How Many Parameters Can Be Maximally Estimated From a Set of Measurements?

Alexandra G. Konings, *Student Member, IEEE*, Kaighin A. McColl, *Student Member, IEEE*, María Piles, *Member, IEEE*, and Dara Entekhabi, *Fellow, IEEE*

***Abstract*—Remote sensing algorithms often invert multiple measurements simultaneously to retrieve a group of geophysical parameters. In order to create a robust retrieval algorithm, it is necessary to ensure that there are more unique measurements than parameters to be retrieved. If this is not the case, the inversion might have multiple solutions and be sensitive to noise. In this letter, we introduce a methodology to calculate the number of (possibly fractional) "degrees of information" in a set of measurements, representing the number of parameters that can be retrieved robustly from that set. Since different measurements may not be mutually independent, the amount of duplicate information is calculated using the information-theoretic concept of total correlation (a generalization of mutual information). The total correlation is sensitive to the full distribution of each measurement and therefore accounts for duplicate information even if multiple measurements are related only partially and nonlinearly. The method is illustrated using several examples, and applications to a variety of sensor types are discussed.**

***Index Terms*—Algorithm design, mutual information, retrieval algorithms, total correlation.**

I. Introduction

REMOTELY sensed measurements using visible, microwave, or other spectral observations of geophysical parameters are generally not a direct observation of the quantity of interest; the raw observations need to be converted to the geophysical variables in the so-called retrieval process. In most applications, multiple geophysical parameters influence the observations. These additional parameters may not be known. If so, it is often advantageous to retrieve multiple parameters at once during a single inversion. If multiple parameters are to be retrieved, however, additional measurements may be needed. These may be obtained by increasing the types of measurements made, e.g., using additional electromagnetic frequencies (spectral channels), or incidence angles and polarizations in the case of radar or radiometers. The same measurement type can also be repeated and combined, e.g., by using multiple observations over the same pixel or multiple nearby pixels. Whatever the source of the additional data, the multiple observations are rarely completely independent. This is demonstrated by the success of dimensionality-reduction methods in various areas of remote sensing, e.g., [1]–[3].

It is not possible to retrieve more unknown parameters than the number of measurements. If the set of measurements is strongly correlated, a simple integer count of the number of measurements may be overcounting the number of unknowns that can be retrieved. It is therefore necessary to be able to derive the (possibly fractional) degrees of freedom that can be obtained by using a certain set of measurements accounting for the duplicate information. Particularly in the atmospheric sounding community, this is commonly done by decomposing the signal into fractional "degrees of freedom of the signal" and "degrees of freedom of the noise" using a method by Rodgers [4]. This method assumes that the measurements are linearly related to the retrieval parameters. For many nonlinear retrieval processes, a single linearization may not be appropriate for use in designing an algorithm that is expected to be applied to large regions or even globally. Furthermore, Rodgers' method assumes that all errors are additive and have a Gaussian distribution. Not all sources of error are additive and Gaussian (for example, speckle noise in radar measurements is multiplicative, or non-Gaussian when transformed to additive decibel units), and error magnitudes may depend on ancillary variables whose global distribution is not Gaussian. The construct of Rodgers' method limits it to Gaussian variables (since only the covariance is used to characterize their probability densities), and an extension to the non-Gaussian case is not possible. In this letter, we present an alternative methodology for those cases when Rodgers' method is not appropriate. We introduce a framework that calculates the maximum number of fractional degrees of freedom [here termed the degrees of information (DoI)] in a set of measurements. The calculation depends on the full probability density function (pdf) of each of the contributing measurements rather than just their covariance and is thus expected to better capture the total amount of information in the measurements. To do this, we propose the use of the normalized total correlation, a generalization of the normalized mutual information. These information-theoretic measures are discussed in Section II.

Information-theoretic concepts have found a wide variety of applications in remote sensing, e.g., [5]–[9]. Here, mutual information and its generalization are used for a specific application in model selection: determining how many unknown parameters can be maximally retrieved from a given data set. The proposed method is independent of, and does not attempt to influence, the exact choice of parameters to be retrieved,

Manuscript received September 22, 2014; accepted November 10, 2014. Date of publication January 8, 2015; date of current version February 6, 2015.

