

## Article

# Quantifying Information Content in Multispectral Remote-Sensing Images Based on Image Transforms and Geostatistical Modelling

Ying Zhang <sup>1</sup>, Jingxiong Zhang <sup>2,\*</sup> and Wenjing Yang <sup>3</sup>

<sup>1</sup> College of Geo-Exploration Science and Technology, Jilin University, Changchun 130026, China; zhangying\_cc@jlu.edu.cn

<sup>2</sup> School of Geodesy and Geomatics, Wuhan University, Wuhan 430079, China

<sup>3</sup> College of Geography and Environment, Shandong Normal University, Jinan 250358, China; ywjsdnu@163.com

\* Correspondence: jxzhang@whu.edu.cn; Tel.: +86-27-6877-8683

Received: 18 January 2020; Accepted: 6 March 2020; Published: 9 March 2020



**Abstract:** Quantifying information content in remote-sensing images is fundamental for information-theoretic characterization of remote sensing information processes, with the images being usually information sources. Information-theoretic methods, being complementary to conventional statistical methods, enable images and their derivatives to be described and analyzed in terms of information as defined in information theory rather than data per se. However, accurately quantifying images' information content is nontrivial, as information redundancy due to spectral and spatial dependence needs to be properly handled. There has been little systematic research on this, hampering wide applications of information theory. This paper seeks to fill this important research niche by proposing a strategy for quantifying information content in multispectral images based on information theory, geostatistics, and image transformations, by which interband spectral dependence, intraband spatial dependence, and additive noise inherent to multispectral images are effectively dealt with. Specifically, to handle spectral dependence, independent component analysis (ICA) is performed to transform a multispectral image into one with statistically independent image *bands* (not spectral bands of the original image). The ICA-transformed image is further normal-transformed to facilitate computation of information content based on entropy formulas for Gaussian distributions. Normal transform facilitates straightforward incorporation of spatial dependence in entropy computation for the aforementioned double-transformed image bands with inter-pixel spatial correlation modeled via variograms. Experiments were undertaken using Landsat ETM+ and TM image subsets featuring different dominant land cover types (i.e., built-up, agricultural, and hilly). The experimental results confirm that the proposed methods provide more objective estimates of information content than otherwise when spectral dependence, spatial dependence, or non-normality is not accommodated properly. The differences in information content between image subsets obtained with ETM+ and TM were found to be about 3.6 bits/pixel, indicating the former's greater information content. The proposed methods can be adapted for information-theoretic analyses of remote sensing information processes.

**Keywords:** information content; spectral dependence; spatial dependence; independent component analysis; normal transform; variograms; additive noise

## 1. Introduction

Remote-sensing images have become major sources of spatial information. These images are formed by recording the reflected radiance or energy from a scene. A typical digital image consists of a

two-dimensional array of pixels, each representing the average reflectance, emittance, or backscattering of the surface within the instantaneous field of view. The images are used, often in combination with other images and ancillary data (e.g., maps), to detect the presence of certain phenomena, map their spatial extents, and estimate certain biophysical variables, as reviewed by Zhang et al. [1]. Statistical models and methods are conventionally employed for descriptions and analyses of source images, ancillary data, and image or map derivatives (as intermediate or end results) in remote sensing information processes. It is increasingly recognized that an information-theoretic perspective [2–4] is complementary to conventional statistical methods for characterizing remote sensing information processes (e.g., measurement, data fusion, feature selection, image classification, and information retrievals) [5–11], because source images, data processing, information retrievals, and resultant information products can then be described and analyzed based on a unifying norm of information bits as formally defined in information theory [2,3], along with statistical measures of variability. Performance limits of models and methods for remote sensing information processing can also be usefully evaluated via information theory [12–16].

To build an information-theory-based framework for analyzing remote sensing information processes, it is important to develop methods for quantifying information content in multispectral images, which are usually starts of remote sensing information chains, and whose information content sets the upper bounds [3,16] for the amount of information available in subsequent information processes in remote sensing. Such methods should preferably be transferable for use in analyses of subsequent information processes in remote sensing, such as image classification. Thus, the aim of this paper is to develop suitable methods for quantifying information content in multispectral images as a basis for information-theoretic analyses of remote sensing information processes systematically.

Remote-sensing images (optical ones, in particular) record reflectance measurements, which are related to the surface's geophysical properties but subject to atmospheric effects and other factors, as mentioned above. To quantify information in multispectral images, we need to look into how images are acquired from a (remote) measurement system and then analyze how information flows through this measurement process (also an image formation process [6]) from the underlying surface phenomenon (source or input) through the sensor (channel) to a resultant image (output).

As discussed in [12], a measurement system includes geospatial fields or objects to be measured, a measurement mechanism, and an observer. Below, we adopt a geostatistical framework for describing the quantities involved. Suppose that a vector-valued random field (RF) [17]  $\mathbf{S}$  refers to parameters characterizing what is measured over a problem domain  $D$  (e.g., the underlying multispectral intensity field over  $D$ , denoted as  $\mathbf{S} = \{\mathbf{S}(x), x \in D\}$ ). The measurement mechanism maps  $\mathbf{S}$  to another vector-valued RF  $\mathbf{Z}$  (e.g., the observed multispectral intensity image data (i.e., measurements),  $\mathbf{Z} = \{\mathbf{Z}(x), x \in D\}$ ). When considering a specific location (i.e., a pixel  $x$ ),  $\mathbf{Z}(x)$  and  $\mathbf{S}(x)$  are random vectors. For a particular band  $b$  in the context of multispectral images, we will have either scalar RFs  $Z_b$  and  $S_b$  over  $D$  or random variables  $Z_b(x)$  and  $S_b(x)$  at a specific pixel  $x$ . Due to inherent uncertainty in the measurement mechanism, the measurements  $\mathbf{Z}$  may contain errors (i.e., noises). These noises are typically additive for optical images and often assumed to be Gaussian and independent of their signal components [12,16]. The observer observes  $\mathbf{Z}$  and determines the information about  $\mathbf{S}$  from  $\mathbf{Z}$ .

In languages of information theory, the aforementioned measurement process can be studied via entropy, mutual information, and other information measures [3]. Entropy indicates how easily images (as messages) can be compressed [18,19]. On the other hand, mutual information measures the information about phenomena on the Earth's surface from images [20,21], thus being more relevant to the problem here. For a multispectral image  $\mathbf{Z}$  of underlying signal  $\mathbf{S}$  described above, mutual information between  $\mathbf{Z}$  and  $\mathbf{S}$  (denoted by  $I(\mathbf{S}, \mathbf{Z})$ ) quantifies the amount of information that  $\mathbf{Z}$  provides about  $\mathbf{S}$ . The greater this mutual information, the more information  $\mathbf{Z}$  conveys about  $\mathbf{S}$ , as discussed also in the context of radar remote sensing by [12]. An information-theoretic framework for quantifying information content in images thus lies in properly estimating mutual information between the images and the underlying surface properties.

However, quantifying information content in  $\mathbf{Z}$  about  $\mathbf{S}$  is far from straightforward, as we need to consider inter-band spectral dependence (as  $\mathbf{Z}$  and  $\mathbf{S}$  are both multispectral), intra-band spatial dependence (as  $\mathbf{Z}$  and  $\mathbf{S}$  are both considered over  $D$  as a whole, not just at a particular pixel  $x$ ), and noise (due to uncertainty in the measurement mechanism, as mentioned above). Spectral and spatial dependence imply information redundancy and should be taken into account for objective evaluation of image information content.

There have been research efforts to address these issues to some extent. To handle spectral dependence (redundancy), principal component analysis (PCA) was used to decorrelate image bands for estimating information content in image data [22,23]. However, PCA does not take image data's spatial structure into account, meaning that any permutation of image pixels would produce the same principal components. As opposed to PCA, maximum noise fraction (MNF) may be used to decompose multispectral satellite images into fraction image bands that are orthogonal linear transforms of the original image bands and show decreasing image quality with increasing component number [24,25].

There has also been work done to address spatial dependence specifically. For example, in [21], spatial redundancy was estimated by use of the difference operator which replaces each value except the first of a sequence by its difference from the preceding value, while PCA was used to handle spectral redundancy, as mentioned above. Razlighi et al. [26] proposed a new model, named quadrilateral Markov random field (MRF), to compute spatial entropy for images (single-band) and spatial mutual information for two images (again, single-band), in which spatial redundancy is accounted for.

To cope with both spectral and spatial redundancy, Wang et al. [27] described methods for spectral and spatial decorrelation in lossless data compression of remote-sensing images whereby optimal band combination and band ordering can be efficiently computed based on the statistical properties of Landsat-TM data. Experiments on several Landsat-TM images show that using both the spectral and the spatial nature of the remotely sensed data results in significant improvement (with higher compression ratios) over spatial decorrelation alone. More recently, for decorrelation of multispectral images, Aiazzi et al. [28] proposed a reversible compression method based on differential pulse code modulation (DPCM). They extended a one-dimensional fixed DPCM to the case of a three-dimensional adaptive prediction for reversible inter-band compression. The modeling of prediction errors from DPCM was based on the generalized Gaussian density (GGD) function. The scatter plot method was used for noise estimation. Both noise variance and entropy model parameter were used for estimating the amount of “useful” information content in multispectral images.

However, decorrelation procedures implemented by Price [22], Wang et al. [27], and Aiazzi et al. [28] may not lead to spectral or spatial independence, as uncorrelatedness is not equivalent to independence [29]. Alternative methods that can effectively handle the issue of dependence (implying information redundancy) are required for accurate estimation of information content. As will be elaborated below, this paper proposes using independence component analysis (ICA) [29] to decompose a multispectral image (observed, with inter-band spectral dependence) into independent components (sources) whose mixtures recover the observed image. This allows for computing information content (entropy) in the original image as a sum of information content in transformed image *bands* (though with an offset term). With ICA transformation, there is no assurance that the resulting transformed data pertain to spectral bands. Thus we denote transformed image bands in italic (i.e., *bands*) to differentiate them from the original image bands. In the text, this should be noted in use of bands for transformed image data. This method for handling spectral dependence is an improvement over existing methods in theoretic terms, since the decomposed components are meant to be statistically independent. It is interesting to note that informational criteria are employed for ICA decomposition [29]. Rationales for using ICA lie also in its widening applications in image-based information processing, such as feature selection and classifier optimization [30] and multivariate spatial analyses and modeling, especially when in combination with geostatistics [31] and geospatial information [32,33].

As for spatial dependence modeling, the quadrilateral MRF model proposed by Razlighi et al. [26] is not directly applicable for multispectral images. In addition, there are assumptions made in the

proposed MRF model, which may become restrictive in certain cases. There seem to be merits in modeling spatial correlation from a geostatistical perspective, which is less restrictive and can also provide an approach to estimate noise variance in images by way of estimating variogram model parameters, as magnitude of the nugget effect might be treated as an estimate of the noise variance.

To facilitate analytical computation of information content, the aforementioned ICA-decomposed image *bands* are further normal-transformed. This also allows spatial dependence to be accommodated easily without having to go through complicated modeling procedures as in [26].

Based on the review and reasoning above, this paper proposes a strategy based on image transformations and geostatistics for quantifying information content in multispectral images. The transformations are for generating independent components (i.e., *bands*) from an original image and for transforming these independent *band* data to normal distributions. The double transformations (i.e., independence- and normality-transformations) facilitate easy computation of the transformed image's (joint) entropy through evaluation of individual transformed image's *bands* entropies followed by summation, while the original image's entropy value and those of the transformed images are determined properly. Geostatistics is applied to quantify spatial correlation through variogram modeling (estimating variogram model parameters, to be precise; this should be noted in the remainder of the text) in the double-transformed image *bands* so that their entropies can be computed using Gaussian models with spatial dependence explicitly accommodated. In addition, variogram modeling provides estimates of nugget variance (in original image *bands*), which can be used for estimating mutual information between the original image and its signal component. Thus, inter-band spectral dependence, intra-band spatial dependence, and (additive) noise are well dealt with in the proposed strategy.

The paper is organized as follows. In Section 2, the methods for estimating various information measures concerning multispectral images are described. Experiments with Landsat TM+ and TM image subsets are presented in Section 3 where experimental data and results are described along with experimental procedures. This is followed by some discussion before the conclusions of the paper.

## 2. Methods

As mentioned in Section 1, the proposed strategy for estimating information content in multispectral images consists of the following steps: (1) ICA transform for estimating statistically independent *bands* from a given original multispectral image; (2) normal transform to acquire independent and normal-distributed image *bands*, with proper adjustment for entropies before and after the normality transform; (3) estimation of information content in ICA and normal-transformed image *bands* with accommodation for spatial correlation, which is quantified through geostatistical modeling; and (4) estimating noise variance in the original image *bands* as nugget effects in their variograms so that mutual information between the original image and its signal component can be computed. We describe the procedures in this section after a brief description of the assumed image data model and, more fundamentally, the underlying assumptions concerning stationarity.

Remote-sensing images (optical ones, in particular) record reflectance measurements related to the surface's geophysical properties but are subject to atmospheric effects and other factors, as mentioned in Section 1. We treat these factors of uncertainty as having generated, in the observed spectral data, the component of noise, which is usually assumed additive and independent of the signal component [5,12,16]. Consider a  $B \times P \times Q$  multispectral image ( $P$  is the number of columns,  $Q$  is the number of rows, and  $B$  is the number of bands) with additive noise:

$$\mathbf{Z}_{B \times P \times Q}(x) = \mathbf{S}_{B \times P \times Q}(x) + \mathbf{N}_{B \times P \times Q}(x) \quad (1)$$

where  $\mathbf{Z}_{B \times P \times Q}(x) = (Z_1(x), \dots, Z_B(x))^T_{P \times Q}$ ,  $\mathbf{S}_{B \times P \times Q}(x) = (S_1(x), \dots, S_B(x))^T_{P \times Q}$ , and  $\mathbf{N}_{B \times P \times Q}(x) = (N_1(x), \dots, N_B(x))^T_{P \times Q}$  represent the multispectral intensity measurements (i.e., the image data), their

signal components, and the noise components, respectively. In the remainder of this paper, we use  $\mathbf{Z}$ ,  $\mathbf{S}$ , and  $\mathbf{N}$ , without indicating the pixel locator  $x$ , for compact notation.

