-poster at a given location on a particular day can be amalgamated into a single ‘Birder user day’ (BuD) for visitation rate assessment. A recent special issue of the Ecosystem Services journal is dedicated to future opportunities for use of crowd-sourced VGI data in cultural ecosystem service assessments (Langemeyer et al., 2023).

The research described in this report explores use of Flickr- and eBird-derived VGI data to identify potential drivers of visitation rate for locations across the Basin.

5. Data

5.1. Spatial framing: River Regions and sub-catchments

The cross-sectional 'individuals' in the visitation rate analyses in this research are Australian Hydrological Geo-spatial Fabric contracting catchments (CCs), grouped into River Regions (Figure 1). Hereafter, contracting catchments are referred to as 'sub-catchments' of River Regions. The overall objective of Research Question 12.2 is to determine how monthly biophysical time-varying and time-fixed driving factors within sub-catchments of a River Region influence visitation rate as proxied by monthly per-sub-catchment counts of BuDs or PuDs.

Analyses of drivers affecting visitation rate is undertaken separately per River Region. This is because it proved infeasible to conduct analyses at whole of Basin, or Northern Basin vs. Southern Basin scale due to the size and complexity of the resulting models. A whole-of-Basin model would contain 73 monthly data points (see Figure 2 and Section 5.2.1) for each of 3170 sub-catchments: more than 230,000 data points in all.

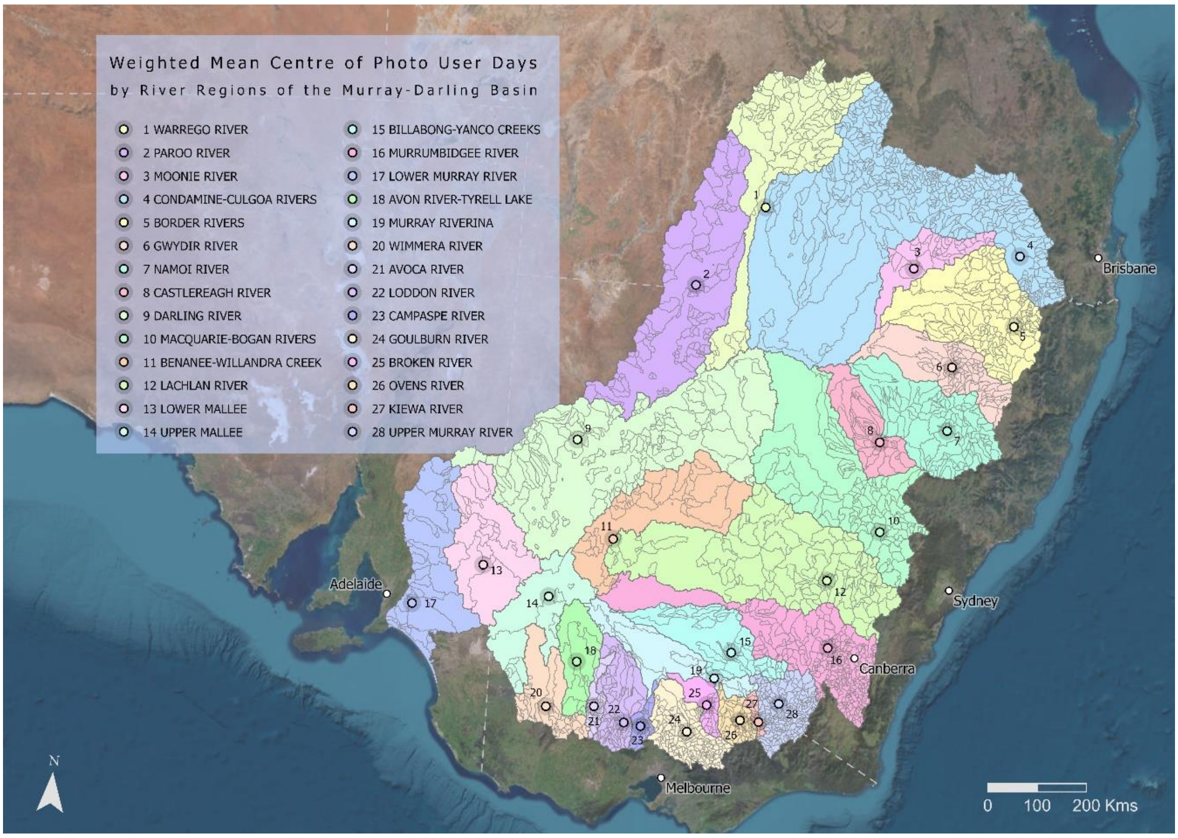

Figure 1: River Regions and their constituent sub-catchments in the Murray Darling Basin. River Regions are denoted by colours and identified via key number. Sub-catchments are shown in grey outline within each River Region.

Sub-catchment size varies considerably between headwaters and the mid and lower catchment. In River Regions such as the Murrumbidgee, Lachlan and Macquarie-Bogan that extend from headwaters to lower catchment, sub-catchments in the lower reaches are considerably larger than those in the headwaters. In contrast, River Regions such as the Upper Mallee and Lower Mallee in the lower floodplain of the Basin typically contain fewer, but larger, sub-catchments.

The number of sub-catchments per River Region is reported in Table 1.

Table 1: Number of sub-catchments per River Region in the Murray Darling Basin

River Region	Sub-catchment count
Avoca River	26
Avon River-Tyrell Lake	18
Benanee-Willandra Creek	44
Billabong-Yanco Creeks	47
Border Rivers	220
Broken River	26
Campaspe River	39
Castlereagh River	47
Condamine-Culgoa Rivers	529
Darling River	177
Goulburn River	207
Gwydir River	103
Kiewa River	23
Lachlan River	140
Loddon River	83
Lower Mallee	25
Lower Murray River	54
Macquarie-Bogan Rivers	145
Moonie River	40
Murray Riverina	25
Murrumbidgee River	349
Namoi River	121
Ovens River	84
Paroo River	117
Upper Mallee	17
Upper Murray River	256
Warrego River	152
Wimmera River	56
Basin total	3170

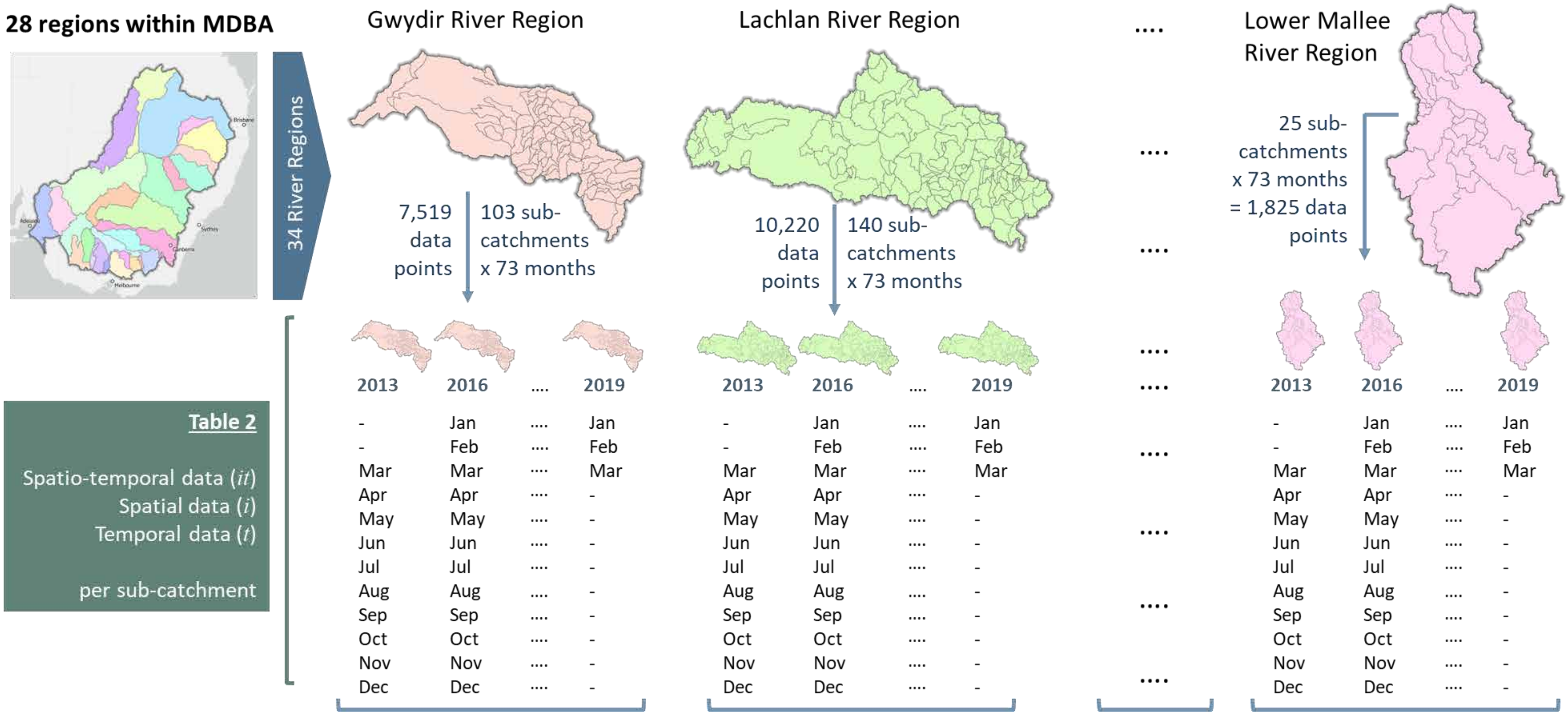

Figure 2: Data concept

5.2. Data for analysis

This section describes the data used in analyses to identify linkages between BuDs or PuDs as VGI metrics of visitation rate and potential time-varying and time-fixed driving factors, including proxies for environmental condition.

5.2.1. Visitation proxies: Birder user Days (BuDs) and Photo user Days (PuDs)

Following the literature, monthly counts of eBird-derived BuDs and Flickr-derived PUDs per sub-catchment in a given River Region are used as a proxy for visitation rate. BuD and PuD data were collated for the period March 2013 to March 2019: 73 consecutive months. This period corresponds with the availability of remotely-sensed NDVi data as a metric of environmental greenness (see Figure 7 and Table 2).

This analysis uses VGI from birdwatchers who post their geo-located, time-stamped bird sighting checklists on the eBird citizen science website (https://ebird.org/home). The eBird site operators (Cornell Lab of Ornithology) facilitate and encourage use of eBird data for scientific research and have granted permission for eBird data to be used in this project (see authorisation email in Appendix 1).

eBird checklists are tagged with an anonymous username unique to each birdwatcher ('eBirder') who posts their bird sighting checklists on eBird. eBirders post checklists for separate bird watching sessions. Posted checklists report bird sightings by species, location (latitude and longitude), date and time. The initial data set downloaded from eBird for this research contains 623,000 bird sightings in checklist posts for locations in the Murray Darling Basin between January 2000 and January 2024 (eBird, 2024).

For use in this research, separate bird sightings are grouped into checklist posts by individual eBirders at distinct geographic locations on different days. These distinct postings are termed 'Birder User Days' (BuDs), following the established naming convention for 'Photo User Days' (PuDs) for the equivalent processing of Flickr photo posts by individual Flickr users at different geographic locations on different days.

Summed monthly PuD and BuD counts per sub-catchment across the Basin as a whole for the 73 months March 2013 to March 2019 are shown in Figure 3 and Figure 4. These figures show some similarity in the spatial distribution of total BuD and PuD counts across the Basin. Note, however, that the total BuD count is considerably larger than the total PuD count. January and July BuD and PuD counts per sub-catchment across the Basin are shown in Figure 5 and Figure 6. The difference in the quantity of BuDs relative to PuDs is also evident here.

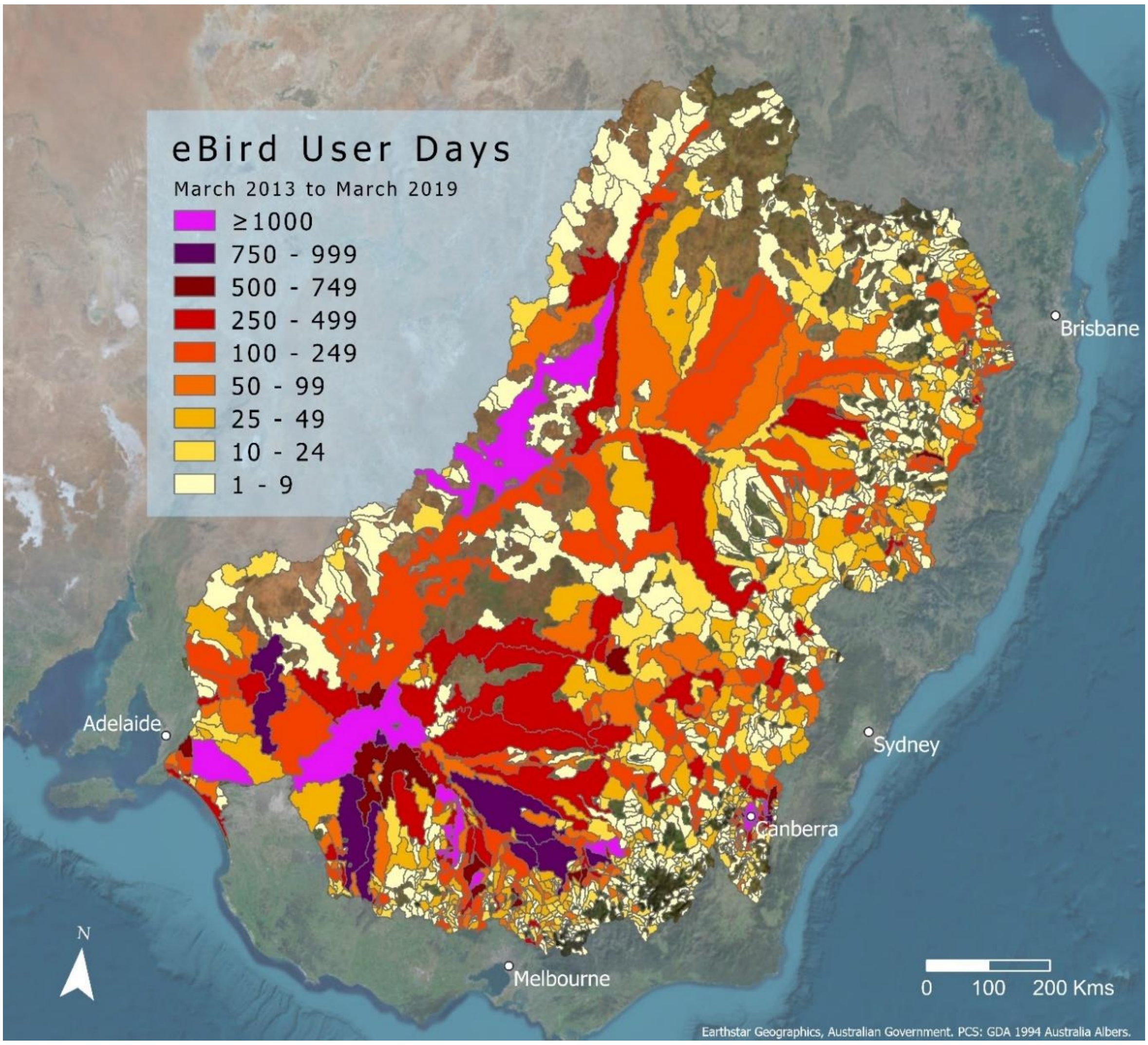

Figure 3: Summed monthly BuD counts per sub-catchment across the Basin for the 73 months March 2013 to March 2019.

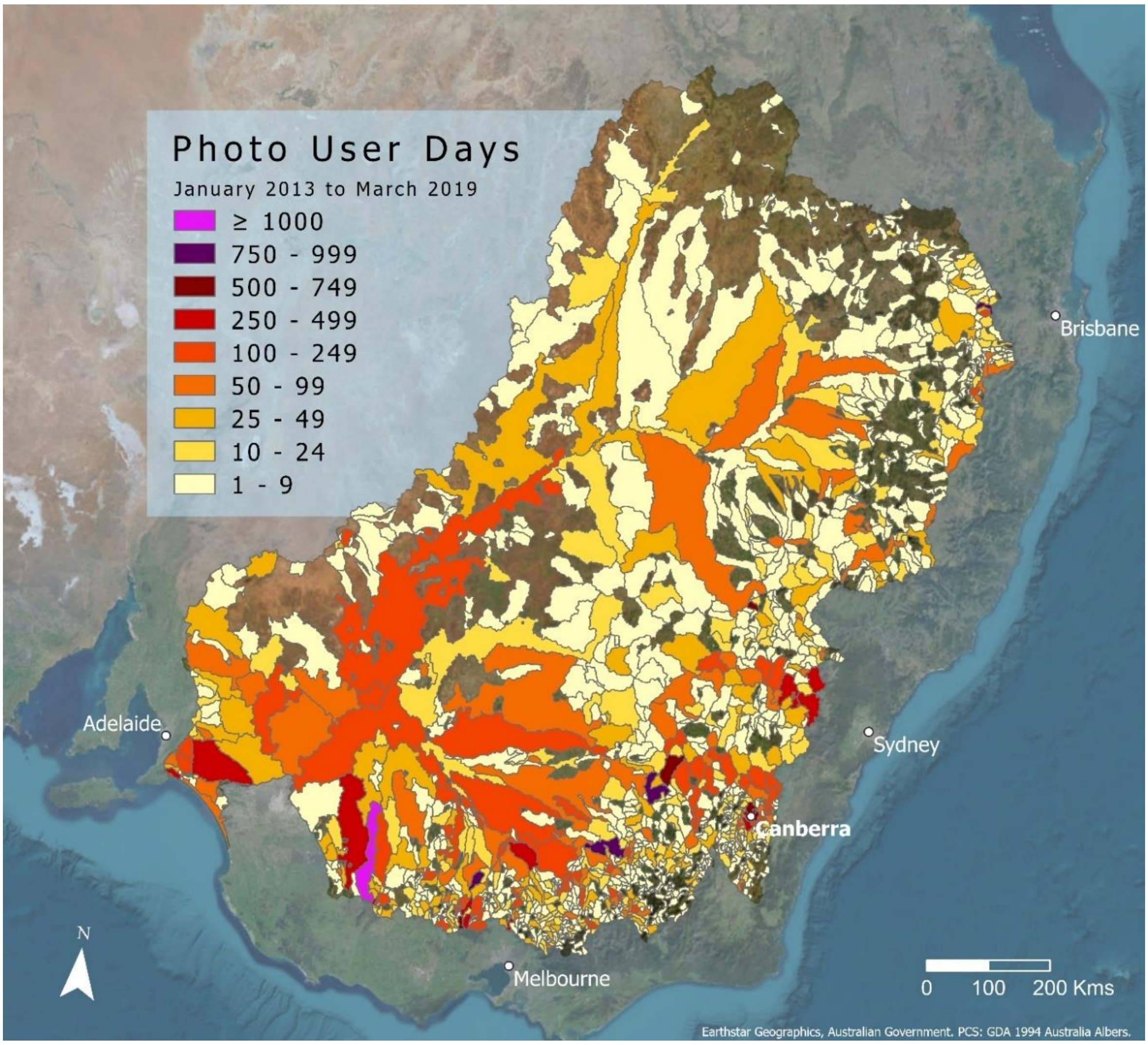

Figure 4: Summed monthly PuD counts per sub-catchment across the Basin for the 75 months January 2013 to March 2019.

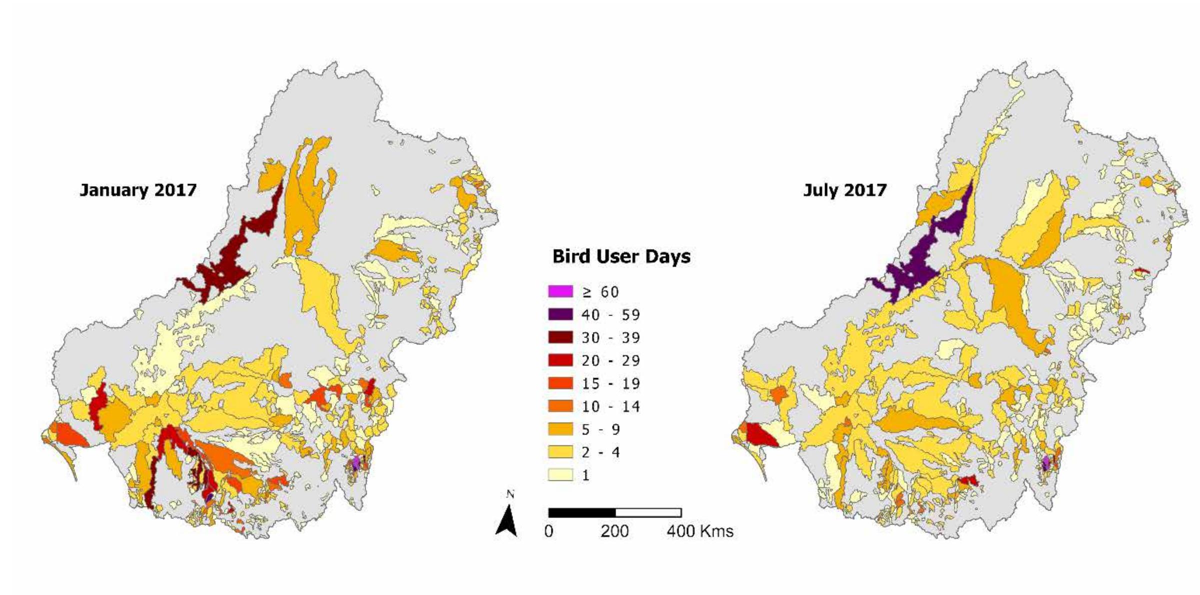

Figure 5: BuD count per sub-catchment across the Basin for January and July 2017.

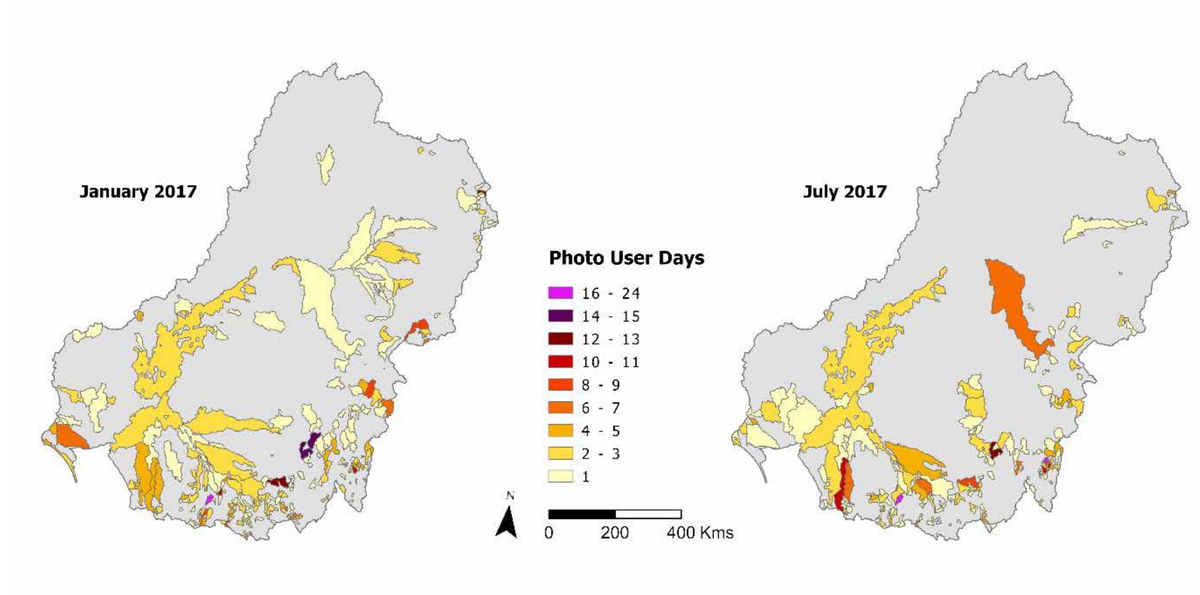

Figure 6: PuD count per sub-catchment across the Basin for January and July 2017.

5.2.2. Potential drivers of visitation across space and time

Multiple papers have used VGI to explore factors that could potentially influence visitation rate at a particular site. For a recent review see Teles da Mota and Pickering (2020). Sonter et al. (2016) and Wilkins, Howe & Smith (2021a) provide useful additional perspectives. Drawing on the above sources and the broader travel and destination choice literature, the data listed in Table 2 were collated on factors that could potentially affect visitation rate to locations across the Murray Darling Basin.

Figure 7 to Figure 15 on subsequent pages illustrate spatial and temporal variation in these data across the Basin.

Table 2: Drivers potentially influencing visitation rate to sub-catchments within River Regions across the Basin.

Potential driver	Unit of measurement	Data derivation
Location-specific, monthly time varying factors per sub-catchment		
'Greenness' proxied via Average Normalized Difference Vegetation Index (NDVI) per month	Numeric (-1 to +1)	Monthly average NDVI value from BOM
Inundation area as % of sub-catchment area* Monthly total inundated area detected using Modified Normalised Difference Water Index (MNDWI)	%	MNDWI via the method described in Gould et al. (2023)
Maximum daily temperature for the month	°C	Daily maximum temperature from SILO
Total rainfall for the month	mm	Monthly rainfall from SILO
Location-specific, time-invariant factors per sub-catchment		
Travel time from nearest major population centre	minutes	Calculated via road network and assumed travel speed **
Distance from nearest road	km	Straight line distance from centroid
Distance from nearest local population centre	km	Straight line distance from centroid
Distance from nearest Important Bird Area	km	Straight line distance from centroid
Distance from nearest Ramsar wetland	km	Straight line distance from centroid
Distance from nearest wetland (DIWA)	km	Straight line distance from centroid
Distance from nearest protected area (CAPAD)	km	Straight line distance from centroid
Average resident population 2013:2019	individuals	From ABS estimated resident population grid, updated annually
Area of irrigated farmland	ha	From land use mapping
Universal time varying factors		
Month of year: Jan to Dec	categorical	
Year: 2013 to 2019	numeric or categorical	
Month index: 1 to 73	numeric	

* Percentage inundation area only reported for sub-catchments for which surface water coverage was observed over 95% or more of their total area.

** Major population centres used: Adelaide, Brisbane, Canberra, Central Coast, Geelong, Gold Coast, Melbourne, Newcastle, Sunshine Coast, Sydney, Toowoomba, Wollongong.

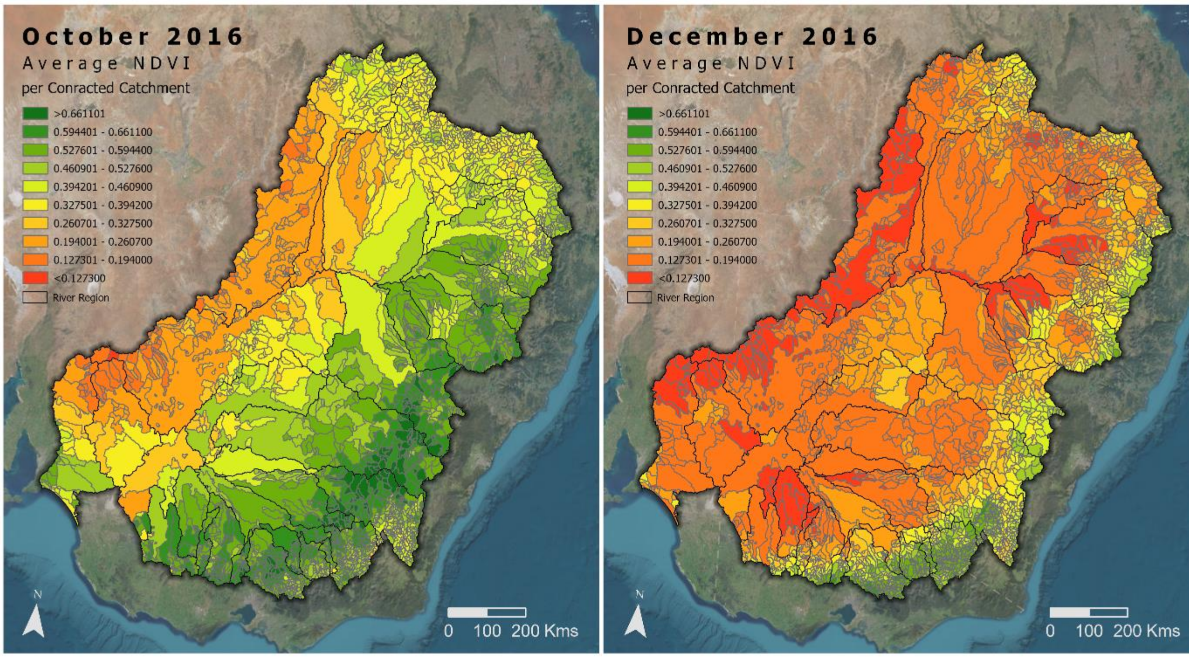

Figure 7: Average NDVI 'environmental greenness' for sub-catchments across the Basin in October and December 2016.