A. G. Konings, K. A. McColl, and D. Entekhabi are with the Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: konings@mit.edu).

M. Piles is with the Remote Sensing Laboratory, Departament de Teoria del Senyal i Comunicacions, Universitat Politècnica de Catalunya, 08034 Barcelona, Spain, and also with SMOS Barcelona Expert Center, 08003 Barcelona, Spain.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LGRS.2014.2381641

only the number of parameters. Although similar information-theoretic concepts (e.g., appropriately chosen combinations of joint and/or conditional entropies) could be used to determine which parameters the observations provide the most information about, the choice of retrieval parameters may be driven by diverse scientific questions or other design factors. The framework presented therefore determines the maximum DoI in the data independently of which particular parameters are to be retrieved or any particular retrieval algorithm. Indeed, it may not always be practical to introduce as many parameters as there are DoI, but the method determines an upper bound.

This letter is organized as follows. In Section II-A, the normalized total correlation is introduced and shown to be confined between 0 and $N-1$. The fractional "DoI" contained in an N-dimensional observation set is linked to the normalized mutual information. The DoI are dependent on the precision of the measurements through a bin size parameter Δ, whose derivation is discussed in Section II-B. This derivation is illustrated with an example using microwave radiometer observations. Additional examples are described in Section III. Finally, in Section IV, the application of the metric to a number of different common measurement types in remote sensing is discussed. This latter section is meant to be illustrative rather than exhaustive.

II. DoI

If measurements are made of two independent random variables X and Y, they can be used to retrieve two unknowns. The set of measurements can be said to contain two "DoI." If a third measurement Z is added that can be perfectly predicted from one of the other two random variables, the measurements still contain only two DoI. In reality, it is more likely that X and Y are related but not completely independent, and Z is similarly correlated to some degree with either X and Y. Depending on how closely related the three variables are, there could be enough information in the correlated random variables to retrieve either one or two unknowns. Some measure of the total amount of information contained in a set of measurements is therefore needed. The measure should be independent of the nature of the relationship between the variables (i.e., not restricted to linear relationships) and generalizable to an arbitrary number of dimensions (number of measurement channels). The DoI provides such a measure and is introduced in Section II-A. Section II-B discusses the bin size parameter necessary to calculate DoI.

A. Definition of Degrees of Information

The Shannon entropy, one of the central tenets of information theory, is the expected value of the information content derived from a single observation of a discrete random variable X. It can also be interpreted as the uncertainty of a variable [10]. The Shannon entropy can be expressed as

$$H(X) = \sum_x p(x) \log p(x) \tag{1}$$

where $p(x)$ is the probability mass function (pmf) of X. If the random variable has a narrow distribution, an observation will, on average, provide less information than if it has a very broad distribution. The $H(X)$ of a discrete random variable is nonnegative. For multiple variables X_i, the joint entropy is

$$H(X_1, \ldots, X_N) = \sum_{x_1} \cdots \sum_{x_N} p(x_1, \ldots, x_N) \log p(x_1, \ldots, x_N). \tag{2}$$

The individual $p(X_i)$ are referred to as the "marginal pmfs," and the individual $H(X_i)$ are referred to as the "marginal entropy(ies)" of each product.

The mutual information is a well-known measure of the reduction in uncertainty between independent and joint measurements of two random variables X and Y. Mathematically, this can be written as

$$I(X;Y) = \sum_y \sum_x p(x,y) \log \frac{p(x,y)}{p(x)p(y)}. \tag{3}$$

By comparing the joint and marginal probability distributions, the mutual information quantifies the degree to which simultaneous consideration of the two variables changes their distribution. That is, it quantifies nonlinearly how dependent the two variables are. When X and Y are independent, $I(X;Y) = 0$. The $I(X;Y)$ is maximized when X and Y are dependent (i.e., perfectly correlated). From the definitions, it can easily be shown that

$$I(X;Y) = H(X) + H(Y) - H(X,Y). \tag{4}$$

For proofs of these and other information-theoretic properties used in this section, the reader is referred to an introductory information theory textbook, such as [10].

