

Geographically weighted modelling of surface salinity in Florida Bay using Landsat TM data

ZHIXIAO XIE*, CAIYUN ZHANG and LEONARD BERRY

Department of Geosciences, Florida Atlantic University, Boca Raton, FL, 33431, USA

(Received 4 January 2012; in final form 9 May 2012)

An effective remote-sensing approach is needed for surface salinity monitoring in Florida Bay, a typical estuarine and coastal ecosystem (ECE). Yet, the non-stationary nature of surface salinity makes it difficult to model with conventional regression methods. A geographically weighted regression (GWR) approach was proposed to model surface salinity from Landsat Thematic Mapper (TM) imagery in this study. The models were constructed and validated with spatiotemporally matched field-surveyed salinity and TM imagery collected in February 1999. The GWR models reported high coefficient of determination (R^2) values and low root mean square errors (RMSEs) in validation. A 1999 model was also used to hindcast the surface salinity with TM imagery collected in December 1998 and validated with surface salinity collected at that time. The validation reported a reasonably low RMSE. It suggests a GWR approach, with field survey and remotely sensed data, may be useful in modelling and predicting the spatial variation pattern of surface salinity in Florida Bay, and could potentially serve as a less costly alternative or a supplement to field survey currently undertaken for salinity monitoring in the coastal areas of the Greater Everglades.

1. Introduction

Estuarine and coastal ecosystems (ECEs) contain some of the marine world's most important ecosystems and represent significant resources for human activities. They are also among the most heavily used and threatened natural systems globally and are the frontier of impacts of climate change and sea level rise. In ECE, salinity is a key physical characteristic as it affects water quality, vegetative communities and most animal species (Marshall *et al.* 2008). Hence, salinity monitoring and modelling are important in the assessment of ecological resources in Florida Bay (CROGEE 2002), a typical and important ECE in the South Florida, USA.

Salinity assessment in Florida Bay currently relies on field observation and salinity modelling. The former includes automatic data collection at gauge stations, as well as field survey. Modelling the spatial and temporal patterns makes it possible to extrapolate to parts or the entire Florida Bay from these sample points and snapshots. Models can be grouped into either statistic based or mechanistic based (Marshall *et al.* 2008). The accuracy of these models is limited by the data available to describe patterns of salinity and the driving processes, e.g. little or no salinity data can be collected for the

*Corresponding author. Email: xie@fau.edu

shallow regions inaccessible to research boats, and their applications are still in preliminary stages (Marshall *et al.* 2008). Other limitations on the current practices are that field surveys are time consuming, labour-intensive and expensive.

Remote sensing has been commonly accepted as an effective approach in monitoring environmental resources in large areas for its coverage, limited cost and repeatability. Different remote sensors have been used in the literature on sea surface salinity mapping, including Thematic Mapper (TM) (Khorram 1985), Moderate Resolution Imaging Spectroradiometer (MODIS) (Palacios *et al.* 2009), microwave sensors (Goodberlet *et al.* 1997, D'sa *et al.* 2000, Wilson *et al.* 2001), and the very recent Aquarius (NASA 2012). Of them, microwave sensors are the most commonly used because the passive microwave signature (emissivity) of water is a function of salinity (Blume *et al.* 1978). There are also attempts to apply Landsat remote-sensing data in salinity mapping in ECE. For example, Khorram (1985) used the Landsat Multispectral Scanner (MSS) to estimate water quality (including salinity) in San Francisco Bay. Lavery *et al.* (1993) used Landsat TM to monitor salinity and other water quality in the Peel-Harvey estuarine system in Western Australia. Vuille and Baumgartner (1993) used TM for salinity study in the North Chilean Altiplano. The principle behind the use of TM for salinity remote sensing is not clear, since reflectances in the optical wavelength of Landsat bands have no direct physical links to salinity. However, it has been suggested that some indirect links may exist, because studies have shown clear relations between some other biophysical properties and salinity in ECE, and the former are probably measurable through Landsat sensors (Carder *et al.* 1993, Ferrari and Dowell 1998). Application of TM in salinity modelling could be very useful because of the relatively fine spatial resolution (30 m), which may help capture the local variations of salinity in ECE. Other attractive features of TM include the long-term continuity, free archive access and a certain degree of preprocessing. The general practice in literature is using regression to describe the relationship between field-measured salinity and remotely sensed response.

