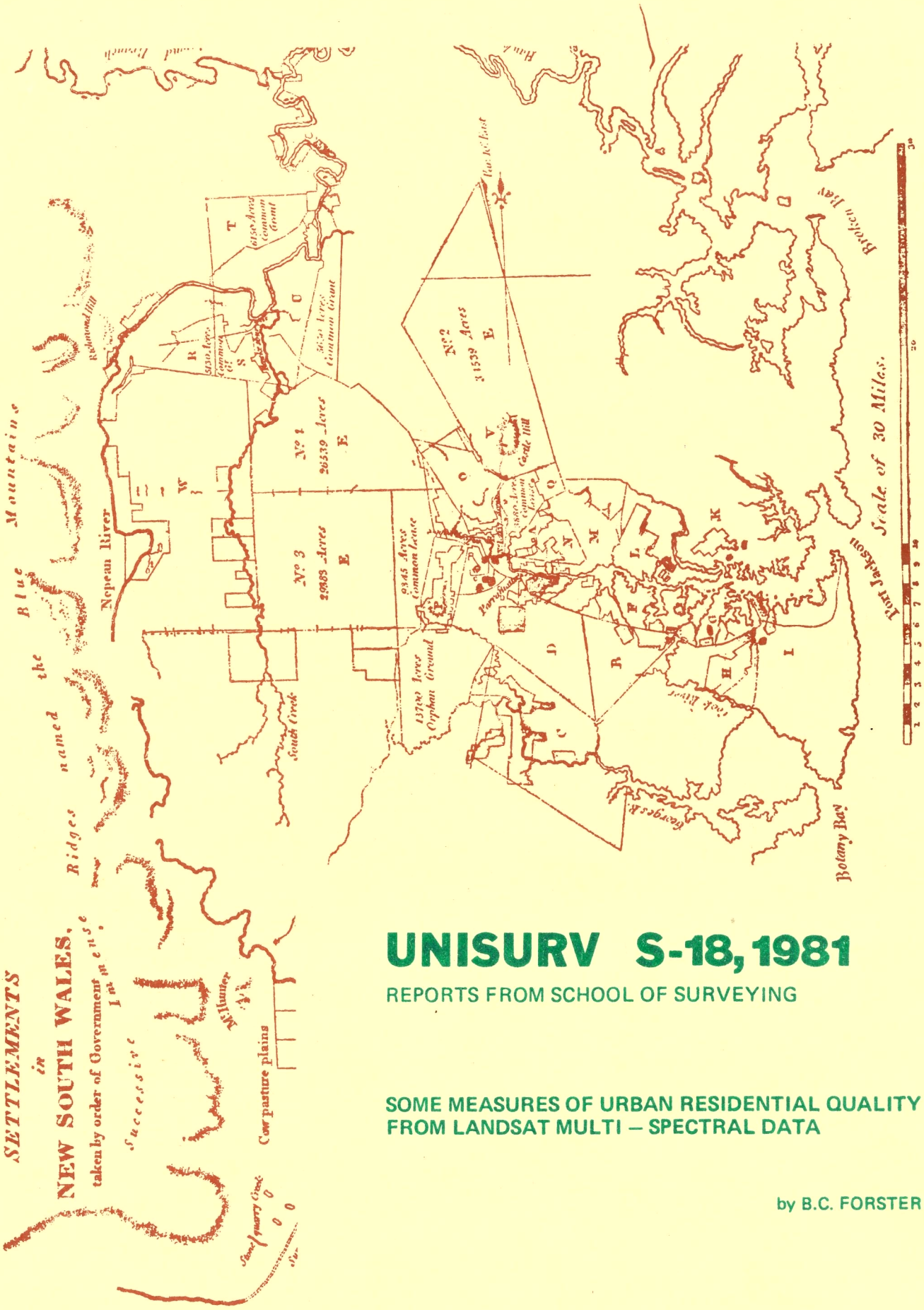


UNISURV S-18, 1981

REPORTS FROM SCHOOL OF SURVEYING

SOME MEASURES OF URBAN RESIDENTIAL QUALITY FROM LANDSAT MULTI – SPECTRAL DATA

by B.C. FORSTER

UNIVERSITY OF NEW SOUTH WALES,
KENSINGTON. N.S.W. AUSTRALIA.

UNISURV REPORT S 18, 1981

SOME MEASURES OF URBAN RESIDENTIAL QUALITY FROM LANDSAT MULTI-SPECTRAL DATA

by

B.C. FORSTER

Received January, 1981

SCHOOL OF SURVEYING,
THE UNIVERSITY OF NEW SOUTH WALES,
P.O. BOX 1,
KENSINGTON, N.S.W., 2033, AUSTRALIA.

ABSTRACT

This study examines the application of Landsat multispectral data to the analysis of urban residential quality in the Sydney (Australia) metropolitan area. A historical review of the use of remotely sensed data in urban areas indicates a number of scene and sensor related limitations to the clustering techniques currently used for classifying land use from Landsat data.

Linear models are developed relating the proportion of various surface cover types to the spectral response from a single pixel area while accounting for the degrading effects of the atmosphere and the integrating effect of the sensor point spread function. The atmospheric effects are modelled for the conditions prevailing at the time of the selected Landsat Sydney scene and include the effect of spatially varying background reflectance. Two models are developed to account for the effect of the point spread function, one using the deconvolved reflectance at the target pixel, the other incorporating the convolved sampled cover from a 3 x 3 pixel array.

Data for the study was obtained from seventy ground truth sites over the study area and from the Landsat computer compatible tapes of the Sydney scene of December 1972. The coefficients of the developed linear equations are calculated from the sampled data using multiple linear regression techniques. These coefficients allow the prediction of the reflectance of a number of residential cover types and the prediction of the proportion of individual cover types contained within a pixel and an extended 5 x 8 pixel block. Of a number of transformations examined, ratio type variables incorporated into the linear equations are found to be more significant than reflectance only variables and for extended areas, variables based on reflectance standard deviation are also found to be significant. The potential to predict the average number of houses per pixel over a ground truth area is an important result of this work. The multidimensional structure of the reflectance data is examined using factor analysis and is found to be essentially represented by two overlaid orthogonal systems, one representing a grass/tree structure, the other a vegetative/non-vegetative structure.

National Library of Australia

Card No. and ISBN

0 85839 029 9

The equations relating cover proportions and housing density to the Landsat derived reflectance data are applied to the analysis of average relative house values, as one measure of residential quality. House size is found to be the primary predictive variable for house value and it is shown that this variable can also be predicted from Landsat data. A best multiple correlation coefficient of 0.73 between house value and Landsat derived data shows that residential areas can be simply disaggregated on the basis of value. A residential quality index based on house size and cover proportions is developed and related to Landsat derived reflectance data at the pixel level with a multiple correlation of 0.87. It is concluded, in part, that the potential of Landsat data as a tool in urban analysis is considerable, particularly when used in association with other readily available data.

TABLE OF CONTENTS page

ACKNOWLEDGEMENTS

To Professor P.V. Angus-Leppan and Professor R.B. Forrest for supervising this thesis and constructive criticism of the initial draft.
To Mr. K. Bullock for assistance in collection of ground data.

To Mr. F. Bird, Valuer General, for allowing access to the records of his department.

To Dr. A. Green, CSIRO, North Ryde, for allowing access to computer compatible tapes of the Sydney scene.

1. REMOTE SENSING IN URBAN LAND CLASSIFICATION

1.1 Introduction

Remote sensing is the science of collecting information about an object or area without visiting it. A number of sensing devices can be used including photographic cameras, multispectral scanners, thermal scanners and microwave instruments; which when mounted on a balloon, aircraft, rocket or satellite can provide a great deal of information about the earth - its surface, environment and resources.

Crowded into less than one percent of the exposed land surface of the earth, urban areas are the focus and ultimate expression of human activities. They are the most modified portion of the earth's surface and their physical expression in the landscape is both complex and diverse. The benefits of remote sensing in urban planning, urban land inventory, urban management and urban research have been reported elsewhere, particularly by BRANCH (1948 and 1971) and BOWDEN (1975) and will not be considered further.

With the launching of the Landsat series of satellites in 1972, 1975 and 1978 a new remote sensing tool became available for urban analysis. While the multispectral scanners of Landsat were primarily developed for agricultural and resource studies, a considerable amount of research has recently been directed towards urban land classification.

The purpose of the research which is the basis of this thesis is to extend the current application of Landsat data in the classification of urban residential areas and to test the potential of Landsat derived variables as a surrogate for determining urban residential quality. The selective but detailed review of previous research which follows will point out past achievements, stress changes in methodology and sensors, indicate current problems and provide a foundation for the chapters that follow.

1.2 Historical Overview - Nadar to Landsat

1.2.1 Conventional Black and White Photographic Sensors

The use of remote sensing in urban areas dates back to 1858 when Tournachon (later known as "Nadar") used a camera carried aloft in a balloon to study parts of the city of Paris. In his photographs of the village of Petit Bicetre the houses could be clearly seen (AM. SOC. of PHOTO, 1975, 27). Increasingly since then the imaged morphology or form of the urban space has been used directly or as a surrogate for functional identification and classification.

Until the second world war remote sensing techniques had remained basically the same since the acquisition of the first aerial photographs. Either a single photograph or stereo pair was used for each portion of the area of interest. Photography was essentially black and white and was of medium to large scale because of flying height limitations. Documented use of this conventional aerial photography for urban area analysis dates from the period immediately following the first world war. BRANCH (1948) makes reference to a number of early writings on the use of aerial photography in urban planning. These include PICK (1920), BRITTENDEN (1920), HAYLER (1920), LEWIS (1922) and MATTHES (1927). Other early writers include LEE (1920) and JOERG (1923).

The second world war brought about a dramatic increase in the awareness of the potential of remote sensing in urban areas. A textbook by BRANCH published in 1948 was the first attempt to examine the full potential of aerial photography in urban analysis. This book has recently been revised (BRANCH, 1971).

Various papers and articles followed which attempted to develop or explore in greater depth the applications and methods originally suggested by BRANCH. WITENSTEIN (1954, 1955, 1956) and WRAY (1960) dealt with the planning use of aerial photography. WITENSTEIN elaborated some of BRANCH's concepts concerning the role of aerial photography in programmes of urban planning and administration. He gave examples which showed that most of the measurements involved in the quantitative analysis of an urban area could be made directly from the imagery. WRAY developed methodology for implementing part of the analytical process outlined by WITENSTEIN. A photomosaic of Peoria, Illinois was used for extensive statistical analysis, involving many physical parameters of the urban area. These included residential and non-residential building and floor space. As a consequence of these studies, areas were grouped to form functional units or urban districts of related land use, function, age and characteristics of development. This work laid the foundations for later contributions to the USGS Census Cities Project (WRAY, 1970). Other studies involving dwelling unit estimation include HADFIELD (1963) and BINSELL (1967).

The relationship between land use classification systems and photo interpreted data was examined by GWYNN (1968). Both GWYNN's work and the work by NUNNALLY et al (1968) and NUNNALLY (1970) suggested that classification systems based on conventional concepts of land use required modification if the full potential of remote sensing as a source of land use data was to be achieved.

The use of conventional black and white photography in socio-economic and demographic studies, as suggested by BRANCH, was further extended by GREEN (1957) and GREEN and MONIER (1959). The goal of these studies was to predict socio-economic variables from photo interpreted data. Extensive literature was cited as evidence for social values attached to housing and residential communities and, by extension, that *observable physical data have meaningful sociological correlates*. Four photo data items were used - concentric zonal location, housing density, number of single family homes and residential desirability. This latter measure was determined from internal and adjacent land uses, each sub-area being classified as having favourable, neutral or unfavourable attributes. Socio-economic data included medium income, rental value, racial composition, educational and occupational status and crime rates. High correlations were found to exist between selected items. The generality of GREEN's approach, his aggregate treatment of data and the choice of data items have been criticized by WITENSTEIN (1957, 97) and WELLAR (1968a). Nevertheless it is still significant that GREEN's scale of residential desirability accounted for 78% of the variation in his socio-economic status scale.

MUMBOWER and DONOGHUE (1967) substantially revised GREEN's approach. Their investigation to determine photographic means of upgrading census type data and defining poverty areas was more correct statistically, in that disaggregated units were first examined - where verified evidence of poverty existed - and these were later aggregated to make more general statements. Photo interpreted surrogates for poverty included residential location, structure type, density, adjacent land use and transportation routes, presence of litter, absence of landscaping and off-street parking.

A further study involving the mapping of urban poverty areas was that by METIVIER and McCOY (1971). Using large scale (1:6000) black and white photographs two methods were employed. The first delimited substandard housing on the basis of the visual appearances of single family residential areas on the photographs. Similar variables to those used by MUMBOWER and DONOGHUE (1967) were identified. Housing density was considered the most effective variable for identifying substandard housing areas. The second approach was to map measured values of single family housing density to determine whether a single quantifiable index obtained from aerial photographs could be used as a surrogate for income and house value. A high negative correlation was found between housing density and both median income and average house values, while density and percent non-white population were

positively correlated. WESTERLUND (1972, 39) suggests that this one variable approach, using housing density, has value for initial analysis in areas where these correlations have been verified. A technique for automated interpretation of housing density using the spacing component to produce diffraction patterns, was further suggested.

The use of aerial photographs as auxilliary census data was studied by EYRE et al (1970) and SHIN - YI HSU (1970). Although a number of difficulties were encountered, the techniques devised were considered most useful for intercensal extrapolation of census data.

Work in Britain by COLLINS and BUSH (1969), COLLINS and EL - BEIK (1971) and GIBSON (1976) has further extended the use of conventional aerial photographs in the acquisition of urban land use information and for derelict land studies, by applying previously researched techniques to real world problems.

Research at the International Institute for Aerial Survey and Earth Science (I.T.C.), Netherlands, has led to some interesting applications of conventional black and white photography, particularly since the establishment in 1968 of a new department for research and training in urban survey with aerial photography. DE BRUIJN (1976, 185 - 186) suggests however that techniques developed elsewhere, especially for American cities, have not been particularly applicable to the more compact and varied urban areas of Europe.

All of the previously cited works relied exclusively on visual interpretation using contextual clues of site and association, no trend towards the computerized pattern recognition techniques of today being apparent. This was to be expected, as at large scales and in a recognisable one dimensional intensity mode the eye/brain combination was a superior analyser of aerial photographs. Nevertheless over the same period initial, theoretical statements of the problem were offered by LATHAM (1959), ROSENFELD (1962a, 1965, 1966) and COLWELL (1965). ROSENFELD (1962b) also confirmed the utility of classifying land use by automatic visual texture measurements. Large scale black and white photographs were spot scanned and the density information converted to time-varying video signals. From these ROSENFELD extracted a measure of contrast frequency which allowed differentiation between urban residential areas, industrial areas, cultivated fields and woods.

1.2.2 Advanced Photographic Techniques

The sixties and early seventies saw a gradual change in the conventional approach to remote sensing with the introduction of colour, colour infrared, multiband , small scale and space photography. This was due to the increasing availability of this type of data, many of the techniques having been wartime and post-war military developments which were only declassified in the early sixties. Declassification provided a major stimulus to the investigation of non-military applications.

A further stimulus was the increasing involvement of the United States National Aeronautics and Space Administration (NASA). NASA accepted a major responsibility for developing civilian applications of remote sensing. The Earth Resources Program, initiated in 1965, involved the development of aircraft and spacecraft sensing systems, and the investigation of earth sciences/geographic applications in conjunction with many other Federal agencies and contractors.

Most American research in urban and regional applications of advanced photographic techniques have however been sponsored by the Department of the Interior, through the Earth Resources Observation System Program (EROS). This program is administered by the U.S. Geological Survey. Their Geographic Applications Program (GAP), a component of EROS, has produced a very extensive collection of studies related to urban applications of remote sensing.

The advent of advanced techniques broughtabout changes in emphasis in urban applications research. There appeared to be fewer studies that dealt specifically with the use of these types of data in the traditional urban comprehensive planning situation than was the case with conventional aerial photography (WESTERLUND, 1972, 56). Also as a natural result of the trend towards smaller scales and greater coverage, advanced technique research has emphasized regional rather than intra-city application, with the exception of housing quality studies (WESTERLUND, 1972, 57).

Application of multiband and colour infrared photography to the evaluation of housing quality, neighbourhood environment, and socio-economic criteria was pursued by the Universities of Northwestern and California - Riverside in a joint research program during 1968-1970. Much of this work built directly upon the earlier studies of GREEN (1957) and MUMBOWER and DONOGHUE (1967). WELLAR (1968a) used large scale aerial photography in nine spectral bands to evaluate the quality of housing and

neighbourhood environment. Residential blocks in five neighbourhoods of Chicago and immediate suburbs were examined. Over twenty variables extracted from the imagery were used in the housing quality studies. It was found that twenty of the twenty four items used by the American Public Health Association as criteria for housing environment appraisal, could be consistently interpreted from the photography.

A study using estimated income, house value, and other socio-economic criteria, based on correlated surrogates observed on colour infrared imagery, was described by MOORE (1969a) and THROWER, MULLENS and SENGER (1968). Three Los Angeles County neighbourhoods were studied. These areas were characterized by low to lower-middle income residents and poor housing. The imagery ranged in scale from 1:6000 to 1:60 000. Eighteen data items were interpreted and verified by reference to an extensive file of ground truth areas. A condensed variable set of nine categories was ranked on a quality scale and census tracts were ranked according to the quality of each variable. Tracts were also ranked according to a selected set of socio-economic variables. Various significant correlations were found between each data set; for example the results indicated a satisfactory ability to differentiate between income levels using interpreted surrogates.

In a related study, MOORE (1969b, 1970) used existing documentary data on housing characteristics for the same three residential areas in Los Angeles. Remote sensing data was not used in the study, however a reduced set of seven environmental variables, determined using discriminant analysis of the documented data, satisfactorily assigned housing at the block level to quality classes. It was argued that based on the work of WELLAR (1968a) these environmental variables could be interpreted from large scale multi-band aerial photography. The possibility of assigning quality classes at the parcel level from environmental characteristics alone, was however rejected.

HORTON and MARBLE (1969) using infrared photography tested MOORE's seven environmental variables by estimating their values for several 10% block samples in a number of areas of Los Angeles County. They concluded that for quick look surveys of blight and decay an acceptable error level was achieved. BOWDEN (1970) further reviewed and analysed the results of these residential studies.

In a study of metropolitan Boston LINDGREN (1971) used colour infrared imagery at a scale of 1:20 000. Objectives of the study were to determine whether dwelling unit estimates could be made from medium scale

imagery and further whether at this scale colour infrared offered advantages over panchromatic or normal colour photography. As in previous studies residential structure counts were highly accurate (99.5%). Dwelling unit estimates were less accurate giving a 3.1% underestimation overall and up to 10% in given blocks. This result was to be expected since areas were primarily multi-family. These estimates compared well with previous work at larger scales. LINDGREN concludes that accurate dwelling unit estimates can be made from imagery at a much smaller scale than have been used previously and that colour infrared imagery is more effective than panchromatic or colour because of its haze penetration capabilities and because of the greater contrast achieved between vegetation and buildings. He recommended still further tests at smaller scales. EYRE (1971) in a more comprehensive regional study, indicates that the type and density of residential development can be estimated within reasonable limits from 1:60 000 scale colour and colour infrared photography.

The ability to evaluate dwelling units counts from smaller scale photography was demonstrated by DUEKER and HORTON (1972) using colour infrared photography at a scale of 1:50 000. They found that the average error in the suburban single family dwelling unit count when compared to census data was 15%. This compares favourably with the earlier results of HADFIELD (1963) and BINSELL (1967) using much larger scale imagery. Regression models were developed to estimate population from dwelling unit counts.

In a report by BOWDEN (1968) a study is cited using 1:60 000 colour infrared photography. Visual interpretation techniques were used to demonstrate that a hierarchy of residential patterns could be categorized, ranging from category one being widely spaced homes of large size and much vegetation, to category twelve being of poorest quality. It was concluded that better socio-economic areas were easier to dissect than the poorer areas because vegetation was better developed and aided greatly in classification. Delimitation of low socio-economic status areas was found to be a much more difficult task using only image derived information.

In a later study by TUYAHOV et al (1973) various levels of housing and environmental quality were assessed and mapped using conventional black and white photography at a scale of 1:23 000 and sub-orbital colour infrared photography at a scale of 1:190 000. Eighteen indicators or image signatures of different socio-economic classes were identified. Those considered potentially observable from satellite imagery included housing

density, geographic pattern and uniformity, surrounding amounts of vegetation, and percentage of roof cover. Using census tract data the authors carried out a factor analysis, the results of which were used to verify the delimitation of socio-economic classes as determined from the imagery. They concluded in part that the location, spatial arrangement, density and character of the housing units have ecological meaning in that exterior characteristics are surrogate measures for interior conditions - a statement not unlike that used as a premise for the work of GREEN (1957), sixteen years before.

In the broader area of urban land use classification a number of studies were also carried out in this period. Typical of urban land use interpretation experiments using low altitude colour photography was a study by HANNAH (1969). This study prepared a land use map of a portion of Asheville, North Carolina from 1:8000 normal colour and colour infrared photography, comparing the results with a similar map prepared from field investigation. Only 1.5% of interpreted land uses, of 1713 parcels, were misrepresented on the photo derived map. Using smaller scale imagery (1:120 000) over Boston SIMPSON (1970) was able to identify eleven major land use classes and thirteen subclasses of features. Using imagery at 1:500 000 GRIFFITHS et al (1971) report that only generalized land use information can be interpreted. TUELLER and LORAIN (1971) in a study of the Lake Tahoe basin indicate that urban cover - primarily housing - could be readily distinguished on 1:108 000 colour infrared photography but with greater difficulty at a scale of 1:432 000. For combined analysis with colour, colour infrared and one or more multispectral bands, they conclude that the red band gives the optimum contrast between cultural and natural features.

Increasingly, research in the land use area was being directed to sub-orbital and satellite imagery. WRAY (1970) describes the objectives of the U.S. Geological Survey Census Cities Program which in part was to determine the capability of high altitude and satellite imagery. WELLAR (1968b, 1969) discusses the applied aspects of hyperaltitude photography in urban and transportation research based on detailed analysis of Gemini photographs over various urban centres at scales of 1:283 000 to 1:1 000 000. From observation he suggests that automatic methods of urban signature discrimination of urban boundaries, central business districts, residential areas,transportation systems and areas of new construction might provide a means of monitoring gross changes in urban and regional land use. Work

by RUSHTON and HULTQUIST (1970) and HULTQUIST et al (1971) at the University of Iowa describe the application of Apollo imagery for monitoring the location and extent of built up areas.

Typical of several efforts at regional land use analysis from manned spacecraft photography, was a study by MACPHAIL and CAMPBELL (1970). They analyzed 17 frames of Gemini 5, Apollo 6 and Apollo 9 photography of the El Paso region of Texas - New Mexico. Imagery was normal colour film at print scales of approximately 1:800 000. Four thematic maps were prepared based on interpretation of the photography. The broad classes of features identified were transportation, drainage pattern, land forms and gross land use. Various subclasses within each were also identified. In the urban land use context only urban irrigated and suburban irrigated were separated.

A similar regional project was undertaken at the University of California, Los Angeles by THROWER et al (1970) using photography from four Gemini and Apollo missions over the south western area of the United States. Objectives of the study were :

(i) to determine what land uses were visible on satellite photography,

(ii) to devise a land use classification system compatible with data obtainable from such photography, and

(iii) to construct land use maps at scales of 1:250 000 and 1:1 000 000

Using visual interpretation a number of land uses and colour textural associations were identified. Included in these were large urban centres which were discernible from the overall shape and size, the presence of street patterns and their greyish coarse textured signature. No sub-classification below this class was indicated. An extensive verification procedure was followed, the results of which indicated in part that while agricultural patterns checked well, large urban centres were somewhat less reliable and further that overlap in the range of reflectance characteristics associated with different land uses varied from place to place making consistent interpretation difficult. In the second part of the study a classification scheme was developed that limited urban areas to one class, settlements, with no sub-classes.

THROWER et al conclude that satellite photography has utility for small scale generalized land use mapping. The use of automated techniques for high reliability categories is suggested but they also indicate that other categories require a greater man-machine interface. The essential

problem of extracting land use data from space imagery was well stated by the authors "*.....at orbital altitudes a single photographic resolution cell represents the integration of a variety of spectral responses associated with a number of phenomena and their condition .... consequently, generalising a variety of such cells into land use categories requires sophisticated interpretation and inference.*" This was one of the first statements of this important problem.

In a similar vein SIMONETT (1969) suggests that major land use categories within urban areas, such as newer versus older residential areas, can only be interpreted with imagery of resolution 100 feet or less and further (SIMONETT and COINER, 1971) that the ability of a remote sensor to provide a given informational element from within the context of a specific environment is related to the spatial (and temporal) complexity of that environment judged by land use mix. Discussing the expected resolution of 320 feet for the Landsat satellite sensor systems, they suggested that in heterogeneous or urban areas its value would be limited.

Other researchers have also developed urban land use classification schemes for thematic mapping from orbital imagery that suggest the level to which satellite imagery can be interpreted. These include WRAY (1971) who gave a gross category of residential and subcategories of single family and multifamily dwellings and ANDERSON et al (1972) who defined residential, strip and clustered settlement and mixed level II categories, among others, within a level 1 category of urban and built up land.

1.2.3 Non-Photographic Sensors

The development of non-photographic remote sensing systems, primarily thermal infrared, multispectral scanners and radar, and research into their application has generally paralleled that of advanced photographic techniques. However the amount of research into urban applications using these new sensors had, until the early 1970's been comparatively small. This was due to the reduced availability of the data and the lower quality of the earlier unclassified systems. Research with these tools was mainly in the investigation of earth science and natural resource applications.

Identification of rural and urban cultural features, usually transportation and settlement features of a fairly gross nature were the principal research application of thermal infrared and side looking airborne radar (SLAR). Two papers, WELLAR (1968c) and MOORE (1968) offer preliminary examinations of the potential utility of both these sensors

systems for urban analysis. Later work by LEWIS (1968), SIMPSON (1969), LEWIS et al (1969) and BRYAN (1974) have further developed application of radar type sensors. All papers point out the strong returns observed from urban linear features caused by multiple reflections. This results when the radar beam is perpendicular to the surfaces of a line or group of buildings on the ground. BRYAN (1974) lists several types of urban land use that were easily and accurately identified. These included residential areas, linear transportation features, large water features and some commercial districts. All were identified at a 5% level of confidence. Wavelengths used were X and L Band and polarizations HH and HV. LEWIS (1968) concluded that multiple-polarized radar imagery was superior to any single polarization scheme for detection of cultural features.

The Willow Run Laboratories (now ERIM) pioneered development of multi-channel imagery systems that spanned the visible to the far infrared spectrum. Together with the Laboratory for Agricultural Remote Sensing (LARS) at Purdue University, they applied these systems to the investigation of crops and soils with some of the first applications of automatic discrimination procedures based on spectral signature, for example see HOFFER (1967), WIEGAND et al (1968), HOFFER and GOODRICK (1971). Work by COLWELL (1971) extended this procedure to urban features. Data was recorded in two separate sets using medium altitude multispectral scanning. Ten bands were used in the 0.4 μm to 1.0 μm region of the spectrum for the first data set and three infrared bands for the second. The prediction of surface cover type for urban watershed management was the main aim of the study. Using training site data, automated recognition procedures were applied to give a recognition rate of 85% for non-vegetated surfaces using the ten channel data set. Most of the non-recognitions were dark toned rooftops not included in the training sets. A high level of recognition was also achieved for such categories as bare soil, gravel, asphalt, a variety of roofing material, lawn, trees and water. However a slightly lower level of recognition was achieved with the three channel infrared data.

1.2.4 Landsat MSS Applications

The greatest advance in non-photographic Earth orbital imagery was undoubtedly the launch of the first Earth Resources Technology Satellite, ERTS 1 (now renamed Landsat 1) on July 23rd 1972. The concept and specifications of this general purpose earth observation satellite were developed within the Earth Resources Observation Satellite Program, now

the Earth Resources Observation Systems (EROS) Program of the U.S. Department of the Interior. Using two independent imagery systems - three vidicon cameras and a four channel multispectral scanner - the satellite began returning pictures within three days. However, following the failure of the vidicon system, research has concentrated on the application of the multispectral scanner system. Wavelength ranges covered by the four channels are 0.5 - 0.6 μm (green) 0.6 - 0.7 μm (red), 0.7 - 0.8 μm (infrared), 0.8 - 1.1 μm (infrared). Two further Landsat satellites were launched in 1975 and 1978. (See Appendix 2 for a more detailed decription of the Landsat system).

The repetitive, synoptic, regional overview of Landsat stimulated much research into urban and regional applications. Many studies particularly in the United States were commenced with great expectation, broad land use classes were mapped and urban boundary change detection was attempted. While visual interpretation of Landsat imagery can be used for classification purposes, the complexity of the urban landscape and the vast amounts of data has led to a greater use of computerized pattern recognition techniques. In general the methods used were essentially those developed for agricultural studies where the spectral and spatial responses were considered sufficiently simple and consistent to encourage their development (WESTERLUND, 1972, 137).

Basically two approaches have been used, supervised and unsupervised training. Each assumes that a number of classes exist in the scene that will each give a unique spectral signature. The response vectors from the same type of area will tend nearly to be the same and will form a cloud or cluster in n-dimensional space (n = 4 for Landsat spectral data), each cluster being contained in one region and representing a particular class. During the classification process unknown data points are tested to see into which cluster and therefore which class they fall. In the case of unsupervised methods there is no prior knowledge of the identity of the classes, and clusters are not forced. Following complete classification, training areas are ground truthed to determine the true identities of the different classes. With supervised methods the classes are defined beforehand and cluster statistics are determined from responses sampled over ground truth areas within each class.

Early research using Landsat data in urban areas was reported by ERB (1974). Imagery over Houston, U.S.A., for August 1972 was supported by aerial photographs as collateral ground truth data. Both unsupervised and

supervised clustering techniques were used to derive a number of land use cover classes. Considerable effort was expanded in an attempt to develop distinct spectral signatures for residential areas in four general categories:

(i) single family - with dense cover of trees,

(ii) single family - with less dense cover of trees, larger lots, many open fields,

(iii) new single family - with very small or no trees, small lots and over all highly reflective,

(iv) multi-family apartments - with extensive paved or roof covered surfaces.

ERB reports that separating these categories spectrally with a reasonable degree of accuracy proved almost impossible, because of confusion with other uses. One example of confusion given was where a pixel (picture element) straddled a bright paved surface and adjacent dark vegetation, giving an integrated response similar to the residential response. It was concluded that the best accuracies for classifying residential areas would be achieved using only two level II categories, single family residential and mixed urban. The latter category included all other land uses not included in the commercial/industrial/transport category and by necessity included some residential areas that could not be separated from adjacent non-residential areas. ZOBRIST et al (1976) report a more detailed breakdown having large buildings, strip cluster development, single family residential and multiple family trailer courts as separate classes. Other categories defined by ERB (1974) were vegetation (woody) being predominantly trees, vegetation (non-woody) including grasses and shrubs, and water. Overall accuracies obtained using an unsupervised classification technique when compared to ground truth data obtained from high altitude aircraft photography were:

Commercial etc.	94.2%
Residential	66.8%
Mixed	51.1%
Vegetation (woody)	95.1%
Vegetation (non-woody)	56.2%
Water	87.7%

Based on ERB's figures however, if the residential and mixed urban were combined into one category, those mixed urban areas would have a 71.9% correct classification in this combined category and residential areas 89.8% accuracy.

Using supervised training techniques with four classes of ground truth sites, a further analysis was carried out by ERB. Eighteen training sites were aggregated into four general urban land use categories, vegetation, residential, commercial/industrial and water. Thirty-five randomly located sample areas were classified on the basis of the response statistics derived from the training sites. These results were compared via regression analysis with conventionally interpreted data. Correlation coefficients computed were, for commercial/industrial 0.75, vegetation 0.73, residential 0.72 and water 0.80. Generally vegetation and residential classes were over classified when low percentages occurred and were under classified when high percentages occurred.

ERB (1974) suggests that the residential land use category is one of the most complex and least consistent categories to be delineated by spectral classification, because of the spectrally heterogeneous inter-mixing of small vegetated and non-vegetated surfaces. He concludes that *urban classification is difficult because of the integrated response from pixels of mixed surface cover proportions and more particularly where vegetation is a major component of the urban scene.* He further concludes that Landsat data is only capable of providing generalized classifications that would have only limited application in urban land use studies. The solution of the mixed pixel problem has been considered by a number of researchers, including HORWITZ et al (1971), SMEDES et al (1975) and JENSEN (1978), however none of these studies were undertaken primarily in an urban situation and generally only dealt with simple mixtures of two surfaces.

ELLEFSEN et al (1973) in a study over the San Francisco Bay area supported ERB's (1974) contention when they suggested that integrated response from different ground cover classes in a single land use class was a significant factor affecting urban classification results. In addition they suggested that regional environmental characteristics affecting landscaping, differential weathering of man-made surface materials and local varieties of building and paving materials tend to reduce identification reliability. However they conclude that for ninety percent of the urban area mapped, broad land use classifications were correct.

Improved techniques were adopted by TODD et al (1973) and TODD and BAUMGARDNER (1973) in an attempt to overcome some of the urban classification problems. In the former study over Milwaukee County, Wisconsin, U.S.A., two types of spectral response were defined - *homogeneous* and *heterogeneous*.

They hypothesized that there were at least three basic homogeneous groupings of earth phenomena and that the predominantly heterogeneous urban classes of land use comprised various proportions of a minimum of two or three spectrally diverse groups of phenomena. The broad classes of urban phenomena were then defined on the basis of the relative proportions contained in each, rather than specifically being based on land use or function. Classes defined were roads/central business district (C.B.D.), inner city, zone of transition, industrial, suburban, wooded suburban and grass. Using supervised clustering methods overall identification accuracy was reported as greater than 90%, with most misclassifications occurring on the rural fringe.

In the latter study over Marion County, Indiana, TODD and BAUMGARNER (1973) considered a number of new approaches. Response data was averaged over a number of contiguous picture elements prior to classification. This procedure reduced the variance of clusters and made spectral characteristics more representative. Stratification of the data set into urban and rural land uses prior to classification also improved the analysis. However the most innovative technique introduced was the inclusion of numerical spectral characteristics, other than single spectral class, into the clustering process. Parameters such as mean, range, standard deviation and correlation coefficients (between spectral bands) were considered. Classes considered were commercial/industrial, multifamily (older) residential, single family (newer) residential, wooded areas, grassy areas, water,clouds and cloud shadow. While all of these classes were separable using only spectral data, another two residential classes could be separated using additional parameters. These were a transitional residential area located between areas of multifamily and single family classes and a vegetative residential class consisting of scattered residential developments built after the second world war and housing upper income families. Most useful parameters in this separation were the means and standard deviations in the infrared bands. Of the other parameters only one, the coefficients of correlation between bands 6 and 7, proved helpful. Its value was reported as being 0.95 or greater for grassy areas whereas for other land use areas it was 0.83 or less. By evaluation of these additional parameters the authors claimed that recognition over all classes was virtually 100% correct.

SWAIN (1976) also tested a number of auxiliary techniques with less positive results, for application in land use classification. Prior to

cluster analysis a processor was applied to the Landsat data to group homogeneous data sets and hold them as such in the clustering procedure. Parameters allowed for variation of the number of pixels grouped in a block, the degree of homogenity, and for further groups aggregation. Non-grouped pixels were treated singularly in the clustering procedure. Use of this processor improved overall classification from 85.7% to 92.8% in one case and 80.8% to 92.5% in another case. However only four basic classes were defined, these being rural, urban, water and forest. Two sets of texture features were also tested by SWAIN. The first set derived from grey tone angular spatial relationships, the grey tones being defined by quantifying the data range into eight equally occurring levels. These included measures of homogenity, linear dependencies, and randomness over a 15 x 15 pixel sub-image, and averaged over a number of angular directions. The second set of texture measures were derived by generating a discrete power spectrum for a 16 x 16 pixel data set. Seven sampling rings were used and the average power spectrum value in each ring was determined and scaled to a 0 - 255 range. This procedure effectively gave the average power of the spectral response in seven different frequency ranges. Thus a total of 15 dimensions could be used in the clustering process (4 spectral, 4 grey tone, and 7 power spectrum). Virtually no improvement in classification occurred using any of the combinations when compared to the use of spectral data alone. In the case of the power spectrum texture measures, used in conjunction with spectral data, overall accuracies were upwards of 15% less accurate. The grey tone texture features were considered least effective when the number of boundaries increased and field sizes decreased. This would suggest that these methods would have least application in urban areas.

An extensive study by CARTER (1977), over selected urban areas in England, also applied a number of processing algorithms prior to a supervised clustering procedure. Normalisation was carried out by dividing each individual intensity value of a pixel by the sum of the intensity values in all four channels. This procedure was introduced as a first order correction to reduce variations due to shadowing and atmospheric effects. Ratioing techniques were also applied. It was considered that this procedure would give enhanced output for healthy vegetation using the ratio of Landsat bands 7 and 5 (infrared and red) but in addition would also reduce signal fluctuations caused by source noise and atmospheric conditions. A low pass filtering technique was also applied to the data, consisting of a simple digital filter which replaced the intensity value of a pixel by a weighted

average value obtained from the eight surrounding pixels. To improve separability of clusters that overlapped in the four dimensional Landsat space, a principal component transformation was applied. This technique essentially rotates and shifts the axes so that a new set of mutually orthogonal axes is created with the origin at the mean of the data set and having the property that the variance is maximised along the first new axis and decreases along each subsequent axis. Finally a simple density slicing technique was applied to an infrared band as a means of separating urban and rural classes.

Results given by CARTER show a 7% increase in the accuracy of recognition of urban areas using normalised data when compared to classification using original data, however the accuracy of classifying rural areas was reduced. Using the principal component transformation of the normalised data practically all the class separation was evident on the first principal component axis, thus the dimensionality of the problem was reduced without loss of classification accuracy.

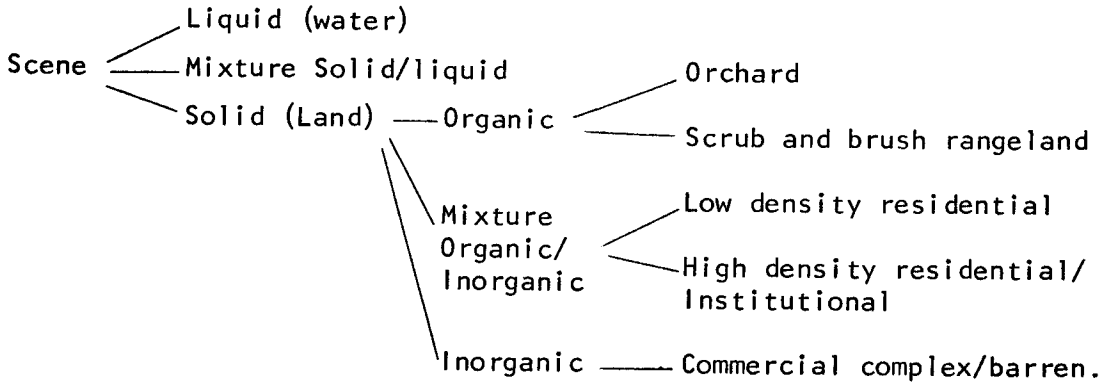
The reported results of including two ratio channels (Band 7/Band 5 and Band 4/Band 5) with the four original channels reduced the accuracy of classification of the urban areas by 5% and increased that of the rural areas by 2%. When the two ratio channels were included with the four normalised channels results almost identical to that using only the four normalized channels were obtained. A reduced classification of 75% was reported for the density slicing tests, while pre-smoothing or low pass filtering of the data before classification caused a very high misclassification rate. CARTER reports that urban areas were typically classified with 80% accuracy compared to the ground truth data, but questions in some cases the appropriateness of the ground truth classifications, which were derived from other sources.

Rectification errors (i.e. between ground data and satellite data) were considered a further problem affecting classification accuracy. Residual coordinate errors ranged from 34 metres to 218 metres along scan, and from 55 metres to 261 metres across scan. The requirement of high spatial relationship between ground truth and satellite data in urban areas has also been stressed by FORSTER (1980a).

FRIEDMAN et al (1979) have also used ratioing techniques and a form of principal component transformation for a binary classification of land cover into urban and non-urban classes. They constructed a ratio transformation of Landsat data by dividing Band 5 by Band 7 but suggested, on

the basis of previous work by FRIEDMAN (1978), that a better ratio would be Band 5 divided by Band 6. Severe striping in Band 6 prevented their use of this ratio. Using a histogram of the Band 5/Band 7 ratio the distinction between urban and non-urban surfaces was reported as not particularly effective. From examination of the principal component analysis they determined that the land-water interface was highlighted by the first transformed axis while urban and non-urban surfaces were more distinguishable on the second axis. Because the transformations were used to classify data, essentially for urban change detection, no absolute measure of accuracy was given.

Following on work by SWAIN et al (1975), JENSEN (1978) applied layered classification logic to a primarily urban scene. Basic pattern recognition essentially used a clustering type algorithm, however the layered classifier filtered the unknown data points through a sequence of tests, each of which were dependent on the outcome of previous tests. JENSEN considered that such a decision tree type approach would define a more logical classification procedure than a single stage analysis. The layered classification logic was given as follows for development of urban classes -



Using conventional classification methods the percentage identification accuracy for low density residential, high density residential/ institutional and commercial complex/barren were reported as 82%, 94% and 87% respectively. After application of the layered classification these improved to 90%, 98% and 90%.

An alternative approach to the inclusion of transformed spectral data was reported by ELLEFSEN et al (1974). In this study pattern recognition techniques were applied to a combined multispectral and multitemporal data set. In a first test application, over the Phoenix metropolitan area, results using data from October 1972 and May 1973 showed

a marked improvement in rural-urban separation over a single date classification. The authors reported that unintentionally this procedure also resulted in the uncovering of a number of sub-classes, for example, older and newer residential areas, a level of discreteness beyond some earlier studies. They suggested that Landsat data held promise for such distinctions as *the quality of housing and the nature of open space.*

In a recent, extensive study of the Sydney metropolitan area BAILEY (1979) makes a clear distinction between land use and land cover or landscape. A number of urban "landscape" classes were defined, a total of fourteen in all, which included for example, houses with many trees, separate houses/small factories, units/factories/terraces/bare earth, short grass, beaches, inner city/shopping/commercial, among others. She suggests that it is the ratio between the elements of the urban landscape which causes groupings, chiefly the ratio of the size of the buildings and/or other man-made surfaces, and a ratio of vegetation to building and other man-made surfaces. A total of eight hundred sites were classified. Careful scrutiny of each of the site classifications showed that 65.5% had been absolutely correctly classified, a further 18.4% cells were mislocated due to registration errors, although the landscape classifications were correct. Thus a total of 83.9% were correctly classified. As the results of the study were for a larger study to quantify the variations within the urban micro-climate, which in part are caused by urban landscape, there was no specific need to define urban land use classes per se.

Another study over the Sydney metropolitan area by FORSTER (1980b) also attempted to disaggregate urban residential cover classes and include the effects of atmosphere and sensor point spread function. This latter effect causes localized scene degradation and has been theoretically examined by DYE(1975). While being of little consequence in agricultural areas, single pixel classification in heterogeneous areas can be markedly affected.

While CARTER (1977) attempted to reduce atmospheric effects by normalising, very few studies up to date have attempted to quantify these effects and apply them in an urban situation, a necessary requirement for signature extension. A number of researchers have however examined the theoretical problems notably TURNER et al (1971), TURNER et al (1972) and TURNER (1975). One recent example of the application of atmospheric corrections to an urban scene is given by KAWATA et al (1978), however only a qualitative expression of the improvement in pattern recognition and

image interpretation was reported. Corrected images included the city of Kanazawa (Japan) and improvements could be seen particularly for small or thin structures such as rivers and roads.

Since the launching of the first Landsat satellite virtually all research in urban areas has been directed towards urban land use or land cover classification, which is in marked contrast to the studies of housing density, housing quality, poverty areas, demography and other socio-economic factors, which were so apparent with larger scale aerial photographs. A number of researchers have however pointed to the potential of Landsat for image surrogate application. ERB (1974) suggests that a measure of vegetative cover can be used to determine density and age. TODD et al (1973) discuss associations between the classified imagery and published socio-economic data. One association suggested was that between medium family income and various land use classes. Inner city classification compared well with the distribution of low income areas, while wooded suburb classification was essentially correlated with high income areas. They suggest extension of these concepts to developing countries where many urban areas have little or no census data.

A more extensive study of the relationship between Landsat data and socio-economic data was carried out by LANDINI and MCLEOD (1979). Basic data for the study was twofold. Land use data over the city of Los Angeles had been derived from Landsat classification. This data was in a form compatible with the city's land use and population files. The seven classes of land use were single family residential, multi-family residential, commercial/industrial, open space, open vacant, other and no data (cloud). The Landsat data was merged with the estimated population and housing data.

The Landsat data items were used as independent variables in a multiple regression analysis to predict each of the socio-economic variables. On a city wide basis correlations were found to be very low ranging from multiple correlation coefficients of 0.27 to 0.46. A sub-area was selected for further analysis, this being the urban/rural fringe, operationally defined as containing those census tracts with less than 1.56 person per acre, based on the U.S. Census Bureau's definition of urban place. The same multiple regression analysis was applied to this reduced area. The relationship between nine population and housing items and the Landsat land use data was shown to have considerably improved. Multiple correlation coefficients ranged from a low of 0.87 for estimated occupied

multi-family dwelling units to a high of 0.97 for total estimated population. Single family land use and open vacant use were shown to be the most significant Landsat data items in predicting total population. The authors conclude in part, that one of the major difficulties to be overcome in the use of Landsat data for socio-economic studies is the development of techniques to increase the number of land use categories. They suggest that it would be more useful to focus on one particular land use type, and further *"that the discrimination of more categories of residential land use would hold the greatest benefit, particularly if this could be expanded to encompass an environmental quality and housing quality index"* (LANDINI et al, 1979, 104).

Results given by MORRISON (1978) confirm the difficulties of population estimation from Landsat data. However good results over Tokyo have been reported by MURAI (1974) who concludes that Landsat Band 7 correlates with high population density and Band 5 with lower population density.

1.3 Summary and Overview

It is quite apparent from the historical background that the regional overview and availability of Landsat imagery has stimulated much research into urban and regional applications. Many studies, particularly in the United States, have been completed that show that broad land use classes can be mapped with better than 80% reliability. However, only limited success has been reported in breaking the residential general class into subclasses, and as suggested by LANDINI et al (1979) this is the area where greatest benefit would derive. Where successes have been reported they have generally been simple dichotomous divisions, such as newer and older housing classes and these generally are land cover classes rather than land use classes. Earlier work with photographic sensors at large scales indicates, however, that residential areas contain a much broader spectrum of classes.

Associated with the greater use of Landsat data has been the increasing application of computerized methods compared with the virtually exclusive use of visual interpretation methods with large scale photographic sensors. In developing urban classes from Landsat data, researchers have relied on the assumption that the area of study is comprised of a number of unique internally homogeneous classes and that cluster or some other form of grouped analysis can be used to identify these unique classes via

ground truth areas. A number of researchers have begun to question this approach which essentially was developed for the recognition of homogeneous agricultural land use classes. BILLINGSLEY (1972) referring to geologic applications suggests that clustering methods *"may prove to be of less value where the data forms a relatively unclustered continuum."* Many writers have pointed out the problem of classifying responses which have derived from a varying mixture of urban surfaces. While a number of attempts have been made to overcome this problem, in general, a marked improvement in classification accuracy has not been forthcoming.

Unlike earlier work with large to medium scale photographic sensors very little progress has been made using Landsat data in image surrogate applications to socio-economic differentiation of urban areas. A number of researchers have inferred or shown for selected areas that a relationship between Landsat data and socio-economic variables does exist, a relationship proved conclusively in pre-Landsat research, for aerial photographs.

Using data derived from larger scale imagery there was little need to account for the effects of the atmosphere through which the reflected energy passed. However for data derived from orbital sensors atmospheric effects are significant, yet there is very little evidence to suggest that these effects have been considered important in urban area studies. A related problem that also causes a degradation of image quality is the effect of the sensor point spread function. While this effect was examined by DYE (1975) for the Landsat multispectral scanner sensors, no evidence exists that it has been accounted for in any urban studies apart from a recent paper by FORSTER (1980b). The effects of low registration accuracy between ground truth areas and Landsat data, an insignificant problem with larger scale photographic sensors or homogeneous agricultural areas, would also appear to be a source of classification error.

There appear therefore to be limitations to the current methods being used to analyse Landsat data over urban areas, leading to a restriction in its potential to distinguish sub-classes of the residential class and its application to socio-economic problems. These limitations are due to a number of interrelated causes:

(i) As the sensor becomes further removed from the scene the interpreter becomes further removed from the contextual clues of site and association so essential

in manual interpretation, with the result that only cover classes and not use classes can be inferred.

(ii) Urban areas are heterogeneous. In such areas the radiation received from a single ground element will comprise radiation from a number of objects or areas which individually may have distinct spectral signatures. Their additive response may not be representative of any one class and a single pixel classification may be incorrect. Urban residential areas typically exhibit this problem.

(iii) The point spread function of the sensor integrates the response from the observed and surrounding pixels. While the recorded response derives predominantly from the observed pixel it is also partially derived from the surrounding pixels. Over homogeneous agricultural areas this is not a problem, except at the interface of two separate classes, but in urban areas it can significantly affect the signature from a single cover class if the surrounding cover is dissimilar.

(iv) Within broad general urban land use classes particularly residential, a continuum of cover classes exist which cannot be easily broken into discrete nominal classes. For example, in one area residential density may be low with few established large trees, another may have a similar density but mature vegetation, while a third may be of high density with little or no vegetation. Between each of these, intermediate examples occur. Thus the response from these areas is not amenable to cluster or similar analysis, because either clusters will have considerable overlap or be extended linear clusters making it extremely difficult or impossible to determine the probability of a feature value being part of one or other classification.

(v) As the altitude of a sensor increase, the effects of the atmosphere on the energy transmitted from the ground to the sensor became more pronounced. The radiation reflected by the surface will be attenuated by the atmosphere as it passes through it to the sensor and an extraneous component of scattered radiation will be added to the transmitted component. This additive component is also dependent on background

reflectance which is relatively constant from homogeneous areas but is not so from spatially heterogeneous areas. In urban areas therefore the degradation of the recorded response due to atmospheric effects will be spatially variable increasing the difficulties of classification.

(vi) The low resolution of the Landsat sensors is the primary or a contributing factor in most of the above. If the sensor resolution was greater the amount of mixing of surface cover in each pixel would be reduced, contextual clues would be enhanced, control point selection would be more accurate allowing for more accurate data registration, and the relative effect of the point spread function would be reduced.

Based on the preceding comments it is considered that a more detailed analysis of the effects of mixed cover response in residential areas and the relationship of cover variables to socio-economic data is a significant area of research.

1.4 Aims and Description of the Research

The aims of this research were:

(i) To determine the reflectance of various urban residential cover surfaces using Landsat digital response data.

(ii) To develop equations to predict the percentage area of particular urban residential cover surfaces contributing to the total reflectance of a pixel.

(iii) To determine the relationship between cover percentages as predicted from Landsat data and average residential house price as one measure of urban residential quality.

These aims necessitated the sampling of ground cover and satellite response from a number of ground truth sites distributed over an urban area; the city of Sydney, Australia, was chosen for this purpose. Residential cover was sampled at the pixel level from seventy areas each containing a block of forty pixels. The equivalent Landsat response was

also determined for each of these pixels. Details of the data collection procedures are given in Chapter 5. The three chapters prior to Chapter 5 examine the theoretical basis for the research. Chapter 2 determines the relationship between satellite response, ground reflectance and ground cover, accounting for the effects of the atmosphere. Chapter 3 details the necessary multivariate statistical procedures required to relate the ground and satellite data sets and Chapter 4 examines various transformations, including band on band ratios, that can be applied to the satellite derived data to improve the satellite/ground relationship.