Although a number of different generalizations of the mutual information exist, the total correlation C [11] captures the amount of information shared between *any* of the measurements in a set. Like the mutual information, the total correlation is the Kullback–Leibler divergence between the joint and the marginal entropies

$$C(X_1, X_2, \ldots, X_N) = \int_{x_1} \cdots \int_{x_N} p(x_1, \ldots, x_N) \times \log \frac{p(x_1, \ldots, x_N)}{p(x_1), \ldots, p(x_N)} dx_N, \ldots, dx_1 \tag{5}$$

$$= \sum_{i=1}^{N} H(X_i) - H(X_1, \ldots, X_N). \tag{6}$$

We further define the normalized total correlation $C_n(X_1, \ldots, X_N)$ as

$$C_n(X_1, \ldots, X_n) = \frac{C(X_1, \ldots, X_N)}{H(X_1, \ldots, X_N)} = \frac{\sum_{i=1}^{N} H(X_i) - H(X_1, \ldots, X_N)}{H(X_1, \ldots, X_N)}. \tag{7}$$

To prove that C_n takes a value between 0 and $N-1$, we use the basic property that

$$H(X_1, \ldots, X_N) \leq \sum_{i=1}^{N} H(X_i) \Longrightarrow \; C_n \geq 0. \tag{8}$$

Since $\max_i(H(X_i)) \leq H(X_1, \ldots, X_N)$, multiplying by N gives

$$N \max_i (H(X_i)) \leq NH(X_1, \ldots, X_N). \tag{9}$$

Furthermore, since $\sum_{i=1}^{N} H(X_i) \leq N \max_i H(X_i)$ by definition

$$\sum_{i=1}^{N} H(X_i) \leq NH(X_1, \ldots, X_N) \implies \frac{\sum_{i=1}^{N} H(X_i)}{H(X_1, \ldots, X_N)} \leq N. \tag{10}$$

By inserting this into (7), it becomes clear that $C_n \leq N - 1$. The C_n therefore takes a value between 0 and $N - 1$. When the X_i are independent, $C_n = 0$. When they have a one-to-one relationship, $C_n = N - 1$.

The higher the normalized total correlation between the measurements, the less information they contain. The total DoI between the X_i is then given by

$$\text{DoI} = N - C_n(X_1, \ldots, X_N). \tag{11}$$

Since additional measurements cannot *remove* information from the first one, DoI ≥ 1. Since, as mentioned earlier, mutual information and entropy are nonnegative, DoI $\leq N$. Thus, DoI $\in [1, N]$, as expected.

In the limit where $C(X_1, \ldots, X_N)$ is maximized, $H(X_1, \ldots, X_N) = H(X_i)$ for all i. Thus, it is possible to derive an alternative normalization using the minimum $H(X_i)$. The normalization with $H(X_1, \ldots, X_N)$ used here is chosen because it is more conservative since $\min(H(X_i)) \leq H(X_1, \ldots, X_N)$. Note that, in two dimensions, Le Hegarat-Mascle *et al.* [12] calculated the mutual information between two remote sensing images, but these were normalized by the entropy of one of the two images so that the resulting measure is not symmetric.

The aforementioned properties were derived based on the assumption that the X_i are discrete variables. In remote sensing, many measurements are continuous rather than discrete. For continuous variables, several of the aforementioned lemmas are false, and mutual information does not have an effective upper bound. Nevertheless, while remote sensing measurements may *appear* to be continuous by taking on an arbitrarily large number of values, the number of possible measurements is in practice limited by the finite accuracy or precision of the instruments. That is, small fluctuations in measurements below some accuracy threshold do not provide any physical information. For a certain bin size Δ, the continuous measurements can be binned into discrete classes by rounding them to the nearest interval of Δ. The resulting constant-bin histograms can be used directly to estimate the pmfs necessary to evaluate C_n.

