UNIVERSITAT
JAUME·I

Título artículo / Títol article:	Effect of denoising in band selection for regression tasks in hyperspectral datasets
Autores / Autors	Pedro Latorre Carmona, Jose Martínez Sotoca, Filiberto Pla, José Bioucas Dias, Carme Julià Ferré
Revista:	IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING
Versión / Versió:	Versió post-print
Cita bibliográfica / Cita bibliogràfica (ISO 690):	LATORRE-CARMONA, Pedro, et al. Effect of denoising in band selection for regression tasks in hyperspectral datasets. *IEEE Journal of selected topics in applied Earth observations and remote sensing*, 2013, vol. 6, no 2, p. 473 – 481.
url Repositori UJI:	http://hdl.handle.net/10234/87989

Effect of denoising in band selection for regression tasks in hyperspectral datasets

Pedro Latorre Carmona, Jose Martínez Sotoca, Filiberto Pla, José Bioucas Dias, Carme Julià Ferré

Abstract—This paper presents a comparative analysis of six band selection methods applied to hyperspectral datasets for biophysical variable estimation problems, where the effect of denoising on band selection performance has also been analyzed. In particular, we consider four hyperspectral datasets and three regressors of different nature ($\varepsilon-$SVR, Regression Trees, and Kernel Ridge Regression). Results show that the denoising approach improves the band selection quality of all the tested methods. We show that noise filtering is more beneficial for the selection methods that use an estimator based on the whole dataset for the prediction of the output than for methods that use strategies based on local information (neighboring points).

Index Terms—Feature Selection, Regression, Noise, hyperspectral datasets

I. INTRODUCTION

HYPERSPECTRAL sensors usually operate in the portion of the spectrum extending from the visible region through the nearinfrared and midinfrared range (wavelengths between 0.3 and 2.5 μm) acquiring hundreds of narrow bands. This very high spectral resolution has fostered applications in many areas such as remote sensing, medical imaging, product quality inspection, fine arts, just to cite a few. A special effort has been made in the use of hyperspectral imaging for remote sensing, and in particular: (a) to infer the value of biological or chemical variables related to the health status of forests and vegetation zones, as well as others that impact their life cycle [2], [13], (b) to analyze the production capability of crops [29], (c) to estimate surface temperature (for instance, [21]), and (d) to determine the health status of oceans and seas [15]. One important issue in hyperspectral data classification and regression is band redundancy [8], [27], [12], [28]. In this work, we consider the band selection problem in the sense of dimensionality reduction.

In regression problems, some covariates may be measured indirectly, and/or under some imprecision or uncertainty. Models dealing with these problems are called errors-in-variables models and have been extensively studied in the last thirty to fourty years [1], [4]. However, only recently this analysis has shifted to the feature selection problem for these models, despite the fact of the concurrent negative effect given by noise and the detection of irrelevant or redundant features.

Pedro Latorre Carmona, Jose Martínez Sotoca and Filiberto Pla are with the Institute of New Imaging Technologies, Universidad Jaume I, Campus del Riu Sec, s/n, 12071 Castellón de la Plana, SPAIN. Emails: [latorre,sotoca,pla]@uji.es

José Bioucas Dias is with Instituto de Telecomunicações and Instituto Superior Técnico, Technical University of Lisbon, 1049-001 Lisbon, Portugal. E-mail: bioucas@lx.it.pt

Carme Julià Ferré is with the Computer Vision Center, Edifici O, Campus UAB, 08193 Bellaterra, Spain. Email: carme.julia@urv.cat

For the best of our knowledge, only the work of Ma et al [18] analyzes the effect of measurement errors in the feature selection capability of generally non-linear feature selection methods for regression, from a theoretical point of view, using penalized estimating equations. Shah et al [25] and Liang et al [17] consider the case of the analysis of the stability of the selection of variables when the variables are affected by measurement errors. Fuchs et al [9] developed a variable selection method that accounted for errors in the predictor as well as in the regressor variables. However, only the linear regressor case is considered. Zhao et al [30] analyzed the consistency of the Lasso method in its ability to select the correct prediction model when measurement errors are also present. Finally, Byeon et al [6] proposed a feature selection method that accounted for the detection of noisy samples in general classification and regression datasets. Other simple linear or partially linear models have also analyzed the effect of noise of feature selection [31], [32].

From a practical point of view, noise elimination in hyperspectral remote sensing is usually made using multiple regression theory-based approaches. The high correlation between neighboring spectral bands is the main reason underlying the good performance of the multiple regression theory in this field. Furthermore, the high spectral correlation means that the hyperspectral vectors belong to low dimensional subspaces, which is beneficial to noise reduction [7].