As mentioned in Section 1, we adopt a geostatistical framework for describing  $\mathbf{Z}$ ,  $\mathbf{S}$ , and  $\mathbf{N}$ . In geostatistics, the data is assumed to be a nonrandom sample from one realization of a random function (RF) [17,34]. Moreover, the RF is assumed to satisfy a stationarity assumption. For an RF, there are at least three possible forms of stationarity; strong (meaning that the probability distribution is invariant with respect to translation and rotation), second order (meaning that the covariance function exists and is invariant with respect to translation), and intrinsic (order zero) (meaning the variance of first-order increments exists and is invariant with respect to translation). Intrinsic is the weakest, second order the next weakest, and strong the strongest [34]. For the RF, there is not just a single probability distribution (except in the case of strong stationarity) but rather one for each point in the underlying space. Strong stationarity does not imply the existence of either first or second-order moments.

For this research, we assumed strong stationarity for each band as well as for the transformed *bands* and also the existence of second-order moments for first-order increments (which is implicit in the use of transformations to normality described later on). This is necessary because estimation of entropy for an image (band) requires knowledge of its probability distribution function being stationary across the image domain (or subdomains if local stationarity is assumed, as is the case for this research). More importantly, with limited experimental image datasets employed in this research, no statistical tests would be possibly performed for stationarity; trying to use the data sets to discern whether strong stationarity is a reasonable assumption is simply not possible [34].

Given the image data model and assumptions concerning stationarity and existence of second-order moments above, we can compute mutual information  $I(\mathbf{S}, \mathbf{Z})$ , which, as mentioned previously, quantifies the amount of information that image  $\mathbf{Z}$  provides about signal  $\mathbf{S}$  and can be computed as:

$$I(\mathbf{S}, \mathbf{Z}) = h(\mathbf{Z}) - h(\mathbf{Z}|\mathbf{S}) = h(\mathbf{Z}) - h(\mathbf{N}) \quad (2)$$

where  $h$  represents (differential) entropy,  $h(\mathbf{Z})$  denotes the entropy of the image data, and  $h(\mathbf{Z}|\mathbf{S})$  represents the conditional entropy given the knowledge of signal  $\mathbf{S}$  (which equals  $\mathbf{N}$ 's entropy due to independence between  $\mathbf{S}$  and  $\mathbf{N}$ ) [2,3]. As noises in individual bands are Gaussian and independent, their entropy can be easily computed as:

$$h(\mathbf{N}) = \sum_{b=1}^B h(\mathbf{N}_b) = \frac{1}{2} \sum_{b=1}^B [\log_2(\det(2\pi e \Sigma_{\mathbf{N}_b}))] \quad (3)$$

where  $\Sigma_{\mathbf{N}_b}$  is band  $b$ 's noise variance.

Due to spectral dependence among image bands and spatial dependence among pixels, quantifying of  $h(\mathbf{Z})$  and hence  $I(\mathbf{S}, \mathbf{Z})$  (Equation (2)) is hardly trivial. A sensible strategy is to decompose a given multispectral image into statistically independent *bands*, which can then be analyzed separately with respect to quantifying information. Such decomposition can be based on ICA, which is developed for blind signal separation [29]. The ICA-transformed image's (joint) entropy can then be computed as the sum of individual bands entropies, with the original image's entropy evaluated easily, as elaborated below.

Let us denote the  $B$ -dimensional ICA transform of an original multispectral image ( $B$ -band) by  $\mathbf{Z}' = [Z'_1, Z'_2, \dots, Z'_B]^T$  (for a pixel at location  $x$  as in Equation (1)). In ICA, a weight matrix  $\mathbf{W}$  is determined so that the individual *bands* of the resulting ICA-transformed image  $\mathbf{Z}'$  are statistically independent:

$$\mathbf{Z}' = \mathbf{W}\mathbf{Z} \quad (4)$$

Suppose that we have applied an ICA transform (with a weight matrix  $\mathbf{W}$ ) to  $\mathbf{Z}$  and obtained ICA-transformed image  $\mathbf{Z}'$ . For any nonsingular matrix  $\mathbf{W}$ , according to information theory, we have [2,3]:

$$h(\mathbf{Z}) = h(\mathbf{Z}' = \mathbf{W}\mathbf{Z}) - \log_2 |\det(\mathbf{W})| = \sum_{b=1}^B h(\mathbf{Z}'_b) - \log_2 |\det(\mathbf{W})| \quad (5)$$

where  $||$  indicates absolute value and  $\det$  determinant, and  $-\log_2 |\det(\mathbf{W})|$  is the difference between joint differential entropies  $h(\mathbf{Z})$  and  $h(\mathbf{Z}')$ . In Equation (5), the joint entropy of  $\mathbf{Z}'$  can be computed as  $h(\mathbf{Z}') = \sum_{b=1}^B h(\mathbf{Z}'_b)$  [3], as individual *bands* in the transformed image  $\mathbf{Z}'$  are assumed to be statistically independent.

Estimation of individual *bands* entropy for ICA-transformed image data is not straightforward except for Gaussian data [2,3]. To accommodate non-normal distribution, a sensible way is to transform non-normal data to normal data with relations between entropy values before and after the transform determined properly. For normality transform, Johnson transformation is recommended [35], as it is able to transform a continuous univariate vector to a random vector from standard normal distribution. The algorithm actually fits the data to one of the following three functional forms ( $S_b$ ,  $S_l$  and  $S_u$ ):

$$\begin{aligned} S_b : \mathbf{Z}''_b &= \gamma + \eta * \ln((\mathbf{Z}'_b - \varepsilon) / (\lambda + \varepsilon - \mathbf{Z}'_b)) \\ S_l : \mathbf{Z}''_b &= \gamma + \eta * \ln(\mathbf{Z}'_b - \varepsilon) \\ S_u : \mathbf{Z}''_b &= \gamma + \eta * a \sinh((\mathbf{Z}'_b - \varepsilon) / \lambda) \end{aligned} \quad (6)$$

where  $\mathbf{Z}''_b$  is a normal-transformed image band,  $\mathbf{Z}'_b$  is the corresponding non-normal image *band* (ICA-transformed, Equation (4)), and  $\gamma$ ,  $\eta$ ,  $\varepsilon$ , and  $\lambda$  are the necessary parameters. The relationships between  $\mathbf{Z}'_b$  and  $\mathbf{Z}''_b$  in Equation (6) can be summarized as  $\mathbf{Z}''_b = G(\mathbf{Z}'_b)$ . The R package *jtrans* (<https://cran.r-project.org/web/packages/jtrans/>) uses an Anderson-Darling test for normality after these transforms. The fit with the largest  $p$ -value is accepted (normality is the Null Hypothesis, and a large  $p$ -value implies a small Type II error probability).

We can derive the general expression relating entropy values before and after transformation as [36]:

$$h(\mathbf{Z}'_b) = h(\mathbf{Z}''_b) - \int f(z''_b) \log_2 g'(z'_b) dz''_b = h(\mathbf{Z}''_b) - \Delta h_b \quad (7)$$

where  $h(\mathbf{Z}'_b)$  and  $h(\mathbf{Z}''_b)$  denote the entropy values (for band  $b$ ) before and after normal transform, respectively. In Equation (7),  $\Delta h_b$  is the difference between differential entropy values before and after the transform and can be computed by noting that the relationship between the probability density functions (PDFs) of  $\mathbf{Z}''_b$  and  $\mathbf{Z}'_b$ ,  $f(z''_b)$  and  $f(z'_b)$ , is  $f(z'_b) = |g'(z'_b)| f(z''_b)$ , for monotonic transformations. Entropy of an ICA- and normal-transformed image single-band data  $\mathbf{Z}''_b$  is  $\frac{1}{2} \log_2 (\det(2\pi e \Sigma_{\mathbf{Z}''_b}))$ , where  $\Sigma_{\mathbf{Z}''_b}$  is the covariance matrix of  $\mathbf{Z}''_b$  [3]. Thus, the relation between  $h(\mathbf{Z}')$  and  $h(\mathbf{Z}'')$  is:

$$h(\mathbf{Z}') = h(\mathbf{Z}'') - \Delta h = \frac{1}{2} \sum_{b=1}^B \left[ \log_2 (\det(2\pi e \Sigma_{\mathbf{Z}''_b})) \right] - \sum_{b=1}^B \Delta h_b \quad (8)$$

While noise  $\mathbf{N}_b$  may well be assumed to be spatially independent as indicated in Equation (3), there exists inter-pixel spatial dependence in  $\mathbf{Z}''_b$ . Variogram models describe spatial relationships in spatial data including remote-sensing images [37–39]. Experimental variogram for the aforementioned ICA- and normal-transformed image *band* data  $\mathbf{Z}''_b$  can be written as:

$$\gamma_{\mathbf{Z}''_b}(h) = \frac{1}{2N_h} \sum_{i=1}^{N_h} [\mathbf{Z}''_b(x) - \mathbf{Z}''_b(x+h)]^2 \quad (9)$$

where  $\mathbf{Z}''_b(x)$  and  $\mathbf{Z}''_b(x+h)$  are  $\mathbf{Z}''_b$  values at pixels  $x$  and  $x+h$ , respectively,  $h$  is the lag, and  $N_h$  is the number of pairs of pixels at lag  $h$ . Variogram model parameters will then be used to build a

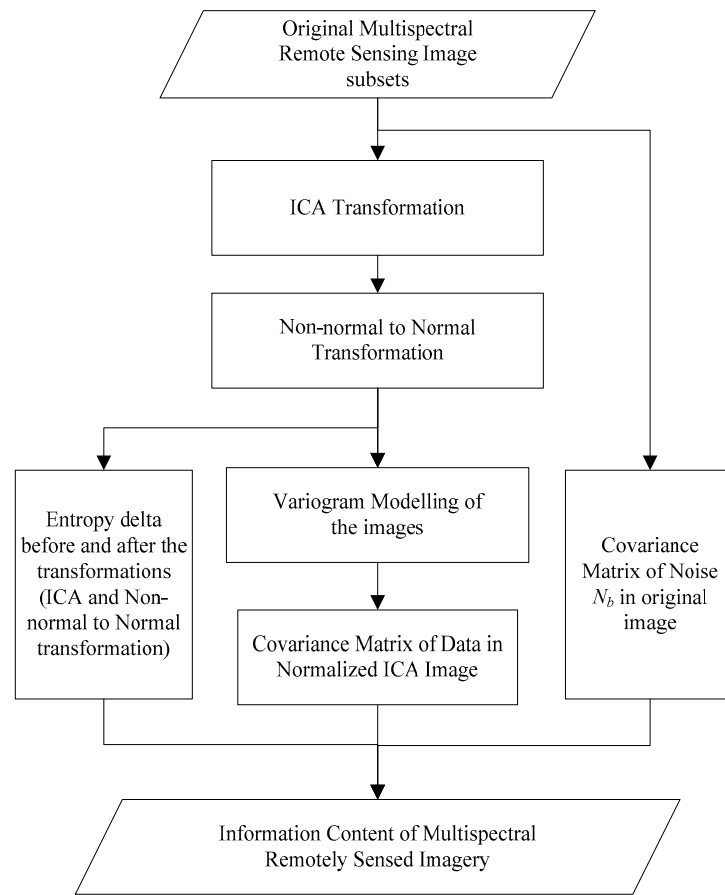
covariance function to populate the variance and covariance matrix  $\Sigma_{Z_b''}$  (for individual *band* data  $Z_b''$ ) for estimating  $h(Z_b'')$  (Equation (8); if the variogram has a sill then  $sill - \gamma(h)$  is a covariance function). Since pixel data are not pure point-support data but are of finite areal support, locations  $x$  and  $x + h$  refer actually to pixel centroids. Based on variogram model parameter estimation, we can also get estimates for nugget effects, which can be used as estimated noise variances [40], allowing for evaluation of noise entropy. It should be noted that, while one possible interpretation of the value of the sill of a nugget component of a variogram is a noise variance, it is not the only possible interpretation, especially with data on pixels. However, we restrict our discussion by treating pixel data as quasi-point data, since we consider the information content of images with respect to their signal components (see Equation (1)), which are themselves on finite areal support.

Joint entropy of the original multispectral image  $Z$  can be computed by combining Equations (5) and (8). To compute the amount of information transmitted by the original image data  $Z$  about its signal component  $S$  (i.e., mutual information between  $Z$  and  $S$ ), we rewrite Equation (2) as:

$$\begin{aligned} I(S, Z) &= \sum_{b=1}^B \left[ h(Z_b'') \right] - \log_2 |det(W)| - \Delta h - h(N) \\ &= \frac{1}{2} \sum_{b=1}^B \log_2 \left[ 2\pi e \det(\Sigma_{Z_b''}) \right] - \log_2 |det(W)| - \Delta h - h(N) \end{aligned} \quad (10)$$

where Equation (3) is used to estimate joint entropy of noise  $h(N)$ .

A flowchart is presented in Figure 1. As shown in Figure 1, first, ICA transform is conducted to decompose the original (multispectral) images into images with independent *bands*. This is done separately for each of the original images. Second, normal transform is carried out to obtain ICA- and normal-transformed images. Before doing this, normality test is needed to check the non-normality of the individual ICA *bands*. This is also done separately for each *band* of each ICA-transformed image. Third, for each *band* in each of the ICA- and normal-transformed (i.e., double-transformed) images, variogram modelling is performed so that covariance matrices corresponding to the individual double-transformed image *bands* can be built with estimated variogram model parameters. Finally, information content in individual *bands* of the double-transformed image can be computed (by taking advantages of multivariate normal distributions) with spectral and spatial dependences accounted for. Joint entropy of the original multispectral image can be estimated based on Equations (5) and (8), with mutual information  $I(S, Z)$  estimated using Equation (10).

