8-6-2005

Simulated vs. Actual Landsat Reflectance Spectra of Bare Soils

Chandrapalsinh Ghanshyamsinh Chavda

SIMULATED VS. ACTUAL LANDSAT REFLECTANCE SPECTRA OF BARE SOILS

By

Chandrapalsinh G. Chavda

A Thesis
Submitted to the Faculty of
Mississippi State University
in Partial Fulfillment of the Requirements
For the Degree of Master of Science
in Biological Engineering
in the Department of Agricultural and Biological Engineering

Mississippi State, Mississippi

August 2005

SIMULATED VS. ACTUAL LANDSAT REFLECTANCE SPECTRA OF BARE SOILS

By

Chandrapalsinh G. Chavda

Approved:

Dr. J. Alex Thomasson
Professor of Biological and Agricultural Engineering
Texas A & M University

(Director of Thesis)

Dr. Lori Mann Bruce
Associate Professor of Electrical and Computer Engineering

(Minor Professor)

Dr. William D. Batchelor
Head and Professor of Agricultural and Biological Engineering

(Graduate Program Director)

Dr. S. D. Filip To
Associate Professor of Agricultural and Biological Engineering

(Committee Member)

Dr. Nicholas H. Younan
Professor of Electrical and Computer Engineering

(Minor Graduate Program Director)

Dr. Kirk H. Schulz

Dean of the Bagley College of Engineering

Name: Chandrapalsinh G. Chavda

Date of Degree: August 6, 2005

Institution: Mississippi State University

Major Field: Biological Engineering

Major Professor: Dr. J. Alex Thomasson

Title of Study: SIMULATED VS. ACTUAL LANDSAT REFLECTANCE SPECTRA OF BARE SOILS

Page in Study: 248

Candidate for Degree of Master of Science

Simulated Landsat reflectance spectra of soil samples were compared to actual Landsat radiance values of soils in two fields (1 and 3) near Vance, Mississippi. The simulated reflectance spectra were calculated by combining Landsat spectral sensitivity with laboratory-based spectrophotometer reflectance values. The actual radiance data were obtained by extracting pixel values from Landsat images. Simple linear regression (SLR) yielded significant linear relationships for 1997 field-1 and 2001 field-3 data. Multiple linear regression (MLR) and weighted linear regression (WLR), which indirectly accounted for moisture content and spatial resolution, respectively, yielded improvement in R^2 for most of the studied bands. The analyses generally satisfied the normality and constant variance assumptions, and removal of outliers improved the validity of the assumptions and R^2. It was concluded that indirect measures of soil moisture content and spatial uncertainty can substantially improve the relationship between remotely sensed bare-soil spectra and laboratory spectra.

DEDICATION

Dedicated to Lord Swaminararyan and my brother Ruturajsinh (Sadhu Shreyassetudas)

ACKNOWLEDGEMENT

I take this opportunity to acknowledge the contributions of those rare individuals without whose support and motivation this marathon thesis would not have been realized.

First and foremost, I would like express my deepest gratitude to my advisor, Dr. J. Alex Thomasson, whose constant support, motivation, and guidance have played a key role throughout my M.S. program. I also thank my committee members, Dr. Lori Mann Bruce and Dr. S. D. Filip To, for serving on my thesis committee, giving special effort and time to review my thesis, and providing useful suggestions and comments. I especially thank Dr. S. D. Filip To for providing me necessary assistance during my last two semesters of study and much needed flexibility while working for his project.

I am also extremely grateful to Dr. Patrick Gerard for his time and efforts. His valuable suggestions have greatly improved the statistical analysis part of this thesis. I sincerely thank Mr. James R. Wooten for helping me with outlier analysis and other members of the Kimbrough Precision-Agriculture/Remote-Sensing Engineering Laboratory (PARSEL), Ms. Swapna Gogineni, Mr. F. Paul Lee, and Dr. Ruixiu Sui for providing useful suggestions and help.

This acknowledgment would not be complete without expressing my sincere gratitude to my friends. I am extremely thankful to Naveen Kumar for writing a Perl

script for preparing outputs, reviewing some of my work, and providing motivation, useful suggestions, and guidance during the M.S. program. I am also grateful to Vijay Shah for his comments, constant support, and motivation throughout my thesis work. I express my thanks to my friends Anand Kumar and Rushabh Doshi for providing necessary help and making my stay at MSU memorable.

Last but not least, I express my humble thanks to my parents, family members and especially to Mr. Bharatbhai Patel, without their selfless support and motivation, I would not have been able to achieve new heights in my career.

TABLE OF CONTENTS

CHAPTER Page

CHAPTER Page

LIST OF TABLES

TABLE Page

LIST OF FIGURES

FIGURE Page

FIGURE Page

FIGURE Page

FIGURE Page

FIGURE Page

FIGURE Page

FIGURE Page

FIGURE Page

FIGURE Page

FIGURE Page

CHAPTER I

INTRODUCTION

Background

The development of advanced technologies like remote sensing, GPS (Global Positioning System), and GIS (Geographical Information Systems) has enabled agricultural researchers and practitioners to gather information about field conditions and their variability in higher quantity, more quickly, and on a site-specific basis. The benefit of this advancement is that farm inputs can be optimized by managing in-field variability in an environmentally friendly manner in order to maximize the productivity and profitability of the farm. Taken together, these practices are known as precision agriculture.

Managing soil variability is an important aspect of precision agriculture, as soil fertility plays an important role in farm productivity. Traditional methods used for mapping soil properties are laborious, time-consuming, and imprecise, the result of which is that farmers generally ignore in-field variability and use a uniform field-management approach. However, soil physical and chemical properties vary with distance, time, and depth. In order to make sound decisions about soil variability, it is important to have timely and precise information about soil properties. Mapping soil physical and chemical properties of a large farm on a site-specific basis is a daunting

task. However, remote sensing has provided great hope as a data source, and research is ongoing to develop practical methods for implementing this technology in precision farming. According to several studies (Stoner et al. (1980), Ben-Dor and Banin (1994), Thomasson et al. (2001), Abdel-hamid (1993) etc.), soil properties have shown good correlations with reflectances in certain visible and near-infrared bands of the electromagnetic spectrum. Soil reflectance data can be collected by optical sensors mounted on satellites and aerial or ground-based platforms. As soil properties vary with time, repeated data collection efforts are essential. Satellite sensors are best equipped to address this issue, as they are continuously in low earth sun-synchronous orbit, whereas aerial and ground-based platforms require a special service arrangement, which increases the costs of this management method.

Landsat is the principal series of earth resource monitoring satellites in the U.S. The latest in the Landsat series of satellite sensors are the Landsat 5 TM (Thematic Mapper) and the Landsat 7 ETM+ (Enhanced Thematic Mapper plus), launched in 1984 and 1999, respectively, and still in operation. The TM and ETM+ sensors provide an important agricultural perspective, as they have relatively good spatial, spectral and temporal resolution. Their seven bands cover important parts of the electromagnetic spectrum from the visible to near-infrared region. Inexpensiveness, ready accessibility and frequent availability make Landsat data important for agricultural applications, especially in a state like Mississippi, whose Delta region consists of highly variable, alluvially derived soils, and over 80% of the state's row crops are grown in this region (Thomasson et al., 2001). Thus, it is worth evaluating the effectiveness of satellite sensors like TM and ETM+ for soil characterization in

Mississippi. Several studies have been conducted for soil remote sensing, soil reflectance comparison and other related topics. The pertinent literature is reviewed in the following section.

Literature Review

Soil characterization with laboratory-based reflectance data

Thomasson et al. (2001) studied the relationships between soil properties and soil reflectance spectra using a laboratory spectrophotometer. They collected 724 soil samples, each representing 0.4 ha, across two fields located in the northern Delta region of Mississippi. A laboratory spectrophotometer was used to obtain reflectance spectra from the soil samples in the range of 250 to 2500 nm. The objectives were to examine relationships between soil reflectance spectra and soil properties and analyze soil reflectance spectra for sources of variation in the data. Based on multiple linear regression and correlation analysis, it was observed that 50-nm band averages of soil reflectance spectra were significantly (5% level) correlated with soil properties, and R^2 values of 0.50 or better were obtained between multiple-linear models of reflectance spectra and the properties, Ca, Mg, Clay, and pH. However, single 50-nm bands did not exhibit strong correlation with any of the studied soil properties, and results were not consistent between the two fields. According to their analysis regarding coefficient of variation, the regions from 400 to 800 and 950 to 1500 nm exhibited the highest discriminatory power for the given instrument. In other words, these spectral regions contained the highest amount of information with which to estimate soil properties.

Hummel et al. (2001) studied the capability of an NIR soil sensor at predicting soil-moisture and organic-matter content of surface and sub-surface soils. Three 5.56-cm diameter soil cores of 1.5-m depth were obtained at 16 sites across Illinois. The cores were divided into eight segments of 10 to 20 cm in depth. Six soil moisture levels were considered for the study, and total soil organic carbon was measured. A prototype sensor having a wavelength range of 1623 to 2467 nm was used to obtain spectral reflectance of the soil samples under laboratory conditions. The spectral reflectance data were obtained at 6.6-nm spacing with each reflectance having a 45-nm pass band; in total, 129 data points were collected. The data were normalized and transformed to optical density [OD = $\log_{10}$ (1/normalized reflectance)]. Stepwise multiple linear regression (SMLR) results indicated that the standard error of prediction (SEP), which is the standard error of the estimate in the validation data, was 5.31% for soil moisture and 0.62% for soil organic matter, and the regression models included four and nine wavelength bands, respectively. It was noted that commercialization of an NIR soil sensor for soil moisture prediction would likely be easy compared to soil organic matter prediction, because fewer wavelengths bands would be needed (four vs. nine).

Abdel-Hamid (1993) studied spectral properties of several Egyptian soils under field conditions. Laboratory analysis of approximately 280 surface soil samples, collected from 26 sites, representing four great groups of four sub-orders of two orders in soil taxonomy, was carried out in order to determine the chemical and physical properties of the soils. An Exotech spectroradiometer with spectral bands corresponding to Landsat MSS bands 4 (0.5 to 0.6 μm), 5 (0.6 to 0.7 μm), 6 (0.7 to

0.8 μm), and 7 (0.8 to 1.1 μm) was used to obtain soil reflectance data. Simple correlation between six soil parameters (organic matter, CEC, iron oxide, $CaCO_3$, total soluble salts and particle size distribution) and soil reflectance in four individual bands was carried out. The results indicated significant negative correlation between soil reflectance and organic matter content, clay content and CEC for all the studied bands. Total soluble salts, $CaCO_3$ and sand content showed positive correlation in all the bands. Iron content exhibited positive correlation with band 5 and negative correlation with band 6, and since MSS band 6 is centered at 0.75 μm, these results indicate an iron absorption band at 0.75 μm.

Color is a useful soil attribute, and it is widely used by soil scientists for field description, identification, characterization, and classification of soils (Bigham and Ciolkosz, 1993, reported by Mattikalli (1997)). Studies have shown that a good relationship exists between visual soil color and soil reflectance spectra; hence, the importance of color is as an indirect measure of other important features that are not so obvious or accurate. Mattikalli (1997) studied soil color modeling using laboratory spectral measurements with visible and NIR bands similar to those of Landsat sensors. Seventy-six soil samples, collected by Jeyasingh (1986), and representing a wide variety of soil composition, particle size distribution, and color, were used for the study. Reflectance values of the soil samples were obtained in the laboratory with a Landsat ground truth radiometer, representing the same spectral bands as Landsat MSS. Soil color measurements were carried out with standard Munsell charts, and were transformed into the RGB color coordinate system. Simple linear regression analysis between soil color components and soil reflectance values yielded a

correlation coefficient of about 0.5. The optimal rotational transformation technique (developed by Jeyasingh, 1986), which uses an optimization procedure to transform multispectral reflectance data to a new coordinate system, such that the correlation between transformed reflectance and soil color is maximized, was used. Multiple linear regressions of transformed reflectance data and soil color yielded R-values of more than 0.8; also, the hue color component was predicted with acceptable accuracy. The author noted that the study had potential application for mapping soil and geological materials of inaccessible regions with space-borne or airborne sensors, an effort which would be very costly and difficult with traditional methods.

Stoner et al. (1980) measured spectral reflectance of Chalmers silty clay loam and Fincastle silt loam soils in the field under various moisture and crop residue conditions, and in the laboratory under controlled moisture equilibria. An Exotech Model 20C spectroradiometer was used to obtain data in the laboratory and field. Results obtained with the ratio technique, which divides the spectral response of a given soil by that of another identically treated soil and provides the greatest magnitude of difference, indicated that the transition region at visible to near-infrared wavelengths (0.6 to 0.8 μm) was useful for characterizing spectral differences between the studied soils. Also, laboratory measured moist-soil reflectance was proportional to that of field measured bare moist soil spectra of the same soils in the spectral region of 0.52 to 1.75 μm.

Ben-Dor and Banin (1994) applied the near-infrared analysis (NIRA, 1.0 to 2.5 μm) approach to soils with spectra in the visible to near-infrared (VNIR, 0.4- to 1.1-nm) region. They noted that the NIRA is an empirical approach not requiring

assumptions about physical and chemical factors, except that concentrations of constituents are proportional to some linear combination of spectral absorption features. Ninety-one soil samples representing 11 groups of the semi-arid and arid climate zones of Israel were analyzed with various chemical analysis methods, and $CaCO_3$, Fe_2O_3, Al_2O_3, SiO_2, K_2O, LOI (loss on ignition), Fed (free iron oxide) and organic matter were determined. Reflectance spectra of the soils were measured with a LICOR spectrometer at a sampling interval of 1 nm. The original spectrum (having 700 spectral data points for each soil sample) was divided into G equal bands (G = 6, 8, 15, 71, and 350) to obtain several compressed spectral data sets consisting of fewer data points. The spectral analysis for each spectral set was carried out in two stages (calibration and validation) with the so-called VNIRA (visible and near infrared analysis) approach. Since VNIRA is empirical, it allows many combinations of samples and data manipulations to obtain optimal predictive performance. The lowest SEP (standard error of prediction) and highest R^2 criteria from the validation set were selected to indicate the optimal data manipulation for obtaining the best analytical performance. In the calibration stage, the prediction equation was developed with simple linear regression of chemical constituents and spectral responses (reflectance, absorption, and their derivatives R', A', A''). Twenty-five bands from these spectral responses that provided the highest correlations were further examined with multiple regression analysis. With the unknown soil sample subgroup (the one not used in the calibration stage), validation of the prediction equation was carried out, and the equation yielding the highest performance was used in further analysis. The overall analyses indicated that the VNIR region is suitable for obtaining quantitative

information about soil chemistry. Except for Fed, all constituents required from 15 to 350 spectral bands for optimal prediction (lowest SEP and highest R^2). Furthermore, the authors stated that while the VNIRA approach is not as precise as chemical analysis, it could be useful for rapid soil characterization and remote-sensing analysis.

The above studies, conducted primarily in the laboratory, indicate that a significant relationship exists between certain soil properties and soil reflectance spectra. The authors noted that the visible through near-infrared region is promising for soil characterization. However, laboratory studies are time-consuming and not useful for real-time field analysis. Thus, satellite sensors, which are already available and provide rapid and repetitive coverage, could be useful and should be considered for practical application in characterization of earth-surface features including soil.

Remote-sensing data of earth-surface features other than soil

Zhou and Li (2003) compared in-situ and MODIS-derived spectral reflectance of snow and sea ice in order to validate MODIS derived sea ice data products. The study was carried out at 22 locations on the Amundsen and Bellingshausen seas, Antarctica. Among 22 daily measurements, three measurements from three Julian days, 65, 70, and 78 were measured under clear sky condition, and were used for detailed comparison and analysis in this study. Spectral albedo and directional reflectance of sea ice and snow were obtained at each site with a ground-based spectroradiometer, which covers 16 of the 20 MODIS visible and near-infrared (VNIR, from 330 to 1060 nm) bands. Similar data were obtained from a MODIS satellite image of the region, which was georectified to match the location of the ground-based data. The 6S radiative transfer model was used to obtain atmospherically corrected MODIS data of

the study site. The ground-based and MODIS reflectance data were taken virtually simultaneously. The in-situ data were combined in proportion to corresponding surface types (degree of ice concentration, by aerial coverage of ice type over the visible area, expressed in tenths) to simulate the 1-km pixel size of the MODIS image. Discrepancy or difference between in-situ and MODIS reflectance data was calculated for each band. It was found that when the surface was homogenous in one ice type (10/10 ice concentration) the discrepancy range was 0.2 to 11.6% with an average of 4.8%. When the surface had one dominant ice type, the discrepancy ranged from 0.8 to 16.9% with an average of 6.2%. These two situations yielded the best agreement between the two data types. In the case of pixels' having more than one ice-type, a discrepancy range of 0.8 to 25.3% with an average of 13% was observed. It was noted that with an inhomogeneous ground site and very variable topography, the discrepancy could be as large as 30% (Hall et al., 1990).

Hall et al. (1990) compared in-situ and satellite-derived (Landsat TM) reflectances of Forbindels glacier, Greenland. Their objective was to assess the capability of TM sensors for obtaining realistic reflectance measurements. Four sites were selected for the study. The SE-590 portable spectrometer, having 252 discrete sensitivity bands in the 0.4- to 1.1-μm range and covering a 10-cm diameter spot, was used to collect in-situ reflectances at 25 equally spaced points within a 30-m by 30-m area. The collected Landsat TM data of the sites were converted to reflectances, and were corrected for atmospheric effects. The average in-situ reflectance data of each site were compared to Landsat TM average reflectance data. Based on results of the analysis, the authors concluded that atmospherically corrected, nadir-viewing,

Landsat TM reflectance provided good agreement with surface reflectance, and could be used for obtaining physically useful reflectances of ice and snow.

Brivio et al. (2001) carried out validation of satellite data for quality assurance in a lake monitoring application. In-situ reflectance and Landsat 5 TM image data were obtained simultaneously over two lake sites during four different times of year. The Landsat data were corrected with two image-based atmospheric correction methods – the cosine and tau-mean model techniques – to obtain atmospherically corrected satellite reflectance. The comparison between in-situ reflectance and Landsat reflectance datasets (the two with atmospheric correction and one uncorrected, apparent reflectance) indicated that the RMSE (root mean square error) between the two atmospherically corrected Landsat datasets and in-situ reflectance was close to 0.01 for each TM band, which was approximately 80% less than with apparent reflectance. It was noted that the large variability in the results of the apparent reflectance model may be due to the influence of scattering on shorter wavelengths and absorption on longer wavelengths. Hence, it was concluded that using Landsat TM data with either image-based atmospheric correction method is valid for lake water monitoring.

Lee and Cohen (2002) compared hyperspectral AVIRIS (224 channels) data with Landsat ETM+ (6 bands, excluding thermal) data for Leaf Area Index (LAI) estimation. They collected LAI, AVIRIS and ETM+ data over two sites and processed the sensor data to obtain surface reflectance. Four datasets were created from AVIRIS reflectance data: simulated ETM+, six channels matching the center wavelengths of ETM+ bands, a group of channels selected with stepwise regression,

and principal components based on 187 channels. These AVIRIS-based datasets were compared to ETM+ data with multiple regression and canonical correlation. The authors concluded that, among all the datasets, 7 to 14 stepwise-selected AVIRIS channels worked best for estimating LAI. Furthermore, they recommended that channel selection be carried out with an integrated approach of theory, data processing and empirical validation.

Generally, onboard calibration systems use preflight calibration parameters to calibrate aerial and satellite multispectral sensors. Degradation of onboard calibration systems is difficult to determine, so Thome (2001) used a reflectance-based vicarious (meaning not dependent on on-board calibration systems) calibration method, an independent approach for absolute radiometric calibration of the ETM+ sensor of Landsat 7. A spectroradiometer (350 to 2500 nm) was used to obtain surface reflectance over a 480-m x 120-m area at three southwestern U.S. sites having fairly uniform reflectance. A total of 640 reflectance samples having a 0.3-m diameter, with ten samples per satellite image pixel, were collected and averaged to obtain the single spectral reflectance of an entire site. Atmospheric characterization and surface reflectance data were used as inputs to radiative-transfer code that computed hyperspectral at-sensor radiance. The code's output was converted to absolute radiance by multiplying it with a supplied solar irradiance curve corrected for changes in earth-sun distance. The DN values extracted from the georectified ETM+ images of the sites were averaged to obtain one value for the entire site. Sensor gain was calculated with the following equation:

$G = DN - O/L$

Where G = Gain

DN = average DN for a given spectral band

O = offset in DN from the calibration file

L = band averaged predicted at-sensor spectral radiance

Analysis of the four best datasets at each site indicated that gain values from all dates were within 5% of each other, indicating stability of ETM+ and quality of the vicarious calibration. Comparison with similar past work at the same sites indicated that this method had absolute uncertainties from 3 to 5%. Comparisons with pre-launch laboratory-based gains agreed to within less than 7% in all the cases, bands 1 through 5 having lower gains than pre-launch and band 7 having higher gain than pre-launch. For shorter wavelengths these biases were attributed to treatment of atmospheric aerosols, while at longer wavelengths they were attributed to the assumed solar irradiances used to convert relative radiance to absolute radiance.

The studies in this section used in-situ data to compare and or validate orbital sensors for earth-surface applications. In general, the authors noted that remote sensors could be used to retrieve reasonably accurate information about earth surface features. Since remote sensing can provide accurate information on earth-surface features, and since information on soil characteristics can be determined with spectral data, it is important to consider remote sensing for soil characterization.

Satellite-based data for soil characterization

Coleman et al. (1993) studied the possibilities of spectral differentiation of surface soils and soil properties from space platforms. They obtained Landsat-5 TM (June 1985) and ground-based Barnes Modular Multiband Radiometer (MMR) spectral data

(1991) of soils over Madison County, Alabama. Seventy-two sample areas, each comprising plots of 3x3 pixels within the Landsat image, with each plot falling within the boundary of a soil-mapping unit, were located across the study area. Furthermore, nine soil samples were collected from each plot, and soil properties like organic matter, iron oxide content, and soil particle fraction were determined. The objectives were to evaluate the effectiveness of Landsat data in differentiating among soils having similar characteristics, and to identify useful combinations of spectral bands. Discriminant analysis and stepwise regression methods were used for the analysis, which obtained 97.2% overall accuracy in differentiating among eight soil types. According to the method of Average Squared Canonical Correlation (ASCC), all the bands taken together obtained the highest correlation with soil type (correlation significant at the 0.0001 level). Also, a significant relationship was observed between the soil spectral and physical properties considered in the study, but the amount of variation explained was low. The Barnes MMR sensor had better results, and the relatively poor performance of the Landsat TM sensor was attributed to atmospheric interference. Furthermore, 30-m resolution was observed to be too coarse for generating equations to predict soil properties.

Rios and Monger (2002) studied the usefulness of Landsat Thematic Mapper (TM) data in classifying soil types in arid and semi-arid regions. A study area of southern New Mexico was selected. A soil map of this study area developed by Gile et al. (1981) was digitized with ARC/INFO. It was converted to a raster grid of 30-m cell size and georeferenced. Four approaches were used: (a) simple, in which Landsat TM bands 1 through 7 except band 6 were considered; (b) technical, in which a three-

band composite image was used, wherein the first band was the first principal component of the three TM visible bands, the second band was the raw TM near-infrared band, and the third band was the first principle component of the two TM mid-infrared bands; (c) scaled, based on normalized band ratios; and (d) complex, in which a known transform like NDVI (Normalized Difference Vegetation Index), SAVI (Soil Adjusted Vegetation Index), NDTVI (Normalize Difference Tillage Index), SVI (Simple Vegetation Index), and Albedo were used. Supervised classification was carried out, and an error matrix approach was used to test the agreement between classified TM data and soil maps. A method called transformed divergency (Td) was used to find separation between soil spectral classes. Other methods including confusion matrix, Kappa coefficient, and KHAT statistics were obtained to measure the accuracy of the classification in each approach. The results indicated that the technical and simple approaches achieved the highest overall accuracies of 66.8% and 70.6%, respectively, and the simple approach suggested bands 2, 4, and 7 as the best bands for identifying soil mapping units. The authors observed the following. 1) KHAT differences were the smallest with the simple and technical approaches. 2) The technical approach compressed all the data without losing much information, whereas the simple approach suggested three bands, which to some extent compromised the final classification results. 3) Two classes in the technical approach obtained accuracies of 100% and 99%, in contrast with the simple approach, in which only one class obtained accuracy as high as 89%. Thus, the technical approach was suggested as the best method among all the studied methods, and appeared to be a good option for soil mapping in arid and semi-arid regions.

Agbu et al. (1990a) evaluated field soil maps and spectral maps independently in terms of how they helped to (1) minimize variation within, and maximize variation among, mapping units, and (2) characterize mapping units with respect to significant soil properties. They determined soil-property composition and variability within mapping units of field soil maps and computer classified SPOT images. They then statistically compared the two map types to determine the degree to which mapping units were homogenous within a type and different between types in terms of significant soil properties. Two areas of 3108-ha size each, located in east central Illinois and having contrasting variability, were sampled for soil properties at 402-m intervals. Aerial photos from 1978 were used to refine or add detail to the field soil map. Georegistered SPOT data from 1987 were used to create four spectral data sets: (1) three original multispectral bands; (2) the first two principal components of the spectral bands, which contained 98% of the variability; (3) three ratio-transformed bands (red/green, near-infrared/red, and near-infrared/green); and (4) texture-feature-transformed bands, which accounted for the similarity of a pixel in a subset of the image to a block of surrounding pixels. The data sets were classified with unsupervised maximum-likelihood classification, and classification clusters were grouped to form spectral classes. Mapping units in the field soil map were defined in terms of soil properties affecting use and management, and were delineated from the 20 soil series mapped in the study area; whereas for the SPOT image they were directly based on spectral properties. Various statistical approaches were used, and the authors concluded that field soil maps generally performed better than image data. However, according to F-ratio analysis, the SPOT texture-feature map was close to

the field soil map in terms of variation among and within map units, and it was best among all spectral map products in most of the analyses. Hence, it was found that computer classification of SPOT digital data can potentially be useful for soil investigation, especially in supporting soil surveys or when soil surveys are not possible.

In a related study, Agbu et al. (1990b) further studied soil property relationships in east central Illinois with SPOT data. Their objectives were (1) to determine relationships among high-resolution satellite data and soil properties that could be useful in soil classification and map-unit delineation, and (2) to develop models based on soil properties to predict satellite spectral response. They also explored relationships among subsurface soil properties and spectral response. Two areas of contrasting variability in map-unit composition, each 3108-ha in size, were sampled with a manual probe at 402-m intervals to measure soil properties. SPOT data of bare soils from April 1987 were used for the study. In addition to the three multispectral bands (Green, Red and NIR), spectral indices including Brightness Index (NIR + Red + Green; useful in improving the visual output and classification), NDVI (NIR - Red / NIR + Red; useful for distinguishing vegetation from bare soil), and Ratio Index (NIR / Red; which normalizes the effect of varying soil types and is sensitive to biomass and vegetation cover), were obtained for the analysis. One study area was used for determining relationships between soil properties and spectral reflectance, and the other was used to test the developed prediction models. Multiple regression analysis of data from the first area indicated that many surface and some subsurface soil properties exhibited correlation with SPOT spectral data. The models of NIR, Red and

Green with certain soil properties yielded R^2 values of 0.20, 0.17, and 0.23, respectively. It was noted that Brightness Index was useful for surface soil properties, and Ratio Index seemed better for subsurface soil properties. The usefulness of measured soil properties in predicting SPOT spectral responses was tested with the validation area. The actual and predicted spectral responses of this area yielded R (correlation coefficient) values from about 0.25 to 0.32 for all SPOT spectral bands and indices considered for the study. The authors observed that landscape position and percent slope were not important site characteristics for predicting satellite spectral reflectances, and that relationships existed between subsurface soil properties and SPOT spectral responses. Also, models developed on one field area were useful in predicting spectral bands from soil properties (and vice-versa) in another field area in the same region.

Thompson et al. (1983) simulated Landsat Multispectral Scanner response to soils with laboratory reflectance measurements on soils obtained by Stoner et al. (1980a) from 39 of the 48 contiguous U.S. states and representing 246 soil series. Initially, radiance at the top of atmosphere was calculated from the laboratory soil reflectance with the approach reported by Jackson et al. (1983). Calculations were carried out for clear (meteorological range = 100 km) and turbid (meteorological range = 10 km) atmospheres. With tables of calibration constants and the solar constant presented by Richardson et al. (1980) for bands 4, 5, 6 and 7, Landsat digital counts (DC) were calculated for both atmospheric conditions. Soil brightness and greenness were obtained by converting Landsat digital counts into the tasseled cap transformation of Kauth and Thomas (1976). It was observed that the brightness and greenness values

of the laboratory soil spectra obtained with the Malila and Gleason (1977) transformation had similar form and shape to those of Landsat MSS data. Furthermore, Landsat digital counts were found to be within the range of soil values seen in Landsat data obtained during 1976, 1977, and 1978 over different regions in the USA, USSR and Australia. Greenness and brightness vector space was able to separate reflectance curve forms representing genetically homogenous soil properties, and organic matter content could be stratified into 0 to 2% and greater than 2% categories with better than 80% accuracy. The authors noted that this technique would be useful in developing a better understanding of the relationships of spectral and physical-chemical properties of soils, accounting for the affect of soil on crop-spectral relationships, and conducting sensitivity analyses of the effect of soils on spectral models.

Dematte and Nanni (2002) evaluated terrestrial, orbital and airborne sensors for soil characterization, discrimination, elemental content estimation, and analysis of weathering with spectral data of two soils formed from the basic rocks in the Parana region of Brazil. Eight soil samples in a 0- to 20-cm layer were collected from the two soil types. Granulometric and chemical analyses were carried out on the samples. An IRIS (Infrared Intelligent Spectroradiometer), having spectral resolution of 2 nm from 450 to 1000 nm and 4 nm from 1000-3000 nm, was used to obtain spectral data of dried, ground and sieved soil samples. Three spectral readings were obtained for each soil sample, and its mean curve was used for the analysis. With these data, Landsat TM and AVIRIS reflectances were simulated. Landsat 5 image data of the same site were used to obtain DN values of the soils. The DN values of each band were then

converted to reflectance values. Reflectance values based on the image data were calculated with the following equations of Markham and Barker (1987):

$L_{s\,(\lambda)} = n_1 + m_1 DN_{\lambda}$, and

$\rho_{app} = L_{s\lambda}\pi / E_{0,\lambda} = L_{s\lambda} \times d^2 / E_{0,\lambda} \cos(\theta_z)$

Where, $Ls_{(\lambda)}$ is the radiance that arrives to the detector of the sensor

$DN_{(\lambda)}$ = Digital number by band of the TM

N and m are calibration coefficients of the TM;

ρ_{app} is the apparent reflectance of the 'top of the atmosphere'

$E_{0,\lambda}$ is the solar irradiance exo-atmospheric by band of the TM

θ_z is the solar angle; and d is the earth-sun distance (astronomical unit)

The authors noted that DN band variation might not correspond well to reflectance variation. Moreover, according to Ephiphanio and Formaggio (1988), soil spectral behavior based on only DN values might lead to a different interpretation from that based on reflectance. This possibility was observed in the Dematte and Nanni (2002) study in that similarity was found between laboratory reflectance curves and respective Landsat image reflectance curves, whereas DN curves from Landsat were found to vary differently than laboratory reflectance. However, the soils could be discriminated by DN levels (Dematte and Garcia 1995, and Nanni and Rocha 1997), but the data were not representative of a real spectral curve. Furthermore, the behavior of Landsat TM and laboratory reflectance curves was found to be similar, except for intensity variations in bands 1 and 2, which could be due to atmospheric interference. The simulated and original Landsat reflectance curves showed opposite behavior in the band 5 to band 7 spectral regions, which could be due to water vapor for which

correction was not made, and also the calibration coefficient used to convert DN values into "top of atmosphere" reflectance may have had some error. The simulated AVIRIS data had a similar curve shape to that of IRIS data and were better than Landsat data as they could be used to identify gibbisite mineral. It was concluded that iron forms, granulometry, and mineralogy influenced the spectral data of the soils. Band 7 was found to be the best at discriminating the soils, and laboratory spectral data were found to be useful for simulating the orbital data.

The studies in this section indicate that many soil properties, including iron forms, granulometry, and mineralogy, exhibit a significant relationship with satellite sensor data. However, the amount of variation explained with satellite data has typically been low, a situation attributed at least partially to atmospheric effects. Laboratory reflectance curves have been found to be similar to those from satellite sensors, and laboratory data have been found to be useful for simulating the orbital data.

Factors influencing soil reflectance

Jacquemoud et al. (1992) observed that soil spectra acquired with single-direction illumination and viewing angle were useful for classification but provided little information regarding factors (particle size and measurement condition) that influence the interaction of solar energy with soils. They also noted that roughness, which varies in the field, is an important factor influencing directional reflectance of bare soils. The authors modeled soil spectral and bi-directional reflectance in order to provide optical constants for a wide range of soil types. The spectral and directional reflectances of 26 soils were measured in the laboratory with a spectroradiometer

(450 to 2400 nm in 1000 narrow bands) and a radiometer simulating Landsat TM channels (TM2 to TM5, and TM7). The modeled bi-directional and spectral reflectances were obtained with the SOILSPECT radiative transfer model (Hapke, 1981) with input parameters of single scattering albedo, phase function, and variable characteristics of soil roughness. R values of 0.996 and 0.997 were obtained between measured and simulated bi-directional and spectral reflectances, respectively. It was concluded that phase function and roughness are wavelength-independent and are mainly functions of refractive indices of soil components, whereas single scattering albedo is dependent on only wavelength and soil moisture content.

Epema (1992) studied atmospheric influences on reflectance of bare soil surfaces in southern Tunisia. The selected location included parts of the playa and footslopes, the playa being a clay plain that is temporarily flooded to form a lake or swamp after exceptional rainfall, and the footslopes being a lower part of a hillslope merging with the alluvial plain. Field measurements were made during November 1987 and April 1988 with a Barnes Modular Multiband Radiometer (MMR), which has bands compatible with those of Landsat TM. Landsat data of the same region were obtained on 18 December 1987 and 8 April 1988. The reflected signal was measured over the plots every hour with the MMR. Variation in the study area was used as the criterion for selection of plots. The Verhoef (1985, 1990) atmospheric model, which assumes surfaces as Lambertian and requires the ratio of diffuse to total irradiance for two or more bands as input, was used. Based on this model, planetary (top of atmosphere) reflectances for TM bands were predicted from ground reflectance obtained by MMR for three different atmospheric conditions: 1) the most clear day of April, 2) the most

hazy day without clouds and still suitable for TM imagery (23 April), and 3) the day of Landsat overpass (8 April). The comparison between Landsat TM and predicted Landsat TM data indicated that atmospheric correction is useful at shorter wavelengths, especially for areas of low ground reflectance. Furthermore, for band 1 the difference between predicted and actual Landsat TM for the atmosphere types considered was always less than 10%. The author also compared the predicted ground reflectance of TM bands, based on the idea of planetary reflectance, to that of in-situ measured ground reflectance with the MMR for a specific atmosphere (8 April). Less than 10% difference was observed in the comparison, and the fit of the data was found to be good. It was noted that stable atmospheric conditions are a prerequisite for good ground reflectance measurements.

Price (1990) stated that measurements with higher spectral resolution generate computational problems, and inversion of an associated high-dimensional matrix during analysis is affected by roundoff errors, noise, and spectral redundancy in the data, causing problems for standard approaches like principal component analysis (PCA). Price studied more than 500 U.S. soils by examining their reflectance spectra from 550 to 2320 nm with 10-nm resolution. A procedure developed for identifying independent spectral variability in the thermal band (Price, 1975) was used to assess the need for higher resolution in characterizing soil spectra. It was concluded that four broadband spectral measurements at 930 - 1130 nm, 2030 - 2310 nm, 630 - 740 nm, and 1610 - 1800 nm, with the known basis vectors (fitting function determined by a best fit across the data set, and obtained with complex analysis of the Gram-Schmidt procedure and PCA), are sufficient for spectral discrimination of the studied soils.

While overall the literature review indicates that remotely sensed data have potential for soil characterization, it also tends to indicate that they cannot be used directly for soil property identification, because the correlations between satellite spectral data and soil properties have typically been low. There are several factors, aside from the fact that spectral data lack full explanatory power for soil properties, that cause this to be the case: 1) soil roughness, 2) directional reflectance characteristics, 3) atmospheric effects, 4) vegetation and topography, 5) geometric and radiometric distortion., 6) sensor characteristics like spectral and spatial resolution, and 7) soil moisture The first four of these were considered in the studies in this section, and appear to be factors that can be dealt with reasonably well. All seven will be discussed briefly below.

With respect to roughness, the contrast of absorption features decreases and soil reflectance increases as particle size decreases (Aztberger, 2002; Bowers and Hanks, 1965; Hunt, 1980; Stoner and Baumgardner, 1980). Furthermore, roughness, which varies in the field, is an important factor influencing directional reflectance of bare soils (Jacquemoud et al., 1992).Soil surfaces are not lambertian surfaces, and apart from wavelength, soil reflectance spectra depend on illuminating and viewing directions (Aztberger, 2002). Particulate materials having low absorption such as desert and beach sands have strong forward scattering, and highly absorbing materials such as clayey and loamy soils have less forward scattering but strong backward scattering, as observed by Coulson (1966) and Coulson et al. (1965).

Atmospheric constituents cause scattering (Mie and Rayleigh), which influences the reflected energy. Huete (2002) noted that the atmosphere has a strong effect on

the spectral signature of dark surfaces, and little effect on brighter surfaces, such as sand. As mentioned previously, Epema (1992) found that atmospheric corrections could successfully account for these effects.

It is well understood that vegetation strongly influences soil reflectance spectra. Under the right experimental conditions (i.e., bare soil), vegetation can be neglected. Topographic parameters like elevation influence soil moisture content, geometric distortions, and directional characteristics, which in turn affect the soil radiance.

Satellite images should be geo-rectified in order to account for distortion caused by the earth's rotation and curvature, satellite or aircraft motion, altitude, viewing perspective and surface elevation effects (Morgan and Ess, 1997). Radiometric distortions are caused by atmospheric variations; variations in sensor viewing angles, which occur during scanning; variation in illumination of the areas viewed; poor sensor performance; and variation in image processing procedures. Hence, radiometric correction should be carried out to account for the above distortions (Morgan and Ess, 1997). With high-quality satellite data, these factors can generally be neglected.

Sensor characteristics such as spectral and spatial resolution and instrument noise also play an important role in soil surface reflectance spectra. Instrument noise can be assumed to be minimized in high-quality satellite data, and spectral resolution has been dealt with in studies such as that of Price (1990). On the other hand, spatial resolution has not been dealt with adequately with respect to soil characterization.

Soil moisture has a major influence on soil radiance (Huete, 2002). Huete, 2002 reported that Idso et al. (1975) observed that the decrease in soil reflectance is

proportional to the thickness of the water film around the soil particles, which increases the absorption and decreases the reflected energy.

Many of the factors affecting the usefulness of remotely sensed data in soil characterization either can be neglected or have been dealt with adequately in research studies. Two key factors that remain are spatial resolution of the sensor and soil moisture content. While the ultimate goal of the current study is the development of a model that would allow Landsat data to be used for soil characterization, it must be kept in mind that even laboratory spectral data cannot be universally used in a detailed way for this purpose. What is hoped for is a method to use Landsat data to accurately predict soil spectra, from which general and practical categorizations of soils may be made.

Objectives

The specific objectives of this study are as follows:

1. To calculate the correlation between predictor data based on laboratory reflectance measurements and Landsat data.
2. To indirectly consider the possible influence of soil moisture content on the correlation by including elevation and soil texture data in the analyses.
3. To indirectly consider the influence of spatial resolution by accounting for soil sample location relative to Landsat image pixels in the analyses.

CHAPTER II

MATERIALS AND METHODS

Sample Collection and Processing

Soil Samples

Location

The study site is located in the northern Mississippi Delta region near Vance, Mississippi and comprises two fields known as field 1 and field 3 (Figure 1). The approximate areas of fields 1 and 3 are 111 and 162 ha, respectively. They are located roughly 1.6 km away from each other in two different Mississippi counties, Field 1 in southern Quitman County and field 3 in northern Tallahatchie County.

Soil Types

The soils in the Mississippi Delta region were formed as part of the alluvial floodplain of the Mississippi river. Alluvial soils are formed from sediments deposited by rivers and its tributaries over the years. Soils developed from alluvial parent material generally are highly variable, since sediments are derived from various sources (Buscaglia, 2000) and are laid down unevenly. The soil variability in the two fields has been observed in association with the USDA-NRCS soil survey, which has listed several soil types in each field (Table 1).

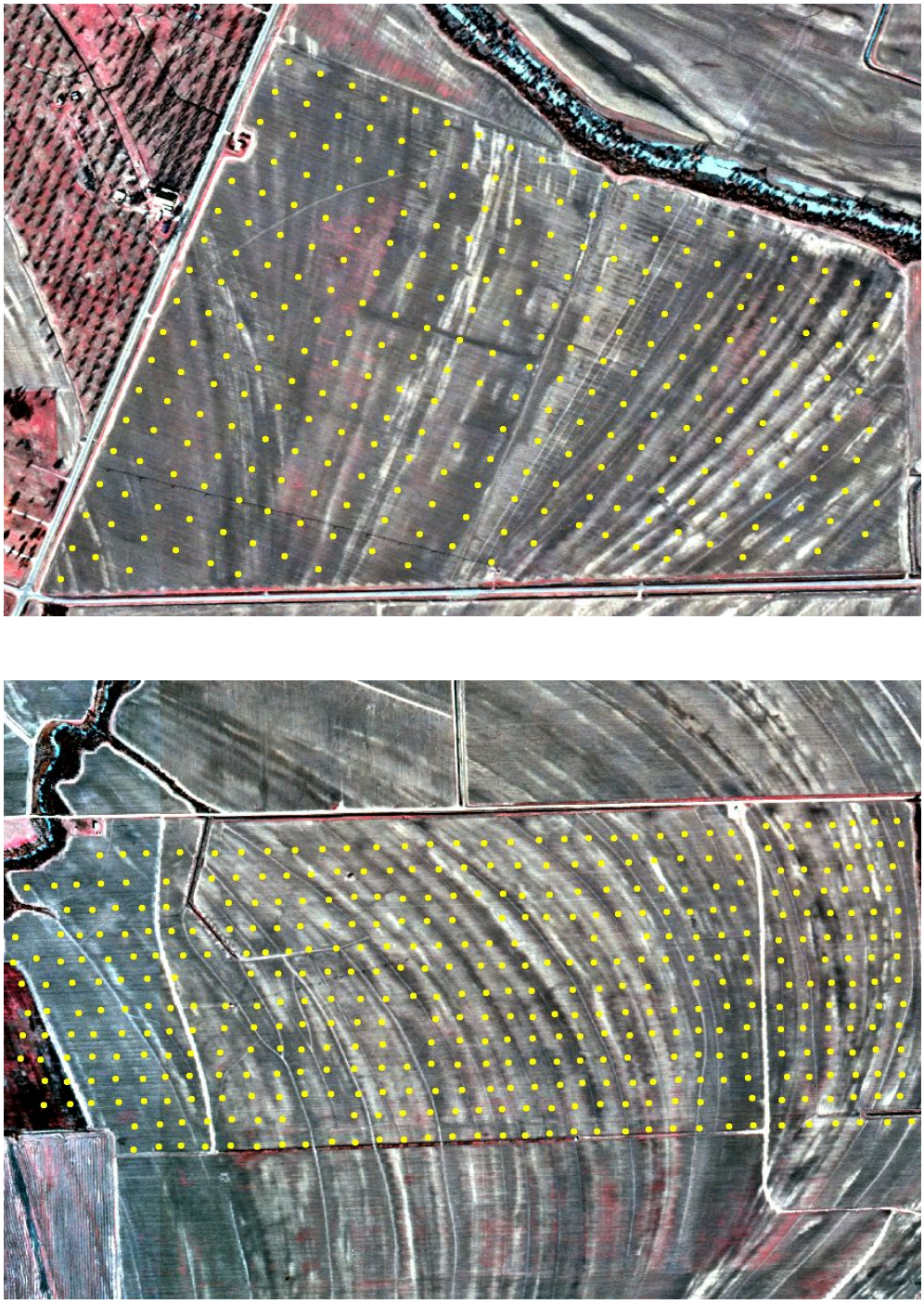


Figure 1: Aerial images of field 1 (top) and field 3 (bottom) with overlaid location of sample points collected manually and recorded with GPS (Source: Al-Rajehy, 2002)

Table 1: NRCS/USDA soil survey classification of the study sites (Source: Al-Rajehy, 2002)

Soil Type	Family	Subgroup
Souva silt loam, nearly level phase	Souva: Not given	Not given
Dowling clay and silty clay	Dowling: Not given	Not given
Dundee fine sandy loam, nearly level phase	Dundee: Fine – silty, mixed, thermic	Aeric Ochraqualfs
Dubbs fine sandy loam, nearly level phase	Dubbs: Fine – silty, mixed, thermic	Typic Hapludalfs
Alligator clay, depressional	Alligator: Fine – montmorillonitic, acid, thermic	Vertic Hapludalfs
Dundee silt loam, 0 to 2% slopes	Dundee: Fine – silty, mixed, thermic	Aeric Ochraqualfs
Dundee and Tensas silt loams, 0 to 3% slopes	Dundee: Fine – silty, mixed, Thermic Tensas: Fine – montmorillonitic, thermic	Aeric Ochraqualfs Aeric Ochraqualfs
Dubbs very fine sandy loam, 0to 2% slopes	Dubbs: Fine – silty, mixed, thermic	Typic Hapludalfs
Forestdale silty clay loam, 0 to 3% slopes	Forestdale: Fine – mixed, thermic	Typic Ochraqualfs

Sample Collection

Core Samples

In order to obtain laboratory reflectance data from bare soils, soil samples were collected from the two fields. Thomasson et al. (2003) collected nine hundred and sixty-nine soil samples in the years 1999 and 2000 to represent the 273 ha of the study site. In field 1, 274 soil samples were collected in July 1999, and 264 in May 2000 (total 538 samples). Also, 431 samples were collected from field 3 in 1999. Soil samples were collected with a 2.5-cm diameter manual probe where the soil was bare. At each sample location, approximately 600 g of soil was obtained with the probe from the soil surface to a depth of 15 to 20 cm.

Procedure

The soil samples were collected on roughly a 0.4-ha grid. A DGPS (differential global positioning system) receiver was used to mark the location of each sample. A single soil core from the probe was considered as one sub-sample. Five sub-samples were obtained at arbitrary points within a 3-m radius of each sample location. The five sub-samples were mixed to create one representative sample for each location. Each composite sample was stored in a plastic bag for further preparation and analysis.

Sample Processing

The collected soil samples were air-dried in a non-environmentally controlled laboratory, having ambient conditions, with approximate average temperature of 25 to 30 °C. The air-dried samples were ground and sieved with a 2-mm mesh sieve before reflectance measurements were made.

Data Collection

Soils Data

Reflectance

Soil reflectance spectra were measured in the laboratory with a Cary 500 UV/VIS/NIR spectrophotometer, equipped with a diffuse-reflectance accessory that incorporates an integrating sphere (Figure 2). The spectrophotometer is sensitive from 250 to 2500 nm. The geometry of the integrating sphere enables it to collect almost all the reflected radiation in an integrated manner by removing any directional

preferences, and to present an integrated signal to the detector. The schematic diagram of Figure 3 illustrates the process of spectrophotometer reflectance.

Figure 2: Cary 500 UV/Vis/NIR laboratory spectrophotometer (Source: Al-Rajehy, 2002)

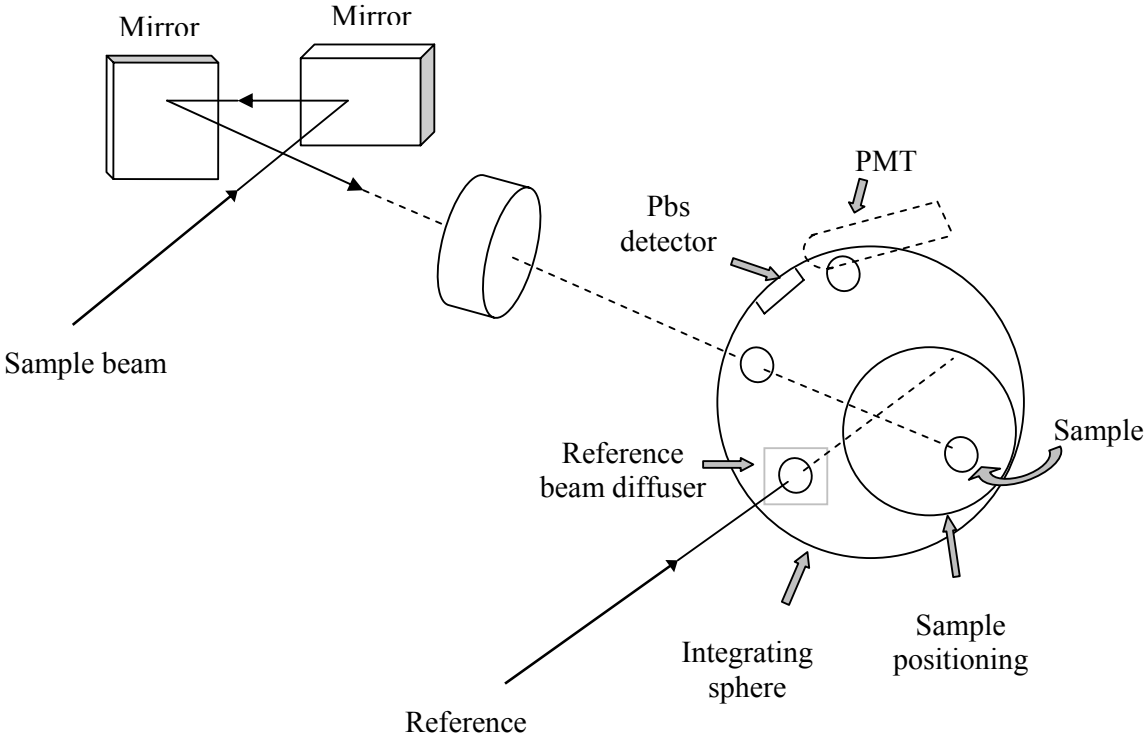


Figure 3: Schematic diagram of soil reflectance process in spectrophotometer

The sample holder was designed and fabricated such that soil samples could be presented to the spectrophotometer. The sample holder is 25 mm in height by 25 mm in diameter, with an interior diameter of 20 mm. It has a removable cap on one end and a polished sapphire glass window at the other end, having thickness of 1 mm and diameter of 22 mm. This type of window was selected because its transmission is roughly constant (around 85% for a 1-mm thick window) from 200 to 6000 nm. Approximately 6.6 g of soil from each collected soil sample was placed in the sample holder until it was full, and to retain the soil the removable cap was placed on the open end of the holder.

The sample holder along with a soil sample was then mounted on the sample port such that the sample holder window was pressed against the port. A computer program provided by the spectrophotometer manufacturer was used to operate the spectrophotometer in order to collect reflectance spectra. The integrating sphere collected reflected energy off the soil sample surface. The standardized reflectance value of each soil sample was calculated as a ratio of the flux reflected by the soil sample to that reflected by a reference disk under identical geometrical and spectral-illumination conditions. The spectrophotometer collected 970 reflectance values for each soil sample. After each reflectance measurement, the cap was removed, the sample holder was emptied, and the optical window was cleaned before adding the next soil sample.

The spectrophotometer was initially calibrated before collecting reflectance spectra of soil samples in a given data collection session (i.e., once per day). In order to obtain baseline reflectance data, a manufacturer-provided, secondary-white-

standard, PTFE (polytetrafluoroethylene) disk calibrated relative to a perfectly diffuse reflector was used. After instrument warm-up, baseline data were recorded with the reference PTFE disk covering the sample port of the diffuse reflectance accessory. A sapphire glass, similar to the one placed on the sample holder as an optical window, was placed on top of the PTFE disk during baseline data collection in order to account for light attenuation caused by the optical window in the sample holder. Soil reflectance data were collected at wavelengths from 250 nm to 2500 nm. Spectral resolutions of 1 and 4 nm were selected for the 250- to 792-nm and 792- to 2500-nm ranges, respectively. Soil reflectance data were obtained over 0.1 s of sampling duration, which was set as a parameter in the computer program before running the data-collection operation.

Texture

Moisture content is an important soil property, and it influences the energy reflected by soils. However, texture, which can be classified according to the proportions of sand, silt and clay, influences the water holding capacity, not to mention water movement and infiltration in the soils (Donahue et al., 1983). Hence, moisture content in the field is affected by texture. Sand and silt particles are considered relatively inert (Reed et al., 2000), but clay particles, because of their electrostatic charges, along with O.M. (organic matter), are considered the seat of most chemical, physical and biological processes (Charman and Murphy, 2000). Furthermore, the clay fraction determines the reactivity and hence the amount of control the clay exerts on water availability (Kutilek, 1973, as reported by Williams et al., 1983, and Foth, 1984). So, in order to account for the influence of moisture

content, soil texture data collected by Thomasson et al. (2003) were used. The hydrometer method, which determines the approximate proportions of clay (particle diameters < 2.0 μm), silt (2.0 to 50 μm), and sand (50 to 2000 μm) particles in a soil, was used to determine texture. From the collected texture data, only clay (%) data were used for this study, since they are most closely related with soil moisture content.

Elevation

Elevation, as well as the proportion of clay in a soil, influences moisture content at a given location. Therefore, elevation data of both fields were collected with a laser altimeter. All these data had been collected by Thomasson et al. (2003) and stored at PARSEL (the Precision-Agriculture/Remote-Sensing Engineering Laboratory, in Mississippi State University's Department of Agricultural and Biological Engineering). Since inherent soil properties and elevation are relatively static, data in this previously developed database were deemed acceptable and used for this study.

Satellite Data Acquisition

Landsat Description

The Landsat satellite program was started in 1972 with a mission to provide continuous-coverage, high-resolution, multispectral data of the earth's surface. The latest in the Landsat series of satellite sensors are the Landsat 5 TM (Thematic Mapper) and the Landsat 7 ETM+ (Enhanced Thematic Mapper plus), launched in 1984 and 1999, respectively, and still in operation. Both satellites operate in sun-synchronous orbit with 16-day temporal, and 30-m spatial, resolution. Table 2

provides the bandwidth (spectral resolution) of all the bands, except band 6, for both satellites. Band 6 was not considered in this study because it is a thermal band with low spatial resolution, and thus did not fit together with the other spectral data available in the study.

Table 2: Spectral bandwidths of ETM+ and TM sensors

Sensor	Band1 (μm)	Band2 (μm)	Band3 (μm)	Band4 (μm)	Band5 (μm)	Band7 (μm)
Landsat 7 (ETM+)	0.45 - 0.52	0.53 - 0.61	0.63 - 0.69	0.78 - 0.90	1.55 - 1.75	2.09 - 2.35
Landsat 5 (TM)	0.45 - 0.52	0.52 - 0.60	0.63 - 0.69	0.76 - 0.90	1.55 - 1.75	2.08 - 2.35

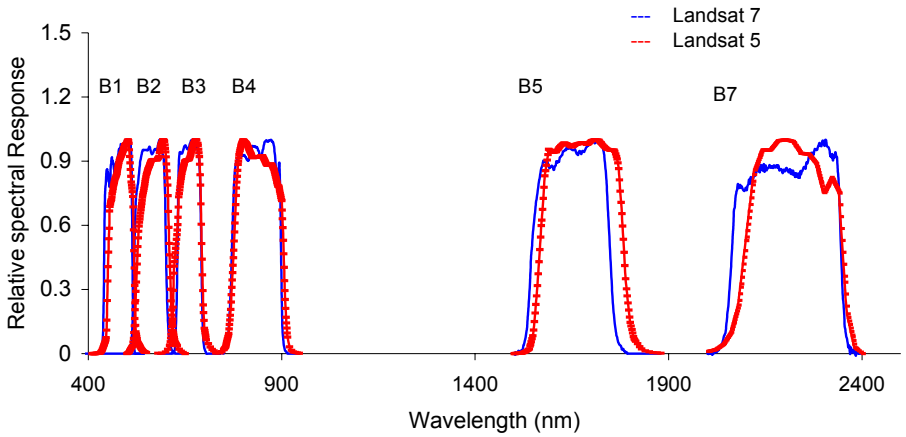


Figure 4: Landsat 7 and Landsat 5 sensor sensitivity with respect to wavelength in each band

Spectral sensitivity data of the TM and ETM+ sensors were available in the Landsat 7 Users Handbook webpage, at the NASA website (http://ltpwww.gsfc.nasa.gov/IAS/handbook/handbook/htmls/chapter8/chapter8.html, downloaded March 24, 2005). These data were used in order to calculate simulated

Landsat reflectance of bare soils. The sensitivity with respect to each band of TM and ETM+ sensors is presented in Figure 4

Image Dates

Geo-rectified Landsat Images of fields 1 and 3 that were acquired before the growing season, during bare-soil condition, were used in order to obtain actual Landsat radiance of bare soils. Available were Landsat 7 images of both fields for year 2003, and Landsat 5 images of fields 1 and 3 for years 1997 and 1999, respectively. The images used had been acquired by Thomasson et al. (2003) and stored in the PARSEL database.

Data Analysis

Data Preparation

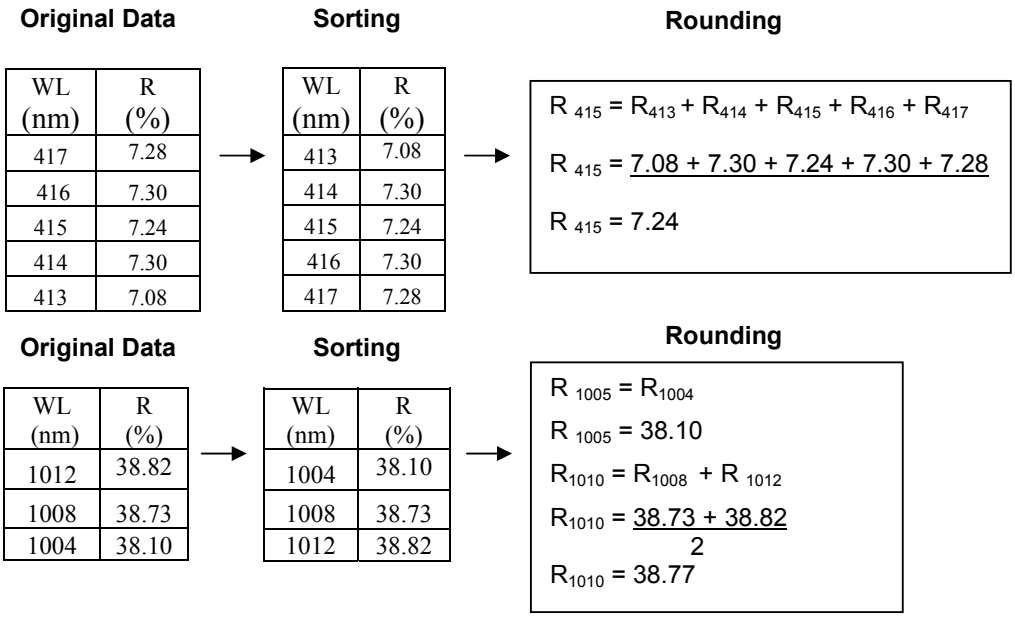
Satellite Data

A computer program written in C and available in the PARSEL database was used to extract DN (digital number) values of each pixel of interest in the Landsat images. The Landsat images and UTM (universal transverse mercator) coordinates of both fields were supplied as input. The program accepts multiple images as input and calculates the center position of each pixel in each image as an input parameter. The program uses the nearest-neighbor method to locate the UTM position near to the center of each pixel, and it assigns that pixel value (DN) to the coordinates for that position. With this algorithm, the program generates a tabular output of DN values associated with pixel center positions in UTM coordinates. The expected reflectance

and actual radiance data of both fields were aligned by matching UTM coordinates pixel by pixel. Similarly, soil texture and elevation data were added to the data table, with a particular UTM location for each data point.

Laboratory Data

The spectrophotometer-based soil reflectance data were collected at wavelengths from 250 to 2400 nm. However, satellite sensitivity data were available from roughly 400 nm, and this study concerns visible and NIR soil reflectance, so spectrophotometer data from 250 to 399 nm were removed from consideration. The visible spectrophotometer data were recorded at 1-nm increments, whereas NIR data were recorded at 4-nm increments, so in order to be consistent and simplify calculations, all spectrophotometer data were rounded at 5-nm increments. In order to achieve rounding, SAS 8.0 statistical software was used. A SAS program was written to receive the original spectrophotometer data as input, sort the data according to wavelength, and then round all wavelength values to the nearest 5 nm. The mean of the five visible data values was calculated for each 5-nm increment, since a multiple of five was used as a center value for averaging. In the case of NIR data, all the wavelength values were likewise rounded to the nearest 5 nm. However, since the NIR data were recorded at a 4-nm increment, either there were two values to be averaged for a 5-nm-increment wavelength, or there was only one value to be used. An example of both visible and NIR rounding processes is provided in figure 5.



Where, WL = Wavelength, and R = Reflectance

Figure 5: Example of rounding process of soil reflectance data at visible (above) and NIR (below) wavelengths

Similarly, Landsat 7 and Landsat 5 sensitivity data were rounded at 5-nm increments with SAS software, because the available data do not have constant spectral resolution, and it is important that there be similar wavelengths between spectrophotometer and satellite data so as to allow correlations to be determined at the same wavelengths. In an effort to simulate the response that would be expected by the satellite sensors, the rounded soil reflectance data were multiplied by the rounded Landsat sensitivity data for each band. The resultant products obtained at every 5-nm wavelength were integrated for each Landsat band in order to obtain expected reflectance, as measured by the satellite sensor, of soil samples collected at a specified ground location. This reflectance was the expected soil reflectance representative of a

30-m pixel in Landsat sensor (TM, ETM+) images. Once again, the position of each soil sample was available in UTM coordinates from the prior DGPS measurements, and the centroid position of each pixel in each image was calculated by the C-language data-handling program. The distance (UTM, easting and northing) between sample location and pixel-centroid position was also generated by the program. An example of the final data set, which was carried out in a two-step process, is provided in Tables 3 and 4.

Table 3: Example of the final data table, Step one.

UTM Coordinates		**Location ID**	**Wavelength, nm**	**Sensor Sensitivity (A)**	**Spectrophotometer Reflectance, % (B)**	**Expected Reflectance, % (A x B)**
Easting	Northing			Band1 ---- Band7	Band1 ----- Band7	Band1 ----- Band7
742632	3772507	2a11	435 | | | | | 2400	0.018 ---- - 0.006	13.8 ----- 64.5	0.25 ------ 0.44
Expected Reflectance, Integration ΣBand1 ---- ΣBand7						223.7 ------ 2018

Table 4: Example of the final data table, Step two

UTM Coordinates		**Location ID**	**Expected Reflectance (Integrated)**	**Actual Radiance, DN**	**Clay %**	**Elevation**	**Inverse Distance (pixel centroid and soil sample location)**
Easting	Northing		Band1	Band1			
742632	3772507	2a11	223.7	99	0.36	-0.11	0.05

Statistical Analysis

In order to determine relationships between actual radiance and expected reflectance of bare soils, statistical analysis, particularly regression and correlation calculations, were carried out. Three types of linear regression analyses, simple, multiple, and weighted, were conducted with SAS 9.1 statistical software.

Simple Linear Regression

In the first analysis, an SLR (simple linear regression) model of actual radiance and expected reflectance was developed. The model uses a least-squares estimation procedure to determine a linear relationship between two random variables. The F statistic and its associated p value were obtained in order to test the model's significance, and the $\alpha = 0.05$ significance level was used for the study. The coefficient of determination (R^2) was calculated as a measure of the strength of the linear relationship between actual radiance and expected reflectance. The SLR model used for the analysis was as shown in Equation 1:

$$(\text{Actual radiance})_i = \beta_0 + \beta_1 (\text{Expected reflectance})_i + \varepsilon_i$$

Letting (Actual radiance) = Y and (Expected reflectance) = X, the equation is

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_I \quad (1)$$

Where, actual radiance is the dependent variable (Y), expected reflectance (X) is the independent variable, β_0 and β_1 are the regression parameters representing intercept and slope of the linear relationship, ε is the random error, and the subscript i indicates particular observational unit. The model tests the hypothesis that there is no significant linear relationship (H0: β_1 (slope) = 0) between dependent and independent

variable, and if this hypothesis is rejected by way of the F statistic and its p value (p value less than $\alpha = 0.05$), the alternative hypothesis that there exists a significant linear relationship (Ha: $\beta_1 \neq 0$) is accepted.

Multiple Linear Regression

Figure 6 indicates that soil moisture can strongly influence the reflectance of soils, a phenomenon that has been noted in several research studies. Although the images were taken during bare-soil conditions or nearly so, it is expected that the fields would have had considerable moisture variation during that time. In order to account for variations in soil moisture content, measurable field parameters influencing moisture content that vary with location but not with time were considered. As mentioned previously, soil texture (taken as clay content) and elevation had been measured, and for the purpose of accounting for moisture content they were added to the former SLR model to yield an MLR (multiple linear regression) model, possibly improving the relationships between satellite-measured radiance and laboratory-measured reflectance. Thus, the MLR model will have three independent variables: expected reflectance, % clay, and elevation (equation 2). Similar to the SLR model, the MLR model also uses the least-squares estimation procedure to determine the linear relationship between model variables.

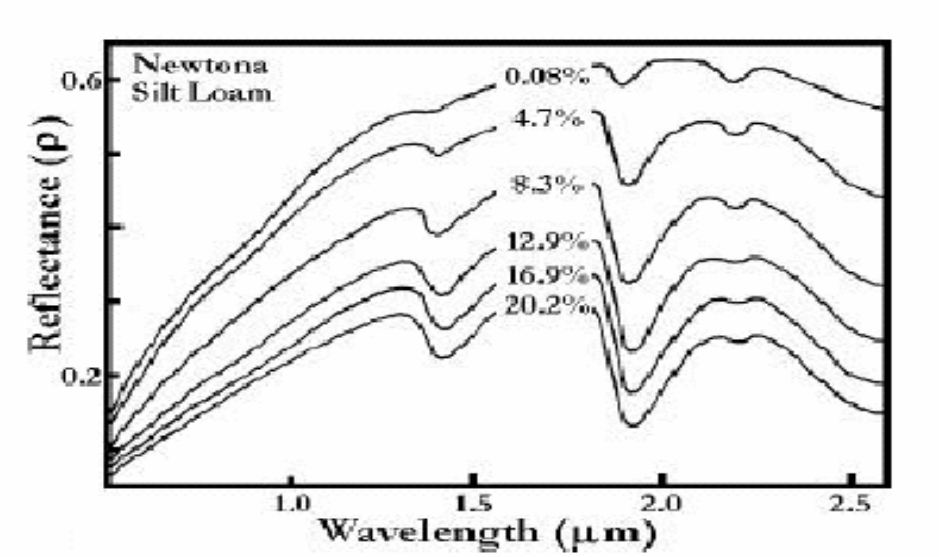


Figure 6: Soil reflectance for different moisture levels (Irons 1991; modified, reported by Ataberger 2002, used with permission)

In this case,

$(\text{Actual radiance})_i = \beta_0 + \beta_1 (\text{Expected reflectance})_i + \beta_2 (\text{clay})_i + \beta_3 (\text{elevation})_i + \varepsilon_i$

Similar to SLR, the Actual radiance = Y, Expected reflectance = X_1, clay = X_2, and elevation = X_3, so the equation is

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_I \quad (2)$$

where β_0, β_1, β_2, and β_3 are regression parameters, and ε is the random error.

The MLR model allows one to test the linear relationship for the null hypothesis H0: $\beta_1 = \beta_2 = \beta_3 = 0$ at $\alpha = 0.05$. If the null hypothesis is rejected, it is concluded that the dependent variable has a significant relationship with at least one of the independent variables. Failure to reject the null hypothesis indicates that no significant relationship exists between the dependent and independent variables. A t-test for each regression parameter (β_1, β_2 and β_3) was also conducted to test the significance of each individual

variable relative to the dependent variable. The R^2 value tends to increase by adding any variable in the model, so adjusted R^2 results were obtained, which unlike R^2 do not always increase as variables are added to the model. Adjusted R^2 is a rescaling of R^2 by degrees of freedom; thus, it is ratio of mean squares rather than sums of squares as with R^2 (Rawlings et al., 1998). Hence, if the adjusted R^2 values vary similarly to the R^2 values, it can be concluded that an increase in R^2 is caused by the variables that were added to the model.

Weighted Linear Regression

In another analysis, the distance from the pixel centroid to the location of the collected soil sample within that pixel was considered. The inverse of this measured distance was applied as a weight to the both the MLR and SLR equations. The assumption here is that the centroid of the 30-m square pixel is the location most representative of the pixel; however, the collected soil sample data point can be anywhere in the pixel. Therefore, applying inverse distance weights to sample points provides a means of considering the quality with which the sample data point represents the pixel. It is expected that this could improve the prediction of the SLR and the MLR models. An F-test and a t-test were conducted to test the significance of the model and its individual variables.

Weights (W = 1 / [distance between pixel centroid and sample location]) were applied to both SLR and MLR equations, so the new models were named WSLR and WMLR, respectively. When the weight option is used in the PROC Reg procedure of SAS, the weighted residual sum of squares,

$\sum w_i (Y_i - \hat{Y}_i)^2$ is minimized. Where, w_i = weight of the i^{th} observation. Y_i = value of

the i^{th} dependent variable, and $\hat{Y}_i$ = predicted value of the ith dependent variable. X_i = independent variable of i^{th} observation. Hence, the new WSLR and WMLR equations are:

$$\text{WSLR: } w_i Y_i = \beta_0 w_i + \beta_1 w_i X_i + w_i \varepsilon_I \quad (3)$$

$$\text{WMLR: } w_i Y_i = \beta_0 w_i + \beta_1 w_i X_{i1} + \beta_2 w_i X_{i2} + \beta_3 w_i X_{i3} + w_i \varepsilon_I \quad (4)$$

Regression Assumptions

The following standard assumptions about regression analysis were adopted from Rawlings et al., 1998.

1) The data pertaining to the independent variable are measured without errors.
2) The random errors are normally distributed.
3) The random errors have zero mean.
4) The random errors have constant variance.
5) The random errors are pairwise independent.

All the regression models used (SLR, MLR, WSLR, and WMLR) are based on these assumptions, and in each case, the data were evaluated to determine whether they satisfied the assumptions or not. Assumption one was taken for granted as being true. Assumption three, which assumes that the random errors have zero mean, is satisfied with the least-squares regression method itself. Hence, the three other assumptions, i.e. normality, constant variance, and pairwise independence, were tested for the SLR, MLR and WLR models.

Normality

Assumption two (normally distributed random errors) was evaluated by plotting normal probability plots. The PROC UNIVARIATE procedure in SAS was used to obtain these plots, wherein ordered residuals are plotted against the normal ordered statistics for the appropriate sample size (Rawlings et al., 1998). The residual, analogous to error, is the deviation of actual data from data predicted by the regression model. The expected result from a normal probability plot is a straight line passing through zero. Great deviation from the line indicates non-normality. Rawlings et al. (1998) noted that the F-test is generally regarded as reasonably robust against non-normality, but the confidence interval estimates of the parameters are not correct in the non-normal case. Standardized residuals – the ratio of residuals to the standard deviation – were used for evaluating the normality and constant variance assumptions.

Constant variance

Standardized residuals were plotted against the predicted value of the dependent variable in order to evaluate assumption four (constant variance of random errors). If there is constant variance, there will be a random scattering of data points above and below the line, e (residual) = 0, with nearly all the data points being within the range of $e = \pm 2s$ (Rawlings et al., 1998). The SAS procedure, PROC GPLOT was used to obtain the plots. Rawlings et al. (1998) noted that if there is heterogeneous variance, some data points will have more influence over the F-test, t-test, and confidence intervals. Hence, the results of these tests could be incorrect.

Pairwise independence

Assumption five holds that the random errors are uncorrelated, which means any two observations of the dependent variable do not have any correlation. If this assumption is not valid, similar to the constant variance assumption, the F-test, t-test, and confidence intervals are not valid (Rawlings et al., 1998). The data points in this study were collected at a 30-meter distance in the field, and it is possible that they may have spatial correlation. The PROC VARIOGRAM procedure of SAS was used to detect spatial correlation in the field. SAS by default assumes isotropy, which means that spatial dependence, if any, is the same in all directions. With PROC VARIOGRAM, two important parameters, the lag distance and maximum lag, are needed to carry out semi-variogram analysis. The lag distance is the distance between two data points in the field. As the distance between two data points in the studied fields was not constant, an approximate distance of 50 meter was considered as the lag distance. Furthermore, SAS by default assumes lag tolerance of ½ (lag distance), which means that all the data points of 50 ± ½ (50) will be in one class, and the next class will have 2 x lag distance ± ½ (50) and so on. Maximum lag is the maximum number of lag classes, and each class has a total number of data pairs at the specific lag distance (e.g., 100 pairs at 50 m, 20 pairs at 100 m, and so on…). Here, a maximum lag value of 40 was determined by trial and error.

As the assumption of pairwise independence relates to the prediction errors, the residual of the dependent variable was used as the variable for calculating the semi-variograms. The regular and robust semi-variograms were obtained with this method and were plotted against the lag distance. The specific point on the plotted curve, from

where onwards the curve is approximately a straight line parallel to the lag distance axis, is known as the range. Distances greater than the ranges are associated with no spatial correlation, and distances less than the range are associated with spatial correlation. The PROC MIXED procedure of SAS was also used to consider spatial correlation and determine its influence on the F-test and model significance. A two-part PROC MIXED analysis obtained the results with and without considering the spatial correlation of the data. The AIC (Akaike Information Criterion), AICC (Akaike's Information Corrected Criterion), and BIC (Schwartz Bayesian Information Criterion) statistics were obtained. These statistics measure relative goodness of fit. Values, obtained for a model not accounting for spatial correlation and one accounting for spatial correlation, were compared, and if the former had values lower than the latter, it was assumed that spatial correlation existed in the data (Littell et al., 1996). This assumption was tested for only the WMLR analysis. For the other analyses, it was assumed that there was no spatial correlation in the field.

Outliers

An observation that is inconsistent with rest of the dataset in terms of the dependent variable is known as an outlier. To detect outliers in the dataset, residuals (standardized) were plotted against predicted values of the dependent variable (actual radiance) with the PROC GPLOT procedure in SAS. The Bonferroni correction approach, which adjusts the significance level (α) by dividing it with the total number of observations (i.e., $\alpha = \alpha/n$) was used to identify outliers in the observation. According to the Bonferroni correction, the observation (i) is an outlier if

$$| r_i | > t_{n-p-1,\, \alpha/2n} ,$$

where, r_i = residual (standardized in this case) of the i^{th} observation

t = t-statistic

n = total number of observations

p = number of independent parameters in the model

It is important to note that the Bonferroni correction is conservative and will tend to identify relatively few points as outliers. Once outliers are identified, it is important to verify that the observations are in reality outliers. Hence, field conditions at the samples points identified as outliers were checked in the images. It is known that vegetation and soils have distinctly different reflectance spectra, so if there were vegetation at some location in the field during the time of soil radiance data collection, data from that specific location would not match with the pattern of other data points. The same would be true if there were standing water at a particular location in the field. All statistical analyses were performed with SAS software.

CHAPTER III

RESULTS AND DISCUSSION

Simple linear relationship between simulated and actual Landsat radiance spectra of soils

Simple linear regression (SLR) analysis was carried out between simulated Landsat reflectance data and actual Landsat radiance (DN) data of soils in fields 1 and 3. Equation 1 describes the linear relationship of these two variables.

Field 1

Table 5 includes the statistical parameters, F and p, along with R^2 for the SLR analysis of field-1 data for years 1997 and 2001. The first step in evaluating the data was to determine if a linear relationship existed between the dependent and independent variables. Hence, scatter plots of data points of dependent variable (actual radiance) versus independent variable (expected reflectance) of all the bands were created.

Table 5: Statistical parameters of SLR analysis of field-1 data from 1997 and 2001

Model: B1 = B0 + B1Band1+e 1997 (Landsat 5)				**Model: B1 = B0 + B1Band1+e 2001 (Landsat 7)**		
Bands	**F**	**P**	**R^2**	**F**	**P**	**R^2**
Band1	27.30	<.0001	0.0912	1.91	0.1678	0.0070
Band2	34.19	<.0001	0.1117	1.17	0.2802	0.0043
Band3	34.04	<.0001	0.1112	13.05	0.0004	0.0458
Band4	24.14	<.0001	0.0815	2.32	0.1292	0.0084
Band5	49.36	<.0001	0.1536	0.12	0.7310	0.0004
Band7	47.90	<.0001	0.1497	0.21	0.6463	0.0008

Figures 7 and 8 are scatter plots of field-1 data from 1997 and 2001, respectively. If a linear relationship were evident, the data points would show an increasing or decreasing trend in the plots. The scatter plots of all the bands (bands 1 through 7 except band 6) in the 1997 data (Figure 7) appear to have an increasing trend. Furthermore, the F and p values (Table 5) indicate statistically significant linear trends. All the bands have p values < 0.0001, which strongly rejects the null hypothesis, H0: $\beta_1 = 0$ (slope is zero) at $\alpha = 0.05$. Thus, it can be reasonably stated that the data in all the studied bands have linear relationships. The highest correlations were with bands 5 and 7, which had R^2 values of 0.153 and 0.149 respectively. These were followed by bands 3 and 2 with R^2 values of approximately 0.11. Bands 1 and 4 had R^2 values of less than 0.10. Clearly, the correlations observed with all the bands were low.

A similar analysis was carried out for the 2001 data from field-1. The scatter plot of all the studied bands appears to indicate a non-linear trend except possibly for band 3 (Figure 8), which seems to have a non-zero slope. The F and p values (Table 5) suggest the same conclusion, as all the p values are greater than $\alpha = 0.05$ except for band 3. Hence, the null hypothesis cannot be rejected for bands 1, 2, 4, 5, and 7. The R^2 values (Table 5) tend to indicate no correlation except for band 3, which has a very low R^2 value of 0.04. As the 2001 data generally indicate no linear relationship between actual radiance and expected reflectance, no further analysis of this dataset was carried out, and only the 1997 dataset was used for other statistical analysis of field 1.

