CSIRO LAND & WATER | CSIRO HEALTH & BIOSECURITY
www.csiro.au

# Weed data aggregation and risk assessment for the Pilbara region of Western Australia

## Interim report 4 on data aggregation and data synthesis

31st July 2017

Bruce L. Webber, Kathryn L. Batchelor, Tommaso Jucker & John K. Scott

For: Department of Biodiversity, Conservation and Attractions, Western Australia
Roy Hill Holdings Pty Ltd
Atlas Iron Limited
Pilbara Corridors – Rangelands NRM (WA)

## Citation

Webber BL, Batchelor, KL, Jucker, T and Scott JK (2017). Weed data aggregation and risk assessment for the Pilbara region of Western Australia. Interim report 4 on data aggregation and data synthesis. CSIRO, Australia.

# SCOPE

This document reports on the scope of work completed in the five months following from the previous report submitted on the 28th February 2017.

# PROGRESS SUMMARY

## Project logistics

As raised in Interim Report 3, the project required re-planning due to the last minute unavailability of the spatial statistician assigned to the project. An alternative scientist, Tommaso Jucker, has joined the project and has been leading progress on the spatial analysis component of the work. Contract and milestone revisions were put in place to (i) extend the project and final report submission date until 31st December 2017 and (ii) revise interim reporting to 31st July 2017 and 31st October 2017. The contract revision underpinning this process was finalised on time with the Department of Parks and Wildlife (now Department of Biodiversity, Conservation and Attractions).

## Data sourcing and aggregation

Data sourcing and aggregation is fast reaching a point beyond which we will not be able to incorporate any further data into the analysis and synthesis. Having said that, however, we are continuing to actively pursue new data sets and sources. Up to this point, we have collated data gleamed from 683 reports, 14 databases from amenable mining companies (supplied via DBCA), 2 Department of Agriculture Rangeland Surveys and the Pilbara Mesquite Management Committee (PMMC). Data was also supplied by botanist Dr Malcolm Trudgen. Collectively, there are over 720,000 records for weeds (Figure 1). In addition, Rio Tinto Ltd have agreed to supply their weed data in electronic form which, at the time of writing, had not been incorporated. It is anticipated their data will fill some of the vacant spaces in overall weed distribution as Rio Tinto is one of the largest miners of the Pilbara and under-represented in captured reports (~10% of reviewed data compared to BHP at 35%). Also, there is a chance of additional data being identified following the five public presentations scheduled for delivery in August (see Appendix 1).

Data was also collected on locations where weeds were noticeably absent (Figure 2). 273 flora and vegetation reports contained raw quadrat data and these were filtered by presence/absence of weeds. In many cases these quadrats were in close proximity to quadrats containing weeds, and the distribution generally follows the locations of mines and infrastructure.