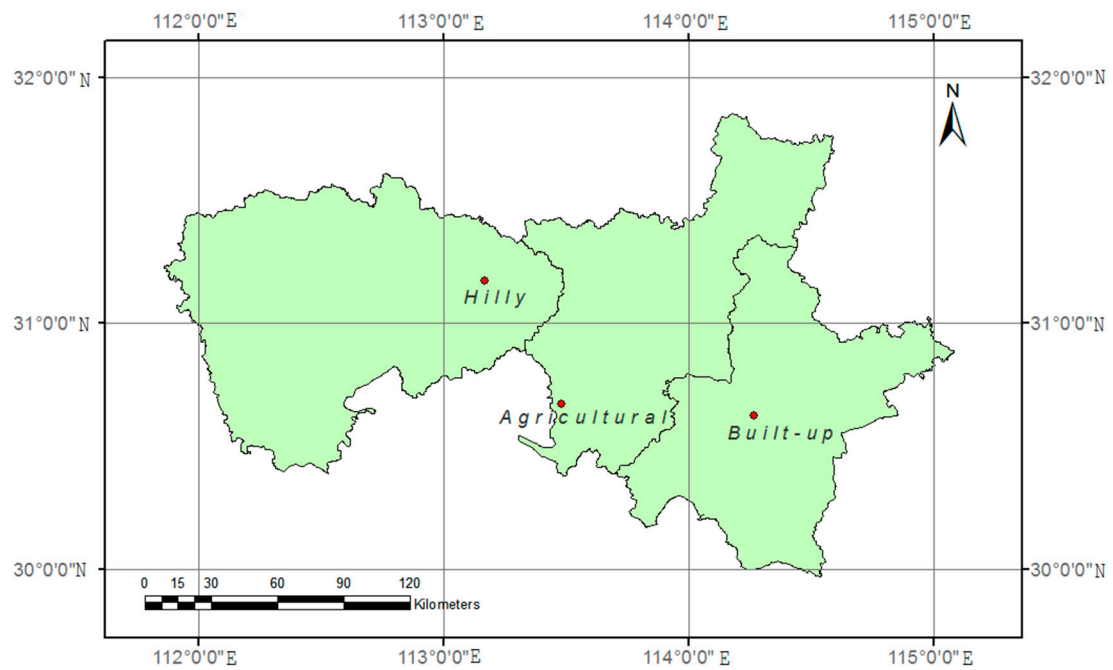


**Figure 1.** The flowchart describing the methods for quantifying information content in multispectral images.

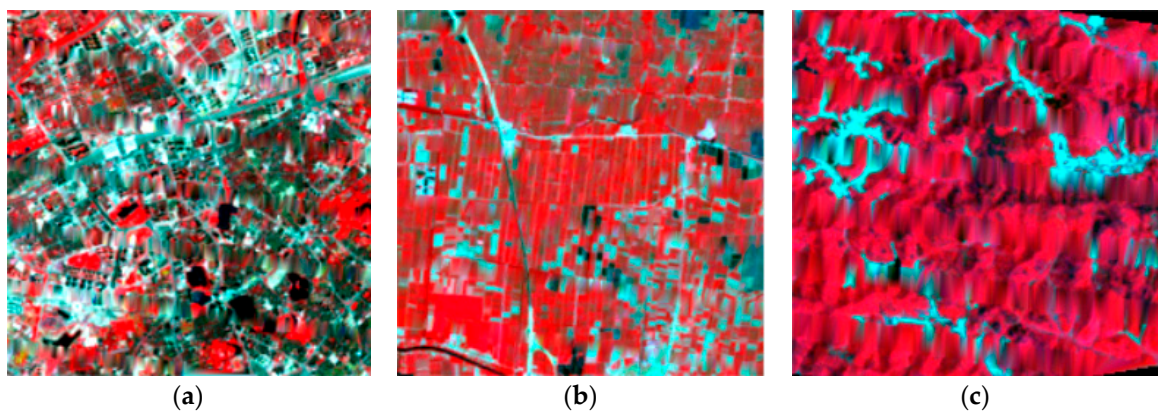
### 3. Experiments

#### 3.1. Experimental Datasets

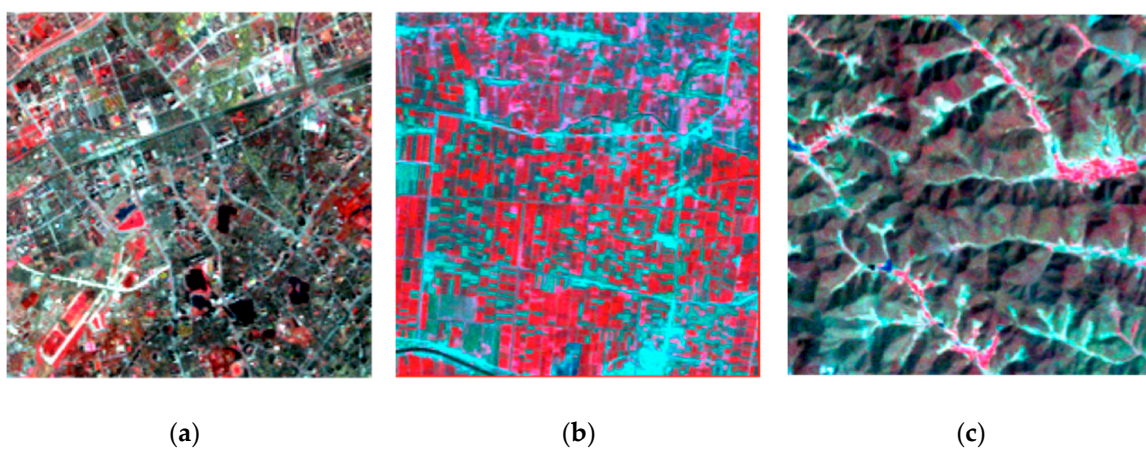
Three sites labeled built-up, agricultural, and hilly were selected for this study, as shown in Figure 2. In this research, two datasets (each with three multispectral remote-sensing image subsets dimensioned 200 by 200 pixels each) were used to validate the performance of the proposed methods. One dataset includes three Landsat ETM+ image subsets over three sites featuring different dominant land cover types (i.e., built-up, agricultural, and hilly) to showcase differences between information content in (subset) images of different broad cover types, as shown in Figure 3a–c, respectively, while the other consists of three Landsat TM image subsets over the same three sites, as shown in Figure 4a–c, respectively. Image subsets in Figures 3 and 4 are shown in standard false color, with bands 4, 3, and 2 displayed in red, green, and blue, respectively. These two datasets were used for comparing the (estimated) amounts of information content obtained by the two different sensors. The reason why data acquired from these two sensors were chosen is that they are both 8-bit, likely more comparable for validating the proposed methods, with ETM+ improved upon TM [41].



**Figure 2.** The three study sites: built-up, agricultural, and hilly.



**Figure 3.** Three Landsat ETM+ image subsets over sites with different dominant land cover types: (a) built-up, (b) agricultural, and, (c) hilly.



**Figure 4.** Three Landsat TM image subsets over the same sites as in Figure 3 with different dominant land cover types: (a) built-up, (b) agricultural, and, (c) hilly.

Specifically, the three Landsat ETM+ image subsets were selected from a Landsat ETM+ image (path and row number P123R39), flown over Wuhan City and surrounding areas, Hubei Province, China, on April 29, 2017. In these Landsat ETM+ image subsets (shown in Figure 3), bands 1–5 and 7 were actually used in the experiment, with band 6 excluded due to its different spatial resolution. The three Landsat TM image subsets were selected from a Landsat TM image (again, path and row number P123R39) acquired on March 14, 2009, as shown in Figure 4. Again, in these Landsat TM image subsets (shown in Figure 4a–c, respectively), bands 1–5 and 7 with the same 30 m spatial resolution were actually used.

### 3.2. Experiments with the Landsat ETM+ Dataset

The experiment procedures for the case study follow what is shown in Figure 1. With ICA transformation, for each image subset  $\mathbf{Z}$ , six image bands were used, resulting in six ICA-transformed image subset  $\mathbf{Z}'$  bands (i.e.,  $\mathbf{Z}'_1 \sim \mathbf{Z}'_6$ , note that these bands' order of appearance and content do not correspond to those of the original image subsets unless specially treated as in Falco et al. [30]). The weight matrices concerned in ICA were used for computing the differences between joint entropies of the original and ICA-transformed image subsets, as shown later.

For each of the six bands in each of the three ICA-transformed image subsets  $\mathbf{Z}'$ , normal transform was undertaken. This generated an ICA- and normal-transformed image subset  $\mathbf{Z}''$  (with bands denoted  $\mathbf{Z}''_1 \sim \mathbf{Z}''_6$ , corresponding to aforementioned ICA-transformed image bands  $\mathbf{Z}'_1 \sim \mathbf{Z}'_6$ ) for each ICA-transformed image subset  $\mathbf{Z}'$ . Table 1 shows the results of normal transform for three image subsets. Each row represents the specific transform parameters for a  $\mathbf{Z}'$  image band. For the  $S_u$  transform, there are four parameters to be evaluated as described in Equation (6).

**Table 1.** Normal transform parameters (see Equation (6)) and entropy difference values for the three image subsets.

Image Subsets	ICA- and Normal-Transformed Image $\mathbf{Z}''$ Bands	Transform Type	Parameters				$\Delta h$
			$\gamma$	$\eta$	$\epsilon$	$\lambda$	
Built-up	1	$S_u$	3.50	1.92	−0.73	0.51	0.13
	2	$S_u$	1.77	2.25	−1.97	1.53	0.09
	3	$S_u$	0.60	1.10	−1.32	0.54	0.17
	4	$S_u$	−0.20	1.69	−5.14	1.36	0.24
	5	$S_u$	0.06	1.40	−0.94	1.03	0.05
	6	$S_u$	−0.42	1.45	12.70	0.97	0.10
Agricultural	1	$S_u$	0.73	1.04	0.59	0.50	0.09
	2	$S_u$	−0.02	1.15	0.11	0.52	0.34
	3	$S_u$	−0.71	1.15	0.11	0.52	0.49
	4	$S_u$	−0.49	1.12	−0.40	0.56	0.19
	5	$S_u$	0.03	2.01	0.04	1.72	0.05
	6	$S_u$	0.24	1.93	0.23	1.60	0.06
Hilly	1	$S_u$	−0.42	0.50	−0.45	0.08	0.03
	2	$S_u$	−0.30	2.71	−0.15	0.55	1.53
	3	$S_u$	−0.56	2.77	0.54	2.45	0.05
	4	$S_u$	−0.16	1.22	−0.09	0.72	0.19
	5	$S_u$	−0.50	2.88	−0.42	2.40	0.14
	6	$S_u$	−0.06	0.91	−0.05	0.43	0.11

Values of  $\Delta h$  (Equation (7)) are also shown in Table 1 (the last column). As shown in Table 1,  $\Delta h$  values are generally small, except for band  $\mathbf{Z}''_2$  in the agricultural and hilly image subsets and band  $\mathbf{Z}''_3$  in the agricultural image subset, indicating largely modest effects of  $\mathbf{Z}'$  image data skewedness. Nevertheless, these differences should be accommodated for objectively quantifying information content in remote-sensing images.

When computing information content in a double-transformed image subset  $\mathbf{Z}''$  (with bands  $\mathbf{Z}_1'' \sim \mathbf{Z}_6''$ ), spatial dependence should be accommodated. In this experiment, spatial correlation was quantified through variogram modeling. After computing the experimental variograms for individual bands of double-transformed image subsets (three of them), theoretic models were fitted to experimental variograms. For this, nested structure models combining nugget effect models, spherical models, and exponential models were used because they provided the best fits among a standard set of variogram models in a weighted least squares sense [37]. The experimental variograms and the corresponding models fitted of the three land cover types are shown in Appendix A (Figures A1–A3) due to space limitation. Resultant variogram parameters are shown in Tables 2–4, indicating variable spatial correlation in terms of partial sills and ranges of spatial correlation. It should be noted that ranges are actually effective ranges for exponential models (i.e., three times  $a$  (not  $a$  itself) in models like  $\exp(-h/a)$ ). Nugget effects are quite small compared with the vertical axes of the variograms shown in Figures A1–A3. Estimated variogram model parameters were then used to calculate spatial covariances for  $\mathbf{Z}''$  image bands (i.e.,  $\Sigma_{\mathbf{Z}''}$  in Equation (8)), allowing for computing entropy  $h(\mathbf{Z}'')$ .

Entropies of double-transformed image data  $\mathbf{Z}_b''$  were estimated using  $(1/2) \log_2[\det(2\pi e \Sigma_{\mathbf{Z}_b'')}]$  (see Equation (8)), based on their variogram models. Joint entropies for  $\mathbf{Z}''$  image subsets were computed as sum of individual  $\mathbf{Z}''$  bands' entropies. Further, using results obtained previously regarding the difference between  $h(\mathbf{Z})$  and  $h(\mathbf{Z}')$  (i.e.,  $-\log_2|\det(\mathbf{W})|$  in Equation (5)) and  $\Delta h$  (shown in Table 1), joint entropies for original image subsets were estimated using Equations (5) and (8), as shown in Table 5.

To estimate mutual information  $I(\mathbf{S}, \mathbf{Z})$  for the original image data, we need to estimate original image noise's entropy  $h(\mathbf{N})$  (Equation (1)). For this, variogram modelling was also performed on original image bands  $\mathbf{Z}_1 \sim \mathbf{Z}_6$  for each image subset. Resultant estimates for nugget effects were then input to Equation (3) as estimates of noise variance for computing  $h(\mathbf{N})$ . Resultant estimates of noise variance are shown in Table 4, with estimates for  $h(\mathbf{N})$  shown in Table 5. Given estimates of  $h(\mathbf{Z})$  and  $h(\mathbf{N})$  shown in Table 5, mutual information  $I(\mathbf{S}, \mathbf{Z})$  was then estimated for the three image subsets (Equation (2)), as also shown in Table 5.

As shown in Table 5, information content for the “built-up” image subset is greater than that for “agricultural” and “hilly” image subsets, while there is slight difference between the latter two image subsets. This is consistent with our intuitive understanding: image scenes of built-up areas change more rapidly (having finer textures) than other types of scenes, thus containing a greater amount of information content.

For comparison, information content was also calculated by skipping one or more of the processing steps (i.e., image transformations and geostatistical modeling) involved in arriving at the estimation of  $I(\mathbf{S}, \mathbf{Z})$  reported in Table 5. We describe relevant results below.

First, we may come up with estimation of information content without considering spectral dependence, spatial dependence, and non-normality (thus becoming a naïve method). This means that we estimate information content simply by computing mutual information of original image subsets' bands separately (using Gaussian distribution's differential entropy formula and assuming spatial independence) and summing up. Results are listed in Table 6 (Method 1, labeled for convenience), where  $I(\mathbf{S}, \mathbf{Z})$  estimates shown in Table 5 (i.e., labeled as Method 5) are duplicated for comparison.