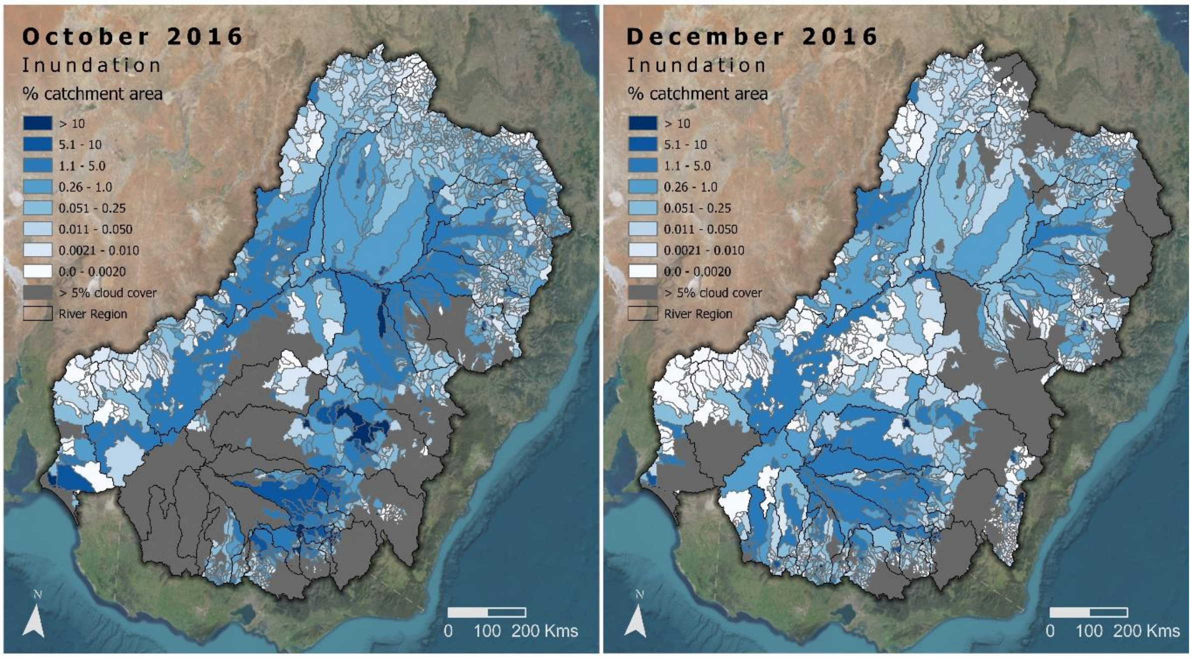

Figure 8: Percentage inundation area for sub-catchments across the Basin in October and December 2016. Sub-catchments with more than 5% cloud cover are shown in grey. These are marked as 'missing data' in analyses.

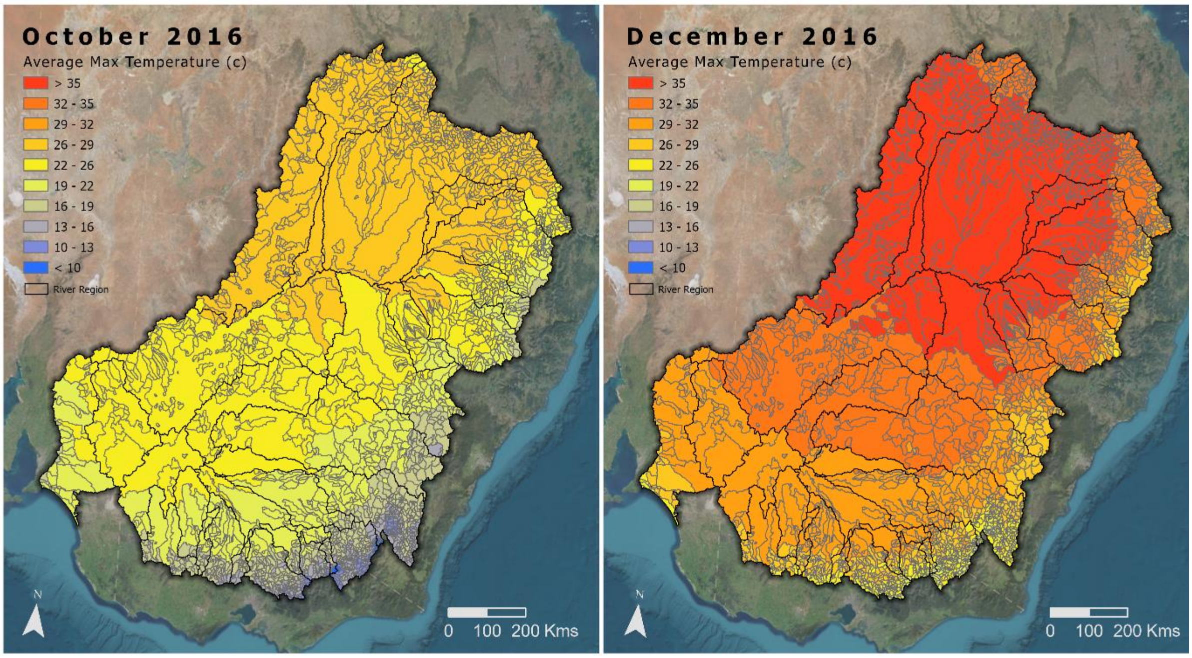

Figure 9: Maximum daily temperature per month for sub-catchments across the Basin in October and December 2016.

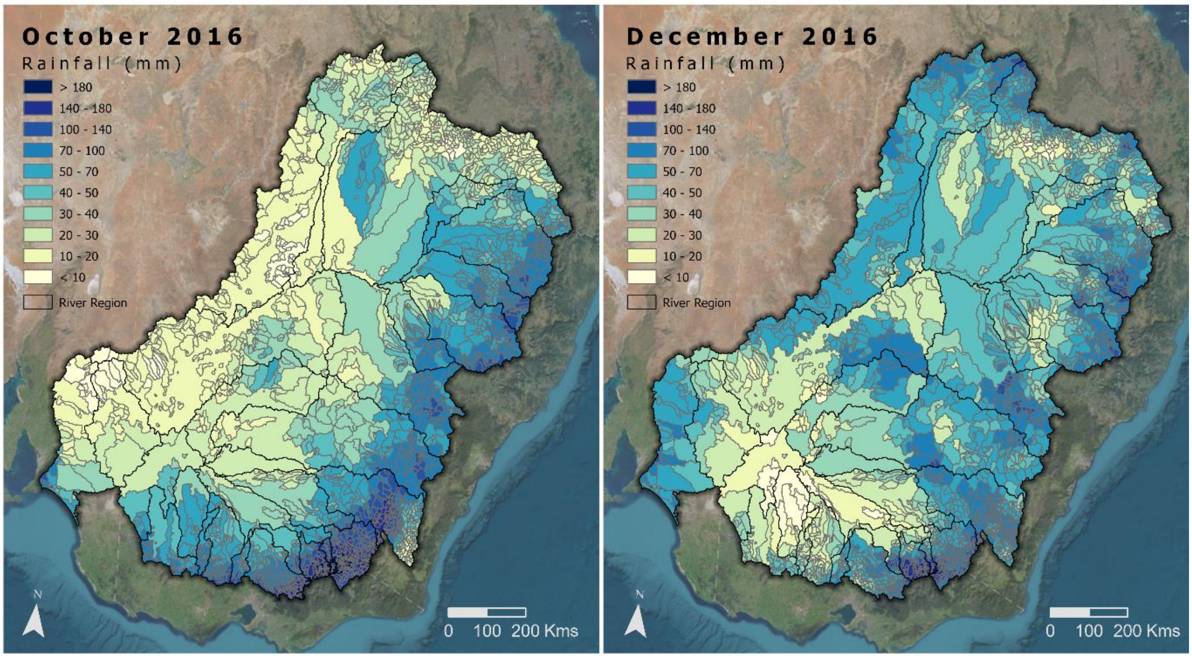

Figure 10: Total monthly rainfall per sub-catchment across the Basin in October and December 2016.

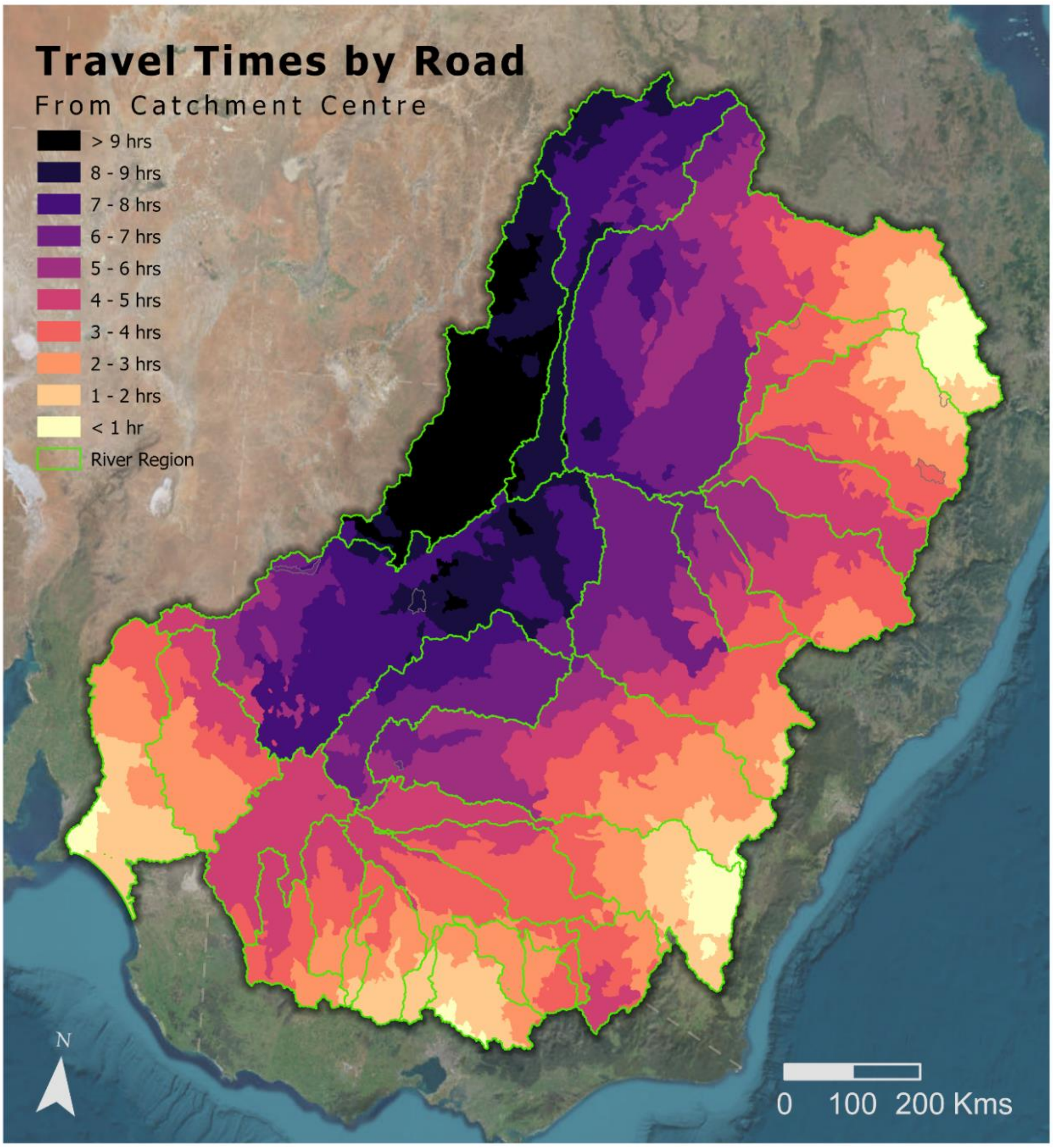

Figure 11: Travel time from sub-catchment geographic centroid to nearest major city.

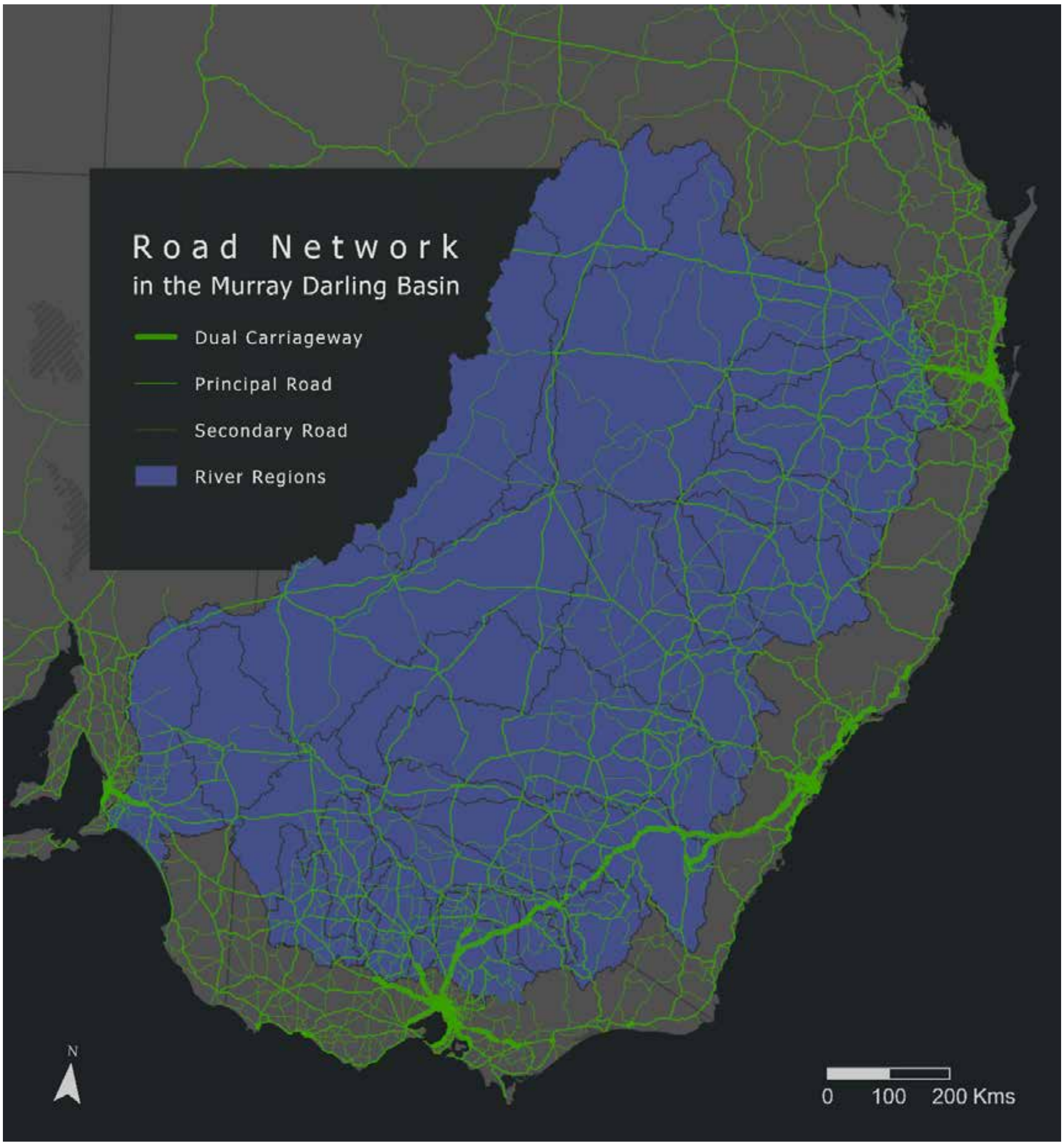

Figure 12: Road network used for travel time calculation.

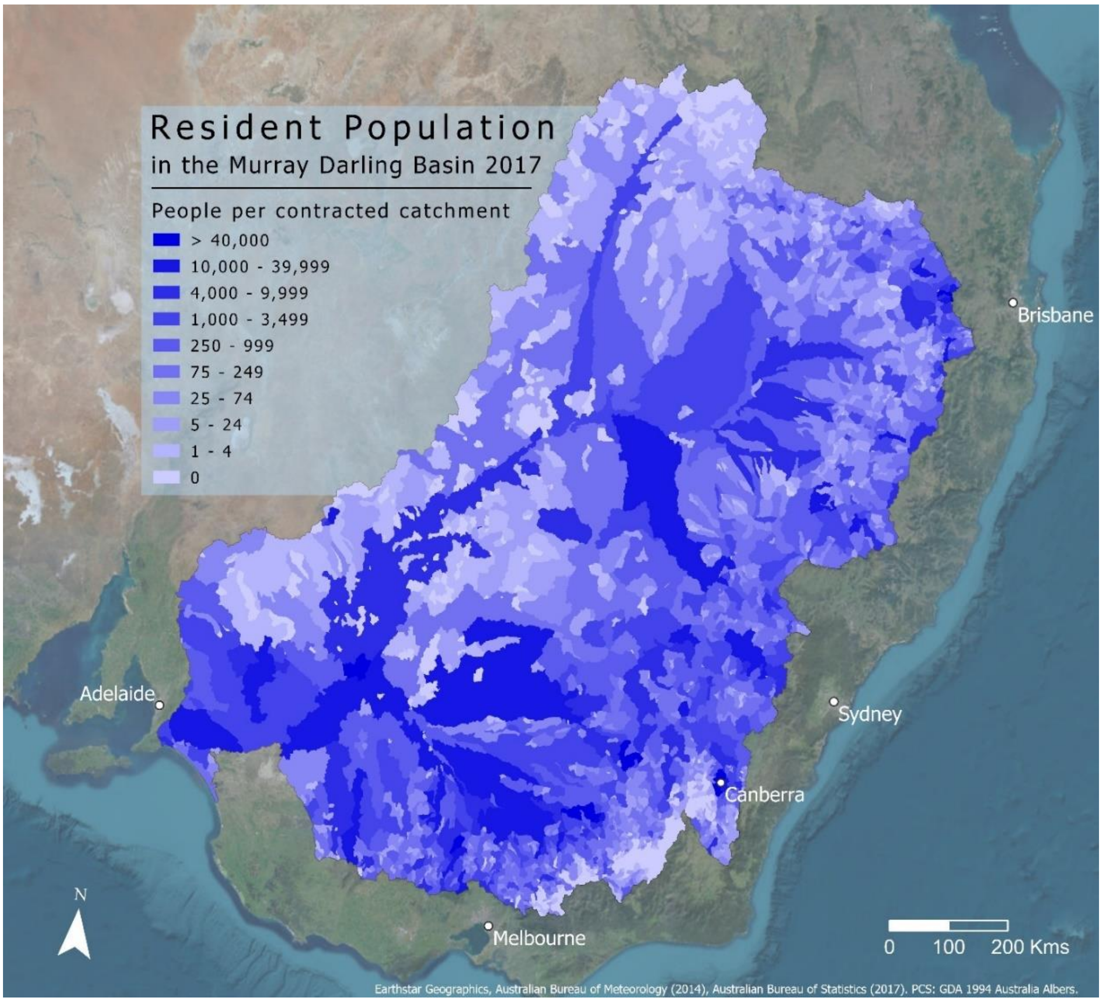

Figure 13: Total population per sub-catchment in 2017

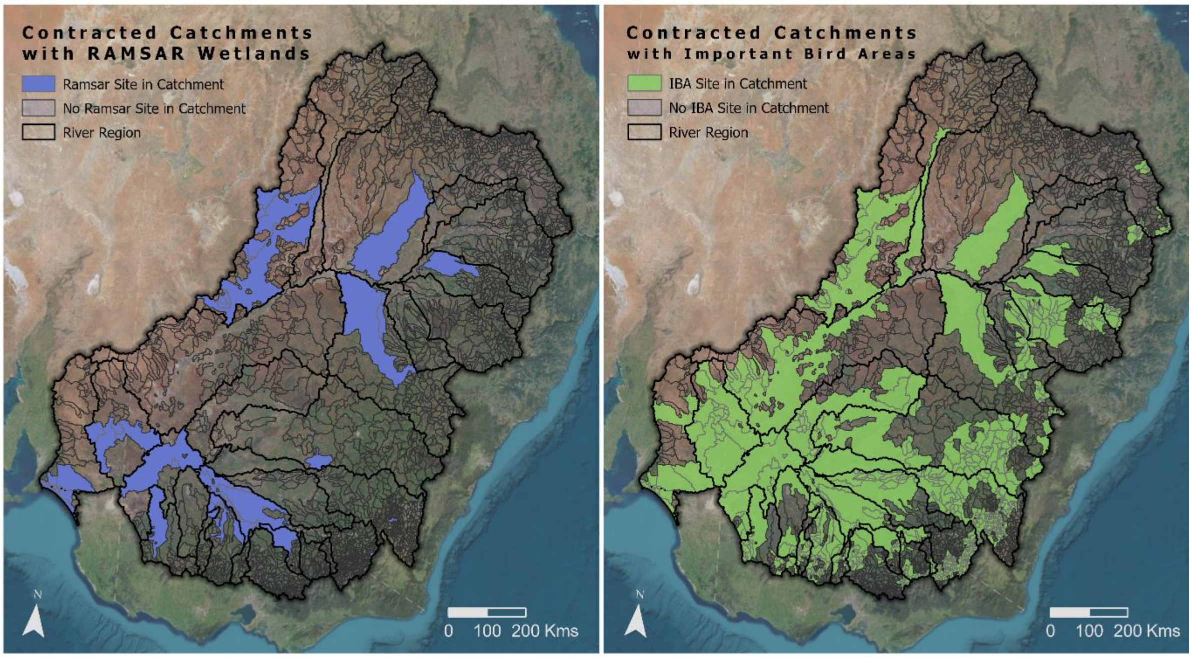

Figure 14: Sub-catchments containing (a) Ramsar-listed wetlands and (b) Important Bird Areas.

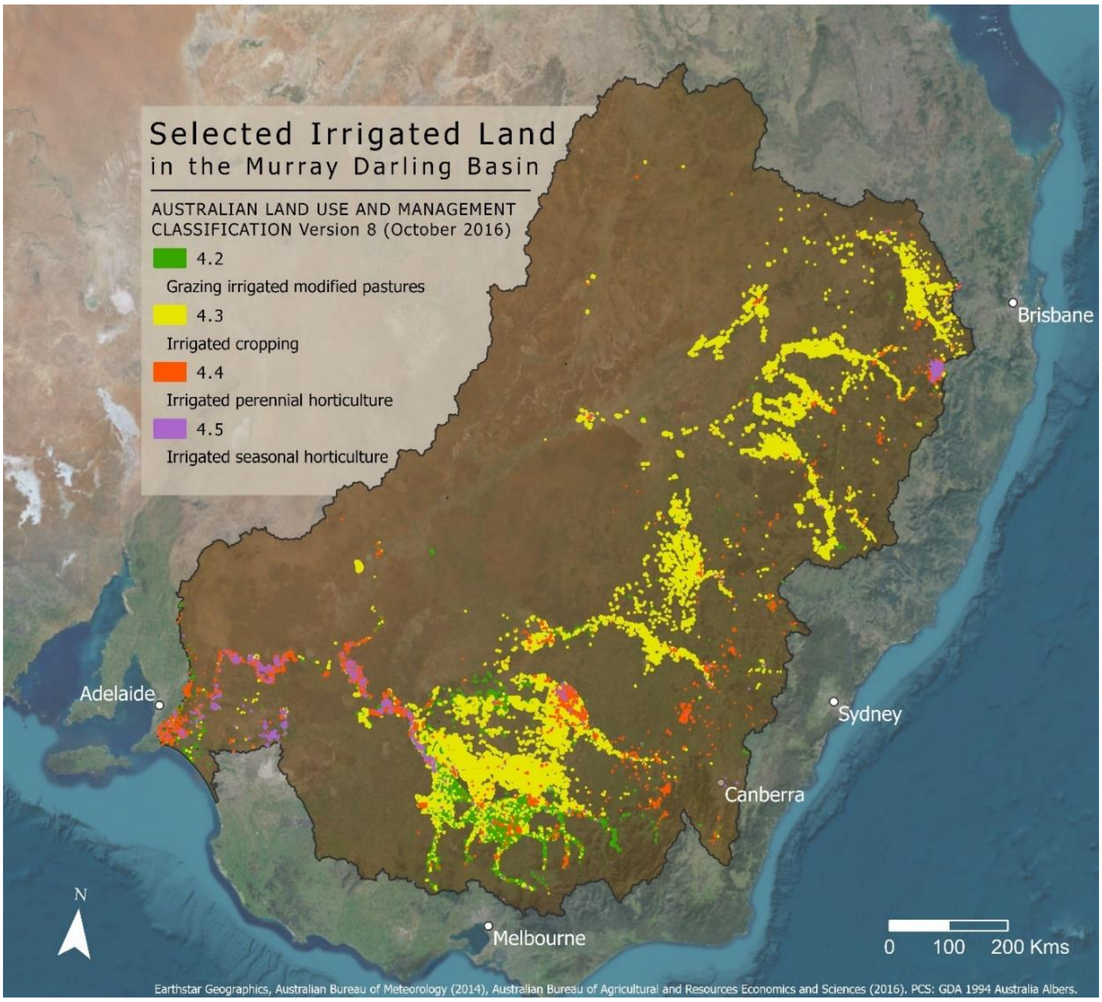

Figure 15: Irrigated land areas across the Basin.

Appendix 2 provides further details of the data sources from which spatio-temporal and spatial data on potential drivers of visitation rate were compiled.

6. Modelling method

6.1. Count data regression on spatio-temporal data

Monthly BuD or PuD counts for sub-catchment i in month t (over the month-index sequence t =1 (March 2013) to t =73 (March 2019)) are denoted BuD_{it} and PuD_{it} to indicate spatial-temporal variation. Either BuD_{it} or PuD_{it} can be used as the dependent for analysis. BuD counts provide a larger dataset, but even though similarities are evident between Figure 3 and Figure 4, visitation patterns may differ between BuDs and PuDs, so separate analyses are undertaken for each visitation proxy.

BuD or PuD counts can only be zero or positive whole numbers. Many sub-catchments within most River Regions report zero BuD or PuD counts for at least some months over the 73-month analysis timeframe. Some sub-catchments report zero BuD or PuD counts for all 73 months. These characteristics require that count data regression models are used to identify potential drivers of variation in visitation rate as proxied by BuD or PuD counts. The high proportion of zero counts further requires that count data models which can accommodate ‘zero inflation’ should be used when the number of zero counts in the observed data cannot be modelled adequately by standard count data models (Cameron and Trivedi, 2013a; Zuur et al., 2017; Zuur and Ieno, 2018).

The spatio-temporal nature of the BuD or PuD count data require that count data regression models must also be able to handle potential temporal and spatial dependency in sub-catchment count data within a River Region. Temporal dependencies arise when successive BuD or PuD counts from a given sub-catchment are more similar than counts from different sub-catchments. Spatial dependencies arise when BuD or PuD counts from neighbouring sub-catchments are more similar than counts from sub-catchments that are more distant. Methods for handling temporal and spatial dependencies in count data regression are described by Anselin (1988), Zuur et al. (2009), Cameron & Trivedi (2013b), LeSage & Pace (2014), Zuur et al.(2017), Zuur & Ieno (2018) and Jung & Glaser (2022).

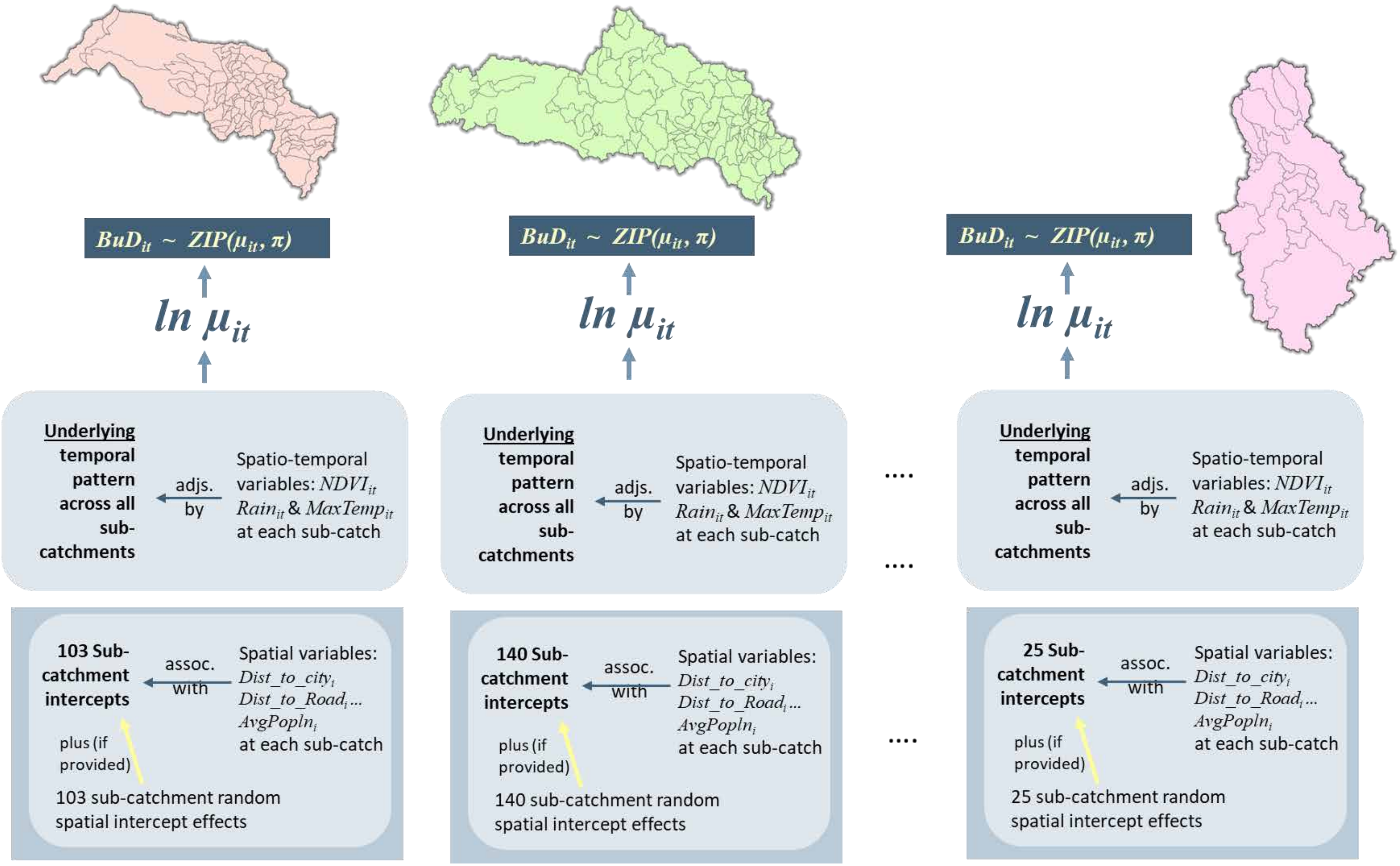

Figure 16: Modelling concept.

The starting point for count data regression modelling is a statistical distribution that expresses the probability of zero or positive integer outcomes (i.e., 'counts') arising. The simplest such distribution is the Poisson distribution in which the probability P of observing a BUD count of 0, 1, 2, 3 …etc., is defined to be:

$$P(BuD_{it} = y \text{ which can be } 0{,}1{,}2{,}3 \ldots . | \mu_{it}) = \frac{\mu_{it}^{y} \times e^{-\mu_{it}}}{y!}$$

Equation 1

where μ is the mean or 'expected value' of the Poisson distribution.