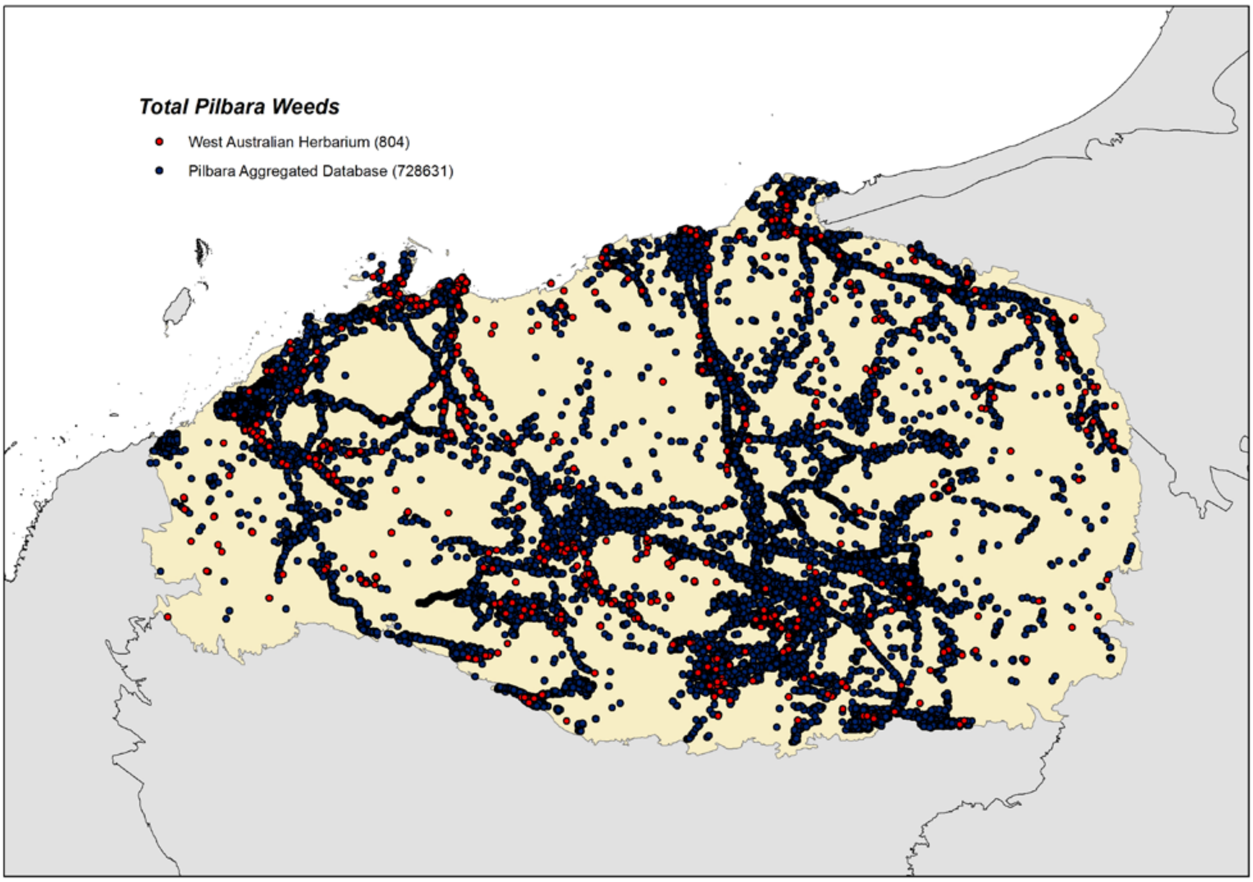


**Figure 1**. Point distribution of weeds, and weed absence (zero records) for the Pilbara bioregion as of July 2017. Dark blue points are CSIRO accumulated data sourced from flora and vegetation reports, contributed mining company databases and the Pilbara Mesquite Management Council. Red points are weed records supplied by the WA Herbarium in 2014, representing knowledge at the beginning of the project.

## Data verification

Data errors came from a variety of sources; data entry in source documents, wrong GPS co-ordinates, missing data, incorrect OCTR translation and incomplete/misapplied identifications. To identify data entry and OCTR translation errors, weed distributions were mapped and if a species was located >50km from the nearest point from the same document, points either revised or deleted if its error couldn't be resolved. Wrong GPS co-ordinates were often due to the wrong AMG Map Zone supplied in the report and these were obvious when the distribution is well outside the Pilbara area. 159 weeds species names are mentioned in the reports, but once names were compared with DBCA's Florabase, 127 current names remain. Some errors in data entry in the vegetation reports cannot be isolated, and some ground-truthing will be required to confirm if needed.

Verifying the extensive data supplied by the PMMC is not complete at this point in time. Extensive data curation has been required to make sense of a rather haphazard collection of tens of thousands of digital files and folders. While the resulting point distributions have been examined by staff familiar with the work, understanding the nature of the data collected (i.e. herbicide log vs observation log) requires further discussion with on ground staff in the Pilbara.