Second, information content was estimated by considering spatial dependence only. For this, variogram parameters estimated based on the original image (Table 4) were used to construct the spatial covariance matrices (assuming existence of sill for a variogram model) for computing joint entropies of the original image subsets and hence mutual information; the rest were the same as in Method 1. Results are shown in Table 6 (Method 2).

**Table 2.** Variogram model parameters (nug: nugget, exp: exponential, sph: spherical; units for ranges are number of pixels) for ICA- and normal-transformed image bands ( $Z_1'' \sim Z_6''$ ).

Bands	Built-Up					Agricultural					Hilly				
	Nug		Exp		Sph	Nug		Exp		Sph	Nug		Exp		Sph
		Sill		Range	Sill		Sill		Range	Sill		Sill		Range	Sill
1	0.02	0.40	35.61	0.51	4.78	0.01	0.83	7.70	0.13	36.18	0.08	0.85	14.44	0.25	68.14
2	0.08	0.76	5.41	0.20	17.31	0.01	0.80	31.21	0.45	4.13	0.40	0.60	38.46	0.51	6.61
3	0.02	0.83	12.05	0.15	63.15	0.22	0.66	11.02	0.14	52.08	0.04	0.55	22.94	0.38	6.35
4	0.01	0.96	4.05	0.11	20.56	0.15	0.76	6.14	0.11	28.52	0.01	0.69	5.62	0.27	14.37
5	0.06	0.88	4.91	0.17	191.30	0.11	0.30	41.18	0.50	5.84	0.19	0.45	6.24	0.34	14.67
6	0.10	0.42	23.75	0.49	4.72	0.25	0.55	4.24	0.18	14.40	0.27	0.44	4.87	0.30	12.52

**Table 3.** Variogram model parameters (nug: nugget, exp: exponential, sph: spherical; units for ranges are number of pixels) for ICA-transformed image bands ( $Z_1' \sim Z_6'$ ).

Bands	Built-Up					Agricultural					Hilly				
	Nug		Exp		Sph	Nug		Exp		Sph	Nug		Exp		Sph
		Sill		Range	Sill		Sill		Range	Sill		Sill		Range	Sill
1	0.04	0.69	21.62	0.23	4.74	0.02	0.68	24.80	0.24	4.34	0.01	0.46	83.57	0.63	9.24
2	0.01	0.38	31.86	0.56	4.60	0.03	0.35	40.94	0.53	5.92	0.04	0.52	98.35	0.23	6.42
3	0.01	0.90	5.12	0.08	22.71	0.01	0.58	7.69	0.30	17.93	0.07	0.49	25.26	0.40	6.16
4	0.14	0.25	20.66	0.59	3.83	0.06	0.29	26.17	0.59	4.95	0.02	0.44	2.55	0.53	9.29
5	0.13	0.82	7.06	0.07	0.28	0.18	0.29	19.80	0.47	4.34	0.11	0.60	4.75	0.20	14.46
6	0.04	0.82	5.18	0.12	18.85	0.27	0.50	5.02	0.18	16.22	0.25	0.36	4.34	0.32	14.65

**Table 4.** Variogram model parameters (nug: nugget, exp: exponential, sph: spherical; units for ranges are number of pixels) for original image bands ( $Z_1 \sim Z_6$ ).

Bands	Built-Up					Agricultural					Hilly				
	Nug		Exp		Sph	Nug		Exp		Sph	Nug		Exp		Sph
		Sill		Range	Sill		Sill		Range	Sill		Sill		Range	Sill
1	0.16	90.36	33.46	87.68	4.70	0.37	3.65	23.80	28.42	6.32	0.31	44.64	72.13	23.15	6.60
2	0.22	106.70	35.34	113.48	4.81	0.26	35.54	29.33	46.23	6.00	0.13	45.77	15.01	30.43	114.56

Table 4. Cont.

Bands	Built-Up					Agricultural					Hilly				
	Nug	Exp		Sph		Nug	Exp		Sph		Nug	Exp		Sph	
		Sill	Range	Sill	Range		Sill	Range	Sill	Range		Sill	Range	Sill	Range
3	0.21	229.84	32.43	222.89	4.94	0.17	93.51	29.57	130.54	6.00	0.31	145.76	15.72	111.37	170.07
4	0.36	92.32	17.45	50.66	4.37	0.45	98.77	37.34	55.39	6.29	0.23	143.89	63.52	158.89	8.54
5	0.21	463.29	7.85	60.80	36.81	0.10	207.50	18.55	105.67	5.01	0.11	136.79	33.07	135.10	7.50
6	0.16	574.04	7.65	12.34	36.02	0.44	218.71	19.61	151.30	5.57	0.30	142.89	87.95	204.42	8.94

Table 5. Estimates of information content in original image subsets with spectral dependence, spatial dependence, and non-normality accommodated (bits/pixel).

Image Subsets	Built-Up	Agricultural	Hilly
Joint entropy $h(\mathbf{Z}'')$	9.88	9.73	9.84
$\Delta h$	0.78	1.22	1.94
Joint entropy $h(\mathbf{Z}')$	9.10	8.51	7.90
Difference between $h(\mathbf{Z})$ and $h(\mathbf{Z}')$	14.68	13.79	14.11
Joint entropy $h(\mathbf{Z})$	23.78	22.30	22.01
Noise joint entropy $h(\mathbf{N})$	5.56	6.51	5.57
Mutual information	18.22	15.79	16.44

Table 6. Estimates of information content for the three image subsets using different methods (bits/pixel).

Methods	Experimental Image Subsets		
	Built-Up	Agricultural	Hilly
1. Not considering spectral and spatial dependences and non-normality	31.68	28.17	29.35
2. Considering spatial dependence only	27.62	23.08	22.69
3. Considering spectral dependence (via ICA) only	21.42	19.58	20.84
4. Considering spectral (via ICA) and spatial dependences	18.77	16.16	17.04
5. Considering spectral (via ICA) and spatial dependences and non-normality	18.22	15.79	16.44

Third, information content was estimated by considering spectral dependence only. For this, sill estimates as part of variogram parameters estimated for the ICA-transformed image subsets (Table 3) were used for computing joint entropies of the ICA-transformed image subsets (hence those of the original image subsets, using Equation (5), assuming spatial independence), with the rest of the steps being the same as in Method 1. Results are shown in Table 6 (Method 3).

Fourth, information content was estimated by considering both spectral and spatial dependence. This means that the processing steps were the same as those for Method 5, except for normal transformation (and hence calculation for  $\Delta h_b$  in Equation (7)). For this, variogram parameters estimated based on the ICA-transformed image subsets (Table 3) were used to construct the spatial covariance matrices for computing joint entropies of the ICA-transformed image subset (hence those of the original image subsets). Results are shown in Table 6 (Method 4).

As shown in Table 6, information redundancy due to spectral dependence is greater than that due to spatial dependence, as results obtained by Method 3 are smaller than those by Method 2. In other words, results obtained by Method 3 indicate substantial information redundancy due to inter-band spectral dependence. Comparing results obtained by Method 4 and Method 5, over-estimation of information content by assuming normality (i.e., no treatment for non-normality) is about 0.5 bits/pixel. Method 5 (representing a full treatment for spectral dependence, spatial dependence, and non-normality) provides the most objective (and accurate) estimates of information content. The differences between the results obtained by Method 5 and those by Method 1 are outstanding, being about 12.9 bits/pixel.

### 3.3. Experiments with the Landsat TM Dataset and Comparisons Between Landsat ETM+ and TM Datasets

The experimental procedures for the TM dataset are the same as for the ETM+ dataset (see Figure 1). Table 7 shows the results of information content estimation for ICA-based methods for the three TM image subsets, with intermediate results not shown (nor in Appendix A). There are distinct differences between the results obtained by Method 1 and Method 5, being about 8.9 bits/pixel.

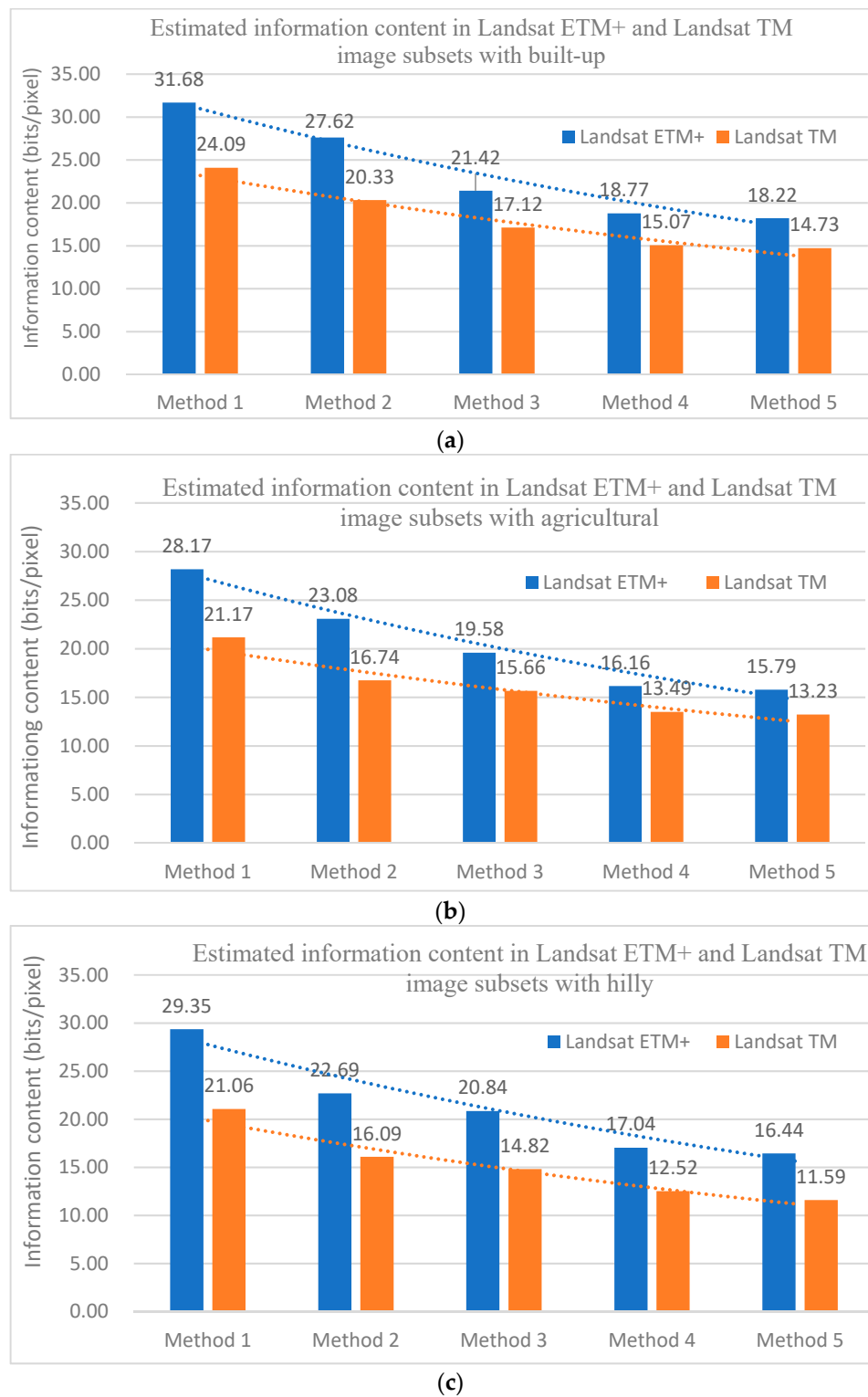
**Table 7.** Estimates of information content for the three TM image subsets using different methods labeled as in Table 6 (bits/pixel).

Methods	Experimental Image Subsets		
	Built-Up	Agricultural	Hilly
1. Not considering spectral dependence, spatial dependences, and non-normality	24.09	21.17	21.06
2. Considering spatial dependence only	20.33	16.74	16.09
3. Considering spectral dependence (via ICA) only	17.12	15.66	14.82
4. Considering spectral (via ICA) and spatial dependences	15.07	13.49	12.52
5. Considering spectral (via ICA) and spatial dependences and non-normality	14.73	13.23	11.59

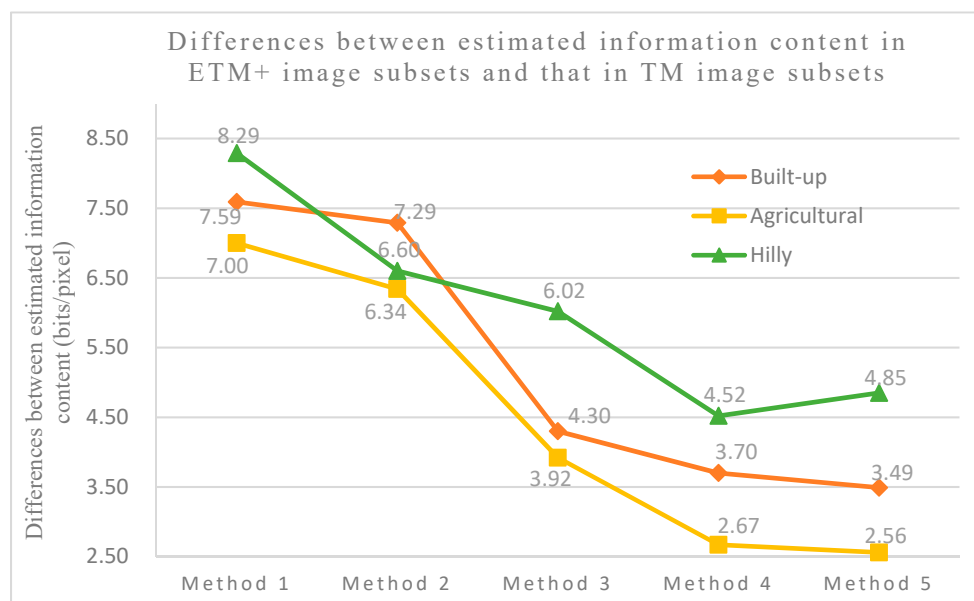
The general patterns of information content estimates with respect to different land cover types are similar to those in the Landsat ETM+ image subsets. Among the different methods for estimating information content, Method 5 (full treatment) provides the most accurate estimates as it accounts for information redundancy due to spectral and spatial dependences and over-estimation due to non-normality. Comparing information content estimates for image subsets acquired with the two sensors, as shown in Tables 6 and 7, information content estimates for Landsat ETM+ image subsets are much greater than those for Landsat TM image subsets, especially for the “hilly” image subsets.

