Effects of Spatial Reference Systems on the Accuracy of Spatial Predictive Modelling along a Latitudinal Gradient

A. J. Turner[1], J. Li [1] and W. Jiang [1]

[1] *Geoscience Australia, GPO Box 378, Canberra, ACT 2601, Australia*
Email: Andrew.Turner@ga.gov.au

Abstract: Spatial predictive information is essential for applications such as modelling, planning, risk assessment and decision making. Previous studies have examined the complex factors that affect the performance of various spatial interpolation methods including sampling density, data variation, spatial structure of data, spatial distribution of samples, data quality, secondary variables and interaction among these factors. Another potential source of error is the Spatial Reference System (SRS) according to which the spatial information is stored. All SRSs generate spatial distortions of reality in some form, and this may impact on the accuracy of spatial predictions. Preliminary investigations into this issue found negligible differences in the predictive errors using different SRSs. However, these studies have tested regions restricted to the latitude extents of the Australian continent. The latitudinal gradient needs to be investigated for its effect on SRSs and any resulting impact on interpolation performance.

This study examines the effects of ten different SRSs on spatial predictions using three datasets in regions spanning a long latitudinal gradient from approximately 10°S to 70°S. The study areas are located off the northern coast of Australia, southern coast of Tasmania, and on a section of the Antarctic coast. The tested SRSs consist of two geographic coordinate systems (i.e., WGS84 and GDA94), six map projections selected for minimal spatial distortion for datasets located in Australia (i.e., Equal-Area Azimuthal, Equidistant Azimuthal, Stereographic Conformal Azimuthal, Albers Equal-Area Conic, Equidistant Conic and Lambert Conformal Conic), and two map projections selected for minimal spatial distortion for datasets located in Antarctica (i.e., Australian Antarctic Lambert and Australian Antarctic Polar Stereographic). Continuous spatial predictions of seabed sand sediment information were built in the SRSs using two interpolation methods: inverse distance weighting (IDW) and ordinary kriging (OK). Prediction accuracy was measured using leave-one-out cross validation, and assessed in a number of error measures including relative mean absolute error and paired Mann-Whitney tests.

The results show insignificant differences in predictive errors between all tested SRSs over the studied latitudinal gradient. This is the case for comparisons between WGS84 and GDA94, geographic coordinate systems and map projections, and map projection classifications. Insignificant difference was also observed between the map projections selected for Australia compared to the map projections selected for Antarctica, even when the Australian map projections were tested on the Antarctic study area and vice versa. There was a high level of variation in map distortion occurring between SRSs, but this did not equate to variation in interpolation error. It is argued that this is due to the nature of projecting data; distortion is applied consistently to all data points and thus the spatial relationship between data points is preserved in terms of their relative distance within the study area, resulting in similar prediction errors.

The negligible levels of variation between SRSs in terms of predictive error suggest that WGS84 is sufficient for spatial predictive modelling. This removes the need for extensive projection selection and the associated transformation tasks, improving modelling efficiency.

Keywords: *Spatial Reference System (SRS), map projection, spatial predictive modelling, inverse distance weighting (IDW), ordinary kriging (OK), predictive accuracy*

1. INTRODUCTION

Spatial data of environmental properties are often collected as sparsely distributed point samples. This is especially true in deep marine regions, where data collection is limited by time and resource constraints. Spatial interpolation techniques use point samples to predict unsampled locations, building a continuous data layer. These continuous predictive layers are essential for applications such as modelling, planning, risk assessment and decision making (Whiteway et al. 2007, Li et al. 2011, Pitcher et al. 2008). Therefore, it is important to seek improvements to the performance of spatial interpolation techniques which will lead to superior technical and scientific advice. Previous studies have examined the complex factors affecting the performance of various spatial interpolation methods. Li and Heap (2008) provide an in-depth discussion into the effects of sampling density, data variation, and sampling design on spatial interpolation methods. Further explored factors are the spatial structure of data, spatial distribution of samples, data quality, secondary variables, the correlation between primary and secondary variables, and interaction among these factors (Li and Heap, 2008, Li et al. 2011).