B. Dependence on the Bin Size Parameter

Using an inappropriate bin size Δ may introduce errors in the estimation of the pmfs and, thus, in the C_n and DoI. If the bin size used is too small, the frequency counts in the bins will be sensitive to noise fluctuations in the data set. If the bin size used is too large, the estimated marginal and joint pmfs may mischaracterize (or even miss altogether) certain peaks in the distribution. Several different approaches have been proposed in the statistical literature to determine the optimal bin width to accurately estimate the pmf with a finite sample. Among these,

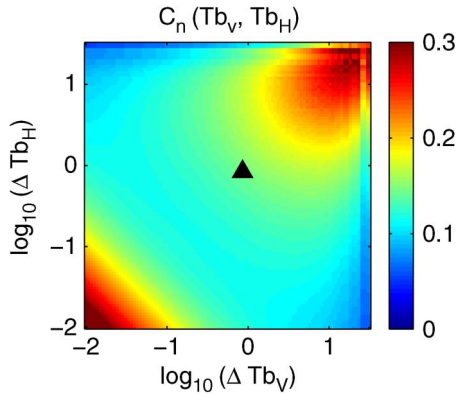

Fig. 1. Normalized total correlation C_n between Aquarius Tb_V and Tb_H as a function of the bin sizes ΔTb_V and ΔTb_H. For large bin sizes relative to the dynamic range of the variables, the C_n suddenly drops when the number of bins is so low that even the approximate shape of the joint pmf is distorted by the wide bins. The black triangle corresponds to the bin sizes recommended by Scott's rule.

Sturges' rule for calculating a bin width based on the range of the data and the number of points is the oldest and the most common. It has been shown to work well for applications of mutual information-based image registration [13] and feature selection [14]. However, it is known to lead to overly large bin size estimates that oversmooth the histogram, particularly for large samples sizes (which are expected in remote sensing) [15]. It can also be sensitive to outliers. A better approach is the so-called Scott's rule, which calculates the bin size Δ_{x_i} from the standard deviation σ_{X_i} of the data instead of their range

$$\Delta_{X_i} = \frac{3.5\sigma_{X_i}}{n^{1/3}} \tag{12}$$

where n is the number of points in the sample. The dependence on $n^{1/3}$ has been shown to be optimal for minimizing L_p error norms [15].

The use of Scott's rule is illustrated using a 2-D example for ease of visualization. Horizontally and vertically polarized measurements of L-band brightness temperatures (Tb_V and Tb_H, respectively) from the Aquarius satellite are used [16]. The data span the period from September 1, 2011 to August 31, 2012 over land and across the globe. Aquarius has three beams with three different incidence angles; only the middle beam is used here. Fig. 1 shows the bin size dependence of the $C_n(Tb_V, Tb_H)$. Since the range and shape of the distribution are similar between the two variables, it is not surprising that the dependence on Δ_{Tb_V} and Δ_{Tb_H} is approximately symmetric. Applying Scott's rule to each of the Tb_V and Tb_H separately leads to two different bin sizes that can be used to determine C_n.

The $C_n(Tb_V, Tb_H) = 1.13/7.87 = 0.14$ at the optimum bin size. This is much lower than the Pearson's correlation coefficient between the values, $r = 0.92$. The joint pmf shown at the bottom of Fig. 2 illustrates why. Although the shapes of the marginal distributions are similar, the long tail in the joint pmfs adds a significant amount of uncertainty between the two polarizations. By contrast, it reduces the Pearson correlation coefficient relatively little because most points fall on or near the diagonal line. This demonstrates the value of nonparametric measures of the degrees of freedom in measurements rather than relying on potentially misleading Gaussian assumptions.

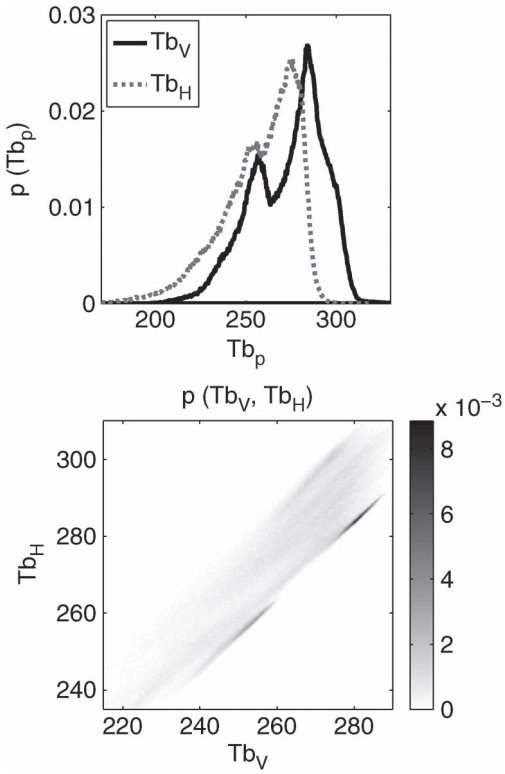

Fig. 2. (Top) Marginal and (bottom) joint pdfs for observed vertically and horizontally polarized brightness temperatures (Tb_V and Tb_H, respectively) from the Aquarius satellite. Note that the edges of the joint pdf tails extend beyond the region shown; the figure is zoomed in for clarity.