In Chapter 6 the coefficients necessary to convert percentage cover to reflectance are determined and the derived reflectances are compared with those from other sources. Likely combinations of bands for the prediction of percentage cover are also examined in Chapter 6. The various ratios and transformations examined in Chapter 4 are applied to the prediction of cover percentages in Chapter 7 and these results are then used in Chapter 8 to examine the relationships between reflectance derived variables and residential quality. Final conclusions are given in Chapter 9.

2. REFLECTED RADIANCE FROM MIXED SURFACES

2.1 Introduction

In this chapter the mathematical relationship between mixed cover reflectance and the radiation received at the sensor are examined. A simple model is developed initially, with the more complex effects of atmosphere and sensor point spread function being added to it.

An individual Landsat pixel is approximately 57 x 79 metres, that is about 0.4 ha. Contained within such an area for a typical residential neighbourhood are five houses more or less, depending on the housing density, with associated areas covered by concrete, road, grass, trees etc. The total radiation at a given wavelength, reflected from such an area, is dependent upon the amount of each surface cover contained within it, and the nature of the reflectance of each surface.

In different areas of a city the cover mixture changes due to higher or lower housing density, smaller or larger houses, wider or narrower roads, more or less vegetation and presence or absence of concrete driveways and footpaths, and consequently the reflected radiation also varies. Therefore when the relationship between reflected radiance and cover mixtures is determined, statements about the residential physical space can be made.

2.2 Definition of Basic Radiation Quantities

Radiant Energy, Q. The energy carried by electromagnetic radiation. Unit is the joule, J.

Radiant Flux, Φ. The time rate of flow of radiant energy. Unit is the watt, W.

$$\Phi = \frac{dQ}{dt}$$

Incident Flux, Φ_i. The radiant flux incident on a surface.

Reflected Flux, Φ_r. The radiant flux reflected by a surface.

Irradiance, E. Radiant flux incident upon a surface per unit area of that surface. Unit is watt per square metre, Wm^{-2}

$$E = \frac{d\Phi_i}{dA}$$

Exitance, M. Radiant flux leaving a surface per unit area of that surface. Unit is watt per square metre, Wm^{-2}

$$M = \frac{d\Phi_r}{dA}$$

Radiant Intensity, I. Radiant flux leaving a small source per unit solid angle in a specified direction. Unit is watt per steradian, Wsr^{-1}

$$I = \frac{d\Phi}{d\omega}$$

Radiance L. The radiant flux per unit solid angle leaving an extended source in a specified direction per unit projected source area in that direction. Unit is watt per steradian per square metre, $Wm^{-2}\ sr^{-1}$.

$$L = \frac{d^2\Phi}{d\omega(dA\cos\theta)} = \frac{dI}{dA\cos\theta}$$

where θ = angle between the line of sight and the normal to the surface considered.

2.3 Reflectance

The hemispheral reflectance, R_h, is defined by the dimensionless ratio of the reflected exitance, M, of a surface to the irradiance, E, on that surface i.e.

$$R_h = \frac{M}{E}$$

As E and M both have units of Wm^{-2}, R_h is therefore a dimensionless quantity. Typically, reflectance varies with wavelength for a particular surface material to produce a characteristic spectral signature. For example, see figure 2.1 for typical earth surfaces (ENTRES, 1974, 43).

The surface of the earth can be considered either smooth or rough. If it is smooth specular reflection is said to occur, if it is rough then scattered or diffuse reflection occurs. Surface roughness is a function of the wavelength of the incident electromagnetic radiation. Rayleigh's criterion of surface roughness is given as

$$h < \frac{\lambda}{8\cos\theta}$$

where h = variations above a plane, in wavelengths.

λ = wavelength

θ = incident angle

When the wavelength of the incident electromagnetic radiation is smaller than the height variations on the surface, the surface is a diffuse reflector and reflects energy in accordance with Lambert's cosine law of radiation. This states that the intensity of the reflected electromagnetic

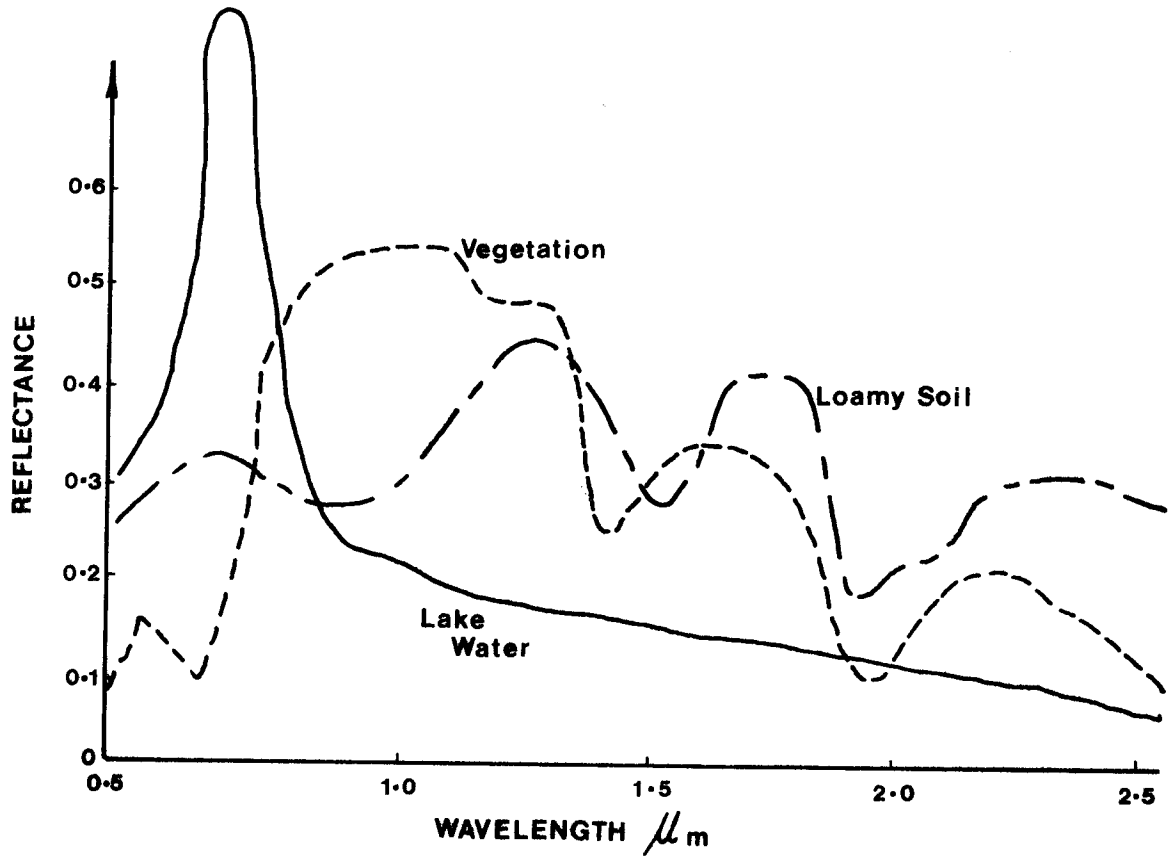


FIGURE 2.1: Spectral reflectance in the visible and near infrared regions of typical earth surfaces (after ENTRES, 1974).

radiation is directly proportional to the intensity of radiation incident on the surface, times the cosine of the angle of incidence. In addition the reflected radiation is unpolarized, phase incoherent and its amplitude fluctuates in a random manner. When the wavelength is greater than the height variations of the surface, the surface is a specular reflector and reflects in accordance with Snell's Law. Here the angle of the incident plane electromagnetic wave equals the angle of the reflected wave. The radiation reflected is phase coherent, polarized and the amplitude fluctuations are small (AM. SOC. of PHOTO., 1975, 80-81).

A mirror is a typical example of a specular reflector and matte white paper that of a diffuse reflector. A surface is usually neither completely smooth or rough and reflection between these two extremes can occur, for example see figure 2.2.

For a Lambertian surface the radiance measured by a sensor will be constant for any angle of reflectance θ, to the surface normal. Radiance depends upon the projected area viewed and for a constant solid viewing angle, the surface area is the projected area divided by cos θ. However from Lambert's cosine law of radiation, reflected intensity varies directly as cos θ, therefore the signal at any viewing angle is constant.

Consider an elemental Lambertian surface dA (see figure 2.3) irradiated by E Wm^{-2}. The radiant flux reflected from dA in any direction θ to the surface normal is given by

$$d\Phi_r = L \cos \theta \, dA \, d\omega$$

Because the elementary solid angle subtended by the area dA is given by

$$d\omega = \frac{2\pi r \sin \theta \, (r \, d\theta)}{r^2}$$

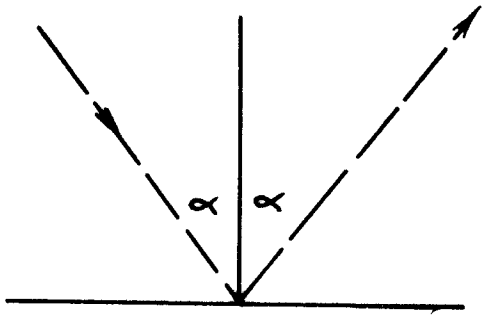
then the total radiant flux in watts reflected into the hemisphere is therefore

$$d\Phi_r = 2\pi \int_0^{\frac{\pi}{2}} L \, dA \cos \theta . \sin \theta \, d\theta \qquad (2.1)$$

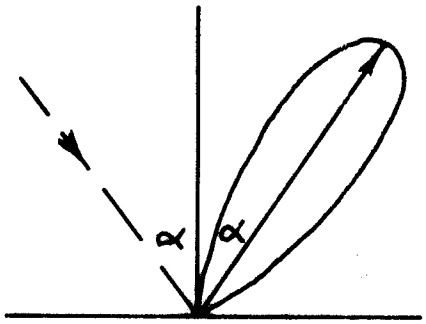
$$= \pi L \, dA$$

The radiant flux incident at the surface is given by

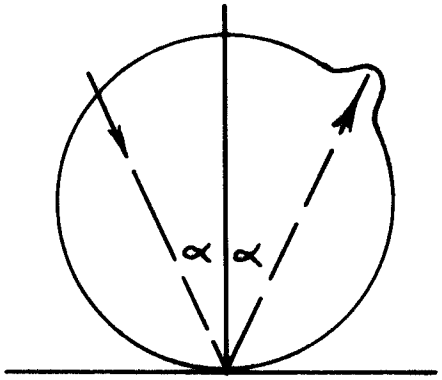
$$d\Phi_i = E \, dA$$



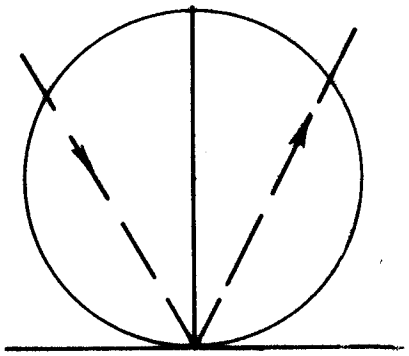
(a) Perfect Specular Reflector



(b) Near - perfect specular reflector



(c) Near - perfect Diffuse Reflector



(d) Perfect diffuse reflector

FIGURE 2.2: Distribution of reflected radiant flux from different types of reflecting surfaces. (AM. SOC. of PHOTO., 1975, 237).

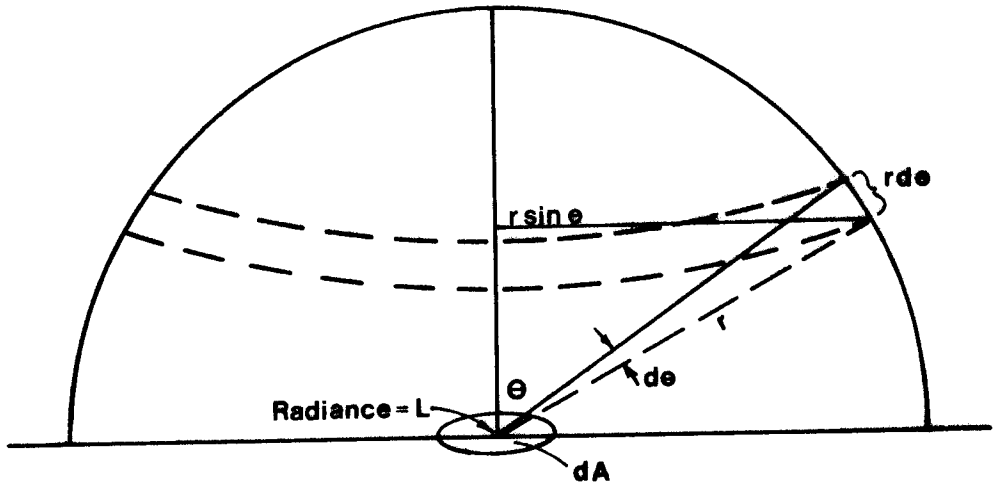


FIGURE 2.3: Geometry of hemispherical interception of radiant flux from a Lambertian surface dA of radiance L.

The ratio of the total reflected radiant flux to the radiant flux incident at the surface, defines the diffuse reflectance of the surface, therefore

$$\frac{d\Phi_r}{d\Phi_i} = R_h = \frac{\pi L}{E} \qquad (2.2)$$

Wavelength dependence is indicated mathematically by

$$R_h(\lambda) = \frac{\pi L_\lambda}{E_\lambda}$$

This is an important relationship because if irradiance can be determined and radiance measured by the sensor, then reflectance can be simply calculated for a diffusely reflecting surface.

If a surface is not a perfect diffuse reflector then the radiance will vary with angle of view and will be greatest in the direction of the specular reflectance component. Assuming the total radiant energy reflected into the hemisphere is the same as for a comparable diffuse surface, then the radiance in any other direction, significantly different to the specular direction will be less than that calculated given Lambertian conditions. Thus the calculated hemispherical reflectance based on this radiance value will be less than the true hemispherical reflectance. For surfaces typically found in residential areas how much specular component exists and how will this affect the calculated hemispherical reflectance? It is generally considered that "....*reflectances of natural surfaces at high solar - elevation angles .....approximate those of diffuse reflectors.*" (AM. SOC. of PHOTO., 1975, 90). Very little data for both the visible and infrared portions of the spectrum exists, however the reflectance at various receiving angles for a number of surfaces in the visible portion of the spectrum have been tabulated (AM. SOC. of PHOTO., 1975, 254-255). Values for selected surfaces are given in Table 2.1. These surfaces were considered likely to occur in residential areas.

For a nadir viewing sensor the radiance measured is proportional to the reflectance in that direction. For aged white concrete, for example, this is 26.6% from Table 2.1.

For constant irradiance E

$$L_\theta = \frac{R_\theta E}{\pi}$$

where the subscript θ indicates directional dependence. From equation

TABLE 2.1: Directional reflectance of urban type surfaces in the visible spectral range.
(AM. SOC. PHOTO., 1975)

Description	Solar Zenith Angle	Azimuth of path of sight relative to sun	Nadir angle of view						
			0^o	15^o	30^o	45^o	60^o	75^o	80^o	85^o
1. Asphalt, oily with dust film blown onto oil.	42.0	0	0.061	0.057	0.058	0.060	0.068	0.090	0.104	0.127
		180		0.067	0.080	0.101	0.090	0.086	0.086	0.088
2. 'White' concrete aged.	42.2	0	0.266	0.263	0.254	0.254	0.266	0.298	0.320	0.374
		180		0.289	0.313	0.343	0.367	0.350	0.343	0.320
3. Calm water, infinite optical depth.	41.5	0	0.022	0.023	0.030	-	0.057	0.139	0.267	0.461
		90		0.022	0.022	0.023	0.029	0.071	0.121	0.214
		180		0.021	0.021	0.022	0.027	0.072	0.125	0.254
4. Grass, lush green closely mowed thick lawn.	40.4	0	0.100	0.096	0.098	0.108	0.120	0.149	0.168	
	39.6	90		0.103	0.110	0.121	0.138	0.159	0.168	
	39.9	180		0.109	0.109	0.119	0.122	0.125	0.125	
5. Macadam, washed off and scrubbed	48.5	0	0.113	0.115	0.119	0.128	0.148	0.194	0.229	
	60.1	90		0.110	0.109	0.116	0.122	0.139	0.147	
	46.0	180		0.126	0.141	0.156	0.166	0.172	0.176	
6. Dirt, hard packed, yellowish	53.2	0	0.243	0.230	0.229	0.239	0.252	0.300	0.330	
	56.5	90		0.243	0.258	0.260	0.276	0.300	0.304	
	51.1	180		0.272	0.313	0.370	0.422	0.432	0.434	
7. Mixed green forest deciduous (oak) and evergreen (pine)	39.0	0	0.036	0.033	0.029	0.021	0.021	0.034		
	37.0	180		0.041	0.049	0.049	0.082	0.263		
8. Grass, dry meadow, dense, mid-summer	45.0	0	0.096	0.090	0.096	0.095	0.108	0.129		
	45.0	90		0.078	0.089	0.101	0.111	0.130		
	45.0	180		0.116	0.131	0.143	0.153	0.170		
	45.0	270		0.107	0.121	0.134	0.137	0.132		

(2.1), approximating and substituting for L_θ

$$\Delta\Phi_h = 2\,E \sum_{0}^{\frac{\pi}{2}} R_\theta \cos\theta.\sin\theta\,\Delta\theta\,\Delta A$$

This equation was used to calculate an approximate value of total radiance reflected into the hemisphere. An approximate value of hemispherical reflectance R_h was then calculated from

$$R_h = \frac{\Delta\Phi_h}{E\,d\,A}$$

As the calculated value of $\cos\theta.\sin\theta\,\Delta\theta$ is constant for constant θ, for any azimuth from the nadir, and assuming $R_{90^o} = R_{270^o}$, then an average value of R_θ was calculated from

$$R_\theta = \frac{R_{0^o} + 2R_{90^o} + R_{180^o}}{4}$$

When only R_{0^o} and R_{180^o} are given R_θ was taken as the mean. The value of $\Delta\theta$ used in the calculations was 0.2619^c (15^o in radians). Reflectance values for 80^o and 85^o nadir view were not used in the calculations. Grass, dry, was averaged over four azimuths.

Calculated values of R_h are compared with R_{0^o} in Table 2.2. Neglecting atmospheric effects, R_{0^o} would be the calculated value determined from the radiance measured at a nadir sensor.

In all cases nadir reflectance was less than hemispherical reflectance although only by a few percentage points and manmade surfaces were not markedly different in this regard when compared to natural surfaces. One exception must however be metal surfaces (no data available) which from observation reflects with a large specular component. This type of surface can be considered to have a similar reflectance distribution as water in Table 2.1. A number of further comments can be made.

(1) For infrared radiation, which has a longer wavelength than visible light, there is an increased probability that the specular component would increase, based on Rayleigh's criterion of roughness.

(2) At high solar elevations, for either visible or infrared radiation, surfaces more closely approximate diffuse reflectors, again based on Rayleigh's criterion.

TABLE 2.2: Comparison of hemispherical and nadir reflectance of urban type surfaces.

Surface Description	Hemispherical Reflectance (%)	Reflectance at Nadir (%)
1. Asphalt, oily with dust film blown onto oil.	8.6	6.1
2. "White" concrete, aged.	30.0	26.6
3. Calm water, infinite optical depth.	4.2	2.2
4. Grass, lush green, closely mowed thick lawn.	11.7	10.0
5. Macadam, washed off and scrubbed.	12.9	11.3
6. Dirt, hard packed, yellowish.	28.0	24.3
7. Mixed green forest, deciduous (oak) and evergreen (pine).	5.4	3.6
8. Grass, dry meadow, dense, midsummer.	12.0	9.6

(3) In urban areas, manmade surfaces are generally weathered which should reduce the specular reflectance component.

(4) For a constant solar elevation angle, the measured radiance and calculated reflectance should also be constant for the same surface material, albeit a lower value than hemispherical reflectance, within the imaged scene.

It was considered that no evidence substantially precluded adopting the Lambertian assumption for all urban materials. In view of the advantages to be gained by the use of this simplified assumption, reflectance of a particular surface material was assumed to depend only on wavelength. In all further sections the symbol R is used for hemispherical reflectance, unless specifically stated.

2.4 Reflectance from Heterogeneous Surfaces

If within one pixel, response is derived from a number of surface types, then each can be considered as the source of separate wave fronts. Suppose two such waves are incident upon a detector so that the total radiant flux, Φ, received due to these waves can be measured. If the time variation of amplitude of one wave can be expressed by the function $f_1(t)$ and the other by $f_2(t)$ then the radiant flux delivered by the first will be,

$$\Phi_1(t) = K f_1^2(t)$$

and by the second,

$$\Phi_2(t) = K f_2^2(t)$$

where K is a constant.
Because of the quantitised nature of electromagnetic energy it is necessary to average the radiant flux delivered over a period of time, T, to obtain a reading, therefore

$$\Phi_1 = K \overline{f_1^2(t)}$$

and

$$\Phi_2 = K \overline{f_2^2(t)}$$

The combined wave at the detector has an amplitude of $f_1(t) + f_2(t)$. Therefore the radiant flux delivered is,

$$\Phi(t) = K\,[f_1(t) + f_2(t)]^2$$

$$= K\,[f_1^2(t) + 2f_1(t)\,f_2(t) + f_2^2(t)]\ ,$$

with an average flux after time T of

$$\Phi \text{ total} = K\,[\,\overline{f_1^2(t) + 2f_1(t)\,f_2(t) + f_2^2(t)}\,]$$

which is equivalent to

$$\Phi \text{ total} = K\,\overline{f_1^2(t)} + K\overline{f_2^2(t)} + 2K\overline{f_1(t)f_2(t)}$$

therefore

$$\Phi \text{ total} = \Phi_1 + \Phi_2 + 2K\overline{f_1(t)f_2(t)} \qquad (2.3)$$

When the two waves are incoherent their amplitudes are randomly related and the final term in equation (2.3) is insignificantly small or zero, so that the radiant flux of the combined wave equals the sum of the radiant flux of the two separate waves. Alternatively if the two waves are coherent and a regular or systematic relationship exists between the two amplitudes, then the received radiant flux may be greater or less than the sum of the individual radiant flux separately incident on the detector. As diffusely reflected solar radiation is incoherent then the measured radiant flux received is the simple sum. This conclusion also applies to more than two surface types contributing to the total radiant flux received at the detector.

Consider a heterogeneous surface of area A in the instantaneous field of view of a sensor, comprised of n different surfaces each of area ΔA_i, i = 1 to n, with an incident radiant flux density, E. For diffusely reflecting surfaces the radiant flux reflected into an infinitely large hemisphere by each individual surface is given by

$$\Delta\Phi_i = \pi\, L_i\, \Delta\, A_i$$

The total radiant flux is therefore

$$\Phi = \pi \sum_{i=1}^{i=n} L_i\, \Delta\, A_i$$

and the diffuse reflectance of the total surface, R, is given by

$$R = \frac{\pi\, \Sigma\, L_i\, \Delta\, A_i}{EA}$$

But $L_i = \frac{R_i E}{\pi}$ from (2.2)

therefore

$$R = \frac{\Sigma R_i \Delta A_i}{A} \qquad (2.4)$$

In addition, L, the total radiance from the surface is

$$L = \frac{E \Sigma R_i \Delta A_i}{\pi A} \qquad (2.5)$$

Therefore the contribution of individual surface reflectance to the hemispherical reflectance and the contribution of measured radiance to the total radiance at a detector is directly proportional to the percentage of area covered by each surface.

In a recent field study by MILTON (1978) the mixed reflectance of two surface types was analysed. The first surface, A, had an incomplete grass cover and significant amounts of bare surface were visible when viewed vertically. The second surface, B, had a complete vegetative cover, but the data was collected as the growing season was ending and a significant proportion of the vegetation was brown and dry.

When measured reflectance was plotted against sampled percentage green cover, there was an initial decline followed by a levelling off for surface A, as percentage grass increased. This was apparent for all Landsat equivalent spectral ranges, suggesting a non-linear response with respect to percentage area. Surface B however, gave results that closely approximated those expected from a simple, linear response. MILTON (1978) explained the non-linear effect as being partly due to the initial masking of a highly reflective background by a low reflecting green vegetation and partly to the effect of increasing amounts of low reflecting shadow falling on the exposed surface. With surface B the influence of shadows was largely removed due to the almost complete cover, and so the total reflectance was due to only two surface types. It is suggested that had MILTON measured the percentage cover of a third surface, 'shadow', results would have conformed to a simple proportions model. Laboratory results given by TUCKER (1977), for a mixture of dead and live grass, also support the theory.

2.5 Atmospheric Effects

2.5.1 Introduction

Up to this point the effects of the atmosphere between scene and detector have not been considered. With the advent of satellite remote

sensing systems the atmospheric contribution to the measured ground response has come under increasing examination (SABATINI et al, 1970, TURNER et al, 1971, TURNER et al, 1972, SHARMA, 1972, TURNER, 1975, DANA, 1975, HERZOG et al, 1975, KAWATA et al, 1978).

It is only recently that computer programs have been developed that adequately deal with realistic vertical distribution profiles for molecular and particulate constituents in the framework of CHANDRASEKHAR's (1960) radiative transfer theory. One such physical model has been developed by TURNER et al (1972). Quantifying the contribution of background surface reflectance (albedo) in increasing path radiance was an important result of this study. This arises from the scattering of radiation emanating from areas outside the instantaneous field of view of the sensor and was first noted by NALEPKA et al (1971) for the case of multispectral remote sensing by aircraft over agricultural areas. With the diversity of surfaces found in urban and closely adjoining areas the effect of a spatially varying albedo and its varying contribution to path radiance are clearly of importance if subtle changes are to be observed.

The apparent radiance of ground materials, as measured by a remote sensor, differs from the intrinsic surface radiance because of the presence of the intervening atmosphere. The surface characteristics and the atmospheric conditions will determine, in general, the radiance measured at a given solar position.

The atmosphere affects the visible and near infrared remote sensor signals in a number of ways. Firstly it modifies the spectral and spatial distribution of the radiation incident on the surfaces being observed. In addition the radiation reflected by the surface is attenuated by the atmosphere as it passes through it to the sensor and an extraneous component of scattered radiation called path radiance is added to the transmitted component.

The characteristics of the surface determine the specular or diffuse nature of the surface reflectance and the amount of incident global irradiance that is reflected by the surface. The surface characteristics also modify the amount of sky irradiance which augments the incident global irradiance and the amount of upwelling path radiance adding to the intrinsic surface radiance.

The two dimensional model of figure 2.4 schematically shows the parameters and interactions which ultimately determine the radiance reaching a satellite sensor, where $E_{\theta\lambda}$ = spectral solar irradiance at the

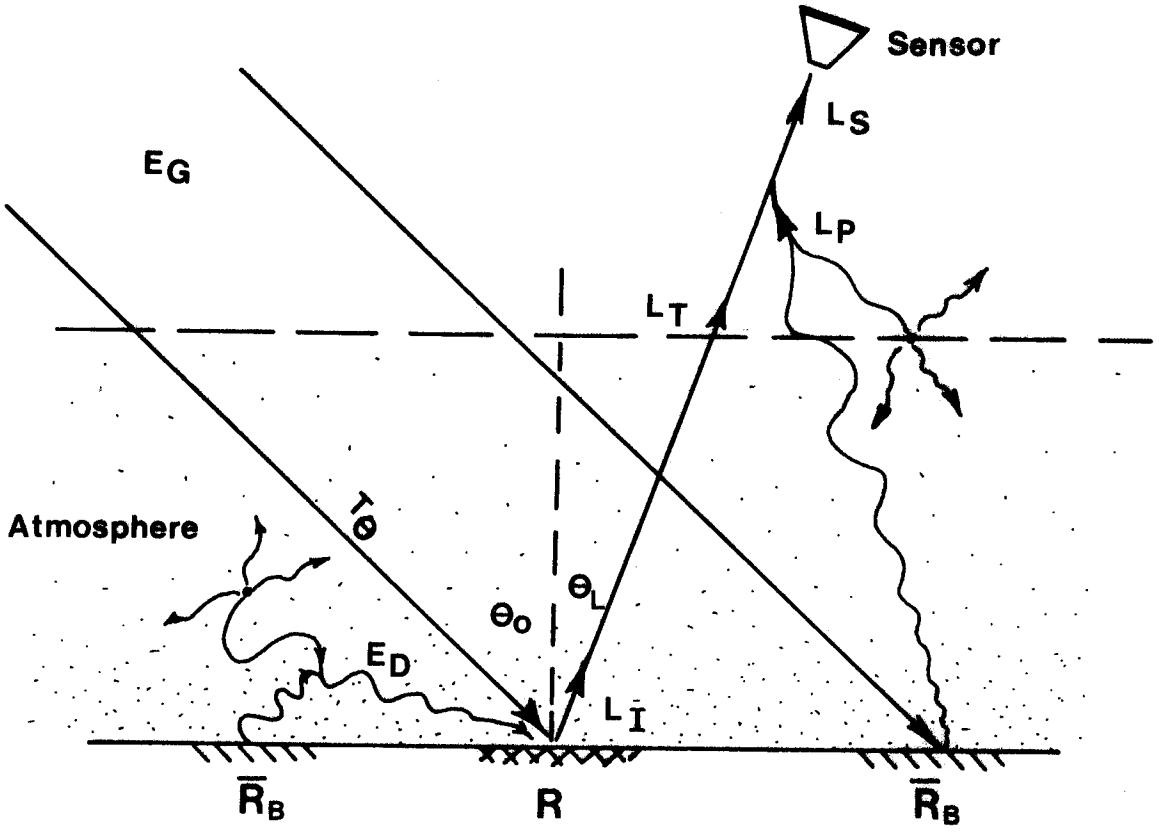


FIGURE 2.4: Parameters and interactions which determine the radiance reaching a satellite sensor.

top of the atmosphere, $Wm^{-2}\ \mu m^{-1}$.

E_o = solar irradiance at the top of the atmosphere, Wm^{-2}.
E_{D_λ} = spectral diffuse sky irradiance, $Wm^{-2}\ \mu m^{-1}$.
E_D = diffuse sky irradiance, Wm^{-2}.
E_{G_λ} = spectral global irradiance incident on the surface, $Wm^{-2}\ \mu m^{-1}$.
E_G = global irradiance incident on the surface, Wm^{-2}.
τ = normal atmospheric optical thickness.
T_θ = atmospheric transmittance at an angle θ to the zenith.
θ_o = solar zenith angle.
θ_v = nadir view angle of the satellite sensor (or scan angle).
μ = cos θ.
R = average target reflectance.
R_i = reflectance of an individual component of the target.
$\overline{R}_B$ = average background reflectance.
L_S = total radiance at the sensor, $Wm^{-2}\ sr^{-1}$.
L_S^* = total radiance received by the sensor, $Wm^{-2}\ sr^{-1}$.
L_I = intrinsic radiance of the target, $Wm^{-2}\ sr^{-1}$.
L_P = path radiance resulting from multiple scattering, $Wm^{-2}\ sr^{-1}$.
K = radiance per bit of sensor count rate.
C = sensor count rate.

In a given spectral interval the solar irradiance reaching the earth's surface is

$$E_G = \int_{\lambda_1}^{\lambda_2} (E_{\theta\lambda}\ T_{\theta_o} \cos\theta_o + E_{D_\lambda})\, d\lambda\ [Wm^{-2}] \qquad (2.6)$$

A fraction of the irradiance is reflected by the surface in the direction of the satellite sensor. For a Lambertian surface the ratio of the radiation reflected in the direction of the satellite sensor to the total radiation reflected into the whole upper hemisphere, above the reflecting surface, is given by $(\pi)^{-1}$ from (2.2). Thus the reflected radiance at the sensor from the surface alone is

$$L_T = \frac{1}{\pi} \int_{\lambda_1}^{\lambda_2} R\ T_{\theta_v}\ (E_{o_\lambda}\ T_{\theta_o} \cos\theta_o + E_{D_\lambda})\ d\lambda\ [Wm^{-2}sr^{-1}] \qquad (2.7)$$

and the total radiance at the sensor can be expressed as

$$L_S = (L_T + L_P) = CK\ [Wm^{-2}\ sr^{-1}] \qquad (2.8)$$

(strictly speaking $L_s^* = CK$, but this is considered further in section 2.6).

Of the variables determining the radiance received by the sensor, only R is associated with the target itself. Hence in order to quantitatively evaluate reflectance, the atmospheric transmittance, the path radiance, the absolute amount of radiance at the satellite and the global irradiance incident on the surface must be known.

In practice the integration indicated in the equations is replaced by summation of finite differences, and the wavelength dependences of the parameters in the equations are replaced by mean values averaged over the spectral interval of the particular sensor. Therefore from (2.6) and (2.7)

$$E_G = (E_{o\lambda}\, T_{\theta_o} \cos\theta_o + E_{D\lambda})\, \Delta\lambda \quad [Wm^{-2}]$$

and

$$L_T = \frac{R\, T_{\theta_v}\, (E_{o\lambda}\, T_{\theta_o} \cos\theta_o + E_{D\lambda})\, \Delta\lambda}{\pi} \quad [Wm^{-2}sr^{-1}]$$

which reduce to

$$E_G = E_o\, T_{\theta_o} \cos\theta_o + E_D \tag{2.9}$$

and

$$L_T = \frac{R\, T_{\theta_v}\, (E_o\, T_{\theta_o} \cos\theta_o + E_D)}{\pi} \tag{2.10}$$

when E_o and E_D are calculated over the narrow band wavelength interval.

2.5.2 Transmittance

The atmospheric transmittance T_θ, can be given as

$$T_\theta = e^{-\tau \sec\theta}$$

where τ = normal atmospheric optical thickness and θ can represent θ_o or θ_v. The normal optical thickness equals the sum of all the attenuating constituents. The optical thickness is made up of components due to molecular (Rayleigh) scattering by the permanent gases, τ_m, aerosol (Mie) scattering due to particulates, τ_p, and selective absorption, τ_a, all of which are wavelength dependent, hence

$$\tau(\lambda) = \tau_m + \tau_p + \tau_a$$

where $\tau_a = \tau_{H2O} + \tau_{O2} + \tau_{O3} + \tau_{CO2}$

Molecular scattering attenuation has a $^{1}/\lambda^{4}$ dependence. Aerosol scattering is also wavelength dependent however the relationship is not constant and depends on the size and vertical distribution of particulates. ANGSTROM (1961) suggests a $^{1}/\lambda^{1.3}$ dependence and this is supported by ROBINSON (1966) for continental regions. Generally aerosol scattering occurs within the first few kilometres of the earth's surface. Absorption attenuation by atmospheric constituents such as water, oxygen, ozone and carbon dioxide is highly wavelength dependent. Except for a narrow oxygen absorption band at 0.76 μm the spectral range 0.5 to 1.0 μm is dominated by water and ozone absorption attenuation.

2.5.3 Path Radiance

The detector above the atmosphere measures the upward radiance in the direction of its line of sight. Since the nadir scan angle of the Landsat satellite is small only the upward normal radiance is usually considered. As a further consequence of this the angle between the solar azimuth angle and the nadir scan azimuth angle need not be considered.

When the sensor "sees" a target against a background surface the following three types of radiant energy reach the detector -

(i) radiant energy reflected by the target and then directly transmitted by the atmosphere,

(ii) radiant energy scattered diffusely by the atmosphere, after having interacted with the background, and

(iii) radiant energy scattered diffusely by the atmosphere.

It is the radiant energy (i) which carries the direct information on the target, and the combined effect of (ii) and (iii) is the path radiance.

The details of radiative transfer theory, developed over the past 70 years, are much too involved to be presented here. Usually the major difficulties of solving the equation of radiative transfer involves essentially two functions, the single scattering albedo, which simply indicates the amount of scattering that takes place, and equals unity if no absorption attentuation occurs, and the single scattering phase function which denotes the fraction of radiation which is scattered from its initial forward direction to some other direction. For Rayleigh scattering the phase function is a simple dipole-like distribution, but for scattering by aerosol particles the energy is usually distributed heavily in the forward

direction, that is, it is highly anisotropic. If there is very little haze present the scattering approximates Rayleigh scattering and 50% of the total scattered radiation is in the forward direction. However for heavy haze 95% of total scattered radiation is in the forward direction (TURNER et al 1972, 898-899).

At a wavelength of 0.55 μm a haze free atmosphere scatters approximately 10% of incident radiation, but for a reduced visual range the scattered radiation is a greater proportion of incident radiation. Thus forward scattering is dominated by aerosol scattering, and back scattering is mainly due to Rayleigh scattering.

TURNER et al (1972) have derived a number of correction algorithms for determining path radiance in various atmospheres, and for varying solar elevation and background reflectance. For a nadir view angle these are of the form

$$\frac{L_p}{E_o} = F(\mu_o,\tau_o) + G(\mu_o,\tau_o)\ H(\tau_o,\bar{R}_B) \qquad (2.12)$$

where

$$F(\mu_o,\tau_o) = \frac{1}{4\pi[\mu_o+(1-\eta)\tau_o]} \left\{ (1-\eta)(\tau_o-1)\left[P(\mu_o)+P(-\mu_o)\right] + \mu_o P(-\mu_o) + (1-\eta)\left[P(\mu_o)+P(-\mu_o)\right] e^{-\tau_o} - \mu_o P(-\mu_o)\ e^{-\tau_o} \right\}$$

$$G(\mu_o,\tau_o) = \frac{\mu_o^{\,2}}{2\pi[\mu_o+(1-\eta)\tau_o]} \left\{ 1 + 4(1-\eta) - \left[1+4(1-\eta)(\tau_o+1)\right]e^{-\tau_o} \right\}$$

$$H(\bar{R}_B,\tau_o) = \frac{\bar{R}_B}{1+2(1-\eta)(1-\bar{R}_B)\tau_o}$$

$$\tau_o = \tau_m + \tau_p$$

$$\eta = \frac{0.5\ \tau_m + 0.95\ \tau_p}{\tau_o}$$

$\bar{R}_B$ = average background reflectance.
μ_o = $\cos \theta_o$.
$P(\mu_o)$ = single scattering phase function at θ_o.
$P(-\mu_o)$ = single scattering phase function at $(180^o - \theta_o)$.

These equations do not account for the absorption effects of water and ozone. These effects are approximately accounted for by substituting $E_o e^{-(\tau_{03} + \tau_{H20})}$ for E_o, since the scattering and absorbing processes can be treated as independent events (TURNER et al, 1972, 896).

The single scattering phase function, $P(\mu_o)$, is the combined effect of scattering by aerosol particles and Rayleigh scattering. Letting $P_p(\mu_o)$ and $P_m(\mu_o)$ be the phase functions for aerosol and Rayleigh scattering respectively and assuming Rayleigh scattering occurs first then $P(\mu_o)$ can be approximately given by

$$P(\mu_o) = \frac{P_m(\mu_o)(1-e^{-\tau_m}) + P_p(\mu_o)(e^{-\tau_m} - e^{-\tau_o})}{e^{-\tau_o}}$$

where

$$\int_o^{4\pi} P_m(\mu_o) \frac{d\omega}{4\pi} = \int_o^{4\pi} P_p(\mu_o) \frac{d\omega}{4\pi} = 1$$

The value of $P_m(\mu_o)$ can be determined from

$$P_m(\mu_o) = \frac{3}{4}(1 + \cos^2\theta_o)$$

derived first by Rayleigh and $P_p(\mu_o)$ can be determined from figure 2.5 (after TURNER et al, 1971) for Deirmendjian's continental type aerosol. As the aerosol scattering phase function changes only a small amount with wavelength, the function for $\lambda = 0.7$ μm was used for all wavelengths.

The average background reflectance, $\bar{R}_B$, is usually obtained by collecting ground truth information for the region considered and weighting the reflectance according to their respective areas (TURNER et al 1972). TURNER (1975) considers that a 3km square grid surrounding a central target pixel will contribute 90% of the path radiance due to background reflectance.

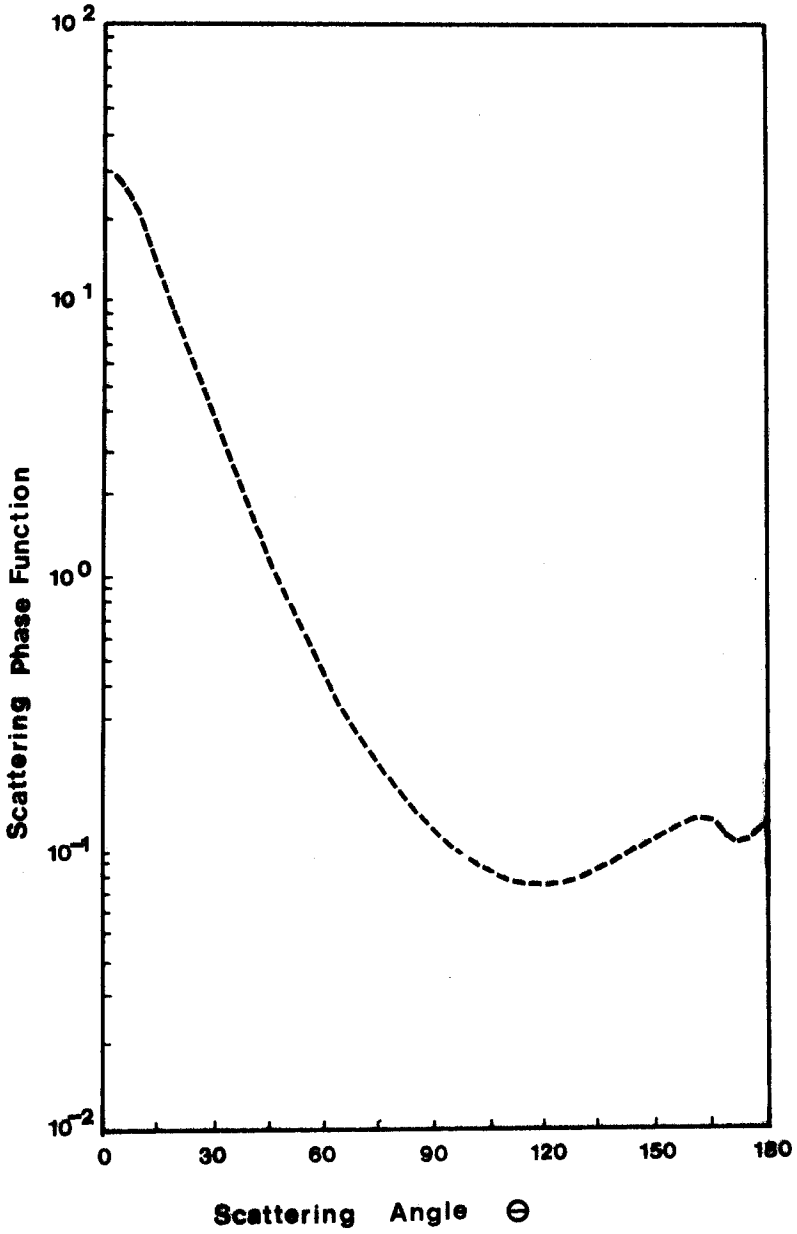


FIGURE 2.5: Single Scattering Phase Function for Deirmendjian's Continental Type Aerosol and Wavelength 0.7 micrometers (after TURNER et al, 1971).

Strictly speaking the contribution to path radiance of a pixel of reflectance R, a distance r from the target, is reduced as r increases. Radiation which originates outside the field of view is attenuated as it propogates to a scattering point on the nadir line of the sensor. It is scattered toward the sensor in accordance with the characteristics of the single scattering phase function and is then further attenuated by the atmosphere along the path to the sensor. The radiance which results from an integration over all pixels and all scattering points along the path, target to sensor, is the path radiance due to background reflectance. Such a calculation has been carried out by TURNER (1975) for simulated data and KAWATA et al (1978) for actual data.

2.5.4 Global Irradiance

The attenuated solar irradiance reaching the earth's surface is augmented by downwelling sky irradiance. This results from similar scattering effects that give rise to path radiance. The total downwelling normalized irradiance can be written as (TURNER et al 1972)

$$\frac{E_G}{E_0} = I(\mu_o,\tau_o)\,[1+J(\tau_o)H(\bar{R}_B,\tau_o)]$$

where

$$I(\mu_o,\tau_o) = \frac{\mu_o^2}{\mu_o+(1-\eta)\tau_o}$$

$$J(\tau_o) = 2(1-\eta)\tau_o$$

and η and H are as for path radiance. When haze increases or visual range decreases the relative contribution of the direct solar irradiance is reduced. This reduced contribution also occurs as the background reflectance increases and the solar zenith angle increases.

As for path radiance, these equations do not account for the absorption effects of water and ozone. Thus the contribution of downwelling sky irradiance to global irradiance can be given by

$$E_D = E_0 e^{-(\tau_{H2O}+\tau_{O3})} I(\mu_o,\tau_o)\,[1+J(\tau_o)H(\bar{R}_B,\tau_o)] - E_o T_{\theta_o} \cos\theta_o \quad (2.13)$$

2.5.5 Target Reflectance

From (2.9) and (2.10)

$$L_T = \frac{RT_{\theta_V}(E_o\ T_{\theta_o}\cos\theta_o + E_D)}{\pi}$$

$$= \frac{R\ T_{\theta_V}\ E_G}{\pi}$$

and LT + LP = CK from (2.8) therefore

$$R = \frac{\pi\ (CK-L_p)}{E_G\ T_{\theta_V}}$$

But

$$R = \frac{\Sigma\ R_i\ \Delta A_i}{A} \quad \text{from (2.4)}$$

therefore

$$\frac{\Sigma R_i\ \Delta A_i}{A} = \frac{\pi\ (CK-L_p)}{E_G\ T_{\theta_V}} \qquad (2.14)$$

From Appendix 1 this equation was solved for each band giving

$$\left(\frac{\Sigma\ R_i\ \Delta A_i}{A}\right)_4 = 0.634\ C_4\ (1-0.00105\ \overline{R}_{B_4})-6.5-0.236\overline{R}_{B_4}$$

$$\left(\frac{\Sigma\ R_i\ A_i}{A}\right)_5 = 0.578\ C_5\ (1-0.00072\ \overline{R}_{B_5})-4.4-0.195\overline{R}_{B_5}$$

$$\left(\frac{\Sigma\ R_i\ A_i}{A}\right)_6 = 0.604\ C_6\ (1-0.00050\ \overline{R}_{B_6})-2.6-0.151\overline{R}_{B_6}$$

$$\left(\frac{\Sigma\ R_i\ A_i}{A}\right)_7 = 0.769\ C_7\ (1-0.00026\ \overline{R}_{B_7})-1.2-0.109\overline{R}_{B_7} \qquad (2.15)$$

Where R_i and $\overline{R}_B$ are given as percentages.

Differentiating average target reflectance, R, with respect to average background reflectance, $\overline{R_B}$, and assuming a high value for C of 100, it can be seen that the accuracy of determination of $\overline{R}_B$ is substantially less than that required for R. In the worst case of band 4 an error in $\overline{R_{B4}}$ of ± 10% would only cause an error of ± 3% in R_4. The effect in the other bands would be smaller, for example, a similar error of ±10% in $\overline{R_{B7}}$ would cause an error of ±1.3% in R_7. In most urban scenes C is considerably less than 100 and so the difference in the accuracy requirements of background and target reflectance are further accentuated.

2.6 Sensor Point Spread Function

An individual Landsat pixel is a radiometric measurement arising from a two dimensional, spatially extended region of the field of view. The contribution from each portion of the region is proportional to the product of the point spread function (psf) and the radiance, and the psf is continuously shifted relative to the scene by the scanning process using both mirror and spacecraft motion (DYE, 1975). The continuous video signal prior to sampling is the result of a convolution of the psf with the scene, and after sampling is the result of a discrete convolution.

Convolution can be described (JENNISON, 1961, 6) as *"the operation whereby a structure under observation is smeared or spread out by the response or resolution of an instrument or mathematical operation."* In the case of Lansat imagery the "structure" is the radiance profile of the ground scene, and the "instrument" the onboard multispectral scanner.

The formula for the convolution of an object and a two dimensional point spread function is

$$K(x,y) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x-z,\ y-w)\,g(z,w)\,dz\,dw$$

where K(x,y) is the ordinate at point x, y after convolution and f and g are the two functions to be convolved. Thus the convolution can extend beyond the bounds of one function taken singly due to the product of the overlap in the first function. Each elemental area dz dw of the function g must be multiplied by the relevant value of f, i.e. f(x-z, y-w), and the total effects then added. By the law of conservation of energy the influence of each element is proportional to its volume, where 'volume'

is the product of radiance and area. That is, the volume of each element after spreading must equal the volume of the element before spreading. To obtain the correct relative influence of each element the evaluated ordinate of the psf is multiplied by the volume of the element (TRINDER, 1970, 40-41).

The Landsat psf is determined by the blur circle and detector size in the across-scan direction, and the effect of a bandpass filter is added in the along-scan direction (DYE, 1975). The two dimensional psf's are shown in figure 2.6 after DYE and have as their product the joint two dimensional psf of the Landsat multispectral scanner.

If the radiance distribution of the scene, including atmospheric effects, can be described by $L_s\ (x-z,\ y-w)$, where x is the along scan direction and y the across scan direction, then the total radiance received by the sensor is

$$L^*_{s(x,y)} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} L_s(x-z,y-w)\ G(z,w)\ dz\ dw$$

where $G(z,w)$ is the two dimensional Landsat psf and L^*_s is the total radiance received by the sensor.

Dividing a Landsat pixel into a 3 x 3 segment array, and limiting the convolution effects to a 3 x 3 pixel array, as an approximation to the psf, L^*_s can be given by (from Appendix 1)

$$\begin{aligned} L^*_s = {} & 0.03\ L_{s(x-1,y-1)} + 0.07\ L_{s(x,y-1)} + 0.02\ L_{s(x+1,y-1)} \\ & + 0.18\ L_{s(x-1,y)} + 0.45\ L_{s(x,y)} + 0.13\ L_{s(x+1,y)} \\ & + 0.03\ L_{s(x-1,y+1)} + 0.07\ L_{s(x,y+1)} + 0.02\ L_{s(x+1,y+1)} \end{aligned} \quad (2.16)$$

The family of curves represented by (2.16) were solved for L_s of the central pixel in terms of L^*_s values, assuming adjoining pixel values were equal to their nearest neighbour in the array and pixel values outside this modified 4 x 4 array were zero, to give (from Appendix 1)

$$\begin{aligned} L_{s(x,y)} = {} & 0.09\ L^*_{s(x-1,y-1)} - 0.35\ L^*_{s(x,y-1)} + 0.07\ L^*_{s(x+1,y-1)} \\ & - 0.78\ L^*_{s(x-1,y)} + 2.76\ L^*_{s(x,y)} - 0.60\ L^*_{s(x+1,y)} \\ & + 0.09\ L^*_{s(x-1,y-1)} - 0.35\ L^*_{s(x,y+1)} + 0.07\ L^*_{s(x+1,y+1)} \end{aligned} \quad (2.17)$$

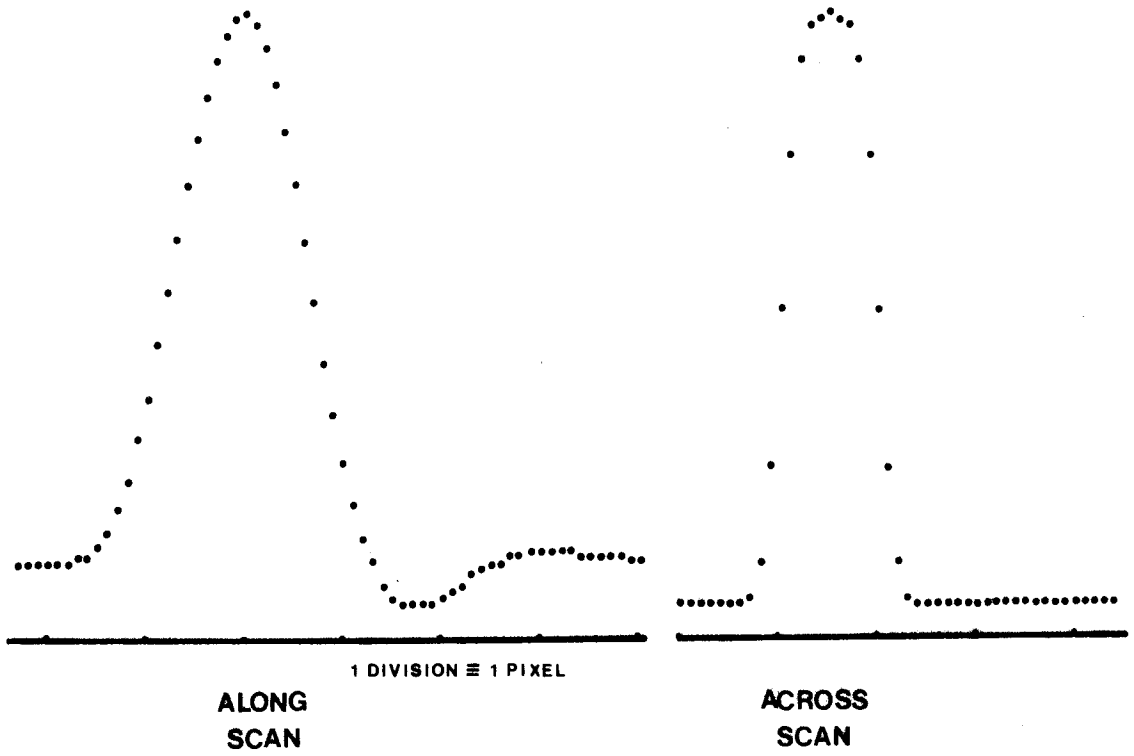


FIGURE 2.6: Landsat point spread function (after DYE, 1975).

