Are Spatial Modelling Methods Sensitive to Spatial Reference Systems for Predicting Marine Environmental Variables?

Wenping Jiang[a] and Jin Li[a]

[a] *Geoscience Australia, GPO Box 378, Canberra, ACT 2601, Australia*
Email: wenping.jiang@ga.gov.au

Abstract: Spatially continuous information is often required for environmental planning and conservation. Spatial modelling methods (SMMs) are essential for generating such information from point data. The accuracy of spatial predictions is crucial for evidence-based decision making but is potentially affected by many factors. For example, spatial reference systems may alter the features of spatial data and thus are likely to affect predictions from SMMs. However, the degree to which such systems can affect spatial predictions has not been examined. It is also not clear whether different types of SMMs respond differently to the choice of a spatial reference system.

In this study, we aim to test how sensitive SMMs are to different spatial reference systems. On the basis of a review of spatial reference systems, we selected eight systems that are suitable for spatial predictions of marine environmental variables in the continental Australian Exclusive Economic Zone (AEEZ). These systems include two geographic coordinate systems (WGS84 and GDA94) and six map projections (Lambert Equal-Area Azimuthal, Equidistant Azimuthal, Stereographic Conformal Azimuthal, Albers Equal-Area Conic, Equidistant Conic and Lambert Conformal Conic). Two most commonly used spatial interpolation methods, i.e. inverse distance squared (IDS) and ordinary kriging (OK), were applied to a seabed sediment dataset that was projected using the eight systems. The accuracies of the methods were assessed using leave-one-out cross validation in terms of their predictive errors (mean absolute error and relative mean absolute error). The spatial prediction maps were also generated and visualised for comparison. The differences in the predictive errors resulted from WGS84 and the map projections were compared using paired Mann-Whitney test for both IDS and OK. The data manipulation and modelling work were implemented in ArcGIS and R.

Results from this study show that whether the data is projected on spherical surfaces based on the geographic coordinate systems or on planar surfaces based on the map projections, the accuracies of the SMMs (IDS and OK) in predicting seabed sediment data in the southwest region of AEEZ are similar and the differences are considered negligible, in terms of both predictive errors and prediction map visualisations. Thus, it is concluded that the SMMs examined are not sensitive to the spatial reference systems tested for spatial predictions of seabed sediment data in the southwest region of AEEZ. However, a few factors may potentially alter the degree to which spatial reference systems affect spatial predictions, e.g. search window size, data density, data spatial distribution and dataset location. Hence further work is required to test different datasets located in other regions and to test a variable search window size.

The outcomes of this study have significant implications for spatial predictions in environmental science. The results suggest that spatial predictions using datasets with a density comparable to or greater than that in this study may use WGS84 directly and may not have to project data on a certain spatial reference system. This would greatly increase data processing efficiency. The findings are applicable to spatial predictions of both marine and terrestrial environmental variables.

Keywords: *Spatial reference systems, inverse distance squared (IDS), ordinary kriging (OK), geostatistics, Australian Exclusive Economic Zone (AEEZ)*

1.   INTRODUCTION

Spatially continuous data are important information for planning, risk assessment and decision making in environmental management and conservation. However, they are often difficult and expensive to acquire, especially for mountainous and deep marine regions (Li et al., 2011b). Therefore, spatial modelling methods (SMMs) become essential tools for generating spatially continuous data from point data collected at sparsely and unevenly distributed locations. Previous studies have examined a number of factors that affect the performance of SMMs. These factors include sample density, data variation (Li et al., 2010, Li et al., 2011c, Li and Heap, 2011), spatial structure of data, sampling design, spatial distribution of samples, data quality, correlation of primary and secondary variables, and interaction among some of these factors (Li and Heap, 2008). Apart from these factors, a spatial reference system used to define sample locations is another potential factor and is worth investigating.