The resulting value of $\text{DoI} = 2 - 0.14 = 1.86$ allows the calculation of the number of overpasses that must be combined to calculate a certain number of parameters from a multitemporal time series using these data. Algorithms using dual-pol radiometric data at L-band from N overpasses can retrieve a maximum of 1.86 × N parameters. Any dependent information between observations at different times is due to the autocorrelation in the physical properties to be retrieved, which is generally neglected in the retrieval process. Thus, the DoI from a single set of dual-polarized measurements is multiplied by N. For example, combining data from two overpasses leads to $\text{DoI}_{2-pass} = 1.86(2) = 3.72$, which is only enough information to robustly retrieve three parameters, even if four measurements are used (two polarizations on two overpasses each). Indeed, a two-overpass time-series algorithm can be applied to these data to robustly retrieve three parameters for each pixel: a single constant vegetation optical depth and the dielectric constant during both overpasses. Additional combinations of overpass numbers and retrieved parameters are also possible [17].

III Example DoI Calculations

In this section, the DoI calculation is illustrated for several additional measurements. Table I shows the $C_n(X_1, \ldots, X_N)$ for several data sources and compares different rearrangements of the same time series. The examples of Table I are discussed one by one hereinafter.

Noisy Linear Relationship: We first consider the case of two linearly related time series of unit slope, e.g., $Y = X$, both distributed normally around 0 with a standard deviation of 1. The X and Y are jointly sampled but are subject to independent normally distributed noise with a standard deviation of 0.1 to produce the series x and y. The mutual information between them measures the respective dependence of variables based on their joint distributions. The addition of independent noise to all values strongly reduces the amount of redundancy between the final measurements; $C_n(X, Y) = 0.21$.

TABLE I
DoI for Several Examples

Datasets	Δ_{opt}	N	C_n	DoI
Noisy linear	(0.10, 0.10)	2	0.21	1.79
$(\sigma_{HH}, \sigma_{VV})$	(0.20 dB, 0.20 dB)	2	0.28	1.72
$(\sigma_{HH}, \sigma_{HV})$	(0.20 dB, 0.30 dB)	2	0.19	1.81
$(\sigma_{VV}, \sigma_{HV})$	(0.20 dB, 0.30 dB)	2	0.18	1.82
$(\sigma_{HH}, \sigma_{VV}, \sigma_{HV})$	(0.20 dB, 0.20 dB, 0.30 dB)	3	0.40	2.60
(σ_{VV}, Tb_V)	(0.20 dB,0.87 K)	2	0.03	1.97
(Tb_V, Tb_H)	(0.87 K, 0.81K)	2	0.14	1.86

Aquarius Multipolarization Backscattering Data σ_{HH}, σ_{VV}, and σ_{HV}: Aquarius makes coincident radar and radiometric measurements. There is a higher normalized total correlation between pairs of two copolarized backscattering coefficients $(C_n(\sigma_{HH}, \sigma_{VV}) = 0.28)$ than between a combination of copolarized and cross-polarized backscattering coefficients $(C_n(\sigma_{HH}, \sigma_{HV}) = 0.19)$. This can be understood by noting that the cross-polarized backscatter is essentially independent of the soil moisture, unlike the copolarized backscatter. Some total correlation remains because both the copolarized and cross-polarized backscatters are sensitive to vegetation and soil roughness. A set of cross-polarized and copolarized data thus carries more information than two different copolarizations, as reflected in the higher DoI. When adding a third polarization, the DoI increases by less than one, as expected from the nonzero C_n between all pairs of polarizations. The total C_n increases when combining all three polarizations, suggesting that the mutual information between different pairs of polarizations is in different parts of the pmf (e.g., different spatial regions or seasons). The total DoI is 2.60.