Because L_s^* is directly porportional to C, the sensor count rate, then the value of C that would have been measured had convolution not occurred, Ca, can be given by

$$C_{a(x,y)} = 0.09\ C_{(x-1,y-1)} + \ldots. + 2.76\ C_{(x,y)} + \ldots 0.07\ C_{(x+1,y+1)} \tag{2.18}$$

Alternatively $C_{(x,y)}$ can be given as a function of reflectance values by substituting

$$\frac{1}{K}\left(\frac{E_G\ T_{\theta_V}}{\pi}\ \frac{\Sigma\ R_i\ \Delta A_i}{A} + L_p\right)$$

for L_s in (2.16) and C for L_s^*. The atmospheric effects are essentially the same for adjoining pixels and so

$$C_{(x,y)} = \frac{E_G\ T_{\theta_V}}{K}\left[0.03\left(\frac{\Sigma\ R_i\ \Delta A_i}{A}\right)_{(x-1,y-1)} + \ldots. + 0.02\left(\frac{\Sigma\ R_i\ \Delta A_i}{A}\right)_{(x+1,y+1)}\right] \tag{2.19}$$

All equations (2.15) to (2.19) are wavelength dependent.

2.7 Summary

The various factors affecting reflected radiance from mixed surfaces can now be brought together. The count values of equations (2.15) do not allow for the effects of psf convolution given in equations (2.18). This can be achieved by substituting band dependent values of C_{a_4}, C_{a_5}, C_{a_6}, and C_{a_7} for C_4, C_5, C_6 and C_7 in (2.15).

As C_a for each band can be calculated from the measured count values at the sensor and assuming a satisfactory value for $\overline{R}_B$ can be determined, then four linear equations of the form

$$(\Sigma\ R_i\ A_i)_n = B_m \qquad m = 4 \text{ to } 7 \tag{2.20}$$

can be derived for each pixel, where A_i is the percentage area of a particular surface and B is the solution of the right hand side of (2.15) with C_a substituted for C, and the whole multiplied by 100.

From (2.19) and letting $W\ (\Sigma R_i\ A_i)$ equal the weighted sum of the reflectance from the 3 x 3 pixel array and

$$D = \left(C_{(x,y)} - \frac{L_p}{K} \right) \frac{100\ K\pi}{E_G\ T\theta_v}$$

which can be solved using (2.15), then a further set of four linear equations can be given

$$W\ (\Sigma\ R_i\ A_i)_n = D_m \quad m = 4 \text{ to } 7 \tag{2.21}$$

The essential difference between (2.20) and (2.21) is that for (2.20) the actual reflectance of the central pixel is related to the deconvolved sensor count of that pixel, whereas for (2.21) the measured sensor count is related to the weighted reflectance of the central and surrounding pixels. It is from these pixels that the convolved response originated.

Both (2.20) and (2.21) are equivalent in form to

$$y = a + b_1x_1 + b_2x_2 + ... + b_kx_k$$

The theoretical background for the solution of this type of equation for the coefficients a to b_k when x and y have many sampled values is considered in Chapter 3.

3. MULTIPLE REGRESSION ANALYSIS FOR THE PREDICTION OF REFLECTANCE AND COVER PERCENTAGES

3.1 Introduction

Both of the models developed in Chapter 2 (2.20) and (2.21) are of a basic linear form relating a function of band response to cover percentage variables. Linear equations of this form are suitable for analysis by multiple linear regression techniques. This is a general statistical technique through which the relationship between a dependent or criterion variable and a set of independent or predictor variables can be analysed. The regression strategy involves the selection of equation coefficients in such a way that the sum of the squared residuals is smaller than any possible alternative values.

This method has been used in a number of other remote sensing applications where the energy response from a surface has been related to various attributes of that surface. Studies relating various water quality parameters to sensor response or between - band response ratios have been reported (WHITLOCK, 1977, JOHNSON, 1975 and 1978, ROGERS et al, 1975, RITCHIE et al, 1974, SCARPACE et al, 1979, KHORRAM , 1979). In general these studies treated sensor response or response ratios as the independent variables to predict water quality as the dependent variable. Other researchers have used regression analysis in vegetation and soil studies to predict, for example, tree stress,green biomass, soil salinity or soil moisture content (RICHARDSON et al, 1977, TUCKER et al, 1977, THOMAS et al, 1977, TUCKER, 1978, PETERSON et al, 1979, LILLESAND et al, 1979).

In the present study multiple linear regression techniques were used to estimate reflectance values for various urban surface types and to predict the percentage cover of individual surfaces within a picture element. This chapter briefly reviews the method, the data requirements and the various measures of precision.

3.2 Multiple Linear Regression

3.2.1 Basic Theory

Consider the following linear equation

$$y = a + \sum_{i=1}^{i=k} b_i x_i$$

For a multiple regression analysis the task is to estimate the a and b_i coefficients. Because of observation error and a limited number of sample pairs drawn from a larger population, a precise estimate of the coefficients is usually not possible. A stochastic disturbance must be introduced and hence the problem of estimating the coefficients is a statistical and not merely a mathematical problem. In this case the multiple regression equation is represented by

$$y' = a + \sum_{i=1}^{i=k} b_i x_i$$

where y' is the estimated value of the dependent variable. The difference between the actual and the estimated value of y for each case is termed the residual, i.e. the error in prediction, which may be interpreted as resulting from the effect of unspecified independent variables and/or a totally random element. The y intercept is given by a, and b_i, i=1 to k, are the partial regression coefficients.

If there are n sample pairs the sum of the squares of the residuals (SS) is given by

$$SS = \sum_{i=1}^{n} (y_i - y_i')^2$$

The a and b_i coefficients are selected in such a way that the sum of the squared residuals is minimized. The least squares criterion implies that any other values for a and b_i would yield a larger $\Sigma(y_i - y_i')^2$. Selection of the optimum a and b_i coefficients using the least squares criterion also implies that the correlation between the actual y values and the y' estimated values is maximized while the correlation between the independent variables and the residual values $(y_i - y_i')$ is reduced to zero.

The actual calculation of a and b_i requires a set of simultaneous equations derived by differentiating $\Sigma(y_i - y_i')^2$ and equating the partial derivatives to zero. This can be shown in matrix form as

$B = (X'X)^{-1}X'Y$, from the normal equations $X'XB = X'Y$

where

$$B = \begin{bmatrix} a \\ b_1 \\ b_2 \\ \vdots \\ b_k \end{bmatrix} \qquad Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

and

$$X = \begin{bmatrix} 1 & x_{11} & x_{21} & \cdot & \cdot & x_{k1} \\ 1 & x_{12} & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & x_{1n} & x_{2n} & \cdot & \cdot & x_{kn} \end{bmatrix}$$

Thus values for the a and b_i coefficients can be estimated using matrix transpose, multiplication, and inversion procedures.

The partial regression coefficients, b_i, have specific meanings, for example b_1 represents the expected change in y with a change of one unit in x_1 when x_2 through x_k are held constant. Equally important is that the combined effects are additive. For a one unit change on each of two variables, x_1 and x_2, the expected change in y would be $(b_1 + b_2)$.

3.2.2 Data Assumptions

To obtain the best unbiased estimates of the coefficients at least six assumptions should be satisfied.

(i) Errorless Observations.

Each value of x_i and y is observed without error. This is a very difficult assumption to satisfy. However the bias introduced into the model is small provided that the residual error or variance is small. For most purpose it is frequently assumed that measurement error is much less significant than errors resulting from incorrect equation specification, so the former is generally ignored (MALINVAND, 1966, 362).

(ii) Variable Linearity.

The relationship between y and each of the independent variables x_i are linear in the parameters of the specific functional form chosen. If it is assumed that the linearity assumption is not satisfied in a particular instance the input data are generally transformed to yield new data which satisfy this assumption more closely. Ordinary linear regression can then be applied to these transformed data.

(iii) Zero Residual Mean.

The distribution of (y-y') has a mean of zero. This means that positive and negative variations are equally likely to occur around the regression line. It implies that no important explanatory variable has been omitted in the specification of the regression model and that the chosen x's represent the major control on the variability of y (MATHER, 1976, 43). The bias introduced into the model when this assumption is not satisfied is small, provided that the residual variance is small (MALINVAND, 258-260).

(iv) Residual Homocedasticity.

The variance of the residuals is constant and independent of x_i. This condition is called homocedasticity, and assumes that the dispersion of the residual terms around their zero mean is of the same order irrespective of the size of x_i. If the variance is not constant, but independent of x_i the estimates of the regression coefficients are still unbiased. However, if the variance is not constant and is also correlated with x_i then the estimates of the coefficients are seriously biased. For both cases the usual methods of statistical inference are invalid (Ibid, 254-257).

(v) Serially Independent Residuals

The residual values are serially independent. If this assumption is not satisfied autocorrelation is said to be present and either the form of the model is incorrect, one or more relevant variables have been omitted or the residuals are truly autocorrelated. The presence of autocorrelation in one dimensional data, such as the values corresponding to a time series, may be tested by the Durbin - Watson d statistic. Testing for autocorrelation in the case of two

dimensional spatial data is more difficult. The effects of autocorrelated residuals on the least squares estimate are the same whether the pattern is one or two dimensional. The estimates of b_i will not be efficient, that is will not have minimum variance, and the variance will be biased, resulting in confidence intervals that are too narrow and giving incorrect values for significance tests. Even if the residuals are autocorrelated, the estimates of b_i will still be unbiased and consistent, although it is likely that their numerical values for a given sample will be incorrect (MATHER, 84).

(vi) Uncorrelated Independent Variables.

The independent variables x_i are linearly independent of each other. If this assumption is not satisfied and the independent variables are correlated, the individual coefficients for each variable are not identifiable. The imprecision in the estimates of the regression coefficients is generally revealed by the occurrence of high standard errors. This assumption becomes particularly critical when it is desired to obtain explanation in contrast to prediction from the independent variables. If the independent variables are generally correlated then *"... this makes it difficult if not impossible to untangle the variance accounted for in the dependent variable and to attribute portions of it to individual independent variables."* (KERLINGER et al, 1973, 296). It should be noted further that if one of the independent variables is a perfect linear function of one or more of the other variables the normal equations become unsolvable, because the determinant, $|X'X|$, will equal zero. However see section 3.3.5 when the regression is forced through the origin.

To make inference from a sample to a population then a further assumption is also necessary.

(vii) Normally Distributed Variables.

For a fixed x_i model the dependent variable, should have a normal distribution. A random x_i model requires that all variables are normally distributed which implies that the residuals should be normally distributed. This assumption may however

be frequently relaxed because the statistical tests for significance are relatively insensitive to departures from normality, particularly so when the sample is large, of the order of thirty samples or more.

3.2.3 Discussion of data assumptions with respect to Landsat response data and sampled cover data

Assumption (i) - Errorless Observations.

Landsat count values can have an error of the order of ±1 count and can be assumed to be randomly distributed. Cover data sampled over a pixel will also be in error and the size of the error will depend on the number of samples taken in each pixel. The percentage standard error of the sampled proportion of a particular cover type can be given by $\left(\frac{pq}{n} \right)^{\frac{1}{2}}$ where p is the percentage of the cover type being sampled, q is the percentage remaining and not of that cover type and n is the number of samples taken. It can be seen from this formula that the maximum error will occur when p = q = 50%. For 20 samples this would give a standard error of ±11% when p = 50%, and ±7% when p = 10%. These sampling errors will be randomly distributed about the true value of a particular percentage when a large number of pixels are sampled.

For the purpose of this research it was assumed that both Landsat and cover related measurement errors were less significant than errors resulting from incorrect equation specification.

Assumption (ii) - Variable Linearity.

From equations (2.20) and (2.21) it has been shown that a function of sensor response will be linearly related to the percentage of surface area of a particular reflectance. Thus (ii) will be satisfied when sensor response is taken as the dependent variable and percentage area as the independent variable. From a physical viewpoint A_i is the independent variable and C (or a function of C) is the dependent variable, however in many cases A_i must be treated as the dependent variable.

Consider the following family of linear equations.

$$y_i = a_j + \sum_{i=1}^{k} b_{ij} x_i$$

If j also ranges from 1 to k then the equations can be considered as k equations in k unknowns and these can be solved simultaneously to give equations of the form

$$x_i = a_i^* + \sum_{j=1}^{k} b_{ji}^* y_i$$

which in terms of multiple linear regression is

$$x_i' = a_i^* + \sum_{j=1}^{k} b_{ji}^* y_i$$

where x_i' is the estimated value of x_i. Therefore for this form of the equation the independent variables will also be linearly dependent on the dependent variable. Landsat provides only four independent measures, the response in bands 4, 5, 6 and 7 so it would seem that for exact equation specification only four cover types should be sampled. However the condition that all cover percentages must sum to 100% provides the means for the estimation of a fifth cover type and further the inclusion of additional cover variables when these represent only a small percentage of the total cover, can be considered as unspecified independent variables adding to the residual when a particular cover variable is treated as the dependent variable. Other procedures for inclusion of non-linear variables are considered in Chapter 4.

Assumption (iii) - Zero Residual Mean.

As the equations relating Landsat response and cover percentages were theoretically derived in Chapter 2, it is known that no important explanatory variables have been omitted. In addition because both Landsat response and cover data are measured values with an assumed random error distribution, it can be concluded that the distribution of the residuals should have a mean of zero. This will apply whether Landsat response (or reflectance) or a cover percentage is the dependent variable.

Assumption (iv) - Residual Homocedasticity

The assumption of homocedasticity is not satisfied when one cover variable is treated as the independent variable. When an individual cover variable comprises a sampled 100% then the residuals at that percentage will only be due to small sampling and measurement errors about

the regression. However when the same variable is sampled at 0%, the dependent response variable can take on any values determined by the respective proportions of all the other cover variables and so the residuals will be of an equivalent order. The residuals will reduce as the percentage of the particular cover variable increases and so will be linearly dependent on the cover percentage, (see figure 3.1). As additional cover variables are included in the regression equation these will contribute explanation to the total regression by reducing the initially observed residual distribution until ultimately with all cover variables in the equation the assumption of homocedasticity would be valid.

Assumption (v) - Serially Independent Residuals.

The serial independence of Landsat MSS data has been examined by a number of writers (TUBBS et al, 1978 and CRAIG, 1979) who show that significant autocorrelation exists. It follows that as the dependent variable of a regression equation is linearly related to the residuals, that the residuals from using Landsat response data as the dependent variable would also be autocorrelated. CRAIG concludes that Landsat response data sampled at a spacing closer than every 10th pixel would yield dependent and therefore redundant samples. Any denser sampling would not increase the degrees of freedom of the sample. He suggest therefore, that in general, a sample size of N x M will only yield approximately

$$\left(\frac{N \times M}{50}\right) + 2$$

degrees of freedom, since this is the number of pixels spaced at 10 steps which can be located in the sample. These conclusions are partially supported by STEINER et al (1975) who indicate that for typical imaged scenes there is an overwhelming probability that two neighbouring points will have the same response.

A great deal of the autocorrelation however, will be due to the spreading caused by the point spread function of the imaging system. DYE (1975) uses a total of seven along and five across scan pixels to contain this effect for Landsat data. It is considered that the use of the approximately deconvolved response data of equation (2.18) would considerably reduce autocorrelation effects. For the reverse case of equation (2.19) where measured response is a function of the weighted

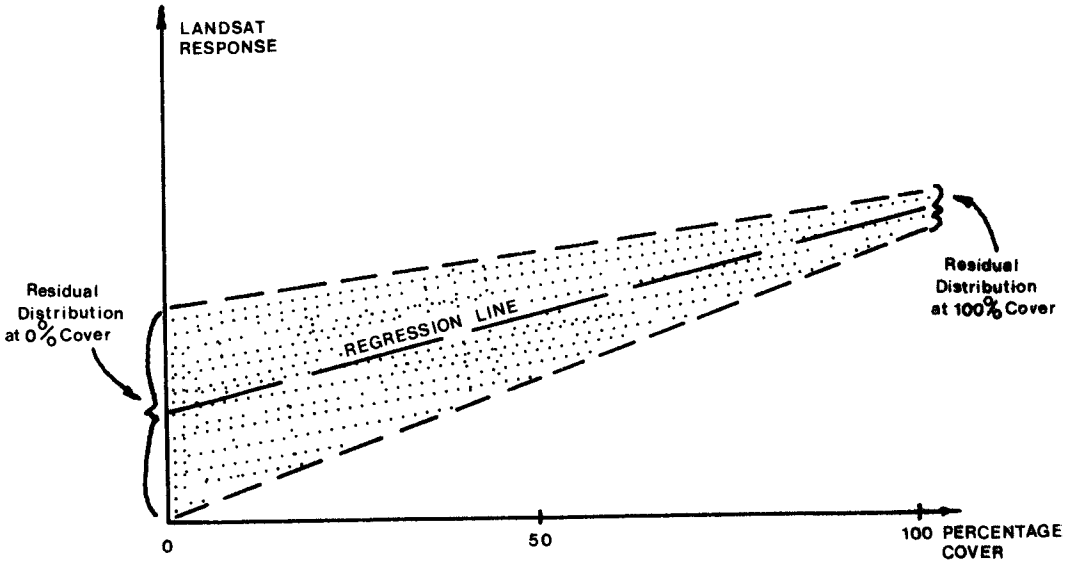


FIGURE 3.1: Residual distribution about a computed regression line for one cover variable entering the regression equation.

reflectance of the central and surrounding pixels, autocorrelation effects will also be reduced. Consider the simple situation of a low reflectance surface S_1, surrounded by a high reflectance surface, S_2, having measured response counts of C_1 and C_2 and actual response counts of Ca_1 and Ca_2 respectively. For surface S_1 the measured C_1 value would be higher than generally predicted resulting in a negative residual whilst for the immediately surrounding surface, S_2, the opposite would apply and a positive residual would result. Thus the residuals would be serially dependent. If however the weighted percentage cover was related to the measured response count, C_1, the percentage of surface, S_1, related to it as a data pair would be reduced, effectively reducing the residual, and the effects of autocorrelation (see figure 3.2). This concept can be considered to apply irrespective of the number of dependent variables.

If cover variables are treated as the dependent variables then it might be expected, due to the regular nature of urban residential development, that autocorrelation would exist. However this would be purely spatial autocorrelation, appearing as groups of positive and negative residuals around a trend surface patterning, unlike the response data which is due to the measurement system. If the surface data is sampled independently then there would be no reason to suspect that the data was serially dependent.

Assumption (vi) - Uncorrelated Independent Variables.

The linear independence of the independent variables can be tested using bivariate correlation analysis. Earlier unpublished research by the writer using sampled cover data indicated very little correlation between individual cover variables, the highest correlation found being of the order of 0.4, however this is not the case for Landsat MSS data. Typically, high correlations exist between bands 4 and 5, and bands 6 and 7, the two visible and the two infrared bands respectively.

Although prediction was the primary aim of this study it is nevertheless undesirable to eliminate independent variables, which may have some explanatory value, solely because their significance is reduced by intervariable correlation. Various transformations can be applied that reduce this effect while still retaining the information content of the data. One such simple transformation would create two new difference and sum variables e.g. (Band 4 - Band 5) and (Band 4 + Band 5).

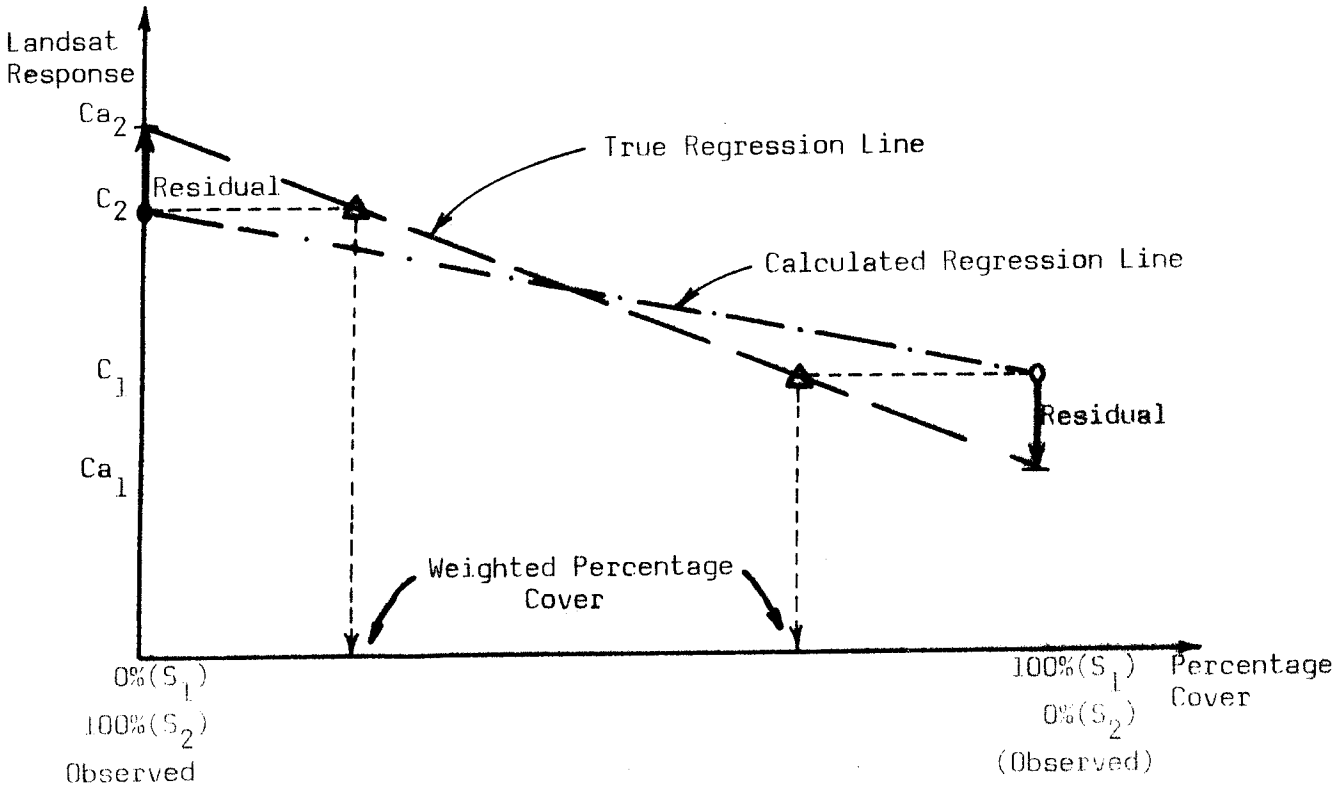


FIGURE 3.2: Schematic diagram of the relationship between the calculated regression line using uncorrected observed percentage cover and observed Landsat response, and the regression line using weighted percentage cover and observed Landsat response, for two adjoining cover surfaces, S_1 and S_2. The serially dependent residuals will be reduced and approach the true regression line when the weighted percentage cover is used.

More complex procedures such as principal component analysis could also be used to create an identical number, n, of linearly independent new variables (orthogonal in n dimensional space) onto which the original data are transformed. This procedure has been used by LODWICK (1979) for comparing multi-temporal Landsat data, and DONKER et al (1976) and AUSTIN et al (1978) for rural land cover classification.

Assumption (vii) - Normally Distributed Variables.

The underlying statistical nature of Landsat MSS data is not known. The data is generally accepted to be approximately normally distributed however MCCLOY (1978) *"is not aware of any definitive work that has shown that this hypothesis is valid"*. The normality or otherwise of percentage cover data is also unknown. It might be expected that cover types generally occupying a small percentage of the residental surface would have a skewed distribution due to the infrequent occurrence of higher percentage values. However for both response and cover variables a large number of independent causes would be contributing small effects to the residuals and so it could be argued from the central limit theorem that the assumption of normality was justified. Nevertheless givena sufficiently large sample the effects of non-normality are reduced.

It can be concluded that response data and percentage cover data are satisfactory data for analysis by multiple regression analysis, provided

(i) the use of correlated independent variables are restricted to transformed linearly independent combinations or an approximation thereof when explanation in contrast to prediction is required.

(ii) the data samples are sufficiently large.

3.3 Measures of Precision

3.3.1 Multiple Correlation coefficient, r and r^2

The total variation or sums of squares in y can be partitioned into two independent components, one that is explained by the regression and another that is unexplained

$$SS_y = SS_{reg} + SS_{res}$$

$$\Sigma(y - \bar{y})^2 = \Sigma(y' - \bar{y})^2 + \Sigma(y - y')^2$$

where SS_{reg} and SS_{res} are the sum of squares of the regression and the residuals respectively.

The ratio r^2 is defined as

$$r^2 = \frac{SS_{reg}}{SS_y}$$

therefore as r^2 approaches 1.0, then SS_{res} must approach zero. Thus r^2 is a statistical measure of the adequacy of the least squares fitting process. It measures the proportion of total variation about the mean value of y that is explained by the regression. Often it is expressed as a percentage after multiplication by 100, and referred to as percentage explanation or simply as explanation.

The sample r^2 given above is a biased estimator of the population r^2 if n is small. An unbiased estimator is given by $\bar{r}^2$ the adjusted r square,

$$\bar{r}^2 = \frac{n-1}{(n-k-1)} (1 - r^2)$$

where

k = the number of partial regression coefficients.

The square root of r^2, r is defined as the correlation coefficient or multiple correlation coefficient for a multiple regression equation and is also a statistical parameter used as a measure of adequacy. A correlation coefficient of 0.9 means that 81 percent of the total variation about the mean value is explained by the regression equation. Care must be taken when using r as a measure of equation precision. When the number of estimated coefficients in the regression equation equals the number of experimental observations, an exact solution for the number of coefficients is obtained. In this case r will equal one yet the coefficients will be in error unless the experimental data is without error. Therefore r is not a good measure of precision when the number of estimated coefficients approaches the number of experimental observations.

The significance of the correlation, r, between the two variables can be tested by the use of the student's t distribution, using the following formula,

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

where n = the number of data pairs studied and where the degrees of freedom are (n-2). The null hypothesis postulated is that there is no correlation between the variables. For regression analysis the two variables can be considered as the observed and calculated value of the dependent variable, and r as the multiple correlation coefficient.

3.3.2 Standard Error of the regression and coefficients

The variance of the regression, S , may be expressed as

$$S = \frac{\Sigma(y-y')^2}{n-(k+1)}$$

where k+1 equals the number of estimated coefficients (k partial regression coefficients plus the y intercept coefficient). The variance is a measure of the deviation between the predicted values and the measured values. The square root of the variance is the standard error, S.E., and is a second measure of the least squares estimation process. The smaller the value, the more precise the fitted equation.

The variance, and thus the standard error of the coefficients is determined from the variance covariance matrix which can be given by

$$\begin{bmatrix} \text{var}(a) & \text{cov}(a,b_1) & \text{cov}(a,b_2) \ldots\ldots & \text{cov}(a,b_k) \\ \text{cov}(b_1,a) & \text{var}(b_1) & \text{cov}(b_1,b_2) \ldots\ldots & \text{cov}(b_1,b_k) \\ \text{cov}(b_2,a) & \text{cov}(b_2,b_1) & \text{var}(b_2) \quad \ldots\ldots & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \text{cov}(b_k,a) & \text{cov}(b_k,b_{k-1}) & \ldots\ldots\ldots\ldots\ldots & \text{var}(b_k) \end{bmatrix}$$

and $SE_{bi} = (\text{var } b_i)^{\frac{1}{2}}$

3.3.3 F-Test

The F-test is a third method of evaluating the adequacy of the least squares estimation process. The F-ratio is defined as

$$F = \frac{SS_{reg}\ (n-k-1)}{SS_{res}\ (k)}$$

or

$$F = \frac{r^2 (n-k-1)}{(1-r^2)\ k}$$

The F ratio is distributed approximately as the F distribution with degrees of freedom k and (n-k-1). This allows a critical F value to be determined based on the degrees of freedom as well as the desired confidence limits. The F-test requires that the F value calculated from the regression must be greater than the critical F value if the regression process is to be judged significant within the confidence limits. Obtaining a statistically significant regression does not necessarily mean that the resulting equation will be useful for predictive purposes (DRAPER et al, 1966, 64). A calculated F value four times the critical F value is suggested if the regression equation is to be regarded as a satisfactory predictor.

3.3.4 F-test for the coefficients

The most common strategy used in testing the coefficients involves a decomposition of the explained sum of squares due to regression, the SS_{reg}, into components attributable to each independent variable in the equation (NIE et al, 1975, 336). The increment in r^2 (or in the explained sum of squares)due to the addition of a given variable is taken as the component or variation attributable to that variable. If the variables entered into the equation are considered to have a non-causal relationship, each variable is treated as if it had been added to the regression equation in a separate step after all variables had been included. The F ratio is then given by

$$F = \frac{\text{incremental SS due to } X_i}{SS_{res}/(n-k-1)}$$

or

$$F = \frac{\text{incremental } r^2\ (n-k-1)}{(1 - r^2)}$$

The degrees of freedom for each F ratio are 1 and (n-k-1), and given the confidence limits, a critical F value can be determined.

3.3.5 Effect of forcing the regression through the origin

When it is known a priori that the constant term in the regression equation is zero, the regression may be forced through the origin where

$$y = 0, x_1 = 0, x_2 = 0 \ldots . x_k = 0$$

Here in the normal equations

$$X = \begin{bmatrix} x_{11} & \cdots & x_{k1} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ x_{1n} & \cdots & x_{kn} \end{bmatrix} \quad \text{and} \quad | X^1X |$$

will not equal zero if $x_1 + x_2 + \ldots + x_k = G$, where G is a non-zero constant.

In these circumstances there are two alternative ways of computing r^2. The first method computes r^2 unadjusted for the mean of the dependent variable, and was proposed by THEIL (1971, 176), therefore

$$r_u^2 = \frac{\Sigma(y')^2}{\Sigma(y)^2}$$

$$= \frac{SS_{reg_u}}{SS_{y_u}}$$

and $\bar{r}_u^2 = \frac{n-1}{(n-k-1)} (1-r_u^2)$

where

SS_{reg_u} = sum of squares of the regression unadjusted for the mean of y,

and

SS_{y_u} = total sum of squares unadjusted for the mean of y.

The second method is the standard approach given previously, that is

$$r^2 = \frac{SS_{reg}}{SS_y}$$

where both SS_{reg} and SS_y are adjusted for the mean.

The choice between r^2 and r_u^2 depends upon whether the measured performance of the regression is desired in terms of the variance of y or the second moment around zero of y. In some cases when the regression is forced through the origin, the computed value of r^2 can be negative.

This means that a better prediction of y is obtained by using the mean, $\bar{y}$, as the estimate, and the assumption that $y = 0$, $x_1 = 0$, $x_2 = 0$, ... , $x_k = 0$ would therefore be suspect.

3.3.6 Selection of the most significant variables

Particularly in the case of predicting percentage cover as the dependent variable with various combinations of response as the independent variables, it is necessary to isolate a subset of available variables that will yield an optimal prediction equation with as few terms as possible. There are several approaches to this problem (NIE et al, 345).

(i) Forward (stepwise) inclusion - independent variables are entered only if they meet certain statistical criteria.

(ii) Backward elimination - independent variables are eliminated one by one from a regression equation that initially includes all variables.

(iii) Stepwise solution - in addition to forward inclusion variables are deleted that no longer meet the pre-established statistical criteria.

(iv) Combinatorial solution - all possible combinations are examined.

3.3.7 Recommended regression procedure and tests

Of the four methods given in 3.3.6 for variable inclusion only the first three were available to the writer in an existing statistical package. Of these, (iii), stepwise solution, was considered the most flexible. Here the variable selected is that which explains the greatest amount of variance unexplained by those variables already in the equation. The selection process is based solely upon the F ratio which each variable would have if it was the only variable added to the regression at that step. After each step the variables in the equation are re-examined and any variable with an F ratio less than a minimum value is considered eligible for removal. A minimum F value of 3.85 was chosen when the degrees of freedom of the residuals was greater than 1000, and 4.00 when they were between 50 and 1000, these being the 5% confidence limits. For the regression as a whole four times the F ratio at the 1% confidence

limits (for the number of degrees of freedom of the particular equation) was chosen as the limiting criterion.

In all further regression calculations, the adjusted r square, $\bar{r}^2$, multiple correlation coefficient, r, standard error, S.E., and F-ratio of the total regression are given. Where the regression is forced through the origin both $\bar{r}^2$ and $\bar{r}_u^2$ are provided. For the coefficients, the standard errors are given.

The stepwise regression procedure and all statistical tests were available on an existing statistical package, SPSS - 6000 version 70, (NIE et al, 1975) and this was used for all multiple linear regression analysis.

4. AN EXAMINATION OF REFLECTANCE RATIOS, TRANSFORMATIONS AND TEXTURAL VARIABLES

4.1 Introduction

Various ratios are routinely used as a first order correction to reduce brightness variations due to atmospheric effects, change in solar zenith angle, shadowing and also variations due to source noise. It is considered that ratioing will result in the cancellation of in-phase fluctuations of the original response variables (MAXWELL, 1976). Particularly in vegetation studies ratios are also used to improve the correlation between vegetation parameters such as biomass and independent response variables. Other transformations in addition to ratios are used to obtain new orthogonal independent variables and to maximize the separation of classes along new coordinate axes. A reduction in the number of variables can also result from these transformations.

Typical ratios and transformations include

(i) Band on band ratios e.g. $\frac{\text{Band 5}}{\text{Band 7}}$

(ii) Band difference on band sum ratios e.g. $\frac{\text{Band 7 - Band 5}}{\text{Band 7 + Band 5}}$

(iii) Individual band on the sum of all band ratios e.g.

$\frac{\text{Band 6}}{\text{Bands (4+5+6+7)}}$, termed normalisation.

(iv) Band difference, band sum transformation.

(v) Kauth-Thomas transformation.

(vi) Principal component transformations.

In this chapter the results of an examination of each ratio and transformation are given. Typical vegetation and soil data were used to examine their characteristics and their inclusion as variables in multiple regression models.

The use of textural variables over extended areas, as an aid to classification, was mentioned in Chapter 1. A simple variance (or standard deviation) measure was examined with respect to its application to the prediction of cover or cover related related variables in residential areas. Details of this examination are also given in this chapter.

4.2 Band on Band Ratios

Chlorophyll absorbs electromagnetic energy most efficiently at wavelengths of 0.4 to 0.5 μm and 0.65 to 0.69 μm. Green vegetation also has a very high reflectance in the near infrared from 0.75 to 1.2 μm. Using these unique absorption and reflectance characteristics MILLER and PEARSON (1971) and TUCKER (1973) have shown that a ratio of the near infrared and chlorophyll absorption bands is well correlated with the amount of green biomass within a scene (MAXWELL, 1976). The second of the two absorption bands and the high reflectance bands are closely approximated by Landsat bands 5 and 7, respectively. The ratio of band 7 to band 5 was therefore considered as a potential predictor of healthy vegetation. MAXWELL (1976) and CARTER (1977) have also suggested the ratio band 5 to band 4 as a good predictor of vegetation because band 4 contains neither of the primary clorophyll absorption bands. CARTER considered that these ratios would give an enhanced output for healthy vegetation and hence better discrimination against urban areas. RICHARDSON et al (1977) used band 5 on band 7, as one of eight vegetation indices in a comparison to determine their ability to distinguish vegetation from a soil background. Using skylab colour imagery, PIECH et al (1978), have used a *blue to green* reflectance ratio to determine lake water quality.

Essentially the band on band ratio is a transformation from rectangular to polar coordinates, therefore

$$B_i = \frac{V_{ij} R_{ij}}{(R_{ij}^2 + 1)^{\frac{1}{2}}}$$

and

$$B_j = \frac{V_{ij}}{(R_{ij}^2 + 1)^{\frac{1}{2}}}$$

where

$$V_{ij} = (B_i^2 + B_j^2)^{\frac{1}{2}}$$

$R_{ij} = \frac{B_i}{B_j}$, and B_i, B_j are the response in bands i and j respectively.

Each of B_i and B_j can be approximated by a n^{th} order polynomial.

$$B_i = a_o + a_1 V_{ij} + a_2 R_{ij} + a_3 V_{ij} R_{ij} + a_4 V_{ij}^2 + \ldots\ldots$$

The four Landsat bands can be represented by a number of different but equivalent functions of V_{ij} and R_{ij}. These are

$$B_4 = \frac{V_{45} R_{45}}{(R_{45}^2 + 1)^{\frac{1}{2}}} = \frac{V_{54}}{(R_{54}^2 + 1)^{\frac{1}{2}}}$$

$$= \frac{V_{46} R_{46}}{(R_{46}^2 + 1)^{\frac{1}{2}}} = \frac{V_{64}}{(R_{64}^2 + 1)^{\frac{1}{2}}}$$

$$= \frac{V_{47} R_{47}}{(R_{47}^2 + 1)^{\frac{1}{2}}} = \frac{V_{74}}{(R_{74}^2 + 1)^{\frac{1}{2}}}$$

$$B_5 = \frac{V_{54} R_{54}}{(R_{54}^2 + 1)^{\frac{1}{2}}} = \frac{V_{45}}{(R_{45}^2 + 1)^{\frac{1}{2}}}$$

$$= \frac{V_{56} R_{56}}{(R_{56}^2 + 1)^{\frac{1}{2}}} = \frac{V_{65}}{(R_{65}^2 + 1)^{\frac{1}{2}}}$$

$$= \frac{V_{57} R_{57}}{(R_{57}^2 + 1)^{\frac{1}{2}}} = \frac{V_{75}}{(R_{75}^2 + 1)^{\frac{1}{2}}}$$

$$B_6 = \frac{V_{67} R_{67}}{(R_{67}^2 + 1)^{\frac{1}{2}}} = \frac{V_{76}}{(R_{76}^2 + 1)^{\frac{1}{2}}}$$

$$= \frac{V_{65} R_{65}}{(R_{65}^2 + 1)^{\frac{1}{2}}} = \frac{V_{56}}{(R_{56}^2 + 1)^{\frac{1}{2}}}$$

$$= \frac{V_{64} R_{64}}{(R_{64}^2 + 1)^{\frac{1}{2}}} = \frac{V_{46}}{(R_{46}^2 + 1)^{\frac{1}{2}}}$$

$$B_7 = \frac{V_{76} R_{76}}{(R_{76}^2 + 1)^{\frac{1}{2}}} = \frac{V_{67}}{(R_{67}^2 + 1)^{\frac{1}{2}}}$$

$$= \frac{V_{75} R_{75}}{(R_{75}^2 + 1)^{\frac{1}{2}}} = \frac{V_{57}}{(R_{57}^2 + 1)^{\frac{1}{2}}}$$

$$= \frac{V_{74} R_{74}}{(R_{74}^2 + 1)^{\frac{1}{2}}} \qquad \frac{V_{47}}{(R_{47}^2 + 1)^{\frac{1}{2}}} \tag{4.1}$$

and each of these can be represented by a polynomial expression.

Which form of the function is used to represent B_i in terms of V_{ij} and R_{ij}, is dependent upon the spectral characteristics of the target being examined. When choosing an appropriate combination for each band, the fourth combination must include i and j values previously used, otherwise the V_{ij}, and R_{ij} values of that function will be redundant i.e. predictable from the other V_{ij} and R_{ij} values. If for the case of predicting percentage vegetated cover the ratios of bands 5 and 4, and bands 7 and 5 are considered statistically significant then the following combinations are appropriate.

$$B_4 = \frac{V_{54}}{(R_{54}^2 + 1)^{\frac{1}{2}}}$$

$$B_5 = \frac{V_{75}}{(R_{75}^2 + 1)^{\frac{1}{2}}} \quad ,$$

for B_6 any combination can be taken for example

$$B_6 = \frac{V_{76} R_{76}}{(R_{76}^2 + 1)^{\frac{1}{2}}}$$

and for B_7

$$B_7 = \frac{V_{75} R_{75}}{(R_{75}^2 + 1)^{\frac{1}{2}}} \quad \text{must be chosen.}$$

These functions, in polynomial form, can then be used in a predictive equation of the form

$$\begin{aligned} A(\text{vegetation}) = {} & K_o + K_1 V_{54} + K_2 R_{54} + K_3 V_{54} R_{54} + K_4 V_{54}^2 + \ldots \\ & + K'_1 V_{75} + K'_2 R_{75} + K'_3 V_{75} R_{75} + K'_4 V_{75}^2 + \ldots \\ & + K''_1 V_{76} + K''_2 R_{76} + K''_3 V_{76} R_{76} + K''_4 V_{76}^2 + \ldots \end{aligned}$$

However for a vegetated surface it was assumed that V_{54} and V_{75} varied only with atmospheric and other effects, and were relatively constant for a particular scene and independent of A (vegetation), whereas R_{54} and R_{75} were considered to vary only with percentage

change in vegetation. Because bands 6 and 7 are highly correlated, the value of R_{76} was assumed to be relatively constant. The term V_{76} also varies with scene brightness effects such as atmosphere and shadowing, and is normally neglected, however as both are infrared bands it was expected that the value of V_{76} would increase with increasing vegetation and therefore should remain in the predictive equation. Given these assumptions the percentage of vegetation within a target area can be approximately given by

$$A(\text{vegetation}) = K^*_o + K^*_1 R_{54} + K^*_2 R_{75} + K^*_3 V_{76} + K^*_4 R^2_{54} + K^*_5 R^2_{75} + K^*_6 V^2_{76} + \ldots\ldots \quad (4.2)$$

which is in the linear form required for multiple linear regression.

From figure 2.1 vegetation has approximate reflectances of 15%, 10%, 30% and 50% in bands 4, 5, 6, and 7 respectively and soil has approximate reflectances of 27%, 32%, 30% and 27% in the same bands. The response in each band was calculated from (2.15) for the known atmospheric conditions (Sydney, Dec 12th 1972 - approximately 15 km visibility) with an assumed background reflectance equal to the target reflectance, and also for no atmospheric conditions, both at a solar zenith angle of 38^o. These response count values are shown in Table 4.1. Figure 4.1 illustrates the relationship between bands 7 and 5, and shows that R_{75} is relatively constant for different atmospheric conditions and varies significantly as vegetation percentage changes. In addition V_{75} is relatively constant for constant atmospheric conditions and is independent of vegetation percentage.

The band 5 and band 4 relationship is shown in figure 4.2 and that of bands 6 and 7 in figure 4.3. On examination of these diagrams it can be seen that some of the assumptions used to derive (4.2) are not valid. While R_{54} is relatively stable with varying atmospheric conditions, V_{54} is not constant for constant atmospheric conditions and appears as an equally good predictor of vegetation percentage. For bands 7 and 6 the assumption of high correlation is not borne out and V_{76} is no more significant a predictor than R_{76} or band 7 alone. Thus for the target surfaces of vegetation and soil only the R_{75} value is a good predictor of percentage vegetation, and may be used alone in a regression analysis with other non-specified terms contributing to the residual. Alternatively

TABLE 4.1: Approximate Landsat response count values for vegetation and soil, for no atmosphere at 38^{o} solar zenith angle and for a 15 km visibility atmosphere at 38^{o} solar zenith angle.

Band	100% Vegetation		100% Soil	
	No Atmosphere	With Atmosphere	No Atmosphere	With Atmosphere
4	38	40	68	65
5	27	28	86	76
6	72	62	72	62
7*	99	73	54	39

* (Doubled band 7 values).

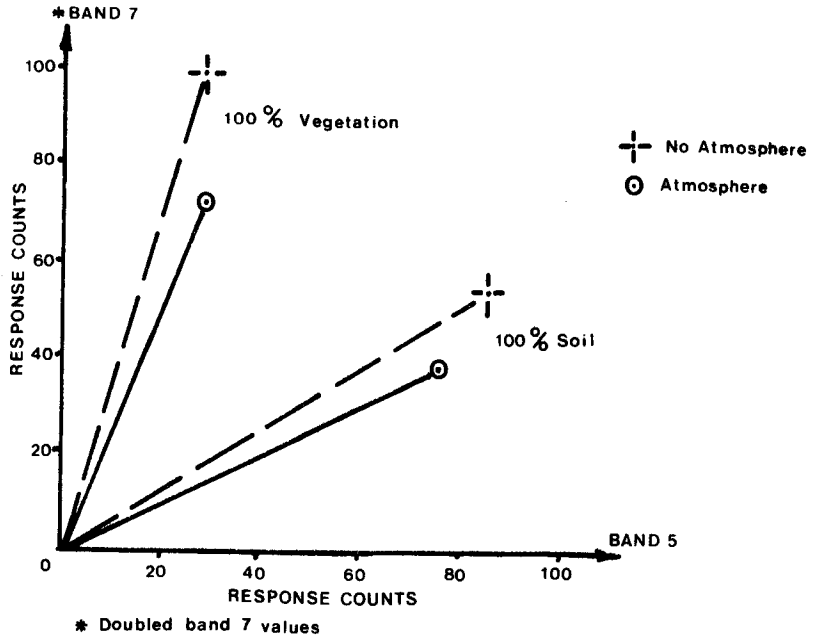


FIGURE 4.1: Band 7 versus band 5 for approximate response counts of vegetation and soil, with and without atmosphere. (Atmosphere approximately 15km visibility).

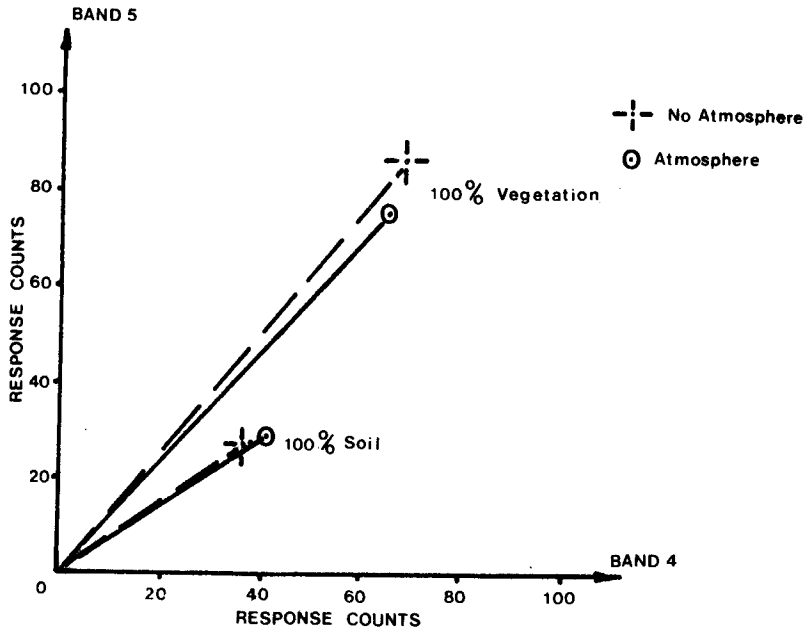


FIGURE 4.2: Band 5 versus band 4 for approximate response counts of vegetation and soil, with and without atmosphere. (Atmosphere approximately 15km visibility).

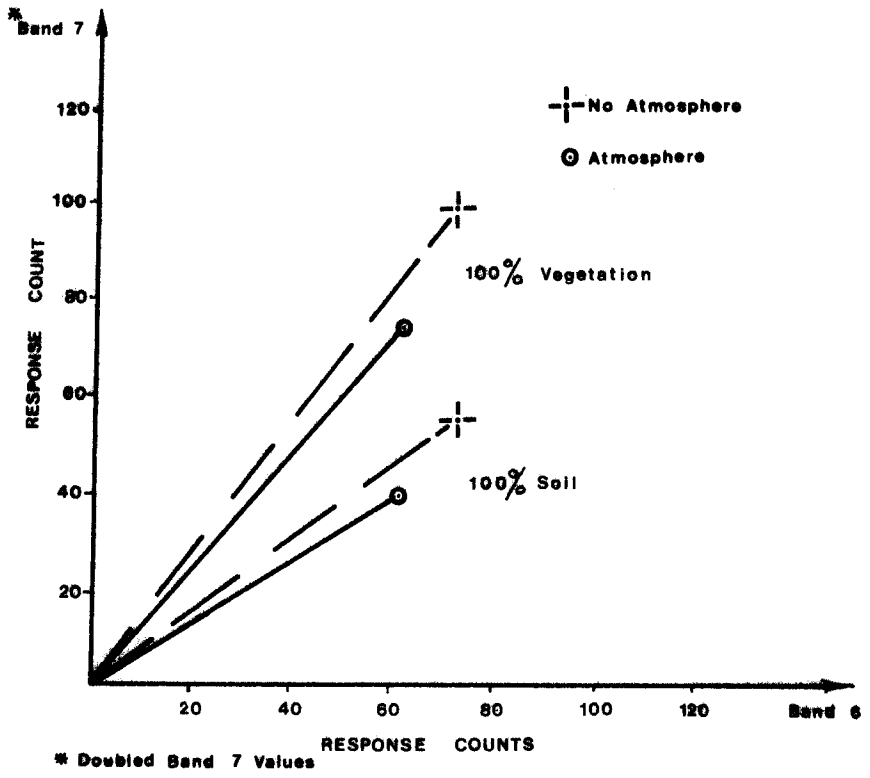


FIGURE 4.3: Band 7 versus band 6 for approximate response counts of vegetation and soil, with and without atmosphere. (Atmosphere approximately 15km visibility).

other R_{ij} and V_{ij} values could be added to a regression analysis and their significance tested at each step. For surfaces other than vegetation and soil the appropriate V_{ij} and R_{ij} terms should be determined from a similar analysis.