The work presented in this paper is, for the best of our knowledge, the first one that conducts an experimental study about the denoising effect on the selection capability of different band selection techniques for regression tasks applied to real hyperspectral datasets. The denoising method applied is the *HySime* method proposed by Bioucas-Dias et al. in [5], which detects the regions of the spectrum that are more affected by noise. Most noise reduction methods applied to remote sensing use the spatial distribution of the image pixels. However, *HySime* does not use this kind of information. This strategy allows to solve those cases where only information about a few (non necessarily spatially related) number of pixels in the image is available. The main objectives and contributions of this paper are:

- To analyze the effect of data denoising on band selection performance.
- To evaluate the performance of different band selection techniques.

II. METHODOLOGY

This section summarizes the basic concepts of the band selection for regression techniques applied in the comparative analysis. The first four techniques correspond to filter

approaches while the last two ones are wrapper approaches. A brief description of the methods follows below:

- **CMI**$_{Dist}$: This method is based on the application of a hierarchical clustering strategy based on Ward's linkage to find clusters of bands using a conditional mutual information distance [16]. For each cluster C_i, it selects the variable with the highest mutual information with respect to the output variable.

- **CMI**$_{Kras}$: This method uses a similar strategy to the previous technique (**CMI**$_{Dist}$), but using another mutual information estimator proposed by Kraskov et al in [14]. This method is based on the use of the digamma function on the k-nearest neighbours for each point in the dataset.

- **Rossi [22]**: This method adds variables using a sequential forward selection technique. Selection of variables is based on the estimation of the mutual information using the estimator of [14].

- **FSR**: Forward Stepwise Regression is a classical linear regression model. The significance of each variable is determined from its t-statistics with the null hypothesis that the correlation between the regressor and the predictor variable is 0. The significance of variables is sorted calculating the p-values of the t-statistic.

- **Elastic Net (EN)** [33]: It proposes the use of a regularization which is a weighted sum of the l_1-norm and the square of the l_2-norm of the coefficient vector formed by the weights of each variable.

- **PS-FS**: It is based on a Particle Swarm Optimization (*PSO*) strategy (Particle-Swarm Feature Selection, PS-FS, [20]). It uses a Neural Network regressor which makes band selection using two particle swarms, one to select a fixed number of variables, and the other to select the best variable set size.

III. DATASET DESCRIPTION

A synthetic hyperspectral regression dataset was generated using the PROSAIL radiative transfer code [11]. This code combines the PROSPECT code to generate the reflectance curve at leaf level of vegetation, with the SAIL code, that generates the Bidirectional Reflectance Function (*BRF*) [24] at canopy level as a function of the leaf biochemical constituent values. 4000 *BRF* curves were generated by PROSAIL, varying the leaf chlorophyll content in the range $[20, 80] \frac{\mu g}{cm^2}$. Afterwards, gaussian noise with $\sigma = \{0.005, 0.01, 0.015, 0.02\}$ was added to the spectral curves of the dataset.

On the other hand, four real hyperspectral datasets were considered: three of them correspond to data points acquired from hyperspectral images obtained from a field campaign (*SEN2FLEX2005* [19]) carried out in Barrax, Albacete, Spain, in 2005. The last dataset is a problem related to fruit quality assessment.

- *AHS*. The dataset consists of the radiance values of image pixels that were taken by the Airbone Hyperspectral Scanner (*AHS*) sensor, during five days in June and July 2005. Corresponding thermal measurements for these pixels were also acquired. This dataset only uses the *AHS* last 10 bands, which are in the infrared/thermal range.

Band 71 is centered at $8.02 \mu m$, and band 80, at $12.85 \mu m$. Since the number of points in the dataset was comparable to the number of input bands, its number was increased using also the points of a 9×9 window around each experimental point. The final training and test datasets are formed by 971 data points each.

- *CASI-AHS-CHLOR*. It consists of the reflectance values of image pixels that were taken by the *CASI* and the *AHS* sensors, on July 13, 2005 during the *SEN2FLEX2005* campaign. The Compact Airborne Spectrographic Imager (*CASI*) sensor images are formed by 144 bands between 0.37 and $1.05 \mu m$. For the *AHS* sensor, the first 63 bands, between 0.46 and $2.49 \mu m$ were considered. Therefore, the input dimensionality of this dataset is 207 bands. The training and test sets are formed by 2205 and 2139 data points, respectively. Corresponding chlorophyll measurements for these pixels were also acquired. The training dataset contained chlorophyll measurements from 7 different crop areas, but there were only of two different crop types (alfalfa and corn) at different growth stages.