Regression techniques have been widely used in remote sensing (Foody 2003). However, the commonly adopted regression models (simple ordinary least squares linear regression or OLS, multiple linear regression, etc.) have their limitations in geospatial data and modelling because relations between geospatial data are often non-linear and spatial processes are often non-stationary, particularly when considering the large area of coverage generally provided by remote sensing (Foody 2003). The surface salinity in ECE can be thought of as a non-stationary process due to complex interactions between land, ocean and the relatively shallow depth. This non-stationary process is actually evidenced by previous remote-sensing salinity modelling studies in ECE, which found these models are site-specific and different empirical models need to be developed for different areas. In this study, we adopted a geographically weighted regression (GWR) approach to model the surface salinity in Florida Bay with the spatiotemporally matched Landsat TM and field-surveyed salinity data. In addition, a GWR model developed for one time may be used as a predictive model to predict (or hindcast) surface salinity for another time when remotely sensed data are available. This makes GWR more appealing than common spatial interpolation methods which are often used in interpolating surfaces from samples, but without the predictive capability. Examining the feasibility of using GWR for surface salinity predictive modelling was therefore another major research objective.

2. Data and methods

2.1 Study area

Florida Bay is a large estuary covering approximately 2200 km² in south Florida, bounded on the north by the Everglades, on the east and south by the Florida Keys and on the west by the Gulf of Mexico (CROGEE 2002) (see figure 1). Florida Bay is very shallow, with average water depth less than 1 metre (CROGEE 2002). It is a place with impact from both land and the Gulf. Freshwater, which carries both dissolved and particulate matter, enters into the Bay from the north, while its western side is affected by the west-to-east (or NW-to-SE) current from the Gulf of Mexico. Although it appears to be a large, open system, Florida Bay is characterized by high level of spatial heterogeneity. The area within the bay is subdivided by shallow carbonate banks into numerous semi-isolated basins especially in the central and eastern zones (Fourqurean and Robblee 1999). The interconnected carbonate mudbanks function as barriers to water circulation across basins and lead to marked spatial differences in water salinity (Hall *et al.* 2007). The salinity is highly dependent on local rainfall and evaporation and changes with two obvious seasons. During the dry season, cells of hypersaline water are common. In the wet season, the salinity is much lower due to water discharge from land.

2.2 Data

Two major types of data are used in this study, field-surveyed surface salinity data by the US Geological Survey (USGS) and the TM imagery. The USGS conducted salinity surveys of Florida Bay approximately every other month from November 1994 to December 2001 (SOFIA 2012a). Starting in the summer of 1996, the USGS used a digital measuring system for salinity survey. Based on the metadata (SOFIA 2012b), the salinity surveys were performed using boat-mounted flow-through systems equipped with a YSI[®] water quality monitor. Position is determined using a global positioning

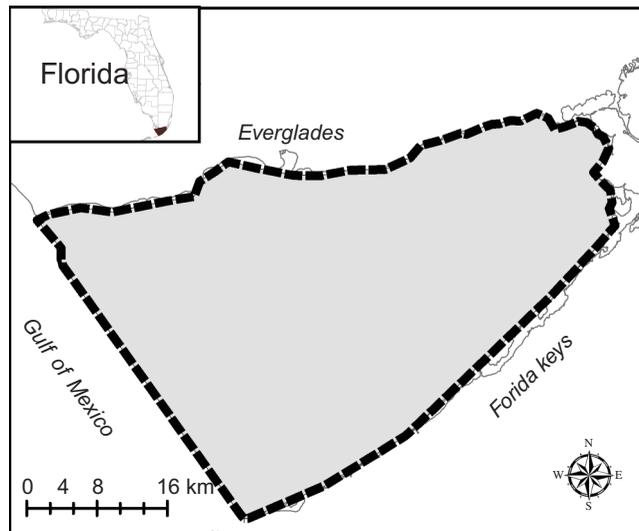


Figure 1. The study area in Florida Bay.

system (GPS) unit which interfaces with the YSI water quality monitor. Data collection occurs every 5 seconds and is stored in the YSI 650 data acquisition system (YSI Inc., Yellow Springs, OH, USA). With the new digital measuring system, each survey collected several thousand surface measurements. The measurement unit is part per thousand (ppt).

