REGULARISATION IN MULTI- AND HYPERSPECTRAL REMOTE SENSING CHANGE DETECTION

Allan Aasbjerg Nielsen

Technical University of Denmark
Informatics and Mathematical Modelling
Building 321, DK-2800 Kgs. Lyngby, Denmark
phone +45 4525 3425, fax +45 4588 1397
e-mail aa@imm.dtu.dk, www.imm.dtu.dk/∼aa

Abstract

Change detection methods for multi- and hypervariate data look for differences in data acquired over the same area at different points in time. These differences may be due to noise or differences in (atmospheric etc.) conditions at the two acquisition time points. To prevent a change detection method from detecting uninteresting change due to noise or arbitrary spurious differences the application of regularisation also known as penalisation is considered to be important. Two types of regularisation in change detected by the multivariate alteration detection (MAD) transformation are considered: 1) ridge regression type and smoothing operators applied to the estimated weights in the MAD transform; and 2) pre-processing (before applying the MAD transformation) by noise reducing orthogonal transformations where the number of retained transformed variables can be considered a regularisation parameter. Regularisation by the former methods smooth the weights given to the individual bands in the MAD transformation and thus it penalises weights that fluctuate wildly as a function of wavelength; regularisation by the latter methods tends to smooth in the image domain. Also, regularisation may be necessary to prevent numerical instability especially when working on hyperspectral data.

Key words: regularisation, hyperspectral data, canonical correlation analysis, multivariate alteration detection (MAD) transformation.

1. Introduction

Change detection methods for multi- and hyperspectral data ideally find differences in data acquired over the same geographical region at different points in time. These differences may be due to not only actual change on the ground but also to noise or differences in (atmospheric etc.) conditions at the two acquisition time points. To prevent a change detection method from detecting uninteresting change due to noise or arbitrary spurious differences the application of regularisation also known as penalisation is considered to be important. In this paper two types of regularisation in change detected by the canonical correlation analysis based multivariate alteration detection transformation are considered. Regularisation may be necessary to prevent numerical instability especially when working on hyperspectral data.

2. Regularisation by smoothing the weights

In ordinary least squares (OLS) regression [17]

$$y = X\theta + e \tag{1}$$

(y is $n \times 1$, X is $n \times p$ and θ is $p \times 1$, where n is the number of observations and p is the number of parameters) we solve for θ by minimising $e^T e$ which leads to the normal equations

$$(X^T X)\hat{\theta} = X^T y \tag{2}$$

or (formally)

$$\hat{\theta} = (X^T X)^{-1} X^T y. \tag{3}$$

To avoid possible (near) singularity problems in $X^T X$ we may minimise $e^T e + \frac{k}{2}\theta^T\theta$ instead. This leads to

$$(X^T X + kI)\hat{\theta} = X^T y \tag{4}$$

where I is the $p \times p$ unit matrix. We see that by doing this we punish or penalise [20, 21, 16] high values of the elements of θ. In other words with increasing k the elements of θ tend to become closer to zero. k can be chosen (stipulated) or estimated from the data by cross-validation. This type of regression is termed ridge regression [8].

More generally we may penalise other characteristics of θ than size by minimising

$$e^T e + \frac{k}{2}(L\theta)^T(L\theta) \tag{5}$$

where L is some matrix. This leads to

$$(X^T X + k\Omega)\hat{\theta} = X^T y \tag{6}$$

with $\Omega = L^T L$. In the above simple case we have $\Omega = I$ $(= L = L_0 = L_0^T L_0)$.

Say instead we wanted to force all elements of θ to be equal. This can be done by setting $L_1\theta = 0$ where L_1 is $(p-1) \times p$ with

$$L_1 = \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & -1 \end{bmatrix} \tag{7}$$

leading to the desired

$$L_1\theta = \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & -1 \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_p \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \tag{8}$$

i.e., $\theta_1 = \theta_2,\ \theta_2 = \theta_3, \ldots, \theta_{p-1} = \theta_p$.

Rather than forcing the elements of θ to be equal we may want to penalise curvature in the elements of θ. For this to make sense some ordering of the elements of θ is assumed; in remote sensing this ordering could be by wavelength. The desired minimum curvature can be achieved by setting $L_2\theta = 0$ where L_2 is $(p-2) \times p$ with

$$L_2 = \begin{bmatrix} 1 & -2 & 1 & 0 & \cdots & 0 \\ 0 & 1 & -2 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -2 & 1 \end{bmatrix} \tag{9}$$

leading to $p \times p$

$$\Omega = L_2^T L_2 = \begin{bmatrix} 1 & -2 & 1 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -2 & 5 & -4 & 1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & -4 & 6 & -4 & 1 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -4 & 6 & -4 & 1 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 1 & -4 & 6 & -4 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 1 & -4 & 6 & -4 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 1 & -4 & 5 & -2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 1 & -2 & 1 \end{bmatrix} \tag{10}$$

which is penta-diagonal.