The change in the values of V_{ij} and R_{ij} can be generalised further, and are predominantly due to change in solar zenith angle and increasing haze. To a first order of correction, the radiance incident at the Landsat sensor changes directly as the cosine of the solar zenith angle, and affects the response from each band equally. Therefore no change in R_{ij} occurs for otherwise constant atmospheric conditions, while V_{ij} is reduced with increasing solar zenith angle. However for increasing haze the changes in V_{ij} and R_{ij} are not so simple. Consider the response B_i from a surface of reflectance R. If there were no atmosphere the amount of irradiance reflected and received at the sensor would be a simple proportion of $E_o \cos \theta_o$. If atmosphere was then introduced there would be an initial reduction in the amount of radiance received from the same surface, but eventually as path radiance increased due to haze the received radiance would approach that due to clouds which can be thought of as thick haze. Theoretically for the extreme of an infinitely thick atmosphere all irradiance at the top of the atmosphere would be reflected. These effects are shown in figure 4.4 for a vegetated surface, a soil surface, a surface with zero reflectance in both bands, a surface with zero reflectance in one band and 100% reflectance in the other (and vice versa), and for a surface of 100% reflectance in each band. The values shown were calculated for a solar zenith angle of 38^o in bands 5 and 7 of Landsat, although similar diagrams could be determined for other conditions. Note that for a soil surface, R_{75} has little variation over a wide range of atmospheric conditions whereas for vegetation, R_{75} changes considerably as haze increases. For both soil and vegetation cover, V_{75} increases rapidly as haze increases, and their individual V_{75} and R_{75} values approach equality i.e. a complete lack of contrast.

4.3 Band Difference, Band Sum Ratios and Transformations

Both the ratios and the transformations using band difference and band sum are considered together because of their common features. RICHARDSON et al (1977) have used both the ratio and transformation in modified form. The ratios used by RICHARDSON et al were

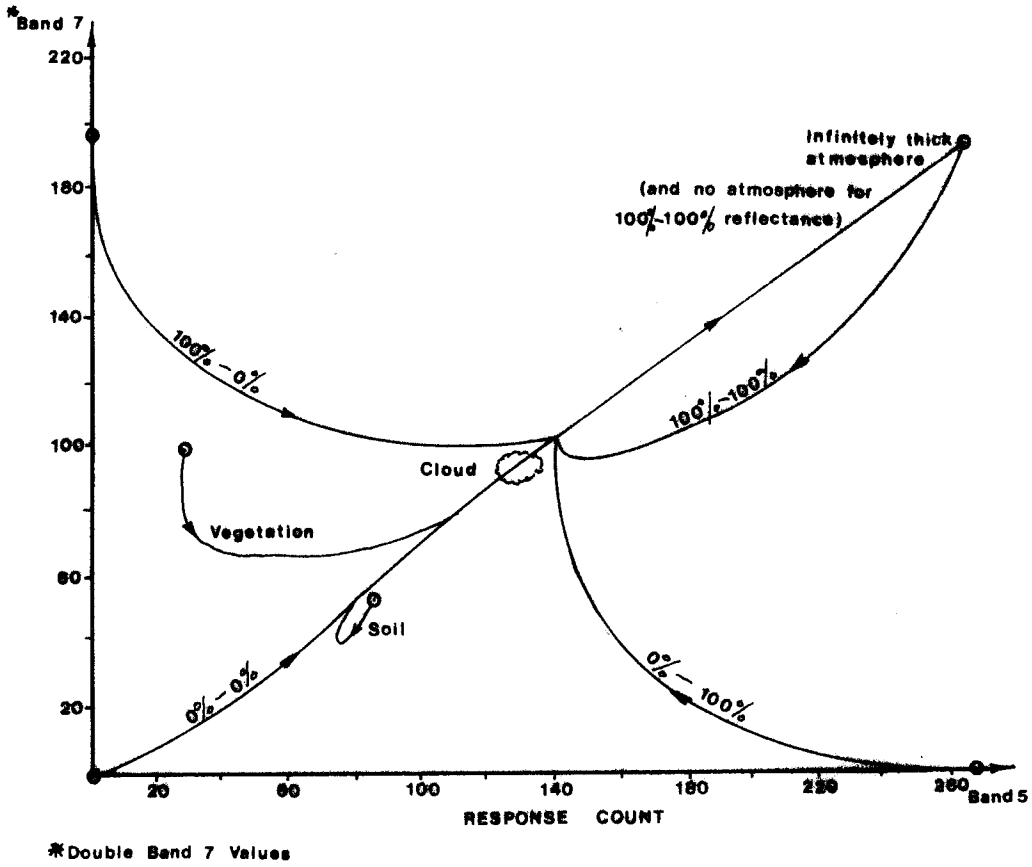


FIGURE 4.4: Schematic paths for change from no atmosphere to an infinitely thick atmosphere for a number of surface reflectance conditions in Landsat bands 5 and 7.

$$\left(\frac{B_7 - B_5}{B_7 + B_5} + 0.5 \right)^{\frac{1}{2}}$$

and

$$\left(\frac{B_6 - B_5}{B_6 + B_5} + 0.5 \right)^{\frac{1}{2}}$$

and were first suggested by ROUSE et al (1973). Each was considered to be a *transformed vegetation index*, the $(B_7 + B_5)$ and $(B_6 + B_5)$ being considered as *normalizing* terms, while the 0.5 term was added to eliminate negative values (DEERING et al, 1975).

RICHARDSON et al (1977) also suggested a vegetation index model using the perpendicular distance of a vegetation candidate signature from a constant R_{57} soil background line, but considered the same measure could be achieved by a modified band difference transformation

$$(2.40B_7 - B_5)$$

where 2.40 was the assumed slope of the soil background line.

It can be shown that the basic band difference, band sum transformation is an affine transformation of the form

$$\begin{bmatrix} B_{(i-j)} \\ B_{(i+j)} \end{bmatrix} = K \begin{bmatrix} \cos\theta, & -\sin\theta \\ \sin\theta, & \cos\theta \end{bmatrix} \begin{bmatrix} B_i \\ B_j \end{bmatrix}$$

where $B_{(i-j)} = (B_i - B_j)$ and $B_{(i+j)} = (B_i + B_j)$. For the most commonly used transformation $\theta = 45^\circ$ and $K = 1.4142$ and for the example quoted by RICHARDSON et al (1977) i.e. $(2.40\ B_7 - B_5)$, $\theta = 22^\circ\ 37'$ and $K = 2.6$. Each band response, B_i, can be represented by equations in terms of $B_{(i-j)}$ and $B_{(i+j)}$, and for the simple case of $\theta = 45^\circ$ and $K = 1.4142$

$$B_i = \tfrac{1}{2} (B_{(i-j)} + B_{(i+j)})$$

Substituting B_i's in this form into a predictive percentage cover equation will not alter the linearity of the model, and percentage area as the dependent variable will still be linearly related to the independent $B_{(i-j)}$, $B_{(i+j)}$ variables. Similar to the case of band on band ratios,

a number of combinations of $B_{(i - j)}$ and $B_{(i + j)}$ can be used to represent the individual B_i's, which combination chosen being dependent on the spectral nature of the cover being predicted, and the avoidance of redundant combinations.

Using the count value data of Table 4.1, for bands 7 and 5, the basic relationship between the transformed values is shown in figure 4.5 with the dotted lines schematically showing the effect of increasing haze and the solid lines that of increasing solar zenith angle. It can be seen that virtually all of the change in vegetation cover can be linearly predicted by $B_{(7 - 5)}$, however unlike the ratio B_{75}, $B_{(7 - 5)}$ varies more under changing haze (or brightness) conditions.

The $(B_i - B_j)$ on $(B_i + B_j)$ ratios are now seen as a ratio of the new variables $B_{(i - j)}$ and $B_{(i + j)}$, and represent one of the polar co-ordinate terms implied by that ratio. They are approximately constant for varying brightness effects and have the added advantage over the simple B_7 on B_5 ratio for vegetation cover prediction, of having an essentially linear response to percentage vegetation change. Again the choice of which ratios to include in a multiple regression analysis for predicting various cover percentages will depend on the spectral characteristics of the cover in question.

4.4 Band on the Sum of All Band Ratios

CARTER (1977) considers this to be a *normalizing* ratio, where each individual intensity value of a pixel is divided by the sum of the intensity values in all four Landsat bands. He considers that if the overall light intensity changes due to shadowing, haze etc, then as a first order correction normalisation should reduce variations in the overall light intensity. The correction assumes that the reduction in the received signal is a similar fraction for all bands which is not always true. After normalisation only three of the normalised bands are necessary to carry all the information (CARTER, 1977, 28-29).

The sum of all bands were considered as a fifth band, S, with wavelength interval 0.5 - 1.1 μm, the concept of polar coordinates being applied in the same way as for ratios under section 4.2. Letting V_{is} and R_{is} be the band on band sum equivalent of V_{ij} and R_{ij} and using bands 5, 6 and 7, then percentage area of a particular surface can be given by

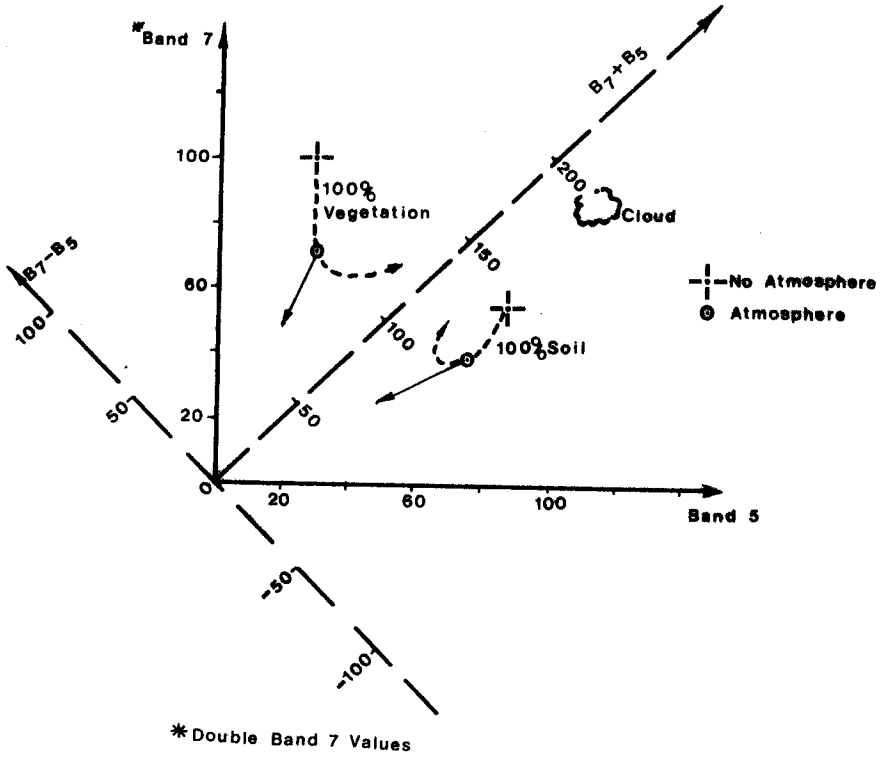


FIGURE 4.5: (Band 7 - Band 5) versus (Band 7 + Band 5) for approximate response of vegetation and soil, with and without atmosphere, and also the effect of increasing solar zenith angle and increasing haze. (Atmosphere approximately 15km visibility).

$$A = K_o + K_1 V_{5S} + K_2 R_{5S} + K_3 V_{5S} R_{5S} + K_4 V_{5S}^2 + \ldots$$
$$+ K'_1 V_{6S} + K'_2 R_{6S} + K'_3 V_{6S} R_{6S} + K'_4 V_{6S}^2 + \ldots$$
$$+ K''_1 V_{7S} + K''_2 R_{7S} + K''_3 V_{7S} R_{7S} + K''_4 V_{7S}^2 + \ldots$$

Essentially the argument for the use of the normalised band values R_{is}, is that they are more likely to be invariant under varying atmospheric conditions than the standard band values. The invariance of R_{is} was tested using the data given in Table 4.1. The four normalised values were computed for each band, for vegetation and soil and for the two atmospheric conditions. These computed values are shown in Table 4.2.

To test the relative stability of the ratios with that of the measured band values, the difference between the various values with and without atmospheric effects and divided by their mean value was calculated. Each was converted to a percentage value which represents the average change due to atmospheric effects. If the change in a ratio value is less than that of its corresponding unnormalised band value then the argument given for their use would be valid for the given set of conditions.

These average changes are shown in Table 4.3. It can be seen that in general the ratio variation is substantially less, however in the case of vegetation in the non-infrared bands the variation is virtually equivalent. This is due to the low reflectance of vegetation in bands 4 and 5, with path radiance contributing a substantial proportion of the observed response. Here the sum of all band responses will have a relatively smaller change in the denominator when compared to B_4 or B_5 in the numerator. For the surface covers examined and the given range of atmospheric conditions, it would be preferable to use the normalised bands 5, 6 and 7 as the non-redundant variable set in a multiple regression analysis, in order to reduce atmospheric effects.

In addition to the ratio variables, R_{is}, in the predictive equation, terms involving the variables, V_{is}, are also included. Which of these variables should remain in the regression analysis will depend on their significance in predicting cover percentage and their invariance to atmospheric change.

Diagrams illustrating the relationships between each band and the new band S are shown in figure 4.6, for 100% vegetation and 100% soil cover for the two given atmospheric conditions. For the purpose of these diagrams B_s, the sum of all bands, has been divided by four and the

TABLE 4.2: Normalised Landsat band response for 100% vegetation and 100% Soil cover under varying atmospheric conditions (38^{o} Solar zenith angle, no atmosphere and 15 km visibility)

Normalised Band	Vegetation		Soil	
	No Atmos.	Atmos.	No Atmos.	Atmos.
4	0.160	0.196	0.243	0.269
5	0.114	0.137	0.307	0.314
6	0.305	0.304	0.257	0.256
7 *	0.420	0.358	0.193	0.161

* Doubled Band 7 values

TABLE 4.3: Average percentage change of normalised band response and band response under two atmospheric conditions for 100% vegetation and 100% soil cover.

Band	Vegetation		Soil	
	% change Normalised Band	% change Band	% change Normalised band	% change band
4	20.2	18.4	10.2	27.8
5	18.4	19.4	4.6	35.6
6	0.4	39.0	0.4	39.0
7*	16.0	52.6	18.0	54.2

* Doubled band 7 values

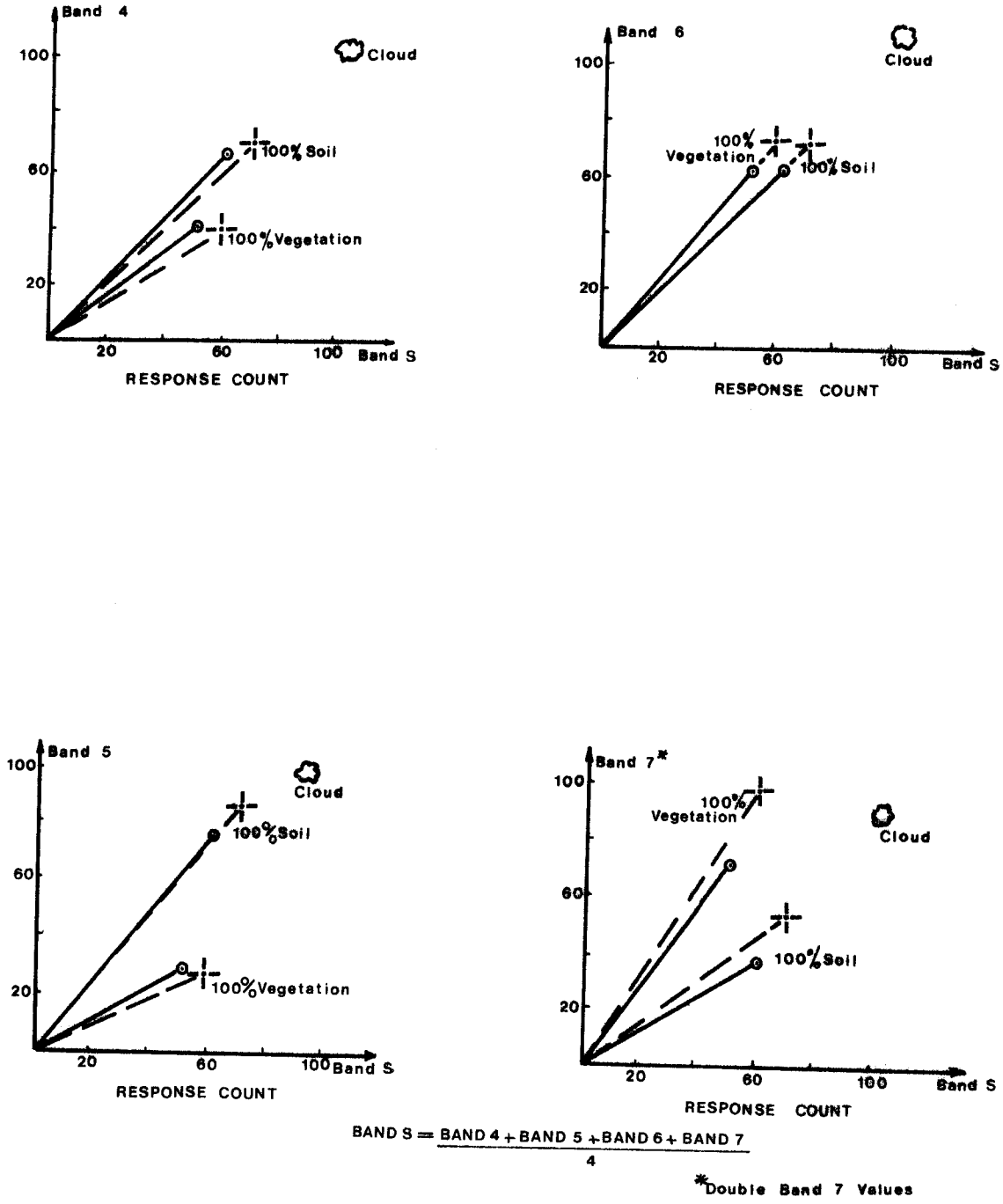


FIGURE 4.6: Landsat bands 4, 5, 6 and 7 versus sum of all bands divided by 4, for approximate response counts of vegetation and soil with and without atmosphere. (Atmosphere approximately 15km visibility).

approximate response of cloud is shown to indicate the direction of change of the ratios due to increasing haze. It can be seen that neither V_{6s} or R_{6s} are particularly useful in predicting cover percentage, and R_{4s} appears only marginally better. The other variables V_{4s}, V_{5s}, R_{5s}, V_{7s} and R_{7s} however all appear reasonably effective variables for predicting vegetation percentage, although the latter variable, R_{7s}, appears least so and in addition is significantly affected by atmospheric effects.

4.5 Kauth-Thomas Transformation

The Kauth-Thomas transformation is based on the assumption that the ratio R_{ij} of soils is relatively invariant to changes in solar zenith angle, haze or type of soil, and when various mixtures of soil and vegetation are plotted in the visible and near infrared Landsat space they will form a roughly triangular shape above a constant soil brightness line. It is considered that location within this triangular shape represents to a degree, vegetative state of development as modified by such factors as soil reflectance, stress of various kinds, mixtures of vegetation and so on (HENDERSON et al, 1975).

Such considerations led KAUTH (1975) to define axes of maximum variation in the Landsat data and to ascribe physical interpretation to these axes. Four orthogonal axes were defined in the four dimensional Landsat space describing *brightness*, *greeness*, *yellowness* and *non-such*, this latter direction containing primarily noise variation. The relationship of the first three axes to Landsat bands 4, 5 and 6 are shown schematically in figure 4.7.

Figure 4.8 after KAUTH et al (1978) expands the illustration to three dimensions termed the *tasseled cap*. Soil samples fall near a line and in a predominantly planar region surrounding that line. Plants start out on bare soil and grow towards the maximum greeness point. Some plant canopies have large amounts of shadow shifting the observation towards the origin. Trees, with a reasonable proportion of shadow are shown in figure 4.8 as the *badge of trees*. Different amounts of shadowing in various canopies creates a region called the *green arm* or *green fold*. When a plant canopy reaches its maximum green development, it will eventually return to the soil by yellowing, withering or removal. The various return paths in the Landsat signal space are the *tassels* of the *cap* (Ibid, 707). For most purposes the response information from

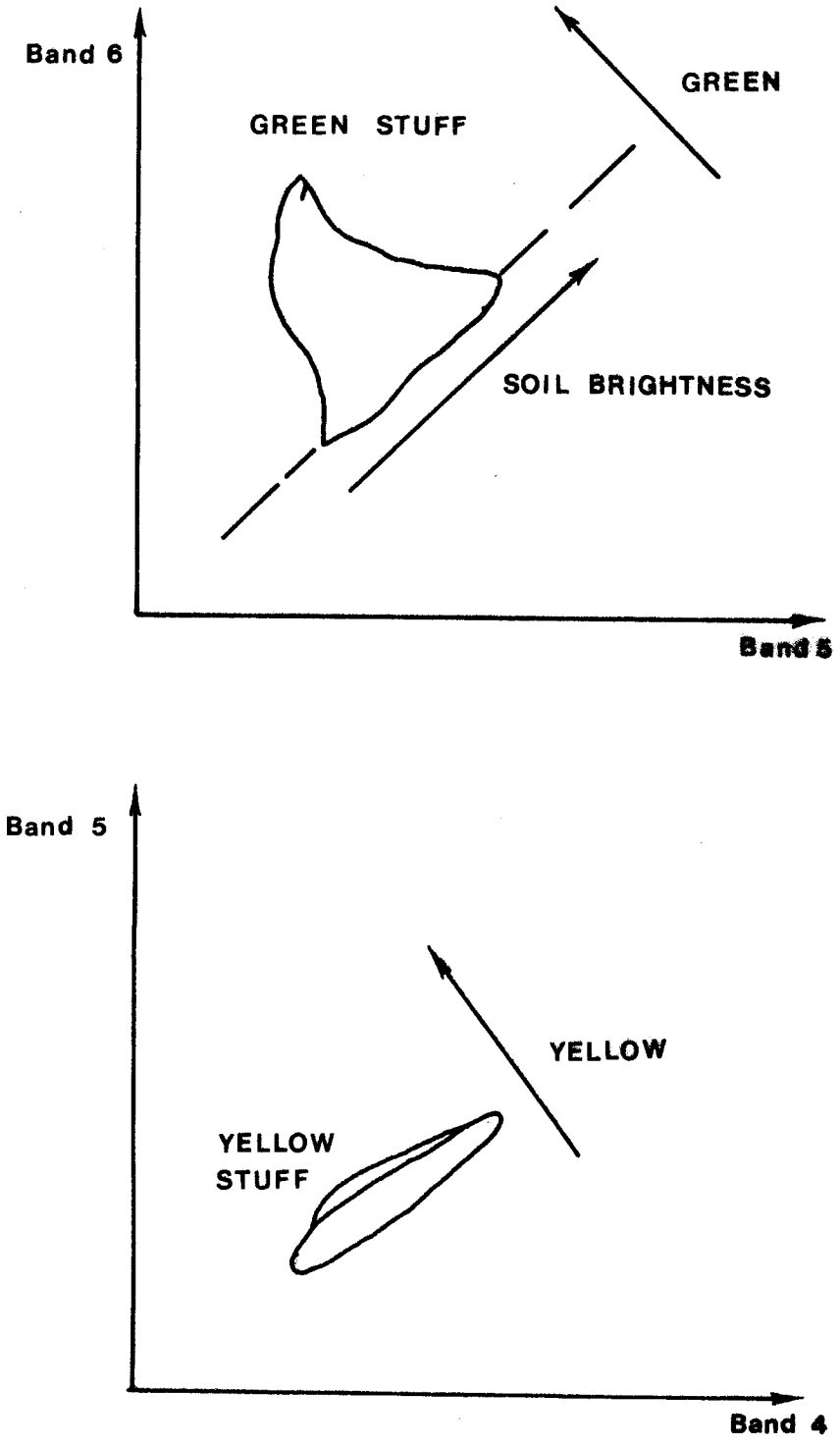


FIGURE 4.7: Graphical relationship of brightness, greeness and yellowness axes of Kauth-Thomas transformation to bands 4, 5 and 6 of Landsat and scatter plot of mixed soil, vegetation response.

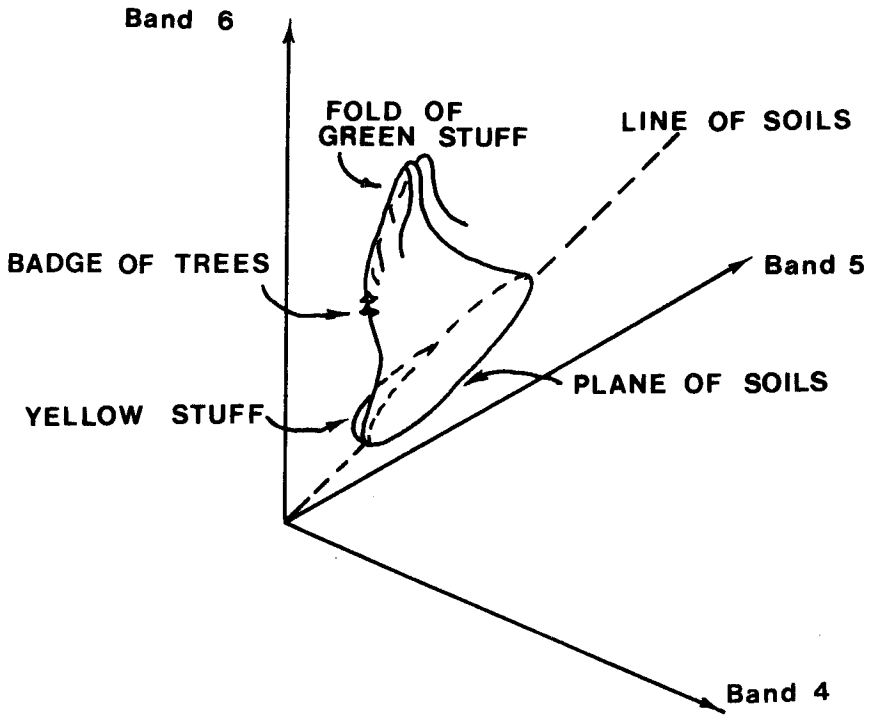


FIGURE 4.8: The Tasseled Cap (after KAUTH et al, 1978).

agricultural areas is substantially contained in a plane defined by unit vectors in the brightness and greeness directions but with a small amount in the yellowness direction perpendicular to the plane. The unit vectors describing these direction and the non-such direction together form an orthogonal (rotation) matrix called the Kauth-Thomas or Tasseled Cap transform.

If X is the Landsat response vector, and Z the transformed vector then (KAUTH and THOMAS, 1976)

$$Z = R X$$

where

$$R = \begin{bmatrix} 0.433 & 0.632 & 0.586 & 0.264 \\ -0.290 & -0.562 & 0.600 & 0.491 \\ -0.824 & 0.533 & -0.050 & 0.185 \\ 0.223 & 0.012 & -0.543 & 0.809 \end{bmatrix}$$

and a transformed vegetation index can be given by (RICHARDSON et al, 1977)

$$GVI = -0.290\ B_4 - 0.562\ B_5 + 0.600\ B_6 + 0.491\ B_7$$

and soil background index by

$$SBI = 0.433\ B_4 + 0.632\ B_5 + 0.586\ B_6 + 0.264\ B_7$$

(Note B_7 is not the doubled band 7 count range referred to in earlier sections, use of these values would require the last column of R to be divided by 2).

It can be seen that this transformation is very similar to the $(B_5 + B_7)$, $(B_7 - B_5)$ transformation considered previously. The new variables are linear functions of the Landsat MSS response and can be used in a predictive multiple linear regression without any further transformation. For a mixed vegetation and soil surface response the indices should be uncorrelated but as for $(B_5 + B_7)$ and $(B_7 - B_5)$ the response values will be affected by atmospheric and solar zenith angle changes. A plot of the first two transformed axes equivalent to SBI and GVI are shown in figure 4.9 with vegetation and soil from Table 4.1 transformed into the two dimensional space.

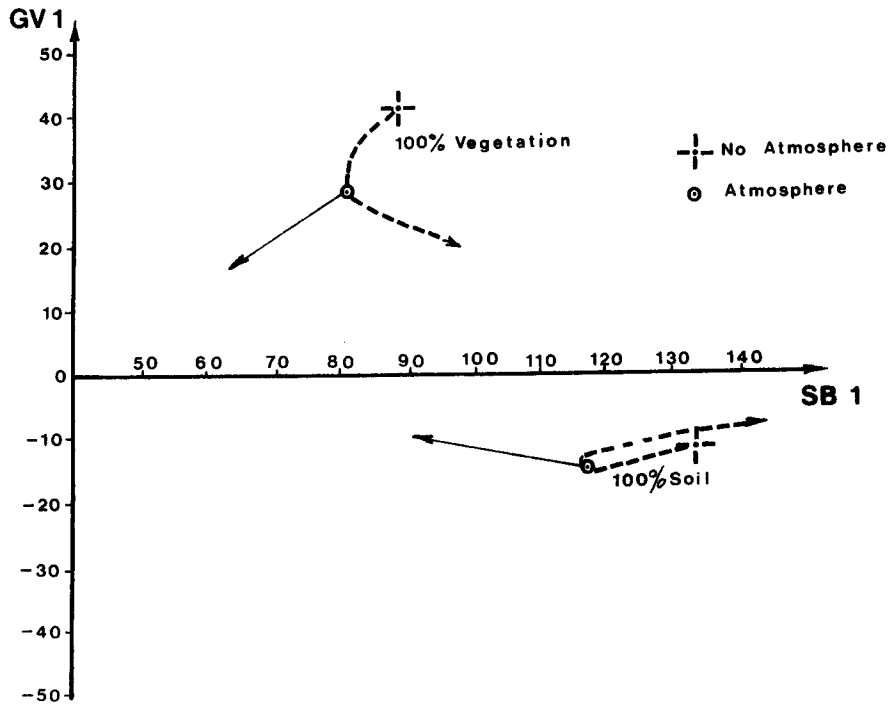


FIGURE 4.9: Plot of first two Kauth-Thomas transformed axes equivalent to GVI and SBI for approximate response counts of vegetation and soil, with and without atmosphere. The approximate direction of change for increasing solar zenith angle (solid line) and increasing haze (dotted line) are also shown. (Atmosphere approximately 15km visibility).

The Kauth-Thomas transformation has been used by RICHARDSON et al (1977) for distinguishing vegetation from soil background information, by THOMPSON et al (1979) to detect and monitor agricultural vegetative water stress over large areas and in LACIE (Large Area Crop Inventory Experiment) to enhance crop growth characteristics (HEYDORN et al 1978, KAUTH et al, 1978). How such a transformation could be applied to urban data has not yet been investigated, but reflectance data on urban residential surfaces (FORSTER, 1980b) suggests that manmade materials could have similar response properties to soil and an urban brightness axis and an orthogonal greeness axis could be defined. It was also shown by FORSTER (1980b) that the band 5, band 4 difference was correlated with percentage concrete cover. This suggested an equivalence of concrete percentage to the yellowness axis of the Kauth-Thomas transformation.

4.6 Principal Component Transformation

The principal component transformation is essentially the same as the Kauth-Thomas transformation in that four new orthogonal axes are defined. Unlike the latter transformation, however, the principal component method allows natural axes to evolve from the underlying data set without a priori conditions, such as the defined soil brightness axis, being imposed and provides new variables which are pairwise uncorrelated. Each principal component is a linear combination of the observed variables and these linear functions are chosen to be orthogonal. An infinite number of orthogonal bases of a vector space can exist, so in order to provide a unique set of coefficients the first principal component is defined as that linear combination of variables which has the maximum variance of all linear functions derivable from the given variables. The second principal component is the linear combination of variables having the maximum variance of all linear functions of the given variables that are orthogonal to the first principal component, continuing to as many principal components as there are independent variables (MATHER 1976, 215-216). The principal components themselves are the eigenvectors of the variance - covariance matrix of the variables observed.

The Landsat response data consists of measurements which are correlated with each other. Principal component analysis allows this data set to be transformed into one where the variables are mutually

uncorrelated. In addition the method allows the determination of the number of linearly independent sources of variation within the Landsat data set that can effectively summarize the data without a significant loss of information. Because of correlation this will be generally less than the number of variables.

A number of researchers have used principal component analysis, for example DONKER and MULDER (1976), AUSTIN and MAYO (1978) and LODWICK (1979). Generally it has been found that the first two principal components of a Landsat data set contribute of the order of 98% of the total variance. The first principal component, PC_1, is seen as representing a general brightness component, and the second, PC_2, a general greenness component or as DONKER et al suggest *"the designation colour is a better one."*

It was of interest to compare the matrix of eigenvectors from DONKER et al (1976) and AUSTIN et al (1978), and these are as follows

Eigenvectors

		1.	2.	3.	4.
	4	0.186	0.524	0.827	0.083
MSS	5	0.360	0.735	-0.560	0.129
Bands	6	0.766	-0.241	0.040	-0.594
	7	0.499	-0.356	0.035	0.789

(from DONKER et al) and

	1.	2.	3.	4.
4	0.135	0.538	0.831	-0.049
5	0.281	0.757	-0.519	0.281
6	0.831	-0.157	-0.065	-0.529
7	0.460	-0.336	0.190	0.799

(from AUSTIN et al)

It can be seen that the matrices are very similar and realizing that the first is based on a data set from the south east region of the Netherlands and the second from the south coast of New South Wales, Australia, the underlying locational invariance of the data sets is even more remarkable. LODWICK (1980) considers that it is preferable to use standardized variables to derive the eigenvectors which then results

in the following matrix

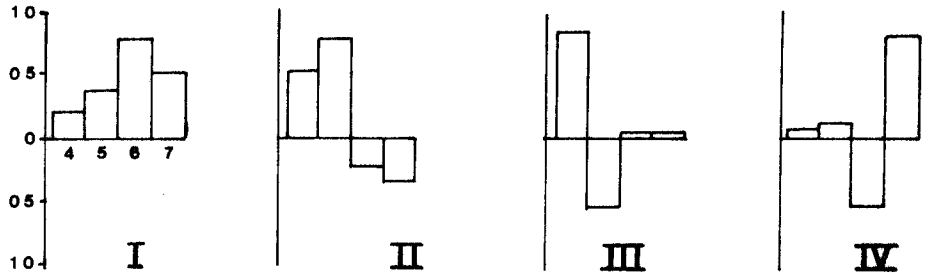
		Eigenvectors			
		1	2	3	4
	4	0.5	0.5	0.7	0
MSS	5	0.5	0.5	-0.7	0
Bands	6	0.5	-0.5	0	-0.7
	7	0.5	-0.5	0	0.7

which in terms of signs and structure is the same as the previous matrices, a band summation, visible minus infrared, difference of the visible, difference of the infrared. This structure is also essentially in the same form as the Kauth-Thomas transform. The loadings of the four spectral bands on the principal components from each of DONKER et al (1976), AUSTIN et al (1978) and LODWICK (1980) and the loadings of the Kauth-Thomas transform on its transformed axes are compared in figure 4.10. Apart from a sign reversal on axes II and III of the Kauth-Thomas transform the essential characteristics of the four examples is the same. As for the Kauth-Thomas transform, the principal component transformed variables will not be invariant under changing atmospheric conditions or changing solar zenith angle.

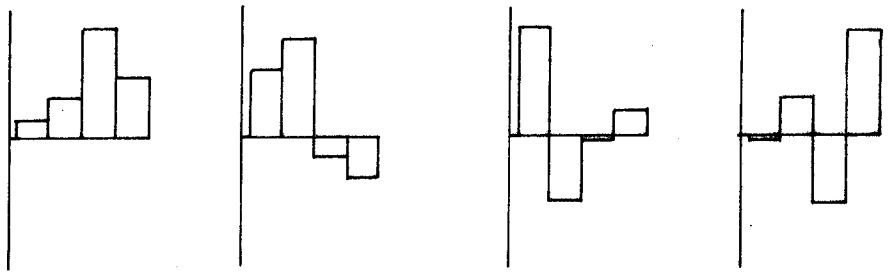
All of the principal component examples come from predominantly rural areas, but it might be expected, as suggested in section 4.5 for the Kauth-Thomas transform axes, that similar principal components could be found in the four dimensional Landsat response space from urban residential areas.

4.7 Standard Deviation or Variance over Extended Areas as a Measure of Texture

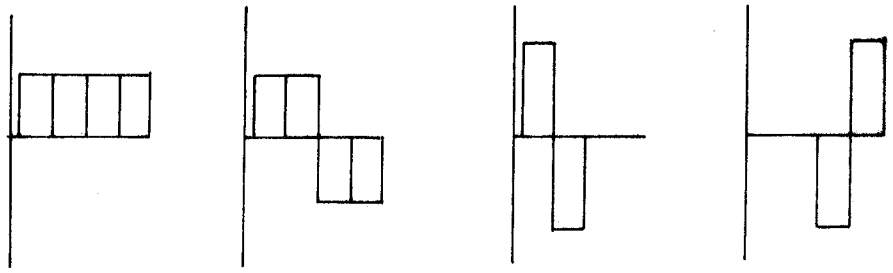
In certain circumstances cover information over extended areas may be preferable to that in individual pixels. It could be expected that the variance would reduce by using average reflectance and average cover data, so that higher multiple correlations would be achieved in multiple regression analysis. The application to extended areas also raises the possibility of using textural type variables to predict surface cover. TODD and BAUMGARDNER (1973) have found that the standard deviations in the infrared bands aid in separating residential areas.



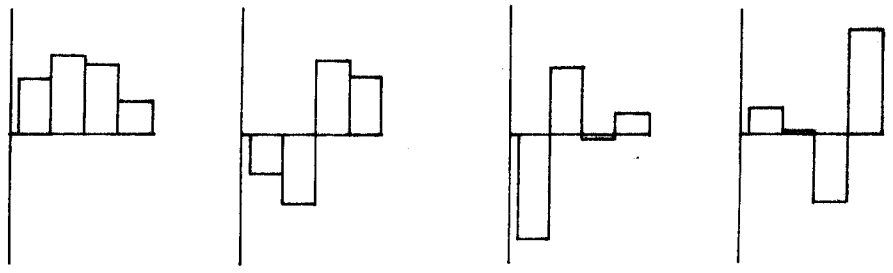
(a) DONKER and MULDER



(b) AUSTIN and MAYO



(c) LODWICK



(d) KAUTH – THOMAS

FIGURE 4.10: Comparison of loadings on Landsat MSS bands for principal components (a, b, c) and Kauth-Thomas axes (d).

If a homogeneous, extended vegetative surface was viewed by the sensor there would be little variance expected in the response signal. A similar situation would exist for a total manmade surface. When a mixture of the two was viewed, the distribution of the various surfaces in an urban residential situation would not be random but would tend to form linear, block oriented clusters. The variance from such an extended surface would therefore tend to be greater than from a homogeneous surface. The variance (or standard deviation) over the extended area would be small for a homogeneous surface, rise to a high depending on the patterning of the residential mixture and fall to a low value from another homogeneous surface, suggesting that in addition to the reflectance data from each band, that variance and a quadratic term in variance (or standard deviation) would also give additional information as to the percentage amounts of various cover.

A cover related variable that is useful in urban studies is housing density, where housing density is the number of houses per unit area. This variable cannot be predicted in a single pixel situation except via its correlation with housing cover, as there is no reflectance measure than can discriminate whether the amount of house cover is due to a small number of large houses or a large number of small houses. However over an extended area it might be expected that the variance of the reflectance in each band, and particularly in the infrared bands, might be greater the lower the number of houses per unit area. The increase in variance can be explained as follows. The reflectance from a pixel containing a large number of small houses with associated vegetation, roads and concrete, will not vary greatly from adjoining pixels containing substantially the same mixture, and so the variance will be small. However with a small number of larger houses with larger rear yards and greater separation the houses might predominantly cause the reflectance from one pixel, and the reflectance from adjoining pixels might be substantially due to vegetation which separates the houses, producing a larger variance.

It is suggested then that a mixed residential surface, where the component surface are small relative to the pixel dimensions, will tend to have a small variance over an extended area, but higher than a completely homogeneous surface. A mixed residential surface, where the component surfaces are relatively large, will tend to have an overall higher variance.

Apart from the work of TODD et al (1973) there is little evidence to suggest the combinations of these textural variables that are significant. However following upon the analysis of ratios and transformations, given previously, a visible infrared difference and also a measure of brightness variation, could be applicable. For these reasons it was considered that the standard deviation in each band, the square of the standard deviation in each band (variance), the difference in the visible and infrared standard deviation and the average standard deviation in all bands, were potential predictor variables. Although these variables are not primary predictors they can nevertheless be used as additional independent variables to the band ratios and transformations in multiple linear regression analysis with cover related dependent variables.

4.8 Summary and Conclusions

Essentially all of the ratios and transformations attempt to distinguish between a measure of relative colour and a measure of intensity or brightness. Each has particularly advantageous characteristics. The various ratios are relatively invariant under varying atmospheric or sun angle conditions and also reduce noise effects. Others vary in a linear manner with vegetation response, as do the(band 7 - band 5), (band 7 + band 5) transform and ratio, or provide uncorrelated variables for further analysis. As the radiation response from various residential surfaces are reduced to their respective spectral reflectance, with the effects of atmosphere accounted for, the advantage of invariance under varying atmospheric conditions is not of itself advantageous to this study. Two advantages of a ratio or a transformation, when being used in a multiple regression analysis to predict percentage cover, are their linear response to a change in that percentage and their ability to create essentially uncorrelated variables. This suggested that the (band - band), (band + band) transformation and the principal component transformation were those to be preferred, although a transformation similar to the Kauth-Thomas transformation but with urban axes defined was also considered to be useful. Neither of these, however, will reduce noise effects unlike the various ratios.

Because the (band on band) or (band on sum of band) ratios are not linearly related to cover percentage change these appeared less useful (with the exception of the (band 7 - band 5) on (band 7 + band 7) ratio which is essentially linearly related to vegetation change). With these

ratios a number of polynomial terms is needed to linearise a relationship which in other transformations is accomplished with fewer variables. In addition, to have an exactly defined predictive model, each ratio, which essentially measures the direction of an observation vector, should be used in conjunction with the magnitude of that vector, the V_{ij} term previously defined, which with atmospheric changes accounted for can act as a predictor variable in its own right. It seemed preferable to use the minimum number of variables for example $(B_i - B_j)$, $(B_i + B_j)$, type variables rather than V_{ij}, R_{ij} and polynomial terms in a predictive equation. Nevertheless stepwise multiple regression analysis can isolate those variables of greatest significance, and with only limited previous research in urban areas available to point to particular combinations it was decided to test all ratios and transformations previously examined, bearing in mind their limitations and further to attempt to derive other empirical relationships. For extended areas, the use of standard deviation or variance as additional predictor variables appeared feasible. It was considered that these values, calculated for each band, should be tested for significance. All results are given in Chapter 7.

5. DATA COLLECTION AND PREPROCESSING

5.1 Introduction

The aims of the research, given previously in Chapter 1, section 1.4, necessitated the collection of four separate sets of data.

(i) Ground control points were required for the calculation of the transformation parameters between ground and Landsat coordinates. These parameters allowed the registration of selected ground truth areas and equivalent parts of the Landsat scene.

(ii) Percentage cover data sampled over each ground truth area were required to be related via regression analysis to the measured Landsat MSS response.

(iii) Landsat digital response data over the selected ground truth and surrounding areas were required to determine background reflectance and to be related via regression analysis to the percentage cover data.

(iv) House value data, as one measure of environmental quality, were required to be related via percentage cover data to Landsat response.

Seventy residential ground truth areas were selected over the Sydney metropolitan area. Each ground truth area contained forty Landsat pixels, in a block of eight along scan pixels and five across scan pixels. Ground control points were selected around each ground truth area and their coordinates in both the Landsat coordinate system and ground coordinate system were determined.

The calculated transformation parameters allowed the registration of the selected ground truth areas with the Landsat response data on a pixel by pixel basis. Ground truth areas needed to be spatially related to the Landsat data at the sub-pixel level if the correlation between ground and satellite data was to be fruitfully examined. Ground and Landsat coordinates were related using a fifth order polynomial interpolation formulae, resulting in standard errors of the order of 30 metres. These transformation procedures however were considered peripheral to the main study and for this reason are given in Appendix 3.

The percentages of various urban cover surfaces, contained in each ground equivalent pixel area, were sampled from large scale aerial photographs. Twenty sampling points were used per pixel. The sampling procedure is illustrated in figure 5.1 and further details and the justification for the sampling design are given in the next section.

5.2 Sample Design

One aim of the research was to determine whether variables derived from Landsat digital data can be used to predict the percentage of various cover classes over an extended area. The relationships between the various data sets must first be determined from samples which can then be used as predictors for the whole population.

When two variables are studied a simple measure of the relationship is the correlation coefficient. For this study it was considered that a correlation of less than 0.3 was of limited value and that its significance should be such as to be able to reject the null hypothesis at the 1% level. From section 3.3.1 a calculated value of n of approximately 60 is required to achieve this significance. Any relationships having higher correlations than 0.3 or determined from more samples would necessarily be significant at the 1% level or better.

A systematic distribution of these samples was decided upon so as to obtain an adequate representation of the population. There is little chance with this method that a large contiguous part of the population would fail to be represented, a number of studies also report other advantages in addition to its convenience. COCHRAN (1963, 223-224) reports a number of studies which indicate that systematic sampling shows a consistent gain in precision over stratified random sampling particularly for data where variation would be nearest to continuous. HOWARD (1970, 227-228) considers for photo-ecological studies, that systematic sampling can provide as good or sometimes a better estimate of the mean for a specified number of samples when compared to random sampling.

It was found that seventy sample areas at approximately four kilometre intervals adequately covered the Sydney metropolitan area. Samples were taken as being those residential areas at or nearest the four kilometre grid intersections. Although this was ten more samples than previously considered necessary on the basis of significance, the figure of seventy was adopted because of the administrative convenience of whole number grid intervals, and to provide a margin of safety particularly

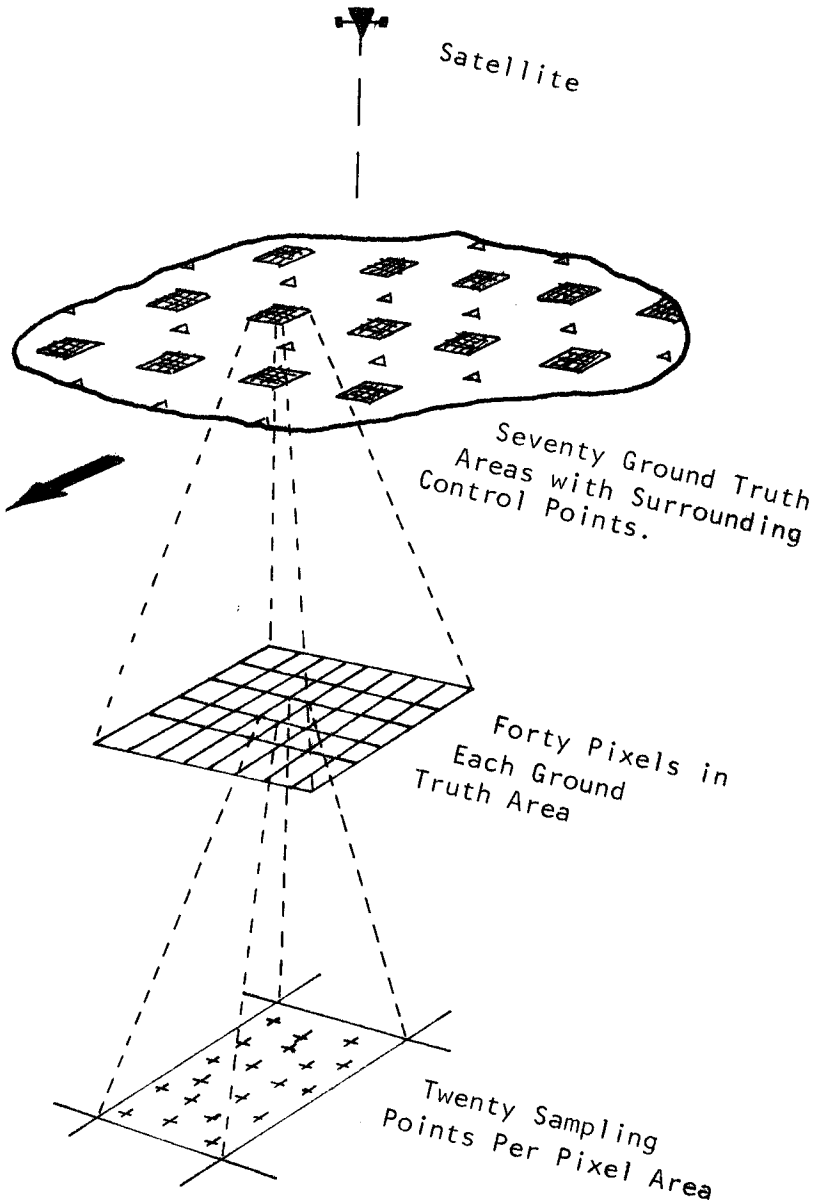


FIGURE 5.1: Schematic diagram of the sampling procedure.

when more than two variables were being related with a subsequent loss of degrees of freedom.

Because the relationship between cover percentages as sampled from aerial photographs and Landsat digital response was to be initially determined at the pixel level, sampled cover characteristics at this level had to be obtained. It was considered that a maximum standard error of the estimate of the percentage cover of approximately 10% would be adequate when an individual cover comprised 50% of the pixel. This figure can be approximately obtained when twenty sample points are used per pixel. Increasing the number of samples to say 50 would only marginally improve the standard error to 7%, and for the additional work involved this was considered uneconomic. In addition the error will not bias the computed regression coefficients although it will increase the variance and so reduce the explanation of the regression.

A stratified unaligned systematic sample with a pixel divided into 20 grid cells was considered appropriate because this method shows the best results when used on cyclic phenomena (BERRY, 1962) and so should tend to reduce the systematic effects of a regular urban pattern. An overall cover of sample points is achieved with each points position in its cell being essentially random (see figure 5.2).

A cluster of forty such pixels, at each sampling site, was selected so that the effects of the point spread function of the sensor could be contained within the sampling area. DYE (1975) suggested a 7 x 5 array of pixels would adequately contain this effect, although for more approximate work a 3 x 3 array was considered sufficient. The slightly larger array of eight pixels along scan and five across scan was selected to compensate for any positional error of the sampled array relative to the Landsat response array and also to increase the number of possible 3 x 3 arrays within each sample area.

5.3 Sampling of Cover Data

Seventy residential areas at approximately four kilometre intervals were selected for study. An area of approximately 400 x 500 metres was sampled for cover percentages around each point. These areas contained forty Landsat pixels, five across scan and eight along scan.

Data relating to cover characteristics were sampledfrom black and white panchromatic aerial photographs taken in 1971. These were at a scale of 1:15 000 and were enlarged to a scale of 1:2 000 over each sample area, to act as a sampling base. Large scale colour or colour

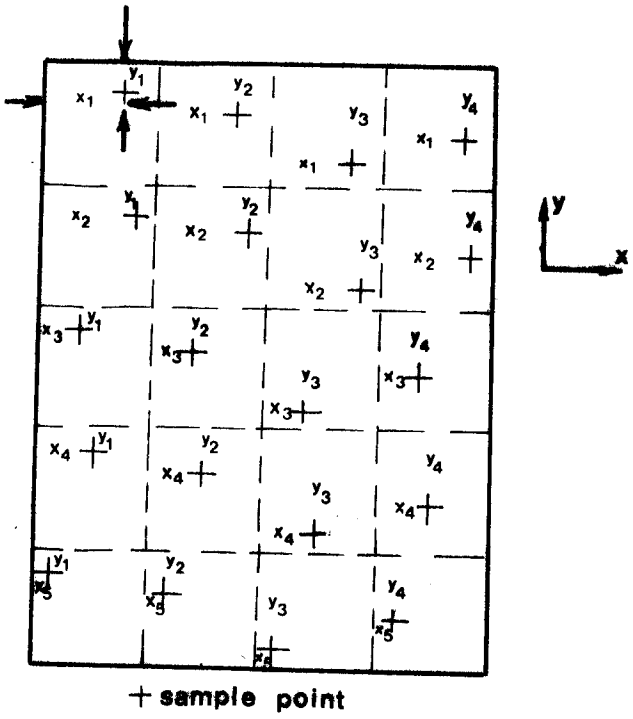


FIGURE 5.2: Location of sampling points within a pixel using a stratified unaligned systematic sampling procedure. x_i and y_i coordinates are randomly selected.

infrared photographs were preferred, but were not available. A sampling overlay at a scale of 1:2 000 was prepared. This consisted of an eight pixel along scan by five pixel across scan grid, with each row of the grid stepped four metres along scan to approximately account for sensor delay inherent in the Landsat system, as illustrated in figure 5.3. In each pixel twenty sampling points were marked.