The Landsat system collects TM images over this region every 16 days. Geometrically corrected and geographically projected TM data (WRS-2 Path 15, Row 42) are available at no cost from the USGS (<http://edcscns17.cr.usgs.gov/EarthExplorer>). Unfortunately, the salinity surveys were not necessarily collected on the dates of the Landsat satellite overpasses. For most of the dates with coincidental satellite and salinity data, the TM data were contaminated by cloud and not useful.

For this study, we were able to get two sets of temporally matched salinity survey data and TM imagery, collected in December 1998 and February 1999, respectively, under relatively cloud-free condition.

2.3 Methods

GWR was used in this study to (1) model the relationship between field-surveyed surface salinity and TM data from 1999 and (2) hindcast the surface salinity in 1998 with the 1998 TM data and the 1999 model.

GWR is designed to incorporate the spatial dependency (non-stationarity) into regression models. To achieve this goal, a local regression model is constructed at every location and different relationships (regression coefficients) are possible at different points (Brunsdon *et al.* 1996, Fotheringham *et al.* 2002). GWR is therefore a truly local approach capable of capturing non-uniform spatial dependency (O'Sullivan and Unwin 2010).

The early application of GWR in remote sensing may be traced back to Foody (2003), who demonstrated the technique in examining the normalized difference vegetation index (NDVI)–rainfall relationship. Some recent examples include modelling net primary production of Chinese forest ecosystems (Wang *et al.* 2005), the estimation of leaf area index (LAI) over a tropical rainforest (Propastin 2009), tree diameter modelling from airborne lidar (Salas *et al.* 2010). To best of our knowledge, no previous remote-sensing study has applied GWR in salinity modelling in ECE. But GWR could be very useful in this regard, because ECE salinity may show strong local variations, as ECE waters are generally shallow and have a higher degree of optical complexity than clear oceanic waters (Liu *et al.* 2003). While a global regression may be identifiable between variables with conventional regression techniques, GWR is better suited to identify location variations in the relationships (Foody 2003). In addition, it is necessary to use a large sample size to capture spatial variation. When the sample size is large, GWR produces significantly better results than OLS (Chen *et al.* 2011).

For this study, the GWR model can be described as

$$Y_i = \beta_{0i} + \beta_{1i}X_{1i} + \beta_{2i}X_{2i} + \dots + \beta_{ni}X_{ni} + \varepsilon_i \quad (1)$$

where Y_i is the surface salinity at location i , X_{ji} is the spectral response for location i in TM band j and β_{ji} is the regression coefficient for location i in TM band j . Note that instead of being the same everywhere, coefficients β now vary in terms of locations (i).

To construct a GWR, it is critical to adequately specify relevant parameters. GWR relies on weighted linear regression to construct a local model for each location. For

each local model, certain field samples are included and spatially weighted, depending on their proximity to the location. The weighting is specified using either a Gaussian or biweight kernel function (Fotheringham *et al.* 2002), for which the bandwidth is usually more important (O'Sullivan and Unwin 2010). The kernel bandwidth can be fixed or adaptive (to accommodate variable sample density). For this study, an adaptive kernel with Akaike information criterion (AICc) bandwidth method is adopted.

The models were first constructed and validated with the 1999 TM and field-surveyed surface salinity. TM bands tend to be correlated, with high correlation between visible bands (TM bands 1, 2 and 3), and between mid-infrared bands (TM bands 5 and 7). GWR models cannot be successfully computed when there is either severe global or severe local multicollinearity, i.e. redundancy among model explanatory variables. It is very likely that not all useful information in the TM imagery will be successfully included in a GWR model when raw TM bands are used due to the potential multicollinearity issue. However, it may be preferable not to arbitrarily exclude certain bands from a model, because salinity modelling succeeded with very different TM band combinations in the literature, e.g. individual TM band 4 or 7 (Lavery *et al.* 1993), combination of TM bands 5 and 6 (Khorram 1985), combination of TM bands 1, 4 and 7 (Vuille and Baumgartner 1993) and combination of TM bands 1, 3 and 4 (Zhang *et al.* 2012). On the other hand, principal component analysis (PCA) is a technique for transforming raw remotely sensed data into a smaller set of uncorrelated variables carrying most of the information (Jensen 2005). Hence, a PCA was conducted for the six TM bands (excluding the thermal bands) and selected PCA components were used in various experiments.