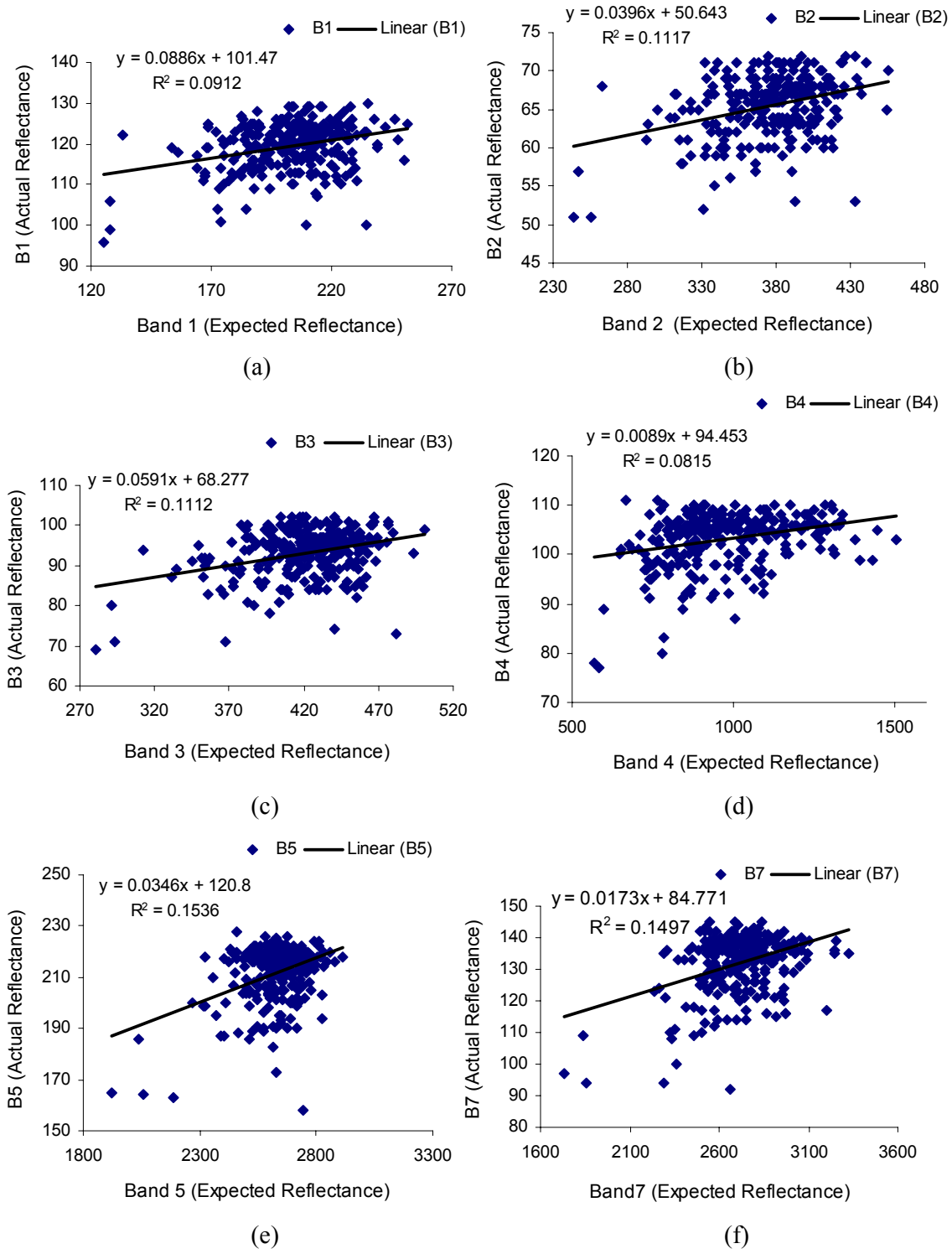


Figure 7: Linear relationship in bands 1 through 7 of 1997 field-1 data

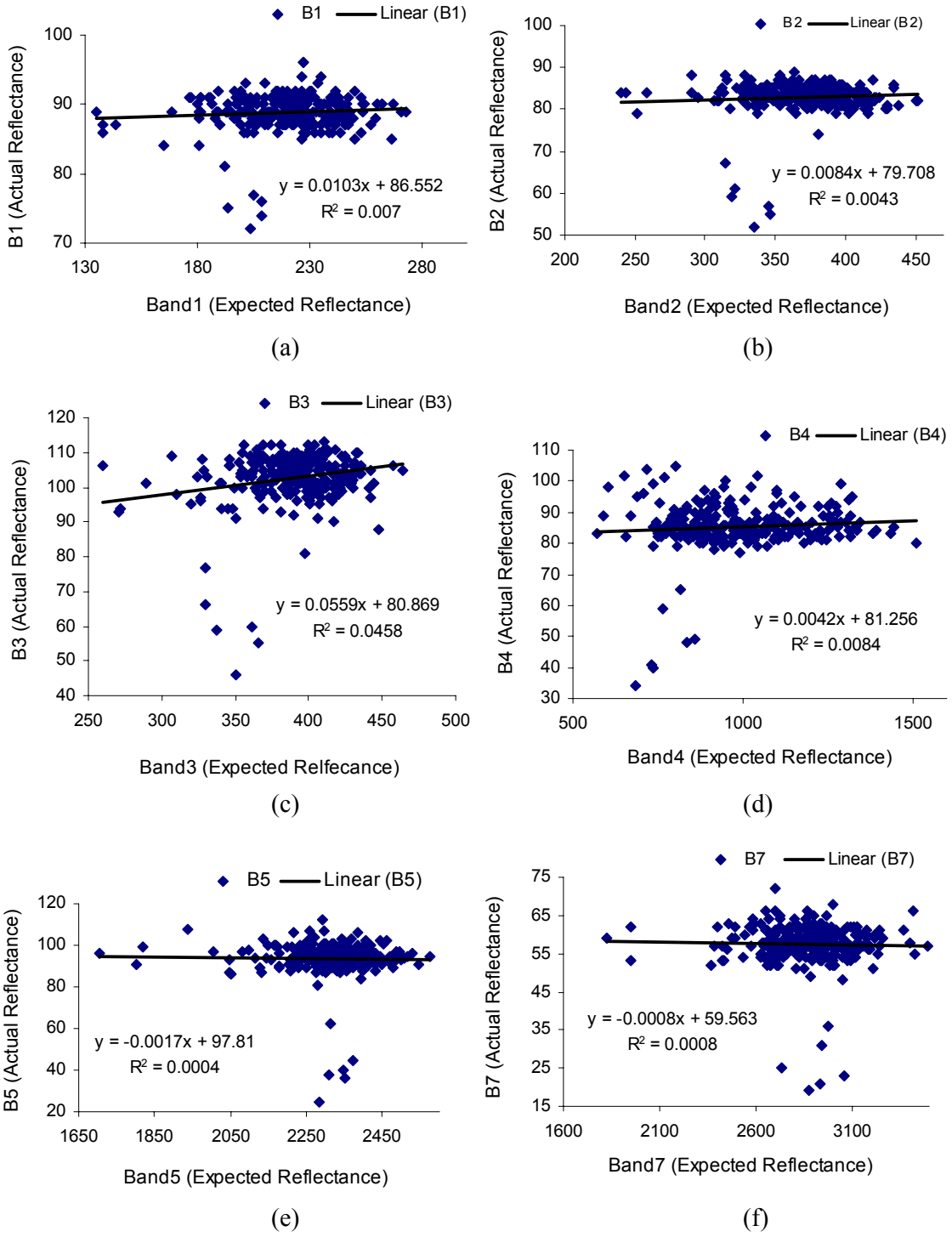


Figure 8: Linear relationship in bands 1 through 7 of 2001 field-1 data

Field 3

Table 6 includes the statistical parameters, F and p, along with R^2 for the SLR analysis of field-3 data for years 1999 and 2001.

Table 6: Statistical parameters of SLR analysis of field-3 data from 1999 and 2001

Model: B1 = $\beta_0 + \beta_1$Band1+ ε 1999 (Landsat 5)				**Model: B1 = B0 + B1Band1+e 2001 (Landsat 7)**		
Bands	**F**	**P**	**R^2**	**F**	**P**	**R^2**
Band1	2.93	0.0881	0.0097	102.96	<.0001	0.2562
Band2	12.53	0.0005	0.0402	100.84	<.0001	0.2522
Band3	8.91	0.0031	0.0289	116.48	<.0001	0.2804
Band4	0.14	0.7085	0.0005	3.15	0.0771	0.0104
Band5	0.00	0.9442	0.0000	46.24	<.0001	0.1339
Band7	0.31	0.5801	0.0010	32.76	<.0001	0.0988

In scatter plots of dependent vs. independent variables of 1999 field-3 data (Figure 9), bands 1, 2, and 3 appear to have linear trends in the data, while bands 4, 5, and 7 appear not to have linear trends. The p values of only bands 2 and 3 are less than 0.05, and for all the other bands it is greater than 0.05. Hence, the null hypothesis, H0: $\beta_1 = 0$ (slope is zero), cannot be rejected for the bands other than 2 and 3, which can be concluded to have a significant linear relationship, albeit with low correlation, having R^2 values less than 0.05. Since most of the bands had no significant linear relationship, and the bands having a significant linear relationship had a very low R^2 value, the 1999 data from field 3 were considered sparingly in further analysis.

The scatter plots of the 2001 dataset of field 3 provided better results (Figure 10) with all bands except band 4, indicating a linear relationship. Similar results were observed in the statistics, as p values less than 0.05 were observed for all bands except band 4 (p =0.077). Thus, the null hypothesis was rejected for bands 1, 2, 3, 5, and 7,

and it was concluded that a significant linear relationship exists between actual radiance and expected reflectance in these bands. In addition, all these bands had R^2 values of approximately 0.10 or greater, with bands 1, 2 and 3 having R^2 values of 0.25, 0.25, and 0.28, respectively.

The foregoing preliminary analysis of the datasets indicates that 1997 field-1 data and 2001 field-3 data had linear relationships, whereas most of the bands in the other data (1999 field-3 and 2001 field-1) exhibited no linear trends, and so these latter data were used sparingly in further analysis. It is not clear why only one set of data for each field appeared to be good, but it is postulated that problems with the timing of the image acquisitions (e.g., vegetation or standing water in the field) or atmospheric problems existed in the weaker sets of Landsat data. Nevertheless, one good set of data remained for each field.

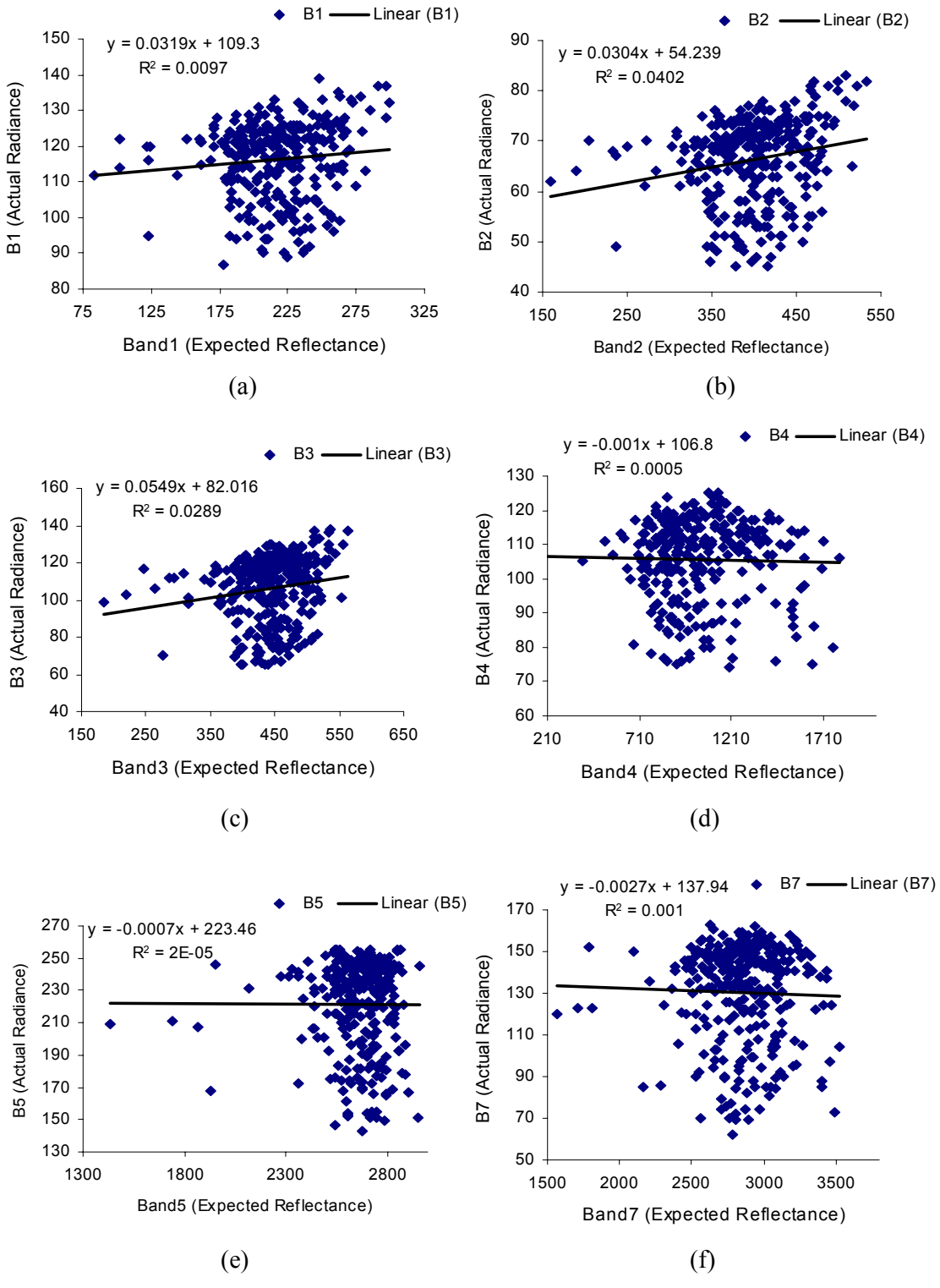


Figure 9: Linear relationship in bands 1 through 7 of 1999 field-3 data

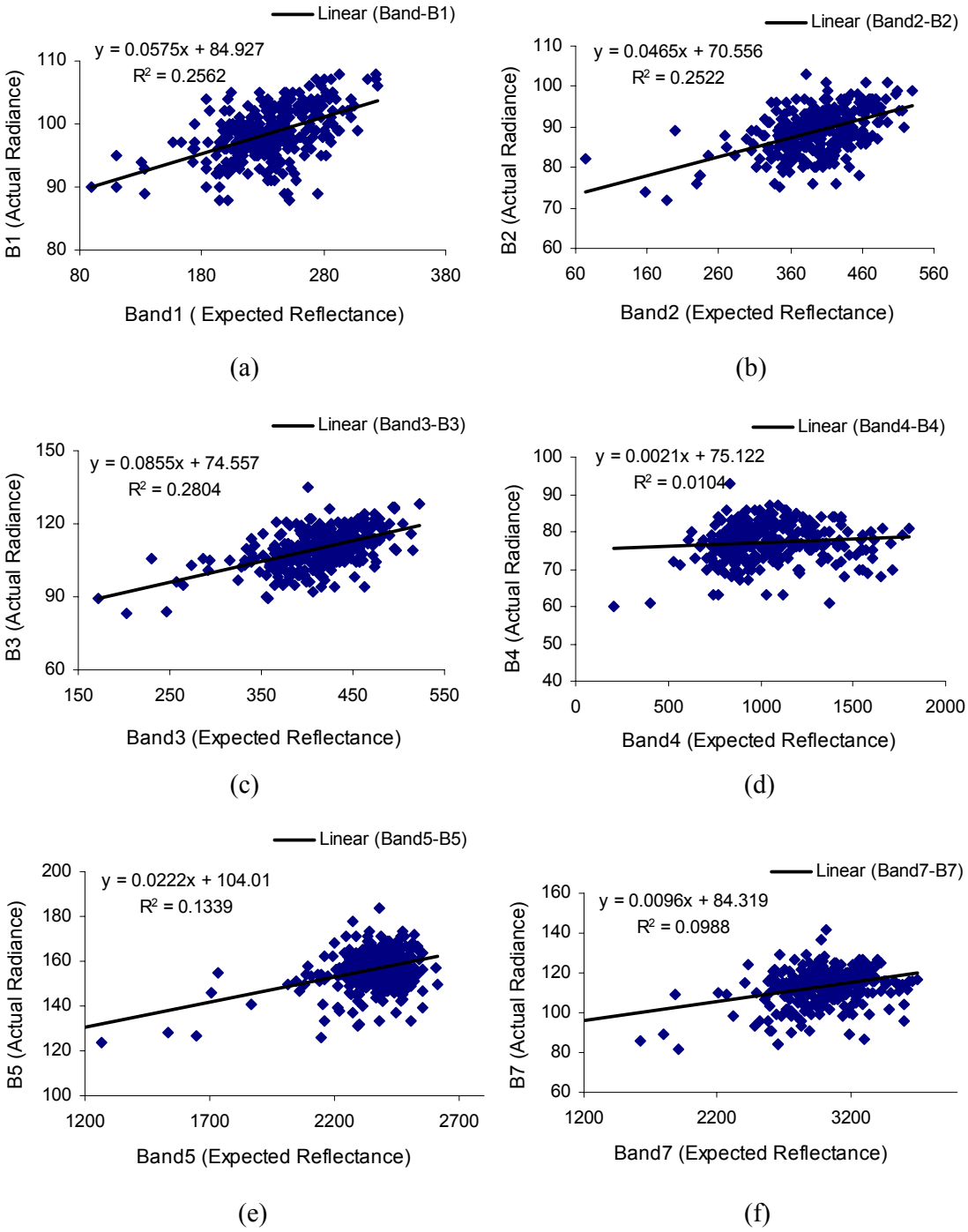


Figure 10: Linear relationship in bands 1 through 7 of 2001 field-3 data

Assumptions

Normality

The normal probability plots of the SLR model with 1997 field-1 spectral data (Figure 11) indicate that residuals in bands 1 through 3 align linearly along the normal reference line, however the upper and lower tails deviate somewhat from the reference line. It is observed in all the bands that a few data points in the lower tails are far away from the reference line, indicating possible outliers. Bands 4 through 7 have curve shapes with both upper and lower tails bending downwards, indicating skewness to the left side of the normal bell shaped curve. Hence, it can be seen that bands 1 through 3 appear to be normal, and bands 4 through 7 are skewed. Irrespective of this, it is important to note that the F-test is robust against non-normality; so minor non-normality is not of great concern with respect to regression analysis in this study. Similar analysis of 2001 field-3 data (Figure 12) indicates that bands 1 through 7 are normally distributed, as the residuals are well aligned along the normal reference line. An outlier is indicated in band-3, as one data point at the upper tail appears to be relatively far from the straight line. The normal probability plots of band-3 in 2001 field-1 data, and bands 2 and 3 in 1999 field-3 data, which had linear relationships, are provided in figure 13. It is apparent that the residual data points do not align well with the normal reference line in any of these plots. Certain data points at the lower tail in the 2001 field-1 band-3 data are far from the normal reference line, implicating them as possible outliers. The 1999 field-3 data in bands 2 and 3 have S-shaped curves; Thus, non-normality is indicated in these three plots.

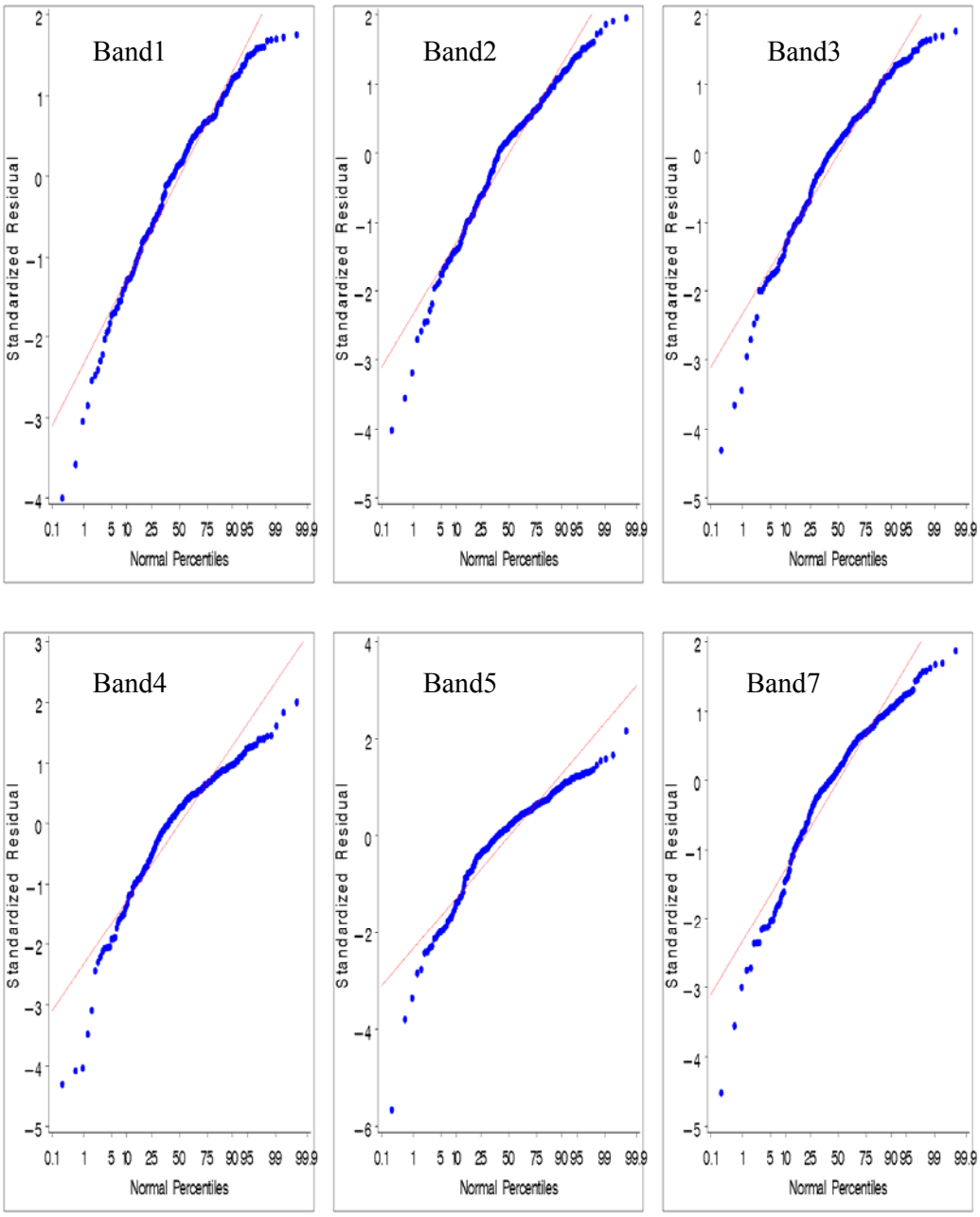


Figure 11: Normal probability plots for bands 1 through 7 of 1997 field-1 data (SLR)

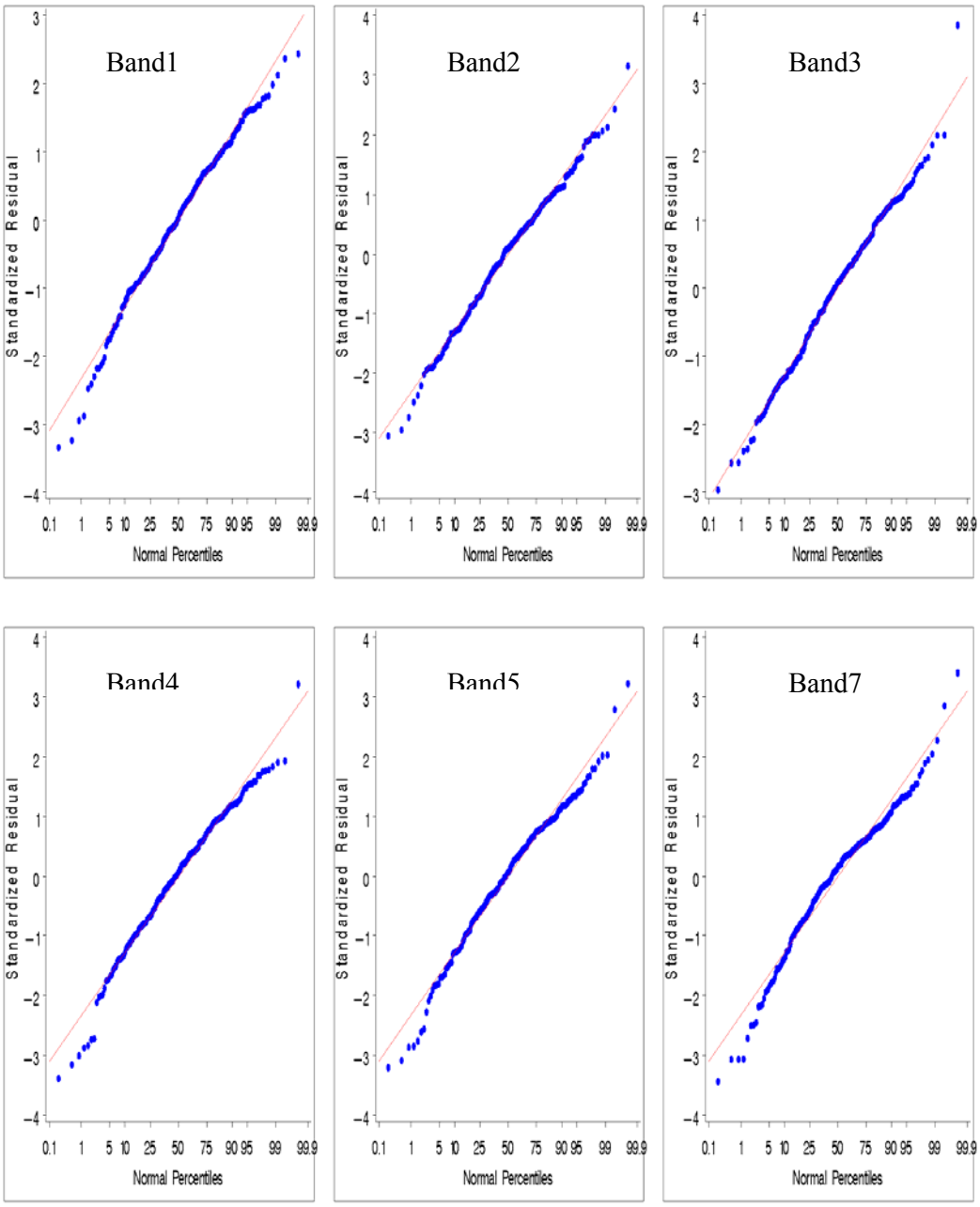


Figure 12: Normal probability plots for bands 1 through 7 of 2001 field-3 data (SLR)

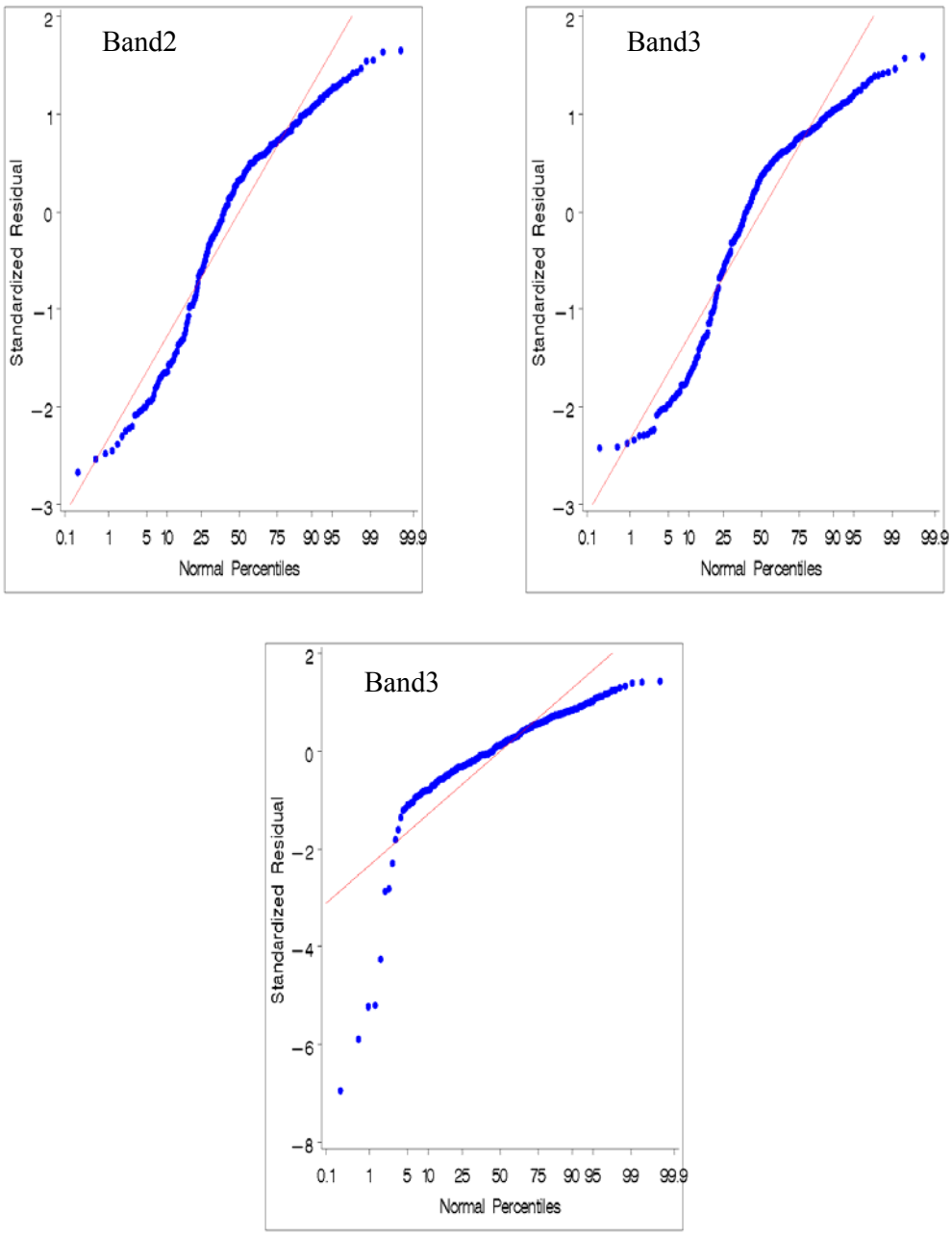


Figure 13: Normal probability plots for band 3 of 2001 field-1 data (bottom) and bands 2 and 3 of 2001 field-3 data (top), (SLR)

Constant Variance

Plots of standardized residuals vs. predicted radiance (dependent variable) for 1997 field-1 data of bands 1 through 7 (Figure 14) indicate that the residuals are randomly distributed about zero. However, a few data points appear to be possible outliers, as they are located far from the groupings of most data points. Since no pattern is observed in the plots, the residuals are taken as having homogenous variance. Similar analysis of 2001 field-3 data (Figure 15) indicates similar results; again, no pattern is observed. Possible outliers are evident in these data also.

The bands of 2001 field-1 and 1999 field-3 data that exhibited linearity between simulated and actual Landsat data were also analyzed for the constant variance assumption. In each case (bands 2 and 3, field-3, 1999; and band-3, field-1, 2001) the residuals are randomly distributed about zero. However, a few data points in the 2001 field-1 band-3 data are far from the groupings of most data points and seem to be outliers. No outliers are observed in the 1999 field-3 data from bands 2 and 3. Overall, it can be concluded that homogenous variance exists in all three bands.

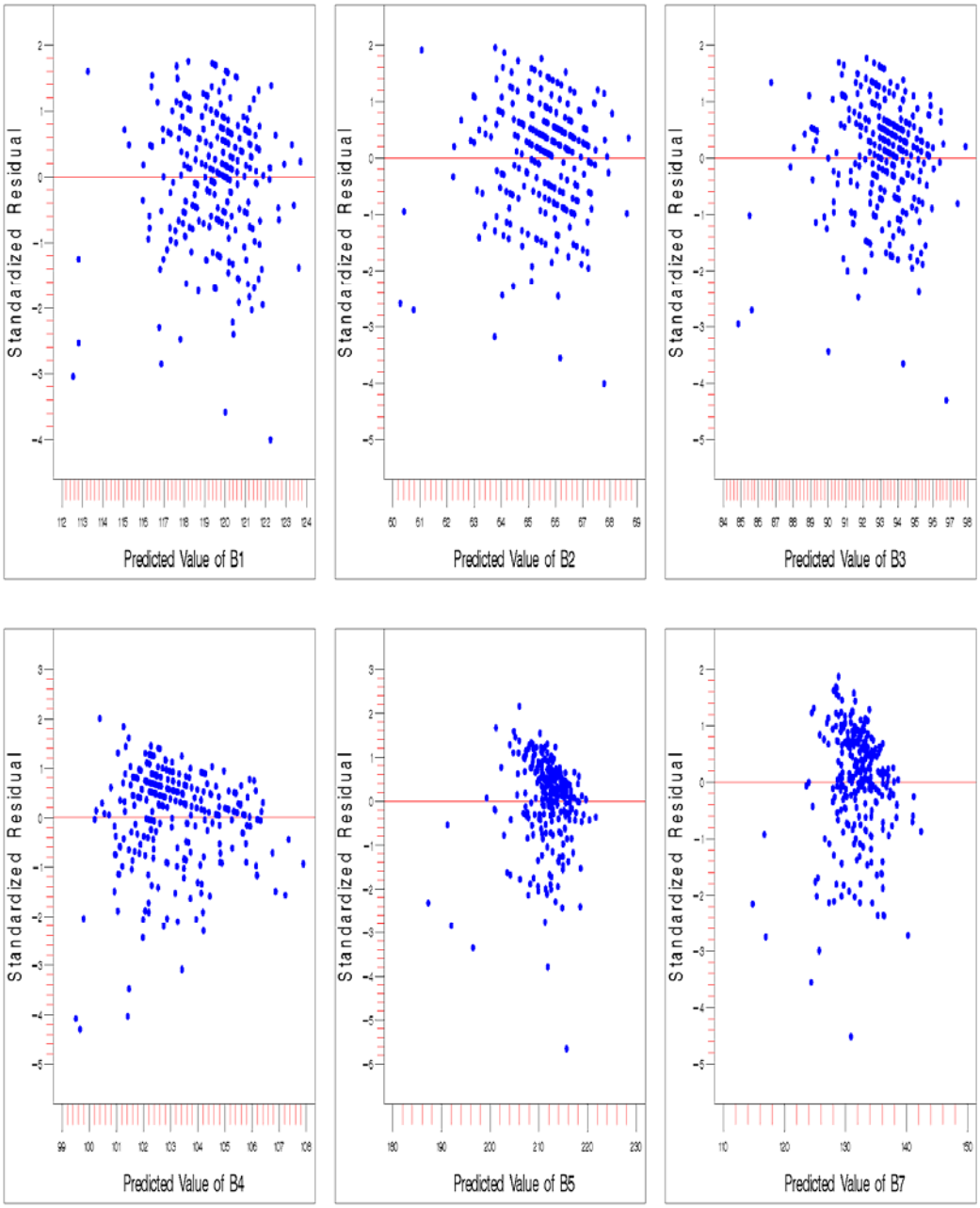


Figure 14: Plots of standardized residual vs. predicted value of dependent variable of bands 1 through 7 of 1997 field-1 data (SLR)

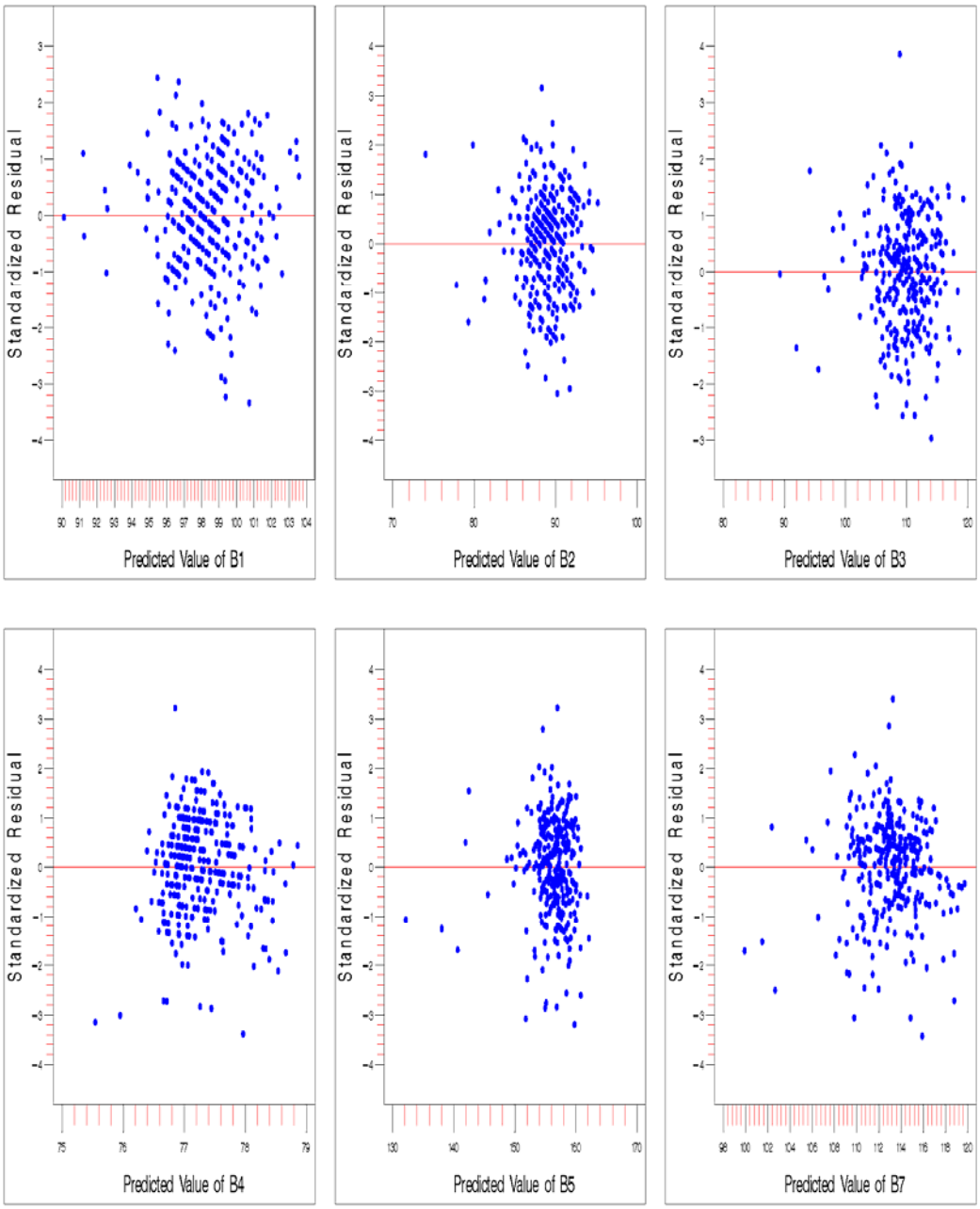


Figure 15: Plots of standardized residual vs. predicted value of dependent variable of bands 1 through 7 of 2001 field-3 data (SLR)

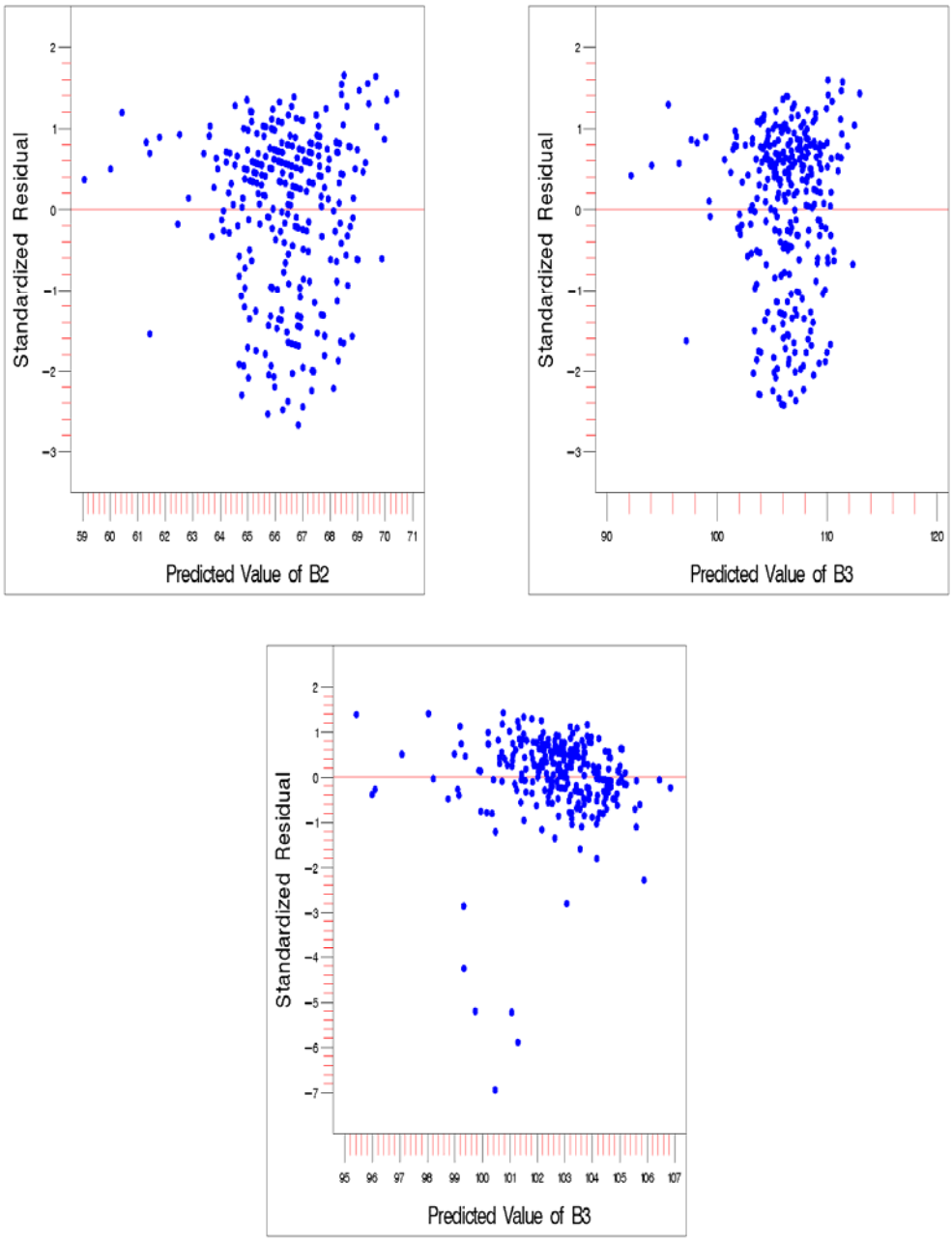


Figure 16: Plots of standardized residual vs. predicted value of dependent variable for band-3 of 2001 field-1 data (bottom) and bands 2 and 3 of 1999 field-3 data (top) (SLR)

Outlier analysis

Detection

The outliers' boundary values were obtained with the Bonferroni correction approach described in chapter II. According to this approach, residuals greater than $|t_{n-p-1,\ \alpha/2n}|$ are consider outliers. Hence, for

Field 1: $t_{274-1-1,\ 0.05/\ 2\ (274)} = t_{272,\ 0.00009} = 3.29$ and -3.29

Field 3: $t_{301-1-1,\ 0.05/\ 2\ (301)} = t_{299,\ 0.00008} = 3.29$ and -3.29

Figures 17 and 18 are plots of standardized residual vs. predicted value of the dependent variable for 1997 field-1 and 2001 field-3 data, respectively. Two red lines at ±3.29 standardized residual values are the outlier boundaries, and data points outside these boundaries are considered as outliers. Therefore, all bands in 1997 field-1 and 2001 field-3 data, except bands 2 and 5 in 2001 field-3 data, include outliers. Similarly, figure 19 indicates that only band 3 in the 2001 field-3 data has outliers. The Reweight and Refit options of the SAS REG procedure were used to remove the outliers prior to developing regression models. New outliers were detected after primary removal of outliers; and the same procedure was repeated until all the outliers were removed.

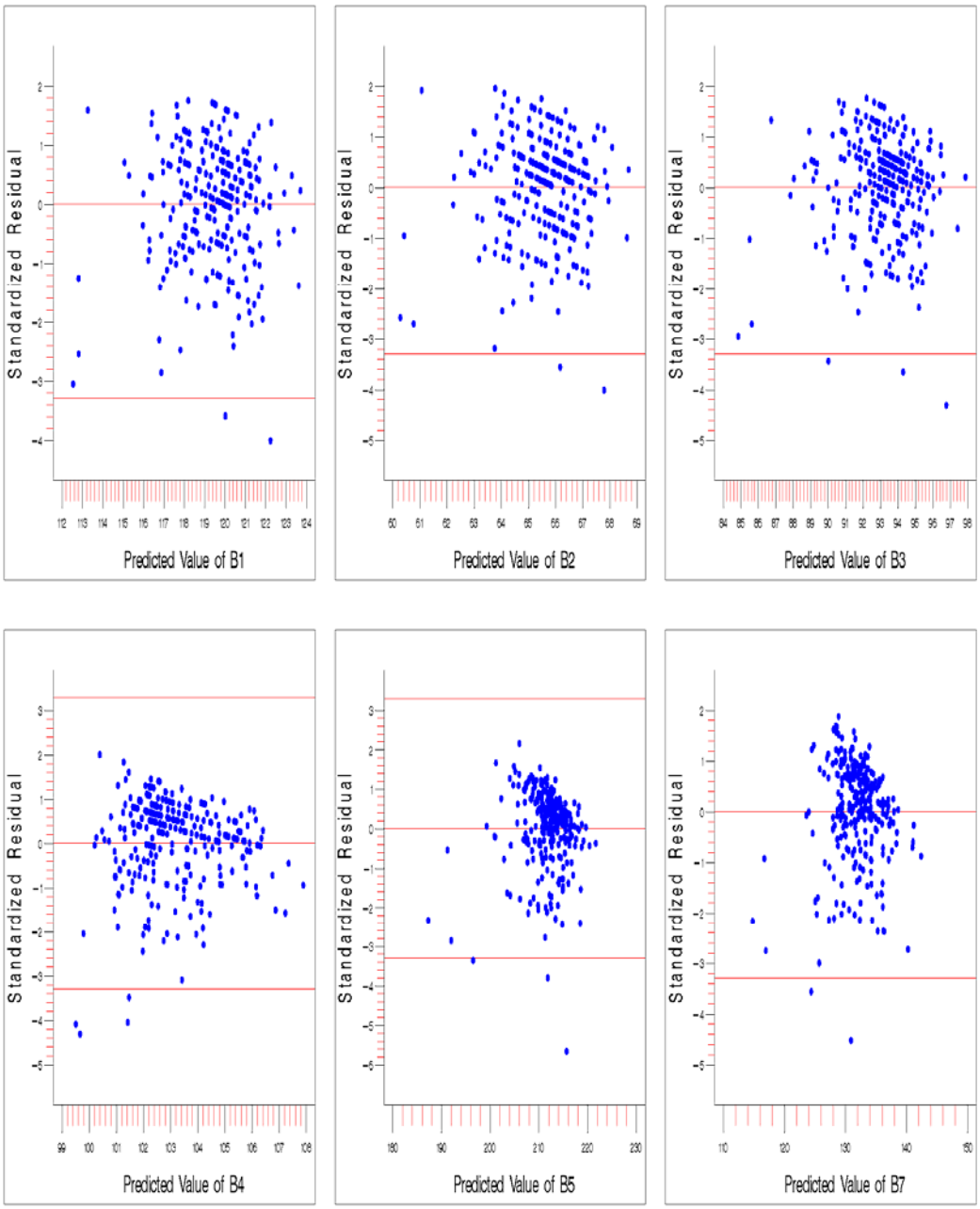


Figure 17: Outlier detection plots of standardized residual vs. predicted value of dependent variable for band 1 through 7 of 1997 field-1 data (SLR)

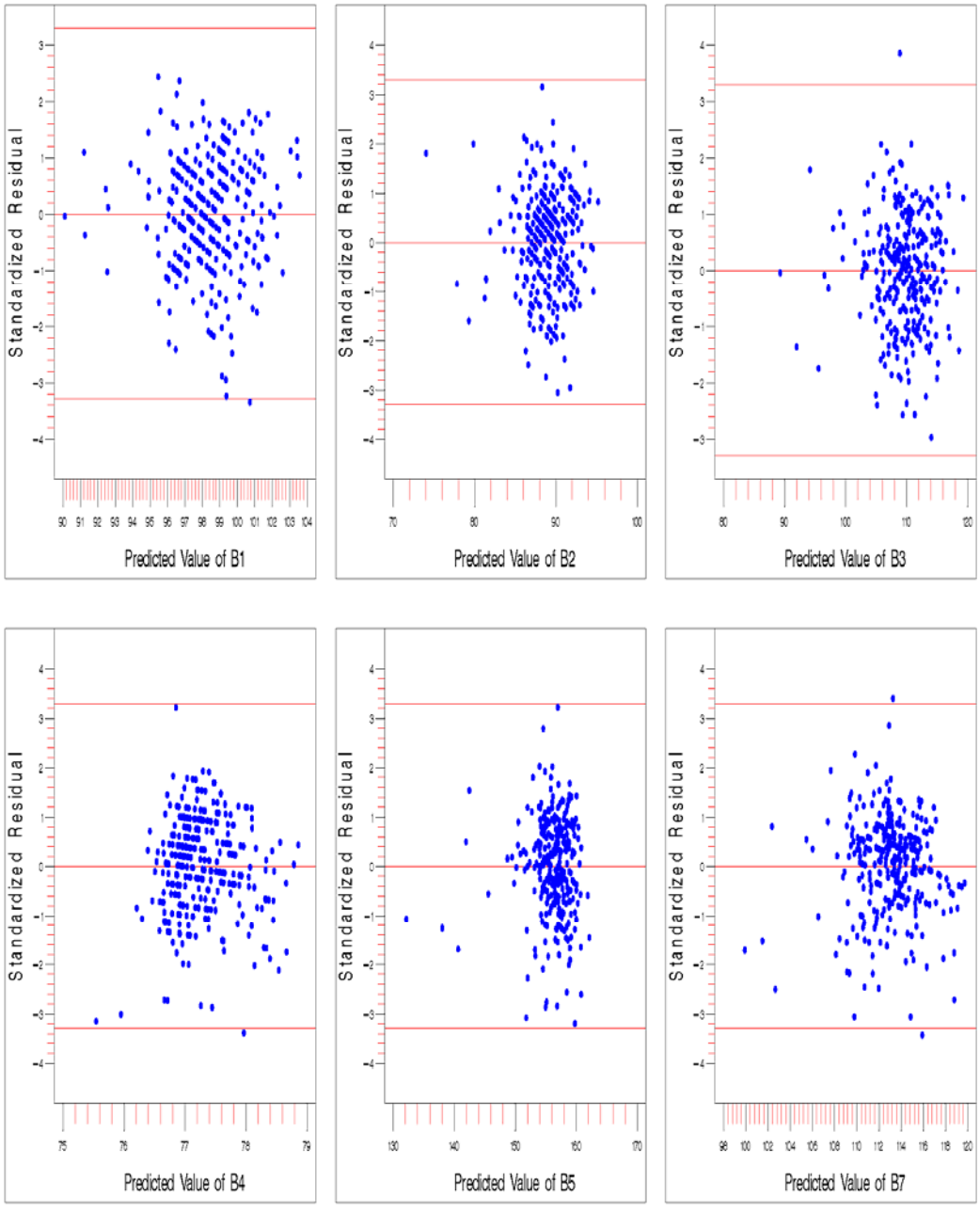


Figure 18: Outlier detection plots of standardized residual vs. predicted value of dependent variable of bands 1 through 7 of 2001 field-3 data (SLR)

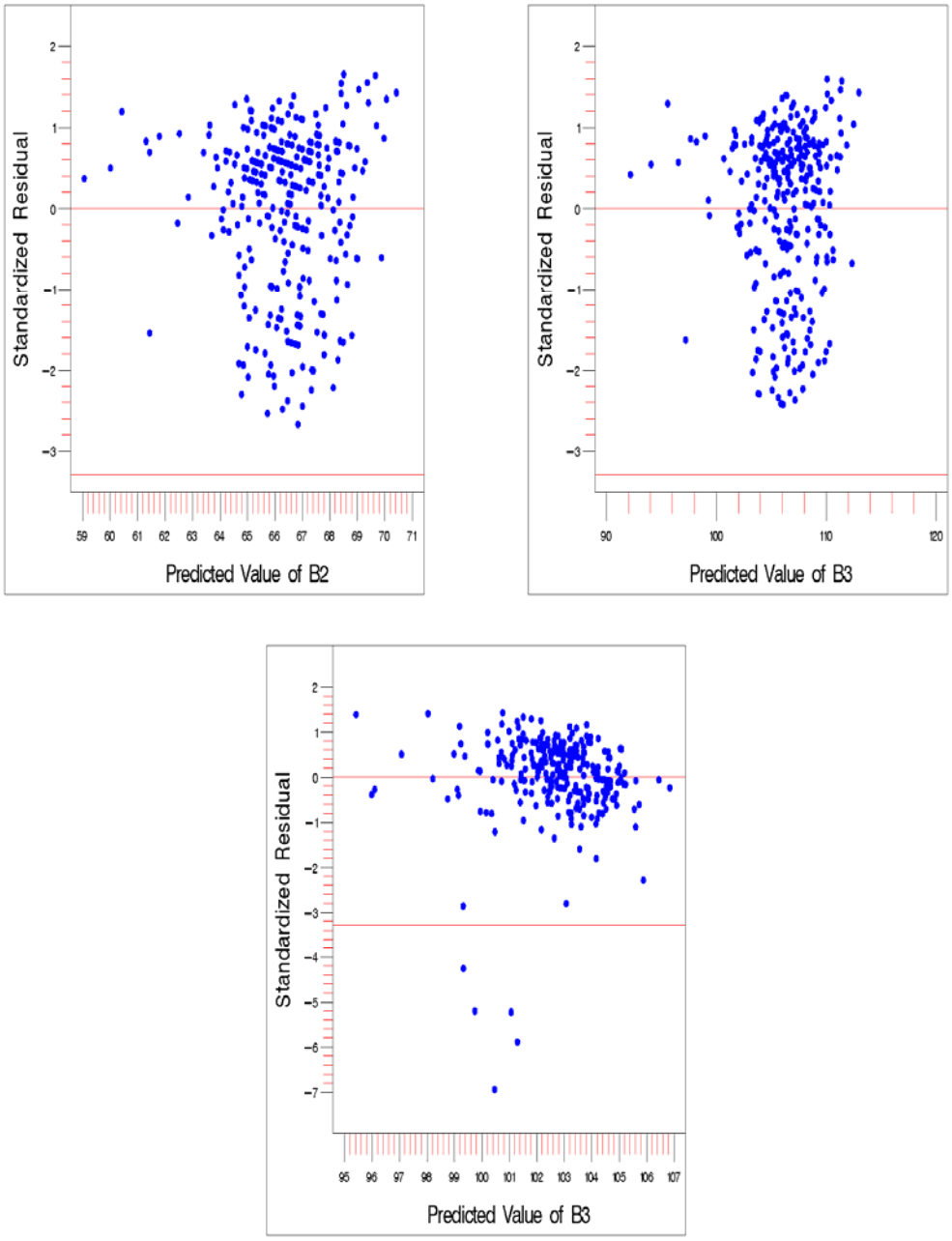


Figure 19: Outlier detection plots of standardized residual vs. predicted value of dependent variable for bands 2 and 3 of 1999 field-3 data (top) and band 3, of 2001 field-1 data (bottom), (SLR)

Removal

Field-1

Figures 20, 22, 24, and 26 are plots of standardized residuals versus predicted values for all bands, before and after removal of outliers. Normal probability plots of the data before and after outlier removal are provided in figures 21, 23, 25, and 27. The results with data including outliers are given in addition to results with data excluding outliers, which have Roman numerals to designate the number of iterations of outlier removal. The constant variance (standardized residual vs. predicted value) plots for bands 1 and 2 seem to have improved after removal of outliers (Figure 20). Improvement in normal probability plots was also observed, as the data points lying far off the reference line at the lower tails were removed as outliers, and thus the curves appear closer to reference line (Figure 21).

The constant variance plots of bands 3 and 4 (Figure 22) and their corresponding normality plots (Figure 23) also indicated improvement in their respective plots after removal of outliers. Bands 3 and 4 required one and two iterations, respectively, to remove the outliers. Band-5 and -7 constant variance plots (Figure 24 and 26) and their corresponding normality plots (Figure 25 and 27) provided similar results as those of the previously discussed bands.

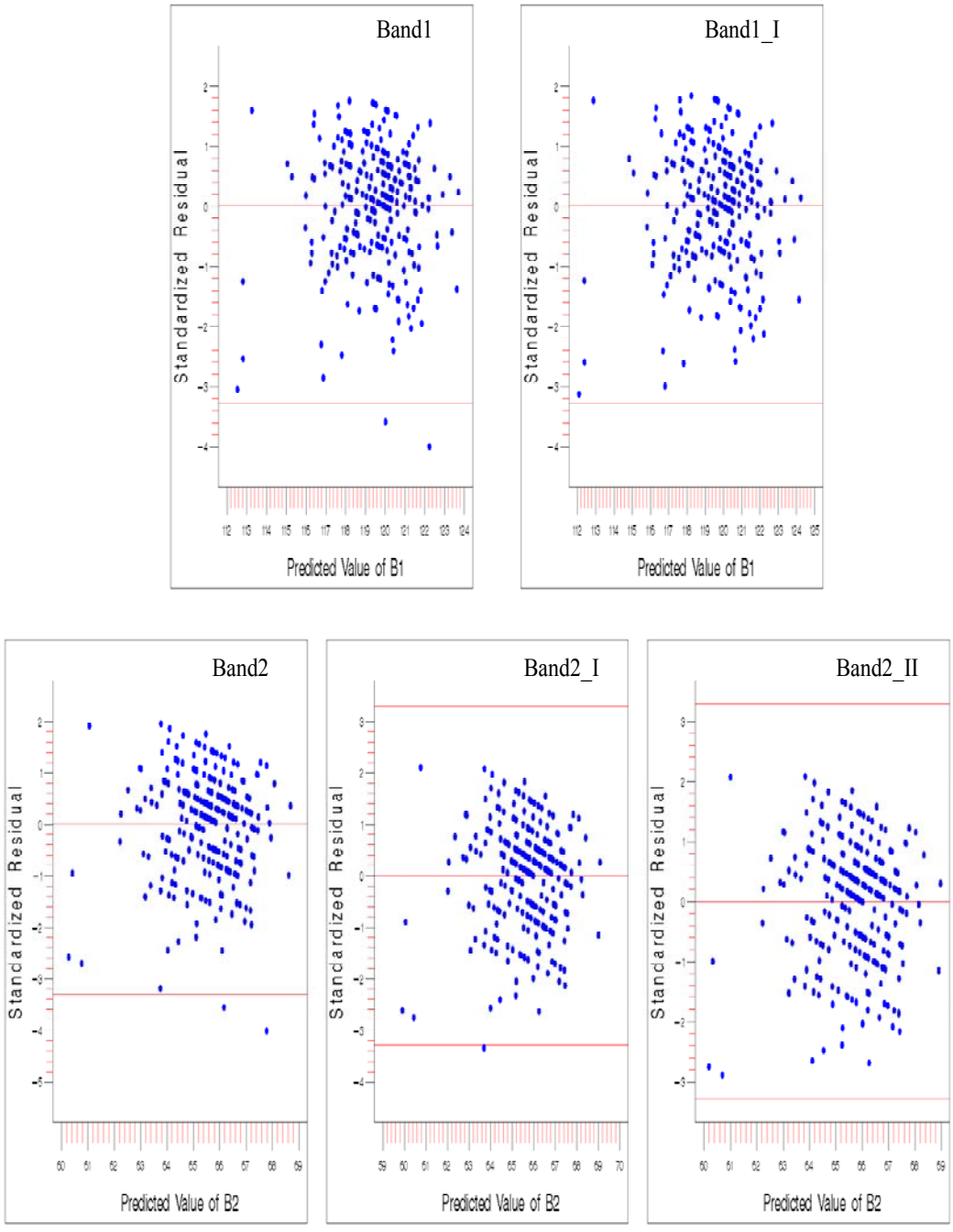


Figure 20: Outlier removal iterations (I, II …) plots of standardized residual vs. predicted value of dependent variable for bands 1 and 2 of 1997 field-1 data (SLR)

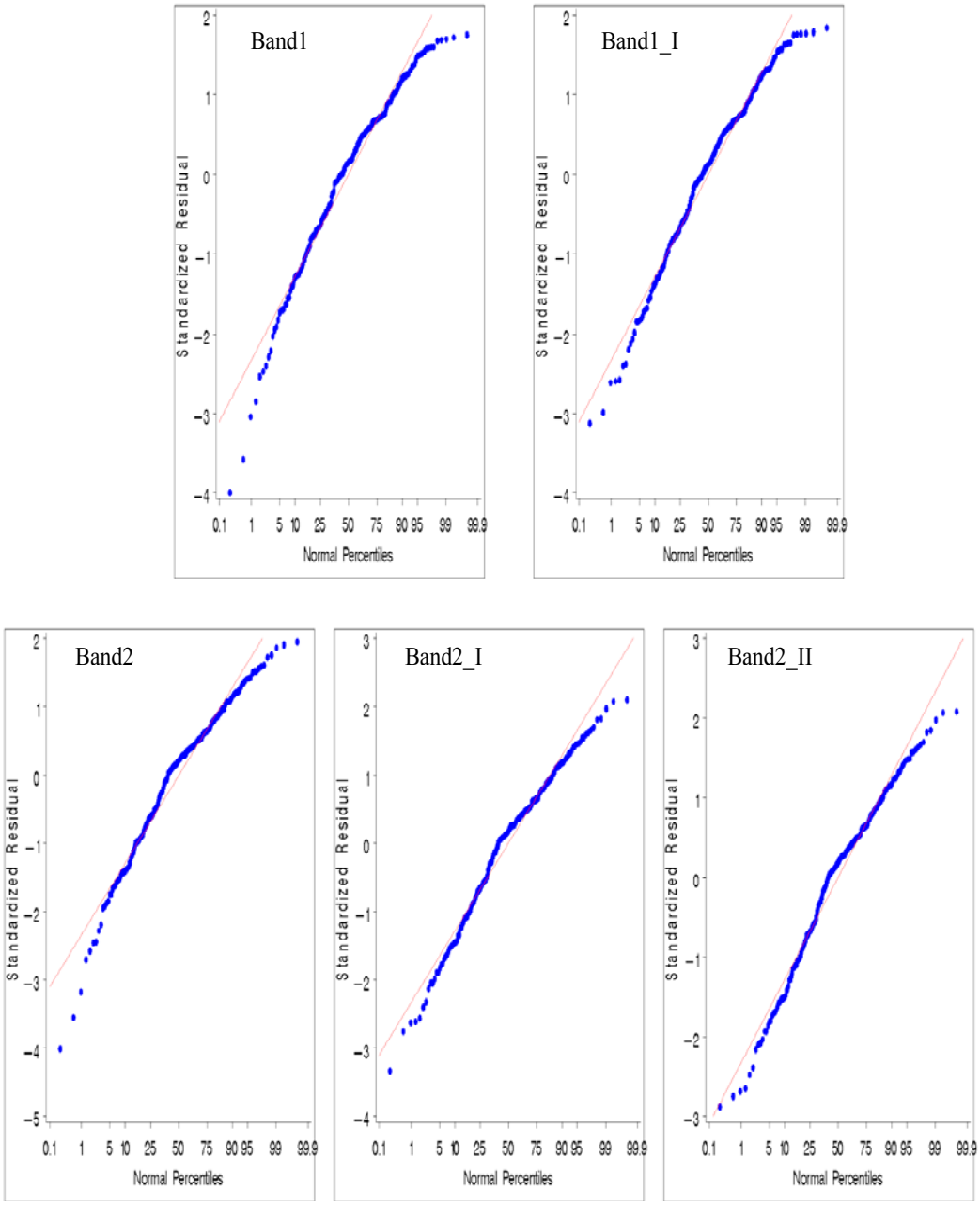


Figure 21: Normal probability plots corresponding to outlier removal iteration (I, II …) plots of bands 1 and 2 of 1997 field-1 data (SLR)

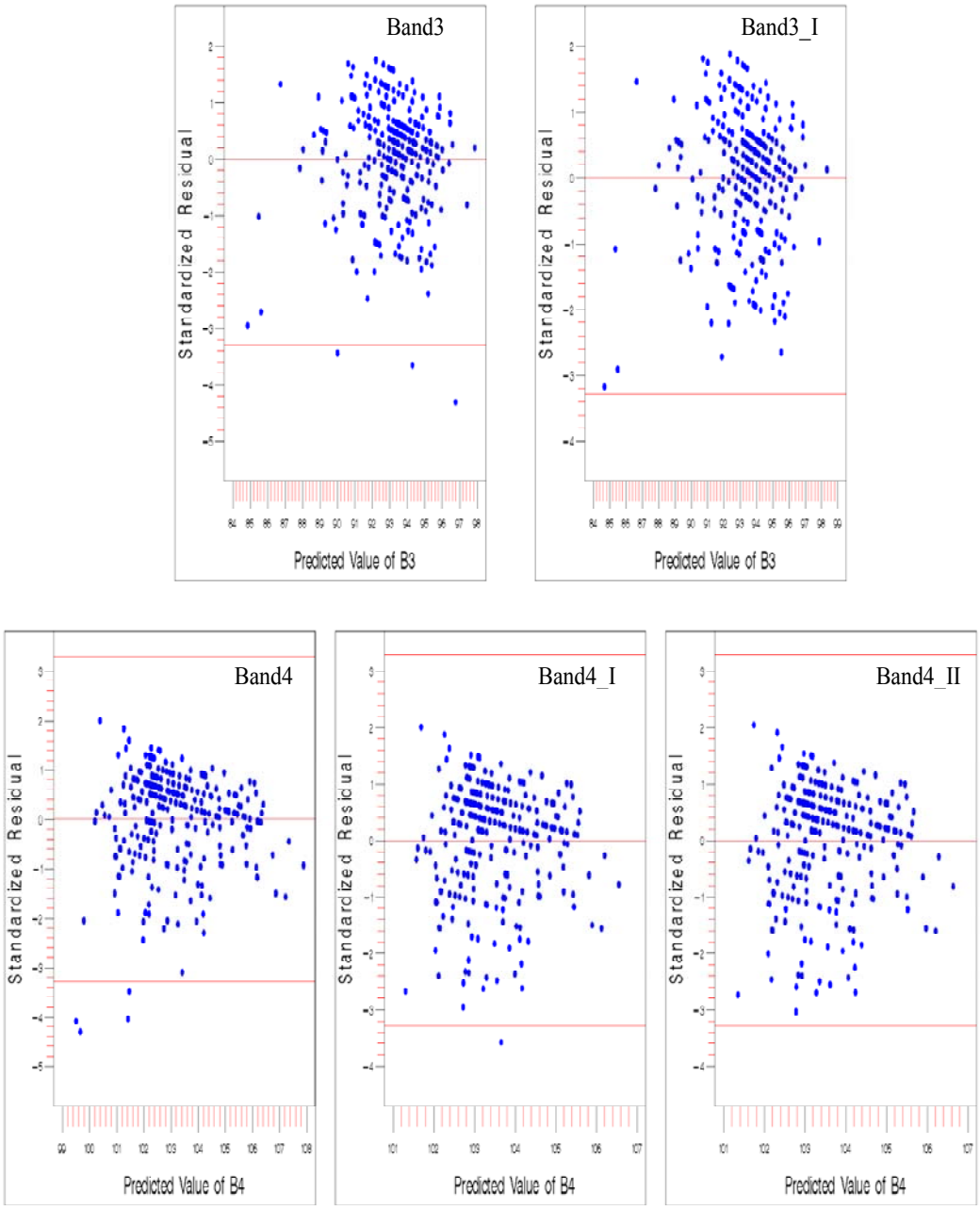


Figure 22: Outlier removal iterations (I, II …) plots of standardized residual vs. predicted value of dependent variable for bands 3 and 4 of 1997 field-1 data (SLR)

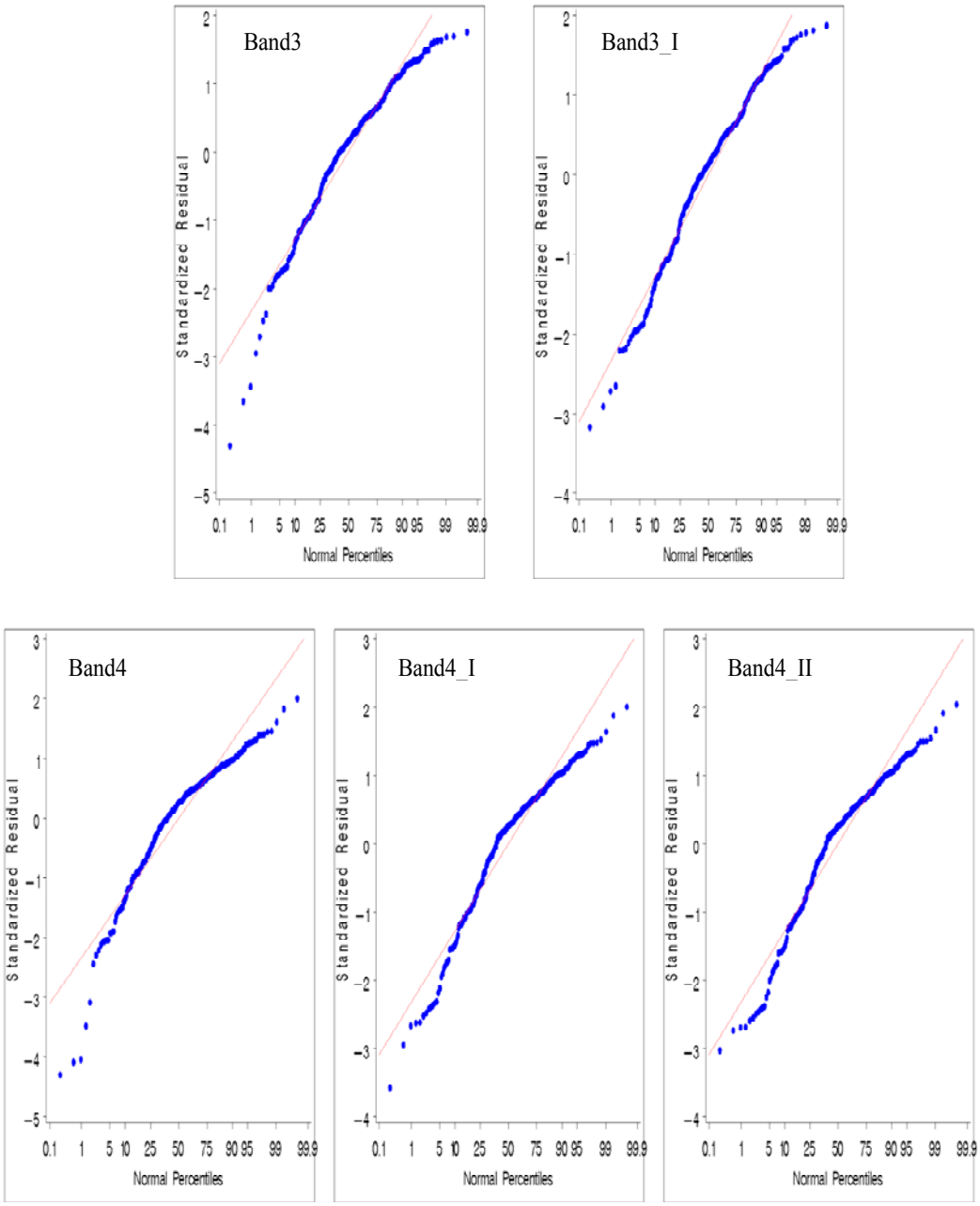


Figure 23: Normal probability plots corresponding to outlier removal iteration (I, II …) plots of bands 3 and 4 of 1997 field-1 data (SLR)

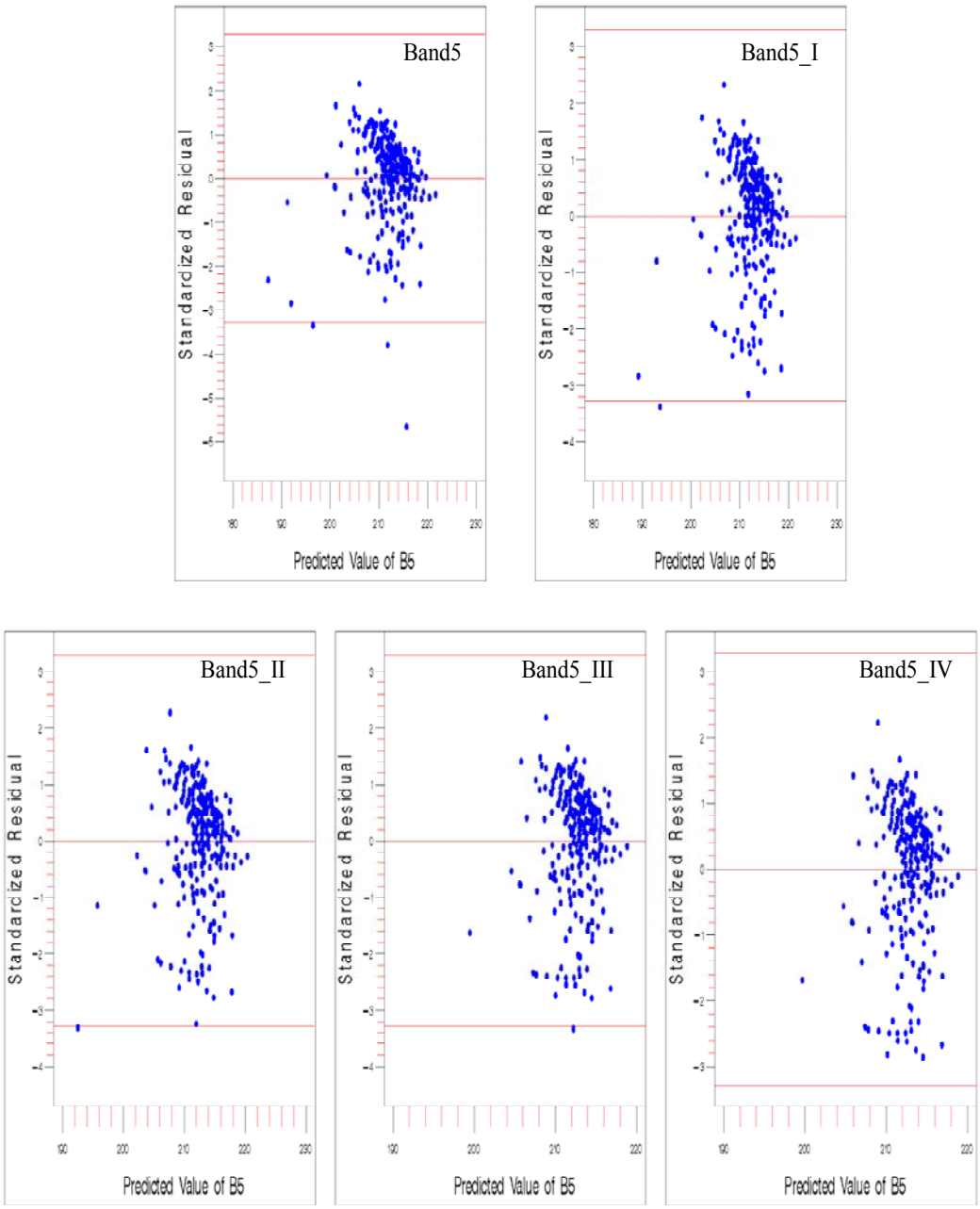


Figure 24: Outlier removal iterations (I, II …) plots of standardized residual vs. predicted value of dependent variable for band 5 of 1997 field-1 data (SLR)

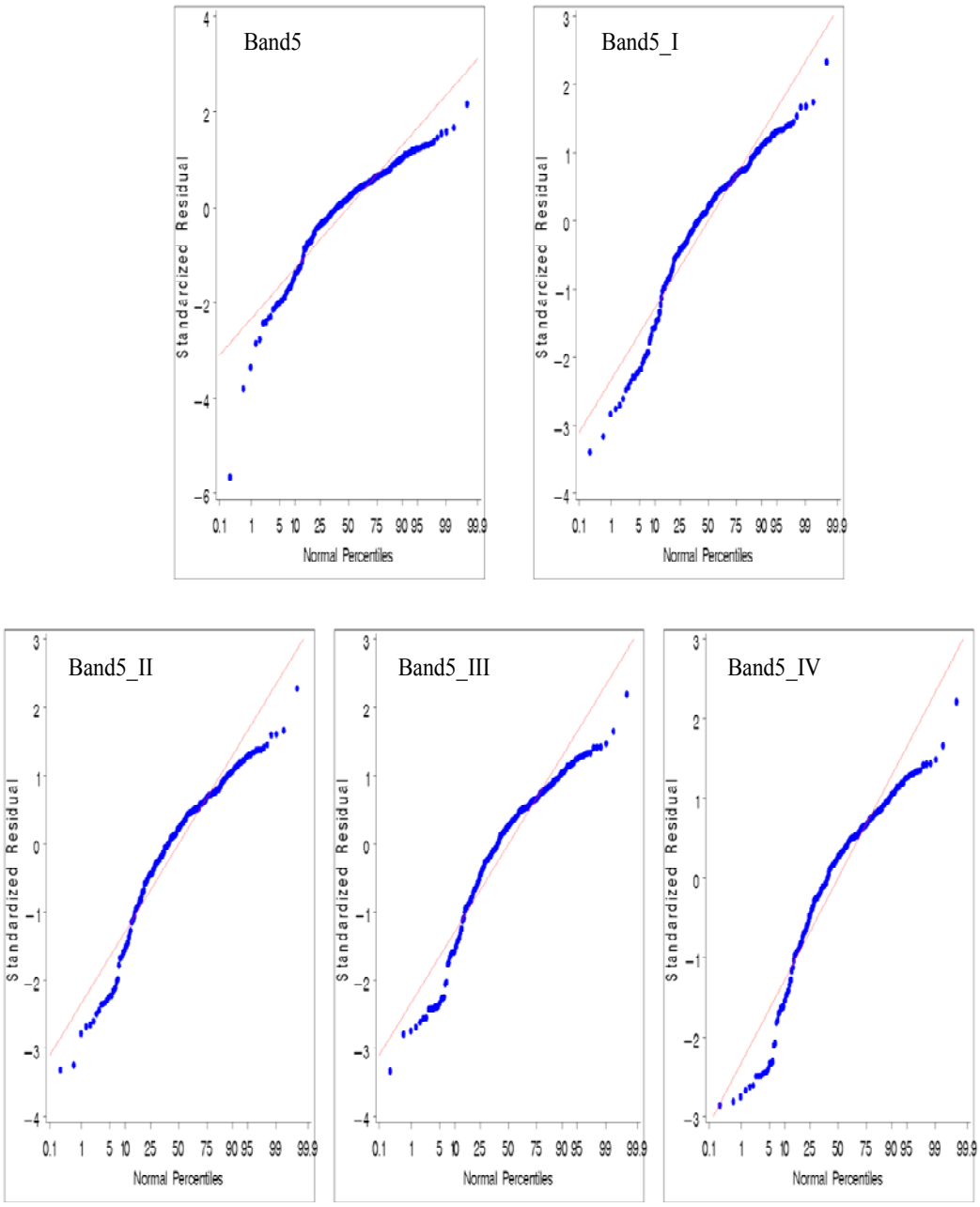


Figure 25: Normal probability plots corresponding to outlier removal iteration (I, II …) plots of band 5 of 1997 field-1 data (SLR)

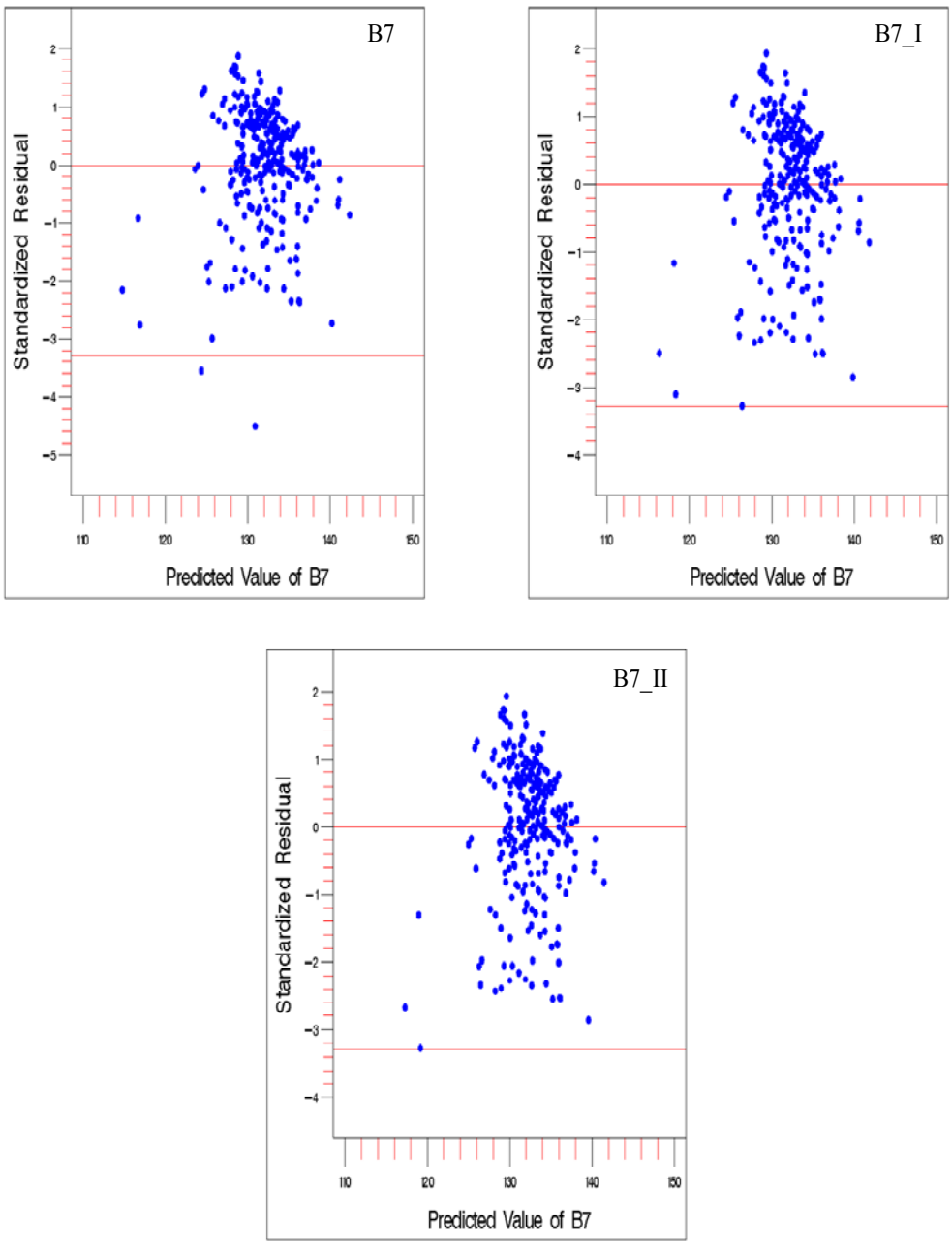


Figure 26: Outlier removal iterations (I, II …) plots of standardized residual vs. predicted value of dependent variable for band 7 of 1997 field-1 data (SLR)

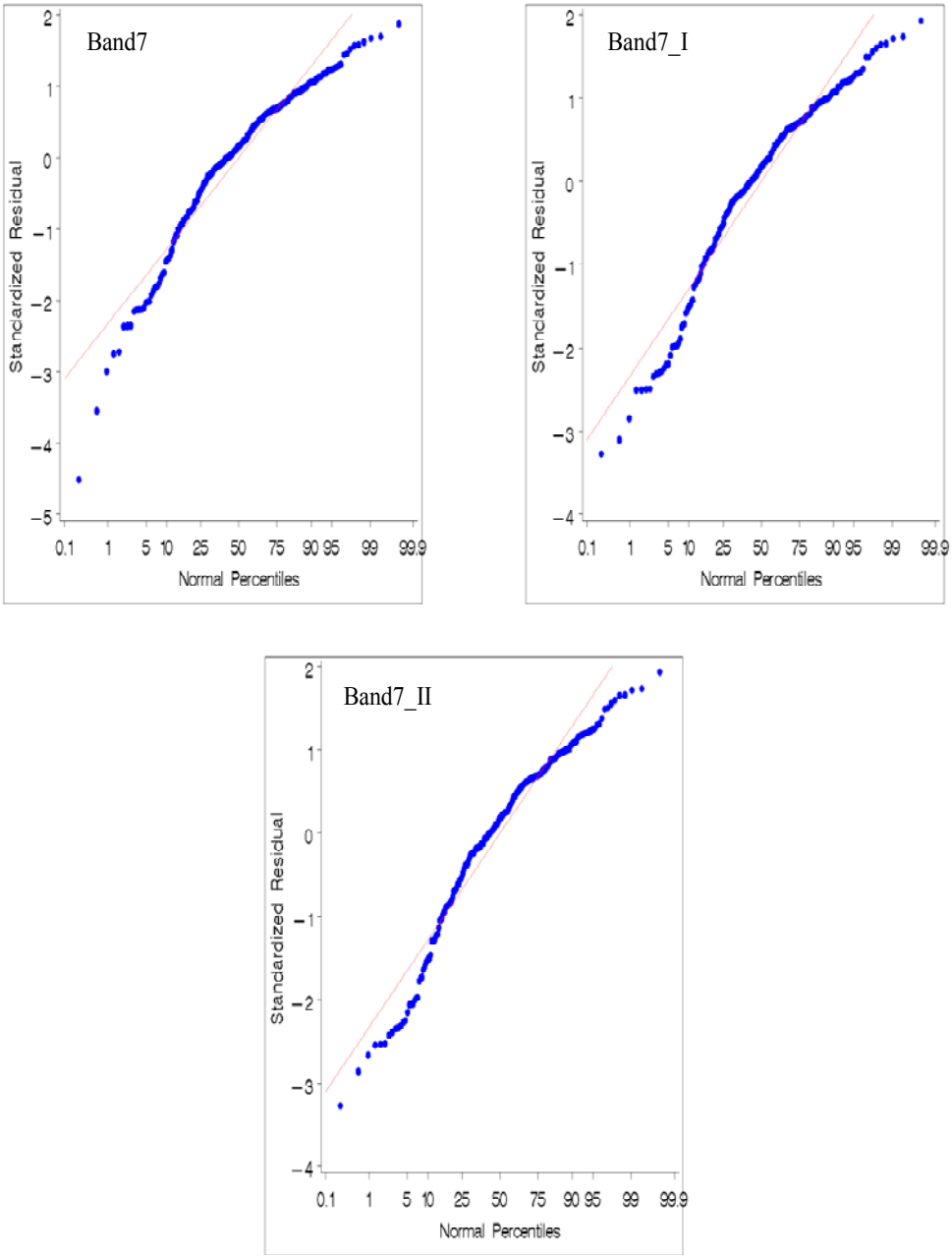


Figure 27: Normal probability plots corresponding to outlier removal iteration (I, II …) plots of band 7 of 1997 field-1 data (SLR)

The outlier removal process for band-3 of 2001 field-1 data is described in figure 28. It is observed that after the first iteration of outlier removal, the plot of constant variance has improved, as the distribution of the residuals across the reference line looks better, but some outliers can still be observed. The second iteration of band 3 removed all the outliers, and the plot indicates better distribution of residuals. Hence, band 3, which previously had outliers and non-constant variance, now appears to have constant variance. The band-3 normality plot, which prior to removal of outliers (Figure 29) had non-normal distribution, has normal distribution after removal of outliers, as the curve seems to be well aligned with the reference straight line.