Of course we can combine these different ways of penalising the elements of θ to obtain a desired structure or desired characteristics of the solution by using

$$\Omega = w_0 L_0^T L_0 + w_1 L_1^T L_1 + w_2 L_2^T L_2 + \cdots. \tag{11}$$

2.1 Canonical correlation analysis

In a canonical correlation analysis (CCA) [7, 4, 1] based change detection scheme termed multivariate alteration detection (MAD) [9, 12, 13, 10, 14, 3, 2, 15, 11] for geometrically co-registered $p \times 1$ data X from one point in time and $q \times 1$ data Y from another point in time we solve the eigenproblem

$$\begin{bmatrix} 0 & \Sigma_{12} \\ \Sigma_{21} & 0 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \rho \begin{bmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} \tag{12}$$

or

$$\begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = (\rho + 1) \begin{bmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} \tag{13}$$

to obtain the desired change detector. Σ_{11} is the variance-covariance matrix of X, Σ_{22} is the variance-covariance matrix of Y and Σ_{12} is the covariance matrix between the two, $\Sigma_{21} = \Sigma_{12}^T$. a is the eigenvector containing the weights with which to multiply X from the one point in time and b is the eigenvector containing the weights with which to multiply Y from the other point in time. To do change detection we form the canonical variates $U = a^T X$ and $V = b^T Y$ and the MAD change detector as the difference $U - V$ between them. More well known expressions for the CCA problem are the coupled eigenproblems

$$\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\, a = \rho^2 \Sigma_{11}\, a \tag{14}$$
$$\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\, b = \rho^2 \Sigma_{22}\, b. \tag{15}$$

If we wish to apply regularisation in this case we could solve

$$\begin{bmatrix} 0 & \Sigma_{12} \\ \Sigma_{21} & 0 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \rho \begin{bmatrix} \Sigma_{11} + k_1\Omega & 0 \\ 0 & \Sigma_{22} + k_2\Omega \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} \tag{16}$$

where k_1 and k_2 determine the amount of regularisation.

3. Regularisation by orthogonal transformation pre-processing

The change detected by the MAD method is invariant to separate linear (affine) transformations in the originally measured variables such as

1. changes in gain and offset in the measuring device used to acquire the data;
2. data normalisation or calibration schemes that are linear (affine) in the gray values of the original variables; or
3. orthogonal or other affine transformations such as principal component (PC) [6] or maximum autocorrelation factor (MAF) transformations [18, 19, 5].

This characteristic can be utilised in a regularisation-type reduction of redundancy in hyperspectral data by means of orthogonal transformations of the data at the two points in time separately before change detection by the MAD method.

4. Results

To illustrate the two regularisation techniques we use 126 channel HyMap data covering a small agricultutal area in Waging-Taching in Bavaria, Germany. The original data are shown in Figures 2 to 3.

To illustrate the effect of regularisation by smoothing the weights, i.e., the eigenvectors in the CCA Figure 1 shows eigenvectors corresponding to the leading canonical variates from the HyMap data. Both non-regularised and regularised eigenvectors are shown. In the regularised case we have applied $k_1 = k_2 = 0.001$ as an example.

To illustrate the effect of regularisation by orthogonal transformations Figures 4 to 8 show principal components (PCs), maximum autocorrelation factors (MAFs) and MAD variates based on 40 MAFs. It is obvious that the MAFs offer a much better representation of the joint signal in all the original spectral bands than do the PCs, see also [15, 11].

5. Conclusions

Two types of regularisation in multi- and hypervariate change detection are described and applied to hyperspectral image data. The first type of regularisation is based on ridge regression type and smoothing operators applied to the estimated weights in the change detection transformation. The second type of regularisation is based on pre-processing (before applying the MAD transformation) by noise reducing orthogonal transformations where the number of retained transformed variables can be considered a regularisation parameter.

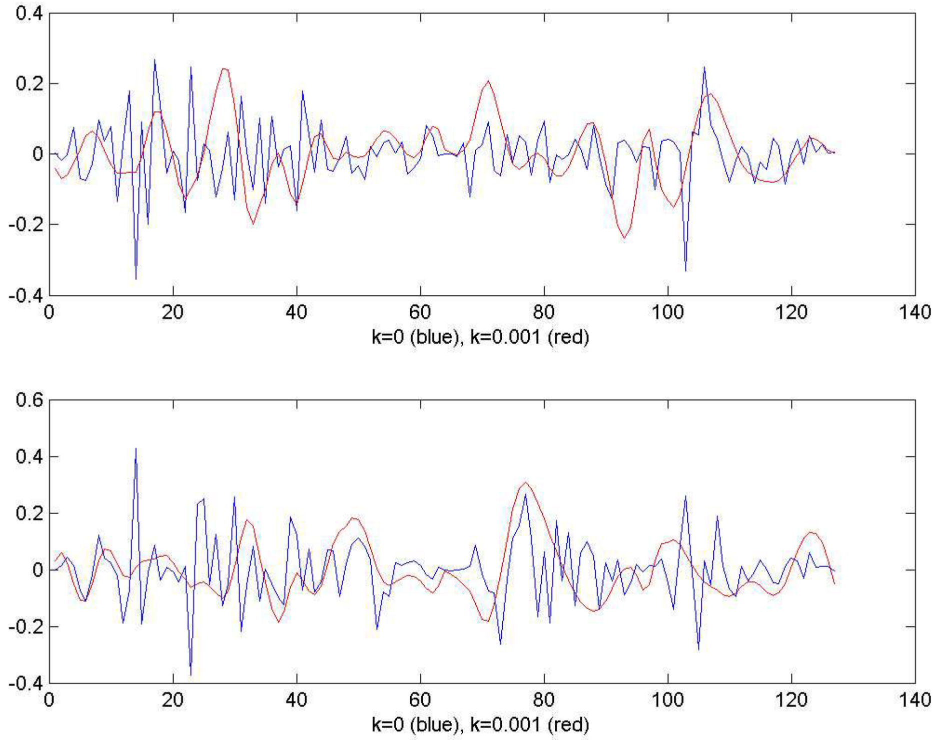

Figure 1: Eigenvectors with and without regularisation ($k_1 = k_2 = 0.001$) for an example with 126 channel HyMap data.

