An Analysis of the Information Dependence Between MODIS Emissive Bands

Srikanth Gottipati[a] and Irina Gladkova[a] and Michael Grossberg[a]

[a]CCNY, NOAA/CREST, 138th Street and Convent Avenue, New York, NY 10031, USA

ABSTRACT

Multispectral, hyperspectral and ultraspectral imagers and sounders are increasingly important for atmospheric science and weather forecasting. The recent advent of multipsectral and hyperspectral sensors measuring radiances in the emissive IR are providing valuable new information. This is due to the presence of spectral channels (in some cases micro-channels) which are carefully positioned in and out of absorption lines of CO2, ozone, and water vapor. These spectral bands are used for measuring surface/cloud temperature, atmospheric temperature, Cirrus clouds water vapor, cloud properties/ozone, and cloud top altidude etc.

The complexity of the spectral structure wherein the emissive bands have been selected presents challenges for lossless data compression; these are qualitatively different than the challenges offered by the reflective bands. For a hyperspectral sounder such as AIRS, the large number of channels is the principal contributor to data size. We have shown that methods combining clustering and linear models in the spectral channels can be effective for lossless data compression. However, when the number of emissive channels is relatively small compared to the spatial resolution, such as with the 17 emissive channels of MODIS, such techniques are not effective. In previous work the CCNY-NOAA compression group has reported an algorithm which addresses this case by sequential prediction of the spatial image. While that algorithm demonstrated an improved compression ratio over pure JPEG2000 compression, it underperformed optimal compression ratios estimated from entropy. In order to effectively exploit the redundant information in a progressive prediction scheme we must, determine a sequence of bands in which each band has sufficient mutual information with the next band, so that it predicts it well.

We will provide a covariance and mutual information based analysis of the pairwise dependence between the bands and compare this with the qualitative expected dependence suggested by a physical analysis. This compression research is managed by Roger Heymann, PE of OSD NOAA NESDIS Engineering, in collaboration with the NOAA NESDIS STAR Research Office through Mitch Goldberg, Tim Schmit, Walter Wolf.

1. INTRODUCTION

Multispectral, hyperspectral and ultraspectral imagers and sounders are increasingly important for atmospheric science and weather forecasting. The recent advent of multipsectral and hyperspectral sensors measuring radiances in the emissive IR are providing valuable new information. This is due to the presence of spectral channels (in some cases micro-channels) which are carefully positioned in and out of absorption lines of CO2, ozone, and water vapor. These spectral micro-channels enable determination of the vertical temperature and moisture structure of the atmosphere as well as characterization of some of the trace gases.

The complexity of the spectral structure wherein the emissive bands have been selected presents challenges for lossless data compression; these are qualitatively different than the challenges offered by the reflective bands. For a hyperspectral sounder such as AIRS, the large number of channels is the principal contributor to data size. We have shown that methods combining clustering and linear models in the spectral channels can be effective for lossless data compression. However, when the number of emissive channels is relatively small compared to the spatial resolution, such as with the 17 emissive channels of MODIS, such techniques are not effective. In previous work[1] the CCNY-NOAA compression group has reported an algorithm which addresses this case by sequential prediction of the spatial image. While that algorithm demonstrated an improved compression ratio

Further author information: (Send correspondence to I. Gladkova)
E-mail: gladkova@@cs.ccny.cuny.edu, Telephone: 1 212 650 6261

Satellite Data Compression, Communication, and Processing IV
edited by Bormin Huang, Roger W. Heymann, Joan Serra-Sagrista
Proc. of SPIE Vol. 7084, 70840J, (2008) · 0277-786X/08/$18 · doi: 10.1117/12.800817

over pure JPEG2000 compression, it significantly underperformed optimal compression ratios estimated from entropy, see Table 1. In order to effectively exploit the redundant information in a progressive prediction scheme we must determine a sequence of bands in which each band has sufficient mutual information with the next band, so that it predicts it well. In this paper we will provide a covariance and entropy based analysis of the pairwise dependence between the bands and compare this with the qualitative expected dependence suggested by a physical analysis.

Table 1. Average MODIS compression ratios for sample granules using SPIE07 algorithm[1] and estimated entropies at different resolutions.

Resolution	SPIE07	Entropy estimated
250m	3.03	3.25
500m	3.20	3.55
1km (day)	4.04	4.32
1km (night)	3.28	5.23

2. MODIS

The MODerate resolution Imaging Spectroradiometer (MODIS) is a key instrument aboard the Terra (EOS AM) and Aqua (EOS PM) polar satellites. Terra's orbit around the Earth is timed so that it passes from north to south across the equator in the morning, while Aqua passes south to north over the equator in the afternoon. The MODIS instrument provides high radiometric sensitivity (12 bit) in 36 spectral bands ranging in wavelength from 0.4 μm to 14.4 μm. These 36 distinct spectral bands are divided into four separate Focal Plane Assemblies (FPA): Visible (VIS), Near Infrared (NIR), Short- and Mid-Wave Infrared (SWIR/MWIR), and Long-Wave Infrared (LWIR). Each FPA focuses light onto a certain section of detector pixels, which are relatively large, ranging from 135 μm to 540 μm square. The large number and variety of detector pixels are what make the wide variety of MODIS data possible. When light hits a detector pixel, it will generate a distinct signal depending on the type of light it is sensitive to. The signals that the pixels generate are what scientists process and study to learn about Earth's land surfaces, water surfaces, and atmosphere. There are 10 detector elements along track for each of the 1 km bands, 20 for each of the 500 m bands, and 40 for the 250 m bands. Table 2 gives a brief description of the 17 emissive IR bands. These bands are used for measuring surface/cloud temperature, atmospheric temperature, Cirrus clouds water vapor, cloud properties/ozone, and cloud top altidude etc. They have 17 spectral bands ranging in wavelength from 3.6 μm to 14.4 μm.