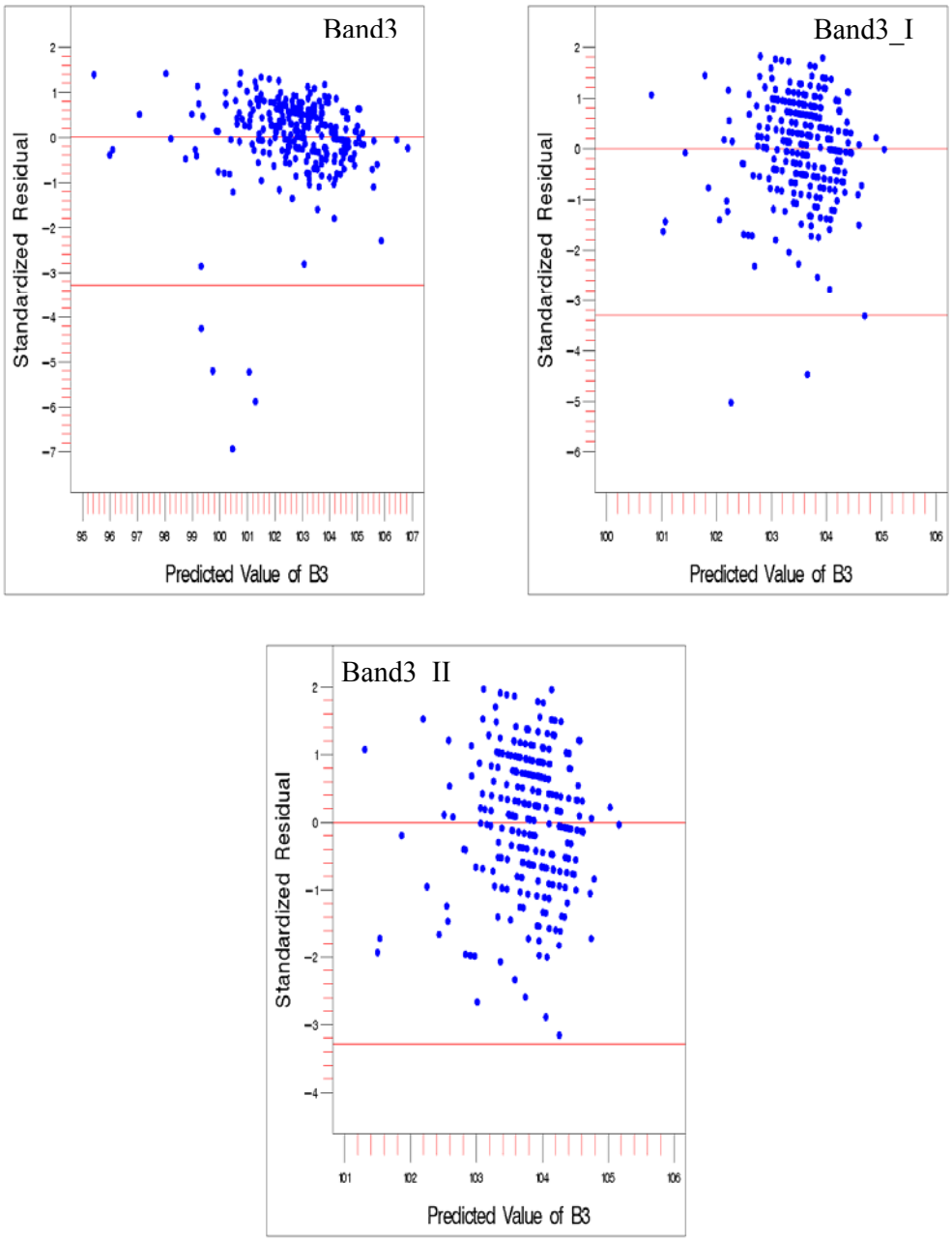


Figure 28: Outlier removal iterations (I, II …) plots of standardized residual vs. predicted value of dependent variable for band 3 of 2001 field-1 data (SLR)

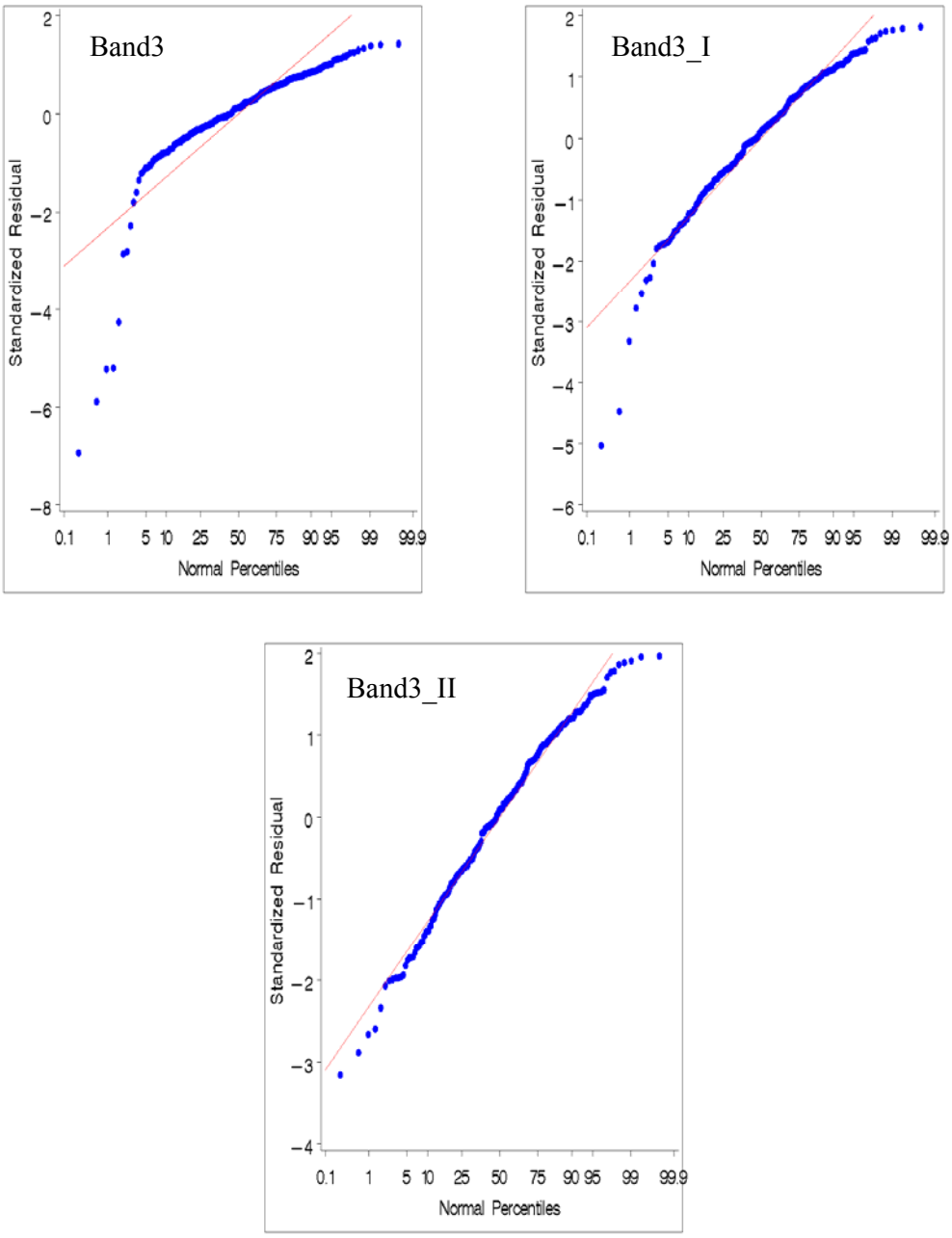


Figure 29: Normal probability plots corresponding to outlier removal iteration (I, II …) plots of band 3 of 2001 field-1 data (SLR)

With the SLR models, the R^2 values of band-1 and band-2 data (Figure 30) improved from 0.09 to 0.11 and 0.11 to 0.14 respectively after the first iteration. Similarly, the R^2 value of band-3 data improved from 0.11 to 0.13, but the R^2 value of band-4 data decreased from 0.08 to 0.04 after removal of outliers (Figure 31). In the case of band-5, the R^2 value (Figure 32) improved from 0.15 to 0.16 after the first iteration but gradually decreased afterwards, whereas for band-7 it remained unchanged. The significance level of the correlations remained high (p values < 0.0001) in all the bands after removal of outliers, except band-4, where the significance level decreased noticeably (p value changed from <0.0001 to 0.0005) upon removal of outliers.

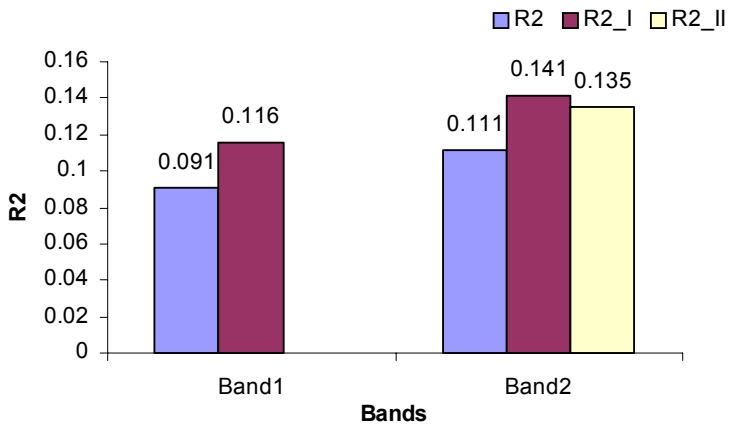


Figure 30: Changes in R^2 values with outlier removal at each iterations (I, II.. ..) for bands 1 and 2 of 1997 field-1 data

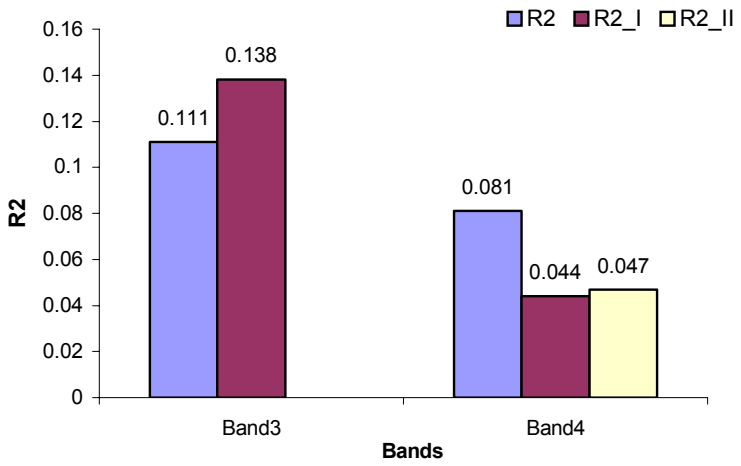


Figure 31: Changes in R^2 values with outlier removal at each iterations (I, II.. ..) for bands 3 and 4 of 1997 field-1 data

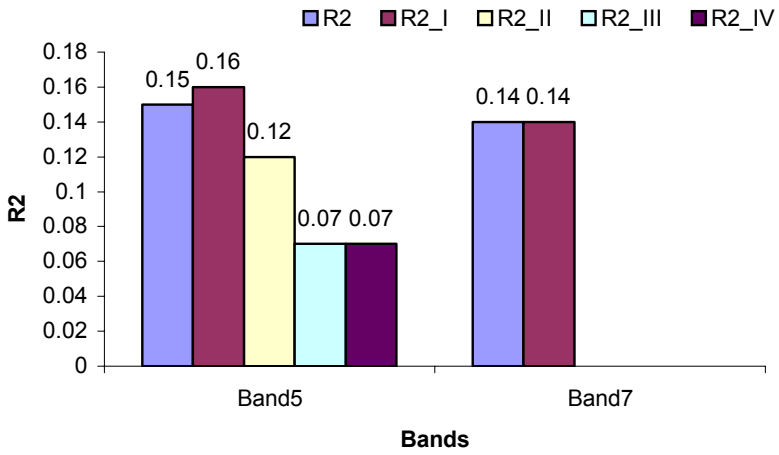


Figure 32: Changes in R^2 values with outlier removal at each iterations (I, II.. ..) for bands 5 and 7 of 1997 field-1 data

The R^2 value for band-3 of 2001 field-1 data decreased from 0.04 to 0.01 (Figure 33), and the p-value, still indicating a significant linear relationship, decreased from 0.0004 to 0.04.

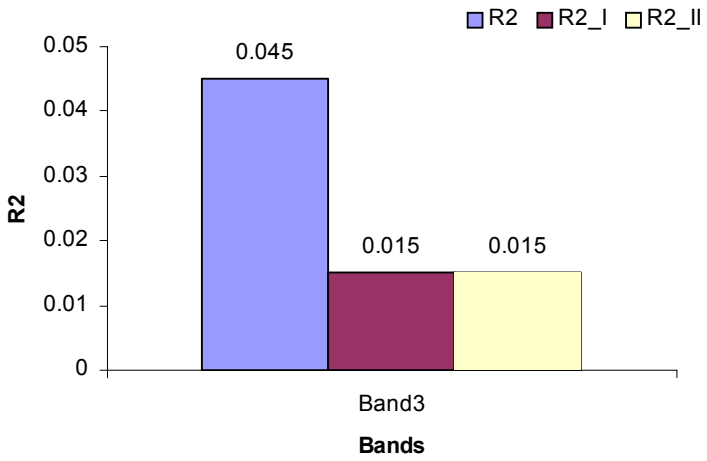


Figure 33: Changes in R^2 values with outlier removal at each iterations (I, II.. ..) for band 3 of 2001 field-1 data

Field-3

The outlier removal process for 2001 field-3 data is illustrated in figures 34 and 36. Bands 2 and 4, which had no outliers, were not analyzed. The normal probability plots corresponding to Figures 34 and 36 are available in Figures 35 and 37, respectively. Band 1, which required two iterations to remove all the detected outliers, indicated improvement in constant variance (Figure 34). Band 3 required one iteration to remove outliers, and similar to band 1, improvement in the constant variance plot was noted (Figure 34). The normal probability plots of bands 1 and 3 also have improved, as after removal of outliers the data points lying far off the reference line are no longer in evidence, and overall the data curve seems closer to the reference line (Figure 35). The outlier removal plots of bands 4 and 7 are given in Figure 36. The constant variance plots of band 4 suggest that after removal of outliers no major improvement occurred except that an outlier was no longer in evidence (Figure 36). This situation was also apparent in the normal probability plot of band 4 (Figure 37). However, in the case of band 7, both constant variance and normal probability plot appeared to improve (Figures 36 and 37).

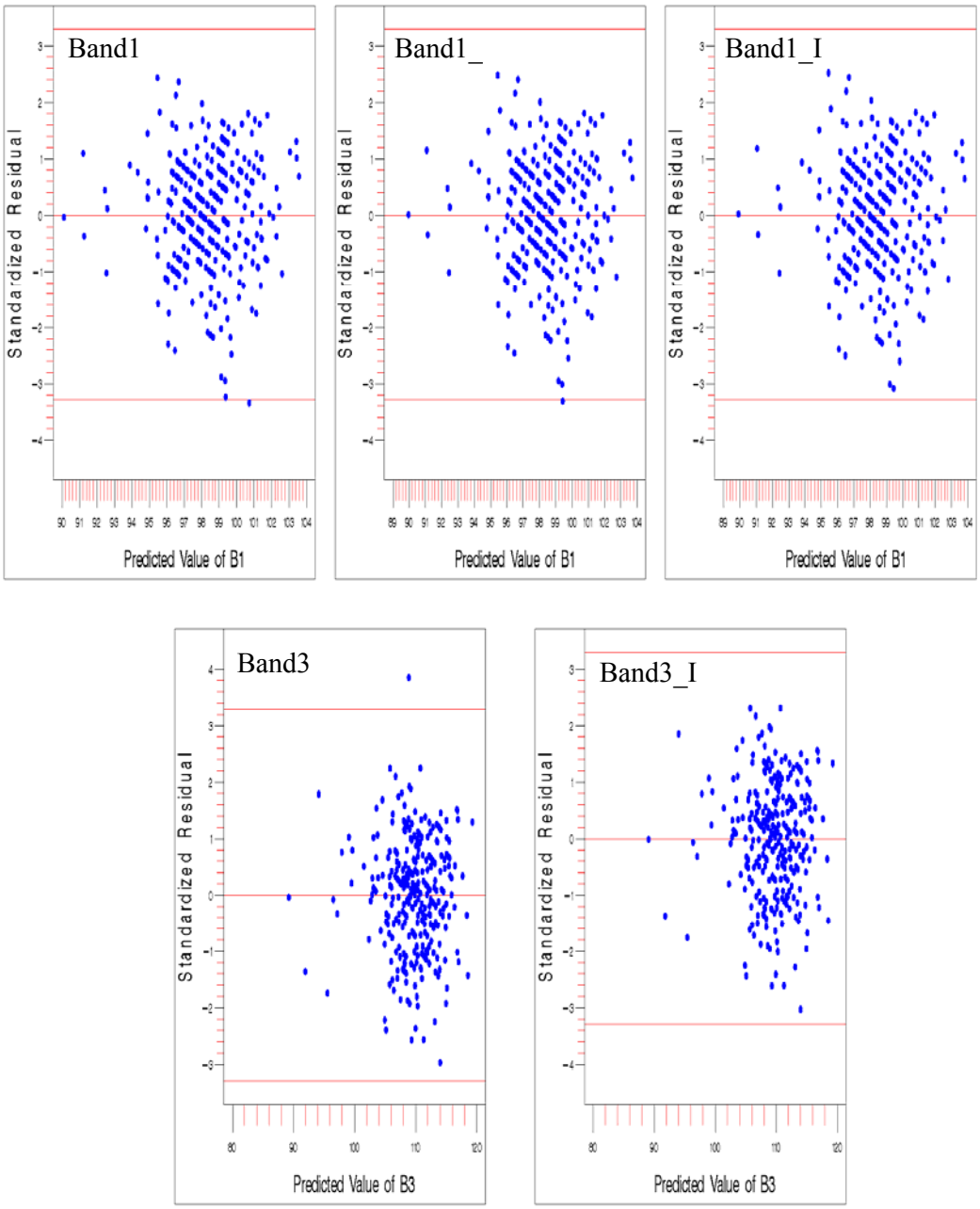


Figure 34: Outlier removal iterations (I, II …) plots of standardized residual vs. predicted value of dependent variable for bands 1 and 3 of 2001 field-3 data (SLR)

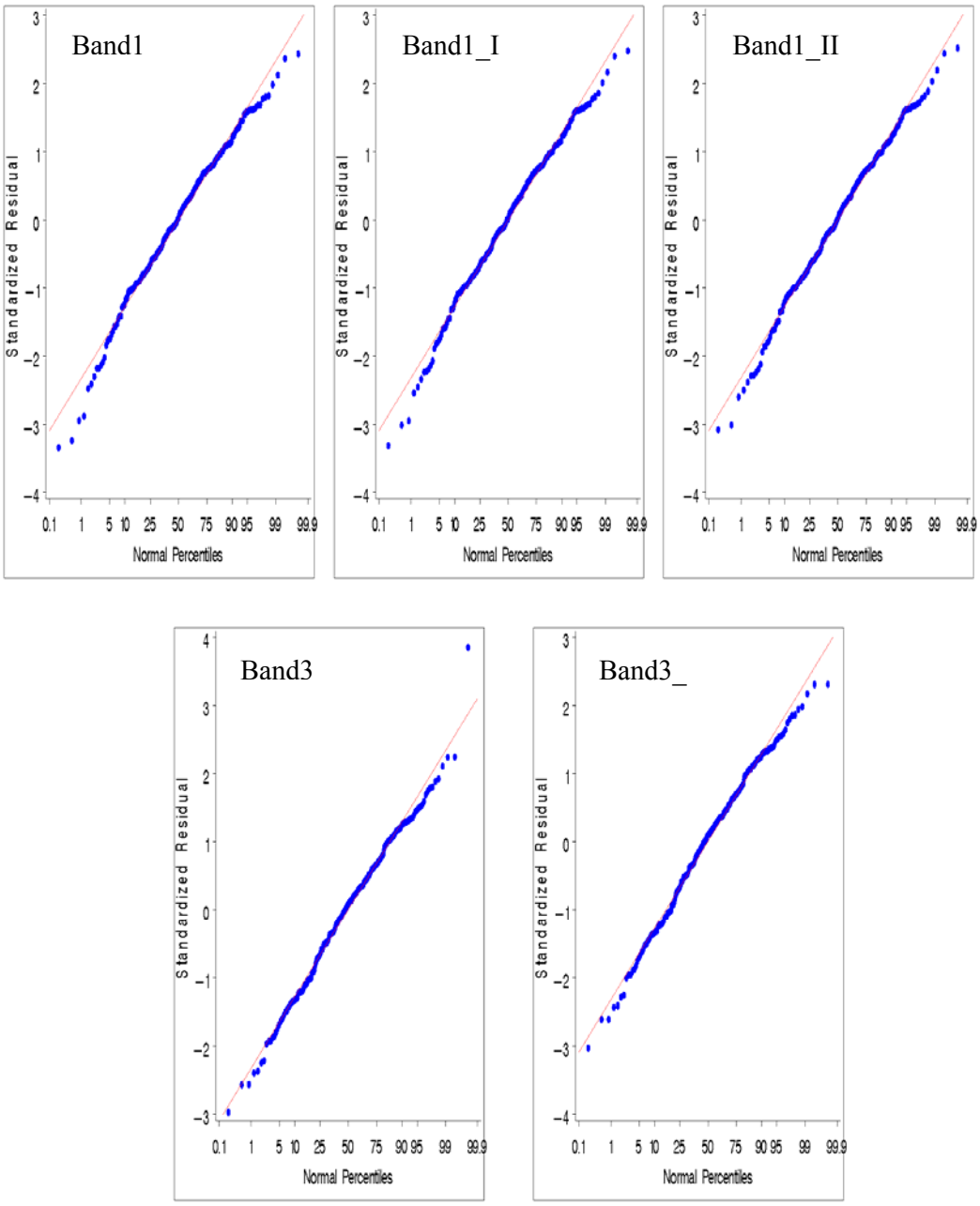


Figure 35: Normal probability plots corresponding to outlier removal iteration (I, II …) plots for bands 1 and 3 of 2001 field-3 data (SLR)

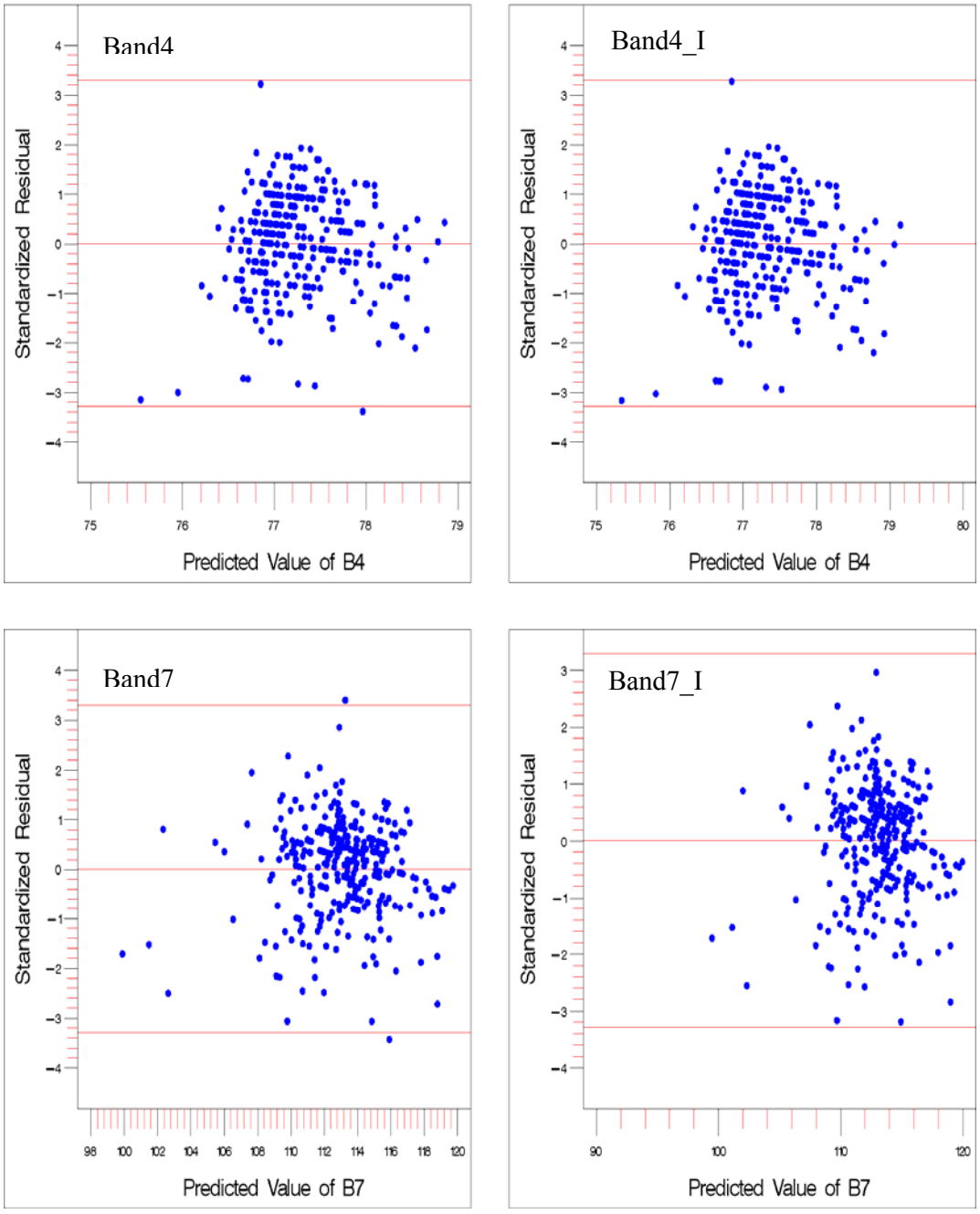


Figure 36: Outlier removal iterations (I, II …) plots of standardized residual vs. predicted value of dependent variable for bands 4 and 7 of 2001 field-3 data (SLR)

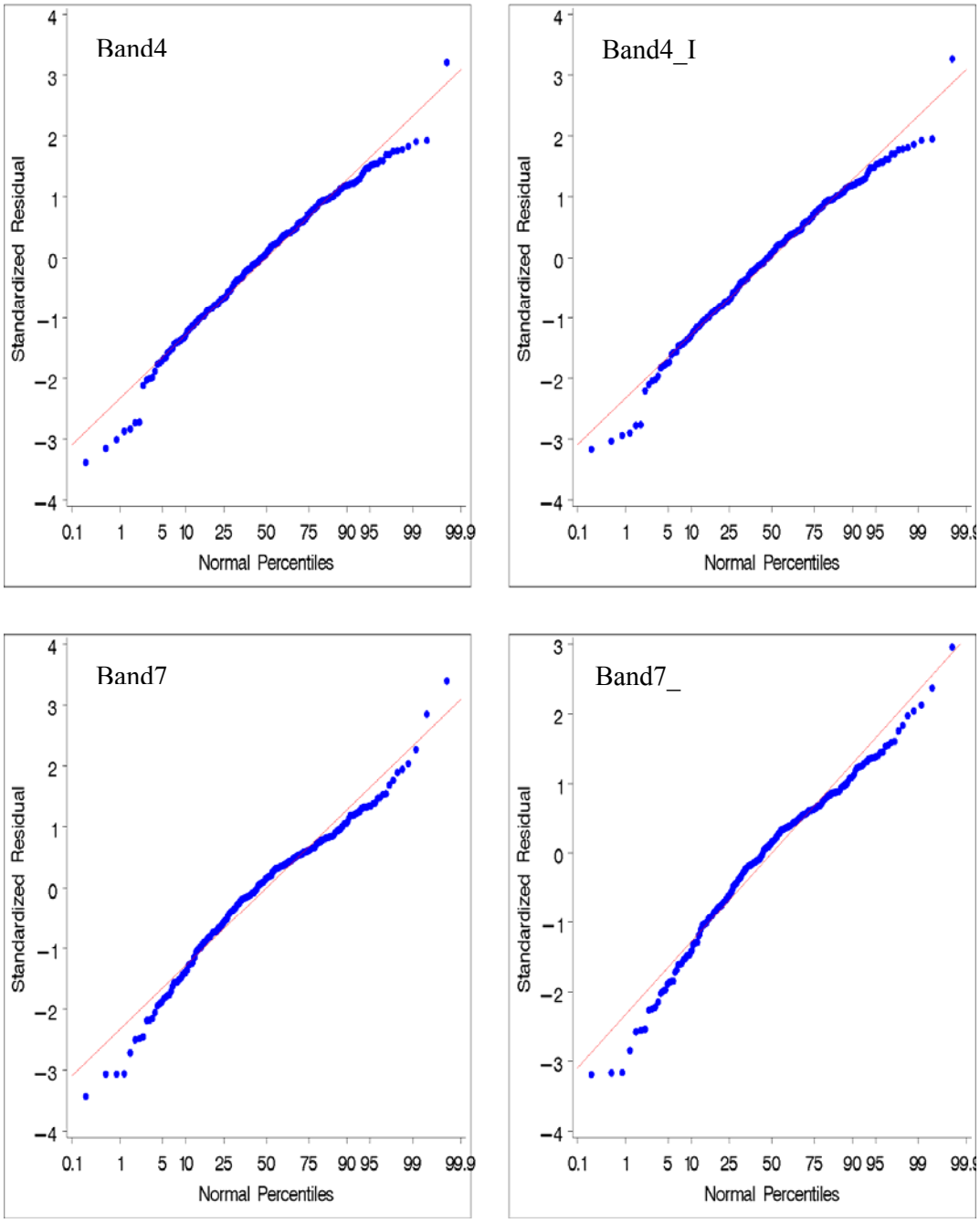


Figure 37: Normal probability plots corresponding to outlier removal iteration (I, II …) plots for bands 4 and 7 of 2001 field-3 data (SLR)

The R^2 value of band 1 gradually improved from 0.25 to 0.27 to 0.28 at each successive iteration of the outlier removal process (Figure 38). The R^2 value of band 3 also improved, from 0.28 to 0.29 (Figure 38). The R^2 value of band 4 remained the same (0.01), and for band 7, it improved from 0.09 to 0.11 (Figure 39). The change in p value for band 4 from 0.07 to 0.03 suggests that after removal of outliers the model changed from non-significant to significant, whereas for all the other bands (1, 3, and 7), the p values did not change and the significance levels remained high (<0.0001).

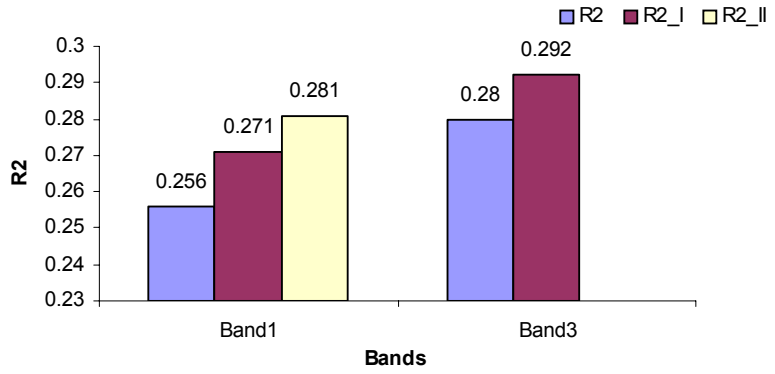


Figure 38: Changes in R^2 values with outlier removal at each iterations (I, II.. ..) for bands 1 and 3 of 2001 field-3 data

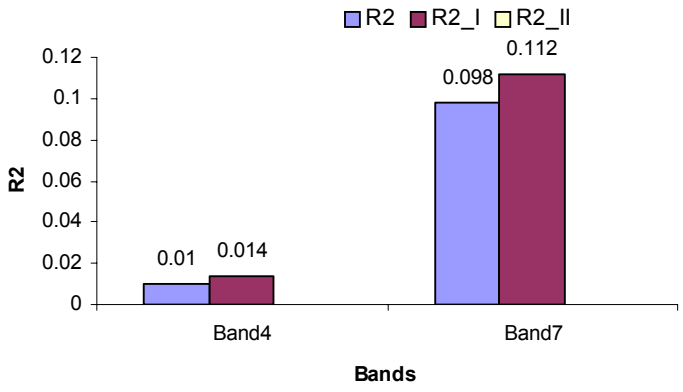


Figure 39: Changes in R^2 values with outlier removal at each iterations (I, II.. ..) for bands 4 and 7 of 2001 field-3 data

Validation

Field-1

The outliers of field-1 that were detected and removed are presented in Tables 7 (1997) and 8 (2001). The first column indicates the bands from which the outliers have been removed. The second column is observation number of the outlier data point in the original dataset. The third and fourth columns are the UTM coordinates of the outliers; the fifth column indicates which outlier removal iteration during which the specific data point was removed as an outlier. The final column has actual reflectance values at the location, including the DN value and the mean and standard deviation of reflectance for the respective band. It is important to check the validity of these data points as outliers before removing from the dataset. Table 7 makes it clear that observation 274 is detected as an outlier in all the bands except band 4. Similarly, points 11 and 32 are detected in all the bands except 1 and 7. Bands 4 and 5 have the highest number of outliers in the 1997 field-1 data. The actual reflectance values in the last three columns indicate that the DN value of all the outliers is at least 2 standard deviations (s.d.) lower than the average value of that band. According to the weather information available from the weather station at Stoneville, Mississippi, no precipitation was observed from April 10 to 17 (image collected on 17th April) of 1997, so it was unlikely, as a possible reason for the outliers, that the soil would have been very wet due to precipitation. On the other hand, field 1 is irrigated, and it is possible that the image was collected during or immediately after a significant irrigation event. Figure 40 indicates that point 274 is located at the corner of the field, which has unusual topography, and so the site may not have been pure bare soil

(Wooten, 2005; personal communication). Points 10 through 13 and 32 are located on a green patch in the false color image according to Wooten (2005; personal communication), and there may have been drainage and/or grass at the site at the time of image collection. Hence, there may have been water or higher moisture content. Furthermore, no healthy green vegetation was observed, as no red patches were visible in the false color image. However, the possibility of dry vegetation or plant residues was not ruled out.

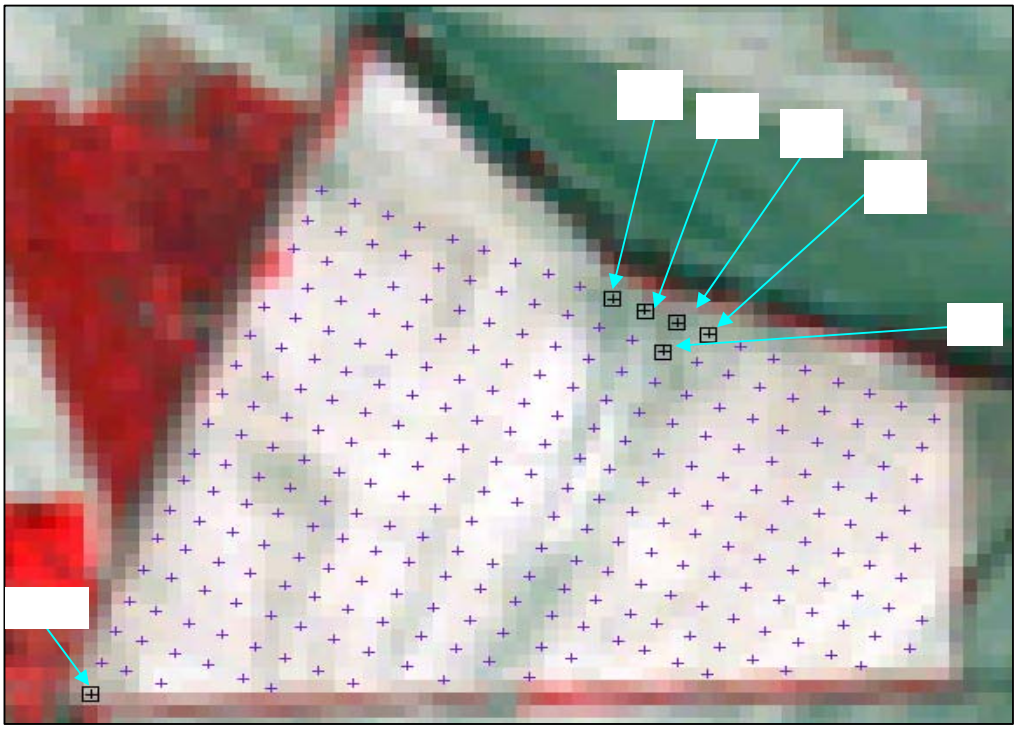
Band-3 of 2001 field-1 data eight data points that were identified as outliers (Table 8). Point 274 was again identified as one of the outliers, confirming the previous conclusion about it. The DN values suggest that all the outlier values were at least 2 s.d. lower than the average value of band-3, except for point 274. According to data from a weather station at Vance, Mississippi, between 18 and 22 May precipitation amounts of 0.10, 0.31, 0.01, 0.30, and 0.06 inches were observed, including on the image collection date of 22 May. The false color image reveals that, except for point 274, all the other outliers are located on a spot that appears black (Figure 41), which is an indication of standing water at the site. Hence, it can be confirmed that the detected outliers were in fact outliers.

Table 7: Detected outliers of all the studied bands of 1997 field-1 data and their corresponding actual radiance values.

Bands	Obs. No.	Easting	Northing	Outlier Iterations	Actual Radiance values		
					DN	Mean	Std. Dev
Band1	32	745886.477	3774943.212	I	100	119.45	5.86
	274	744858.6	3774263.226	I	100		
Band2	11	745853.4	3775025.156	II	52	65.54	3.93
	32	745886.477	3774943.212	I	53		
	274	744858.6	3774263.226	I	53		
Band3	11	745853.4	3775025.156	I	71	93.13	5.89
	32	745886.477	3774943.212	I	74		
	274	744858.6	3774263.226	I	73		
Band4	10	745796.438	3775049.2	I	77	103.16	5.54
	11	745853.4	3775025.156	I	80		
	12	745911.001	3775001.461	I	78		
	13	745968.42	3774977.651	II	87		
	32	745886.477	3774943.212	I	83		
Band5	10	745796.438	3775049.2	III	165	211.81	11.10
	11	745853.4	3775025.156	I	163		
	12	745911.001	3775001.461	II	164		
	13	745968.42	3774977.651	IV	183		
	32	745886.477	3774943.212	I	173		
	274	744858.6	3774263.226	I	158		
Band7	11	745853.4	3775025.156	I	94	131.93	9.34
	274	744858.6	3774263.226	I	92		

Table 8: Detected outliers for band 3 of 2001 field-3 data and their corresponding actual radiance values.

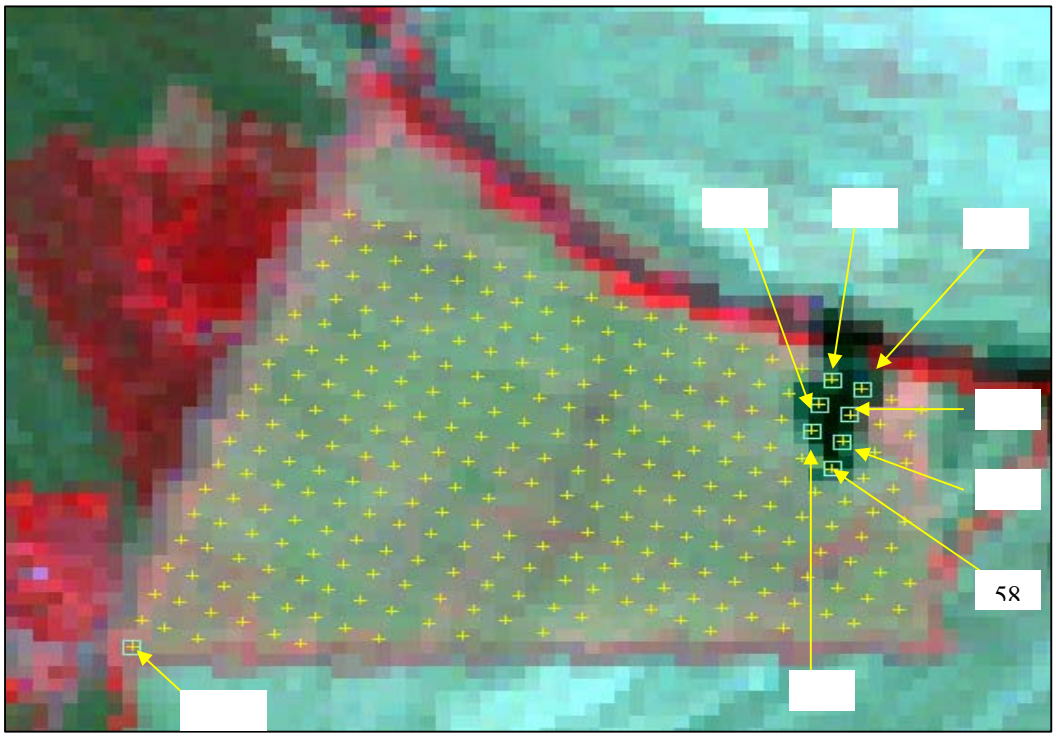
Bands	Obs. No.	Easting	Northing	Outlier Iterations	Actual Radiance values		
					DN	Mean	Std. Dev
Band3	17	746199.299	3774882.447	II	77	102.68	8.060
	18	746256.624	3774858.747	I	59		
	37	746175.402	3774824.991	I	60		
	38	746233.099	3774801.19	I	46		
	57	746161.302	3774763.683	I	55		
	58	746220.214	3774739.359	I	66		
	77	746200.159	3774676.231	II	81		
	274	744858.6	3774263.226	II	88		



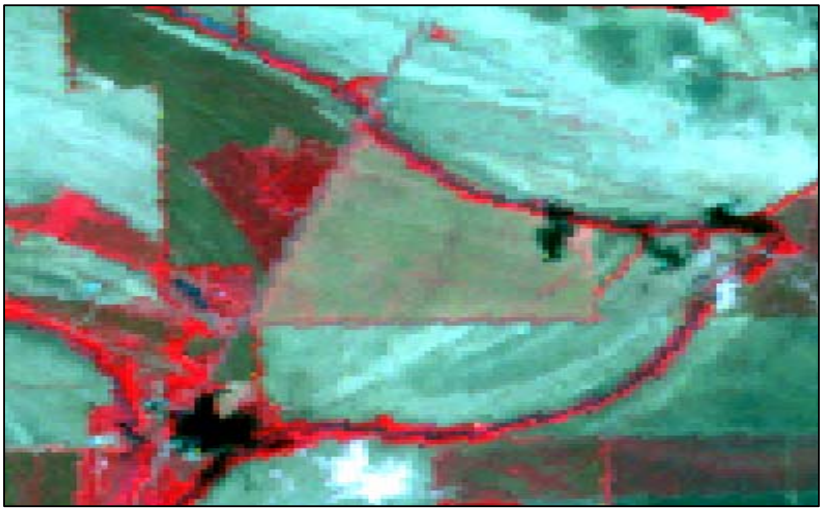
(a)

(b)

Figure 40: Position of outliers on the close-up (a) and distant (b) false color Landsat images of 1997 field-1 data (SLR)



(a)



(b)

Figure 41: Position of outliers on the close-up (a) and distant (b) false color Landsat images of 2001 field-1 data (SLR)

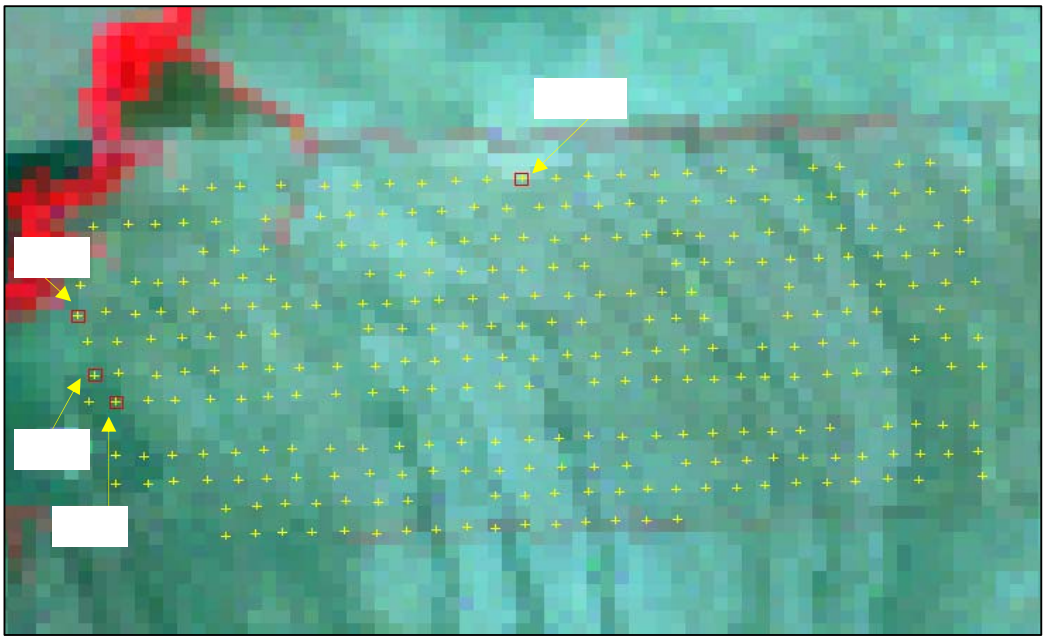
Field-3

Year 2001 data from field-3 included four outliers in bands 1, 3, 4, and 7 (Table 9). Point 152 was detected in bands 1 and 7, and point 282 was detected in bands 3 and 7. Point 282 has a DN value at least 2 s.d. greater than the average value for the respective band (3 or 7), whereas all the other outliers are at least 2 s.d. lower than the average for the respective band.

Table 9: Detected outliers for bands 1, 3, 4, and 7 of (2001) field-3 data and their corresponding actual radiance values.

Bands	Obs. No.	Easting	Northing	Outlier Iterations	Actual Radiance values		
					DN	Mean	Std. Dev
Band1	152	741454.316	3772824.257	I	89	98.40	4.078
	202	741419.572	3772954.314	II	88		
Band3	282	742318.416	3773253.088	I	135	109.68	7.982
Band4	123	741496.779	3772766.256	I	61	77.282	5.054
Band7	152	741454.316	3772824.257	I	87	113.06	8.89
	282	742318.416	3773253.088	I	142		

The precipitation levels associated with 2001 field-3 data were to the same as those for 2001 field-1 data, since the same image was used for both fields. Figure 42 reveals that the point 282 is located on a swampy section of the field and may possibly include some live grass (Wooten, 2005; personal communication). The other three observations are located at the western end of the field, where various topographical changes occur and may influence the data (Wooten, 2005; personal communication). According to Fewell planting company, the farm owner and manager, cotton was planted in this field on April 28, 2001, so small cotton plants may been contributing to the upwelling radiance during the time of image collection.



(a)

(b)

Figure 42: Position of outliers on the close-up (a) and distant (b) false color Landsat images of 2001 field-3 data (SLR)

Influence of Soil Moisture Content on Landsat Radiance Data

Multiple linear regression (MLR) analysis was carried out in an attempt to account for the influence of soil moisture content on soil radiance data collected with the Landsat satellite. Soil texture (clay content) and elevation, which are both related to soil moisture, were added to the simple linear regression model relating Landsat simulated reflectance to actual radiance of the soils. Equation 2 describes the linear relationship among the variables. All the MLR models were tested to determine whether they satisfy the statistical assumptions.

Field 1

Table 10 includes values of F and p, as well as R^2 and adjusted R^2 for all the studied bands.

Table 10: Statistical parameters of MLR analysis of field-1 dataset from 1997 and 2001

Model: B1 = β_0 + β_1Band1 + β_2Clay + β_3Elevation+ ε 1997 (Landsat 5)					**Model: B1 = β_0 + β_1Band1 + β_2Clay + β_3Elevation+ ε 2001 (Landsat 7)**			
Bands	**F**	**P**	**R^2**	**Adj. R^2**	**F**	**P**	**R^2**	**Adj. R^2**
Band1	34.79	<.0001	0.2788	0.2708	0.96	0.4119	0.0106	-0.0004
Band2	42.27	<.0001	0.3196	0.3120	0.70	0.5544	0.0077	-0.0033
Band3	46.93	<.0001	0.3427	0.3354	5.18	0.0017	0.0544	0.0439
Band4	41.25	<.0001	0.3143	0.3067	2.85	0.0379	0.0307	0.0199
Band5	52.28	<.0001	0.3674	0.3604	4.16	0.0066	0.0442	0.0336
Band7	49.59	<.0001	0.3552	0.3481	2.26	0.0815	0.0245	0.0137

It can be seen in Table 10 that all the MLR models from 1997 field-1 data have significant relationships, as did the SLR models for that set of data. Thus, the null hypothesis H0: $\beta_1 = \beta_2 = \beta_3$ is rejected; indicating that at least one of the independent

variables ($Band_y$, Clay, Elevation) has a significant linear relationship with the dependent variable (B_y). The t-tests and their p-values provided in table 11 indicate that, for 2001 field-1 data, elevation has a significant linear relationship with only bands 3 and 7 (p = 0.05 and 0.02, respectively), while clay and expected reflectance have a significant relationship with all the bands (p = <0.0001).

Table 11: t and p values of MLR analysis for 1997 field-1 data

Model: $B1 = \beta_0 + \beta_1 Band1 + \beta_2 Clay + \beta_3 Elevation + \varepsilon$ 1997 (Landsat 5)												
Variables	*Band1*		*Band2*		*Band3*		*Band4*		*Band5*		*Band7*	
	t	**P**	**t**	**P**	**t**	**P**	**t**	**P**	**t**	**P**	**t**	**P**
Expected reflectance	6.30	<.0001	6.80	<.0001	6.86	<.0001	4.51	<.0001	6.92	<.0001	5.64	<.0001
Clay	-8.37	<.0001	-9.05	<.0001	-9.6	<.0001	-9.53	<.0001	-9.38	<.0001	-8.79	<.0001
Elevation	-0.22	0.82	-1.50	0.13	-1.9	0.05	-0.06	0.95	1.08	0.28	2.31	0.02

Compared to the SLR model, the R^2 values were higher with the MLR model in 1997 field-1 data, being greater than 0.25 for all the studied bands, with band 5 having the highest R^2 value at 0.36 (Table 10). Furthermore, adjusted R^2 had similar values to those of R^2 (Table 10), indicating that the increase in R^2 is caused by the added variables themselves. However, as elevation had a non-significant relationship with all bands except 3 and 7, the increase in R^2 appears generally to be caused by the inclusion of clay. The comparison between SLR and MLR models for 1997 field-1 data (figure 43) indicates a substantial increase in R^2 with MLR for all studied bands.

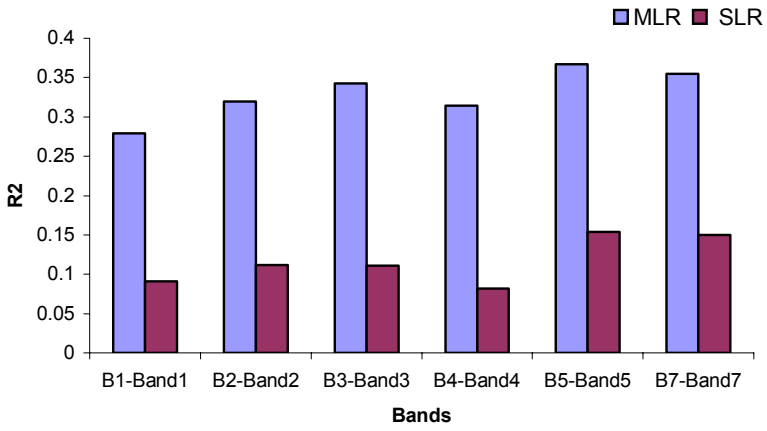


Figure 43: Comparison of R^2 values between MLR and SLR models for 1997 field-1 data.

Table 12: t and p-values of MLR analysis of dependent variables for 2001 field-1 data

Model: B1 = β_0 + β_1Band1+ β_2Clay + β_3Elevation+ ε, 2001 (Landsat 7)												
Variables	*Band1*		*Band2*		*Band3*		*Band4*		*Band5*		*Band7*	
	t	**P**	**t**	**P**	**t**	**P**	**t**	**P**	**t**	**P**	**t**	**P**
Expected reflectance	1.46	0.14	1.11	0.26	3.68	0.0003	1.31	0.19	0.18	0.86	0.01	0.98
Clay	0.38	0.70	0.83	0.40	1.12	0.26	0.33	0.74	1.84	0.06	1.55	0.12
Elevation	0.94	0.34	0.55	0.58	1.19	0.23	-2.42	0.01	-2.85	0.004	-1.92	0.056

MLR analysis of 2001 data from the same field indicates that models for bands 3, 4, and 5 had significant relationships (p = 0.001, 0.03, and 0.006 respectively). Thus, the H0: $\beta_1 = \beta_2 = \beta_3$ is rejected for these three bands, indicating that at least one of the independent variables (expected reflectance, clay, or elevation) had a significant linear relationship with actual radiance. The other three bands, for which the null hypothesis could not be rejected, indicate that there is no linear relationship between actual radiance and the group of dependent variables. The t tests and p values (Table 12) indicate that with band 3 expected reflectance had a significant linear relationship (p = 0.0003) with actual radiance, but neither clay (p = 0.26) nor elevation (p =0.23)

was significantly related. With bands 4 and 5, only elevation had a significant linear relationship ($p = 0.01$ and 0.004). Thus, for 2001 field-1 data the MLR results have similar results to those with SLR. The fact that clay and elevation were generally not significant suggests that the 2001 field-1 data may have been affected by other influences like the presence of vegetation, plant residues, or possibly atmospheric distortion. Therefore, these data were not studied further.

Field 3

Statistics from the MLR analysis of 1999 and 2001 field-3 data are provided in table 13. Results of the analysis show that all the models were significant, having p values of less than 0.0001, except for band 4 in the 1999 data ($p = 0.0037$). Thus, the null hypothesis, H0: $\beta_1 = \beta_2 = \beta_3$, is rejected, indicating that at least one of the independent variables had a significant linear relationship with actual radiance. The t tests for 1999 field-3 data (Table 14) indicated that bands 1 through 3 had significant linear relationships ($p < 0.0001$), but bands 4, 5, and 7 had no significant linear relationship ($p > 0.05$) with actual radiance. Furthermore, clay had a non-significant linear relationship ($p > 0.05$), while elevation did have a significant linear relationship ($p < 0.0009$).

Table 13: Statistical parameters of MLR analysis for 1999 and 2001 field-3 data.

Model: B1 = β_0 + β_1Band1+ β_2Clay + β_3Elevation+ ε 1999 (Landsat 5)					**Model: B1 = β_0 + β_1Band1+ β_2Clay + β_3Elevation+ ε 2001 (Landsat 7)**			
Bands	**F**	**P**	**R^2**	**Adj. R^2**	**F**	**P**	**R^2**	**Adj. R^2**
Band1	8.23	<.0001	0.0768	0.0675	64.65	<.0001	0.3950	0.3889
Band2	12.19	<.0001	0.1096	0.1006	63.73	<.0001	0.3916	0.3855
Band3	13.99	<.0001	0.1238	0.1149	68.23	<.0001	0.4080	0.4020
Band4	4.58	0.0037	0.0442	0.0346	32.34	<.0001	0.2462	0.2386
Band5	14.69	<.0001	0.1292	0.1204	68.37	<.0001	0.4085	0.4025
Band7	17.67	<.0001	0.1514	0.1429	52.62	<.0001	0.3470	0.3404

Table 14: t and p values of MLR analysis for 1999 field-3 data.

Model: B1 = β_0 + β_1Band1+ β_2Clay + β_3Elevation+ ε, 1999 (Landsat 5)												
Variables	*Band1*		*Band2*		*Band3*		*Band4*		*Band5*		*Band7*	
	t	**P**	**t**	**P**	**t**	**P**	**t**	**P**	**t**	**P**	**t**	**P**
Expected reflectance	3.62	<.0001	5.41	<.0001	5.20	<.0001	0.19	0.84	1.38	0.16	0.41	0.68
Clay	-0.04	0.96	0.06	0.95	0.15	0.88	-0.91	0.36	-1.67	0.096	-1.85	0.065
Elevation	4.59	<.0001	4.76	<.0001	5.62	<.0001	3.39	0.0008	6.08	<.0001	6.59	<.0001

Compared to SLR results, the MLR results with 1999 field-3 data had an additional band (band-1) with a significant linear relationship. Since clay was not significant individually, the improvement would appear to be primarily because of elevation. The models with 1999 field-3 data had R^2 values of greater than 0.10 for bands 2, 3, 5, and 7, with band 7 having the highest R^2 value at 0.15. Bands 1 and 4 had R^2 values less than 0.10 (Table 13). Interestingly, the correlations evident in the infrared bands appear to be due more to elevation than expected reflectance or clay, as neither of the latter was significant in the model. In the case of visible bands, correlations can be attributed mainly to a combination expected reflectance and elevation. Consideration of adjusted R^2, which has similar values to those of R^2, suggests that improvement in R^2 is real and related to the addition of elevation in the

model. Comparison of MLR models with SLR models for 1999 field-3 data indicates that improvement in R^2 values in the visible bands is due mainly to elevation.

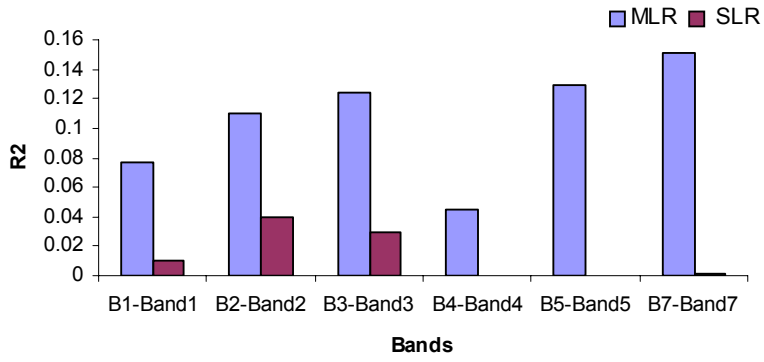


Figure 44: Comparison of R^2 values of MLR and SLR model for 1999 field-3 data

Since clay is considered to be more important than elevation as a parameter explaining moisture content, and since clay was not significant in the model, 1999 field-3 data were not studied for further analyses.

Results with 2001 field-3 data indicate that all the bands have significant MLR models ($p < 0.0001$) (Table 13), meaning that the null hypothesis (H0: $\beta_1 = \beta_2 = \beta_3$) can be rejected, and it can thus be concluded that at least one independent variable has a significant linear relationship with actual radiance. T tests and p values (Table 15) indicate that all bands except band 4 have a significant linear relationship with actual radiance, which is a similar conclusion to that of the SLR analyses. The two added variables, clay (p-value = <0.0001) and elevation (p-value < 0.05), are significant in the MLR models for all the studied bands. The highest R^2 value was 0.40, observed with bands 3 and 5, while the lowest was 0.24, observed with band 4. All the other

bands had R^2 values greater than 0.30 (Table 13). The adjusted R^2 values have similar results to those of R^2, suggesting that the increase in R^2, particularly evident with band 4, is because of the addition of clay and elevation into the model.

Table 15: t and p-values of MLR analysis of dependent variables of field-3, 2001

Model: B1 = β_0 + β_1Band1+ β_2Clay + β_3Elevation+ ε, 2001 (Landsat 7)												
Variables	*Band1*		*Band2*		*Band3*		*Band4*		*Band5*		*Band7*	
	t	**P**	**t**	**P**	**t**	**P**	**t**	**P**	**t**	**P**	**t**	**P**
Expected reflectance	9.39	<.0001	8.9	<.0001	10.0	<.0001	1.38	0.1691	7.7	<.0001	5.4	<.0001
Clay	-8.12	<.0001	-8.1	<.0001	-7.9	<.0001	-8.99	<.0001	-11.7	<.0001	10.6	<.0001
Elevation	-2.68	0.0078	-2.6	0.0078	-2.2	0.023	-4.86	<.0001	-2.4	0.015	-2.2	0.022

Compared to SLR results, the R^2 values in the MLR models generally improved significantly with the addition of clay and elevation (Figure 45).

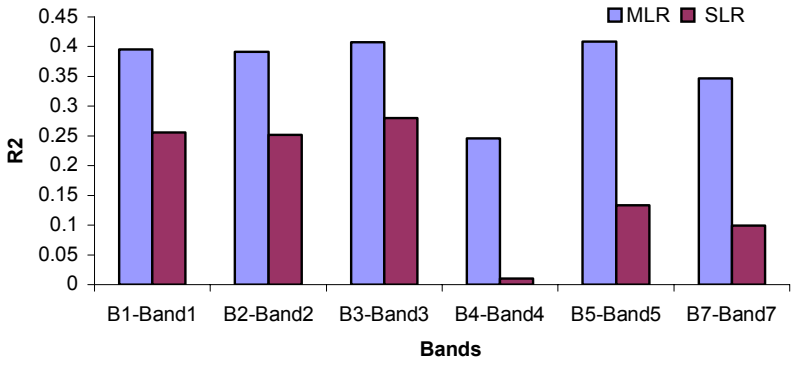


Figure 45: Comparison of R^2 values of MLR and SLR model for 2001 field-3 data

Assumptions

Normality

The normal probability plots of MLR models for 1997 field-1 data (Figure 46) show that the standardized residuals from bands 1 through 3 are aligned with the reference line, indicating that the residuals are normally distributed. A few data points at the lower tail are far from the reference line, indicating possible outliers in the dataset. Residuals in bands 4 through 7 are also aligned with the reference line (Figure 46). Band 7 data form a curve shape similar to that with the SLR model; however, unlike with SLR, the upper and lower tails are closer to the reference line. The curve shape is not visible with bands 4 and 5, as was the case with the SLR model. However, at the lower tails a few data points lie far from the reference line, again indicating possible outliers. Overall, compared to SLR, improvement in normality is observed in all the bands with the MLR model. Similar analysis for 2001 field-3 data indicates that the residuals of all the studied bands are well aligned with the reference line (Figure 47). The few data points that lie far from the reference line in bands 2 through 7 suggest possible outliers. Compared to SLR, normality appears to be better with MLR.

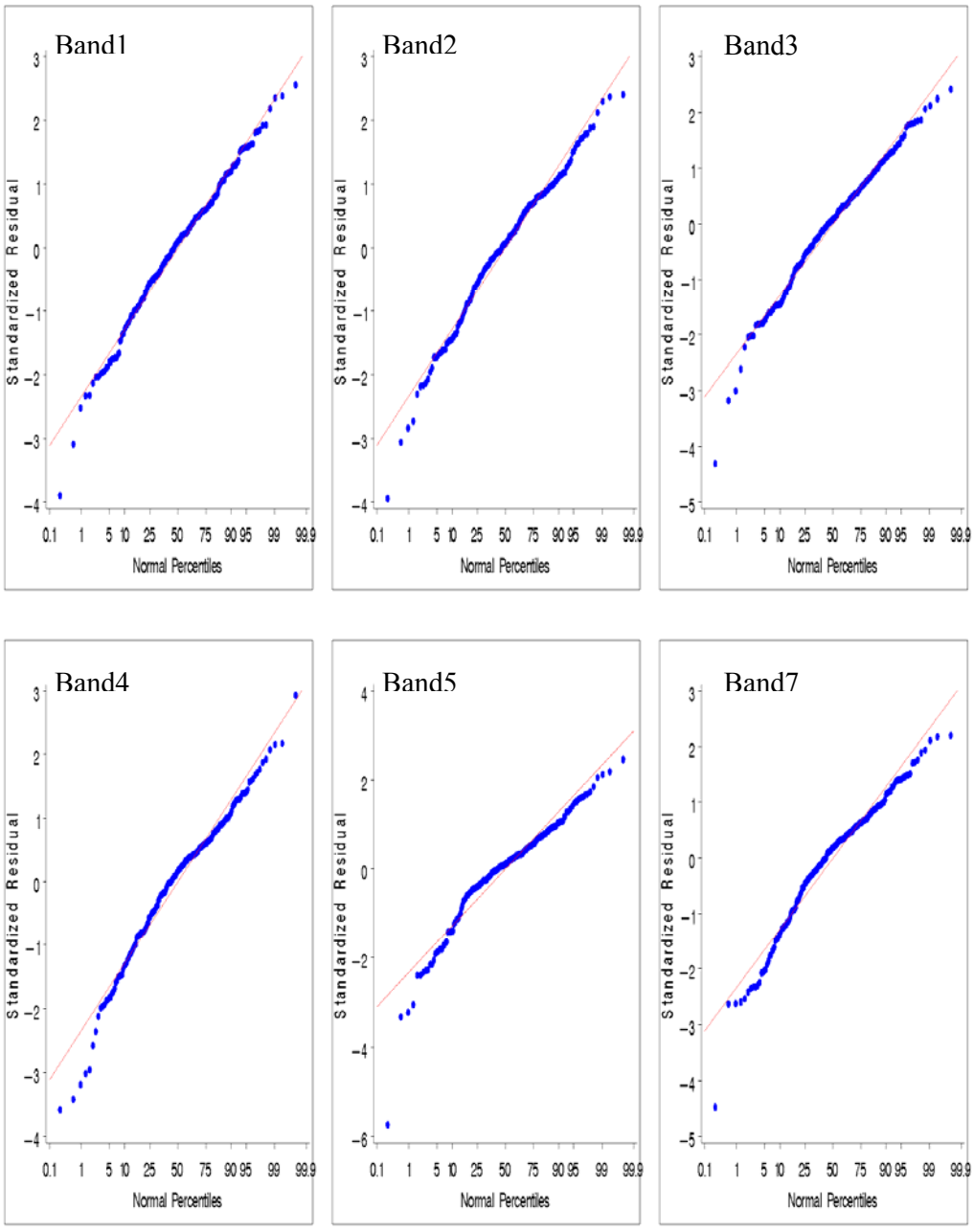


Figure 46: Normal probability plots of bands 1 through 7 of 1997 field-1 data (MLR)

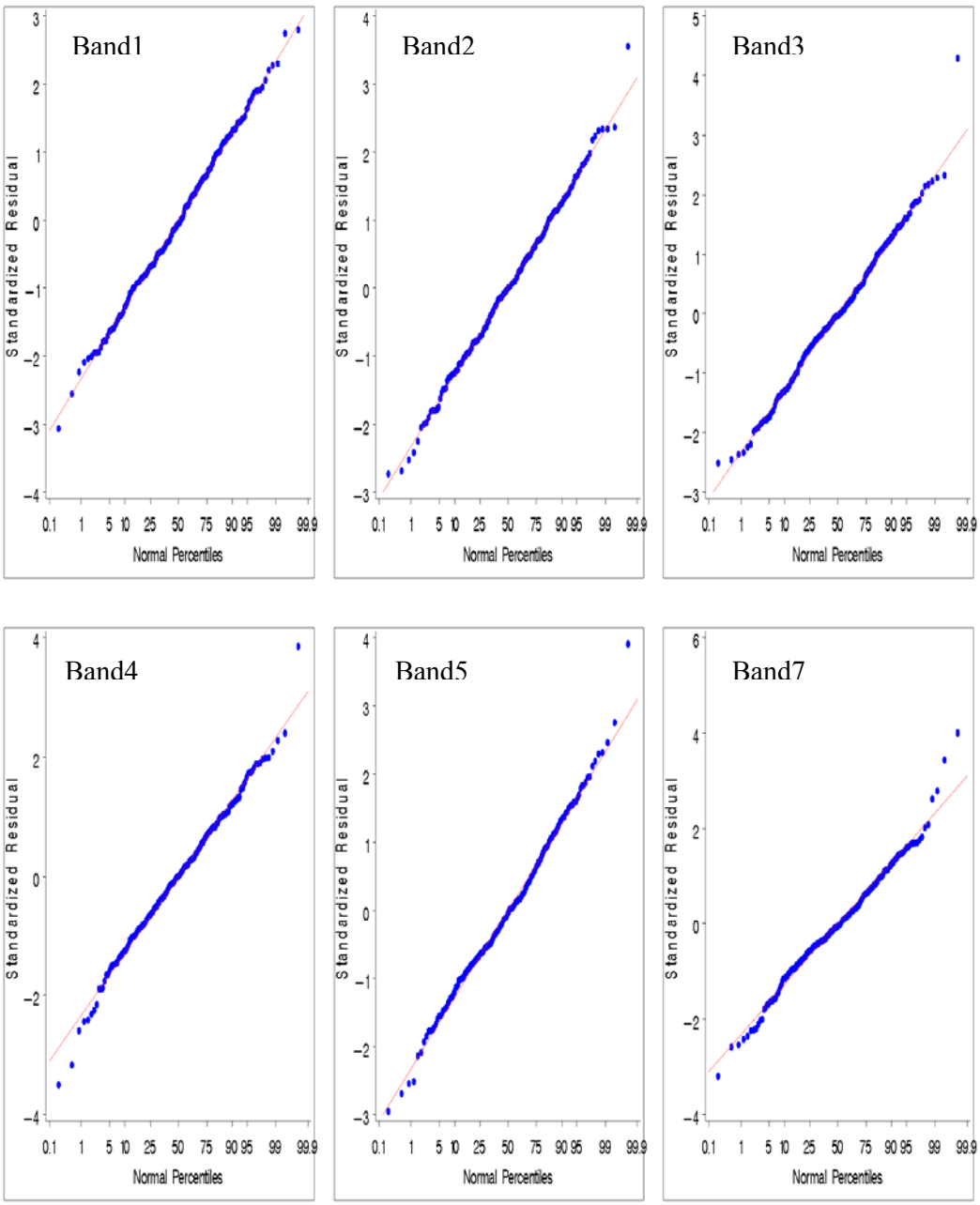


Figure 47: Normal probability plots of bands 1 through 7 of 2001 field-3 data (MLR)

Constant Variance

Plots of standardized residual vs. predicted value of the dependent variables for 1997 field-1 and 2001 field-3 data are provided in figures 48 and 49 respectively. The residuals in all the studied bands of both fields appear to be well distributed above and below the reference line (zero residual), suggesting that the residuals have constant variance. However, a few data points are far from the rest of the data points, indicating the likely presence of outliers. Overall, the conclusion with respect to constant variance is similar for both SLR and MLR, but careful comparison of the plots indicates an improvement with MLR that is most prominently visible in the band-4 data from both fields.

