

# Hyperspectral anomaly detection and un-mixing in fat-tailed clutter.

M. Bernhardt, J.P. Heather, M.I. Smith.  
Waterfall Solutions Ltd,  
Parallel House, 32 London Road,  
Guildford, Surrey, GU1 2AB.

## Abstract

*It is now established that hyperspectral images of many natural backgrounds have statistics with fat-tails. In spite of this, many of the algorithms that are used to process them appeal to the multivariate Gaussian model. In this paper we consider biologically motivated generative models that might explain observed mixtures of vegetation in natural backgrounds. The degree to which these models match the observed fat-tailed distributions is investigated. Having shown how fat-tailed statistics arise naturally from the generative process, the models are put to work in new anomaly detection and un-mixing algorithms. The performance of these algorithms is compared with more traditional approaches.*

Keywords: Keywords: Hyperspectral, Un-mixing, Anomaly Detection.

## Introduction

There is now a body of work that suggests that many hyperspectral natural backgrounds do not have Gaussian statistics [1-3]. This leads to the question of what distributions to use instead of the Gaussian. The most consistent finding is that the tails of the clutter distributions are fatter than those of a Gaussian. The main approach taken to-date has been to retain the analytically attractive ‘Elliptically Contoured’ (EC) feature of the Gaussian, but to also allow more general radial forms - see [4, 5] for detailed treatment of EC distributions. There has been relatively little work undertaken to test real hyperspectral data to see if it is indeed EC and this is a more difficult question to answer in these high dimensional data-sets. In terms of the tails, a variety of forms have been proposed, including power laws. It is worth noting though that power laws derived, e.g. for the multivariate Student-T distribution, often turn out to be quite high powers. This begs the question as to whether the tails of the data distribution are really power laws or whether they are in fact exponential. It is

known that it is hard to differentiate between exponential tails and high power-law tails in finite data-sets. In this paper we take a slightly different approach and attempt to understand how the generic statistical features observed in hyperspectral vegetation backgrounds come about.

This is done by formulating a number of plausible models for the competition and growth between different vegetation types as they compete for resources (light, space, nutrients etc.). These models are loosely based around work that has been carried out by theoretical biologists in an attempt to explain the evolution of the natural landscape [6-11], but is somewhat simplified for our purposes. Many of these papers adopt a very ‘high level’, taking the statistical mechanical philosophy from physics where one finds that in large collections (of plants in this case) the details of interactions between individuals are washed-out and only key properties like symmetry determine the large-scale observations. Indeed, Self-Organised-Criticality [12] is used to motivate some of the biological models. Following the spirit

of these papers we illustrate a number of models which can be solved and lead to exponentially fat tails. Of note is the fact that the distributions predicted by these models are not EC. An interesting question (not covered here) would be to determine how ‘distinguishable’ they are from EC distributions in these very high dimensional data sets (within the confines of finite data). The distributions arising from these models are then used to develop anomaly detection and un-mixing algorithms. The results of the algorithms on the public-domain AVIRIS data sets [13] are shown.

### Generative models for vegetation backgrounds

The models developed in this section describe the changing abundances of various species of vegetation under some rather stylised competition rules. In order to make a connection to the hyperspectral domain we assume that the standard linear mixture model can be used to convert from abundance to spectra.

In terms of the un-mixing problem the constraints on the abundance coefficients are simply:

$$\begin{aligned} a_i &\geq 0 \\ a_i &\leq 1 \\ \sum_i a_i &= 1 \end{aligned}$$

When there are  $d+1$  end-members, these constraints define a  $d$  dimensional simplex. The maximum entropy distribution over this volume is simply the uniform distribution. For moderate to large values of  $d$  (more so for large values), one expects that the abundance will be concentrated mainly on a few end-members with most end-members having very small values. This is because the neighbourhood of the point  $(1/d, 1/d, \dots, 1/d)$  is very small compared to the volume of the overall simplex.