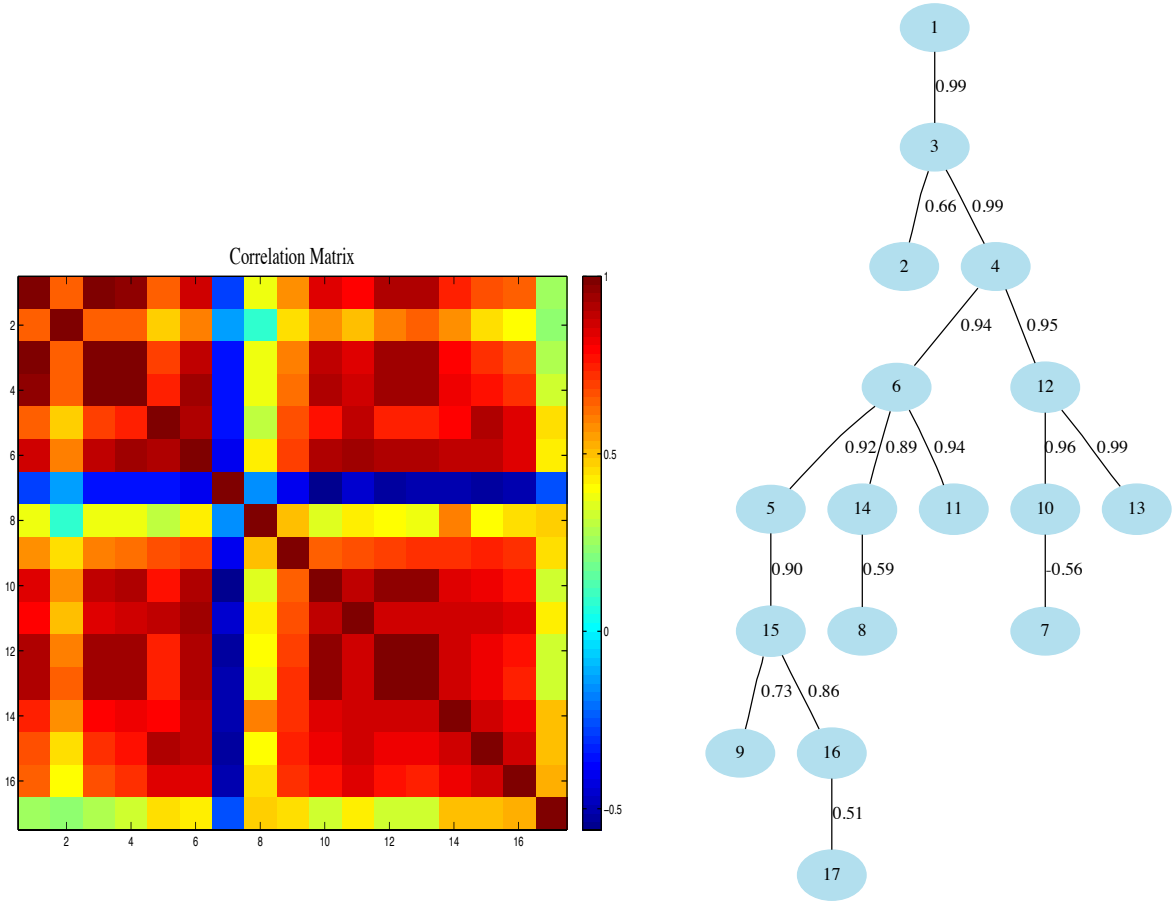
3. LINEAR CORRELATIONS

In probability theory and statistics, correlation (often measured as a correlation coefficient), indicates the strength and direction of a linear relationship between two random variables. In general statistical usage, correlation or co-relation refers to the departure of two variables from independence. A number of different coefficients are used for different situations. The best known is the Pearson product-moment correlation coefficient, which is obtained by dividing the covariance of the two variables by the product of their standard deviations.

Figure 1-(a) shows the correlation matrix for the 17 emissive channels of a sample of 7 MODIS granules. Notice that there is a very low linear correlation between channel 7 and all the other channels. Figure 1-(b) shows the spanning tree with the maximum possible total correlation among various channels such that each channel is included in the tree. From this graph one can see that different subgraphs do make sense as they belong to the same class in Table 2. However there does exist some cross correlation between channels belonging to different classes.

Table 2. MODIS IR Emmisive channels

Primary Use	Band	Bandwidth	Spectral Radiance
Surface/Cloud Temperature	20	3.660 - 3.840	0.45(300K)
	21	3.929 - 3.989	2.38(335K)
	22	3.929 - 3.989	0.67(300K)
	23	4.020 - 4.080	0.79(300K)
Atmospheric Temperature	24	4.433 - 4.498	0.17(250K)
	25	4.482 - 4.549	0.59(275K)
Cirrus Clouds Water Vapor	26	1.360 - 1.390	6.00
	27	6.535 - 6.895	1.16(240K)
	28	7.175 - 7.475	2.18(250K)
Cloud Properties	29	8.400 - 8.700	9.58(300K)
Ozone	30	9.580 - 9.880	3.69(250K)
Surface/Cloud Temperature	31	10.780 - 11.280	9.55(300K)
	32	11.770 - 12.270	8.94(300K)
Cloud Top Altitude	33	13.185 - 13.485	4.52(260K)
	34	13.485 - 13.785	3.76(250K)
	35	13.785 - 14.085	3.11(240K)
	36	14.085 - 14.385	2.08(220K)



(a) (b)

Figure 1. Figure (a) shows the correlation matrix for a sample of MODIS granules. The correlation matrix determines the strength of linear dependence among the 17 MODIS IR emissive channels. Figure (b) shows the spanning tree with the maximum possible total correlation among various channels and such that each channel is included in the tree.