Acknowledgement

The author wishes to thank Andreas Müller, DLR, Oberpfaffenhofen, Germany, for the geometrically and radiometrically calibrated HyMap data.

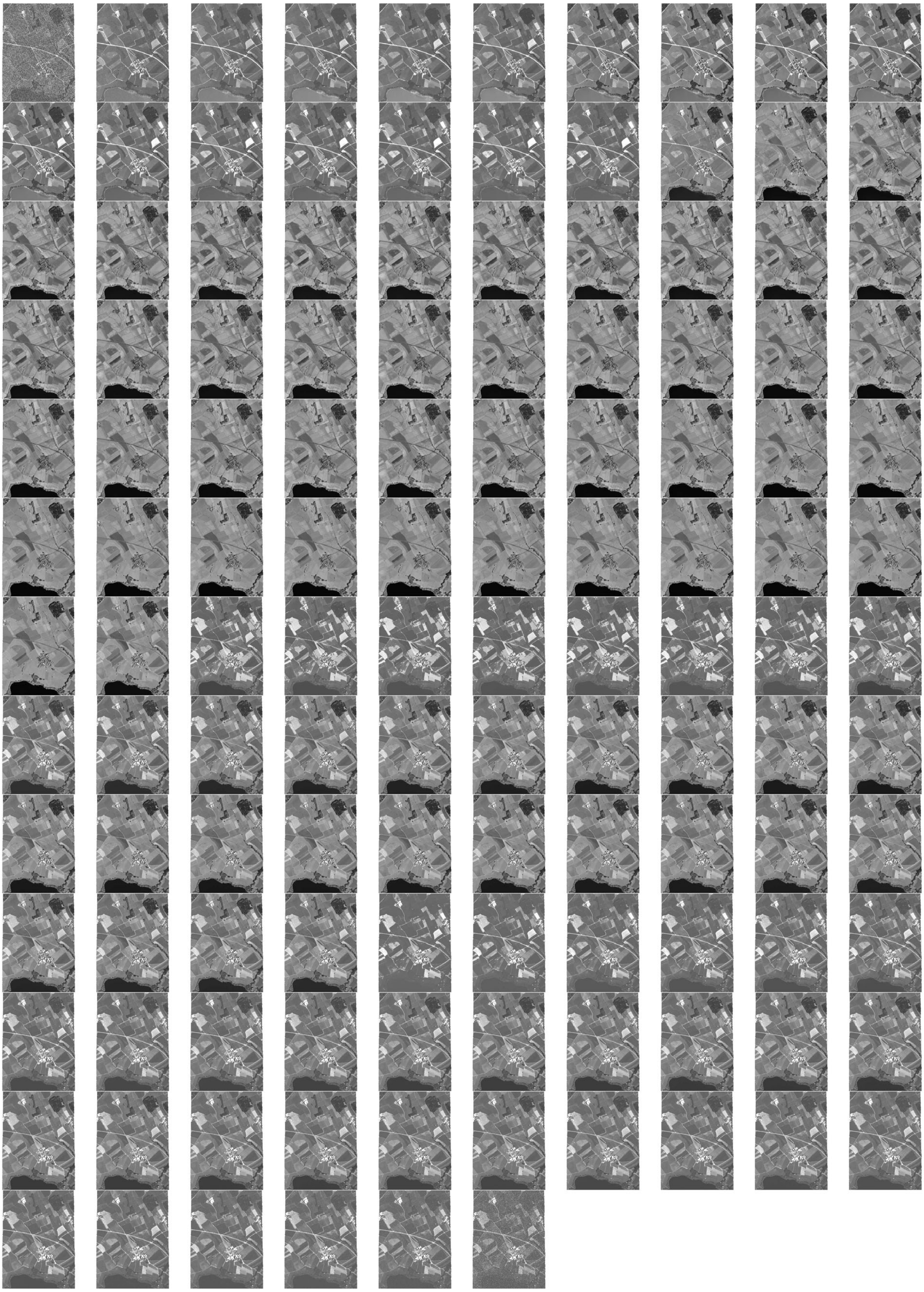

Figure 2: HyMap data from 30 June 2003 at 8:43 UTC covering an agricultural region near Lake Waging-Taching in Germay, spectral bands 1 to 126 row-wise.