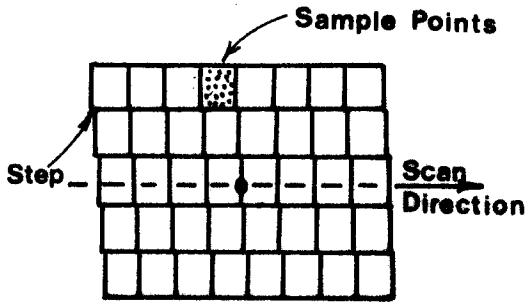


FIGURE 5.3: Sampling grid used to estimate ground cover percentages showing step for sensor delay. Each cell represents one Landsat pixel.

The centroid coordinates of each of the sample areas were transformed into their Landsat equivalents using the derived transformation parameters (see Appendix 3) and the small distances to the nearest pixel centre (across scan) and midway between pixels (along scan) were determined i.e. the centre of a 5 x 8 block of pixels. These adjusted centroids and the direction of the across scan track were marked up on 1:10 000 planimetric maps covering each sample area, and then transferred to each enlarged photograph by comparison of map and photo detail.

The sampling overlay was registered with each photo and the following data were sampled over each pixel.

House percentage cover (H)
Other Building percentage cover (O)
Road percentage cover (R)
Concrete percentage cover (C)
Tree percentage cover (T)
Grass percentage cover (G)
Water percentage cover (W)
Soil percentage cover (S)
Number of houses per pixel (N)

House roofs in Sydney are predominantly red/brown tile and should have similar reflectance characteristics. Other buildings separate from the main dwellings, consisted of small buildings, predominantly with weathered iron roofs. In addition this class of cover also included a small percentage of commercial, industrial, and multi-family units that encroached on the predominantly single family dwelling areas. Roads were of asphalt construction, and concrete included footpaths, drives, parking areas etc. Tree and grass percentages are self explanatory, however grass at the time of overflight (southern summer) is relatively dry and trees are predominantly native with relatively few deciduous trees. Water percentage included swimming pools, and in coastal areas sea water where the sampling grid extended slightly over the sea.

Soil percentage was a catch-all class for any area that was not covered with vegetation or man-made structures and included soil, exposed rock and sand. These cover types have quite separable signatures normally, but in this study they amounted to only a few percent of the total cover and could be considered together.

The number of houses per pixel was sampled so that housing density could be tested for its correlation with band response and also to allow the calculation of house size, house percentage divided by the number of houses, which from previous experience is known to be a significant variable for predicting house value (FORSTER, 1975).

5.4 Landsat Digital Data

The equivalent Landsat response in each band and for each individual pixel was obtained from the computer compatible tapes of Landsat scene No. 1141 - 23140, the *Sydney Scene* of December 12th 1972, using the line and pixel coordinates of each to access the data. Pixel and line values have their origin in the northwest corner of the Landsat frame, positively increasing west to east for pixel values and north to south for line values. The Sydney sub-scene begins at approximately line 1500 and pixel 1300 and covers a square of approximately 35 km sides, or approximately 600 pixels east/west by 450 pixels north/south. A computer card deck listing each pixel in blocks of forty was prepared, with the appropriate cover and response data related to it, forming a total data file of 2800 pixels.

Where an interswath discontinuity occurred, care had to be taken in selecting the appropriate pixel responses to relate to the sampled cover data. While the ground and Landsat coordinates were registered at the centroid the occurrence of the discontinuity within the sampling frame of 5 x 8 pixels meantthat the appopriate pixel response selected on the opposite side of the discontinuity to the centroid had to be stepped 50m (or approximately onepixel width) west when the discontinuity was south of the centroid and 50m east when it was north of the centroid.

5.5 Background Reflectance Data

The average background reflectance, $\overline{R}_{B_i}$, i = 4 to 7, required to calculate target reflectance(see section 2.5.5), can be obtained by collecting surface cover information over an area surrounding a central target pixel and weighting the reflectance according to the respective relative areas of each cover type (TURNER et al, 1972, TURNER, 1975). These sampled proportions and the known reflectance of each cover type are then used to obtain the average background reflectance. This procedure was rejected for a number of reasons -

(i) The reflectance of individual cover surfaces was unknown prior to the calculation of the background reflectances.

(ii) The sampling of cover surfaces around each individual pixel would be very time consuming.

(iii) The method was not particularly suitable for use with a computer because each pixel would require a resampled and recalculated background.

An alternative approach can be suggested from an examination of the nature of background reflectance. The effect of the reflectance of a background pixel on the path radiance at a target pixel reduces the further the pixel is from the target pixel, due to the increasing length of path travelled by the energy from the background pixel to the sensor, which produces increased scattering and absorption. Increasing the angle made between the energy path and the sensor nadir also means that less energy is scattered in the direction of the sensor due to the effect of the single scattering phase function. Thus the major effect of background reflectance is due to those areas immediately surrounding the target pixel.

It was therefore considered that little error would result if the average reflectance over each ground truth area was used as the background reflectance for each pixel within that ground truth area. It was further considered that the average response count over each ground truth area, converted to reflectance, should be used in preference to determining the background reflectance from sampled cover percentages. The average response count over a ground truth area was easily derived from the computer stored digital data and it therefore required only the calculation of the relationship between count value and reflectance to effect the conversion.

When average reflectance over each ground truth area is used the target and background reflectance can be considered equal and (A 1.3) which relates target reflectance to response count and background reflectance, can be inverted to give background reflectance, $\overline{R}_{B_i}$, as a function of response count, C_i. The following equations can therefore be derived

$$\overline{R}_{B_4} = \frac{C_4 - 10.25}{(1.950 + 0.00105C_4)}$$

$$\overline{R}_{B_5} = \frac{C_5 - 7.61}{(2.067 + 0.00072C_5)}$$

$$\overline{R}_{B_6} = \frac{C_6 - 4.30}{(1.907 + 0.00050C_6)}$$

$$\overline{R}_{B_7} = \frac{C_7 - 1.56}{(1.442 + 0.00026C_7)} \qquad (5.1)$$

It is suggested for use in an operational mode that the background reflectance be precomputed at ten pixel intervals (along or across scan) and these be held in a look up table form against Landsat line and pixel coordinates. In the present study, however, background reflectance was calculated only for each ground truth area.

5.6 Determination of Corrected Data Sets

Using the background reflectances calculated for each band for each ground truth area, the raw Landsat response counts from each pixel were converted to reflectance values using (A 1.3). A new data file relating cover percentages to pixel reflectance was then created, containing data for each of the 2800 (70 x 40) pixels sampled.

Two further data files were prepared which accounted for the effects of the sensor point spread function. Reflectance values for each pixel were deconvolved using (2.18) and substituting reflectance for sensor count rate. Because this calculation uses the reflectance values of the eight surrounding pixels, pixels on the edge of a ground truth area cannot be converted and so the final file contained only data for 1260 (70 x 18) pixels (see figure 5.4). This file related cover values to the deconvolved reflectance values for each pixel. A second file was created by convolving the percentage cover data of each pixel with its neighbours in the same manner as given in (2.19). Again the file contained 1260 pixels but now related convolved percentage cover values to each pixel's reflectance value.

These three files, atmospherically corrected reflectance only data, deconvolved reflectance and convolved cover data, related to their respective cover or reflectance, are the main data files used in Chapters 6 and 7. A further file relating average cover and average reflectance variables, over each ground truth area, was also used in Chapter 7.

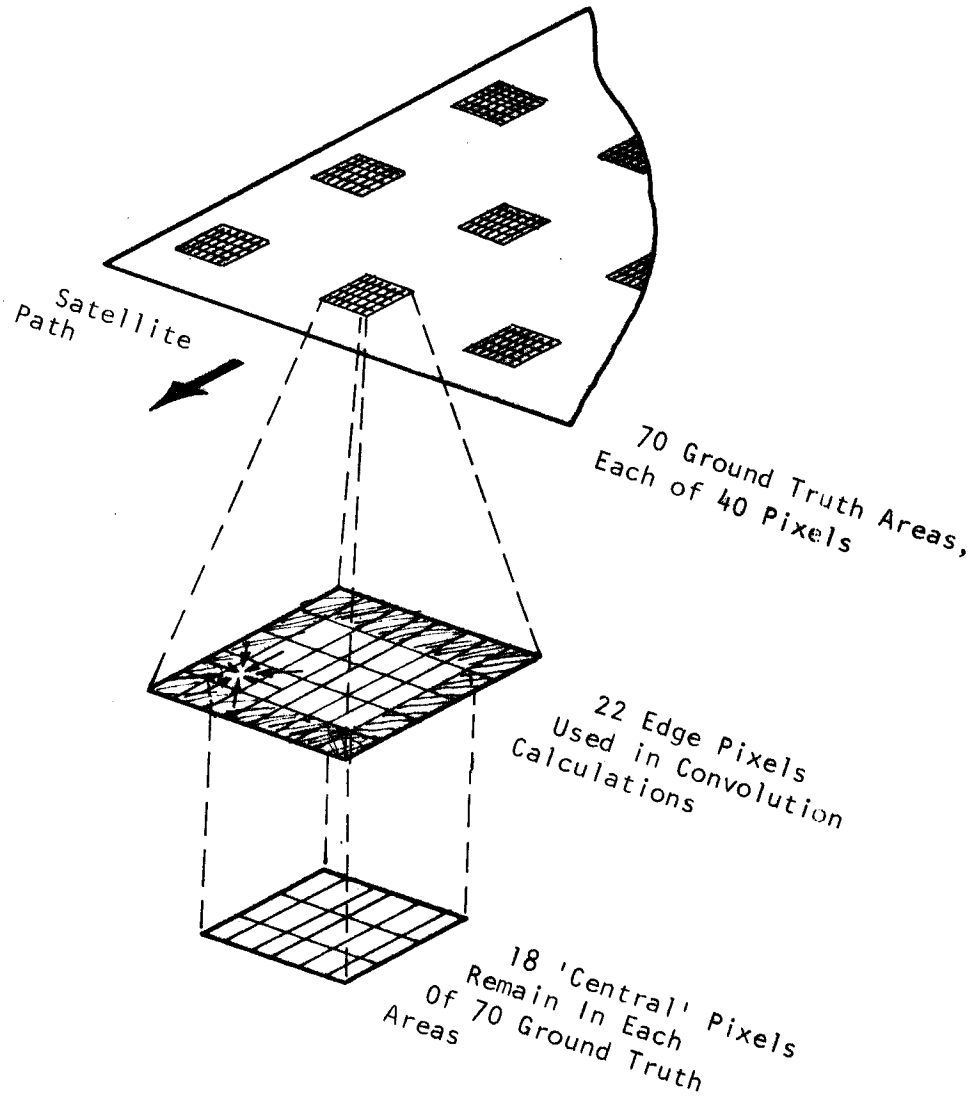


FIGURE 5.4: Relationship between edge pixels used in convolution calculations and remaining central pixels.

6. REFLECTANCE PREDICTION

6.1 Introduction

The theoretical relationship between reflectance and surface cover was described in Chapter 2, where it was shown that the reflectance from a mixed cover surface was a function of the proportionate amounts of individual cover contained in that mixture. The results of using this theoretical relationship and the sampled data to determine the coefficients necessary to convert percentage cover to reflectance, are described in this chapter. From these coefficients the reflectance from a surface containing 100% of an individual cover were extrapolated, and are compared with reflectance values from other sources in the latter half of the chapter.

Three models were studied, using reflectance in each band as the dependent variables and percentage cover as the independent variables in a multiple regression analysis. The first related count values (converted to reflectance) to percentage cover without accounting for the effects of point spread convolution. The latter two models firstly related deconvolved reflectance to cover data, and then reflectance to convolved cover data.

For each, both the standard regression procedure and the forced origin regression procedure were used. Essentially these are the same except that for one the constant is unrestricted in value and for the other the constant is forced to a value of zero. Stepwise regression was used so that the best subset of variables was output at each stage. All cover variables were allowed to freely enter the equations, provided they satisfied the minimum F ratio criteria given in section 3.3.7, and for the standard regression procedure, the least significant variable was not included.

The change in average reflectance due to a small change in the percentage of a particular cover surface was examined, and used to determine likely band combinations for predicting the percentage amounts of cover contributing to an observed reflectance. The results of this examination are also contained in this chapter.

6.2 Results of Calculations

The coefficients computed for each band, each model and the two regression procedures are shown in Tables 6.1 through 6.6. Cover variables are listed in order of their entry into the regression equation. Previously defined statistical parameters are also tabulated.

For an individual model and band the two regression procedures gave the same multiple correlation coefficient, r. The multiple correlation coefficients for different models and different bands within those models, however, were not the same. The convolved percentage cover model gave the highest multiple correlation coefficient in all four bands ranging from an r value of 0.72 for band 5 to 0.56 for band 6. Lowest multiple correlation coefficient of all three models came from the deconvolved model for band 6 with a low r value of 0.39. The deconvolved model had lower r values in all four bands when compared to the other two models. However it can be seen from the F ratios that the regressions for all models and each procedure were highly significant being well within four times the 1% confidence limits chosen as the limiting criterion in 3.37.

Some cover variables, particularly with the standard regression procedure, were not significant, the F - ratio for them to enter the equation falling below the stated criterion. It should be noted however, that when the regression is forced through the origin, the coefficients, which represent the absolute change from zero, will be statistically insignificant if the coefficient approaches zero. For the standard regression procedure a variable causing a change in the dependent variable approximately equal to the equation constant will have a relative coefficient approaching zero, and it will also be statistically insignificant. In addition the convolved and deconvolved models had a reduced data set (as the perimeter values of each ground truth area are not included) which caused a reduction in the number of pixels containing soil and water cover, as these covers predominantly occurred within the perimeter zone. Using a reduced data set would limit the significance of these variables.

Estimates of the 100% reflectance for each variable in each band were calculated by substituting for the appropriate variable. Where the coefficient was insignificant a zero value was assumed. Estimates of the reflectance for various mixtures can also be calculated by inserting

TABLE 6.1

Reflectance in each band as a function of percentage cover (standard regression procedure)

Band	Variable	Coefficient	S.E.	r	$\bar{r}^2$	S.E.	F
4	Tree	-0.061	0.002	0.61	0.37	1.7	335
	Other	0.045	0.004				
	Concrete	0.051	0.007				
	Road	0.015	0.003				
	Water	-0.029	0.007				
	Grass	*	*				
	House	*	*				
	Soil	*	*				
	Constant	14.7	0.1				
5	Tree	-0.127	0.005	0.65	0.42	2.2	341
	Grass	-0.059	0.005				
	Water	-0.120	0.009				
	Concrete	0.042	0.009				
	Road	-0.037	0.006				
	House	-0.018	0.005				
	Soil	*	*				
	Other	*	*				
	Constant	18.1	0.4				
6	Grass	0.092	0.004	0.54	0.29	2.6	224
	Water	-0.128	0.010				
	House	0.051	0.005				
	Concrete	0.108	0.010				
	Tree	0.019	0.004				
	Soil	*	*				
	Other	*	*				
	Road	*	*				
	Constant	18.3	0.3				
7	Grass	0.185	0.006	0.63	0.40	3.8	309
	Tree	0.095	0.007				
	Water	-0.170	0.016				
	House	0.073	0.008				
	Concrete	0.125	0.016				
	Other	-0.037	0.010				
	Road	*	*				
	Soil	*	*				
	Constant	22.2	0.5				

* not significant

TABLE 6.2

Reflectance in each band as a function of percentage cover (forced origin regression procedure)

Band	Variable	Coefficient	S.E.	r	$\bar{r}^2$	$\bar{r}_u^2$	S.E.	F
4	Grass	0.143	0.001	0.61	0.37	0.99	1.7	208
	House	0.154	0.002					
	Road	0.160	0.003					
	Other	0.191	0.004					
	Tree	0.085	0.002					
	Concrete	0.193	0.006					
	Soil	0.177	0.010					
	Water	0.115	0.007					
5	House	0.162	0.003	0.65	0.42	0.97	2.2	252
	Grass	0.122	0.002					
	Road	0.144	0.003					
	Other	0.182	0.004					
	Concrete	0.222	0.008					
	Tree	0.054	0.002					
	Soil	0.175	0.012					
	Water	0.062	0.008					
6	Grass	0.275	0.002	0.53	0.28	0.99	2.6	137
	House	0.234	0.003					
	Tree	0.202	0.003					
	Road	0.183	0.004					
	Other	0.182	0.005					
	Concrete	0.290	0.009					
	Soil	0.191	0.014					
	Water	0.054	0.010					
7	Grass	0.407	0.003	0.63	0.40	0.99	3.8	229
	House	0.296	0.005					
	Tree	0.317	0.004					
	Road	0.222	0.006					
	Concrete	0.346	0.014					
	Other	0.185	0.008					
	Soil	0.221	0.021					
	Water	0.052	0.015					

* Not significant

TABLE 6.3

Deconvolved reflectance in each band as a function of percentage cover (standard regression procedure).

Band	Variable	Coefficient	S.E.	r	$\bar{r}^2$	S.E.	F
4	Tree	-0.096	0.006	0.44	0.19	3.0	75
	Grass	-0.024	0.006				
	Concrete	0.052	0.017				
	Soil	0.163	0.059				
	House	*	*				
	Water	*	*				
	Other	*	*				
	Road	*	*				
	Constant	16.4	0.3				
5	Tree	-0.134	0.009	0.54	0.29	3.6	88
	Grass	-0.039	0.009				
	Concrete	0.116	0.022				
	Soil	0.237	0.073				
	Water	-0.163	0.060				
	House	0.027	0.012				
	Other	*	*				
	Road	*	*				
	Constant	15.4	0.7				
6	Grass	0.137	0.010	0.39	0.15	4.0	38
	House	0.092	0.013				
	Concrete	0.150	0.025				
	Tree	0.046	0.011				
	Water	-0.274	0.067				
	Soil	0.319	0.082				
	Other	*	*				
	Road	*	*				
	Constant	15.0	0.8				
7	Grass	0.092	0.014	0.52	0.27	5.7	92
	Road	-0.181	0.017				
	Water	-0.562	0.095				
	Other	-0.147	0.024				
	House	-0.046	0.016				
	Concrete	*	*				
	Tree	*	*				
	Soil	*	*				
	Constant	33.9	1.0				

* Not significant

TABLE 6.4

Deconvolved reflectance in each band as a function of percentage cover (forced origin regression procedure)

Band	Variable	Coefficient	S.E.	r	$\bar{r}^2$	$\bar{r}_u^2$	S.E.	F
4	House	0.168	0.005	0.44	0.19	0.96	3.0	36
	Grass	0.140	0.003					
	Road	0.161	0.007					
	Concrete	0.215	0.016					
	Other	0.158	0.012					
	Tree	0.067	0.005					
	Soil	0.333	0.059					
	Water	0.134	0.049					
5	House	0.182	0.007	0.54	0.29	0.93	3.6	64
	Grass	0.115	0.004					
	Road	0.155	0.009					
	Concrete	0.269	0.019					
	Other	0.154	0.014					
	Soil	0.391	0.072					
	Tree	0.020	0.007					
	Water	*	*					
6	Grass	0.287	0.005	0.39	0.15	0.97	4.0	27
	House	0.242	0.008					
	Tree	0.197	0.007					
	Road	0.152	0.010					
	Concrete	0.297	0.022					
	Other	0.150	0.016					
	Soil	0.465	0.081					
	Water	*	*					
7	Grass	0.431	0.007	0.51	0.26	0.97	5.7	55
	House	0.293	0.011					
	Tree	0.339	0.010					
	Road	0.161	0.014					
	Concrete	0.318	0.031					
	Other	0.193	0.022					
	Soil	0.424	0.114					
	Water	*	*					

* Not significant.

TABLE 6.5

Reflectance in each band as a function of convolved percentage cover (standard regression procedure)

Band	Variable	Coefficient	S.E.	r	$\bar{r}^2$	S.E.	F
4	Tree	-0.103	0.004	0.66	0.43	1.4	243
	Concrete	0.065	0.012				
	Grass	-0.016	0.003				
	Soil	0.133	0.037				
	Other	*	*				
	Water	*	*				
	Road	*	*				
	House	*	*				
	Constant	16.1	0.2				
5	Tree	-0.159	0.006	0.72	0.52	1.8	224
	Grass	-0.055	0.006				
	Concrete	0.109	0.017				
	Road	-0.034	0.011				
	Water	-0.120	0.036				
	Soil	0.139	0.048				
	Other	*	*				
	House	*	*				
	Constant	17.6	0.4				
6	Grass	0.098	0.007	0.56	0.30	2.2	80
	Road	-0.082	0.010				
	Concrete	0.151	0.019				
	House	0.057	0.009				
	Water	-0.221	0.044				
	Soil	0.258	0.061				
	Other	-0.047	0.013				
	Tree	*	*				
	Constant	19.3	0.5				
7	Grass	0.096	0.010	0.68	0.47	3.2	221
	Road	-0.230	0.015				
	Other	-0.183	0.019				
	Water	-0.384	0.063				
	House	-0.032	0.012				
	Tree	*	*				
	Concrete	*	*				
	Soil	*	*				
	Constant	34.1	0.7				

* Not significant

TABLE 6.6

Reflectance in each band as a function of convolved percentage cover (forced origin regression procedure)

Band	Variable	Coefficient	S.E.	r	$\bar{r}^2$	$\bar{r}_u^2$	S.E.	F
4	House	0.160	0.004	0.66	0.43	0.99	1.4	119
	Grass	0.146	0.002					
	Road	0.159	0.006					
	Concrete	0.226	0.012					
	Other	0.169	0.008					
	Tree	0.059	0.003					
	Soil	0.295	0.038					
	Water	0.149	0.028					
5	House	0.180	0.005	0.72	0.52	0.98	1.8	166
	Grass	0.120	0.003					
	Road	0.142	0.007					
	Concrete	0.283	0.015					
	Other	0.161	0.010					
	Soil	0.338	0.048					
	Tree	0.015	0.004					
	Water	*	*					
6	Grass	0.291	0.003	0.55	0.30	0.99	2.2	68
	House	0.250	0.006					
	Tree	0.193	0.005					
	Concrete	0.343	0.018					
	Road	0.111	0.009					
	Other	0.147	0.013					
	Soil	0.444	0.059					
	Water	*	*					
7	Grass	0.438	0.005	0.69	0.47	0.99	3.2	138
	House	0.305	0.009					
	Tree	0.333	0.007					
	Concrete	0.386	0.026					
	Other	0.150	0.019					
	Road	0.107	0.013					
	Soil	0.483	0.085					
	Water	*	*					

* Not significant

the appropriate cover percentages. Calculated reflectance values for the various bands, models and procedures are given in Table 6.7.

6.3 Comparison of Results

The difference in coefficients for the standard regression procedure and the forced origin regression procedure can be seen as a difference between a relative change model and an absolute change model.

The relative change model can be considered as a closed ground cell response, that is whenever the percentage cover of a variable is zero there will always be a response from 100% of the cell due to the remaining cover variables. Because of the inter-dependence of the variables, an increase of one percent in one variable must result in the loss of one percent from the combined sum of the other variables. This will be different for different variables and for different values of that variable. The resultant measured change would therefore be due to the combined effect.

In comparison the absolute change model can be considered as an open ground cell response, that is, when all cover variables are zero the cell is effectively transparent and can be considered as a black cell from which no radiance emanates. As individual cover variables occupy an increasing proportion of the ground cell the response will therefore be due to the summation of the responses from each individual cover, weighted by the proportion of the cell that it occupies.

Theoretically the deconvolved reflectance model should have achieved a higher multiple correlation than the atmospherically corrected only model, because the relationship between ground cover and reflectance was more correctly represented. The lower correlation actually obtained is considered to be due to the deconvolution procedure enchancing not only the reflectance but noise effects as well, leading to larger residuals and hence a reduced correlation. This will not be the case for the convolved cover model which is effectively the result of a smoothing operation. In addition limiting the deconvolution calculation to a 3 x 3 array will also introduce some error. A number of noise suppression techniques were applied and while these marginally improved the correlation it was considered that they so altered the theoretically correct deconvolution procedure as to make their application inappropriate.

TABLE 6.7

Estimated Reflectance for various types of urban cover

Band	House	Other	Road	Concrete	Tree	Grass	Water	Soil
4 (a)	14.7*	19.2	16.2	19.8	8.6	14.7*	11.8	14.7*
(b)	15.4	19.1	16.0	19.3	8.5	14.3	11.5	17.7
(c)	16.4*	16.4*	16.4*	21.6	6.8	14.0	16.4*	32.7
(d)	16.8	15.8	16.1	21.5	6.7	14.0	13.4	33.3
(e)	16.1*	16.1*	16.1*	22.6	5.8	14.5	16.1*	29.4
(f)	16.0	16.9	15.9	22.6	5.9	14.6	14.9	29.5
5 (a)	16.3	18.1*	14.4	22.3	5.4	12.2	6.1	18.1*
(b)	16.2	18.2	14.4	22.2	5.4	12.2	6.2	17.5
(c)	18.1	15.4*	15.4*	27.0	2.0	11.5	-0.9	39.1
(d)	18.2	15.5	15.4	26.9	2.0	11.5	0.0*	39.1
(e)	17.6*	17.6*	14.2	28.5	1.7	12.1	5.6	31.5
(f)	18.0	16.1	14.2	28.3	1.5	12.0	0.0*	33.8
6 (a)	23.4	18.3*	18.3*	29.1	20.2	27.5	5.5	18.3*
(b)	23.4	18.2	18.3	29.0	20.2	27.5	5.4	19.1
(c)	24.2	15.0*	15.0*	30.0	19.6	28.7	-12.4	46.9
(d)	24.2	15.0	15.2	29.7	19.7	28.7	0.0	46.5
(e)	25.0	14.6	11.1	34.4	19.3*	29.1	-2.8	45.1
(f)	25.0	14.7	11.1	34.3	19.3	29.1	0.0*	44.4
7 (a)	29.5	18.5	22.2*	34.7	31.7	40.7	5.2	22.2*
(b)	29.6	18.5	22.2	34.6	31.7	40.7	5.2	22.1
(c)	29.3	19.2	15.8	33.9*	33.9*	43.1	-22.3	33.9*
(d)	29.3	19.3	16.1	31.8	33.9	43.1	0.0	42.4
(e)	30.9	15.8	11.1	34.1*	34.1*	43.7	-4.3	34.1*
(f)	30.5	15.0	10.7	38.6	33.3	43.8	0.0*	48.3

*Coefficient assumed to be zero.

(a) Reflectance and cover, standard regression procedure, (b) Reflectance and cover, forced origin procedure, (c) Deconvolved reflectance and cover, standard regression procedure, (d) Deconvolved reflectance and cover, forced origin procedure, (e) Reflectance and convolved cover, standard regression procedure, (f) Reflectance and convolved cover, forced origin procedure.

The equivalence of the overall multiple correlation coefficients for each model for the two different regression procedures indicates that the assumption of a zero constant for the forced regression procedure was valid, that is the additive effect of path radiance had been eliminated. Had this assumption not been valid, then the correlation coefficient would have been somewhat less for the forced origin procedure or could have taken a negative value.

A comparison of the reflectance values of Table 6.7 show essentially the same result for either the standard or the forced origin regression procedure, except for soil and water cover variables which show some marked variation due to the limited nature of the data. The forced origin regression procedure however allows the calculation of all cover reflectances (apart from soil and water) without the need to assume that an insignificant coefficient has a zero value. For this reason this procedure is to be preferred.

A subtle but marked difference exists between the reflectance values calculated from the two models accounting for sensor point spread and the atmospherically corrected only model. In almost all cases the reflectance using the former two models are significantly more similar than either of them is with the latter model. For example in band 4 the standard error of the difference for the reflectance values were found to be significantly different at the 5% level on each of four occasions, when comparing the atmospherically corrected reflectance values with the reflectance values of each of the other two models, yet no values of these latter two models were significantly different at the same level. This similarity indicates that while these two models used different approaches to the sensor point spread problem, the underlying theory was correct. In all further work described in this chapter the results of the convolved cover model using the forced origin regression procedure are used.

A comparison of the spectral signatures of each of the covers, grass, trees, house, concrete, road and other is given in figure 6.1. Most immediately obvious is that the discrimination of the individual covers is greatest in the infrared bands and also the typical vegetative signatures of grass and trees. On closer examination of figure 6.1 it can be seen that grass and trees, house and concrete, and other and road reflectances closely parallel each other indicating three distinct surface classes with subclasses being differentiated by brightness only. These three classes might be considered as vegetation, urban residential

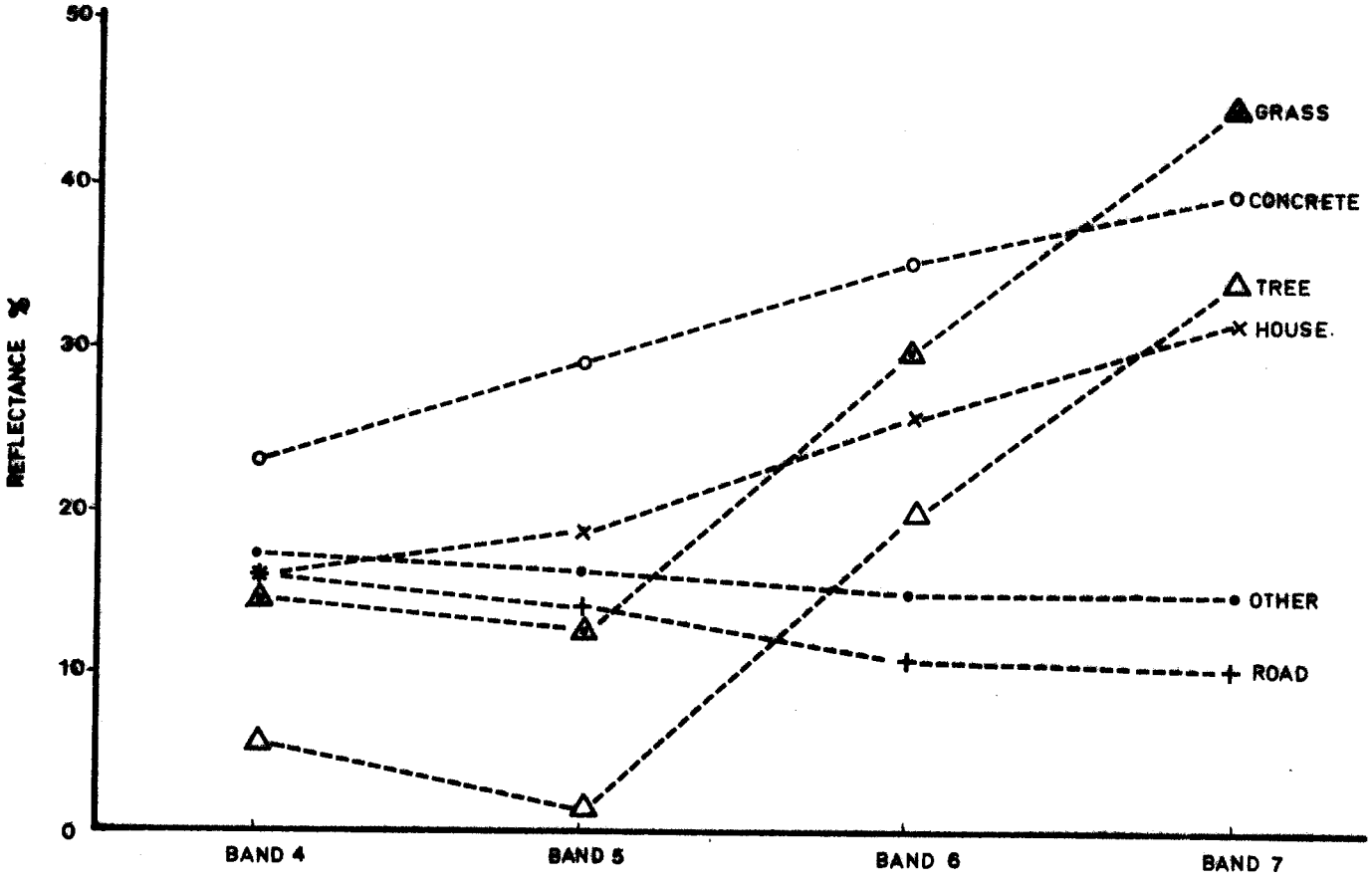


FIGURE 6.1: Percentage reflectance of various residential cover surfaces in the four Landsat bands. (Note that lines between point reflectance values in each band are illustrative only and do not represent continuous reflectance values).

and non-residential. These classes and subclasses were examined further. The results of this examination are given in Chapter 7.

Very little urban reflectance data was available to make an external comparison of the results shown in Table 6.7. BAILEY (1979) however had used Landsat data over the Sydney area to define a number of different landscape types. Although many of these types or classes included a mixture of surface covers some were sufficiently homogeneous to allow a comparison. Additionally, because the data derived from homogeneous areas, the effects of the sensor point spread function and incorrect pixel positioning were minimal. The response counts of the training set categories of bush, units, large expanses of concrete and short grass, from BAILEY (1979), are shown in Table 6.8. Units are the Australian equivalent of flats, condominiums, apartments etc and typically have large tiled roof areas with little open space between them and are here considered as equivalent to a homogeneous cover of house as defined for this study. Bush refers to large expanses of Australian native trees, mainly eucalyptus.

TABLE 6.8

Landsat response count values of various homogeneous urban cover (after BAILEY, 1979)

Cover	Band 4	Band 5	Band 6	Band 7
Bush	25.8	18.4	34.2	37.2
Units	45.4	43.8	50.8	46.8
Concrete	87.2	90.8	75.8	60.0
Grass	40.0	33.6	62.3	65.6

*Doubled Band 7 values

The count values of Table 6.8 were approximately converted to reflectance values using (5.1). The Sydney scene used by BAILEY was the same as used for this study so that atmospheric conditions were also the same. In addition the cover was sufficiently extensive to assume a background reflectance equivalent to the target reflectance. These computed reflectance values and the reflectance values derived from Table 6.6 and indicated by the letter (f) in Table 6.7, are given in Table 6.9.

TABLE 6.9

Comparison of percentage reflectance of various urban cover with those derived from BAILEY (1979)

Cover	Band 4		Band 5		Band 6		Band 7	
		*		*		*		*
Bush	5.9	7.9	1.5	10.8	19.3	15.5	33.3	24.6
Units	16.0	17.6	18.0	17.2	25.0	24.1	30.5	31.1
Concrete	22.6	37.7	28.3	39.0	34.3	36.8	38.6	40.1
Grass	14.6	14.9	12.0	12.4	29.1	29.9	43.8	43.9

* Derived from BAILEY (1979)

Tree cover of the present study has a lower reflectance than bush in the visible bands but is higher in the near infrared bands. It is considered that this is due to the increased shadow from an extensive stand of trees as represented by bush as compared to the single or smaller groups of trees found in residential areas. The overall effect of shadow is to flatten or average out the typical vegetative signature of trees. Shadow strongly reduces the response from the infrared and so bush would also tend to have lower values in these bands due to this effect.

Both results show that the reflectance of concrete is quite high across all bands, BAILEY's results however show even higher reflectances than those of the present study, particularly in bands 4 and 5. Her results show a relatively constant reflectance across all bands compared to a steadily increasing reflectance as indicated by this study. The average cover of concrete in the present study was approximately 5% with a standard deviation of 5%, indicating that any additive response from concrete was due to very small areas of that cover. It is assumed that the reflectance of such small amounts of cover are still substantially affected by the average background reflectance and so the measured reflectance is seen as being approximately the average of the true reflectance of concrete from an extended surface and the background average reflectance. It could also be further suggested that differences will arise from extrapolating from regression coefficients derived from a small

percentage of concrete cover to a 100% equivalent total cover value.

For the two surfaces, grass and house, where neither shadow nor proportion of cover can cause any adverse reflectance changes, the results compare extremely well with BAILEY's grass and units. Average differences across all bands being 0.4 percent reflectance for grass cover, and 1.0 percent reflectance for house (units) cover.

The results of the present study are also compared with those given in Table 2.2 which showed the hemispherical and nadir reflectance of urban type surfaces in the visible portion of the spectrum. Assuming the surface description of Table 2.2, of concrete, macadam, forest and grass (dry) are approximately equivalent to the present study cover classes of concrete, road, bush and grass, and further that their reflectances are averaged for hemispherical and nadir reflectances then the following comparison with the average 'visible' reflectance (average of band 4 and band 5 reflectance for convolved cover, forced regression procedure) from 4 (f) and 5 (f) of Table 6.7 can be made:-

From Table 2.2		From Table 6.7	
concrete	28.3%	concrete	25.5%
macadam	12.1%	road	15.0%
forest	4.5%	bush	3.9%
grass (dry)	10.8%	grass	13.3%

The results are clearly in good agreement even given their diverse sources. The marginally lower reflectance for concrete of the present study was also noted in the BAILEY (1979) comparison.

6.4 Cover Change and Its Effect

While all models gave reasonable estimates of the total reflectance when each cover comprised 100% of the ground cell, it was difficult to visualize the effect on reflectance of a small change in any individual cover variable because of the interdependence of the total variable set.

Each individual ground cell can be considered as being comprised of one hundred *one percent cells*. If each of these smaller cells contained the sampled average proportion of cover variables, then the reflectance from the total cell would equal the sample average reflectance. If now one of these smaller cells was completely replaced by an individual cover, the subsequent change in reflectance would give a measure of the combined

change from an average background reflectance due to an increase of one unit in a particular variable and the corresponding decrease of one unit of average reflectance.

The average response count in each band over all sampled areas were as follows :-

Band 4	38.8
Band 5	35.0
Band 6	49.8
Band 7	48.8 (Doubled)

These were converted to average reflectance values using (5.1) to give 14.3, 13.1, 23.6, 32.5 percentage reflectance for each of bands 4, 5, 6 and 7 respectively.

Using the coefficients of Table 6.6, new coefficients can be determined that represent the combined change from an average background reflectance. For band 4 the average reflectance is 14.3%, so that as the average background reduces by one percent the reflectance will reduce by 0.143%. However if this reduction in average background is now replaced by one percent of house cover for example, then the reflectance will be increased by 0.160 percent. The combined effects will therefore produce an increase in reflectance of 0.017 percent. Similarly, coefficients for all bands and all cover variables were derived. These coefficients are shown for all cover variables, apart from soil and water, in Table 6.10.

TABLE 6.10

Coefficients, of various urban cover types, representing the change in reflectance due to a change of 1% in that cover from a surface containing the average of all cover percentages.

Band	Average Reflectance	House	Other	Road	Concrete	Tree	Grass
4	14.3	+0.017	+0.026	+0.016	+0.083	-0.084	+0.003
5	13.1	+0.049	+0.030	+0.011	+0.152	-0.116	-0.011
6	23.6	+0.014	-0.089	-0.125	+0.107	-0.043	+0.055
7	32.5	-0.020	-0.175	-0.218	+0.061	+0.008	+0.113

These new equations gave insight into the desirable band combinations required to predict particular cover characteristics. For example, band 4 or band 5 reflectance should be a good predictor of the change in tree percentage from average, because all other cover variables cause a relatively small change or are in the opposite direction. A similar suggestion can be made for the grass percentage in band 7.

Possible support for these findings comes from RICHARDSON et al (1977) who found higher correlations between crop cover and band 4 and 5 individually than any vegetation index model tested. Crop cover and these bands were negatively correlated, as are tree cover and bands 4 and 5 as indicated by the large negative coefficient of tree cover in Table 6.10. RICHARDSON et al also tabulated results showing that band 6 had a higher positive correlation with leaf area index than all vegetation index models, and that band 7 by itself was only marginally less significant. As they point out, individual Landsat bands may sometimes correlate better with various vegetation parameters than transformed composite bands used as vegetation indices.

The band 7, band 5 difference should be a good predictor of green content (i.e. the sum of grass and tree percentages), a result indicated by a number of researchers and discussed in Chapter 4, and is due to the high reflectance of vegetation in band 7 and the low reflectance in band 5. The following equation was determined from the results of Table 6.10,

$$R_7 - R_5 = 19.4 - 0.069H - 0.205O - 0.229R - 0.091C + 0.124T + 0.124G$$

where R_i = reflectance in Band i, i = 4 to 7.

The equation shows that any increase in tree and grass percentages over their average values will increase the difference between the band 7 and band 5 reflectances, whereas an increase in any non-vegetated surface will cause a decrease.

Less obvious effects can also be seen, as noted by FORSTER (1980b). The difference between bands 4 and 5 should correlate with concrete percentage. The equation of difference is given by

$$R_4 - R_5 = 1.2 - 0.032H - 0.004O + 0.005R - 0.069C + 0.032T + 0.014G$$

In this equation concrete percentage has the largest coefficient, with all other large changes (apart from house percentage) being negligible or in the opposite direction. As each house has associated with it concrete driveways and footpaths, the house and concrete combined percentage might be considered as a *residential percentage*, with the difference in reflectance between bands 4 and 5 acting as a significant measure. Normally these two bands are considered so highly correlated that little information is contained in them, although the yellowness axis of the Kauth-Thomas transform derived for rural (mainly agricultural) areas also correlates well with this difference. The sum of these two bands should also positively correlate with concrete percentage and negatively correlate with tree percentage.

The other two highly correlated bands are band 6 and band 7. Their difference also produced interesting results and is given by

$$R_7 - R_6 = 8.9 - 0.034H - 0.086\ 0 - 0.093R - 0.046C + 0.051T + 0.058G$$

Although not as effective as the band 7 minus band 5 relationship the opposite effect of vegetated and non-vegetated surfaces is equally apparent and could be considered as an additional predictor variable for determining their respective cover percentages.

Care must be taken in using equations derived from Table 6.10 because the probability of seemingly insignificant coefficients having an effect will depend on the variance of their cover percentages.

6.5 Summary and Conclusions

All of the models proved to be very significant when tested using multiple linear regression analysis, however the convolved percentage cover model had the highest overall correlation with reflectance in each of the four bands. For all models the forced origin regression procedure gave more significant estimates of the coefficients, which allowed for more reliable estimates of the reflectance of individual cover types. Further, because the form of the linear function connecting all cover variables did not affect the solution of the normal equations when the regression was forced through the origin, all cover variables could be entered into the regression. For these reasons the convolved percentage cover model using the forced origin regression procedure is to be preferred when predicting individual cover reflectance.

The calculated reflectances of grass and trees, in each band, indicated a signature typical of vegetated surfaces. The individual grass and tree signatures were of the same general form but could be distinguished by the level of brightness in each band, which was also the case for the surface cover types of road and other, and house and concrete. A comparison of reflectance values with two external sources showed very similar results and where differences did occur these were satisfactorily explained.

It was found that equations relating reflectance to changes in surface cover percentages from their average amounts were more explanatory than those using total cover percentages and gave insight into the combined variables needed to predict individual cover percentages. Predictably the (band 7 - band 5) difference was closely related to percentage of vegetative cover, however the possible significance of bands 4 or 5 or their sum for determining tree percentage, band 7 for determining grass percentage and the difference of the infrared bands for determining vegetative content, were unexpected results. The (band 4 - band 5) difference and (band 4 + band 5) sum were suggested as possible predictors for concrete percentage, however this tentative conclusion was placed in some doubt by the difference in 100% concrete reflectance with that indicated by BAILEY (1979). Nevertheless this band difference and sum and their relationships to concrete percentage and also house plus concrete percentage were considered to be worth further examination.

The work described in this chapter was extended by testing the various ratios and transformations discussed in Chapter 4 for their ability to predict individual cover percentages and by examining the suggested combined cover percentages of vegetation, urban residential and urban non-residential surfaces. These results are given in Chapter 7.

7. DERIVATION OF PREDICTIVE EQUATIONS FOR COVER PERCENTAGES FROM REFLECTANCE DATA

7.1 Introduction

In Chapter 4 various ratios and transformations were examined. This chapter describes the application of these variables to the prediction of cover percentages. Initially the reflectance and cover data of the atmospherically corrected only model, the deconvolved reflectance model and the convolved cover model were tested via regression analysis for multiple correlation. The most predictive model was subsequently used to determine the coefficients of the most significant ratio and transformation variables. For each cover variable the results are compared. Conclusions are reached as to the optimum combination of predictive variables on the basis of level of correlation and the number of variables required.

The underlying structure of the reflectance data is described using results obtained from factor analysis (essentially equivalent to principal component analysis). Various factors and rotated factors were determined and correlated with cover data to give them meaning.

This chapter also describes the application of multiple regression analysis to the data from extended areas, including textural data. Derived equations relate cover percentages and a measure of housing density to the independent reflectance variables.

7.2 Definitions of Variables

This section defines all the cover variables, reflectance ratios and transformations, and textural variables that are used in subsequent sections.

H = House percentage per pixel
O = Other building percentage per pixel
R = Road percentage per pixel
C = Concrete percentage per pixel
T = Tree percentage per pixel
G = Grass percentage per pixel
GREEN = T + G
NONRESID = R + O
RESID = H + C

NONGREEN = H + C + R + O

Ri = R_i, pixel reflectance in bands i = 4 to 7

Dij = Ri - Rj

Sij = Ri + Rj

Rij = $\frac{Ri}{Rj}$

Rij2 = $(Rij)^2$

Vij = $(Ri^2 + Rj^2)^{\frac{1}{2}}$

Vij2 = $(Vij)^2$

RVij = (Rij)(Vij)

DSij = $\frac{Ri - Rj}{Ri + Rj}$

DSij2 = $(DSij)^2$

DSVij = (DSij)(Vij)

S = Σ Ri

RiS = $\frac{Ri}{S}$

ViS = $(Ri^2 + S^2)^{\frac{1}{2}}$

ViS2 = $(ViS)^2$

RViS = (RiS)(ViS)

Fi = F_i, unrotated factor i, i = 1 to 4

FRi = FR_i, rotated factor i, i = 1 to 4

Ai = A_i, Average reflectance over ground truth area, bands i = 4 to 7

SDi = SD_i, standard deviation of reflectance over ground truth area in bands i = 4 to 7

SDi2 = $(SDi)^2$

SUMSD = Σ SDi

SDvr = (SD4 + SD5) - (SD6 + SD7)

Ha to Ga = Average cover percentage over ground truth area, house to grass.

AGREEN = Ta + Ga

ARESID = Ha + Ca

ANONRESID = Oa + Ra

ANONGREEN = Ha + Ca + Ra + Oa

Na = Average number of houses per pixel over ground truth area.

7.3 Comparison of Predictive Models

Cover percentage of house, other, road, concrete, tree and grass were predicted using reflectance only, band difference and band sum, band on band ratios (with polynomials) and normalised ratios with polynomials. Each of these was determined for the atmospherically corrected reflectance model, the deconvolved reflectance model and the convolved cover model. The variables entered into the regression were restricted to four so that the degrees of freedom would be equivalent, allowing a more genuine comparison of their level of explanation. Multiple correlation (r), explanation ($\overline{r^2}$), F - ratio and equation standard error are shown for each of the covers and models in Tables 7.1 to 7.5.

In all cases the convolved cover percentage model has larger correlation coefficients for each band than the other two models. In general the atmospherically corrected only model has larger correlation coefficients than the deconvolved reflectance model, however the results for these two models are comparable for the cover types, house, road and concrete.

Of the ratios and transformations used, the band on band reflectance ratios in particular and the normalised and the difference on sum reflectance ratios generally, are shown to be more predictive for all three models. All the regressions using reflectances, reflectance ratios and transformations were however highly significant, the F-ratios being well within four times the 1% confidence limits for each model.

Conclusions based on the results of this section may be stated as follows. The most significant cover predictions are achieved using the convolved cover model and ratio type transformations, indicating the superiority of this combination. The better results for the various ratios, for all models, follows from the study made in Chapter 4 where these were considered to reduce noise and brightness effects, so reducing the variance of the data. The lower correlations for the deconvolved response model again, as in Chapter 6, indicate the enhanced effects of noise. The comparable results for house, road and concrete prediction using the atmospherically corrected only model suggests however that the noise effects are not as significant when dealing with linear or grouped cover where rapid changes in reflectance, significantly greater than noise, occur across boundaries. Because of the significantly better results

TABLE 7.1

Regression statistics of equations predicting cover percentage using reflectance only as the dependent variables

(a) Atmospherically Corrected Data

	House	Other	Road	Concrete	Tree	Grass
r	0.36	0.42	0.44	0.33	0.59	0.52
$\bar{r}^2$	0.13	0.17	0.19	0.11	0.34	0.27
F-ratio	102	294	225	86	490	354
S.E.	11.7%	8.1%	19.6%	4.9%	13.5%	14.7%

(b) Deconvolved Reflectance Data

	House	Other	Road	Concrete	Tree	Grass
r	0.34	0.21	0.46	0.29	0.49	0.43
$\bar{r}^2$	0.12	0.04	0.20	0.09	0.24	0.19
F-ratio	83	29	104	60	135	146
S.E.	11.4%	6.8%	9.9%	4.9%	12.1%	14.1%

(c) Convolved Cover Data

	House	Other	Road	Concrete	Tree	Grass
r	0.55	0.49	0.62	0.39	0.68	0.57
$\bar{r}^2$	0.30	0.23	0.39	0.15	0.46	0.33
F-ratio	137	129	201	58	357	206
S.E.	7.3%	4.4%	5.3%	3.1%	8.1%	10.0%

TABLE 7.2

Regression statistics of equations predicting cover percentage using band difference and band sum reflectance as the independent variables

(a) Atmospherically Corrected Data						
	House	Other	Road	Concrete	Tree	Grass
r	0.36	0.42	0.44	0.33	0.59	0.52
$\bar{r}^2$	0.13	0.17	0.19	0.11	0.34	0.27
F-ratio	135	294	330	114	736	266
S.E.	11.7%	8.1%	10.8%	4.9%	13.5%	14.7%
(b) Deconvolved Reflectance Data						
	House	Other	Road	Concrete	Tree	Grass
r	0.34	0.22	0.45	0.30	0.49	0.44
$\bar{r}^2$	0.12	0.05	0.20	0.09	0.24	0.19
F-ratio	83	33	157	40	202	148
S.E.	11.4%	6.8%	9.9%	4.9%	12.1%	14.1%
(c) Convolved Cover Data						
	House	Other	Road	Concrete	Tree	Grass
r	0.55	0.49	0.62	0.39	0.68	0.57
$\bar{r}^2$	0.30	0.23	0.39	0.15	0.46	0.33
F-ratio	182	129	268	74	536	207
S.E.	7.3%	4.4%	5.3%	3.1%	8.1%	10.0%

TABLE 7.3

Regression statistics of equations for predicting cover percentage using band on band reflectance ratios as the independent variables

(a) Atmospherically Corrected Data						
	House	Other	Road	Concrete	Tree	Grass
r	0.40	0.45	0.46	0.34	0.67	0.55
$\bar{r}^2$	0.16	0.20	0.21	0.11	0.45	0.30
F-ratio	137	176	184	92	580	307
S.E.	11.4%	8.0%	10.7%	4.9%	12.4%	14.4%
(b) Deconvolved Reflectance Data						
	House	Other	Road	Concrete	Tree	Grass
r	0.38	0.29	0.45	0.31	0.56	0.46
$\bar{r}^2$	0.14	0.08	0.20	0.09	0.31	0.21
F-ratio	51	39	81	65	144	83
S.E.	11.2%	6.7%	9.9.%	4.9%	11.5%	14.0%
(c) Convolved Cover						
	House	Other	Road	Concrete	Tree	Grass
r	0.57	0.53	0.65	0.42	0.73	0.61
$\bar{r}^2$	0.32	0.28	0.42	0.17	0.53	0.37
F-ratio	150	251	302	67	355	245
S.E.	7.2%	4.2%	5.2%	3.1%	7.6%	9.7%

TABLE 7.4

Regression statistics of equations for predicting cover percentage using band difference on sum reflectance ratios as the independent variables

(a) Atmospherically Corrected Data						
	House	Other	Road	Concrete	Tree	Grass
r	0.38	0.42	0.45	0.33	0.67	0.55
$\bar{r}^2$	0.15	0.18	0.20	0.11	0.45	0.30
F-ratio	120	149	181	88	566	300
S.E.	11.5%	8.1%	10.7%	4.9%	12.4%	14.5%
(b) Deconvolved Reflectance Data						
	House	Other	Road	Concrete	Tree	Grass
r	0.37	0.30	0.45	0.29	0.54	0.46
$\bar{r}^2$	0.13	0.09	0.20	0.08	0.29	0.21
F-ratio	97	32	106	59	172	83
S.E.	11.3%	6.6%	9.9%	4.9%	11.7%	13.9%
(c) Convolved cover Data						
	House	Other	Road	Concrete	Tree	Grass
r	0.58	0.53	0.65	0.40	0.73	0.61
$\bar{r}^2$	0.33	0.28	0.42	0.16	0.54	0.37
F-ratio	158	167	228	60	368	189
S.E.	7.1%	4.2%	5.2%	3.1%	7.5%	9.7%

TABLE 7.5

Regression statistics of equations for predicting cover percentage using normalised reflectance ratios as the independent variables

(a) Atmospherically Corrected Data

	House	Other	Road	Concrete	Tree	Grass
r	0.38	0.45	0.47	0.34	0.68	0.54
$\bar{r}^2$	0.15	0.20	0.22	0.11	0.46	0.30
F-ratio	121	176	194	89	586	295
S.E.	11.5%	8.0%	10.6%	4.9%	12.3%	14.5%

(b) Deconvolved Reflectance Data

	House	Other	Road	Concrete	Tree	Grass
r	0.36	0.27	0.45	0.29	0.56	0.45
$\bar{r}^2$	0.13	0.07	0.20	0.09	0.31	0.20
F-ratio	95	26	159	60	140	104
S.E.	11.3%	6.7%	9.9%	4.9%	11.5%	14.0%

(c) Convolved Cover Data

	House	Other	Road	Concrete	Tree	Grass
r	0.57	0.53	0.65	0.41	0.70	0.61
$\bar{r}^2$	0.32	0.28	0.42	0.17	0.49	0.37
F-ratio	298	125	300	65	300	188
S.E.	7.2%	4.3%	5.2%	3.1%	7.9%	9.7%

using the convolved cover data, only this model was considered further. It should be noted that cover percentages predicted from the convolved cover model represent a weighted value of the target pixel and adjoining pixels and so can be considered as a smoothing effect.