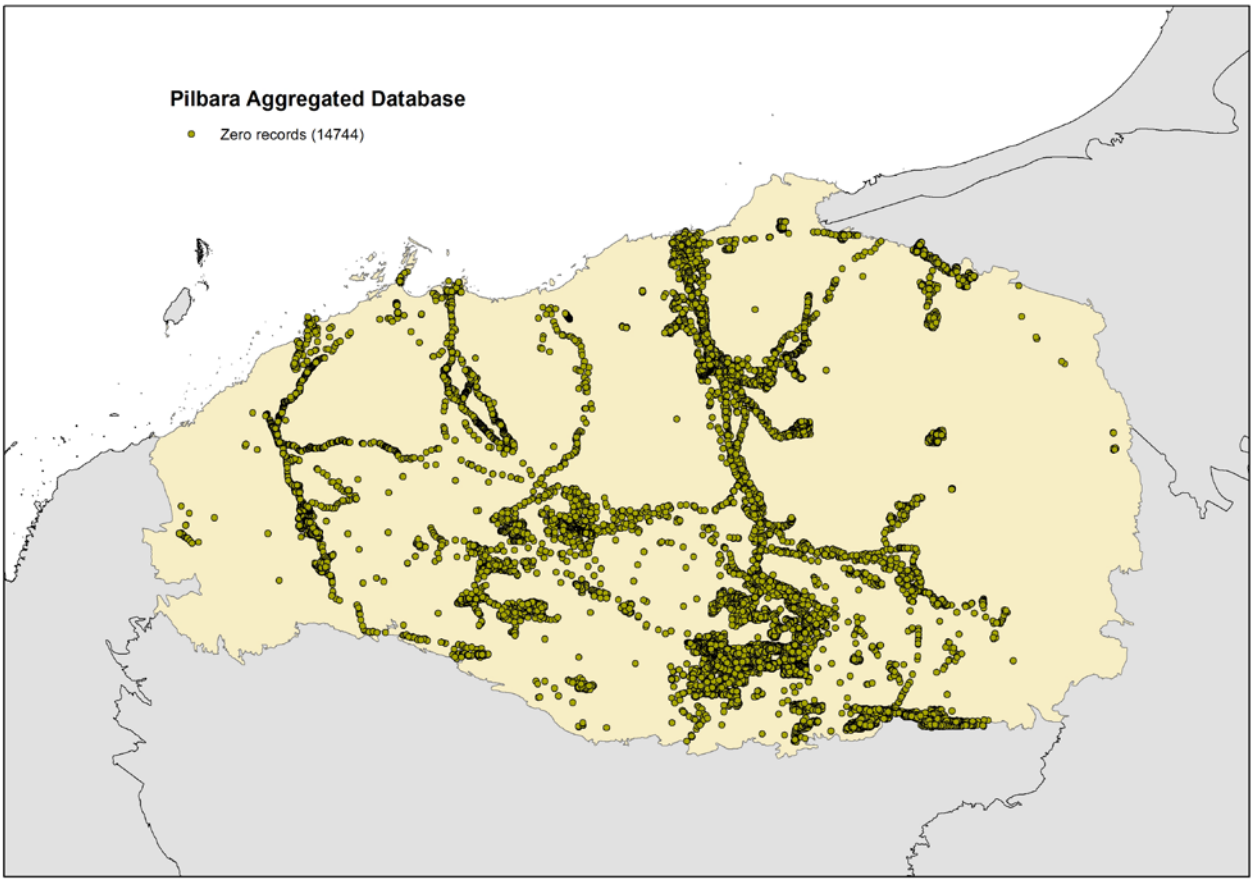


**Figure 2**. Point distribution of quadrats where weeds were absent (zero records) for the Pilbara bioregion.

## Data synthesis

The compiled data were used to model the probability of occurrence for the 20 most common weeds in the database in relation to human activities (road and rail density, distance to mines, distance to human settlements), natural and human-mediated disturbance (fire frequency, persistent vegetation cover, protected area cover) and environmental and climatic suitability (NDVI, distance to water bodies, mean annual temperature and rainfall). Poisson point process models were chosen for this purpose, as they provide an intuitive way to explicitly account for the high degree of spatial variability in sampling effort we observe in the database. Appropriate layer selection and source for each of the chosen covariates remains a work in progress, given the multiple sources of layers available.

The models revealed a number of clear and consistent patterns in terms of weed prevalence (Figure 3). For instance, weeds were much more likely to be recorded in areas with high road and rail line densities, as well as in the proximity of mine sites and water courses. Additionally, weeds were generally less common in areas which have persistent vegetation cover and on land tenure falling under some form of protection.  Using the models, we were also able to identify areas of the Pilbara bioregion that are characterized by very high weed prevalence, as well as those that currently are less affected by weed encroachment (Figure 4).

