**Greenspace Data in ALSWH cohorts: Description**

Tafzila Mouly, Luke Knibbs and David Fitzgerald, May 2021

**Background**

As part of ALSWH EoI#699 (Knibbs, Mishra, Fitzgerald, Hockey, Mouly), two indicators of greenspace - annual average normalized difference vegetation Index (NDVI), and annual average non-photosynthetic vegetation fractional cover - have been estimated and linked to the geocoded residential addresses of ALSWH participants from 1994 to 2018. Additionally, the percentages of parkland in different buffers have been estimated based on census land-use classifications available since the 2011 census (2011 and 2016). These estimates are available for researchers using the ALSWH data.

See the Greenspace Structure and Accuracy sections below for further descriptions of the method to produce the greenspace data.

All greenspace metrics have been estimated for four circular buffers with radii of 100m, 500m, 1000m and 3000m around the residential address (Figure 1). These buffers aim to capture near-home and more distal greenspaces. The values of each greenspace metric represent the average in a given buffer.****

**Figure 1 Different Circular buffer around Geocoded Address**

**Overview and usage** Please note, there are no costs payable for non-commercial research-related use of these estimates. It is expected that any academic publications (journal article, conference paper, book chapter, etc.) that use the greenspace data offer the data contributors (A/Prof Luke Knibbs, [l.knibbs@uq.edu.au] and Dr Tafzila Mouly) the opportunity to be a co-author. Apart from recognition of the work of the contributor, this is also a way to ensure the data are applied and interpreted in the most valid way at the analysis planning stage, which benefits users and saves time. The contributors may choose not to take up the option of co-authorship, but can still offer useful content knowledge and feedback.

These are the variables in the Green Space data, estimates in each of the four buffers:

1. Annual average NDVI from 1994 to 2018 with quarterly for 1973-78 cohort
2. Annual average fractional cover (non-photosynthetic vegetation) from 1994 to 2015

With quarterly data for 1973-78 cohort

1. The percentage of parklands

**Normalized difference vegetation Index (NDVI)**

The normalized difference vegetation index or NDVI is the most commonly used satellite-derived vegetation index [1]. It is a ratio of the difference of spectral reflectance between red and near-infrared (NIR) band and summation of spectral reflectance of these two bands. NDVI has been calculated from Landsat Surface Reflectance Tier 1 images with a spatial resolution of 30 m (Landsat 5, 7 and 8), and aggregated to annual averages for each year. The data is cloud, shadow, snow and water masked. The values range from 0 to 1, with higher values indicating more green vegetation.

The data has annual NDVI estimates from 1994 to 2018 for all ALSWH cohorts. Additionally for the 1973-78 cohort there are average quarterly estimates for the survey years of the 1973-78 cohort. The particular three-month quarters used were chosen to capture the period where a majority of the participants’ responded, as shown in Table 1.

***Table 1 Three-month average NDVI for 1973-78 Cohort***

|  |  |  |
| --- | --- | --- |
| **Survey** | **Three Month Period** | **Survey Coverage Percentage** |
| **1** | July-September, 1996 | 95.27% |
| **2** | March-May, 2000 | 69.15% |
| **3** | March-May, 2003 | 73.23% |
| **4** | March-May, 2006 | 73.38% |
| **5** | April-June, 2009  | 70.72% |
| **6** | May -July, 2012 | 54.11% |
| **7** | April-June, 2015 | 74.84% |
| **8** | June-August, 2018 | 71.92% |

**Fractional Cover**

The annual average fractional cover of non-photosynthesising vegetation (spatial resolution: approximately 30 m) has been estimated in each buffer. These data were derived from data hosted by the Terrestrial Ecosystem Research Network (TERN), Geoscience Australia, and the Joint Remote Sensing Research Program for estimating the seasonal fractional cover data for the whole Australian [2] continent. The data is freely available on the website of the Terrestrial Ecosystem Research Network (TERN). They have estimated the proportion of photosynthetic vegetation (PV), non-photosynthetic vegetation (non-PV), and bare soil using Landsat images. We included only non-PV fractional cover, which includes the dead trees, litter, and wood. The inclusion of the non-PV band only was due to the other bands (PV and bare soil) being highly correlated with NDVI, while non-PV band may capture other aspects of the environment not already captured by NDVI. Annual fractional cover was calculated for the period between 1st December of the preceding year to 30th November of that year.