4. ENTROPY APPROACHES

While the correlation matrix among various channels does give us a measure of how each channel is correlated to the other, the correlation method itself is based on the assumptions of linearity, i.e., one is forced to assume that two channels A and B are linearly related. While such an assumption gives us a good first approximation of the degree of dependence between two channels, one should look at further measures of dependence between two channels. In this section we look at two such measures which are well studied in the field of information theory, mutual information and relative entropy.

Since entropy is based on the probability distribution, entropy estimation is related to estimation of the probability. While the imager data are quantized, the bit depth of the data is 12 bits. Since we are averaging over many measurements, we can approximate the discrete entropy by computing the differential entropy of a continuous probability density function.

The non-parametric Kozachenko-Leonenko (K-L) entropy estimator[2–4] is described here. Let $\rho(x, y)$ denote the Euclidean distance between two points $x, y \in \mathbf{R}^m$. For a given sample $X_1, \ldots, X_N$, and a given X_i in the sample, from the $N-1$ distances $\rho(X_i, X_j),\ j = 1, \ldots, N,\ j \neq i$, we form the order statistics

$$\rho^{(i)}_{1,N-1} \leq \rho^{(i)}_{2,N-1} \leq \cdots \leq \rho^{(i)}_{N-1,N-1}. \tag{1}$$

Therefore, $\rho^{(i)}_{1,N-1}$ is the nearest-neighbor distance from the observation X_i to some other X_j in the sample, $j \neq i$, and similarly $\rho^{(i)}_{k,N-1}$ is the k-th nearest neighbor distance from X_i to some other X_j. Then an asymptotically unbiased and consistent estimator for Shannon's entropy $H(X)$ is given by

$$\hat{H}(X) = \frac{1}{N}\sum_{i=1}^{N} \log\left[(N-1)\exp[-\Psi(k)]V_m(\rho^{(i)}_{k,N-1})^m\right], \tag{2}$$

where $\Psi(z) = \Gamma'(z)/\Gamma(z)$ is the digamma function and V_m is the volume of a unit ball $\mathcal{B}(0, 1)$ in $\mathbf{R}^m$.

The K-L entropy estimator performs poorly (in terms of bias) in high dimensions due to a combination of computational complexity and lack of data. In a d-dimensional space one needs 2^d data points in order to get an unbiased entropy estimate. However, we live in a data-poor situation. Even in a data-rich situation due to the computational complexity of nearest-neighbor computations one is forced to use a random (smaller size) sample data. The CCNY-NOAA compression group[5] has developed a technique to remove the bias from the K-L non-parametric entropy estimator to estimate the differential entropy of data in high dimensions. We have done this by introducing a novel method to remove the bias of the K-L entropy estimator when the bias is asymptotically gaussian. We fit the data to a guassian mixture model and compute its true entropy and also compute the K-L entropy estimate of randomnly sampled data using the same guassian mixture model. The difference between these two entropies gives us the magnitude of the bias in the K-L entropy estimate which enables bias correction for the K-L entropy estimate of the data.

The nearest-neighbors are computed using the approximate nearest-neighbor method.[6,7] Computing exact nearest neighbors in dimensions much higher than 8 seems to be a very difficult task. Few methods seem to be significantly better than a brute-force computation of all distances. However, it has been shown that by computing nearest neighbors approximately, it is possible to achieve significantly faster running times.

4.1 Mutual information

Intuitively, mutual information measures the information that two random variables X and Y share: it measures how much knowing one of these variables reduces our uncertainty about the other. For example, if X and Y are independent, then knowing X does not give any information about Y and vice versa, so their mutual information is zero. At the other extreme, if X and Y are identical then all information conveyed by X is shared with Y: knowing X determines the value of Y and vice versa. As a result, the mutual information is the same as the uncertainty contained in Y (or X) alone, namely the entropy of Y. The mutual information of two random variables X and Y is given by

$$I(X, Y) = H(X) + H(Y) - H(X, Y), \tag{3}$$

where each term can be estimated by using equation (2).

Figure 2 shows the mutual information matrices $I(c_i, c_j),\ \forall i, j \in \{1, 2, \ldots, 17\}$, where c_i represents channel i, for a MODIS granule with varying window size from 1×1 to 10×10. As we increase the window size the mutual information matrices stabilize. Figure 3 shows the spanning tree with maximum possible total mutual information and includes all the channels for MODIS granule. Notice that this graph is quite different from that shown in Figure 1-(b).