- *FGVC*. It consists of the reflectance values of image pixels that were taken by the *CASI* and the *AHS* sensors on July 13, 2005 during the *SEN2FLEX2005* campaign. Output values consist of the Fraction of Green Vegetation Cover (*FGVC*), defined as the fraction of horizontal area associated with the photosynthetically active green vegetation. Digital Hemispherical Photographs (*DHPs*) were adquired at Nadir over eight different places in the test site, consisting of sugarbeet, corn and onion crops. The training and test sets are formed by 1347 and 1309 data points, respectively. The input dimensionality is 207 bands.

- *Orange Juice* [3]. It is divided into 150 points for training and 68 points for testing obtained in the near-infrared spectra (from $1.10 \mu m$ to $2.50 \mu m$, acquiring 700 bands in total), and the corresponding level of saccharose of different orange juice samples. This dataset was provided by Prof. Marc Meurens[1]. Since the number of input variables is higher than the input space dimensionality, a piecewise cubic spline interpolation was applied on the training and testing datasets which allowed a reduction in the number of bands to 100.

IV. EXPERIMENTAL RESULTS

In order to analyze the quality of the band selection algorithms, regardless of the type of regressor used, three different regression methods were applied: Support Vector Regression ($\varepsilon-$ *SVR*) with a radial basis function, Regression Trees (*RT*) and Kernel Ridge Regression (*KRR*)[23]. The regression schemes were trained using the same training set. In each case, the subset of selected bands is validated using the Root Mean Square Error (*RMSE*) value given by the regressor. For *SVR*, the selection of its working parameters (C, σ, ϵ), for each one of the datasets, with and without noise, was made applying an exhaustive grid search using equally spaced steps in the logarithmic space of these parameters. In the case of Kernel
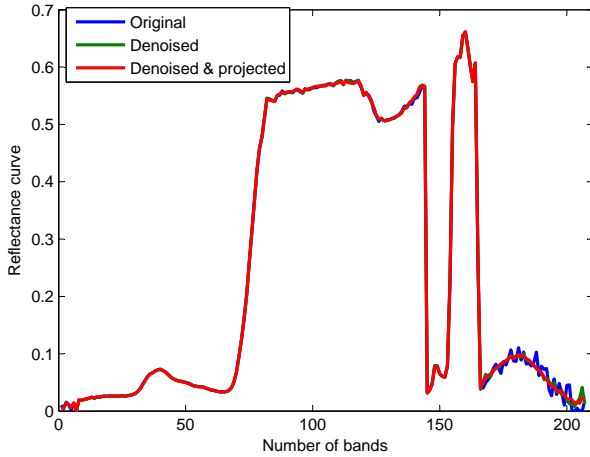
---

[1]http://www.ucl.ac.be/mlg

Fig. 1. Denoising results on one hyperspectral curve of the *CASI-AHS-CHLOR* dataset. Blue: noisy reflectance curve obtained by the *CASI* and *AHS* sensors; Red: result given by *HySime*. Green: reconstruction of the projected spectral signal
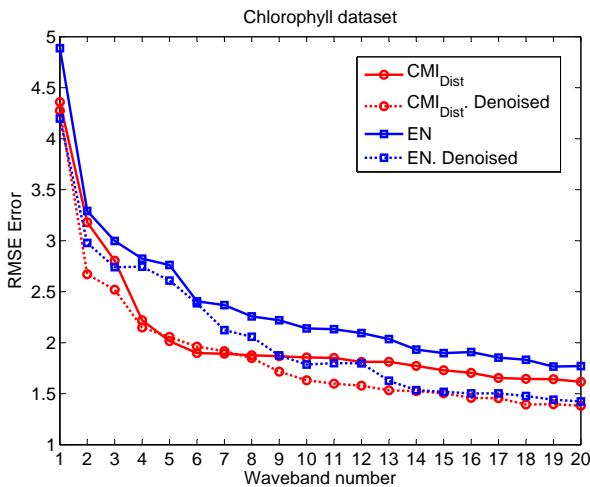

Fig. 2. Denoising effect on the $RMSE$ error for the the $CASI-AHS-CHLOR$ dataset and the EN and CMI_{Dist} feature selection methods.

ridge Regression, the same strategy as the one used for *SVR* was applied, i. e., a grid search over the logarithmic scale space of the (λ, σ) parameters, where λ is the regularization parameter, and σ is the kernel parameter. Figure 1 shows, in blue, the atmospherically corrected curve obtained by the *CASI* and *AHS* sensors, and in red the result given by *HySime*, for one pixel of the *CASI-AHS-CHLOR* dataset. In this case, the denoising effect is stronger in bands above 160. *HySime* [5] starts by estimating the signal and the noise correlation matrices and then it selects the subset of eigenvectors that best represents the signal subspace in the minimum mean square error sense. It works minimizing the power of the signal projection error and the power of the noise projection.

