# A KERNEL VERSION OF MULTIVARIATE ALTERATION DETECTION

*Allan A. Nielsen*[a,b] *and Jacob S. Vestergaard*[b]

Technical University of Denmark
[a]DTU Space – National Space Institute
[b]DTU Compute – Applied Mathematics and Computer Science
DK-2800 Kgs. Lyngby, Denmark

## ABSTRACT

Based on the established methods kernel canonical correlation analysis and multivariate alteration detection we introduce a kernel version of multivariate alteration detection. A case study with SPOT HRV data shows that the kMAD variates focus on extreme change observations.

## 1. INTRODUCTION

Based on a kernel extension of Hotelling's original, linear canonical correlation analysis (CCA) [1, 2, 3] and multivariate alteration detection (MAD) [4, 5, 6, 7], a kernel version of the MAD transformation termed kMAD is introduced here. Previously kernel versions of principal component analysis (PCA) [8], maximum autocorrelation factor (MAF) and minimum noise fraction (MNF) analyses [9, 10, 11] were described. The kMAD method is based on kCCA. In this context the following issues are important

- the choice of parameter(s) in the applied kernel,
- the choice of regularization parameters – the inherent Q-mode formulation of kCCA and kMAD makes regularization a must and not an option (as it is with, say, linear CCA, kPCA and kMNF),
- choice of the number of canonical variates (CVs) or MAD variates (MADs) with relevant change information,
- whether we should focus on CVs (or MADs) associated with high or low canonical correlations, $\rho$, and
- the pixels used in the so-called training data (to comply with the idea in the iMAD method these should be the no-change pixels).

Some of these issues are left for further work.

kCCA is based on Q-mode formulation of CCA which means that in principle we have as many CVs as we have training samples, i.e., potentially thousands. Therefore we use scrambling and scree plots [12] to determine the number of change relevant kCVs (or kMADs).

In linear MAD we argue that the change relevant MADs are the ones associated with low canonical correlations, $\rho$. This is because these MAD variates contain maximum variance (variance is the traditionally used measure of high dynamics, in this case change). Here, with the potentially very large number of MAD variates, we argue that the change relevant kMADs are associated with high $\rho$.

Good general references to kernel methods are [13, 14, 15], the latter with a remote sensing focus.

AAN's e-mail address is alan@dtu.dk and his homepage is http://www.compute.dtu.dk/∼alan.

## 2. THE KERNEL CCA TRANSFORMATION

Regularization in CCA was first suggested in [16]. In a regularized version of kernel CCA we solve the eigensystem

$$\begin{bmatrix} 0 & K_xK_y \\ K_yK_x & 0 \end{bmatrix} \begin{bmatrix} c \\ d \end{bmatrix} = R \begin{bmatrix} (1-\lambda_x)K_x^2 + \lambda_x K_x & 0 \\ 0 & (1-\lambda_y)K_y^2 + \lambda_y K_y \end{bmatrix} \begin{bmatrix} c \\ d \end{bmatrix} \tag{1}$$

with $c^T[(1-\lambda_x)K_x^2+\lambda_x K_x]c = d^T[(1-\lambda_y)K_y^2+\lambda_y K_y]d = 1$ and $\rho = c^T K_x K_y d/\sqrt{c^T K_x^2 c\, d^T K_y^2 d}$. Here $K_x$ and $K_y$ are kernelized, centered versions of the data, $c$ and $d$ are the eigenvectors sought, $R$ is the eigenvalue, and $\rho$ is the canonical correlation. $\lambda_x$ and $\lambda_y$ are the regularization parameters the choice of which is an issue. Normally we would choose (positive) $\lambda_x = \lambda_y$ and either simply stipulate a (small) value, choose a value that makes the terms in the right-hand-side matrix elements approximately equal, or we would set up a cross-validation scheme to choose a value that is optimal in some sense.

### 2.1. Implementation Issues

The above symmetric general eigenvalue problem is $2n$ by $2n$ and we may re-write to two coupled $n$ by $n$ equations

$$K_x K_y d = R[(1-\lambda_x)K_x^2 + \lambda_x K_x]c \quad (2)$$
$$K_y K_x c = R[(1-\lambda_y)K_y^2 + \lambda_y K_y]d \quad (3)$$

(which shows that $R = c^T K_x K_y d = d^T K_y K_x c$).