7.4 More Detailed Examination of Cover Variable Prediction Using the Convolved Cover Model

Significant variables of reflectance only, band difference, band sum reflectance, band on band reflectance ratios, band difference on sum reflectance ratios and normalised reflectance ratios are shown in Tables 7.6 to 7.10 for the convolved cover model data. Significant coefficients were again limited to a maximum of four as in section 7.3. While in some cases additional significant coefficients could have been extracted only a marginal increase in the multiple correlation coefficient would have resulted due to variable intercorrelation. Various statistics are given in the tables and the order of entry of each variable into the regression is shown in roman numerals after the coefficient. Little difference in predictability separated the various ratios for each of the dependent cover variables although all were better predictors than the reflectance only and band difference, band sum reflectance variables.

House cover percentage is predictable mainly from the difference in reflectance between bands 5 and 7, indicated by R5 and R7 having opposite signs and being the first two variables to enter the equation, and the primary influence of D57, R57 and DS57 in their respective equations. Brightness type variables are not influential in the prediction as indicated by the equal number of positive and negative signs in the reflectance variables, the opposite signs of S56 and S46, the lower significance of V452 and V45, and the absence of any significant V variables in the normalised reflectance regression.

As suggested in the previous chapter the prediction of concrete percentage is strongly influenced by the sum and difference of reflectance in bands 4 and 5 and this is supported by R4 and R5 having opposite signs, S45 and D45 being the two most significant variables in their respective equation, and RV54 and V45 being the primary variables in their respective equations. It should be noted that the product variable RV54, being the product of the ratio of reflectances 5 and 4 and the polar length coordinate V54, is a measure of the

TABLE 7.6

Coefficients for predicting residential cover percentages from the convolved cover model using reflectance data only as the independent variables.

Variables	House	Other	Road	Concrete	Tree	Grass
R4	-0.88 (iii)	+0.35 (iii)	+0.73 (ii)	-0.18 (iv)	-1.90 (ii)	+1.60 (iii)
(S.E.)	(0.22)	(0.13)	(0.16)	(0.09)	(0.24)	(0.30)
R5	+1.79 (i)	+0.30 (ii)	+0.45 (iii)	+0.51 (i)	-1.55 (i)	-1.49 (ii)
(S.E.)	(0.16)	(0.09)	(0.12)	(0.07)	(0.17)	(0.21)
R6	+0.61 (iv)		-0.34 (iv)	+0.26 (iii)		
(S.E.)	(0.18)		(0.14)	(0.08)		
R7	-0.87 (ii)	-0.43 (i)	-0.63 (i)	-0.18 (ii)	+0.37 (iii)	+1.50 (i)
(S.E.)	(0.11)	(0.03)	(0.08)	(0.05)	(0.05)	(0.07)
Constant	+27.5	+9.9	+25.8	+0.2	+50.5	-15.2
(S.E.)	(2.4)	(1.4)	(1.7)	(1.0)	(2.6)	(3.3)
Correlation (r)	0.55	0.49	0.62	0.39	0.68	0.57
Explanation (r^2)	0.30	0.23	0.39	0.15	0.46	0.33
F-ratio	137	129	201	58	357	206
S.E.	7.3%	4.4%	5.3%	3.1%	8.1%	10.0%

TABLE 7.7

Coefficients for predicting residential cover percentages from the convolved cover model using band difference and band sum reflectances as the independent variables

Variables	House	Other	Road	Concrete	Tree	Grass
D45				-0.35 (ii)		
(S.E.)				(0.08)		
D46			+0.54 (ii)			
(S.E.)			(0.11)			
D47					-0.37 (ii)	+0.80 (iii)
(S.E.)					(0.05)	(0.25)
D56						
(S.E.)						
D57	+0.72 (i)	+0.54 (i)	+0.56 (i)			-2.30 (i)
(S.E.)	(0.05)	(0.04)	(0.06)			(0.24)
D67				+0.12 (iii)		
(S.E.)				(0.04)		
S45				+0.19 (i)	-1.54 (i)	+0.80 (ii)
(S.E.)				(0.03)	(0.06)	(0.09)
S46	-0.78 (iii)		0.10 (iii)			
(S.E.)	(0.20)		(0.05)			
S47		+0.35 (ii)				
(S.E.)		(0.13)				
S56	+1.13 (ii)					
(S.E.)	(0.18)					
S57		-0.24 (iii)				
(S.E.)		(0.11)				
S67						
(S.E.)						
Constant	+26.5	+9.9	+25.8	+1.0	+50.6	-15.2
(S.E.)	(2.3)	(1.4)	(1.7)	(0.9)	(2.3)	(3.3)
r	0.55	0.49	0.62	0.39	0.68	0.57
r^2	0.30	0.23	0.39	0.15	0.46	0.33
F-ratio	182	129	268	74	536	207
S.E.	7.3%	4.4%	5.3%	3.1%	8.1%	10.0%

TABLE 7.8

Coefficients for predicting residential cover percentages from the convolved cover model using band on band reflectance ratios as the independent variables.

	House	Other	Road	Concrete	Tree	Grass
R47			+42.8 (i)			
(S.E.)			(2.4)			
R57	+93.7 (i)					
(S.E.)	(10.8)					
R65			-1.42 (iii)			
(S.E.)			(0.53)			
R75					+26.8 (i)	
(S.E.)					(2.6)	
RV45			-0.394 (ii)			
(S.E.)			(0.067)			
RV54				+0.424 (i)		
(S.E.)				(0.037)		
RV75					-0.581 (ii)	
(S.E.)					(0.072)	
RV76						+0.648 (i)
(S.E.)						(0.051)
R462	-16.2 (ii)					
(S.E.)	(3.0)					
R472		+10.9 (ii)				
(S.E.)		(2.9)				
R572	-31.4 (iii)	+13.2 (i)		-10.5 (iii)		-44.4 (ii)
(S.E.)	(10.0)	(2.9)		(1.7)		(4.2)
R742					+1.93 (iii)	-1.93 (iii)
(S.E.)					(0.25)	(0.20)
R762				-1.83 (ii)		
(S.E.)				(0.45)		
V452	-0.0083 (iv)					
(S.E.)	(0.0031)					
V572					+0.0094 (iv)	
(S.E.)					(0.0026)	
V672				-0.0017 (iv)		
(S.E.)				(0.0004)		

TABLE 7.8 contd.

	House	Other	Road	Concrete	Tree	Grass
Constant	+1.0	0.0	+5.4	+5.4	-23.4	+19.8
(S.E.)	(2.2)	(0.3)	(2.1)	(1.2)	(3.3)	(3.3)
r	0.57	0.53	0.65	0.42	0.73	0.61
$\bar{r}^2$	0.32	0.28	0.42	0.17	0.53	0.37
F-ratio	150	251	302	67	355	245
S.E.	7.2%	4.2%	5.2%	3.1%	7.6%	9.7%

TABLE 7.9

Coefficients for predicting residential cover percentages from the convolved cover model using band difference on sum reflectance ratios as the independent variables.

Variables	House	Other	Road	Concrete	Tree	Grass
DS46	-31.5 (ii)					
(S.E.)	(5.2)					
DS47		+66.7 (i)	+69.8 (i)		-71.1 (iii)	-77.8 (iv)
(S.E.)		(5.8)	(7.6)		(7.1)	(14.8)
DS57	+63.9 (i)			-9.8 (iv)		
(S.E.)	(4.3)			(2.6)		
V45	+2.67 (iv)			+0.44 (i)		
(S.E.)	(0.57)			(0.06)		
V67						+5.47 (i)
(S.E.)						(0.36)
DSV45		-0.54 (iii)		-0.69 (ii)		
(S.E.)		(0.13)		(0.12)		
DSV47						+5.05 (iii)
(S.E.)						(0.55)
DSV57					+4.86 (ii)	
(S.E.)					(0.52)	
DSV67				+0.47 (iii)		
(S.E.)				(0.10)		
V452	-0.072 (iii)					
(S.E.)	(0.015)					
V462			-0.0058 (ii)			
(S.E.)			(0.0011)			
V472					+0.0175 (iv)	
(S.E.)					(0.0034)	
V562						-0.0717 (ii)
(S.E.)						(0.0058)
DS452			-94.0 (iii)			
(S.E.)			(18.0)			
DS472		58.0 (ii)	+38.0 (iv)			
(S.E.)		(8.0)	(11.0)			
DS572					+228 (i)	
(S.E.)					(17)	

TABLE 7.9 contd.

	House	Other	Road	Concrete	Tree	Grass
Constant	+19.9	+21.7	+39.7	-4.3	-4.9	-89.3
(S.E.)	(6.3)	(1.0)	(1.3)	(1.7)	(1.7)	(7.2)
r	0.58	0.53	0.65	0.40	0.73	0.61
$\bar{r}^2$	0.33	0.28	0.42	0.16	0.54	0.37
F-ratio	158	167	228	60	368	189
S.E.	7.1%	4.2%	5.2%	3.1%	7.5%	9.7%

TABLE 7.10

Coefficients for predicting residential cover percentages from the convolved cover model using normalised reflectance ratios as the dependent variables.

Variables	House	Other	Road	Concrete	Tree	Grass
R4S					-186 (ii)	+129 (iv)
(S.E.)					(20)	(26)
R5S	+220 (i)				-268 (iv)	+684 (iii)
(S.E.)	(11)				(45)	(79)
R6S		-579 (iii)		+30.4 (ii)		
(S.E.)		(263)		(9.6)		
R7S		-331 (i)				
(S.E.)		(51)				
RV5S			-0.65 (iii)	+1.49 (i)	+0.83 (i)	
(S.E.)			(0.11)	(0.18)	(0.45)	
RV6S					-1.80 (iii)	
(S.E.)					(0.23)	
RV7S						+1.16 (i)
(S.E.)						(0.09)
V5S2						
(S.E.)						
V7S2				-0.0010 (iii)		
(S.E.)				(0.0002)		
R4S2			+251 (i)			
(S.E.)			(36			
R5S2			+447 (ii)	-239 (iv)		-2540 (ii)
(S.E.)			(47)	(51)		(251)
R6S2	+227 (ii)	+968 (iv)				
(S.E.)	(39)	(463)				
R7S2		+357 (ii)				
(S.E.)		(65)				
Constant	-28.3	+165.2	+3.5	-9.1	+122.2	-68.5
(S.E.)	(4.5)	(32.8)	(1.2)	(3.0)	(7.2)	(8.1)
r	0.57	0.53	0.65	0.41	0.70	0.61
$\bar{r}^2$	0.32	0.28	0.42	0.17	0.49	0.37
F-ratio	298	125	300	65	300	188
S.E.	7.2%	4.3%	5.2%	3.1%	7.9%	9.7%

difference between bands 5 and 4 as indicated by figure 7.1. When comparing the spectral signature of house and concrete in Chapter 6 their parallel nature was noted, this is also indicated by the signs of the reflectance-only coefficients being the same. Concrete percentage has somewhat lower coefficients than house percentage due to the higher reflectance of concrete in all bands.

In the same manner as house percentage, other percentage is primarily predicted by a difference between the visible and infrared reflectances, R4 and R5 having positive coefficients and R7 a negative coefficient. Additionally D57, R572 and DS47 are the first variables to enter their respective equations. Little influence of brightness type variables is seen, S47 and S57 having opposite signs in the band difference, band sum equation with no other brightness type variables being significant.

The difference in visible and infrared reflectance primarily contributes to the prediction of road percentage. R4 and R5 have opposite signs to R6 and R7, D46 and D57 are positive and are the first two variables to enter the band difference, band sum equation, both R47 and R65 are significantly represented and DS47 is a primary entering variable, all indicating the visible infrared difference. Brightness variables are marginally significant and suggest that the average low reflectance of road, represented by V462 and S46, contributes to the prediction of percentage cover. For reflectance only the equivalence of the coefficient signs between road and other percentage support the parallel nature of their spectral signatures as suggested in Chapter 6 (apart from, R6 which is insignificant for other percentage). Because road surface has a lower reflectance than other reflectance across all bands, the magnitude of the reflectance coefficients are greater for road percentage.

Predictably, grass percentage is primarily determined by the visible/infrared difference although having a sign reversal when compared to other non-vegetative surfaces,R7 (positive) and R5 (negative) entering the reflectance only model as the first two variables. In other regression equations D57, R572 are significant variables. However in the regression using difference on sum reflectance ratios the primary variable is V67, representing a brightness measure in the infrared and RV76 is a primary variable in the band on band reflectance ratios. The product variable, RV76, essentially represents an infrared difference. This latter result was predicted in Chapter 6 where it was suggested that the difference in

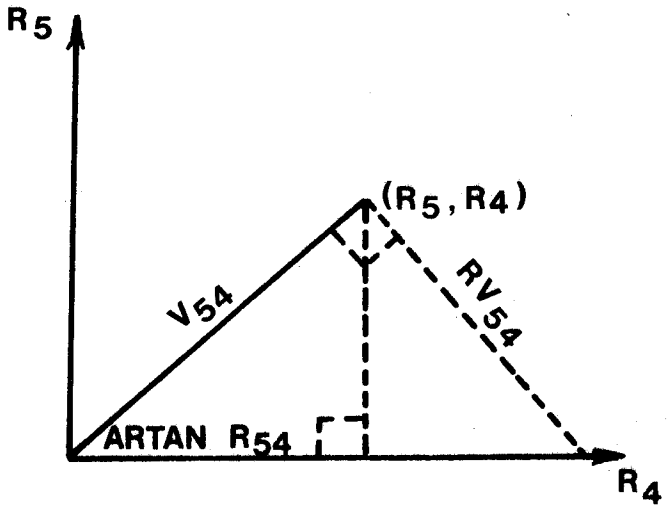


FIGURE 7.1: Geometric relationship between V54, R54 and RV54.

the infrared band reflectances may act as an additional predictor of vegetative content. The emergence of infrared brightness as a grass predictor variable had not been suggested previously. Such a result would seem reasonable given the high reflectance of grass in the infrared bands, however the two secondary coefficients to V67 both reinforce a visible/infrared difference.

The suggested influence of the sum of the visible band reflectances on tree percentage prediction is borne out by the reinforcing negative effect of R4 and R5 and the primary influence of S45. In the ratio type variables the typical visible/infrared difference also plays a primary predictive role as indicated by the primary effect of R75 and DS572, and the secondary effects of DSV57 and RV75. Of all cover variables, tree percentage cover was most significantly predicted, which is probably due to the low visible reflectance of tree cover and the substantial difference between its visible and infrared reflectance. Such a result may not be typical of all tree cover. Most trees in the Sydney area are eucalyptus, which have a particularly dark foliage while still retaining the high,typical vegetative response, in the infrared.

In the previous discussion of influential predictor variablesthe normalised reflectance ratios, polynomials and products, were only mentioned on one occasion. The addition of another spectral band i.e. 0.5 to 1.1 μm, which is dependent on the other four bands, causes these variables to have somewhat different effects to the equivalent variables using only bands 4 to 7. As a general rule the most significant normalised reflectance variable for predicting a particular cover percentage was found to be a variable that includes the reflectance in the band where the greatest difference between the individual total cover reflectance and the average reflectance for that band occurs. The only exception was the primary predictor variable for road cover, which includes the reflectance in a band with a very small difference from the average in that band. The average reflectance for each band over all ground truth areas was calculated as 14.4%, 13.2%, 23.8% and 32.9% for bands 4, 5, 6 and 7 respectively, and in band 4 road reflectance is only 1.5% higher than the average. In this case the percentage of cover is essentially being predicted by the variation in the summed denominator, which for 100% road cover has its lowest value. The R4 numerator will vary least of all because the road reflectance is close to the average

in band 4 and the range of reflectance is least with a standard deviation of 1.8 compared to the next lowest in band 5 of 2.5. Therefore the numerator will have a small variation while the denominator has its maximum variation from 0 to 100% road, making the ratio R45 (or R452) the most significant predictor. Concrete cover which has the highest sum of reflectance for 100% cover also displays somewhat unique results. Here the brightness variable V5S2 is a significant secondary predictor, whereas the equations for the other cover variables do not include brightness type variables.

While one criterion for choosing a variable set to predict cover percentages must be the level of prediction, a further criterion should be the number of variables required for that prediction, given the large data sets and the need to limit computations. The multiple correlation coefficients of each of the regression equations were compared as each significant variable entered the equation. The improved multiple correlation coefficient (r) for each additional variable, for each ratio and transformation variable set and for each cover variable are shown in Tables 7.11 to 7.16. As each primary variable involved a computation with at least two reflectance variables, except in the case of reflectance only, this latter equation was compared only on the basis of its multiple correlation after the entry of two reflectance variables.

On examination of these tables it can be seen that the band on band ratios are generally superior to the other transformations, having in most cases higher multiple correlations on the entry of the primary predictor variable and the final variable. One particular exception in this regard is the prediction of road percentage cover, where r is relatively low on the entry of the first variable, having a value of 0.40 compared to the primary difference on sum variable which has a high of 0.63. Nevertheless on the entry of the second variable the position is reversed, the band on band ratio variables then having a slightly increased r value. In the case of road and concrete percentage cover prediction the second level reflectance only model gives comparable results, and in the case of grass cover gives better results than the other models at the primary variable entry level. While the band on band ratios are superior in most cases, this superiority is only marginal and was not considered sufficient to give unreserved support for their adoption.

TABLE 7.11

Multiple correlation coefficient for predicting House cover for each variable entered into the regression equation.

Step Entered	Reflectance Only	Diff. and Sum	Band Ratios	Diff. on Sum	Normalised
1	0.46	0.50	0.54	0.55	0.55
2	0.54	0.54	0.56	0.57	0.57
3	0.55	0.55	0.57	0.57	
4	0.55	0.55	0.57	0.58	
5			0.58	0.58	
6			0.59	0.59	
7					

TABLE 7.12

Multiple correlation coefficient for predicting Other cover for each variable entered into the regression equation.

Step Entered	Reflectance Only	Diff. and Sum	Band Ratios	Diff. on Sum	Normalised
1	0.40	0.48	0.53	0.50	0.49
2	0.48	0.48	0.54	0.53	0.53
3	0.49	0.49		0.53	0.53
4					0.53

TABLE 7.13

Multiple correlation coefficients for predicting Road cover for each variable entered into the regression equation.

Step Entered	Reflectance Only	Diff. and Sum	Band Ratios	Diff. on Sum	Normalised
1	0.54	0.62	0.40	0.63	0.61
2	0.62	0.62	0.65	0.64	0.63
3	0.62	0.62	0.65	0.64	0.65
4	0.63			0.65	

TABLE 7.14

Multiple correlation coefficient for predicting Concrete cover for each variable entered into the regression equation.

Step Entered	Reflectance Only	Diff. and Sum	Band Ratios	Diff. on Sum	Normalised
1	0.14	0.36	0.38	0.36	0.38
2	0.38	0.38	0.39	0.38	0.38
3	0.39	0.39	0.40	0.39	0.40
4	0.39	0.39	0.42	0.40	0.41
5			0.43	0.42	
6			0.43		

TABLE 7.15

Multiple correlation coefficients for predicting Tree cover for each variable entered into the regression equation.

Step Entered	Reflectance Only	Diff. and Sum	Band Ratios	Diff. on Sum	Normalised
1	0.65	0.66	0.68	0.67	0.65
2	0.66	0.68	0.71	0.71	0.68
3	0.68		0.73		0.69
4			0.73		0.70
5			0.73		0.72
6					0.73
7					0.74

TABLE 7.16

Multiple correlation coefficients for predicting Grass cover for each variable entered into the regression equation.

Step Entered	Reflectance Only	Diff. and Sum	Band Ratios	Diff. on Sum	Normalised
1	0.55	0.54	0.55	0.53	0.55
2	0.56	0.57	0.57	0.57	0.57
3	0.58	0.58	0.61	0.60	0.60
4			0.61	0.61	0.61
5				0.62	0.62

It can be seen from Tables 7.11 to 7.16 that in all cases one variable primarily contributes to the total multiple correlation. On the basis of significant correlation and simplicity of computation, it was considered that the following equations could be used in future work to adequately predict percentage cover from one variable.

$$H = 43.2 + 44.2\ DS\ 57 \quad (7.1)$$

$$O = 0.5 + 23.3\ R572 \quad (7.2)$$

$$R = 32.1 + 48.0\ DS47 \quad (7.3)$$

$$C = -3.1 + 0.29\ S45 \quad (7.4)$$

$$T = -11.9 + 10.3\ R75 \quad (7.5)$$

$$G = -14.0 + 1.56\ R7 \quad (7.6)$$

Where more accurate results are required the following equations derived from Tables 7.6 to 7.10 and using a maximum of four significant variables only, were found to be the most highly predictive for each cover variable.

$$H = 19.9 - 31.5\ DS46 + 63.9\ DS57 + 2.67\ V45 - 0.072\ V452 \quad (7.7)$$

$$O = 10.9\ R472 + 13.2\ R572 \quad (7.8)$$

$$R = 39.7 + 69.8\ DS47 - 0.0058\ V462 - 94.0\ DS452 + 38.0\ DS472 \quad (7.9)$$

$$C = 5.4 + 0.424\ \ RV54 - 10.5\ R572 - 1.83\ R762 - 0.0017\ V67 \quad (7.10)$$

$$T = -23.4 + 26.8\ R75 - 0.581\ RV75 + 1.93\ R742 + 0.0094\ V572 \quad (7.11)$$

$$G = -89.3 - 77.8\ DS47 + 5.47\ V67 + 5.05\ DSV47 - 0.0717\ V562 \quad (7.12)$$

Results published by RICHARDSON and WIEGAND (1977) show a range of correlation coefficient values when predicting crop and shadow cover from various single transformed response variables. These ranged from 0.564 to 0.716 for crop cover, and from 0.324 to 0.466 for shadow cover. Untransformed response was correlated more highly in both cover cases having $r = -0.809$ for crop cover and band 5, and $r = -0.518$ for shadow cover and band 5. This is the reverse of the situation found in the present study. The levels of correlation are comparable in both studies, however the data used by RICHARDSON et al (1977) were from extended fields which suggests

that the correlations achieved for intrapixel cover prediction are in comparison, very good.

7.5 Examination of Reflectances Using Factor Analysis

The reflectance data of the convolved percentage cover model was subjected to a factor analysis without rotation, which is equivalent to a principal components analysis when the communalities are set to one i.e. all variables are retained in the analysis. The factor score coefficient matrix was determined, with each factor nomalised to one. These coefficients represent the weights to be given to each variable so that composite scales can be built that represent the theoretical dimensions associated with the respective factors. The coefficient matrix, with the respective percentage of variance contributed by each factor, is shown in Table 7.17.

TABLE 7.17

Unrotated normalised factor score coefficient matrix for reflectance data

	F1	F2	F3	F4
R4	0.565	-0.397	0.723	-0.018
R5	0.548	-0.430	-0.655	0.295
R6	0.529	0.473	-0.176	-0.683
R7	0.318	0.659	-0.135	-0.668
% Variance	50.7	43.3	3.8	2.2
Cumulative %	50.7	94.0	97.8	100.0

Although not directly comparable to the results given in section 4.6, the basic band sum, visible/infrared difference, visible difference and infrared difference are immediately apparent. AUSTIN et al (1978) indicated that their first principal component represented approximately 90% ofthe variance and a similar proportion was determined by DONKER et al (1976). This represents a major difference from the results of this study where, from Table 7.17, the first factor contributes to only 50.7% of the variance while the second factor contributes 43.3%. In addition the final

two factors contribute a larger percentage of the variance, a total of 6%, compared to less than 1% for both AUSTIN et al and DONKER et al.

Each of these previous studies used data sets derived from rural areas, essentially composed of grass, trees and crops, with limited non-vegetative cover. The signatures of these various vegetative surfaces basically parallel each other, as was the case for tree and grass cover in this study, and are primarily different only on the basis of brightness, and so brightness or the sum of all band responses was the predominant contributor to variance. In an urban scene a large proportion of the cover is non-vegetative which has a different visible/infrared signature to vegetation. While brightness is still a major contributor to the variance, visible/infrared difference is also important. Because of the great diversity of an urban scene when compared to a rural scene the third and fourth factors may also be marginally significant contributors to the variance.

The correlations between the four derived factors and the various percentage cover variables were calculated to determine whether the factors represented some underlying structure of the cover data. This correlation matrix is given in Table 7.18. Immediately apparent is that non-vegetative surfaces primarily correlate with the second factor, while the vegetative surfaces, particularly tree cover, are relatively equally correlated with the first two factors. This led to the hypothesis that the sum of the vegetative surfaces would be positively correlated with F2. The correlations between GREEN (where GREEN = T + G) and all factors were calculated as follows

	F1	F2	F3	F4
GREEN	-0.19	0.76	0.15	0.09

These results confirmed the hypothesis. It followed therefore that F2 represented a continuum from 100% GREEN to 100% NONGREEN.

TABLE 7.18

Correlation matrix of unrotated factors and cover variables

	F1	F2	F3	F4
House	+0.20	-0.46	-0.22	-0.08
Other	+0.05	-0.48	-0.02	-0.04
Road	+0.01	-0.62	0.00	-0.04
Concrete	+0.27	-0.24	-0.15	-0.05
Tree	-0.47	+0.49	+0.01	+0.02
Grass	+0.19	+0.51	+0.18	+0.10

Of the non-vegetative surfaces both house and concrete percentage show a small but significantly better correlation with F1 and F3 than the other non-vegetative surfaces. This follows from their higher overall brightness, essentially represented by F1 and their previously discussed correlation with the difference of the visible bands essentially represented by F3. The sum of these two surfaces was suggested, in Chapter 6, as a possible comparable axis of differentiation to yellowness of the Kauth-Thomas transformation. A new variable RESID defined as being equal to (H + C) was tested for correlation with the four factors with the following results -

	F1	F2	F3	F4
RESID	0.27	-0.48	-0.24	-0.08

The summation of these two cover variables reinforced their negative correlation with F2, which effectively discounted the possibility of RESID acting as a completely separate orthogonal dimension in the same manner as yellowness.

The negative and positive correlation of tree and grass cover respectively with F1 and their positive correlation with F2 suggested that these two variables were essentially orthogonal. A low correlation of -0.13 between these two variables confirmed this. It followed therefore that rotated factors could represent a tree and grass dimension. An orthogonal varimax rotation was applied to the reflectance data. This method of rotation is most widely used (NIE et al, 1975) and defines a simple factor as one with only ones and zeros in the column. The rotated matrix of the normalised factor scores is given in Table 7.19. It should be noted that the rotated R1 will no longer contribute to the maximum variance. Rotated factors are given the symbol FRi, i = 1 to 4.

A quite different data structure resulted. As can be seen from Table 7.19 the first factor weighs heavily on the sum of the visible reflectance or visible brightness, the second factor on the sum of the infrared or infrared brightness, while the third and fourth factors represent a visible and an infrared difference as was found for the unrotated factors. The correlations between all cover variables and GREEN, NONGREEN and RESID and rotated factors FR1, FR2, FR3 and FR4 were calculated. The correlation matrix is given in Table 7.20.

TABLE 7.19

Varimax rotated normalised factor score coefficient matrix for reflectance data.

	FR1	FR2	FR3	FR4
R4	0.727	-0.036	-0.698	-0.080
R5	0.663	-0.002	0.714	-0.173
R6	-0.151	0.638	-0.017	0.707
R7	0.100	0.769	0.046	-0.682
% variance	47.7	45.9	3.7	2.7
Cumulative %	47.7	93.6	97.3	100.0

TABLE 7.20

Correlation matrix of rotated factors with cover variables and derived cover variables

	FR1	FR2	FR3	FR4
House	0.42	-0.25	0.21	0.15
Other	0.32	-0.36	0.02	0.07
Road	0.37	-0.50	0.01	0.06
Concrete	0.34	-0.03	0.15	0.11
Tree	-0.66	0.12	-0.03	-0.08
Grass	-0.13	0.52	-0.16	-0.14
GREEN	-0.58	0.50	-0.15	-0.16
RESID	0.48	-0.23	0.23	0.17
NONGREEN	0.58	-0.48	0.16	0.16

Tree cover percentage has a correlation of -0.66 with FR1 and grass cover percentage a correlation of 0.52 with FR2. Both have low correlations with all other factors. The first factor was therefore considered to represent a tree cover dimension and the second factor a grass cover dimension. All other non-vegetative surfaces and the derived cover variables GREEN, RESID and NONGREEN are essentially equally correlated with FR1 and FR2.

The variable GREEN has correlations of -0.58 and 0.50 with FR1 and FR2 respectively. These results suggested that this total vegetative cover variable could be represented by an axis at approximately 45^{o} to the orthogonal negative direction of FR1 and the positive direction of FR2. That is ((FR2) + (-FR1)) should correlate highly with GREEN showing that this direction parallels the original F2. A correlation coefficient of 0.76 was determined which is the same as previously calculated for GREEN and F2.

The positive correlation of FR3 with house and concrete percentage as opposed to the very low correlation with road and other percentage suggested that this factor marginally represents RESID cover, although as previously mentioned cannot be considered as a separate orthogonal RESID axis. Grass cover is negatively correlated with FR3, with tree cover having a very low correlation, so that the opposite direction of this dimension could also marginally suggest amount of grass cover, however this cover is adequately represented by FR2 and so this possible dimension was not considered further. The fourth factor shows essentially the same correlations with cover as FR3, however as this factor contains primarily noise effects (DONKER et al 1976, KAUTH, 1975) it was not considered further.

The total residential cover can now be represented by two overlaid orthogonal systems each having a common dimension. The first system separates GREEN and NONGREEN cover, and the second system separates grass and trees, each having a common dimension marginally representing RESID cover.

Using the factor scores of Table 7.19 as weighting coefficients and substituting for the 100% reflectance of each cover from Table 6.7 the coordinates of each cover in the rotated FR1, FR2, system were calculated. These values are plotted in figure 7.2. It can be seen that house cover represents the reflectance of an average non-vegetative surface and that the 100% non-vegetative surface reflectances define an axis approximately

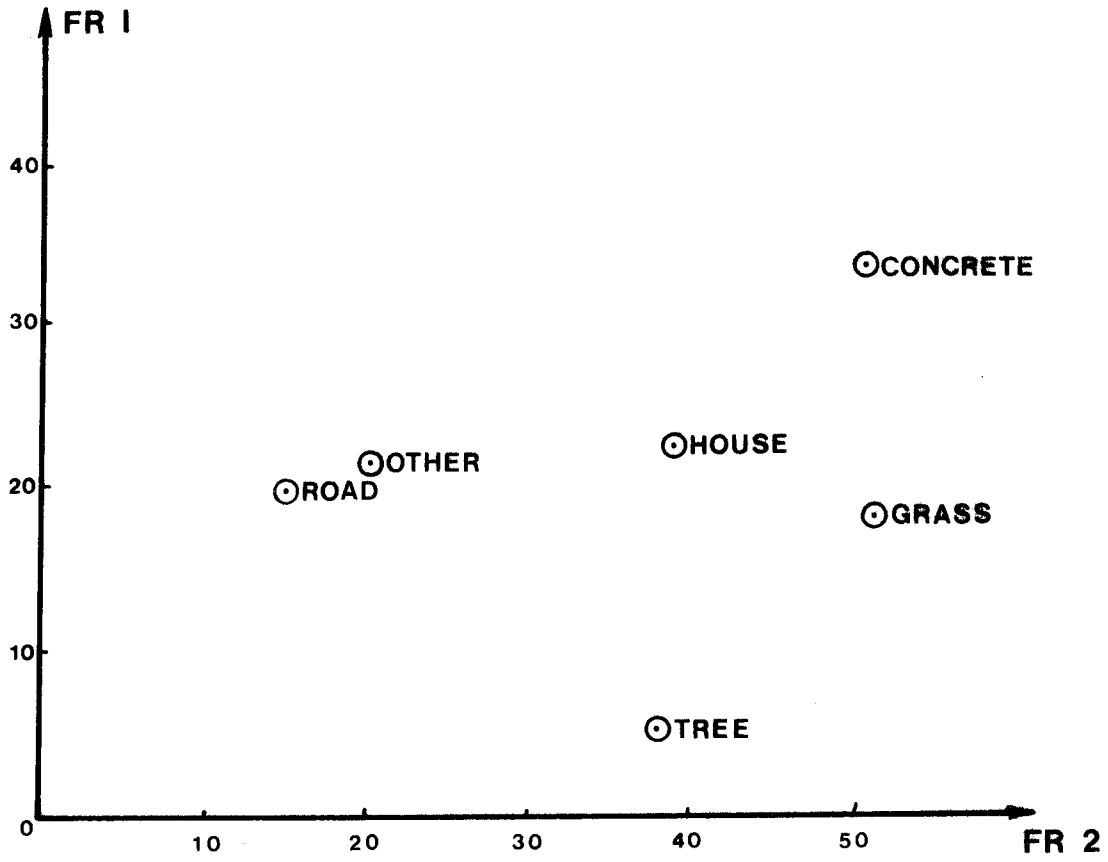


FIGURE 7.2: Location of various residential surfaces in the rotated Factor 1 and Factor 2 system.

at right angles to the GREEN dimension discussed previously and approximately parallel to a line defining the 100% vegetative surfaces. Concrete lies at one end of this axis and road and other surfaces at the opposite end. Neglecting the cover surface 'other' for simplicity, all of these relationships can be shown schematically in figure 7.3.

While this type of analysis gives insight into the underlying structure, indices derived from the factor scores will be no more predictive than the equations for individual cover derived previously, and will be less predictive when compared to the ratio type variables. For this reason a GREEN index, giving the predicted percentage of vegetative cover was determined regressing GREEN as the dependent variable against band on band reflectance ratios. This gave the following results -

$$GREEN = 72.6 - 110.2\ R57 + 13.5\ R76 + 0.452\ RV57 \qquad (7.13)$$

where

$r = 0.82$

$\overline{r}^2 = 0.67$

$S.E. = 8.8\%$

$F = 849$

This represents the most predictive and significant equation derived for single pixel areas given in this chapter.

7.6 Application to Extended Areas

In certain circumstances cover information over extended areas may be preferable to that determined in individual pixels. It would be expected that variance would be reduced by using average reflectance data so that higher multiple correlations could be achieved between average cover and average reflectance. In addition as suggested in Chapter 4 the use of textural variables may prove significant. The use of extended areas means that the effects of the sensor point spread function can be neglected and the effects of background reflectance are substantially reduced, because the average reflectance is essentially its own background.

The average reflectance in each band and the reflectance standard deviation over each ground truth area were determined. Average reflectances were transformed to their band on band ratio equivalents. These reflectance variables were regressed with various average cover variables as the dependent variables, to give the results shown in equations 7.14 to 7.19.

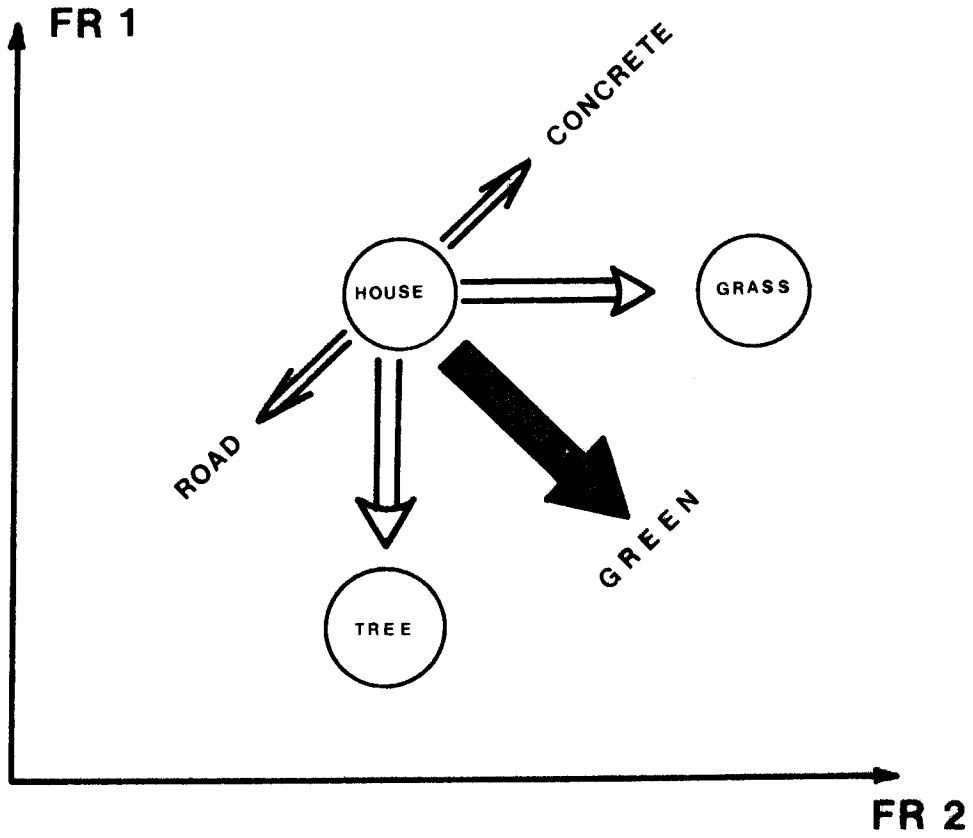


FIGURE 7.3: Schematic relationship between various urban residential cover types with respect to rotated factors FR1 and FR 2.

Statistical parameters are shown after each equation, and the order of entry of each variable is given by their order in the equation.

$$Ha = 136.6 - 57.3\ R76 - 105.5\ R46 - 0.643\ SUMSD + 82.8\ R57 \quad (7.14)$$
$$(r = 0.81,\ \bar{r}^2 = 0.63,\ S.E. = 3.9\%,\ F = 28)$$

$$Oa = -36.3 + 63.4\ R47 + 4.11\ SD72 - 11.3\ SD7 + 12.1\ R64 \quad (7.15)$$
$$(r = 0.88,\ \bar{r}^2 = 0.76,\ S.E. = 2.2\%,\ F = 50)$$

$$Ra = -9.3 + 51.5\ R47 \quad (7.16)$$
$$(r = 0.89\ \bar{r}^2 = 0.79,\ S.E. = 2.5\%,\ F = 234)$$

$$Ca = -17.7 + 24.5\ R54 \quad (7.17)$$
$$(r = 0.63,\ \bar{r}^2 = 0.39,\ S.E. = 1.7\%,\ F = 42)$$

$$Ta = 70.8 - 3.80\ RV46 + 7.20SD4 \quad (7.18)$$
$$(r = 0.91,\ \bar{r}^{\,2} = 0.82,\ S.E. = 4.2\%,\ F = 145)$$

$$Ga = -30.4 + 3.18\ RV76 - 1.60\ RV64 - 0.98\ SD52 - 0.41\ RV64 \quad (7.19)$$
$$(r = 0.81,\ \bar{r}^2 = 0.63,\ S.E. = 6.5\%,\ F = 27)$$

All equations are significant at four times the 1% level of confidence and represent excellent results. All multiple correlation coefficients are substantially greater than their single pixel equivalents. In all cases apart from road and concrete cover, the standard deviation variables are seen to be significant predictors, justifying their inclusion.

In addition to the basic cover variables a number of derived cover variables were also used as the dependent variables. These results are given in the following equations.

$$AGREEN = -315.0 + 282.1\ R67 \quad (7.20)$$
$$(r = 0.92,\ \bar{r}^2 = 0.84,\ S.E. = 5.7\%,\ F = 337)$$

$$ARESID = -77.9 + 143.5\ R54 - 2.16\ V47 + 2.18\ RV65 - 0.66\ RV74 - 0.43\ SD62 \quad (7.21)$$
$$(r = 0.85,\ \bar{r}^{\,2} = 0.70,\ S.E. = 4.2\%,\ F = 31)$$

$$\text{ANONGREEN} = -22.2 + 92.0\ R47 \qquad (7.22)$$
$$(r = 0.92,\ \bar{r}^2 = 0.84,\ \text{S.E.} = 3.7\%,\ F = 340)$$

$$Na = -5.3 + 24.0\ R57 - 0.62\ SD7 + 0.14\ RV64$$
$$-2.6\ SD4 + 0.64\ SD42 - 0.23\ RV56 \qquad (7.23)$$
$$(r = 0.91,\ \bar{r}^2 = 0.81,\ \text{S.E.} = 0.6,\ F = 45)$$

These again represent highly significant equations. Both AGREEN and ANONGREEN can now be predicted by a single variable.

An important result at this level of aggregation is that the prediction of vegetative content is no longer dependent on the visible/infrared comparison so apparent in single pixel analysis, but on the band 6, band 7 reflectance ratio. The potential influence of these two bands had been suggested in Chapter 6. The visible infrared influence on the prediction of man made materials, represented by ANONGREEN, is nevertheless still apparent. It is suggested that the strong slope of vegetation reflectance as it increases through the infrared, more critically differentiates it from man made materials. A further results of interest is that the standard deviation is no longer a significant predictor variable, except for a marginal influence on ARESID, and is due to the surfaces represented by the composite variables being more irregularly distributed throughout the ground truth area and not having the strong patterning of individual cover variables.

The equation predicting the average number of houses per pixel over the ground truth area is very significant but quite complex, as it represents a number of interacting effects. It is strongly dependent on the standard deviation in band 7 and a non-linear effect of the standard deviation in band 4, as suggested in Chapter 4. The equation can be simply modified to give the density of housing per hectare or per acre. Because it is not a cover variable, the potential to predict such a variable is an important result.

7.7 Summary and Conclusions

Of the three models examined relating cover percentage to reflectance, it was found that the convolved percentage cover model gave superior results for cover prediction, for all cover and for all reflectance ratios and transformations. The ratio type transformed reflectance variables were better predictors of cover percentage than the linearly related

reflectance variables, and of these the band on band reflectance ratios in terms of correlation and number of variables required, appeared marginally superior, although this was not conclusive. All reflectance, ratio and transformed variable regressions were however significant predictors of cover percentage. Multiple correlations achieved ranged from 0.42 for concrete cover to 0.73 for tree cover. It is concluded that significant single variable and multiple variable cover percentage predictive equations can be derived using multiple regression methods, and that a GREEN index representing percentage of total vegetative cover, with a multiple correlation coefficient of 0.82, is most reliably predicted.

The multidimensional structure of the reflectance data was found to be essentially represented by two overlaid orthogonal systems, one representing a grass/tree structure, the other a vegetative/non-vegetative structure. These systems illustrate the structure but do not improve cover prediction.

Multiple correlation coefficients were substantially improved between cover and reflectance type variables when regression analysis was used with extended data, and was due to the reduction of variance and the addition of textural variables. The ability to predict the average number of houses per pixel from reflectance data is considered an important result of the extended data analysis.

8. THE APPLICATION OF COVER RELATED VARIABLES AS SURROGATE PREDICTORS OF RESIDENTIAL QUALITY

8.1 Introduction

The direct prediction of cover percentages at the pixel level, or cover percentages and housing density at the extended area level, can obviously have direct application in many urban studies. Of greater interest is the ability to predict characteristics of residential areas seemingly unrelated directly to these cover variables. The work of LANDINI et al (1979) predicting socio-economic variables from urban cover and MURAI (1974) and MORRISON (1978) estimating population density, are good examples of surrogate variable application.

In this chapter the relationship between average house price, as one measure of residential quality, and cover related variables are examined. Initially the procedures for collecting price and other housing characteristics data are briefly mentioned. Some background on the determinants of house price is given, followed by a multiple regression analysis of price and housing characteristics data which had not been collected via remote sensing. The results of this analysis are compared with a similar analysis using cover related variables determined from aerial photographs and finally with an analysis using Landsat derived data. Multiple correlations are compared and conclusions reached as to the advantages or otherwise of the sensor related models.

The chapter concludes with a brief analysis of a total residential quality index and a discussion of other potential applications of cover data to urban residential analysis.

8.2 Determinants of House Price

In recent years the attention of researchers from many different fields including economics, geography, sociology, and planning has been directed to examining from which attributes of housing the consumer derives utility and whether these can be ranked in any meaningful way. It is argued that if one house or group of houses has more desirable attributes than another, this higher valuation by the consumer will be reflected in a higher market price. Most of these studies can be grouped into one of four areas (FORSTER, 1975).

(i) Housing characteristic models are locationally insensitive models and ignore the locational and environmental characteristics of a neighbourhood. They assume that house prices are determined solely by the characteristics of the house itself, such as number of rooms, age, type etc. Although it is obvious that where a house is located should make a considerable difference to its price, it can be a suitable approach when analysing small areas of comparable housing stock.

(ii) The most common models are those predicting house price on the basis of the trade off between housing costs and transport. These accessibility models, of which ALONSO (1964) is the most well known, predict that house prices decline with distance from the central business district and also that house prices will be higher in areas having above average accessibility. The predictions derived from these models however are fairly general and are not particularly suitable for determining the price of individual properties. Apart from this, the fact that many houses of equal accessibility have widely differing prices indicates that many other factors, in addition to accessibility influence the price.

(iii) Environmental or area preference models reflect researchers attempts to explain house price differences in terms of neighbourhood attractiveness. RICHARDSON (1971) and others have put forward the hypothesis that the environmental atrributes of an area determine the residential site choice. It follows that these attributes attract competition for houses in a particular area and their prices are increased. In a similar way unfavourable attributes tend to cause prices to be lower.

Clearly the difficulty with these models is defining the variables that measure environmental attractiveness. As GOODCHILD (1974) has shown, preference towards environmental attributes varies by socio-economic class and that which may be considered a favourable attribute by one group may not be considered as such by another. This model would also infer either, that in a given area houses are homogeneous or that differences between houses are not reflected as a difference in price. Neither of these inferences would be valid, and so while these environmental models may reflect an underlying area price, they cannot predict individual prices.

(iv) Universal models incorporate all of these factors that influence house price. These are usually disaggregated into locational and environmental effects with some measure of the socio-economic nature of the neighbourhood and housing characteristics.

It should be noted that all of the models mentioned are static or relative models, theoretically reflecting the conditions prevailing at any one time. Dynamic models would in addition, have to include variables that measured the changing influence of the economy, government decisions, and demographic factors. However while it is considered that the absolute relationship between property prices could vary quite quickly in times of rapid price increases, the underlying relative relationships are not as prone to rapid change (TIMMS, 1971, 120-121). What are considered as 'good' neighbourhoods or 'poor' neighbourhoods do not change quickly but are subjected to historically quite long movements.

In this study the average house price level over each ground truth area was only considered. These prices were examined using generalised universal models, which include some measures of location, environment and average housing characteristics. Data were derived essentially from three main sources; actual house sales data, data sampled from aerial photographs, and satellite derived data.

8.3 House Price Data Collection and Transformation

The streets included in each ground truth area were listed and the records of the Valuer General's department were accessed using these locational identifiers. The Valuer General's department is a government instrumentality that collects property sales data and assesses property value. For each area the recorded house sales were extracted for the period from the beginning of 1968 until the end of 1972. The total data set extracted for each house sold, comprised, sale price in Australian dollars, date of sale, number of rooms, area of house block, building material, either brick, timber or fibrocement, type of house, either detached, semi-detached (attached on one side only) or terrace (attached on both sides), presence or absence of swimming pool, number of stories, and age of house.

For each ground truth area additional data were derived from available maps. These comprised distance to the central business district (C.B.D.), distance to the nearest coast or major waterway, distance to the nearest railway station, average height above sea level, and a measure of terrain variability. This latter variable was taken as the sum of the number of 20 metre contours crossed in an east-west and north-south direction over a one kilometre square centred on the ground truth area.

The house sales data extracted from the Valuer General's records covered a period from January 1968 to December 1972, and so was not comparable because of increasing sale prices, inflation over that period ranging between 5% and 15%. A series of prices corrected to the beginning of January 1973 were calculated for each property using

$$P73 = P_T (1 + g)^{(73-T)} \qquad (8.1)$$

where

P73 = Price 1st January 1973

P_T = Price at time T

g = Average annual inflation rate, decimals of one.

and different values of g, ranging from 0.05 to 0.15 in steps of 0.01.

If the correct average annual inflation rate had been applied then the regression between P73 and time would have a correlation of zero. Each of the P73 values were correlated with time, the minimum zero value being achieved when g = 0.12. These new set of prices were given the variable name $(P73)_{12}$.

While a value of g = 0.12 was the average correction to be applied it was expected that some areas had increased in price at a different annual rate, and at different rates within years. To test this the $(P73)_{12}$ values in each area were used as the dependent variables in a multiple regression analysis with age of property, number of rooms, area of house block, type of house (1 for terrace, 2 for semi-detached and 3 for attached) and building material (1 for fibro, 2 for timber and 3 for brick) and in addition (73 - T) and $(73 - T)^2$ to test for non-linear effects of time. The former non-temporal variables were included to account for any price differences due to individual property variations. If the average annual rate of g = 0.12 was applicable to a particular area then the time variables would not be significant variables, and no further adjustment would be required. If the time variables were significant indicating a variation from the average g value, then each individual price in that particular area was corrected by the appropriate amount using the coefficients of the time variables. This occurred in about ten percent of all cases. The adjusted $(P73)_{12}$ values were then averaged for each area.

