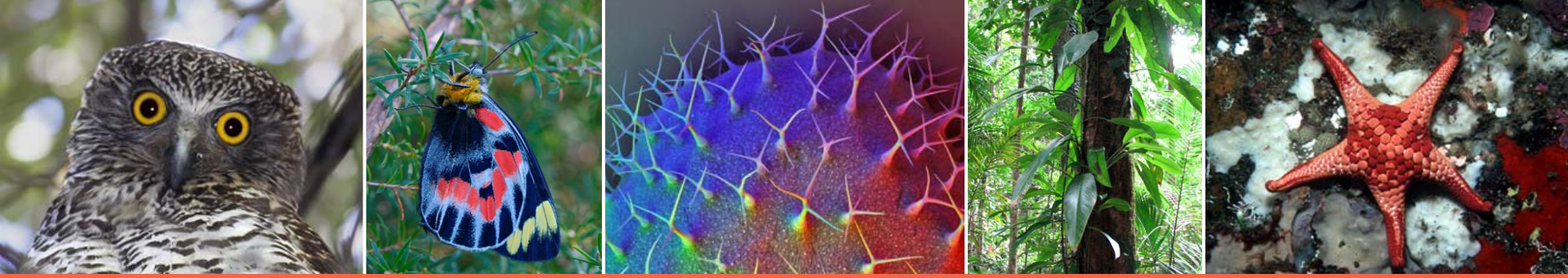


www.ala.org.au

Data Quality

ALA-ERIN Collaboration

Simon Bennett

ALA / ERIN DSEWPAC, 14 Oct 2011

Australian Government
Department of the Environment, Water, Heritage and the Arts

The Atlas is funded by the Australian Government under the National Collaborative Research Infrastructure Strategy and further supported by the Super Science Initiative of the Education Investment Fund

Overview

- ALA Data Quality Wiki
- Spatial Precision Issues
- Spatial Outlier Detection -
  - Jackknife
  - Convex Hull Trimming
  - Alpha Hulls
  - Environmental Surface Issues
  - Maxent Pair wise Analysis
- Python Tool

ALA Data Quality Wiki

ala-dataquality - Atlas of Living Australia - Data Quality - Google Project Hosting - Windows Internet Explorer

http://code.google.com/p/ala-dataquality/

ausavim@gmail.com | My favorites | Profile | Sign out

ala-dataquality
Atlas of Living Australia - Data Quality

Search projects

Project Home | Wiki | Administer

Summary | Updates | People

Tip: Project owners, see our Getting Started guide for steps to configure your project.

Project Information

Starred by 1 user
Activity High
Project feeds

Code license
Mozilla Public License 1.1

Labels
ALA, biodiversity, Australia, TWDG

Members
ausavim, brynk.ala, milo_nicholls@hotmail.com, moyesyside, leebelbin, dhobern
14 committers

Your role
Owner

Links

External links
Data Quality Checks Spreadsheet
alageospatialportal
alabiocache
alabie
ALA Home
gbif-ecat

Groups
ALA Data Quality Discussion

Atlas of Living Australia - Data Quality Portal

Welcome to the Atlas of Living Australia Data quality checks and issues pages.

- The Data Quality Checks Spreadsheet is a summary of the data quality checks implemented and planned for the Atlas of Living Australia.
- We have also set up a Wiki page for each check to document more detail where needed and for you to provide comments on each check.
- Please post your comments of a broader nature, along with any suggestion for content inclusion to the general comments page.

Links to key ALA data quality documents and resources on the internet are provided below.

Key ALA Data Quality Documents

- Data Quality Checks Spreadsheet
- Data Quality Wiki
- Completeness model (previously named 'quality model')
- Darwin Core Key Spatial Concepts
- ALA continues to focus on data quality

Key Data Quality References and Tools

- Darwin Core: Quick Reference Guide
- Tann,J & P. Flemons 2008. ALA Review of Online and Desktop Tools.
- Chapman, A. D. 2005. Principles of Data Quality
- Chapman, A.D. 2005. Principles and Methods of Data Cleaning – Primary Species and Species-Occurrence Data
- Chapman, A. D. 2005. Uses of Primary Species-Occurrence Data
- Diva GIS
- Georeferencing Calculator: Online Tool
- Georeferencing Calculator: Online manual
- Georeferencing Calculator: MaNIS/HerpNet/ORNIS Georeferencing Guidelines
- Georeferencing Calculator: Georef-calculator on google code. See Wiki for User Manual
- BioGeomancer WorkBench
- Biogeomancer-core - Google code
- speciesLink Data Cleaning
- speciesLink spOutlier
- Maps and the GDA: Geocentric Datum of Australia: Info Sheet