The data has annual Fractional Cover estimates from 1994 to 2015 for all ALSWH cohorts. Additionally there are average quarterly estimates for the 1973-78 cohort during the survey years. The particular three-month quarters used were not exactly the same quarters as for the NDVI, but were the calendar seasons, autumn or winter, that closely matched the quarters used for the NDVI (and identified in Table 1.) These are shown in Table 2.

**Table 2 Three-month average fractional cover (non-photosynthetic vegetation) 1973-78 cohort**

|  |  |  |
| --- | --- | --- |
| **Survey** | **Three Month Period** | **Survey Coverage Percent** |
| **1** | June-August, 1996 | 87 |
| **2** | March-May, 2000 | 69 |
| **3** | March-May, 2003 | 73 |
| **4** | March-May, 2006 | 73 |
| **5** | March-May, 2009  | 63 |
| **6** | June -August, 2012 | 44 |
| **7** | March-May, 2015 | 60 |

**Parkland**

The Australian Bureau of Statistics introduced mesh blocks as the smallest spatial unit for the 2011 and 2016 censuses [3]. It is not a traditional land use map with definitive land areas. Instead, mesh blocks are given land use categories based on their primary use of most of their area. Each mesh block is coded as one of ten land use categories – "residential, commercial, industrial, parkland, education, hospital/medical, transport, primary production, water and other." The percentage of parklands in each buffer (i.e. total parkland area / buffer area \* 100) was estimated. Parklands include natural reserves and conserved areas, public open spaces and sporting facilities irrespective of whether they are enclosed or open.

**Greenspace Data Structure**

There were annual greenspace data – NDVI, fractional cover and parkland – derived for all the addresses of the respondent during the survey. If we had three addresses for a respondent, for example, there would be three complete records of all the annual data for that one respondent. Therefore, many of the annual values would be for years when the respondent was not at the address. To correct this the data were compressed into one record per respondent. This was done to ensure the annual residential greenspace data matched the years the respondent was at that residence. The compression is described thoroughly in the **Compressing the Air Pollution** **data** documentation. The following example gives a brief summary of the method. Say we were notified of a new of address at survey 4 and we last had contact with the respondent at survey 2 living at an different address, then the midpoint year between the survey 4 and survey 2 responding dates would be calculated. We would use the annual greenspace data for the survey 2 address up to but not including the calculated mid-point year. We would then use the annual greenspace data for the survey 4 address from the mid-point year onwards. Any subsequent changes of address would be treated the same way. By this method multiple greenspace records per respondent were compressed into one record per respondent.

It follows that if we lost contact with respondents for long periods of time we would not know if they changed address or not. Their greenspace values could be based on an address where they may not be living any more. To account for this whenever a respondent withdrawals from the survey the greenspace values are set to missing after the withdrawn years. Similarly, whenever a respondent dies the greenspace values are set to missing after the death year. However, some women have not withdrawn but also not responded to recent surveys and so we cannot be sure of their addresses. These women will still have the greenspace values for recent years but the researcher may decide not to use these values since they are based on an address which may have changed.

**Accuracy**

At each ALSWH survey wave, every respondent’s address and date of response was recorded, where it was known. These addresses are confidential and are never released to researchers. For the purpose of this work the longitude and latitude of these addresses were derived using Google Maps and then later used to estimate annual greenspace. The process of determining the accuracy and displaying it on the data was as follows.