Table I shows the accumulated *RMSE* results for the first 10 and 20 selected bands for two of the feature selection methods (CMI_{Dist} and CMI_{Kras}) for the synthetic dataset with different gaussian noise levels, using *SVR* as regressor. We

considered these two methods because they use the same band grouping strategy, but they use a different mutual information estimator. The idea is to analyse the behaviour of the estimator in relation to the noise when a denoising algorithm is applied to a synthetic problem.

The statistical significance of the difference in performance of the two feature selection methods was assessed using a nonparametric Wilcoxon Rank-Sum test [26]. In this test, we consider that the difference in the results is statistically significant if the (two-sided) *p*-value is < 0.01. We also use the $(+)$ sign to indicate a statistically significant difference (and the $(-)$ sign otherwise). We can see from Table I that CMI_{Kras} works better, compared to CMI_{Dist} for the synthetic dataset, before the application of the *Hysime* denoising method. On the other hand, we can also see that denoising improves *RMSE* error for both methods and that denoising gives better results for CMI_{Dist} compared to CMI_{Kras}.

Table II summarizes the performance results over the different band selection approaches, where the *HySime* method was applied for the four real hyperspectral datasets. Results in rows $K = 5$, $K = 10$, $K = 15$ and $K = 20$ show the accumulated *RMSE* (approximate area under the *RMSE* curve) in the ranges from 1 to 5, from 1 to 10, from 1 to 15, and from 1 to 20 bands, respectively. These four intervals of subsets of bands have been considered to be the approximated transitory period to reach a stable reduction of error for most of the datasets and regression algorithms used. Results obtained with the best of the band selection methods for 20 bands are similar to those obtained with the total amount of bands.

In order to analyze the statistical significance of the results, Friedman and Quade Tests [10] were applied on the results with a confidence level $p = 0.005$. This kind of technique measures the significance of the statistical difference of the results of the algorithms, using rankings of results obtained by them. Both statistical methods use the Fisher distribution to discern the statistical significance for the six methods and over the first $K = 5, 10, 15, 20$ variables.

This distribution follows $(N_M - 1)$ and $(N_M - 1) \cdot (N_B - 1)$ degrees of freedom, where N_M is the number of methods, and N_B the number of feature subsets on which the ranking is applied. Therefore, for the different rows in the table ($K = 5, 10, 15, 20$), we obtain the values: $F(5, 20) = 4.76$, $F(5, 45) = 3.91$, $F(5, 70) = 3.70$, and $F(5, 95) = 3.61$. The table shows the statistical significance being positive $(+)$ when the value of the test is greater than the Fisher distribution and negative $(-)$ otherwise.

Table II shows that the CMI_{Dist} method has the best performance in all cases except for the *KRR* and *RT* regressors on the *AHS* and *Orange Juice* datasets. The *EN* method is the second best one.

Analyzing the *Rossi* method in relation to the CMI_{kras} method, we can see that although both methods use the same nearest-neighbour information estimator, the CMI_{kras} method estimates the conditional mutual information which allows to analyze the relationships of information between features. This in theory should make it superior to the *Rossi* method. Nevertheless, looking at the results we see that this is accomplished when validated with the $\varepsilon-$SVR and *KRR*

regression algorithms, but not when using *RT*.

PS-FS is a *wrapper* type method based on a Neural Network regressor to make an optimal search, and where the error of the regressor acts as the search criteria. It obtains good results for all the datasets when using $\varepsilon-$SVR and *RT* as regressors. *FSR* gives the worst results in most of the cases.

Figure 2 shows the denoising effect on the *RMSE* error for the *CASI-AHS-CHLOR* dataset for two of the feature selection methods used (CMI_{Dist} and *EN*). We can see that the error decreases for the first 20 selected bands for both methods. This behaviour is also observed for the other methods (though not shown in the plot). Figure 2 also shows that denoising effect is higher for the *EN* method, allowing the performance for the denoised case and both methods to be similar.

