**Information on the gap-filled environmental & plot-level meteorological data**

Date: January 20, 2015. Updated May 14, 2020.

Contact: Ed Ryan ([edmund.ryan@manchester.ac.uk](mailto:edmund.ryan@manchester.ac.uk)); Elise Pendall ([e.pendall@westernsydney.edu.au](mailto:e.pendall@westernsydney.edu.au))

**Notes for all datasets:**

* The plot numbering is the same as given at the site and in the database, e.g. 1\_4ACN means plot 1 which is the 4th plot of the five plots belonging to the ACN treatment. Letters of the treatment: 1st: A or C = Ambient / Elevated CO2; 2nd: C or H = Control / Heated; 3rd: N or D or S = No / Deep / Surface irrigation.
* ‘Day’ (e.g. Day355) is the day number from 01/01/2006. The day column in the datasets are given in date format.
* Each flag column indicates whether the measurement was made (flag=0) or gapfilled (flag≥1). Flag information specific for each dataset is given below.

**Names of gap-filled Datasets:**

1. PHACE\_Greenness\_gapfilled\_fromDay425.

Vegetation greenness was quantified every 2-4 weeks between March and October, resulting in 71 separate greenness measurement days spanning 2007-2012; photographs were taken for all 30 plots on each measurement day using a 2 m high camera stand and a 1 m2 ground frame. SamplePoint software was used to quantify parts of the photograph that were not vegetation (i.e. soil or litter) and to classify the different species of grass. Greenness was quantified by converting the image pixels within each photo to a matrix of numbers using Matlab R2011a and quantifying the hue, saturation, and value scales for detection of green. Greenness varies from 0 (absence of green biomass) and 1 (plot is completely covered with green biomass).Units are given as a percentage, i.e. range = 0-100%. Flag definition: 0 = measured / no gap-filling required; 1 = gap-filled using linear interpolation as described above.

1. PHACE\_SWC\_Sentek\_DSM1\_gapfilled\_Allyears.
2. PHACE\_SWC\_Sentek\_DSM1\_gapfilled\_Allyears.
3. PHACE\_SWC\_Sentek\_DSM1\_gapfilled\_Allyears.

Soil Water Content (SWC) was measured daily and at a depth of 5-15cm (DSM1), DSM2 (SWC at 15-25cm depth) and DSM3 (SWC at 35-45cm depth) using a Sentek probe. SWC is given as are a percentage, i.e. range = 0-100%. Flag definition: 0 = measured / no gap-filling required; 1 = gap-filled using mean of data of plots that (a) belong to the same treatment, and (b) represent the distribution of soil water contents to the original plot (R2>0.9 and median absolute difference < 1%); 2 = gap-filled using cubic spline interpolation.

1. PHACE\_Temp\_DST1\_gapfilled\_fromDay185.
2. PHACE\_Temp\_DST2\_gapfilled\_fromDay185.
3. PHACE\_Temp\_DST3\_gapfilled\_fromDay185.

For the temperature data at each plot, DST1 refers to air temperature, while DST2 and DST3 correspond to soil temperature measurements at 3cm and 10cm depths respectively. Temperature is measured hourly and given in Kelvin. Flag definition: 0 = measured / no gap-filling required; 1 = gap-filled using mean of data of plots that (a) belong to the same treatment, and (b) represent the distribution of temperatures to the original plot (R2>0.9 and median absolute difference < 0.5oC); 2 = gap-filled using air temperature from HOBO met station (DST1) or by cubic spline interpolation (DST2 and DST3).

1. PHACE\_VPD\_gapfilled\_fromDay185.

Vapor Pressure Deficit (VPD) was calculated using the HOBO Relative humidity data gap-filled using NARR data, and the plot level air temperature data (Temp\_DST1). We use the formula: VPD=SVP – (RH/100)\*SVP, where Saturated Vapor Pressure (SVP) is given by:

SVP =

VPD is given on the hourly time-scale and units are in KPa. Note that it is only computed for up to 2012; please contact E. Ryan if you need 2013 as well. Flag definition: 0 = measured / no gap-filling required; 1 = both air temperature and RH needed to be gap-filled; 2 & 3 = gap-filled due to air temperature being gap-filled by a neighboring plot (2) or using air temperature from HOBO met station (3); 4 & 5 = gap-filled due to RH being >100 (4) / <0 (5); 6 = gap-filled due to RH needing to be filled but when 0<RH<100.

**Description of the: (a) gap-filling methods employed for the soil water content (SWC), plot level air temperature (Temp\_DST1) and soil temperature at two depths (Temp\_DST2 and Temp\_DST3); (b) linear interpolation of vegetation greenness (greenness) data.**

Gap-filling was required for the plot level soil water content (SWC), plot level soil temperature (Temp\_DST2 and Temp\_DST3) and plot level air temperature (Temp\_DST1). The first stage of gap-filling for the daily SWC data involved replacing missing data with that of a plot within the same treatment as long as the correlation coefficient of SWC between the two plots was greater than 0.9 and the Median Absolute difference (MAD) must be less than 0.01 (SWC data range from 0.03 to 0.31). For days where gaps still remained, the second stage of gap-filling involved using a cubic spline interpolation ([Lyche and Schumaker 1973](#_ENREF_2)), assuming that the precipitation for each day of missing data was not greater than 2mm ([Chapin III and Matson 2011](#_ENREF_1)); if the precipitation was greater than 2mm then it is likely the cubic spline method would not accurately capture the change in SWC resulting from the precipitation event. For this dataset, none of the missing SWC data occurred on days with high precipitation so this was not a problem and all of the gaps could be filled. The accuracy of the cubic spline method was tested on a gap-free subset of the SWC dataset of length 450 days, by inserting numerous gaps which span the range of gap lengths observed in the dataset. The R2 between the gap-filled data and corresponding true values was 0.996 and a MAD of 0.0008. The soil temperature data (Temp\_DST2 and Temp\_DST3) were each filled in exactly the same way except that the MAD threshold was 1oC. It was found that the cubic spline interpolation was needed far less often than with the SWC data because the soil temperature data from plots within the same treatment very closely matched one another. The plot-level air temperature (Temp\_DST1) were gap-filled by taking the mean of the temperature in the other plots belonging to the same treatment. It was not necessary to only use those plots where correlated and matched well with the plot to be gap-filled as there was generally much less variation and bias within plots of the same treatment. Remaining gaps were replaced with NARR data.

We differentiate between estimates on non-measurement days by linear interpolation used for the greenness data and the gap-filling carried out on the other environmental and micrometeorological data. Adoption of the other gap-filling methods is appropriate for greenness because there are significantly less frequent measurements made over the experimental period that the other data (table 1). We use linear interpolation because greenness is generally smooth in time unlike SWC and Tsoil which can greatly vary on daily and hourly time-scales due to variation in meteorology. Cubic spline interpolation is not appropriate here because there was insufficient data during 2007-2012 to justify such a method.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | SWC (DSM1) | SWC (DSM1) | SWC (DSM1) | AirT (DST1) | Tsoil (DST2) | Tsoil (DST3) | Greenness |
| % of data points present | 94.2% | 89.1% | 88.9% | 92.7% | 97.5% | 97.5% | 3.5% |

Table 1. Percentage of data present for the plot level environmental datasets (SWC and Tsoil) and plot-level micrometeorological data (AirT) and the greenness dataset.