To examine the impact of spatial autocorrelation of data on validation, different scenarios of subsetting data were tested, including (1) the 1st out of every three survey points was used for GWR model training, and the remaining two-thirds was used for validation; (2) the 1st and the 5th of every 10 for training and validation, respectively; (3) the 1st and the 13th of every 25 for training and validation, respectively; (4) the 1st and the 25th of every 50 for training and validation, respectively; (5) the 1st and the 50th of every 100 for training and validation, respectively; (6) the 1st and the 100th of every 200 for training and validation, respectively; and (7) the 1st and the 150th of every 300 for training and validation, respectively. Scenario (1) was also used to build an OLS linear regression model for comparison.

To test the feasibility of using a GWR model to hindcast surface salinity, selected 1998 PCA data were used as input to a 1999 GWR model. The modelled salinity was validated with the salinity survey in 1998. Prior to PCA, the 1998 TM imagery was georeferenced to and spectrally calibrated to the 1999 TM imagery with multi-date image normalization using regression (Jensen 2005).

3. Results

3.1 The 1999 PCA and GWR models

For the 1999 TM data, the first three PCA components account for a total cumulative loading of 94% and they were used as the explanatory variables in salinity modelling. PCA component matrix indicates that the combination of the three PCAs contains significant information from all six bands. As described in section 2.3, different scenarios for subsetting data were tested. For comparison, an OLS linear regression model was also built. The OLS model resulted in an extremely low coefficient of determination

(R^2) (0.046) and no validation was performed. Of the GWR models, there is a clear decreasing trend for model R^2 when sparser data were used as the training data. When the number of training data is larger than 100, R^2 for the models is above 0.9. When the number of training data drops to 53, the model R^2 drops to 0.72. The validation root mean square errors (RMSEs) steadily arise with the decreasing R^2 . These are consistent with the GWR design that a large number (in hundreds) of data are needed for a well-performed GWR.

The spatial pattern of modelled salinity is also essential for model assessment. Figures 2(a)–(d) present the modelled surface salinity with selected scenarios. It appears that the modelled surface salinity generally corresponds well to the surveyed salinity. But the models with better performance measures (higher adjusted R^2 and lower validation RMSE) capture more spatial structure details and show reasonably smooth spatial transitions. It suggests that the more data, the more detailed spatial variation that could be captured in a GWR model.

3.2 Hindcast 1998 surface salinity with 1998 TM data and a 1999 GWR model

In this test, a 1999 GWR model was developed with all the 1999 surface salinity survey data and the first three PCA components of the 1999 TM data. Then, the model was applied to the first three PCA components of the 1998 TM data. The modelled results were validated with the field-surveyed salinity in 1998 (total 14,376 points). The RMSE (3.532 ppt) is higher than the results with the 1999 GWR models for 1999 data. However, by examining the field-surveyed salinity and the model prediction errors, it was found there is a cluster of lower surface salinity in one small bay at the north. The cluster may be the result of some local events. Without the cluster, the validation RMSE was much lower (2.53 ppt). Figures 2(e) and (f) show the 1998 field-surveyed surface salinity, TM imagery and modelled surface. The modelled surface shows a reasonably similar pattern as the surveyed data.

4. Discussion and conclusions

ECE surface salinity modelling with remote sensing is challenging but potentially very rewarding. Although spectral signatures in TM imagery do not directly respond to surface salinity, surface salinity affects many aspects of biophysical characteristics in ECE and there may exist some indirect links. Literature has shown that the inverse relationship between coloured dissolved organic material (CDOM) and salinity makes TM band 1 (blue) a useful band due to a strong absorption of CDOM in the ultraviolet and blue portion of the visible spectrum (D'sa *et al.* 2000). TM bands 3 (red) and 4 (near-infrared) are believed to be indirectly linked to salinity for their correlation with water depth (Jensen 2005) and seagrass vegetation characterization (Zhang *et al.* 2012). This study suggests that the indirect links may be reflected by the combined spectral information in visible, near-infrared as well as mid-infrared TM bands, because no single band stands out in the PCA component matrix. These indirect links need to be further investigated with field survey and laboratory analysis before a sound conclusion can be drawn.

