URBAN DENSITY MONITORING USING HIGH RESOLUTION SPACEBORNE SYSTEMS

B.C. Forster
Director and Associate Professor

C. Jones
Research Assistant

Centre for Remote Sensing and
School of Surveying
The University of New South Wales
Sydney
Australia

Commission VII

The relationship between residential housing density and SPOT, and Landsat TM data, has been analysed using data drawn from thirty ground truth sites over the city of Sydney, Australia. Both radiometric response and response standard deviation were regressed against housing density, to determine the most significant variables. Data were analysed at two area levels, 60 by 60 metres, and 180 by 180 metres. The highest correlation achieved for the smaller sites was 0.60 using a two variable equation of Landsat TM band 4 and SPOT XS band 2. For the larger sites correlation ranged from 0.8 to 0.9 using one and five variables. SPOT XS band 2 was also the most significant variable in these regression equations.

1. INTRODUCTION

Many previous studies have used data from the Landsat MSS series of satellites to analyse urban areas, Forster (1981, 1983, 1985), Jensen et al (1983), Landini and McLeod (1979). The low spatial resolution of Landsat MSS data is however, a major limitation to its use as a primary data source for urban information.

The launch of Landsat TM and SPOT satellite remote sensing systems have provided a significant advance in the ability to repetively acquire information at ground resolutions more appropriate to the urban environment. SPOT digital image data, in both P and XS modes, and Landsat TM data, in all bands, have been acquired over the city of Sydney, Australia, and are being used to predict residential housing density, as part of a more ambitious research program to predict inter-census population change.

All of the remotely sensed data were acquired during the 1985 to 1986 time frame. Thirty sites covering residential areas of different densities have been sampled over the city of Sydney using air photographs, and these results related to the SPOT and TM digital image data, using regression analysis techniques. Earlier work on this research has been reported in Forster et al (1987a, 1987b).

2. RESIDENTIAL HOUSING DENSITY

Housing density represents a useful input to inter-census population change studies, but is not directly observable from spaceborne sensor systems because of their spatial resolution. Even with SPOT's 10 metre resolution P mode data, this is still the case because the average low density housing is about the same size as the pixel, and cannot be positively discriminated without an IFOV nearer to 5 metres. Nevertheless the 10 and 20 metre resolution of SPOT, and the 30 metre resolution of Landsat TM offer the potential of using spectral and textural variables for housing density studies.

It can be hypothesised that radiometric variance over a pixel neighbourhood, as a measure of texture, will increase as the pixel size approaches the size of the residential elements, and then reduce as the size of the pixel continues to decrease. It might be assumed then, for a constant pixel size, that as housing density varies from very low, ie. near homogeneous vegetation cover, to very high, ie. near homogeneous synthetic cover, that variance will pass through a maximum. Near infrared response data should allow the discrimination of vegetated and non-vegetated classes of housing density, and it is therefore suggested that a combination of textural data and spectral data should allow the prediction of housing density.

3. SAMPLING PROCEDURE

Thirty ground truth sites were selected over the Sydney Metropolitan area, comprising single dwelling residential areas ranging from high to low density, with dimensions of 180 metres by 180 metres. The sampling procedure has been designed to allow the collection of

(a) housing density

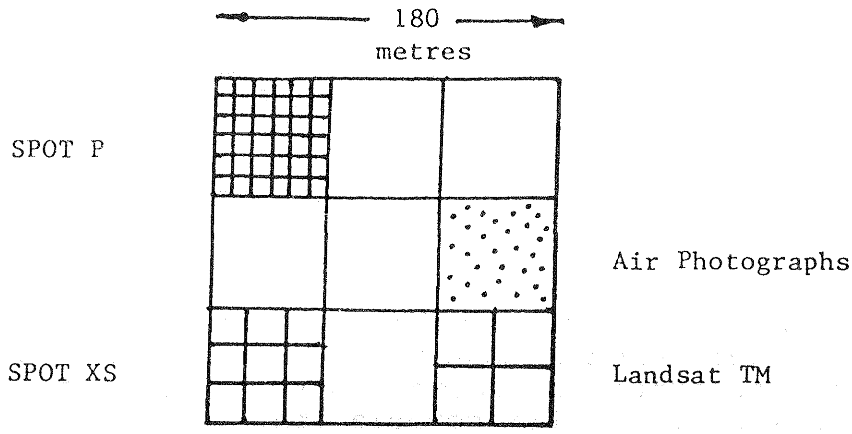
(b) percentage of cover types, house, road, grass, tree and other (including bare soil and non-residential land use)

(c) radiometric response and response standard deviation, over a 60 metre by 60 metre neighbourhood for data from SPOT P and XS modes, and Landsat TM (excluding TM Band 6, in the thermal infrared).