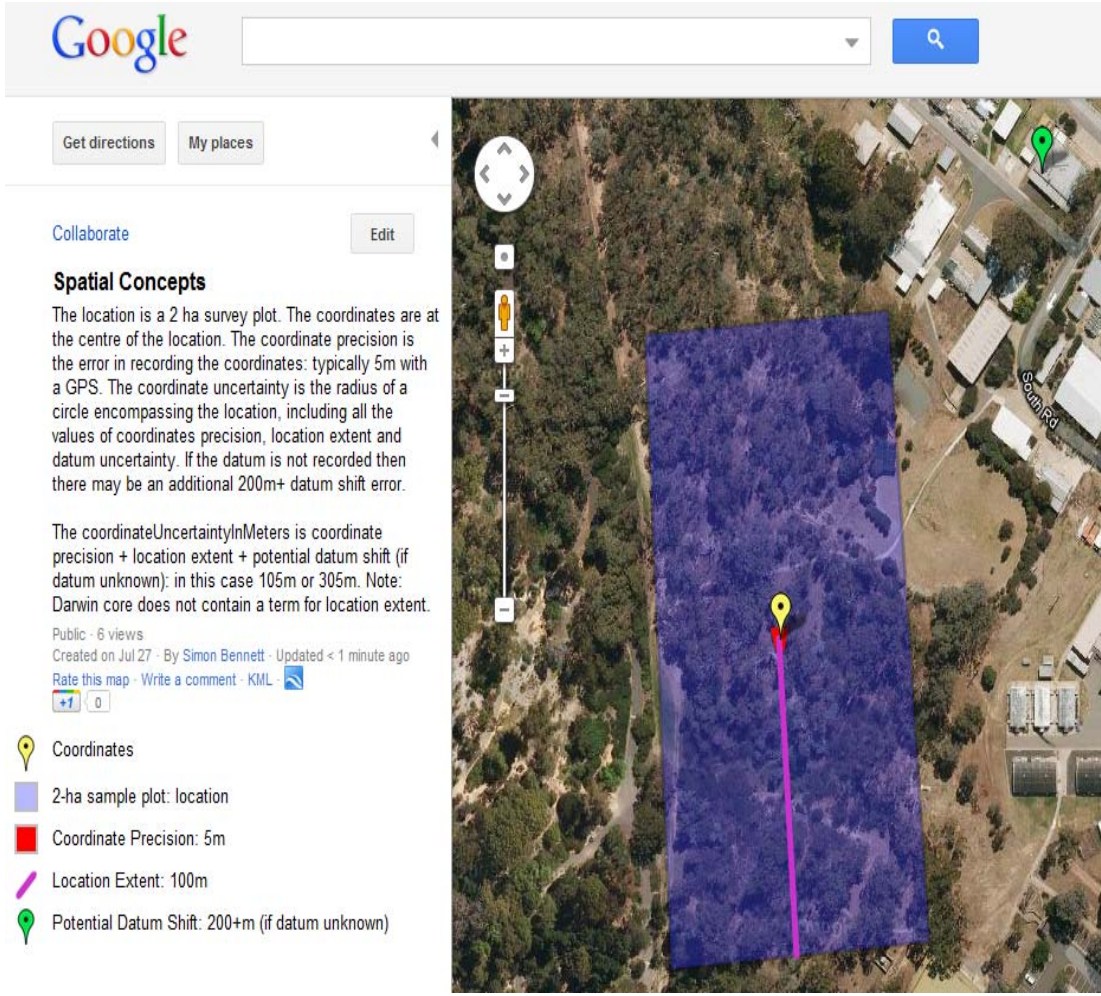
©2011 Google - Terms - Privacy - Project Hosting Help

Powered by Google Project Hosting

- Data Quality One Stop Shop
  - Links to key documents
  - List of data quality checks
  - Methodologies
  - Discussion
- http://code.google.com/p/ala-dataquality/