Equations 2 and 3 may be written

$$K_x K_y[(1-\lambda_y)K_y^2 + \lambda_y K_y]^- K_y K_x c = R^2[(1-\lambda_x)K_x^2 + \lambda_x K_x]c \quad (4)$$
$$K_y K_x[(1-\lambda_x)K_x^2 + \lambda_x K_x]^- K_x K_y d = R^2[(1-\lambda_y)K_y^2 + \lambda_y K_y]d \quad (5)$$

where $^-$ denotes the inverse or if needed the Moore-Penrose inverse and $K_y K_x = K_y^T K_x^T = (K_x K_y)^T$. As shown in the appendix of [10] which concentrates on the case where the right-hand-side is not full rank we may solve one of these two $n$ by $n$ symmetric generalized eigenvalue problems for example Equation 4 to obtain eigenvectors $c$ and eigenvalues $R^2$ and the other by insertion into Equation 3 to obtain eigenvectors $d$.

The training data used when the number of observations is high may either be selected subjectively, by iteratively trying to home in on no-change observations (much like the iMAD method), by random selection or by replacing actual observations with a high number of cluster centers from a k-means analysis, [17]. Here we simply use a random sample.

## 3. THE KERNEL MAD TRANSFORMATION

The MAD variates are the differences between corresponding pairs of canonical variates, [4, 5, 6, 7]. In linear MAD we argue that the change relevant MADs are the ones associated with low canonical correlations, $\rho$. This is because these MAD variates contain maximum variance. Here, with a potentially very large number of kMAD variates, we argue that the change relevant kMADs are associated with high $\rho$. This is based on a combination of the observation that intuitively differences between very similar variates are interesting for change detection and experiments which show that very high order MAD variates appear very noisy and without information. To determine the number of kMADs with relevant change information we use scrambling, [12] pp. 537–538. A so-called scree plot of $\rho$ with and without scrambling will give a good indication of the desired number of kMADs associated with high values of $\rho$ to choose, see the next section.

### 3.1. Choice of kernel and regularization parameters

The kernel width $\sigma$ can be chosen simply as the mean or the median of distances between observations in the original feature space, here we use the mean. As mentioned in the introduction regularization is here a must. The regularization parameters are chosen as small positive numbers, here we use $\lambda_x = \lambda_y = 10^{-6}$.

Alternatively, the parameters may be found by cross-validation.

## 4. CASE STUDY

In this case 1,000 training samples and a Gaussian kernel with scale parameter equal to the mean distance between the observations in the relevant feature space are used. To establish which number of high canonical correlation associated kCVs to retain and to find the standard deviations in the scree plots below we use 100 randomly generated sets of training observations for time point two, all to go with the same randomly generated set of training observations for time point one.

The data used here are SPOT HRV data from 5 February 1987 and 27 January 1989 covering large pineapple fields (to the north and east, in bright red), small scale coffe plantations (to the north and west, in darker red), and the town of Thika (to the south) in the Kiambu district north of Nairobi in Kenya, see Figure 1. The images are 512 rows by 512 columns, 20.0 m pixels, and three spectral bands.

The scree plot in Figure 2 shows that with the 1,000 training samples chosen for the analysis we may expect relevant change signal in the first approximately 70 kMADs.

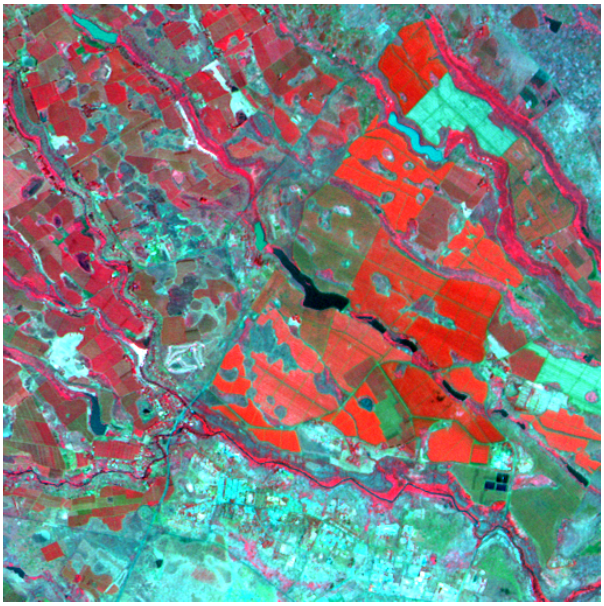
Figure 3 shows the three iMADs and the first three kMADs as RGB. All change variates are stretched between mean minus and mean plus six standard deviations of the variables in question. This leaves no-change regions in gray and change areas in saturated colours (including black and white, if present). We see that 1) the iMADs show change in water regions, in Thika town, and in the large pineapple fields where some fields go from bare soil to healthy pineapple and vice versa, that 2) the kMADs focus on the extreme change observations here associated primarily with water and the town, and that 3) the kernel variates show a much better discrimination between change and no-change compared with the linear variates.