A majority of SMMs treat the spatial dimension as planar although spatial data are located on the non-planar surface of the Earth. In a geographic coordinate system, longitude and latitude are not uniform measurement units; however, SMMs assume that they are in some planar system and ignore the changes in distance along the latitude. Li et al. (2010, 2011c) propose in their series of spatial modelling studies that data should be appropriately projected such that the distance in the x axis reflects approximately the same distance in the y axis. However, as Tissot (1881) pointed out, there is no map projection that can transform the non-planar Earth surface onto a planar map surface without introducing distortions in spatial properties, i.e. distance, area, shape and direction. The projection of data from geographic coordinates to a map is, therefore, subject to significant variation and error. Results from a perceptual land-area estimation study (Battersby, 2009) show that the impact of projection distortion on perceived land areas can be substantial. It has also been found that large interpolation errors can be produced when interpolating over large areas of the Earth (Robeson, 1997). Willmott et al. (1985) conducted a sensitivity study to investigate errors on small-scale climate maps caused by the common practice of interpolation and contouring, using data drawn from 100 irregularly-spaced weather stations over the western half of the northern hemisphere. They developed two algorithms based on Shepard's (1984, 1968) well-known local-search interpolation function to perform the interpolation and contouring process both on the surface of sphere and in Cartesian planar space. They found that planar interpolation methods can produce interpolation errors as large as 10 °C in terms of mean annual air temperature. A similar study (Usery and Seong, 2000) in which four equal-area map projections were compared for generating regional and global raster data, indicates that the accuracy varies with the projection type, the latitude of the location, and the raster resolution.

In this study, we aim to examine the effects of spatial reference systems on the predictive accuracy of SMMs in predicting spatial pattern of seabed sediment properties for the southwest region of the continental Australian Exclusive Economic Zone (AEEZ). The main objective is to quantify how sensitive SMMs are to different spatial reference systems, and to determine whether different types of SMMs respond differently to the choice of a spatial reference system.

2.   MATERIAL AND METHOD

2.1.   Dataset

The study was undertaken using point datasets of seabed sediment properties for the continental AEEZ stored in the Marine Samples Database (MARS) at Geoscience Australia. Details of data collection, processing and cleaning can be found in Li et al. (2010), Li et al. (2011a) and Li et al. (2011d). After data filtering and quality control, 4,817 samples with mud content information were identified as available for modeling. Samples in the southwest region of AEEZ were then selected to test the sensitivities of SMMs to spatial reference systems in predicting the mud content of seabed. A total of 177 cleaned samples with a density of 0.34 samples per 1000 km^2 were used (Li et al., 2011a). The region covers an area of 523,000 km^2 with water depths ranging from 0 to 5,539 m, and comprises four geomorphic provinces (Heap and Harris, 2008). Spatial distribution of the samples is shown in Fig. 1.

2.2.   Selection of spatial reference systems

There are two types of spatial reference systems, namely geographic coordinate systems and projected coordinate systems, also known as map projections. A geographic coordinate system defines locations on the Earth using a three-dimensional spherical surface that approximates the shape of the Earth while map projections provide various mechanisms to project the Earth's spherical surface onto a two-dimensional planar surface for creating maps (Snyder and Voxland, 1989).
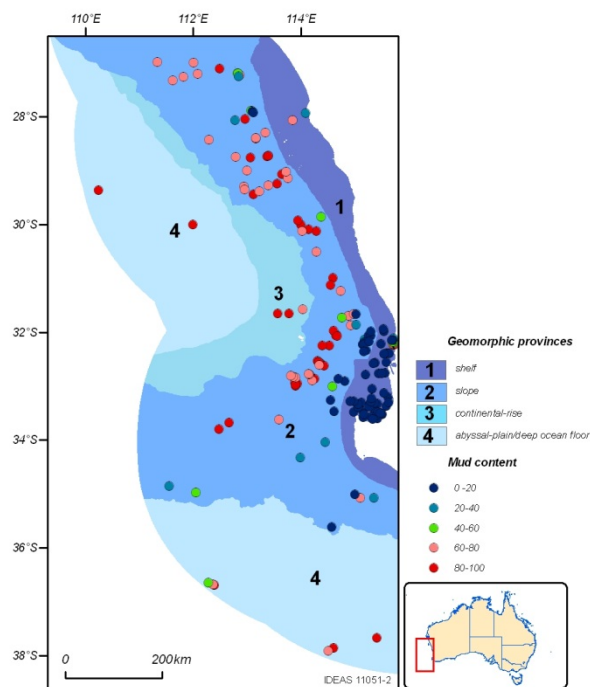
Figure 1. Spatial distribution of mud samples in the southwest region of AEEZ (Li et al., 2011b).