The sampling framework for an individual ground truth site is illustrated in figure 1.

The 180 metre primary sampling cell was divided into 9 secondary cells. Air photography was used to sample the percentage of cover type and the number of houses in each secondary cell. This data comprised the ground truth information. The remotely sensed data has been registered to the ground so that each secondary cell contains approximately

4 Landsat TM pixels, 9 SPOT XS mode pixels, and 36 SPOT P mode pixels. For each, the mean and standard deviation (as a measure of textural variation) of the radiometric response were determined. These values were analysed with the ground truth data for all thirty sites, using multiple linear regression techniques, to determine the optimum data combination to predict housing density in each secondary cell.



IN EACH 60 X 60 M CELL

- 36 SPOT P
- 9 SPOT XS
- 4 LANDSAT TM

PLUS AIR PHOTOGRAPHS FOR GROUND TRUTH

Figure 1. Illustration of the sampling procedure

4. RESULTS FROM REGRESSION ANALYSIS

4.1 Standard Deviation as a Measure of Housing Density

The basic premise underlying the initial examination of the data was that the variation in the spectral response would be an indicator of the housing density. This was tested using regression analysis with number of houses as the dependent variable and standard deviation (S.D.) and $(S.D.)^2$, for all bands of both SPOT and TM, as the independent variables. Table 1 gives the results of this analysis. Because of the low correlation only the most significant variable is given.

TABLE 1. Coefficients and Correlation (R) of Regression Equations for Housing Density Versus Standard Deviation of Response Variables

Sensor Band	Constant	SD	SD2	R
SPOT P	3.710		0.009	0.144
XS1	3.305	0.243		0.156
XS2	3.700		0.027	0.178
XS3		Not Significant		
TM1	3.796		0.015	0.157
TM2		Not Signifcant		
TM3		Not Significant		
TM4	4.565	-0.135		0.142
TM5	4.307		-0.005	0.142
TM7		Not Significant		

It can be seen from the results of Table 1 that standard deviation of the various spectral response has only a low correlation with number of houses, with the results from SPOT XS Band 2 having the highest correlation.

When the standard deviation of all spectral response variables were regressed against number of houses, the following regression equation resulted,

$$\text{No. of Houses} = 4.075 - 0.015\ \text{TM4}\ (SD)^2 + 0.020\ \text{TM1}\ (SD)^2 - 0.008\ \text{TM5}\ (SD)^2 + 0.029\ \text{XS2}\ (SD)^2 \quad \ldots\ (1)$$

with a correlation of R = 0.335

Due to the low correlation of house numbers with standard deviation, those sites with high positive and negative residuals, greater than 2 houses per secondary site, were reexamined on the original sampled air photographs. In addition because the sampled air photographs were taken some years before the satellite imagery selected, colour air photographs that were temporally coincident, but of a smaller scale, were also examined to determine if any temporal change in land cover had occurred. A total of eight sites with large positive residuals and five sites with large negative residuals were examined. As a general rule it was found that those sites with positive residuals had few trees and a road passing through them, while those with negative residuals were centred on the rear gardens of house allotments with a high tree content. Some minor change was also noted with the more recent photographs.

4.2 Inclusion of Response Variables in Regression Analysis

Previous research using Landsat MSS (Forster, 1983) has shown that a combination of visible and infrared bands in a regression equation were good predictors of tree and road percentages. For this reason it was decided to incorporate spectral response variables in addition to standard deviation variables into the regression equations for predicting number of houses.

All SPOT and TM spectral response variables and their associated standard deviations were regressed against house number, resulting in the following equation.

$$\text{No. of Houses} = 0.498 - 0.061\ TM4 + 0.222\ XS2 \quad \ldots.\ (2)$$

, with R = 0.600. SPOT band XS2 was the first variable to enter the equation, and thus was most significant. None of the standard deviation variables were found to be significant. An examination of the residuals showed that very low and very high density housing sites had the largest residuals, with high density sites being lower than expected and low density being higher than expected. In an attempt to account for this non-linear effect second and third order variables of SPOT XS2 were entered into the regression, with the variable $(XS2)^3$ only being significant, to give the following equation -

$$\text{No. of Houses} = 1.682 + 7.982 \times 10^{-5}\ (XS2)^3 \quad \ldots.\ (3)$$

, with R = 0.588

4.3 Regression of the Average Response over all Primary Sampling Sites

The relatively low correlation achieved using the initial sampling sites, suggested that a number of errors unrelated to housing density could be affecting the results. These were considered to be -

(a) Errors of registration. Particularly with the high spatial resolution SPOT data, an error of half a pixel in the registration with respect to the photo sampled areas could lead to a significant change in the house numbers attributed to these response variables.