Using Google Maps

Each address was put into Google Maps where its longitude and latitude were calculated. Google maps gave an accuracy value for each longitude and latitude estimate. The accuracy values were:

|  |  |
| --- | --- |
| ROOFTOP | the addresses for which Google has location information accurate down to street address precision.  |
| RANGE\_INTERPOLATED | the addresses that reflect an approximation (usually on a road) interpolated between two precise points (such as intersections). An interpolated range generally indicates that rooftop geocodes are unavailable for a street address |
| GEOMETRIC\_CENTER | geometric centers of a location such as a polyline (for example, a street) or polygon (region). |
| APPROXIMATE | the addresses that are characterized as approximate |
| missing | there was no address or no longitude and latitude could be found |

These accuracy values were included in the data. There is an accuracy variable for each cohort wave – AccuracyWave1, AccuracyWave2, etc. These variables can take five values including ‘missing’ (as listed above.)

To understand how these accuracy values are recorded in the data it is helpful to look at some examples in the greenspace data (Table 3) and the geographic relocation data (Table 4)

**Table 3 Examples of accuracy values for two records in the 1973-78 cohort Greenspace data**

Greenspace Data (only showing accuracy values and SurveyAddressIndicator)

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ID | SurveyAddress Indicator | accuracyW1 | accuracyW2 | accuracyW3 | accuracyW4 | accuracyW5 | accuracyW6 | accuracyW7 | accuracyW8 |
| 1 | 124567 | ROOFTOP | ROOFTOP | missing | ROOFTOP | RANGE\_INTERPOLAT | GEOMETRIC\_CENTER | ROOFTOP | missing |
| 2 | 1 | ROOFTOP | missing | missing | missing | missing | missing | missing | missing |

**Table 4 Examples of Geographic Relocation values for two records in the 1973-78 cohort Greenspace data**

Geographic Relocation data

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ID | Wave | cohort | Moved | dist\_km | dist\_m | EstYearMove | EstDateMove |
| 1 | 1 | YNG |  |  |  |  |  |
| 1 | 2 | YNG | 0 |  |  |  |  |
| 1 | 4 | YNG | 0 |  |  |  |  |
| 1 | 5 | YNG | 1 | 0.74 | 736 | 2008 | 13DEC2007 |
| 1 | 6 | YNG | 1 | 6.89 | 6889 | 2011 | 12DEC2010 |
| 1 | 7 | YNG | 1 | 24.43 | 24434 | 2014 | 21DEC2013 |
| 2 | 1 | YNG |  |  |  |  |  |

ID 1 in table 3 has SurveyAddressIndicator ‘124567’, so we had an address from the eight surveys except waves 3 and 8. Her address’s accuracy was ‘ROOFTOP’ at waves 1, 2 and 4. The accuracy changed in later waves as she changed addresses, as can be seen in Table 4. The first change of address occurred in 2008 and so her greenspace values for 1996 to 2007 were based on an address with ROOFTOP accuracy. From 2008 to 2010 the greenspace values were based on an address with accuracy RANGE INTERPOLAT, and from 2013 to 2018 there were based on the address with ROOFTOP accuracy. Since we did not know her address at wave 8 the greenspace values at wave 8 were based on the assumption she did not change address at wave 8. Similarly, we did not know her address at wave 3 so we assumed her address at that time was the same as in the previous survey. This assumption becomes more problematic for the second ID.

For ID 2 we only had an address at wave 1 in 1996. We did not know if she ever changed address since that time and so there is only one record in the Geographic Relocation file. All of her greenspace values, which were from 1996 to 2018, were based on her 1996 address. If she changed address since that time all the greenspace values will be based on a wrong address for the years following her change of address.

**Greenspace Variable List**

**Data files**

There is a separate green space file for each ALSWH cohort.

|  |  |  |
| --- | --- | --- |
| File name | Number of Records | Cohort |
| GreenSpaceOLD | 12,432 | 1921-26 |
| GreenSpaceMID | 13,712 | 1946-51 |
| GreenSpaceYNG | 14,245 | 1973-78 |
| GreenSpaceNYC | 16,957 | 1989-95 |

Each file has one record for each ALSWH woman in the cohort for whom we had an address.