Two widely used geographic coordinate systems are chosen for this study: the World Geodetic System of 1984 (WGS84) and the Geodetic Datum of Australia 1994 (GDA94). WGS84 serves as a standard framework for locational measurement worldwide (Kennedy, 1989) while GDA94 was developed specifically for Australia. Both WGS84 and GDA94 use the Earth's centre of mass as the origin, referred to as geocentric datums.

Selection of an appropriate map projection is a complex process involving an evaluation of map projection alternatives based on the characteristics of map projections and the types of analysis to be performed. By applying Young's rule (1920), Snyder's decision tree (1987) and Mailing's property-use table (1992), we selected Lambert Equal-Area Azimuthal, Stereographic Conformal Azimuthal, Equidistant Conic and Equidistant Azimuthal map projections, plus Lambert Conformal Conic and Albers Equal-Area Conic recommended by Geoscience Australia (www.ga.gov.au). Basically, they are equidistant, equal-area and conformal projections with conic or azimuthal developable surfaces respectively. Thus, we can test these map projections in terms of the spatial properties preserved and the geometric construction methods used.

2.3. SMMs – Inverse distance squared and ordinary kriging

In this study, we examined two types of SMMs commonly compared in environmental studies (Li and Heap, 2008, 2011), i.e. inverse distance weighting (IDW) method and ordinary kriging (OK). IDW estimates the values of a spatial variable at unsampled points using a linear combination of values at sampled points weighted by an inverse function of the distance from the point of interest to the sampled points (ESRI, 2000). The main factor affecting the accuracy of IDW is the value of the power parameter (p) (Isaaks and Srivastava, 1989). The most popular choice of p is 2; and the resulting method is often called inverse distance squared (IDS) (Collins and Bolstad, 1996).

OK is the most general and widely used geostatistical interpolation method. Kriging assumes that the distance between sample points reflects a spatial correlation that can be used to explain variation in the surface. To make a prediction, kriging creates the variograms and covariance functions to estimate the statistical dependence values derived from the spatial autocorrelation (ESRI, 2000). The empirical variogram model applied in this study was the spherical model, selected as the best fit for spatial prediction of mud content data for the southwest region of AEEZ (Li et al., 2010). More details on kriging method can be found in the work of Krige (1951), Burrough and McDonnell (1998), Webster and Oliver (2001), Pebesma (2004) and Li and Heap (2008).

The modelling processes include data transformation, data projection, spatial interpolation and leave-one-out cross validation. For both IDS and OK, the search window size, i.e. the number of nearest samples used for making predictions, was set a value of 20, as suggested by Li et al. (2010). All the modelling work was undertaken using geostatistical and spatial analyst extensions in ArcGIS.

2.4. Assessment of accuracy

We applied leave-one-out cross validation to compare model performance based on different spatial reference systems. Several error measures for assessing accuracy have been considered. Mean absolute error (*MAE*) is among the best overall measures of model performance as it summarises the mean difference in the units of observed and predicted values (Willmott, 1982). Relative mean absolute error (*RMAE*) was applied because it is not sensitive to the changes in unit/scale (Li and Heap, 2008). Moreover, since the data are not normally distributed, paired Mann-Whitney tests were used to compare the predictive errors in terms of *RMAE* between WGS84 and different map projections for both IDS and OK. The testing work was implemented in R (R, Development Core Team, 2012).

3. RESULTS

IDS and OK were applied to the mud content dataset projected on different spatial reference systems. The observed and predicted values were compared using leave-one-out cross validation. The calculated values of *MAE* and *RMAE* are given in Table 1, along with the resulting P-values from paired Mann-Whitney tests.