Another factor which may influence spatial interpolation performance is the spatial reference system (SRS), in which the interpolating data is stored. SRSs are the frameworks which define locations on the surface of the earth, and may be either a geographic coordinate system (GCS), or a projected coordinate system (PCS), also known as a map projection (Jiang and Li, 2013). A GCS uses a three-dimensional spheroid approximately the shape of the earth, with angular units of measure to reference locations on the earth's surface (Esri, 2016a). Map projections are the two-dimensional representations of the earth's surface, used for creating maps (Snyder, 1987).

Interpolation accuracy issues arise for both GCSs and map projections. Practical spatial interpolation methods assume input coordinates are uniform in distance. This is a problem in GCSs, which use longitude and latitude as input coordinates, as one degree of longitude at the equator has a greater distance than one degree of longitude near one of earth's poles (Jiang and Li, 2013). Interpolation concerns also exist for map projections, as all map projections introduce distortions to spatial properties when transforming the non-planar Earth onto a planar map (Tissot, 1881). Different classifications of map projections preserve certain spatial properties, while distorting others. Conformal projections persevere local shape, equal-area projections preserve area, equidistant projections preserve the distances between certain points, and azimuthal projections preserve direction from one point to all other points (Jiang and Li, 2014). The distortion of spatial properties also varies depending on the location on the map. Map projections are constructed with points (or lines) of contact which are tangential to the globe (Kennedy and Kopp, 2000). These points or lines are significant because they define the location of zero distortion. Generally, spatial distortion will increase with greater distance from the point of contact (Kennedy and Kopp, 2000).

Preliminary investigations into these interpolation issues by Jiang and Li (2013) and Buckerfield et al. (2014) found negligible differences in the predictive errors between different SRSs. However, these studies have been restricted to the latitude extents of the Australian continent, and regions at more extreme latitudes have not yet been tested. Thus a further test is warranted.

In this study we test the effect of different SRSs on interpolation predictions along a long latitudinal gradient. To accomplish this, four questions are investigated: (1) is there a difference in interpolation performance between WGS84 and GDA94 over the latitudinal gradient? (2) Is there a difference in interpolation performance between GCSs and map projections over the latitudinal gradient? (3) Is there a difference in interpolation performance between map projection classifications over the latitudinal gradient? (4) Is there a difference in interpolation performance between Australian and Antarctic map projections over the latitudinal gradient?

2. MATERIALS AND METHODS

2.1. Study Datasets

This study was conducted using point samples of seabed sediment properties from the Marine Sediments (MARS) database (www.ga.gov.au/applications/marine-sediments-database) at Geoscience Australia. The MARS database contains detailed information on seabed sediments from over 300 marine surveys conducted in Australia's marine jurisdiction, including the Australian Antarctic Territory (Heap, 2009).

Three study areas spanning a long latitudinal gradient from 9°S to 69.4°S were selected: Northern Australia Region (NAR), Southern Tasmania Region (STR), and Australian Antarctic Region (AAR) (Figure 1). These locations were selected to allow testing over a long latitudinal gradient. Sediment samples in MARS were