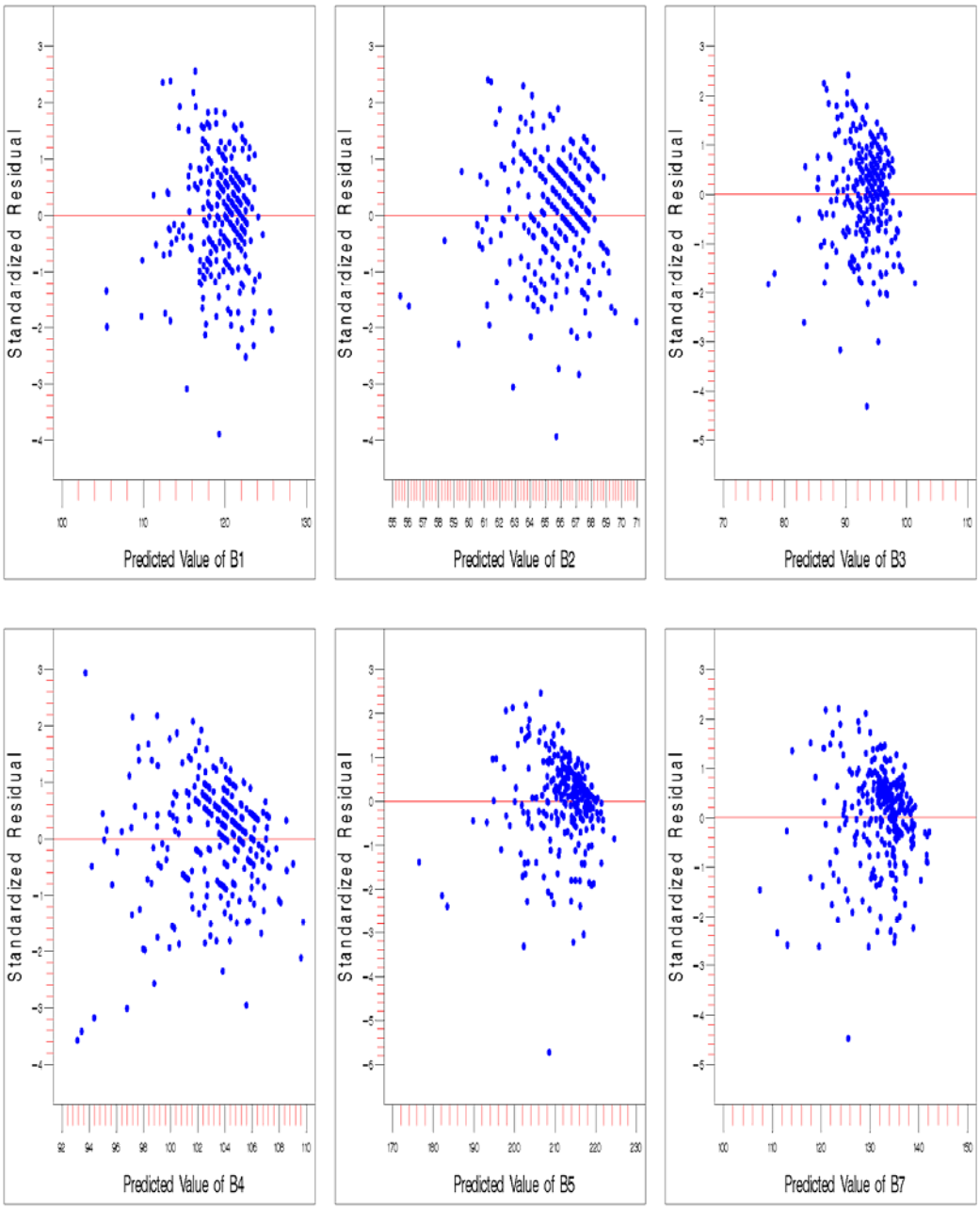


Figure 48: Plots of standardized residual vs. predicted value of dependent variable for bands 1 through 7 of 1997 field-1 data (MLR)

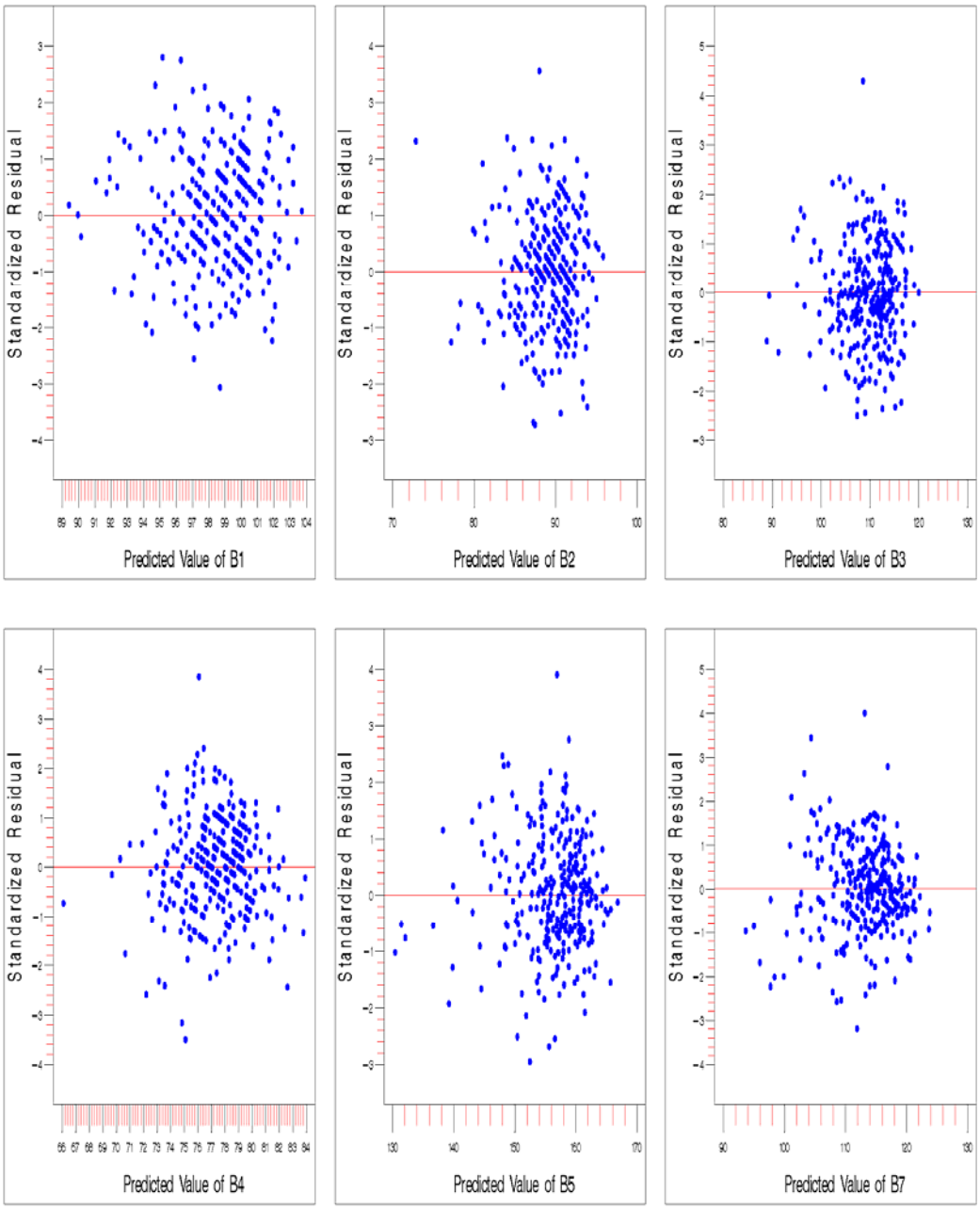


Figure 49: Plots of standardized residual vs. predicted value of dependent variable for bands 1 through 7 of 2001 field-3 data (MLR)

Outlier Analysis

Detection

The outlier boundary values of MLR analysis are similar to those of SLR analysis (3.29 and –3.29). Figures 50 and 51 are plots of standardized residual vs. predicted value of the dependent variable for 1997 field-1 data and 2001 field-3 data, respectively. It can be seen that all bands include outliers except band 1 in the 2001 field-3 data. The detected outliers were removed with the Reweight option, and the model was refitted again with the Refit option of the PROC REG procedure in SAS. New outliers were detected in some of the bands, and the removal procedure was repeated until all the detected outliers were removed.

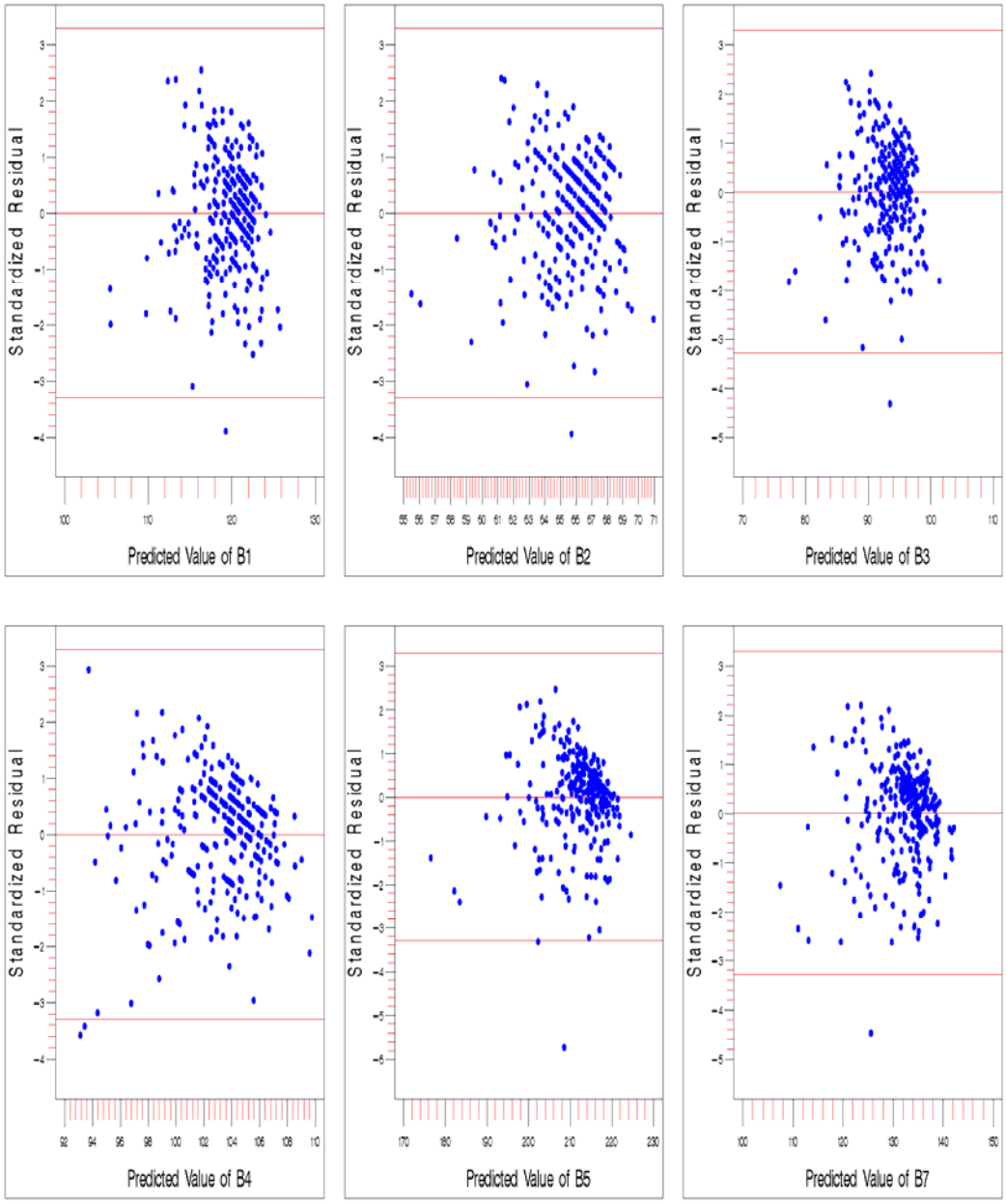


Figure 50: Outlier detection plots of standardized residual vs. predicted value of dependent variable for bands 1 through 7 of 1997 field-1 data (MLR)

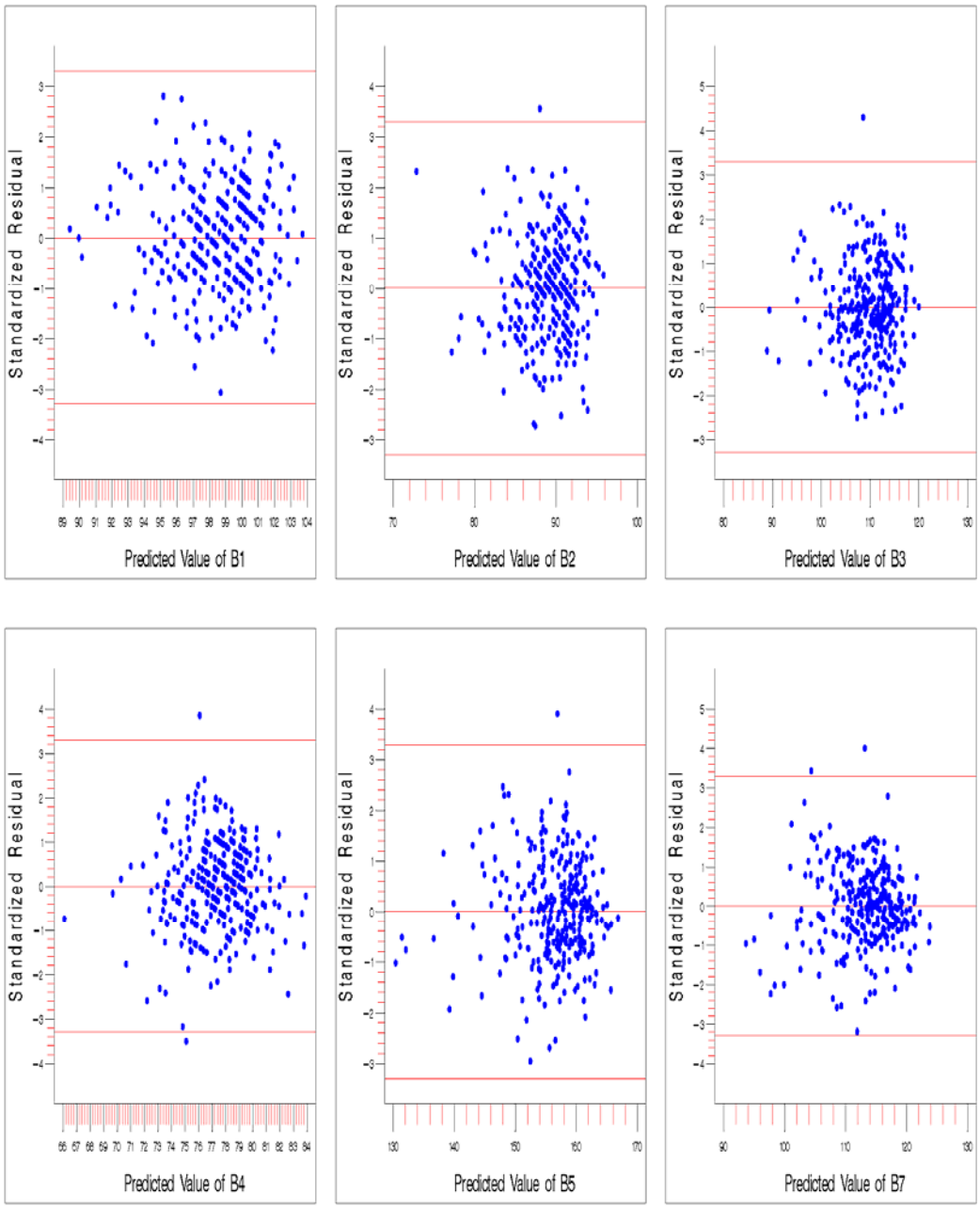


Figure 51: Outlier detection plots of standardized residual vs. predicted value of dependent variable for bands 1 through 7 of 2001 field-3 data (MLR)

Removal

Field-1

Figures 52, 54, and 56 illustrate the outlier removal procedure for 1997 field-1 data. The normal probability plots of the corresponding bands and iterations of 1997 field-1 data are provided in figures 53, 55, and 57, respectively. These plots indicate the changes in normality due to removal of outliers. The results of outlier removal for MLR analysis are similar to those for SLR analysis. Figure 52 indicates that bands 1 and 2 needed only a single iteration to remove all the detected outliers. The corresponding normality plots (Figure 53) indicate improvement also in normality after removal of outliers, as the data curve appears to be better aligned to the reference line, and data points lying far off the line are no longer present. The constant variance plots of bands 3 and 4, which needed two iterations to remove all the detected outliers, indicate similar behavior to that of bands 1 and 2 (Figure 54). The corresponding normality plots of band 3 indicate that with each iteration of outlier removal, an improvement in the normal probability curve was noted. Similar results for normality were observed for band 4 (Figure 55). Band-5 and -7 constant variance plots and normality plots (Figures 56 and 57) also indicate improvement after removal of outliers.

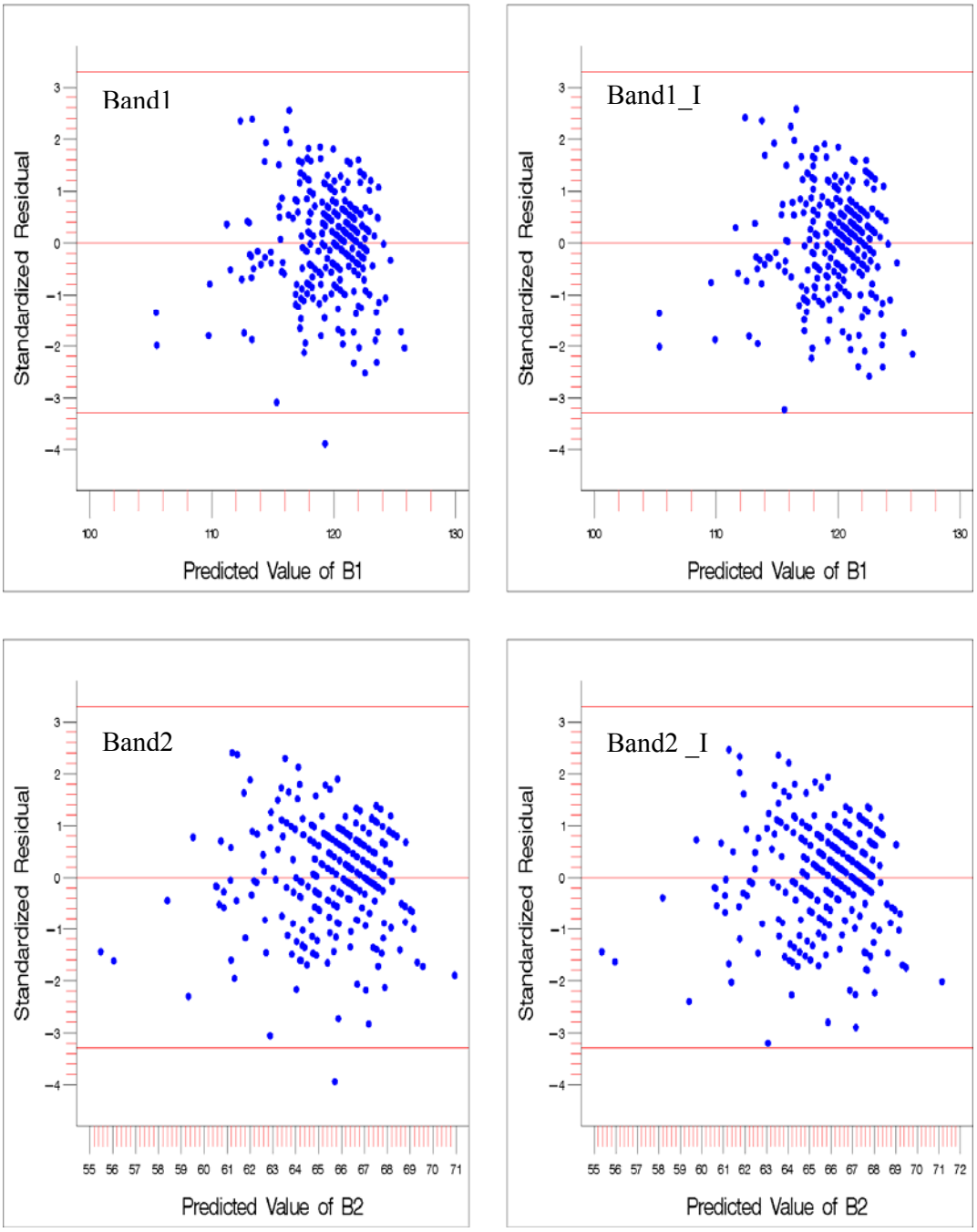


Figure 52: Outlier removal iterations (I, II …) plots of standardized residual vs. predicted value of dependent variable for bands 1 and 2 of 1997 field-1 data (MLR)

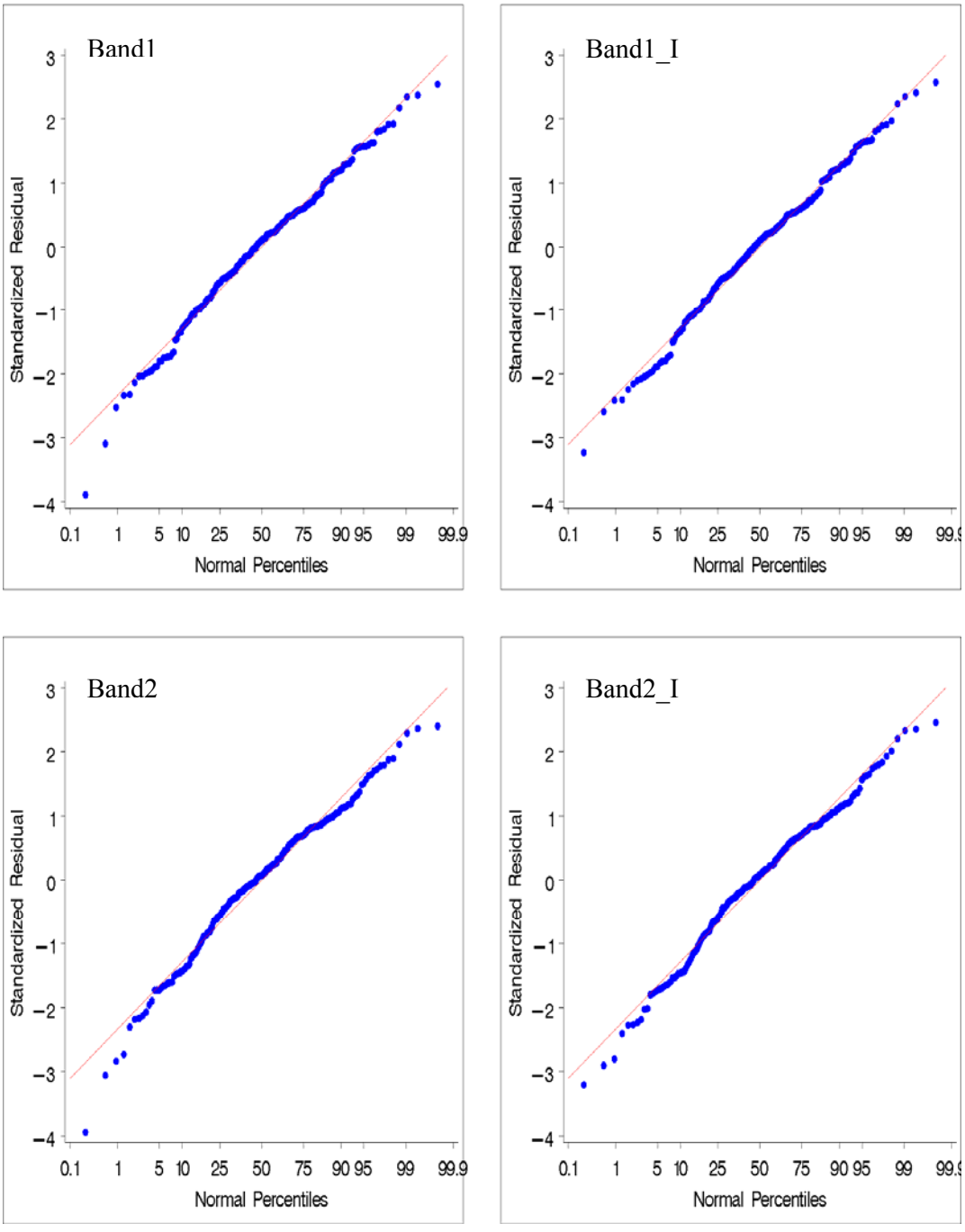


Figure 53: Normal probability plots corresponding to outlier removal iterations (I, II …) plots for bands 1 and 2 of 1997 field-1 data (MLR)

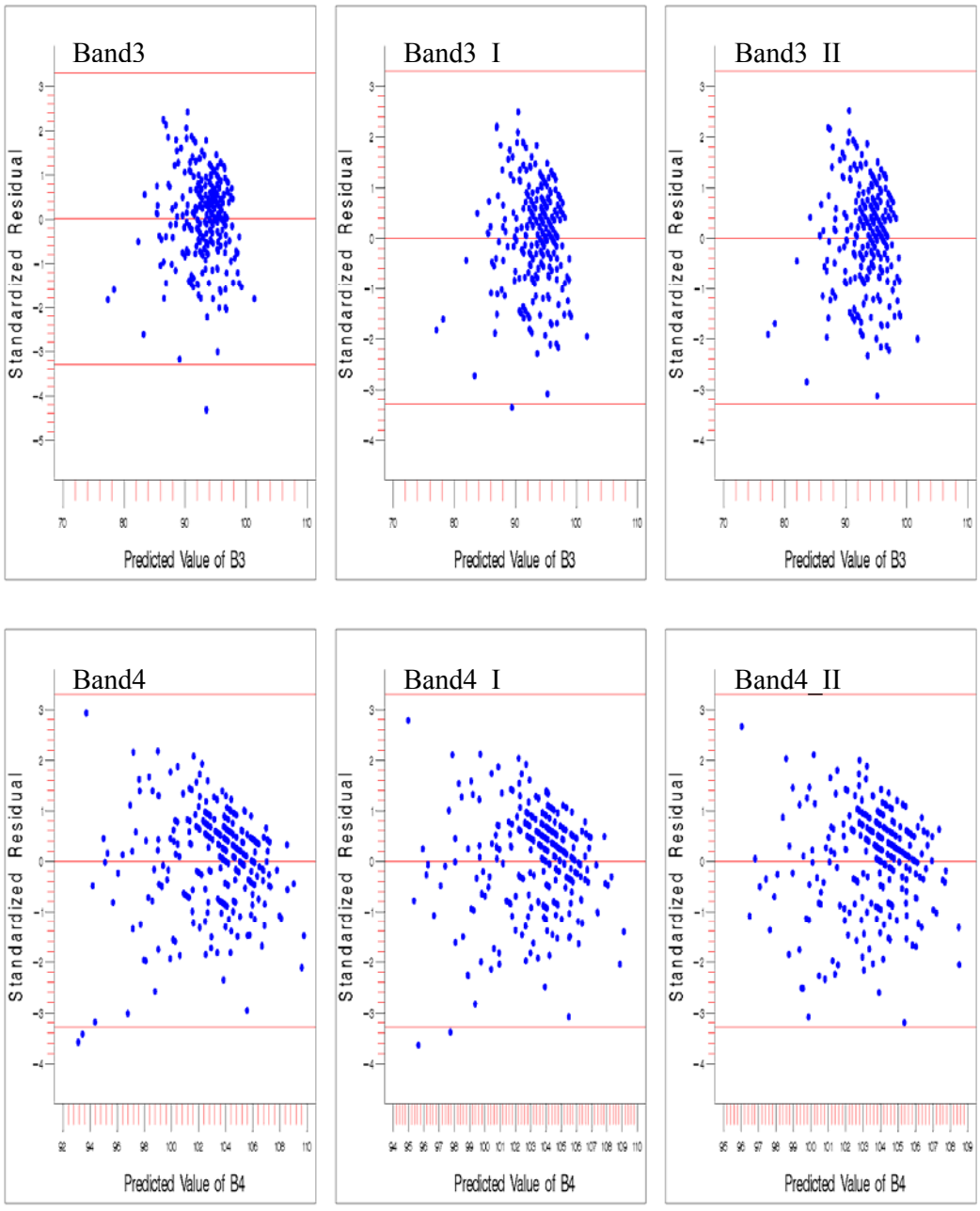


Figure 54: Outlier removal iterations (I, II …) plots of standardized residual vs. predicted value of dependent variable for bands 3 and 4 of 1997 field-1 data (MLR)

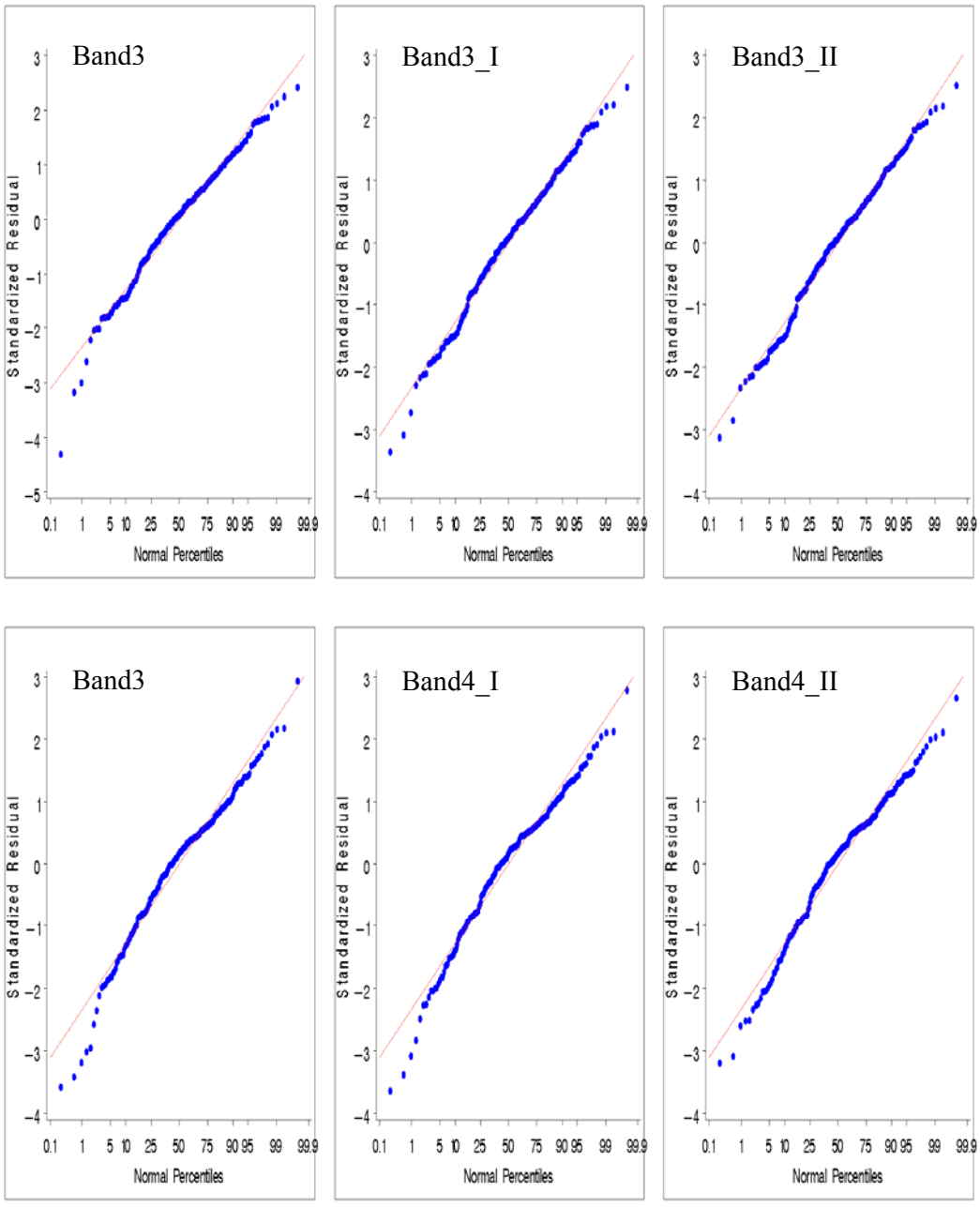


Figure 55: Normal probability plots corresponding to outlier removal iterations (I, II …) plots for bands 3 and 4 of 1997 field-1 data (MLR)

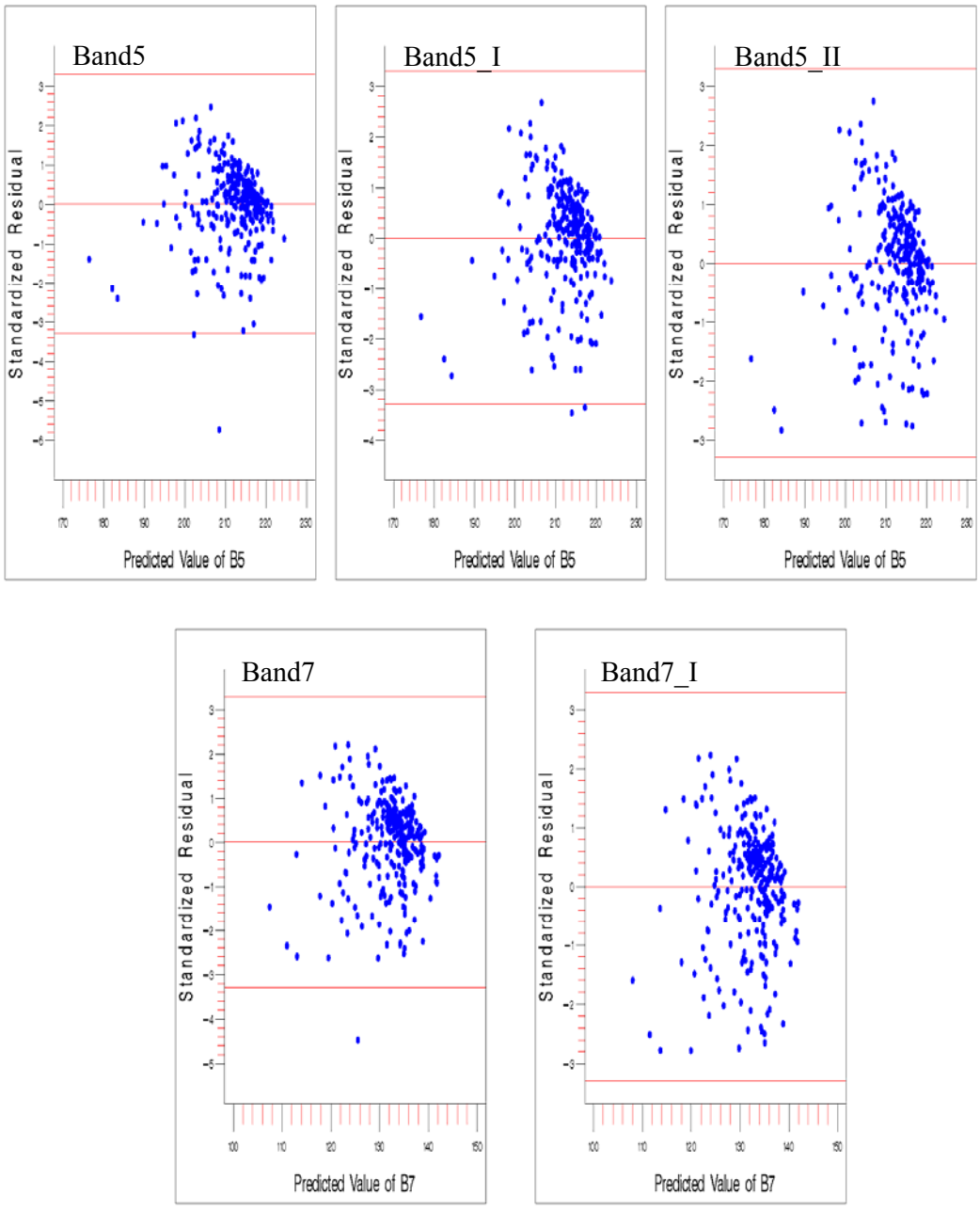


Figure 56: Outlier removal iterations (I, II …) plots of standardized residual vs. predicted value of dependent variable for bands 5 and 7 of 1997 field-1 data (MLR)

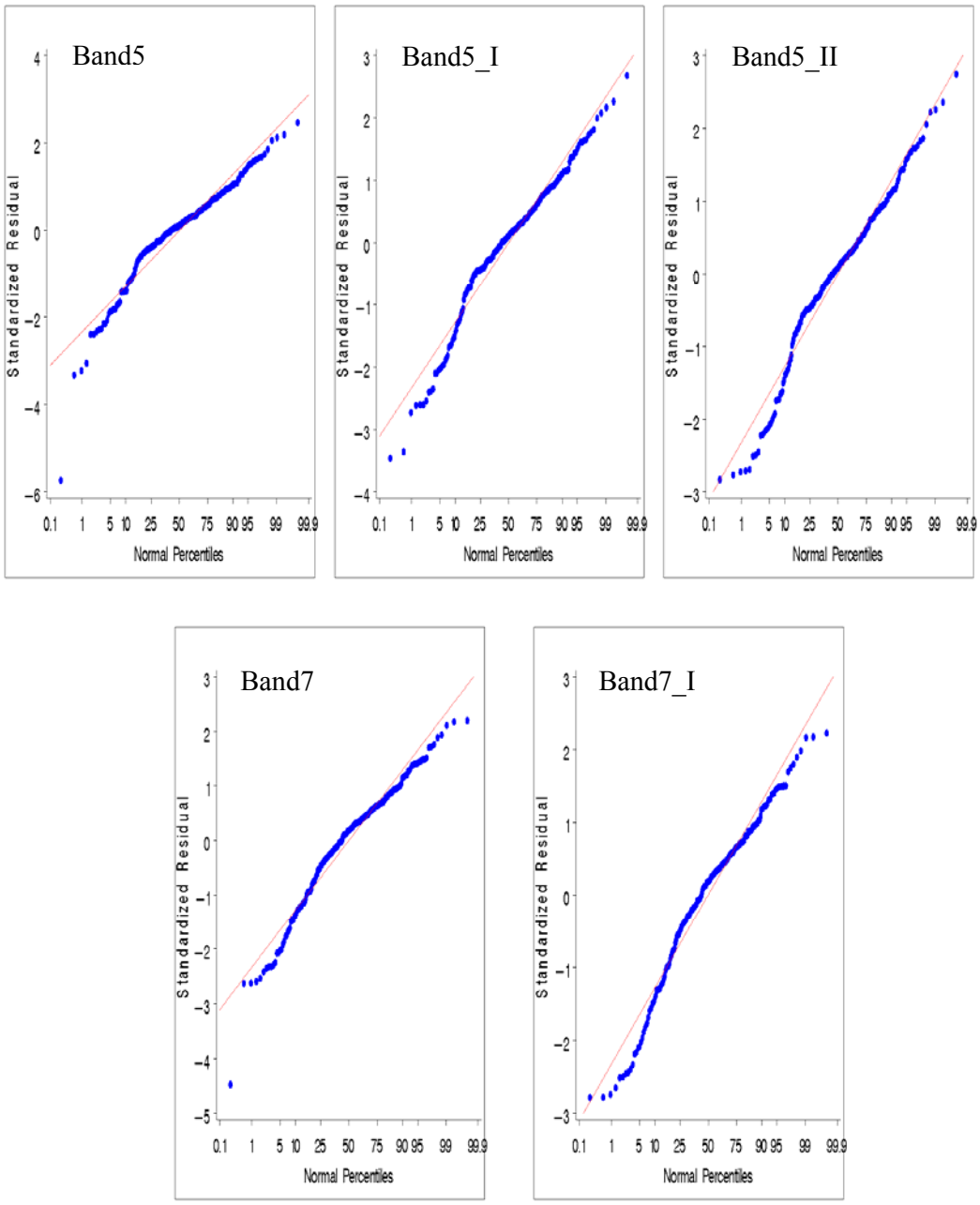


Figure 57: Normal probability plots corresponding to outlier removal iterations (I, II …) plots for bands 5 and 7 of 1997 field-1 data (MLR)

The figures 58, 59, and 60 illustrate the changes in R^2 with the removal of outliers. It can be observed that 1 to 2% improvement in R^2 occurred with bands 1 and 2 (figure 58). Band 3 had a 2% improvement in R^2 after the first iteration and remained same afterwards, whereas band 4 had a gradual decrease in R^2 with each successive iteration (figure 59). In bands 5 and 7, approximately 6% and 1% improvement in R^2 was observed after two and one iterations, respectively. The p values were not influenced by removal of outliers and remained the same (<0.0001) for all the studied bands.

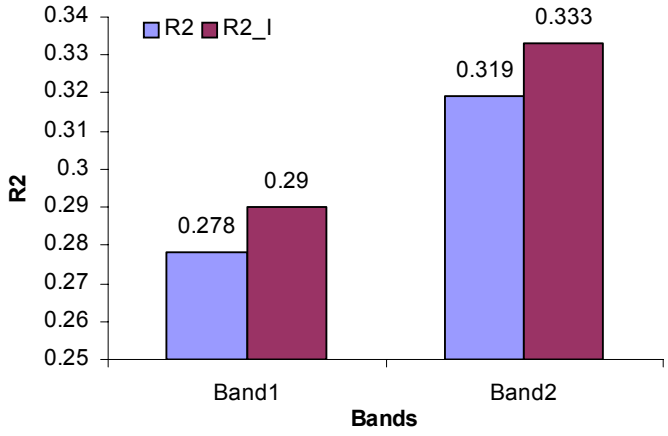


Figure 58: Changes in R^2 values with outlier removal at each iterations (I, II.. ..) for bands 1 and 2 of 1997 field-1 data.

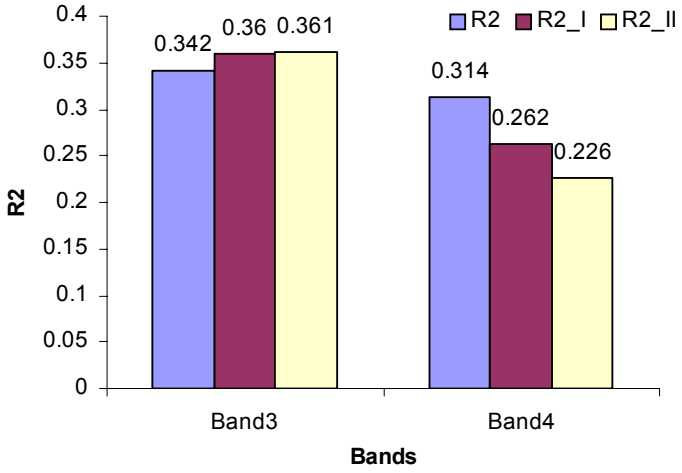


Figure 59: Changes in R^2 values with outlier removal at each iterations (I, II.. ..) for bands 3 and 4 of 1997 field-1 data.

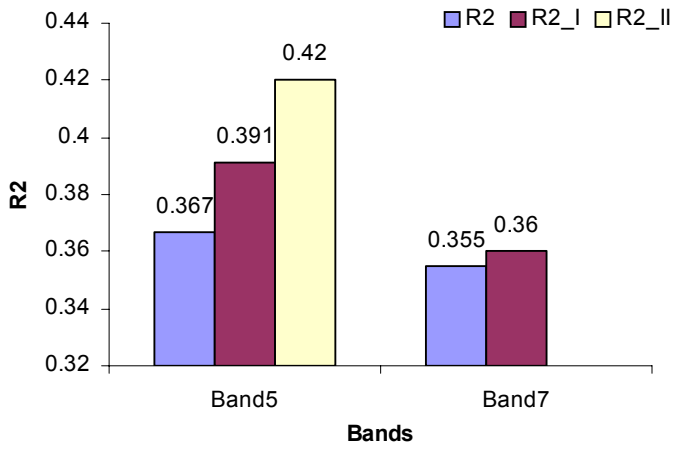


Figure 60: Changes in R^2 values with outlier removal at each iterations (I, II.. ..) for bands 5 and 7 of 1997 field-1 data.

Field-3

The outlier removal process for 2001 field-3 data is illustrated in Figures 61 and 63. The corresponding normality plots are presented in Figures 62 and 64. Similar to previous results, bands 2, 3, and 4 (figure 61), and bands 5 and 7 (figure 63) all had improvement in the constant variance plots. Furthermore, their corresponding normal probability plots also had improvement in normality, as data points lying far off the reference line were no longer observed after removal of outliers, and overall the data curve seems to be better aligned to the reference line.

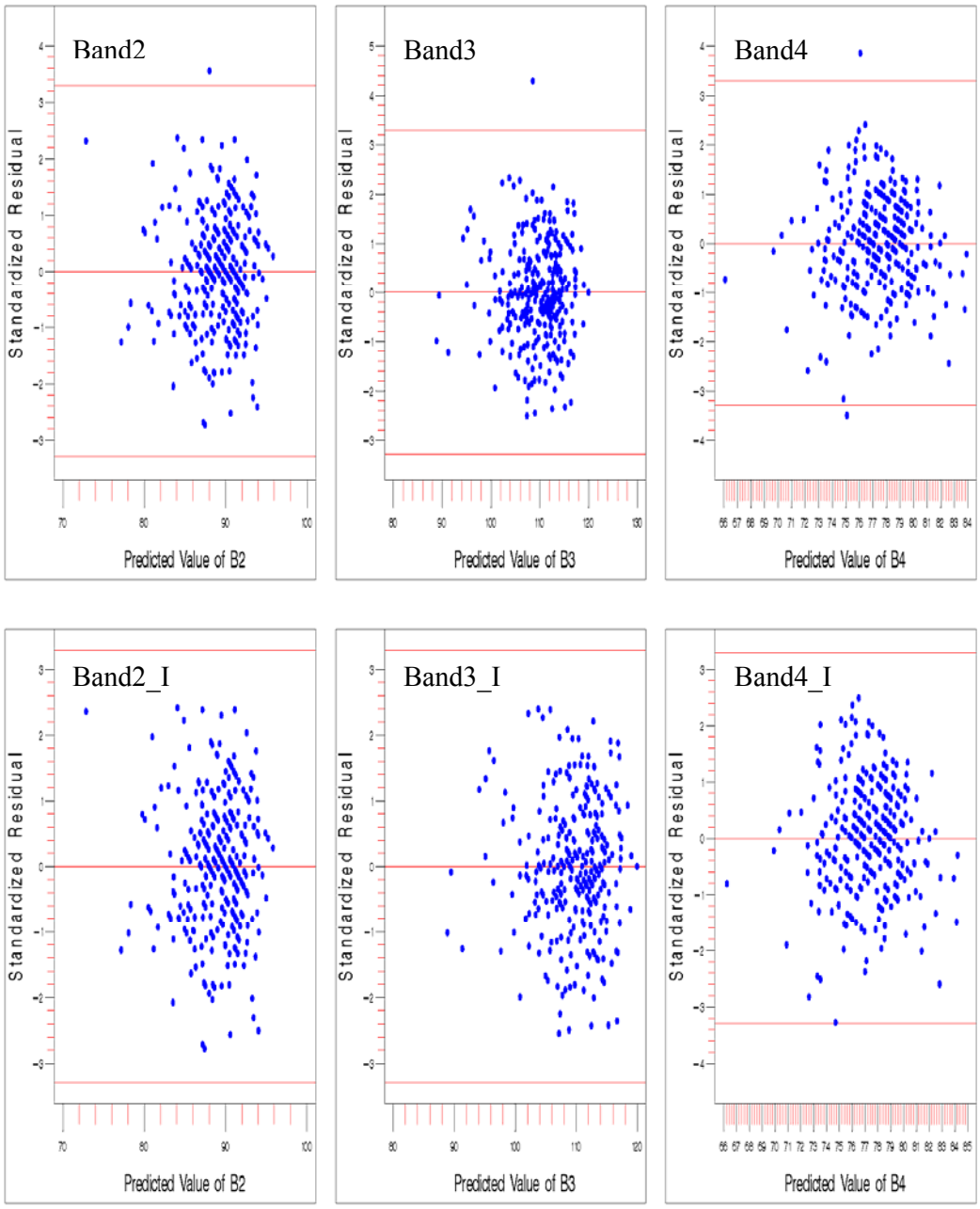


Figure 61: Outlier removal iterations (I, II …) plots of standardized residual vs. predicted value of dependent variable for bands 2 through 4 of 2001 field-3 data (MLR)

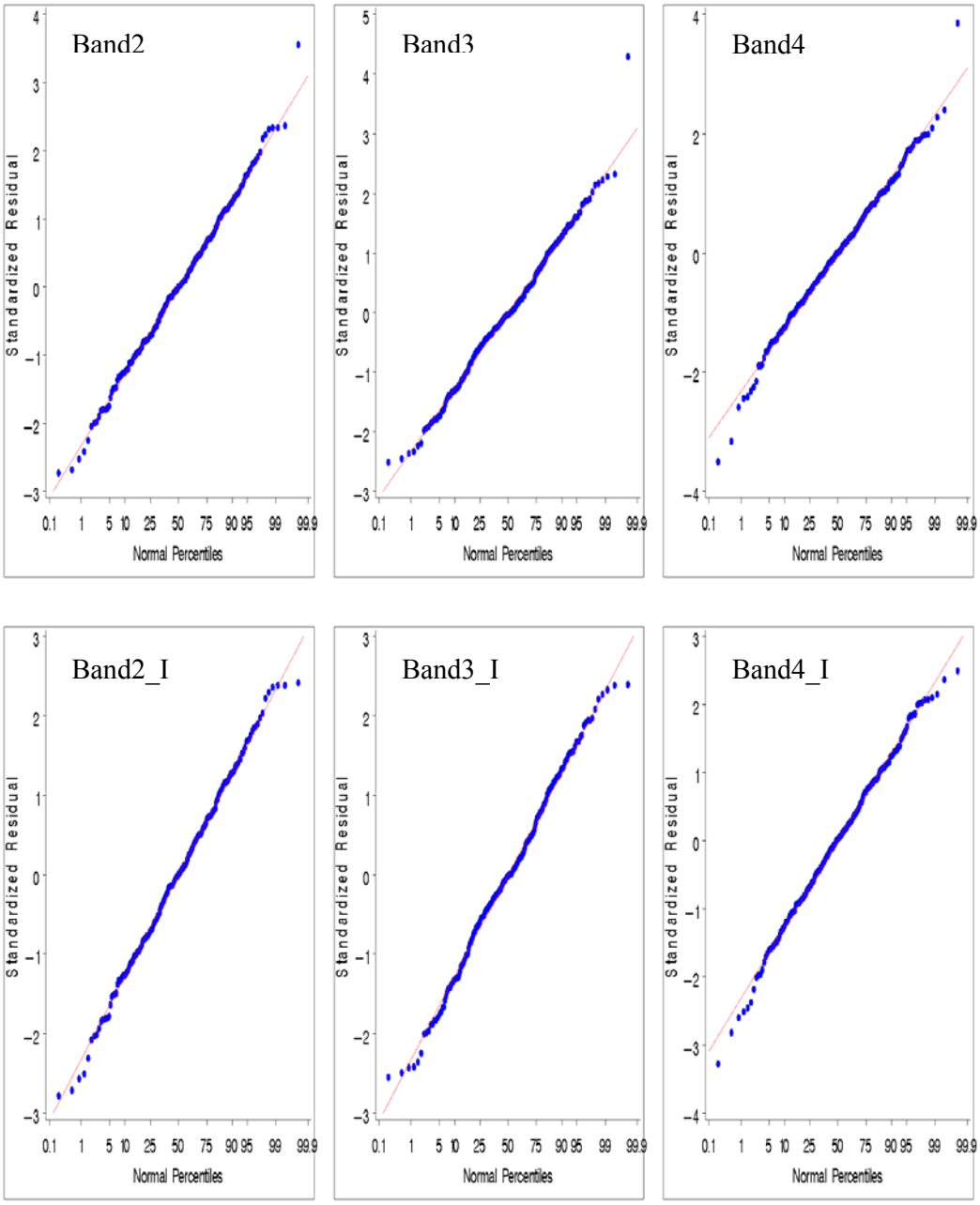


Figure 62: Normal probability plots corresponding to outlier removal iteration (I, II …) plots for bands 1 through 4 of 2001 field-3 data (MLR)

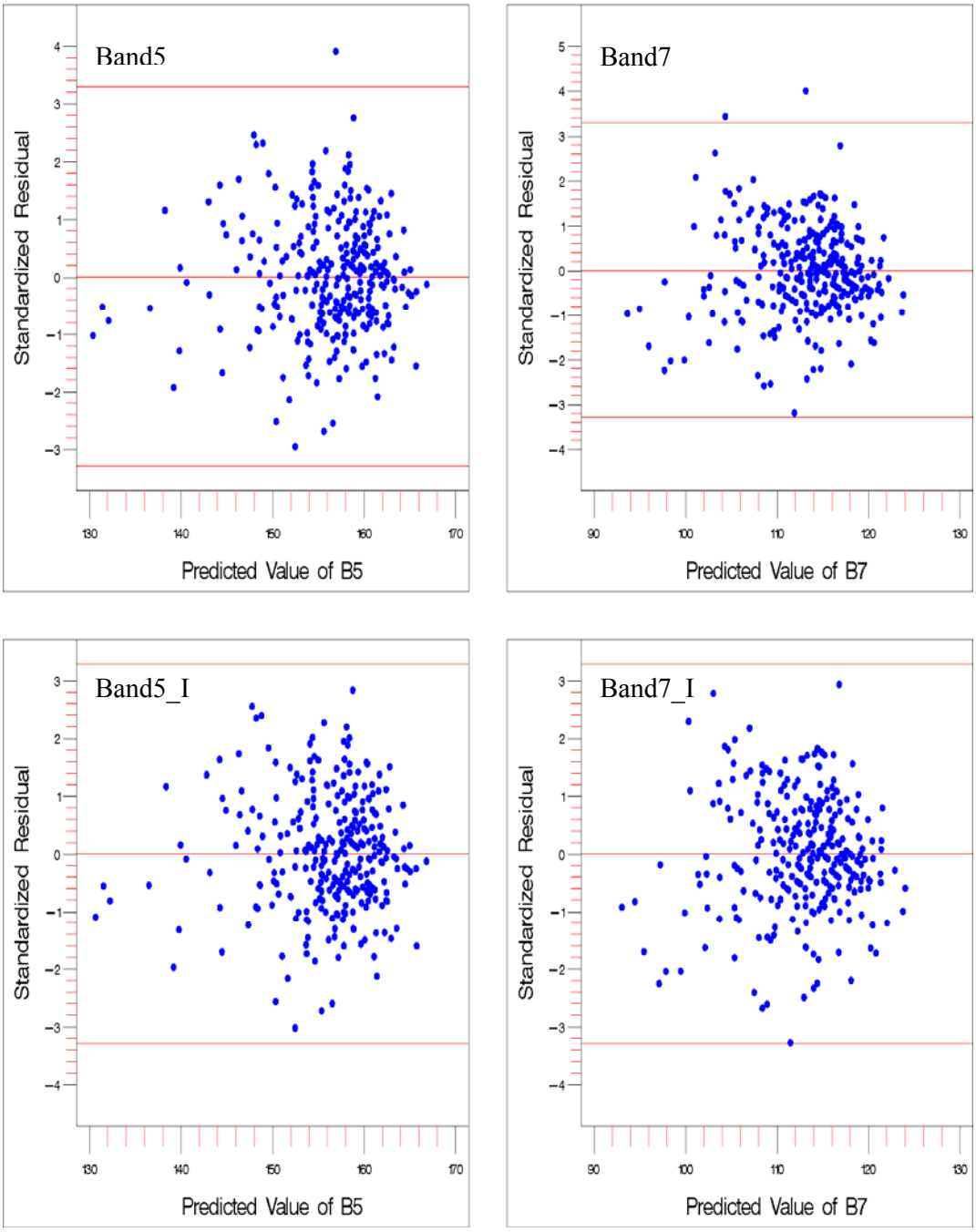


Figure 63: Outlier removal iterations (I, II …) plots of standardized residual vs. predicted value of dependent variable for bands 5 and 7 of 2001 field-3 data (MLR)

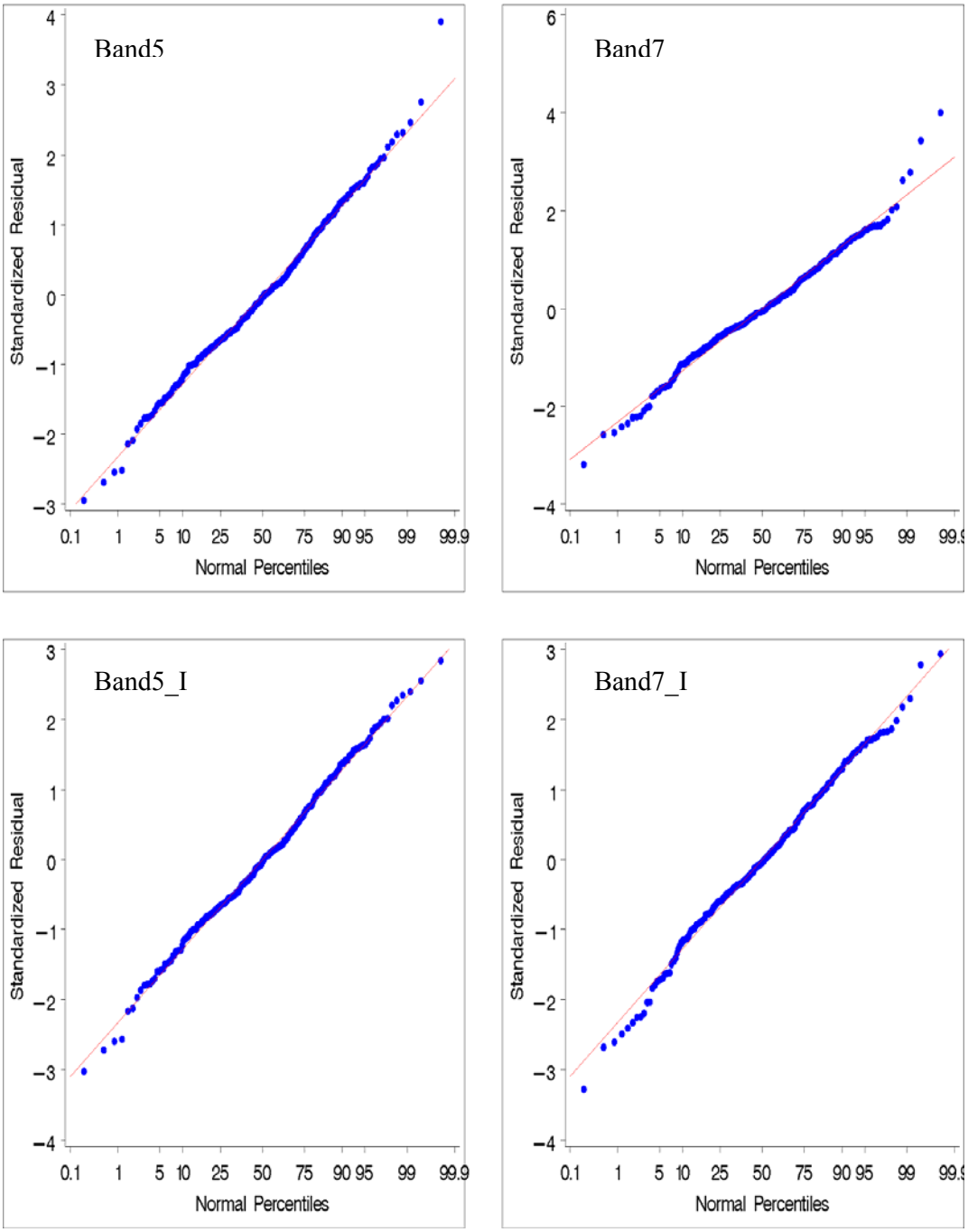


Figure 64: Normal probability plots corresponding to outlier removal iteration (I, II …) plots for bands 5 and 7 of 2001 field-3 data (MLR)

Figures 65 and 66 illustrate the effects of outlier removal on the R^2 values in bands 2, 3, and 4, and bands 5 and 7 respectively. An improvement in R^2 was observed for all the bands. The p values were not influenced by outlier removal, and significance levels remained high ($p<0.0001$).

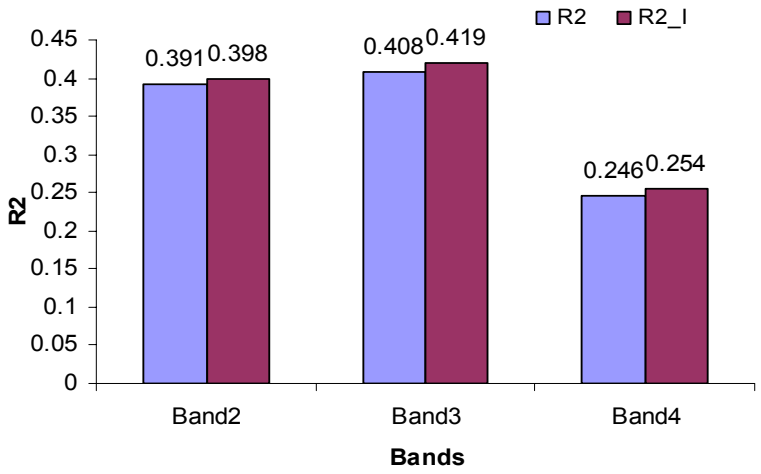


Figure 65: Changes in R^2 values with outlier removal at each iterations (I, II.. ..) for bands 2 through 4 of 2001 field-3 data

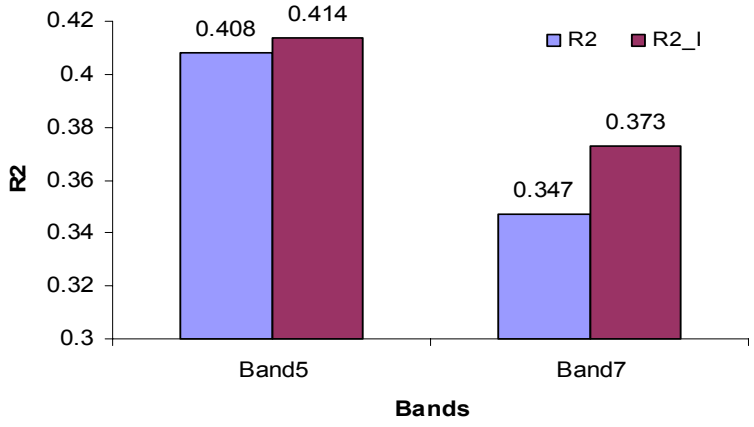


Figure 66: Changes in R^2 values with outlier removal at each iterations (I, II.. ..) for bands 5 and 7 of 2001 field-3 data

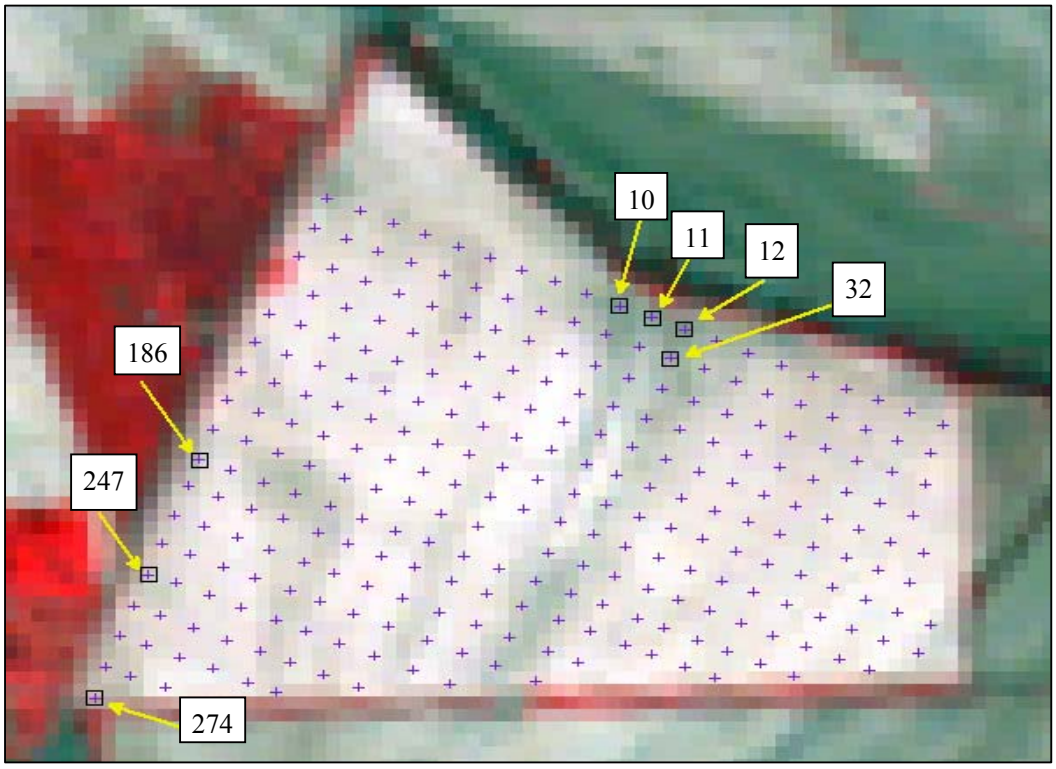
Validation

The MLR outliers of 1997 field-1 data are presented in table 16. As with SLR, point 274 was again observed as an outlier in all bands except band 4. Furthermore, the DN values of all the detected outliers were at least two standard deviations lower than the average value of their respective band. Also as with SLR, bands 4 and 5 had the highest number of outliers. Two new outliers, points 247 and 186, were observed in the MLR analysis, and point 13, which was an outlier in the SLR analysis, was not detected as an outlier in the MLR analysis.

Table 16: Detected outliers of all the studied bands of 1997 field-1 data and their corresponding actual radiance values (DN, Mean, Std. Dev) (MLR).

Bands	Obs. No.	Easting	Northing	Outlier Iterations	Actual Radiance values		
					DN	Mean	Std.Dev
Band1	274	744858.60	3774263.22	I	100	119.45	5.86
Band2	274	744858.60	3774263.22	I	53	65.54	3.93
Band3	32	745886.47	3774943.21	II	74	93.13	5.89
	274	744858.60	3774263.22	I	73		
Band4	10	745796.43	3775049.2	I	77	103.16	5.54
	11	745853.40	3775025.15	II	80		
	12	745911.00	3775001.46	I	78		
	32	745886.47	3774943.21	II	83		
Band5	32	745886.47	3774943.21	I	173	211.81	11.10
	186	745046.34	3774740.40	II	186		
	247	744955.25	3774510.26	II	190		
	274	744858.60	3774263.22	I	158		
Band7	274	744858.60	3774263.22	I	92	131.93	9.34

Figure 67 is a false color Landsat image of the field-1 with the position of the detected outliers and other sample data points overlaid. As, most of the outliers were similar to those of the SLR analysis, the reasons for their being outliers appear to be similar. No different reasons were observed for the new detected outliers (points 247 and 186).



(a)

(b)

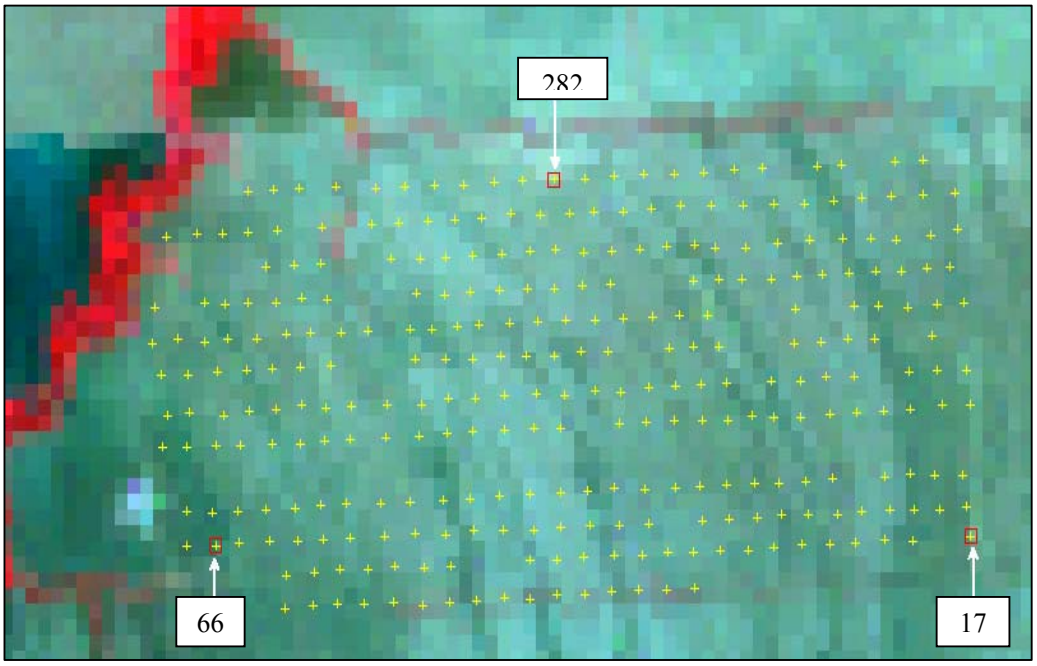
Figure 67: Position of outliers on the close-up (a) and distant (b) false color Landsat images of 1997 field-1 data (MLR)

Outliers from 2001 field-3 data are given in table 17. Point 282, which was an outlier in the SLR analysis, was again observed as an outlier, but in the MLR analysis it was an outlier in all bands, which was not the case for SLR. Also similar to SLR, the DN value of point 282 was high compare to the average DN value of its respective bands. Point 17 also had higher DN values, and point 66 had lower DN values than the average DN values in the respective bands. Two new outliers were observed with MLR, and three outliers detected with SLR were not detected with MLR analysis.

Table 17: Detected outliers of bands 2 through 7 for 2001 field-3 data and their corresponding actual radiance values.

Bands	**Obs. No.**	**Easting**	**Northing**	**Outlier Iterations**	**Actual Radiance values**		
					DN	**Mean**	**Std. Dev**
Band2	282	742318.416	3773253.088	I	103	88.97	5.391
Band3	282	742318.416	3773253.088	I	135	109.68	7.982
Band4	282	742318.416	3773253.088	I	93	77.282	5.054
	66	741561.461	3772587.365	I	60		
Band5	282	742318.416	3773253.088	I	184	56.24	9.013
Band7	282	742318.416	3773253.088	I	142	113.06	8.89
	17	743249.364	3772604.477	I	129		

Figure 68 illustrates the position of outliers in the field; the reasons that point 282 appears as an outlier are similar to those with the SLR analysis. In the case of points 66 and 17, no other specific reasons were noted, but it is worth point out that these points are also at the field boundaries, where it is likely that a Landsat pixel could include radiances from things other than bare soil in the field.



(a)

(b)

Figure 68: Position of outliers on the close-up (a) and distant (b) false color Landsat images of 2001 field-3 data (MLR)

Influence of soil sample location relative to Landsat image pixel

It is assumed that the centroid location of the roughly 30-m square Landsat image pixel is most representative of the pixel, however the soil-sample sites were not laid out in accord with image pixel structure, so they could be anywhere within a pixel. In order to account for the difference between sample location and pixel centroid location, the inverse of the distance between them was applied as a weighting factor to both SLR and MLR models. Equations 3 and 4 describe the relationship of the Weighted Simple Linear Regression (WSLR) and Weighted Multiple Linear Regression (WMLR) models.

WSLR (Weighted Simple Linear Regression)

Tables 18 and 19 include values of F and p, along with R^2 and adjusted R^2, for the WSLR analysis of fields 1 and 3, respectively. As with the SLR analysis, the WSLR analysis of 1997 field-1 data indicates that all the studied bands have a significant relationship between predicted Landsat reflectance and actual radiance. The R^2 values ranged from 0.09 to 0.14, with band 7 having the highest value.

Table 18: Statistical parameters for WSLR analysis of 1997 and 2001 field-1 data

Model: $B1 = \beta_0 + \beta_1 Band1 + \varepsilon$ 1997 (Landsat 5)				**Model: B1 = B0 + B1Band1+e 2001 (Landsat 7)**		
Bands	**F**	**P**	**R^2**	**F**	**P**	**R^2**
Band1	26.81	<.0001	0.0897	2.20	0.1389	0.0080
Band2	27.27	<.0001	0.0911	0.29	0.5912	0.0011
Band3	30.34	<.0001	0.1004	5.85	0.0162	0.0211
Band4	25.63	<.0001	0.0861	0.70	0.4042	0.0026
Band5	42.56	<.0001	0.1353	0.06	0.8022	0.0002
Band7	45.43	<.0001	0.1431	0.00	0.9600	0.0000

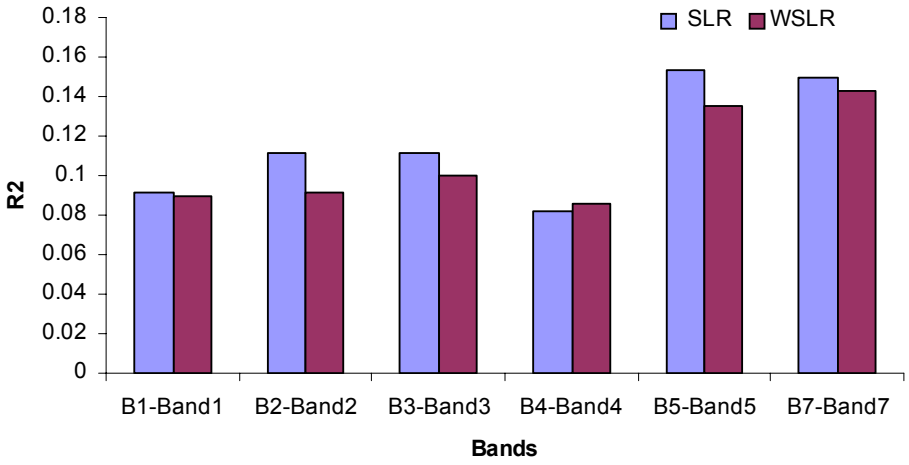


Figure 69: Comparison of R^2 values between SLR and WSLR analyses for 1997 field-1 data

Comparison of WSLR and SLR analyses of 1997 field-1 data (Figure 69) indicates a decrease in R^2 values except with band 4, which had a small improvement. In the case of 2001 field-1 data, WSLR analysis had similar results to those of SLR analysis, with only band 3 having a significant relationship ($p = 0.016$). The R^2 value for band 3 was very low (0.02). Again, a decrease in R^2 values was observed in the WSLR analysis for 2001 field-1 data.

The WSLR analysis of 1999 field-3 data indicates that bands 1 through 3 have significant relationships while the other three bands do not. Comparison to SLR analysis indicates some improvement with WSLR analysis; particularly that one additional band exhibited a significant relationship. Furthermore, an improvement in R^2 values can be observed with all the bands. However, R^2 values are still very low (0.04 to 0.08).

Table 19: Statistical parameters for WSLR analysis of 1999 and 2001 field-3 data

Model: B1 = $\beta_0 + \beta_1$Band1+ ε 1999 (Landsat 5)				Model: B1 = B0 + B1Band1+e 2001 (Landsat 7)		
Bands	**F**	**P**	**R^2**	**F**	**P**	**R^2**
Band1	12.51	0.0005	0.0401	95.00	<.0001	0.2411
Band2	27.53	<.0001	0.0843	121.33	<.0001	0.2887
Band3	19.55	<.0001	0.0614	142.20	<.0001	0.3223
Band4	1.16	0.2816	0.0039	7.11	0.0081	0.0232
Band5	2.47	0.1174	0.0082	73.53	<.0001	0.1974
Band7	2.69	0.1019	0.0089	54.19	<.0001	0.1534

The WSLR analysis of 2001 field-3 data indicates a significant relationship for all the studied bands. Band 4, which did not have significant relationship (p value = 0.07) with SLR analysis does have a significant relationship (p = 0.008) with WSLR analysis. Furthermore, comparison of R^2 values between SLR and WSLR analyses (Figure 70) indicates improvement in all bands except band 1, which had a small decrease.

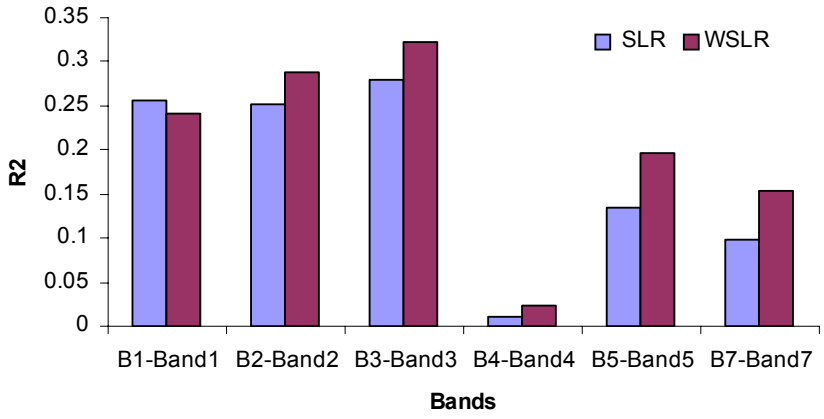


Figure 70: Comparison of R^2 values between SLR and WSLR analyses for 2001 field-3 data

WMLR (Weighted Multiple Linear Regression)

The WMLR analysis of data from fields 1 and 3 was carried out, and F and p, along with R^2 and adjusted R^2 of both fields are provided in Tables 20 and 23, respectively. Results for the 1997 field-1 data indicate a significant relationship and R^2 values greater than 0.35 for all the studied bands (Table 20). These results are similar to those of the MLR analysis, and R^2 values for WMLR follow a similar pattern to those of the MLR analysis, except that increases in R^2 values can be observed with WMLR for all bands (Figure 71).