Table 1. The predictive errors of the IDS and OK methods based on different spatial reference systems in predicting seabed mud content in the southwest region of AEEZ and the resulting P-values from the Mann-Whitney tests between WGS84 and the map projections in terms of *RMAE* for IDS and OK respectively.

Map Projection	MAE		RMAE (%)		Mann-Whitney test (P-value)	
	IDS	OK	IDS	OK	IDS	OK
WGS84	9.83	9.65	21.27	20.88		
GDA94	9.83	9.65	21.27	20.88		
Albers Equal-Area Conic	9.78	9.60	21.15	20.76	0.95	0.32
Lambert Equal-Area Azimuthal	9.74	9.54	21.07	20.63	0.31	0.11
Lambert_Conformal Conic	9.79	9.61	21.17	20.78	0.89	0.49
StereoG_Conformal Azimuthal	9.78	9.61	21.16	20.78	0.89	0.45
Equidistant Conic	9.79	9.61	21.17	20.78	0.97	0.41
Equidistant Azimuthal	9.74	9.54	21.06	20.65	0.56	0.26

Firstly, we compared the predictive errors from IDS and OK predictions using the two geographic coordinate systems, i.e. WGS84 and GDA94. As can be seen, the predictive errors from IDS and OK predictions using WGS84 and GDA94 are exactly the same. Then we compared the predictive errors resulted from the geographic coordinate systems with the selected map projections. It shows that the accuracy of both IDS and OK method slightly increased after the data was projected. However, the differences were less than 1%. The resulting P-values (0.11-0.97) from Mann-Whitney tests also indicated that the predictive errors from IDS and OK predictions based on WGS84 and the six map projections have no statistically significant difference.

We further examined the predictive errors resulted from the six map projections. Firstly, with respect to spatial properties preserved, we compared the equidistant, equal-area and conformal projections with conic developable surfaces. The differences in the predictive errors resulted from these map projections were trivial ($\leq 0.10\%$). Secondly, we compared these projections with azimuthal construction methods. Both IDS and OK methods performed slightly better when the data was projected using the Equal-Area and Equidistant projections than using the Conformal projections. However, the differences in the predictive errors were minor ($\leq 0.73\%$). Secondly, with respect to developable surfaces applied in construction methods, the Equal-Area and Equidistant projections with Azimuthal surfaces produced slightly better accuracy than those with Conic surfaces. However, the greatest differences were only 0.73% and 0.66%, respectively. The differences in the predictive errors resulted from the Conformal projections were even more negligible ($\leq 0.1\%$).

The prediction maps are plotted in Fig. 2 and Fig. 3 for visual examination purpose. The predicted patterns from all spatial reference systems are similar, with no major discrepancies observed.
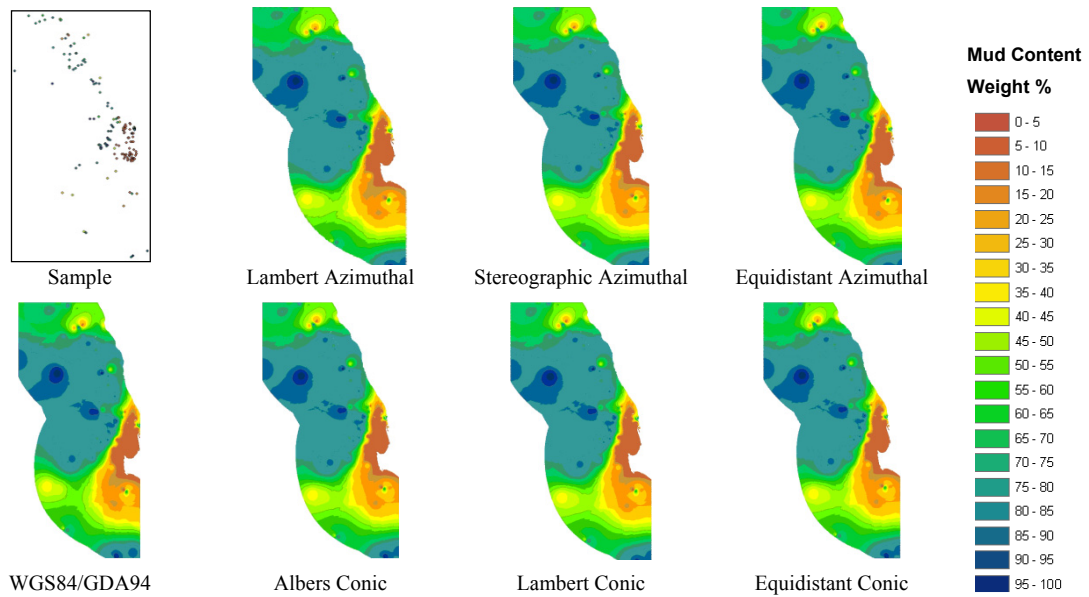