extracted and cleaned for these three regions. The NAR dataset extent covers an area of 181 261 km^2 with a total of 91 point samples. The STR dataset is the smallest study area, with an area of 82 181 km^2 and 36 point samples. The AAR covers an area of 167 854 km^2, with 138 points samples used for interpolation.
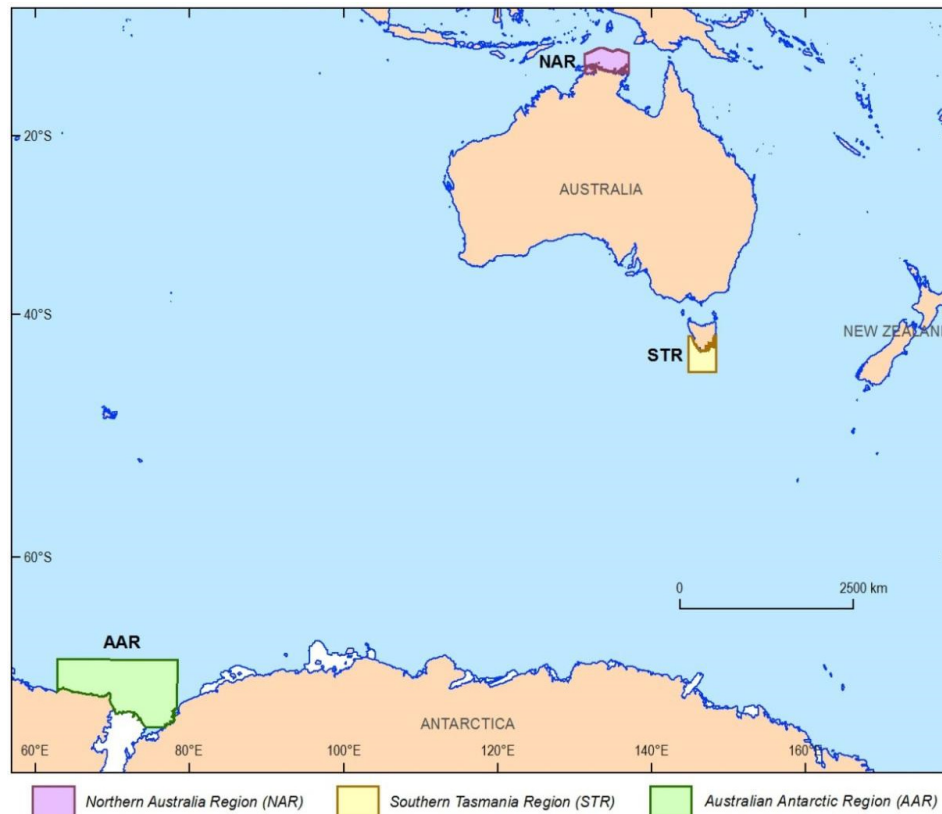

Figure 1. Location of the three study areas (NAR, STR and AAR). Study areas were selected to span over a wide latitudinal gradient and where sufficient MARS data was available.

2.2.   Spatial Reference System Selection

The selection of SRSs in this study was based on the previous work by Jiang and Li (2013). They performed an extensive selection process to find two GSCs and six map projections, suitable for continental Australia with a focus on the southwest Australian Exclusive Economic Zone (AEEZ). Buckerfield *et al.* (2014) used the same eight SRSs for their study focusing on the entire Australian Continent. The two GCSs are six map projections respectively were WGS84, GDA94, Albers Equal-Area Conic, Lambert Equal-Area Azimuthal, Lambert Conformal Conic, Equidistant Azimuthal, Equidistant Conic, and Stereographic Conformal Azimuthal. The six aforementioned map projections have points (or lines) of contact set to minimise distortion over the Australian continent and thus are referred to as Australian map projections in this study. However, to account for the AAR study area, a further selection process using Young's Rule (Young, 1920) and Snyder's decision tree (Snyder, 1987) was utilized to select two additional projections with points of contact set to minimise distortion in Antarctica: Australian Antarctic Lambert and Australian Antarctic Polar Stereographic. These two projections will be referred to as the Antarctic map projections.

2.3.   Spatial Interpolation Method Selection

Two commonly used spatial interpolation methods were selected for this study: inverse distance weighting (IDW) and ordinary kriging (OK). These selections were based on their use in relevant past studies (Jiang and Li, 2013, Buckerfield et al. 2014) and common use in environmental sciences (Li and Heap, 2008, Li et al 2011, Li and Heap, 2011).

2.4.   Model Methods

Three datasets were converted from WGS84 into a further nine SRSs, resulting in a total of 30 datasets. Two interpolation techniques, IDW and OK, were applied to each of the 30 datasets, building 60 spatial predictive models. The workflow for the two interpolation methods varied slightly, but was consistent for all SRSs and