Table 20: Statistical parameters for WMLR analysis of 1997 and 2001 field-1 data

Model: B1 = β_0 + β_1Band1+ β_2Clay + β_3Elevation+ ε 1997 (Landsat 5)					**Model: B1 = β_0 + β_1Band1+ β_2Clay + β_3Elevation+ ε 2001 (Landsat 7)**			
Bands	**F**	**P**	**R^2**	**Adj. R^2**	**F**	**P**	**R^2**	**Adj. R^2**
Band1	48.81	<.0001	0.3516	0.3444	2.17	0.0921	0.0235	0.0127
Band2	56.45	<.0001	0.3854	0.3786	0.37	0.7775	0.0041	-0.0070
Band3	66.03	<.0001	0.4232	0.4168	2.50	0.0598	0.0270	0.0162
Band4	58.62	<.0001	0.3944	0.3877	3.00	0.0312	0.0322	0.0215
Band5	69.91	<.0001	0.4372	0.4309	2.43	0.0654	0.0263	0.0155
Band7	67.85	<.0001	0.4298	0.4235	1.08	0.3595	0.0118	0.0008

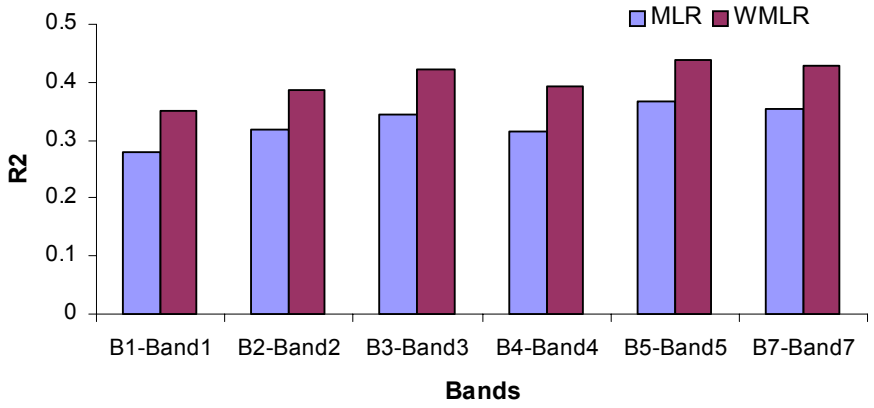


Figure 71: Comparison of R^2 values between MLR and WMLR analysis for 1997 field-1 data

Statistics from the t test (Table 21) indicate that the independent variables, expected reflectance and clay content, have a significant relationship with actual radiance in all the bands (p = < 0.0001). However, elevation is significant in only bands 5 and 7 (P = 0.02 and 0.0008). The results are very similar to those of the MLR analysis, except that elevation's significance in band 3 data with MLR occurs in band 5 with WMLR.

Table 21: t and p-values of WMLR analysis of independent variables for 1997 field-1 data

Model: B1 = β_0 + β_1Band1+ β_2Clay + β_3Elevation+ ε 1997 (Landsat 5)												
Variables	*Band1*		*Band2*		*Band3*		*Band4*		*Band5*		*Band7*	
	t	**P**	**t**	**P**	**t**	**P**	**t**	**P**	**t**	**P**	**t**	**P**
Expected Reflectance	7.05	<.0001	6.95	<.0001	7.48	<.0001	4.96	<.0001	7.62	<.0001	5.85	<.0001
Clay	-10.2	<.0001	-11.3	<.0001	-12.2	<.0001	-11.5	<.0001	-11.5	<.0001	-10.8	<.0001
Elevation	0.95	0.3423	-0.60	0.5464	-0.62	0.5381	0.86	0.3921	2.26	0.0245	3.40	0.0008

The WMLR results for 2001 field-1 data indicate that only bands 3 and 4 had a significant relationship (p = 0.05 and 0.03, respectively). The t statistics reveal that the independent variable clay had a non-significant relationship with actual radiance in all the bands. Furthermore, expected reflectance had a significant relationship in only band 3 (p = 0.01), similar to the case with SLR and MLR analyses. Elevation had a significant relationship in bands 1, 4, and 5. Hence, the significance of the WMLR model in band 4 appears to be due mainly to elevation. Since expected reflectance was significant only in band 3, no further WMLR analysis of this dataset was carried out.

Table 22: t and p values of WMLR analysis of independent variables for 2001 field-1 data

Model: $B1 = \beta_0 + \beta_1 Band1 + \beta_2 Clay + \beta_3 Elevation + \varepsilon$, 2001 (Landsat 7)												
Variables	*Band1*		*Band2*		*Band3*		*Band4*		*Band5*		*Band7*	
	t	**P**	**t**	**P**	**t**	**P**	**t**	**P**	**t**	**P**	**t**	**P**
Expected Reflectance	1.73	0.08	0.66	0.51	2.56	0.010	0.46	0.64	0.16	0.87	0.37	0.71
Clay	0.12	0.90	0.62	0.53	0.54	0.58	-0.04	0.96	1.70	0.09	1.53	0.12
Elevation	2.07	0.039	0.66	0.50	1.18	0.23	-2.88	0.004	-2.07	0.03	-0.92	0.35

According to F statistics of the WMLR analysis for 1999 field-3 data, all the bands have significant relationships, and except for band 4 all bands have R^2 values from 0.10 to 0.16 (Table 23).

Table 23: Statistical parameters for WMLR analysis of 1999 and 2001 field-3 data

Model: $B1 = \beta_0 + \beta_1 Band1 + \beta_2 Clay + \beta_3 Elevation + \varepsilon$ 1999 (Landsat 5)					**Model: $B1 = \beta_0 + \beta_1 Band1 + \beta_2 Clay + \beta_3 Elevation + \varepsilon$ 2001 (Landsat 7)**			
Bands	**F**	**P**	**R^2**	**Adj. R^2**	**F**	**P**	**R^2**	**Adj. R^2**
Band1	12.47	<.0001	0.1119	0.1029	60.74	<.0001	0.3802	0.3740
Band2	18.14	<.0001	0.1549	0.1463	72.69	<.0001	0.4234	0.4176
Band3	18.69	<.0001	0.1588	0.1503	77.09	<.0001	0.4378	0.4321
Band4	5.47	0.0011	0.0524	0.0428	40.16	<.0001	0.2886	0.2814
Band5	18.08	<.0001	0.1544	0.1459	86.35	<.0001	0.4659	0.4605
Band7	19.89	<.0001	0.1673	0.1589	64.39	<.0001	0.3941	0.3880

The t test suggests that expected reflectance in all the bands had a significant relationship with actual radiance, except in band 4 (Table 24). The elevation variable was significant in all the bands, and clay was significant in bands 4, 5, and 7 (Table 24).

Table 24: t and p-values of WMLR analysis of dependent variables for 1999 field-3 data

Model: B1 = β_0 + β_1Band1+ β_2Clay + β_3Elevation+ ε, 1999 (Landsat 5)												
Variables	*Band1*		*Band2*		*Band3*		*Band4*		*Band5*		*Band7*	
	t	**P**	**t**	**P**	**t**	**P**	**t**	**P**	**t**	**P**	**t**	**P**
Expected Reflectance	5.20	<.0001	6.86	<.0001	6.48	<.0001	1.67	0.095	2.85	0.004	2.00	0.046
Clay	-1.80	0.0722	-1.52	0.1289	-1.35	0.1790	-2.68	0.007	-3.46	0.0006	-3.48	0.0006
Elevation	4.19	<.0001	4.40	<.0001	5.37	<.0001	2.36	0.019	5.57	<.0001	5.87	<.0001

It has been shown that in bands 5 and 7 all the independent variables were significant in the WMLR model for actual radiance. This result is a substantial improvement over that of the MLR analysis, wherein none of the infrared bands had significant relationships. Therefore, these two bands will be considered further. The visible bands had two independent variables that were significant, but clay was not, and since clay was considered to be the more important moisture-related parameter, the visible bands in the 1999 field-3 data were not studied further with WMLR. Figure 72 indicates that, compared to MLR, WMLR improved R^2 in all bands.

In the case of 2001 data, the F-statistics indicate that all the bands had significant relationships ($p < 0.0001$). Furthermore, R^2 values greater than 0.35 were observed with all the studied bands except band 4 ($R^2 = 0.28$), and the highest R^2 value (0.46) was associated with band 5 (Table 23). The t test reveals that all the independent variables (expected reflectance, clay, and elevation) were significant in the WMLR model, except expected reflectance in band 4 (Table 25).

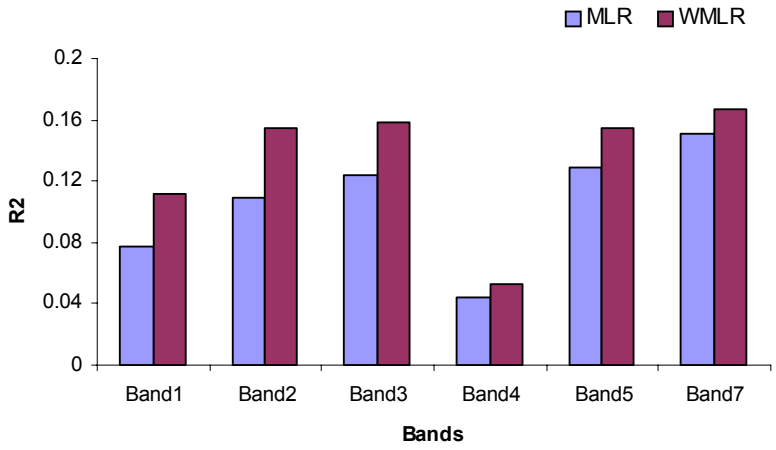


Figure 72: Comparison of R^2 values between MLR and WMLR analyses for 1999 field-3 data

Table 25: t and p-values of WMLR analysis of dependent variables for 2001 field-3 data

Model: B1 = β_0 + β_1Band1+ β_2Clay + β_3Elevation+ ε, 1999 (Landsat 5)												
Variables	*Band1*		*Band2*		*Band3*		*Band4*		*Band5*		*Band7*	
	t	**P**	**t**	**P**	**t**	**P**	**t**	**P**	**t**	**P**	**t**	**P**
Expected Reflectance	8.17	<.0001	9.67	<.0001	10.43	<.0001	1.72	0.0868	8.77	<.0001	6.07	<.0001
Clay	-7.76	<.0001	-8.20	<.0001	-7.58	<.0001	-9.59	<.0001	-12.1	<.0001	-10.6	<.0001
Elevation	-3.79	0.0002	-2.88	0.0042	-3.23	0.0014	-5.85	<.0001	-3.84	0.0001	-4.15	<.0001

These results are similar to those of the MLR analysis. Comparing R^2 values between MLR and WMLR analyses of 2001 field-3 data (Figure 73) indicates substantial improvement with WMLR in all studied bands except band 1.

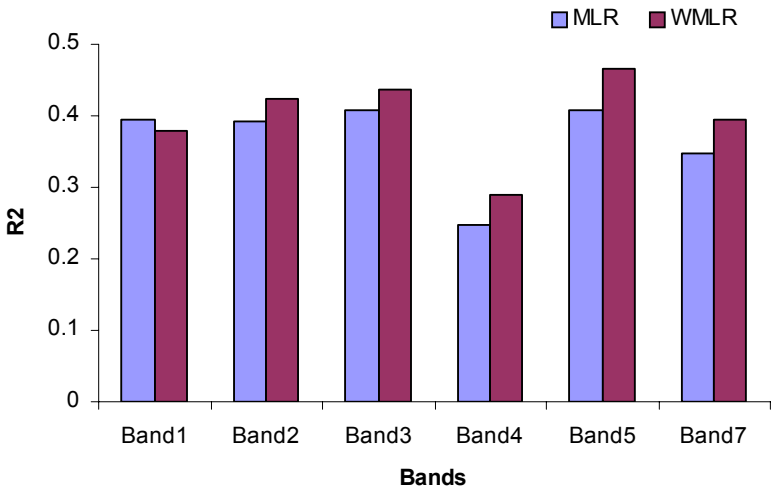


Figure 73: Comparison of R^2 values between MLR and WMLR analyses for 2001 field-3 data

Assumptions

The regression assumptions were tested for both WSLR and WMLR models.

Normality

WSLR

Normal probability plots from the WSLR analysis for 1997 and 2001 field-1 data and 1999 and 2001 field-3 data are given as figures 74 through 77 respectively. The plots for 1997 field-1 data show that in bands 1 through 3, data points are generally aligned with the reference line, while a few at the upper and lower tails are far off the line, implicating them as possible outliers (Figure 74). Data points in bands 4 through 7 are also reasonably well aligned to the reference line (Figure 74). However, in bands 4 and 5, some curvature is observed in the data, but again such behavior is not of much concern, as the F test is robust against non-normality. The normality plot of

2001 field-1 data in band 3 suggests non-normality, as the data points are not aligned with the reference line (Figure 75), but again this is not of much concern. Once again, a few data points lie far from the reference line at the lower tail, likely implicating them as outliers.

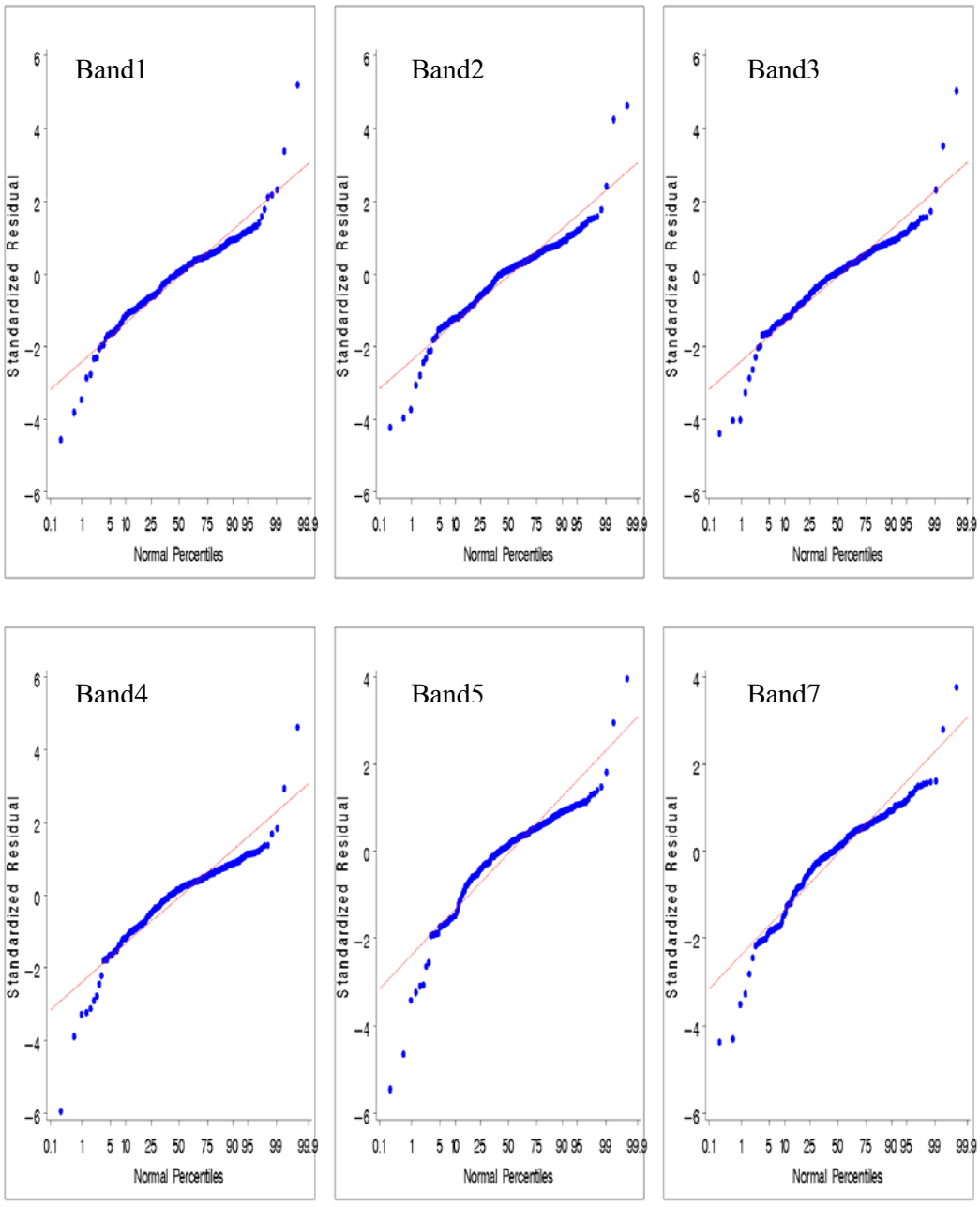


Figure 74: Normal probability plots for bands 1 through 7 of 1997 field-1 data (WSLR)

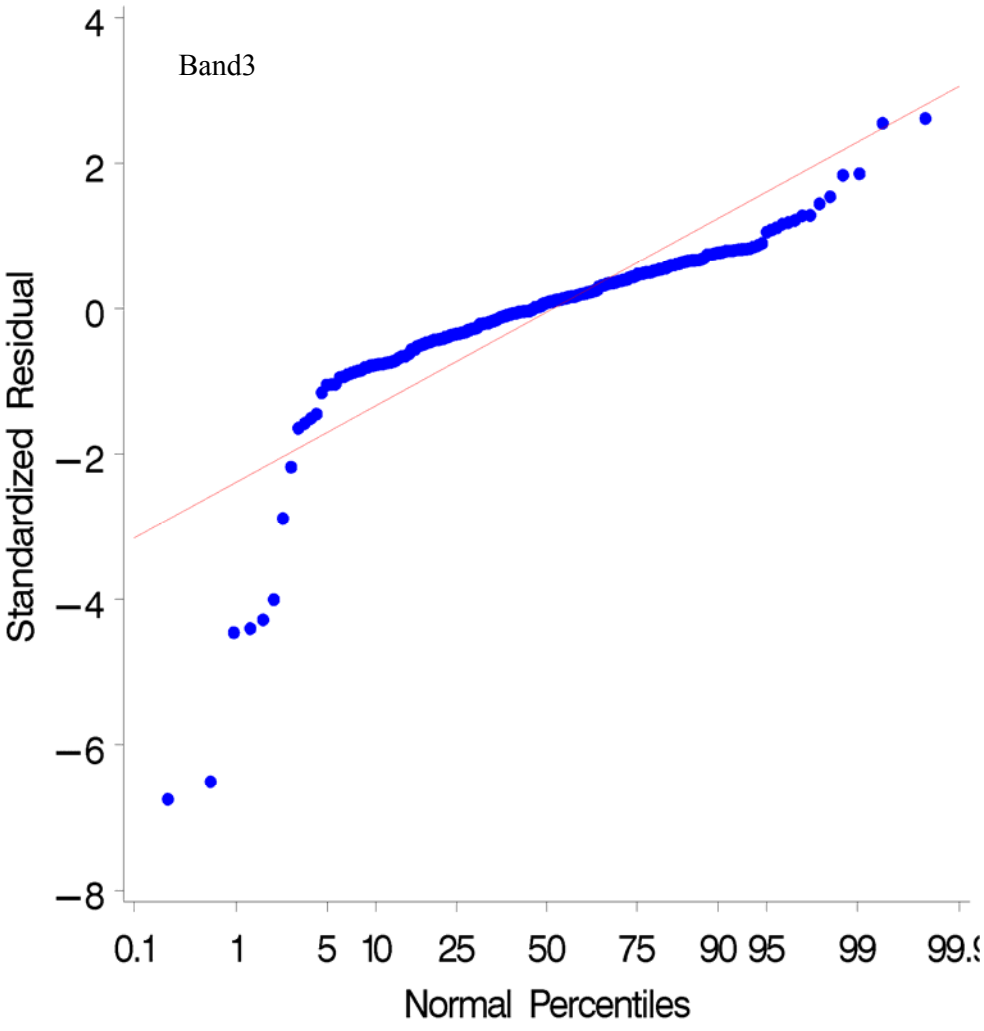


Figure 75: Normal probability plots for band 3 of 2001 field-1 data (WSLR)

The normality plots of 1999 field-3 data in bands 1 to 3 reveal that the data points are fairly well aligned with the reference line, with a few lying farther from the line at the lower and upper tails, implicating them as likely outliers (Figure 76). Normality plots of 2001 field-3 data (Figure 77) indicate that data points in all the studied bands are well aligned to the reference line. Again, a few data points lie farther from the reference line at the lower and upper tails, implicating them as possible outliers.

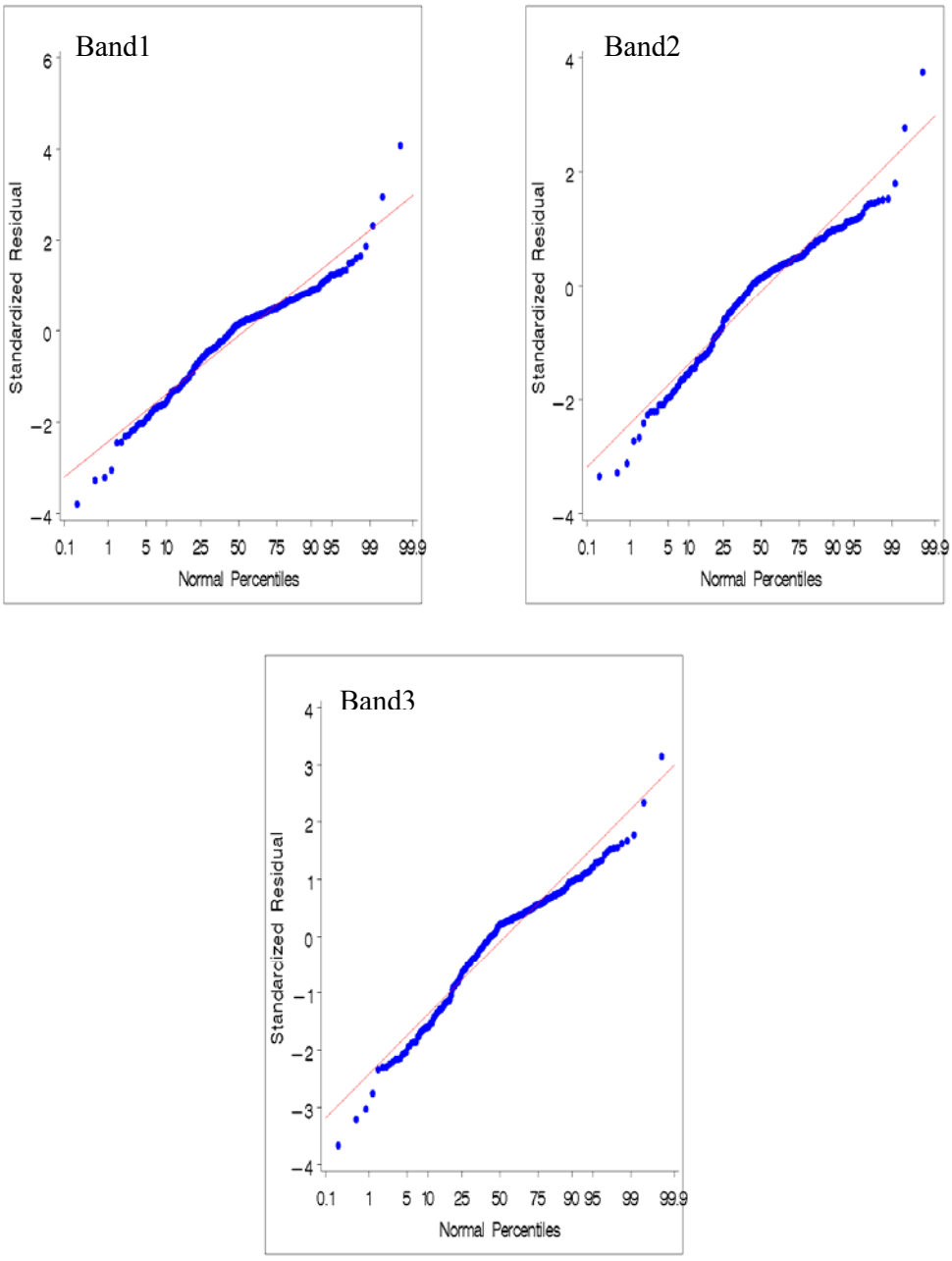


Figure 76: Normal probability plots for bands 1 through 3 of 1999 field-3 data (WSLR)

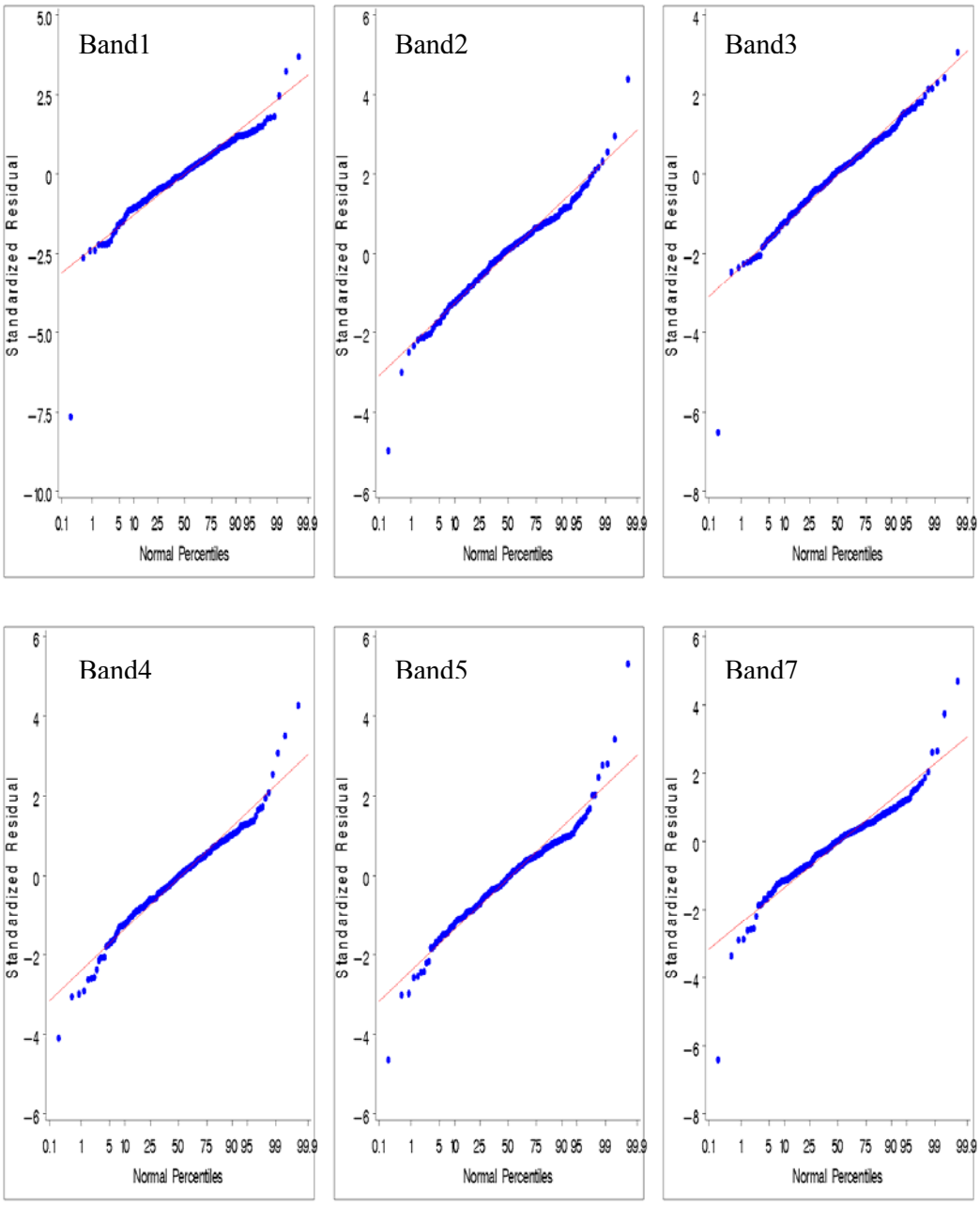


Figure 77: Normal probability plots for bands 1 through 7 of 2001 field-3 data (WSLR)

WMLR

The normal probability plots for WMLR analysis of 1997 field-1 data and 1999 and 2001 field-3 data are presented in figures 78 through 80 respectively. The plots for field 1 indicate that the data points in all the bands are well aligned to reference line. A few points at the upper and lower tails lie far from the reference line, implicating them as possible outliers (Figure 78). The normality plots for 1999 field-3 data in bands 5 and 7 suggest that the data points together exhibit some curvature, but are reasonably well aligned with the reference line, with a few possible outliers lying slightly away from the line (Figure 79). Again, it is important to note that the F statistic is robust against non- normality, so any mild curvature in the data is of not much concern. The 2001 field-3 plots suggest normal distribution in all the bands, as the data points are well aligned with the reference line, a few possible outliers lying far from the line (Figure 80).

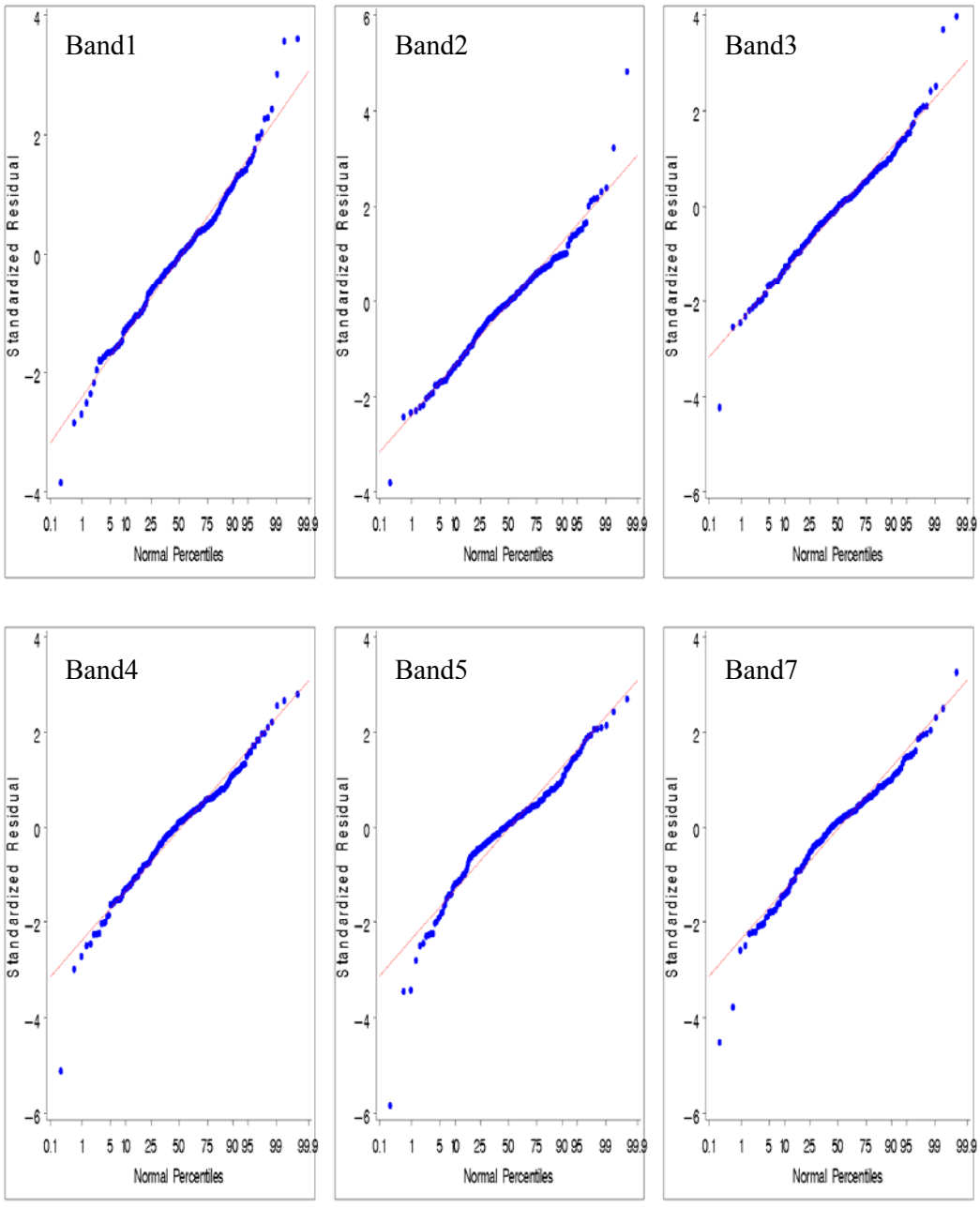


Figure 78: Normal probability plots for bands 1 through 7 of 1997 field-1 data (WMLR)

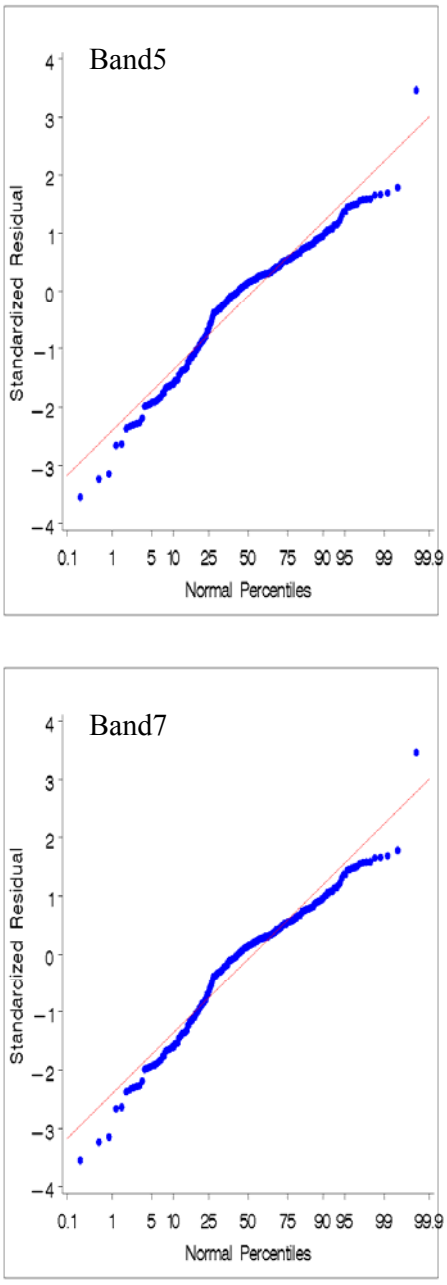


Figure 79: Normal probability plots for bands 5 and 7 of 1999 field-3 data (WMLR)

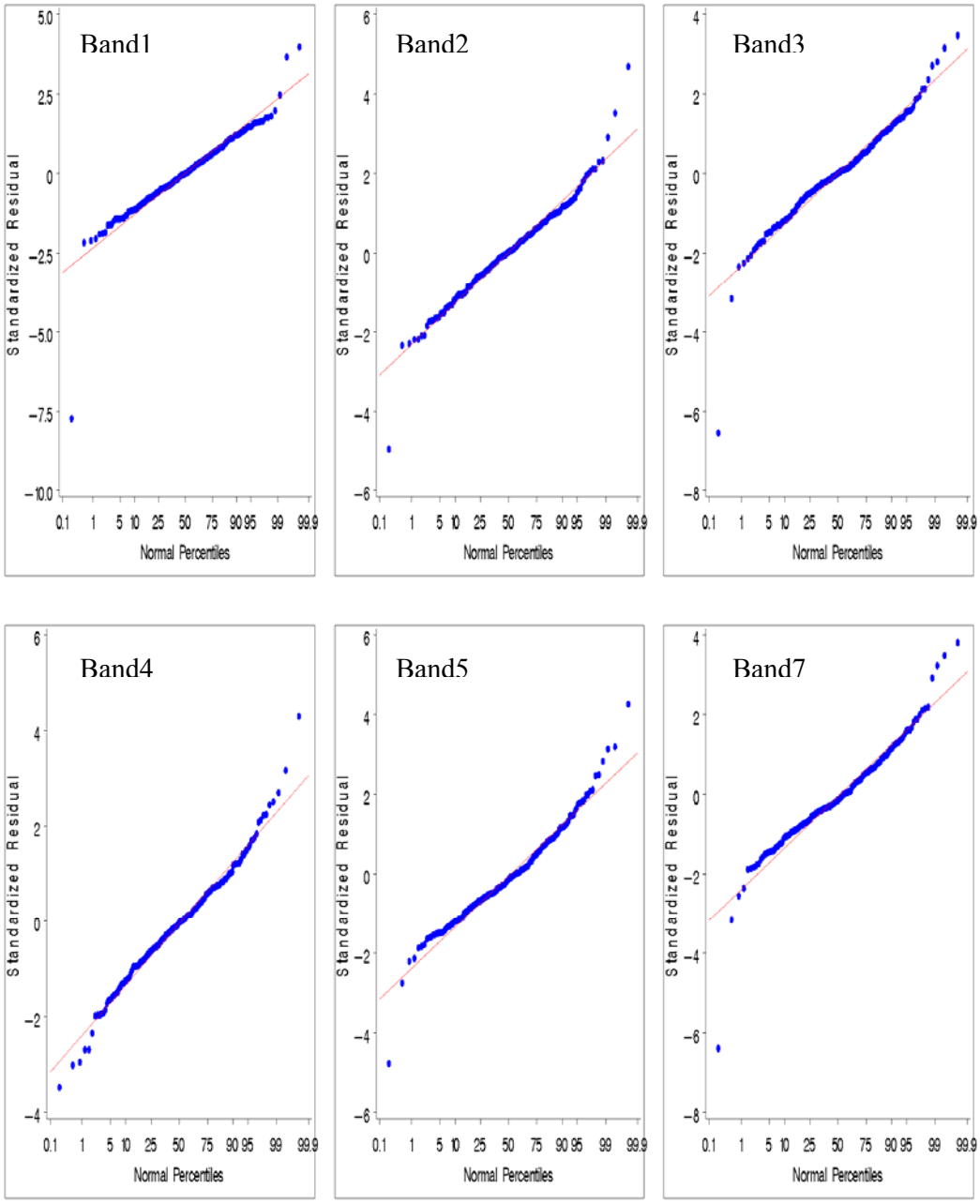


Figure 80: Normal probability plots for bands 1 through 7 of 2001 field-3 data (WMLR)

Constant Variance

WSLR

Plots of standardized residuals vs. WSLR predicted values for 1997 field-1 data and 1999 and 2001 field-3 data are given in Figures 81 through 83, respectively. The plots for 1997 field-1 data suggest that the data points are randomly distributed above and below the reference line (standardized residual = 0) in all the bands, indicating homogenous variance (Figure 81). A few datpoints' lying farther from the line than most indicates the possible presence of outliers. Figures 82 and 83, of 1999 and 2001 field-3 data (bands 1 to 3 in each), also indicate homogenous variance among the respective data. Again, possible outliers appear to be present.

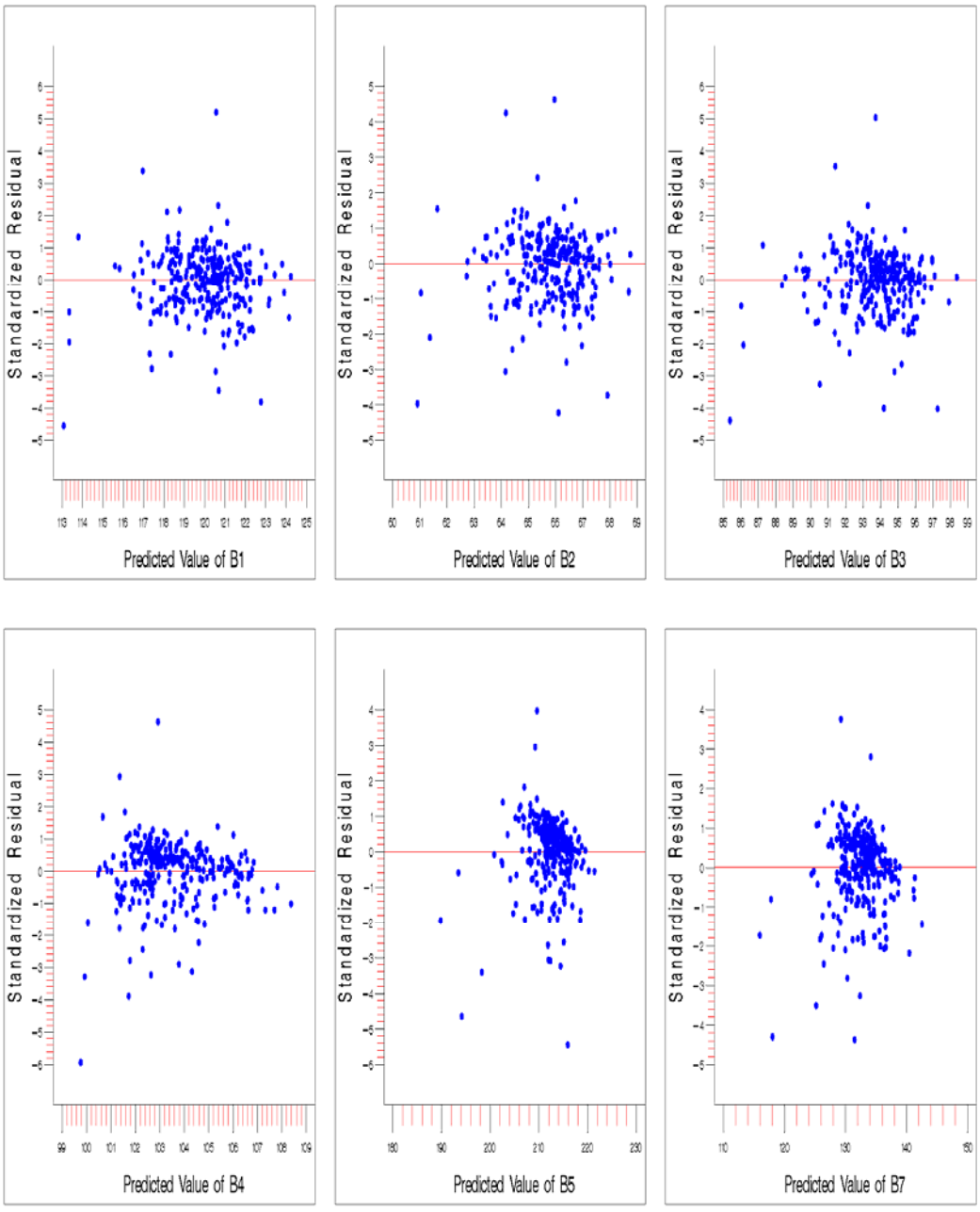


Figure 81: Plots of standardized residual vs. predicted value of dependent variable for bands 1 through 7 of 1997 field-1 data (WSLR)

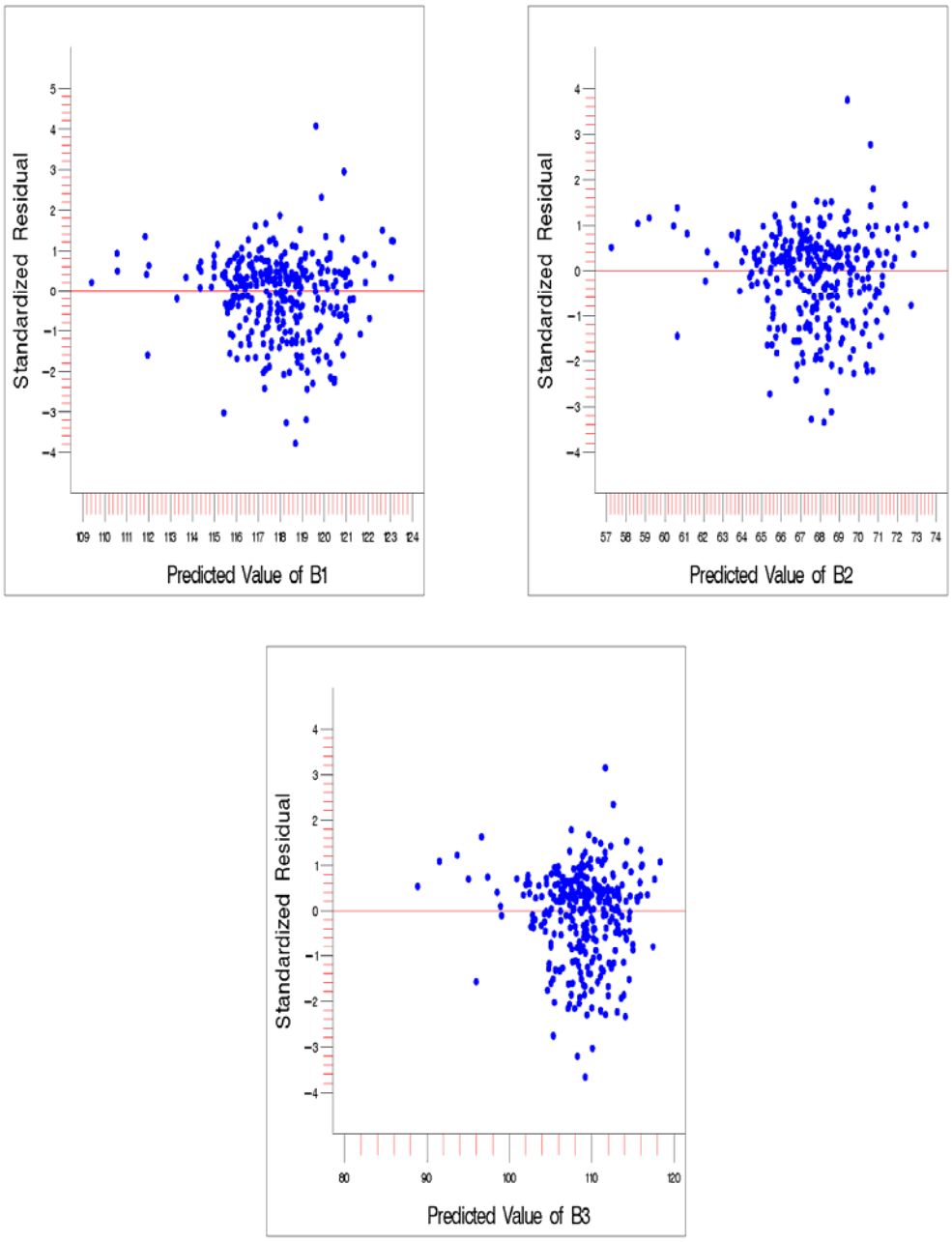


Figure 82: Plots of standardized residual vs. predicted value of dependent variable for bands 1 through 3 of 1999 field-3 data (WSLR)

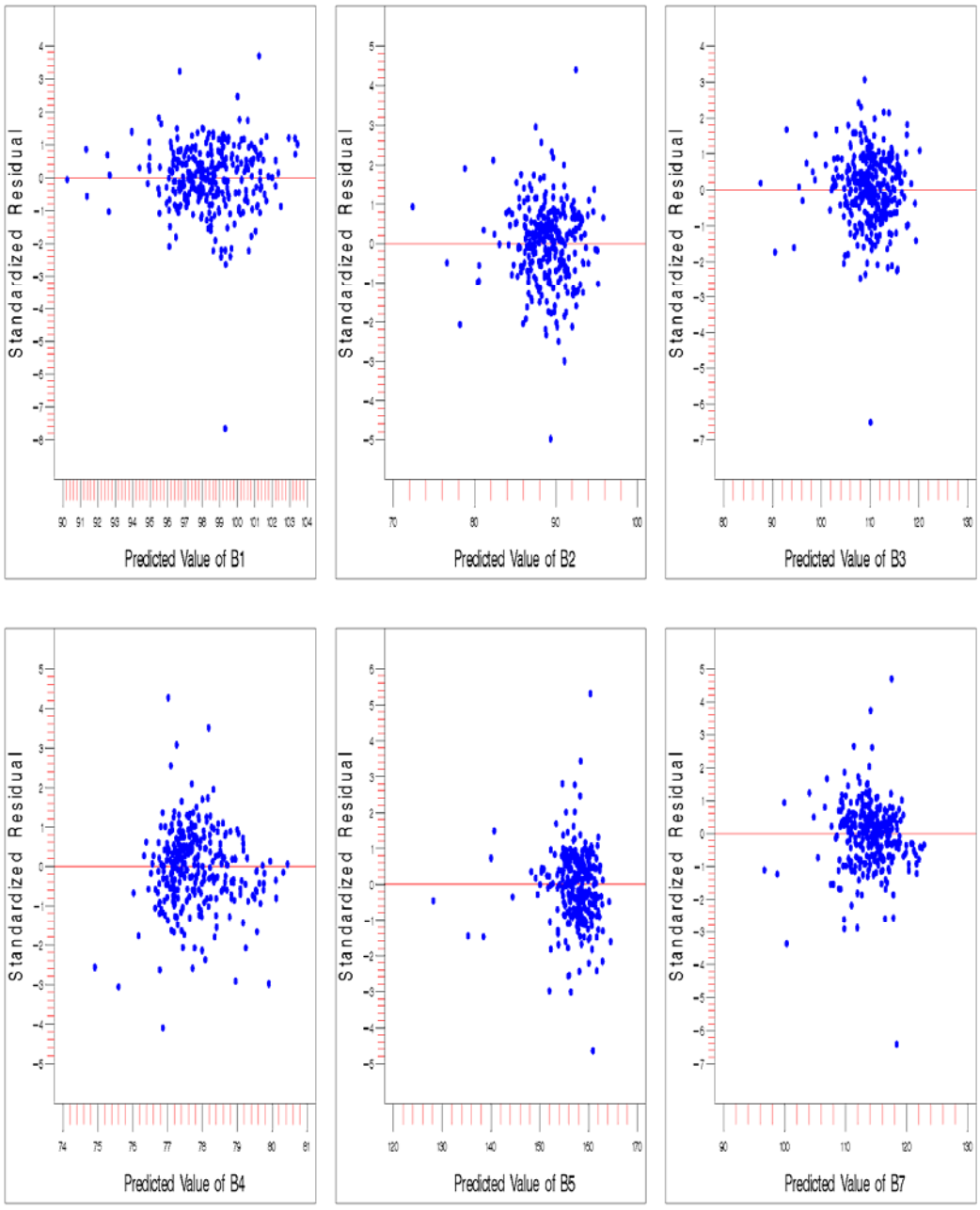


Figure 83: Plots of standardized residual vs. predicted value of dependent variable for bands 1 through 7 of 2001 field-3 data (WSLR)

WMLR

Figures 84 through 86 are plots of standardized residuals vs. WMLR predicted values for 1997 field-1 data and 1999 and 2001 field-3 data, respectively. Figure 84 indicates that 1997 field-1 data points are randomly distributed above and below the reference line (standardized residual = 0) in all the bands, indicating homogenous variance. A few data points lying far from the line appear to be possible outliers. Figures 85 and 86 also suggest constant variance for 1999 field-3 data in bands 5 and 7 and 2001 field-3 data in bands 1 through 7, respectively. The presence of possible outliers is once again observed.

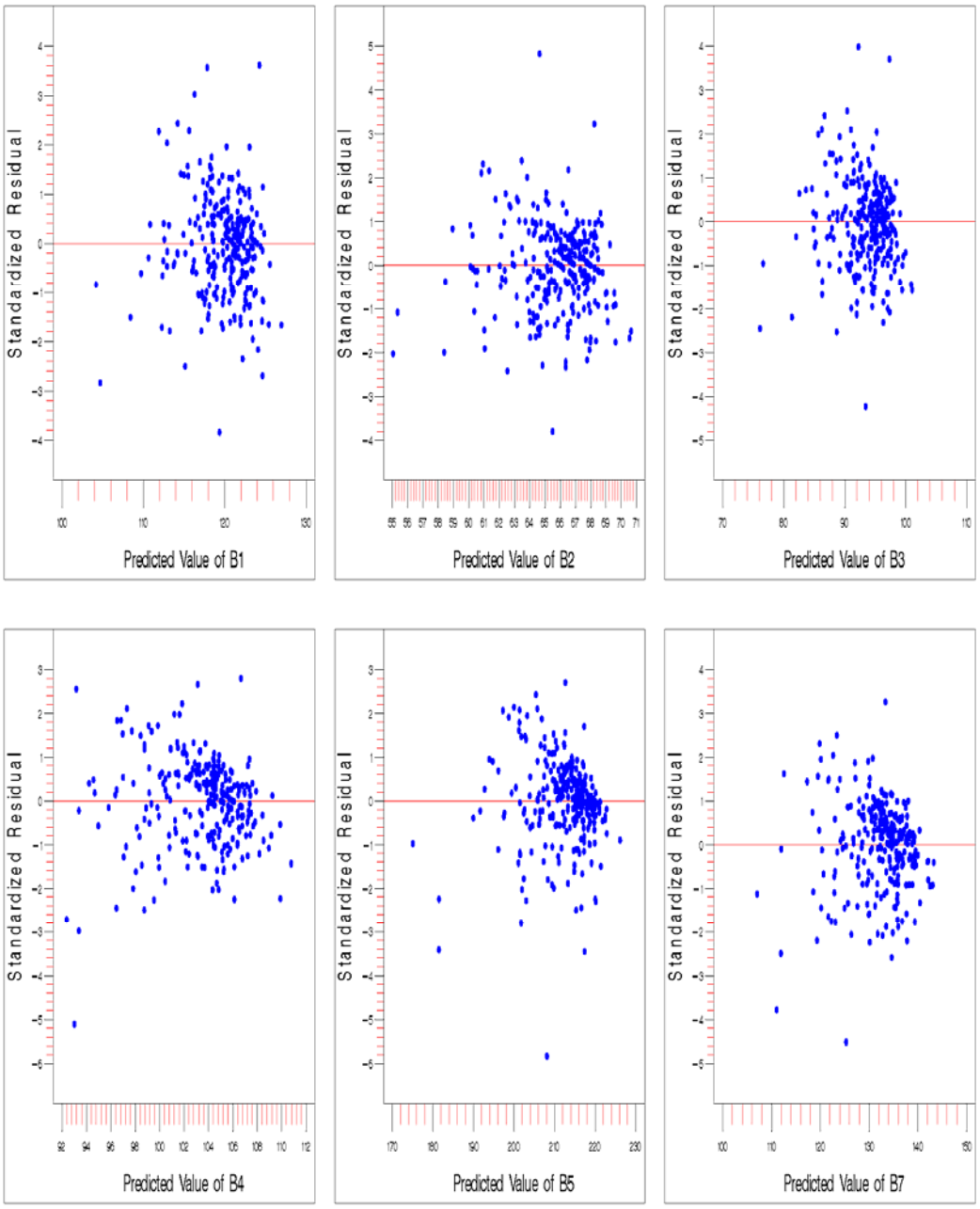


Figure 84: Plots of standardized residual vs. predicted value of dependent variable for bands 1 through 7 of 1997 field-1 data (WMLR)

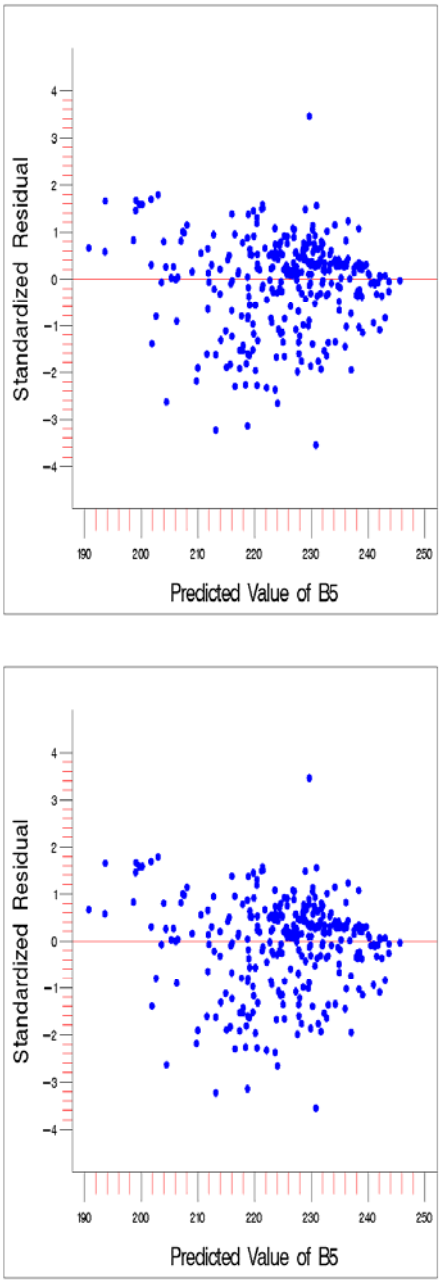


Figure 85: Plots of standardized residual vs. predicted value of dependent variable for bands 5 and 7 of 1999 field-3 data (WMLR)

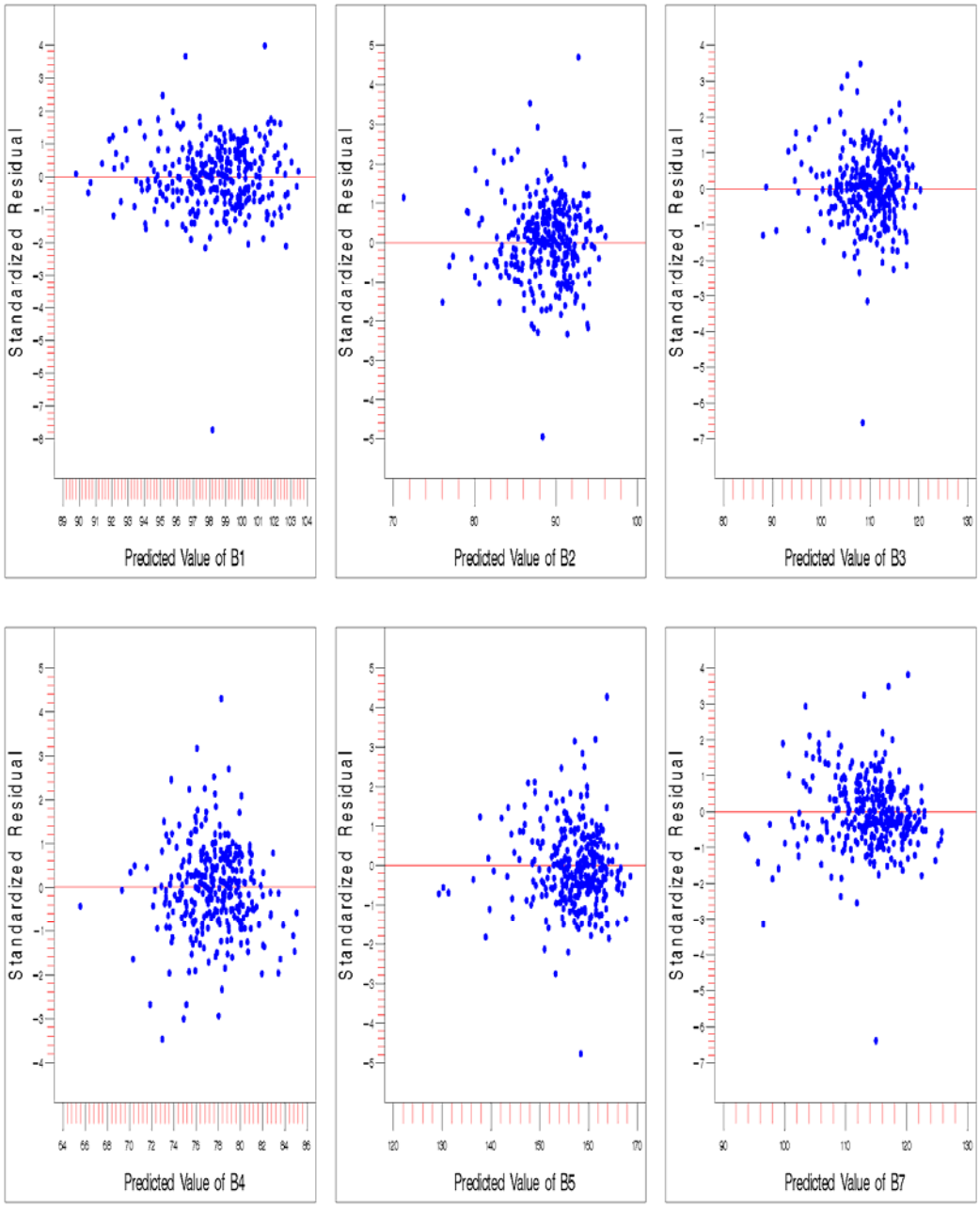


Figure 86: Plots of standardized residual vs. predicted value of dependent variable for bands 1 through 7 of 2001 field-3 data (WMLR)

Outlier Analysis

Detection

WSLR

The outlier boundary values for WSLR and WMLR analyses are similar to those for SLR and MLR analyses (3.29 and –3.29). Figures 87 through 90 are plots of standardized residual vs. predicted value of the dependent variable for 1997 and 2001 field-1 data and 1999 and 2001 field-3 data. The lines marked at the 3.29 and –3.29 standardized residual positions are the outlier boundaries. Data points outside these lines were considered as outliers. According to figures 87 and 88, all the bands in the 1997 field-1 data and band 3 in 2001 field-1 data have outliers, as a few data points in each were lying outside the outlier boundaries. Similarly, the 1999 and 2001 field-3 data also had outliers in all the studied bands (Figures 89 and 90).

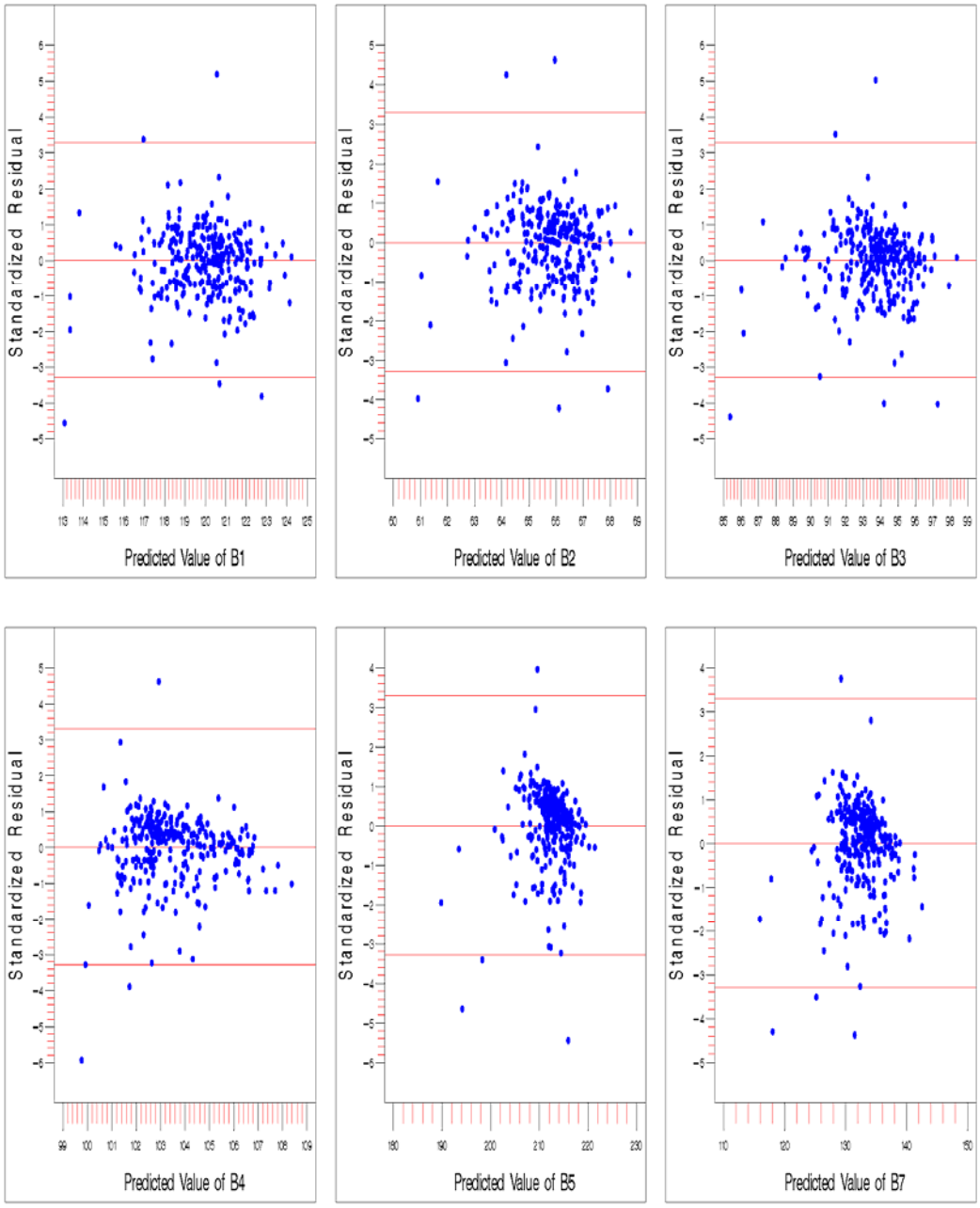


Figure 87: Outlier detection plots of standardized residual vs. predicted value of dependent variable for bands 1 through 7 of 1997 field-1 data (WSLR)

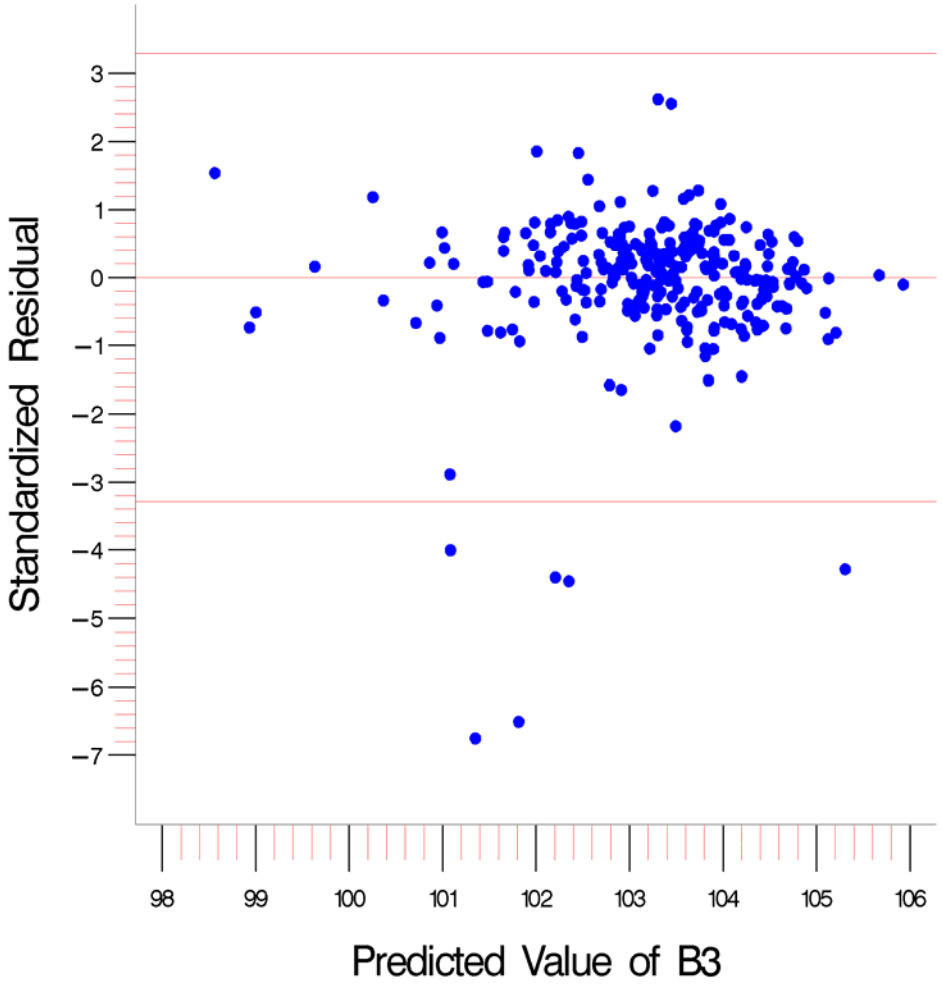


Figure 88: Outlier detection plots of standardized residual vs. predicted value of dependent variable for band 3 of 2001 field-1 data (WSLR)

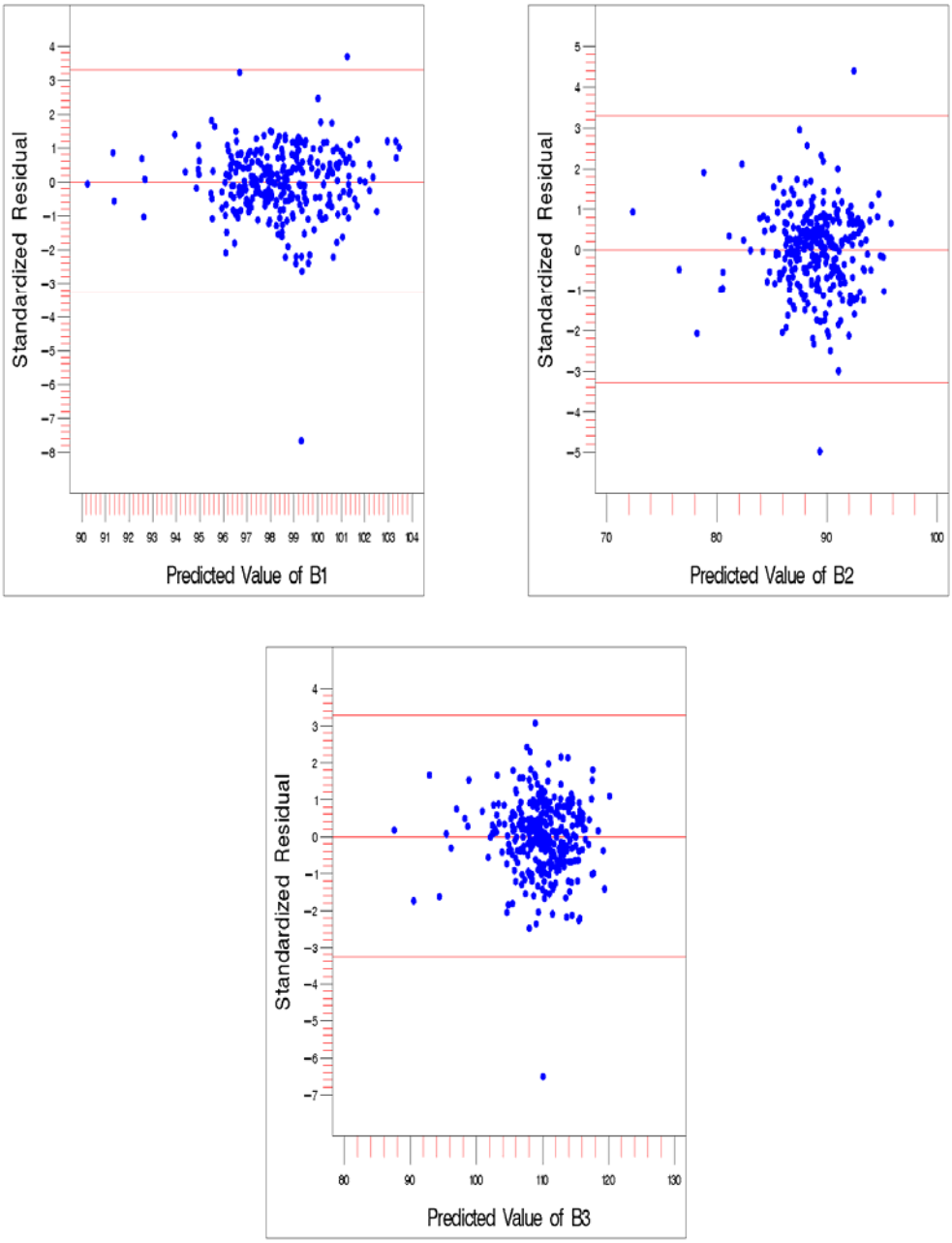


Figure 89: Outlier detection plots of standardized residual vs. predicted value of dependent variable for bands 1 through 3 of 1999 field-3 data (WSLR)

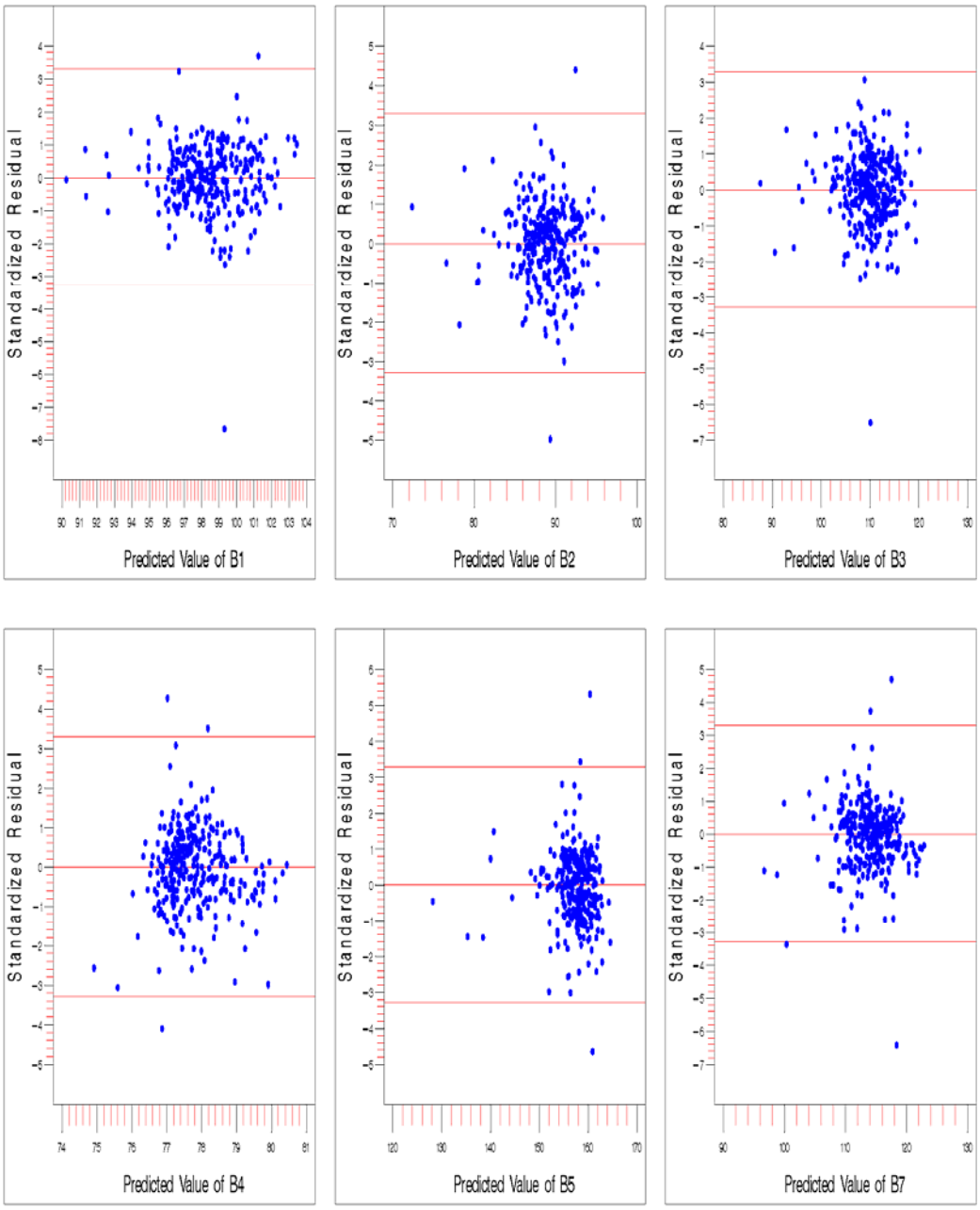


Figure 90: Outlier detection plots of standardized residual vs. predicted value of dependent variable for bands 1 through 7 of 2001 field-3 data (WSLR)

WMLR

Figures 91 through 93 illustrate WMLR outlier detection in 1997 field-1 data and 1999 and 2001 field-3 data, respectively. As can be seen in these figures, a few data points lying outside the outlier boundaries were detected in all the studied bands. These outliers were removed with the process previously described in WSLR section.