Figure 3 shows the distribution of crops for a part of the Barrax test site during the *SEN2FLEX* campaign. This figure also includes the chlorophyll values measured *in situ* (in $\frac{\mu g}{cm^2}$). Chlorophyll measurements were made in seven different locations of the Barrax test site during this campaign. The test data set was formed by the rest of the points in the image. A manual labeling of different crops (alfalfa, bare soil, corn, barley, garlic, wheat, forest, onion and opium poppy) was made during the field campaign. This information was used to identify the image pixels that corresponded to these labeled crops. With these pixels we built a training and a validation set with 3000 data points each one. An analysis of the quality of the samples was performed by classifying the validation set from the training set with the Nearest Neighbor classifier (*NN*) obtaining an accuracy of 97.5%. Later, we gathered both sets and the *NN* classifier was used to assign the labels to the rest of the image pixels with spectral information from both sensors (138541 pixels).

Figure 4(a)-(d) shows the regression results for the whole Barrax test site image, for the case where all bands have been used (Figure 4(a)), and for three of the feature selection methods considered, i. e., *FSR*, *EN* and CMI_{Dist} (Figure 4(b)-(d)). In order to obtain these images, first a validation with the Regression Tree technique was performed with all bands, estimating the values of chlorophyll in the validation set (2139 data points) from the training set (2205 points), with a *RMSE* error of 1.81. Then, both sets were gathered, and *Regression Trees* technique was trained with the data points corresponding to the in situ measured chlorophyll values (Figure 3).

Figure 5 shows the corresponding regression results when the *HySime* denoising method was used. The same strategy was applied to infer the chlorophyll values for the whole image. From both Figures, we can see that:

1) CMI_{Dist} gives the best overall results in comparison with the other three feature selection methods, taking into account the chlorophyll measures inferred using all the sensor bands (Figures 4(a) and 5(a)).
2) The three methods correctly infer the chlorophyll values for the regions that form the training dataset, for the first 20 bands selected. However, all of them fail to give correct chlorophyll estimation results for the rest of the image.
3) The three feature selection methods and the regression strategy used give chlorophyll estimation results that

seem *homogeneous* for the rest of the pixels in the image. We think this is because we had samples of just seven regions and of two types of crops. This makes the regressor to obtain similar chlorophyll values for all these pixels. To make a better estimation of the chlorophyll values for the rest of the pixels, we should have had chlorophyll information about other crop types.

On the other hand, tables III and IV show the performance of the feature selection algorithms (averaged over the three regression methods) on the original and denoised datasets, respectively. Some interesting points deserve our attention:

1) Tables III and IV show that the application of a denoising technique on all the hyperspectral datasets before doing the band selection improves the regression error. This improvement is highest for CMI_{Dist}, *EN* and *PS-FS*. The behaviour observed in CMI_{Dist} is in agreement with the results observed in Table I. In the case of the wrapper techniques (*EN* and *PS-FS*), they use the whole training dataset to do band selection. CMI_{Dist} is a filter-type method that uses information measures estimated using a Watson-Nadaraya kernel. This is a parametric kernel that needs all the training dataset samples in order to assess these parameters correctly.
2) The worse behavior for the CMI_{kras} and *Rossi* methods when the *Hysime* denoising method is applied may be due in part to the strategy of the estimator used to obtain the information measures. The estimator in both cases is a local estimator based on a distance criterion between sample neighbours [14]. Thus noise filtering modifies the distance and the order of the neighbors in the neighborhood, and this may not always be beneficial.
3) *FSR* performs the poorest, and it is also dataset and regressor dependent. This may be partly because *FSR* makes band selection considering a linear regression model.
4) *RMSE* reduction due to noise filtering is not the same for all the real datasets, being higher for *CASI-AHS-CHLOR* and *FGVC* datasets. In the case of the *Orange Juice* dataset, noise reduction effect in the band selection methods is not so evident. This may be because the number of samples is small for a high dimensional space, which implies a high dispersion of the data. For the *AHS* dataset, the number of bands is relatively small and the noise reduction using the *HySime* method does not significantly improve the regression error.

V. CONCLUSIONS

In this paper, the effect of denoising on the selection capability of different band selection techniques for regression tasks applied to one synthetic and four real hyperspectral datasets has been investigated. Regression results on a Barrax test site image, chlorophyll measures in this case, show the capability to correctly estimate them, using several band selection methods, with a reduced number of bands. This seems to indicate the possibility to be able to reduce the spectral range where the sensor would be really useful in order to measure the biophysical parameters of the crops aimed at being identified.