study areas (Figure 2). For the OK models, an arcsine transformation was used to normalize the data, and the resultant predictions were then transformed back for the accuracy assessment (Li et al. 2011).
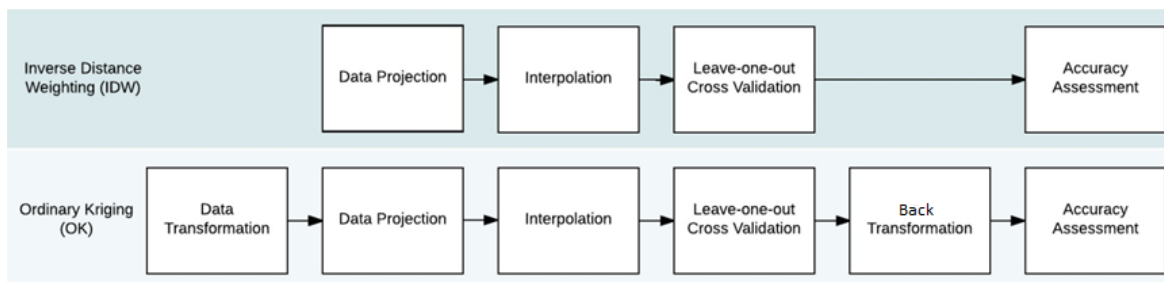

Figure 2. The prediction model workflows varied slightly between interpolation techniques.

The IDW model search window was set at 20 with a power parameter of 2. The OK search window was also set to 20, and a spherical model type was used. Model parameters, such as the lag size, nugget, partial sill and range were optimised using the Geostatistical Analyst Extension inbuilt function. This function is based on minimising the mean square error of the cross validation output (Esri, 2016b).

The interpolation models were tested using leave-one-out cross validation. This test removes one data point at a time, and predicts the value for the removed point based on the remaining data points in the model (Esri, 2007). These results are further explored with a variety of error measures to assist in the accuracy assessment. These measures consist of mean absolute error (MAE), root mean square error (RMSE), relative mean absolute error (RMAE), relative root mean square error (RRMSE) and variance explained by predictive models based on cross-validation (VEcv) (Li and Heap 2008, Li 2016 and Li 2017). Only the results of RMAE were reported in this paper due to space limitations; other measures are available in Turner et al. (2018). The differences in RMAE between WGS84 and all other SRSs were tested using paired Mann-Whitney tests (R Development Core Team, 2016).

3.   RESULTS

The predictive errors of the IDW and OK methods for each tested SRS over the three study areas in terms of RMAE and paired Mann-Whitney test (p-value) are presented in Table 1. The p-value was not included for GDA94 as its RMAE was consistently equal to WGS84.

Table 1. The predictive errors in RMAE and the difference to WGS84 as the calculated p-value of the IDW and OK methods for each tested SRS over the three study areas.

Spatial Reference System	NAR				STR				AAR			
	RMAE (%)		P- value		RMAE (%)		P- value		RMAE (%)		P- value	
	IDW	OK	IDW	OK	IDW	OK	IDW	OK	IDW	OK	IDW	OK
WGS84	31.18	32.73	-	-	25.62	25.42	-	-	45.67	46.37	-	-
GDA94	31.18	32.73	-	-	25.62	25.42	-	-	45.67	46.37	-	-
Albers Equal-Area Conic	31.04	32.64	0.33	0.88	24.87	25.59	0.06	0.74	45.65	45.80	0.84	0.06
Lambert Equal-Area Azimuthal	31.17	32.72	0.36	0.13	24.92	25.50	0.01	0.51	46.12	48.51	0.38	0.13
Lambert Conformal Conic	31.18	32.71	0.16	0.82	24.75	25.66	0.08	0.94	46.03	48.47	0.48	0.10
Equidistant Azimuthal	31.17	32.70	0.15	0.90	24.85	25.56	0.03	0.39	45.99	46.31	0.62	0.20
Equidistant Conic	31.14	32.73	0.39	0.30	24.80	25.56	0.07	0.13	45.83	47.58	0.58	0.08
Stereographic Conformal Azimuthal	31.18	32.78	0.20	0.08	24.75	25.59	0.08	0.51	46.03	48.44	0.67	0.10
Australian Antarctic Lambert	31.18	33.55	0.25	0.96	24.75	25.59	0.08	0.20	46.03	48.35	0.48	0.12
Australian Antarctic Polar Stereographic	31.18	32.66	0.27	0.49	24.75	25.59	0.08	0.23	46.03	48.36	0.48	0.12