Compared with iMAD analysis, kMAD analysis gives a strong discrimination between change and no-change regions with a conspicuously better suppression of the no-change background.

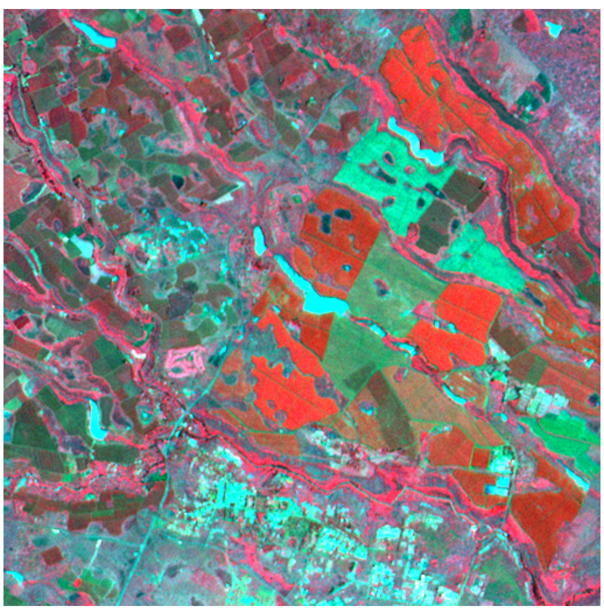
## 5. REFERENCES


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

[2] P. L. Lai and C. Fyfe, “Kernel and nonlinear canonical correlation analysis,” *International Journal of Neural Systems*, vol. 10, no. 5, pp. 365–377, 2000.



(a) 5 February 1987.



(b) 27 January 1989.

**Fig. 1**. SPOT HRV bands 3, 2 and 1 as RGB.

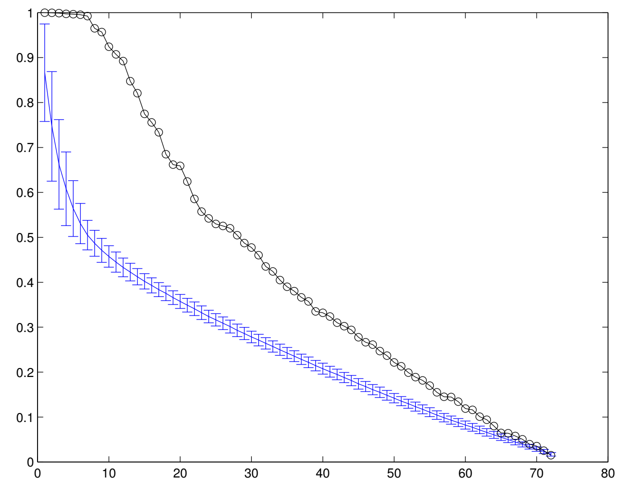



**Fig. 2**. Scree plot of canonical correlations for SPOT HRV data from Thika, Kenya. Scrambled data are shown in the error bar curve (the error bars show the standard deviation for the 100 outcomes of the 1,000 randomly selected training pixels for the second time point).


[3] F. R. Bach and M. I. Jordan, "Kernel independet component analysis," *Journal of Machine Learning Research*, vol. 3, pp. 1–48, 2002.

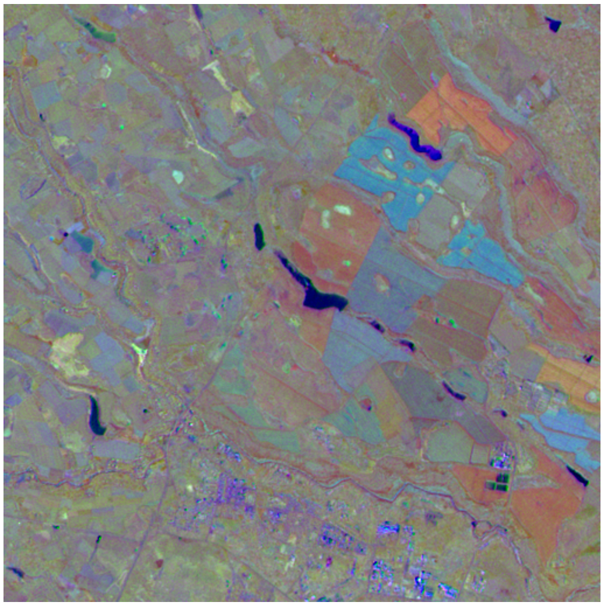
[4] A. A. Nielsen, K. Conradsen, and J. J. Simpson, "Multivariate alteration detection (MAD) and MAF postprocessing in multispectral, bitemporal image data: New approaches to change detection studies," *Remote Sensing of Environment*, vol. 64, no. 1, pp. 1–19, 1998, http://www2.imm.dtu.dk/pubdb/p.php?1220.