Results for the original bands

		$\sigma = 0.005$	$\sigma = 0.010$	$\sigma = 0.015$	$\sigma = 0.020$
CMI_{Dist}	(K = 10)	29.69	42.51	53.03	61.08
CMI_{Kras}	(K = 10)	29.42	40.33	51.57	58.16
Test	(p-value)	1.0 (-)	0.364 (-)	0.427 (-)	0.879 (-)
CMI_{Dist}	(K = 20)	55.08	78.22	97.08	101.25
CMI_{Kras}	(K = 20)	53.13	71.32	84.33	97.75
Test	(p-value)	0.310 (-)	0.035 (-)	0.007 (+)	0.330 (-)

Results for the corresponding denoised bands

		$\sigma = 0.005$	$\sigma = 0.010$	$\sigma = 0.015$	$\sigma = 0.020$
CMI_{Dist}	(K = 10)	19.01	24.85	35.21	43.83
CMI_{Kras}	(K = 10)	30.00	26.90	36.23	44.71
Test	(p-value)	0.002 (+)	0.003 (+)	0.241 (-)	0.364 (-)
CMI_{Dist}	(K = 20)	36.51	49.86	68.86	86.77
CMI_{Kras}	(K = 20)	56.11	52.28	72.40	87.42
Test	(p-value)	$6.8\ 10^{-8}$ (+)	$2.9\ 10^{-5}$ (+)	$5.1\ 10^{-4}$ (+)	0.956 (-)

TABLE I
ACCUMULATED *RMSE* ERROR FOR THE CMI_{Dist} AND CMI_{Kras} METHODS FOR THE PROSAIL SYNTHETIC DATASET, FOR DIFFERENT LEVEL OF GAUSSIAN NOISE ADDED
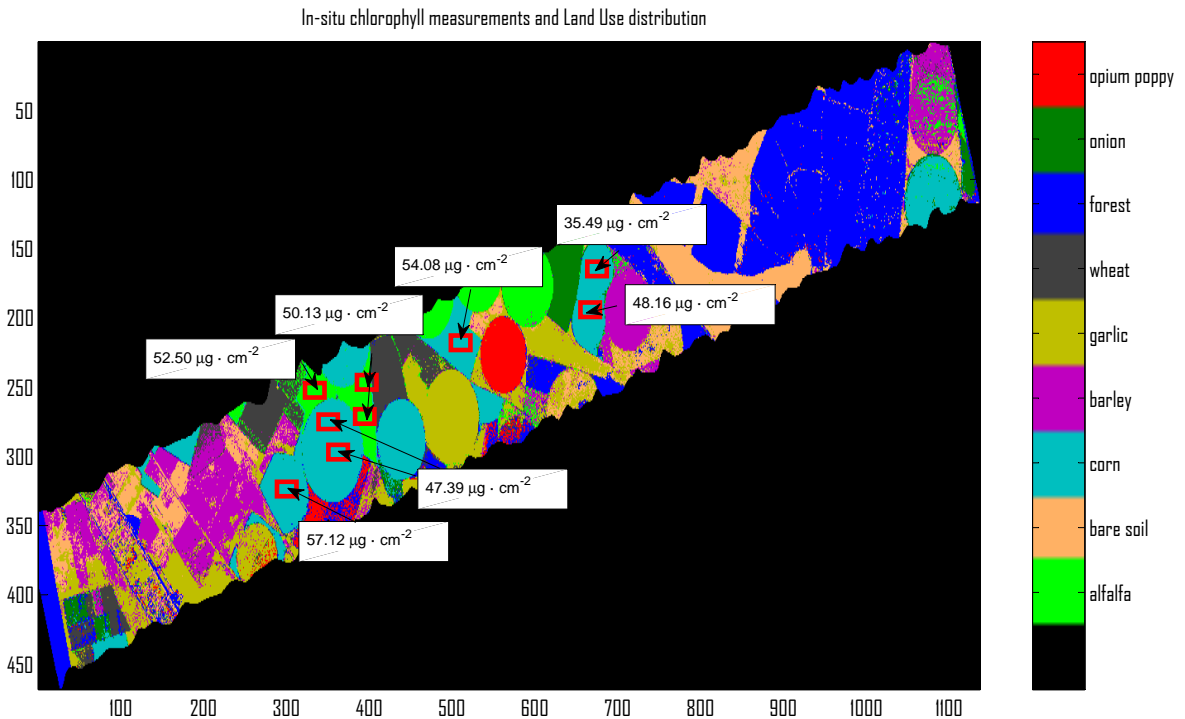

Fig. 3. Distribution of crops and *points* where in situ chlorophyll measurement were made.