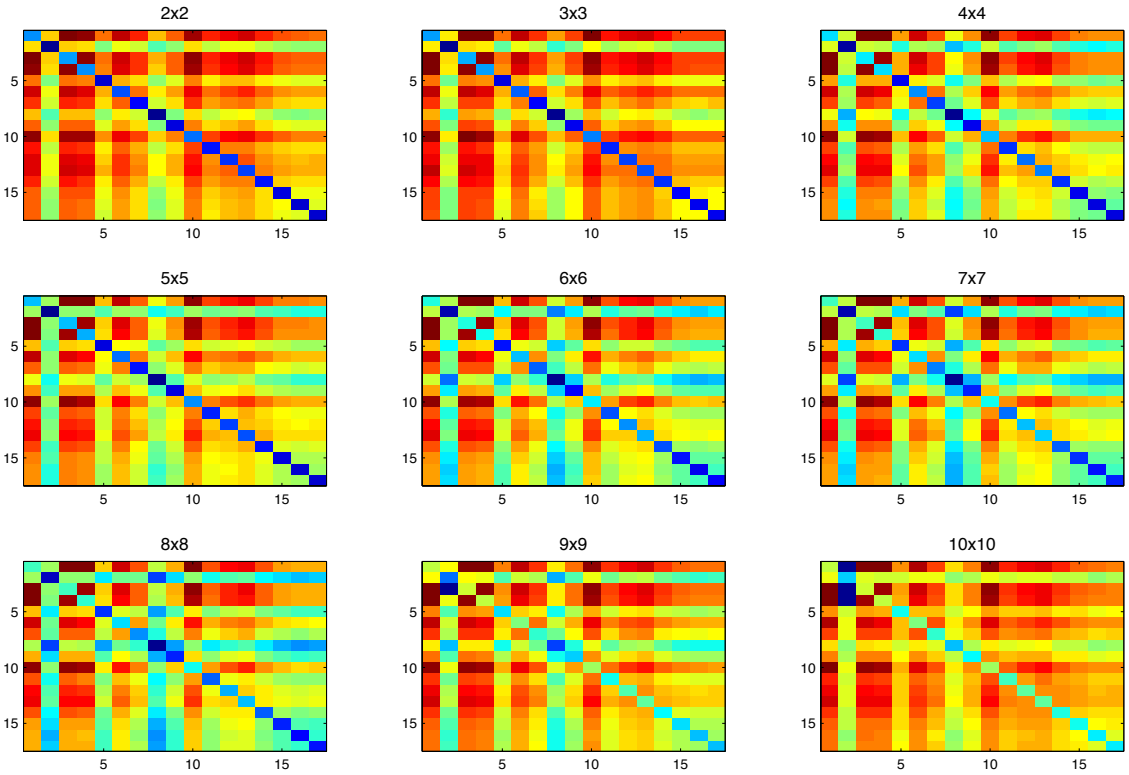


Figure 2. The figure shows the mutual information $I(c_i, c_j),\ \forall i, j \in \{1, 2, \ldots, 17\}$ matrices for MODIS granule MOD01.A2006.176.0955.005.2006214124831 with varying window sizes from 1×1 to 10×10.

4.2 Relative entropy

In probability theory and information theory, relative entropy is a non-commutative measure of the difference between two probability distributions P and Q. In information theory, t measures the expected difference in the number of bits required to code samples from P when using a code based on P, and when using a code based on Q or as the expected extra message-length per datum that must be communicated if a code that is optimal for a given (wrong) distribution Q is used, compared to using a code based on the true distribution P.

If $dP = pd\mu$ and $dQ = qd\mu$ are probability measures over a set X, absolutely continuous with respect to a measure μ then the Kullback-Leibler relative entropy from P to Q is defined as

$$K(P, Q) = \int_X p \log \left(\frac{p}{q} \right) d\mu. \tag{4}$$

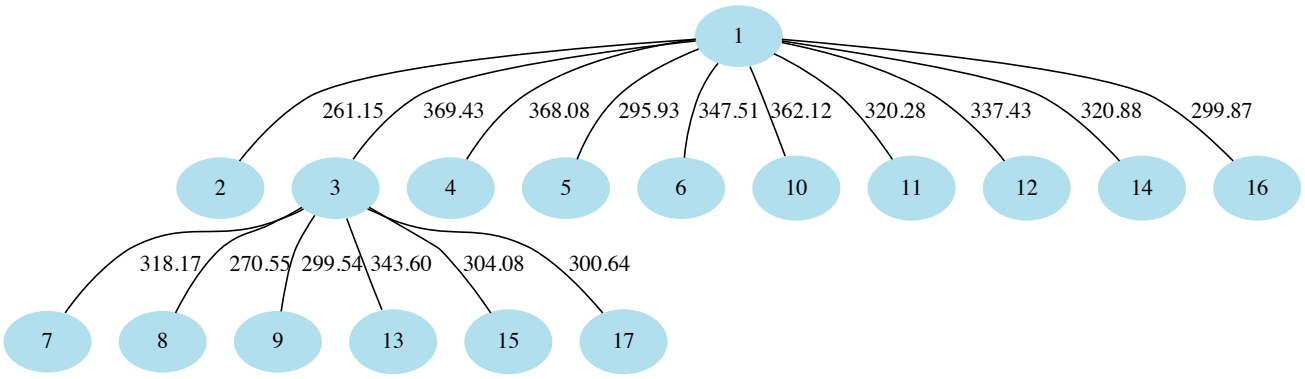


Figure 3. The figure shows the spanning tree with maximum possible total mutual information and includes all the channels for MODIS granule MOD01.A2006.176.0955.005.2006214124831.