The average value of age (time before 1973), number of rooms, and block size for each ground truth area were also determined. For each area

the percentage of houses of each different material, different type, and different number of stories was determined and as well the percentage of properties with swimming pools. A data file was created containing the average adjusted price, average house characteristics, locational and terrain values for each ground truth area.

8.4 House Price Prediction

The average P73 values were used as the dependent variable in a multiple regression analysis with housing characteristics, locational and terrain data. Quadratic values of distance to the C.B.D. and number of rooms were also included with the independent variables as it was considered that these variables may not be linearly related to price. Stepwise multiple regression was used with a minimum F-ratio criterion of 4.0 for each variable to enter the regression.

The most significant variable predicting house price was found to be the number of rooms squared followed by distance to the nearest railway station. This latter variable was only marginally significant with the multiple correlation coefficient, r, increasing from 0.86 to 0.88 as it entered the regression. Overall F-ratio of the equation was 102, which is highly significant, with a standard deviation of $A9,600.

From a consideration of these results it was expected that a measure of house dimensions derived from the photographically sampled variables should also have been a good predictor of average price. Such a variable can be derived by dividing Ha (average percentage of house cover per pixel) by Na (average number of houses per pixel), therefore house size percentage,

$$HZ = \frac{Ha}{Na} \qquad (8.2)$$

where HZ = the average percentage of pixel area covered by one house.

In addition to HZ and HZ^2, all cover variables and locational variables were regressed with P73 as the dependent variable. These latter variables were included because they can be determined from aerial photographs or in some cases Landsat imagery, although in this case they were determined from maps which were originally derived from aerial photographs. Only three variables were significant HZ^2, HZ and Ra (average percentage of road cover per pixel), the latter variable being only marginally significant. Multiple correlation, r, was 0.88. The derived equation, with appropriate statistics, was as follows

$$P73 = 109.6 + 3.75\ HZ^2 - 34.04\ HZ - 0.51\ Ra \qquad (8.3)$$
$$(r = 0.88,\ \bar{r}^2 = 0.77,\ F = 70,\ S.E. - 9.5)$$

where

P73 and S.E. are in thousands of dollars.

This is an equivalent result to that achieved using known house characteristics. It is considered that Ra possibly represents a negative environmental quality effect or may also represent a locational effect as road percentage tends to be higher nearer the C.B.D.

The emergence of house size as the major predictor of house price is an interesting result, as this variable should be predictable from Landsat data. In Chapter 7 average house percentage was related to Landsat derived variables with a multiple correlation of 0.81, and average house numbers per pixel with a multiple correlation of 0.91. A new variable HZ^1, representing the Landsat equivalent of HZ, was determined using (7.14) and (7.23) to give,

$$HZ^1 = (136.6 - 57.3\ R76 - 105.5\ R46 - 0.643\ SUMSD + 82.8\ R57)\ /\ (-5.3 + 24.0\ R57 - 0.62\ SD7 + 0.14\ RV64 - 2.6\ SD4 + 0.64\ SD42 - 0.23\ RV56) \qquad (8.4)$$

The correlation between HZ and HZ^1 was found to be 0.70. While this value was lower than expected it still represented a significant result. The regression equation between HZ and HZ^1 (8.5) also shows that the relationship is essentially correct because the constant is very close to zero, and the coefficient is very close to one as follows-

$$HZ = 0.19 + 0.98\ HZ^1 \qquad (8.5)$$

The new variable HZ^1, $(HZ^1)^2$ and locational variables were regressed with P73 as the dependent variable. Once again house size was the most explanatory variable with the average height above sea level of the ground truth area and distance to water bodies being marginally significant variables. A multiple correlation coefficient value of 0.62 was determined. While the latter two variables are different from those determined previously they still represent a measure of location, although somewhat more environmentally related.

It was considered that an alternative to using the calculated value HZ^1, which is a rather cumbersome combination of reflectance derived variables, was to use reflectance values directly in a regression with P73. The band on band reflectance ratios and polynomial terms, the various standard deviation variables defined in Chapter 7, and the various locational variables were used as the independent variables. The following equation resulted

$$P73 = + 167.6 - 5.84\ RV45 - 0.86\ CBD \qquad (8.6)$$

where

CBD = average distance in kilometres of a ground truth area to the centralbusiness district

r = 0.60

$\bar{r}^2$ = 0.32

F = 16

S.E. = 16.2

Although not as predictive as other extended area equations, the equation is still significant at four times the 1% confidence level.

An examination of the residuals revealed that two areas had predicted prices substantially below their measured values. These represented two unique areas within the data set. Average values for each area were $±116,000 and $A127,000, while the balance of areas ranged in average price from $A14,000 to $A73,000. Equivalent values for 1980 would be approximately three times larger. The predicted prices still indicated however, that both areas were at the top of the price range with values of $A62,000 and $A75,000 respectively, but that many other undefined factors apart from house size and location contributed to their average price. Both areas are commonly known as two of the most exclusive areas in Sydney - Bellevue Hill and Vaucluse.

When these two unique areas were removed from the data set, a substantially higher multiple correlation of 0.73 was achieved, with distance to waterways and average height above sea level re-entering the regression equation as the secondary variables, the derived equation then being.

$$P73 = 82.1 - 2.69\ RV45 + 0.12\ RL - 0.58\ CST \qquad (8.7)$$

where

RL = average height in metres above sea level of a ground truth area, and

CST = average distance in kilometres of a ground truth area to the nearest coast or major waterway.

r = 0.73

$\overline{r^2}$ = 0.50

F = 21

S.E. = 7.7

House prices predicted from Landsat data cannot be considered as a reliable source for detailed analysis as indicated by the standard errors of the latter two equations, however as WESTERLUND (1972, 39) suggests of MUMBOWER and DONOGHUE's (1967) photographically derived housing density, it has benefit for initial analysis. If five arbitrary classes of house price are assumed 0 to $80,000 in steps of $20,000 and greater than $80,000, then the following results can be arrived at. From the regression equation (8.6) 62% of the areas would be correctly classified from the predicted results, 33% would be one class either above or below their predicted class and 5% would be removed by two class divisions. From equation (8.7) 75% would be correctly classified, and 25% would be classified either one above or below their correct class. These results should generally satisfy the need for more residential classes as expressed by LANDINI et al (1979, 104).

8.5 A Total Residential Quality Index

As RICHARDSON (1971) suggests house price differences may be explained in terms of environmental or area preference models. While there are some reservations to this hypothesis it could be assumed that house price or a measure thereof is an indicator of residential environmental quality. In addition many studies have shown a relationship between house price or rent and socio-economic status, classically represented by Hoyt's sector theory (see for example TIMMS, 1971, for a discussion of this theory). It could be suggested then that house size as a surrogate for house price may be a measure of housing quality and social environment.

The percentage of grass and trees in an area can normally be associated with environmental quality indicating open space and aesthetic qualities. NARIGASAWA and FUCHIMOTO (1979) have used a vegetation cover ratio determined from multispectral scanner data as one of a number of environmental themes. GREEN (1957) and GREEN et al (1959) have also suggested the prevalence of industrial land use as a negative residential

quality measure. While there are other factors contributing to residential quality, the purpose of this section is to demonstrate how a general quality index might be achieved.

If it is assumed that

(i) house size is a measure of housing quality and social environment and would be positively related to a residential quality index,

(ii) the total vegetative content is a positive indicator of quality, and

(iii) the sum of road and other building percentages represents the negative influence of encroachment of non-residential land use and likely high noise and pollutant levels, then a residential quality index, RQI, might be represented by

$$RQI = \frac{HZ \times AGREEN}{ANONRESID} \qquad (8.8)$$

This derived variable was used as the dependent variable in a regression with various reflectance variables. The following results were achieved

$$RQI = -129.8 + 674\ R65 + 113.1\ R56 - 1.78\ RV45 \qquad (8.9)$$

$r = 0.87$

$\bar{r}^2 = 0.75$

$F = 64$

$S.E. = 6.7$

which represents a highly significant result.

Equation (8.9) can be normalised to give a suitable scale, say 0 to 100. The means and standard deviation of AGREEN, HZ and ANONRESID were 52 ± 14, 5.4 ± 1.3 and 19 ± 9.0 respectively. If it is assumed that the best and worst possible residential values of these variables were at ± 2 SD, then the highest value of RQI would be 640 and the lowest 2. If the above equation is then divided by 6.4, RQI would have a scale from zero for worst quality to one hundred for best quality, therefore,

$$RQI = -20.3 + 105\ R65 + 17.7\ R56 - 0.28\ RV45 \qquad (8.10)$$

8.4 Other Potential Applications

Very rarely can remotely sensed data be used as the sole source of information about an area. Any survey of an area must draw on all available data and then select a subset which in terms of cost, speed, and reliability can satisfy the defined goals of the survey. It is with this point in mind that potential applications are suggested.

(i) Urban catchment runoff studies require as input, data on rainfall, slopes and surface characteristics. A typical urban drainage system is made up of a number of elements which drain sub-catchment areas. Cover percentage determined from Landsat data at either the single pixel or extended area level could be used to derive aggregate coefficients of runoff. Over each area these coefficients could be used with slope data, derived from topographic maps or digitally held terrain models, and rainfall intensity data to calculate the rate and magnitude of surface runoff.

(ii) A number of researchers have suggested the application of Landsat data to population estimation. While predicted housing cover can be used as a surrogate for population density, it is preferable to use housing density as a measure of the number of families. Size of family also tends to vary between socio-economic groupings. If house size or house price is a surrogate for socio-economic status and is used in conjunction with housing density, both derivable from Landsat data, then the potential exists for a better estimation of population and its areal distribution.

Census data gives very accurate estimates of population within each census district, but the boundaries of districts are generally not related to any stratification within the community or the landscape and so can contain mixtures of land use and housing types. In these circumstances the population densities determined from the aggregated census district data may bear little relationship to actual densities in particular sub-areas. Landsat derived population estimates would help overcome this anomaly.

(iii) Rather than use predicted cover percentages at the pixel or extended area level as a means of residential differentiation it would be preferable to develop a set of simpler classes. Using G, T, RESID and NONRESID as the primary cover classes a sixteen level classification scheme could be developed with above and below average of each of the cover classes acting as a dichotomous division. Further research would be

required to give specific meaning to these classes in their residential context. However it might be assumed that these classes could be used to differentiate varying levels of residential development on the urban fringe in contrast to attempting to define a single urban/rural interface.

(iv) Two major orthogonal factors that are cited from social area analysis of western industrialized cities are family status and social status. These factors appear to be reflected in the physical environment of the Sydney urban area. Percentage grass cover and percentage tree cover are uncorrelated and when combined with measured house size and housing density might suggest themselves as surrogate variables for the orthogonal, family and social status factors.

8.6 Summary and Conclusions

It has been demonstrated that the average area Sydney house price can be substantially predicted from Landsat derived reflectance data with readily available locational variables acting as secondary auxilliary predictors. It cannot be concluded that this type of analysis would be directly applicable to other cities. Nevertheless, while Sydney has many unique characteristics, it is not atypical of western industrialized cities and in population, size, residential density, morphology and climate is quite similar to the western seaboard North American cities of Los Angeles and San Francisco. The emergence of the single variable house size as the primary predictor of price would suggest that these general results could be applicable in other similar cities.

The demonstration of the use of Landsat data for predicting a residential quality index and the suggested applications of reflectance derived cover and cover related data, indicates that the potential of Landsat data as a tool in urban analysis is considerable, particularly when it is used in association with other readily available data.

9. FINAL CONCLUSIONS

9.1 The measured response at the Landsat sensor from urban residential areas is complicated by the mixed nature of the cover surface, the effect of background reflectance on path radiance and the convolving effect of the sensor point spread function.

9.2 Appropriately processed Landsat response data from urban residential areas of mixed surface cover, can be used to determine the reflectance of individual surface cover types.

9.3 The percentage amounts of individual residential surface cover types contributing to the response at the Landsat sensor can be significantly predicted at the single pixel level. Better results are achieved using data from extended areas (5 x 8 pixel block).

9.4 The use of a convolved cover model, where the measured sensor response is related to the percentage of surface cover types in the central and surrounding pixels, is to be preferred for predicting cover reflectance and percentage cover at the pixel level.

9.5 Ratio type variables are significantly better than individual band variables for predicting cover percentages.

9.6 Regression equations relating percentage of surface cover to reflectance derived variables give the following best results (on the basis of multiple correlation coefficient) for various surface types, and for single pixel and extended (5 x 8 pixel block) areas.

Cover	r (pixel)	r (extended)
House	0.58	0.81
Other	0.53	0.88
Road	0.65	0.89
Concrete	0.42	0.63
Tree	0.73	0.91
Grass	0.61	0.81
Green	0.82	0.92

9.7 Variation of reflectance over an extended (5 x 8 pixel block) area, as measured by the standard deviation, is a significant auxiliary variable for predicting surface cover percentages.

9.8 The total residential percentage cover data can be represented by two overlaid orthogonal systems. The first system separates vegetative and non-vegetative surface covers, and the second system separates grass and tree surface covers.

9.9 Relative average house price over extended areas (of approximately 18ha) can be substantially predicted from the average size of houses within that area, as interpreted from aerial photographs.

9.10 The average number of houses and the average house size within a sampled area of approximately 18ha can be significantly predicted from Landsat derived reflectance variables.

9.11 The relative average house price over extended areas of approximately 18ha can be significantly predicted from Landsat derived reflectance variables.

9.12 A residential classification scheme can be developed using relative house prices predicted from Landsat derived reflectance variables.

9.13 A residential quality index based on house size, percentage vegetative cover and percentage road and other building cover can be significantly predicted from Landsat derived reflectance variables.

9.14 The potential of Landsat data as a tool in urban analysis is considerable, particularly when it is used in association with other readily available data.

APPENDIX 1

A1. ATMOSPHERIC AND POINT SPREAD FUNCTION CALCULATIONS

A1.1 Sensor Response

The Landsat bands plotted in figure A1.1 represent averaged curves for the four MSS sensors (HUGHES AIRCRAFT CO., 1972). MILTON (1978) has questioned the validity of the half bandwidth points for MSS band 7 of 0.800 - 1.100 μm as quoted by NASA (1972) as the nominal half band widths. However to retain consistency with other results the published band width of 0.800 - 1.100 μm for band 7 and bandwidths of 0.500 - 0.600 μm, 0.600 - 0.700 μm, 0.700 - 0.800 μm, respectively of bands 4, 5 and 6 were used for interpolation and calculation.

The radiance per bit of count rate, K, is a constant determined by the scanner system. For Landsat 1 the values of K in the four bands are -

$$K_4 = 0.1953 \text{ Wm}^{-2} \text{ sr}^{-1} \text{ bit}^{-1}$$

$$K_5 = 0.1575$$

$$K_6 = 0.1386$$

$$K_7 = 0.3175$$

derived from the ERTS Data Users Handbook for Landsat 1 (National Aeronautics and Space Administration (NASA), 1972) where K_4, K_5, K_6, are the radiance per bit of count rate for bands 4, 5 and 6 and K_7 is for a doubled band 7 scale. The measured count rates at the sensor were converted to radiance values by using these expressions.

A1.2 Solar Irradiance at the Top of the Atmosphere, E_o

A plot of solar spectral irradiance at the top of the atmosphere is shown in figure A1.2. This was taken from SABATINI et al (1970) after JOHNSON (1961) for an atmosphere with zero optical thickness. The values shown in figure A1.2 were derived assuming the mean distance of the earth from the sun. However at aphelion the distance of the earth from the sun is approximately 1.034 times that at perihelion. Hence by simple geometry this represents an increase in the solar irradiance between aphelion and perihelion of approximately 6.8%, or a change from the mean of ± 3.4%.

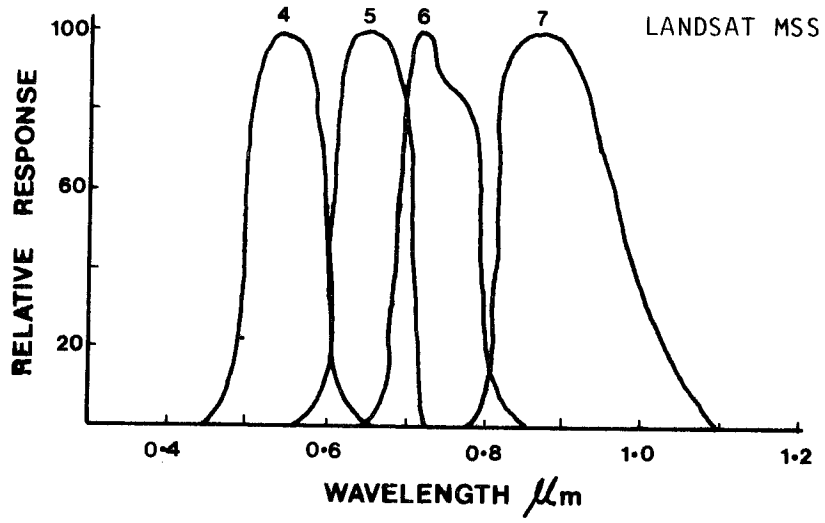


FIGURE A1.1: Spectral response of Landsat MSS.

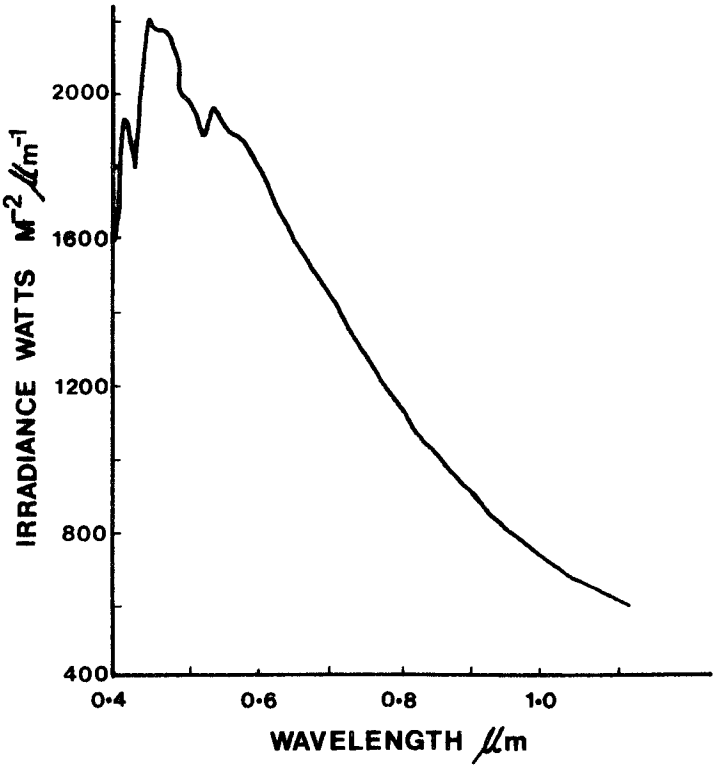


FIGURE A1.2: Solar spectral irradiance at the top of the atmosphere.

Interpolated values for the four MSS bands are given in Table A1.1 at monthly intervals from July 1st to January 1st.

A1.3 Atmospheric Transmittance

Available atmospheric data at the time of satellite overpass (9.09 am local time, 12th December 1972) were obtained from the Bureau of Meteorology for the Sydney area. These were as follows

Temperature	28.7^{o}
Relative Humidity	41%
Wind	North west at 3 knots
Pressure	1014.4 mb
Visibility	15 km

Graphs by ELTERMAN (1968) allow estimates of atmospheric transmittance for a Rayleigh type atmosphere at 0^{o} solar zenith angle. For the mean wavelength of the four Landsat bands the following transmittance estimates were determined, with their equivalent optical thickness.

Band 4	$T = 0.91$	$\tau_m = 0.09$
Band 5	$T = 0.95$	$\tau_m = 0.05$
Band 6	$T = 0.97$	$\tau_m = 0.03$
Band 7	$T = 0.99$	$\tau_m = 0.01$

These figures were given for a standard pressure, p, at sea level, h_o, of 1013 mb. For any other pressure at elevation h, the following relationship applies

$$\tau_m (h) = \tau_m (h_o) \frac{p(h)}{p(h_o)}$$

For a pressure of 1014.4 mb at the time of overpass, the correction was therefore negligible.

It is reasonable to assume that the haze level or visibility of the atmosphere is in most cases determined by the aerosols, which are concentrated heavily in the lower most portion of the atmosphere. Visibility or visual range can be defined as the distance at which the contrast between a black object with radiance L and sky background with radiance L_B is reduced to two percent which is assumed to be the limiting (daylight) contrast between objects. Koschmieder's formula shows that visual range is inversely

TABLE A1.1

Spectral solar irradiance at the top of the atmosphere, July 1st (apehelion) through January 1st (perihelion) [Wm^{-2} μm^{-1}]

Band	July 1st	Aug 1st	Sep 1st	Oct 1st	Nov 1st	Dec 1st	Jan 1st
4	1835	1857	1878	1900	1922	1943	1965
5	1579	1598	1606	1635	1654	1672	1691
6	1251	1266	1280	1295	1310	1324	1339
7	783	792	801	810	819	828	837

proportional to the volume extinction coefficient and hence visual range is a measure of optical thickness for a particular vertical particulate distribution, (MIDDLETON, 1952).

Using this definition of visual range TURNER et al (1972) have graphed aerosol optical thickness as a function of visual range for a number of wavelengths. However visual range is usually defined at a wavelength of 0.55 μm and the graphs do not account for spectral variations in visibility. The observed visibility of 15 km, at the time of over flight, could only be considered comparable in the visible part of the spectrum. Band 4 of Landsat, centrally located in the visible spectrum, was considered to approximate this condition. A value of τ_p of 0.34, for visual range 15km and wavelength 0.55 μm, was interpolated.

However from Chapter 2

$$\tau_p = \frac{K}{\lambda^{1.3}} \quad , \text{ where K is a constant}$$

Thus substituting for $\tau_p = 0.34$ at $\lambda = 0.55$ μm, K equals 0.16. Therefore τ_p was calculated for bands 5, 6, and 7 giving

τ_p (band 5) = 0.28

τ_p (band 6) = 0.23

τ_p (band 7) = 0.17

Since aerosol particulates have a negligible influence on pressure, no pressure correction was required. While there seems to be a positive correlation of scattering coefficient with relative humidity, the effects are important only when the relative humidity exceeds 35% by one estimate (ROSENBERG, 1967) or 60% by another estimate (PUESCHEL et al, 1969). For an observed relative humidity of 41% it was considered negligible. Aerosol optical thickness declines with altitude, however over Sydney with approximately sea level terrain this effect was not considered.

The optical thickness of water vapour is very small for wavelengths less than 0.7 μm but must be accounted for at longer wavelengths. Values of optical thickness are usually tabulated as a function of the mass of liquid water equivalent (of the total vapour) in a vertical column of atmosphere of unit area in cross section. Typical values are 0.5 to 1.0 gcm^{-2} for a mid-latitude winter atmosphere and 2.5 to 3.0 gcm^{-2} for a mid-latitude summer atmosphere.

For a relative humidity of 41% and temperature of 28.7°C a partial pressure of water vapour of 12.7 mm of mercury was calculated. This converted via the ideal gas law to 12.2 gm^{-3}. It was assumed that water vapour density decreased exponentially with height, and an equivalent atmospheric height (scale height) of 2.5 km was used to give a liquid water equivalent mass per unit area of 3.0 gcm^{-2} (AM. SOC. of PHOTO, 1975, 183).

Bands 6 and 7 of Landsat have spectral response functions which span the water vapour absorption bands. SABITINI et al (1970) tabulates the percentage departure of expected radiances from a turbid atmosphere as defined by ELTERMAN (1968) with 2.0 gcm^{-2} of moisture added. These percentage departure values are based on expected radiances from a surface with unit reflectance, neglecting the effects of atmospheric scattering. These values are shown in Table A1.2, for no moisture and 3.0 gcm^{-2} moisture content at 30° solar zenith angle.

TABLE A1.2

Percentage departure of expected radiance from a model atmosphere with 2.0 gcm^{-2} moisture added.

Band	No Moisture (30° Solar Zenith angle)	3.0 gcm^{-2} (30° Solar Zenith angle)
6	+ 6%	- 2%
7	+18%	- 7%

From Table A1.2 for band 6

$$\frac{\pi E_G \, e^{-\tau_o \sec 30^o} \, e^{-\tau_o}}{\pi E_G \, e^{-\tau_3 \sec 30^o} \, e^{-\tau_3}} = \frac{106}{98} \qquad \text{(A1.0)}$$

where E_G is global irradiance and τ_o and τ_3 are normal optical thickness for no moisture and 3.0 gcm^{-2} moisture content respectively. For ELTERMAN's (1968) reference turbid atmosphere, transmittance is 0.88 at 0° zenith angle which gave a value of τ_o of 0.128. Substituting in (A1.0), gave $\tau_3 = 0.165$. But $\tau_o = 0.128$, therefore the optical thickness for 3.0 gcm^{-2} water content alone was 0.04, to the second decimal place. Using a similar analysis for band 7 the equivalent value was 0.11.

The optical thickness of ozone in the visible spectrum is quite small having its greatest value of 0.05 at 0.6 μm for ELTERMAN's (1968) reference atmosphere. The zone of absorption is from approximately 0.5 to 0.7 μm and a representative value of optical thickness for both bands 4 and 5 is 0.03.

The component parts of the total optical thickness at the time of over flight are shown in Table A1.3. They are tabulated for each band of Landsat, with the total optical thickness and equivalent total transmittance also being shown.

TABLE A1.3

Component parts of optical thickness and total transmittance for each Landsat band over Sydney 12th Dec. 1972.

	Band 4	Band 5	Band 6	Band 7
τ_m	0.09	0.05	0.03	0.01
τ_p	0.34	0.28	0.23	0.17
τ_{H2O}	0.00	0.00	0.04	0.11
τ_{03}	0.03	0.03	0.00	0.00
τ_{Total}	0.46	0.36	0.30	0.29
T	0.63	0.70	0.74	0.75

A1.4 Solar zenith and azimuth angle, and nadir scan angle

Solar zenith angle can be calculated using spherical trigonometry and solar ephemeris or from tables or planispheres made for the purpose. Using a planisphere constructed for a Sydney south latitude of 34^{o} a solar zenith angle of 38^{o} and solar azimuth of 84^{o} were determined. These values were for a satellite overflight time of $23^{h}.14$ G.M.T. which for a Sydney longitude of $10^{h}.01$ East gives a local apparent time of $9^{h}.15$ (9 hours 09 minutes) on 12th December 1972.

A nadir scan angle can be considered equal to zero because the Landsat scanners look vertically down from an altitude in excess of 900 km with a total field of view of 11^{o}. Assuming a representative value of 0.4 for optical thickness, transmittance will vary from 0.670 in the vertical to 0.669 at a $5\frac{1}{2}^{o}$ nadir scan angle. This non-vertical effect was considered negligible.

A1.5 Path Radiance

Using the optical thickness values given in Table A1.3, E_{O_λ} for December 12th interpolated from Table A1.1, the wavelength interval for each band, and values of $P(\mu_o)$ and $P(-\mu_o)$ at $\theta_o = 38^o$ from figure 2.5, the functions F, G and H of (2.12) were calculated for each Landsat band. For F and G these calculated values were as follows:

$$F_4 = 0.0106 \qquad G_4 = 0.0453$$
$$F_5 = 0.0071 \qquad G_5 = 0.0361$$
$$F_6 = 0.0047 \qquad G_6 = 0.0292$$
$$F_7 = 0.0023 \qquad G_7 = 0.0208$$

The function H is dependent on $\overline{R}_B$, average background reflectance, and was determined from

$$H = \frac{\overline{R}_B}{1 + J(1-\overline{R}_B)} \qquad \text{where} \quad J = 2(1 - \eta)\,\tau_o$$

and for each band calculated as

$$J_4 = 0.1238$$
$$J_5 = 0.0779$$
$$J_6 = 0.0530$$
$$J_7 = 0.0281$$

For small values of J, H was approximated by

$$H = \overline{R}_B - J\,\overline{R}_B + J\,\overline{R}_B^{\,2}$$

From (2.12) and substituting $E_o\, e^{-(\tau_{O3} + \tau_{H2O})}$ for E_o then L_p was calculated from

$$L_p = E_o\, e^{-(\tau_{O3} + \tau_{H2O})}\,(F + G\,H)$$

which gave

$$L_{p4} = 2.0 + 7.5\,\overline{R}_{B4} + 1.1\,\overline{R}_{B4}^{\,2} \qquad [Wm^{-2}\ sr^{-1}]$$
$$L_{p5} = 1.2 + 5.4\,\overline{R}_{B5} + 0.5\,\overline{R}_{B5}^{\,2}$$
$$L_{p6} = 0.6 + 3.5\,\overline{R}_{B6} + 0.2\,\overline{R}_{B6}^{\,2}$$

$$L_{p7} = 0.5 + 4.5\,\bar{R}_{B7} + 0.1\,\bar{R}_{B7}^{2} \qquad (A1.1)$$

A1.6 Global Irradiance

Using values of J and H from section A1.5 and values of I calculated for the given conditions as follows

$$I_4 = 0.7307$$

$$I_5 = 0.7510$$

$$I_6 = 0.7624$$

$$I_7 = 0.7753$$

and then from (2.13)

$$E_D = E_o\, e^{-(\tau_{H2O} + \tau_{O3})}\, I\,[1 + JH] - E_o\, T_{\theta_o} \cos\theta_o$$

which gave

$$E_{D4} = 52.6 + 15.0\,\bar{R}_{B4} + 2.1\,\bar{R}_{B4}^{2} \quad [Wm^{-2}]$$

$$E_{D5} = 38.6 + 8.8\,\bar{R}_{B5} + 0.7\,\bar{R}_{B5}^{2}$$

$$E_{D6} = 25.8 + 4.9\,\bar{R}_{B6} + 0.3\,\bar{R}_{B6}^{2}$$

$$E_{D7} = 26.0 + 4.4\,\bar{R}_{B7} + 0.1\,\bar{R}_{B7}^{2}$$

and

$$E_{G4} = 138.4 + 15.0\,\bar{R}_{B4} + 2.1\,\bar{R}_{B4}^{2} \quad [Wm^{-2}]$$

$$E_{G5} = 122.4 + 8.8\,\bar{R}_{B5} + 0.7\,\bar{R}_{B5}^{2}$$

$$E_{G6} = 97.4 + 4.9\,\bar{R}_{B6} + 0.3\,\bar{R}_{B6}^{2}$$

$$E_{G7} = 173.0 + 4.4\,\bar{R}_{B7} + 0.1\,\bar{R}_{B7}^{2} \qquad (A1.2)$$

As the quadratic term will add a maximum of less than 2% to E_G when background reflectance is 100% it was neglected in all further calculations.

A1.7 Target Reflectance

From (2.14)

$$\frac{\Sigma R_i \, \Delta A_i}{A} = \frac{\pi (CK - L_p)}{E_G \, T_{\theta V}}$$

Thus for each band the following was derived

$$\left(\frac{\Sigma R_i \, \Delta A_i}{A} \right)_4 = \frac{\pi (0.1953 \, C_4 - 2.0 - 7.5 \, \overline{R}_{B_4} - 0.5 \, \overline{R}_{B_4}^2)}{0.63 \, (138.4 + 15.0 \, \overline{R}_{B_4})}$$

$$\left(\frac{\Sigma R_i \, \Delta A_i}{A} \right)_5 = \frac{\pi (0.1575 \, C_5 - 1.2 - 5.4 \, \overline{R}_{B_5} - 0.5 \, \overline{R}_{B_5}^2)}{0.70 \, (122.4 + 8.8 \, \overline{R}_{B_5})}$$

$$\left(\frac{\Sigma R_i \, \Delta A_i}{A} \right)_6 = \frac{\pi (0.1386 \, C_6 - 0.6 - 3.5 \, \overline{R}_{B_6} - 0.2 \, \overline{R}_{B_6}^2)}{0.74 \, (97.4 + 4.9 \, \overline{R}_{B_6})}$$

$$\left(\frac{\Sigma R_i \, \Delta A_i}{A} \right)_7 = \frac{\pi (0.3175 \, C_7 - 0.5 - 4.5 \, \overline{R}_{B_7} - 0.1 \, \overline{R}_{B_7}^2)}{0.75 \, (173.0 + 4.4 \, \overline{R}_{B_7})}$$

A number of approximations were applied. The approximation $\frac{1}{1+x} = 1 - x$ was used because $x = (\frac{15.0}{138.4})$ is small. Terms that contributed 1% or less to total reflectance when background reflectance was 100% were neglected, which gave in terms of percent reflectance -

$$\left(\frac{\Sigma R_i \, \Delta A_i}{A} \right)_4 = 0.634 \, C_4 \, (1 - 0.00105 \, \overline{R}_{B_4}) - 6.5 - 0.236 \, \overline{R}_{B_4}$$

$$\left(\frac{\Sigma R_i \, \Delta A_i}{A} \right)_5 = 0.578 \, C_5 \, (1 - 0.00072 \, \overline{R}_{B_5}) - 4.4 - 0.195 \, \overline{R}_{B_5}$$

$$\left(\frac{\Sigma R_i \, \Delta A_i}{A} \right)_6 = 0.604 \, C_6 \, (1 - 0.00050 \, \overline{R}_{B_6}) - 2.6 - 0.151 \, \overline{R}_{B_6}$$

$$\left(\frac{\Sigma R_i \, \Delta A_i}{A} \right)_7 = 0.769 \, C_7 \, (1 - 0.00026 \, \overline{R}_{B_7}) - 1.2 - 0.109 \, \overline{R}_{B_7} \qquad \text{(A1.3)}$$

A1.8 Point Spread Function

The purpose of this part of the research was to develop an algorithm that removed the effects of psf convolution. However this idealistic requirement was tempered by the knowledge that any increase in the number of pixels used in the solution array would substantially increase computing time to a point that made the use of the deconvolution algorithm inappropriate for practical application. What was required was a simplified algorithm that substantially reduced the effect, so that the correlation between ground and sensor data could be improved.

The average ordinates for the psf were determined at discrete one third pixel and line intervals using values bilinearly interpolated from figure 2.6. These ordinate values were limited to a 3 x 3 pixel array, a total of 81 values. While this procedure neglected small negative ordinate values two pixels from the central pixel in the along scan direction and small positive values in the opposite direction, it was considered that their exclusion would not substantially alter the results. Values were normalised so that the integral of the psf equalled one and are shown in Table A1.4.

A constant, unit value central pixel surrounded by zero value pixels was assumed, and the psf of Table A1.4 was discretely stepped at 1/3rd pixel and line intervals. The results at each point were averaged across the appropriate pixel to give the results shown in Table A1.5.

These results represent the values recorded at the appropriate pixel positions for a true central pixel response of one. Alternatively they can be considered to represent the weights to be given to each pixel's true response as their contribution to the central pixels's recorded response.

The Landsat MSS records in a six line swath causing a discontinuity to occur at the edge of the first and sixth lines. Because of this effect the results of Table A1.5 are strictly only appropriate for central pixels on the second through 5th lines. However the spatial relationship between ground truth areas and Landsat data was only known to the nearest half pixel. The inclusion of a correction for this effect was therefore considered inappropriate in many cases. Given this and the fact that the error would be small and only affect part of the sample the simplified concept of no discontinuity was assumed.

TABLE A1.4

Ordinate values of Landsat sensor psf for a 3 x 3 pixel array.

0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0.008	0.025	0.054	0.075	0.103	0.070	0.043	0.015	0
0.052	0.178	0.366	0.516	0.590	0.497	0.300	0.105	0
0.060	0.206	0.422	0.605	0.680	0.572	0.347	0.122	-0.009
0.052	0.178	0.366	0.516	0.590	0.497	0.300	0.105	0
0.008	0.025	0.054	0.075	0.103	0.070	0.043	0.015	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0

Scan Direction →

TABLE A1.5

Estimated recorded response from a unit value central pixel.

0.03	0.07	0.02
0.18	0.45	0.13
0.03	0.07	0.02

Scan Direction →

Using the values from Table A1.5, $L_s^*(x,y)$ can be given by

$$\begin{aligned} L_s^*(x,y) = {} & 0.03\, L_s(x-1,y-1) + 0.07\, L_s(x,y-1) \\ & +0.02\, L_s(x+1,y-1) + 0.18\, L_s(x-1,y) + 0.45\, L_s(x,y) \\ & +0.13\, L_s(x+1,y) + 0.03\, L_x(x-1,y+1) + 0.07\, L_s(x,y+1) \\ & +0.02\, L_x(x+1,y+1) \end{aligned} \qquad \text{(A1.4)}$$

where x and y are the pixel and line coordinates of the central pixel, L_s^* is the total radiance received by the sensor, and L_s is the total radiance at the sensor prior to convolution.

Similar equations for adjoining pixels can be formed. Two assumptions were made in forming these equations,

(i) the radiance values of pixels adjoining the central 3 x 3 array were set equal to their nearest neighbour in the array,

(ii) all radiance values of pixels outside this modified 4 x 4 array were set equal to zero.

The assumptions were chosen in preference to assuming all pixels outside the 3 x 3 array were equal to zero which would only realistically occur in extreme cases. The nearest neighbour assumption of (i) was considered to be statistically more probable than a zero assumption.

The equations formed can be shown in matrix notation as,

$$\begin{bmatrix} L_s^*(x-1,y-1) \\ L_s^*(x,y-1) \\ L_s^*(x+1,y-1) \\ L_s^*(x-1,y) \\ L_s^*(x,y) \\ L_s^*(x+1,y) \\ L_s^*(x-1,y-1) \end{bmatrix} = \begin{bmatrix} 0.73 & 0.15 & 0.00 & 0.10 & 0.02 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.21 & 0.52 & 0.15 & 0.03 & 0.07 & 0.02 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.21 & 0.67 & 0.00 & 0.03 & 0.09 & 0.00 & 0.00 & 0.00 \\ 0.10 & 0.02 & 0.00 & 0.63 & 0.13 & 0.00 & 0.10 & 0.02 & 0.00 \\ 0.03 & 0.07 & 0.02 & 0.18 & 0.45 & 0.13 & 0.03 & 0.07 & 0.02 \\ 0.00 & 0.03 & 0.09 & 0.00 & 0.18 & 0.58 & 0.00 & 0.03 & 0.07 \\ 0.00 & 0.00 & 0.00 & 0.10 & 0.02 & 0.00 & 0.73 & 0.15 & 0.00 \end{bmatrix} \begin{bmatrix} L_s(x-1,y-1) \\ L_s(x,y-1) \\ L_s(x+1,y-1) \\ L_s(x-1,y) \\ L_s(x,y) \\ L_s(x+1,y) \\ L_s(x-1,y-1) \end{bmatrix}$$

$$\begin{bmatrix} L^{*}_{s(x,y+1)} \\ L^{*}_{s(x+1,y+1)} \end{bmatrix} \quad \begin{bmatrix} 0.00 & 0.00 & 0.00 & 0.03 & 0.07 & 0.02 & 0.21 & 0.52 & 0.15 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.03 & 0.09 & 0.00 & 0.21 & 0.67 \end{bmatrix} \begin{bmatrix} L_{s(x,y+1)} \\ L_{s(x+1,y+1)} \end{bmatrix}$$

and were solved for $L_{s(x,y)}$ in terms of L^{*}_{s} values to give

$$\begin{aligned} L_{s(x,y)} = {} & 0.09\, L^{*}_{s(x-1,y-1)} - 0.35\, L^{*}_{s(x,y-1)} + 0.07\, L^{*}_{s(x+1,y-1)} \\ & -0.78\, L^{*}_{s(x-1,y)} + 2.76\, L^{*}_{s(x,y)} - 0.60\, L^{*}_{s(x+1,y)} \\ & +0.09\, L^{*}_{s(x-1,y-1)} - 0.35\, L^{*}_{s(x,y+1)} + 0.07\, L^{*}_{s(x+1,y+1)} \end{aligned} \qquad \text{(A1.5)}$$

APPENDIX 2

LANDSAT MSS SYSTEM

The Landsat satellites are earth resource observatories in near polar orbits. Three satellites have been launched, Landsat 1 (formerly ERTS1), Landsat 2 and Landsat 3. A fourth satellite in the series, Landsat D, is scheduled for launch in late 1981.

Two sensors are carried by each satellite - the Multispectral Scanner (MSS) system and the Return Beam Vidicon (RBV) system. The RBV sensor is essentially a high resolution television camera. This system ceased to operate on Landsat 1 shortly after its launching.

The MSS sensor, from which the data for this thesis is derived, continually scans a 185 km swath of the earth's surface producing four spectral bands of digitized imagery every 25 seconds, an area of approximately 185 km square. The earth is scanned by an oscillating mirror carried in the satellite. The intensity values of 6 pixels in the direction of satellite travel, for each of the four bands are sensed simultaneously by the scanner system. Data for an area of 474 m by 185 km is therefore recorded on each sweep. A schematic diagram of the MSS system is shown in figure A2.1.

The spectral regions recorded in each band are 0.5 - 0.6 μm (green), 0.6 - 0.6 μm (red), 0.7 - 0.8 μm (near infrared) and 0.8 - 1.1 μm (near infrared). These bands are labelled bands 4, 5, 7 and 7 respectively.

The scanner has a ground resolution of about 79m . While successive scan lines are spaced approximately 79 m apart on the ground, the pixels in a scan line are sampled every 57 m (approximately). An image on one band therefore consists of 3240 pixels (picture elements) along each scan line (approximately in an east-west direction) and 2340 scan lines. The satellite covers the same area of the earth every 18 days amounting to 20 passes per year.

Data from the Landsat satellite are transmitted directly to a ground receiving station when it is in range (2400 km) or can be recorded on tape for later transfer when out of range. The data from a single scene is distributed on a set of four computer compatible tapes (CCT's). Each CCT contains the data for all four wavebands for a strip approximately 45 km wide and 185 km long. Landsat data for the Australian region is now received directly through a receiving station and can be purchased through a processing and distributing centre in Canberra, Australian Capital Territory.

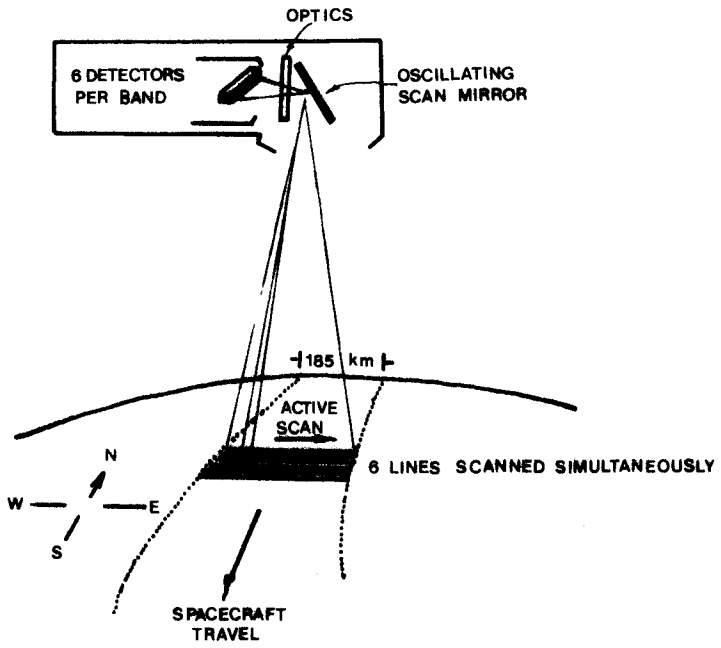


FIGURE A2.1: Schematic diagram of Landsat Multispectral scanner system.

APPENDIX 3.

A3. GEOMETRIC TRANSFORMATION PROCEDURES

A3.1 Introduction

A necessary procedure prior to the main study was the transformation of parts of the Landsat scene into the ground truth areas. As these ground truth areas were to be located at 4 km intervals, sufficient ground control to ensure a tight fit to each area was required and to allow examination of surrounding residuals. Control points were also located so that, if inadequate results were achieved from an overall polynomial transformation, individual affine transformations could be computed. With seventy ground truth sites a total of one hundred ground control points was needed to satisfy these aims.

Unlike cultivated rural areas, urban areas exhibit an extremely heterogeneous surface cover, and considerable interpixel and intra-pixel change can occur. Ground truth areas therefore needed to be spatially related to the Landsat digital image at the sub-pixel level if the correlation between ground and image was to be fruitfully examined.

The number, distribution and registration accuracy of ground control points (G.C.P.) will influence the accuracy of the computed ground image spatial relationship. BERNSTEIN (1976) shows that if G.C.P. registration accuracy is 0.5 of a pixel then 12 G.C.P. are needed to obtain a 50 metre root mean square error and 18 G.C.P. for a 40 metre root mean square error. As CARTER (1977) notes, these theoretical levels are not so easy to obtain.

A3.2 Basic Theory

Current research on the selection of suitable formulae for the geometric correction of Landsat data reveal three basic methods (TRINDER, 1978)

(i) A parametric solution in which an attempt is made to determine the position and attitude of the spacecraft using known formulae describing the image formation process.

(ii) Polynomial interpolation formulae which are a simplified approach to the adjustments, and do not require any knowledge of the image formation process.

(iii) A simple interpolation formula, followed by a least squares linear interpolation procedure.

TRINDER (1978) suggests that there appears to be little if any difference between results obtained using all three methods above. Because of its simplicity and ease of application using an available multiple regression computer package, the polynomial method was chosen for this investigation.

Results by TRINDER (1978) indicate that a 3rd order polynomial is adequate for the transformation of an entire Landsat scene using a limited number of ground coordinates derived from 1:100 000 and 1:250 000 map sources. However it was considered that when using a small portion of a scene with a large number of more accurate ground coordinates, as was the case in this study, a higher order effect could be significant. For this reason the adjustment equations initially chosen were complete 5th order polynomials of the form

$$E = a_o + a_1X + ... + a_5X^5 + a_6Y + ... + a_{10}Y^5 + a_{11}XY + ... + a_{23}X^4Y^3 \qquad (A3.1)$$

$$N = b_o + b_1X + ... + b_5X^5 + b_6Y + ... + b_{10}Y^5 + b_{11}XY + ... + b_{23}X^4Y^3 \qquad (A3.2)$$

where E and N are East and North coordinates respectively, (these coordinates were Integrated Survey Grid (I.S.G.) coordinates, a local system based on a transverse mercator projection, which is used for large scale mapping in the state of New South Wales). X and Y are the pixel and line numbers respectively of the control points and a_o to a_{23} and b_o to b_{23} are the transformation parameters.

In addition to these transformations, a correction for interswath discontinuity was applied. This cannot be corrected by the poynomial transformations. The discontinuity or apparent slip between swaths, is due to both earth rotation and sensor delay effects, and is of the order of 50 m for Australian latitudes (TRINDER, 1978). A 50 m shift was applied to all lines in each consecutive swath of six lines, the remaining linear skew was then accounted for by the polynomial transformation.

As the polynomial equations are essentially a least squares adjustment with E and N as the dependent variables and X and Y (and higher orders) as the independent variables, the problem was treated using multiple regression techniques to compute the transformation parameters.

In addition to the interswath geometric error corrected prior to the regression procedure, other significant geometric errors can exist. These include, the effects of non-linear scan rate of the oscillating scanning mirror and a topographic effect due to height differences of the various G.C.P. A full description of all geometric errors has been documented elsewhere (KRATKY, 1975).

For the selected Landsat sub-scene the topographic effect was found to be insignificant. The maximum height difference was 200 m, and the maximum distance from the satellite nadir was approximately 500 pixel widths. Using $\Delta r = \Delta z\ r/z$ (where Δz is the elevation of the satellite and r is the pixel coordinate along a particular scan line as measured from the satellite nadir), the maximum possible correction was found to be of the order of 0.1 pixel widths or 6 m which was substantially less than sampling error.

The effect of non-linear scan rate can generally be accounted for by higher order polynomials. In addition the selected sub-scene extended east only 30 km from the satellite nadir, and thus the main components of this effect were approximately linear. TRINDER (1978) found that the non-linear scan rate effect was completely eliminated by the 3rd order terms in a complete 3rd order polynomial.

A3.3 Control Point Selection and Distribution

The purpose of the investigation was to allow transformation between parts of the Landsat Sydney scene and the ground truth areas. These ground truth areas were located at approximately 4 km centres and sufficient ground control to ensure a tight fit to each area was required.

Computer shadeprints and a corresponding printout of the intensity values could be produced for any specific area. To produce the shadeprints, a histogram of the intensity values was determined and nine different gray scale values were assigned to equal portions (TRINDER, 1978). The corresponding 'map' of the intensity values was achieved using a character to represent all occurrences of a decimal value.

An examination of various shadeprints from each of the four bands indicated that for urban areas Band 7 was preferable for G.C.P. selection due to the abundance of small parks, intersecting man-made linear features, small water bodies and for coastal cities natural and man-made projections into a water body. All of these features are generally against backgrounds which give significant contrast in this near infrared band.

Control points were selected from within areas of fifty by fifty pixel dimension. While this produced a more dense distribution of control points in the along scan direction, since Landsat pixels are 1.4 times longer in the across scan direction, this was considered an advantage as most non-linear scanning errors tend to have their maximum effect in this direction. The selection of the best G.C.P. in each area also meant that in some cases the points were relatively close together.

Initial control point selection was made using the shadeprints, while final estimation of pixel coordinates was achieved using the character printout. Pixel and line values for each G.C.P. were estimated to 0.5 of a pixel. For small ground features on uniform backgrounds, the displays revealed symmetric distribution of intensities if the pixel coincided exactly with the detail. Interpolation in these cases was relatively easy giving whole or half pixel values. If the ground feature was asymetric with respect to the pixel location, the distribution of intensity values was also asymetric and an estimation of the location to the nearest half pixel was required. This led to a maximum error of ± 0.25 pixel.

The coordinates of the G.C.P. were obtained from 1:10 000 series planimetric maps which were available for the whole of the Sydney region. Coordinates were estimated to the nearest 5 m by referring to the closest grid intersection and measuring the small displacements from that grid intersection.

Preferred control points were features of the order of one or two pixels in size, against a contrasting background. The centre or intersection of features was chosen in preference to the edge where the intensity change takes place. Unless enhanced, edges are smeared due to the low resolution of the Landsat scanner and an unambiguous interpretation is difficult to achieve. For specific examples of urban control points see FORSTER (1980 a).

A3.4 Results of Computation

Standard errors of the order of 30 m were achieved for both equations A3.1 and A3.2. The distribution of vector residuals is shown in Figure A3.1. At map scale each cell represents 2850 by 3950 metres therefore each cell contains 50 by 50 Landsat elements. For convenience each control point is located at the centre of the cell from which it was selected. The vector residuals at each control point are drawn at a fifty times enlargement, and so at this scale the cell represents a 57 by 79 metre picture element.

To reduce the size of the values used in the regression analysis, 250,000 metres and 1,200,000 metres were subtracted from the Zone 56 I.S.G. Easting and Northing coordinates respectively. Line and pixel values were reduced by 1000 units and divided by 10. Thus in the resulting equations X and Y values are fractions in units of 10 pixels or lines. By scaling down the values to be used in the computations, computer round-off error is minimized.

For the easting equation, A3.1, the significant variables, at a 90% confidence level or greater, were X, Y and X^5, all other variables being insignificant. The standard error of the regression as a whole using these three variables was 32.3 m and the equation was significant at the 99.9% level. The order of entry of the variables into the regression computation was X, Y and X^5, with the standard error improving from 1376 m to 33 m as the Y value entered the equation, and only a marginal but significant improvement with the entry of X^5. The coefficient of X^5 was significant at the 98.6% level.