From the other observed results, we can also conclude that the application of the *HySime* denoising technique on the datasets before performing the band selection, reduces the regression error. This also applies when using regressors of different nature ($\varepsilon-$SVR, Regression Trees, and Kernel Ridge Regression). This improvement is highest for three of the feature selection methods used in the experiments (CMI_{Dist}, *EN* and *PS-FS*). Their selection criterion uses the information given by the whole dataset for the prediction of the output. Noise elimination improves their prediction. The worse behavior when *Hysime* is applied, for the CMI_{kras} and *Rossi* methods may be because the estimator in both cases is based on a distance criterion between sample neighbors.

This does not always improve the selection performance after noise filtering. We have also seen that the degree of *RMSE* reduction due to noise filtering is dependent on the complexity and structure of the dataset.

ACKNOWLEDGMENT

CASI-AHS-CHLOR dataset

$\varepsilon-$SVR	CMI_{Dist}	CMI_{Kras}	Rossi	FSR	EN	PS$-$FS	Friedman Test	Quade Test
K = 5	13.67	18.55	19.20	25.80	15.27	**13.55**	22.92 (+)	10.51 (+)
K = 10	**22.75**	30.21	29.24	40.94	25.50	23.13	35.37 (+)	18.22 (+)
K = 15	**30.49**	41.26	37.62	52.72	33.78	31.54	53.31 (+)	21.74 (+)
K = 20	**37.57**	50.68	45.07	63.02	41.12	39.29	72.16 (+)	28.95 (+)
RT	CMI_{Dist}	CMI_{Kras}	Rossi	FSR	EN	PS$-$FS	Friedman Test	Quade Test
K = 5	**11.34**	14.88	15.03	24.69	12.84	12.70	13.77 (+)	7.16 (+)
K = 10	**19.20**	25.16	25.39	42.99	22.42	23.17	23.37 (+)	12.12 (+)
K = 15	**27.86**	32.98	34.52	59.84	31.26	32.62	23.41 (+)	13.05 (+)
K = 20	**35.37**	41.04	43.67	75.09	40.53	41.89	35.88 (+)	18.27(+)
KRR	CMI_{Dist}	CMI_{Kras}	Rossi	FSR	EN	PS$-$FS	Friedman Test	Quade Test
K = 5	**12.18**	15.71	14.52	23.97	14.47	13.46	11.02 (+)	7.70 (+)
K = 10	**20.04**	23.61	21.35	38.95	23.73	21.89	17.49 (+)	8.86 (+)
K = 15	**26.92**	30.44	27.13	52.14	32.77	29.86	34.42 (+)	13.41 (+)
K = 20	33.19	36.84	**32.94**	63.43	41.11	37.35	52.64 (+)	19.82 (+)

AHS dataset

$\varepsilon-$SVR	CMI_{Dist}	CMI_{kras}	Rossi	FSR	EN	PS$-$FS	Friedman Test	Quade Test
K = 5	**10.43**	10.71	11.57	13.30	14.57	11.01	8.11 (+)	6.60 (+)
K = 10	22.71	22.97	24.13	25.42	27.90	23.58	5.52 (+)	6.14 (+)
RT	CMI_{Dist}	CMI_{kras}	Rossi	FSR	EN	PS$-$FS	Friedman Test	Quade Test
K = 5	22.08	**22.04**	24.81	24.41	24.96	23.94	0.42 (-)	0.37 (-)
K = 10	49.16	50.68	52.97	55.89	53.84	52.83	1.07 (-)	1.37 (-)
KRR	CMI_{Dist}	CMI_{kras}	Rossi	FSR	EN	PS$-$FS	Friedman Test	Quade Test
K = 5	13.18	13.60	13.51	14.92	15.07	**12.96**	3.94 (-)	5.85 (+)
K = 10	21.59	22.37	22.37	23.80	24.82	**21.58**	7.84 (+)	9.14 (+)