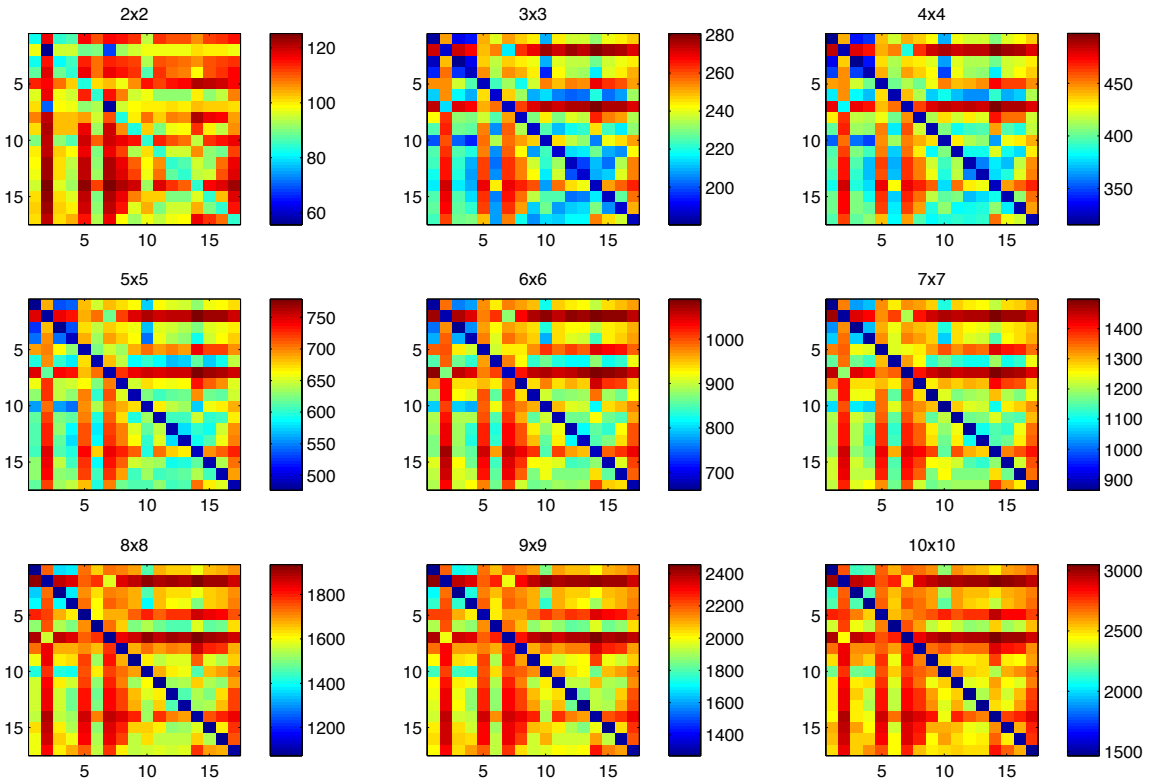


Figure 4. The figure shows the relative entropies for all the 17 channels using distinct windows with sizes ranging from 2×2 to 10×10. We compute relative entropies using windows in order to account for spatial correlations among the channels. As window size increases relative entropy values stabilize.

The K-L relative entropy estimator[3,4] is described here. Given N independent observations $X_1, \ldots, X_N$ distributed with the density p and M observations $Y_1, \ldots, Y_M$ distributed with density q we wish to estimate

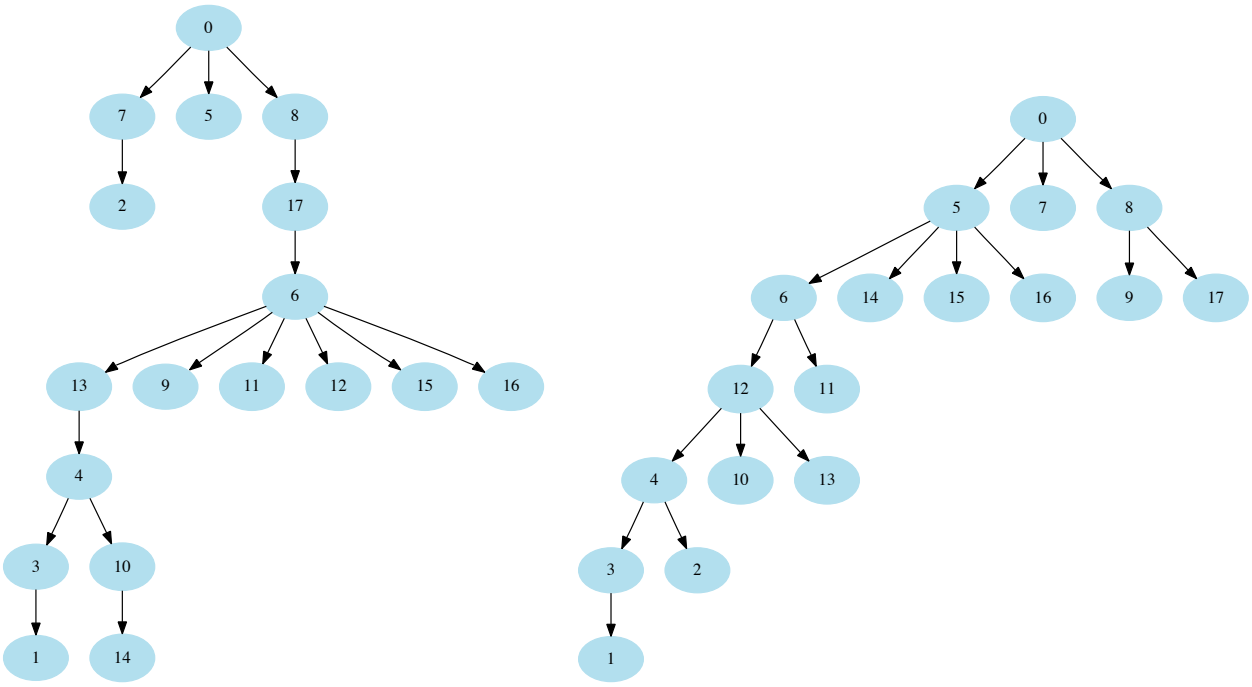
$K(P,Q)$. Given X_i in the sample, $i \in \{1,2,\ldots,N\}$, consider $\rho(X_i, Y_j)$, $j = 1,\ldots,M$, and the order statistics

$$\tilde{\rho}^{(i)}_{1,M} \leq \tilde{\rho}^{(i)}_{2,M} \leq \cdots \leq \tilde{\rho}^{(i)}_{M,M}, \tag{5}$$

so that $\tilde{\rho}^{(i)}_{k,M}$ is the k-th nearest-neighbor distance from X_i to some Y_j, $j \in \{1,\ldots,M\}$. Then an asymtotically unbiased and consistent estimator for relative entropy is given by

$$\hat{K}(P,Q) = m \log \left[\prod_{i=1}^{N} \frac{\tilde{\rho}^{i}_{k,M}}{\rho^{i}_{k,N}} \right]^{1/N} + \log \frac{M}{N-1} \tag{6}$$

We compute the relative entropy matrices for each MODIS granule with varying window sizes. That is, we compute $K(c_i, c_j)$, $\forall i,j \in \{1,2,\ldots,17\}$, where c_i stands for channel i, and for the diagonal elements of the relative entropy matrices we compute the entropies for each channel. Figure 4 shows the relative entropy matrices for a select MODIS granule for varying window sizes ranging from 2×2 to 10×10. Figure 5-(a) shows the minimal spanning tree using Edmond's algorithm[8–10] on the relative entropy matrix. The Edmond's algorithm produces a directed spanning tree such that the sum of all the weights of the edges is the minimum possible and at the same time includes all the channels in the tree at least and at max once.