In order to clearly visualize the differences in information content between Landsat ETM+ and TM image subsets, three bar charts are presented in Figure 5. As shown in Figure 5, for all three different

land cover types, information content estimates for Landsat ETM+ image subsets are consistently greater than those for Landsat TM image subsets. The differences in information content estimates between the two sensors are calculated and shown in Figure 6 for convenience.



**Figure 5.** Estimated information content in Landsat ETM+ and Landsat TM image subsets over sites with different dominant land cover types: (a) built-up, (b) agricultural, and, (c) hilly.



**Figure 6.** Differences in information content between Landsat ETM+ and Landsat TM image subsets, estimated by different methods.

As shown in Figure 6, the differences in information content estimates between the two sensors are the biggest when considering none of spectral dependence, spatial dependences, and non-normality (Method 1). The differences in estimates obtained by Method 2 (considering spectral dependence only) are the second biggest. Comparing Method 2 with Method 3 (considering spatial dependence only), we find that information redundancy due to spectral dependence is greater than that due to spatial dependence. After considering all aspects (spectral dependence, spatial dependence, and non-normality) as by Method 5, the differences in estimated information content in images obtained by the two sensors are still quite outstanding, being about 3.6 bits/pixel on average. In terms of information content, the results indicate that the quality of ETM+ images is better than that of TM images. The fact that ETM+ image subsets contain more information content than TM ones was confirmed by noting greater values of image DN's variance, a conventional statistical measure of image variability, in the former than in the latter, although not reported here. We are not able to provide further explanations, since there does not seem to be published work on formal comparisons of Landsat ETM+ and TM sensor performances and/or image quality, let alone that based on information theory.

#### 4. Discussion

This paper has proposed a strategy based on image transformations and geostatistical modeling for quantifying information content in multispectral remote-sensing images. The proposed strategy is expectedly valuable, with adaptations, for analyzing subsequent information processes in remote sensing. Below, we compare ICA with MNF for handling spectral redundancy, address the issues of stationarity and geostatistical modeling, and give an outlook about applicabilities of the proposed methods, followed by some prospectives about further research.

##### 4.1. Comparing ICA and MNF for Handling Spectral Redundancy

Although spectral redundancy, spatial redundancy, and non-normality are important factors when quantifying information content in multispectral images, we focus on spectral redundancy here. This is based on the understanding that the experiments carried out in this research were actually designed both for testing the full-treatment method (e.g., Method 5 in Tables 6 and 7) and for comparing it with alternatives in which inter-band spectral dependence, intra-band spatial dependence, and/or non-normality is not accounted for (e.g., Methods 1 through 4 in Tables 6 and 7). The other reason is

that comprehensive comparisons between the proposed methods and existing methods (e.g., those reviewed in Section 1) are beyond the scope of this paper.

For handling spectral redundancy, de-correlation methods, such as PCA and MNF, may be used, as reviewed in Section 1. PCA leads to linear orthogonalization of correlated multivariate data whereby spatial structure is not taken into account. MNF transform [24,25], though similar to PCA in the sense of orthogonalization, is better suited for spatial data than PCA, because MNF transform works in a metric space defined by a noise covariance matrix estimated from the data, with noise estimated as the difference between central and neighboring pixels. Thus, as a follow-up to the discussions above and those in Section 1 concerning differences between spectral decorrelation and spectral independence, we performed empirical comparisons between ICA and MNF with respect to spectral redundancy handling, with results described below.

When estimating information content by considering spectral dependence (via MNF) only, estimation was obtained for MNF-transformed image subsets (both ETM+ and TM) following the procedures similar to Method 3 in Tables 6 and 7 except for using MNF rather than ICA for decorrelation-oriented image transformation. Results are shown in Table 8 (Method 6). When quantifying information content by considering both spectral dependence (via MNF) and spatial dependence, experimental variograms were computed with corresponding models fitted. As an example, results of model fitting for the three ETM+ image subsets (built-up, agricultural, and hilly) are shown in Appendix A, Figure A1c, Figure A2c, and Figure A3c, due to space limitation. After variogram modeling, information content was estimated following procedures similar to those of Method 4 in Tables 6 and 7, leading to estimated information content by considering both spectral dependence (via MNF) and spatial dependence. Results are also shown in Table 8 (Method 7).

**Table 8.** Estimates of information content for both ETM+ and TM image subsets using MNF-based methods (bits/pixel).

Methods	Data	Experimental Image Subsets		
		Built-Up	Agricultural	Hilly
6. Considering spectral dependence (via MNF) only	ETM+	23.21	21.33	22.30
	TM	19.90	16.50	16.92
7. Considering spectral (via MNF) and spatial dependences	ETM+	20.39	17.72	18.30
	TM	17.96	14.27	14.23

As shown in Table 8 (Methods 6 and 7), for all these different land cover types, estimates of information content obtained via MNF transform are bigger than those via ICA transform shown in Tables 6 and 7 (Method 3 and 4). For instance, comparing results obtained from Method 6 (Table 8) and Method 3 (Table 6) with ETM+ dataset reveals that over-estimation of information content by MNF as opposed to ICA is 1.79, 1.75, and 1.46 (bits/pixel) for built-up, agricultural, and hilly cover types, respectively. In other words, more spectral redundancy appears to remain in MNF-based estimates of information content than in ICA-based ones, whether considering spectral redundancy only (Method 6) or both spectral and spatial dependence (Method 7). This is because MNF transform can only be used to generate uncorrelated variables, which are not necessarily independent, as mentioned in Section 1. It (MNF) is thus less effective in handling spectral redundancy (and hence less accurate in quantifying information content) in multispectral images than ICA, as shown by empirical evidence above.

#### 4.2. The Issues of Stationarity Assumptions and Geostatistical Modeling

For quantifying information content in multispectral images, we assumed strong stationarity (for each band as well as for the transformed bands) and the existence of second-order moments for first-order increments. The stationarity assumption was paralleled in this research by use of experimental image subsets, which were deliberately selected from sites of relative homogeneity

(i.e., specific land-cover types) so that local stationarity may be assumed therein. This experimental design was taken as a measure of ensuring the appropriateness of applying the proposed methods for estimating information content in this research.

Although discerning whether strong stationarity is a reasonable assumption is not possible with just a dataset or two, we should analyze the image data being assessed to discern whether intrinsic or second-order stationarity are reasonable assumptions, especially when the image data are acquired over large areas with outstanding spatial heterogeneity. For this, useful work has been done, which utilizes hypothesis assumption [42]. With all the experimental image subsets,  $p$ -values more than 0.05 were registered, indicating that the stationarity assumption should be accepted. We should also try to incorporate well-established methods and techniques for stationarity tests and for accommodating non-stationarity in spatial dependence when working with images over large areas.

Anisotropy is an important aspect in geostatistical modeling. Although not reported in the paper, we checked for possible anisotropy in variogram models fitted with the image data used in this research. The directional variograms (with directions 0, 45, 90, 135, 180, 225, 270, and 315 degrees counter-clockwise from positive direction of x-axis tested in experiments) appeared very similar. Thus, we adopted omnidirectional variograms for the experiments.

Data support is yet another issue. The image data were assumed to be point values in this research. However, image data are actually on finite areal support, as all images are subject to sensor's point spread function (PSF) convolution [16,38,39]. Thus, we should describe pixel-support data using variogram models that accommodate their finite support. In turn, it is important to investigate methods for quantifying information content while dealing with sensors' PSFs [16,38,39].

Geostatistical modeling in this research was limited to two-point variograms. However, one of the shortcomings of variogram-based models lies in their lack of modeling complex spatial patterns, which are often present in remote-sensing images, especially those over large areas and/or heterogeneous landscapes. For modelling complex spatial (and spatio-temporal) patterns, multiple-point statistics (MPS) [43,44] is advantageous over conventional two-point variogram-based geostatistics. Thus, MPS is anticipated to contribute to improved estimation of information content in images over complex landscapes where spatial correlation is not adequately described by two-point variograms. Further research on extending the proposed methods with MPS is worthwhile.

#### *4.3. Applicabilities of the Proposed Methods and Some Topics for Further Research*

As information-theoretic description and analyses of images (source or derivative) will play a greater role in remote sensing information engineering and consumption, accurately measuring information content will become more and more important. The value of the work pursued here will be well manifested in the contemporary information era.

In terms of applications, the proposed methods for information quantifying can be used, for example, to select optimal features (for image classification) based on informational criteria (e.g., maximum mutual information and minimal information redundancy) [7,45]. Classifier performance evaluation can also be undertaken based on information-theoretic analyses from source images through feature selection to resultant classifications (e.g., Marinoni et al. [11]). This is particularly useful when classifications are viewed as lossy compressions of images. As close relationships among ICA and maximum autocorrelation factors (MAF) transform is well established [46], research efforts on informational analyses of image and non-image datasets can be joined to improve our understanding about information dynamics with respect to data fusion.

To promote future research, we highlight a few more aspects whereby the methods pursued here may be improved. First, studies on quantifying information content based on combined use of the proposed method and other methods would be usefully explored, although some comparisons were done in this research, as shown in Section 4.1. In other words, synthesis of the methods proposed in this paper and by other researchers likely sheds light on where further improvements may be made. Second, uncertainty in the estimated amounts of information content should be evaluated. This is

hardly an easy task, as there appears to be little literature on this matter. Third, the proposed methods can be usefully extended for use in hyperspectral and time-series image analyses and applications. Last, adaptations to the proposed methods are necessarily made to enhance efficiency in computing with image datasets and to accommodate non-stationarity in spatial dependence, especially for large-area applications, as mentioned in Section 4.2 above.

## 5. Conclusions

For quantifying information content in multispectral remote-sensing images, this paper has proposed an integrated strategy based on image transformations and geostatistical modelling to account for inter-band spectral dependence, intra-band spatial dependence, and non-normality systematically. Specifically, ICA is applied to transform the original multispectral image subsets into ones consisting of independent *bands* to allow for estimating joint entropy as the sum of individual transformed image *bands* entropy. These image *bands* are further normal-transformed to facilitate analytical computing of joint entropies. Spatial correlation in the aforementioned ICA- and normal-transformed image *bands* is quantified via variogram parameters and can be easily accommodated when computing normal distributions' entropy values. (Additive) noise can also be taken care of as noise variance is estimated as nugget effects (in variograms). Theoretically speaking, information redundancy due to spectral and spatial dependences is well handled through ICA transforms and geostatistical modeling, respectively, while information overestimation due to non-normality is taken care of by normal-transform.

Results based on Landsat ETM+ and TM image datasets consistently confirmed the advantages of the proposed methods for full treatment (Method 5, Tables 6 and 7) and for handling spectral dependence (Method 3, Tables 6 and 7 as opposed to Method 6, Table 8). Specifically, differences between information content estimated by full treatment (Method 5) and that by naive treatment (Method 1) range from 7.9 to 13.5 (bits/pixel) (referring to Tables 6 and 7). Overestimation of information content by MNF as opposed to ICA exceeds 1.4 (bits/pixel) (comparing Table 6 with Table 8). It was also shown that information redundancy due to spectral dependence is greater than that due to spatial dependence, with differences between spectral redundancy and spatial redundancy ranging from about 1.1 to 6.2 (bits/pixel). Differences in information content between image subsets obtained with ETM+ and TM were found to be about 3.6 bits/pixel, indicating the former's greater information content.

The proposed method set (Method 5, Tables 6 and 7) for estimating information content in multispectral images is highly recommendable for information-theoretic analyses of remote sensing and geospatial information processes. This assertion is made on the grounds that it (the proposed strategy) provides the most accurate measures of information content (of all methods experimented in this research) and that it is easily implemented with all relevant computational procedures established or extensible.

**Author Contributions:** Y.Z., the principal author, contributed mostly to the literature survey in the fields of information theory and remote sensing. Y.Z. was also responsible for research-related experimentation, results analysis and interpretation, and paper writing. The primary contribution of J.Z., the corresponding author, includes literature review, research design, and paper revision. W.Y. helped with experimentation. All authors have read and agreed to the published version of the manuscript.

**Funding:** The research reported in this paper is supported by the National Natural Science Foundation of China (Grant No. 41471375).

**Acknowledgments:** Comments and suggestions from anonymous reviewers and those concerning geostatistics from an anonymous reviewer for a previous manuscript were greatly appreciated. Michael Goodchild (Emeritus Professor at the University of California, Santa Barbara) and Zongjian LIN (former President of the Chinese Academy of Surveying and Mapping) provided valuable guidance about information theory for Jingxiong Zhang. We pray for the healing of novel coronavirus pneumonia (COVID-19) patients in Wuhan, in China, and elsewhere around the world and the stopping of the plague by the LORD.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A

Empirical variograms and corresponding models fitted for the Landsat ETM+ image subsets