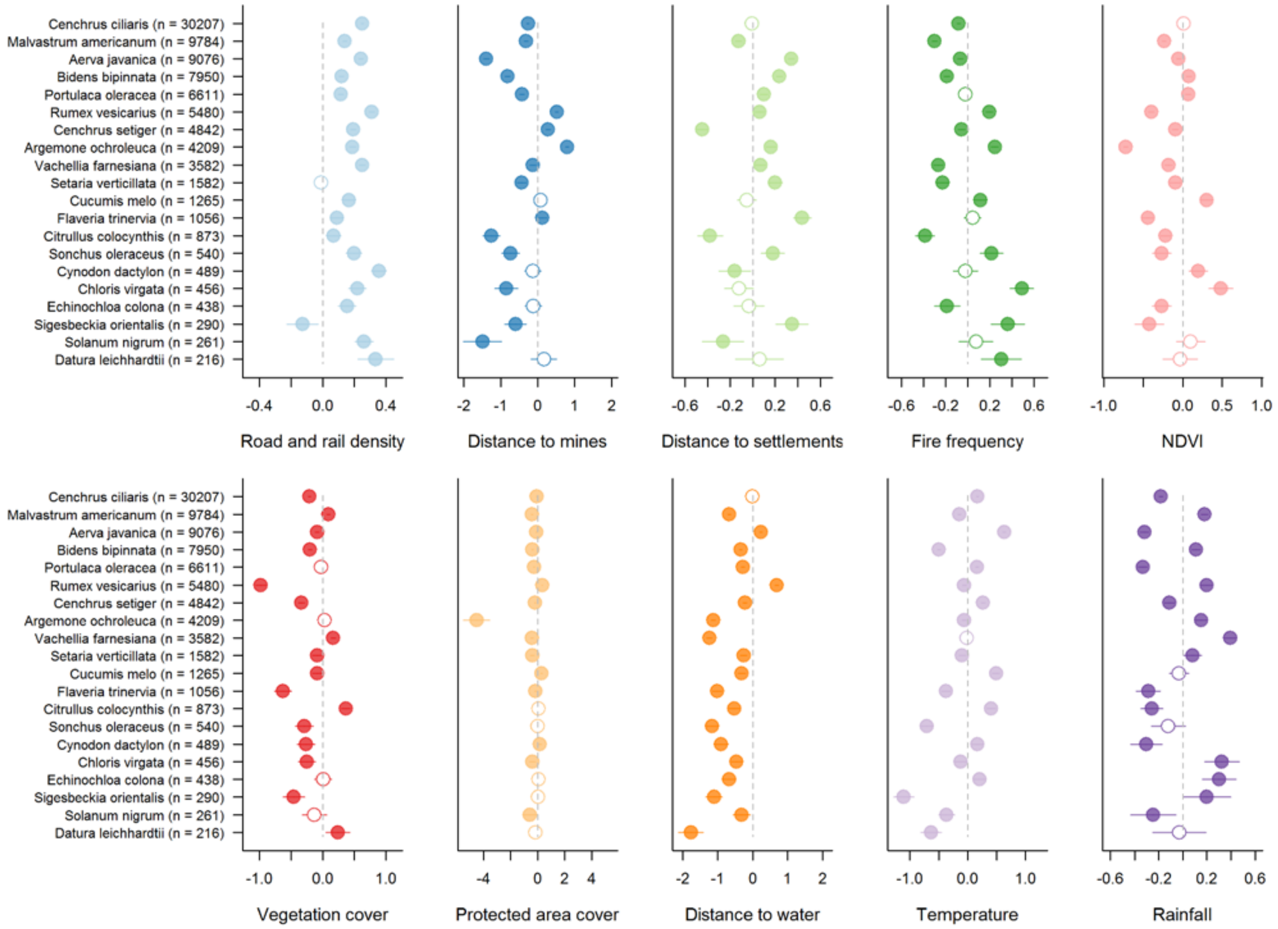


**Figure 3**. Standardized model coefficients showing the response of the top 20 most abundant weeds in the database to the spatial predictors included in the models. Filled circles correspond to statistically significant model coefficients ($P < 0.05$), with error bars corresponding to 95% confidence intervals of the model coefficients.