There is a simple algorithm for generating random numbers that are uniform on a  $d$ -simplex which can be described in pseudo-code terms as:

```

a(1)=Uniform[0,1]
total=a(1)
for i=2 to d
  a(i)=(1-total)*Uniform[0,1]
  total = total + a(i)
next
output RandomPermutation(a)

```

An example of the output of this algorithm is shown below in Figure 1 where the abundance histogram has been plotted:

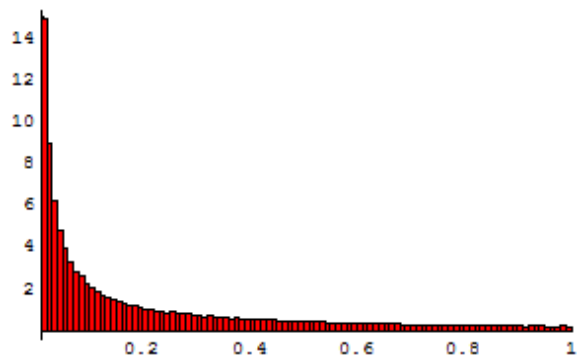


Figure 1: Abundance distribution produced by the maximum entropy model.

The distribution of abundance under this simple maximum entropy model is asymptotically exponential with only a finite number of significant end-members contributing to the mixture.

(2-6)

### A model for competition in vegetation eco-systems

There has been much work in the theoretical biology community studying the behaviour and interaction between species which might be of value in understanding hyperspectral vegetation backgrounds in the context of the linear mixture model [6-11]. Much of this work is fairly complex and does not lend itself easily to the type of analysis required here. To address this we devised a model which is described below.

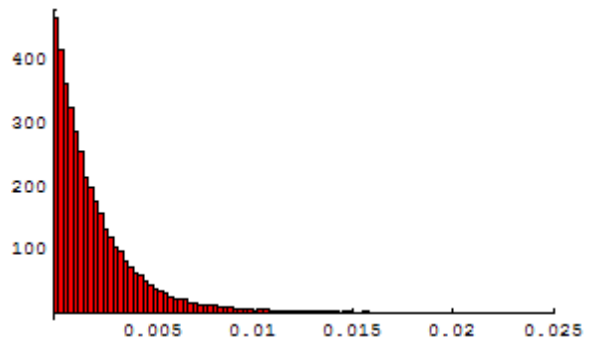
In simple terms, vegetation is subject to growth processes that increase its

abundance, and competition with other vegetation species for resources (water, nutrients, space, light). Lack of resources (and seasonal effects) leads to a reduction in abundance. This dynamic process results in the abundance distributions for vegetation that are observed in nature. These processes therefore have a direct impact on the observed distributions in hyperspectral imagery through the LMM interpretation.

This model, which was originally devised to model wealth re-distribution within an economy [15], attempts to capture some of the processes through which natural vegetation grows and competes for resources. In this model there are a number of species within a pixel. At each time-step two species ( $i$  and  $j$ ) are chosen at random to compete. The proportion of the abundance for the pixel held by the combination of these two species is then redistributed between them at random. Exactly how the re-distribution occurs depends upon a parameter. If this parameter  $\lambda=0$  then the total abundance is uniformly randomly re-distributed, i.e. in theory one species could be totally wiped out. If this parameter is non-zero then it determines a limit on the proportion of each species' abundance that can be captured by the other species in a single iteration. For example, if  $\lambda=0.5$  then only half of the abundance of each of the two individual species is 'up for grabs' in a single iteration. This seems intuitively more realistic, though clearly this model is a grossly simplified caricature of any real biological process. This 'algorithm' can be implemented in terms of the simple set of equations below. Note that these equations conserve the overall abundance between all species (unity).