Figure 2. Predicted spatial distribution of seabed mud content in the southwest region of the continental AEEZ using IDS based on different spatial reference systems.
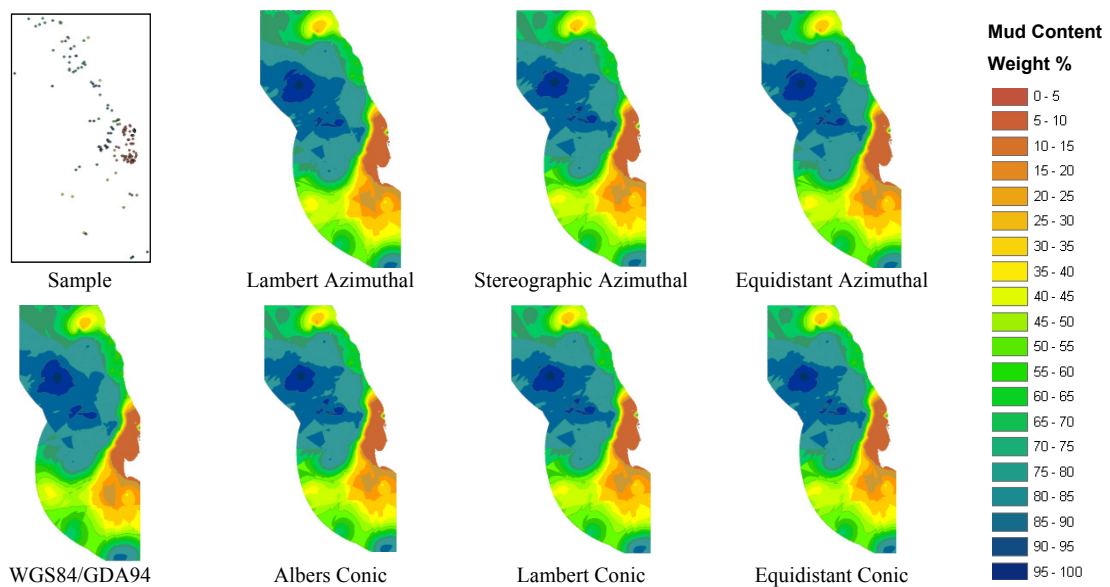

Figure 3. Predicted spatial distribution of seabed mud content in the southwest region of the continental AEEZ using OK based on different spatial reference systems.

4. DISCUSSION

WGS84 is dynamic in nature with site coordinates constantly changing to reflect tectonic movement on a global scale while GDA94 is a static datum fixed at the beginning of 1994 to the International Terrestrial Reference Frame (ITRF) realisation at that time (www.icsm.gov.au/gda/wgs84fact.pdf). Stanaway (2007) points out, for applications that require precise absolute positions, the discrepancy between WGS84 and GDA94 due to plate tectonics needs to be taken into account. However, the SMMs are not expected to be sensitive to the two geographic coordinate systems because they make predictions based on the relative positions between the samples and the samples are time independent in this case. This is confirmed by the findings in this study.