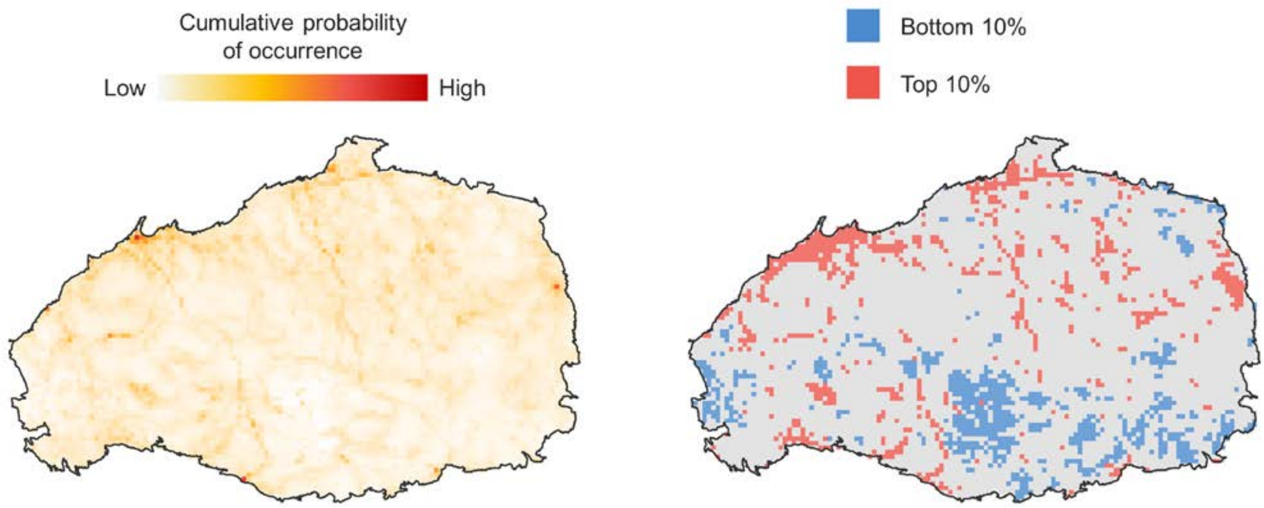


**Figure 4**. Cumulative probability of weed occurrence at 5 x 5 km resolution for the Pilbara bioregion. The map on the left shows continuous values of probability of weed occurrence (from low to high), whereas the map on the right highlights the location of the 5 x 5 km grid cells with the highest (red) and lowest (blue) probability of weed occurrence.

## Database roadshow

Planning for a database roadshow, consisting of a series of presentations and discussions on the project, is well underway. Events will be held in Perth in the week of $14^{th}$ August and across the Pilbara (the primary presentation will be in Karratha, with regional talks as appropriate) in the week of $21^{st}$ August. Logistics assistance will be provided by Jo Kuiper of the PMMC.

## Conclusions

- Solutions have been found to achieve all project milestones by the revised end date of $31^{st}$ December 2017.
- Data aggregation is complete, although additional datasets may be possible to add in to the final data synthesis if supplied in an electronic form (notably from Rio Tinto). Even though it may not be included in the analysis, the project will continue to be active in the collection of data.
- Data verification stage is mostly complete, with the PMMC data to be finalised after the stakeholder meetings in August 2017
- Data synthesis and spatial analyses are well advanced and on track to be completed by the revised end date of $31^{st}$ December 2017.
- Engagement with key stakeholders to share the data and to improve the final stages of refinement will take place via a roadshow in August.

CONTACT US

**t** 1300 363 400
+61 3 9545 2176
**e** csiroenquiries@csiro.au
**w** www.csiro.au

AT CSIRO, WE DO THE EXTRAORDINARY EVERY DAY

We innovate for tomorrow and help improve today – for our customers, all Australians and the world.

Our innovations contribute billions of dollars to the Australian economy every year. As the largest patent holder in the nation, our vast wealth of intellectual property has led to more than 150 spin-off companies.

With more than 5,000 experts and a burning desire to get things done, we are Australia's catalyst for innovation.

CSIRO. WE IMAGINE. WE COLLABORATE. WE INNOVATE.

FOR FURTHER INFORMATION

**CSIRO Land & Water | CSIRO Health & Biosecurity**
Bruce Webber
**t** +61 8 9333 6802
**e** bruce.webber@csiro.au
**w** http://people.csiro.au/Bruce-Webber