For all study areas, WGS84 and GDA94 are shown to have the exact same results (Table 1). The variation between GCSs (WGS94 and GDA94) and map projections was not significant ($p \geq 0.05$) (Table 1). The

Lambert Equal-Area Azimuthal and Equidistant Azimuthal results for the STR dataset display what appear to be significantly different values from WGS84 ($p < 0.05$), however the actual differences in RMAE from WGS84 are not as large as some other SRSs such as Lambert Conformal Conic. Therefore, their variations are considered negligible. The comparison of RMAE for map projection classifications shows insignificant difference for IDW and OK methods over the latitudinal gradient (Table 1). Similarly, the comparison of Australian and Antarctica map projections displays insignificant difference for each interpolation technique over all study areas (Table 1).

The visual assessment on the prediction layers showed that the spatial distortion occurred when applying the Antarctic map projections to the NAR and STR study areas, and when applying the Australian map projections to the AAR study area. While the principal scale remained the same for each prediction map, the scale factor clearly changed. However, despite the apparent spatial distortion, the interpolation results appeared remarkably similar. A subset of the SRSs for the AAR IDW predictions, shown in Figure 3, demonstrates this clearly.
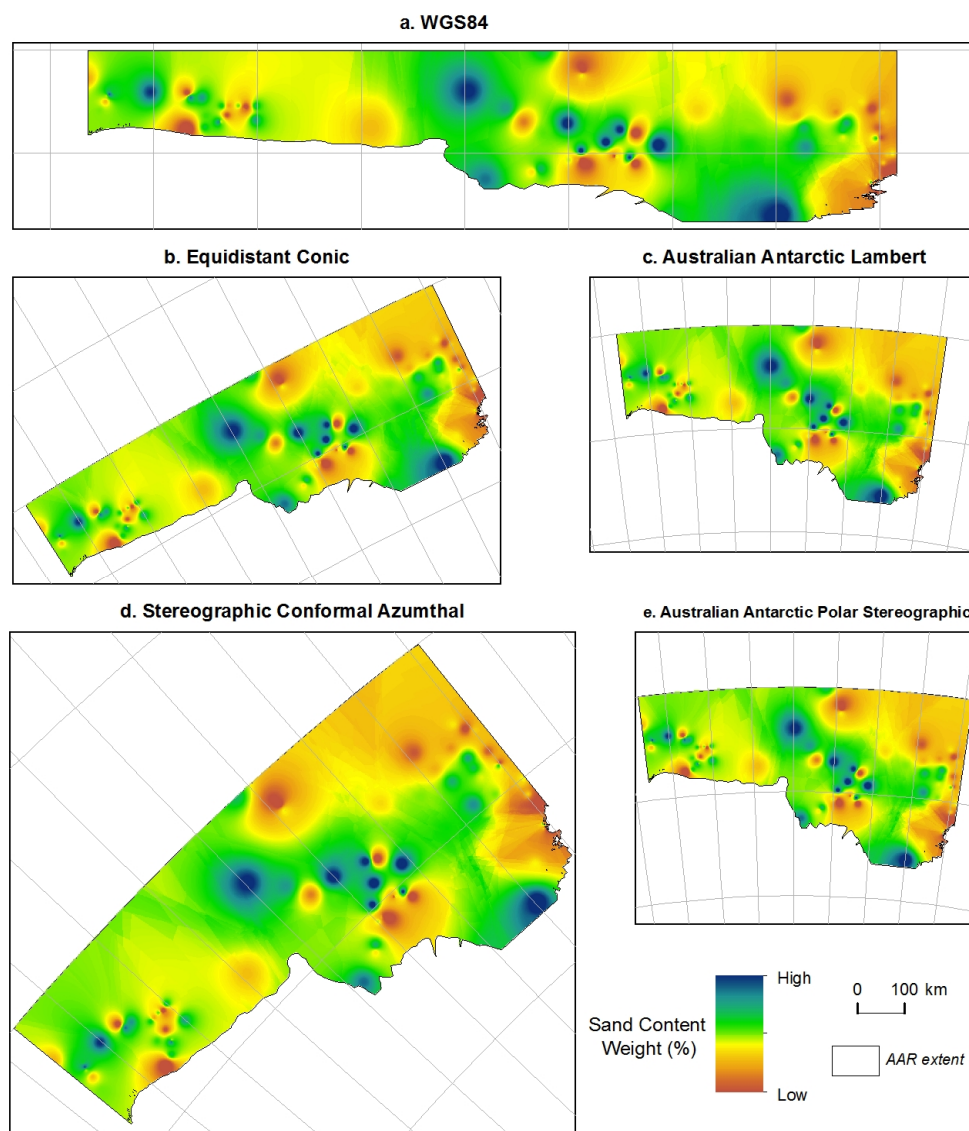