If BUD counts are distributed according to the Poisson distribution then:

$$BuD_{it} \sim Poisson(\mu_{it}) \Rightarrow E(BuD_{it}) = \mu_{it} \quad \text{and} \quad var(BuD_{it}) = \mu_{it}$$

with $E(\cdot)$ denoting the expected value of BUD count and $var(\cdot)$ denoting the variance of BUD counts around their expected value. The Poisson distribution in Equation 1 shows how the probability of observing a particular BuD count (0, 1, 2, 3, …..) in a particular sub-catchment in a particular month over the 73-month timeframe is influenced by the Poisson mean.

The Poisson distribution forms the basis for a Poisson count data model in which BuD_{it} counts are the dependent variable and driving factors enter as independent variables ('covariates') which can influence the estimated Poisson mean (μ_{it}) for each data point (BuD_{it}).

The regression equation for a Poisson count data regression with BuD_{it} as the dependent variable and with spatio-temporal (e.g., NDVi greenness), spatial (e.g., distance to the nearest local population centre) and temporal (e.g., month-of-the-year) covariates influencing the estimated Poisson mean is:

$$BuD_{it} \sim Poisson(\mu_{it})$$
$$E(BuD_{it}) = \mu_{it} \text{ and } var(BuD_{it}) = \mu_{it}$$

Equation 2

$$\begin{aligned} ln(\mu_{it}) = \beta_1 &+ \beta_2 NDVi_{it} + \beta_3 Rain_{it} + \beta_4 MTemp_{it} \\ &+ \beta_5 PC\_Dist_i + \beta_6 Dist\_to\_Road_i + \beta_7 DIWA\_Dist_i \\ &+ \beta_8 IBA\_Dist_i + \beta_9 CAPAD\_Dist_i \\ &+ \beta_{10} CC\_Area_i + \beta_{11} IrrigAreaPc_i + \beta_{12} NDViByIrrP_i + \beta_{13} AvgPopln_i \\ &+ \beta_{14f} Month\_factor_t + f(year_t) \text{ for years 2013 to 2019} \end{aligned}$$

or

$$+ f(month\_index_t) \quad \text{for month-index 1 to 73}$$

Covariates such as month-of-the-year, year and month-index (1, 2 … 73) which influence BuD counts across all sub-catchments in the River Region simultaneously can be introduced as either categorical factors (e.g. $Month\_factor_t, Year\_factor_y$), as a linear trend (e.g., across years), or as non-linear, non-parametric 'smoothers' across year (e.g., $f(year_t)$)) or across month-index (e.g., $f(month\_index_t)$)) (Zuur et al., 2017). Non-linear per year or per month index smoothers provide considerable additional flexibility to accommodate non-linear, common across sub-catchment, time variations in the intensity of visitation.

The BuD and PuD data show strong seasonal patterns of visitation, likely influenced by features such as popular holiday periods and the arrival of migratory bird species. The overall level of BuD or PuD counts also vary between years. This is likely due, at least in part, to varying utilisation of the eBird and Flickr web posting sites through time. The month, year or month-index temporal terms in the count data regression models provide the ability to account for the *average* between-month and inter-annual variation present in visitation counts *across* sub-catchments within a River Region.

The NDVI greenness and percentage inundation area environmental condition covariates also contain strong cyclic variation. Consequently, if the month, year or month-index temporal terms in the regression models provide a poor representation of the *average* seasonal and inter-year variation in BuD or PuD counts the model may use NDVI greenness or percentage inundation area *instead* of the month, year or month-index terms to represent the *average* temporal variations in visitation count *across* sub-catchments. This could lead to a false conclusion that changes in environmental condition are strongly associated with variations in visitation rate when, in fact, cyclic variation in the environmental condition metric is being used to explain some of the seasonal variation in visitation rate due to holiday periods or the arrival of migratory bird species.

The performance of the month, year or month-index temporal terms in the regression models is therefore examined closely during model fitting to ensure they provide an appropriate representation of *average across*-sub-catchment variation in visitation rate. This then leaves a role for NDVi greenness or percentage inundation area, as environmental condition metrics that vary through time *and* across space, to potentially help explain *between* sub-catchment differences in visitation rate.

When Bayesian methods are used to fit a Poisson count data regression model such as Equation 2 to observed BuD_{it} count data for sub-catchments in a River Region, separate ('posterior marginal') distributions are estimated for each parameter (β_1 to β_{13}) associated with a fixed covariate and any temporal terms (β_{14f}) and/or $f(year_t)$ or $f(month\_index_t)$ non-parametric smoothers, as appropriate. This allows important covariate drivers of spatio-temporal variation in BuD_{it} count to be identified.

In the Poisson probability density function the variance of counts around their expected value increases linearly as expected value increases. This makes it difficult for Poisson-based regression models to adequately fit data which include small numbers of unexpectedly large counts. A common approach to resolve this issue is to switch from Poisson-based regression models to regression models that are based on the negative binomial distribution.

The negative binomial distribution allows the variance of counts to rise more rapidly than the mean, with the rate of variance expansion controlled by a size parameter k . The probability of observing a BUD count of 0, 1, 2, 3 …etc., under the negative binomial distribution is defined to be:

$$P(BuD_{it} = y \text{ which can be } 0,1,2,3 \ldots . | k, \mu_{it}) = \frac{\Gamma(y+k)}{\Gamma(k) \times \Gamma(y+1)} \times \left(\frac{k}{\mu_{it}+k}\right)^k \times \left(1 - \frac{k}{\mu_{it}+k}\right)^y$$

Equation 3

where μ_{it} is the mean, k is the size parameter and $\Gamma(\cdot)$ denotes a gamma function.

If BUD count is distributed following a negative binomial distribution then:

$$BuD_{it} \sim NegBin(k, \mu_{it}) \Rightarrow E(BuD_{it}) = \mu_{it} \quad \text{and} \quad var(BuD_{it}) = \mu_{it} + \frac{\mu_{it}^2}{k}$$

The negative binomial distribution forms the basis for a negative binomial count data model. Equation 4 shows the form of a negative binomial count data regression model with BuD_{it} as the dependent variable.

$$BuD_{it} \sim NegBin(k, \mu_{it})$$

$$E(BuD_{it}) = \mu_{it} \text{ and } var(BuD_{it}) = \mu_{it} + \frac{\mu_{it}^2}{k}$$

Equation 4

$$\begin{aligned} ln(\mu_{it}) = \beta_1 &+ \beta_2 NDVi_{it} + \beta_3 Rain_{it} + \beta_4 MTemp_{it} \\ &+\beta_5 PC\_Dist_i + \beta_6 Dist\_to\_Road_i + \beta_7 DIWA\_Dist_i \\ &+\beta_8 IBA\_Dist_i + \beta_9 CAPAD\_Dist_i \\ &+\beta_{10} CC\_Area_i + \beta_{11} IrrigAreaPc_i + \beta_{12} NDViByIrrP_i + \beta_{13} AvgPopln_i \\ &+\beta_{14f} Month\_factor_t \text{ or } f(year_t) \text{ or } f(month\_index_t) \end{aligned}$$

$$k \sim log\ Gamma\left(\bar{k}, \sigma_k^2\right)$$

Bayesian estimation of a negative binomial regression model on BuD_{it} count data produces separate (posterior marginal) distributions for each of the regression parameters β_1 to β_{13}, relevant temporal terms and the mean and variance of the distribution of the size parameter k.

A high proportion of zero counts, beyond that expected under a standard Poisson distribution or a negative binomial distribution, can be accommodated by zero-inflation (Zuur and Ieno, 2018). The probability density function for a zero-inflated Poisson (ZIP) distribution and the regression equation for a zero-inflated Poisson regression model (ZIP model) are shown in Equation 5 and Equation 6 respectively:

$$P(BuD_{it} = y \text{ which can be } 0,1,2,3 \ldots . | \pi, \mu_{it})$$

$$= \begin{cases} \pi + (1-\pi) \times e^{-\mu_{it}} & \text{for } BuD_{it} = 0 \\ (1-\pi) \times \dfrac{\mu_{it}^y \times e^{-\mu_{it}}}{y!} & \text{for } BuD_{it} > 0 \end{cases}$$

Equation 5

$$BuD_{it} \sim ZIP(\mu_{it}, \pi) \Rightarrow E(BuD_{it}) = (1-\pi)\mu_{it} \quad \text{and} \quad var(BuD_{it}) = (1-\pi)\left(\mu_{it} + \pi\mu_{it}^2\right)$$

$$BuD_{it} \sim ZIP(\mu_{it}, \pi)$$

$$E(BuD_{it}) = (1-\pi)\mu_{it} \quad \text{and} \quad var(BuD_{it}) = (1-\pi)\left(\mu_{it} + \pi\mu_{it}^2\right)$$

Equation 6

$$\begin{aligned} ln(\mu_{it}) = \beta_1 &+ \beta_2 NDVi_{it} + \beta_3 Rain_{it} + \beta_4 MTemp_{it} \\ &+\beta_5 PC\_Dist_i + \beta_6 Dist\_to\_Road_i + \beta_7 DIWA\_Dist_i \\ &+\beta_8 IBA\_Dist_i + \beta_9 CAPAD\_Dist_i \\ &+\beta_{10} CC\_Area_i + \beta_{11} IrrigAreaPc_i + \beta_{12} NDViByIrrP_i + \beta_{13} AvgPopln_i \\ &+\beta_{14f} Month\_factor_t \text{ or } f(year_t) \text{ or } f(month\_index_t) \end{aligned}$$

$$logit(\pi) = \gamma$$

The ZIP model produces a separate estimate for the probability of observing a zero count (π) which moderates the standard probabilities derived from the Poisson mean (μ_{it}) to better accommodate the 'inflated' proportion of zero counts. The ZIP probability density function (Equation 5) shows that because the π term *inflates* the probability of a zero count above that determined by the standard Poisson mean, the probability of observing non-zero counts must be *deflated* accordingly to ensure that the total probability across zero and all non-zero counts still sums to one.

The probability of observing a zero count can be estimated in the regression model as a simple intercept (π) which applies to all sub-catchments across all month-indices, or as being dependent on temporal (π_t), spatial (π_i) or spatio-temporal (π_{it}) drivers[1]. Zero-inflated negative binomial models (ZINB models) can be used when Poisson or ZIP models are unable to adequately fit BuD_{it} or PuD_{it} data which contain modest numbers of high counts as well as an inflated proportion of zero counts (Zuur and Ieno, 2018).

Individual sub-catchments in a River Region are the cross-sectional individuals in BuD_{it} or PuD_{it} count data regressions. Sub-catchment-specific adjustments to the estimated Poisson, negative binomial or ZIP mean – and thus to the predicted probability of observing a given count – can be introduced as random terms, drawn from a zero-centred distribution with a variance determined through the model fitting. Sub-catchment-specific random terms α_i are introduced into the ZIP model in Equation 7:

$$BuD_{it} \sim ZIP(\mu_{it}, \pi, \sigma_{ID}^2)$$
$$E(BuD_{it}) = (1-\pi)\mu_{it} \quad \text{and} \quad var(BuD_{it}) = (1-\pi)\left(\mu_{it} + \pi\mu_{it}^2\right)$$

Equation 7

$$\begin{aligned} ln(\mu_{it}) = \beta_1 &+ \beta_2 NDVi_{it} + \beta_3 Rain_{it} + \beta_4 MTemp_{it} \\ &+\beta_5 PC\_Dist_i + \beta_6 Dist\_to\_Road_i + \beta_7 DIWA\_Dist_i \\ &+\beta_8 IBA\_Dist_i + \beta_9 CAPAD\_Dist_i \\ &+\beta_{10} CC\_Area_i + \beta_{11} IrrigAreaPc_i + \beta_{12} NDViByIrrP_i + \beta_{13} AvgPopln_i \\ &+\beta_{14f} Month\_factor_t \ \text{ or } \ f(year_t) \ \text{ or } \ f(month\_index_t) + \alpha_i \end{aligned}$$

$$logit(\pi) = \gamma \qquad \text{and} \qquad \alpha_i \sim N(0, \sigma_{ID}^2)$$

where the sub-catchment-specific random terms α_i follow a normal distribution with mean zero and variance σ_{ID}^2. The ID denotes that this random term applies across all sub-catchments in the River Region i = 1 …..n, which are indexed by ID number. Sub-catchment-specific random terms provide considerable additional flexibility to accommodate between-sub-catchment variation in the temporal average of observed BuD_{it} or PuD_{it} counts.

To ensure that inference regarding the statistical significance (in a frequentist modelling setting) or 'significant importance' (in a Bayesian modelling setting) of potential drivers is robust and reliable, the ultimate goal of increasing model sophistication is to ensure that any remaining idiosyncratic disturbances ('errors') around predictions of BuD_{it} or PuD_{it} counts are independently and identically distributed (i.i.d.). Plots of BuD count per sub-catchment, per month for some River Regions show distinct clumping (see e.g., Figure 17). If the ability of an estimated model to predict

[1] In the R-INLA software that is used to fit count data regression models for River Regions (see Section 6.3) it is currently only possible to estimate the zero count probability as a simple intercept term (π).

zero counts is not as good as its ability to predict higher counts this will lead to spatial clustering of prediction errors if there are spatial clusters of zero counts in sections of the River Region. This departure from independently and identically distributed errors could potentially lead to incorrect identification of 'important' or 'not important' driving factors of visitation.

The problem of spatially clumped errors ('positive spatial auto-correlation of errors') can be addressed by including sub-catchment-specific spatial weighting terms in the model. An approach for handling this problem by introducing catchment-specific conditional (spatial) auto-regressive (CAR) terms into the model is explained by Cressie & Kapat (2008) and Song and De Oliveira (2012) and exemplified by Zuur & Ieno (2018, chap. 24). A ZIP count data model with sub-catchment-specific CAR spatial terms in shown in Equation 8.

$$BuD_{it} \sim ZIP(\mu_{it}, \pi, \sigma^2_{CAR})$$

$$E(BuD_{it}) = (1-\pi)\mu_{it} \quad \text{and} \quad var(BuD_{it}) = (1-\pi)\left(\mu_{it} + \pi\mu^2_{it}\right)$$

Equation 8

$$\begin{aligned} ln(\mu_{it}) = \beta_1 &+ \beta_2 NDVi_{it} + \beta_3 Rain_{it} + \beta_4 MTemp_{it} \\ &+\beta_5 PC\_Dist_i + \beta_6 Dist\_to\_Road_i + \beta_7 DIWA\_Dist_i \\ &+\beta_8 IBA\_Dist_i + \beta_9 CAPAD\_Dist_i \\ &+\beta_{10} CC\_Area_i + \beta_{11} IrrigAreaPc_i + \beta_{12} NDViByIrrP_i + \beta_{13} AvgPopln_i \\ &+\beta_{14f} Month\_factor_t \ \text{ or } \ f(year_t) \ \text{ or } \ f(month\_index_t) + \alpha_i \end{aligned}$$

$$logit(\pi) = \gamma \qquad \text{and} \qquad \alpha_i \sim N(0, \sigma^2_{CAR})$$

knowing neighbour weightings w_{ij} between sub-catchment i and all other sub-catchments $j \neq i$

Neighbourhood weightings w_{ij} determine the extent to which the sub-catchment-specific random effect for sub-catchment i is influenced by the random effects from other sub-catchments $j \neq i$. The more direct the spatial adjacency, the stronger the influence. In this study neighbourhood weighting w_{ij} is set to 1 for sub-catchments $j \neq i$ that share a border with focus sub-catchment i . Neighbourhood weightings w_{ij} are set to 0 for all other sub-catchments.

6.2. Assessing model suitability and model fit

Following Zuur et al. (2017) and Zuur & Ieno (2018), candidate function forms and potential covariates for use in count data regression models to predict BuD_{it} or PuD_{it} for sub-catchments within a River Region are assessed on the following attributes:

1. Ability to accommodate the proportion of zero counts in the data (see Subsection 6.2.2).
2. Ability of the temporal-smoothing approach to accommodate any temporal correlation in the data by adequately representing average across-sub-catchment seasonal variation in BuD or PuD counts (see Subsection 6.2.3).
3. Ability of sub-catchment-specific random effects to accommodate spatial autocorrelation in the data (see Subsection 6.2.4).
4. After identifying a model form that performs satisfactorily in assessments 1, 2 and 3, select covariates for inclusion in the final model based on model fitting parsimony (see Subsection 6.2.5).

Appropriate functional form and temporal and/or spatial correlation of regression errors (Steps 1 to 3) need to be assessed because departure from independently and identically distributed (i.i.d.) regression errors can lead to unreliable conclusions regarding the importance (or otherwise) of covariates that potentially drive sub-catchment BuD_{it} or PuD_{it} counts within a River Region. Assessment steps 1 to 4 are applied as the sophistication of count data models increases in the sequence shown in Table 3, from lowest (M1) to highest (M8) complexity. Model assessment need not proceed all the way to M8 if a less complex model performs adequately for assessment steps 1 to 3.

Prior to proceeding through model assessment steps 1 to 4, covariate data are screened for outliers, collinearity and correlations (see Subsection 6.2.1). Steps 1 to 3 are then assessed for models with the full suite of available covariates.

Table 3: Model assessment and fitting sequence for count data regression models to describe BuD or PuD counts for sub-catchments within a River Region.

Model	Model description (REs denotes random effects)
1	Non-spatial Poisson: no REs, month by factor (f_month), year via smoother *f(year)* [7 knot]
2	Non-spatial ZIP: no REs, month by factor (f_month), year via smoother *f(year)* [7 knot]
3	Non-spatial ZINB: no REs, month by factor (f_month), year via smoother *f(year)* [7 knot]
4	Non-spatial ZIP: sub-catch REs, month by factor (f_month), year via smoother *f(year)* [7 knot]
5	Non-spatial ZIP: sub-catch REs, month-index via smoother *f(m_index)* [18 knot]
6	Non-spatial ZINB*:* sub-catch REs*,* month-index via smoother *f(m_index)* [18 knot]
7	Spatial ZIP: spatial CAR sub-catch REs, smoother *f(m_index)* [18 knot]
8	Spatial ZINB: spatial CAR sub-catch REs, smoother *f(m_index)* [18 knot]

6.2.1. Data screening

Within each River Region the dependent variables (BuDs and PuDs) and independent variables (covariates) were examined for outliers using a Cleveland dotplot (Cleveland, 1993; Zuur et al., 2017). Independent variables were tested for collinearity via variance inflation factors (VIFs), using a rejection threshold of ≥ 3 (after Zuur et al. (2017, sec. 9.5)). If VIFs suggested collinearity, Pearson's correlation factors were used to identify which variables were likely to be causing the problem. Collinear variables were dropped until VIFs for all remaining variables were less than 3. Correlations beyond ±0.5 were noted for further evaluation during regression fit testing.

6.2.2. Ability to handle zero inflation

Zuur et al. (2017, sec. 8.6) and Zuur & Ieno (2018) explain how the parameter distributions from fitted Bayesian models can be used to conduct the simulations listed in Table 4 to assess whether the functional forms of the sequence of models in Table 3 are appropriate to accommodate the level of zero inflation present in BuD or PuD count data from sub-catchments within a River Region.

Table 4: Simulations used to assess the ability of a model to accommodate zero inflation in BuD or PuD counts for sub-catchments in a River Region.

Simulation	Reference & Application
Number of zero counts produced by fitted model (Poisson, NB, ZIP, ZINB)	Zuur and Ieno (2018, chap. 18)
Dispersion statistic produced by fitted model (Poisson, NB, ZIP, ZINB)	Zuur and Ieno (2018, chap. 18)
Distribution of counts produced by fitted model (Poisson, NB, ZIP, ZINB)	Zuur and Ieno (2018, chap. 18)

Simulations are conducted as follows. Bayesian estimation is used to fit the chosen model (one of M1 to M8) to BuD_{it} or PuD_{it} data from sub-catchments in a River Region over the 73-month timeframe. This produces estimated posterior marginal distributions for regression parameters for each spatio-temporal and spatial covariate and relevant temporal term(s). If the model includes terms such as the k size parameter in a negative binomial model, the zero-inflation parameter π in a ZIP model, or sub-catchment-specific random effects, the variance of these distributions will also be estimated.

Simulation draws 5000 sets of samples from the fitted distributions for all model parameters. Each set of parameter samples is then applied to sub-catchment-specific, month-specific covariate data from the River Region to calculate μ_{it} (and if necessary k, π, α_i) for each sub-catchment in each month. The calculated μ_{it} (and if necessary k, π, α_i) are used to generate random draws of BuD or PuD counts from the relevant form of distribution (Poisson, negative binomial, ZIP, or ZINB) to produce simulated count data for each sub-catchment in each month.

The number of counts of 0, 1, 2, 3 etc. BuDs or PuDs across all sub-catchments can then be summed within each 73-month simulation. A histogram of the total number of zero counts across all 5000 simulations can then be compared with the total number of zero counts in the BuD or PuD data from the River Region. Similarly, a plot of the average number of 0, 1, 2, 3 etc. BuD or PuD counts across all 5000 simulations can be compared with the number 0, 1, 2, 3 etc. BuD or PuD counts in the data. These plots provide a visual indication of whether the functional form of the regression model is appropriate for accommodating the level of zero inflation present in the River Region's BuD or PuD count data (see Figure 27).

6.2.3. Ability to handle temporal correlation

The temporal autocorrelation function (ACF) provides a formal approach for detecting temporal correlation among regression residuals in fitted models (Kleiber and Zeileis, 2008; Zuur et al., 2009, Section 6.1). Significant temporal autocorrelation in regression residuals across a substantial proportion of sub-catchments in a River Region indicates that more sophisticated temporal modelling is required to produce i.i.d. regression errors and provide reliable inference regarding covariate drivers of BuD or PuD counts. The sophistication of temporal modelling increases through the modelling sequence in Table 3 from models M1 to M4 (which use month as a categorical factor plus a 7-knot non-linear smoother on year) to models M5 to M8 (which use an 18-knot non-linear smoother on month-index (1 to 73)).

Temporal autocorrelation is detected if the magnitude of the autocorrelation function of a lagged residual for a sub-catchment exceeds a critical value of $1.96/\sqrt{n}$ where n is the number of data points in the modelled time series (here n = 73). The ACF test was implemented using the `acf()`

function in R on the Pearson residuals for predicted BuD_{it} or PuD_{it} counts across the 73-month-index modelling timeframe for each sub-catchment in a River Region separately.

Pearson residuals for each data point are defined as the observed data minus the predicted expected value divided by the square root of the variance of the predicted values (Cameron and Trivedi, 2005). Thus, for a ZIP model, the Pearson residuals r_{it}^{P} are:

$$r_{it}^{P} = \frac{BuD_{it} - (1-\pi)\mu_{it}}{\sqrt{(1-\pi)\left(\mu_{it} + \pi\mu_{it}^{2}\right)}}$$

where the expected value and variance of expected values predicted by the ZIP model are those shown in Equation 6. The proportion of sub-catchments in the River Region for which the acf test indicates that temporal autocorrelation is present is reported.

6.2.4. Ability to handle spatial correlation

The explanation in this section draws on Moraga (2023, sec. 8.1).

Moran's global I index (Moran, 1950) is commonly used to produce a test statistic to assess the significance of spatial autocorrelation (i.e. clumping) of metrics among polygons (here sub-catchments) in areal data (Moraga, 2023). Moran's I index is used here to assess spatial autocorrelation (i.e., lack of spatial independence) among Pearson's residuals from fitted regression models in the M1 to M8 sequence of Table 3. Lack of spatial independence among regression residuals indicates that more sophisticated spatial modelling is required to produce i.i.d. residuals and reliable inference.

The global Moran's I statistic is given by:

$$I = \frac{n\sum_i\sum_j w_{ij}(Y_i - \bar{Y})\left(Y_j - \bar{Y}\right)}{\left(\sum_{i\neq j} w_{ij}\right)\sum_i (Y_i - \bar{Y})^2}$$

where n is the number of sub-catchments in the River Region, Y_i is the observed value of interest (here the month-index-specific Pearson residual from the regression model) in sub-catchment i, and $\bar{Y}$ is the mean of Pearson's regression residuals in month-index t across sub-catchments in the River Region. The w_{ij} spatial weights report the importance of spatial adjacency between sub-catchments i and j as the inverse of the number of spatial neighbours of sub-catchment i . Sub-catchments $j \neq i$ are defined to be spatial neighbours if they share a common border, following the 'queen' definition. Spatial weights w_{ij} for sub-catchments that are not neighbours of sub-catchment i are set to zero.

Under a null hypothesis of no spatial autocorrelation, observations Y_i are distributed i.i.d. and the Moran's I statistic is distributed asymptotically normal with a mean of:

$$E[I] = \frac{-1}{n-1}$$

and a variance calculated from the number of sub-catchments in the River Region and combinations of the spatial weights (see Moraga (2023, Section 8.1)).

Moran's I values generally range between -1 and 1. Values significantly higher than $E[I] = -1/(n-1)$ indicate positive spatial autocorrelation i.e., significant clustering of regression residuals among neighbouring sub-catchments. Values significantly lower than $E[I] = -1/(n-1)$ indicate negative spatial autocorrelation i.e., significant dispersion of regression residuals among neighbouring sub-catchments. The asymptotic normal distribution enables a statistical test to be defined via a z-score:

$$z = \frac{I - E(I)}{\sqrt{Var(I)}}$$

The Moran's I-derived z-test was used to assess spatial autocorrelation across sub-catchments among the Pearson residuals from fitted models of monthly BuD or PuD counts within a River Region. The test hypotheses were:

$H_0: I = E[I]$ (no spatial autocorrelation)

$H_{1a}: I > E[I]$ (positive spatial autocorrelation)

$H_{1b}: I < E[I]$ (negative spatial autocorrelation)

The null hypothesis of no spatial autocorrelation was rejected if the p-value of the z score for the month-index concerned was less than $\alpha = 0.05$. The test was run using the `moran.test()` function from the **spdep** package in R on the Pearson residuals for predicted BuD_{it} or PuD_{it} counts for all sub-catchments in a River Region with non-zero BuD or PuD in each of the 73 months in the model timeframe. The proportion of $\alpha < 0.05$ test results (one from each month-index in the modelling timeframe) is reported.

The Moran's I test is evaluated only across those sub-catchments with non-zero BuD or PuD counts in the month-index concerned because, by construction, the expected values from fitted count data models can never be less than zero (see model equations in Section 6.1); hence the Pearson residuals from sub-catchments with zero BuD or PuD counts will always be negative. This will trigger rejection of the null hypothesis of no spatial autocorrelation from well-fitting models if zero counts in the data are clumped across neighbouring sub-catchments within the River Region. Clumping of zero BuD or PuD counts is common in the data from River Regions across the Murray Darling Basin (for example, see Figure 17).

6.2.5. Model fit and covariate selection

Once the appropriate function form of model has been established through the assessments described in the preceding three subsections, the model fit metrics in Table 5 are used to inform which covariates should be included. Model fit metrics for use when models are estimated via INLA methods (see Section 6.3) are described in Gómez-Rubio (2020, Section 2.4). Zuur et al. (2017, sec. 8.6) explain how these metrics can be applied.

Table 5: Metrics used to inform covariate inclusion once the appropriate functional form of count data regression model has been established via assessments 1, 2 and 3.

Metric	Reference	Application
Model maximum likelihood (i.e., probability of observed data under a given fitted model)	Chib (1995) Chib & Jeliazkov (2001)	Gómez-Rubio (2020 Section, 2.4.1)
Deviance information criterion (DIC) (a complexity-penalised goodness of fit metric)	Spiegelhalter et al. (2002)	Zuur et al. (2017, sec. 8.6.2)
Watanabe-Akaike information criterion (WAIC) (another complexity-penalised goodness of fit metric)	Watanabe (2013) Gelman et al. (2014)	Zuur et al. (2017, sec. 8.6.2)
Conditional predictive ordinates (CPO) (a 'leave one out' cross-validation), summarised via the sum of the natural logs of CPO values across all data points: $-\sum_{i=1}^{n} ln(CPO_i)$. CPO values for individual data points can be plotted across the data sequence to provide a visual indication of model fit.	Pettit (1990)	Zuur et al. (2017, sec. 8.6.3) Gómez-Rubio (2020 Section, 2.4.2)

6.3. Model estimation

The count data regression models used to estimate monthly BuD or PuD counts per sub-catchment within River Regions across the Basin run on large datasets (approaching 39,000 data points for the River Region with the highest number of sub-catchments) and must be able to handle potential temporal correlation among counts from a given sub-catchment across months, potential spatial autocorrelation between counts from neighbouring sub-catchments in the same month, and very high levels of zero inflation. Rapid model estimation is necessary to make model estimation and refinement tractable. For all these reasons Bayesian model estimation via integrated nested Laplace approximation (INLA) was used (Lindgren et al., 2011; Rue et al., 2017, 2009). Models were estimated using the R-INLA package available from www.r-inla.org. Default R-INLA diffuse priors for covariates and default R-INLA priors for hyperparameters were used for all model estimations and models converged rapidly without obvious problems.

Explanatory examples from Gómez-Rubio (2020), Moraga (2023, 2019), Zuur et al. (2017) and Zuur and Ieno (2018) were extremely useful.

7. Results

This section reports results from fitting count data models to BuD_{it} or PuD_{it} data for River Regions across the MDB (Figure 1), using the model fitting methodologies described in Section 6. Models for BuD_{it} and PuD_{it} are fitted separately, with either $NDVI_{it}$ or $WatPc95_{it}$ proxying (loosely) 'environmental condition' and 'riparian condition', respectively (Table 2, Figure 7 and Figure 8). After data screening (Subsection 6.2.1) other spatio-temporal and spatial covariates (Table 2), temporal drivers (either month as a factor + year as a 7-knot smoother, or month-index as an 18-knot smoother over the full 73 month-index time frame), and sub-catchment-specific non-spatial or spatial random effects (Equation 7 and Equation 8) are included in models of increasing sophistication (Table 3).

Results are presented in full for the Lachlan River Region in the following subsections. Abbreviated results are then presented for other River Regions in subsections thereafter.

7.1. Lachlan catchment: detailed results

The Lachlan River is a major tributary of the Murrumbidgee and thence the Murray. The Lachlan River Region comprises 140 sub-catchments which vary considerably in area from the headwaters to the lower floodplain (Figure 1). Sub-catchments in the upper reaches of the Lachlan are relatively accessible from Canberra and Sydney, but sub-catchments on the lower floodplain are more remote, but still relatively accessible from regional population centres (Figure 11).