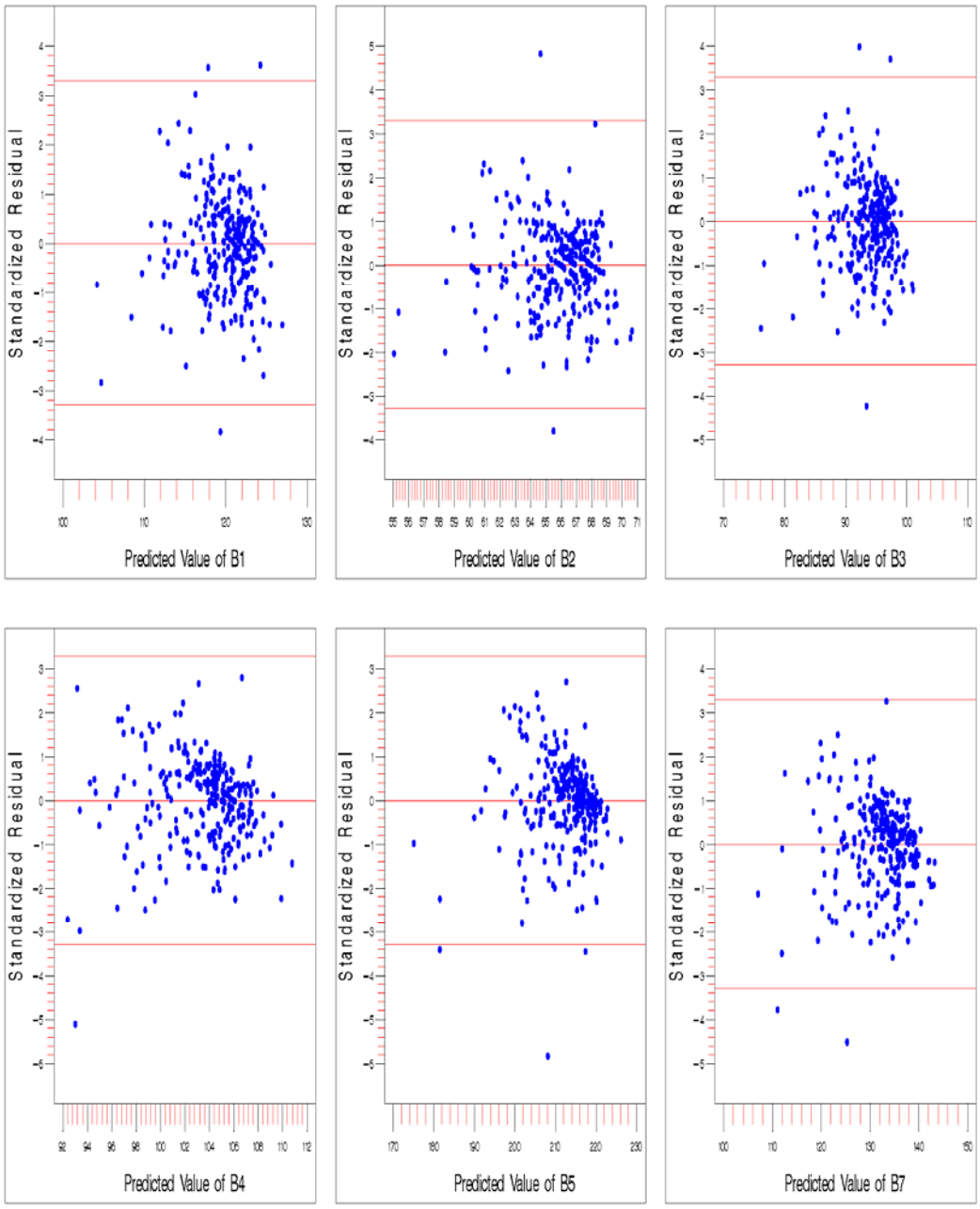


Figure 91: Outlier detection plots of standardized residual vs. predicted value of dependent variable for bands 1 through 7 of 1997 field-1 data (WMLR)

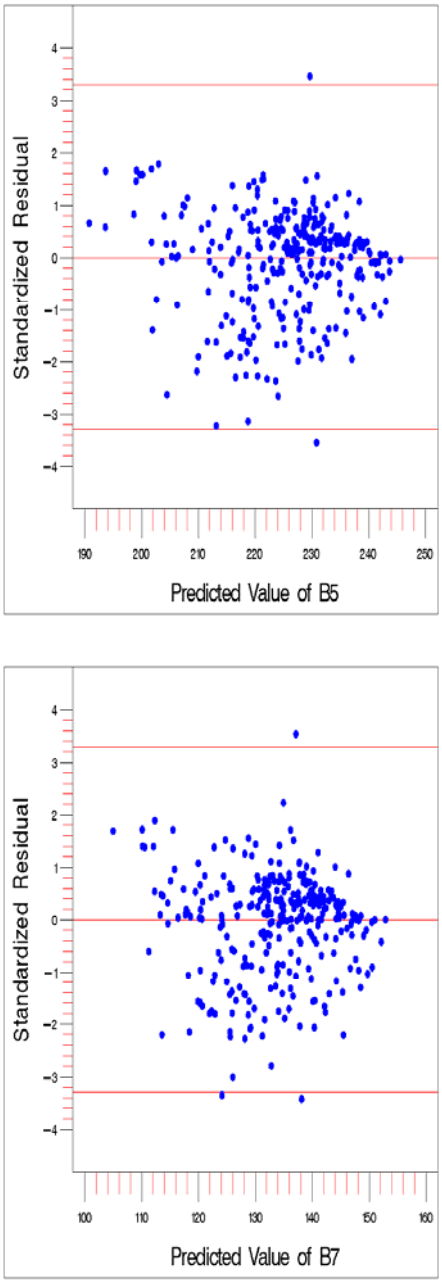


Figure 92: Outlier detection plots of standardized residual vs. predicted value of dependent variable for bands 5 and 7 of 1999 field-3 data (WMLR)

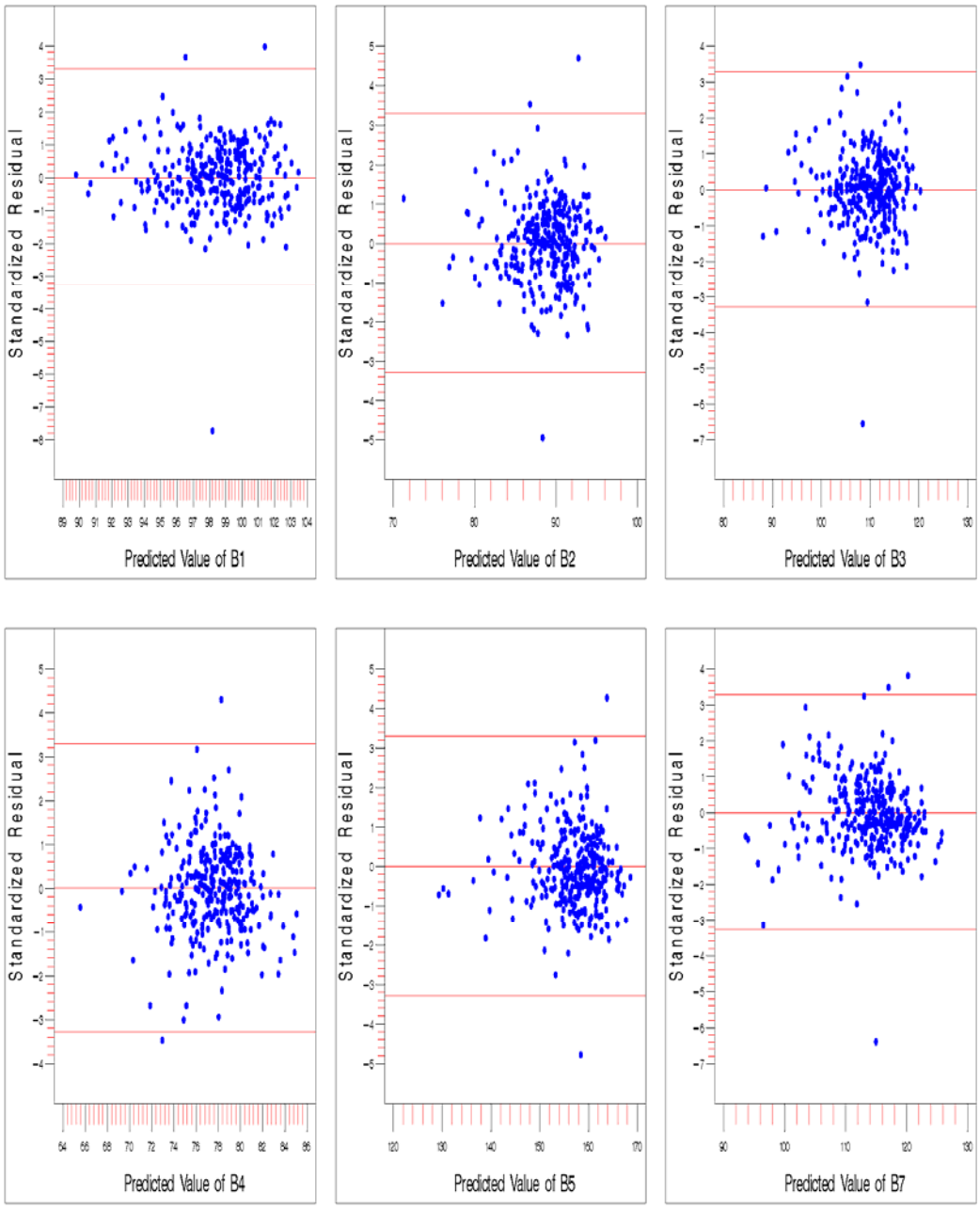


Figure 93: Outlier detection plots of standardized residual vs. predicted value of dependent variable for bands 1 through 7 of 2001 field-3 data (WMLR)

Removal

WSLR

Field-1

The outlier removal process for 1997 and 2001 field-1 data is illustrated in Figures 94, 96, 98, 100, 102, and 104. The notations describing the outlier iterations are similar to those described in the previous sections (SLR, MLR). The corresponding normal probability plots of the data before and after outlier removal are provided in Figures 95, 97, 99, 101, 103, and 105. The constant variance plots for bands 1 and 2 of 1997 field-1 data illustrate that, after removal of outliers, the distribution of the residuals about the reference line is improved. Both bands required two iterations to remove all the outliers (Figure 94). The corresponding normality plots (Figure 95) also appear to improve, as after removal of outliers the data points are seen to be well aligned to the reference line. Furthermore, data points lying far from the reference line are no longer in evidence. The constant variance and normality plots for band 3 (Figures 96 and 97) suggest similar conclusions to those for bands 1 and 2. The band-4 outlier removal process needed three iterations to remove the detected outliers. The band-4 constant variance plots (Figure 98) appear to be improved without the outliers, and the normality plots (Figure 99) appear to have improved also, as the data points lying at the lower and upper tails are no longer visible after the removal of outliers. Furthermore, the curvature of the data has become more nearly linear and closer to the reference line. Band 5 data needed four iterations to remove the outliers, and with each iteration the constant variance and normal probability plots appear to be

improved (Figures 100 and 101). With band 7, three iterations were required to remove the outliers, and similar observations were noted (Figures 102 and 103). Band 3 of 2001 field-1 data required two iterations to remove the outliers, and, similar conclusions to those for 1997 field-1 data were drawn (Figures 104 and 105)

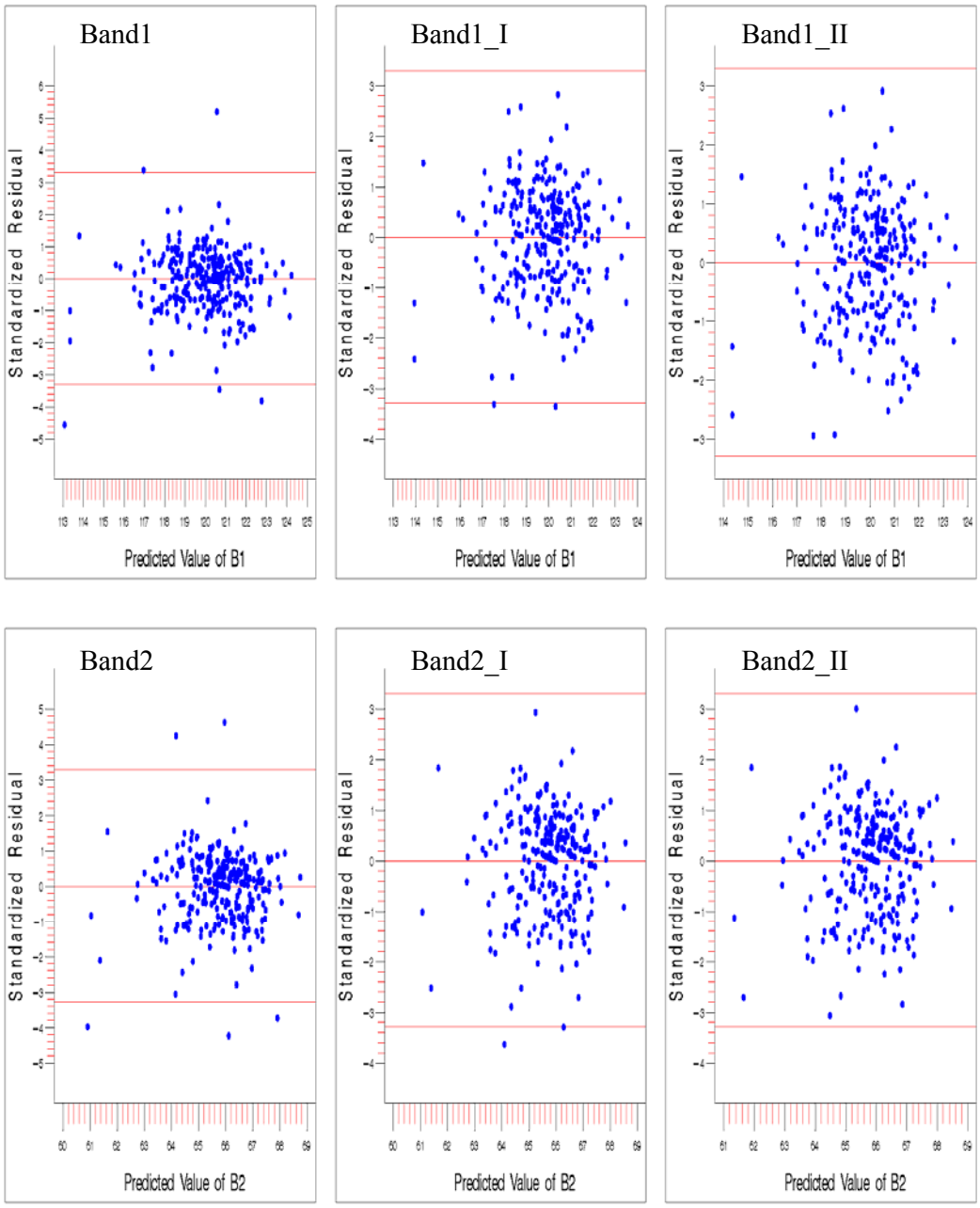


Figure 94: Outlier removal iteration (I, II …) plots of standardized residual vs. predicted value of dependent variable for bands 1 and 2 of 1997 field-1 data (WSLR)

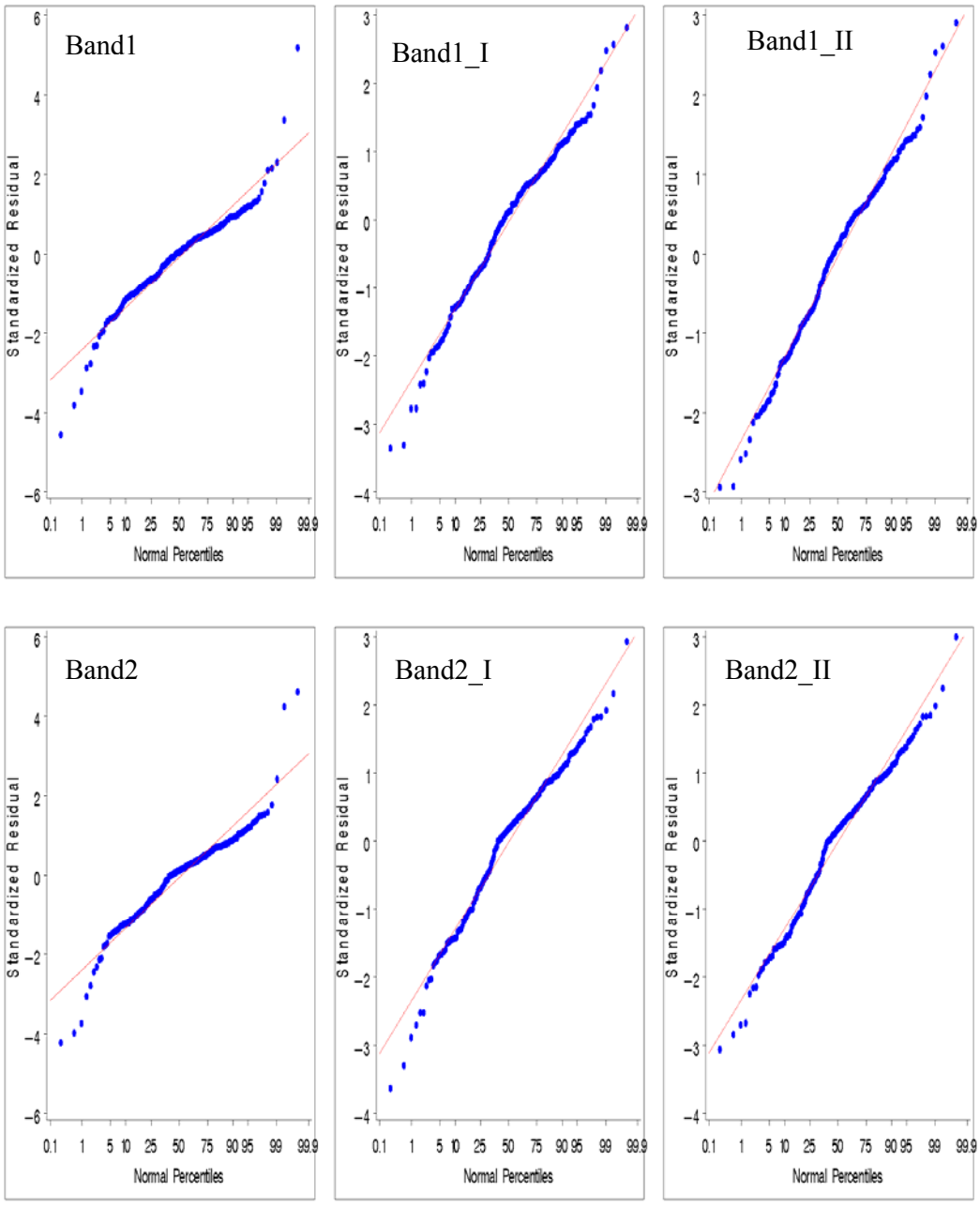


Figure 95: Normal probability plots corresponding to outlier removal iteration (I, II …) plots for bands 1 and 2 of 1997 field-1 data (WSLR)

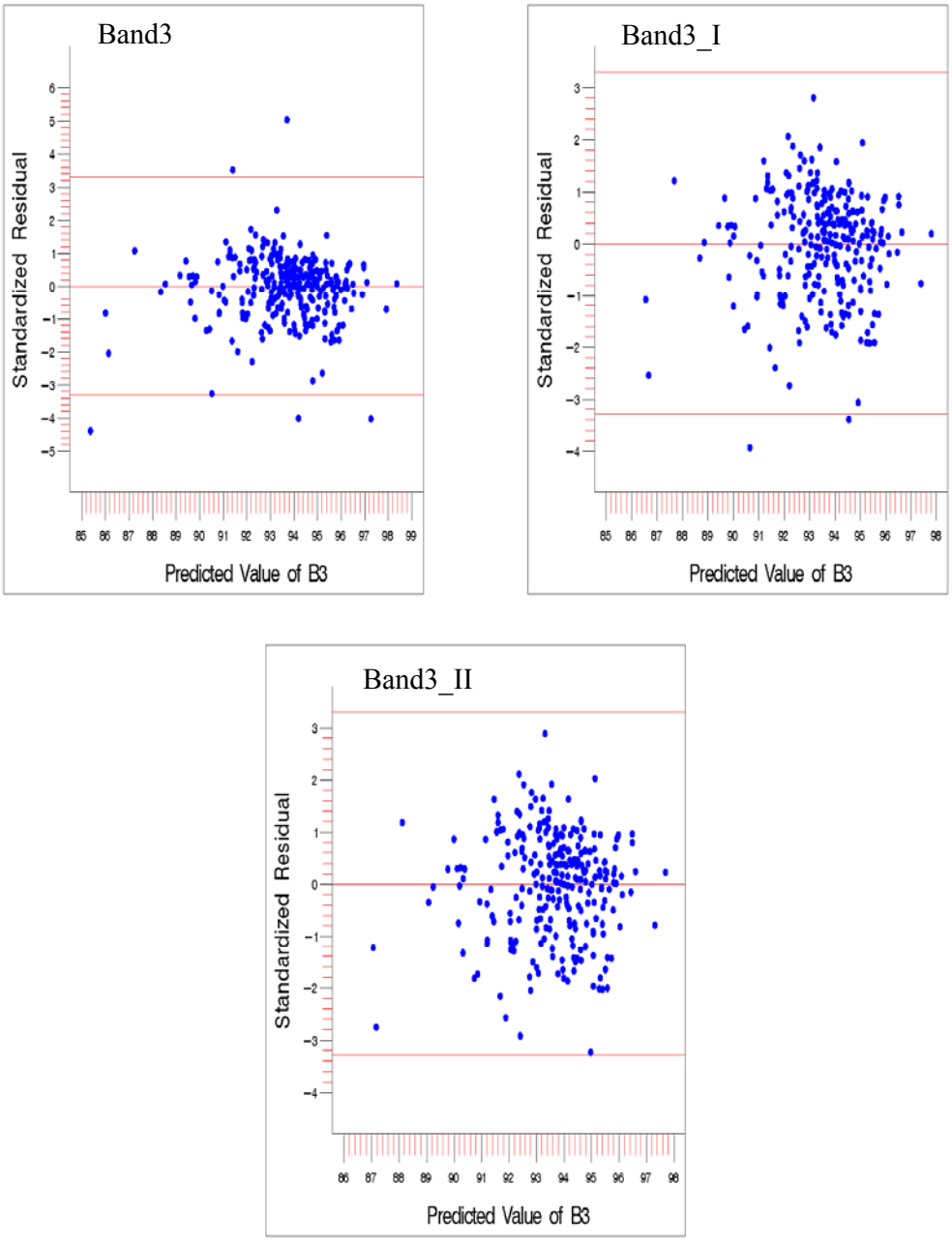


Figure 96: Outlier removal iteration (I, II …) plots of standardized residual vs. predicted value of dependent variable for band 3 of 1997 field-1 data (WSLR)

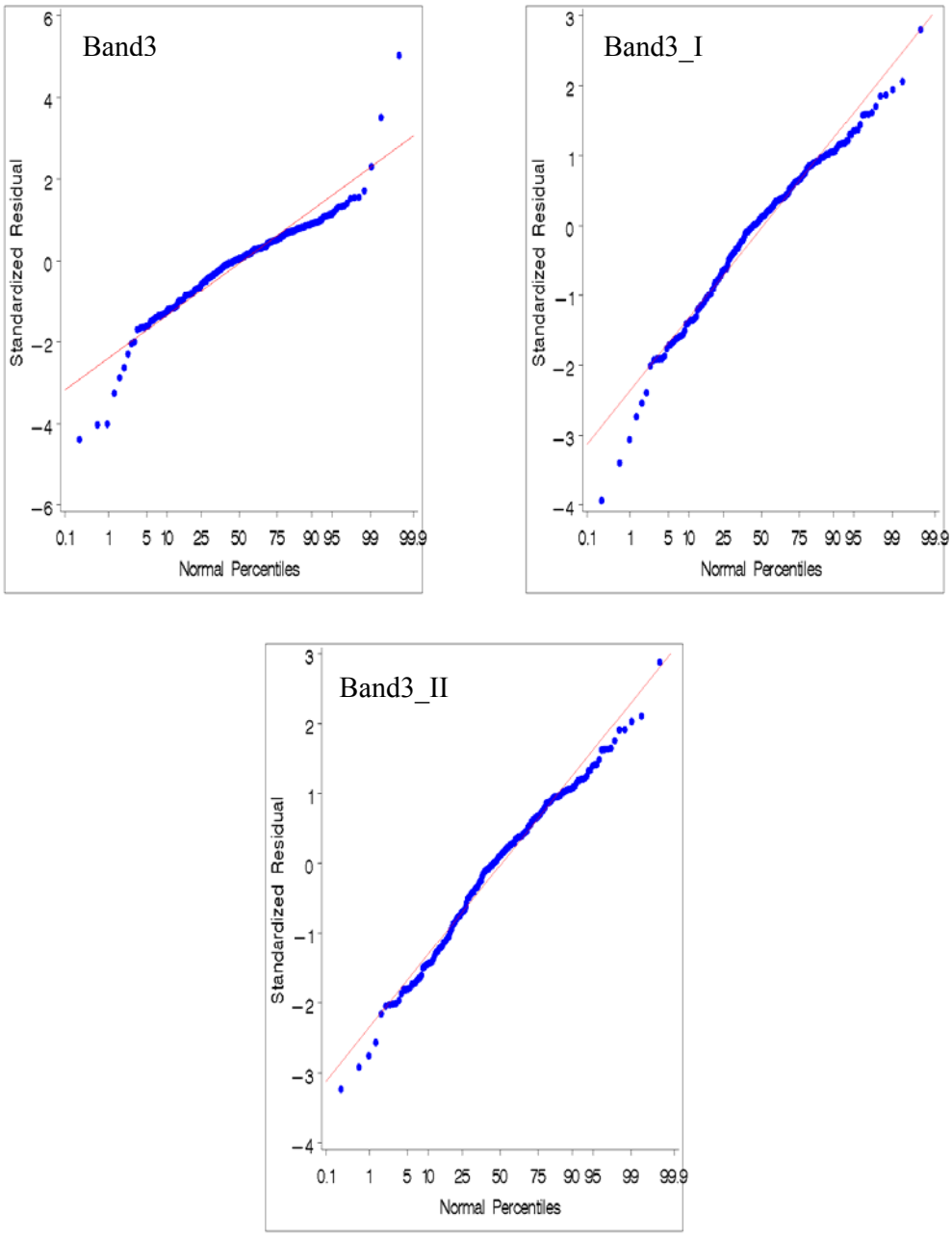


Figure 97: Normal probability plots corresponding to outlier removal iteration (I, II …) plots for band 3 of 1997 field-1 data (WSLR)

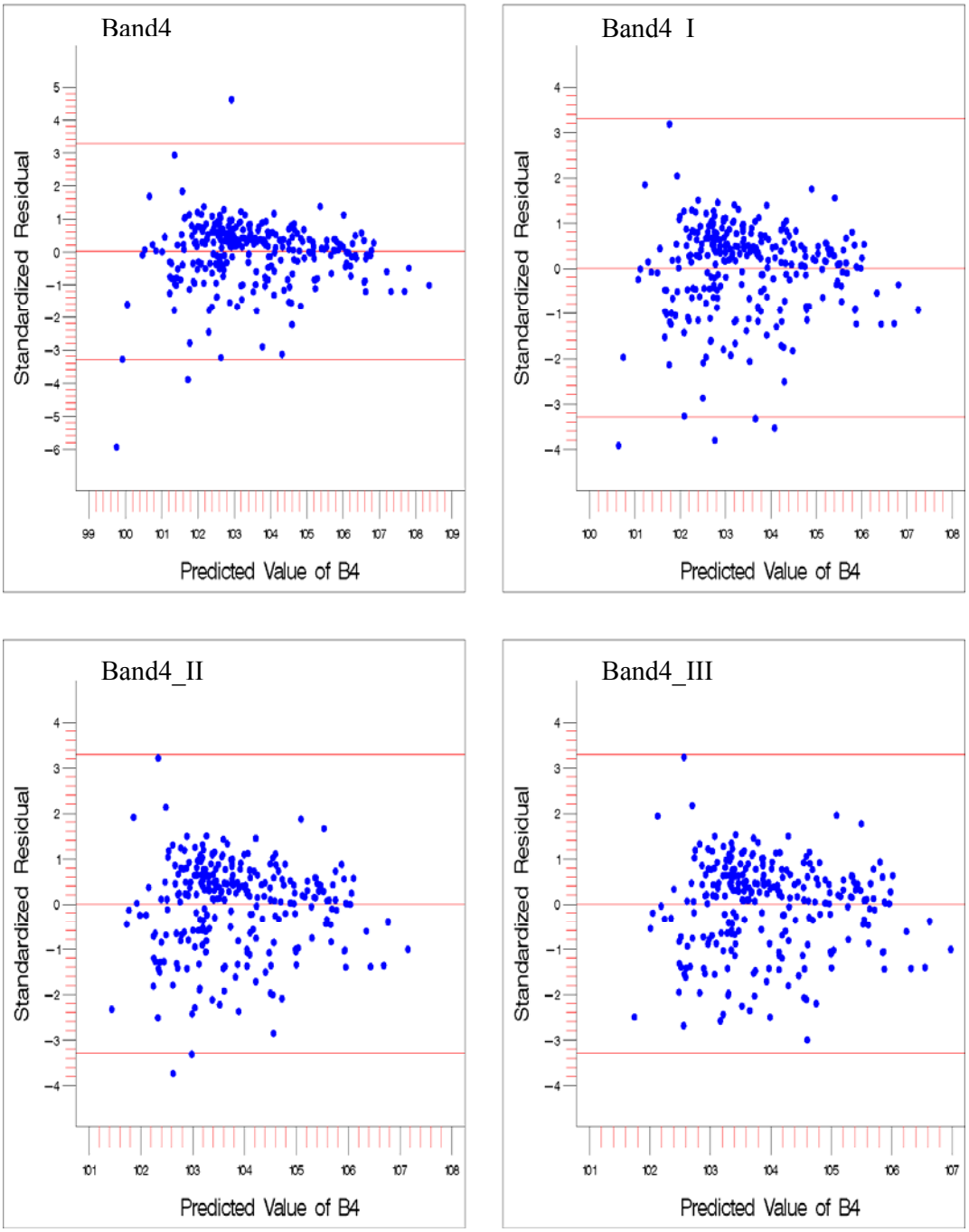


Figure 98: Outlier removal iteration (I, II …) plots of standardized residual vs. predicted value of dependent variable for band 4 of 1997 field-1 data (WSLR)

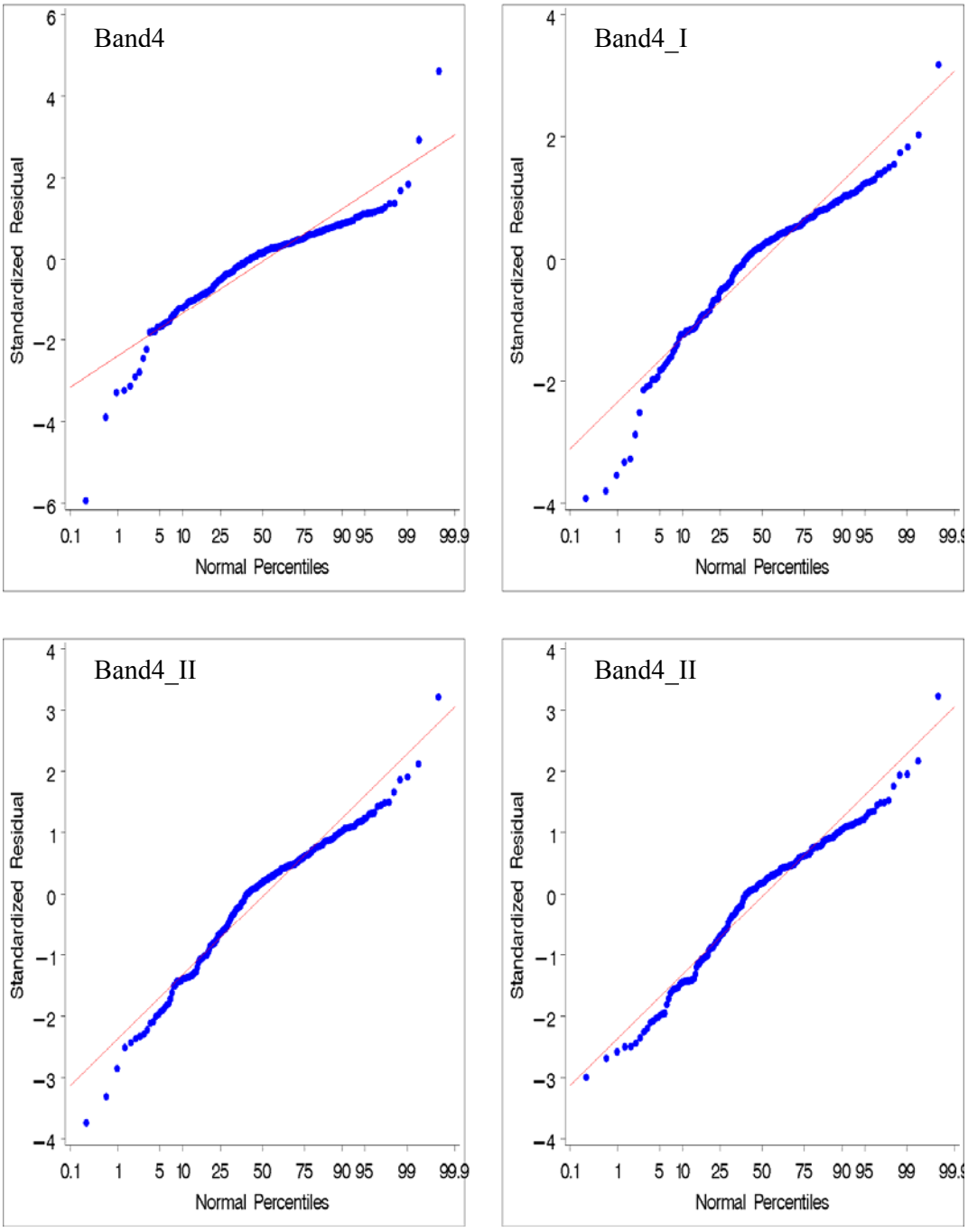


Figure 99: Normal probability plots corresponding to outlier removal iteration (I, II …) plots for band 4 of 1997 field-1 data (WSLR)

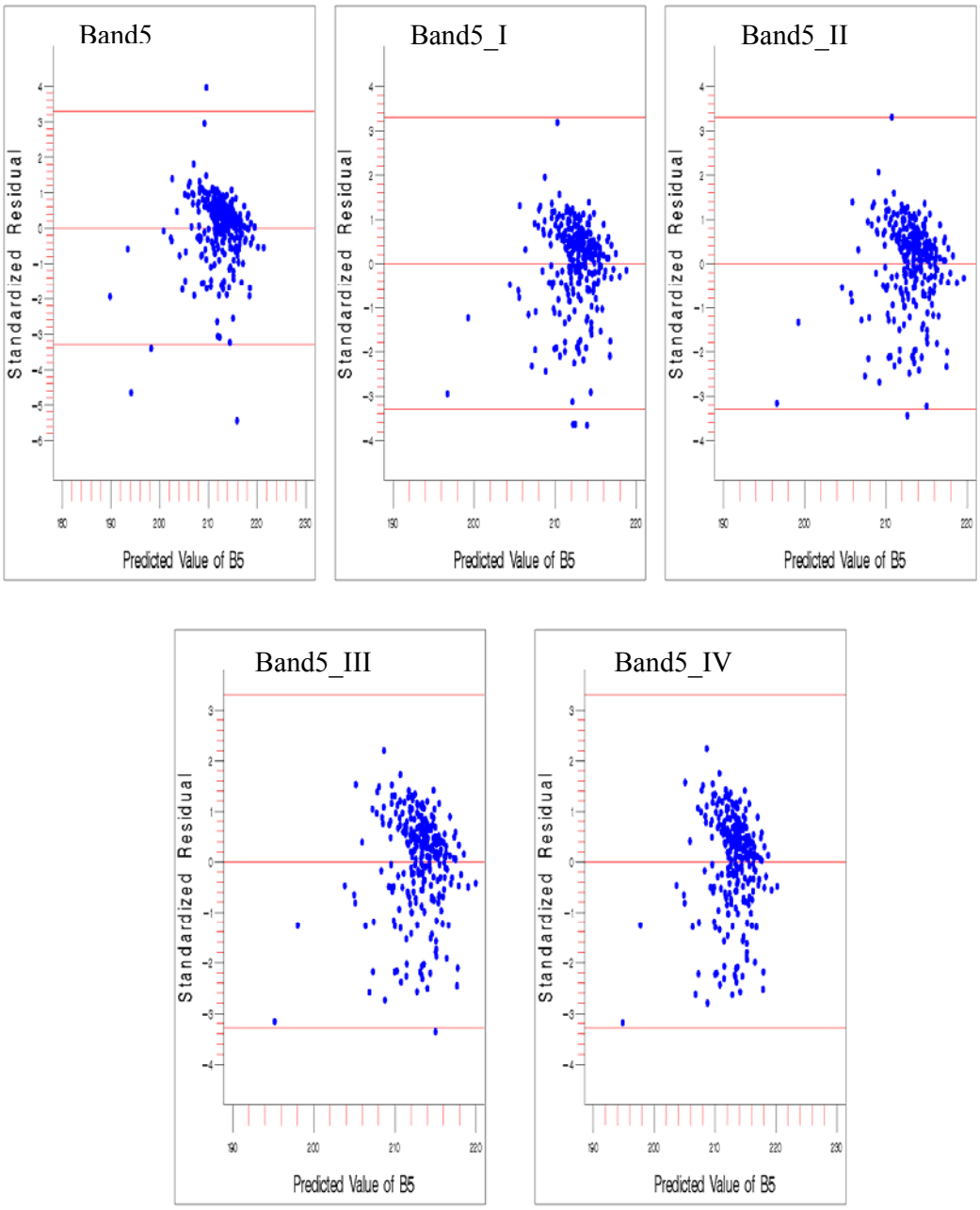


Figure 100: Outlier removal iteration (I, II …) plots of standardized residual vs. predicted value of dependent variable for band 5 of 1997 field-1 data (WSLR)

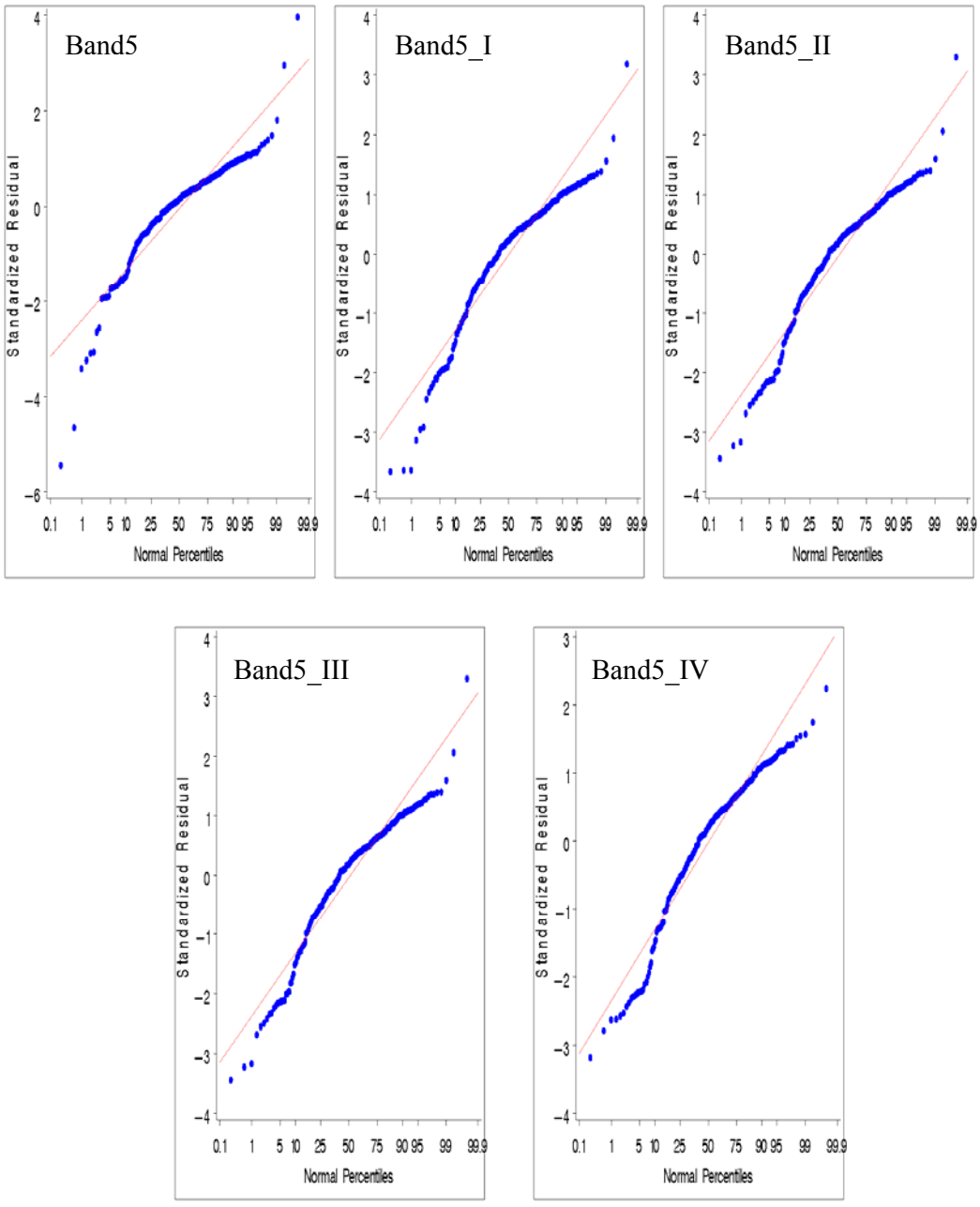


Figure 101: Normal probability plots corresponding to outlier removal iteration (I, II …) plots for band 5 of 1997 field-1 data (WSLR)

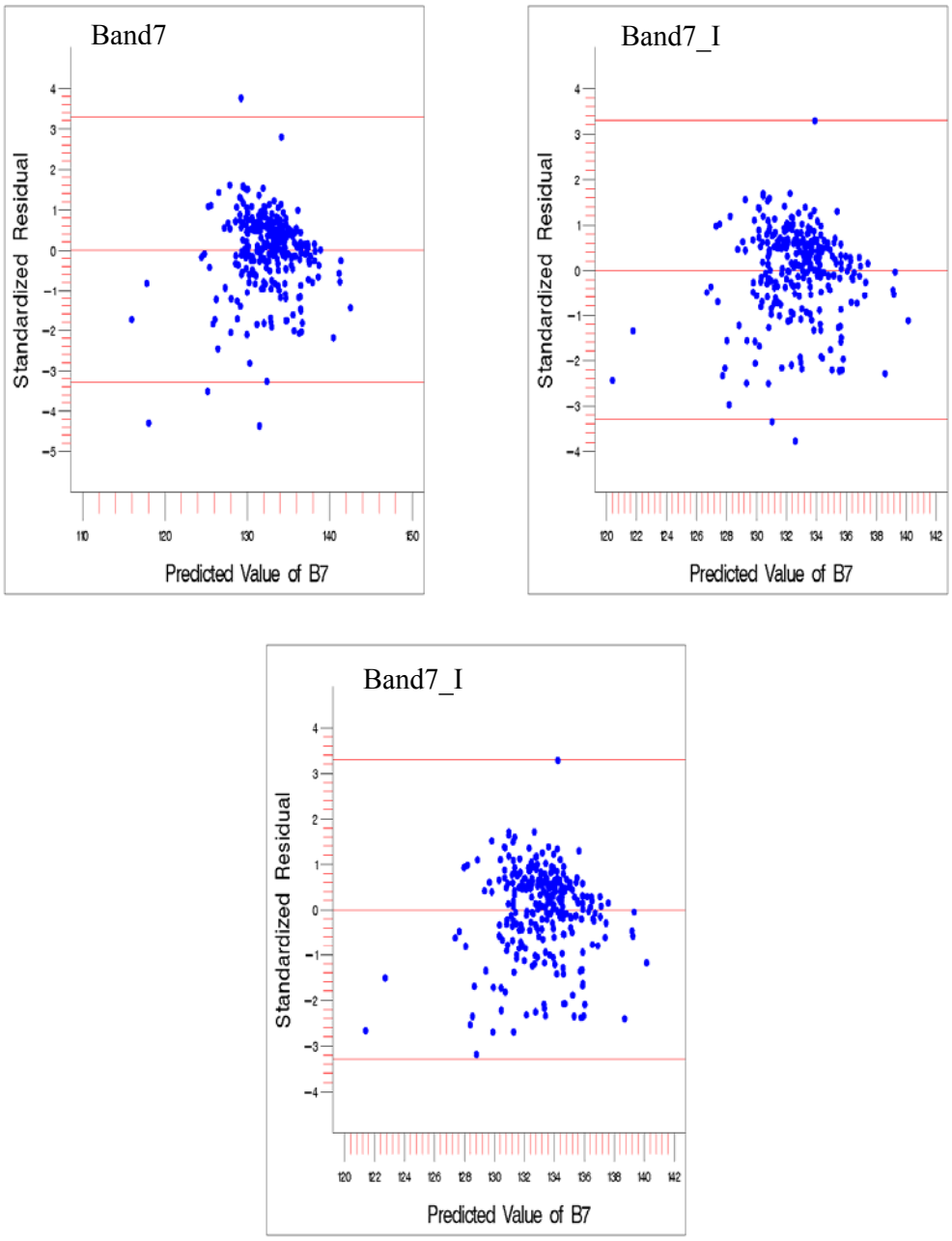


Figure 102: Outlier removal iteration (I, II …) plots of standardized residual vs. predicted value of dependent variable for band 7 of 1997 field-1 data (WSLR)

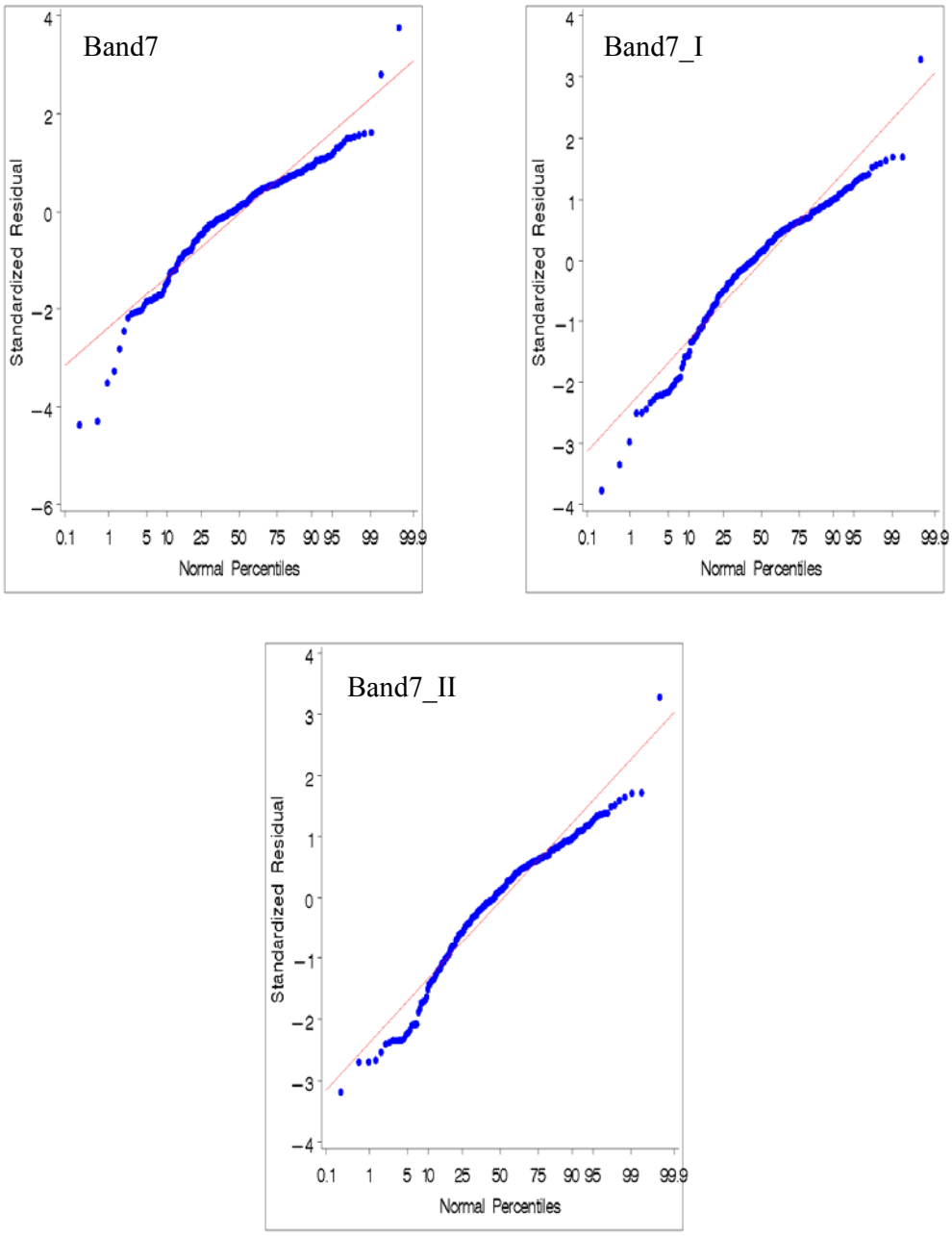


Figure 103: Normal probability plots corresponding to outlier removal iteration (I, II …) plots for band 7 of 1997 field-1 data (WSLR)

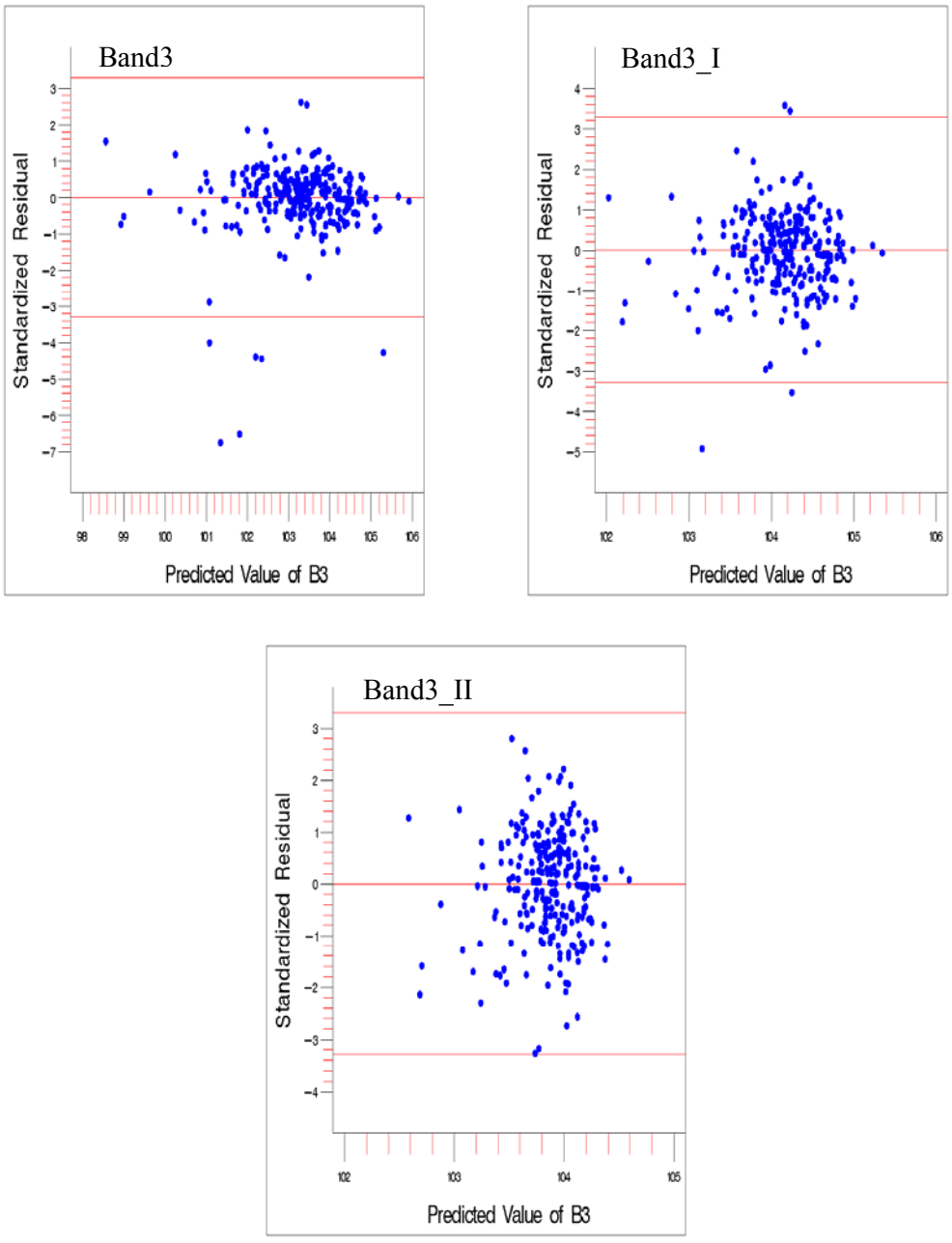


Figure 104: Outlier removal iteration plots of standardized residual vs. predicted value of dependent variable for band 3 of 2001 field-1 data (WSLR)

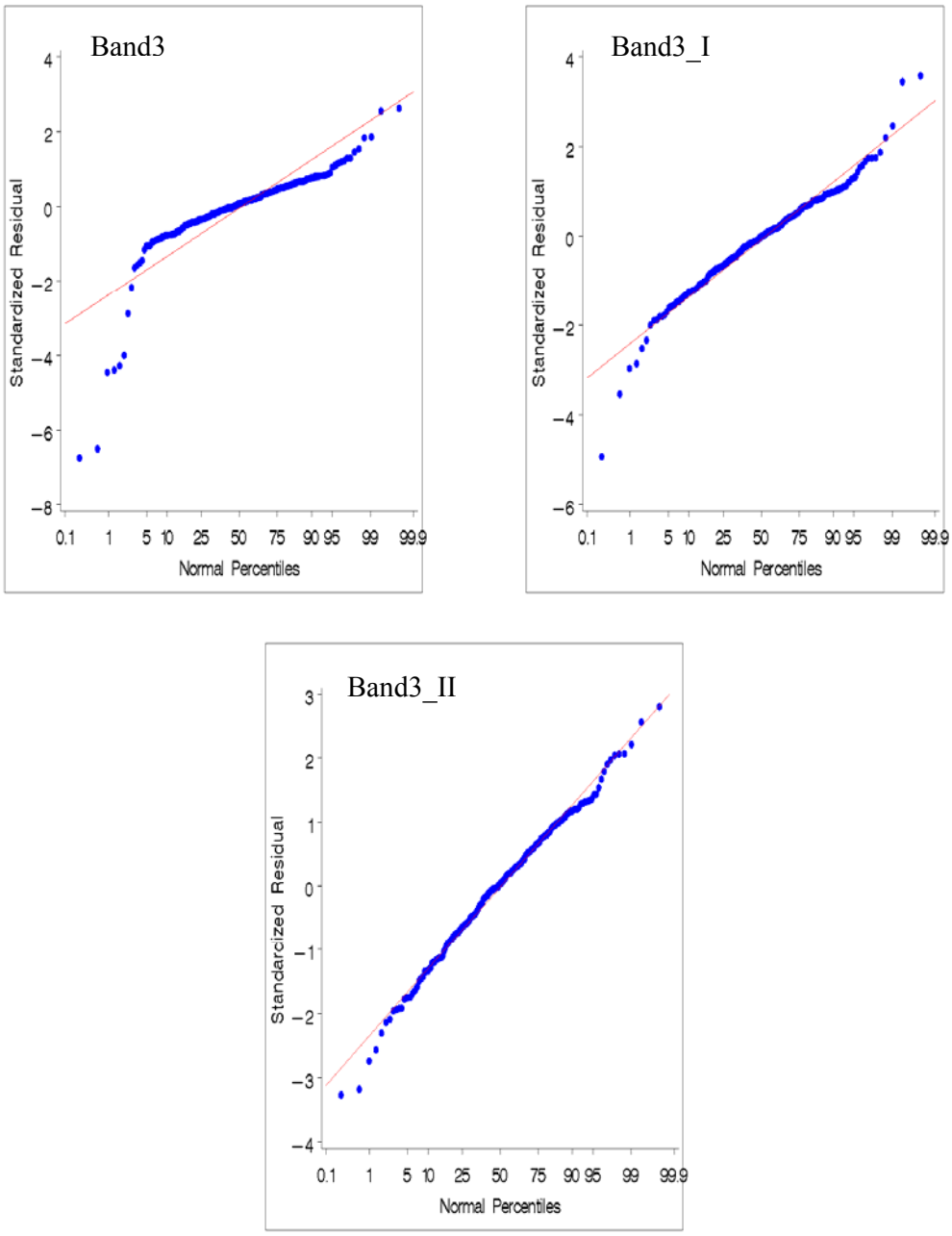


Figure 105: Normal probability plots corresponding to outlier removal iteration (I, II …) plots for band 3 of 2001 field-1 data (WSLR)

The R^2 values for bands 1 through 3 (Figure 106) are as follows: no improvement occurred after removal of outliers in band 1, a slight increase occurred for band 2, and a slight decrease occurred for band 3. However, after the first iteration, the R^2 value for both bands 2 and 3 remained constant. Figure 107 suggests that with the removal of outliers a decrease in R^2 is observed in bands 4 through 7. In band 7, R^2 remained constant after the first iteration. However, in band-5 after the first iteration, R^2 gradually increased at each iteration, but it remained low compared to the model including outliers. The significance levels of the models remained high ($p<0.0001$) for all the studied bands. The results for band 3 of 2001 field-1 data indicate that after outlier removal the R^2 value decreased (Figure 108). Furthermore, contrary to the results with 1997 field-1 data, a decrease in the p value was observed with band 3 of 2001 field-1 data, from significant (0.01) to non-significant at each iteration (0.10, 0.26).

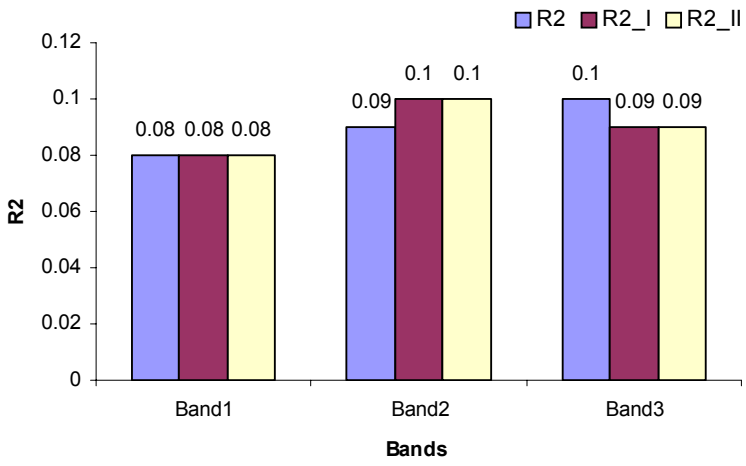


Figure 106: Changes in R^2 values with outlier removal at each iterations (I, II.. ..) for bands 1 through 3 of 1997 field-1 data (WSLR)

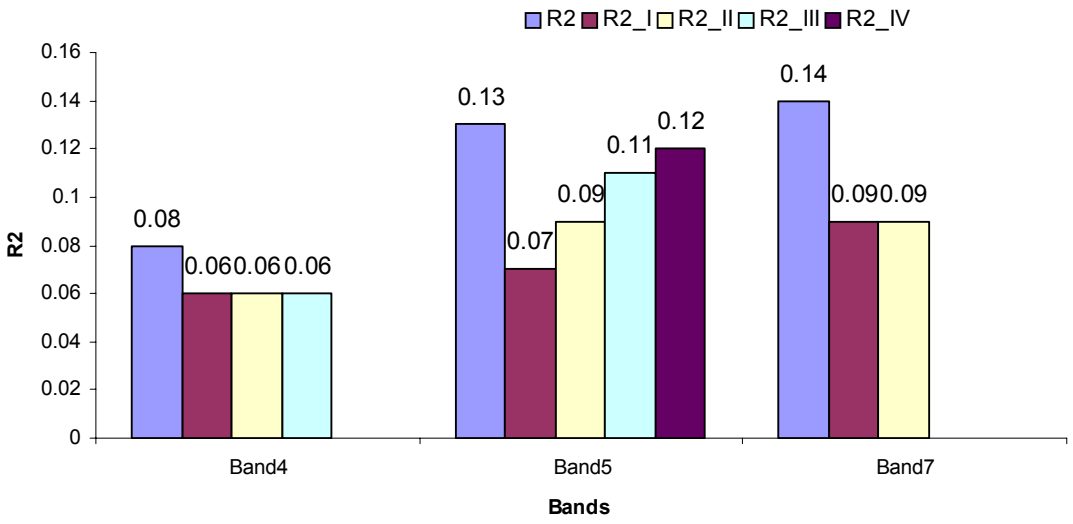


Figure 107: Changes in R^2 values with outlier removal at each iterations (I, II.. ..) for bands 4 through 7 of 1997 field-1 data (WSLR)

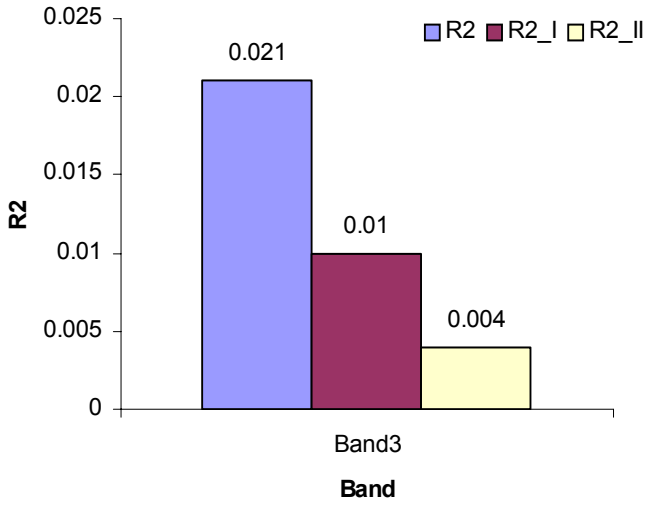


Figure 108: Changes in R^2 values with outlier removal at each iterations (I, II.. ..) for band 3 of 2001 field-1 data (WSLR)

Field-3

The outlier removal process for 1999 and 2001 field-3 data is illustrated in Figures 109, 111, 113, 115, and 117. The normal probability plots of the outlier removal process for 1999 data are provided as Figures 110 and 112, and for 2001 data are provided as Figures 114, 116, and 118. The constant variance plots for band 1 of 1999 field-3 data, which required three iterations to remove all the detected outliers, indicate that after removal the overall pattern of distribution remained essentially the same (Figure 109). The corresponding normality plots illustrate that data points lying far off the reference are no longer in evidence after removal of outliers, and overall the curve seems closer to the reference line (Figure 110). Similar observations were noted for bands 2 and 3 (Figures 111 and 112).

The outlier removal process for bands 1 and 2 of 2001 field-3 data in, illustrated in Figure 113, revealed that after outlier removal the constant variance assumption appeared to be improved. The corresponding normality plots also improved, and the data curve without the outliers appears to be well aligned with the reference line (Figure 114).

Bands 3 and 4, which needed one and two iterations to remove outliers, respectively, produced similar results regarding constant variance and normality plots to those of bands 1 and 2 (Figures 115 and 116). Similar observations were noted for bands 5 and 7 as well, which needed one and three iterations, respectively (Figures 117 and 118).

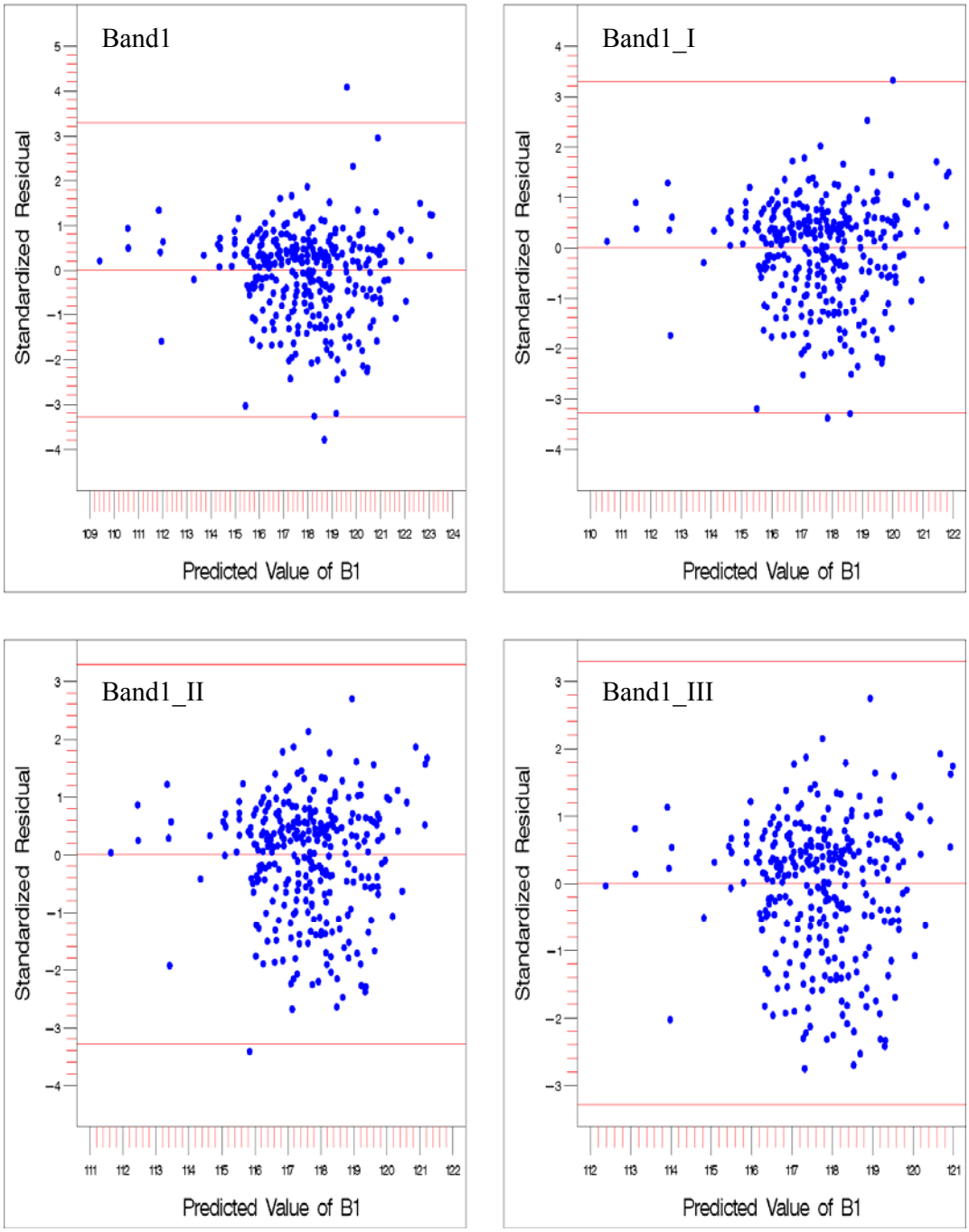


Figure 109: Outlier removal iteration plots of standardized residual vs. predicted value of dependent variable for band 1 of 1999 field-3 data (WSLR)

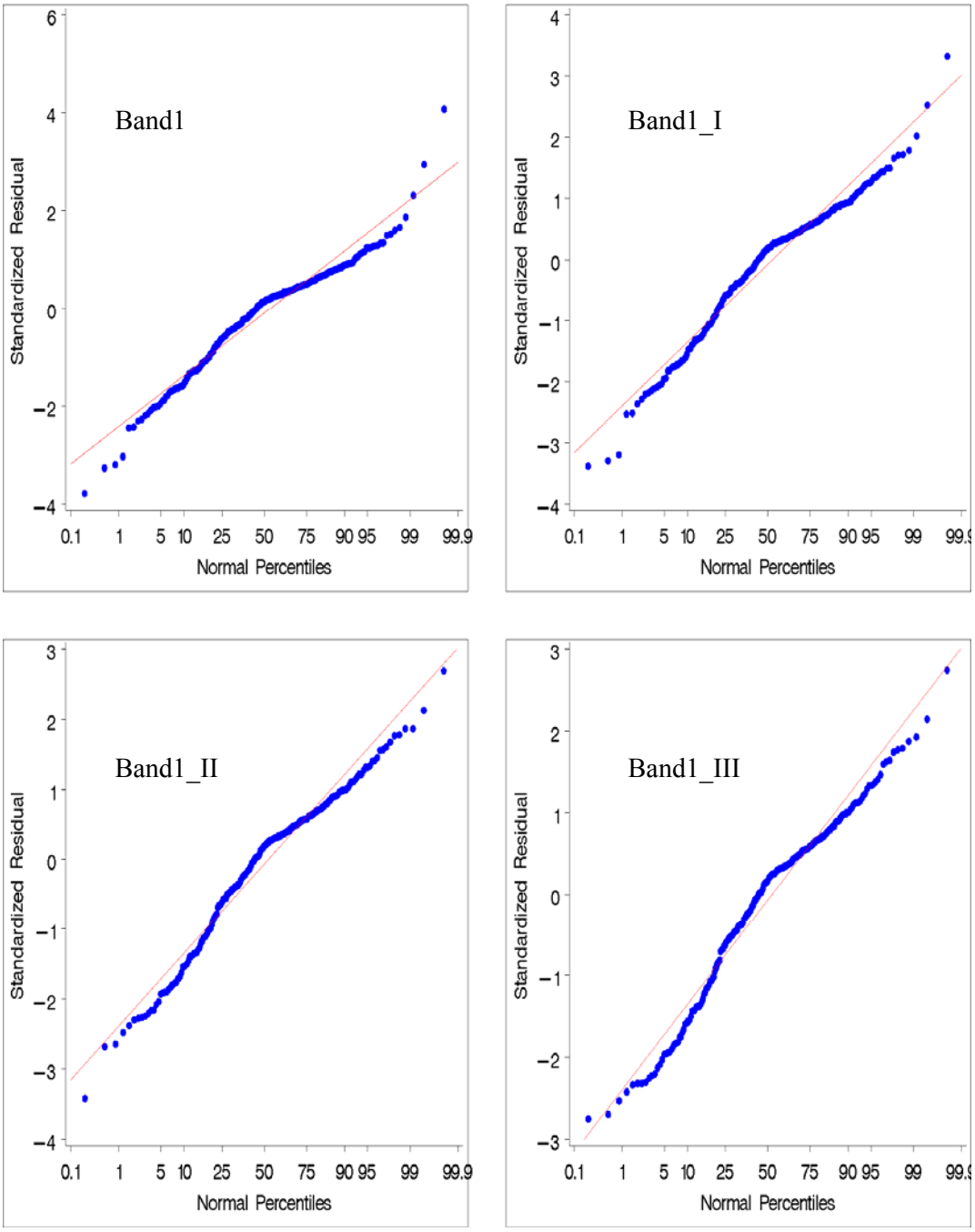


Figure 110: Normal probability plots corresponding to outlier removal iterations (I, II …) plots for band 1 of 1999 field-3 data (WSLR)

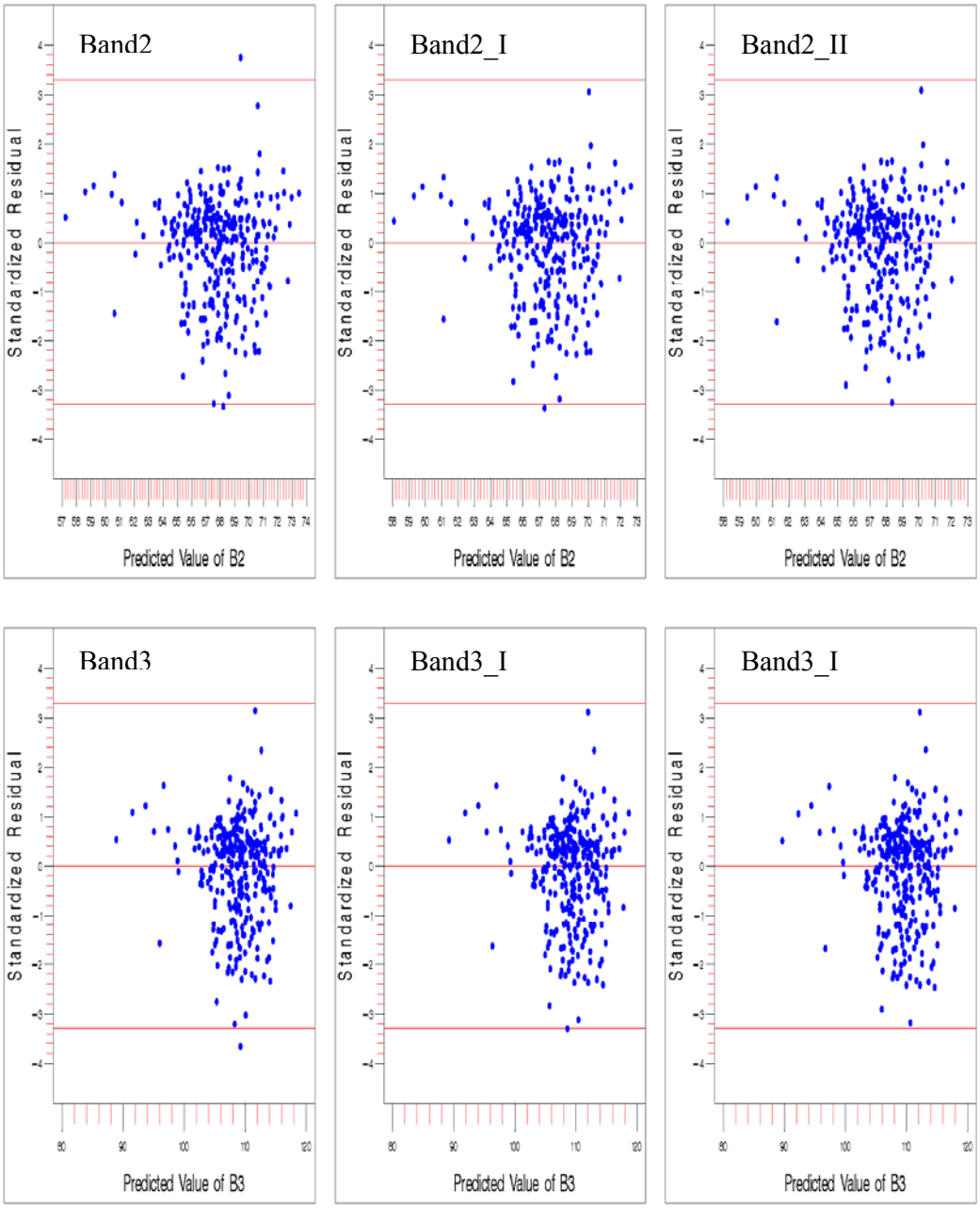


Figure 111: Outlier removal iteration plots of standardized residual vs. predicted value of dependent variable for bands 2 and 3 of 1999 field-3 data (WSLR)

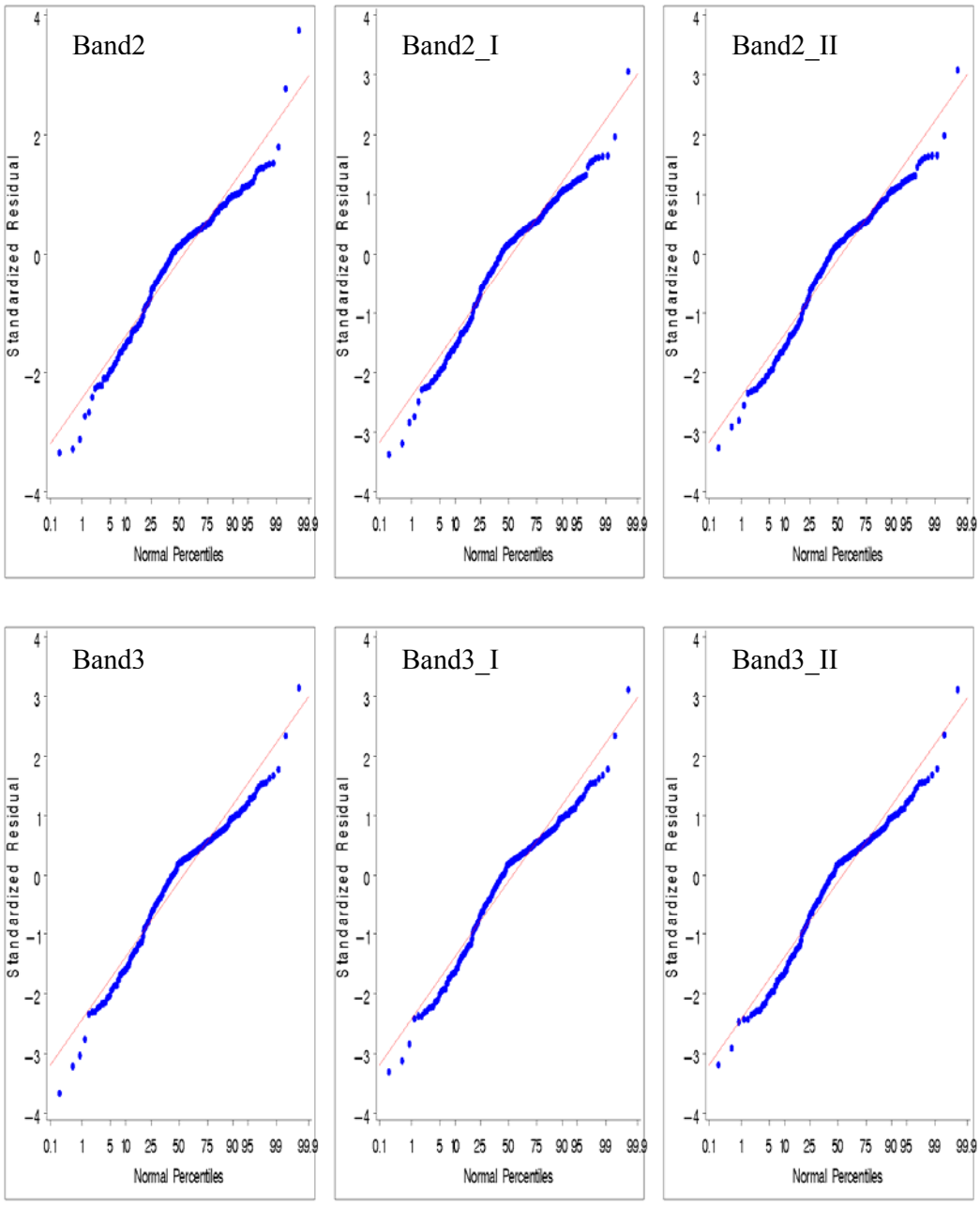


Figure 112: Normal probability plots corresponding to outlier removal iterations (I, II …) plots for bands 2 and 3 of 1999 field-3 data (WSLR)

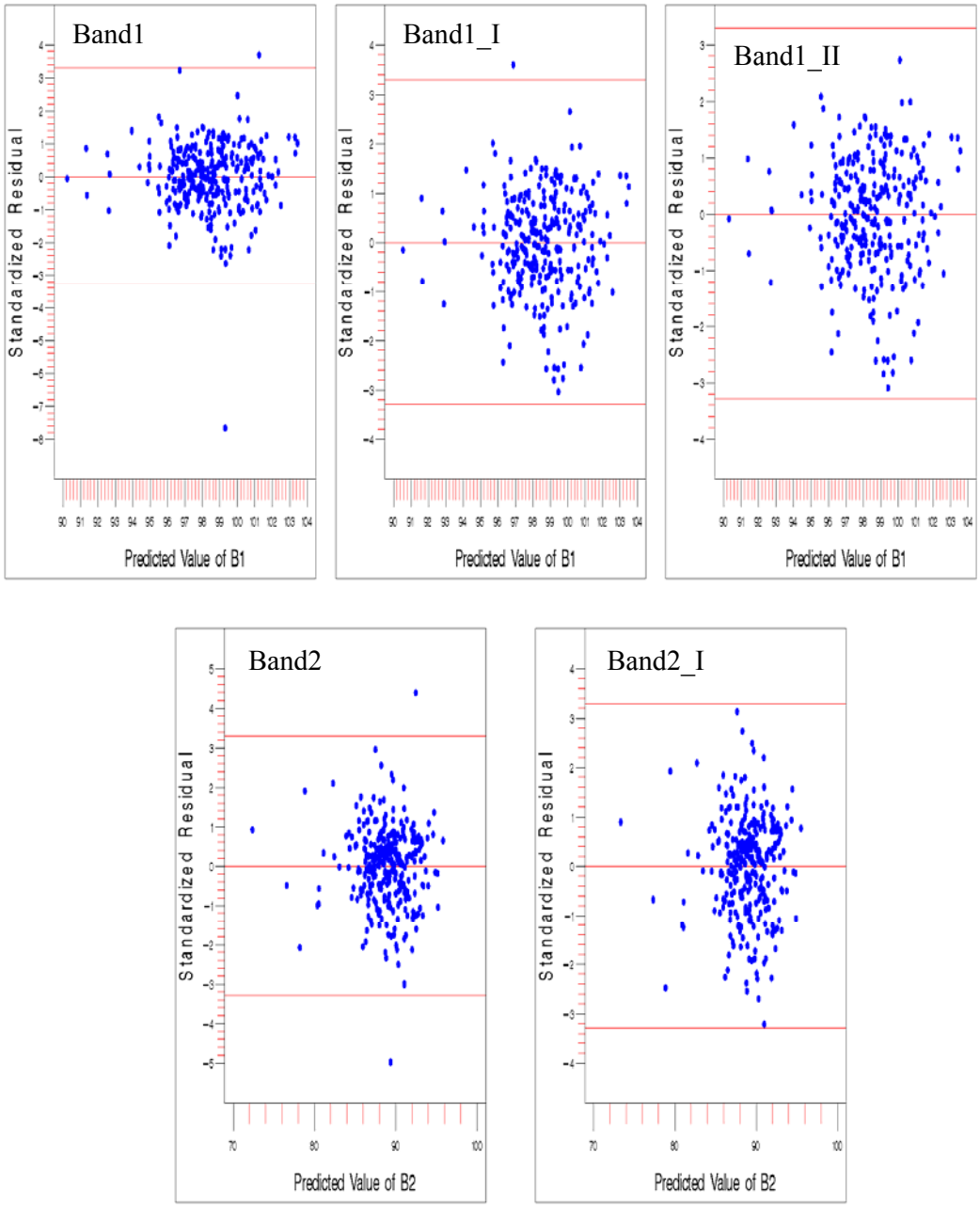


Figure 113: Outlier removal iteration plots of standardized residual vs. predicted value of dependent variable for bands 1 and 2 of 2001 field-3 data (WSLR)