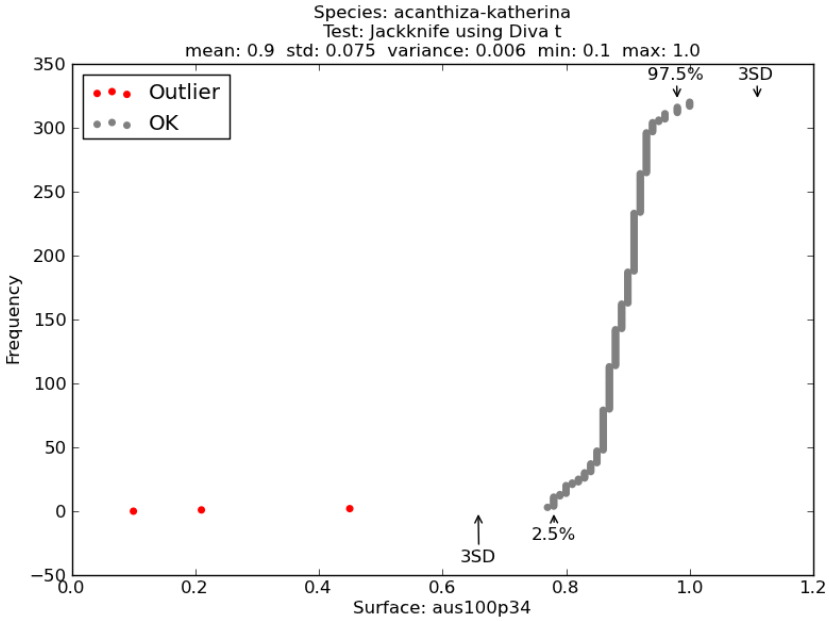
Spatial Precision Issues

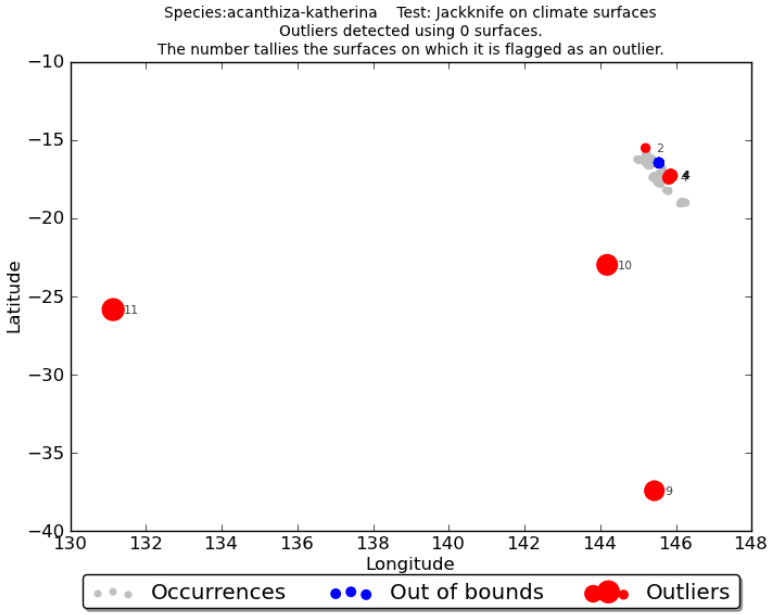
- Spatial precision and uncertainty, extent – What does it all mean?
- coordinateUncertaintyInMeters is overall error not just extent of location.
- Darwin core has not have an extent of location term – even though recommended georeferencing tools use it.
- All is now clear – but re-education needed.



Outlier - Jackknife

- http://code.google.com/p/ala-dataquality/wiki/DETECTED_OUTLIER_JACKKNIFE

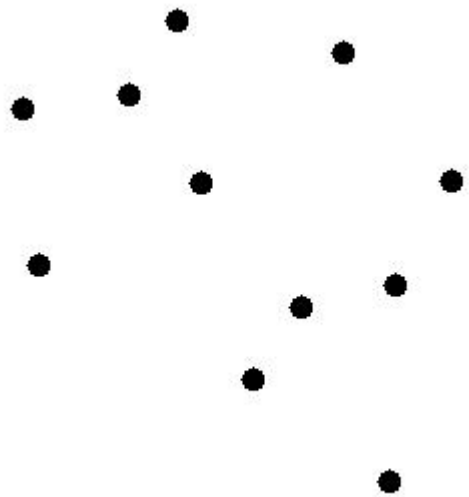




- Presently being implemented as an ALA Data quality check
- Prototyped in Python Data Quality Workbench – provided to ERIN

Outlier – Convex Hull

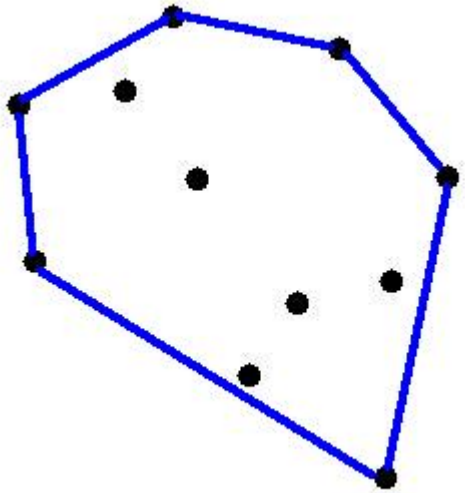
- With Ross Cunningham, ANU Fenner School