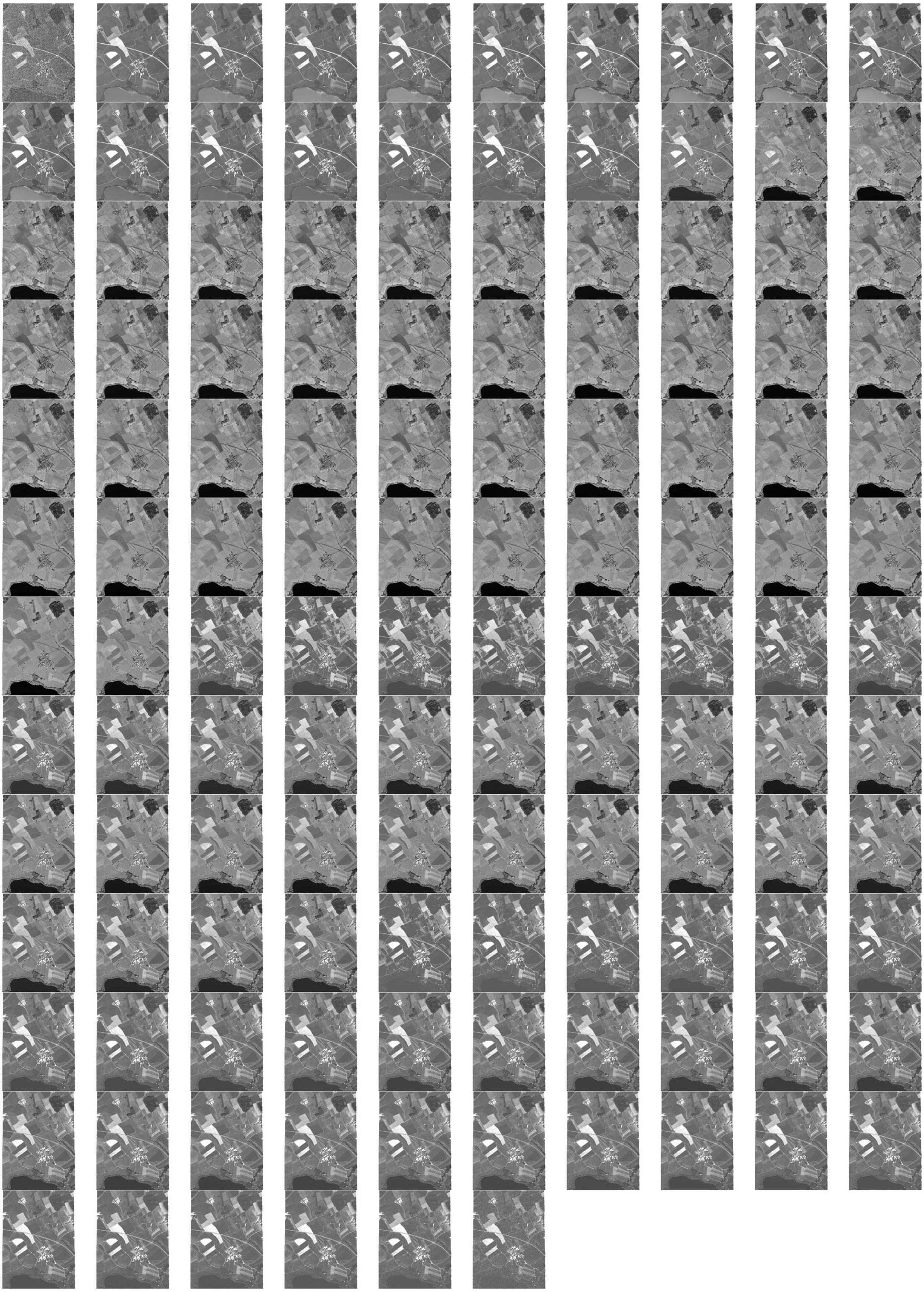

Figure 3: HyMap data from 4 August 2003 at 10:23 UTC covering an agricultural region near Lake Waging-Taching in Germay, spectral bands 1 to 126 row-wise.