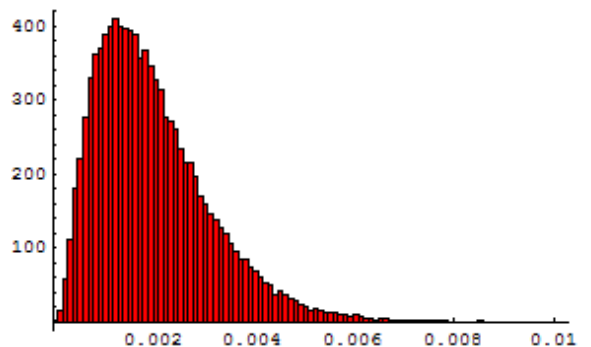
$$\begin{aligned}
 a_i &\rightarrow a_i + \Delta \\
 a_j &\rightarrow a_j - \Delta \\
 \Delta &= (1 - \lambda)(\varepsilon a_j - (1 - \varepsilon)a_i) \\
 \varepsilon &= \text{RandomUniform}(0,1)
 \end{aligned}$$

This model was iterated for a large number of steps and the properties of the limiting distribution were analysed numerically. The graph in Figure 2 shows a plot of the distribution of abundance after running this algorithm with  $\lambda=0$ .



**Figure 2: Abundance distribution for vegetation competition model when  $\lambda=0$ .**

When the value of  $\lambda$  is increased from zero, the distribution changes from an exponential to a Gamma. This is shown in the graph in Figure 3.



**Figure 3: Abundance distribution for vegetation competition model when  $\lambda=0.4$ .**

In [15] it is shown that the model results in a Gamma distribution.

## Anomaly detection

$$\lambda = \frac{\langle x \rangle}{\alpha}$$

The work described in previous sections of this paper seems to indicate that the Gamma distribution (of which the exponential is a special case) is a good model for the abundance of individual materials. This is used as the motivation for developing the following algorithm which (by analogy with Principal Components) we have termed the Gamma Components Analysis (GCA) Algorithm.

The idea is to find the directions in the data along which the projection of the data is most like a Gamma distribution. If the hypothesis that the data should be better described by these statistics than Gaussian is true, then this should lead to a better representation of the data. Also, our approach does not require the data to be EC.

First we address the problem of computing how good a fit to a Gamma distribution the data along a particular projection is, and then we describe how the best directions are found.

Minus the log of the Gamma distribution or the negative log-likelihood is given by:

$$L(x) = \frac{x}{\lambda} + \alpha \log x - \alpha \log \lambda + \log \Gamma(\alpha)$$

The maximum likelihood estimators can be derived from this. The estimator of  $\alpha$  is given by the solution of:

$$-\log \alpha + \Psi(\alpha) - \langle \log x \rangle = 0$$

where the angle braces denote the sample average and the Digamma function has been introduced (the logarithmic derivative of the Gamma function). Thus  $\alpha$  depends solely upon the sample averaged logarithm of the data. The parameter  $\lambda$  can be found from:

This allows the straightforward numerical determination of the parameters of the distribution. The ‘goodness of fit’ can be obtained by summing the log-likelihood over all the data once the parameters have been found. This solves the problem of finding the goodness of a particular direction in the data space, but does not tell us how to find good directions. This is clearly a hard non-linear optimisation problem. Instead of trying to solve this problem using, for example, steepest descent (which is clearly computationally difficult) we opt for an approximate solution. The following methodology has been applied in a variety of data estimation areas, including nonlinear PCA [16, 17] and is a kind of Monte-Carlo approach. The logic of the algorithm is as follows:

1. Choose  $n$  directions at random
2. Pick the best one (on the basis described above)
3. Project out this direction from the data
4. Repeat until enough directions found (data well approximated)
- 5.

Each time  $n$  directions are found and the best one taken, the following statement is true with probability  $p$ :

*The direction chosen is amongst the top  $1 - (1 - p)^{\frac{1}{n}}$  of possible directions.*

As an example to clarify this, imagine we found  $n=100$  directions. What can be asserted with 95% certainty? Answer: the direction chosen is amongst the top 3% of possible directions.

A good way to choose directions along which the data has a large projection is to pick two data points at random and use the line between them as a direction.

## Un-mixing

In this section we attempt to make use of the assumption that abundances are Gamma distributed in an un-supervised un-mixing algorithm. To do this we draw on previous work [18, 19] which used Non-Negative Matrix Factorisation to un-mix spectra.