Aquarius Multi-Instrument Data σ_{VV} and Tb_V: Because radar and radiometric measurements are affected differently by soil and vegetation scattering, the C_n between coincident brightness temperature Tb_v and backscattering coefficient σ_{VV} data is low, $C_n(Tb_v, \sigma_{VV}) = 0.03$. Other combinations of backscatter and brightness temperatures had even lower total correlation and thus contain more DoI.

IV. Applications to Particular Remote Sensing Observations

The DoI framework can be applied to a variety of remote sensing observations and used to determine how many geophysical parameters can be maximally retrieved. Note that, in hyperspectral imagery, the determination of the number of parameters that can be retrieved from unmixing algorithms is known as the "intrinsic dimensionality" problem and has been well studied (e.g., [18]–[20]). The high number of dimensions in these images (generally more than 100) makes total correlation computationally expensive to calculate for such images. Instead, the primarily application of this method

is to monospectral, multispectral, and lidar data, as outlined hereinafter and shown by example in Section III.

Microwave Radiometry: Radiometric measurements are made at a certain incidence angle, frequency, and polarization. For a given incidence angle and frequency then, DoI ≤ 2 (DoI ≤ 4 if the radar is fully polarimetric). Additional information can be obtained by measuring the same pixel at multiple incidence angles. This concept is used by the soil moisture retrieval algorithm of the European Space Agency's Soil Moisture Ocean Salinity satellite [21], among others. The DoIs can provide a framework to calculate how many geophysical and biophysical variables can be determined from a collection of correlated multiangular measurements. Similar principles apply for multitemporal retrieval algorithms, which combine measurements made at different times under the assumptions that at least one of the retrieval parameters is constant over the time period between the observations [17] or for multifrequency algorithms.

Radar: Whether the data are obtained using a real or synthetic aperture, the return from radar systems can generally be described by a maximum of eight parameters—the phase and amplitude of the backscattered waves in two possible transmit polarizations and two possible receive polarizations. (Radar altimetry applications, which are based on the signal return time, provide an exception.) Thus, the number of DoIs in a single set of measurements can be no more than eight, even though radar scattering is sometimes expressed in a 16-element Mueller matrix.

As in passive microwave applications, polarimetric, multi-incidence angle, and multitemporal methods [22] can be used to increase the number of geophysical variables that can be retrieved. The DoI can be used to determine how many polarizations, angles, or temporal samples are needed.

Lidar: The DoI framework may not be as useful for discrete pulse lidars as for other measurement types because different returns view different parts of the canopy. However, DoI can be informative when applied to waveform-recording lidars, whether used to retrieve canopy biophysical parameters or atmospheric composition information. Unlike in radar systems, the incidence angle does not vary, and multi-incidence angles cannot be used to increase the DoI in the system. Instead, lidar observations at multiple wavelengths and depolarizations could be used to infer multiple properties.

V. Conclusion

When designing retrieval algorithms, the first choice to be made is the number of parameters to be retrieved from the measurements. The DoI framework presented in this letter provides a method for estimating how many parameters can maximally be retrieved, depending on the amount of duplicate information present in the joint pmf of the measurements. The use of the entire joint pmf allows accounting for the fact that less-commonly occurring measurements add a lot of uncertainty to the retrieval, and leads to a better estimate of the uncertainty in the data.

The DoI in the measurements are independent of the type of retrieval algorithm, whether it is statistical, physical, or some combination thereof. Once the DoI is obtained, the maximum number of independent parameters that can be retrieved is given by the floor of the DoI. Generally, the presence of noise implies that not all the information in the measurements can be used for parameter retrieval. The true information content of a set of measurements is thus below the DoI. Most retrieval algorithms retrieve each parameter independently (i.e., all combinations of parameter values are possible solutions). In this case, the degrees of freedom needed for the retrieval is exactly equal to the number of parameters to be retrieved.

References

[1] N. Renard and S. Bourennane, "Dimensionality reduction based on tensor modeling for classification methods," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 4, pp. 1123–1131, Apr. 2009.