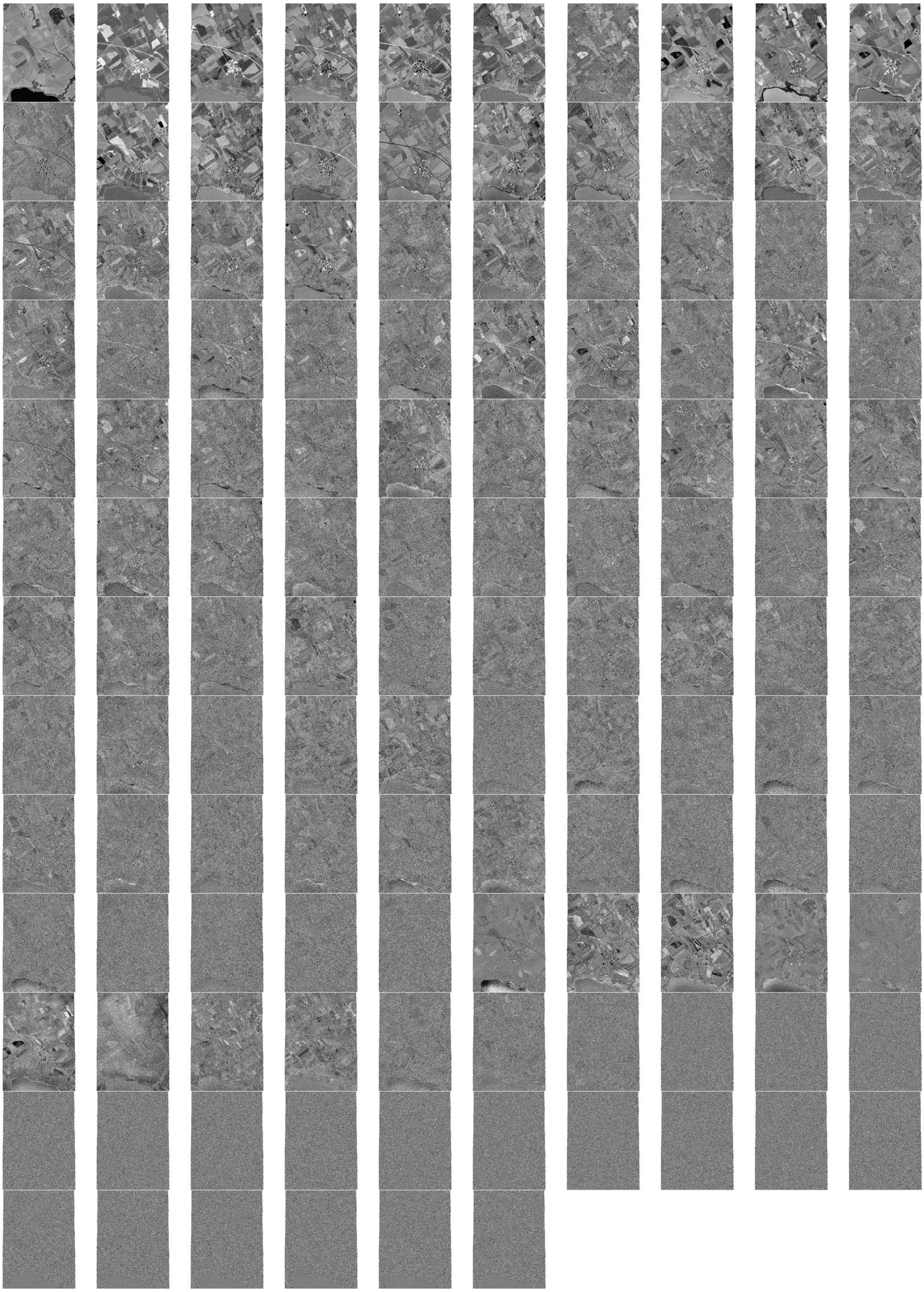

Figure 4: HyMap data from 30 June 2003 at 8:43 UTC covering an agricultural region near Lake Waging-Taching in Germay, principal components 1 to 126 row-wise.