The hypothesis that the hyperspectral data is a linear mixture of end-member spectra can be written as:

$$V \approx WH$$

The approach chosen here is to find  $W$  and  $H$  together without going through a separate phase of end-member determination. This approach is called non-negative matrix factorisation and has been described in [18]. There are a number of advantages to this approach:

1. The end-members are found automatically from the data.
2. The end-members do not have to be points in the data space (unlike Simplex Volume Maximisation, SVM).
3. It can be extended to include different distributions, i.e. not just Gaussian.

The mixing coefficients are automatically found to be positive (as the name of the algorithm suggests), but are not constrained to be less than one. In addition, the sum to one constraint is not present in the description of [18]. The algorithm has been modified to include these additional constraints.

For the Gaussian noise case the algorithm derived in [18] reduces to the following simple updating procedure:

$$W_{ik} \leftarrow W_{ik} \frac{(VH^T)_{ik}}{(WHH^T)_{ik}}$$

$$W_{ik} \leftarrow W_{ik} \frac{(VH^T)_{ik}}{(WHH^T)_{ik}}$$

Modifying the distribution has the effect of changing these iterative update rules. The arguments of [18] that lead to the update equations above for the Gaussian distribution can be extended for the case of the Gamma distribution. In this case the update equations become:

$$W_{pq} \leftarrow W_{pq} \frac{\sum_j \frac{V_{pj}H_{qj}}{(WH)_{pj}^2}}{\sum_k \frac{H_{qk}}{(WH)_{pk}}}$$

$$H_{pq} \leftarrow H_{pq} \frac{\sum_i \frac{V_{iq}W_{ip}}{(WH)_{iq}^2}}{\sum_k \frac{W_{kp}}{(WH)_{kq}}}$$

It is also possible to modify the algorithm so as to include the sum-to-unity constraint for the abundances.

## Results

Both real and synthetic hyperspectral data have been used for the purpose of evaluating the performance of the proposed anomaly detection and un-mixing algorithms.

The real data was a set of spectral cubes from the well-known NASA AVIRIS sensor. AVIRIS has 224 bands (of varying quality) and covers the 0.5 $\mu$ m to 2.5 $\mu$ m wavelength range. Both the raw data and the atmosphere compensated reflectance data are available. Unfortunately ground truth information (i.e. the physical properties) for the observed scene is not available, making AVIRIS data unsuitable for assessing the performance of the un-mixing algorithms. However, it has been used to trade-off the proposed gamma

components anomaly detection algorithm against the baseline RX algorithm [20, 21].