By integrating field survey and remotely sensed data, GWR appears to be an appropriate approach to modelling the non-stationary and complex spatial structure of surface salinity in ECE environment. The conventional regression techniques,

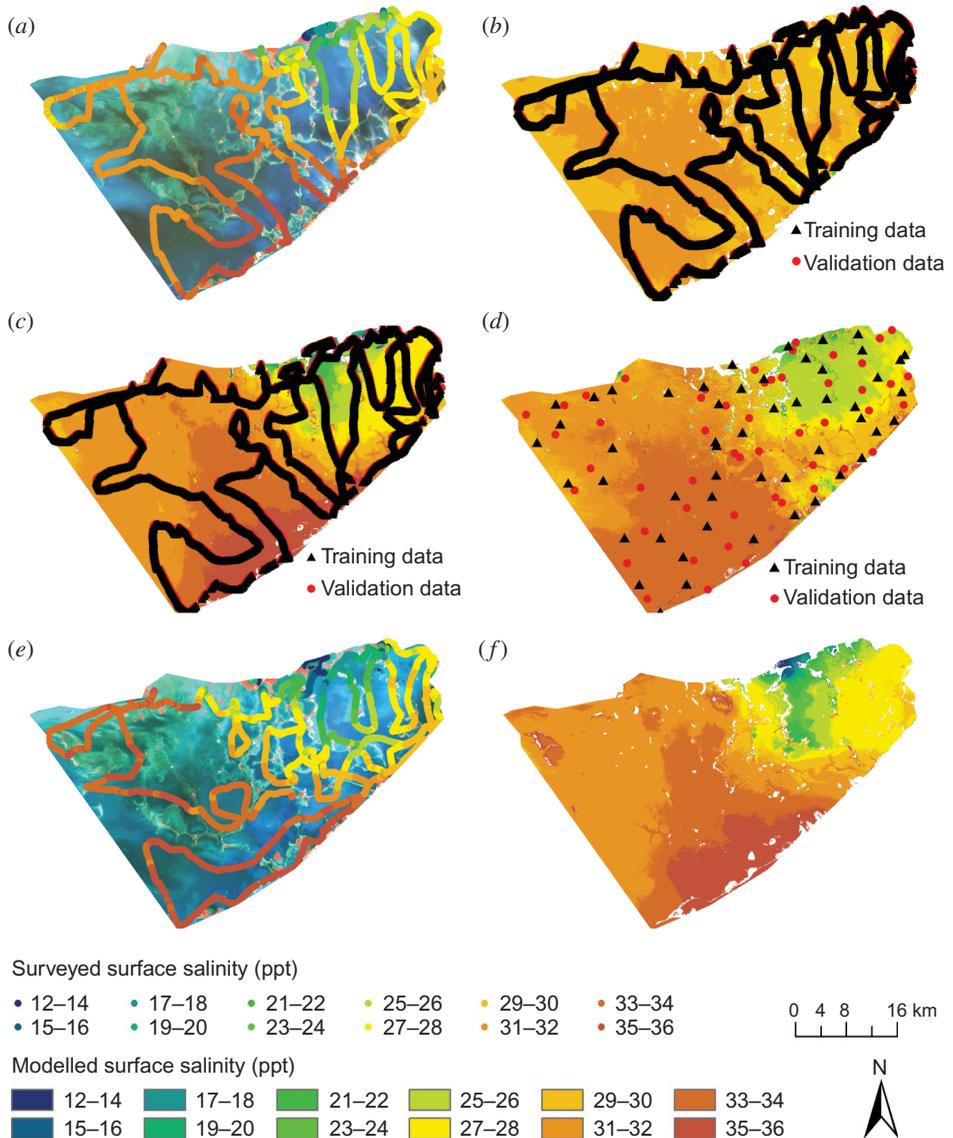


Figure 2. (a) The field-surveyed surface salinity data and the TM imagery (TM bands 4, 3 and 2 as red, green and blue) for 1999. (b)–(d) The 1999 salinity data subsets for training and validation and modelled surfaces: (b) one-third data for training and two-thirds for validation, OLS model; (c) one-third for training and two-thirds for validation, GWR model; (d) 1/200 for training and 1/200 for validation, GWR model. (e) The field-surveyed surface salinity data and the TM imagery (TM bands 4, 3 and 2 as red, green and blue) for 1998. (f) The hindcasted surface salinity with 1998 TM data as input to a 1999 GWR model. The modelled salinity outliers are mostly sighted at areas with high vegetation coverage and/or dry areas, or with cloud cover, and they were removed.