The following subsections report detailed results from fitting the M1:M8 sequence of count data models to Lachlan River Region data. Results are reported for:

- BuDs as the dependent variable, with NDVi as an environmental condition proxy
- BuDs as the dependent variable, with percentage inundation area (WatPc95) as an environmental condition proxy
- PuDs as the dependent variable, with NDVi as an environmental condition proxy
- PuDs as the dependent variable, with percentage inundation area (WatPc95) as an environmental condition proxy

All models include other spatio-temporal and spatial covariates, and temporal terms with either month as a categorical variable and year as a non-parametric temporal smoother, or with month-index (1 : 73) as a non-parametric temporal smoother.

7.1.1. BuDs and PuDs as dependent variables

Total BuD and PuD counts per Lachlan sub-catchment for the years 2013 to 2018 are shown in Figure 17 and Figure 18 . (There are only 10 months of data in 2013 and 12 months of data for 2014 to 2018[2]). BuD counts per sub-catchment across months of 2017 as an indicative year are shown in Figure 19. PuD counts per sub-catchment across months of 2014 as an indicative year are shown in Figure 20.

[2] Total annual BUD and PUD counts from 2019 are not shown because only 3 months of data are available for that year.

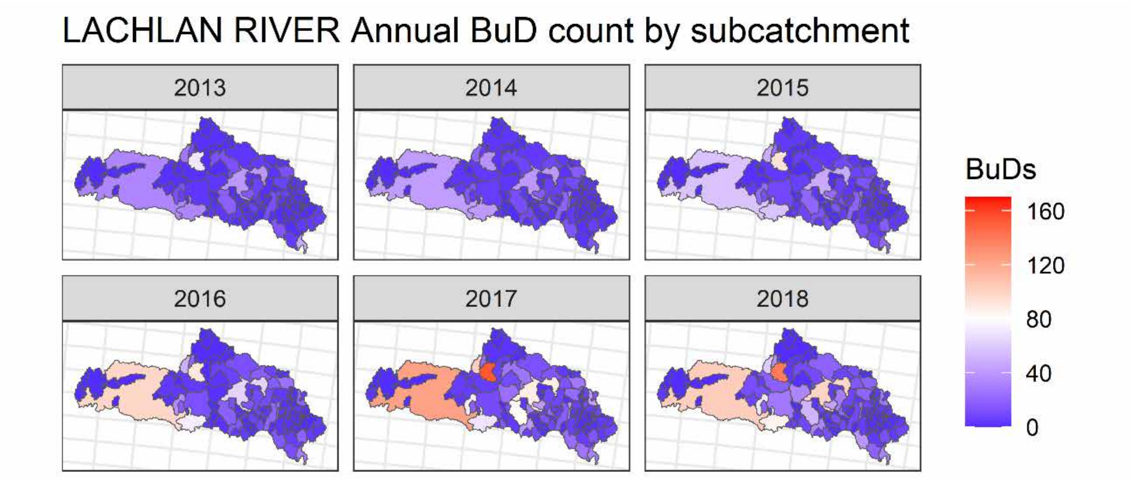

Figure 17: Lachlan River Region total annual BuD counts per sub-catchment for the years 2013 – 2018. (Only 10 months of data (March to December) shown for 2013).

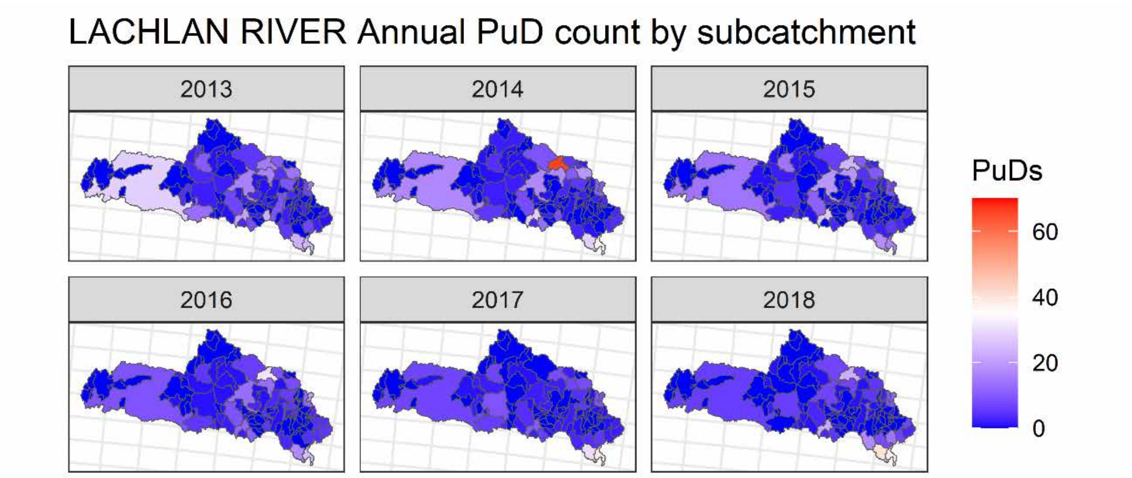

Figure 18: Lachlan River Region total annual BuD counts per sub-catchment for the years 2013 – 2019. (Only 10 months of data (March to December) shown for 2013).

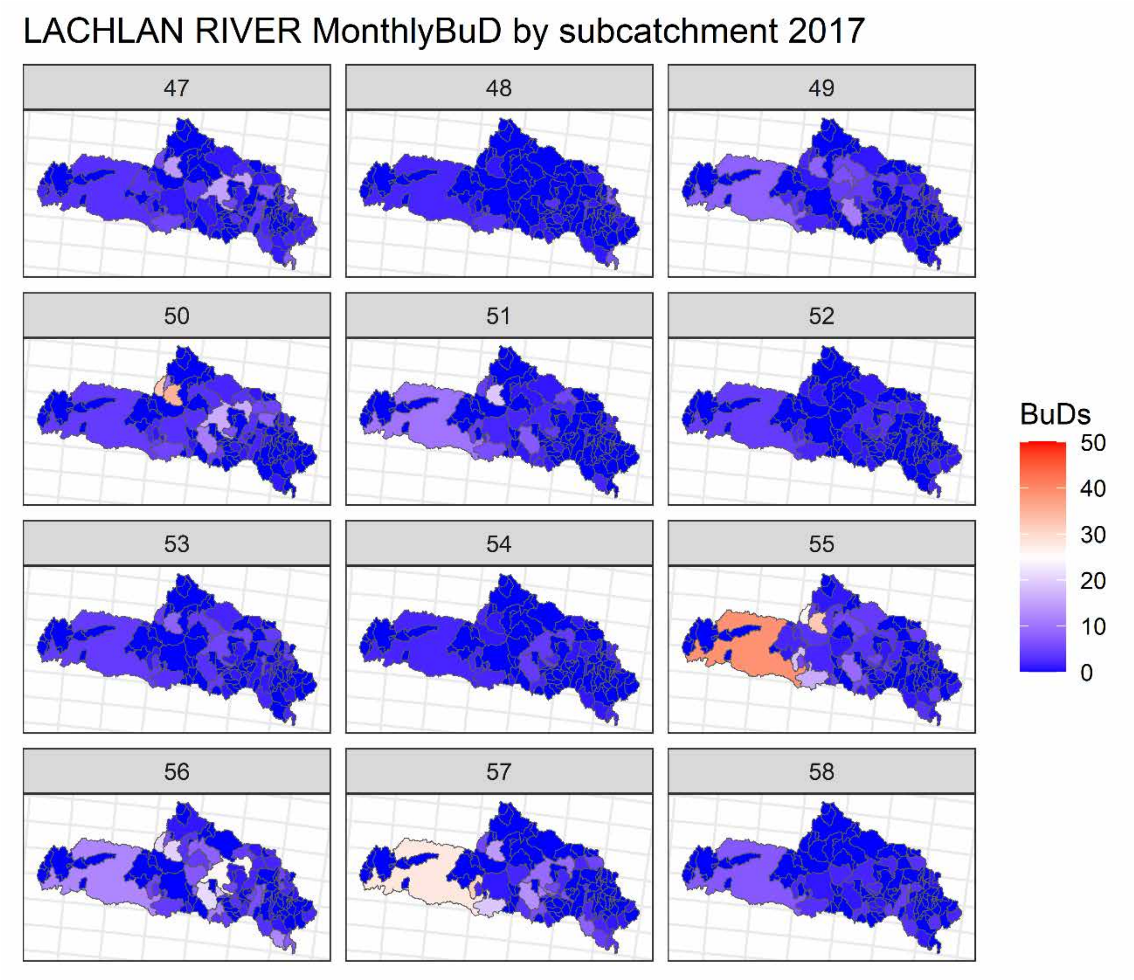

Figure 19: Lachlan River Region monthly BuD counts January = month-index 47 to December = month-index 58 for the year 2017.

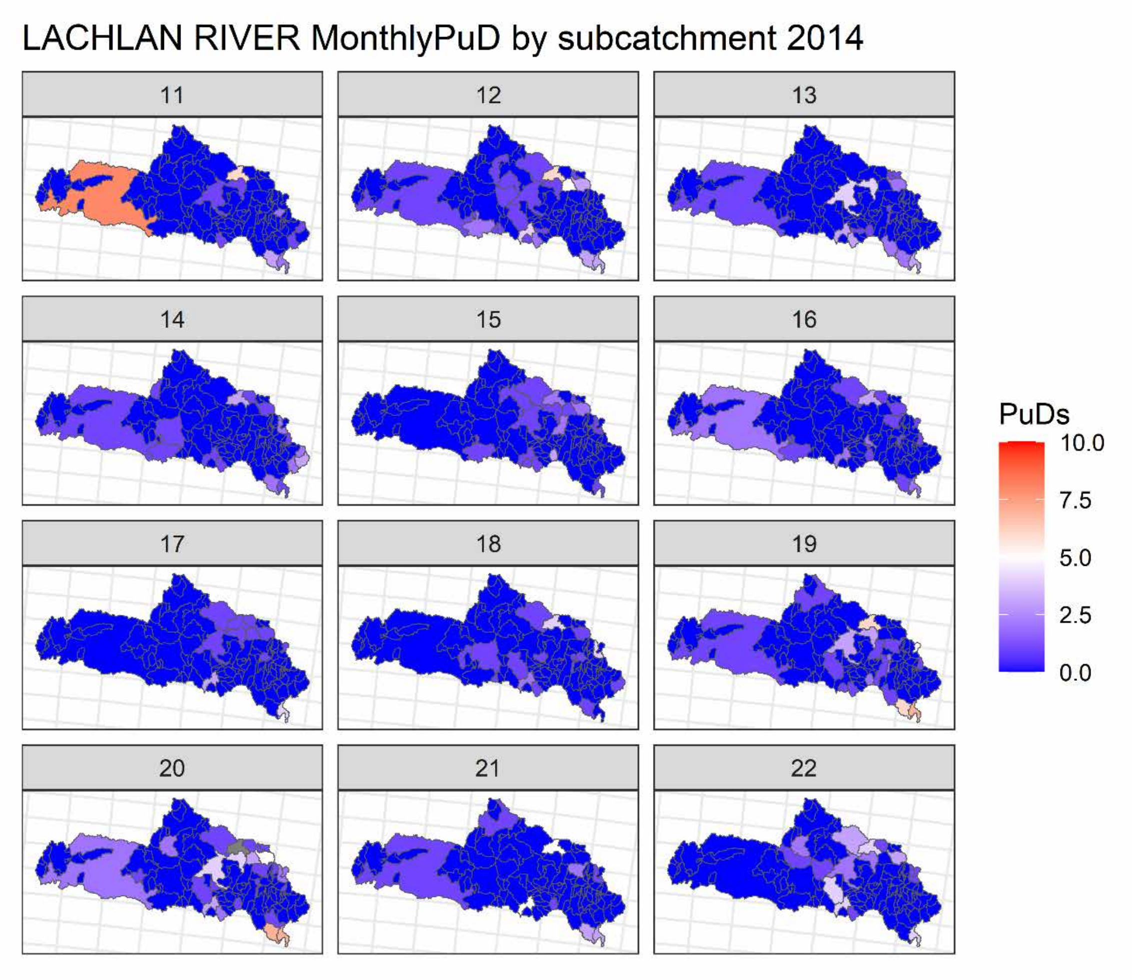

Figure 20: Lachlan River Region monthly PuD counts January = month-index 11 to December = month-index 22 for the year 2014.

There are very high proportions of zero BuD and PuD counts for sub-catchments across the Lachlan River Region between March 2013 and March 2019. The full 10220 data point (= 140 sub-catchments x 73 months) BUD dataset contains 81% zero counts and the full PuD dataset contains 89% zero counts. When data points for which percentage inundation area is missing are removed, dataset size reduces to 7227 observations in total, of which 81% of BuD counts and 91% of PuD counts are zero. This suggests that zero-inflated form regression models will likely be required.

There is considerable spatial variation in BuD and PuD counts across the Lachlan River Region and temporal variation across months of the year and between years. BuD counts are considerably higher than PuD counts. Total annual BuD counts are generally increasing over 2013-2018, whereas total annual PuD counts are generally declining over the same period. Individual sub-catchments in the mid and lower sections of the River Region are spatial hotspots of BuD counts, whereas individual sub-catchments in the mid and upper sections of the River Region are spatial hotspots of PuD counts. There are sizeable clumps of sub-catchments with zero or very low BuD and PuD counts across the River Region. In contrast, sub-catchments that are hotspots of BuD or PuD visitation tend to be somewhat isolated. It will be challenging for the count data regression models to replicate this spatio-temporal variation variation accurately.

7.1.2. Data screening

Data screening was conducted on the dependent and independent variables. Cleveland dotplots were used to identify outliers and provide an initial visual assessment of correlation (Figure 21, Figure 22 and Figure 23). VIFs were used to test for collinearity (Table 6) in combination with cross-correlation plots (Figure 24, Figure 25 and Figure 26). No significant collinearity was present among the spatio-temporal independent variables or among the spatial independent variables in the Lachlan data.

Plots of Pearson's correlations (Figure 24, Figure 25 and Figure 26) indicated that strong correlations were present between the distance to the nearest local population centre, distance to the nearest point on the road network and the travel time variable, and also between the distance to the nearest Ramsar site and the nearest DIWA wetland (although these did not refer to the same site for the Lachlan data). Modest negative correlation was present between NDVis and maximum temperatures. However, apart from the inevitable collinearity between the irrigated area variable and the interaction term between irrigated area and NDVi, all VIFs were lower than 3.0. This suggested that although correlations were present these should not cause collinearity problems for model estimation. All independent variables were therefore used in the full model to guard against potential omitted variable bias.

Table 6: Variance inflation factors (VIFs) for (a) spatio-temporal independent variables, (b) spatial independent variables, and (c) sub-catchment areas and populations as independent variables.

(a)		(b)		(c)	
Variable	VIF	Variable	VIF	Variable	VIF
NDVi	1.42	PC_DIST_km	1.99	CCArea_km2	2.78
Rain	1.05	Dist2Road_DIST_km	1.40	IrrigAreaPc	4.67
MTemp	1.40	Min_mins	2.02	NDViByIrrigPc	4.48
WatPc95	1.01	RAMSAR_DIST_km	2.01	AvgMthlyPop	2.85
		DIWA_DIST_km	2.66		
		IBA_DIST_km	1.35		
		CAPAD_DIST_km	1.23		

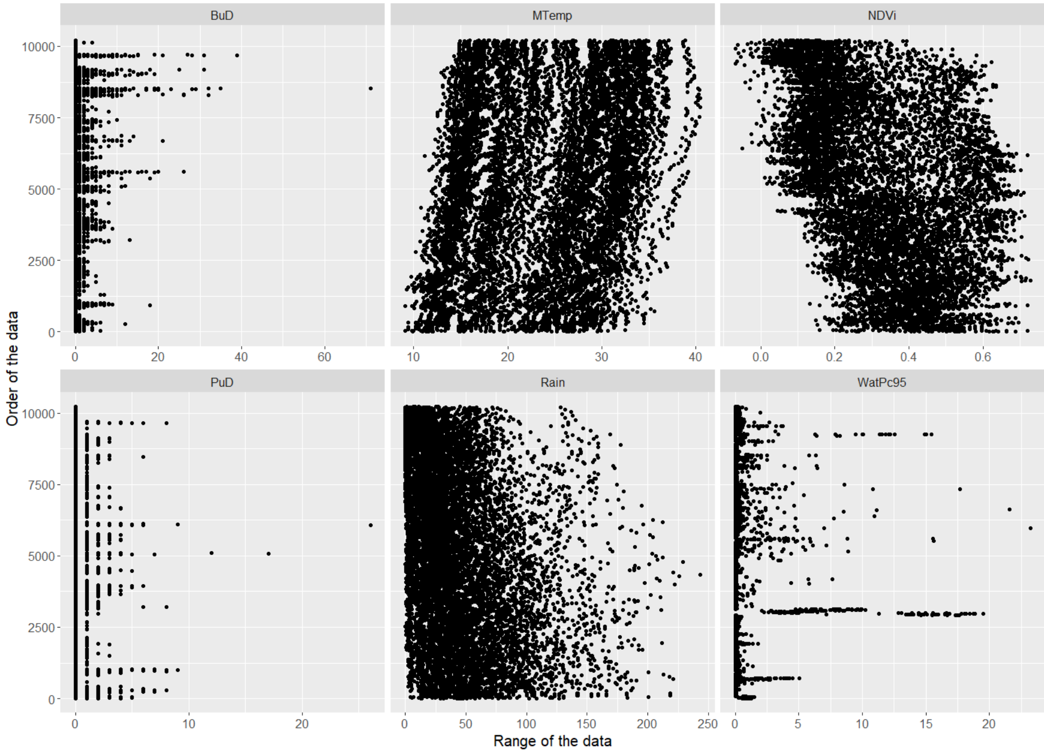

Figure 21: Cleveland dotplots showing BuDs, PuDs and spatio-temporal independent variables plotted against order of the data (1 : (140 x 73 =10220)).

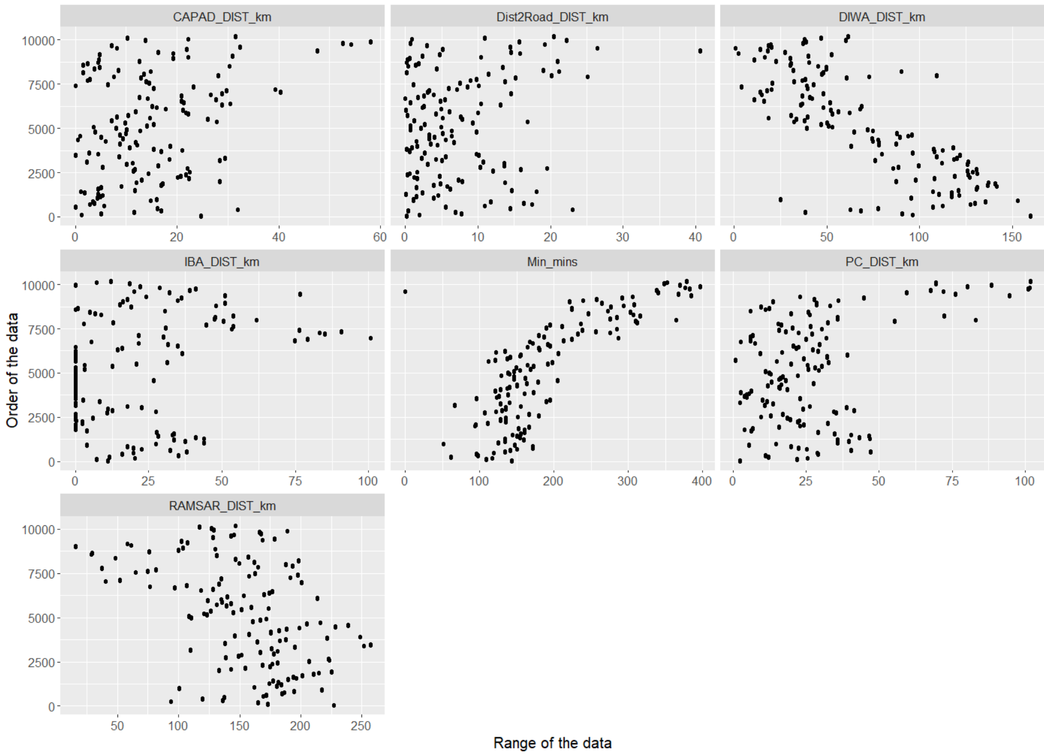

Figure 22: Cleveland dotplots showing spatial independent variables plotted against order of the data (1 : (140 x 73 =10220)).

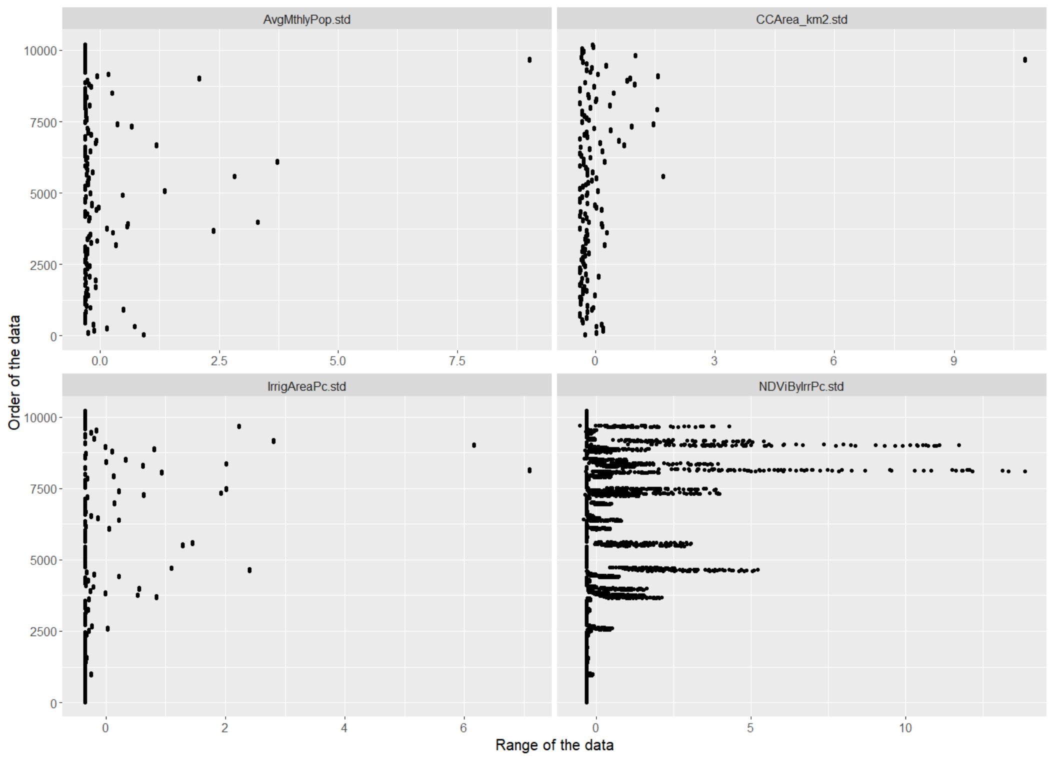

Figure 23: Cleveland dotplots showing sub-catchment area, percentage irrigated area per sub-catchment, average resident population and the sub-catchment area x percentage irrigated area interaction term plotted against order of the data (1 : (140 x 73 =10220)).

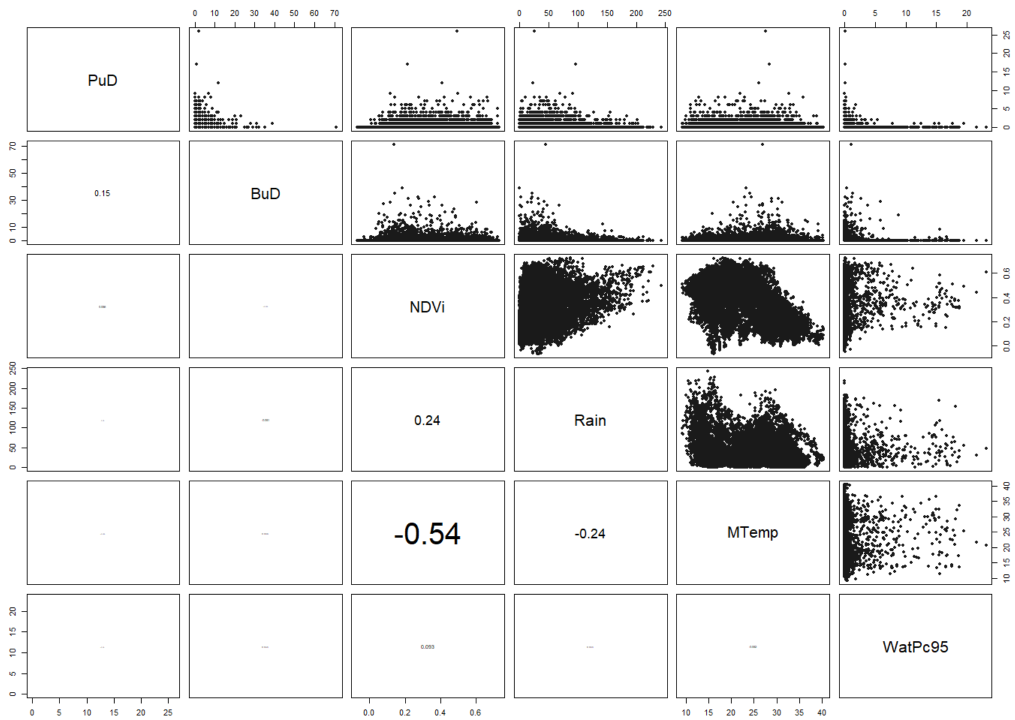

Figure 24: Pearsons correlations plots for spatio-temporal independent variables.

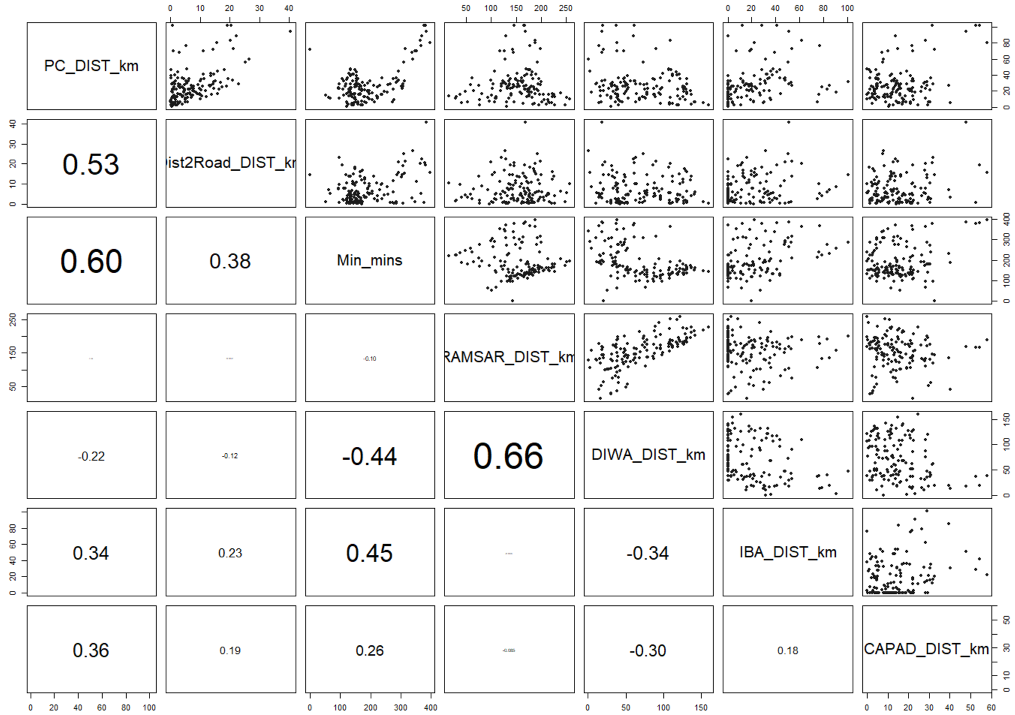

Figure 25: Pearsons correlations plots for spatial independent variables.

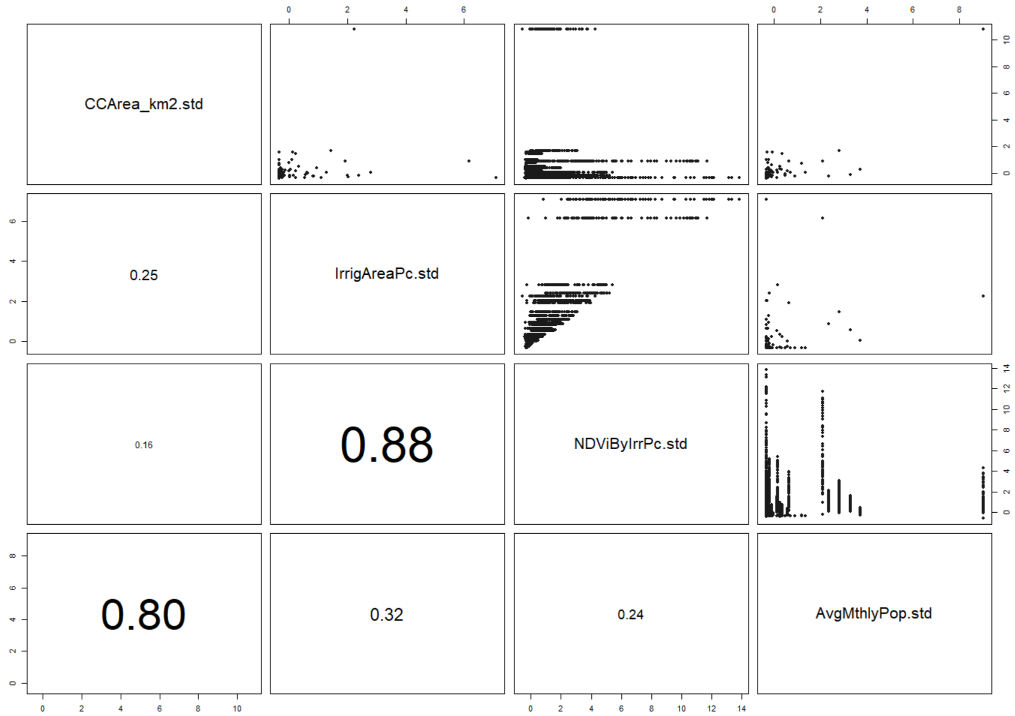

Figure 26: Pearsons correlations plots for sub-catchment area, percentage irrigated area per sub-catchment, average resident population and the sub-catchment area x percentage irrigated area interaction term.