More variables were required to adequately define the northing transformation, (A3.2). Here the significant variables were, in order of entry, Y, X, XY^2, X^5 and X^3. The resulting standard errors as each new variable was entered was 1848 m, 26.34 m, 26.29 m, 26.38 m and 24.79 m. Little if any improvement in the standard error was achieved as variables XY^2 and X^5 were entered even though their coefficients were significant at the 90% level or better. Least significant however was XY^2 in the final variable set, and this could possibly be removed with little change in the results. Overall equation significance was greater than 99.9%.

The final results for each equation were

$$E = 43564.7 + 569.19\ X - 111.69\ Y - 0.27\ X^5(10^{-7})$$

and

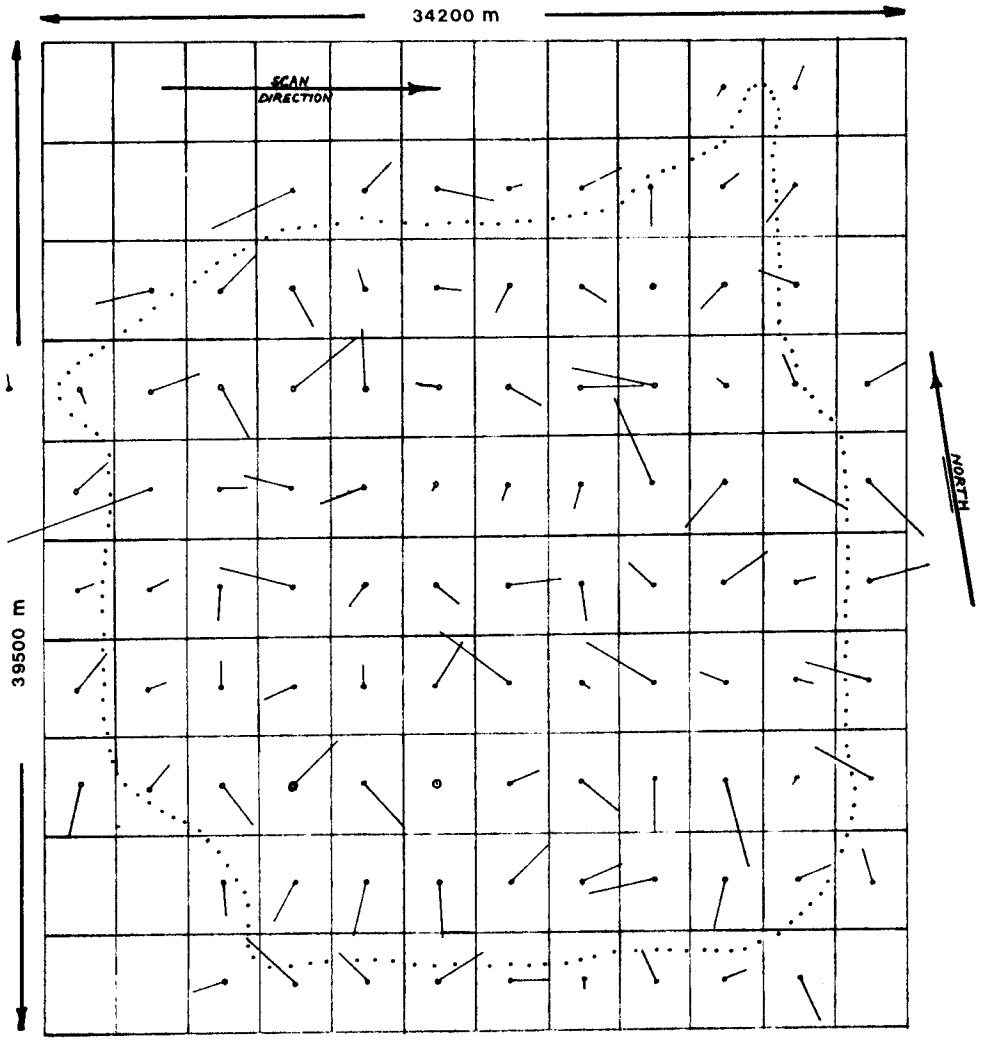


FIGURE A3.1: Vector residual distribution of 100 G.C.P. The centres of each cell represent the approximate relative position of the control points on the ground, being separated by 2850m along scan and 3950m across scan. The dotted line indicates the outer limits of the area in which ground truth sites were sampled.

$N = 126705.3 - 794.07\ Y - 108.57\ X + 0.14\ X\ Y^2(10^{-3}) + 0.18\ X^5\ (10^{-6}) - 0.15\ X^3\ (10^{-2})$, where E and N are in metres and X and Y are in units of ten pixels or lines respectively. Similar equations with X and Y as the dependent variables and polynomials in E and N as the independent variables were also derived, for use when transforming Integrated Survey Grid (I.S.G.) coordinates into the image coordinate system.

Because the use of a fifth order polynomial can lead to extensive warping especially at the edges of the control block, the derived equations were only used to predict coordinates of points well within the perimeter.

The azimuth of the across scan track was computed from the derived polynomial equations and a calculated value of $11^{\circ}\ 08'$ was determined. As the scene in question extended from the satellite nadir, east for approximately 30 km and given a circular standard error of ± 30 m, the standard error of calculated azimuth was approximately $\pm\ 0^{\circ}\ 05'$.

REFERENCES

ALONSO, W., 1964. *Location and Land Use: Toward a General Theory of Land Rents*. Harvard University Press.

AMERICAN SOCIETY OF PHOTOGRAMMETRY, 1975. *Manual of Remote Sensing*. Reeves, R.G. (ed.), Falls Church, Va.

ANDERSON, J., HARDY, E. and ROACH, J., 1972. A Land Use Classification System for Use with Remote Sensor Data. *Circular 671*, U.S. Geological Survey, Washington, D.C.

ÅNGSTROM, A.K., 1961. Techniques of Determining the Turbidity of the Atmosphere. *Tellus*, Vol. 13, No. 2.

AUSTIN, M.P. and MAYO, K.K., 1978. Some Results from Principal Component Analysis on Landsat Imagery. Paper presented at the *US-Australia Workshop on Image Processing Techniques for Remote Sensing*, Canberra.

BAILEY, M.A., 1979. The Structure of the Urban Landscape as Defined by Data from Landsat. Paper presented at the *First Australasian Landsat Conference*, Sydney.

BERNSTEIN, R., 1976. Digital Image Processing of Earth Observational Sensor Data. *IBM Journal of Research and Development*, 20., No. 1, 40-57.

BERRY, B.J.L., 1962. Sampling, Coding and Storing Flood Plain Data. *Agr. Handbook* No. 237, U.S. Department of Agriculture.

BILLINGSLEY, F.C., 1972. Some Digital Techniques for Enhancing ERTS Imagery. Paper presented at the *Remote Sensing Symp.*, Sioux Falls.

BINSELL, R., 1967. *Dwelling Unit Estimation from Aerial Photography*. Remote Sensing Laboratory, Department of Geography, North Western University, Evanston.

BOWDEN, L.W., 1968. Multi-sensor Signatures of Urban Morphology, Function and Evolutions. *Status Report* II, Tech. Rep. No. 2, Department of the Interior, Washington, D.C.

BOWDEN, L.W., 1970. *Remote Sensing of Urban Environments in Southern California*. University of California, Riverside.

BOWDEN, L.W., 1975. Urban environments: Inventory and Analysis, in Reeves, R.G. (ed.), *Manual of Remote Sensing*. American Soc. of Photogrammetry, Vol. II, 1815-1823.

BRANCH, M.C., 1948. *Aerial Photography in Urban Planning and Research*. Harvard City Planning Series, No. 14, Harvard University Press, Cambridge.

BRANCH, M.C., 1971. *City Planning and Aerial Information*. Harvard University Press, Cambridge.

BRITTENDEN, J.E., 1920. Surveying for Town Planning by the Use of Aerial Photographs. *The Garden Cities and Town Planning Magazine*, Vol. 10(4), 86-87.

BRYAN, M.L., 1974. Extraction of Urban Land Cover Data from Multiplexed Synthetic Aperture Radar Imagery. *Proceedings, 9th International Symposium on Remote Sensing of Environment* , Ann Arbor, 271-288.

CARTER, P., 1977. Automated Land Use Mapping Using Aerial/Satellite Data. *Report No. RRL77/543, Image Analysis Group,* Atomic Energy Research Establishment, Harwell, England.

CHANDRASEKHAR, S., 1960. *Radiative Transfer.* Dover Publications Inc., New York.

COCHRAN, W.G., 1963. *Sampling Techniques.* (Second Ed.), John Wiley and Sons, New York.

COLLINS, W.G. and BUSH, P.W., 1969. The Use of Aerial Photographs in the Study of Derelict Land. *Journal of Town Planning Institute,* 55, 246-255.

COLLINS, W.G. and EL-BEIK, ALY. H.A., 1971. The Acquisition of Urban Land Use Information from Aerial Photographs of the City of Leeds (Great Britain). *Photogrammetria,* Vol. 27, No. 2, 71-92.

COLWELL, R.N., 1965. The Extraction of Data from Aerial Photographs by Human and Mechanical Means. *Photogrammetric Engineering,* 20 (6), 211-228.

COLWELL, J.E., 1971. Multispectral Remote Sensing of Urban Features. *Report of Contract USGS 14-08-001-11068,* Office of Remote Sensing, Water Resources Division, U.S.G.S.

CRAIG, R.G., 1979. Autocorrelation in Landsat Data. Paper presented *13th International Symposium on Remote Sensing of Environment,* Ann Arbor.

DANA, R.W., 1975. Solar and Atmospheric Effects on Satellite Imagery Derived from Aircraft Reflectance Measurements. *Proceedings of the 10th International Symposium on Remote Sensing of Environment,* Vol. II, Ann Arbor, 683-694.

DE BRUIJN, C.A., DE HAAS, W.G.L., HOFSTEE, P., HIJLAND, A.B.M., POLLÉ, V.F.L., 1976. Urban Survey with Aerial Photography: A Time for Practice. *Proceedings of Commission Vii, 13th Congress I.S.P.,*Helsinki, 185-224.

DEERING, D.W., ROUSE, J.W., HASS, R.H. and SCHELL, J.A., 1975. Measuring "Forage Production" of Grazing Units from LANDSAT MSS Data. *Proceedings of the 10th International Symposium on Remote Sensing of Environment,* Ann Arbor, 1169-1178.

DONKER, N.H.W. and MULDER, N.J., 1976. Analysis of MSS Digital Imagery with the Aid of Principal Component Transformation. Paper presented at *Comm. VII, 13th Congress of I.S.P.,* Helsinki.

DRAPER, N.R. and SMITH, H., 1966. *Applied Regression Analysis.* John Wiley and Sons, New York.

DUEKER, K.J. and HORTON, F.E., 1972. Utility of Remote Sensing Data for Urban Land Use Planning. *Proceedings of Operational Remote Sensing Seminar, The American Society of Photogrammetry,* Houston, 178-190.

DYE, R.H., 1975. Restoration of Landsat Images by Discrete Two-Dimensional Doncovolution. *Proceedings of the 10th International Symposium on Remote Sensing of Environment*, Vol. 11, Ann Arbor, 725-730.

ELLEFSEN, R., SWAIN, P.H., WRAY, J.R., 1973. Urban Land Use Mapping by Machine Processing of ERTS-1 Multispectral Data: A San Francisco Bay Area Example. *Lars Information Note 032973*, Purdue.

ELLEFSEN, R., GAYDOS, L., SWAIN, P., WRAY, J.R., 1974. New Techniques in Mapping Urban Land Use and Monitoring Change for Selected U.S. Metropolitan Areas, an Experiment Employing Computer Assisted Analysis of ERTS-1 MSS Data. *Proceedings of Symposium on Remote Sensing and Photo Interpretation, International Society for Photogrammetry, Commission VII*, Banff, Canada, 51-63.

ELTERMAN, L., 1968. UV, Visible and IR Attenuation for Altitudes to 50km. *Project AFCRL-68-0153*, Air Force Cambridge Research Laboratories, Cambridge, Mass.

ENTRES, S.L., 1974. Fundamentals of Remote Sensing of the Earth in *Proceedings of the 1st Technical Session of the Remote Sensing Society* (Ed. COLLINS, W.G. and VAN GENDEREN, J.L.) Remote Sensing Society, 43.

ERB, R.B., 1974. ERTS 1 Urban Land Use Analysis. Report for period July 1972 - June 1973, in *The ERTS-1 Investigation (ER-600)*, Vol. V, NASA TM X-58121.

EYRE, A.L., ADOLPHUS, B., AMIEL, M., 1970. Census Analysis and Population Studies. *Photogrammetric Engineering* 36(5), 460-466.

EYRE, L.A., 1971. High Altitude Color Photos. *Photogrammetric Engineering* 37(11), 1149-1153.

FORSTER, B.C., 1975. *The Derivation of an Urban House Price Model.* M.Sc. Thesis, University of Reading, U.K.

FORSTER, B.C., 1980a. Urban control for Landsat data. *Photogrammetric Engineering and Remote Sensing*, Vol. 46, No. 4, 539-545.

FORSTER, B.C., 1980b. Urban Residential Ground Cover Using Landsat Digital Data. *Photogrammetric Engineering and Remote Sensing*, Vol. 46, No. 4, 547-558.

FRIEDMAN, S.Z., 1978. *Use of the Ratio Threshold Classifier in Mapping Urban/Non-Urban Land Cover from Landsat MSS Imagery.* M.Sc. Thesis, University of Wisconsin, Madison, Wisconsin, U.S.A.

FRIEDMAN, S.Z. and ANGELICI, G.L., 1979. The Detection of Urban Expansion from Landsat Imagery. *Remote Sensing Quarterly*, Vol. 1, No. 1, 58-79.

GIBSON, L.J., 1976. *The Use of Aerial Photography for Spoiled and Derelict Land Studies: with Special Reference to Glamorgan.* Unpublished Ph.D. Thesis, University of Aston, 283p.

GOODCHILD, B., 1974. Class Differences in Environmental Perception: an Exploratory Study. *Urban Studies*, Vol. 11, No. 2, 157-169.

GREEN, N.E., 1957. Aerial Photographic Interpretation and the Social Structure of the City. *Photogrammetric Engineering* 23 (1), 89-96.

GREEN, N.E. and MONIER, R.B., 1959. Aerial Photographic Interpretation of the Human Ecology of the City. *Photogrammetric Engineering* 25(5), 770-773.

GRIFFITHS, T.B., HOWARD, W.A. and KRACHT, J.B., 1971. Developing Remote Sensing Display Modes to Satisfy Urban Planning Data Input Needs. *Interagency Report USGS-207*, Department of Geography, University of Denver, Denver.

GWYNN, A.P., 1968. *Aerial Photographic Interpretation and Land Use Classification*. Regional Planning Council, Baltimore.

HADFIELD, S.M., 1963. *An Evaluation of Land Use and Dwelling Unit Data Derived from Aerial Photography*. Chicago Area Transportation Study, Chicago.

HANNAH, J.W., 1969. A Feasibility Study for the Application of Remote Sensors to Selected Urban and Regional Land Use Planning Studies. *Technical Report No. 11*, East Tennessee State University, Johnson City.

HAYLER, G.W., 1920. The Airplane and City Planning. *The American City*, Vol. 23(6), 575-579.

HENDERSON, R.G., THOMAS, G.S. and NALEPKA, R.F., 1975. Methods of Extending Signatures and Training without Ground information. *ERIM 109600-16-F*, Ann Arbor, Michigan.

HERZOG, J.H. and STURM, B., 1975. Preprocessing Algorithms for the Use of Radiometric Corrections and Texture/Spatial Features in Automatic Land Use Classification. *Proceedings of the 10th International Symposium on Remote Sensing of Environment*, Ann Arbor, Vol. 11, 705-724.

HEYDORN, R.P., TRICHEL, M.C. and ERICKSON, J.D., 1978. Methods of Segment Wheat Area Estimation. *Proceedings of Technical Sessions, The LACIE Symposium*, Houston, Vol. 11, 621-632.

HOFFER, R.M., 1967. Interpretation of Remote Multispectral Imagery of Agricultural Crops. *Research Bulletin No. 831*, Laboratory for Remote Sensing, Purdue University, Lafayette.

HOFFER, R.M. and GOODRICK, F.E., 1971. Variables in Automatic Classification Over Extended Remote Sensing Test Sites. *Proceedings of the 7th International Symposium on Remote Sensing of Environment*, Ann Arbor, Vol. III, 1967-1982.

HORTON, F.E. and MARBLE, D.F., 1969. Housing Quality in Urban Areas: Data Acquisition and Classification Through the Use of Remote Sensor Imagery. Paper presented at the *NASA-MSC Program Review*, Houston, September, 1969.

HORWITZ, H., NALEPKA, R.F. and HYDE, P.D., 1971. Estimating the Proportions of Objects within a Single Resolution Element of a Multispectral Scanner. *Proceedings of the 7th International Symposium on Remote Sensing of Environment*, Ann Arbor, 1307-1320.

HOWARD, J.A., 1970. *Aerial Photo-Ecology*. Faber and Faber, London.

HSU, SHIN YI, 1970. Inter-Census Population Estimation by the Use of Photos and USGS Topographic Maps; a Sample Study of Population Change in the Atlanta Area. Paper presented at *ASP-ACSM Annual Convention*, Washington, D.C., March, 1970.

HUGHES AIRCRAFT CO., 1972. Multispectral Scanner Systems for ERTS: Four Band Scanner System. *NASA-CR-132758*, Vol. 1.

HULTQUIST, N.B., RUSHTON, G. and SCHMITT, R.P., 1971. Identifying and Forecasting Change in a Regional System of Urban Places. *Proceedings of the 7th International Symposium on Remote Sensing of Environment*, Ann Arbor, Vol. II, 1553-1562.

JENNISON, R.C., 1961. *Fourier Transforms and Convolutions for the Experimentalist*. Permagon Press, London.

JENSEN, J.R., 1978. Digital Land Cover Mapping Using Layered Classification Logic and Physical Composition Attributes. *The American Cartographer*, Vol. 5, No. 2, 121-132.

JOERG, W.L.G., 1923. The Use of Airplane Photography in City Geography. *Annals of the Association of American Geographers*, XIII, 211.

JOHNSON, W.B., 1961. *Satellite Environment Handbook*. Standford University Press.

JOHNSON, R.W., 1975. Quantitative Suspended Sediment Mapping Using Aircraft Remotely Sensed Data. *Proceedings of the Earth Resources Survey Symposium*, Houston.

JOHNSON, R.W., 1978. Mapping of Chlorophyll a Distributions in Coastal Zones. *Photogrammetric Engineering and Remote Sensing*, Vol. 44, No. 5, 617-624.

KAUTH, R.J., 1975. Soil Reflectance. *Memo TF3-75-5-190*, NASA Johnson Space Center.

KAUTH, R.J. and THOMAS, G.C., 1976. The Tasselled Cap - A Graphic Description of the Spectral - Temporal Development of Agricultural Crops as seen by Landsat. *Proceedings of the Symposium on Machine Processing of Remotely Sensed Data*, Lars, Purdue, 6/29-7/72.

KAUTH, R.J., LAMBECK, P.F., RICHARDSON, W., THOMAS, G.S. and PENTLAND, A.P., 1978. Feature Extraction Applied to Agricultural Crops as seen by Landsat. *Proceedings of Technical Sessions, The LACIE Symposium*, Houston, Vol. II, 705-721.

KAWATA, Y., HABA, Y., KUSAKA, T., TERASHITA, Y. and UENO, S., 1978. Atmospheric Effects and their Correction in Airborne Sensor and Landsat MSS Data. *Proceedings of the 12th International Symposium on Remote Sensing of Environment*, Vol. II, Manila, 1241-1257.

KERLINGER, F.N. and PEDHAZUR, E.G., 1973. *Multiple Regression in Behavioural Research*. Holt, Rinehart and Winston, New York.

KHORRAM, S., 1979. Remote Sensing Analysis of Water Quality in the San Francisco Bay - Delta. *Proceedings of the Thirteenth International Symposium on Remote Sensing of Environment*, Ann Arbor.

KRATKY, V., 1975. Precision Processing of ERTS Imagery. Paper presented at *ACSM-ASP Semi-Annual Convention*, San Francisco.

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 Quarterly*, Vol. 1, No. 1, 80-105.

LATHAM, J.P., 1959. A Study of the Application of Electronic Scanning and Computer Devices to the Analysis of Geographic Phenomena. *Final Report*, Office of Naval Research, Washington.

LEE, W.T., 1920. Airplanes in Geography. *The Geographic Review*, X, 313.

LEWIS, A.J., 1968. Evaluation of Multiple Polarized Radar Imagery for the Detection of Selected Cultural Features. *Technical Letter NASA-130*, National Aeronautics and Space Administration.

LEWIS, A.J., MAC DONALD, H.C. and SIMONETT, D.S., 1969. Detection of Linear Cultural Features with Multipolarized Radar Imagery. *Proceedings 6th International Symposium on Remote Sensing of Environment*, Ann Arbor, 879-893.

LEWIS, N.P., 1922, A New Aid in City Planning. *The American City*, Vol. 26(3).

LILLESAND, J.M., EAV, B.B. and MANION, P.D., 1979. Quantifying Urban Tree Stress through Microdensitometric Analysis of Aerial Photography. *Photogrammetric Engineering and Remote Sensing*, Vol. 45, No. 10, 1401-1410.

LINDGREN,D.T., 1971. Dwelling Unit Estimation with Color-IR Photos. *Photogrammetric Engineering* 37(5), 373-378.

LODWICK, G.D., 1979. Measuring Ecological Changes in Multitemporal Landsat Data Using Principal Components. Paper presented *13th International Symposium on Remote Sensing of Environment*, Ann Arbor.

LODWICK, G.D., 1980. Private Communication.

McCLOY, K., 1978. Mapping Landcover in the South East of South Australia. Paper presented at the *U.S.-Australia Workshop on Image Processing Techniques for Remote Sensing*, Canberra.

MAC PHAIL, D.D. and CAMPBELL, L.F., 1970. The El-Paso (Texas-New Mexico) Study Area: A Comprehensive Analysis of Gemini 4 and Apollo 6 and 9 Space Photography. Paper presented at *AAG Commission on Geographic Applications of Remote Sensing*, East Tennessee State University, Johnson City.

MALINVAND, E., 1966. *Statistical Methods of Econometrics*. Trans. by A. Silvery Amsterdam, Worth-Holland Pub. Co.

MATHER, P.M., 1976. *Computational Methods of Multivariate Analysis in Physical Geography*. John Wiley and Sons, London.

MATTHES, G.H., 1927. Aerial Surveys for City Planning. *American Society of Civil Engineering, Transactions*, Vol. 91, 314-325.

MAXWELL, E.L., 1976. Multivariate System Analysis of Multispectral Imagery. *Photogrammetric Engineering and Remote Sensing*, Vol. 42, No. 9, 1173-1186.

METIVIER, E.D. and McCOY, R.M., 1971. Mapping Urban Poverty Housing from Aerial Photography. *Proceedings 7th International Symposium on Remote Sensing of Environment*, Ann Arbor, University of Michigan, Vol. 2, 1563-1570.

MIDDLETON, W.E.K., 1952. *Vision Through the Atmosphere*. Univ. of Toronto Press.

MILLER, L.D. and PEARSON, R.L., 1971. Aerial Mapping Program of the IBP Grassland Biome: Remote Sensing of the Productivity of Shortgrass Prairie as Input into Biosystem Models. *Proceedings of the 7th International Symposium on Remote Sensing of Environment*, Univ. of Michigan, Vol. 1, 165-207.

MILTON, E.J., 1978. A Portable "Landsat" Radiometer for Ground Data Collection in Remote Sensing. Paper presented at the *5th Annual Conference of the Remote Sensing Society*, Durham.

MOORE, E.G., 1968. Side Looking Radar in Urban Research: A Case Study. *Technical Letter NASA-138*. National Aeronautics and Space Administration.

MOORE, E.G., 1969a. *Collection of Housing Quality Data by Remote Sensor: The Los Angeles Experiment*. Remote Sensing Laboratory, Dept. of Geography, Northwestern University, Evanston.

MOORE, E.G., 1969b. *Application of Housing Aggregates to Quality Classes: The Potential of Photographic Sensors*. Remote Sensing Laboratory, Dept. of Geography, Northwestern University, Evanston.

MOORE, E.G., 1970. *Application of Remote Sensors to the Classification of Areal Data at Different Scales: A Case Study in Housing Quality*. Remote Sensing Laboratory, Dept. of Geography, Northwestern University, Evanston.

MORRISON, A., 1978. Landsat Digital Data for Population Estimation: Testing the Technique with Gridsquare Census Data in the Scottish Highlands. Paper presented at the *5th Annual Conference of the Remote Sensing Society*, Durham, England, Dec. 1978.

MUMBOWER, L. and DONOGHUE, J., 1967. Urban Poverty Study. *Photogrammetric Engineering* 33 (6), 610-618.

MURAI, S., 1974. Estimation of Population Density in Tokyo Districts from ERTS-1 Data. *Proceedings of the 9th International Symposium on Remote Sensing of Environment*, Vol. 1, Ann Arbor, 13-22.

NALEPKA, R.F., HORWITZ, H.M. and THOMSON, N.S., 1971. Investigation of Multispectral Sensing of Crops. *Report No. 31650-30-T*, Willow Run Laboratories, The University of Michigan, Ann Arbor, Michigan.

NARIGASAWA, K. and FUCHIMOTO, M. 1979. Urban Environmental Survey by Remote Sensing. Presented *Thirteenth International Symposium on Remote Sensing of Environment*, Ann Arbor.

NASA 1972. ERTS Data Users' Handbook. *Document No. 7154249*, NASA, Goddard Space Flight Center, Greenbelt, Maryland.

NIE, N.H., HADLAI HULL, C., JENKINS, J.G., STEINBRENNER, K. and BENT, D.H., 1975. *Statistical Package for the Social Sciences*. (Second Edition). McGraw Hill, New York.

NUNNALLY, N.R. and WITMER, R.E., 1968. A Strategy for Developing Classification of Land Use as Interpreted from Remote Sensing Imagery. *Technical Report No. 1*, Atlantic University, Florida.

NUNNALLY, N.R., 1970. A Land Use Interpretation Experiment. *Technical Report 69-5*, East Tennessee State University Association of American Geographers.

PETERSON, J.B., BARRETT, F.R. and BECK, R.H., 1979. Predictability of Change in Soil Reflectance on Wetting. *Proceedings of the Symposium on Machine Processing of Remotely Sensed Data*, Purdue University, 264-268.

PICK, M., 1920. Aerial Photography. *The Garden Cities and Town Planning Magazine*, Vol. 10(4), 71-86.

PIECH, K.R., SCHOTT, J.R. and STEWART, K.M., 1978. The Blue-to-Green Reflectance Ratio and Lake Water Quality. *Photogrammetric Engineering and Remote Sensing*, Vol. 44, No. 10, 1303-1310.

PUESCHEL, R.F., CHARLSON, R.J. and AHLQUIST, N.C., 1969. On the Anomalous Deliquescence of Sea-Spray Aerosols. *Journal Applied Meteorology*, Vol. 8, 995-998.

RICHARDSON, A.J. and WIEGAND, C.L., 1977. Distinguishing Vegetation from Soil Background Information. *Photogrammetric Engineering and Remote Sensing*, Vol. 43, No. 12, 1541-1552.

RICHARDSON, H.W., 1971. *Urban Economics*. Penguin.

RITCHIE, J.C., McHENRY, J.R., SCHIEBE, F.R. and WILSON, R.B., 1974. The Relationship of Reflected Solar Radiation and the Concentration of Sediment in the Surface Water of Reservoirs. Paper presented *Conference on Earth Resources Observations and Information Analysis System*, Tullahoma, Tennessee.

ROBINSON, N., 1966. *Solar Radiation*. Elsevier Publishing Co., New York.

ROGERS, R.H., SHAH, N.J., McKEON, J.B., WILSON, C., REED, L., SMITH, V.E. and THOMAS, N., 1975. Application of Landsat to the Surveillance and Control of Eutrophication in Saginaw Bay. Paper presented *Tenth International Symposium on Remote Sensing of Environment*, Ann Arbor.

ROSENBERG, G.V., 1967. The Properties of an Atmospheric Aerosol from Optical Data. *Izv. Atms. and Oceanic Phys.*, Vol. 3, 545-551, in Engl. Transl.

ROSENFELD, A., 1962a. An approach to Automatic Photographic Interpretation. *Photogrammetric Engineering* 28, No. 4, 660.

ROSENFELD, A., 1962b. Automatic Recognition of Basic Terrain Types from Aerial Photographs. *Photogrammetric Engineering*, Vol. 28, No. 1, 115-132.

ROSENFELD, A., 1965. Automatic Imagery Interpretation. *Photogrammetric Engineering* 31(2), 240-242.

ROSENFELD, A., 1966. Image Processing, in *Five Papers on Remote Sensing and Urban Information Systems*. Dept. of Geography, Northwestern University, Evanston.

ROUSE, J.W., HASS, R.H., SCHELL, J.A. and DEERING, D.W., 1973. Monitoring Vegetation Systems in the Great Plains with ERTS. *Proceedings of the Third ERTS Symposium, NASA SP-351*, Vol. I, 309-317.

RUSHTON, G. and HULTQUIST, N., 1970. Remote Sensing Techniques for Evaluating Systems of Cities: A Progress Report. *Technical Report No. 2*, Institute of Urban and Regional Research, University of Iowa, Iowa City.

SABATINI, R.R. and RABCHEVSKY, G., 1970. Use of Ground-Truth Measurements to Monitor ERTS Sensor Calibration. *Technical Report No. 16*, NASA Contract No. NAS 5-10343.

SCARPACE, F.L., HOLMQUIST, K.W. and FISHER, L.T., 1979. Landsat Analysis of Lake Quality. *Photogrammetric Engineering and Remote Sensing*, Vol. 45, No. 5, 623-633.

SHARMA, R.D., 1972. Enhancement of Earth Resources Technology Satellite (ERTS) and Aircraft Imagery Using Atmospheric Corrections. *Proceedings of the 8th International Symposium on Remote Sensing of Environment*, Ann Arbor, Vol. 1, 137-152.

SIMONETT, D.S., 1969. Questions on Thematic mapping: Some Preliminary Observations in *The Utility of Radar and Other Remote Sensors in Thematic Land Use Mapping from Spacecraft*. (SIMONETT, D.S. (Ed.)) Annual Report, Technical Letter NASA-140.

SIMONETT, D.S. and COINER, J.C., 1971. The Susceptibility of Environments to Low Resolution Imaging. *Proceedings of the 7th International Symposium on Remote Sensing of Environment*, Vol. 1, Ann Arbor, 373-394.

SIMPSON, R.B., 1969. Geographic Evaluation of Radar Imagery of New England. *Interagency Report NASA-163*, National Aeronautics and Space Administration.

SIMPSON, R.B., 1970. *Production of a High Altitude Land Use Map and Data Base for Boston.* Department of Geography, Dartmouth College, Hannover.

SMEDES, H.W., HULSTROM, R.L. and RANSOM, K., 1975. The Mixture Problem in Computer Mapping of Terrain: Improved Techniques for Establishing Spectral Signatures, Atmospheric Path Radiance and Transmittance. *Proceedings of NASA Earth Resource Survey Symposium*, Houston, National Aeronautics and Space Administration, 1099-1156.

STEINER, D. and SALERNO, A.E., 1975. Remote Sensor Data Systems, Processing and Management, in *Manual of Remote Sensing*, Vol. 1, 631, R.G. Reeves (ed.) American Society of Photogrammetry, Falls Church, Va..

SWAIN, P., LANDGREBE, D. and HAUSKA, H., 1975. Layered Classification Techniques for Remote Sensing Application. *Proceedings of NASA Earth Resources Survey Symposium*, Houston, 1088-1097.

SWAIN, P.H., 1976. Land Use Classification and Mapping by Machine - Assisted Analysis of LANDSAT Multispectral Scanner Data. *LARS Information Note*, 111276, Purdue.

THEIL, H., 1971. *Principles of Econometrics.* John Wiley and Sons, New York.

THOMAS, J.R. and GERBERMANN, A.H., 1977. Yield/Reflectance Relations in Cabbage. *Photogrammetric Engineering and Remote Sensing*, Vol. 43, No. 10, 1257-1266.

THOMPSON, D.R. and WEHMANEN, O.A., 1979. Using Landsat Digital Data to Detect Moisture Stress. *Photogrammetric Engineering and Remote Sensing*, Vol. 45, No. 2, 201-207.

THROWER, N.J., MULLENS, R.H. and SENGER, L.W., 1968. Analysis of NASA Color Infrared Photography of Los Angeles, in *Mission 73 Summary and Data Catalog*, (PASCUCCI, R.F. and NORTH, G.W. (eds)). National Aeronautics and Space Administration.

THROWER, N.J., MULLENS, R.H., SENGER, L.W. and WALTON, K.J., 1970. Satellite Photography as a Geographic Tool for Land Use Mapping of the South Western U.S., 1 July 1968 - 31 January 1970. *Technical Report 69-3*, East Tennessee State University, Johnson City.

TIMMS, D., 1971. *The Urban Mosaic.* Cambridge Geographical Studies, Cambridge University Press.

TODD, W. and BAUMGARDNER, M.F., 1973. Land Use Classification of Marion County, Indiana, by Spectral Analysis of Digitized Satellite Data. *Lars Information Note 101673*, Purdue.

TODD, W., MAUSEL, P. and BAUMGARDNER, M.F., 1973. An Analysis of Milwaukee County Land Use by Machine Processing of ERTS Data. *LARS Information Note 022773*, Purdue.

TRINDER, J.C., 1970. *Accuracy of Pointing to Blurred Photogrammetric Signals.* Ph.D. Thesis, University of N.S.W., Sydney.

TRINDER, J.C., 1978. Rectification of Landsat Digital Data. Paper presented at the *U.S.-Australia Workshop on Image Processing Techniques for Remote Sensing*, Canberra.

TUBBS, J.D. and COBERLY, W.A., 1978. Spatial Correlation and its Effect upon Classification Results in Landsat. Paper presented *12th International Symposium on Remote Sensing of Environment*, Manila.

TUCKER, C.J., 1973. *The Remote Estimation of a Grassland Canopy*. M.S. Thesis, Colorado State University, Fort Collins, Colo..

TUCKER, C.J., 1977. A Spectral Method for Determining the Percentage of Live Herbage Material in Clipped Samples. *NASA Technical Memo. 79010*, GSFC.

TUCKER, C.J. and MILLER, L.D., 1977. Soil Spectra Contributions to Grass Canopy Spectral Reflectance. *Photogrammetric Engineering and Remote Sensing*, Vol. 43, No. 6, 721-726.

TUCKER, C.J., 1978. A Comparison of Satellite Sensor Bands for Vegetation Monitoring. *Photogrammetric Engineering and Remote Sensing*, Vol. 44, No. 11, 1369-1380.

TUELLER, P.T. and LORAIN, G., 1971. Environmental Analysis of the Lake Tahoe Basin from Small Scale Multispectral Imagery. *Proceedings of the 7th International Symposium on Remote Sensing of Environment*, Vol. 1, Ann Arbor, 453-468.

TURNER, R.E., MALILA, W.D. and NALEPKA, R.F., 1971. Importance of Atmospheric Scattering in Remote Sensing or everything you've wanted to know about atmospheric scattering but were afraid to ask. *Proceedings of the 7th International Symposium on Remote Sensing of Environment*, Vol. III, Ann Arbor, 1651-1695.

TURNER, R.E. and SPENCER, M.M., 1972. Atmospheric Model for Correction of Spacecraft Data. *Proceedings of the 8th International Symposium on Remote Sensing of Environment*, Vol. II, Ann Arbor, 895-934.

TURNER, R.E., 1975. Signature Variations due to Atmospheric Effects. *Proceedings of the 10th International Symposium on Remote Sensing of Environment*, Vol. II, Ann Arbor, 671-682.

TUYAHOV, A.J., DAVIES, S. and HOLZ, R.K., 1973. Detection of Blight Using Remote Sensing Techniques, in *Remote Sensing of Earth Resources*, Vol. 2, (ed.) Shahrokhi, F., Space Institute, University of Tennessee, 213-226.

WELLAR, B.S., 1968a. Generation of Housing Quality Data from Multiband Aerial Photographs. *Technical Letter NASA-119*, NASA Manned Spacecraft Center, Houston.

WELLAR, B.S., 1968b. *Hyperaltitude Photography as a Data Base in Urban and Transportation Research*. Remote Sensing Laboratory, Department of Geography, Northwestern University, Evanston.

WELLAR, B.S., 1968c. Thermal Infrared Imagery in Urban Studies. *Technical Letter NASA-135*, NASA Manned Spacecraft Center, Houston.

WELLAR, B.S., 1969. The Role of Space Photography in Urban and Transportation Data Series. *Proceedings of the 6th Symposium on Remote Sensing of Environment*, Vol. II, Ann Arbor, 831.

WESTERLUND, F.V., 1972. Remote Sensing for Planning: a Bibliography and Review of Literature. *Urban Planning/Development Series*, No. 10, Department of Urban Planning, University of Washington, Washington.

WHITLOCK, C.H., 1977. Fundamental Analysis of the Linear Multiple Regression Technique for Quantification of Water Quality Parameters from Remote Sensing. *NASA Report TM X-74600*, National Aeronautics and Space Administration.

WIEGAND, C.L., HEILMAN, M.D. and GERBERMANN, A.H., 1968. Detailed Plant and Soil Thermal Regime in Agronomy. *Proceedings of the 5th International Symposium on Remote Sensing of Environment*, Ann Arbor, 325-342.

WITENSTEIN, M.M., 1954. Photo Sociometrics. The Application of Aerial Photography to Urban Administration and Planning Problems. *Photogrammetric Engineering*, 20(3), 410-427.

WITENSTEIN, M.M., 1955. Uses and Limitations of Aerial Photography in Urban Analysis and Planning. *Photogrammetric Engineering* 21(4), 566-573.

WITENSTEIN, M.M., 1956. A Report on Application of Aerial Photography to Urban Use Inventory, Analysis and Planning. *Photogrammetric Engineering*, 22(4), 656-664.

WITENSTEIN, M.M., 1957. A Review of Aerial Photographic Interpretation and the Social Structure of the City. *Photogrammetric Engineering*, 23(2), 97.

WRAY, J.R., 1960. Photo Interpretation in Urban Area Analysis, in *Manual of Photographic Interpretation* (R.N. Colwell, ed.) American Society of Photogrammetry, Washington, D.C., 667-703.

WRAY, J.R., 1970. Census Cities Project and Atlas of Urban and Regional Change. *Proceedings of Third Annual Earth Resources Program Review*, Vol. 1, Houston, 2-1 - 2-16.

WRAY, J.R., 1971. *An Operational Remote Sensing Land Use Change Detection System for Metropolitan Areas*. Paper presented at the 7th International Symposium on Remote Sensing of Environment, Ann Arbor, Michigan.

ZOBRIST, A.L., BRYANT, N.A. and LANDINI, A.J., 1976. IBIS - A Geographical Information System That Can Use Satellite Data for Urban Analysis. Paper presented *URISA National Conference*, Atlanta, Georgia.

Publications from

THE SCHOOL OF SURVEYING, THE UNIVERSITY OF NEW SOUTH WALES
P.O. Box 1, Kensington, N.S.W. 2033
AUSTRALIA

Reports

1.* G.G. Bennett, "The discrimination of radio time signals in Australia", Uniciv Rep. D-1, 88 pp. (G 1)

2.* J.S. Allman, "A comparator for the accurate measurement of differential barometric pressure", Uniciv Rep. D-3, 9 pp. (G 2)

3. R.S. Mather, "The establishment of geodetic gravity networks in South Australia", Uniciv Rep. R-17, 26 pp. (G 3)

4. R.S. Mather, "The extension of the gravity field in South Australia", Uniciv Rep. R-19, 26 pp. (G 4)

5.* J.S. Allman, "An analysis of the reliability of barometric elevations", Unisurv Rep. 5, 335 pp. (S 1)

6.* R.S. Mather, "The free air geoid for South Australia and its relation to the equipotential surfaces of the earth's gravitational field", Unisurv Rep. 6, 491 pp. (S 2)

7.* P.V. Angus-Leppan (Editor), "Control for mapping" (Proceedings of Conference, May 1967), Unisurv Rep. 7, 329 pp. (G 5)

8.* G.G. Bennett & J.G. Freislich, "The teaching of field astronomy", Unisurv Rep. 8, 30 pp. (G 6)

9.* J.C. Trinder, "Photogrammetric pointing accuracy as a function of properties of the visual image", Unisurv Rep. 9, 64 pp. (G 7)

10.* P.V. Angus-Leppan, "An experimental determination of refraction over an icefield", Unisurv Rep. 10, 23 pp. (G 8)

11.* R.S. Mather, "The non-regularised geoid and its relation to the telluroid and regularised geoids", Unisurv Rep. 11, 49 pp. (G 9)

12.* G.G. Bennett, "The least squares adjustment of gyro-theodolite observations", Unisurv Rep. 12, 53 pp. (G 10)

13.* R.S. Mather, "The free air geoid for Australia from gravity data available in 1968", Unisurv Rep. 13, 38 pp. (G 11)

14.* R.S. Mather, "Verification of geoidal solutions by the adjustment of control networks using geocentric Cartesian co-ordinate systems", Unisurv Rep. 14, 42 pp. (G 12)

* Out of print

15.* G.G. Bennett, "New methods of observation with the Wild GAKI gyro-theodolite", Unisurv Rep. 15, 68 pp. (G 13)

16.* G.G. Bennett, "Theoretical and practical study of a gyroscopic attachment for a theodolite", Unisurv Rep. 16, 343 pp. (S 3)

17. J.C. Trinder, "Accuracy of monocular pointing to blurred photogrammetric signals", Unisurv Rep. 17, 231 pp. (S 4)

18. A. Stolz, "The computation of three dimensional Cartesian co-ordinates of terrestrial networks by the use of local astronomic vector systems", Unisurv Rep. 18, 47 pp. (G 14)

19. R.S. Mather, "The Australian geodetic datum in earth space", Unisurv Rep. 19, 130 pp. (G 15)

20.* J.G. Fryer, "The effect of the geoid on the Australian geodetic network", Unisurv Rep. 20, 221 pp. (S 5)

21.* G.F. Toft, "The registration and cadastral survey of native-held rural land in the Territory of Papua and New Guinea", Unisurv Rep. 21, 441 pp. (S 6)

22. R.S. Mather et al, "Communications from Australia to Section V, International Association of Geodesy, XV General Assembly, International Union of Geodesy and Geophysics, Moscow 1971", Unisurv Rep. 22, 72 pp. (G 16)

23. A.H. Campbell, "The dynamics of temperature in surveying steel and invar measuring bands", Unisurv Rep. S 7, 195 pp.

24. A. Stolz, "Three-D Cartesian co-ordinates of part of the Australian geodetic network by the use of local astronomic vector systems", Unisurv Rep. S 8, 182 pp.

25. Papers by R.S. Mather, H.L. Mitchell & A. Stolz on the following topics:- Four-dimensional geodesy, Network adjustments and Sea surface topography, Unisurv G 17, 23 pp.

26. Papers by L. Berlin, G.J.F. Holden, P.V. Angus-Leppan, H.L. Mitchell & A.H. Campbell on the following topics:- Photogrammetry co-ordinate systems for surveying integration, Geopotential networks and Linear measurement, Unisurv G 18, 80 pp.

27. R.S. Mather, P.V. Angus-Leppan, A. Stolz & I. Lloyd, "Aspects of four-dimensional geodesy", Unisurv G 19, 100 pp.

28. H.L. Mitchell, "Relations between MSL & geodetic levelling in Australia", Unisurv Rep. S 9, 264 pp.

29. A.J. Robinson, "Study of zero error & ground swing of the model MRA101 tellurometer", Unisurv Rep. S 10, 200 pp.

30. Papers by J.S. Allman, R.C. Lister, J.C. Trinder & R.S. Mather on the following topics:- Network adjustments, Photogrammetry and 4-Dimensional geodesy, Unisurv G 20, 133 p.

31. G.J.F. Holden, "An evaluation of orthophotography in an integrated mapping system", Unisurv Rep. S 12, 232 pp.

32. G.J. Hoar, "The analysis precision and optimization of control surveys", Unisurv Rep. S 13, 200 pp.

33. Papers by E. Grafarend, R.S. Mather & P.V. Angus-Leppan on the following topics:- Mathematical geodesy, Coastal geodesy and Refraction, Unisurv G 21, 100 pp.

34. Papers by R.S. Mather, J.R. Gilliland, F.K. Brunner, J.C. Trinder, K. Bretreger & G. Halsey on the following topics:- Gravity, Levelling, Refraction, ERTS imagery, Tidal effects on satellite orbits and photogrammetry, Unisurv G 22, 96 pp.

35. Papers by R.S. Mather, E.G. Anderson, C. Rizos, K. Bretreger, K. Leppert, B.V. Hamon & P.V. Angus-Leppan on the following topics:- Earth tides, Sea surface topography, Atmospheric effects in physical geodesy, Mean sea level and Systematic errors in levelling, Unisurv G 23, 96 pp.

36. Papers by R.C. Patterson, R.S. Mather, R. Coleman, O.L. Colombo, J.C. Trinder, S.U. Nasca, T.L. Duyet & K. Bretreger on the following topics:- Adjustment theory, Sea surface topography determinations, Applications of Landsat imagery, Ocean loading of Earth tides, Physical geodesy, Photogrammetry and Oceanographic applications of satellites, Unisurv G 24.

37. E.G. Anderson, "The Effect of Topography on Solutions of Stokes' Problem", Unisurv Rep. S 14, 252 pp.

38. A.H.W. Kearsley, "The Computation of Deflections of the Vertical from Gravity Anomalies", Unisurv Rep. S 15, 181 pp.

39. Papers by S.M. Nakiboglu, B. Ducarme, P. Melchior, R.S. Mather, B.C. Barlow, C. Rizos, B. Hirsch, K. Bretreger, F.K. Brunner & P.V. Angus-Leppan on the following topics:- Hydrostatic equilibrium figures of the Earth, Earth tides, Gravity anomaly data banks for Australia, Recovery of tidal signals from satellite altimetry, Meteorological parameters for modelling terrestrial refraction and Crustal motion studies in Australia, Unisurv G 25.

40. Papers by R.S. Mather, E.G. Masters, R. Coleman, C. Rizos, B. Hirsch, C.S. Fraser, F.K. Brunner, P.V. Angus-Leppan, A.J. McCarthy & C. Wardrop on the following topics:- Four-dimensional geodesy, GEOS-3 altimetry data analysis, Analysis of meteorological measurements for microwave EDM and Meteorological data logging system for geodetic refraction research, Unisurv G 26, 113 pp.

41. Papers by F.K. Brunner, C.S. Fraser, S.U. Nasca, J.C. Trinder, L. Berlin, R.S. Mather, O.L. Colombo & P.V. Angus-Leppan on the following topics:- Micrometeorology in geodetic refraction, Landsat imagery in topographic mapping, Adjustment of large systems, GEOS-3 data analyses, Kernel functions and EDM reductions over sea, Unisurv G 27, 101 pp.

42. K. Bretreger, "Earth Tide Effects on Geodetic Observations", Unisurv S 16, 173 pp.

43. Papers by S.M. Nakiboglu, H.L. Mitchell, K. Bretreger, T.A. Herring, J.M. Rüeger, K.R. Bullock, R.S. Mather, B.C. Forster, I.P. Williamson & T.S. Morrison on the following topics:- Variations in gravity, Oceanographic and geodetic levelling, Ocean loading effects on Earth tides, Deflections of the vertical, Frequencies of EDM instruments, Land information system, Sea surface topography, Accuracy of Aerial Triangulation and Questionnaire to Surveyors, Unisurv G 28, 124 pp.

44. Papers by F.L. Clarke, R.S. Mather, D.R. Larden & J.R. Gilliland on the following topics:- Three dimensional network adjustment incorporating ξ, η and N, Geoid determinations with satellite altimetry, Geodynamic information from secular gravity changes and Height and free-air anomaly correlation, Unisurv G 29, 87 pp.

45. Papers by K. Bretreger, J.C. Trinder, C.J.H. Smith, S.M. Nakiboglu, T.S. Lim, T.A. Herring, P.V. Angus-Leppan, P.C. Covell & S.U. Nasca on the following topics:- Ocean tide models from altimetry, rectification of LANDSAT data, Numerical tests of the initial value method, Accuracy of ξ, η from horizontal gravity gradients, Radiation effects on metal bands, Errors in short range EDM and Contour lines in Engineering, Aust.J.Geod.Photo.Surv. No. 30, 127 pp.

46. Contributions to the XVII General Meeting of the IUGG, Canberra, 3-15 December, 1979. Papers by J.S. Allman, P.V. Angus-Leppan, L. Berlin, F.K. Brunner, R. Coleman, J. Kahar, A.H.W. Kearsley, H. van Gysen, E.G. Masters, R.S. Mather and C. Rizos on the following topics:- Adjustment of continental networks, Ratio method and Meteorology, Refraction in levelling, Atmospheric turbulence, Combination solutions of geoid, Outer zone effects on geoid, Tide models, MSL and sea surface slopes from altimetry and Geopotential from spherical harmonic models, Aust.J.Geod.Photo.Surv. No. 31, 177 pp.

47. C. Rizos, "The role of the gravity field in sea surface topography studies", Unisurv S 17, 299 pp.

48. Papers by H. van Gysen, C.S. Fraser, C. Rizos and J.C. Trinder on the following topics:- Gravimetric deflections of the vertical, Calibration of non-metric camera, SST studies and ocean circulation and Aerial film granularity and visual performance, Aust.J.Geod. Photo.Surv. No. 32, 121 pp.

49. Papers by B.J. Burford, E.G. Masters, A. Stolz, C.S. Fraser and F.K. Brunner on the following topics:- Dopper in control networks, Crustal motion from LAGEOS, Variance analysis of adjustments and Incremental strain near Palmdale, Aust.J.Geod.Photo.Surv. No. 33, 85 pp.

50. B.C. Forster, "Some measures of urban residual quality from LANDSAT multi-spectral data", Unisurv S 18, 223 pp.

Proceedings

P.V. Angus-Leppan (Editor), "Proceedings of conference on refraction effects in geodesy & electronic distance measurement", 264 pp. Price: $10.00

R.S. Mather & P.V. Angus-Leppan (Eds.), "Australian Academy of Science/International Association of Geodesy Symposium on Earth's Gravitational Field & Secular Variations in Position", 740 pp. Price: $20.00

Monographs

1. R.S. Mather, "The theory and geodetic use of some common projections", (2nd edition), 125 pp. Price: $5.00

2. R.S. Mather, "The analysis of the earth's gravity field", 172 pp. Price: $5.00

3. G.G. Bennett, "Tables for prediction of daylight stars", 24 pp. Price: $2.50

4. G.G. Bennett, J.G. Freislich & M. Maughan, "Star prediction tables for the fixing of position", 200 pp. Price: $8.00

5. M. Maughan, "Survey computations", 98 pp. Price: $5.00

6. M. Maughan, "Adjustment of Observations by Least Squares", 61 pp. Price: $4.00

7. J.M. Rüeger, "Introduction to Electronic Distance Measurement", (2nd edition), 140 pp. Price: $7.00

OTHER PRICES (Surface Mail Postage Inclusive

1. Aust.J.Geod.Photo.Surv. (formerly Unisurv G)

   2 issues annually of approximately 100 pages each issue.

   Subscription for 1980

To Libraries	$16
To Individuals	$11

2. Special Series (Unisurv S)

   Research reports of 200 to 300 pages, published annually on average.

	Post 1979	Pre 1980
To Libraries	$25 per copy	$17 per copy
To Individuals	$18 per copy	$12 per copy