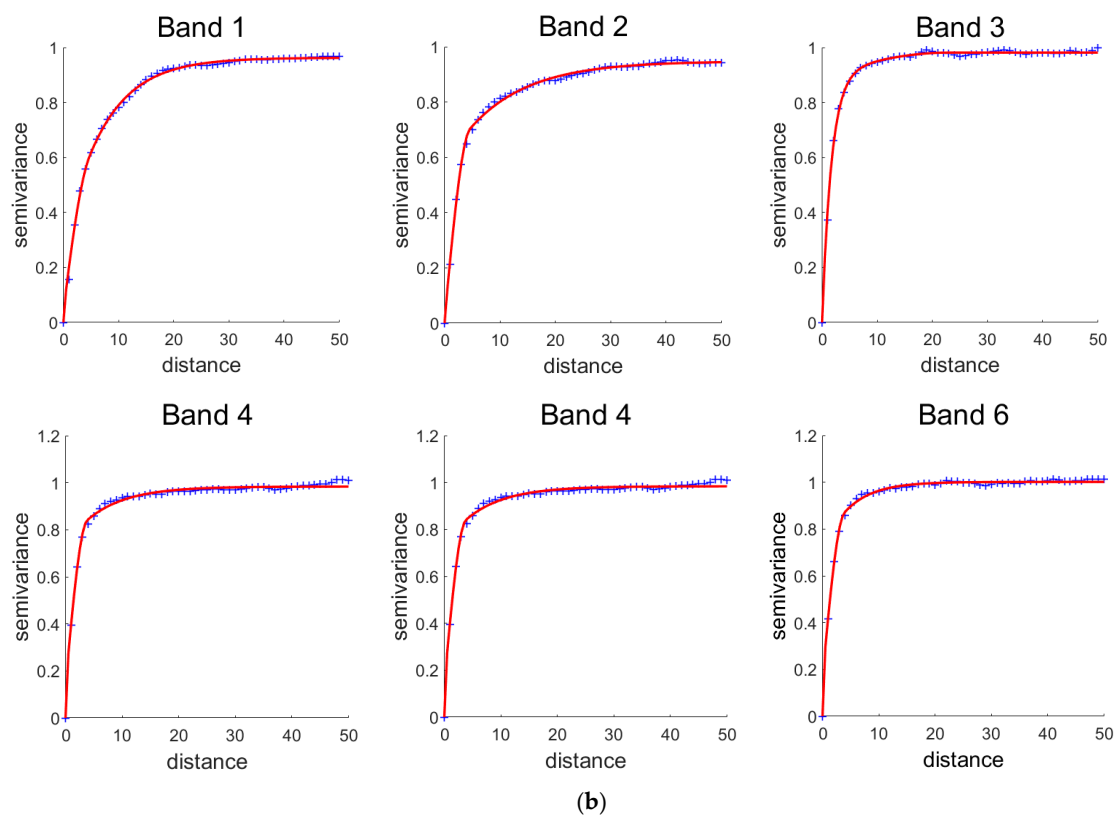
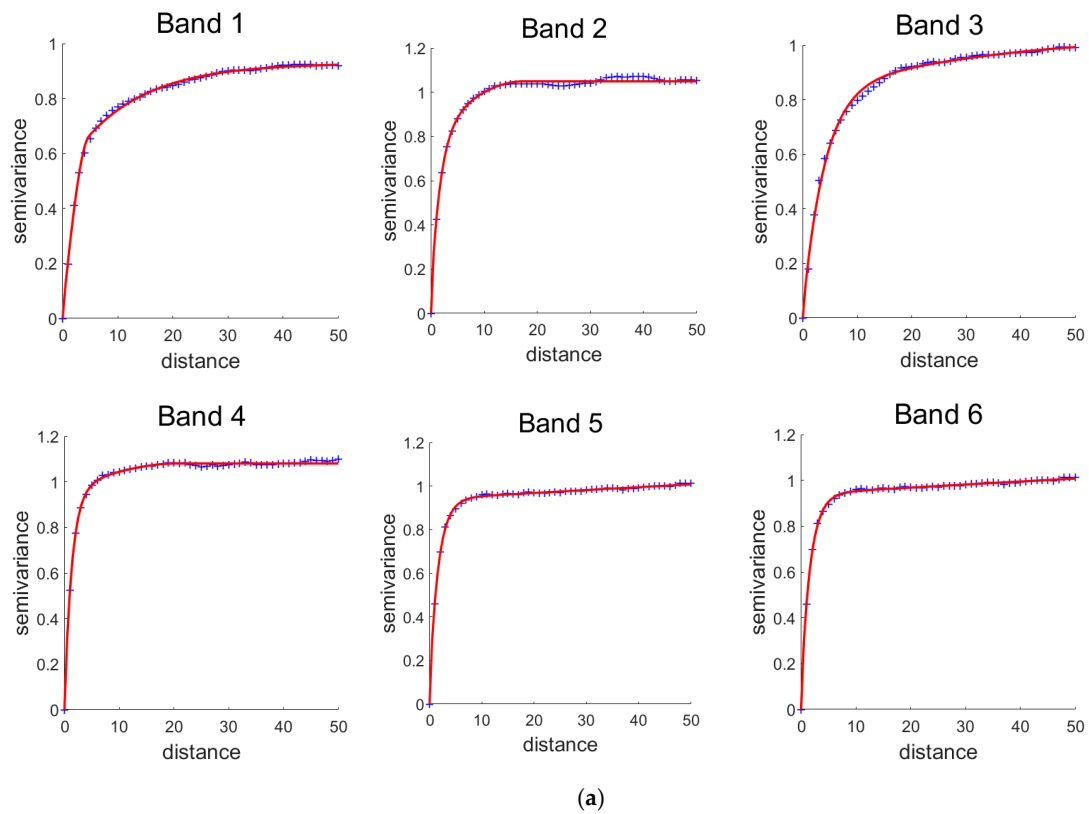
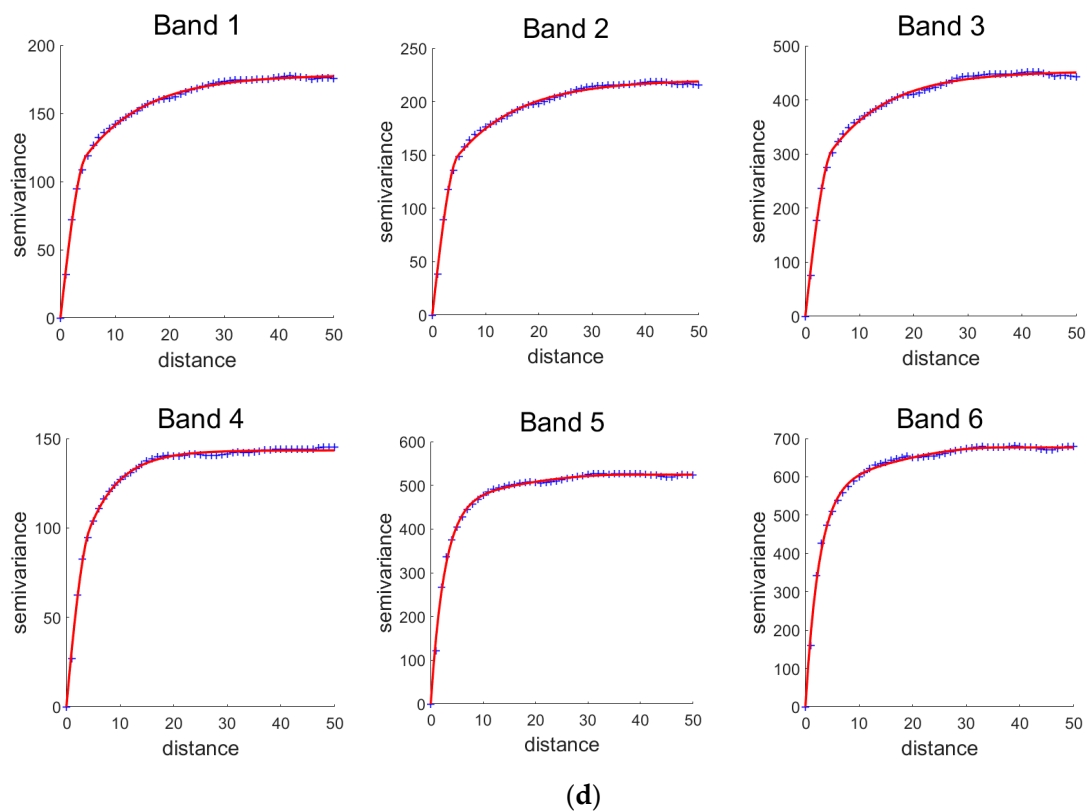
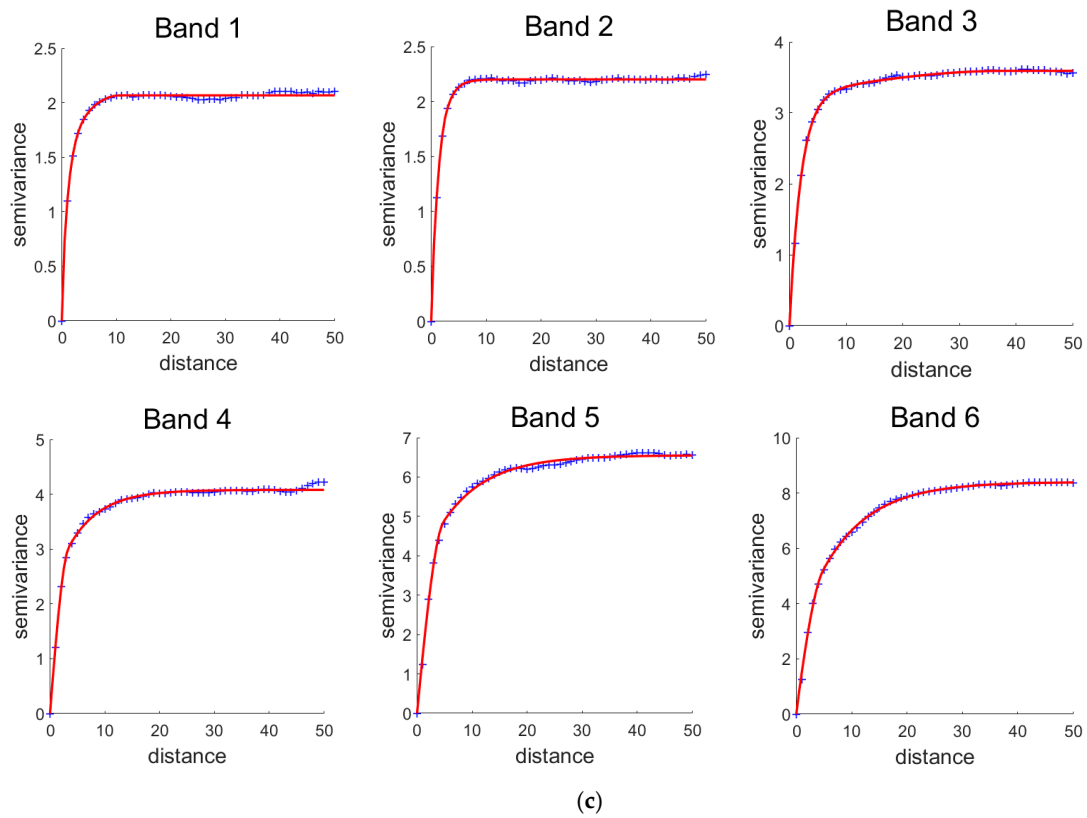


Figure A1. Cont.



**Figure A1.** The experimental variograms (blue cross) and the corresponding models fitted (red lines) for the built-up image subset: (a) independent component analysis (ICA)- and normal-transformed image bands ( $Z''_1 \sim Z''_6$ ), (b) ICA-transformed image bands ( $Z'_1 \sim Z'_6$ ), (c) maximum noise fraction (MNF)-transformed image bands ( $Z'_1 \sim Z'_6$ ), and (d) original image bands ( $Z_1 \sim Z_6$ ).