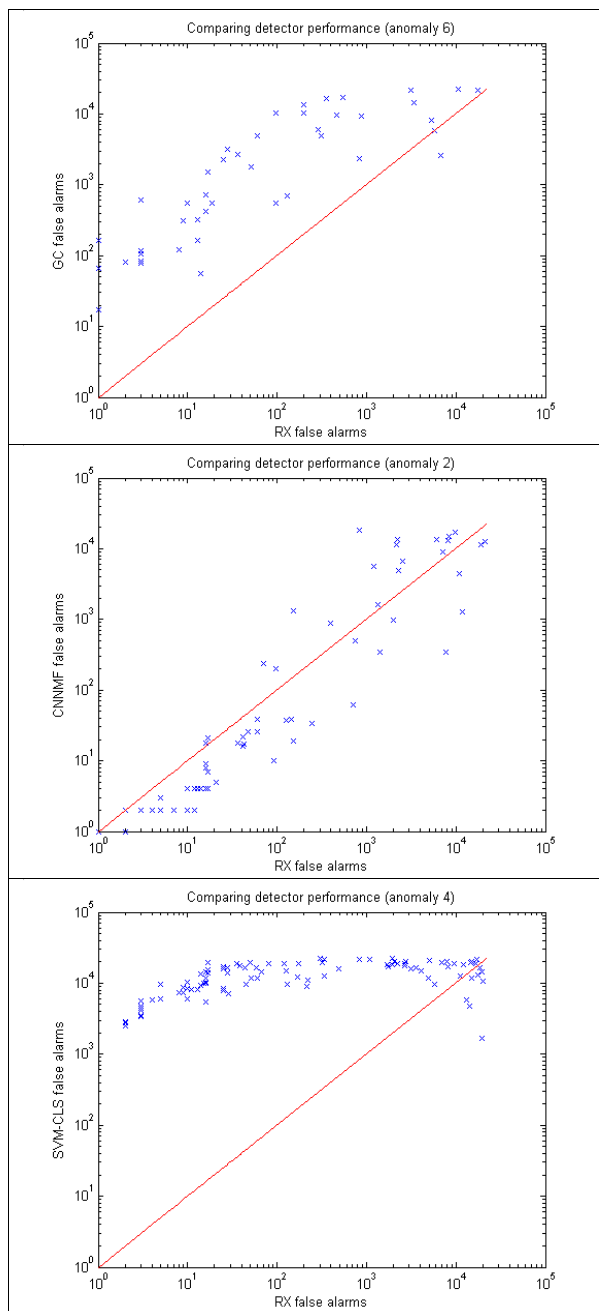
In order to do this regions of natural vegetation were selected from the AVIRIS scene and used as the 'background' for a series of cubes containing embedded anomalies. This ensured that the backgrounds were indeed natural, since the cubes contain both vegetation and urban areas. The anomalies were selected randomly from the urban region of the same AVIRIS scene and embedded at random (known) locations at a range of mixing levels. These embedded cubes then formed the test set for evaluating the anomaly detection algorithms.

The synthetic data was generated by using spectra from a hyperspectral library in the short and mid-wave infra-red. Spectra were made by combining together known end-members in known (random) abundance proportions. The final cubes were modified by removing pixels containing more than a certain abundance fraction of a single material. In other words, we have control over the maximum purity of any material present in the synthetic data. This is very useful since a baseline algorithm (like SVM below) may do very well indeed if pure pixels are present, but its performance will degrade in the more realistic scenario where the purity of the purest pixels is less than 100%.

As a baseline unsupervised un-mixing algorithm we chose SVM to find end-members, followed by fully-constrained-least-squares. As a baseline anomaly detection algorithm we chose the well-known RX algorithm [20, 21].

The experiments can be divided into anomaly detection and un-mixing sets. The anomaly detection experiments were conducted on the AVIRIS data as described above, with anomalies embedded. The anomalies were pixels from an urban area

embedded in the vegetation region, with various weights ranging from nearly zero up to unity. The anomaly detectors that were tested were: the baseline RX algorithm; the Gamma Components algorithm with the Non-Negative Matrix Factorisation (NNMF) un-mixing algorithm; and the baseline un-mixing algorithm. It is always possible to convert an un-mixing algorithm into an anomaly detector by using the un-mixing error as a measure of 'anomalousness'. In order to directly compare the algorithms we considered each anomaly in turn. For each anomaly the algorithm threshold was adjusted such that the anomaly was just detectable. At this threshold setting the number of false alarms was noted. This was repeated over all anomalies and algorithms. We then plotted (for each anomaly) the false alarms due to one algorithm against those of another. These plots are shown in Figure 4. On such a plot two identical algorithms would produce a line a 45 degrees. If the points lie above this line then the algorithm on the horizontal axis is better than the one on the vertical axis. As there were a very large number of such plots we have only reproduced a sample here.



**Figure 4 False alarm plots for Gamma Components (GC), Constrained Non-Negative Matrix Factorisation (CNNMF) and Simplex Volume Maximisation + Constrained Least Squares (SVMCLS) based anomaly detectors versus RX.**