7.1.3. Linear predictor for model fitting

After data screening, the following fixed covariate terms were included in the linear predictor of the log of expected BuD or PuD counts for sub-catchment i in month-index t for fitting the model sequence M1 : M8:

$$
\begin{aligned}
ln(\mu_{it}) = \beta_1 &+ \beta_2 NDVi_{it} + \beta_3 Rain_{it} + \beta_4 MTemp_{it} \\
&+ \beta_5 PC\_Dist_i + \beta_6 Dist\_to\_Road_i + \beta_7 DIWA\_Dist_i \\
&+ \beta_8 IBA\_Dist_i + \beta_9 CAPAD\_Dist_i \\
&+ \beta_{10} CC\_Area_i + \beta_{11} IrrigAreaPc_i + \beta_{12} NDViByIrrP_i + \beta_{13} AvgPopln_i
\end{aligned}
$$

An interaction term between NDVi and percentage irrigated area is included to see whether the impact of landscape greenness on visitation rate increases as the percentage area of irrigated land in the sub-catchment decreases. The percentage area of irrigated land in a sub-catchment is included as a separate term in the regression partly because the interaction term between NDVi and percentage irrigated area is included, and partly because percentage irrigated area could plausibly influence visitation rate directly because there is unlikely to be public access to privately-owned irrigated farmland.

Terms to represent temporal variation in BuD or PuD counts within sub-catchments are also included in the linear predictor as shown below:

$+\beta_{14f} Month\_factor_t + f(year_t)$ for years 2013 to 2019 [in M1 to M4]

or

$+f(month\_index_t)$ for month_index 1 to 73 [in M5 onwards]

Sub-catchment-specific random terms were also included in M4 to M6, and conditional autoregressive (CAR)-weighted sub-catchment-specific spatial random terms were included in M7 and M8 as shown below:

$+\alpha_i \sim N(0, \sigma^2_{ID})$ with ID denoting sub-catchment ID [in M4, M5 and M6]

or

$+\alpha_i \sim N(0, \sigma^2_{CAR})$ with CAR denoting CAR spatial weighing [in M7 and M8]

Separate models were fitted for BuDs and PuDs, with NDVi and WatPc95 introduced (separately) as alternative proxies for environmental condition. The sections immediately following present results from models to predict monthly sub-catchment BuD counts using NDVi as a proxy for environmental condition.

7.1.4. BuDs via NDVi: Model selection via fit simulations and tests for temporal and spatial autocorrelation among the residuals

Modelling approaches across M1 to M8 to accommodate potential zero inflation and temporal and spatial autocorrelation in BuD and PuD data were explored using the methods described in Subsections 6.2.2 to 6.2.4. Results are reported in Table 7.

Plots showing the total number of zero counts produced by 5000 simulations from the estimated distributions of model coefficients were used to assess whether the functional form of the model could accommodate the level of zero inflation in the data (Figure 27). The red dot in these plots shows the number of zero counts in the full set of Lachlan River Region BuD_{it} data (8276 of 10220 data points: 81%).

For the modelling sequence M1 to M8, Table 7 reports overall fit metrics: marginal likelihood, DIC, WAIC, the negative sum of the natural logarithm of CPO scores, the model dispersion statistic, the percentage of (month, sub-catchment) data pairs for which Moran's I test suggested spatial autocorrelation was present in regression residuals (at $\alpha = 0.05$), and the percentage of sub-catchment-specific 73-month time series for which the ACF test suggested temporal autocorrelation was present in the regression residuals (at $\alpha = 0.05$).

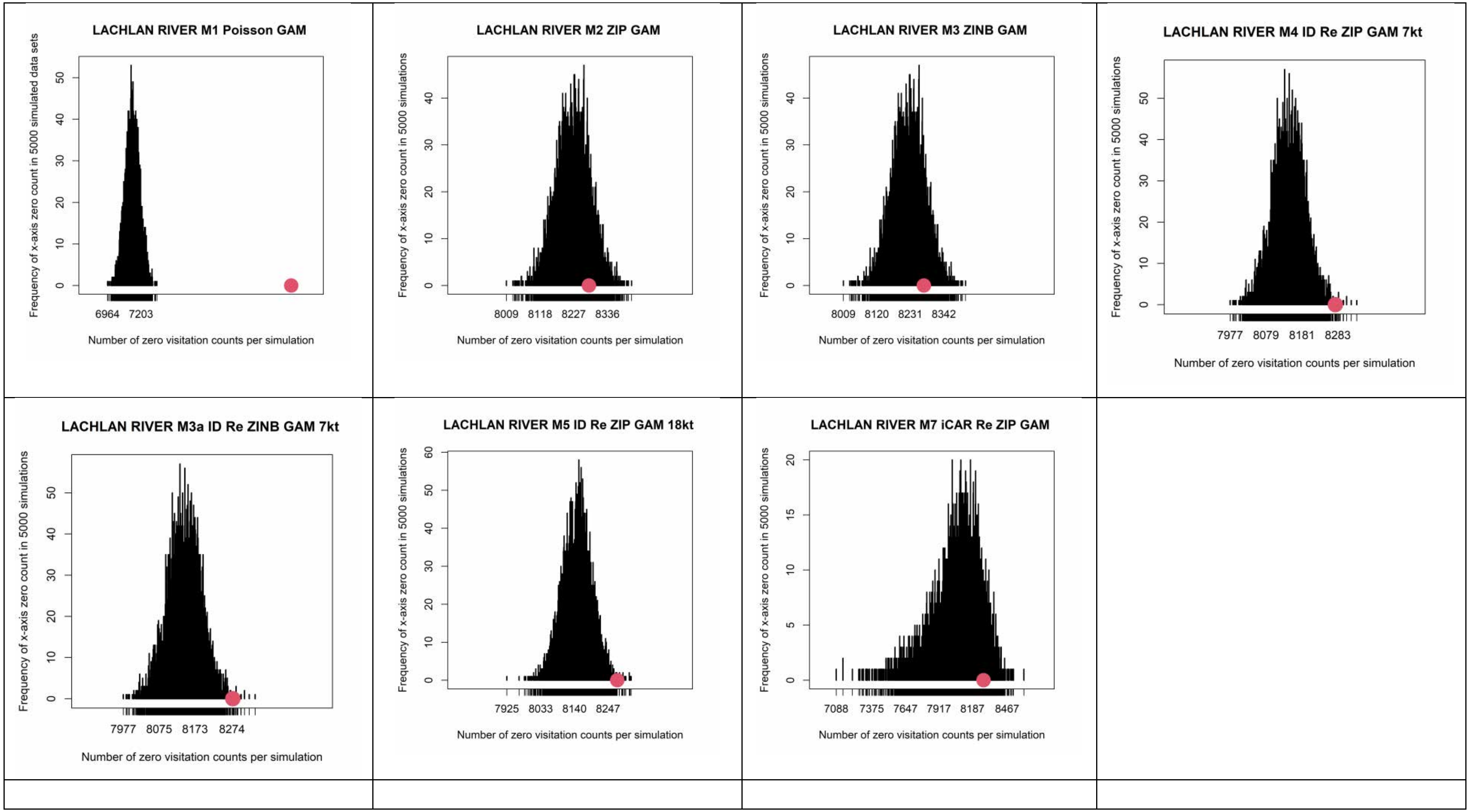

Figure 27: For the Lachlan River Region, histograms showing the spread in total number of zero BuD_{it} counts obtained from 5000 simulation runs using parameter realisations drawn from fitted posterior marginal distributions from models M1 to M4, M3a, M5 and M7. The red dots report the number of zero BuD_{it} counts in the Lachlan River data.

Table 7: BuDs & NDVi: Model form assessments and overall fit metrics across models M1 to M8 for BUD_{it} counts from 140 sub-catchments in the Lachlan River Region over 73 months.

Lachlan River: BuDs & NDVi	**M1**	**M2**	**M3**	**M4**	**M3a**	**M5**	**M6**	**M7**	**M8**
Model form	Poisson	ZIP	ZINB	ZIP	ZINB	ZIP	ZINB	ZIP	ZINB
Random effects	none			ID REs				iCAR REs	
Temporal form	Month as a categorical factor + *f(year)* as a 7-knot smoother					*f(month-index)* as an 18-knot smoother			
Model form assessments									
Zero-inflation adequate?	no	yes	yes	no	no	no		yes	
Moran's I (% fitted points p < 0.05)	39.7	37.0	50.7	23.2	21.9	24.7		26.0	
Temporal ACF (% fitted points p < 0.05)	15.7	15.7	14.3	24.3	14.3	24.3		27.1	
Model fit metrics									
Marginal likelihood	-10157.4	-8640.79	-7537.96	-6973.92	-6493.97	-7300.8		-7339.08	
DIC	19723.7	16817.84	14718.59	13276.12	12359.35	13916.83		14275.3	
WAIC	22444.3	18435.39	14725.19	13427.08	12371.78	14485.97		15173.05	
CPO via $-\sum_{i=1}^{n} ln(CPO_i)$	-9999.2	-8477.6	-7362.65	-6697.13	-6186.25	-7027.17		-31768.1	
Model dispersion	2.98	1.48	1.25	1.08	0.91	1.08		0.970	

Table 8: BuDs & NDVi: Means of estimated marginal posterior distributions for covariates across models M1 to M8 for BUD_{it} counts from 140 sub-catchments in the Lachlan River Region.

Lachlan River: BuDs & NDVi	**M1**		**M2**		**M3**		**M4**		**M3a**		**M5**		**M6**	**M7**		**M8**
Model form	Poisson		ZIP		ZINB		ZIP		ZINB		ZIP		ZINB	ZIP		ZINB
Random effects	none						ID REs							iCAR REs		
Temporal form	Month as a categorical factor + *f(year)* as a 7-knot smoother										*f(month-index)* as an 18-knot smoother					
(Intercept)	**-1.247**	-	-0.006	n.i.	**-0.683**	-	**-2.646**	-	**-2.888**	-	**-2.384**	-		**-2.531**	-	
NDVi.std	**-0.170**	-	**-0.119**	-	**-0.117**	-	**-0.103**	-	**-0.126**	-	**0.127**	+		**0.129**	+	
Rain.std	**-0.086**	-	**-0.071**	-	**-0.089**	-	-0.028	n.i.	-0.038	n.i.	-0.02	n.i.		-0.02	n.i.	
MTemp.std	**-0.338**	-	**-0.281**	-	**-0.759**	-	-0.033	n.i.	-0.095	n.i.	**0.250**	+		**0.248**	+	
PC_DIST_km.std	**-1.29**	-	**-0.895**	-	**-1.255**	-	**-1.679**	-	**-1.655**	-	**-1.659**	-		**-1.571**	-	
Dist2Road_DIST_km.std	**-0.115**	-	**-0.184**	-	**-0.156**	-	-0.173	n.i.	-0.183	n.i.	-0.171	n.i.		-0.095	n.i.	
Min_mins.std	**0.369**	+	**0.37**	+	**0.246**	+	-0.017	n.i.	-0.042	n.i.	0.021	n.i.		-0.042	n.i.	
DIWA_DIST_km.std	**-0.566**	-	**-0.578**	-	**-0.597**	-	-0.125	n.i.	-0.131	n.i.	-0.135	n.i.		0.33	n.i.	
RAMSAR_DIST_km.std	**0.174**	+	**0.219**	+	**0.088**	+	-0.188	n.i.	-0.188	n.i.	-0.204	n.i.		-0.827	n.i.	
IBA_DIST_km.std	**-0.295**	-	**-0.244**	-	**-0.37**	-	-0.126	n.i.	-0.143	n.i.	-0.12	n.i.		-0.399	n.i.	
CAPAD_DIST_km.std	**0.128**	+	**0.05**	+	-0.015	n.i.	-0.066	n.i.	-0.065	n.i.	-0.061	n.i.		-0.044	n.i.	
CCArea_km2.std	**0.257**	+	**0.176**	+	**0.601**	+	**0.854**	+	**0.841**	+	**0.848**	+		**0.739**	+	
IrrigAreaPc.std	**0.193**	+	**0.157**	+	**0.239**	+	0.138	n.i.	0.152	n.i.	0.129	n.i.		0.253	n.i.	
NDViByIrrPc.std	0	n.i.	0	n.i.	-0.003	n.i.	0.003	n.i.	0.016	n.i.	-0.006	n.i.		-0.005	n.i.	
AvgMthlyPop.std	0.155	n.i.	0.032	n.i.	**0.179**	+	0.113	n.i.	0.151	n.i.	0.118	n.i.		0.186	n.i.	
negbin_size_param					0.319	+			1.165	+						
zero_prob_param			0.644	+	0.007	+	0.261	+	0.006	+	0.305	+		0.304	+	
random_effect_variance							1.753	+	1.751	+	1.761	+		2.414	+	

Table 9: BuDs & NDVi: Means of estimated marginal posterior distributions for temporal covariates and smoothers across models M1 to M8 for BUD_{it} counts from 140 sub-catchments in the Lachlan River Region over 73 months.

Lachlan River: BuDs & NDVi	**M1**		**M2**		**M3**		**M4**		**M3a**		**Smoother across month-index 1 to 73**	**M5**		**M6**	**M7**		**M8**
Model form	Poisson		ZIP		ZINB		ZIP		ZINB			ZIP		ZINB	ZIP		ZINB
Temporal form	**Month as a categorical factor + *f(year)* as a 7-knot smoother**											***f(month-index)* as an 18-knot smoother**					
fmonthFeb	**-0.765**	-	**-0.738**	-	**-0.843**	-	**-0.548**	-	-0.603	-	mindex.cr1	**-0.463**	-		**-0.461**	-	
fmonthMar	-0.143	n.i.	**-0.233**	-	**-0.437**	-	0.072	n.i.	0.131	n.i.	mindex.cr2	**-0.346**	-		**-0.347**	-	
fmonthApr	0.036	n.i.	-0.075	n.i.	**-0.531**	-	**0.445**	+	0.43	+	mindex.cr3	**-0.95**	-		**-0.941**	-	
fmonthMay	**-0.846**	-	**-0.784**	-	**-1.963**	-	-0.172	n.i.	-0.398	n.i.	mindex.cr4	-0.112	n.i.		-0.111	n.i.	
fmonthJun	**-1.149**	-	**-1.075**	-	**-2.289**	-	-0.37	n.i.	-0.485	n.i.	mindex.cr5	**-0.383**	-		**-0.383**	-	
fmonthJul	**-0.935**	-	**-0.856**	-	**-2.159**	-	-0.103	n.i.	-0.227	n.i.	mindex.cr6	**-0.24**	-		**-0.238**	-	
fmonthAug	**-1.066**	-	**-1.114**	-	**-2.123**	-	-0.398	n.i.	-0.361	n.i.	mindex.cr7	**0.424**	+		**0.422**	+	
fmonthSep	**0.351**	+	0.185	n.i.	**-0.659**	-	**0.846**	+	**0.747**	+	mindex.cr8	**-0.314**	-		**-0.314**	-	
fmonthOct	**0.912**	+	**0.657**	+	0.211	n.i.	**1.138**	+	**1.129**	+	mindex.cr9	0.158	n.i.		0.158	n.i.	
fmonthNov	**0.472**	+	**0.487**	+	-0.075	n.i.	**0.729**	+	**0.528**	+	mindex.cr10	**0.215**	+		**0.214**	+	
fmonthDec	-0.149	n.i.	**-0.18**	-	**-0.345**	-	-0.039	n.i.	-0.049	n.i.	mindex.cr11	**0.419**	+		**0.419**	+	
year.cr1	**-0.352**	-	**-0.344**	-	**-0.355**	-	**-0.354**	-	**-0.332**	-	mindex.cr12	**0.257**	+		**0.256**	+	
year.cr2	0.029	n.i.	0.045	n.i.	0.016	n.i.	0.046	n.i.	0.07	n.i.	mindex.cr13	**0.999**	+		**0.997**	+	
year.cr3	**0.209**	+	**0.205**	+	**0.155**	+	**0.219**	+	**0.225**	+	mindex.cr14	**-0.29**	-		**-0.29**	-	
year.cr4	**0.655**	+	**0.558**	+	**0.706**	+	**0.615**	+	**0.699**	+	mindex.cr15	**0.764**	+		**0.762**	+	
year.cr5	**0.453**	+	**0.315**	+	**0.518**	+	**0.442**	+	**0.511**	+	mindex.cr16	**0.796**	+		**0.795**	+	
year.cr6	**0.449**	+	**0.38**	+	**0.606**	+	**0.385**	+	**0.554**	+	mindex.cr17	**-0.372**	-		**-0.370**	-	

Figure 27 shows that only the ZIP (M2), ZINB (M3) and ZIP + iCAR random effects (M7) models appear fully able to handle the level of zero inflation in the BuD data, although the ZIP and ZINB models with ID random effects (M4, M5 and M3a) and are almost able to do this.

Results in Table 7 show that Moran's I test suggests that sub-catchment specific random effects (models M4, M3a, M5 and M7) are required to reduce the occurrence of spatial autocorrelation to below 30% of the model time frame (i.e. less than 30% of the 73 monthly snapshots of predicted BuD counts). Temporal autocorrelation does not appear to be a particular problem, affecting less than 30% of the Lachlan's 140 sub-catchments for all models tested.

The model fit metrics (marginal likelihood, DIC and WAIC) in Table 7 indicate that sub-catchment specific random effects improve overall model fit and suggest that temporal representation via a month factor term and the 7-knot year smoother provides a more parsimonious solution than the 18-knot month-index smoother.

(Separate plots of fitted BUD count estimates showed that whilst the overall fitting performance of the ZINB models (M3 and M3a) was good, these models were prone to heavily over-estimating BuD counts at hotspots. This led to the ZINB models being set aside; hence the more sophisticated ZINB models (M6 and M8) were not tested.)

Based on ability to handle zero inflation, performance regarding spatial and temporal autocorrelation, and metrics of model fit, models M4, M5 and M7 are regarded as the most appropriate models for predicting BUD counts across sub-catchments in the Lachlan over the 73-month model timeframe.

7.1.5. BuDs via NDVi: Covariate parameter estimates and temporal representation

Means of the estimated marginal posterior distributions for model covariates are reported in Table 8. Estimates for the zero inflation parameter (π), the negative binomial size parameter (k) and the variance of sub-catchment specific random effects ($\sigma^2_{\alpha_i}$) are also reported, where relevant. Parameters which are important (i.e., the 95% credible interval about thee mean of their estimated posterior distribution does not span zero) are shown in bold, with their sign of action marked as '+' or '-'. Parameters that are not important are marked as 'n.i.'.

Once sub-catchment specific random effects are introduced (M4, M5, M7 of the ZIP models) only a small number of important covariates remain: $NDVi_{it}$, $MTemp_{it}$, $PC\_DIST\_km_i$ and $CCArea\_km2_i$.

In these models the temporal terms and the sub-catchment specific random effects provide the remainder of the fit. The impact of covariates $PC\_DIST\_km_i$ and $CCArea\_km2_i$ remains consistent throughout the model fitting sequence. This suggests that these terms (which represent the 'accessibility' and size of a sub-catchment) provide additional ability to inform the spatial distribution of BUD counts across sub-catchments. In contrast, covariate $NDVi_{it}$ flips sign from negative to positive, and the $MTemp_{it}$ term becomes important, when the 18-knot month-index smoother replaces the factor month plus 7-knot year smoother. This suggests that these spatial-temporal covariates provide additional ability to inform the spatio-temporal distribution of BUD counts.

Table 9 reports the means of estimated posterior marginal distributions for the month factor parameters and the smoother terms (for the 7-knot year smoother and the 18-knot month-index smoother). The month factor parameter estimates are the main interest here. January is the baseline month hence parameter estimates indicate that BUD counts are generally higher across all sub-catchments during September, October and November.

7.1.6. BuDs via NDVi: Plots of fitted values

Results from fitting the model sequence M1 to M7 to monthly sub-catchment BUD counts in the Lachlan, using NDVi as a proxy for ‘environmental greenness’, suggest that ZIP models with a sub-catchment specific random effect and a temporal representation of either month as a factor plus a 7-knot year smoother (M4), or as an 18-knot month-index smoother (M5), are the most appropriate. Predicted expected monthly BuD counts per sub-catchment, aggregated into total BuD counts per sub-catchment per year, from these two models are compared with the Lachlan BUD count data in Figure 28 overleaf.

The plots in Figure 28 show that predictions from models M4 and M5 are very similar to the BUD count data. As noted earlier, most of this explanatory power comes from the sub-catchment specific random effects in combination with the (common across all sub-catchments) temporal representation. However, spatial covariates $PC\_DIST\_km_i$ and $CCArea\_km2_i$ contribute consistently to the fit of these models, and the spatio-temporal covariates $NDVi_{it}$ and $MTemp_{it}$ also play a role, but this is not consistent as the temporal representation changes from month as a factor plus a 7-knot year smoother in M4 to an 18-knot month-index smoother in M5.

The subsections below follow a similar sequence in reporting results from models to predict monthly sub-catchment BuD counts using WatPc95 as a proxy for environmental condition.

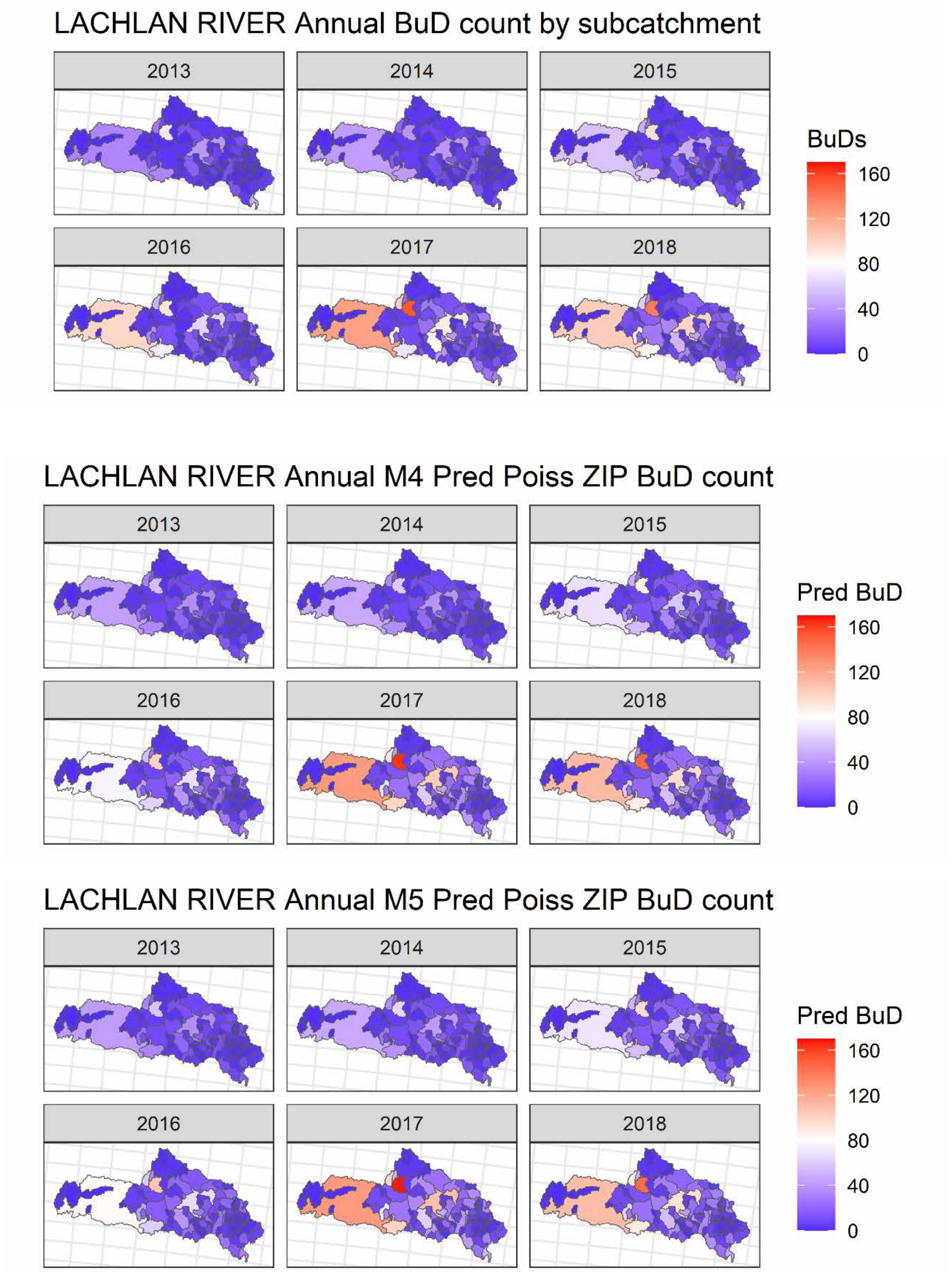

Figure 28: Comparison of total annual eBird-derived BUD count data (top panel) with predicted annual BUD count (summed from predicted monthly BUD counts) from fitted models M4 (mid panel) and M5 (lower panel) when NDVi is used as a proxy for environmental condition.

7.1.7. BuDs via WatPc95: Model selection via fit simulations and tests for temporal and spatial autocorrelation among the residuals

Table 10: BuDs & WatPc95: Model form assessments and fit metrics across simplified models M1 to M8 for BUD_{it} counts from 140 sub-catchments in the Lachlan River Region over 73 months.

Lachlan River: BuDs & WatPc95	**M1***	**M2***	**M3**	**M4****	**M3a**	**M5****	**M6**	**M7****	**M8**
Model form	Poisson	ZIP	ZINB	ZIP	ZINB	ZIP	ZINB	ZIP	ZINB
Random effects	none			ID REs				iCAR REs	
Temporal form	Month as a categorical factor + *f(year)* as a 7-knot smoother					*f(month-index)* as an 18-knot smoother			
Model form assessments									
Zero-inflation adequate?	no	yes		only just		only just		only just	
Moran's I (% fitted points p < 0.05)	omitted	omitted		omitted		omitted		omitted	
Temporal ACF (% fitted points p < 0.05)	omitted	omitted		omitted		omitted		omitted	
Model fit metrics: all covariates									
Marginal likelihood	-7293.53	-6189.07		-5013.31		-5275.32		-5313.5	
DIC	13959.07	11862.83		9399.71		9808.58		9890.34	
WAIC	15069.51	13201.55		9561.69		10890.89		10976.39	
CPO via $-\sum_{i=1}^{n} ln(CPO_i)$	-7135.42	-6030.82		-4761.63		-5024.01		-5325.43	
Model dispersion	3.37	1.61		1.07		1.04		1.03	
Model fit metrics: reduced covariates									
Marginal likelihood	-7280.4	-6177.59		-4963.27		-5223.55		-5267.23	
DIC	13954.09	11857.82		9398.01		9805.8		9804.44	
WAIC	15071.51	13275.66		9571.79		10881.48		10883.97	
CPO via $-\sum_{i=1}^{n} ln(CPO_i)$	-7133.26	-6029.37		-4759.55		-5020.88		-5039.67	
Model dispersion	3.37	1.61		1.07		1.04		1.06	

*M1 & M2: remove Rain & MTemp; **M4, M5 & M7: only retain WatPc95, MTemp, PC_DIST_km and CCArea_km2.

7.1.8. BuDs via WatPc95: Covariate parameter estimates and temporal representation

Table 11: BuDs & WatPc95: Means of estimated marginal posterior distributions for covariates across M1 to M8 for BUD_{it} counts from 140 sub-catchments in the Lachlan over 73 months.

Lachlan River: BuDs & WatPc95	M1		M2		M3		M4		M3a		M5		M6	M7		M8
Model form	Poisson		ZIP		ZINB		ZIP		ZINB		ZIP		ZINB	ZIP		ZINB
Random effects	none						ID REs							iCAR REs		
Temporal form	Month as a categorical factor + *f(year)* as a 7-knot smoother										*f(month-index)* as an 18-knot smoother					
(Intercept)	**-1.764**	-	**-0.43**	-			**-2.68**	-			**-2.274**	-		**-2.407**	-	
WatPc95.std	**0.125**	+	**0.173**	+			0.031	n.i.			**0.137**	+		**0.140**	+	
Rain.std	0.009	n.i.	0.046	n.i.			0.045	n.i.			-0.007	n.i.		-0.007	n.i.	
MTemp.std	-0.010	n.i.	-0.015	n.i.			0.081	n.i.			**0.163**	+		**0.163**	+	
PC_DIST_km.std	**-1.412**	-	**-0.971**	-			**-1.630**	-			**-1.637**	-		**-1.635**	-	
Dist2Road_DIST_km.std	**-0.066**	-	**-0.139**	-			-0.160	n.i.			-0.152	n.i.		-0.068	n.i.	
Min_mins.std	**0.503**	+	**0.412**	+			-0.013	n.i.			-0.017	n.i.		-0.048	n.i.	
DIWA_DIST_km.std	**-0.516**	-	**-0.508**	-			-0.167	n.i.			-0.165	n.i.		0.301	n.i.	
RAMSAR_DIST_km.std	**0.164**	+	**0.184**	+			-0.149	n.i.			-0.139	n.i.		-0.733	n.i.	
IBA_DIST_km.std	**-0.346**	-	**-0.277**	-			-0.119	n.i.			-0.129	n.i.		-0.340	n.i.	
CAPAD_DIST_km.std	**0.131**	+	**0.064**	+			-0.068	n.i.			-0.079	n.i.		-0.065	n.i.	
CCArea_km2.std	**0.287**	+	**0.245**	+			**0.813**	+			**0.821**	+		**0.715**	+	
IrrigAreaPc.std	**0.221**	+	**0.178**	+			0.162	n.i.			0.106	n.i.		0.217	n.i.	
NDViByIrrPc.std	**-0.050**	-	-0.034	n.i.			-0.030	n.i.			0.02	n.i.		0.021	n.i.	
AvgMthlyPop.std	**0.161**	+	0	n.i.			0.134	n.i.			0.126	n.i.		0.205	n.i.	
negbin_size_param																
zero_prob_param			**0.64**	+			**0.254**	+			**0.305**	+		**0.304**	+	
random_effect_variance							1.674				1.679			2.285		

Table 12: BuDs & WatPc95: Means of estimated marginal posterior distributions for temporal covariates and smoothers across models M1 to M8 for BUD_{it} counts from 140 sub-catchments in the Lachlan River Region over 73 months.