(b) As has been shown using Landsat MSS data (Forster, 1981) the sensor point spread function can result in a contribution of approximately 50% of the radiance received at the sensor from the neighbouring pixels of the target pixel. In sites of rapidly changing land cover and with small pixel sizes this will cause an incorrect relationship between interpreted house number values and the measured spectral response.

For both these reasons it was considered appropriate to analyse the regression relationships between the average number of houses over each of the thirty, 180 by 180 metre primary sites, and the average spectral response values.

All average response and response standard variation variables were regressed against average number of houses as the dependent variable. The five most significant variables were allowed to enter the regression equation. While a total of 16 variables were found to be significant, it was considered inappropriate to include these because of the greatly reduced sample size of thirty primary sites. The resulting regression equations, correlation and the order in which variables entered the equation are shown in Table 2.

TABLE 2. Coefficients and Correlation (R) of Regression Equations for Housing Density Versus Response Variables and Standard Deviations over Primary Sites

		Response Variables					
Order of Entry	Constant	XS2	TM5(SD)	XS2(SD)	TM1	XS1	R
1	-3.885	0.263					0.811
2	-3.414	0.216	0.184				0.853
3	-2.957	0.221	0.186	-0.293			0.876
4	-8.643	0.091	0.197	-0.284	0.137		0.893
5	-8.008	0.278	0.189	-0.343	0.194	-0.266	0.906

It should be noted that standard deviation is now a significant contributor to house number. Apart from step 4, the small change in the coefficient values of XS2 and TM5 (SD) indicates that the additional variables are relatively uncorrelated and are accounting for real change in the house number data set. These results compare favourably with those using Landsat MSS data, where a correlation of 0.91 was achieved with 6 variables, but over a primary site of 400 x 450 metres, (5 lines and 8 pixels, Landsat MSS). This represents an area six times the size of that used for the SPOT and TM data.

5. CONCLUSIONS

(i) Response standard deviation is not a primary predictor of housing density.

(ii) SPOT XS band 2 response is the most significant predictor variable of housing density.

(iii) Correlation of housing density with response variables, increases from 0.335 to approximately 0.9 when the sample site is increased from 60 x 60 to 180 x 180 metres.

(iv) The higher spatial resolution SPOT and Landsat TM data has resulted in satisfactory predictive equations over areas six times smaller than that achieved with previous research using Landsat MSS.

6. REFERENCES

Forster, B.C., 1981. Some measures of residential quality from Landsat multispectral data. Unisurv Report, 518, U.N.S.W.

Forster, B.C., 1983. Some urban measurements from Landsat data. Photo. Eng. and Remote Sensing, 49, 1693-1707.

Forster, B.C., 1985. An examination of some problems and solutions in monitoring urban areas from satellite platforms. Int. J. of Remote Sensing, 6, (1), 139-151.

Forster, B.C., Smith, C., Farrington, R. and Jones, C., 1987. Urban research using SPOT digital imagery, Proc. 4th Australasian Rem. Sens. Conf., September 1987, Adelaide, 752-8.

Forster, B.C., Jones, C., Smith, C. and Farrington, R., 1987. Mapping and urban multi-experimental assessment of SPOT (MUMAS) - Urban results. Proc. Inter. Conf. on SPOT 1, Image Utilization, Assessment Results, Paris, November 1987. (to be published)

Jensen, J.R., Bryan, M.L., Friedman, S.Z., Henderson, F.M., Holz, R.K., Lindgren, D., Toll, D.L., Welch, R.A., and Wray, J.R., 1983. Urban/suburban land use analysis. In Manual of Remote Sensing, Vol. 2, ed. Colwell, R.M., 1571-1666.

Landini, A.J., and McLeod R.G., 1979. Using population statistics for a first look at the utility of Landsat data for urban areas. Remote Sensing Quart., 1, 80.

Townshend J., and Justice, C., 1981. Information extraction from remotely sensed data. A user view. Int. J. Remote Sensing, 2, 313.