Each file has these variables:

**IDalias**

**Cohort** indicator: YNG, MID, OLD, NYC

**SurveyAddressIndicator** : A string variable indicating all the waves/ surveys for which we had an address.

**Non-photosynthetic vegetation** **Fractional Cover (FC\_NP)**

These are annually from 1994 to 2015 inclusive. For each of these years there are four buffer radii – 1000m, 100m, 3000, 500m.

* FC\_NP\_Annual\_1000m\_1994 - FC\_NP\_Annual\_1000m\_2015
* FC\_NP\_Annual\_100m\_1994 - FC\_NP\_Annual\_100m\_2015
* FC\_NP\_Annual\_3000m\_1994 - FC\_NP\_Annual\_3000m\_2015
* FC\_NP\_Annual\_500m\_1994 - FC\_NP\_Annual\_500m\_2015

**Normalized difference vegetation Index (NDVI)**

These are annually from 1994 to 2018 inclusive. For each of these years there are four buffer radii – 1000m, 100m, 3000, 500m.

* NDVI\_Annual\_1000m\_1994 - NDVI\_Annual\_1000m\_2018
* NDVI\_Annual\_100m\_1994 - NDVI\_Annual\_100m\_2018
* NDVI\_Annual\_3000m\_1994 - NDVI\_Annual\_3000m\_2018
* NDVI\_Annual\_500m\_1994 - NDVI\_Annual\_500m\_2018

**Parkland Percentage**

Parkland Percentage was recorded at 2011 and 2016 only. For each of these years there are four buffer radii – 1000m, 100m, 3000, 500m.

* Parkland\_Percentage\_1000m\_2011, Parkland\_Percentage\_1000m\_2016
* Parkland\_Percentage\_100m\_2011, Parkland\_Percentage\_100m\_2016
* Parkland\_Percentage\_3000m\_2011, Parkland\_Percentage\_3000m\_2016
* Parkland\_Percentage\_500m\_2011, Parkland\_Percentage\_500m\_2016

**Buffer Areas**

These are constant values showing the area for each buffer.

* Buffer\_Area\_1000m (3.141177001)
* Buffer\_Area\_3000m (28.26183592)
* Buffer\_Area\_500m (0.785191165)

**Accuracy**

There are accuracy values for each survey wave for which we have greenspace data.

* 1946-51 cohort AccuracyW1 – AccuracyW8
* 1921-26 cohort AccuracyW1 – AccuracyW19
* 1973-78 cohort AccuracyW1 – AccuracyW8
* 1989-95 cohort AccuracyW1 – AccuracyW5

**References**

1. Jinru, X. and S. Baofeng. (2017). Significant Remote Sensing Vegetation Indices: A Review of Developments and Applications. Journal of sensors. vol. 2017.
2. TERN AusCover. (2020). Seasonal fractional cover - Landsat, JRSRP algorithm, Australia coverage. Obtained from [http://data.auscover.org.au/xwiki/bin/view/Product+pages/Landsat+Seasonal+Fractional+Cover#Comments](http://data.auscover.org.au/xwiki/bin/view/Product%2Bpages/Landsat%2BSeasonal%2BFractional%2BCover#Comments) , made available by the AusCover facility ([http://www.auscover.org.au](http://www.auscover.org.au/)) of the Terrestrial Ecosystem Research Network (TERN, [http://www.tern.org.au](http://www.tern.org.au/)). Accessed [15/09/2020].
3. Australian Bureau of Statistics (ABS). (2016). Australian Statistical Geography Standard (ASGS) Volume 1 - Main Structure and Greater Capital City Statistical Areas (cat no. 1270.0.55.001). Obtained from [https://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/1270.0.55.001July%202016?OpenDocument](https://www.abs.gov.au/AUSSTATS/abs%40.nsf/DetailsPage/1270.0.55.001July%202016?OpenDocument)