Convex hull peeling

From: Greg Aloupis - Université Libre de Bruxelles

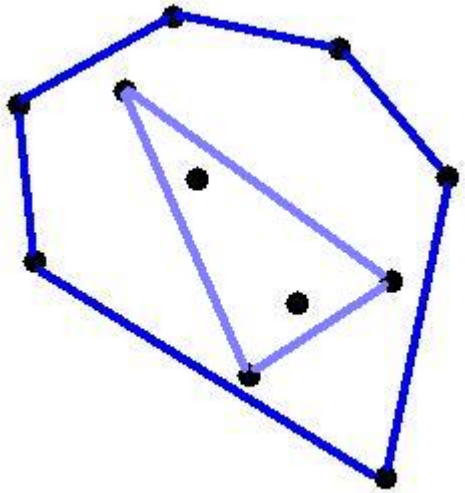
Outlier – Convex Hull



Convex hull peeling

Outliers tend to occur on first few hulls

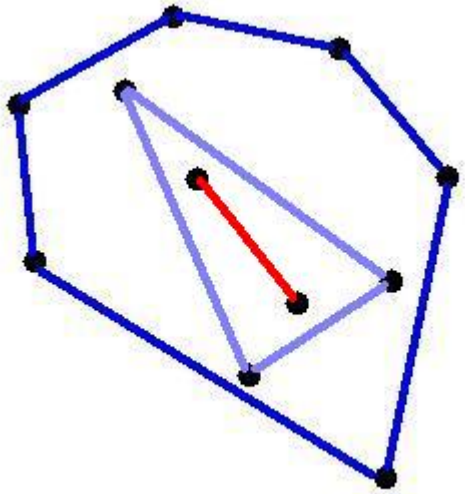
Outlier – Convex Hull



Convex hull peeling