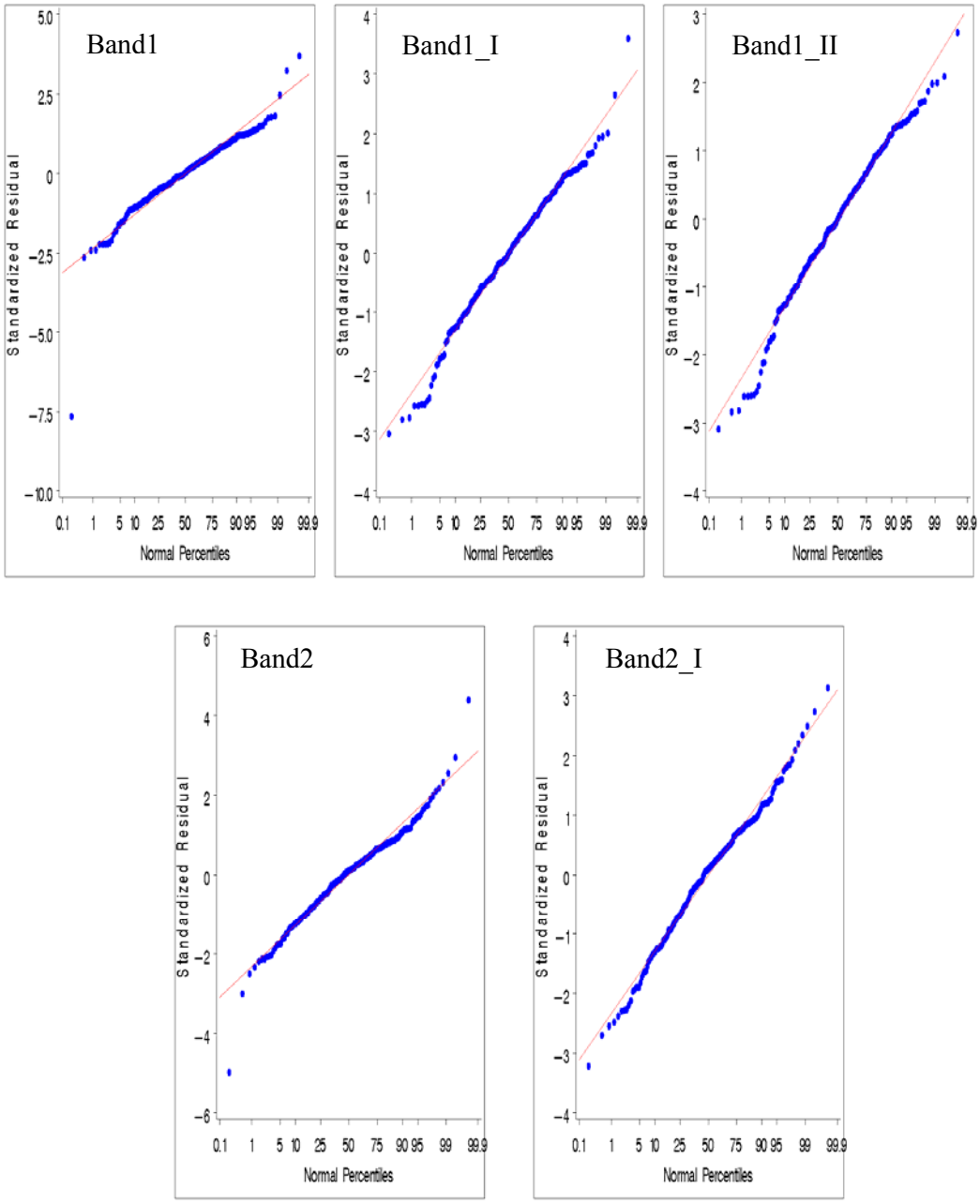


Figure 114: Normal probability plots corresponding to outlier removal iterations (I, II …) plots for bands 1 and 2 of 2001 field-3 data (WSLR)

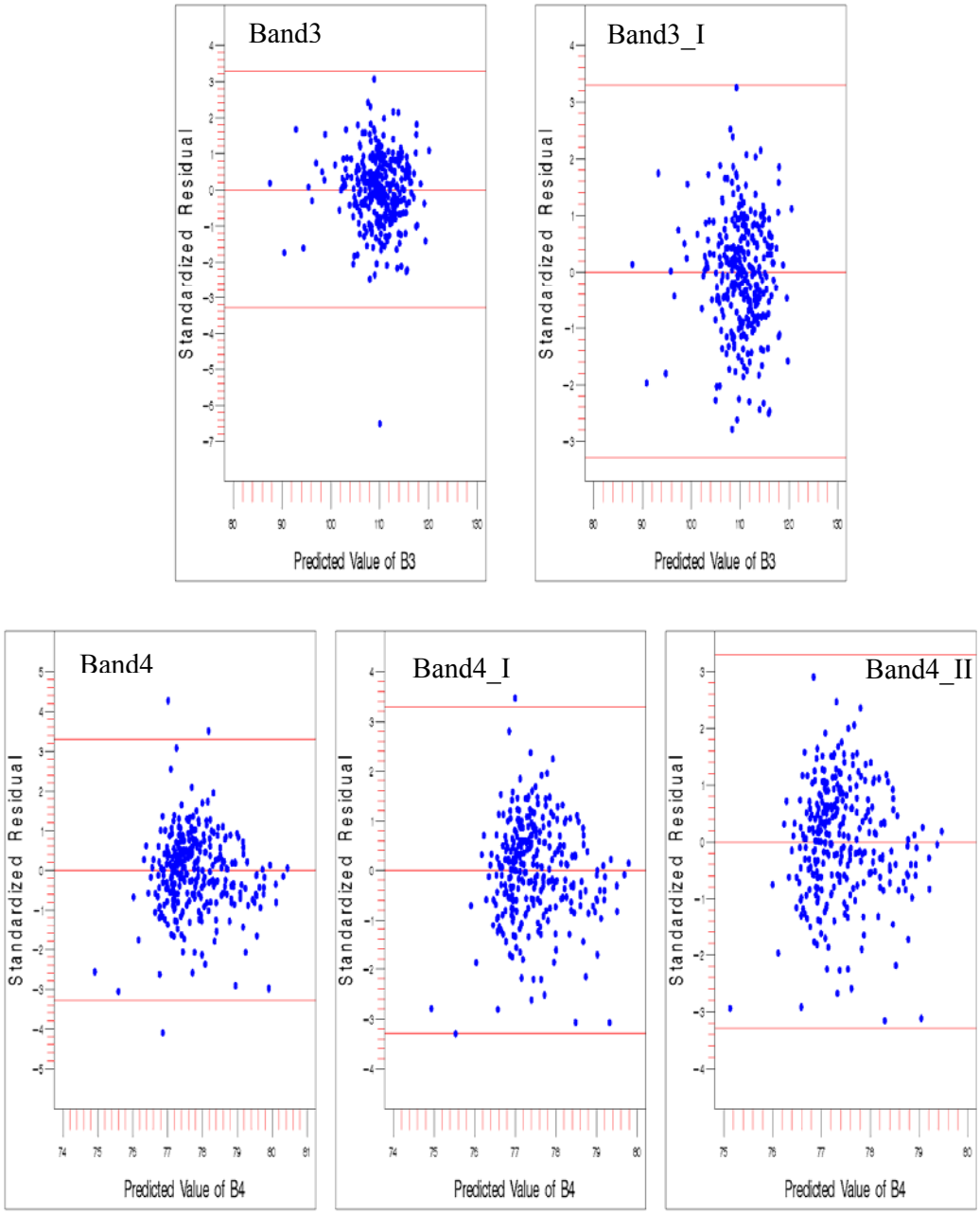


Figure 115: Outlier removal iteration plots of standardized residual vs. predicted value of dependent variable for bands 3 and 4 of 2001 field-3 data (WSLR)

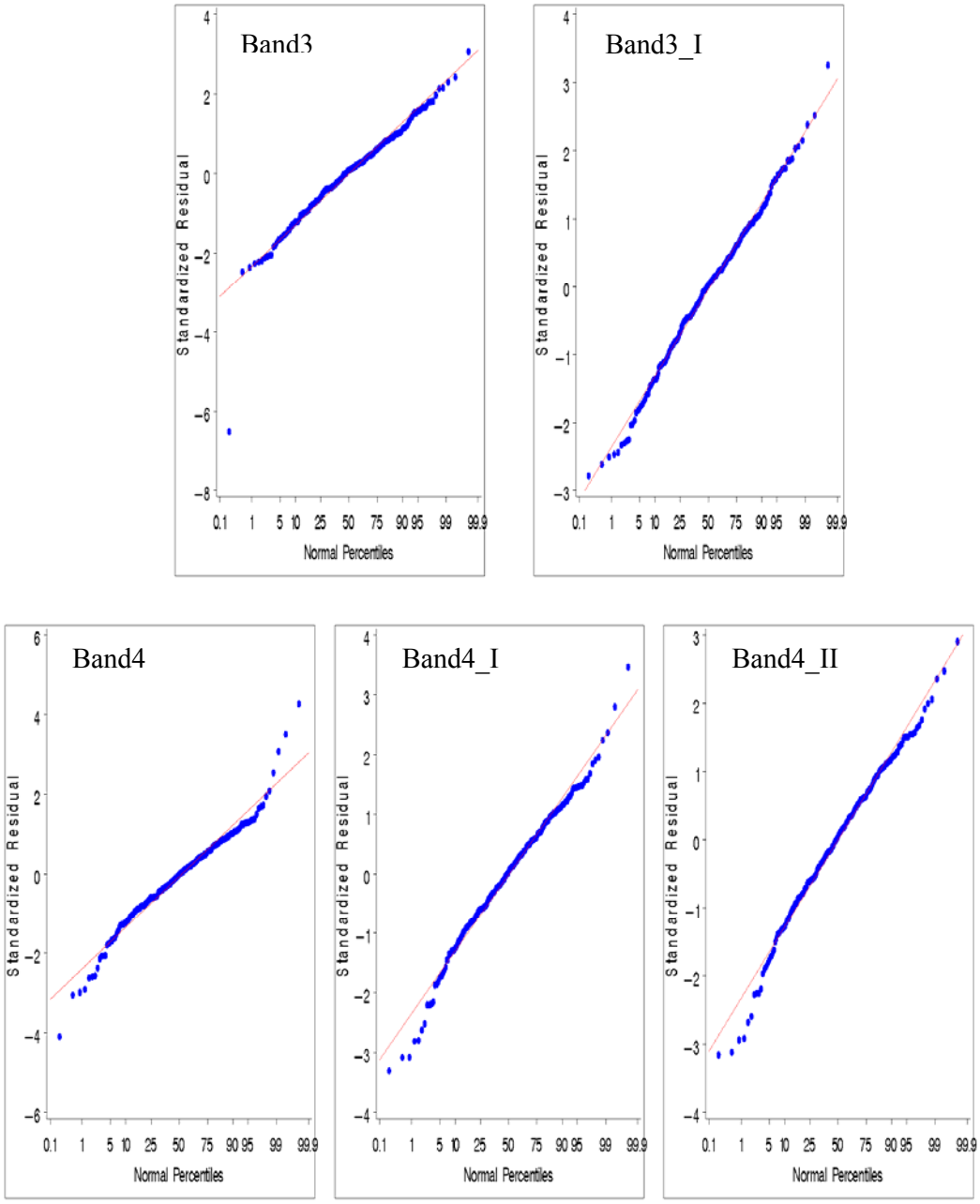


Figure 116: Normal probability plots corresponding to outlier removal iterations (I, II …) plots for bands 3 and 4 of 2001 field-3 data (WSLR)

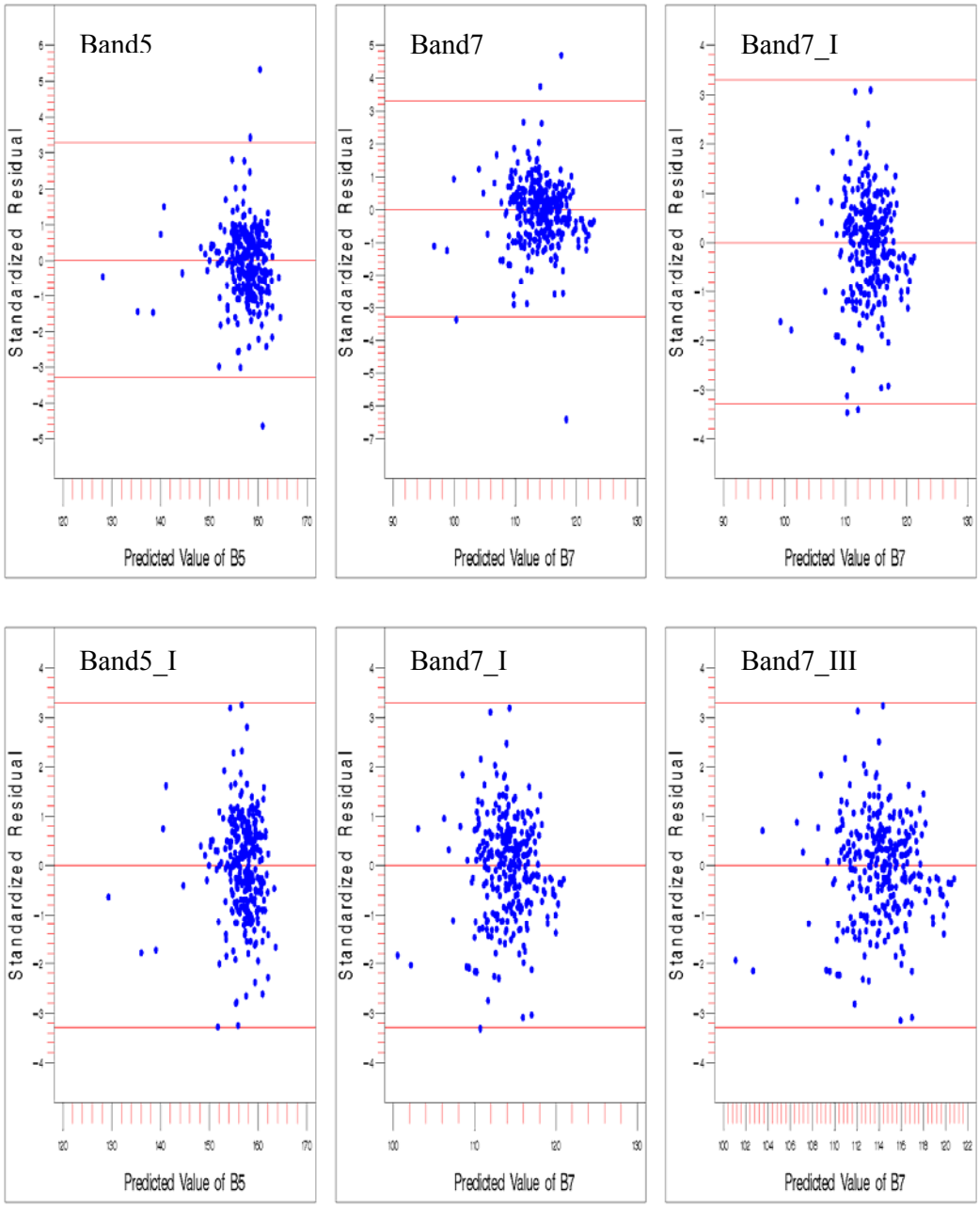


Figure 117: Outlier removal iteration plots of standardized residual vs. predicted value of dependent variable for bands 5 and 7 of 2001 field-3 data (WSLR)

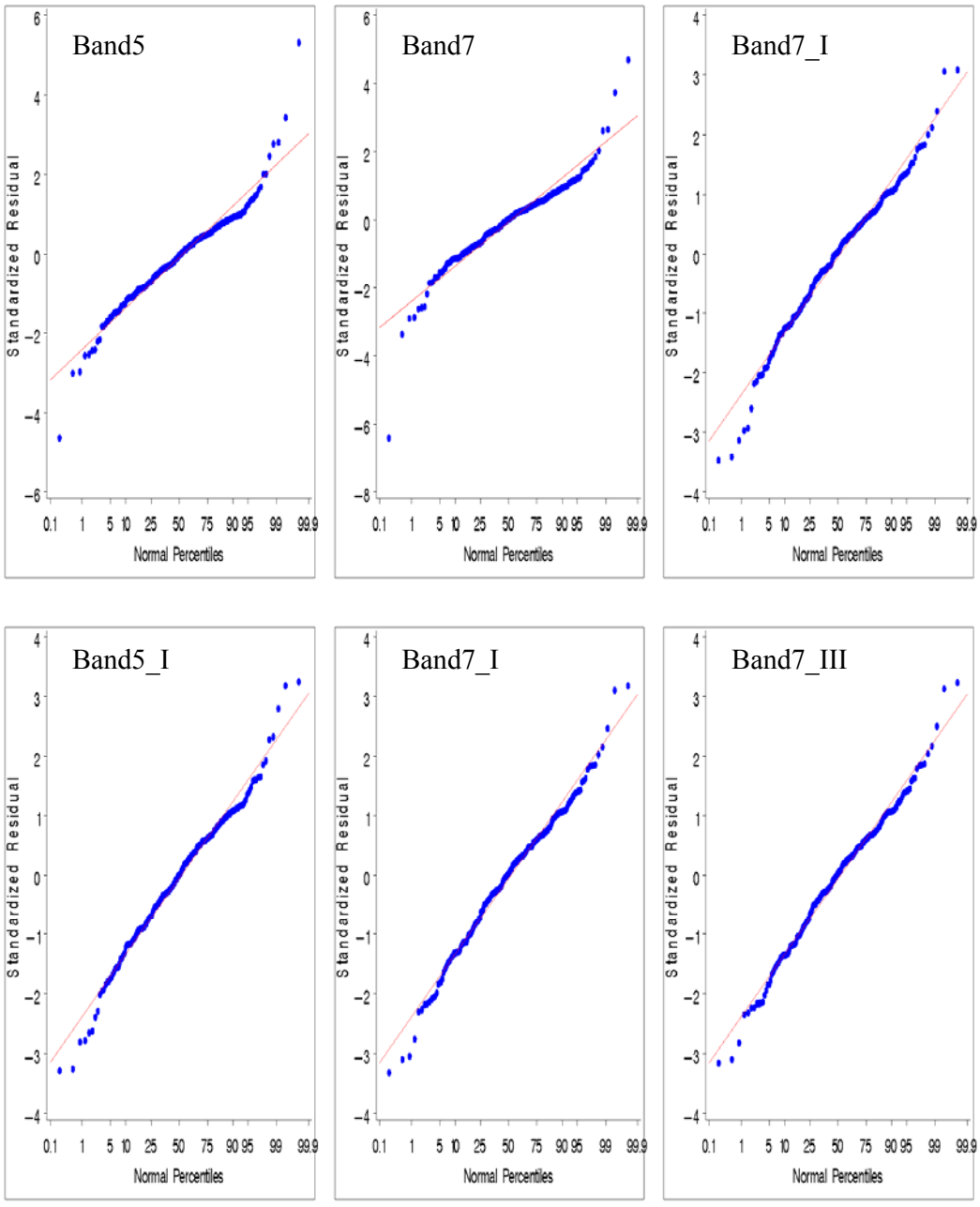


Figure 118: Normal probability plots corresponding to outlier removal iterations (I, II …) plots for bands 5 and 7 of 2001 field-3 data (WSLR)

Figure 119 illustrates the changes in R^2 values for 1999 field-3 data. It was observed that for bands 1 and 2, the R^2 values decreased after removal of outliers. With band 3, the R^2 value did not change. The p values for band 1 changed from 0.0005 to 0.0029 to 0.0087 to 0.016. While a decrease in significance was observed, the band-1 model remained significant. For bands 2 and 3, the p value did not change appreciably ($p < 0.0001$). The changes in R^2 values for 2001 field-3 data are illustrated in Figures 120, 121, and 122. It was observed that the R^2 value for bands 1 and 2 improved from 0.24 to 0.29 and 0.28 to 0.29, respectively, after removal of outliers (Figure 120). The R^2 value for band 3 increased from 0.32 to 0.35 after removal of outliers, whereas for band 4 it decreased from 0.02 to 0.01 (Figure 121). In the case of band 5, R^2 increased from 0.19 to 0.20, and for band 7, it decreased from 0.15 to 0.11 (Figure 122). The significance levels for bands 1, 2, 3, 5 and 7 remained high ($p < 0.0001$) after removal of outliers. However, even though the model for band 4 remained significant, the p value decreased from 0.008 to 0.02.

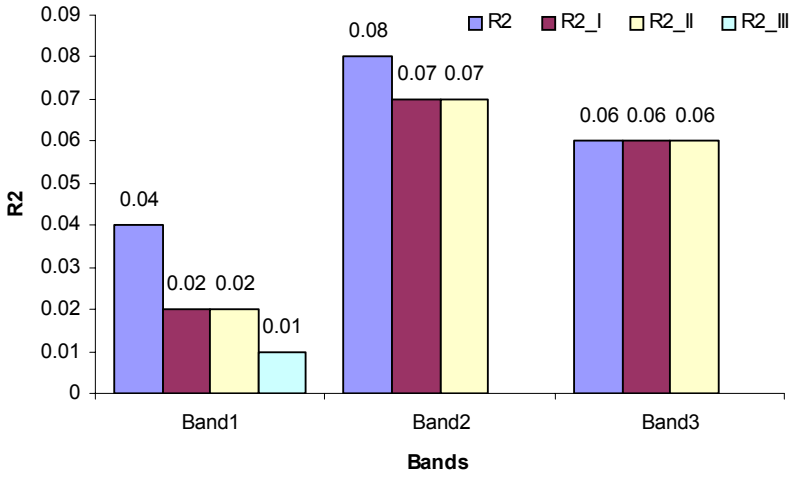


Figure 119: Changes in R^2 values with outlier removal at each iterations (I, II.. ..) for bands 1 through 3 of 1999 field-3 data (WSLR)

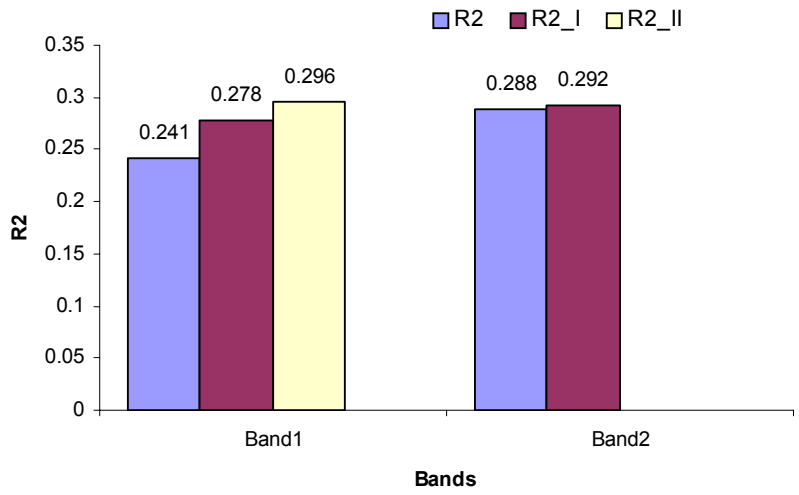


Figure 120: Changes in R^2 values with outlier removal at each iterations (I, II.. ..) for bands 1 and 2 of 2001 field-3 data (WSLR)

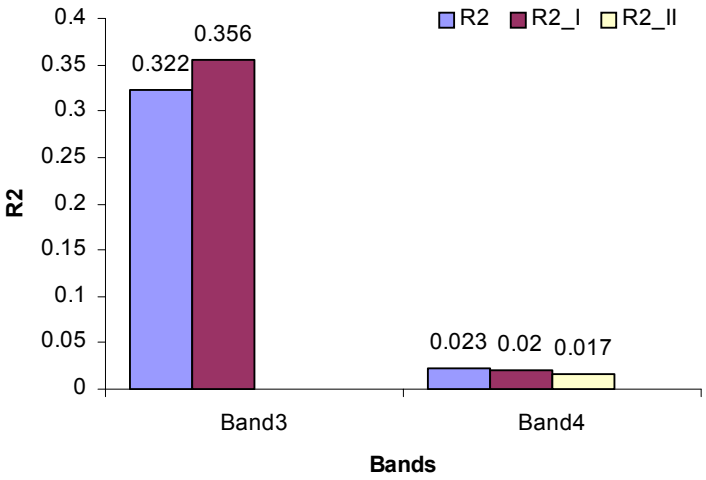


Figure 121: Changes in R^2 values with outlier removal at each iterations (I, II.. ..) for bands 3 and 4 of 2001 field-3 data (WSLR)

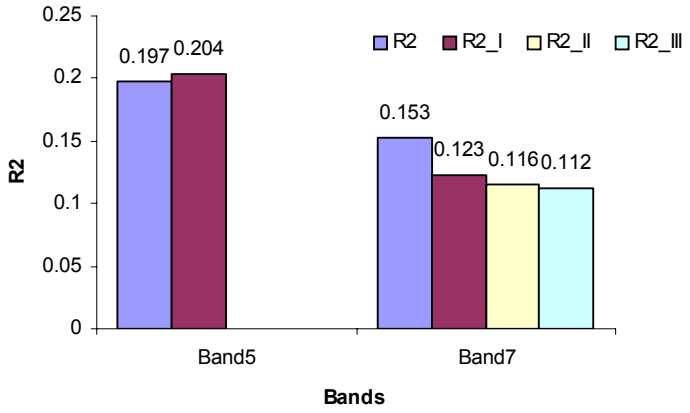


Figure 122: Changes in R^2 values with outlier removal at each iterations (I, II.. ..) for bands 5 and 7 of 2001 field-3 data (WSLR)

WMLR

Field-1

The outlier removal process for WMLR analysis of 1997 field-1 data is illustrated in Figures 123, 125, and 127. The corresponding normality plots with outlier removal iterations are provided as Figures 124, 126, and 128. The notations describing the iterations are similar to those used in earlier sections. The constant variance plots for bands 1 and 2 (Figure 123) suggest that after outliers removal the distribution of residuals about the reference line (standardized residual = 0) is improved. Both bands required two iterations to remove all the detected outliers. The corresponding normality plots in Figure 124 also appear to be improved, as the data points lying far from the reference line are no longer in evidence, and the curve is well aligned to the reference line. Similar observations were noted for bands 3, 4, 5, and 7, which required one, two, one, and one iterations, respectively, to remove all the detected outliers.

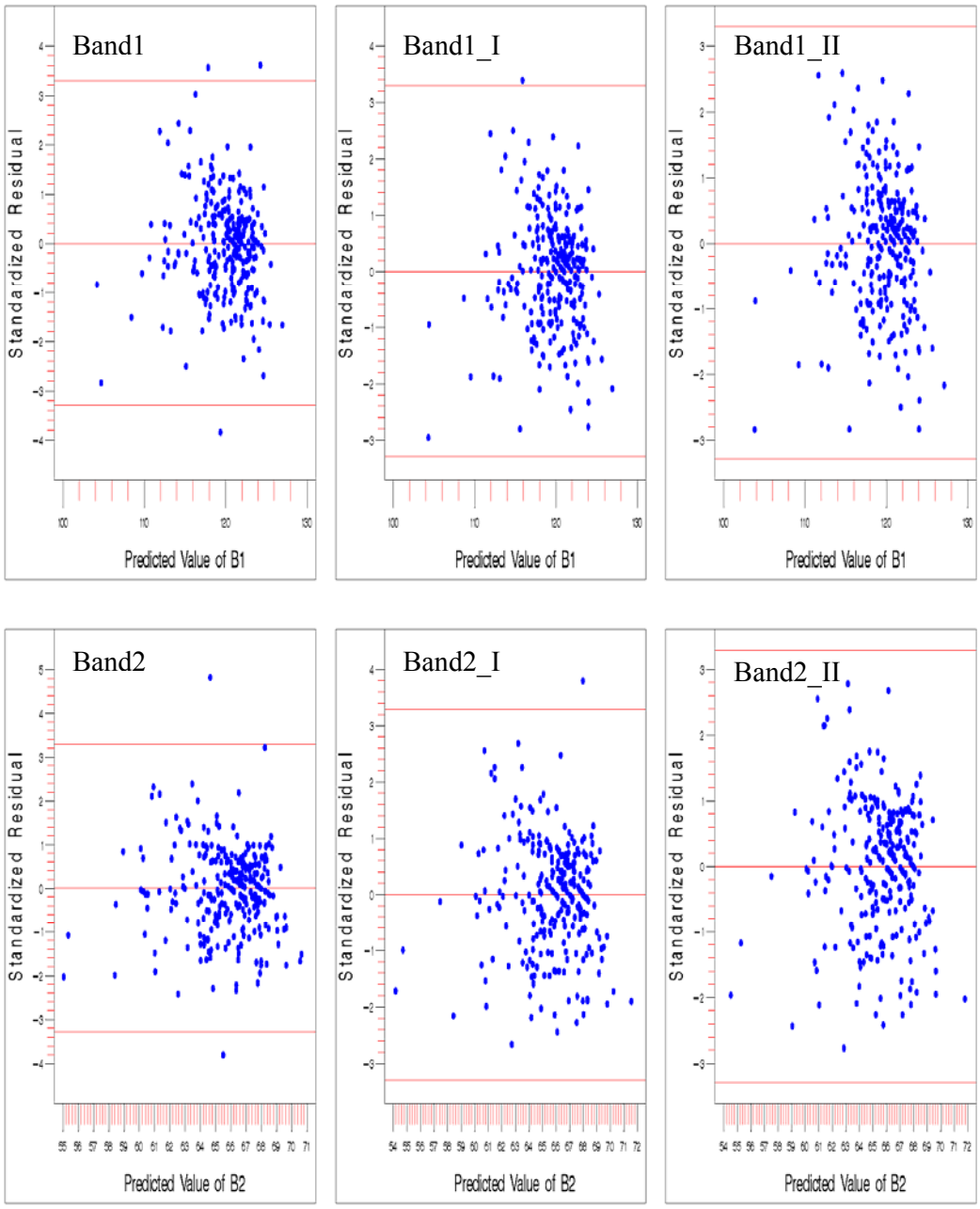


Figure 124: Outlier removal iteration plots of standardized residual vs. predicted value of dependent variable for bands 1 and 2 of 1997 field-1 data (WMLR)

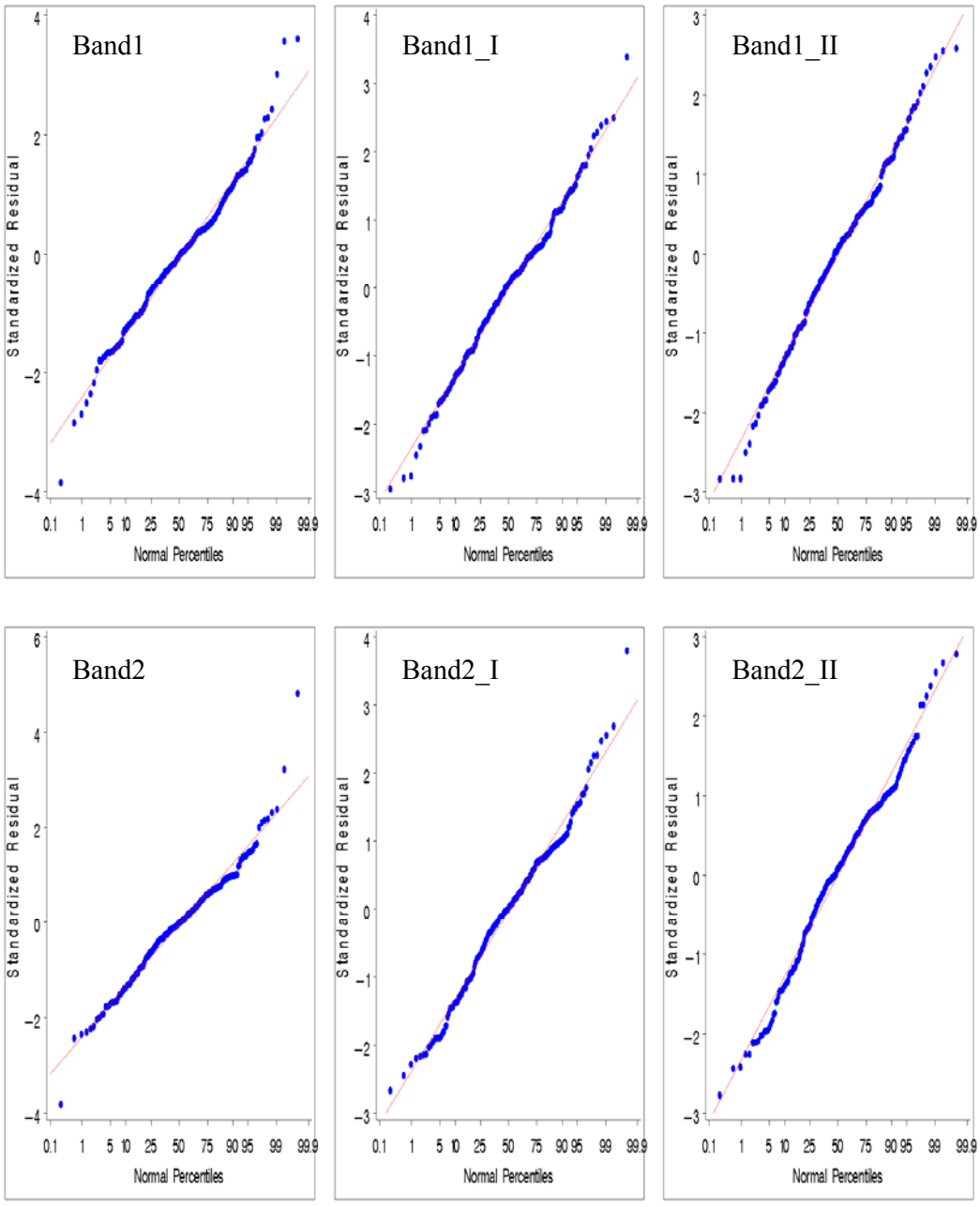


Figure 124: Normal probability plots corresponding to outlier removal iterations (I, II …) plots of bands 1 and 2 of 1997 field-1 data (WMLR)

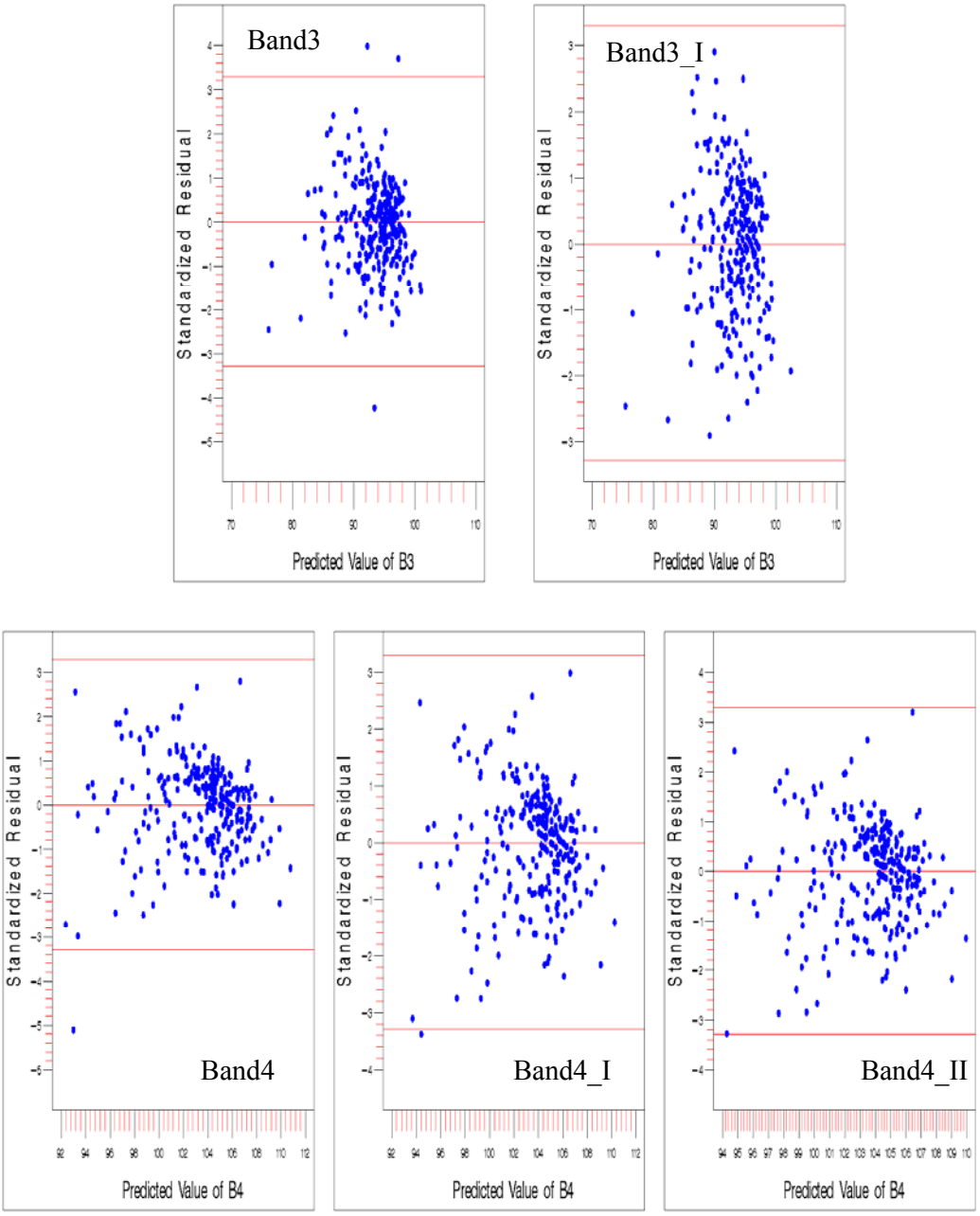


Figure 125: Outlier removal iteration plots of standardized residual vs. predicted value of dependent variable for bands 3 and 4 of 1997 field-1 data (WMLR)

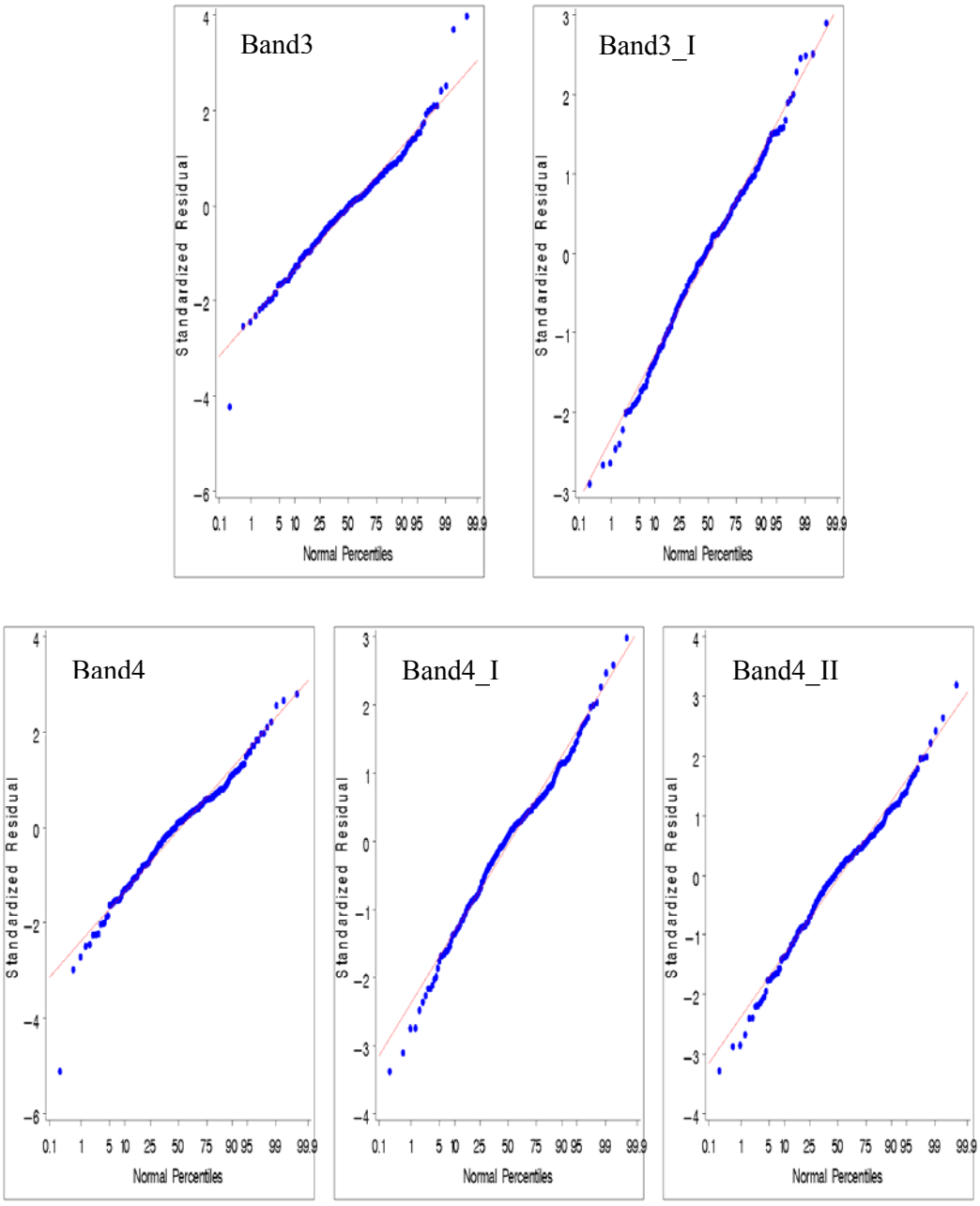


Figure 126: Normal probability plots corresponding to outlier removal iterations (I, II …) plots of bands 3 and 4 of 1997 field-1 data (WMLR)

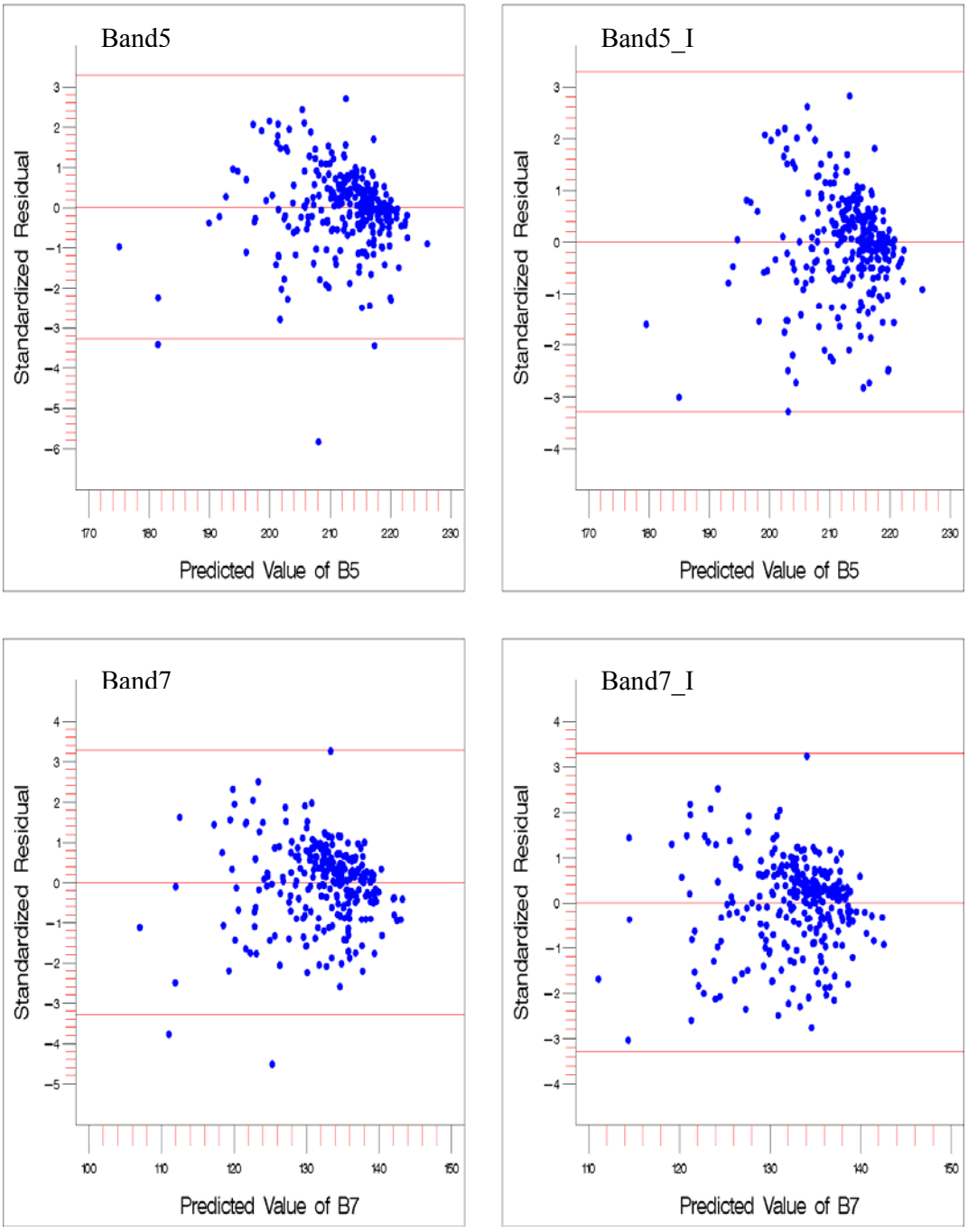


Figure 127: Outlier removal iteration plots of standardized residual vs. predicted value of dependent variable for bands 5 and 7 of 1997 field-1 data (WMLR)

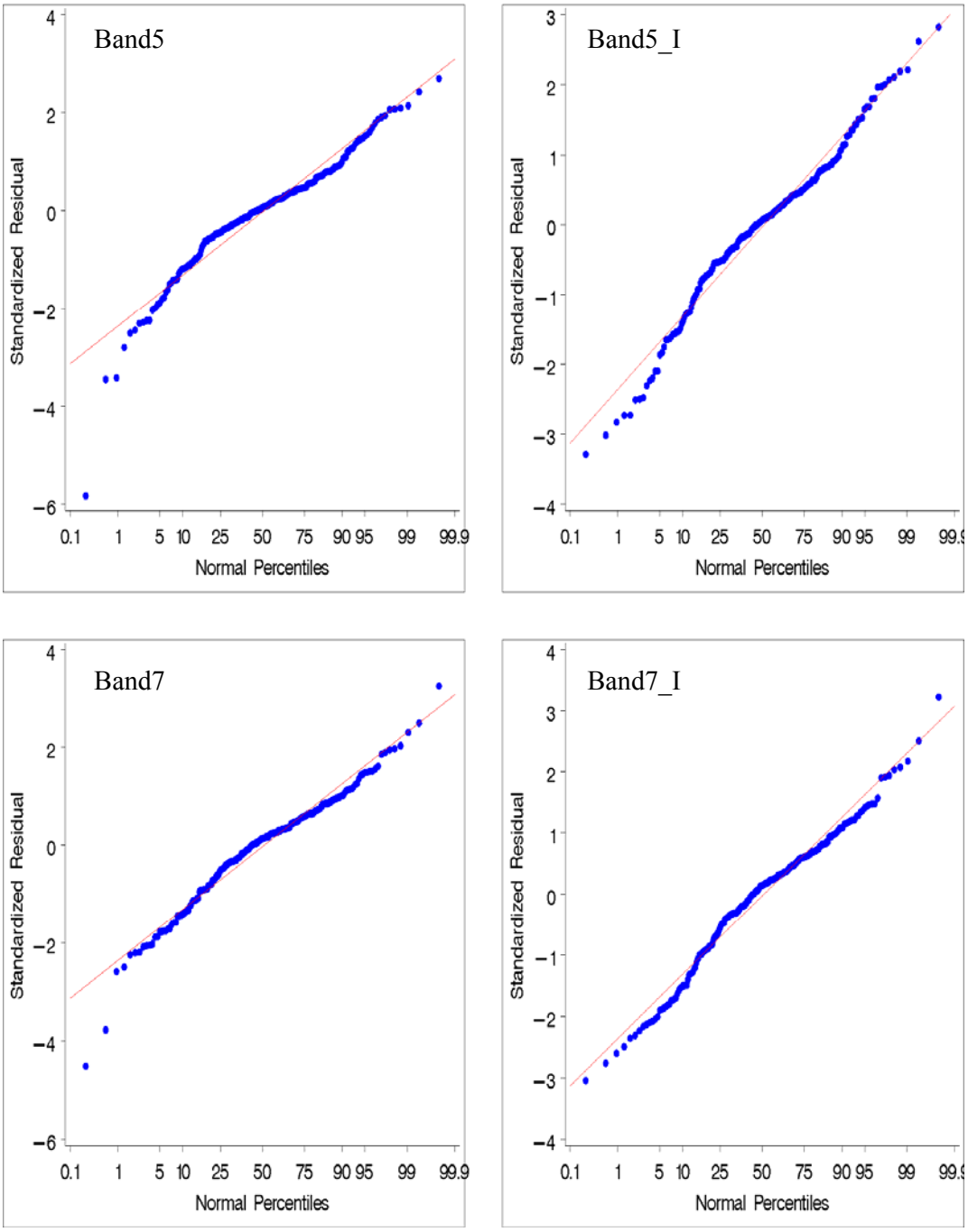


Figure 128: Normal probability plots corresponding to outlier removal iterations (I, II …) plots of bands 5 and 7 of 1997 field-1 data (WMLR)

The R^2 values of bands 1 and 2 (Figure 129) increased after removal of outliers. In band 1, the R^2 value decreased from 0.356 to 0.351 after one iteration and then increased to 0.371 after the second iteration; whereas for band 2 the R^2 value increased from 0.385 to 0.434 after one iteration and then decreased slightly to 0.415 after the second iteration, remaining higher than the original value. The R^2 value in band 3 increased from 0.42 to 0.44 (Figure 130). For bands 4, 5, and 7 the R^2 value decreased from 0.394 to 0.331, 0.437 to 0.409, and 0.429 to 0.385, respectively (Figures 130 and 131). The p values did not change appreciably ($p<0.0001$) in all the bands.

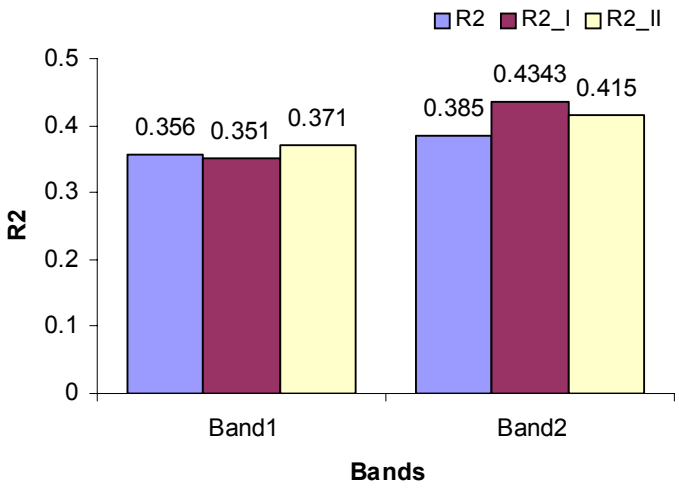


Figure 129: Changes in R^2 values with outlier removal at each iterations (I, II.. ..) for bands 1 and 2 of 1997 field-1 data (WMLR)

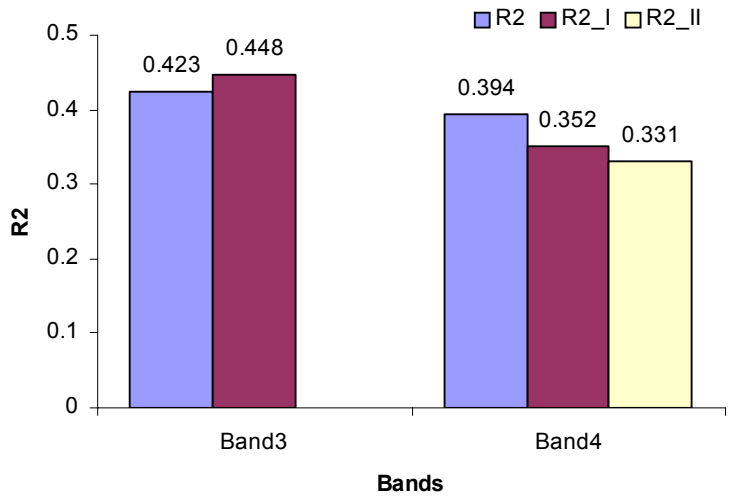


Figure 130: Changes in R^2 values with outlier removal at each iterations (I, II.. ..) for bands 3 and 4 of 1997 field-1 data (WMLR)

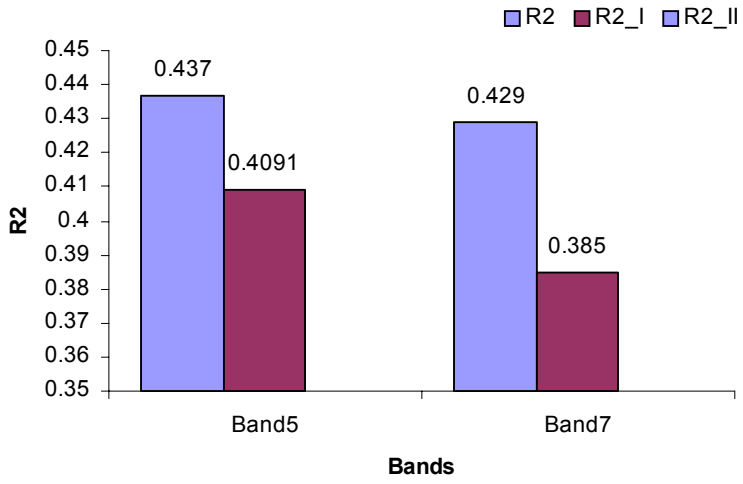


Figure 131: Changes in R^2 values with outlier removal at each iterations (I, II.. ..) for bands 5 and 7 of 1997 field-1 data (WMLR)

Field-3

The outlier removal process for WMLR analysis of field-3 data is illustrated in Figure 132 for 1999 data and in Figures 134, 136, 138, and 140 for 2001 data. The corresponding normality plots of the outlier removal iterations are provided in Figures 133, 135, 137, 139, and 141. The 1999 constant variance plots (Figure 132) show that bands 5 and 7 required three and one iterations, respectively, to remove all the detected outliers. The corresponding normality plots of bands 5 and 7 suggest that, after removal of outliers, the curve of data is closer to the reference line (Figure 133). Overall, improvement in the assumptions of constant variance and normality was observed when outliers were removed.

A look at the 2001 data indicates that bands 1 and 2 required one iteration, band 3 and 4 required two iterations, band 5 required three iterations, and band 7 required two iterations to remove all the detected outliers. In all bands, it was observed that after removal of outliers, the distribution of data points about the reference line (standardized residual = 0) appeared to be improved (Figures 134, 136, 138, and 140). The corresponding normal probability plots indicate improvement in normality, as the few data points lying far from the reference line are no longer in evidence, and overall the curves appear to be well aligned with the reference line (Figures 135, 137, 139, and 141.

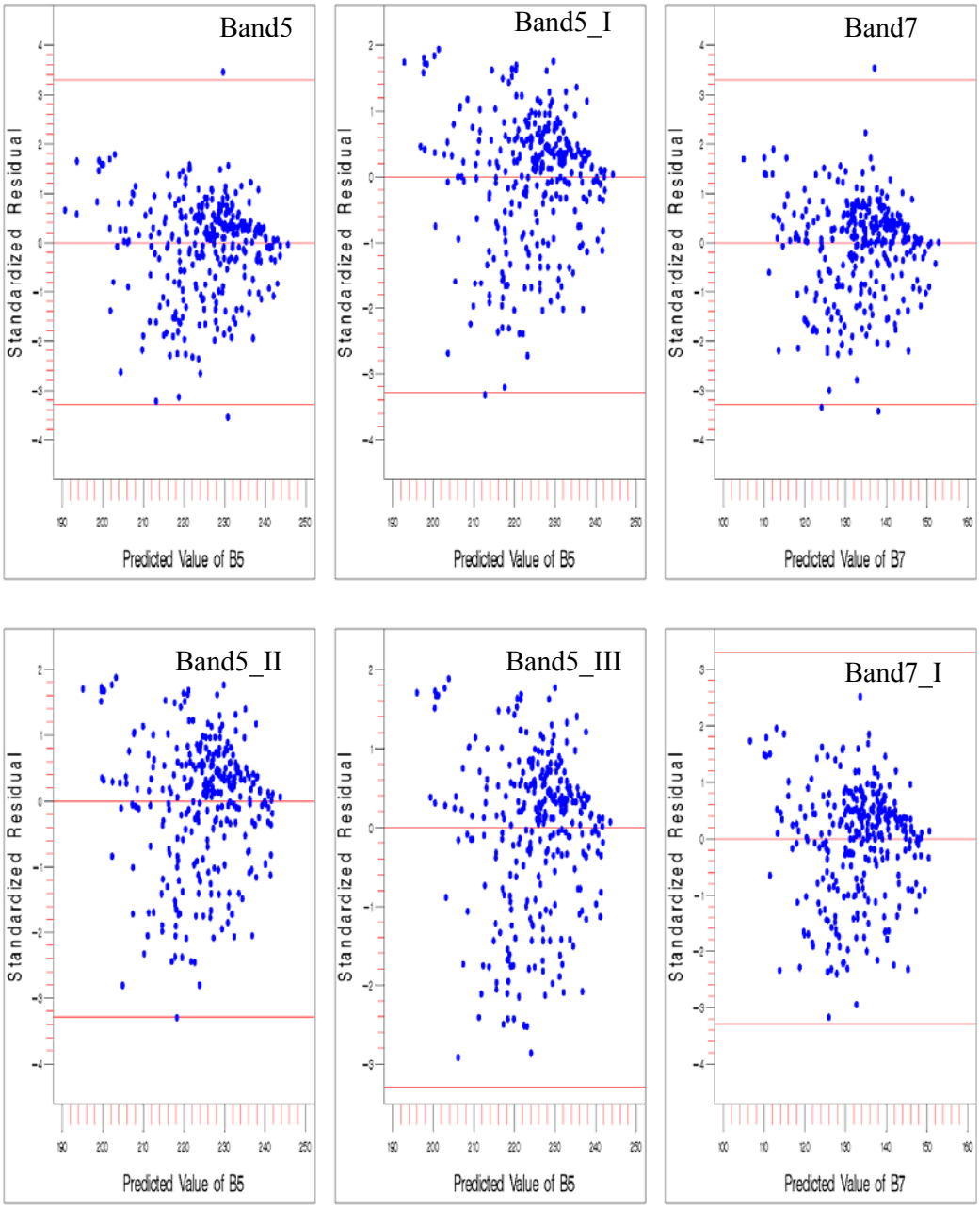


Figure 132: Outlier removal iteration plots of standardized residual vs. predicted value of dependent variable for bands 5 and 7 of 1999 field-3 data (WMLR)

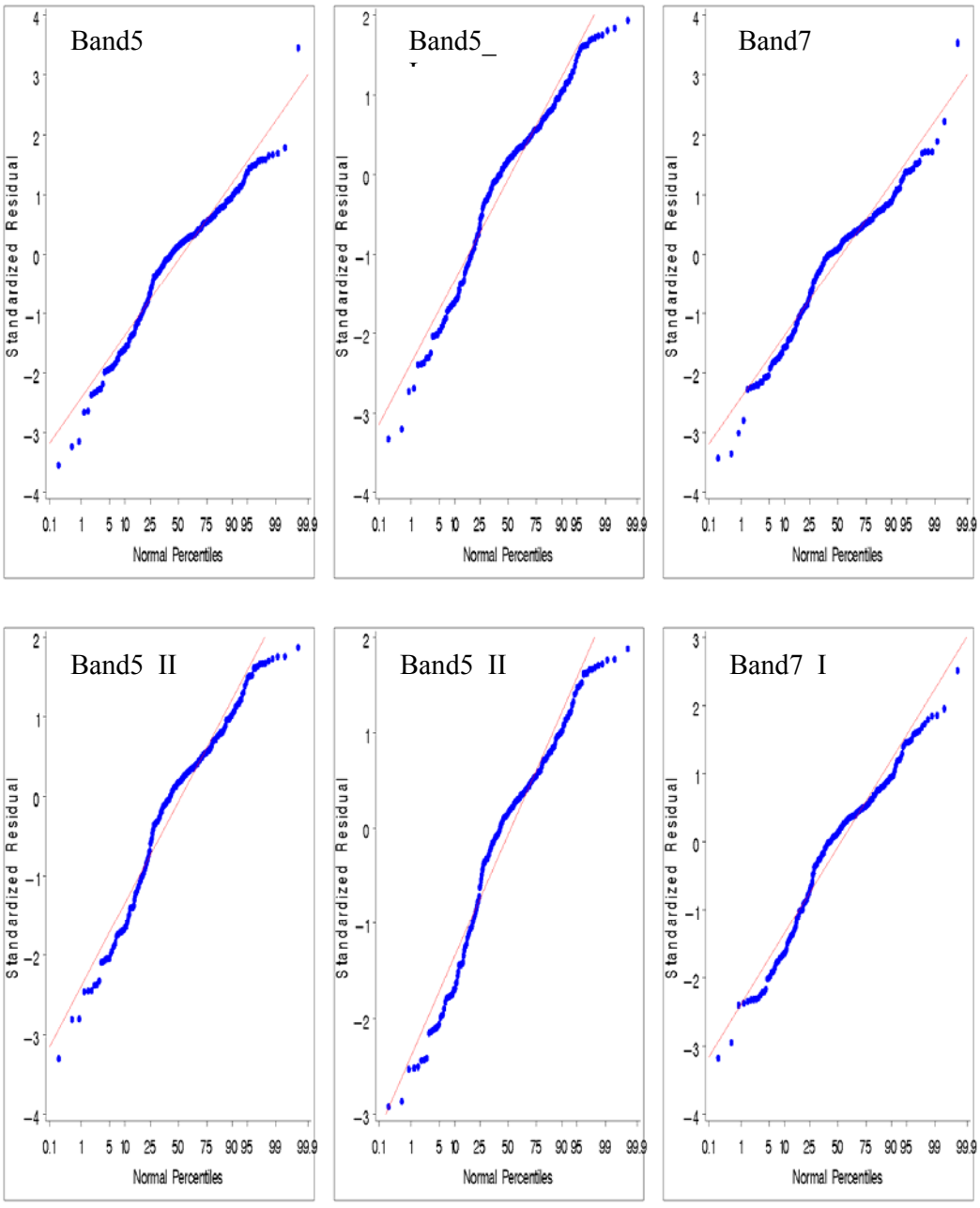


Figure 133: Normal probability plots corresponding to outlier removal iterations (I, II …) plots for bands 5 and 7 of 1999 field-3 data (WMLR)

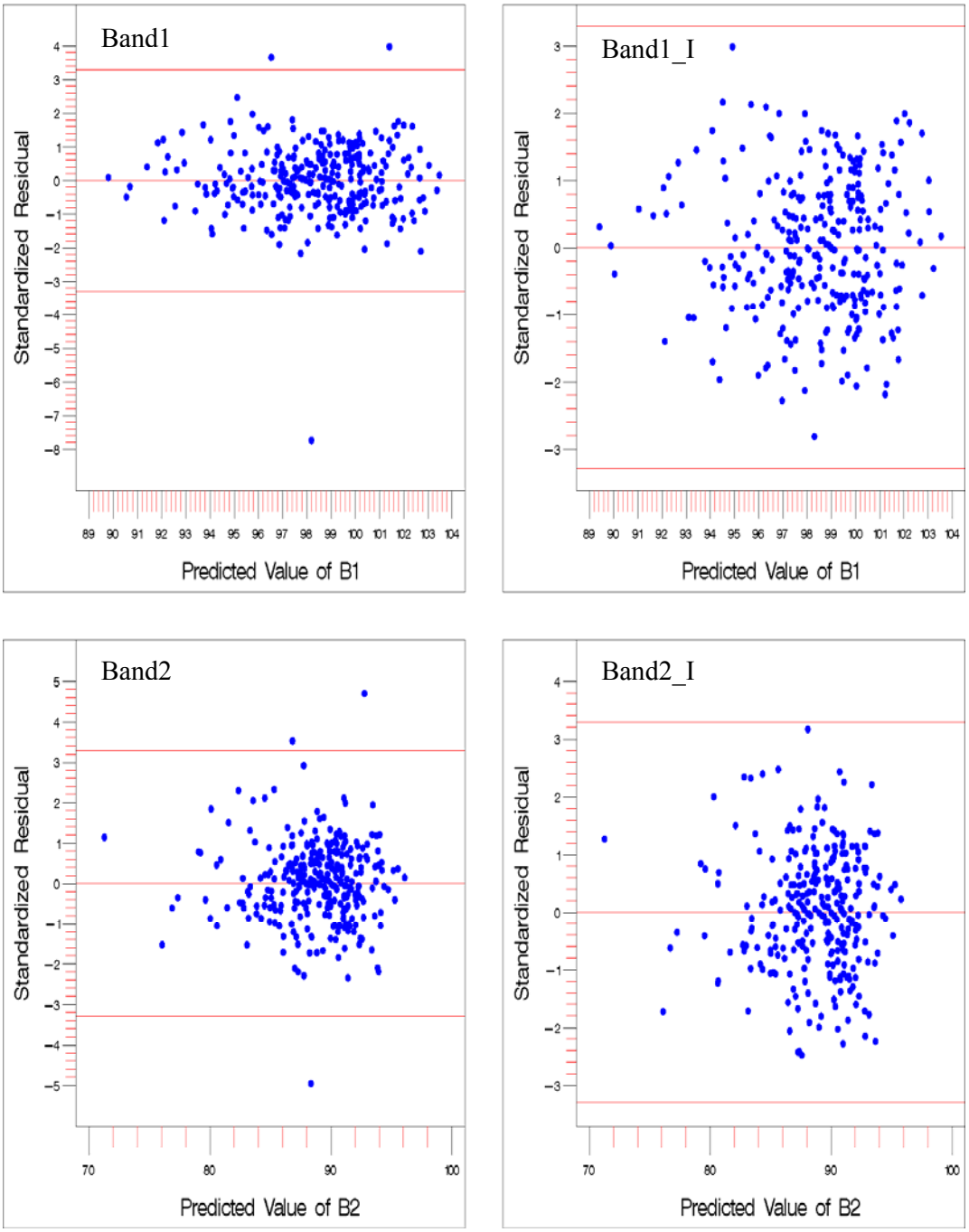


Figure 134: Outlier removal iteration plots of standardized residual vs. predicted value of dependent variable for bands 1 and 2 of 2001 field-3 data (WMLR)

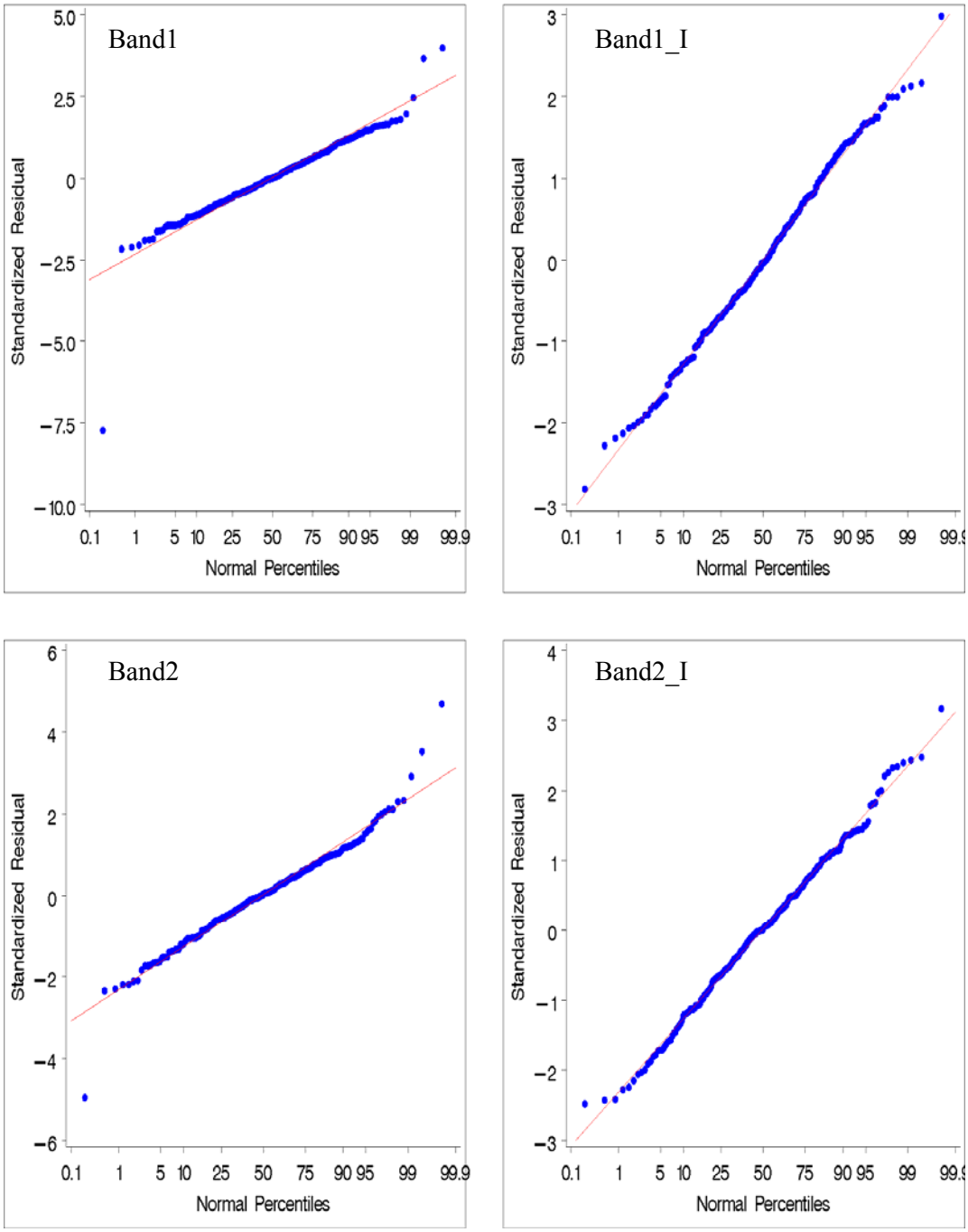


Figure 135: Normal probability plots corresponding to outlier removal iterations (I, II …) plots for bands 1 and 2 of 2001 field-3 data (WMLR)

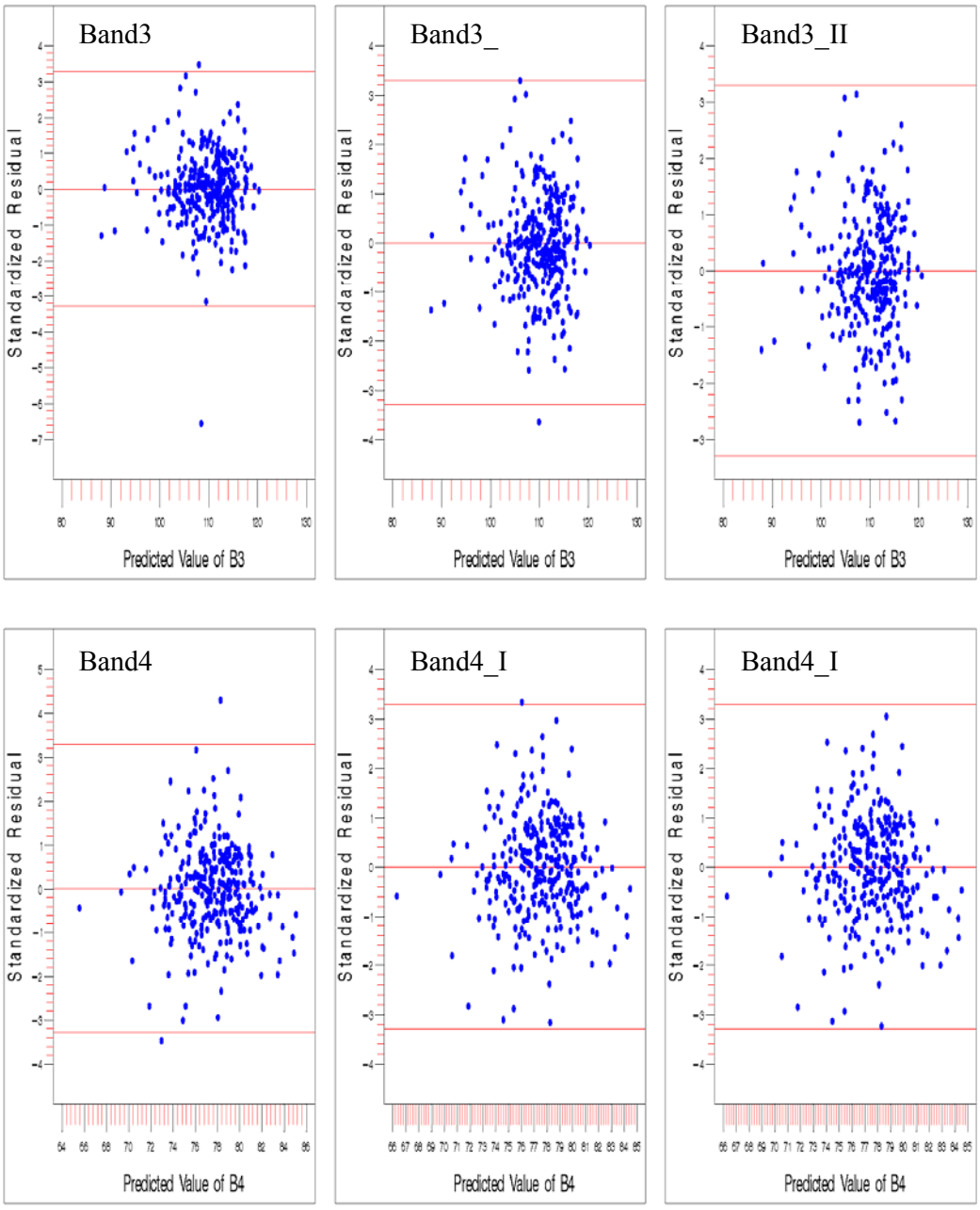


Figure 136: Outlier removal iteration plots of standardized residual vs. predicted value of dependent variable for bands 3 and 4 of 2001 field-3 data (WMLR)

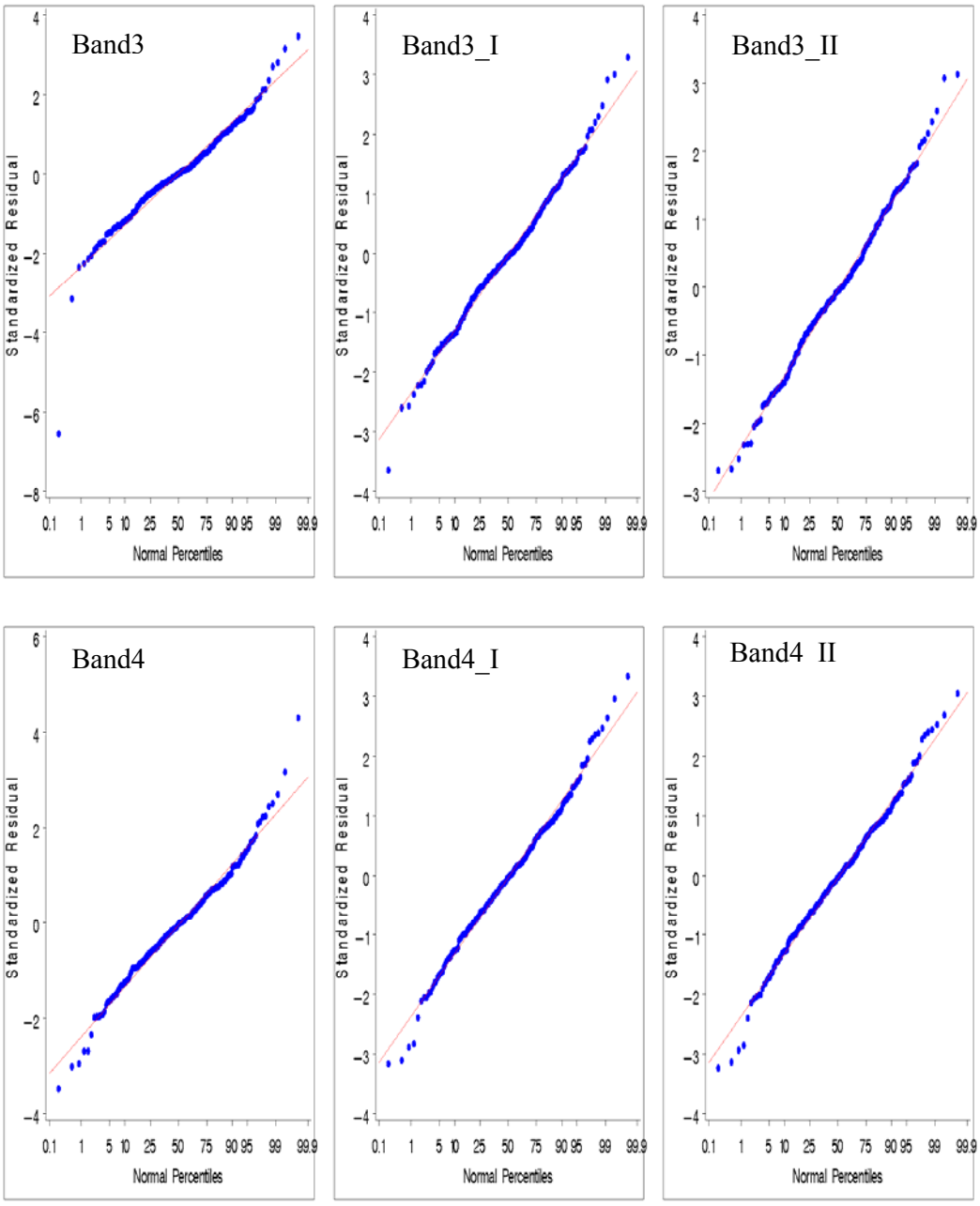


Figure 137: Normal probability plots corresponding to outlier removal iterations (I, II …) plots for bands 3 and 4 of 2001 field-3 data (WMLR)