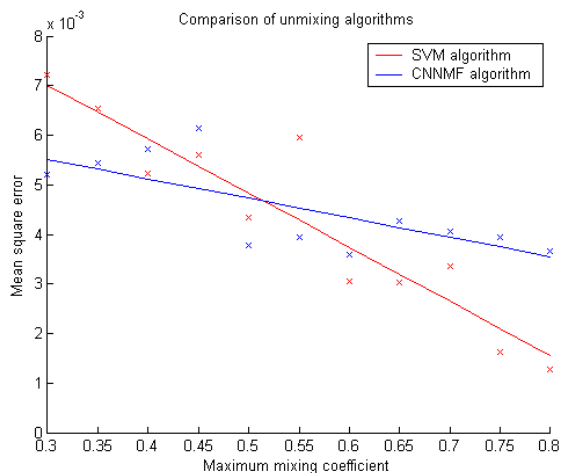
Although we only show example plots, they are typical of all results achieved. The results show that both the Gamma Components algorithm and the Non-

Negative Matrix Factorisation algorithm (run as an anomaly detector) perform as well as RX, but with different false alarms. Interestingly, the baseline un-mixing algorithm SCMCLS performs very poorly as an anomaly detector.

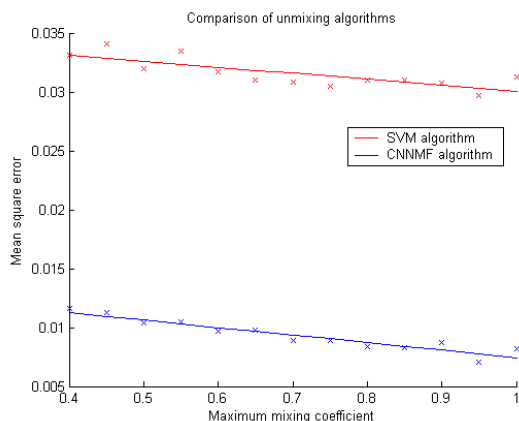
In order to perform a quantitative comparison for the un-mixing algorithms, the two un-mixing algorithms were executed on the synthetic data using a range of maximum mixture coefficients (see Section 6.1). In each case the end-members calculated by each algorithm were compared with the library spectra that were used to generate the synthetic test data. In order to make a quantitative assessment of performance, the end-members output by each algorithm were matched with the library spectra in such a way as to minimise the overall mean square difference between each pair (optimal matching). This overall mean square error was then plotted for a range of maximum mixture coefficients. In other words, multiple synthetic data sets were generated with different allowed maximum purities and each un-mixed. The performance on each was assessed as the mean square error of the optimal assignment of the end-members chosen by the algorithm to the true end-members.

The quantitative results in Figure 5 indicate that the baseline SVM algorithm works very well when large mixture coefficients are present (some nearly pure materials present). This is to be expected given the close relationship between the SVM algorithm and the linear mixture model which was used to generate the synthetic test data. However, when the mixture coefficients are reduced the SVM performance deteriorates rapidly. When maximum mixture coefficient is less than 0.5 the CNNMF un-mixing algorithm outperforms the SVM approach. This is a significant result because in real hyperspectral data it is very unlikely that pure pixels will be present for more than a

handful of the materials present in the scene. Figure 6 shows what happens when the number of end-members is increased to eight. In this case the SVM severely underperforms the CNNMF algorithm. Note that each point on the plot represents a complete un-mixing of a cube (where the highest purity within the cube is plotted on the horizontal axis).



**Figure 5 Comparison of the un-mixing performance of the two algorithms as the maximum mixing coefficient is varied when there are five true end-members. Lines of best fit have been plotted.**



**Figure 6 Comparison of the un-mixing performance of the two algorithms as the maximum mixing coefficient is varied when there are eight true end-members. Lines of best fit have been plotted.**

### Conclusion

The results for anomaly detection are very interesting. Both of the novel anomaly

detectors perform as well as the RX algorithm. RX is a good benchmark algorithm which is expected to do well if the data is EC. The fact that both the GC and CNNMF algorithms are able to achieve similar performance, although they are based upon different principles (and hence are complementary), suggests that there is room for improvement in anomaly detection through a hybrid (or fused) approach. Initial analysis shows that the three algorithms tend to produce different false alarms and so would benefit from fusion.

If the data was not EC then it would be expected that RX would not perform well, and our algorithms better (since they have no EC assumption).