Hull with 50% points inside is inter-quartile range

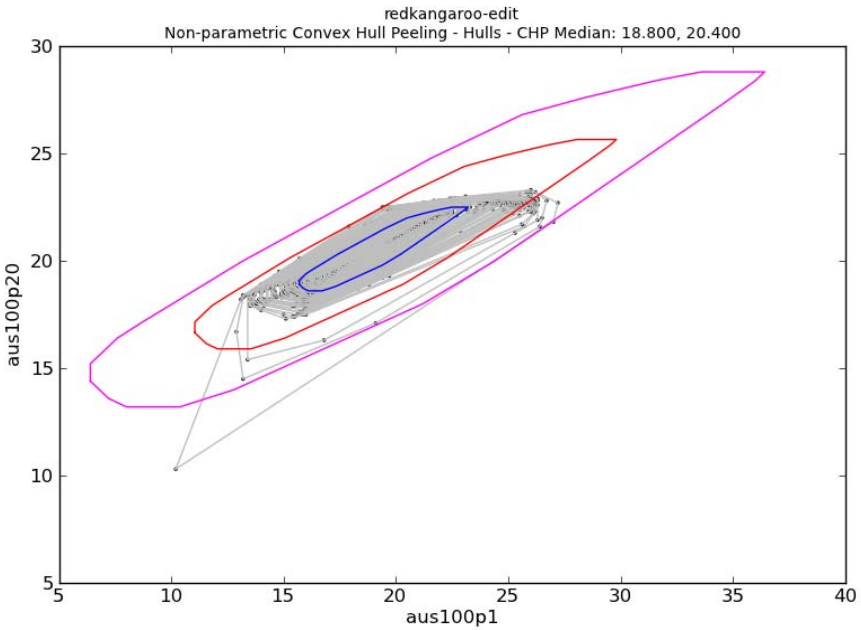
Outlier - Convex hulls

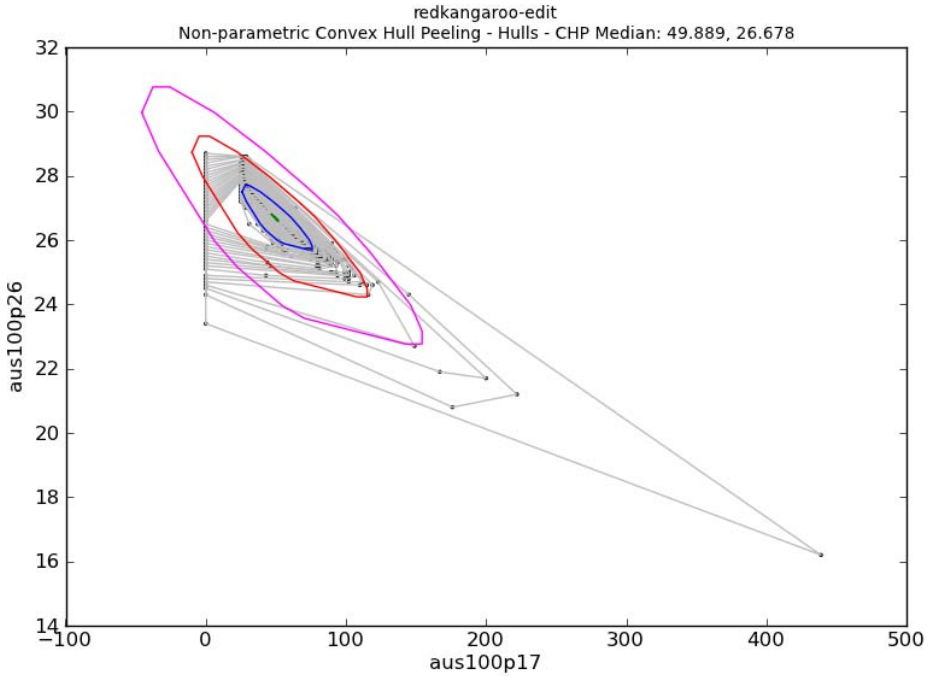


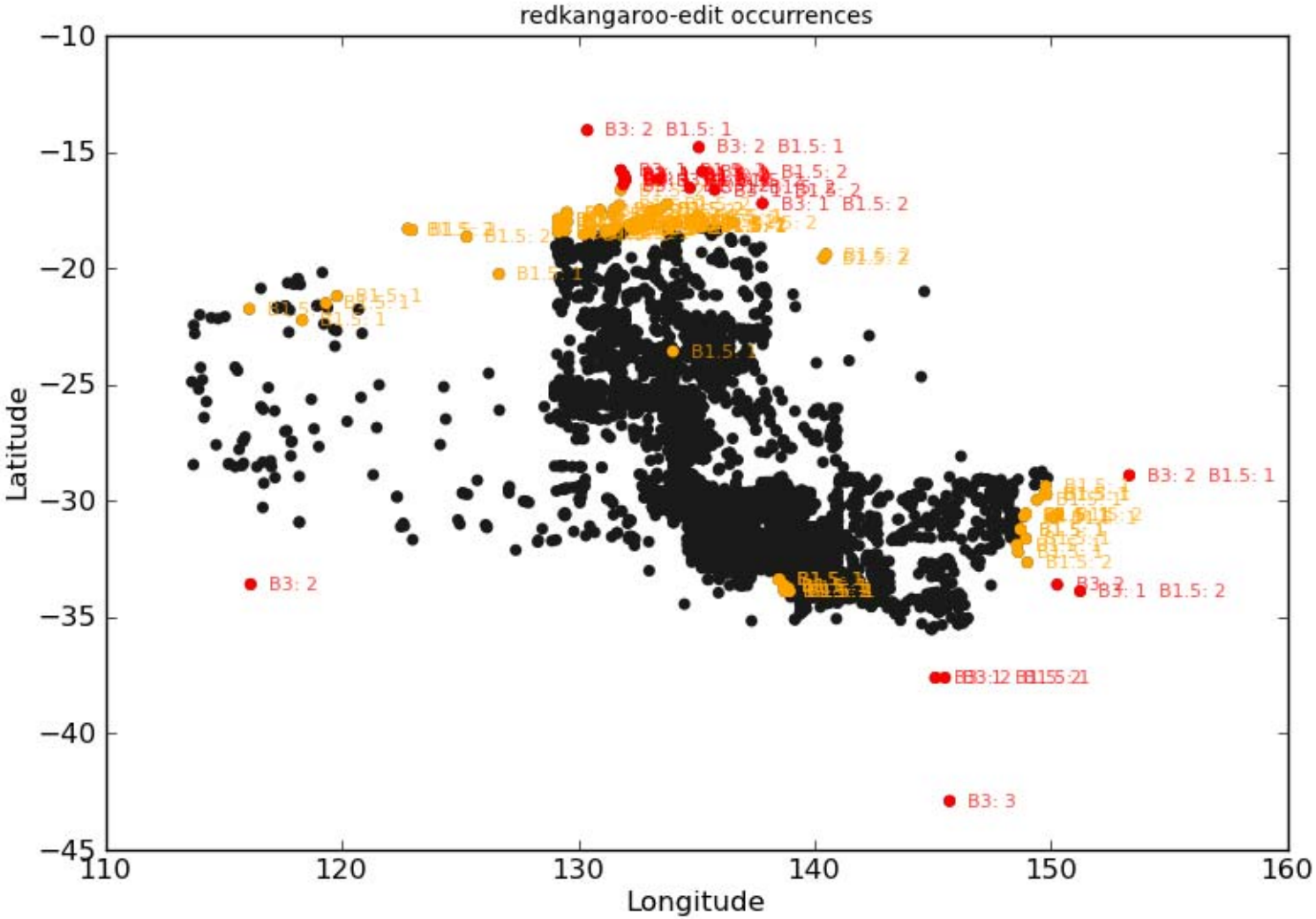
Convex hull peeling

Inner hull is median

Outlier – Convex Hull

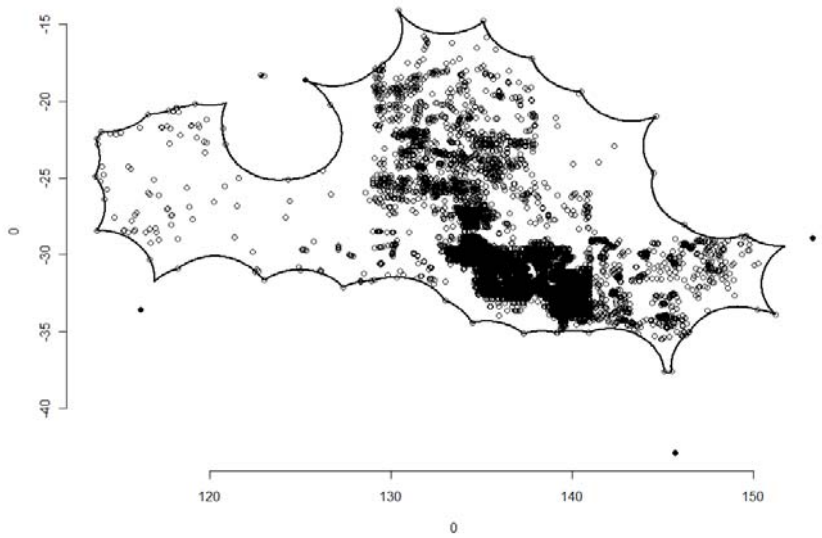




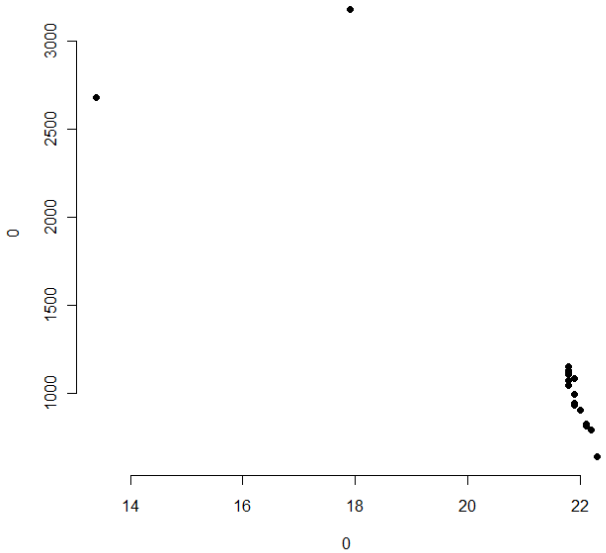

redkangaroo-edit occurrences
Latitude
Longitude
−10
−15
−20
−25
−30
−35
−40
−45
110
120
130
140
150
160
B3: 2 B1.5: 1
B3: 2 B1.5: 1
B3: 1 B1.5: 2
B3: 2 B1.5: 1
B3: 2
B3: 1 B1.5: 2
B3: 3

Outlier – Alpha Hull

- From Burgman and Fox 2003
- Being used by Birds Australia for mass mapping exercise.