[5] A. A. Nielsen, "The regularized iteratively reweighted MAD method for change detection in multi- and hyperspectral data," *IEEE Transactions on Image Processing*, vol. 16, no. 2, pp. 463–478, Feb. 2007, http://www2.imm.dtu.dk/pubdb/p.php?4695.

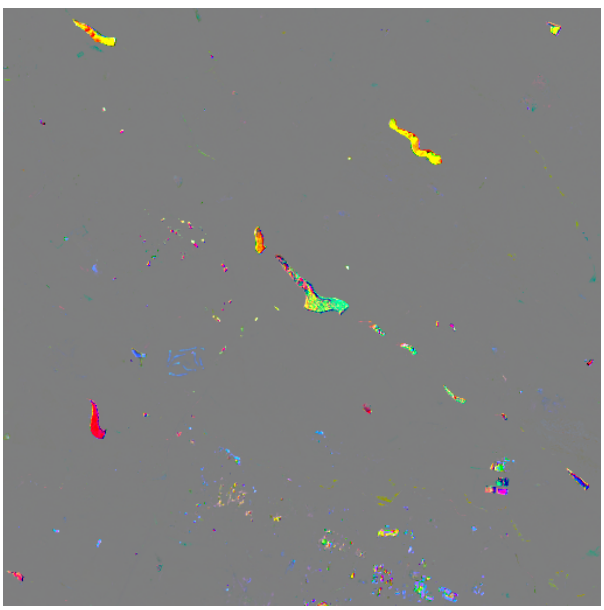
[6] M. J. Canty, A. A. Nielsen, and M. Schmidt, "Automatic radiometric normalization of multitemporal satellite imagery," *Remote Sensing of Environment*, vol. 91, no. 3-4, pp. 441–451, Jun. 2004, http://www2.imm.dtu.dk/pubdb/p.php?2815.

[7] M. J. Canty and A. A. Nielsen, "Automatic radiometric normalization of multitemporal satellite imagery with the iteratively re-weighted MAD transformation," *Remote Sensing of Environment*, vol. 112, no. 3, pp. 1025–1036, Mar. 2008, http://www2.imm.dtu.dk/pubdb/p.php?5362.

[8] B. Schölkopf, A. Smola, and K.-R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem,"



(a) iMAD.



(b) kMAD.

**Fig. 3**. Change variables iMADs and kMADs 1, 2 and 3 as RGB. All variables are stretched over 12 standard deviations.


*Neural Computation*, vol. 10, no. 5, pp. 1299–1319, 1998.

[9] 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.

[10] A. A. Nielsen, "Kernel maximum autocorrelation factor and minimum noise fraction transformations," *IEEE Transactions on Image Processing*, vol. 20, no. 3, pp. 612–624, Mar. 2011, http://www2.imm.dtu.dk/pubdb/p.php?5925.

[11] L. Gómez-Chova, A. A. Nielsen, and G. Camps-Valls, "Explicit signal to noise ratio in reproducing kernel Hilbert spaces," Vancouver, Canada, Jul. 2011, IGARSS, invited contribution, http://www2.imm.dtu.dk/pubdb/p.php?6004.

[12] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, Springer, second edition, 2009.

[13] B. Schölkopf and A. Smola, *Learning with Kernels*, Massachusetts Institute of Technology Press, 2002.

[14] J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*, Cambridge University Press, 2004.

[15] G. Camps-Valls and L. Bruzzone, Eds., *Kernel Methods for Remote Sensing Data Analysis*, Wiley, 2009.

[16] H. D. Vinod, "Canonical ridge and econometrics of joint production," *Journal of Econometrics*, vol. 4, pp. 147–163, 1976.

[17] H. Kwon and N. M. Nasrabadi, "Kernel RX-algorithm: A nonlinear anomaly detector for hyperspectral imagery," *IEEE Transactions on Image Processing*, vol. 43, no. 2, pp. 388–397, Feb. 2005.