[2] G. Licciardi, P. Marpu, J. Chanussot, and J. Benediktsson, "Linear versus nonlinear PCA for the classification of hyperspectral databased on the extended morphological profiles," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 3, pp. 447–451, May 2012.

[3] L. Guanter *et al.*, "Retrieval and global assessment of terrestrial chlorophyll fluorescence from GOSAT space measurements," *Remote Sens. Environ.*, vol. 121, pp. 236–251, Jun. 2012.

[4] C. Rodgers, *Inverse Methods for Atmopsheric Sounding: Theory and Practice*, F. Taylor, Ed. Singapore: World Scientific, 2000.

[5] L. Gueguen, S. Cui, G. Schwarz, and M. Datcu, "Multitemporal analysis of multisensor data: Information theoretical approaches," in *Proc. IEEE IGARSS*, 2010, pp. 2559–2562.

[6] E. Erten, A. Reigber, L. Ferro-Famil, and O. Hellwich, "A new coherent similarity measure for temporal multichannel scene characterization," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 7, pp. 2839–2851, Jul. 2012.

[7] D. Cerra and M. Datcu, "A model conditioned data compression based similarity measure," in *Proc. DCC*, 2008, p. 509.

[8] C. Cariou, K. Chehdi, and S. Le Moan , "BandClust: An unsupervised band reduction method for hyperspectral remote sensing," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 3, pp. 565–569, May 2011.

[9] M. Hossain, X. Jia, and M. Pickering, "Subspace detection using a mutual information measure for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 2, pp. 424–428, Feb. 2014.

[10] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Hoboken, NJ, USA: Wiley, 1991.

[11] S. Watanabe, "Information theoretical analysis of multivariate correlation," *IBM J. Res. Develop.*, vol. 4, no. 1, pp. 66–82, Jan. 1960.

[12] S. Le Hegarat-Mascle, D. Vidal-Madjar, O. Taconet, and M. Zribi "Application of Shannon information theory to a comparison between L- and C-band SIR-C polarimetric data versus incidence angle," *Remote Sens. Environ.*, vol. 60, no. 2, pp. 121–130, May 1997.

[13] P. A. Legg, P. L. Rosin, D. Marshall, and J. E. Morgan, "Improving accuracy and efficiency of registration by mutual information using Sturges histogram rule," in *Proc. Med. Image Understand. Anal.*, 2007, pp. 17–18.

[14] A. Hacine-Gharbi, P. Ravier, R. Harba, and T. Mohamadi, "Low bias histogram-based estimation of mutual information for feature selection," *Pattern Recognit. Lett.*, vol. 33, no. 10, pp. 1302–1308, Jul. 2012.

[15] M. Wand, "Data-based choice of histogram bin width," *Amer. Stat.*, vol. 51, no. 1, pp. 59–64, Feb. 1997.

[16] D. Le Vine, G. S. E. Lagerloef, F. Colomb, S. Yueh, and F. Pellerano, "Aquarius: An instrument to monitor sea surface salinity from space," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 7, pp. 2040–2050, Jul. 2007.

[17] M. Piles *et al.*, "Vegetation optical depth retrieval using time-series of dual-polarized L-band radiometer observations," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, submitted for publication.

[18] M. Hasanlou and F. Samadzadegan, "Comparative study of intrinsic dimensionality estimation and dimension reduction techniques on hyperspectral images using K-NN classifier," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 6, pp. 1046–1050, Nov. 2012.

[19] K. Cawse-Nicholson, S. Damelin, A. Robin, and M. Sears, "Determining the intrinsic dimension of a hyperspectral image using random matrix theory," *IEEE Trans. Image Proc.*, vol. 22, no. 4, pp. 1301–1310, Apr. 2013.

[20] R. Heylen and P. Scheunders, "Hyperspectral intrinsic dimensionality estimation with nearest-neighbor distance ratios," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, pp. 570–579, Apr. 2013.

[21] Y. Kerr *et al.*, "The SMOS soil moisture retrieval algorithm," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 5, pp. 1384–1403, Apr. 2012.

[22] S.-B. Kim *et al.*, "Soil moisture retrieval using time-series radar observations over bare surfaces," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 5, pp. 1853–1863, May 2012.