Lachlan River: BuDs & WatPc95	**M1**		**M2**		**M3**		**M4**		**M3a**			**M5**		**M6**	**M7**		**M8**
Model form	Poisson		ZIP		ZINB		ZIP		ZINB			ZIP		ZINB	ZIP		ZINB
Temporal form	**Month as a categorical factor + *f(year)* as a 7-knot smoother**										**smoothers**	***f(month-index)* as an 18-knot smoother**					
fmonthFeb	**-0.579**	-	**-0.554**	-			**-0.491**	-			mindex.cr1	-0.757	-		-0.756	-	
fmonthMar	**0.203**	+	-0.003	n.i.			0.137	n.i.			mindex.cr2	-0.485	-		-0.484	-	
fmonthApr	**0.622**	+	**0.433**	+			**0.664**	+			mindex.cr3	-1.038	-		-1.034	-	
fmonthMay	-0.109	n.i.	-0.139	n.i.			0.068	n.i.			mindex.cr4	0.001	n.i.		0.001	n.i.	
fmonthJun	-0.254	n.i.	-0.367	n.i.			-0.092	n.i.			mindex.cr5	-0.343	-		-0.342	-	
fmonthJul	-0.050	n.i.	-0.191	n.i.			0.065	n.i.			mindex.cr6	-0.488	-		-0.485	-	
fmonthAug	-0.318	n.i.	**-0.475**	-			-0.185	n.i.			mindex.cr7	0.451	+		0.449	+	
fmonthSep	**1.087**	+	**0.83**	+			**1.161**	+			mindex.cr8	-0.528	-		-0.527	-	
fmonthOct	**1.265**	+	**0.951**	+			**1.214**	+			mindex.cr9	0.104	n.i.		0.104	n.i.	
fmonthNov	**0.749**	+	**0.714**	+			**0.805**	+			mindex.cr10	0.04	n.i.		0.034	n.i.	
fmonthDec	0.102	n.i.	0.004	n.i.			0.026	n.i.			mindex.cr11	0.39	+		0.39	+	
year.cr1	**-0.351**	-	**-0.324**	-			**-0.361**	-			mindex.cr12	0.237	n.i.		0.236	n.i.	
year.cr2	0.079	n.i.	0.11	+			0.066	n.i.			mindex.cr13	1.077	+		1.075	+	
year.cr3	**0.183**	+	**0.16**	+			**0.197**	+			mindex.cr14	-0.4	-		-0.399	-	
year.cr4	**0.864**	+	**0.743**	+			**0.754**	+			mindex.cr15	0.482	+		0.481	+	
year.cr5	**0.640**	+	**0.509**	+			**0.598**	+			mindex.cr16	0.715	+		0.715	+	
year.cr6	**0.660**	+	**0.526**	+			**0.508**	+			mindex.cr17	-0.4	-		-0.399	-	

7.1.9. BuDs via WatPc95: Modelling summary and plots of fitted values

Models M1, M2, M4, M5 and M7 were assessed for their suitability for modelling monthly BUD counts per sub-catchment in the Lachlan River Region when the percentage inundated area (per month) per sub-catchment (WatPc95) is used as a proxy for environmental condition. Percentage inundated areas are only reported for a sub-catchment for months in which at least 95% of the area of the sub-catchment is visible via remote sensing. Months for which WatPc95 data are not available are removed from the analysis; hence the dataset for this analysis is smaller than that for the analysis in which NDVi was the environmental proxy (7227 data points instead of 10220). These breaks in the monthly sequence of data disrupt assessment of spatial and temporal autocorrelation, hence these tests were not conducted for these models. The overall percentage of zero counts in this dataset is 81%.

Findings from sequential modelling of the various functional forms were very similar to those when BUD counts were modelled with NDVi as the environmental proxy (see Table 10, Table 11 and Table 12).

Sub-catchment specific random terms are again important for improving model fit. The spatial covariates $PC\_DIST\_km_i$ and $CCArea\_km2_i$ are consistently important for predicting BuD count irrespective of the form of the model and temporal representation. The spatio-temporal covariates $WatPc95_{it}$and $MTemp_{it}$ both appear as important for predicting BuD count once the 18-knot month-index smoother provides cross sub-catchment temporal variation in the models (M5 and M7). Both $WatPc95_{it}$and $MTemp_{it}$ show a positive associate with BuD count, i.e., increased percentage inundation and higher maximum temperature in a sub-catchment are associated with a increase in predicted sub-catchment BuD count.

These results are surprising because percentage inundation area is much less seasonally cyclic than NDVi and its patterns of variation are much stronger in some sub-catchments than others (which again contrasts with NDVi).

Predicted expected monthly BuD counts per sub-catchment, aggregated into total BuD counts per sub-catchment per year, from models M4 and M5 are compared with the Lachlan BUD count data in Figure 29 overleaf. The plots show that predictions from models M4 and M5 are very similar to the BUD count data. As noted earlier, most of this explanatory power comes from the sub-catchment specific random effects in combination with the (common across all sub-catchments) temporal representation. However, spatial covariates $PC\_DIST\_km_i$ and $CCArea\_km2_i$ contribute consistently to the fit of these models, and the spatio-temporal covariates $WatPc95_{it}$ and $MTemp_{it}$ also play a role, but this is not consistent as the temporal representation changes from month as a factor plus a 7-knot year smoother in M4 to an 18-knot month-index smoother in M5.

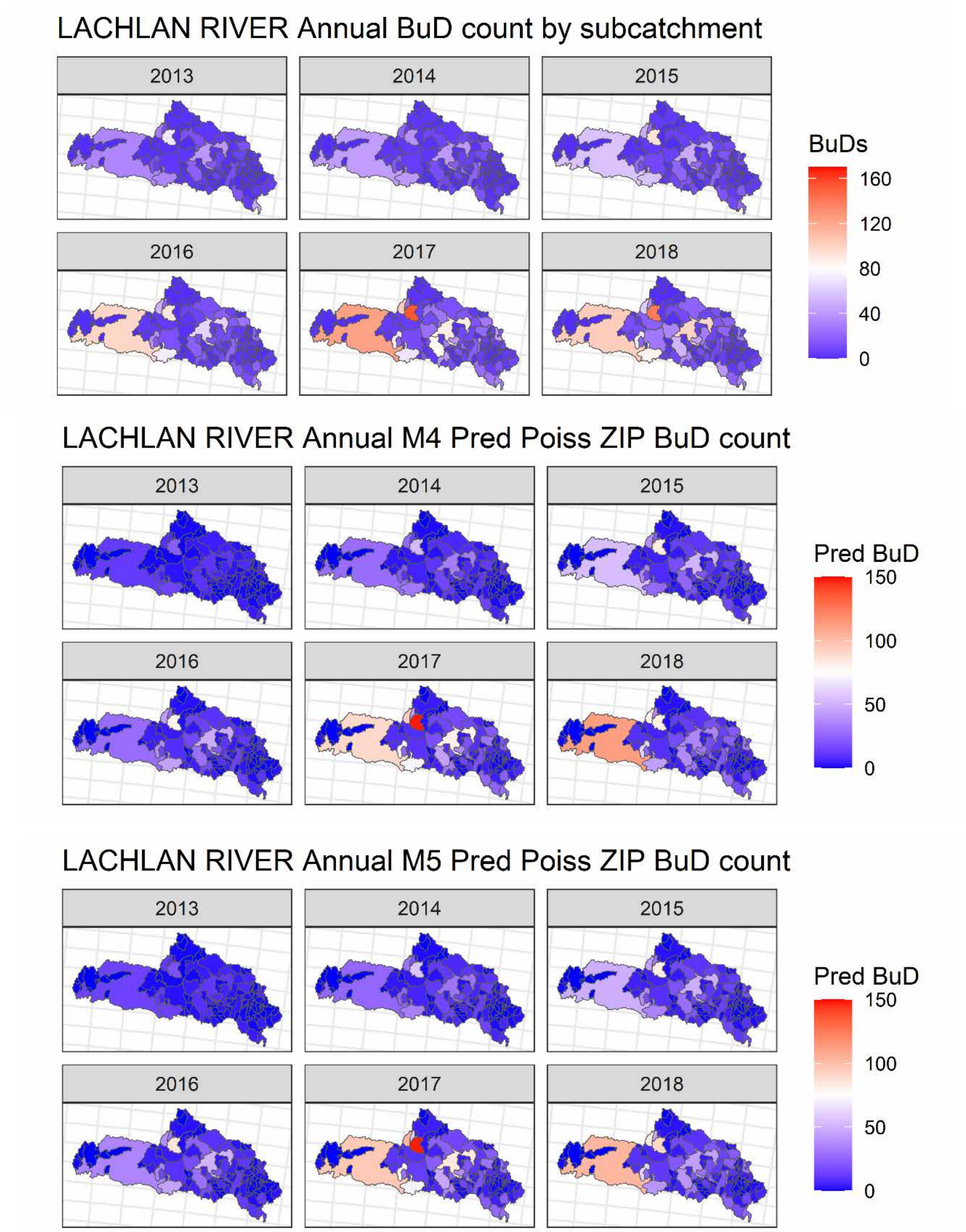

Figure 29 Comparison of total annual eBird-derived BUD count data (top panel) with predicted annual BUD count (summed from predicted monthly BUD counts) from fitted models M4 (mid panel) and M5 (lower panel) when WatPc95 is used as a proxy for environmental condition.

7.1.10. PuDs via NDVi: Model selection via fit simulations and tests for temporal and spatial autocorrelation among the residuals

Table 13: PuDs via NDVi: Model form assessments and overall fit metrics across models M1 to M8 for PUD_{it} counts from 140 sub-catchments in the Lachlan River Region over 73 months.

Lachlan River: PuDs & NDVi	**M1**	**M2**	**M3**	**M4**	**M3a**	**M5**	**M6**	**M7**	**M8**
Model form	Poisson	ZIP	ZINB	ZIP	ZINB	ZIP	ZINB	ZIP	ZINB
Random effects	none			ID REs				iCAR REs	
Temporal form	Month as a categorical factor + *f(year)* as a 7-knot smoother					*f(month-index)* as an 18-knot smoother			
Model form assessments									
Zero-inflation adequate?	no	yes		yes		yes		yes	
Moran's I (% fitted points p < 0.05)	11.0	9.6		13.7		13.7		12.3	
Temporal ACF (% fitted points p < 0.05)	15.7	14.3		31.4		32.9		27.1	
Model fit metrics: all covariates									
Marginal likelihood	-4470.07	-4269.88		-3786.5		-3806.3		-3845.87	
DIC	8601.16	8202.74		7063.42		7112.16		7135.15	
WAIC	8685.37	8232.87		7080.09		7130.64		7151.69	
CPO via $-\sum_{i=1}^{n} ln(CPO_i)$	-4311.24	-4108.15		-3537.4		-3560.9		-3580.6	
Model dispersion	2.03	1.65		0.733		0.750		0.770	

7.1.11. PuDs via NDVi: Covariate parameter estimates and temporal representation

Table 14: PuDs via NDVi: Means of estimated marginal posterior distributions for covariates across M1 to M8 for PUD_{it} counts from 140 sub-catchments in the Lachlan over 73 months.

Lachlan River: PuDs & NDVi	**M1**		**M2**		**M3**		**M4**		**M3a**		**M5**		**M6**	**M7**		**M8**
Model form	Poisson		ZIP		ZINB		ZIP		ZINB		ZIP		ZINB	ZIP		ZINB
Random effects	none						ID REs							iCAR REs		
Temporal form	Month as a categorical factor + *f(year)* as a 7-knot smoother										*f(month-index)* as an 18-knot smoother					
(Intercept)	**-2.122**	-	**-1.3**	-			**-3.32**	-			**-3.461**	-		**-3.499**	-	
NDVi.std	**-0.193**	-	**-0.171**	-			**-0.104**	-			0.041	n.i.		0.038	n.i.	
Rain.std	-0.046	n.i.	-0.025	n.i.			0.028	n.i.			0.049	n.i.		0.052	n.i.	
MTemp.std	**-0.762**	-	**-0.826**	-			-0.131	n.i.			0.139	n.i.		0.167	n.i.	
PC_DIST_km.std	**-1.626**	-	**-1.605**	-			**-1.724**	-			**-1.716**	-		**-1.365**	-	
Dist2Road_DIST_km.std	-0.027	n.i.	-0.075	n.i.			-0.027	n.i.			-0.024	n.i.		-0.014	n.i.	
Min_mins.std	**-1.069**	-	**-0.91**	-			-0.414	n.i.			-0.417	n.i.		-0.396	n.i.	
DIWA_DIST_km.std	**-0.507**	-	**-0.432**	-			-0.23	n.i.			-0.221	n.i.		0.1	n.i.	
RAMSAR_DIST_km.std	**0.291**	+	**0.203**	+			0.149	n.i.			0.133	n.i.		-0.259	n.i.	
IBA_DIST_km.std	**0.345**	+	**0.328**	+			0.23	n.i.			0.239	n.i.		-0.088	n.i.	
CAPAD_DIST_km.std	**-0.175**	-	**-0.165**	-			-0.093	n.i.			-0.095	n.i.		0.039	n.i.	
CCArea_km2.std	**0.294**	+	**0.258**	+			0.4	n.i.			0.408	n.i.		0.319	n.i.	
IrrigAreaPc.std	0.04	n.i.	0.074	n.i.			0.003	n.i.			0.001	n.i.		0.033	n.i.	
NDViByIrrPc.std	-0.009	n.i.	-0.031	n.i.			0.042	n.i.			0.034	n.i.		0.037	n.i.	
AvgMthlyPop.std	**0.456**	+	**0.439**	+			**0.578**	+			**0.578**	+		**0.603**	+	
negbin_size_param																
zero_prob_param			0.462	+			0.227	+			0.238	+		0.237	+	
random_effect_variance							1.462	+			1.471	+		1.943	+	

Table 15: PuDs via NDVi: Means of estimated marginal posterior distributions for temporal covariates and smoothers across models M1 to M8 for PUD_{it} counts from 140 sub-catchments in the Lachlan River Region over 73 months.

Lachlan River: PuDs & NDVi	**M1**		**M2**		**M3**		**M4**		**M3a**			**M5**		**M6**	**M7**		**M8**
Model form	Poisson		ZIP		ZINB		ZIP		ZINB			ZIP		ZINB	ZIP		ZINB
Temporal form	**Month as a categorical factor + *f(year)* as a 7-knot smoother**										**smoothers**	***f(month-index)* as an 18-knot smoother**					
fmonthFeb	-0.187	n.i.	-0.24	n.i.			-0.033	n.i.			mindex.cr1	0.284	n.i.				
fmonthMar	**-0.682**	-	**-0.818**	-			-0.257	n.i.			mindex.cr2	0.186	n.i.				
fmonthApr	**-1.012**	-	**-1.169**	-			-0.206	n.i.			mindex.cr3	-0.271	-				
fmonthMay	**-1.551**	-	**-1.766**	-			-0.305	n.i.			mindex.cr4	0.301	n.i.				
fmonthJun	**-1.741**	-	**-1.993**	-			-0.223	n.i.			mindex.cr5	0.102	n.i.				
fmonthJul	**-2.134**	-	**-2.338**	-			-0.552	n.i.			mindex.cr6	-0.037	n.i.				
fmonthAug	**-1.845**	-	**-2.067**	-			-0.395	n.i.			mindex.cr7	-0.06	n.i.				
fmonthSep	**-1.036**	-	**-1.308**	-			-0.049	n.i.			mindex.cr8	-0.244	n.i.				
fmonthOct	-0.159	n.i.	-0.359	n.i.			**0.42**	**+**			mindex.cr9	-0.1	n.i.				
fmonthNov	**-0.313**	-	**-0.458**	-			0.064	n.i.			mindex.cr10	0.005	n.i.				
fmonthDec	**-0.55**	-	**-0.612**	-			**-0.405**	-			mindex.cr11	-0.576	-				
year.cr1	0.155	+	0.164	+			0.103	n.i.			mindex.cr12	-0.193	n.i.				
year.cr2	-0.174	-	-0.175	-			-0.119	n.i.			mindex.cr13	-0.091	n.i.				
year.cr3	-0.176	-	-0.18	-			-0.157	-			mindex.cr14	-0.275	n.i.				
year.cr4	-0.357	-	-0.289	-			-0.343	-			mindex.cr15	0.277	n.i.				
year.cr5	-0.258	-	-0.12	n.i.			-0.161	-			mindex.cr16	-0.241	n.i.				
year.cr6	-0.109	n.i.	-0.11	n.i.			-0.184	n.i.			mindex.cr17	-0.334	n.i.				

7.1.12. PuDs via NDVi: Modelling summary and plots of fitted values

Count data regression models were constructed to predict monthly Flickr-derived PuD counts per sub-catchment, initially with NDVi as the environmental condition proxy. As noted earlier, the PuD count dataset is dominated by zero counts (91%) and where monthly PuD counts do arise they are usually considerably lower than monthly BuD counts. These features make it more challenging to fit regression models to monthly PuD count data than was the case for monthly BuD count data.

Models M1, M2, M4, M5 and M7 were assessed for their suitability for modelling monthly PUD counts per sub-catchment in the Lachlan River Region when NDVi is used as a proxy for environmental condition. Results are shown in Table 13, Table 14 and Table 15.

Sub-catchment specific random terms are once again important for improving model fit (Table 13). The spatial covariates $PC\_DIST\_km_i$ and $AvgMthlyPop_i$ are consistently important for predicting PuD count irrespective of the form of the model and temporal representation (Table 14). The spatio-temporal covariate $NDVi_{it}$ appears to be negatively associated with PuD count when month as a factor and a 7-knot year smoother are used to represent cross-sub-catchment temporal variation (M4) (Table 14). This association disappears once the 18-knot month-index smoother provides cross sub-catchment temporal variation in the models (M5 and M7). For the ZIP model without sub-catchment specific random terms (M2), distributions of the month as factor terms suggest that PuD counts are higher in January than other months, with the exception of February and October (Table 15).

These results suggest that PuDs proxy a different visitation pattern to BuDs. Resident population in a sub-catchment is consistently positively associated with PuD count, and peak PuD counts arise during the summer holiday period. This was not the case for BuDs.

Predicted expected monthly PuD counts per sub-catchment, aggregated into total PuD counts per sub-catchment per year, from models M4 and M5 are compared with the Lachlan PUD count data in Figure 30 overleaf. The plots show that predictions from models M4 and M5 are very similar to each other and reasonably similar to the PUD count data, although neither of the models can fully predict peak PuD counts from hotspot sub-catchments. As was the case for the BuD models, most of the explanatory power comes from the sub-catchment specific random effects in combination with the (common across all sub-catchments) temporal representation. However, spatial covariates $PC\_DIST\_km_i$ and $AvgMnthlyPop_i$ contribute consistently to the fit of these models. The spatio-temporal covariates $NDVi_{it}$ has a negative association with PuD count when the temporal representation uses month as a factor plus a 7-knot year smoother (M4).

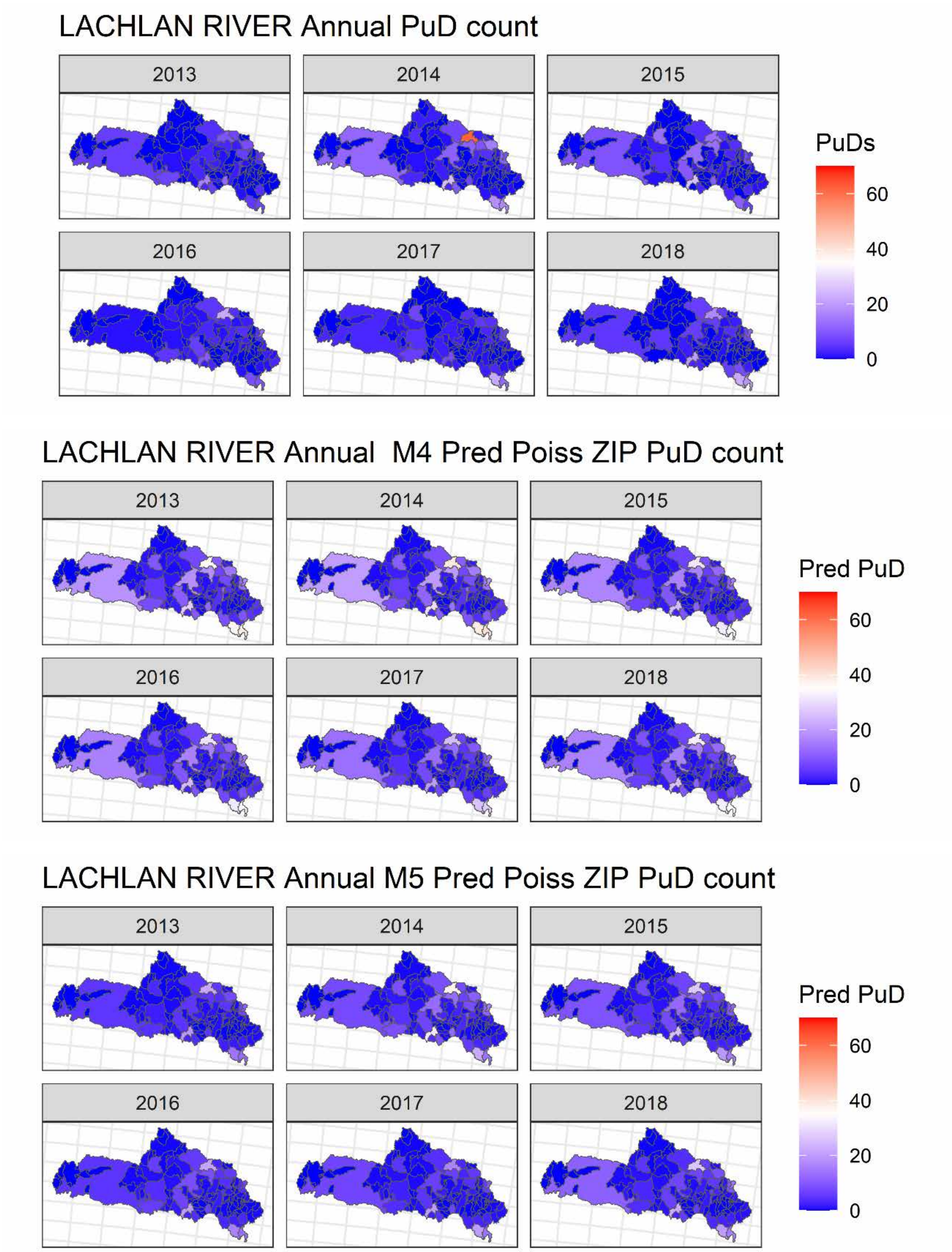

Figure 30: Comparison of total annual Flickr-derived PUD count data (top panel) with predicted annual PUD count (summed from predicted monthly PUD counts) from fitted models M4 (mid panel) and M5 (lower panel) when NDVi is used as a proxy for environmental condition.

7.1.13. PuDs via WatPc95: Model selection via fit simulations and tests for temporal and spatial autocorrelation among the residuals

Table 16: PuDs & WatPc95: Model form assessments and overall fit metrics across models M1 to M8 for PUD_{it} counts from 140 sub-catchments in the Lachlan River Region over 73 months.

Lachlan River: PuDs & WatPc95	**M1**	**M2**	**M2a***	**M4**	**M3a**	**M5**	**M6**	**M7**	**M8**
Model form	Poisson	ZIP	ZIP*	ZIP	ZINB	ZIP	ZINB	ZIP	ZINB
Random effects	none			ID REs				iCAR REs	
Temporal form	Month as a categorical factor + *f(year)* as a 7-knot smoother					*f(month-index)* as an 18-knot smoother			
Model form assessments									
Zero-inflation adequate?	no	yes	yes	yes		yes		yes	
Moran's I (% fitted points p < 0.05)	omitted	omitted	omitted	omitted		omitted		omitted	
Temporal ACF (% fitted points p < 0.05)	omitted	omitted	omitted	omitted		omitted		omitted	
Model fit metrics: all covariates									
Marginal likelihood	-2888.32	-2764.21	-2773.3	-2457.49		-2459.65		-2501.39	
DIC	5437.35	5204.96	5236.53	4452.53		4469.12		4535.25	
WAIC	5737.48	5282.3	5272.59	4486.04		4485.39		4564.41	
CPO via $-\sum_{i=1}^{n} ln(CPO_i)$	-2736.51	-2613.25	-2626.53	-2236.25		-2241.41		-2292.7	
Model dispersion	2.63	2.14	2.00	0.647		0.652		0.611	

Model M2a includes the 18-knot month-index smoother without any sub-catchment specific random effects.

7.1.14. PuDs via WatPc95: Covariate parameter estimates and temporal representation

Table 17: PuDs & WatPc95: Means of estimated marginal posterior distributions for covariates across M1 to M8 for PUD_{it} counts from 140 sub-catchments in the Lachlan over 73 months.

Lachlan River: PuDs & WatPc95	**M1**		**M2**		**M2a**		**M4**		**M3a**		**M5**		**M6**	**M7**		**M8**
Model form	Poisson		ZIP 7kt		ZIP 18kt		ZIP		ZINB		ZIP		ZINB	ZIP		ZINB
Random effects	none						ID REs							iCAR REs		
Temporal form	Month as a categorical factor + *f(year)* as a 7-knot smoother										*f(month-index)* as an 18-knot smoother					
(Intercept)	-2.524	-	-1.616	-	-2.428	-	-3.798	-			-3.695	-		-3.731	-	
WatPc95.std	0.049	n.i.	0.041	n.i.	0.057	n.i.	0.001	n.i.			0.007	n.i.		0.034	n.i.	
Rain.std	0.01	n.i.	0.031	n.i.	0.079	n.i.	0.088	n.i.			**0.13**	+		**0.137**	+	
MTemp.std	**-0.458**	-	**-0.600**	-	-0.226	n.i.	0.092	n.i.			0.228	n.i.		0.273	n.i.	
PC_DIST_km.std	**-1.799**	-	**-1.768**	-	**-1.771**	-	**-1.853**	-			**-1.857**	-		**-1.547**	-	
Dist2Road_DIST_km.std	0.120	n.i.	0.06	n.i.	0.07	n.i.	0.045	n.i.			0.048	n.i.		0.105	n.i.	
Min_mins.std	**-0.952**	-	**-0.799**	-	**-0.888**	-	-0.462	n.i.			-0.469	n.i.		-0.346	n.i.	
DIWA_DIST_km.std	**-0.516**	-	**-0.450**	-	**-0.408**	-	-0.336	n.i.			-0.325	n.i.		0.149	n.i.	
RAMSAR_DIST_km.std	**0.246**	+	**0.177**	+	**0.166**	+	0.213	n.i.			0.208	n.i.		-0.205	n.i.	
IBA_DIST_km.std	**0.284**	+	**0.253**	+	**0.276**	+	0.221	n.i.			0.224	n.i.		-0.036	n.i.	
CAPAD_DIST_km.std	**-0.227**	-	**-0.195**	-	**-0.202**	-	-0.071	n.i.			-0.075	n.i.		0.084	n.i.	
CCArea_km2.std	**0.272**	+	**0.260**	+	**0.276**	+	0.424	n.i.			0.428	n.i.		0.339	n.i.	
IrrigAreaPc.std	0.038	n.i.	0.099	n.i.	0.046	n.i.	0.044	n.i.			0.018	n.i.		0.032	n.i.	
NDViByIrrPc.std	-0.074	n.i.	-0.105	n.i.	-0.054	n.i.	0.017	n.i.			0.043	n.i.		0.045	n.i.	
AvgMthlyPop.std	**0.530**	+	**0.485**	+	**0.478**	+	**0.595**	+			**0.594**	+		**0.611**	+	
negbin_size_param																
zero_prob_param			0.476	+	0.482	+	0.238	+			0.237	+		0.236	+	
random_effect_variance							1.539	+			1.541	+		2.097	+	

Table 18: PuDs & WatPc95: Means of estimated marginal posterior distributions for temporal covariates and smoothers across models M1 to M8 for PUD_{it} counts from 140 sub-catchments in the Lachlan River Region over 73 months

Lachlan River: PuDs & WatPc95	**M1**		**M2**		**M3**		**M4**		**M3a**			**M5**		**M2a**	**M7**		**M8**
Model form	Poisson		ZIP		ZINB		ZIP		ZINB			ZIP		ZIP	ZIP		ZINB
Temporal form	**Month as a categorical factor + *f(year)* as a 7-knot smoother**										**smoothers**	***f(month-index)* as an 18-knot smoother**					
fmonthFeb	-0.107	n.i.	-0.136	n.i.			0.046	n.i.			mindex.cr1	0.122	n.i.	-	0.195	n.i.	
fmonthMar	**-0.362**	-	-0.542	-			-0.033	n.i.			mindex.cr2	0.056	n.i.	+	0.027	n.i.	
fmonthApr	**-0.648**	-	-0.829	-			0.056	n.i.			mindex.cr3	-0.473	-	-	-0.463	-	
fmonthMay	**-1.004**	-	-1.385	-			0.182	n.i.			mindex.cr4	0.709	+	n.i.	0.772	+	
fmonthJun	**-1.193**	-	-1.646	-			0.235	n.i.			mindex.cr5	-0.169	n.i.	n.i.	-0.216	n.i.	
fmonthJul	**-1.405**	-	-1.768	-			0.085	n.i.			mindex.cr6	-0.16	n.i.	-	-0.111	n.i.	
fmonthAug	**-1.342**	-	-1.75	-			0.108	n.i.			mindex.cr7	-0.013	n.i.	-	0.033	n.i.	
fmonthSep	**-0.909**	-	-1.279	-			0.184	n.i.			mindex.cr8	-0.332	n.i.	n.i.	-0.373	n.i.	
fmonthOct	-0.031	n.i.	-0.194	n.i.			0.536	+			mindex.cr9	-0.027	n.i.	-	0.006	n.i.	
fmonthNov	-0.164	n.i.	-0.297	n.i.			0.334	n.i.			mindex.cr10	0.077	n.i.	n.i.	0.108	n.i.	
fmonthDec	**-0.508**	-	-0.534	-			-0.379	-			mindex.cr11	-0.774	-	n.i.	-0.824	-	
year.cr1	**0.206**	**+**	-0.136	n.i.			0.211	+			mindex.cr12	0.006	n.i.	n.i.	0.073	n.i.	
year.cr2	-0.109	n.i.	-0.542	-			-0.086	n.i.			mindex.cr13	-0.379	-	n.i.	-0.377	-	
year.cr3	**-0.171**	-	-0.829	-			-0.055	n.i.			mindex.cr14	-0.614	-	n.i.	-0.658	-	
year.cr4	**-0.354**	-	-1.385	-			-0.352	-			mindex.cr15	0.279	n.i.	n.i.	0.362	n.i.	
year.cr5	-0.137	n.i.	-1.646	-			-0.13	n.i.			mindex.cr16	-0.412	-	n.i.	-0.431	-	
year.cr6	-0.061	n.i.	-1.768	-			-0.093	n.i.			mindex.cr17	-0.44	n.i.	n.i.	-0.472	-	

7.1.15. PuDs via WatPc95: Modelling summary and plots of fitted values

Count data regression models were also constructed to predict monthly Flickr-derived PuD counts per sub-catchment with WatPc95 as the environmental condition proxy. As noted earlier, the dataset is smaller when WatPc95 is used as the condition proxy because months in which cloud cover prevented at least 95% of a sub-catchment's area being visible are removed. The remaining data are heavily dominated by zero counts (91%).