(a) Relative Entropy MST

(b) Prediction based MST

Figure 5. (a) Minimal spanning tree for the relative entropy matrix with window size 10×10 for MODIS granule MOD01.A2006.176.0955.005.2006214124831. (b) Minimal spanning tree for the cost matrix using prediction based technique for the same MODIS granule. The cost matrix elements represent the memory size of the compressed file after prediction. The root node $'0'$ in both the figures is an artificial node which means that all of node $'0'$'s child nodes are to be self compressed without any prediction.

5. PREDICTION BASED TECHNIQUE

In a previous paper,[1] the authors have developed a histogram-matching based prediction technique for multispectral imager compression purposes where, given a channel (image) A, it is used to predict another channel B.

Channels A's predictive performance is measured by the size of the compressed file, $C_{B|A}$, of channel B. When a given channel A predicts itself, $C_{A|A}$ is assigned the size of the compressed file using the best available 2-d compression software. In this way, when every channel is used to predict every other channel in the collection of channels one gets a cost matrix with the respective compressed file sizes.

The performance of the lossless compression technique for the collection of channels depends on the total compressed file sizes for each channel. This requires finding a prediction sequence in such a way that one could do a lossless decompression and at the same time achieve best compression ratios possible. This problem is the same as that of finding a directed minimal spanning tree of the non-symmetric cost matrix $C_{A|B}$.

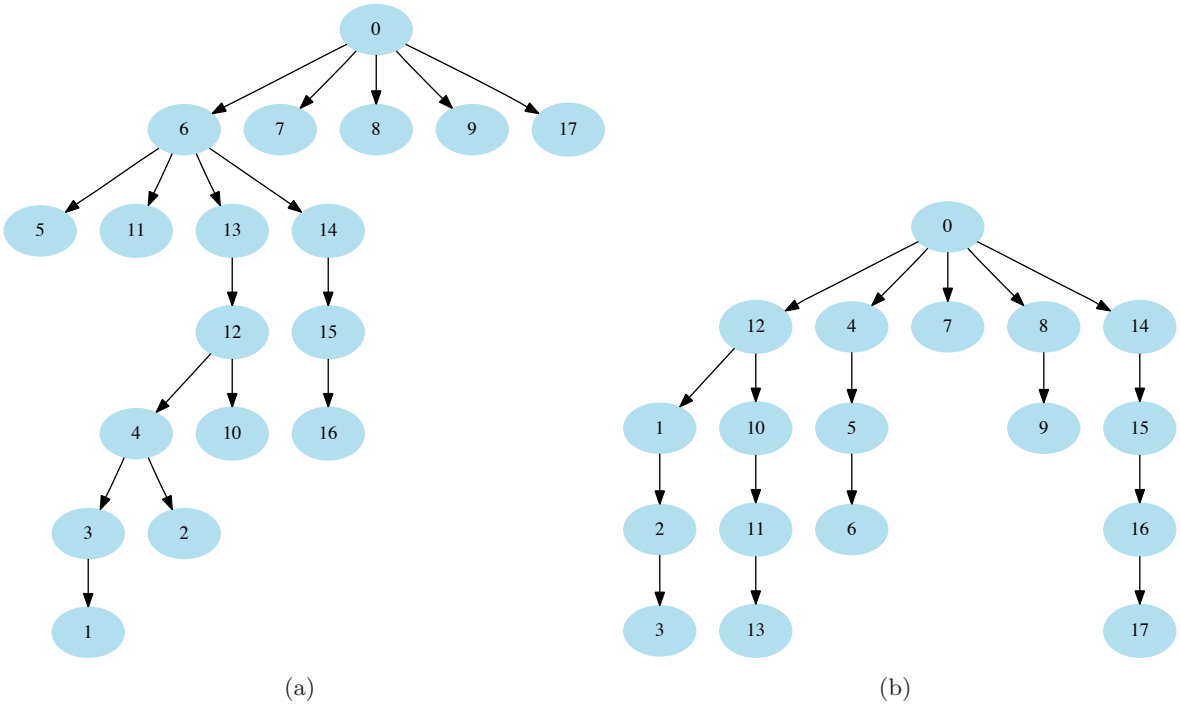
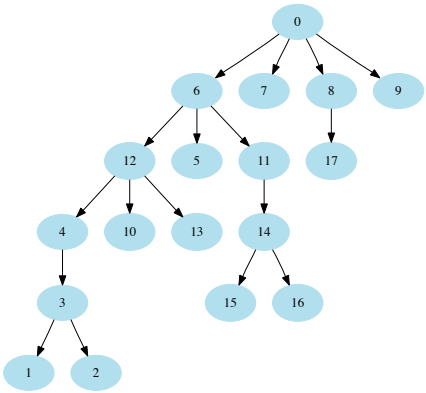
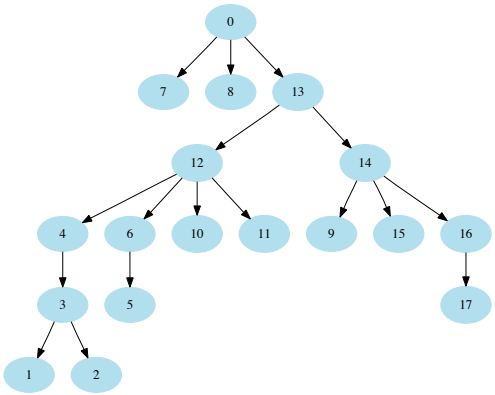


Figure 6. Figure (a) shows the minimal spanning tree for the mean cost matrix of the sample granules shown in Fig. 7. Figure (b) shows the tree which is a result of the physical dependencies among the various spectral channels based upon what they measure. Although the trees are not identical some of the subtrees are identical. For a larger ensemble of MODIS granules, the MST for the mean cost matrix should give us the right physical dependecies among various bands.