The remaining question is whether the unit difference in geographical coordinates or distortions introduced by map projections has more effect on the accuracy of the SMMs. The result from this study indicates that, when the data is projected using the six map projections, the accuracy of both IDS and OK methods marginally increases. However, the improvement of less than 1% is considered negligible. As expected, because spatial interpolation methods focus on local data, the unit difference in geographical coordinates and distortions introduced by map projections result in minor differences in the accuracy of the two spatial interpolation methods. However, data density and search window control the region size formed by samples; hence potentially, they may alter the degree to which spatial reference systems affect spatial predictions. Further work is required to test datasets with different data density and to test a variable search window size.

Among the six map projections, Lambert Equal-Area Azimuthal and Equidistant Azimuthal performed slightly better than the others for both IDS and OK, indicating that Equal-Area and Equidistant projections with Azimuthal surfaces are more suitable than other projections for the spatial predictions of seabed sediment data in the southwest region of AEEZ. This is consistent with Young's rule (1920) and Sear's recommendation (1967) for choosing map projections for the Australia region. Essentially, the choice is dependent on the determination of a circular or elongate shape of the AEEZ. Young's rule (1920) provides a quantitative means to determine shapes while Snyder's decision tree (1987) makes choice based on visualisations.

Another issue is that both the unit difference in geographical coordinates and distortions introduced by map projections are dependent on the location of an area. For example, the unit difference in latitude and longitude gradually increases from the equator towards the North/South Poles, while different map projections also generate spatially uneven distortions. We recommend that further investigation is needed to test other regions in the continental AEEZ.

5. CONCLUSIONS

This study examines how sensitive the two most commonly used SMMs (IDS and OK) are to the spatial reference systems selected for predicting marine environmental variables in the continental AEEZ and whether they respond differently to the choice of a spatial reference system. The main findings are:

1. The SMMs (IDS and OK) examined in this study are not sensitive to the two geographic coordinate systems (WGS84 and GDA94);
2. Whether the data is projected on spherical surfaces based on the geographic coordinate systems or on planar surfaces based on the map projections, the accuracies of the SMMs (IDS and OK) in predicting seabed mud content in the southwest region of AEEZ are similar and the differences are considered negligible, in terms of predictive errors and prediction map visualisations;
3. The slightly better prediction performance from Lambert Equal-Area Azimuthal and Equidistant Azimuthal projections for both IDS and OK indicates that Equal-Area and Equidistant projections with Azimuthal surfaces are more suitable than other projections for spatial predictions of seabed sediment data in the southwest region of AEEZ;
4. The two SMMs (IDS and OK) respond consistently to the choice of a spatial reference system.

The results suggest that spatial predictions using datasets with a density comparable to or greater than that in this study may use WGS84 directly and may not have to project data on a certain spatial reference system. This would greatly increase data processing efficiency. The outcomes of this study have significant implications for spatial predictions in environmental science. The findings are applicable to spatial predictions of both marine and terrestrial environmental variables.

ACKNOWLEDGMENTS

REFERENCES

Battersby, S. E. 2009. The effects of global-scale map projection knowledge on perceived land area. *Cartographica,* 44**,** 33-44.

Burrough, P. A. & McDonnell, R. A. 1998. *Principles of Geographical Information Systems*. Oxford: Oxford University Press.

Collins, F. C. & Bolstad, P. V. 1996. A comparison of spatial interpolation techniques in temperature estimation. *The Third International Conference/Workshop on Integrating GIS and Environmental Modeling*, Santa Barbara, CA. National Center for Geographic Information and Analysis, Santa Barbara.

ESRI 2000. Arc/Info. *ESRI, Inc., Redlands, California*.

Heap, A. D. & Harris, P. T. 2008. Geomorphology of the Australian margin and adjacent seafloor. *Australian Journal of Earth Sciences,* 55**,** 555-585.

Isaaks, E. H. & Srivastava, R. M. 1989. *Applied Geostatistics.* New York: Oxford University Press, 561pp.

Kennedy, M. 1989. *Understanding of map projections.* ESRI: Manual of ArcGIS.

Krige, D. G. 1951. A statistical approach to some mine valuations problems at the Witwatersrand. *Journal of the Chemical, Metallurgical and Mining Society of South Africa,* 52**,** 119-139.