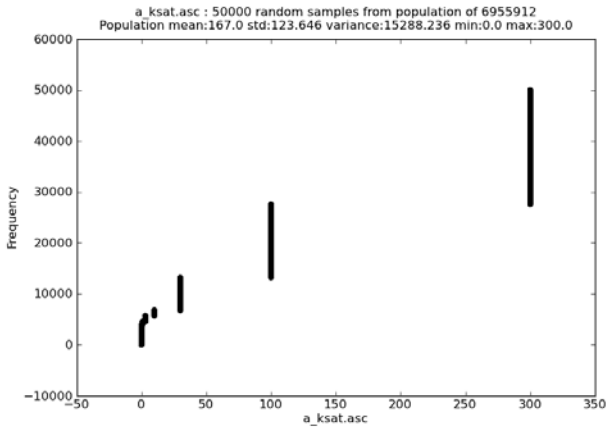
Red Kangaroo: lat/long

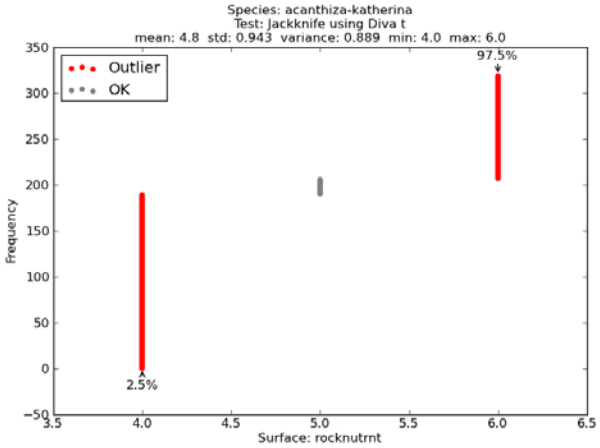


Black Grasswren: p12/20 – falls over

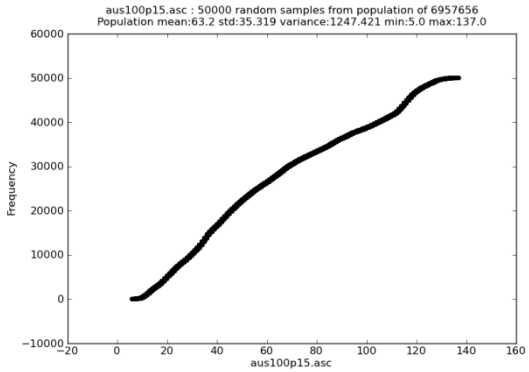
Environmental Surface Issues

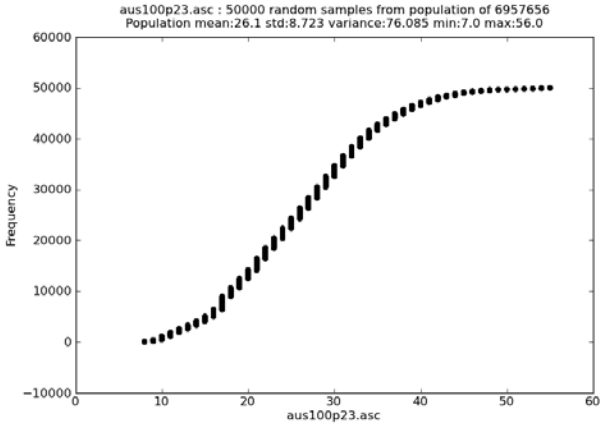
- Many 100's of environmental surface
- The various outlier checks exposing issue surfaces
- Will need an optimal set for each check