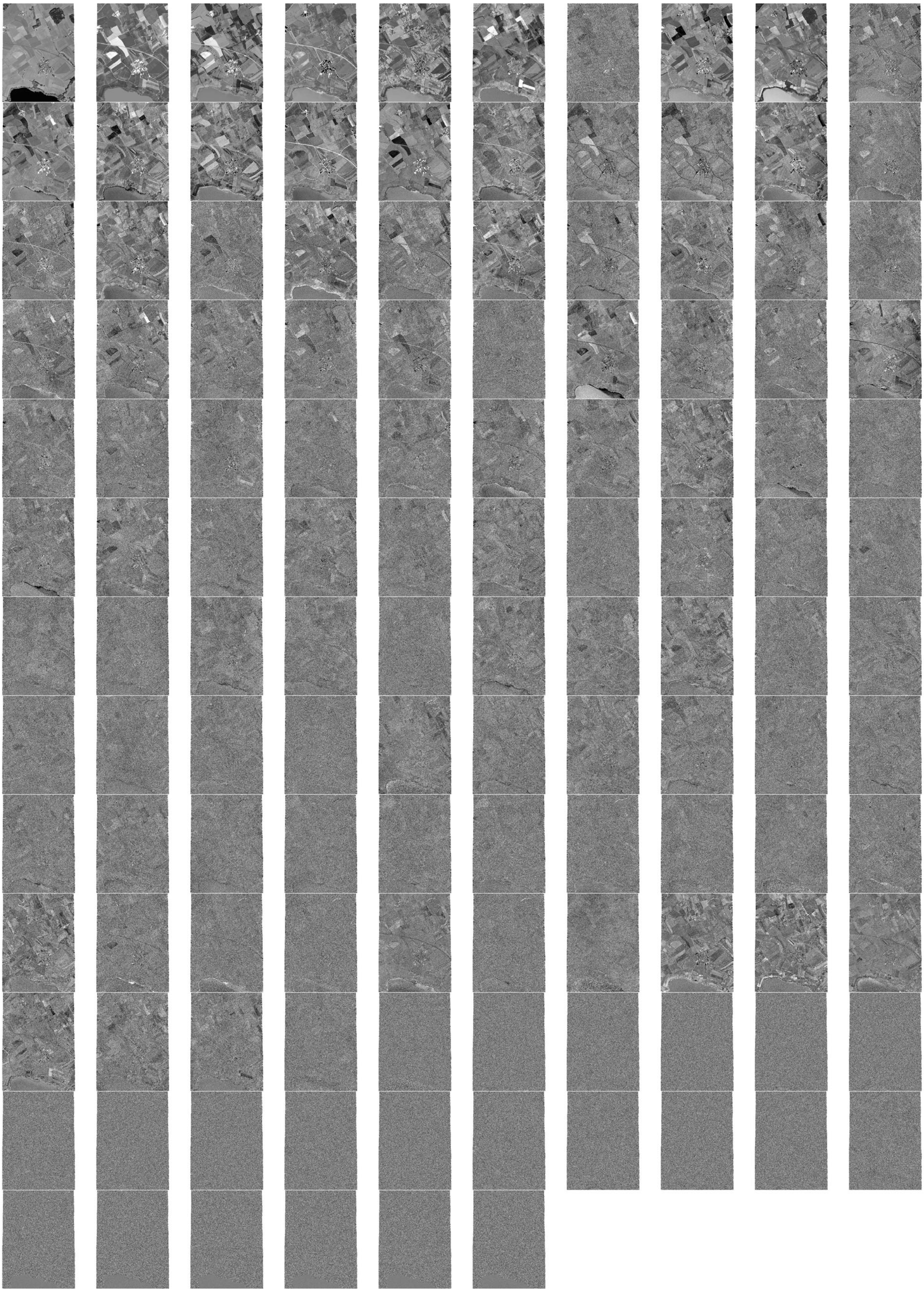

Figure 5: HyMap data from 4 August 2003 at 10:23 UTC covering an agricultural region near Lake Waging-Taching in Germay, principal components 1 to 126 row-wise.

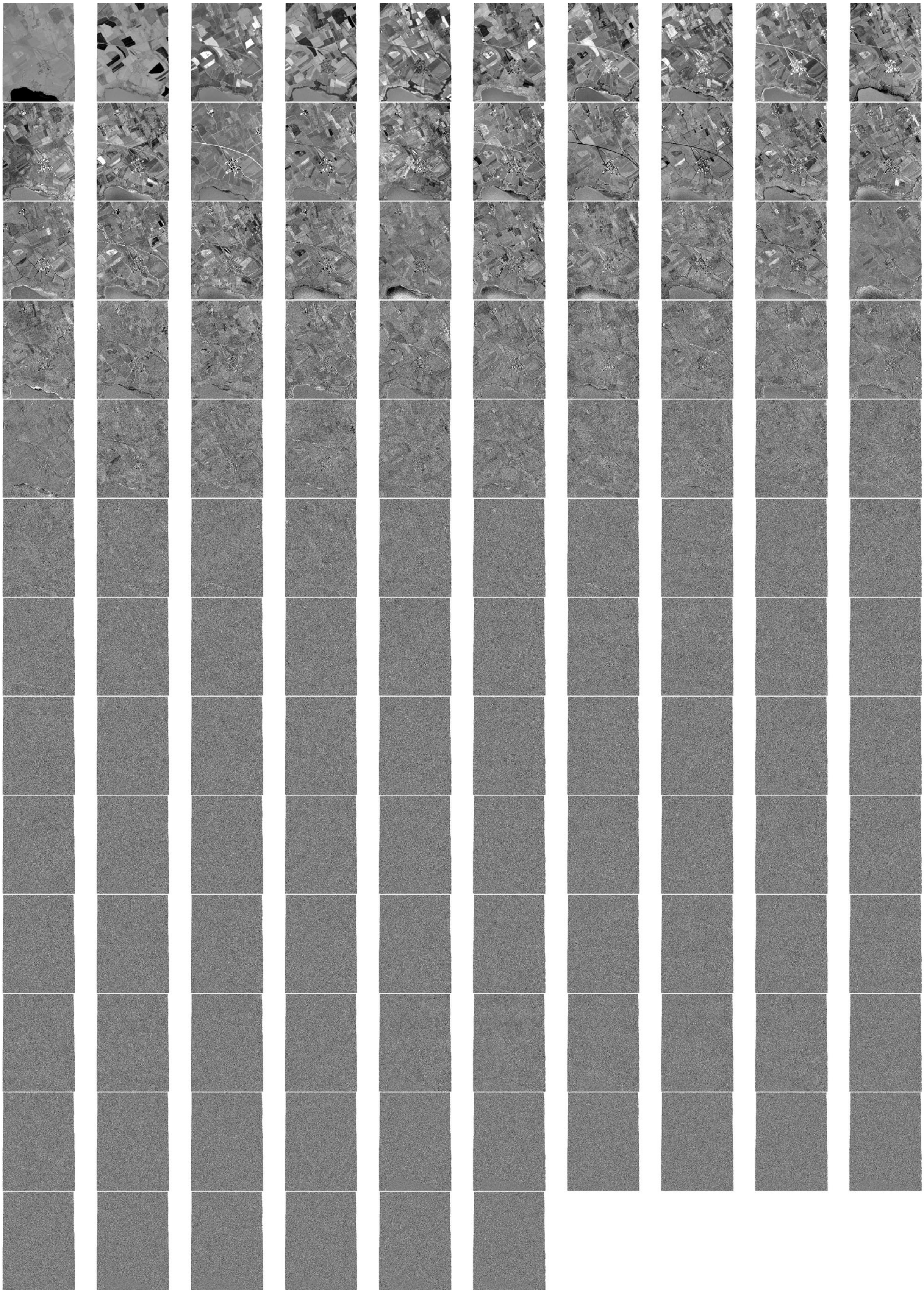

Figure 6: HyMap data from 30 June 2003 at 8:43 UTC covering an agricultural region near Lake Waging-Taching in Germay, maximum autocorrelation factors 1 to 126 row-wise.