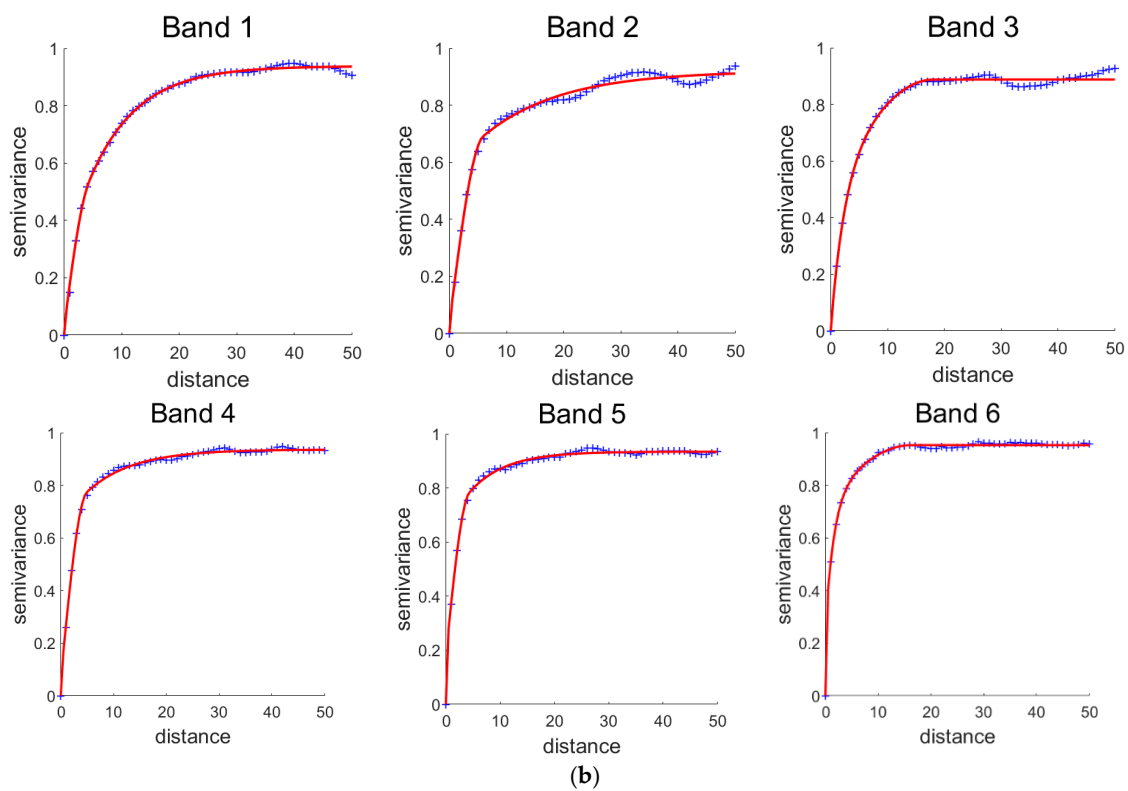
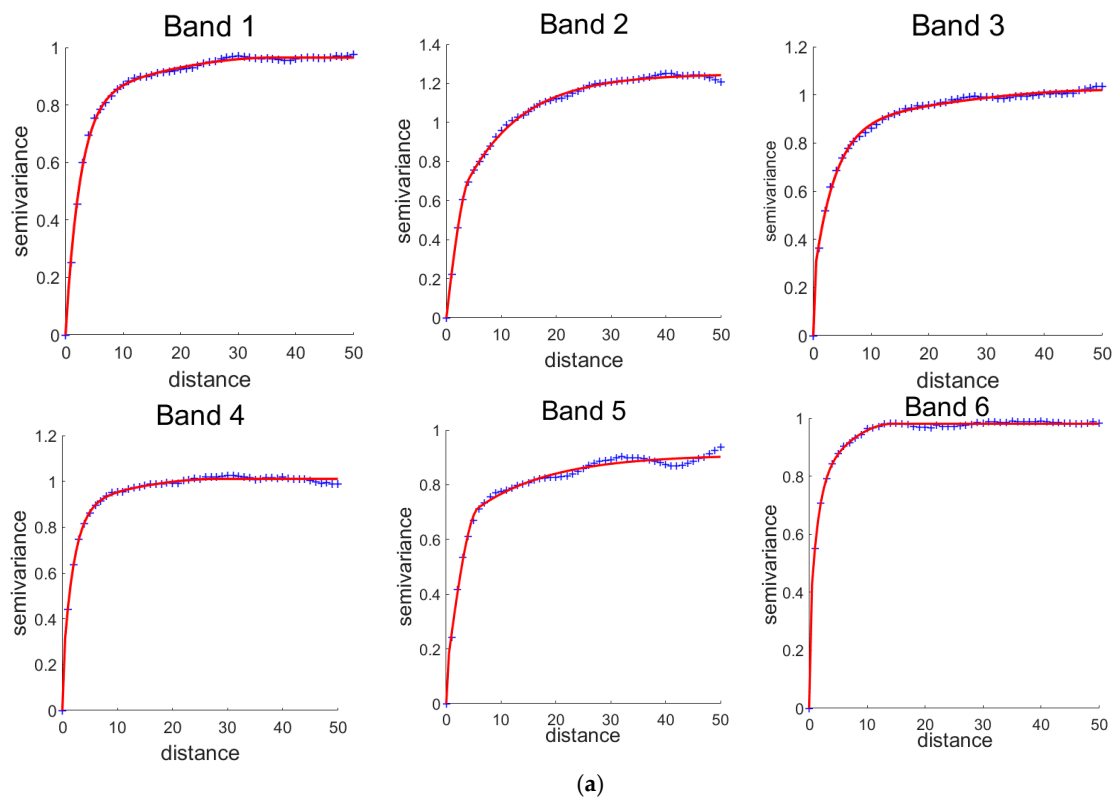
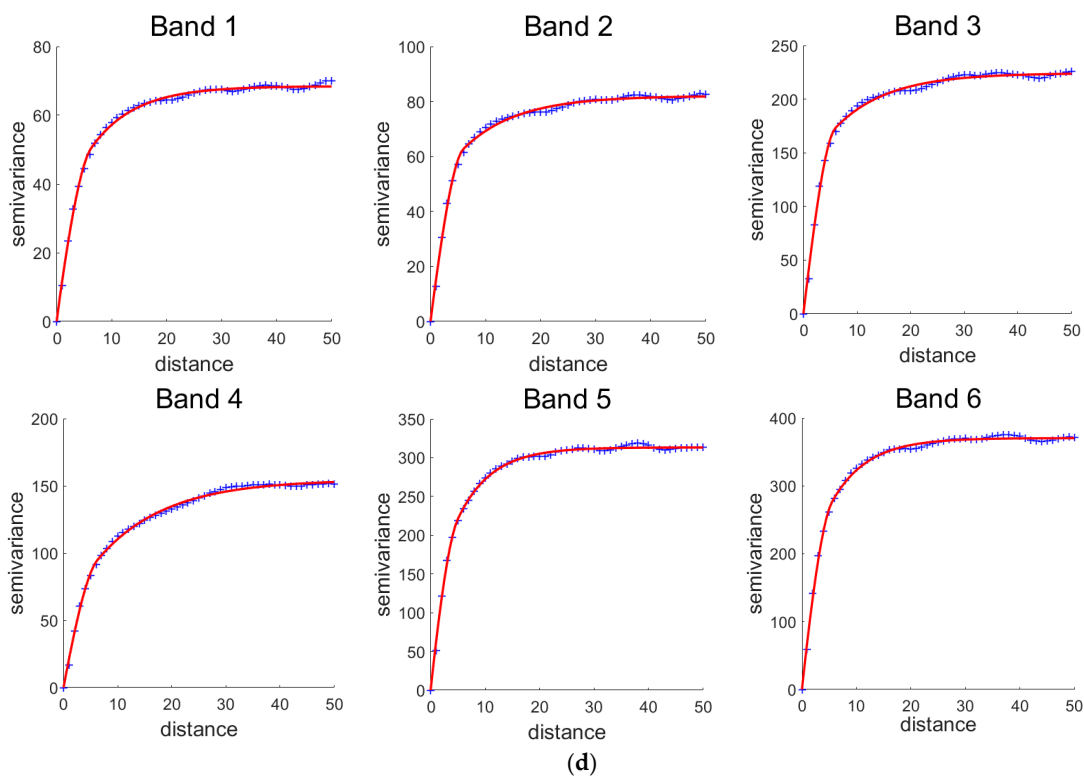
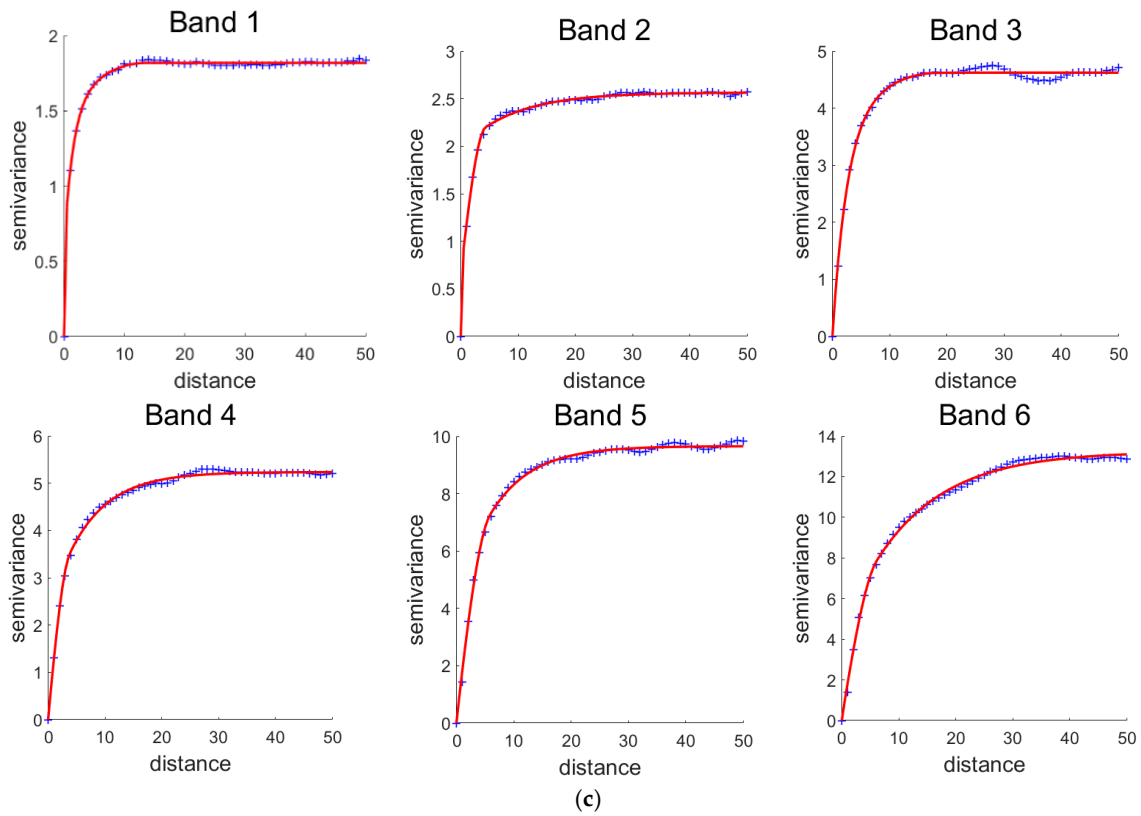


Figure A2. Cont.



**Figure A2.** The experimental variograms (blue cross) and the corresponding models fitted (red lines) for the agricultural image subset: (a) ICA- and normal-transformed image bands ( $Z'_1 \sim Z'_6$ ), (b) ICA-transformed image bands ( $Z''_1 \sim Z''_6$ ). (c) MNF-transformed image bands ( $Z'_1 \sim Z'_6$ ), and (d) original image bands ( $Z_1 \sim Z_6$ ).

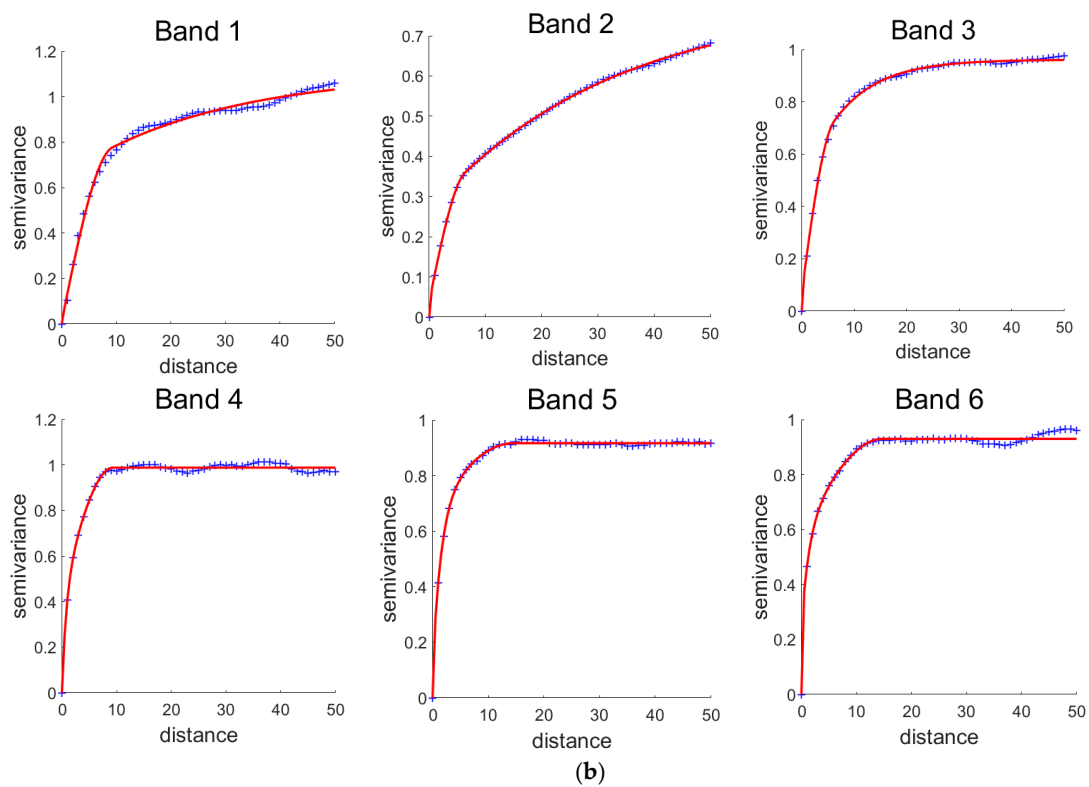
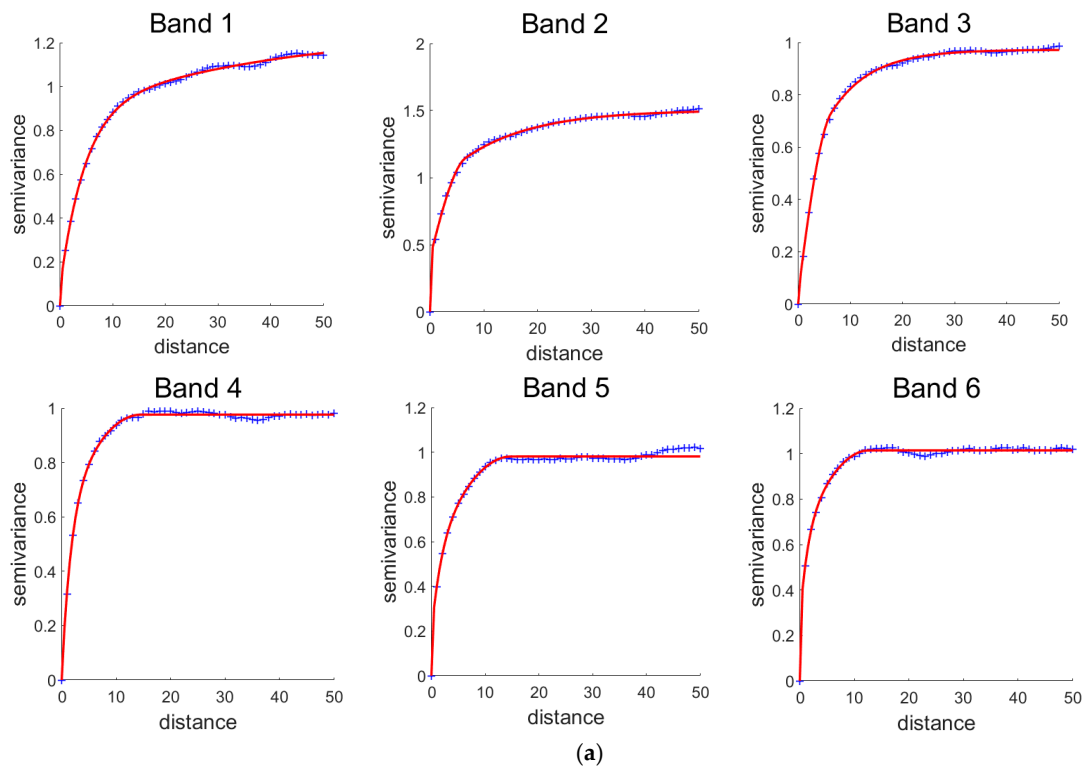
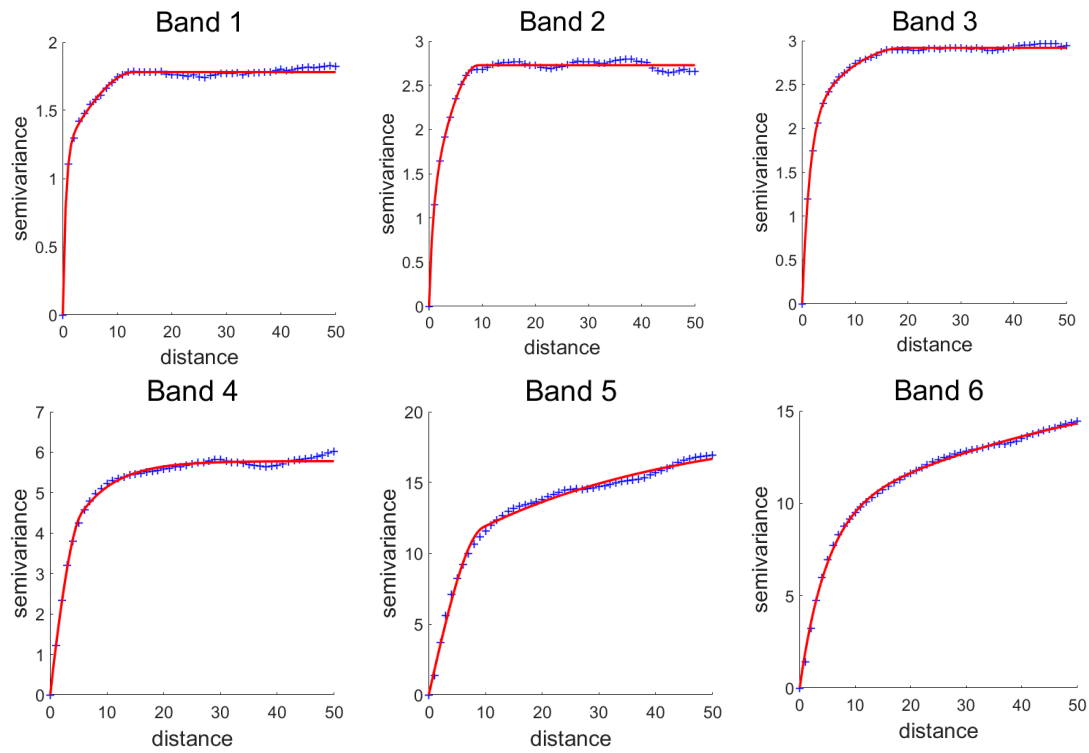
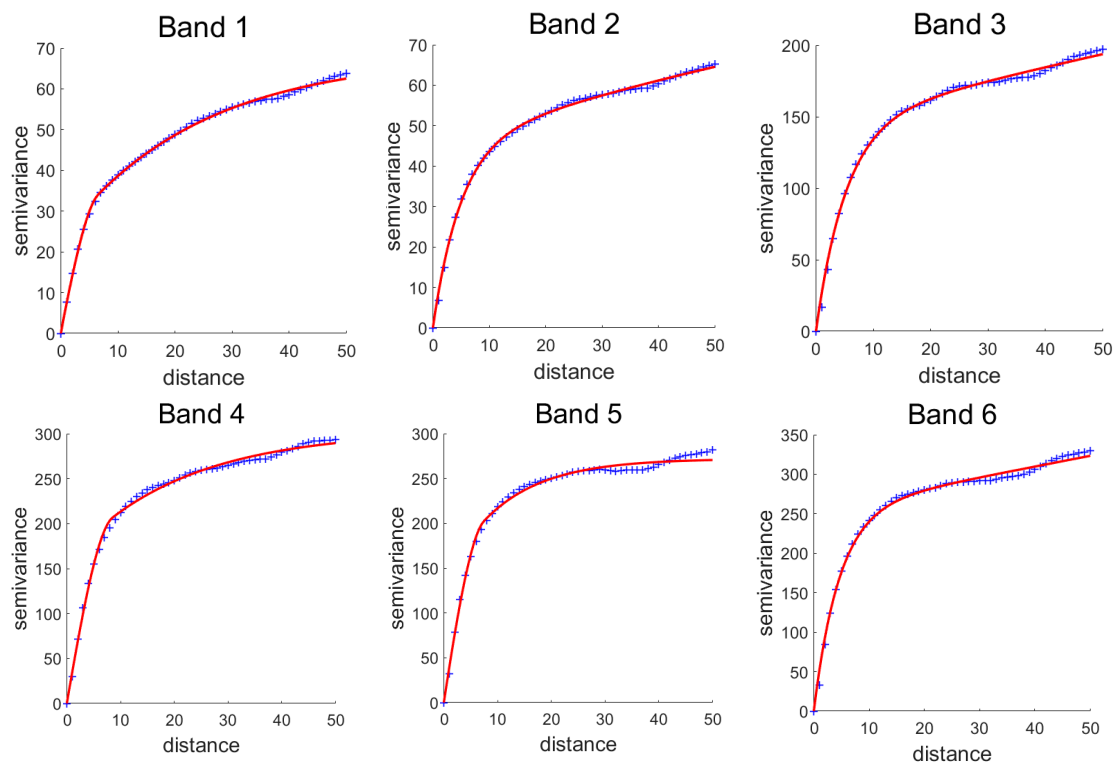