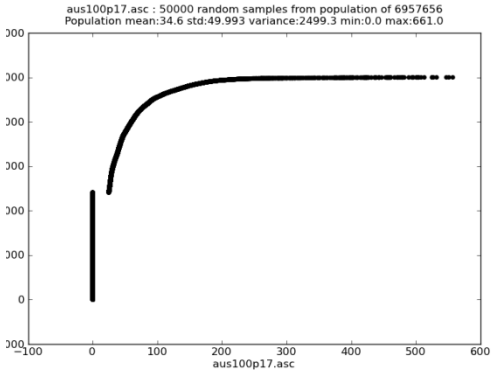


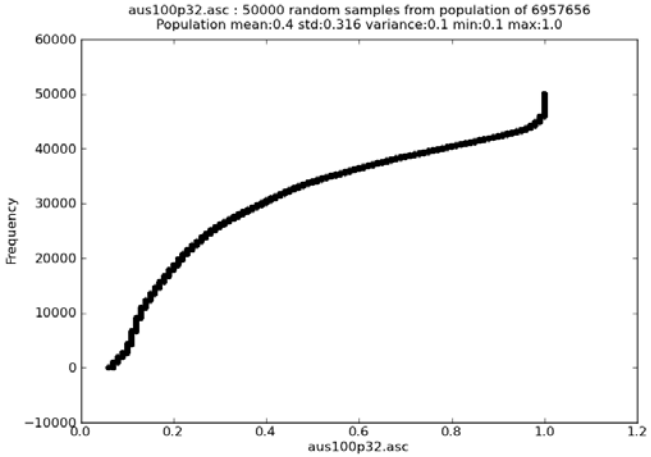


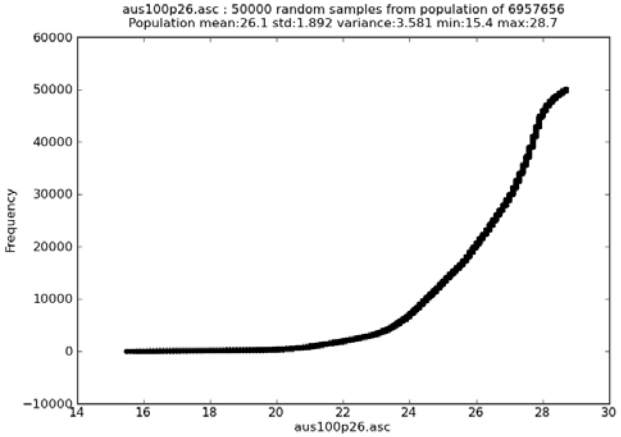
Lee's minimal set



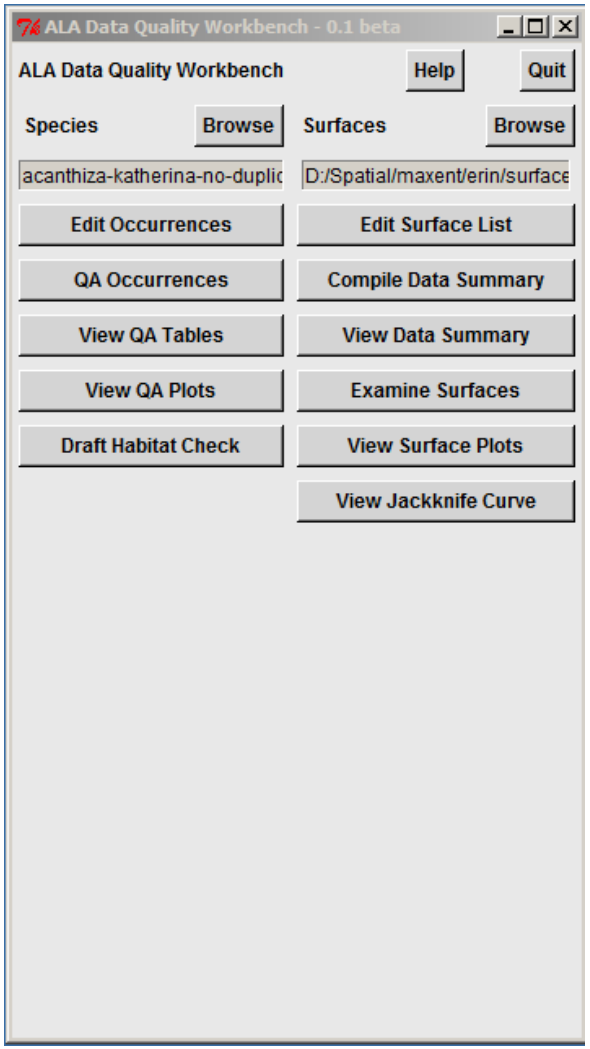








Python Tool


ALA Data Quality Workbench - 0.1 beta
ALA Data Quality Workbench
Help
Quit
Species
Browse
Surfaces
Browse
acanthiza-katherina-no-duplic
D:/Spatial/maxent/erin/surface
Edit Occurrences
Edit Surface List
QA Occurrences
Compile Data Summary
View QA Tables
View Data Summary
View QA Plots
Examine Surfaces
Draft Habitat Check
View Surface Plots
View Jackknife Curve

Outlier – Maxent Pair wise Analysis

With Simon Barry and Warren Muller, CSIRO Mathematics, Informatics and Statistics

Jack-knife occurrence records through Maxent:

a) for all data
b) removing each record in turn,
c) then removing each pair of records in turns

Obtain Maxent prediction for each occurrence for every combination

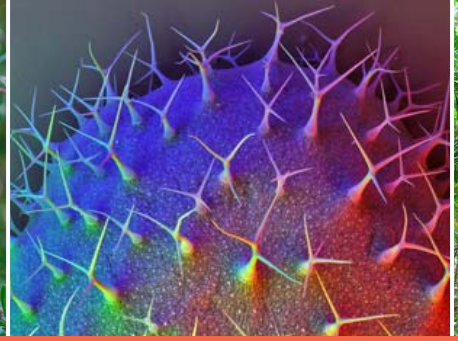
Mountain Thornbill – 189 records became 15000 Maxent runs!!

Output presently being analysed by CSIRO.

Next Steps

- Complete documentation of precision/uncertainty issues
- Propose extent of location be added as Darwin Core field
- Write up progress with outlier detection work and refine methodology e.g. remove outliers to minimise hull area
- Determine optimal environmental surfaces to use with each test
- Continue work with Data Quality Wiki
-





The Atlas of Living Australia Participants

www.ala.org.au

Australian museum — nature culture discover

MUSEUMVICTORIA

Queensland Government

qm queensland museum www.qm.qld.gov.au

Tasmanian Museum & Art Gallery

Tasmania — Explore the possibilities

Australian Government — Department of the Environment, Water, Heritage and the Arts

ABRS

CSIRO

Australian Government — Department of Agriculture, Fisheries and Forestry

THE UNIVERSITY OF ADELAIDE AUSTRALIA

CHAH — Council of Heads of Australasian Herbaria

Southern Cross UNIVERSITY — A new way to think

ATLAS OF LIVING AUSTRALIA — sharing biodiversity knowledge

The Council of Heads of Australian Faunal Collections (CHAFC)

The Council of Heads of Australian Entomological Collections (CHAEC)

The Australian Microbial Resources Research Network (AMRRN)

The Council of Heads of Australasian Museum Directors (CAMD)

An Australian Government Initiative

National Collaborative Research Infrastructure Strategy

The Atlas is funded by the Australian Government under the National Collaborative Research Infrastructure Strategy and further supported by the Super Science Initiative of the Education Investment Fund