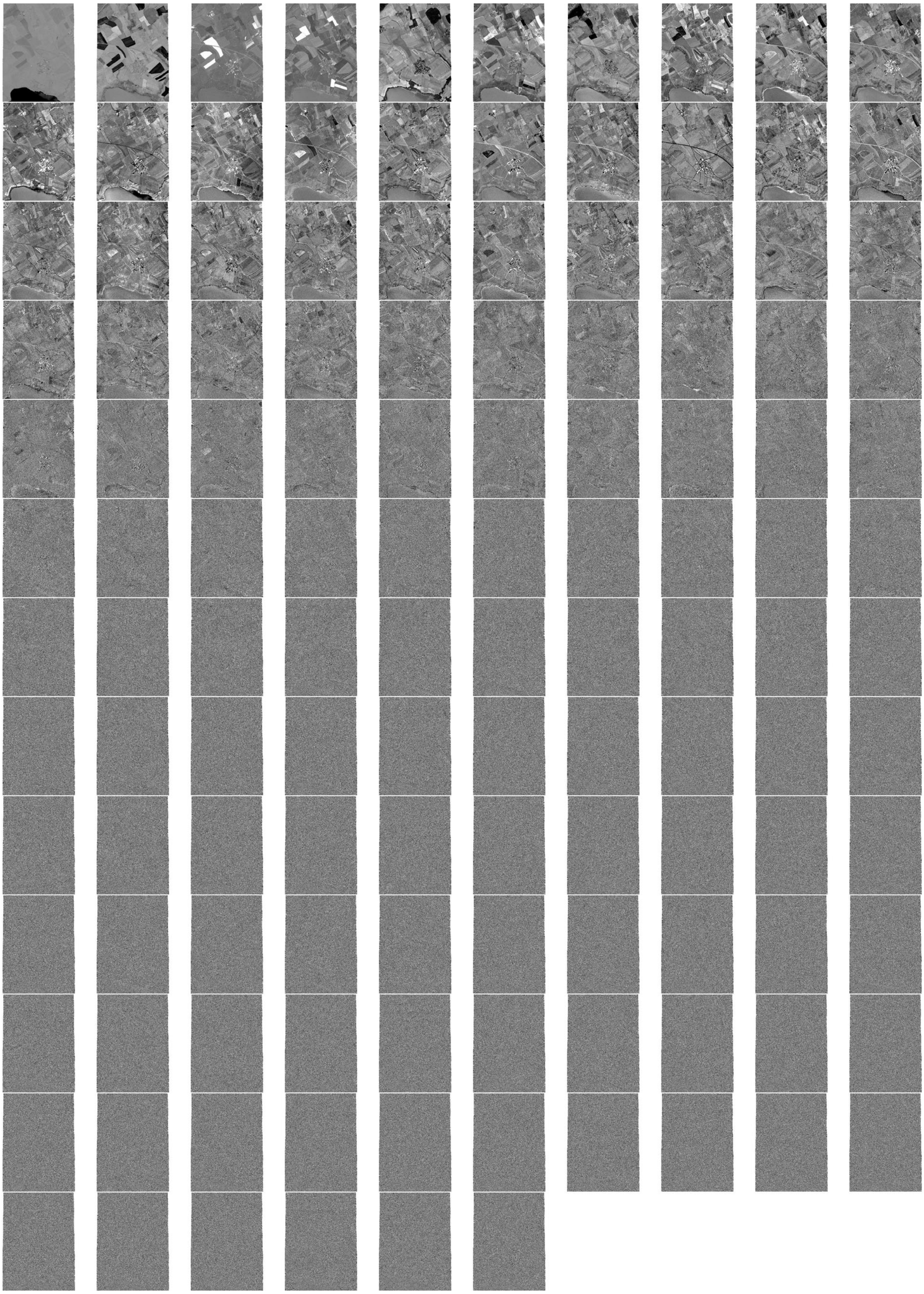

Figure 7: HyMap data from 4 August 2003 at 10:23 UTC covering an agricultural region near Lake Waging-Taching in Germay, maximum autocorrelation factors 1 to 126 row-wise.