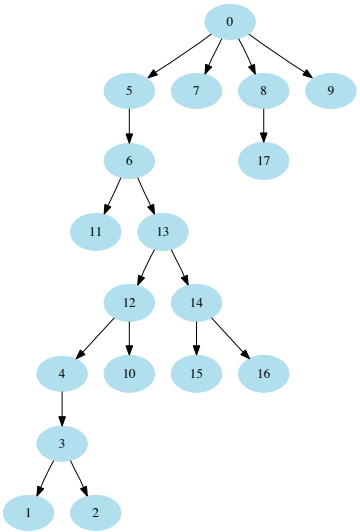
Figure 5-(b) shows the minimal spanning tree for the cost matrix. There are some similarities between the two trees in Figures 5-(a),(b) at the subtree level. The next question that one needs to ask is if the minimum spanning trees for different MODIS granules are the same. We have computed MST's for six more granules as shown in Figure 7 and they are not identical. So we computed the mean cost matrix for all the seven test granules and generated its MST as shown in Figure 6-(a). Figure 6-(b) shows the tree which is a result of the physical dependencies among the various spectral channels based upon what they measure. Although the trees are not identical some of the subtrees are identical. One could only hope that for a larger ensemble of MODIS granules, the MST for the mean cost matrix would give us the right physical dependecies of various bands. We then computed the lossless compression ratios using these MSTs and the results are reported in Table 3. The MSTs improve the compression ratios from the normal linear sequence and also the JPEG2000 ratios. However, one interesting aspect that we have discovered is that the mean MST for seven granules does equally better than the corresponding MSTs for each individual granule. This suggests that we could precompute offline the mean MST for an ensemble of granules and use that for online compression purposes. The physics based tree does better than JPEG2000 but not better than the linear sequence.



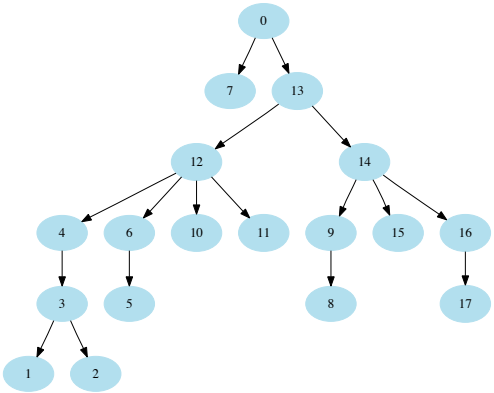
(a) 174.1005.005.2006214124917



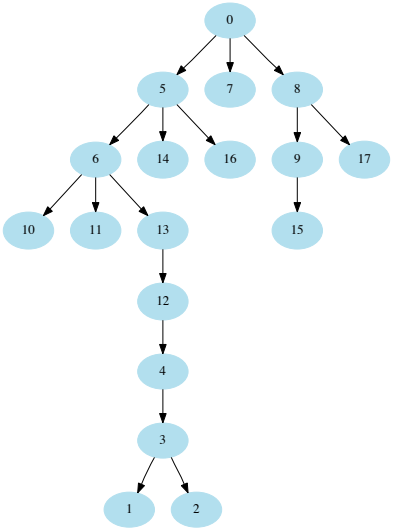
(b) 175.0915.005.2006214125006



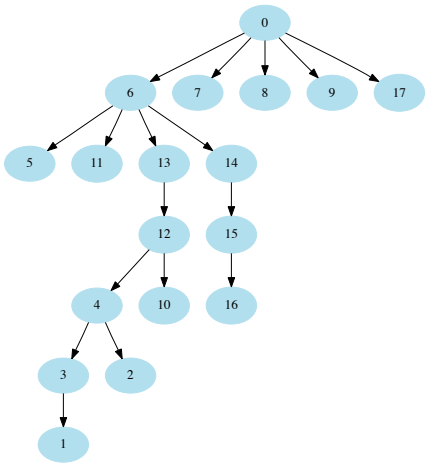
(c) 177.0900.005.2006215131020



(d) 179.0850.005.2006214124349



(e) 180.0930.005.2006214124332



(f) 181.1010.005.2006214124336

Figure 7. For a collection of MODIS granules for different days and times, the figure shows the corresponding directed minimal spanning trees that optimize the lossless compression ratio. Notice that they are not consistent across different granules. However, they are consistent at the subtree level.