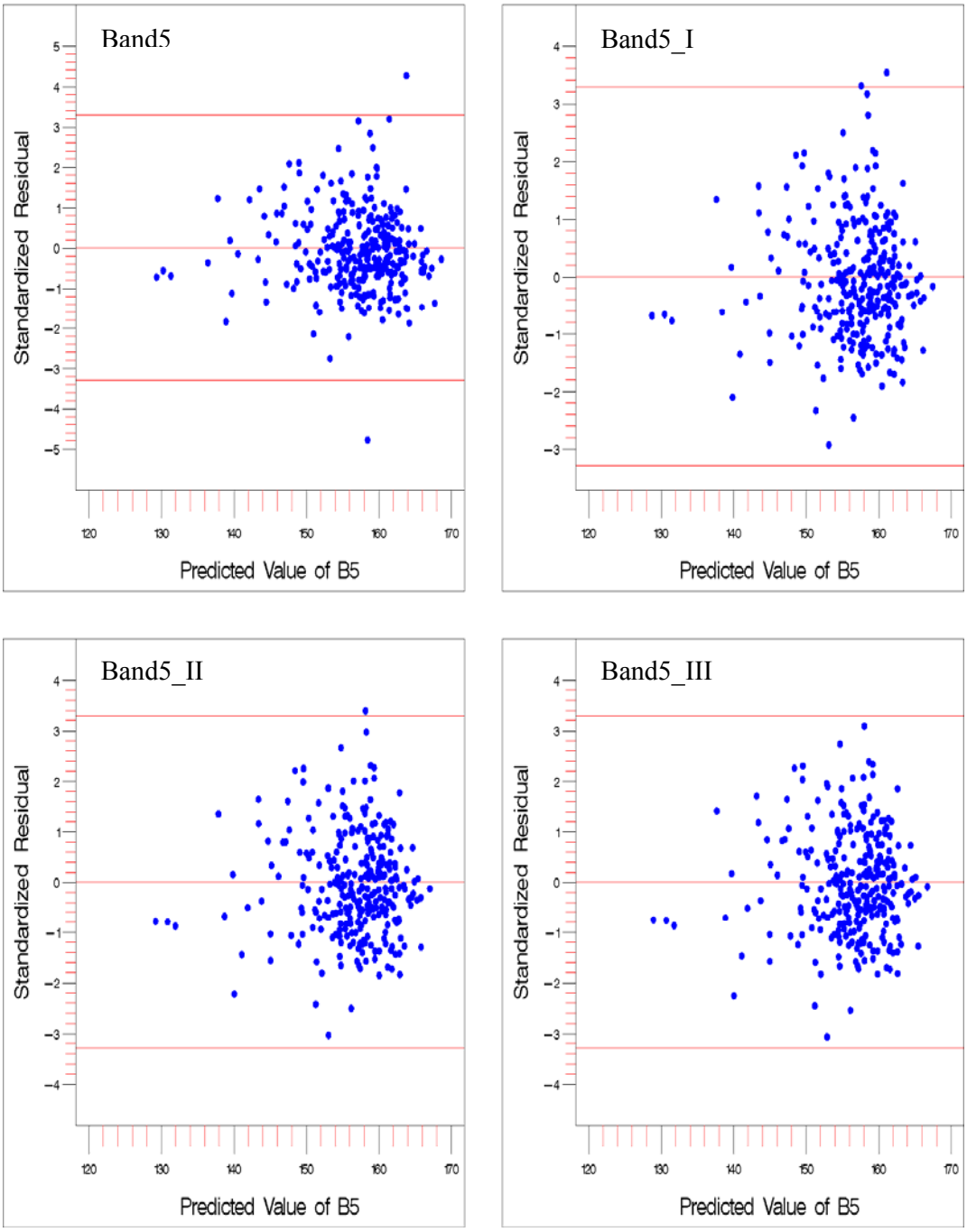


Figure 138: Outlier removal iteration plots of standardized residual vs. predicted value of dependent variable for band 5 of 2001 field-3 data (WMLR)

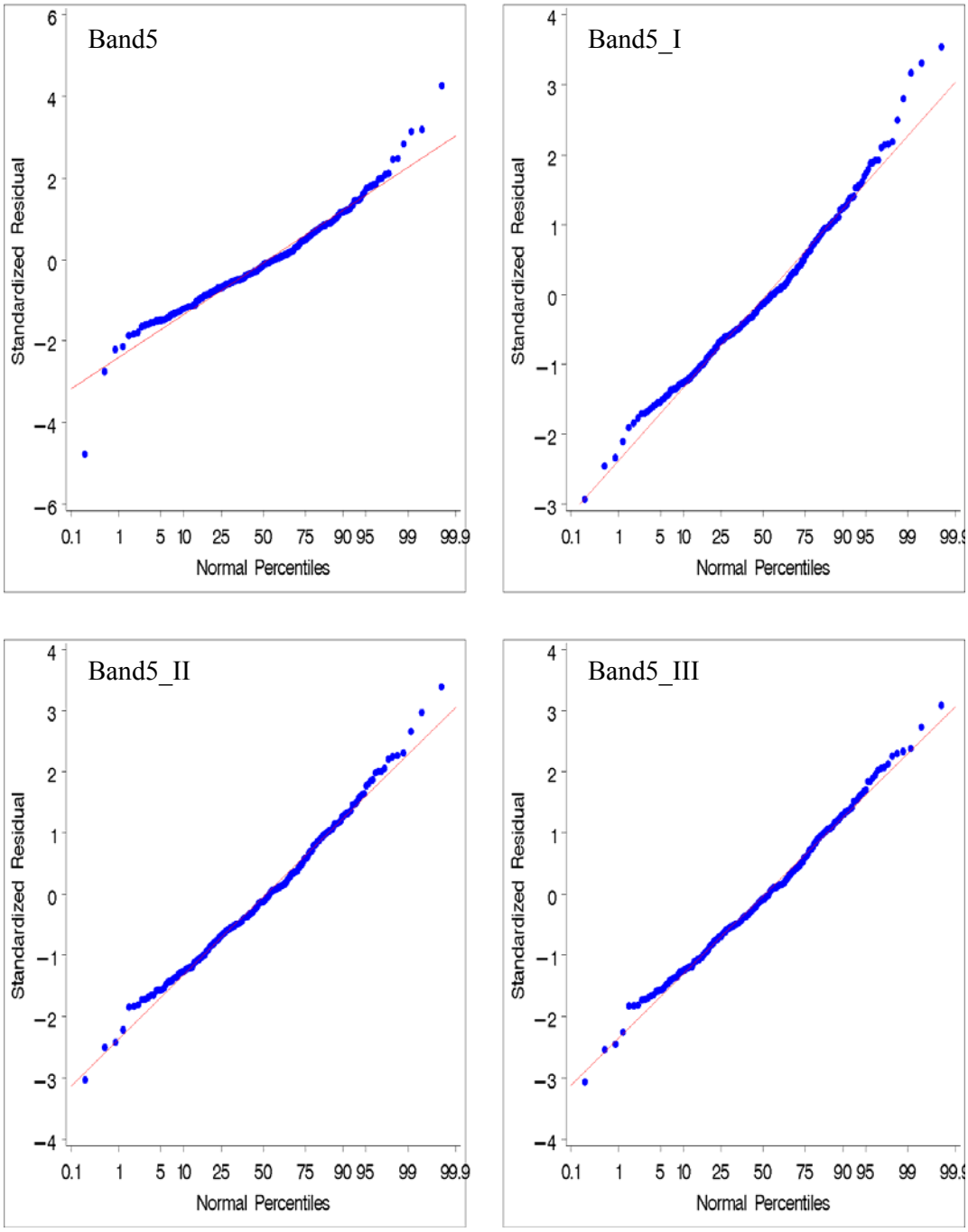


Figure 139: Normal probability plots corresponding to outlier removal iterations (I, II …) plots for band 5 of 2001 field-3 data (WMLR)

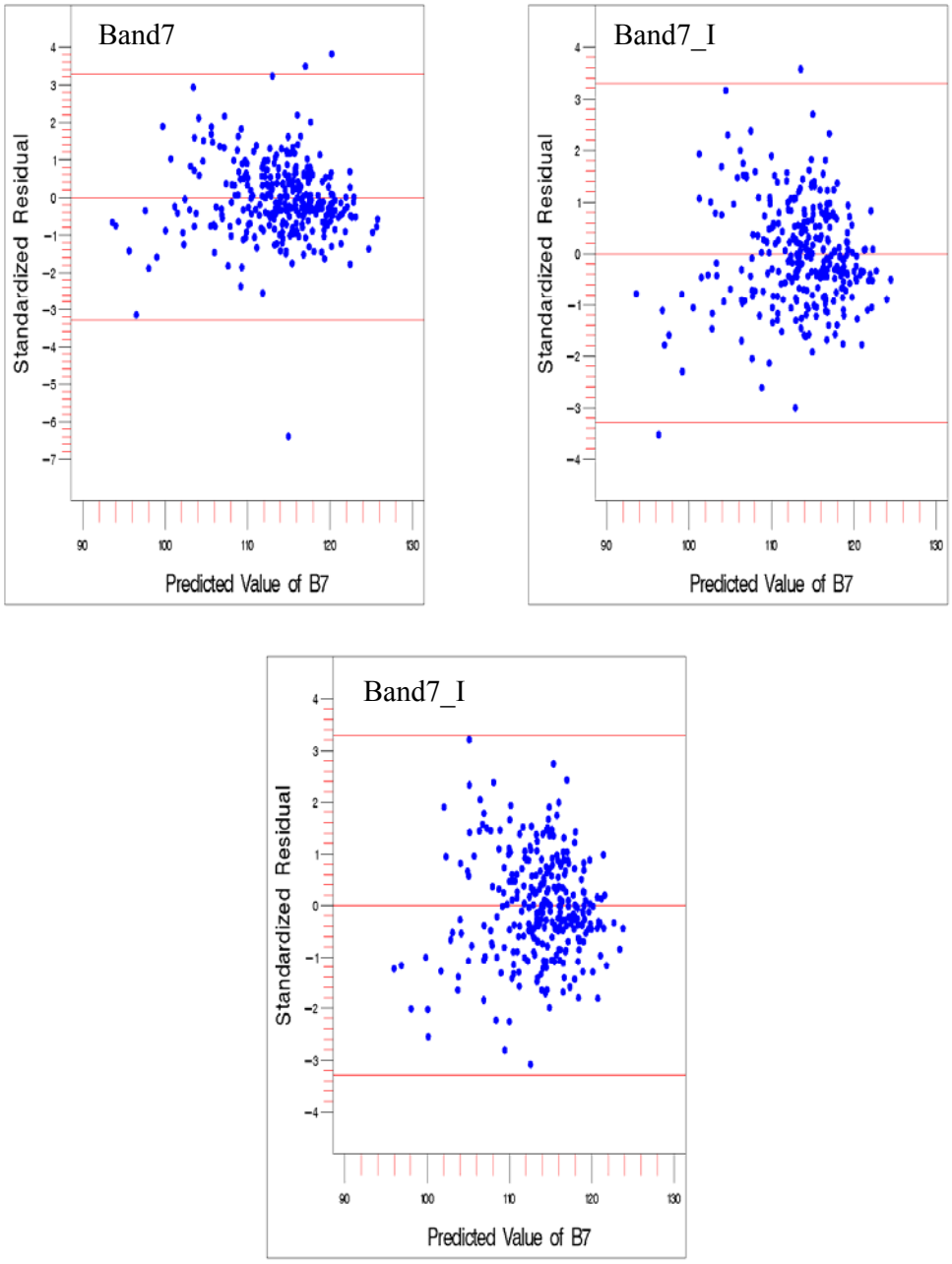


Figure 140: Outlier removal iteration plots of standardized residual vs. predicted value of dependent variable for band 7 of 2001 field-3 data (WMLR)

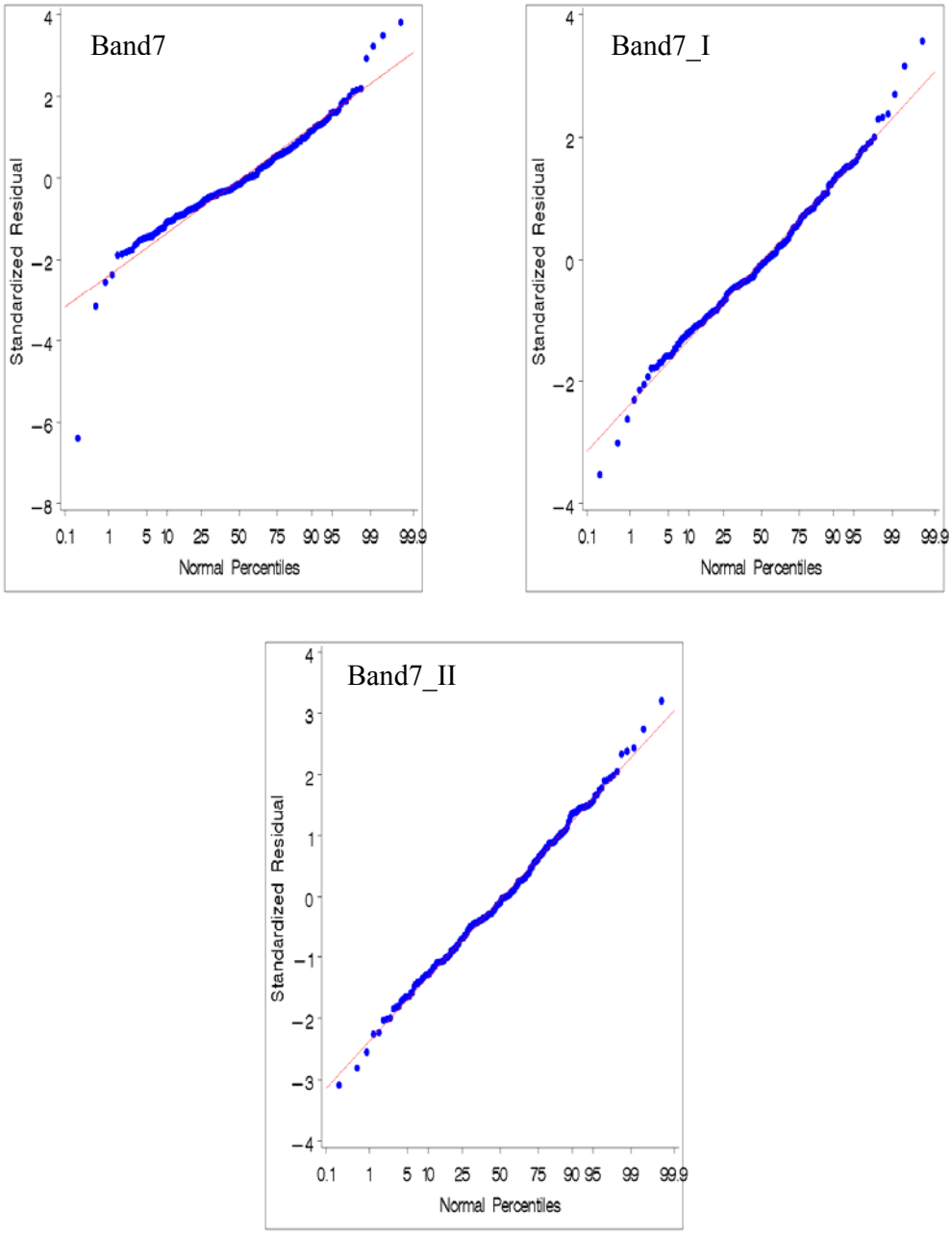


Figure 141: Normal probability plots corresponding to outlier removal iterations (I, II …) plots for band 7 of 2001 field-3 data (WMLR)

Figure 142 indicates that for band 5 the R^2 value increased after one iteration, but it decreased after the second and third iterations. Overall, a decrease of 1% was observed. With band 7, R^2 decreased from 0.167 to 0.163. The significance levels remained high ($p < 0.0001$) after removal of outliers for both bands.

Figure 143 shows that R^2 improved from 0.380 to 0.444 and 0.423 to 0.451 for bands 1 and 2, respectively, for 2001 field-3 data. For band 3, it improved gradually from 0.43 to 0.48 to 0.505 at each successive iteration (Figure 144). With Band 4, R^2 decreased after the first iteration from 0.288 to 0.281 but then increased to 0.293 after the second iteration, giving an overall minor increase (Figure 145). Band 5 had a decrease in R^2 from 0.465 to 0.45 after the first iteration, but then R^2 increased gradually and after the next two iterations, ending with a similar value (0.466) to the original (Figure 145). Band 7 was the only band in the 2001 data that had a decrease in R^2, from 0.39 to 0.359 after two iterations. However, irrespective of such changes in R^2, The p values did not change appreciably, with high significance levels ($p < 0.0001$) in all the bands.

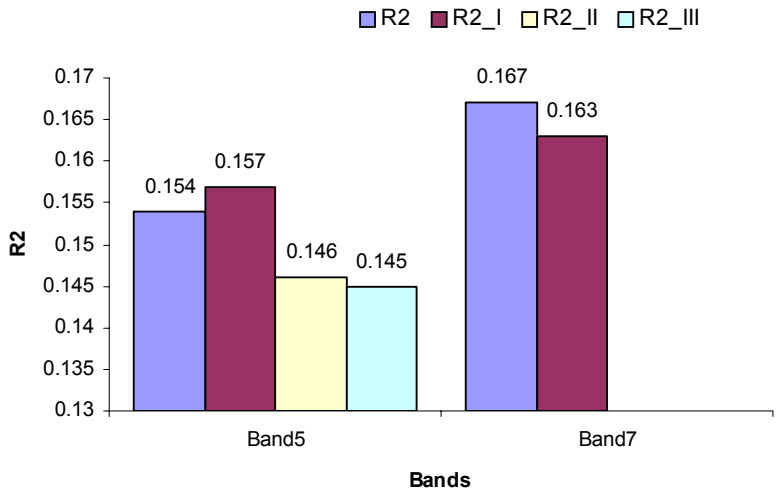


Figure 142: Changes in R^2 values with outlier removal at each iterations (I, II.. ..) for bands 5 and 7 of 1999 field-3 data (WMLR)

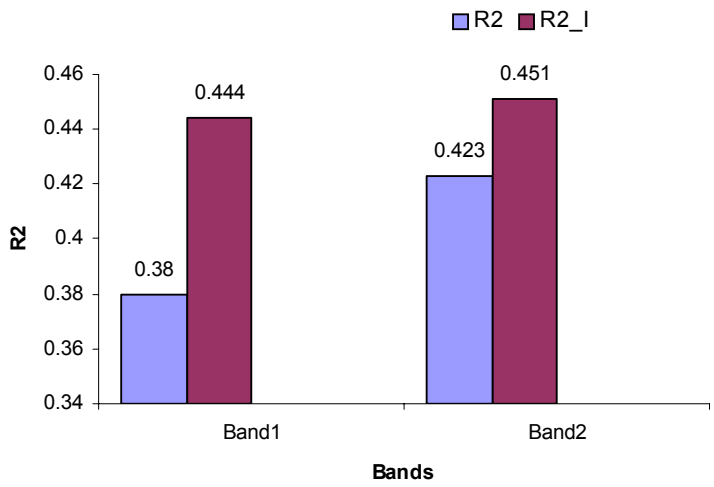


Figure 143: Changes in R^2 values with outlier removal at each iterations (I, II.. ..) for bands 1 and 2 of 2001 field-3 data (WMLR))

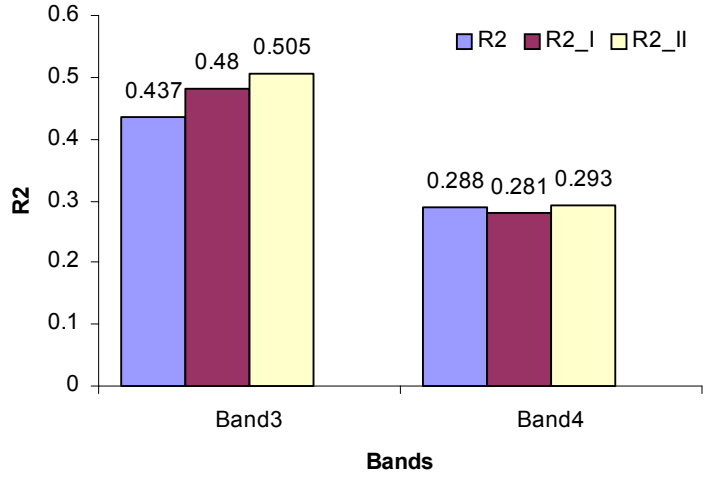


Figure 144: Changes in R^2 values with outlier removal at each iterations (I, II.. ..) for bands 3 and 4 of 2001 field-3 data (WMLR)

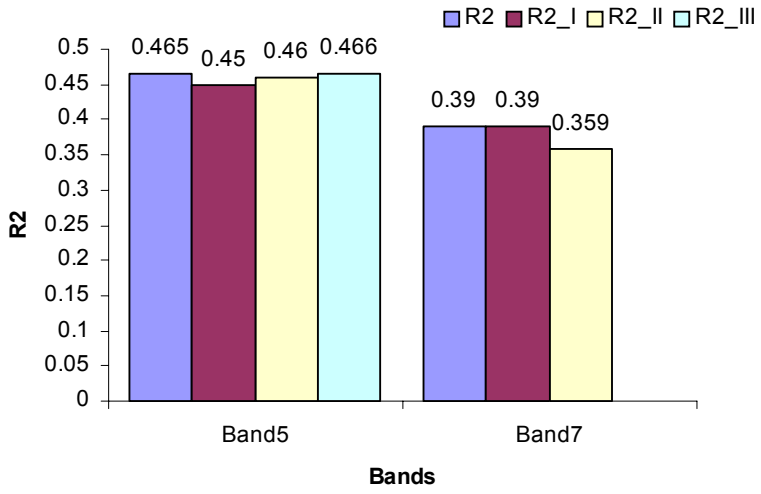


Figure 145: Changes in R^2 values with outlier removal at each iterations (I, II.. ..) for bands 5 and 7 of 2001 field-3 data (WMLR)

Validation

WSLR

Field-1

Tables 26 and 27 illustrate the details of outliers removed from WSLR analysis of 2001 and 1997 field-1 data. Table 26 indicates that band 3 of 2001 field-1 data had 10 outliers. Compared to SLR analysis of the same data, two more points (27 and 219) were detected as outliers. All the other points are same as with SLR analysis and should have the same reasons for being outliers. Points 27 and 219 have higher DN value than the average DN value for band 3, but the values are only approximately one s.d. higher than the average value. The positions of these two outliers on the image (Figure 146) gives no obvious indication of a reason for their being outliers.

Table 26: Detected outliers for Band-3 of 2001 field-1 data and their corresponding actual reflectance values (DN, Mean, Std. Dev), (WSLR)

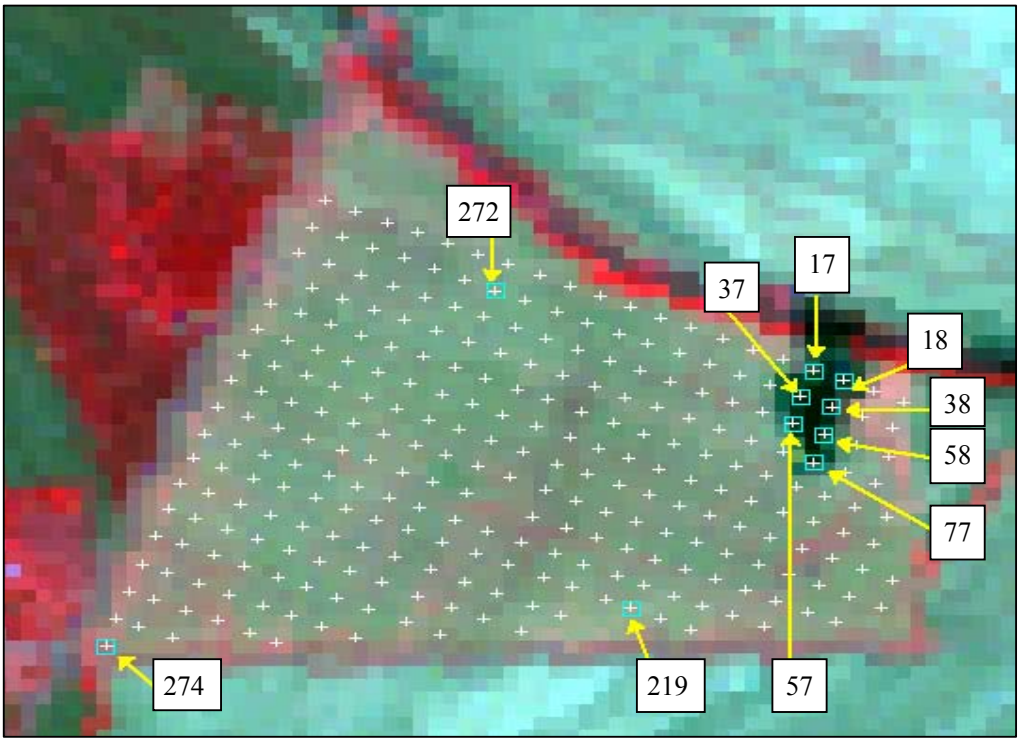
Bands	Obs. No.	Easting	Northing	Iteration	Actual Reflectance Values		
					DN	Mean	Std. Dev
Band 3	17	746199.299	3774882.447	II	77	102.68	8.06
	18	746256.624	3774858.747	I	59		
	27	745596.905	3775061.759	II	109		
	37	746175.402	3774824.991	I	60		
	38	746233.099	3774801.19	I	46		
	57	746161.302	3774763.683	I	55		
	58	746220.214	3774739.359	I	66		
	77	746200.159	3774676.231	II	81		
	219	745853.587	3774348.727	II	110		
	274	744858.6	3774263.226	I	88		

Table 27 includes details of outliers removed from 1997 field-1 data. Compared to SLR analysis, there was a substantial increase in the number outliers with WSLR

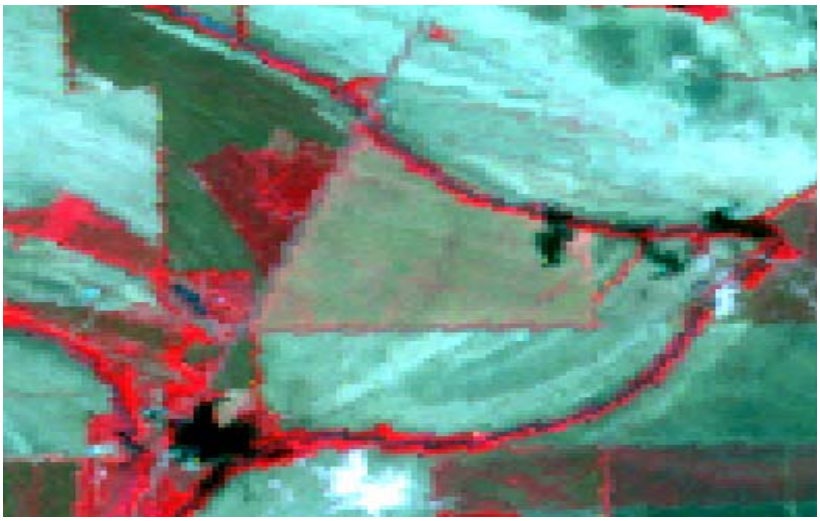
analysis. Newly detected points were 9, 15, 51, 106, 171, and 247 (Table 27). Points 106 and 171 had higher DN values than the average for their respective bands, but they were within 2 s.d. of the mean. Figure 147 reveals no obvious reasons for their being outliers. Points 9, 15, and 51 were located on a site where there was or may have drainage at the time of image collection, and so they have similar reasons for being as outliers as do points 10 through 13 and 32. All the other points were also detected in SLR analysis and should have the same reasons as described in the SLR outlier validation section.

Table 27: Detected outliers for all the studied bands of 1997 field-1 data and their corresponding actual radiance values (DN, Mean, Std. Dev), (WSLR)

Bands	Obs. No.	Easting	Northing	Iteration	Actual Radiance Values		
					DN	Mean	Std. Dev
Band 1	11	745853.4	3775025.156	II	101	119.45	5.86
	12	745911.001	3775001.461	I	96		
	15	746083.645	3774929.377	I	112		
	32	745886.477	3774943.212	II	100		
	106	745436.342	3774854.558	I	125		
	171	745359.874	3774680.629	I	129		
	274	744858.6	3774263.226	I	100		
Band 2	11	745853.4	3775025.156	II	52	65.54	3.93
	12	745911.001	3775001.461	I	51		
	15	746083.645	3774929.377	I	59		
	32	745886.477	3774943.212	II	53		
	106	745436.342	3774854.558	I	71		
	171	745359.874	3774680.629	I	71		
	274	744858.6	3774263.226	I	53		
Band 3	11	745853.4	3775025.156	II	71	93.13	5.89
	12	745911.001	3775001.461	I	69		
	15	746083.645	3774929.377	I	84		
	32	745886.477	3774943.212	II	74		
	106	745436.342	3774854.558	I	100		
	171	745359.874	3774680.629	I	102		
	274	744858.6	3774263.226	I	73		
Band 4	9	745738.189	3775072.989	III	89	103.16	5.54
	10	745796.438	3775049.2	II	77		
	11	745853.4	3775025.156	I	80		
	12	745911.001	3775001.461	I	78		
	13	745968.42	3774977.651	II	87		
	15	746083.645	3774929.377	II	95		
	32	745886.477	3774943.212	III	83		
	51	745813.416	3774904.566	II	93		
	171	745359.874	3774680.629	I	110		
Band 5	11	745853.4	3775025.156	I	163	211.81	11.1
	12	745911.001	3775001.461	I	164		
	13	745968.42	3774977.651	III	183		
	15	746083.645	3774929.377	II	200		
	32	745886.477	3774943.212	II	173		
	51	745813.416	3774904.566	II	191		
	106	745436.342	3774854.558	III	222		
	171	745359.874	3774680.629	I	221		
	247	744955.255	3774510.268	IV	190		
	274	744858.6	3774263.226	I	158		
Band 7	11	745853.4	3775025.156	I	94	131.93	9.34
	12	745911.001	3775001.461	I	94		
	15	746083.645	3774929.377	II	120		
	51	745813.416	3774904.566	II	114		
	106	745436.342	3774854.558	I	143		
	274	744858.6	3774263.226	I	92		

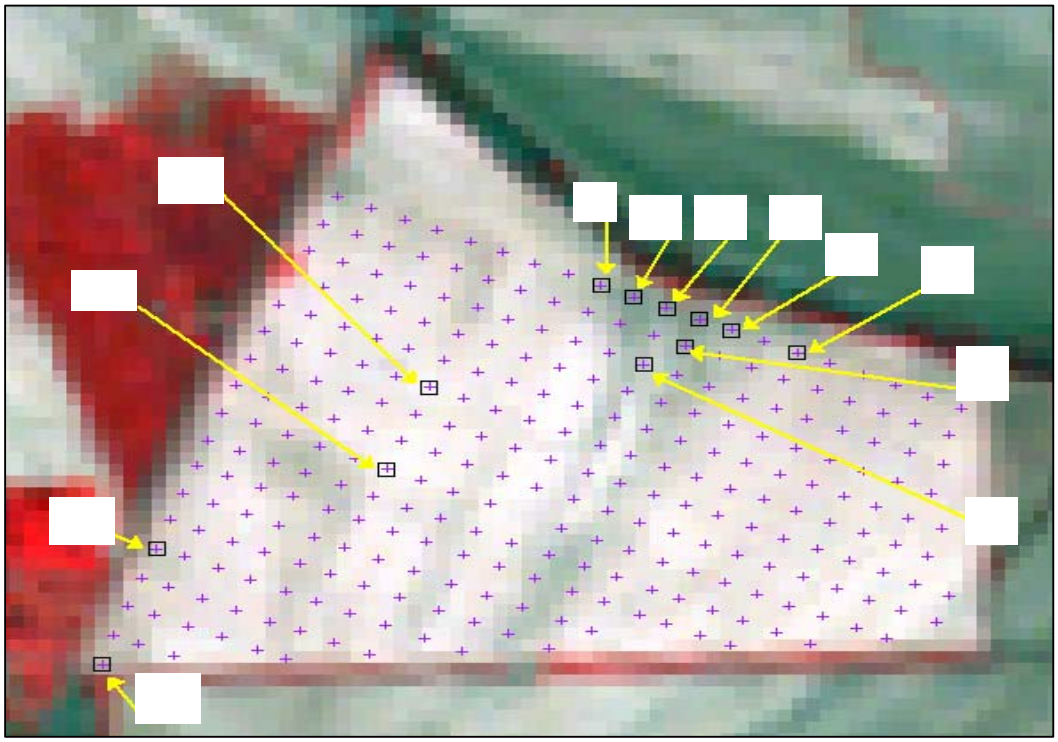


(a)



(b)

Figure146: Position of outliers on the close-up (a) and distant (b) false color Landsat images of 2001 field-1 data (WSLR)



(a)

(b)

Figure147: Position of outliers on the close-up (a) and distant (b) false color Landsat images of 1997 field-1 data (WSLR)

Field-3

The details of outliers in bands 1 through 3 of 1999 field-3 data are described in the table 28. The SLR analysis of the same field, in which only bands 2 and 3 were significant, did not have any outliers. For WSLR, six data points were observed to be outliers, with band 1 including all of them. In band 1, all the outliers except for points 244 and 288 have DN values at least two s.d. lower than the mean value for band-1. Points 244 and 288 have higher values than the mean value and were within the range of two s.d. Band 2, which also had point 244 as an outlier, had similar outliers to those of band 1. The other points in bands 2 and 3 had lower DN values than their respective mean values. Figure 148 indicates that points 56, 98, 212, and 254 are located on a dark green patch, which in a false color image is possibly an indication of higher moisture content at that location. Point 288 neighbors a red pixel that suggests the presence of green vegetation that possibly could have influenced point 228. No obvious reasons for being outliers were observed point 244.

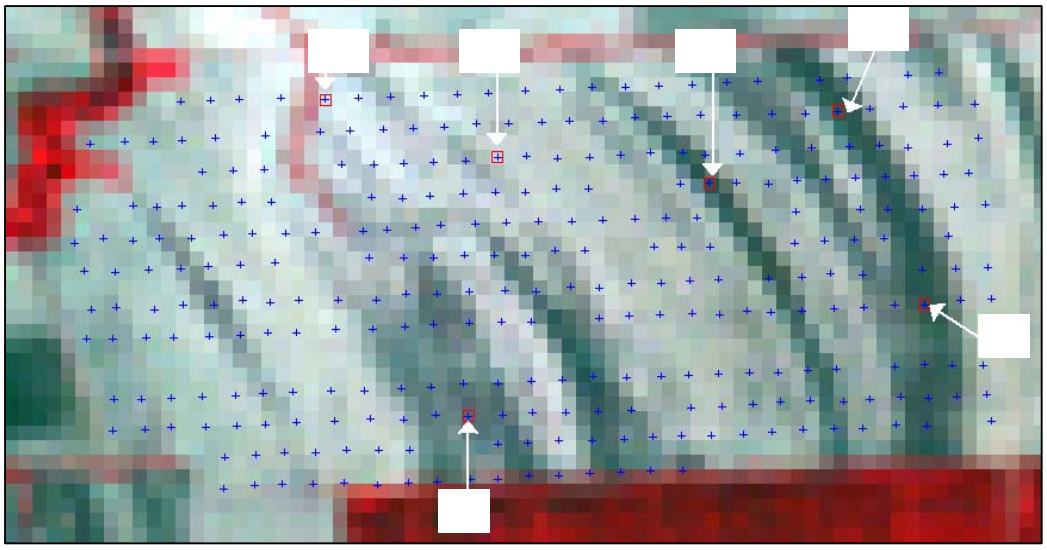
Table 29, which has outlier details for 2001 field-3 data, reveals that compared to SLR a substantial increase in outliers was observed with WSLR analysis. Except for point 202, all the other outliers were not detected in the SLR analysis. All the outliers, except point 67 in bands 4 and 7, point 94 in band 5, and points 95 and 202 in band 7 were within the range of ± 2 s.d of the mean of their respective bands. The image in figure 149 suggests that similar to previous analysis of field-3 data, these outliers were also located at the edges of the field, and might possibly have been influenced by factors described in the previous sections on SLR and MLR analyses.

Table 28: Detected outliers for bands 1 to 3 of 1999 field-3 data and their corresponding actual radiance values (DN, Mean, Std. Dev), (WSLR)

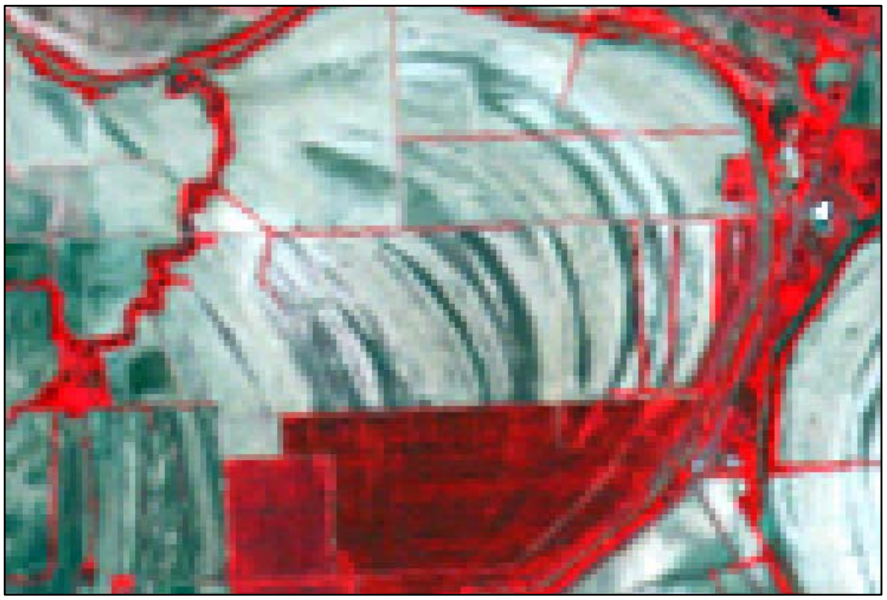
Bands	Obs. No.	Easting	Northing	Iteration	Actual Radiance Values		
					DN	Mean	Std. Dev
Band 1	56	742205.52	3772613.727	I	96	116.23	10.69
	98	743115.043	3772834.088	II	90		
	212	742686.168	3773072.193	III	87		
	244	742263.443	3773124.034	I	129		
	254	742942.729	3773214.238	II	90		
	288	741919.511	3773237.477	II	134		
Band 2	56	742205.52	3772613.727	I	53	66.41	8.34
	98	743115.043	3772834.088	II	46		
	244	742263.443	3773124.034	I	76		
Band 3	56	742205.52	3772613.727	I	75	106.32	17.2
	98	743115.043	3772834.088	II	65		

Table 29: Detected outliers for all the studied bands of 2001 field-3 data and their corresponding actual radiance values (DN, Mean, Std. Dev), (WSLR)

Bands	Obs. No.	Easting	Northing	Iteration	Actual Radiance Values		
					DN	Mean	Std. Dev.
Band 1	267	742096.151	3773180.453	II	105	98.4	4.078
	288	741919.511	3773237.477	I	107		
	296	741521.018	3773153.131	I	89		
Band 2	288	741919.511	3773237.477	I	101	88.97	5.391
	296	741521.018	3773153.131	I	81		
Band 3	296	741521.018	3773153.131	I	94	109.68	7.982
Band 4	67	741495.938	3772585.687	I	63	77.282	5.054
	94	741553.16	3772647.159	II	61		
	243	742320.773	3773125.507	II	84		
	244	742263.443	3773124.034	I	82		
	288	741919.511	3773237.477	I	86		
Band 5	243	742320.773	3773125.507	I	171	156.24	9.013
	244	742263.443	3773124.034	I	170		
	296	741521.018	3773153.131	I	147		
Band 7	67	741495.938	3772585.687	I	82	113.06	8.89
	95	741497.632	3772648.072	II	84		
	175	741439.16	3772898.574	II	96		
	202	741419.572	3772954.314	III	84		
	243	742320.773	3773125.507	I	128		
	244	742263.443	3773124.034	I	126		
	296	741521.018	3773153.131	I	99		

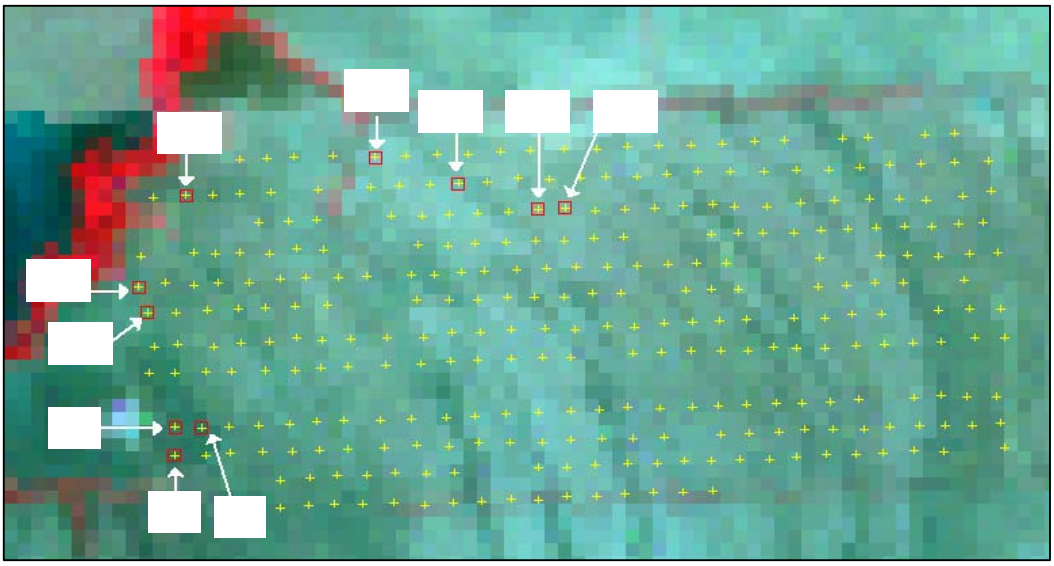


(a)



(b)

Figure 148: Position of outliers on the close-up (a) and distant (b) false color Landsat images of 1999 field-3 data (WSLR)



(a)

(b)

Figure 149: Position of outliers on the close-up (a) and distant (b) false color Landsat images of 2001 field-3 data (WSLR)

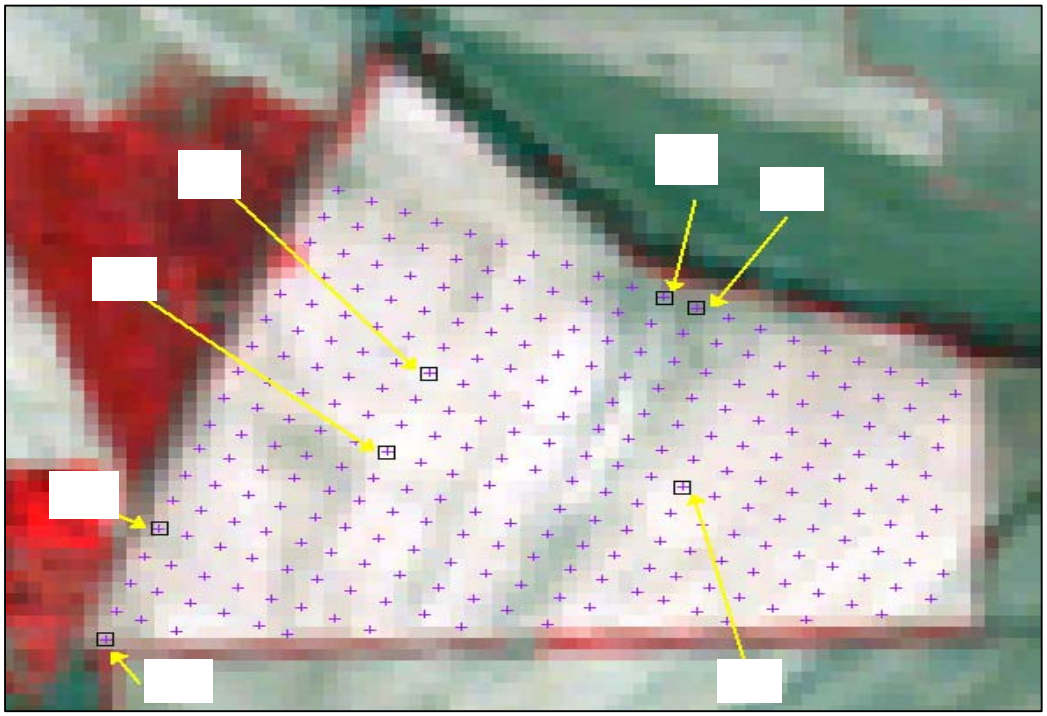
WMLR

Field-1

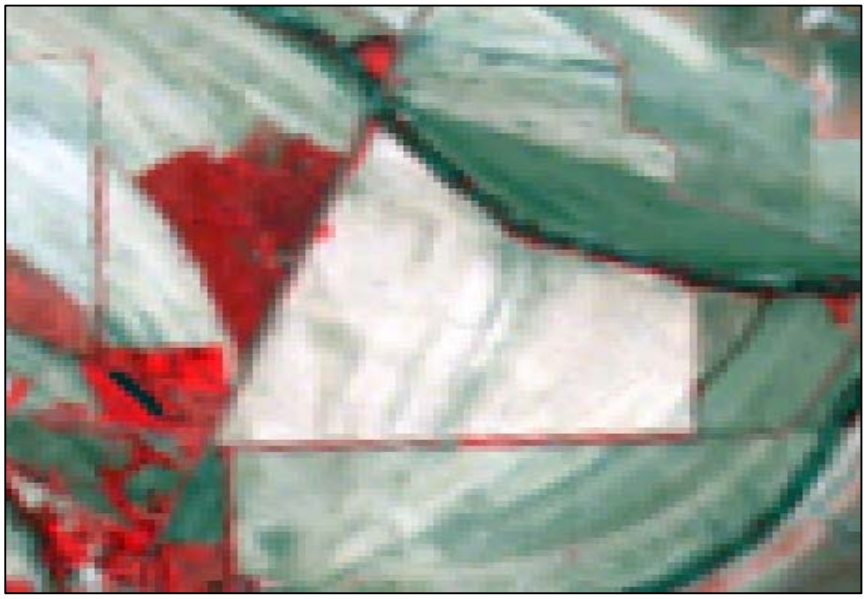
Table 30 includes details of outliers removed from WMLR analysis of 1997 field-1 data. Points 106, 135, and 171, which were not detected in MLR analysis, were detected in WMLR analysis. Observations 32, 10, and 186 were detected in MLR but not in WMLR. The newly detected points (106, 135, and 171) had higher DN values than the mean values of their respective bands but were within the range of 2 s.d. All the other outliers were at least 2 s.d lower than the mean value of their respective bands. From figure 150, no obvious reasons for these points' being outliers were observed, and for the other outliers the reasons were the same as those found in the MLR analysis.

Table 30: Detected outliers for all the studied bands of 1997 field-1 data and their corresponding actual radiance values (DN, Mean, Std. Dev), (WMLR)

Bands	Obs. No.	Easting	Northing	Iteration	Actual Radiance Values		
					DN	Mean	Std. Dev
Band 1	135	745887.443	3774601.69	II	127	119.45	5.86
	106	745436.342	3774854.56	I	125		
	171	745359.874	3774680.63	I	129		
	274	744858.6	3774263.23	I	100		
Band 2	106	745436.342	3774854.56	I	71	65.54	3.93
	171	745359.874	3774680.63	II	71		
	274	744858.6	3774263.23	I	53		
Band 3	106	745436.342	3774854.56	I	100	93.13	5.89
	171	745359.874	3774680.63	I	102		
	274	744858.6	3774263.23	I	73		
Band 4	11	745853.4	3775025.16	II	80	103.16	5.54
	12	745911.001	3775001.46	I	78		
Band 5	12	745911.001	3775001.46	I	164	211.81	11.1
	247	744955.255	3774510.27	I	190		
	274	744858.6	3774263.23	I	158		
Band 7	12	745911.001	3775001.46	I	94	131.93	9.34
	274	744858.6	3774263.23	I	92		



(a)



(b)

Figure 150: Position of outliers on the close-up (a) and distant (b) false color Landsat images of 1997 field-1 data (WMLR)

Field-3

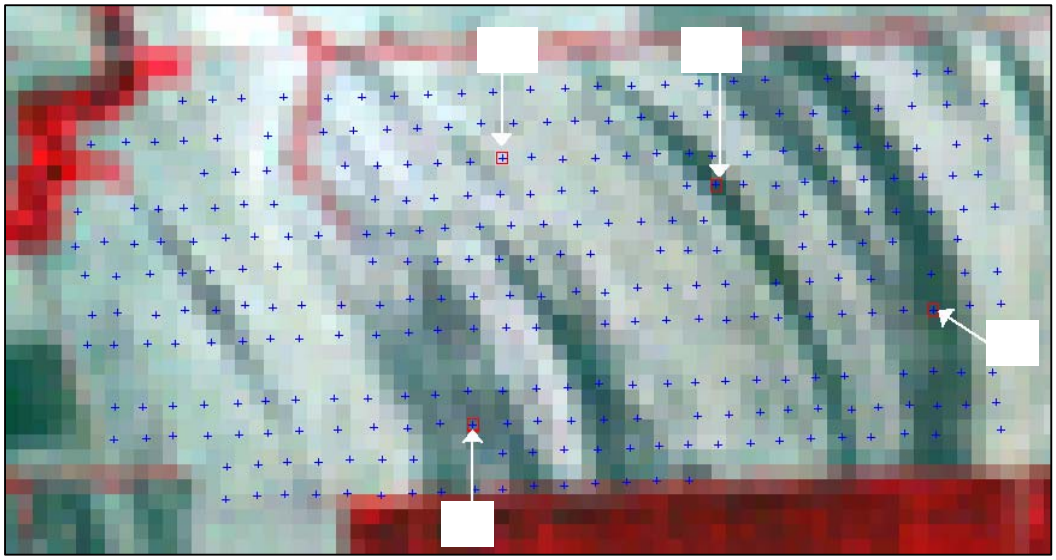
Tables 31 and 32 include details of outliers removed from WMLR analysis of 1999 and 2001 field-3 data, respectively. Four outliers were detected in bands 5 and 7 of the 1999 data. Except for points 212 (bands 5 and 7) and 98 (band 5), the DN values of all the observation were within ± 2 s.d. of the mean value of their respective bands. Points 56, 98, and 212 were located on a dark green patch in the false-color image (Figure 151), suggesting higher moisture content at that location. However, this was not the case with point 244. Outliers in the 2001 data were more numerous with WMLR analysis than with MLR. The DN values of all the outliers, except points 282 (bands 3, 4, 5, and 7) and 67 (bands 4 and 7), were within the approximate range of ± 2 s.d. of the mean of their respective bands. Points 282 and 67 had higher and lower DN values, respectively, than their respective band means. Similar to WSLR analysis, Figure 152 provided no obvious explanation for the outliers detected in the WMLR analysis. As mentioned previously, one- month-old cotton plants or precipitation may have influenced these data points.

Table 31: Detected outliers for all the studied bands of 1999 field-3 data and their corresponding actual radiance values (DN, Mean, Std. Dev), (WMLR)

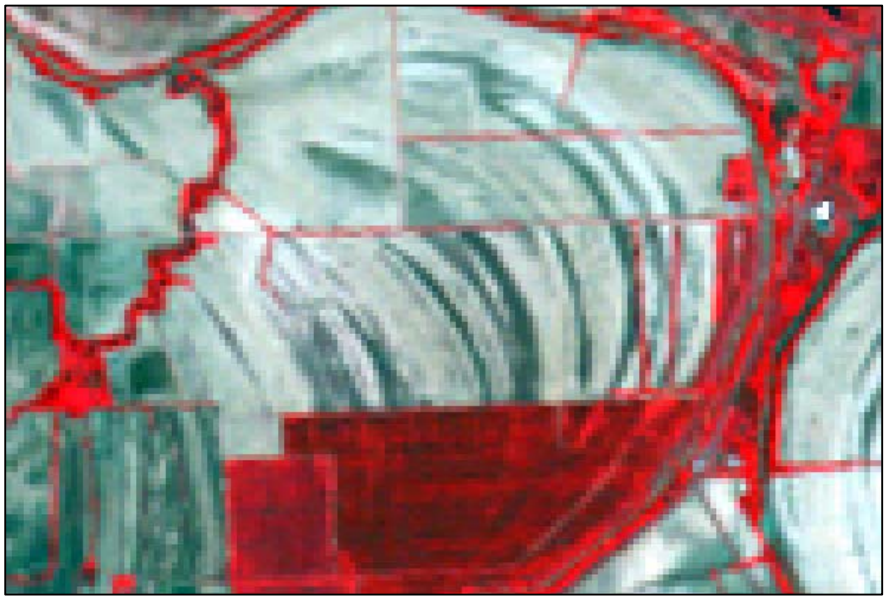
Bands	Obs. No.	Easting	Northing	Iteration	Actual Radiance Values		
					DN	Mean	Std. Dev
Band 5	56	742205.52	3772613.727	II	168	221.66	27.95
	98	743115.043	3772834.088	II	155		
	212	742686.168	3773072.193	I	153		
	244	742263.443	3773124.034	I	248		
Band 7	56	742205.52	3772613.727	I	84	130.15	23.88
	212	742686.168	3773072.193	I	74		
	244	742263.443	3773124.034	I	153		

Table 32: Detected outliers for all the studied bands of 2001 field-3 data and their corresponding actual radiance values (DN, Mean, Std. Dev), (WMLR)

Bands	Obs. No.	Easting	Northing	Iteration	Actual Radiance Values		
					DN	Mean	Std. Dev
Band 1	267	742096.151	3773180	I	105	98.4	4.078
	288	741919.511	3773237	I	107		
	296	741521.018	3773153	I	89		
Band 2	267	742096.151	3773180	I	97	88.97	5.391
	288	741919.511	3773237	I	101		
	296	741521.018	3773153	I	81		
Band 3	182	742660.054	3773006	II	120	109.68	7.982
	203	743202.344	3773093	II	99		
	282	742318.416	3773253	I	135		
	296	741521.018	3773153	I	94		
Band 4	67	741495.938	3772586	I	63	77.282	5.054
	282	742318.416	3773253	II	93		
	288	741919.511	3773237	I	86		
Band 5	243	742320.773	3773126	II	171	156.24	9.013
	244	742263.443	3773124	I	170		
	282	742318.416	3773253	II	184		
	288	741919.511	3773237	III	167		
	296	741521.018	3773153	I	147		
Band 7	67	741495.938	3772586	II	82	113.06	8.89
	282	742318.416	3773253	II	142		
	243	742320.773	3773126	I	128		
	244	742263.443	3773124	I	126		
	296	741521.018	3773153	I	99		

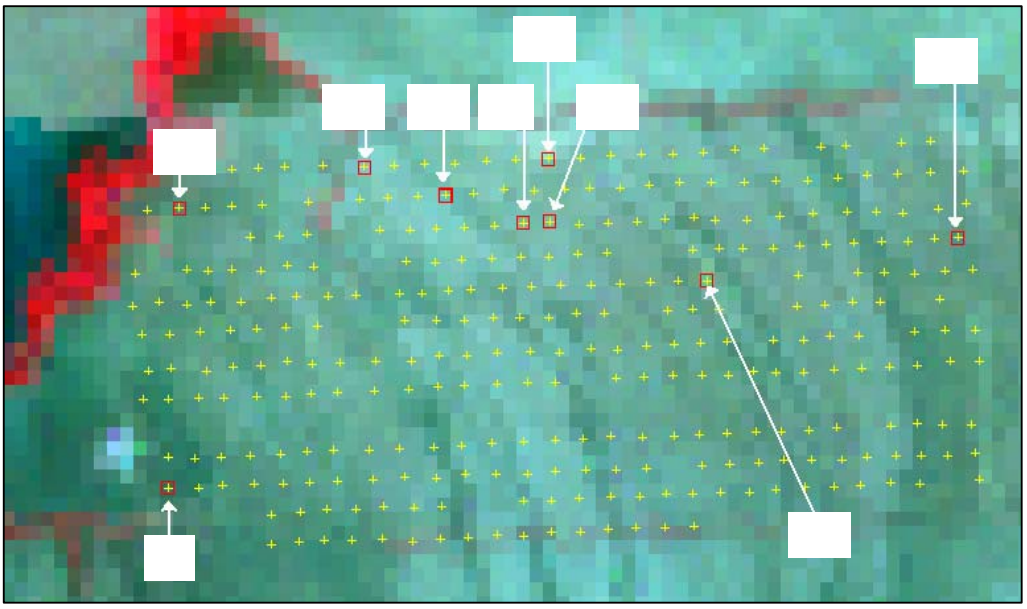


(a)



(b)

Figure151: Position of outliers on the close-up (a) and distant (b) false color Landsat images of 1999 field-3 data (WMLR)



(a)

(b)

Figure152: Position of outliers on the close-up (a) and distant (b) false color Landsat images of 2001 field-3 data (WMLR)

Pairwise independence

Figures 153 and 154 are semi-variogram plots for band-5 of 1997 field-1 and 2001 field-3 data, respectively. Both semi-variograms are given with and without outliers. In Figure 153, the semi-variogram without removal of outliers suggests that the curve approximates a horizontal line from the lag distance value of 700 m onwards, and before that, the line has a certain degree of slope. This shape indicates spatial correlation in the data between 0- and 700-m lag distances. After removal of outliers (Figure 153b), the degree of slope is much more prominent, and the amplitude appears to increase, but the range value of 700-m remained the same. The PROC MIXED analysis in Table 33 presents the fit statistics (-2 Res Log Likelihood, AIC, AICC, and BIC) for the WMLR analysis, both accounting for and not accounting for spatial correlation. It can be seen that for both cases, the fit statistics values were smaller when the spatial correlation was accounted for than when it was not accounted for, again suggesting significant spatial correlation in the model. Furthermore, when the outliers were not removed and the spatial correlation was not accounted for, the F and p values for reflectance, clay, and elevation were 58.10 and <0.0001, 134.03 and <0.0001, and 5.11 and 0.02, respectively. When spatial correlation was accounted for, these values became 40.23 and <.0001, 56.10 and <.0001, and 18.56 and <.0001 respectively. This means that accounting for spatial correlation did not influence the level of significance. When outliers were removed, the F and p values provided similar conclusions to those without outlier removal. In this case, when spatial correlation was not accounted for, the F and p values for reflectance, clay, and elevation were 46.09 and <.0001, 130.3 and <.0001, and 6.78

and 0.0098, respectively, while when spatial correlation was accounted for, the values became 29.60 and <.0001, 48.15 and <.0001, and 27.67 and <.0001, respectively. Again, accounting for spatial correlation did not affect the model's level of significance. Similar results were also observed for other bands

Table 33: Fit statistics for WMLR analysis of 1997 field-1 data

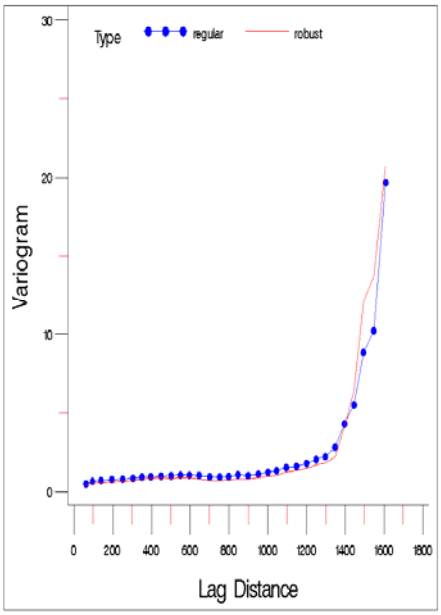
Model: B5 = β_0 + β_1Band5+ β_2Clay + β_3Elevation+ ε, 1997 (Landsat 5)				
Fit Statistics	Outliers not removed		Outliers removed	
	Spatial Correlation		Spatial Correlation	
	Not-Accounted	*Accounted*	*Not-Accounted*	*Accounted*
-2 Res Log Likelihood	1970.5	1844.9	1889.3	1761
AIC (smaller is better)	1972.5	1844.9	1891.3	1761
AICC (smaller is better)	1972.5	1844.9	1891.3	1761
BIC (smaller is better)	1976.1	1844.9	1894.9	1761

In the case of band-5 semi-variogram analysis of 2001 field-3 data, an approximate range value of 750 m was observed both with (Figure 154a) and without (Figure 154b) outliers. The overall amplitude of the curve seemed to decrease after removal of outliers, whereas the overall curve pattern remained the same. Hence, it is apparent that spatial correlation exists between data points 700 and 750 m apart in fields 1 and 3, respectively. Table 34 reveals that in the case of 2001 field-3 data, the results of the PROC MIXED analysis were somewhat similar to those with 1997 field-1 data. Before outlier removal, the F and p values for reflectance, clay, and elevation with spatial correlation unaccounted for were 76.95 and <.0001, 146.3 and <.0001, and 14.76 and 0.0001 respectively. With spatial correlation accounted for, these values became 32.12 and <.0001, 53.44 and <.0001, and 0.96 and 0.3283 respectively. This result indicates that spatial correlation did not affect the level significance for reflectance and clay, but elevation changed from being significant to

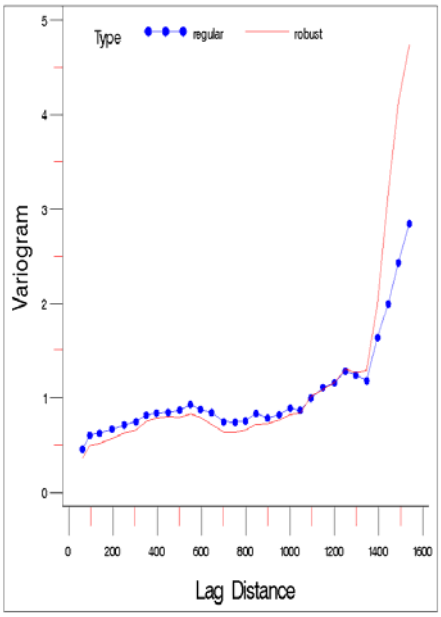
non-significant. Similar results were obtained after the outliers were removed, as the F and p values indicated that reflectance, clay, and elevation were significant when the spatial correlation was not accounted for, and when it was accounted for, elevation was no longer significant. The other bands of the 2001 field-3 data behaved similarly to band-5 except band-4, which started with a non-linear relationship in WMLR analysis and produced unusual and uncharacteristic results.

Table 34: Fit statistics for WMLR analysis of 2001 field-3 data

Model: $B5 = \beta_0 + \beta_1 Band5 + \beta_2 Clay + \beta_3 Elevation + \varepsilon$, 2001 (Landsat 7)				
Fit Statistics	Outliers not removed		Outliers removed	
	Spatial Correlation		Spatial Correlation	
	Not-Accounted	*Accounted*	*Not-Accounted*	*Accounted*
-2 Res Log Likelihood	2070.4	1907.2	1972.3	1794.5
AIC (smaller is better)	2072.4	1907.2	1974.3	1794.5
AICC (smaller is better)	2072.4	1907.2	1974.3	1794.5
BIC (smaller is better)	2076.1	1907.2	1977.9	1794.5

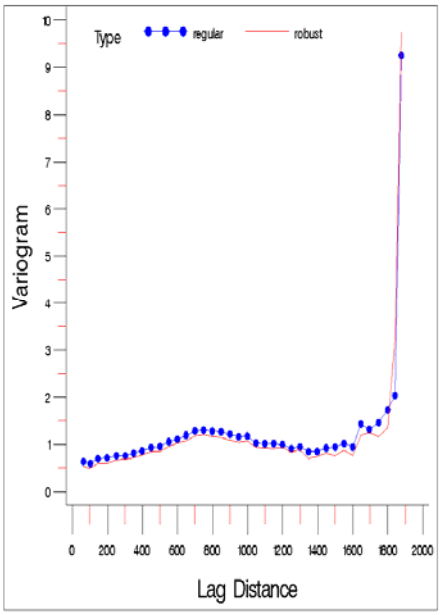


(a)

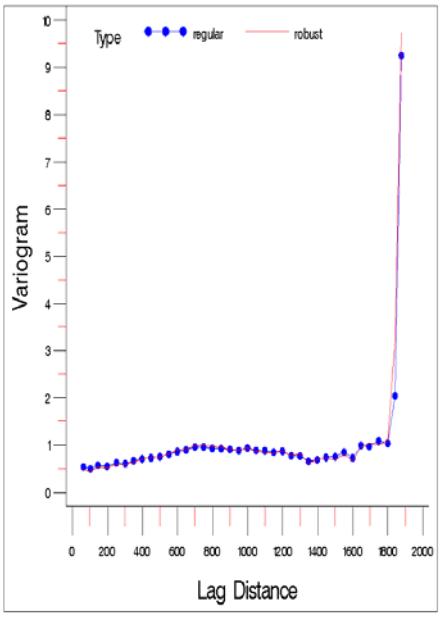


(b)

Figure153: Semi-variogram for band-5 of 1997 field-1 data before (a) and after (b) removal of outliers (WMLR)



(a)



(b)

Figure 154: Semi-variogram for band-5 of 2001 field-3 data before (a) and after (b) removal of outliers (WMLR)

CHAPTER IV

Summary, Conclusions, and Suggestions

Summary

Simulated Landsat reflectance, calculated from laboratory-based reflectance of soil samples, was statistically related with actual Landsat radiance data of bare soils. Two fields (field 1 and field 3) located at Vance, Mississippi, were selected for study. Soil samples, collected from the fields were dried and measured with a spectrophotometer to obtain soil reflectance. Reflectance values were multiplied with Landsat sensitivity values and integrated over the breadth of each Landsat band to obtain simulated Landsat reflectance. Landsat images taken in 1997, 1999, and 2001 were used to obtain radiance data from the soils. Soil texture and elevation data were collected to indirectly account for soil moisture content. Distance between pixel centroid and soil sample location was measured in order to study the influence of spatial resolution. Statistical analysis was then conducted on the collected data with simple linear regression (SLR), multiple linear regression (MLR), and weighted linear regression (WLR).

The SLR analysis determined significant relationships between simulated Landsat reflectance and actual Landsat radiance data for all Landsat bands of 1997 field-1 and 2001 field-3 data, except for band 4 of 2001 field-3 data. With 2001 field-1 and 1999 field-3 data, no significant linear relationship was found for any bands except band-3

of 2001 field-1 data and bands 2 and 3 of 1999 field-3 data. Even though relationships in 1997 field-1 and 2001 field-3 data were generally significant, the R^2 values were very low. The highest R^2 of 0.28 was observed for band-3 of 2001 field-3 data. The bands having a significant relationship satisfied the regression assumptions of normality and constant variance. Furthermore, few outliers were detected, and when removed the normality and constant variance assumptions appeared to be improved. Outlier removal in most bands resulted in about 1 to 3% improvement in R^2. The outliers were validated with available information on field conditions at the time of image collection, and with the images themselves. Most outliers were located at the periphery of the fields and probably were affected by topographical changes, unusual field characteristics at a specific location, drainage and other moisture issues, and dry vegetation such as Bermuda grass. The generally non-significant linear relationships found with 2001 field-1 and 1999 field-3 data can be attributed mainly to the presence of vegetation, possibly dry vegetation.

The MLR analysis, which indirectly accounted for the influence of moisture content by adding clay content and relative elevation as parameters in the SLR models, improved the results for 1997 field-1 and2001 field-3 data. The included variables, except elevation in field-1, had significant linear relationships with actual Landsat radiance from bare soils. The improvement in results for 2001 field-1 and 1999 field-3 data was due to elevation, since clay and reflectance had non-significant relationships with actual Landsat radiance in most of the bands. The highest R^2 value of 0.40 was observed with bands 3 and 5 in the 2001 field-3 data. The validation of regression assumptions for 1997 field-1 and 2001 field-3 data indicated that data in all

the studied bands were generally normally distributed and had constant variance. A few outliers were observed, and they were removed with the Bonferroni correction method. Removal of outliers improved the normality and constant variance assumptions for all the bands. Also, small improvements in R^2 were observed. The removed outliers were validated as described in the SLR section. Overall, the infrared bands had significant improvement with outlier removal; the fact that infrared bands are strongly influenced by moisture content was evident in this work. Even though R^2 values improved considerably after indirectly accounting for moisture content, R^2 values were still below 0.50,

The inverse of the distance measured from pixel centroid to sample location was applied as a weight to both the SLR and MLR models to indirectly consider the influence of spatial resolution. The WSLR analysis produced around 3 to 6% improvement in R^2 for 2001 field-3 data. The WMLR analysis produced around 6 to 8% and 3 to 6% improvements in R^2 for 1997 field-1 and 2001 field-3 data, respectively. Both WSLR and WMLR models satisfied the regressions assumptions of normality and constant variance. When observed outliers were removed, improvement in the normality and constant variance assumptions as well as in R^2 values was noted in most of the bands. The removed outliers were validated as with earlier analyses.

The pairwise independence (no spatial correlation) assumption, which was assumed to be true in previous sections, was tested for band 5 of 1997 field-1 and 2001 field-3 data. The results suggest that spatial correlation existed in both fields, so R^2 values may have been biased by the spatial correlation. However, p values (<0.0001) based on a mixed-model analysis indicated that the regression models

remained significant even when spatial correlation was accounted for. Hence, it can be deduced that if the sample location is closer to the pixel centroid, the sample will likely be more closely representative of the pixel. More importantly, it is reasonable to use inverse distance weighting in developing models for remotely sensed soil radiance. The result here tends to indicate that higher spatial resolution would result in higher correlations between remote-sensing and ground-based sample data.

The results of this study indicate a significant relationship between simulated Landsat reflectance based on laboratory reflectance and actual Landsat radiance from soils. They also indicate that the explanatory power of regression models can be improved substantially by accounting for soil moisture content by using clay content and relative elevation data, as well as by accounting for the poor spatial relationship between some ground samples and image pixels by inverse-distance weighting. The low R^2 values (< 0.5) after accounting for moisture content and spatial relationships can be attributed partially to error sources like vegetation, soil roughness, atmospheric effects, and directional reflectance. However, while the influence of spatial resolution was considered indirectly to an extent, it is likely that inverse-distance weighting could not fully explain the overall variability within a 30-m pixel, which could be another reason for the low R^2 values. Furthermore, clay content and relative elevation are not perfect surrogates for soil moisture content, and so variation in moisture content is likely still another reason. Inconsistency in results from one year's data to another could be due to different field and weather conditions in each year.

Conclusions

In short, the following conclusions were drawn from this study.

1. A significant linear relationship between simulated Landsat reflectance based on laboratory reflectance and actual Landsat radiance from bare soils was generally observed, but the amount of variation explained was low for all bands.
2. Including clay content and relative elevation as indirect soil moisture content parameters in regression models improved the models' ability to explain variation in the data and indicated that moisture content can be accounted for to some extent by including these parameters.
3. Bands 5 and 7 had the highest R^2 values, suggesting that soil moisture content is a major factor in remotely sensed soil variability, since these infrared bands are more strongly influenced by moisture than visible bands.
4. When relating ground-based to remotely sensed soils data, including data on the distance between sample location and pixel centroid is important, as it substantially improved the R^2 values of regression models in this study.
5. The presence of outliers can influence regression assumptions as well as R^2, and validly detected outliers if removed can improve the results.
6. Spatial correlation was observed in both fields, but significance levels of the regression models remained high even when spatial correlation was accounted for.

Suggestions for Future study

In order to improve correlations between expected sensor reflectance and actual remote-sensor radiance data of soils, it is essential to consider all the factors that influence remotely sensed spectral radiance from soils. Hence, aside from soil moisture content, factors like soil roughness, soil directional reflectance, and atmospheric corrections should be considered in the analysis. Also, the spatial resolution of the sensor should be selected in such a way that it can account for the inherent spatial variability in the soils. Furthermore, regression assumptions, especially spatial correlation, should be checked, and if problems with the assumptions are detected, they need to be dealt with before arriving at final conclusion.

REFERENCES

A. Al. Rajehy. 2002. Relationship between soil reflectance and soil physical and chemical properties. Unpublished Thesis, Mississippi State University.

Abdel-Hamid, M.A. 1993. Surface soil reflectance as a criteria of soil classification related to some physical and chemical soil properties. *Egypt. J. Soil Sci.*, 33(2): 149-162.

Agbu, P.A. 1990. Soil property relationship with Spot satellite digital data in East Central Illinois. *Soil Sci. Soc. Am. J.* 54(3): 807-811.

Agbu, P.A., D.J. Fehrenbacher, and I.J. Jansen. 1990. Statistical comparison of SPOT spectral maps with field soil maps. *Soil Sci. Soc. Am. J.* 54(3): 812-818.

Atzberger, C. 2002. Soil optical properties – a review. In: http://www.feut.uni-trier.de:8080/pdf/Kursbegleitung/radiometrie/Atzberger_Review_Soil_Reflectance.pdf, accessed February 27, 2005.

Ben-Dor, E., and A. Banin. 1994. Visible and near-infrared (0.4-1.1 micrometer) analysis of arid and semi-arid soils. *Rem. Sens. Environ.* 48 (3): 261-274.

Brivio, P.A., C. Giardino, and E. Zilioli. 2001. Validation of satellite data for quality assurance in lake monitoring applications. *Sci. Total Environ.* 268(1-3): 3-18.

Buscaglia, H.J. 2000. Spatial analysis of soil fertility and application of remote sensing for site-specific cotton production. Unpublished Dissertation. Mississippi State University.

Charman, P.E.V., and B.W. Murphy. 2000. *Soils: Their Properties and Management*. Melbourne: Oxford University Press.

Coleman, T. L., P.A. Agbu, and O.L. Montgomery. 1993. Spectral differentiation of surface soils and soil properties: Is it possible from space platforms?. *Soil Sci.* 155:283-293.

Dematte, J.A.M., and M.R. Nanni. 2003. Weathering sequence of soils developed from basalt as evaluated by laboratory (IRIS), airborne (AVIRIS) and orbital TM sensors. *Int. J. Rem. Sens.*, 24(23): 4715-4738.

Donahue, R.L., R.W. Miller, and J.C. Shickluna. 1983. *An Introduction to Soils and Plant Growth, 5th Ed*. Englewood Cliffs, N.J.: Prentice-Hall.

Epema, G.F. 1992. Atmospheric condition and its influence on reflectance of bare soil surfaces in southern Tunisia. *Int. J. Rem, Sens.*, 13 (5): 853-868.

Foth, H.D. 1984. *Fundamentals of Soil Science*. 7th edition, John wiley & sons

Littell, R.C. G.A.Milliken, W.W. Stroup, and R.D. Wolfinger 1996. *SAS System for Mixed Models*, Cary N.C.: SAS Institute Inc.

Hall, D.K., R.A. Bindschadler, J.L. Foster, A.T.C. Chang, and H. Siddalingaiah. 1990. Comparison of in-situ and satellite-derived reflectances of Forbindels Glacier, Greenland. *Int. J. Rem. Sens.*, 11(3): 493-504.

Huete, A. 2002. Soil optical properties and monitoring soil processes. In: http://tbrs.arizona.edu/education/553-2002/Soils_lecture.pdf, accessed February 21, 2005.

Hummel, J.W., K.A. Sudduth, and S.E. Hollinger. 2001. Soil moisture and organic matter prediction of surface and subsurface soils using an NIR soil sensor. *Comp. Elec. Agric.* 32(2):149-165.

Jacquemoud, S., F. Baret, and J.F. Hanocq. 1992. Modeling spectral and bi-directional soil reflectance. *Rem. Sen Environ.* 41(2/3):123-132.

Lee, K-S., and W.B. Cohen. 2002. Comparison of AVIRIS and Landsat ETM+ for the estimation of leaf area index. In: Proc. 2002 AVIRIS Workshop. Pasadena, Calif., NASA-JPL.

Mattikalli, N.M. 1997. Soil color modeling for the visible and near-infrared bands of Landsat sensors using laboratory spectral measurements. *Rem. Sens. Environ.* 59(1):14-28.

Morgan, M., and D. Ess 1997. *The Precision Farming Guide for Agriculturists*. Moline, Ill: John Deere.

Price, J.C. 1990. On the information content of soil reflectance spectra. *Rem. Sens. Environ.* 33(2): 113-121.

Rawling, J.O., S.G. Pantula, and D.A. Dickey. 1998. *Applied Regression Analysis, 2nd Ed*. New York: Springer-Verlag.

Reed S, N. Bailey, and O. Onokpise. 2000. Soil science for archaeologists. In: http://www.cr.nps.gov/seac/soilsmanual/cover.htm accessed March 5, 2005.

Rios, J.J.M., and H.C. Monger. 2002. Soil classification in arid lands with thematic mapper data. *Terra* 20(2): 89-100.

Stoner, E.R., M.F. Baumgardner, R.A. Weismiller, L.L. Biehl, and B.F. Robinson. 1980. Extension of laboratory-measured soil spectra to field conditions. *Soil Sci. Soc. Am. J.* 44: 572-574.

SAS 9.1 for Windows, SAS user help, Version 5.0.2195. SAS Institute Inc., Cary, NC, USA.

Thomasson, J.A., R. Sui, M.S. Cox, and A. Al-Rajehy. 2001. Soil reflectance for determining soil properties in precision agriculture. *Trans. ASAE* 44(6): 1445-1453.

Thome, K.J. 2001. Absolute radiometric calibration of Landsat 7 ETM+ using the reflectance-based method. *Rem, Sens. Environ.* 78:27-38.

Thompson, D.R., K.E. Henderson, A.G. Houston, and D.E. Pitts. 1984. Variation in alluvial-derived soils as measured by Landsat Thematic Mapper. *Soil Sci. Soc. Am. J.* 48: 137-142.

Thompson, D.R., D.E. Pitts, and K.E. Henderson. 1983. Simulation of Landsat Multispectral Scanner response of soils using laboratory reflectance measurements. *Soil Sci. Soc. Am. J.* 47: 542-546.

Williams J., R.E. Prebble, W.T.Williams, and C.T. Hignett. 1983. The influence of texture, structure and clay mineralogy on the soil moisture characteristic. *Aus. J. Soil Res.* 21: 15:32.

Wooten, J.R. May through June, 2005. Personal communication on outlier validation. Research Associate I, Department of Agricultural and Biological Engineering, Mississippi State University.

Zhou, X., and S. Li. 2003. Comparison between in situ and MODIS-derived spectral reflectances of snow and sea ice in the Amundsen Sea, Antarctica. *Int. J. Rem. Sens.* 24(24): 5011-5032