The CNNMF un-mixing algorithm has been demonstrated to be robust to impure pixels and outperforms the baseline algorithm as the mixtures become more dilute. When the number of true end-members is increased the both algorithms find it harder to identify the true end-members, but the performance of CNNMF is substantially better than the baseline algorithm over the whole range of dilutions considered.

## References

1. D.G. Manolakis, D. Marden, J.P. Kerekes, G.A. Shaw, Statistics of hyperspectral imaging data, Proc. SPIE vol. 4381, pp308-316, Aug. 2001.
2. D. Marden, D. Manolakis, Algorithms for hyperspectral imaging data exploitation using non-Gaussian elliptically contoured distributions, in Imaging Spectroscopy VIII, S.S. Shen Ed., Seattle, WA, 2002 SPIE.
3. D. Manolakis, G. Shaw, Detection algorithms for hyperspectral imaging applications, IEEE Signal Proc. Magazine, Jan. 2002, pp29-43.
4. M. Rangaswamy, D. Weiner, A. Ozturk. Non-Gaussian random vector identification using spherically invariant random processes, IEEE Transactions on Aerospace and Electronic systems. Vol. 29 no. 1 pp111-124 Jan. 1993.
5. M. Rangaswamy, D. Weiner, A. Ozturk. Computer generation of correlated non-Gaussian radar clutter. IEEE Transactions on Aerospace and Electronic systems. vol. 31 no. 1 pp106-116, Jan. 1995.
6. J. B. Wilson et. Al. Are there assembly rules for plant species abundance? An investigation in relation to soil resources and successional trends. Journal of Ecology, Vol. 84, No. 4 pp527-538, Aug. 1996.
7. J. Bolliger, J.C. Sprott, D.M. Mladenoff. Self-organisation and complexity in historical landscape patterns. OIKOS 100:541-553, 2003.
8. R.V. Solé, S.C. Manrubia, Are rainforests self-organised into a critical state? J. Theor. Bio. No. 173, pp31-40, 1995.
9. D. Storch, K.J. Gaston, J. Cepák, Pink landscapes: 1/f spectra of spatial environmental variability and bird community composition. Proc. Royal Soc. Lond. B No. 269 pp1791-1796, 2002.
10. G. Caldarelli, A. Giacometti, A. Maritan, Randomly pinned landscape evolution. Physical Review E Vol.55 No. 5 ppR4865-R4868.
11. R.V. Solé, D. Alonso, A. McKane. Self-organised instability in complex ecosystems. Phil. Trans. Royal Soc. Lond. B. No. 357, pp667-681, 2002.
12. P. Bak, K. Chen, Self organised criticality, Scientific American. Jan. 1991, pp26-33.
13. <http://aviris.jpl.nasa.gov/>
14. Many papers by E T Jaynes cover this, but see for an interesting example; E. T. Jaynes, Where do we Stand on Maximum Entropy? in The Maximum Entropy Formalism, R. D. Levine and M. Tribus (eds.), M. I. T. Press, Cambridge, MA, p. 15, 1979; or alternatively: E.T. Jaynes, Monkeys, Kangaroos and N,' in Maximum-Entropy and Bayesian Methods in Applied Statistics, J. H. Justice (ed.), Cambridge Univ. Press, Cambridge, p. 26, 1986.
15. M. Patriarca, K. Kaski, A. Chakrabort. A Statistical model with a standard Gamma distribution. Pre-print available on-line from xxx.lanl.gov e-print server: arXiv:cond-mat/0402200v1, 6 February 2004
16. A.J. Smola, P.L. Bartlett, Sparse greedy Gaussian process regression, Proc Neural and Information Processing Systems (NIPS) 2000, pp619-625.

## Acknowledgements

The work reported in this paper was funded by the Electro-Magnetic Remote Sensing (EMRS) Defence Technology Centre, established by the UK Ministry of Defence and run by a consortium of SELEX Sensors and Airborne Systems, Thales Defence, Roke Manor Research and Filtronic.