Models M1, M2, M4, M5 and M7 were assessed for their suitability for modelling monthly PUD counts per sub-catchment in the Lachlan River Region when WatPc95 is used as a proxy for environmental condition. Results are shown in Table 16, Table 17 and Table 18.

Sub-catchment specific random terms again substantially improve model fit (Table 16). Spatial covariates $PC\_DIST\_km_i$ and $AvgMthlyPop_i$ are again consistently important for predicting PuD count irrespective of the form of the model and temporal representation (Table 17). The spatio-temporal covariate $Rain_{it}$ appears to be positively associated with PuD count when the 18-knot month-index smoother is used to represent cross-sub-catchment temporal variation (M5 and M7) (Table 17). This association disappears if the simpler temporal representation is used with sub-catchment specific random terms (M4). The month as factor terms again suggest that PuD counts are higher in January than other months, with the exceptions of February, October and November (Table 15).

As was the case when NDVi was used as the environmental condition proxy, these results suggest that PuDs proxy a different visitation pattern to BuDs. Resident population in a sub-catchment is consistently positively associated with PuD count, and peak PuD counts arise during the summer holiday period, but also during the spring. This latter feature was also present for BuDs.

Predicted expected monthly PuD counts per sub-catchment, aggregated into total PuD counts per sub-catchment per year, from models M4 and M5 are compared with the Lachlan PUD count data in Figure 31 overleaf. The plots show that predictions from models M4 and M5 are very similar to each other and reasonably similar to the PUD count data, although hotspots are again under-predicted. Most of the explanatory power comes from the sub-catchment specific random effects in combination with the (common across all sub-catchments) temporal representation. However, spatial covariates $PC\_DIST\_km_i$ and $AvgMnthlyPop_i$ contribute consistently to the fit of these models. The spatio-temporal covariate $Rain_{it}$ has a positive association with PuD count when the temporal representation uses an 18-knot month-index smoother (M5 & M7).

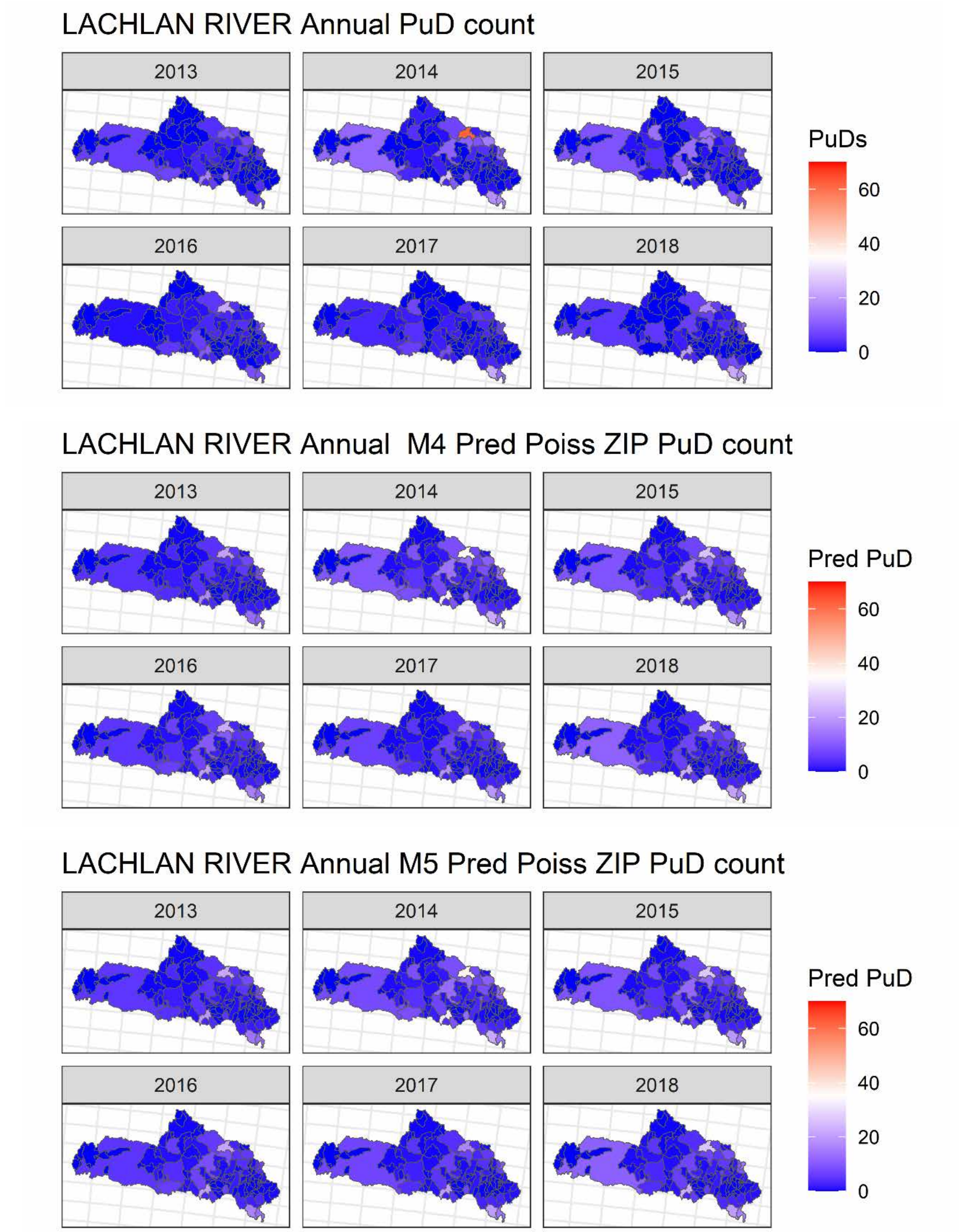

Figure 31: Comparison of total annual Flickr-derived PUD count data (top panel) with predicted annual PUD count (summed from predicted monthly PUD counts) from fitted models M4 (mid panel) and M5 (lower panel) when WatPc95 is used as a proxy for environmental condition.

7.2. Summary results across the Basin: BuDs as dependent variable, NDVi or WatPc95 as the environmental condition proxy

Modelling results from the Lachlan catchment in the previous subsection showed that count data models for eBird-derived BuDs can be developed[3], driven by spatio-temporal environmental variables (monthly NDVi, monthly percentage inundated area, total monthly rainfall, and monthly maximum temperature), sub-catchment specific variables (e.g., distances to the nearest town, travel times to major cities, sub-catchment area, resident population), and a cross-sub-catchment temporal term which tracks seasonality in visitation and between-year changes in usage of the eBird citizen science site.

Modelling to predict monthly BuD counts for sub-catchments in the Lachlan identified that sub-catchment specific random effects (i.e., sub-catchment specific intercept terms (see Figure 16)) will likely be required to assist in accurately predicting monthly counts for sub-catchment visitation hotspots at one extreme, whilst also being able to predict the many sub-catchments that report zero monthly counts. Zero-inflated models will almost certainly be required for all River Regions across the Basin to further assist in this. The best-fitting models for BuD counts in sub-catchments in the Lachlan combined sub-catchment specific random terms with non-parametric temporal smoothers. These two features deliver most of the explanatory power. However, sub-catchment specific non-temporal drivers such as distance from the nearest local population centre and the area of the sub-catchment both consistently provided extra explanatory capability to these already flexible models, almost irrespective of the model form and temporal representation used. The two environmental condition proxies, $NDVi_{it}$ and $WatPc95_{it}$, were often found to be associated with BuD counts, but the direction of this association (positive or negative) was seen to vary depending on the form of temporal smoother used in the model. This suggests that the environmental condition terms add explanatory power to the models, and this may well be by helping to further explain between-sub-catchment temporal variation in BuD count.

In this section models M4 and M5 are used to predict monthly BuD counts in 19 River Regions across the Basin[4], with $NDVi_{it}$ and $WatPc95_{it}$ introduced separately as a proxy for environmental condition. M4 and M5 both include sub-catchment specific random terms but differ in that M4 represents common-across-sub-catchment temporal variation via month as a factor term together with a 7-knot smoother on year (2013 – 2019), whereas M5 uses an 18-knot smoother on month-index (1 to 73) for the same purpose. The intention is to see whether $NDVi_{it}$ and $WatPc95_{it}$ appear consistently as positive or negative drivers of monthly BuD count for catchments across the Basin. Results are summarised in Table 19 overleaf and shown in Figure 32 to Figure 35 following.

---

[3] Count data regression models were similarly constructed for Flickr-derived PuDs; however the available dataset for PuDs contains only approximately one third of the number of data points in the eBird-derived BuDs dataset. Subsequent modelling uses BuDs as the dependent variable.

[4] Models are not constructed for the smaller Victorian sub-catchments on the southern edge of the Basin.

Table 19: Importance and direction of action of NDVi and WatPc95 as potential drivers of monthly BuD counts for River Regions across the Murray Darling Basin. Estimates obtained using count data regression models M4 and M5 (see Table 3 for further details).The + and – signs denote important positive and negative action associated with the environmental proxy concerned (i.e., the 95% credible interval for the parameter's posterior marginal distribution does not include zero); n.i. denotes 'not important'.

Model	**Model M4**		**Model M5**	
Environmental condition proxy	$NDVi_{it}$	$WatPc95_{it}$	$NDVi_{it}$	$WatPc95_{it}$
River Region				
Avon River-Tyrell Lake	n.i.	n.i.	n.i.	n.i.
Benanee-Willandra Creek	-	-	-	-
Billabong-Yanco Creeks	n.i.	+	n.i.	n.i.
Border Rivers	+	+	+	+
Castlereagh River	-	n.i.	n.i.	n.i.
Condamine-Culgoa Rivers	n.i.	n.i.	+	n.i.
Darling River	n.i.	n.i.	n.i.	n.i.
Gwydir River	n.i.	n.i.	n.i.	n.i.
Lachlan River	-	n.i.	+	+
Lower Mallee	n.i.	n.i.	+	n.i.
Lower Murray River	n.i.	n.i.	-	n.i.
Macquarie-Bogan Rivers	-	n.i.	n.i.	+
Moonie River	n.i.	n.i.	n.i.	n.i.
Murray Riverina	n.i.	+	+	+
Murrumbidgee River	-	+	n.i.	+
Namoi River	n.i.	n.i.	+	n.i.
Paroo River	-	n.i.	n.i.	n.i.
Upper Mallee	n.i.	+	+	+
Warrego River	-	n.i.	n.i.	n.i.

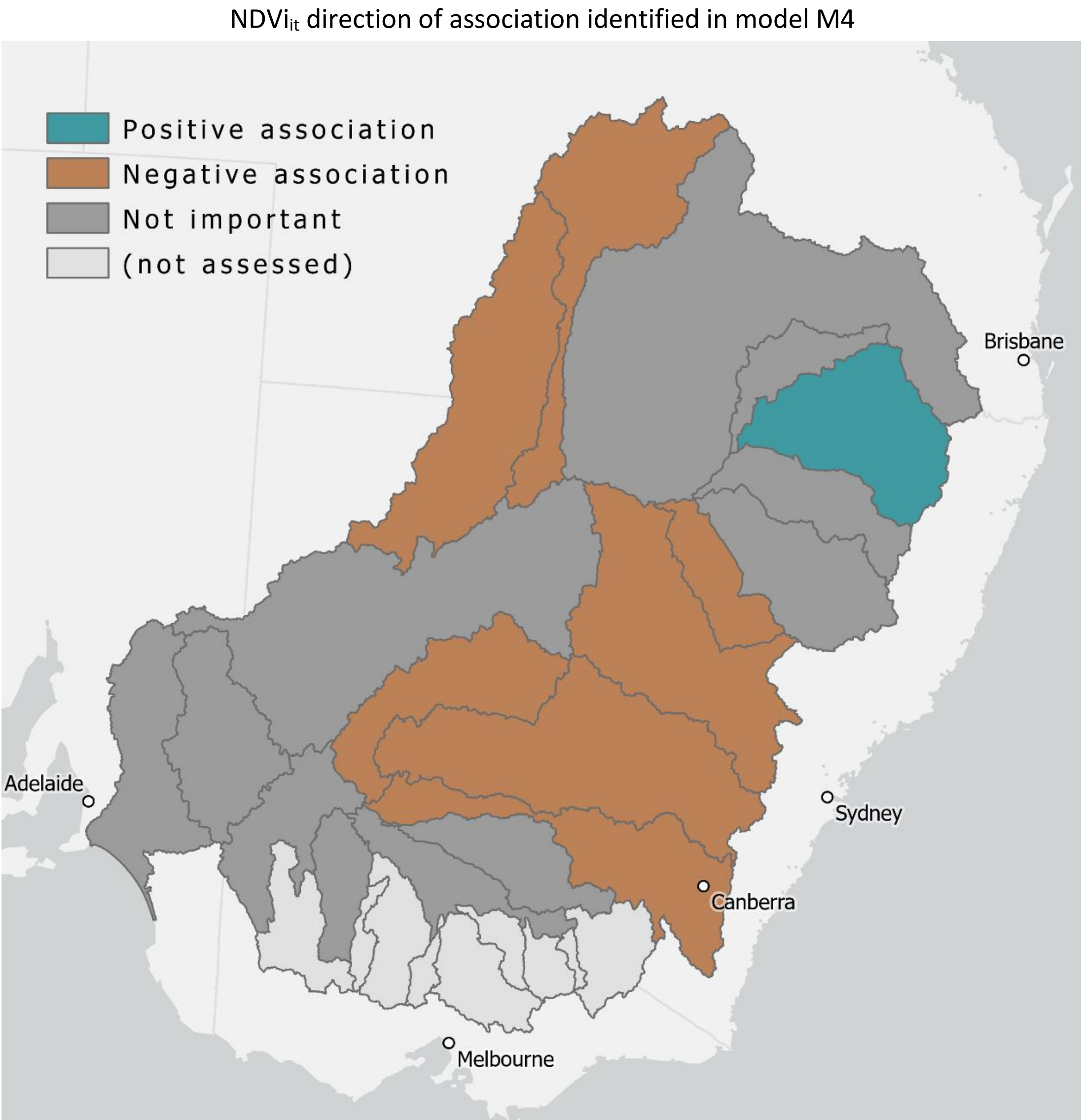

Figure 32: Direction of association between the $NDVi_{it}$ environmental condition proxy and BuD count as identified in count data regression model M4.

NDVi$_{it}$ direction of association identified in model M5

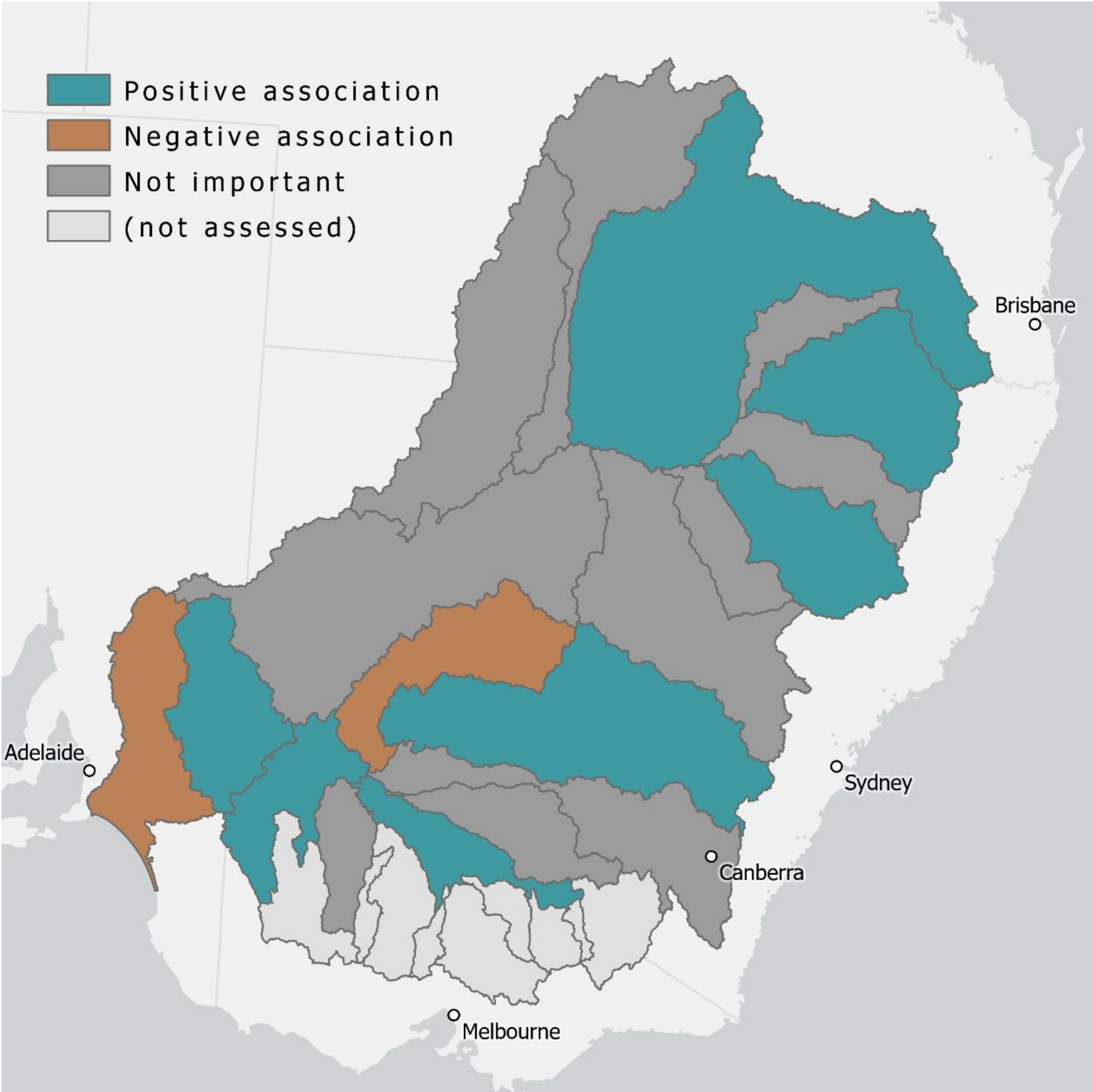

Figure 33: Direction of association between the $NDVi_{it}$ environmental condition proxy and BuD count as identified in count data regression model M5.

WatPc95$_{it}$ direction of association identified in model M4

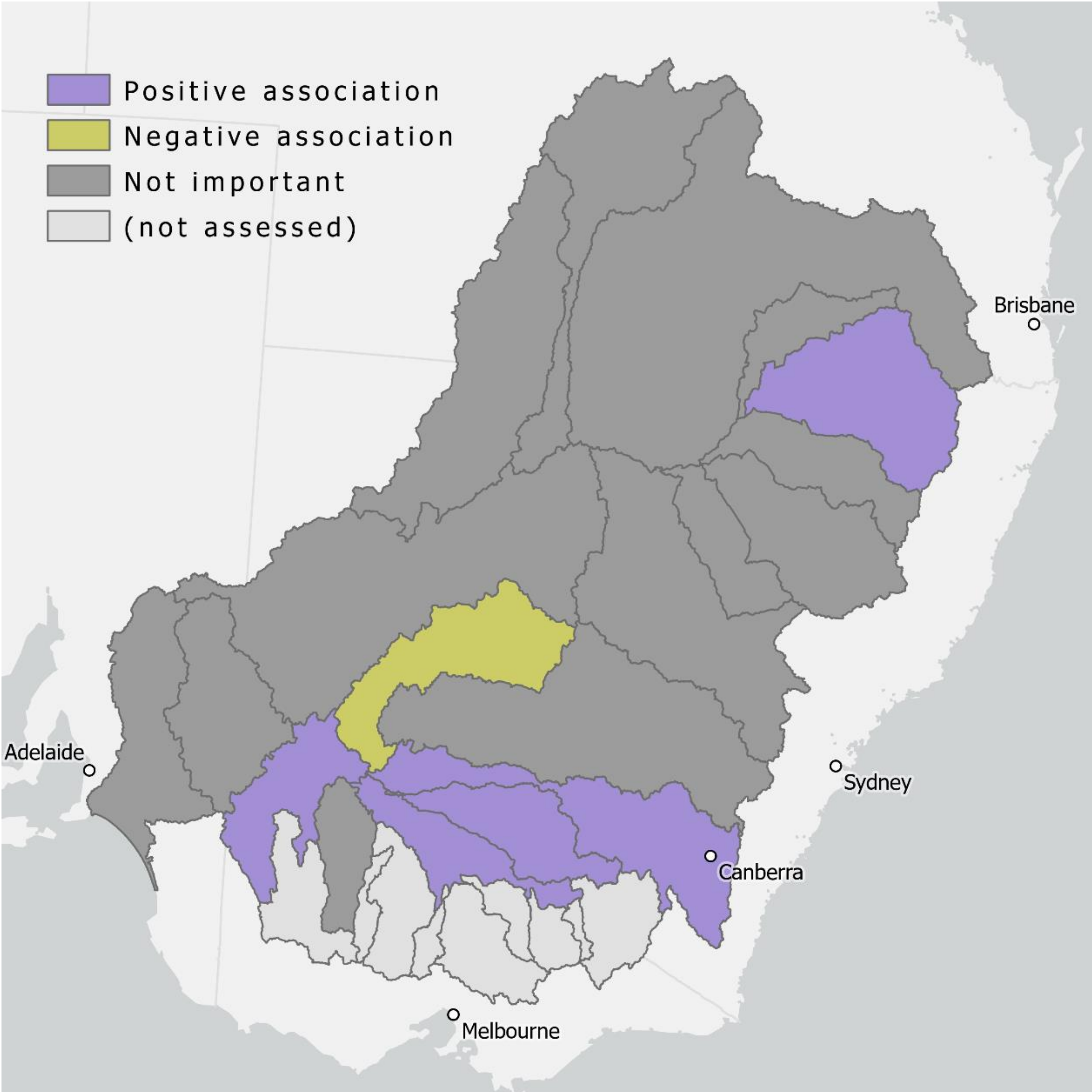

Figure 34: Direction of association between the $WatPc95_{it}$ environmental condition proxy and BuD count as identified in count data regression model M4.

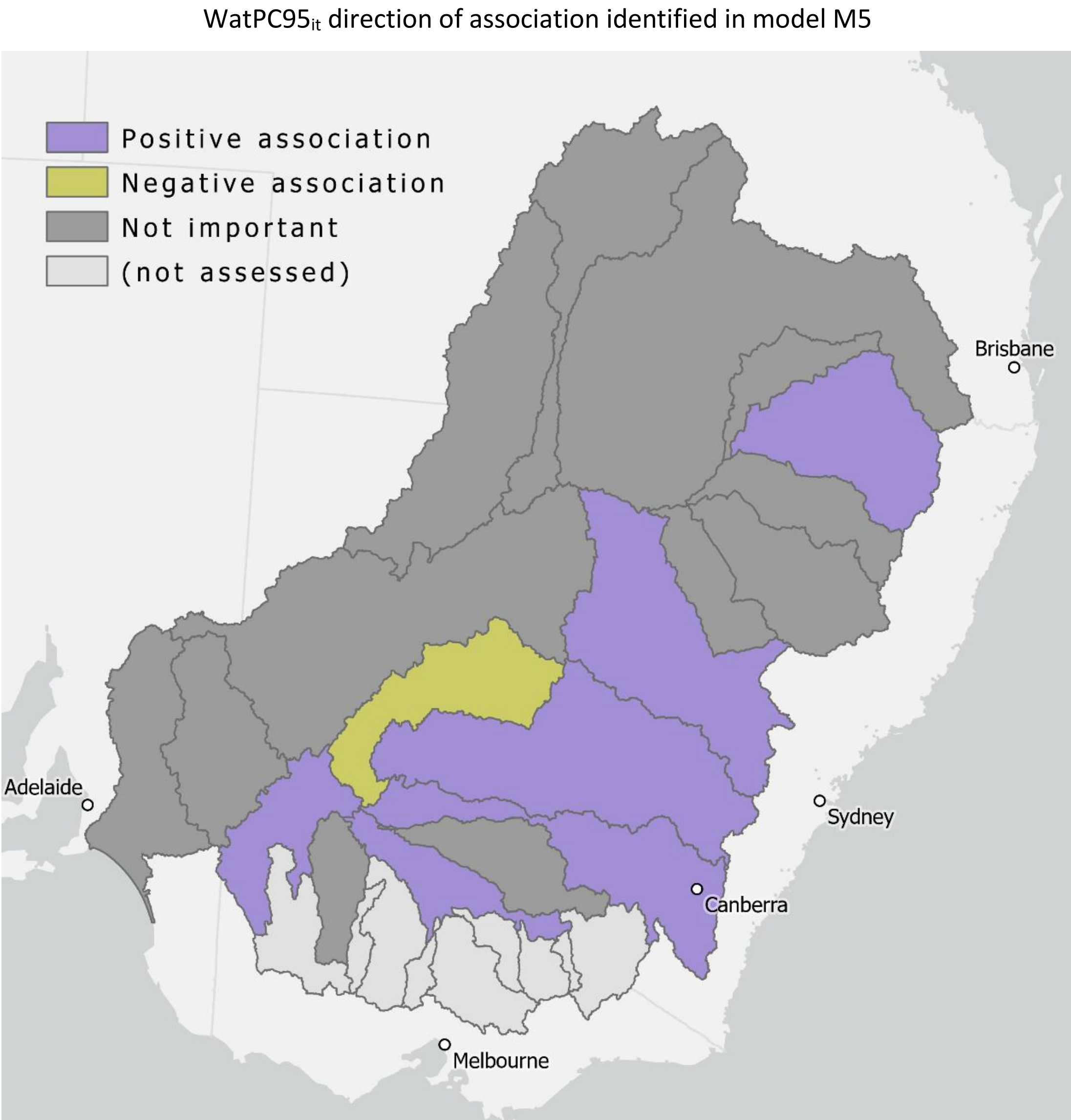

Figure 35: Direction of association between the $WatPc95_{it}$ environmental condition proxy and BuD count as identified in count data regression model M5.

8. Discussion

8.1. BuDs and PuDs as proxies for visitation

BuD data derived from citizen science bird species list postings on the eBird website (eBird, 2021; Sullivan et al., 2009) proved to be an excellent source of data on birding-related visitation across the MDB. PuD data derived from the photo posting site Flickr.com were also readily usable as VGI on visits to locations across the Basin. However, the two data sources clearly reflect different visitation communities as evidenced by seasonal differences in web posting intensity and different usage trajectories over the 2013 – 2019 data collection period, with BuD counts rising and PuD counts falling during this time. The user communities who generate these data are unlikely to be representative of the general population, but methodologies have been developed to allow wider metrics of visitation to be extrapolated from eBird BuD data when augmented by auxiliary surveys (Cameron and Kolstoe, 2022b).

8.2. NDVi and percentage inundated area as potential drivers of visitation rate

Count data regression models of monthly BuD and PuD counts for sub-catchments within River Regions across the Basin required sub-catchment specific random intercept terms and common-cross-sub-catchment temporal terms, together with capacity to handle zero inflation, to ensure that predicted monthly visitation counts were able to track low levels of visitation across many sub-catchments whilst also being able to accommodate localised visitation hotspots. The sub-catchment specific random terms and the temporal smoothers provided the majority of the model fit, to an extent that could be considered 'over fitting' if the data were drawn from a sample of sub-catchments rather than the full population of sub-catchments, as was the case here.

Most spatial and spatio-temporal covariates were discarded as unimportant once sub-catchment specific random terms were introduced. However, for models predicting BuD counts, distance to the nearest local population centre was always retained with an important negative association to BuD count, and the area of the sub-catchment was always retained with an important positive association to BuD count[5]. For models predicting PuD counts, distance to the nearest local population centre was always retained with an important negative association to PuD count, and resident population within the sub-catchment was also retained with an important positive association to PuD count. These relationships were consistent across models, suggesting that distance to the nearest local population centre is a strong inverse proxy for accessibility, catchment area is a strong positive proxy for 'visitation sampling effort' (for BuDs), and resident population is strongly positively associated with PuD visitation.

Associations between monthly NDVi and monthly percentage inundation area in a sub-catchment and monthly BuD counts for that sub-catchment are less clear. These terms were sometimes one of the few covariates retained in models for BuD count once sub-catchment specific random terms were introduced (Table 19 and Figure 32 to Figure 35), but when this happened the direction of the association between percentage inundation area and BuD count often varied when the form of temporal smoothing in the regression model was changed.

A positive association between percentage inundation area and monthly sub-catchment BuD count was robust across models M4 and M5 (with sub-catchment specific random effects but different temporal smoothers) for the Murrumbidgee, Murray Riverina, Upper Mallee and Border Rivers

[5] Models which normalised counts per unit catchment area (effectively normalising per unit sampling effort) were also constructed. These showed very similar performance and usage of covariates to the models described earlier in this report. Sub-catchment area was retained as a term in the regression so that potential interaction between NDVI and the percentage irrigated area in a sub-catchment could be explored. (No important interaction effects of this type were found.)

regions. Model M5, with sub-catchment random effects and the more flexible 1 to 73 month-index temporal smoother, identified positive association between percentage inundation area and BuD count for the Border Rivers, Macquarie-Bogan, Lachlan, Murrumbidgee, Murray Riverina and Upper Mallee regions, but the positive association became unimportant for the Macquarie-Bogan and Lachlan regions when model M4 (with random effects but a simpler smoother) was used (Figure 34 and Figure 35).

Positive association between average monthly NDVi and monthly BuD count were less robust to the change of temporal smoother between models M4 and M5 (Figure 32 and Figure 33). Only the Border Rivers region retained a positive association between NDVi and BuD count as these models were switched. The association between NDVi and BuD count switches from positive to negative for the Lachlan River region when the model is switched from M4 to M5. Uniquely, the Benanee-Willandra Creek region maintains a negative association between BuD counts and both NDVi and percentage inundation area across models M4 and M5.