e.g. OLS may be useful when searching for a general spatial trend at a coarser resolution with sparsely distributed samples, or in relatively small areas where spatial non-stationarity is not obvious, as demonstrated in literature. However, it may fail when finer resolution spatial pattern needs to be modelled. In addition, a PCA

transformation may be a necessary data preparatory step so that the multicollinearity problems can be resolved and a better GWR model can be built with most information utilized in the original data. More importantly, this study provides some initial evidences that a GWR surface salinity model with TM data may be useful to hindcast surface salinity back to the past, and by the same principle to predict into the future. This may supply an alternative to the costly field survey, or as a supplement in the middle of two surveys. It should also be noted that this study only tested the GWR models in dry season. Further tests are needed for salinity modelling in wet season, when heavy rainfall and run-off make the bay environment more complex.

Finally, although this study focused on surface salinity modelling with TM imagery, it is expected that the general GWR approach could be readily applied in modelling other water quality aspects, and with other remotely sensed data in ECE.

Acknowledgements

This research is funded by the USGS through the Cooperative Ecosystem Studies Units (CESU) agreement. The comments from the reviewers and editors are greatly appreciated.

References

- BLUME, H.J., KENDALL, B.M. and FEDORS, J.C., 1978, Measurement of ocean temperature and salinity via microwave radiometry. *Boundary Layer Meteorology*, **13**, pp. 295–308.
- BRUNSDON, C.F., FOTHERINGHAM, A.S. and CHARLTON, M.E., 1996, Geographically weighted regression: a method for exploring spatial non-stationarity. *Geographical Analysis*, **28**, pp. 281–298.
- CARDER, K.L., STEWARD, R.G., CHEN, R.F., HAWES, S. and LEE, Z., 1993, AVIRIS calibration and application in coastal oceanic environments: tracers of soluble and particulate constituents of the Tampa Bay coastal plume. *Photogrammetric Engineering and Remote Sensing*, **59**, pp. 339–344.
- CHEN, G., ZHAO, K., MCDERMID, G.J. and HAY, G.J., 2011, The influence of sampling density on geographically weighted regression: a case study using forest canopy height and optical data. *International Journal of Remote Sensing*, **33**, pp. 2909–2924.
- CROGEE (COMMITTEE ON RESTORATION OF THE GREATER EVERGLADES ECOSYSTEM), 2002, *Florida Bay Research Programs and Their Relationship to the Comprehensive Everglades Restoration Plan* (Washington, DC: The National Academies Press).
- D'SA, E.J., ZAITZEFF, J.B. and STEWARD, R.G., 2000, Monitoring water quality in Florida Bay with remotely sensed salinity and in situ bio-optical observations. *International Journal of Remote Sensing*, **21**, pp. 811–816.
- FERRARI, G.M. and DOWELL, M.D., 1998, CDOM absorption characteristics with relation to fluorescence and salinity in coastal areas of the Southern Baltic Sea. *Estuarine, Coastal and Shelf Science*, **47**, pp. 91–105.
- FOODY, G.M., 2003, Geographical weighting as a further refinement to regression modelling: an example focused on the NDVI–rainfall relationship. *Remote Sensing of Environment*, **88**, pp. 283–293.
- FOTHERINGHAM, A.S., BRUNSDON, C. and CHARLTON, M., 2002. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*, 269 p. (New York: Wiley).
- FOURQUREAN, J.W. and ROBBLEE, M.B., 1999, Florida Bay: a history of recent ecological changes. *Estuaries*, **22**, pp. 345–357.
- GOODBERLET, M.A., SWIFT, C.T., KILEY, K.P., MILLER, J.L. and ZAITZEFF, J.B., 1997, Microwave remote sensing of coastal zone salinity. *Journal of Coastal Research*, **13**, pp. 363–372.