Li, J. & Heap, A. 2008. *A Review of Spatial Interpolation Methods for Environmental Scientists.* Canberra: Geoscience Australia. Record 2008/23.

Li, J. & Heap, A. 2011. A review of comparative studies of spatial interpolation methods: performance and impact factors. *Ecological Informatics,* 6**,** 228-241.

Li, J., Heap, A., Potter, A. & Daniell, J. J. 2011a. *Predicting Seabed Mud Content across the Australian Margin II: Performance of Machine Learning Methods and Their Combination with Ordinary Kriging and Inverse Distance Squared.* Canberra: Geoscience Australia. Record 2011/07, 69pp.

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

Li, J., Heap, A. D., Potter, A., Huang, Z. & Daniell, J. 2011c. 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., Heap, A. D., Potter, A., Huang, Z. & Daniell, J. 2011d. Seabed mud content across the Australian continental EEZ 2011. Available: https://www.ga.gov.au/products/servlet/controller?event=FILE_SELECTION&catno=71977 [Accessed GEOMET: 14815].

Li, J., Potter, A., Huang, Z., Daniell, J. J. & Heap, A. 2010. *Predicting Seabed Mud Content across the Australian Margin: Comparison of Statistical and Mathematical Techniques Using a Simulation Experiment.* Canberra: Geoscience Australia. Record 2010/11.

Mailing, D. H. 1992. *Coordinate systems and map projections.* Oxford: Pergamon Press,1992.

Pebesma, E. J. 2004. Multivariable geostatistics in S: the gstat package. *Computer & Geosciences,* 30**,** 683-691.

R Development Core Team, 2012. R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing.

Robeson, S. M. 1997. Spherical Methods for Spatial Interpolation: Review and Evaluation. *Cartography and Geographic Information Science,* 24**,** 3-20 (18).

Sear, W. J. 1967. The projection story. *Cartography,* 6**,** 64-72.

Shepard, D. 1968. A two-dimensional interpolation function for irregularly spaced data. *In Proceedings of the 23rd National Conference, ACM.*, Princeton, NJ. Brandon/Systems Press Inc, 517-23.

Shepard, D. 1984. Computer mapping: The SYMAP interpolation algorithm. *In:* GAILE, G. L. & WILLMOTT, C. J. (eds.) *Spatial statistics and models.* Dordrecht, Holland: D. Reidel 133-45.

Snyder, J. P. 1987. *Map Projections--A Working Manual.* Geological Survey Professional Paper,1395. Washington, DC: U. S. Government Printing Office.

Snyder, J. P. & Voxland, P. M. 1989. *An album of map projections.* United States Geological Survey Professional Paper,1453. Washington, D.C.: U.S. Government Printing Office.

Stanaway, R. 2007. *GDA94, ITRF, WGS84: What's the difference? Working with dynamic datums* [Online]. Available: http://www.quickclose.com.au/stanawayssc2007.pdf.

Tissot, N. A. 1881. *Memoire sur la representation des surfaces et les projections des cartes geographiques.* Paris, France: Gauthier Villars.

Usery, E. L. & Seong, J. C. 2000. A Comparison of Equal-Area Map Projections for Regional and Global Raster Data. *Proceeding of 29th International Geographical Congress*, Seoul, Korea, 2000.

Webster, R. & Oliver, M. 2001. *Geostatistics for Environmental Scientists.* Chichester: John Wiley & Sons Ltd. 271pp.

Willmott, C. J. 1982. Some comments on the evaluation of model performance. *Bulletin American Meteorological Society,* 63**,** 1309-1313.

Willmott, C. J., Rowe, C. M. & Philpot, W. D. 1985. Small-Scale Climate Maps: A Sensitivity Analysis of Some Common Assumptions Associated with Grid-Point Interpolation and Contouring. *Cartography and Geographic Information Science,* 12**,** 5-16(12).

Young, A. E. 1920. *Some investigations in the theory of map projection.* Technical series no.1. London: Royal Geographical Society.