Figure A3. Cont.



(c)



(d)

**Figure A3.** The experimental variograms (blue cross) and the corresponding models fitted (red lines) for the hilly image subset: (a) ICA- and normal-transformed image bands ( $Z'_1 \sim Z'_6$ ), (b) ICA-transformed image bands ( $Z'_1 \sim Z'_6$ ), (c) MNF-transformed image bands ( $Z'_1 \sim Z'_6$ ), and (d) original image bands ( $Z_1 \sim Z_6$ ).

## References

1. Zhang, J.; Atkinson, P.M.; Goodchild, M.F. *Scale in Spatial Information and Analysis*; CRC Press: Boca Raton, FL, USA, 2014.
2. Shannon, C.E. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423. [\[CrossRef\]](#)
3. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*, 2nd ed.; John Wiley & Sons Inc.: Hoboken, NJ, USA, 2006.
4. Goodchild, M.F. The Nature and Value of Geographic Information. In *Foundations of Geographic Information Science*; Duckham, M., Goodchild, M.F., Worboys, M.F., Eds.; Taylor & Francis: London, UK, 2003; pp. 19–31.
5. Huck, F.O.; Fales, C.L.; Rahman, Z.-U. An information theory of visual communication. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **1996**, *354*, 2193–2248. [\[CrossRef\]](#)
6. O'Sullivan, J.A.; Blahut, R.E.; Snyder, D.L. Information-theoretic image formation. *IEEE Trans. Inf. Theory.* **1998**, *44*, 2094–2123. [\[CrossRef\]](#)
7. Peng, H.; Long, F.; Ding, C. Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, *27*, 1226–1238. [\[CrossRef\]](#) [\[PubMed\]](#)
8. Meynet, J.; Thiran, J.P. Information theoretic combination of pattern classifiers. *Pattern Recognit.* **2010**, *43*, 3412–3421. [\[CrossRef\]](#)
9. Jia, X.; Kuo, B.C.; Crawford, M.M. Feature mining for hyperspectral image classification. *Proc. IEEE.* **2013**, *101*, 676–697. [\[CrossRef\]](#)
10. Vergara, J.R.; Estévez, P.A. A Review of feature selection methods based on mutual information. *Neural Comput. Appl.* **2014**, *24*, 175–186. [\[CrossRef\]](#)
11. Marinoni, A.; Iannelli, G.C.; Gamba, P. An information theory-based scheme for efficient classification of remote sensing data. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 5864–5876. [\[CrossRef\]](#)
12. Bell, M.R. Information Theory and Radar: Mutual Information and the Design and Analysis of Radar Waveforms and Systems. Ph.D. Thesis, California Institute of Technology, Pasadena, CA, USA, 1988. [\[CrossRef\]](#)
13. Sheppard, C.J.R.; Larkin, K.G. Information capacity and resolution in three-dimensional imaging. *Optik* **2003**, *113*, 548–550. [\[CrossRef\]](#)
14. Weidmann, C.; Vetterli, M. Rate distortion behavior of sparse sources. *IEEE Trans. Inf. Theory.* **2012**, *58*, 4969–4992. [\[CrossRef\]](#)
15. Konings, A.G.; McColl, K.A.; Piles, M.; Entekhabi, D. How many parameters can be maximally estimated from a set of measurements? *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 1081–1085. [\[CrossRef\]](#)
16. Zhang, J.; Yang, K.; Liu, F.; Zhang, Y. Information-theoretic characterization and undersampling ratio determination for compressive radar imaging in a simulated environment. *Entropy* **2015**, *17*, 5171–5198. [\[CrossRef\]](#)
17. Chilès, J.P.; Delfiner, P. *Geostatistics: Modeling Spatial Uncertainty*, 2nd ed.; John Wiley & Sons Inc.: Hoboken, NJ, USA, 2012.
18. Chen, T.M.; Staelin, D.H.; Arps, R.B. Information content analysis of landsat image data for compression. *IEEE Trans. Geosci. Remote Sens.* **1987**, *25*, 499–501. [\[CrossRef\]](#)
19. Aiazzi, B.; Baronti, S.; Alparone, L. Lossless compression of hyperspectral images using multiband lookup tables. *IEEE Signal Process. Lett.* **2009**, *16*, 481–484. [\[CrossRef\]](#)
20. Frost, V.S.; Shanmugan, K.S. The information content of synthetic aperture radar images of terrain. *IEEE Trans. Aerosp. Electron. Syst.* **1983**, *5*, 768–774. [\[CrossRef\]](#)
21. Blacknell, D.; Oliver, C.J. Information content of coherent images. *J. Phys. D Appl. Phys.* **1993**, *26*, 1364–1370. [\[CrossRef\]](#)
22. Price, J.C. Comparison of the information-content of data from the landsat-4 thematic mapper and the multispectral scanner. *IEEE Trans. Geosci. Remote Sens.* **1984**, *3*, 272–281. [\[CrossRef\]](#)
23. Horler, D.N.H.; Ahern, F.J. Forestry information content of thematic mapper data. *Int. J. Remote Sens.* **1985**, *7*, 405–428. [\[CrossRef\]](#)
24. Nielsen, A.A.; Larsen, R. Restoration of geris data using the maximum noise fractions transform. In Proceedings of the International Airborne Remote Sensing Conference and Exhibition, Strasbourg, France, 11–15 September 1994; pp. 557–568.

25. Green, A.A.; Berman, M.; Switzer, P.; Craig, M.D. A Transformation for ordering multispectral data in terms of image quality with implications for noise removal. *IEEE Trans. Geosci. Remote Sens.* **1988**, *26*, 65–74. [\[CrossRef\]](#)
26. Razlighi, Q.R.; Kehtarnavaz, N.; Nosratinia, A. Computation of image spatial entropy using quadrilateral markov random field. *IEEE Trans. Image Process.* **2009**, *18*, 2629–2639. [\[CrossRef\]](#)
27. Wang, J.; Zhang, K.; Tang, S. Spectral and spatial decorrelation of landsat-TM data for lossless compression. *IEEE Trans. Geosci. Remote Sens.* **1995**, *33*, 1277–1285. [\[CrossRef\]](#)
28. Aiazzi, B.; Alparone, L.; Barducci, A.; Baronti, S.; Pippi, I. Estimating noise and information of multispectral imagery. *Opt. Eng.* **2002**, *41*, 656–668. [\[CrossRef\]](#)
29. Hyvärinen, A. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Trans. Neural Netw.* **1999**, *10*, 626–634. [\[CrossRef\]](#) [\[PubMed\]](#)
30. Falco, N.; Benediktsson, J.A.; Bruzzone, L. A Study on the effectiveness of different independent component analysis algorithms for hyperspectral image classification. *IEEE J. Sel. Top. Appl. Earth Observ.* **2014**, *7*, 2183–2199. [\[CrossRef\]](#)
31. Boluwade, A.; Madramootoo, C.A. Geostatistical independent simulation of spatially correlated soil Variables. *Comput. Geosci.* **2015**, *85*, 3–15. [\[CrossRef\]](#)
32. Li, Z.; Huang, P. Quantitative measures for spatial information of maps. *Int. J. Geogr. Inf. Sci.* **2002**, *16*, 699–709. [\[CrossRef\]](#)
33. Harrie, L.; Stigmar, H. An evaluation of measures for quantifying map information. *ISPRS J. Photogramm. Remote Sens.* **2010**, *65*, 266–274. [\[CrossRef\]](#)
34. Myers, D.E. To be or not to be . . . stationary? That is the question. *Math. Geol.* **1989**, *21*, 347–362. [\[CrossRef\]](#)
35. Chou, Y.M.; Polansky, A.M.; Mason, R.L. Transforming nonnormal data to normality in statistical process-control. *J. Qual. Technol.* **1998**, *30*, 133–141. [\[CrossRef\]](#)
36. Gong, W.; Yang, D.; Gupta, H.V.; Nearing, G. Estimating information entropy for hydrological data: One-dimensional case. *Water Resour. Res.* **2014**, *50*, 5003–5018. [\[CrossRef\]](#)
37. Cressie, N. Fitting Variogram models by weighted least-squares. *J. Int. Assoc. Math. Geol.* **1985**, *17*, 563–586. [\[CrossRef\]](#)
38. Woodcock, C.E.; Strahler, A.H.; Jupp, D.L.B. The use of variograms in remote sensing: I. Scene models and simulated images. *Remote Sens. Environ.* **1988**, *25*, 323–348. [\[CrossRef\]](#)
39. Woodcock, C.E.; Strahler, A.H.; Jupp, D.L.B. The use of variograms in remote sensing: II. Real digital images. *Remote Sens. Environ.* **1988**, *25*, 349–379. [\[CrossRef\]](#)
40. Asmat, A.; Atkinson, P.M.; Foody, G.M. Geostatistically estimated image noise is a function of variance in the underlying signal. *Int. J. Remote Sens.* **2010**, *31*, 1009–1025. [\[CrossRef\]](#)
41. Phiri, D.; Morgenroth, J. Developments in landsat land cover classification methods: A review. *Remote Sens.* **2017**, *9*, 967. [\[CrossRef\]](#)
42. Nunes, M.A.; Taylor, S.L.; Eckley, I.A. A Multiscale test of spatial stationarity for textured images in R. *R J.* **2014**, *6*, 20–30. [\[CrossRef\]](#)
43. Journal, A.; Zhang, T. The necessity of a multiple-point prior model. *Math. Geol.* **2006**, *38*, 591–610. [\[CrossRef\]](#)
44. Tan, X.; Tahmasebi, P.; Caers, J. Comparing training-image based algorithms using an analysis of distance. *Math. Geosci.* **2014**, *46*, 149–169. [\[CrossRef\]](#)
45. Feng, J.; Jiao, L.; Liu, F.; Sun, T.; Zhang, X. Mutual-information-based semi-supervised hyperspectral band selection with high discrimination, high information, and low redundancy. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 2956–2969. [\[CrossRef\]](#)
46. Larsen, R.; Hilger, K.B. Statistical shape analysis using non-euclidean metrics. *Med. Image Anal.* **2003**, *7*, 417–423. [\[CrossRef\]](#)