Table 3. Compression ratios for MODIS test granules. The prediction scheme used here is based on the histogram-matching technique.[1]

Granule name (MOD01.A2006)	JPEG	Linear sequence	Optimal tree	Mean optimal tree	Physics based tree
174.1005.005.2006214124917	3.2979	3.4623	3.6225	3.6004	3.3307
175.0915.005.2006214125006	3.1696	3.4382	3.6345	3.6068	3.3220
176.0955.005.2006214124831	3.2442	3.4536	3.6716	3.6445	3.3322
177.0900.005.2006215131020	3.1686	3.3934	3.5779	3.5734	3.2723
179.0850.005.2006214124349	2.9442	3.2464	3.4259	3.3833	3.1216
180.0930.005.2006214124332	3.3210	3.5131	3.7174	3.6703	3.3634
181.1010.005.2006214124336	3.1906	3.3754	3.5420	3.5218	3.2590

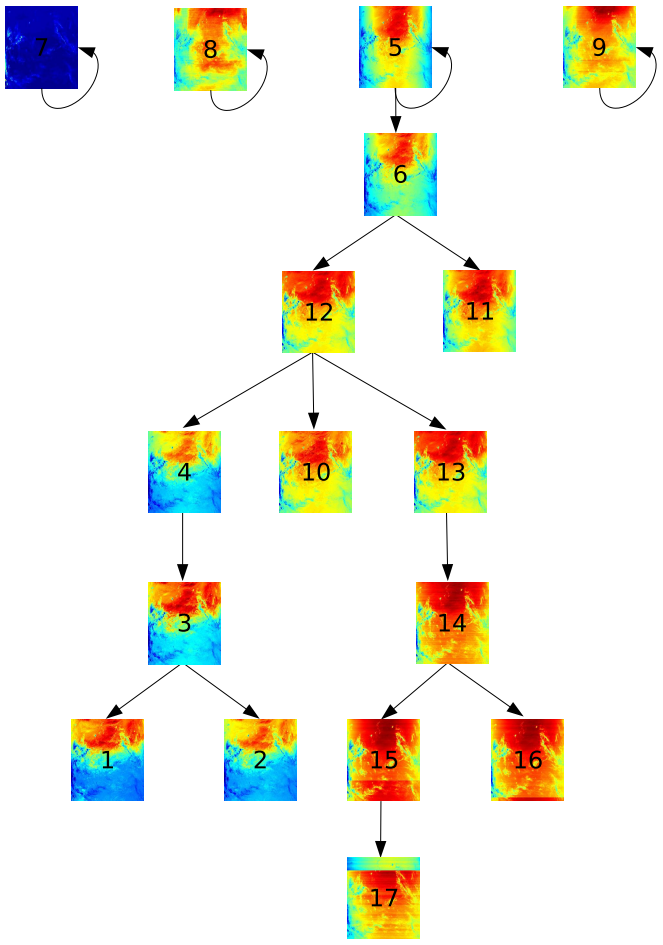


Figure 8. The figure gives a visual representation of the minimal spanning tree for the mean cost matrix of the sample granules shown in Fig. 7. It gives a fair idea of the degree of correlation between the parent and child nodes in the tree. The channels shown here belong to MODIS granule MOD01.A2006.175.0915.005.2006214125006.

6. CONCLUSION

We have presented and analyzed different methods to find band dependencies among the MODIS IR emissive channels. We have used these techniques to find optimal prediction sequence for the purposes of lossless compression. We have found that for lossless compression the optimal predictive sequence depends on the kind of prediction technique used. It is not possible to use the prediction sequence derived from relative entropy method while using a linear prediction method, as the entropy methods can extract nonlinear dependencies among the channels. When using any linear prediction method for predicting one channel from the other, one could use the naive correlation matrix method to find the prediction sequence. Also, we have discovered that the prediction sequences for various granules are not identical but it could be helpful to compute the average prediction sequence offline over an ensemble of granules and implement it in the compression algorithm. This method is recommended as it is time inefficient to compute the cost matrices for finding the prediction sequences. Future research will also focus on devising a real-time computable non-symmetric metric for the purposes of finding band correlations.

7. ACKNOWLEDGMENTS

This compression research is managed by Roger Heymann, PE of OSD NOAA NESDIS Engineering, in collaboration with the NOAA NESDIS STAR Research Office through Mitch Goldberg, Tim Schmit, Walter Wolf.

REFERENCES

1. I. Gladkova, S. Gottipati, and M. Grossberg, "A new lossless compression algorithm for satellite earth science multi-spectral imagers," in *Proceedings of SPIE – Satellite Data Compression, Communications, and Archiving II*, R. W. Heymann, B. Huang, and I. Gladkova, eds., **6683**, Sept 2007.
2. L. F. Kozachenko and N. N. Leonenko, "Sample estimate of the entropy of a random vector," *Problems of Information Transmission* **23**(2), pp. 95–101, 1987.
3. N. N. Leonenko, L. Pronzato, and V. Savani, "A class of renyi information estimators for multidimensional densities," *Annals of Statistics* , 2008.
4. N. N. Leonenko, L. Pronzato, and V. Savani, "Estimation of entropies and divergences via nearest neighbors," *Tatra Mt. Math. Publ.* , 2008.
5. M. Grossberg, S. Gottipati, I. Gladkova, M. Goldberg, and L. Roytman, "An analysis of optimal compression for the advanced baseline imager based on entropy and noise estimation," in *Proceedings of SPIE – Satellite Data Compression, Communications, and Archiving II*, W. Heymann, C. W. Wang, and T. J. Schmit, eds., **6300**, Sept 2006.
6. S. Arya and D. M. Mount, "Approximate nearest neighbor searching," in *Proc. 4th Ann. ACM-SIAM Symposium on Discrete Algorithms (SODA'93)*,
7. S. Arya, D. M. Mount, N. S. Netanyahu, R. Silverman, and A. Y. Wu, "An optimal algorithm for approximate nearest neighbor searching," *Journal of the ACM* , 1998.
8. J. Edmonds, "Optimum branchings," *J. Rsearch of the National Bereau of Standards* , 1967.
9. Y. J. Chu and T. H. Liu, "On the shortest arborescence of a directed graph.," *Science Sinica* , 1965.
10. *Developments in Operations Research*, ch. An algorithm to construct a minimum spanning tree in a directed network. Gordon and Beach, NY, 1971.