- HALL, M.O., MADLEY, K., DURAKO, M.J., ZIEMAN, J.C. and ROBBLEE, M.B., 2007, Florida Bay. In *Seagrass Status and Trends in the Northern Gulf of Mexico: 1940–2002*, L. Handley, D. Altsman and R. DeMay (Eds.), pp. 243–254 (Reston, VA: US Geological Survey Scientific Investigations Report 2006–5287 and US Environmental Protection Agency 855-R-04-003).
- JENSEN, J., 2005, *Introductory Digital Image Processing: A Remote Sensing Perspective*, pp. 296–300 (Upper Saddle River, NJ: Pearson Prentice Hall).
- KHORRAM, S., 1985, Development of water quality models applicable for the entire San Francisco Bay and Delta. *Photogrammetric Engineering and Remote Sensing*, **51**, pp. 53–62.
- LAVERY, P., PATTIARATCHI, C., WYLLIE, A. and HICK, P., 1993, Water quality monitoring in estuarine waters using the Landsat Thematic Mapper. *Remote Sensing of Environment*, **46**, pp. 268–280.
- LIU, Y., ISLAM, M.A. and GAO, J., 2003, Quantification of shallow water quality parameters by means of remote sensing. *Progress in Physical Geographer*, **27**, pp. 24–43.
- MARSHALL, F.E., SMITH, D. and NUTTLE, W.K., 2008, *Task 7: Simulating and Forecasting Salinity in Florida Bay: A Review of Models. Critical Ecosystems Studies Initiative Project Task Report for Everglades National Park* (New Smyrna Beach, FL: Cetacean Logic Foundation).
- NASA, 2012, Aquarius: studying our salty seas from space. Available online at: http://www.nasa.gov/mission_pages/aquarius/main/index.html (accessed 5 April 2012).
- O’SULLIVAN, D. and UNWIN, D.J., 2010, *Geographic Information Analysis*, 2nd ed., 405 p. (New York: John Wiley and Sons).
- PALACIOS, S.L., PETERSON, T.D. and KUDELA, R.M., 2009, Development of synthetic salinity from remote sensing for the Columbia River plume. *Journal of Geophysical Research*, **114**, C00B05.
- PROPASTIN, P.A., 2009, Spatial non-stationarity and scale-dependency of prediction accuracy in the remote estimation of LAI over a tropical rainforest in Sulawesi, Indonesia. *Remote Sensing of Environment*, **113**, pp. 2234–2242.
- SALAS, C., ENE, L., GREGOIRE, T.G., NÆSSET, E. and GOBAKKEN, T., 2010, Modelling tree diameter from airborne laser scanning derived variables: a comparison of spatial statistical models. *Remote Sensing of Environment*, **114**, pp. 1277–1285.
- SOFIA, 2012a, Florida Bay salinity data. Available online at: <http://sofia.usgs.gov/metadata/sflwww/FBsaldat.html> (accessed 5 April 2012).
- SOFIA, 2012b, Coastal gradients salinity surveys. Available online at: http://sofia.usgs.gov/metadata/sflwww/coastal_grads_salsurveys.html (accessed 5 April 2012).
- VUILLE, M. and BAUMGARTNER, M.F., 1993, Hydrological investigation in the North Chilean Altiplano using Landsat-MSS and -TM data. *Geocarto International*, **3**, pp. 35–45.
- WANG, Q., NI, J. and TENHUNEN, J., 2005, Application of a geographically-weighted regression analysis to estimate net primary production of Chinese forest ecosystems. *Global Ecology and Biogeography*, **14**, pp. 379–393.
- WILSON, W.J., YUEH, S.H., LI, F.K., DINARDO, S., YI, C., KOBLINSKY, C., LAGERLOEF, G. and HOWDEN, S., 2001, Ocean surface salinity remote sensing with the JPL Passive/Active L-/S-band (PALS) microwave instrument. In *IEEE 2001 International Geoscience and Remote Sensing Symposium*, 9–13 July 2001, Sydney, NSW, vol. 2, pp. 937–939. Available online at: <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=976686&isnumber=21046&tag=1> (accessed 21 May 2012).
- ZHANG, C., XIE, Z., ROBERTS, C., BERRY, L. and CHEN, G., 2012, Salinity assessment in Northeast Florida Bay using Landsat TM data. *Southeastern Geographer*, in press.