Further investigation could usefully explore why associations between percentage inundation area and BuD count arise; are these River Region-wide interaction artefacts of the temporal modelling or are they meaningful effects associated with specific sub-catchments in a River Region? Given that sizeable changes in percentage inundation area are more likely to arise in sub-catchments in the mid and lower sections of a River Region it is plausible that this positive association could be meaningful when it is identified. This is particularly so because positive associations between percentage inundation area and BuD count occur for the Macquarie-Bogan, Lachlan, Murrumbidgee and Murray Riverina, all of which are major river systems with substantial areas of wetland in their lower sections. Increasing wetland inundation could attract large numbers and variety of birds which could then attract visits from birdwatchers.

8.3. Alternative metrics of environmental condition

Whilst percentage inundation area shows some potential for meaningful positive association with BuD counts in some River Regions, it would be useful to investigate the availability and performance of alternative metrics that might align more closely with environmental condition whilst also still being available across the whole of the Basin. Possible metrics from SEEA Ecosystem Accounts condition accounts and the SEEA Ecosystem Accounts biodiversity stock account could be useful. It might be helpful to explore the availability and performance of the remotely sensed condition metrics that were developed when CSIRO-led Ecosystem Accounts were produced for Gunbower, Pericoota, Koondrook.

The concluding part of the RQ12.2 Extension B (‘eBird’) research will explore whether inundation extent at a modest number of wetland-focused birdwatching sites across the Basin is associated with higher birdwatching visitation rates, as proxied by eBird BuDs. Inundation area at these sites is being accessed from Digital Earth Australia’s Water Observations from Space database (https://www.ga.gov.au/scientific-topics/dea/dea-data-and-products/dea-water-observations). It will be informative to see how count data regression models similar to those applied in this research perform when they are used to model visitation rate at a modest number of focused sites rather than targeted to predict BuD counts for all sub-catchments within large River Regions.

8.4. Alternative modelling options

Zero altered (hurdle) models in R-INLA allow covariates to be included in the ‘probability of non-zero count’ component. Currently R-INLA only allows the ‘probability of zero count’ term in the zero inflated models reported in this study to be represented as a constant, without any covariates. Zero altered (hurdle) models with covariates informing the ‘probability of a non-zero count’ could potentially enable a cleaner separation between covariates which affect the probability of a zero counts and covariates which act to increase or decrease counts at locations where non-zero counts occur.

Ongoing development regarding use of randomised quantile residuals for regression diagnostics in count data models, including zero inflated models, could add insight for model fitting (Bai et al., 2021; Dunn and Smyth, 1996; Feng et al., 2020). Leave group out cross validation (LGOCV) could also be explored as an alternative to CPO as a leave one out cross validation (LOOCV) approach, as this appears to offer advantages for assessing over-fitting in structured models like those used in this study (Adin et al., 2024).

9. Conclusion

The research undertaken in this study has found that eBird-derived BuDs show real promise as citizen science derived metrics of visitation rate that are available across the whole of the Basin. Cross calibration with visitation by the broader outdoor recreational community should be pursued as strong progress in this direction is already evident in the literature (Cameron and Kolstoe, 2022b; Langemeyer et al., 2023).

On reflection, a bigger challenge probably surrounds use of NDVi and percentage inundation area as proxies for environmental condition. The availability and performance of alternative metrics that are more representative of riparian environmental condition could usefully be explored. Remotely sensed condition metrics developed for use in SEEA Ecosystem Accounts could be worth considering(e.g., Harwood et al., 2023).

The concluding part of the RQ12.2 Extension B ('eBird') research will use the regression methodologies developed during this study to explore whether environmental condition at individual birdwatching hotspots across the Basin, as proxied by surface water area from Digital Earth Australia's Water Observations from Space, affects birder visitation rate (proxied by BuDs). Analysis will need to control adequately for additional factors such as migratory bird presence, species diversity and rarity (data on these attributes are available via the species lists posted on eBird).

Possibilities should also be explored for linking predicted increases in visitation rate at sub-catchment resolution from the modelling developed in this study with Tourism Research Australia's LGA-resolution survey-derived data on tourism expenditures. Most LGAs across the Basin have reasonably complete data coverage of overnight expenditures. Coverage of day trip and other components of expenditure is sparse in less populated parts of the Basin. Ideally, results – or some pointers towards an approach to link changes in local visitation rates to changes in tourism expenditures – can be included in the final report deliverable from WERP Theme 4 RQ12.2 (Extension A Final Report – delivery to MDBA on 17th October).

10. References

Adin, A., Krainski, E.T., Lenzi, A., Liu, Z., Martínez-Minaya, J., Rue, H., 2024. Automatic cross-validation in structured models: Is it time to leave out leave-one-out? Spat Stat 62, 100843. https://doi.org/https://doi.org/10.1016/j.spasta.2024.100843

Aither, 2022. Murray-Darling Basin Social and Economic Conditions Report.

Andersen, M.S., Levin, G., Odgaard, M.V., 2019. Economic benefits of reducing agricultural N losses to coastal waters for seaside recreation and real estate value in Denmark. Mar Pollut Bull 140, 146–156. https://doi.org/https://doi.org/10.1016/j.marpolbul.2019.01.010

Anselin, L., 1988. Spatial Econometrics. Kluwer Academic Publishers, Dordrecht.

Australian Bureau of Statistics, 2022. Gross Domestic Product - Production approach [WWW Document]. URL https://www.abs.gov.au/statistics/detailed-methodology-information/concepts-sources-methods/australian-system-national-accounts-concepts-sources-and-methods/2020-21/chapter-9-gross-domestic-product-production-approach-gdpp. (accessed 1.5.23).

Bai, W., Dong, M., Li, L., Feng, C., Xu, W., 2021. Randomized quantile residuals for diagnosing zero-inflated generalized linear mixed models with applications to microbiome count data. BMC Bioinformatics 22, 564. https://doi.org/10.1186/s12859-021-04371-6

Bhatt, P., Pickering, C.M., 2022. Destination image of Chitwan National Park, Nepal: Insights from a content analysis of online photographs. Journal of Outdoor Recreation and Tourism 37. https://doi.org/10.1016/j.jort.2022.100488

Camacho-Valdez, V., Saenz-Arroyo, A., Ghermandi, A., Navarrete-Gutiérrez, D.A., Rodiles-Hernández, R., 2020. Spatial analysis, local people’s perception and economic valuation of wetland ecosystem services in the Usumacinta floodplain, Southern Mexico. PeerJ 2020. https://doi.org/10.7717/peerj.8395

Cameron, A.C., Trivedi, P.K., 2013a. Regression analysis of count data. Cambridge University Press, Cambridge, U.K.

Cameron, A.C., Trivedi, P.K., 2013b. Regression analysis of count data. Cambridge University Press, Cambridge, U.K.

Cameron, A.C., Trivedi, P.K., 2005. Microeconometrics: methods and applications, Cambridge University. Cambridge University Press.

Cameron, T.A., Kolstoe, S.H., 2022a. Using Auxiliary Population Samples for Sample-Selection Correction in Models Based on Crowd-Sourced Volunteered Geographic Information. Land Econ 98, 1–21. https://doi.org/10.3368/le.98.1.040720-0050R1

Cameron, T.A., Kolstoe, S.H., 2022b. Using Auxiliary Population Samples for Sample-Selection Correction in Models Based on Crowd-Sourced Volunteered Geographic Information. Land Econ 98, 1–21. https://doi.org/10.3368/le.98.1.040720-0050R1

Chib, S., 1995. Marginal Likelihood from the Gibbs Output. J Am Stat Assoc 90, 1313–1321. https://doi.org/10.1080/01621459.1995.10476635

Chib, S., Jeliazkov, I., 2001. Marginal Likelihood From the Metropolis–Hastings Output. J Am Stat Assoc 96, 270–281. https://doi.org/10.1198/016214501750332848

Ciesielski, M., Stereńczak, K., 2021. Using Flickr data and selected environmental characteristics to analyse the temporal and spatial distribution of activities in forest areas. For Policy Econ 129. https://doi.org/10.1016/j.forpol.2021.102509

Cleveland, W.S., 1993. Visualizing data. Hobart press.

Cressie, N., Kapat, P., 2008. Some Diagnostics for Markov Random Fields. Journal of Computational and Graphical Statistics 17, 726–749. https://doi.org/10.1198/106186008X340760

Dunn, P.K., Smyth, G.K., 1996. Randomized Quantile Residuals. Journal of Computational and Graphical Statistics 5, 236–244. https://doi.org/10.1080/10618600.1996.10474708

eBird, 2021. eBird: An online database of bird distribution and abundance [web application]. eBird, Cornell Lab of Ornithology, Ithaca, New York. (Accessed: February 2024)) [WWW Document].

Feng, C., Li, L., Sadeghpour, A., 2020. A comparison of residual diagnosis tools for diagnosing regression models for count data. BMC Med Res Methodol 20, 175. https://doi.org/10.1186/s12874-020-01055-2

Gelman, A., Hwang, J., Vehtari, A., 2014. Understanding predictive information criteria for Bayesian models. Stat Comput 24, 997–1016. https://doi.org/10.1007/s11222-013-9416-2

Ghermandi, A., 2022a. Geolocated social media data counts as a proxy for recreational visits in natural areas: A meta-analysis. J Environ Manage 317. https://doi.org/10.1016/j.jenvman.2022.115325

Ghermandi, A., 2022b. Geolocated social media data counts as a proxy for recreational visits in natural areas: A meta-analysis. J Environ Manage 317. https://doi.org/10.1016/j.jenvman.2022.115325

Gómez-Rubio, V., 2020. Bayesian Inference with INLA. Chapman & Hall/CRC Press. , Boca Raton, Florida, USA.

Goodchild, M.F., 2007. Citizens as sensors: the world of volunteered geography. GeoJournal 69, 211–221. https://doi.org/10.1007/s10708-007-9111-y

Guilfoos, T., Thomas, P., Kolstoe, S., 2023. Estimating habit-forming and variety-seeking behavior: Valuation of recreational birdwatching. Am J Agric Econ. https://doi.org/10.1111/ajae.12422

Harwood, T., Lehmann, E., Giljohann, K., Williams, K., Liu, N., Ferrier, S., Ware, C., Donohue, R., Sivanandam, P., Malley, C., Schmidt, B., 2023. 9-arcsecond gridded HCAS 2.3 (2001-2018) base model estimation of habitat condition and general connectivity for terrestrial biodiversity, ecosystem site condition, annual epochs and 18-year trends for continental Australia.

Hausmann, A., Toivonen, T., Fink, C., Heikinheimo, V., Tenkanen, H., Butchart, S.H.M., Brooks, T.M., Di Minin, E., 2019. Assessing global popularity and threats to Important Bird and Biodiversity Areas using social media data. Science of the Total Environment 683, 617–623. https://doi.org/10.1016/j.scitotenv.2019.05.268

Huang, B., Young, M.A., Carnell, P.E., Conron, S., Ierodiaconou, D., Macreadie, P.I., Nicholson, E., 2020. Quantifying welfare gains of coastal and estuarine ecosystem rehabilitation for recreational fisheries. Science of The Total Environment 710, 134680. https://doi.org/https://doi.org/10.1016/j.scitotenv.2019.134680

Jung, R.C., Glaser, S., 2022. Modelling and Diagnostics of Spatially Autocorrelated Counts. Econometrics. https://doi.org/10.3390/econometrics10030031

Kleiber, C., Zeileis, A., 2008. Applied econometrics with R. Springer Science & Business Media.

Kolstoe, S., Cameron, T.A., 2017. The Non-market Value of Birding Sites and the Marginal Value of Additional Species: Biodiversity in a Random Utility Model of Site Choice by eBird Members. Ecological Economics 137, 1–12. https://doi.org/10.1016/j.ecolecon.2017.02.013

Langemeyer, J., Ghermandi, A., Keeler, B., van Berkel, D., 2023. The future of crowd-sourced cultural ecosystem services assessments. Ecosyst Serv 60, 101518. https://doi.org/https://doi.org/10.1016/j.ecoser.2023.101518

LeSage, J.P., Pace, R.K., 2014. The Biggest Myth in Spatial Econometrics. Econometrics. https://doi.org/10.3390/econometrics2040217

Lindgren, F., Rue, H., Lindström, J., 2011. An Explicit Link between Gaussian Fields and Gaussian Markov Random Fields: The Stochastic Partial Differential Equation Approach. J R Stat Soc Series B Stat Methodol 73, 423–498. https://doi.org/10.1111/j.1467-9868.2011.00777.x

Moraga, P., 2023. Spatial Statistics for Data Science: Theory and Practice with R. . Chapman & Hall/CRC Data Science Series.

Moraga, P., 2019. Geospatial Health Data: Modeling and Visualization with R-INLA and Shiny. . Chapman & Hall/CRC Biostatistics Series.

Moran, P.A.P., 1950. Notes on Continuous Stochastic Phenomena. Biometrika 37, 17–23. https://doi.org/10.2307/2332142

Norman, P., Pickering, C.M., 2019. Factors influencing park popularity for mountain bikers, walkers and runners as indicated by social media route data. J Environ Manage 249. https://doi.org/10.1016/j.jenvman.2019.109413

Norman, P., Pickering, C.M., Castley, G., 2019. What can volunteered geographic information tell us about the different ways mountain bikers, runners and walkers use urban reserves? Landsc Urban Plan 185, 180–190. https://doi.org/10.1016/j.landurbplan.2019.02.015

Pettit, L.I., 1990. The Conditional Predictive Ordinate for the Normal Distribution. Journal of the Royal Statistical Society: Series B (Methodological) 52, 175–184. https://doi.org/https://doi.org/10.1111/j.2517-6161.1990.tb01780.x

Rue, H., Martino, S., Chopin, N., 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. J R Stat Soc Series B Stat Methodol 71, 319–392.

Rue, H., Riebler, A., Sørbye, S.H., Illian, J.B., Simpson, D.P., Lindgren, F.K., 2017. Bayesian computing with INLA: a review. Annu Rev Stat Appl 4, 395–421.

Sinclair, M., Ghermandi, A., Moses, S.A., Joseph, S., 2019a. Recreation and environmental quality of tropical wetlands: A social media based spatial analysis. Tour Manag 71, 179–186. https://doi.org/10.1016/j.tourman.2018.10.018

Sinclair, M., Ghermandi, A., Moses, S.A., Joseph, S., 2019b. Recreation and environmental quality of tropical wetlands: A social media based spatial analysis. Tour Manag 71, 179–186. https://doi.org/10.1016/j.tourman.2018.10.018

Sinclair, M., Mayer, M., Woltering, M., Ghermandi, A., 2020a. Using social media to estimate visitor provenance and patterns of recreation in Germany's national parks. J Environ Manage 263. https://doi.org/10.1016/j.jenvman.2020.110418

Sinclair, M., Mayer, M., Woltering, M., Ghermandi, A., 2020b. Valuing nature-based recreation using a crowdsourced travel cost method: A comparison to onsite survey data and value transfer. Ecosyst Serv 45. https://doi.org/10.1016/j.ecoser.2020.101165

Sinclair, M., Vishnu Sagar, M.K., Knudsen, C., Sabu, J., Ghermandi, A., 2021. Economic appraisal of ecosystem services and restoration scenarios in a tropical coastal Ramsar wetland in India. Ecosyst Serv 47. https://doi.org/10.1016/j.ecoser.2020.101236

Song, J.J., De Oliveira, V., 2012. Bayesian model selection in spatial lattice models. Stat Methodol 9, 228–238. https://doi.org/https://doi.org/10.1016/j.stamet.2011.01.003

Sonter, L.J., Watson, K.B., Wood, S.A., Ricketts, T.H., 2016. Spatial and Temporal Dynamics and Value of Nature-Based Recreation, Estimated via Social Media. PLoS One 11, e0162372.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P., Van Der Linde, A., 2002. Bayesian measures of model complexity and fit. J R Stat Soc Series B Stat Methodol 64, 583–639. https://doi.org/https://doi.org/10.1111/1467-9868.00353

Stithou, M., Hynes, S., Hanley, N., Campbell, D., 2012. Estimating the value of achieving "good ecological status" in the Boyne River Catchment in Ireland using choice experiments. Econ Soc Rev (Irel) 43, 397–422.

Sullivan, B.L., Aycrigg, J.L., Barry, J.H., Bonney, R.E., Bruns, N., Cooper, C.B., Damoulas, T., Dhondt, A.A., Dietterich, T., Farnsworth, A., Fink, D., Fitzpatrick, J.W., Fredericks, T., Gerbracht, J., Gomes, C., Hochachka, W.M., Iliff, M.J., Lagoze, C., La Sorte, F.A., Merrifield, M., Morris, W., Phillips, T.B., Reynolds, M., Rodewald, A.D., Rosenberg, K. V, Trautmann, N.M., Wiggins, A., Winkler, D.W., Wong, W.-K., Wood, C.L., Yu, J., Kelling, S., 2014. The eBird enterprise: An integrated approach to development and application of citizen science. Biol Conserv 169, 31–40. https://doi.org/https://doi.org/10.1016/j.biocon.2013.11.003

Sullivan, B.L., Wood, C.L., Iliff, M.J., Bonney, R.E., Fink, D., Kelling, S., 2009. eBird: A citizen-based bird observation network in the biological sciences. Biol Conserv 142, 2282–2292. https://doi.org/https://doi.org/10.1016/j.biocon.2009.05.006

Teles da Mota, V., Pickering, C., 2021. Assessing the popularity of urban beaches using metadata from social media images as a rapid tool for coastal management. Ocean Coast Manag 203. https://doi.org/10.1016/j.ocecoaman.2021.105519

Teles da Mota, V., Pickering, C., 2020. Using social media to assess nature-based tourism: Current research and future trends. Journal of Outdoor Recreation and Tourism 30. https://doi.org/10.1016/j.jort.2020.100295

Teles da Mota, V., Pickering, C., Chauvenet, A., 2022. Popularity of Australian beaches: Insights from social media images for coastal management. Ocean Coast Manag 217. https://doi.org/10.1016/j.ocecoaman.2021.106018

Tenkanen, H., Di Minin, E., Heikinheimo, V., Hausmann, A., Herbst, M., Kajala, L., Toivonen, T., 2017. Instagram, Flickr, or Twitter: Assessing the usability of social media data for visitor monitoring in protected areas. Sci Rep 7. https://doi.org/10.1038/s41598-017-18007-4

Toivonen, T., Heikinheimo, V., Fink, C., Hausmann, A., Hiippala, T., Järv, O., Tenkanen, H., Di Minin, E., 2019. Social media data for conservation science: A methodological overview. Biol Conserv 233, 298–315. https://doi.org/10.1016/j.biocon.2019.01.023

Tourism Research Australia, 2022. National Visitor Survey: methodology [WWW Document]. URL https://www.tra.gov.au/domestic/national-visitor-survey-methodology/national-visitor-survey-nvs-methodology (accessed 1.9.23).

Watanabe, S., 2013. A widely applicable Bayesian information criterion. The Journal of Machine Learning Research 14, 867–897.

Wilkins, E.J., Howe, P.D., Smith, J.W., 2021a. Social media reveal ecoregional variation in how weather influences visitor behavior in U.S. National Park Service units. Sci Rep 11, 2403. https://doi.org/10.1038/s41598-021-82145-z

Wilkins, E.J., Wood, S.A., Smith, J.W., 2021b. Uses and Limitations of Social Media to Inform Visitor Use Management in Parks and Protected Areas: A Systematic Review. Environ Manage 67, 120–132. https://doi.org/10.1007/s00267-020-01373-7

Zuur, A.F., Ieno, E.N., 2018. Beginner's Guide to Spatial, Temporal, and Spatio-Temporal Ecological Data Analysis with R-INLA. Volume II: GAM and zero-inflated models. Highland Statistics Ltd., Newburgh, United Kingdom.

Zuur, A.F., Ieno, E.N., Saveliev, A.A., 2017. Beginner’s Guide to Spatial, Temporal, and Spatial-Temporal Ecological Data Analysis with R-INLA. Volume I: Using GLM and GLMM. Highland Statistics Ltd., Newburgh, United Kingdom.

Zuur, A.F., Ieno, E.N., Walker, N.J., Saveliev, A.A.,, Smith, G.M., 2009. Mixed effects models and extensions in ecology with R : . Springer., New York, USA.

11. Appendix 1: Authorisation to use eBird data

Re: Request to use eBird data on a second research project

eBird <help@ebird.org>

Wed 27/03/2024 07:11

To:

Jim Smart <j.smart@griffith.edu.au>

Hi Jim, thank you for notifying us of your use of eBird data for an additional research purpose. This email represents written approval from the Cornell Lab of Ornithology for the use of eBird data in both your initial project and your new project for the Murray Darling Basin. Please let us know if you have any additional questions or if we can be of further assistance.

We wish you all the best in your research.

Sincerely,

Jenna

Jenna Curtis

eBird Project Leader

Cornell Lab of Ornithology

12. Appendix 2: Data source meta data

Group	File	Quick Description	Origin
Flickr 'Photo User Days'	a_PuDs_x11320	Flickr Photo User Days by 11320m grid (2013_01 - 2019_03).	Flickr, 2021
	a_PuDs_xCC	Flickr Photo User Days by contracted catchment (2013_01 - 2019_03).	Flickr, 2021
	a_PuDs_xRR	Flickr Photo User Days by river region (2013_01 - 2019_03).	Flickr, 2021
eBird 'Bird User Days'	a_BuDs_x11320	eBird Bird User Days by 11320m grid (2013_01 - 2019_03).	Cornell Lab of Ornithology, 2023
	a_BuDs_xCC	eBird Bird User Days by contracted catchment (2013_01 - 2019_03).	Cornell Lab of Ornithology, 2023
	a_BuDs_xRR	eBird Bird User Days by river region (2013_01 - 2019_03).	Cornell Lab of Ornithology, 2023
Environmental and population data (time varying)	b_Inundation_xCC	Inundation (m^2) in contracted catchments by water area, clear (or dry) area, and bad (or clouded) area, as well as various percentage and threshold metrics (2013_03 - 2019_03).	Gould et al, 2023
	b_Inundation_xRR	Inundation (m^2) in river regions by water area, clear (or dry) area, and bad (or clouded) area, as well as various percentage metrics (2013_03 - 2019_03).	Gould et al, 2023
	b_MaxTemp_x11320	Average monthly max temperature (°C) by 11320m grid (2012_12 - 2019_03).	Queensland Government, 2023
	b_MaxTemp_xCC	Average monthly max temperature (°C) by contracted catchment (2012_12 - 2019_03).	Queensland Government, 2023
	b_MaxTemp_xRR	Average monthly max temperature (°C) by river region (2012_12 - 2019_03).	Queensland Government, 2023
	b_NDVI_x11320	Average monthly NDVI by 11320m grid (2013_01 - 2019_03).	Australian Bureau of Meteorology, 2014
	b_NDVI_xCC	Average monthly NDVI by contracted catchment (2013_01 - 2019_03).	Australian Bureau of Meteorology, 2014
	b_NDVI_xRR	Average monthly NDVI by river region (2013_01 - 2019_03).	Australian Bureau of Meteorology, 2014
	b_Population_x11320	Annual population estimates summed using ABS $1km^2$	Australian Bureau of

		population grid by 11320m grid (2013-2019).	Statistics 2018a, 2018b, 2019, 2020
	b_Population_xCC	Annual population estimates summed using ABS 1km^2 population grid by contracted catchment (2013-2019).	Australian Bureau of Statistics 2018a, 2018b, 2019, 2020
	b_Population_xRR	Annual population estimates summed using ABS 1km^2 population grid by river region (2013-2019).	Australian Bureau of Statistics 2018a, 2018b, 2019, 2020
	b_Rainfall_x11320	Average total monthly rainfall (mm) by 11320m grid (2012_12 - 2019_03).	Queensland Government, 2023
	b_Rainfall_xCC	Average total monthly rainfall (mm) by contracted catchment (2012_12 - 2019_03).	Queensland Government, 2023
	b_Rainfall_xRR	Average total monthly rainfall (mm) by river region (2012_12 - 2019_03).	Queensland Government, 2023
Location data (time invariant)	c_Distance_2_Feature_x11320	Euclidean distance (m) to nearest given feature (population centre (PC), CAPAD, DIWA, watercourse (WCL), IBA, RAMSAR) from 11320m grid centroid.	Australian Bureau of Statistics, 2016; Australian Government, 2010, 2015 2022; Geoscience Australia, 2006; Birds Australia, 2009.
	c_Distance_2_Feature_xCC	Euclidean distance (m) to nearest given feature (population centre (PC), CAPAD, DIWA, watercourse (WCL), IBA, RAMSAR) from contracted catchment geometric centroid.	Australian Bureau of Statistics, 2016; Australian Government, 2010, 2015 2022; Geoscience Australia, 2006; Birds Australia, 2009.
	c_Distance_2_Feature_xCC_FWMC	Euclidean distance (m) to nearest given feature (population centre (PC), CAPAD, DIWA, watercourse (WCL), IBA, RAMSAR) from photo user day weighted mean centre by contracted catchment.	Australian Bureau of Statistics, 2016; Australian Government, 2010, 2015 2022; Geoscience Australia, 2006; Birds Australia, 2009;
	c_Distance_2_Road_x11320	Euclidean distance (m) to nearest road type combination from 11320m grid centroid.	Geoscience Australia, 2006

	c_Distance_2_Road_xCC	Euclidean distance (m) to nearest road type combination from contracted catchment geometric centroid.	Geoscience Australia, 2006
	c_Distance_2_Road_xCC_FWMC	Euclidean distance (m) to nearest road type combination from photo user day weighted mean centre by contracted catchment.	Geoscience Australia, 2006
	c_DistanceTime_2_City_x11320	Road distance (km) and travel time (mins) by road between 11320m grid centroid and selected major cities.	Geoscience Australia, 2006
	c_DistanceTime_2_City_xCC	Road distance (km) and travel time (mins) by road between contracted catchment geometric centroid and selected major cities.	Geoscience Australia, 2006
	c_DistanceTime_2_City_xCC_FWMC	Road distance (km) and travel time (mins) by road between photo user day weighted mean centre by contracted catchment and selected major cities.	Geoscience Australia, 2006
	c_MDB_Region_x11320	Various region locations (SWSDL, SRA Valley, IBRA) of 11320 grid centroid.	Murray-Darling Basin Authority, 2013, 2015; Australian Government, 2020.
	c_MDB_Region_xCC	Various region locations (SWSDL, SRA Valley, IBRA) of contracted catchment geometric centroid	Murray-Darling Basin Authority, 2013, 2015; Australian Government, 2020.
	c_MDB_Region_xCC_FWMC	Various region locations (SWSDL, SRA Valley, IBRA) of photo user day weighted mean centre by contracted catchment.	Murray-Darling Basin Authority, 2013, 2015; Australian Government, 2020.

Appendix 2: References

Australian Bureau of Meteorology, 2014. NDVI (Normalised Difference Vegetation Index) – High-resolution gridded monthly NDVI dataset. Accessed August 2022, [Online]. Available at: http://www.bom.gov.au/metadata/catalogue/19115/ANZCW0503900404?template=full

Australian Bureau of Statistics, 2016. Australian Statistical Geography Standard (ASGS 2016). Accessed June 2023. [Online] Available at: https://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/1270.0.55.001July%202016?OpenDocument

Australian Bureau of Statistics, 2018a. Australian Population Grid 2006 to 2016. Accessed August 2023, [Online]. Available at: https://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/3218.02016

Australian Bureau of Statistics, 2018b. Australian Population Grid 2017. Accessed August 2023, [Online]. Available at: https://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/3218.02016-17?OpenDocument

Australian Bureau of Statistics, 2019. Australian Population Grid 2018. Accessed August 2023, [Online]. Available at: https://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/3218.02017-18

Australian Bureau of Statistics, 2020. Australian Population Grid 2019. Accessed August 2023, [Online]. Available at: https://www.abs.gov.au/statistics/people/population/regional-population/2018-19#data-downloads-population-grid-files

Australian Government, 2010. Directory of Important Wetlands in Australia (DIWA) Spatial Database. Accessed August 2022. [Online] Available at: http://www.environment.gov.au/fed/catalog/search/resource/details.page?uuid=%7BED248FC1-7237-4A74-91AC-2DA3FC277E0A%7D

Australian Government, 2015. RAMSAR Wetlands of Australia. Bioregional Assessment Source Dataset. Accessed June 2023. [Online] Available at: https://data.gov.au/dataset/ds-dga-0378c301-c421-49ca-b48b-f300570d84ad/details

Australian Government, 2022. Collaborative Australian Protected Areas Database (CAPAD) 2022 – Terrestrial. Accessed October 2022. [Online] Available at: https://www.arcgis.com/home/item.html?id=ec356a872d8048459fe78fc80213dc70 -DIWA

Birds Australia, 2009. Birds Australia - Important Bird Areas (IBA) 2009. Bioregional Assessment Source Dataset. Accessed 13 March 2019, [Online]. Available at: https://data.gov.au/data/dataset/5d488350-83b6-4e71-8d17-687ad8ff9941

Cornell Lab of Ornithology, 2023. eBird: An online database of bird observations. Accessed January 2024 on application via: https://ebird.org/home

Flickr, 2021. Flickr API: Access to Photo Data. Accessed 2021 on application via: https://www.flickr.com/services/api/

Geoscience Australia, 2006. GEODATA TOPO 250K Series 3. Accessed June 2019. [Online] Available at: https://ecat.ga.gov.au/geonetwork/srv/eng/catalog.search#/metadata/64058

Gould, L., Stewart-Koster, B., Hunt, G., Kennard, M., Sheldon, F., 2023. Predicting floodplain habitat inundation throughout the Murray-Darling Basin to inform the management of environmental flows. Final Report for MD-WERP. Australian Rivers Institute, Griffith University, Brisbane. https://www.griffith.edu.au/australian-rivers-institute.

Murray-Darling Basin Authority, 2013. Surface Water Sustainable Diversion Limit (SDL) Resource Units. Accessed January 2022. [Online] Available at: https://data.gov.au/data/dataset/surface-water-sdl-resource-units

Murray-Darling Basin Authority, 2015. SRA 2 Valley Spatial Layer. Accessed September 2023. [Online] Available at: https://data.gov.au/data/dataset/sustainable-rivers-audit-2-valley-and-zone-spatial-layers