Figure 3. IDW spatial predictions of the AAR study region for five of the ten tested SRSs.

4. DISCUSSION

4.1. WGS84 and GDA94 Performance Comparison

WGS84 and GDA94 are GCSs with slight differences due to their inverse flattening parameters and increasing disparity due to tectonic movement. The study results found no differentiation between the two in

any of the three study regions in terms of predictive accuracy. This suggests that the variation between the two GCSs has no impact on interpolation results at the cell size used in the study.

4.2. GCS and Map Projection Performance Comparison

Common interpolation techniques assume equal distances over changes of latitude and longitude, and this becomes increasingly false at locations further from the equator. This may be a cause of prediction error in GCS interpolation models. If the disparity in distance is a significant cause of error when interpolating in GCSs, then WGS84 and GDA94 would be expected to show higher error compared to the map projections in the study areas further away from the equator. The results did not show this pattern. There was a high level of spatial distortion evident in the GCSs, but it did not significantly impact on the interpolation accuracy.

4.3. Map Projection Classifications Performance Comparison

Different map projection classifications preserve some spatial properties and distort others. The tested map projections covered a range of classifications (i.e., conic, azimuthal, conformal, equidistant and equal-area). The comparison results showed insignificant difference in error values. The use of Young's Rule (Young, 1920) and Snyder's decision tree (Snyder, 1987) to select map projection classifications with minimal distortion was ineffective for the purpose of minimising interpolation error.

4.4. Australian and Antarctic Map Projections Performance Comparison

Interpolating with the Antarctic map projections in the NAR and STR datasets, which are based in Australia, induced a high level of distortion, which was visually evident in the prediction maps. Similarly, the Australian map projections tested in the AAR dataset also displayed high levels of distortion (Figure 3).

The results showed no clear correlation between spatial distortion level and interpolation accuracy. The map projections with minimal distortions did not perform significantly better than the map projections with high distortion, and in many cases the highly distorted map projections resulted in less interpolation error, such as the Australian Antarctic Polar Stereographic projection in the OK model of the NAR dataset (Table 1).

The reason for this may lie within the mechanics of projection transformations and interpolation methods. When projected, the data points in a study area are displaced; the distances between points change, as do the angles. However, the changes that undergo each point are consistent within each projection. A point may move further away from another point, but it will also move away from all other points in a similar proportional manner. Despite the distortions, the spatial relationship between all points remains similar, which produces similar interpolation results. The relationships must change partially to give variation in results between projections, but a change due to distortion may result in an increase or decrease in prediction accuracy. Different levels of distortions are occurring between GCSs and map projection classifications but this does not necessarily correlate to higher or lower interpolation error.

Overall, the variations between interpolation results were negligible and therefore we recommend SRS selection and projection can be removed for interpolation applications. Data points can be interpolated in WGS84 or the SRS they are originally stored in. This removes the projection selection and transformation processes, improving workflow efficiency.

5. CONCLUSIONS

This study examined the variation of interpolation accuracy of IDW and OK methods between different SRSs over three distinct regions spanning a latitudinal gradient. The key findings of the study which answer the proposed questions in the introduction are as follows:

1. It was found that the slight differences between WGS84 and GDA94 did not affect interpolation accuracy along the latitudinal gradient.
2. The results did not show GCSs inducing significantly higher or lower interpolation error than map projections over the latitudinal gradient.
3. The comparison of the error values of the map projections in various classifications showed insignificant difference between all classifications over the latitudinal gradient.
4. There was no significant variation in the performance of Australian and Antarctic map projections over the latitudinal gradient. Map projections with low spatial distortions did not outperform the map projections with high spatial distortion.