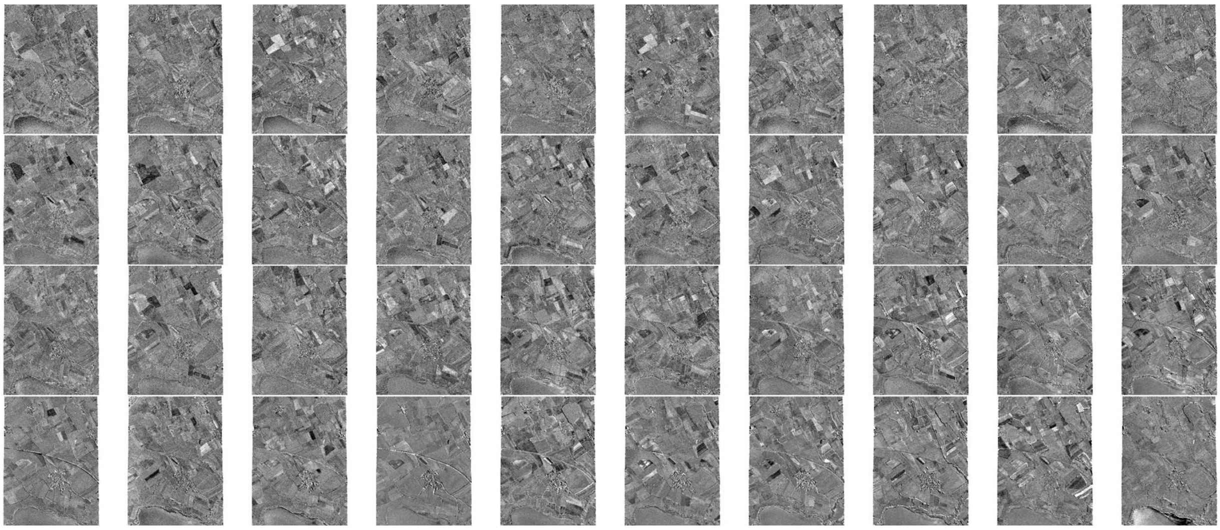

Figure 8: MAD variates 1-40 row-wise, based on 40 MAFs.

References

[1] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, Wiley, third edition, 2003.

[2] M. J. Canty and A. A. Nielsen, “Unsupervised classification of changes in multispectral satellite imagery,” in *Proceedings of the 11th SPIE International Symposium on Remote Sensing*, Maspalomas, Gran Canaria, Spain, 13-16 September 2004, Internet http://www.imm.dtu.dk/∼aa/SPIE2004.pdf.

[3] M. J. Canty, A. A. Nielsen, and M. Schmidt, “Automatic radiometric normalization of multitemporal satellite imagery,” *Remote Sensing of Environment*, vol. 91, pp. 441–451, 2004.

[4] W. W. Cooley and P. R. Lohnes, *Multivariate Data Analysis*, John Wiley and Sons, New York, 1971.

[5] A. A. Green, M. Berman, P. Switzer, and M. D. Craig, “A transformation for ordering multispectral data in terms of image quality with implications for noise removal,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 26, no. 1, pp. 65–74, Jan. 1988.

[6] H. Hotelling, “Analysis of a complex of statistical variables into principal components,” *Journal of Educational Psychology*, vol. 24, pp. 417–441, 498–520, 1933.

[7] H. Hotelling, “Relations between two sets of variates,” *Biometrika*, vol. XXVIII, pp. 321–377, 1936.

[8] A. E. Hoerl and R. W. Kennard, “Ridge regression. Biased estimation for nonorthogonal problems,” *Technometrics*, vol. 12, no. 1, pp. 55–67, 1970.

[9] A. A. Nielsen, *Analysis of Regularly and Irregularly Sampled Spatial, Multivariate, and Multi-temporal Data*, Ph.D. thesis, Informatics and Mathematical Modelling, Technical University of Denmark, Lyngby, 1994, Internet http://www.imm.dtu.dk/∼aa/NielsenPhD.pdf.

[10] A. A. Nielsen, “Multi-channel remote sensing data and orthogonal transformations for change detection,” in *Machine Vision and Advanced Image Processing in Remote Sensing*, I. Kanellopoulos, G. G. Wilkinson, and T. Moons, Eds. Springer, 1999.

[11] A. A. Nielsen, “Iteratively reweighted multivariate alteration detection in multi- and hyperspectral data,” submitted, 2004.

[12] A. A. Nielsen and K. Conradsen, “Multivariate alteration detection (MAD) in multispectral, bi-temporal image data: A new approach to change detection studies,” Tech. Rep. 1997-11, Informatics and Mathematical Modelling, Technical University of Denmark, 1997, Internet http://www.imm.dtu.dk/∼aa/tech-rep-1997-11.

[13] A. A. Nielsen, K. Conradsen, and J. J. Simpson, “Multivariate alteration detection (MAD) and MAF post-processing in multispectral, bi-temporal image data: New approaches to change detection studies,” *Remote Sensing of Environment*, vol. 64, pp. 1–19, 1998.

[14] A. A. Nielsen and A. Müller, “Change detection by the MAD method in hyperspectral image data,” in *Proceedings of the 3rd EARSeL Workshop on Imaging Spectroscopy*, 2003, pp. 115–116, Internet http://www.imm.dtu.dk/∼aa/NielsenMuellerEARSeL2003.pdf.

[15] A. A. Nielsen, A. Müller, and W. Dorigo, “Hyperspectral data, change detection and the MAD transformation,” in *Proceedings of the 12th Australasian Remote Sensing and Photogrammetry Association Conference*, 2004, Internet http://www.imm.dtu.dk/∼aa/Australasian2004.pdf.

[16] J. Ramsay and B. Silverman, *Functional Data Analysis*, Springer, 1997.

[17] J. A. Rice, *Mathematical Statistics and Data Analysis*, Duxbury Press/Wadsworth, second edition, 1995.

[18] P. Switzer and A. A. Green, “Min/max autocorrelation factors for multivariate spatial imagery,” Tech. Rep. 6, Stanford University, 1984.

[19] P. Switzer and S. E. Ingebritsen, “Ordering of time-difference data from multispectral imagery,” *Remote Sensing of Environment*, vol. 20, pp. 85–94, 1986.

[20] H. D. Vinod, “Canonical ridge and econometrics of joint production,” *Journal of Econometrics*, vol. 4, pp. 147–1662, 1976.

[21] B. Yu, I. M. Ostland, P. Gong, and R. Pu, “Penalized discriminant analysis of *in situ* hyperspectral data for conifer species recognition,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 5, pp. 2569–2577, 1999.