The variations between interpolation accuracy results are insignificant. Therefore, for studies at a similar level of precision, data points can be interpolated in WGS84 or the SRS they are originally stored in, which removes projection selection and associated transformation tasks, and improves modelling efficiency.

ACKNOWLEDGMENTS

REFERENCES

Buckerfield, S., Li, J., Sanabria, A. and Jiang, W. (2014). *A Further Study on the Effects of Spatial Reference Systems on the Predictive Accuracy of Spatial Interpolation Methods.* Geoscience Australia (Unpublished).

Esri (2007). Performing cross-validation and validation. http://webhelp.esri.com/arcgisdesktop/9.2/index.cfm?id=3355&pid=3334&topicname=Performing_cross-validation_and_validation.

Esri (2016a). What are geographic coordinate systems? http://desktop.arcgis.com/en/arcmap/latest/map/projections/about-geographic-coordinate-systems.htm.

Esri (2016b). Parameter optimization. http://desktop.arcgis.com/en/arcmap/10.3/guide-books/extensions/geostatistical-analyst/parameter-optimization.htm.

Heap, A. D. (2009). Marine Sediments (MARS) Database: Commonwealth of Australia (Geoscience Australia).

Jiang, W. and Li, J. (2013). Are spatial modelling methods sensitive to spatial reference systems for predicting marine environmental variables? Paper presented at 20th International Congress on Modelling and Simulation, Adelaide, Australia, 1–6 December 2013.

Jiang, W. and Li, J. (2014). *The effects of spatial reference systems on the predictive accuracy of spatial interpolation methods*. Canberra: Geoscience Australia. Record 2014/001.

Kennedy, M. and Kopp, S. (2000). *Understanding Map Projections.* Redlands, California: ESRI.

Li, J. and Heap, A. (2008). *A review of spatial interpolation methods for environmental scientists*. Canberra: Geoscience Australia. Record 2008/23.

Li, J. and Heap, A. (2011). A review of comparative studies of spatial interpolation methods in environmental sciences: Performance and impact factors. *Ecological Informatics*. **6**, 228-241.

Li, J., Heap, A., Potter, A. and Daniell, J. (2011). Application of machine learning methods to spatial interpolation of environmental variables. *Environmental Modelling and Software*. 26, 1647-1659.

Li, J., Heap, A. D., Potter, A., Huang, Z. and Daniell, J. (2011). Can we improve the spatial predictions of seabed sediments? A case study of spatial interpolation of mud content across the southwest Australian margin. *Continental Shelf Research* 31, 1365-1376.

Li, J. (2016) Assessing spatial predictive models in the environmental sciences: accuracy measures, data variation and variance explained. *Environmental Modelling & Software*. 80 1-8.

Li, J. (2017) Assessing the accuracy of predictive models for numerical data: Not r nor r^2, why not? Then what? *PLOS ONE*. 12 (8): e0183250.

Pitcher, C. R., Doherty, P. J. and Anderson, T. J. (2008). *Seabed environments, habitats and biological assemblages.* Collingwood: CSIRO Publishing.

R Development Core Team (2016). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Snyder, J. P. (1987). Map projections - A working manual. *USGS Professional Paper 1395*.

Tissot, N. A. (1881). *Mémoire sur la représentation des surfaces et les projections des cartes géographiques*. Paris, France: Gauthier-Villars.

Turner, A., Li, J. and Jiang, W. (2018) *A further study on the effects of spatial reference systems on the accuracy of spatial predictive methods along a long latitudinal gradient*. Geoscience Australia, Record 2018/xxx (in press).

Whiteway, T., Heap, A. D., Lucieer, V., Hinde, A., Ruddick, R. and Harris, P. T. (2007). *Seascapes of the Australian margin and adjacent sea floor: Methodology and results*. Canberra: Geoscience Australia. Record 2007/11.

Young, A. E. (1920) Some Investigations in the Theory of Map Projections. *Technical Series No.1*, London Royal Geographical Society.