FGVC dataset

$\varepsilon-$SVR	CMI_{Dist}	CMI_{Kras}	Rossi	FSR	EN	PS$-$FS	Friedman Test	Quade Test
K = 5	**0.32**	0.53	0.55	0.57	0.36	**0.32**	19.65 (+)	7.54 (+)
K = 10	**0.54**	0.80	0.87	0.88	0.55	0.55	53.75 (+)	16.64 (+)
K = 15	**0.72**	1.01	1.08	1.14	0.73	0.75	71.20 (+)	25.12 (+)
K = 20	**0.87**	1.19	1.25	1.37	0.90	0.92	52.08 (+)	27.10 (+)
RT	CMI_{Dist}	CMI_{Kras}	Rossi	FSR	EN	PS$-$FS	Friedman Test	Quade Test
K = 5	**0.21**	0.43	0.38	0.44	0.25	0.27	15.13 (+)	7.53 (+)
K = 10	**0.32**	0.74	0.56	0.73	0.39	0.49	50.55 (+)	15.86 (+)
K = 15	**0.40**	1.04	0.73	1.02	0.54	0.69	107.82 (+)	26.75 (+)
K = 20	**0.48**	1.33	0.89	1.30	0.69	0.89	154.97 (+)	37.14 (+)
KRR	CMI_{Dist}	CMI_{Kras}	Rossi	FSR	EN	PS$-$FS	Friedman Test	Quade Test
K = 5	**0.42**	0.60	0.63	0.73	0.44	0.46	28.71 (+)	9.56 (+)
K = 10	0.76	1.01	1.09	1.25	**0.74**	0.88	81.78 (+)	19.58 (+)
K = 15	1.09	1.37	1.47	1.71	**1.02**	1.26	109.74 (+)	30.17 (+)
K = 20	1.40	1.70	1.83	2.14	**1.31**	1.62	162.57 (+)	40.11 (+)

Orange Juice dataset

$\varepsilon-$SVR	CMI_{Dist}	CMI_{Kras}	Rossi	FSR	EN	PS$-$FS	Friedman Test	Quade Test
K = 5	53.92	51.83	52.32	48.86	48.74	**48.47**	5.67 (+)	3.70 (-)
K = 10	**92.97**	100.77	103.11	94.33	94.91	96.59	11.51 (+)	5.50 (+)
K = 15	**133.80**	148.83	154.24	141.48	140.35	144.13	23.89 (+)	10.09 (+)
K = 20	**175.14**	197.37	205.38	188.55	186.02	191.94	40.90 (+)	16.45 (+)
RT	CMI_{Dist}	CMI_{Kras}	Rossi	FSR	EN	PS$-$FS	Friedman Test	Quade Test
K = 5	54.09	45.04	**43.40**	56.51	47.59	53.43	14.82 (+)	10.16 (+)
K = 10	106.15	96.14	**91.10**	110.29	98.22	105.78	16.65 (+)	11.50 (+)
K = 15	159.44	147.64	**142.86**	163.93	149.94	157.98	18.35 (+)	13.21 (+)
K = 20	217.76	200.71	**194.02**	219.80	201.39	210.93	22.46 (+)	14.95 (+)
KRR	CMI_{Dist}	CMI_{Kras}	Rossi	FSR	EN	PS$-$FS	Friedman Test	Quade Test
K = 5	41.63	41.05	**40.07**	42.02	42.04	42.05	1.65 (-)	2.09 (-)
K = 10	79.40	78.28	78.68	78.01	**77.67**	79.94	1.12 (-)	0.77 (-)
K = 15	107.96	107.58	109.01	107.28	**106.01**	108.88	2.10 (-)	0.66 (-)
K = 20	150.96	151.09	154.25	151.67	**148.19**	152.47	4.96 (+)	2.07 (-)

TABLE II

ACCUMULATED $RMSE$ OVER DIFFERENT SUBSETS OF DENOISED BANDS OBTAINED BY SVR, RT AND KRR WITH DIFFERENT BAND SELECTION ALGORITHMS IN REGRESSION TASKS.

REFERENCES

[1] T. W. Anderson, "Estimating linear statistical relationships", *Ann. Statist.*, vol. 12, pp. 1-45, 1984.

[2] G. P. Asner, R. E. Martin, "Airborne spectranomics: mapping canopy chemical and taxonomic diversity in tropical forests", *Front. Ecol. Environ.*, vol. 7, no. 5, pp. 269-276, 2009.

[3] N. Benoudjit, E. Cools, M. Meurens and M. Verleysen, "Chemometric calibration of infrared spectrometers: selection and validation variables by non-linear models", *Chemometrics and Intelligent Laboratory Systems*, vol 70, pp. 47-53, 2004

[4] P. J. Bickel, Y. Ritov, "Efficient estimation in the errors in variables model", *Ann. Statist.*, vol. 15, pp. 513-540, 1987.

[5] J. Bioucas-Dias, J.M.P. Nascimento, "Hyperspectral Subspace Identification", *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 8, pp. 2435-2445, Mar. 2008.

[6] B. Byeon, K. Rasheed, "Using Genetic Algorithms for Simultaneous Noise Removal and Feature Selection in Classification and Regression problems," *International Conference on Artificial Intelligence*, pp. 304-310, 2009.

[7] J. M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot, "Hyperspectral unmixing overview: geometrical, statistical, and sparse regression-based approaches," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol.
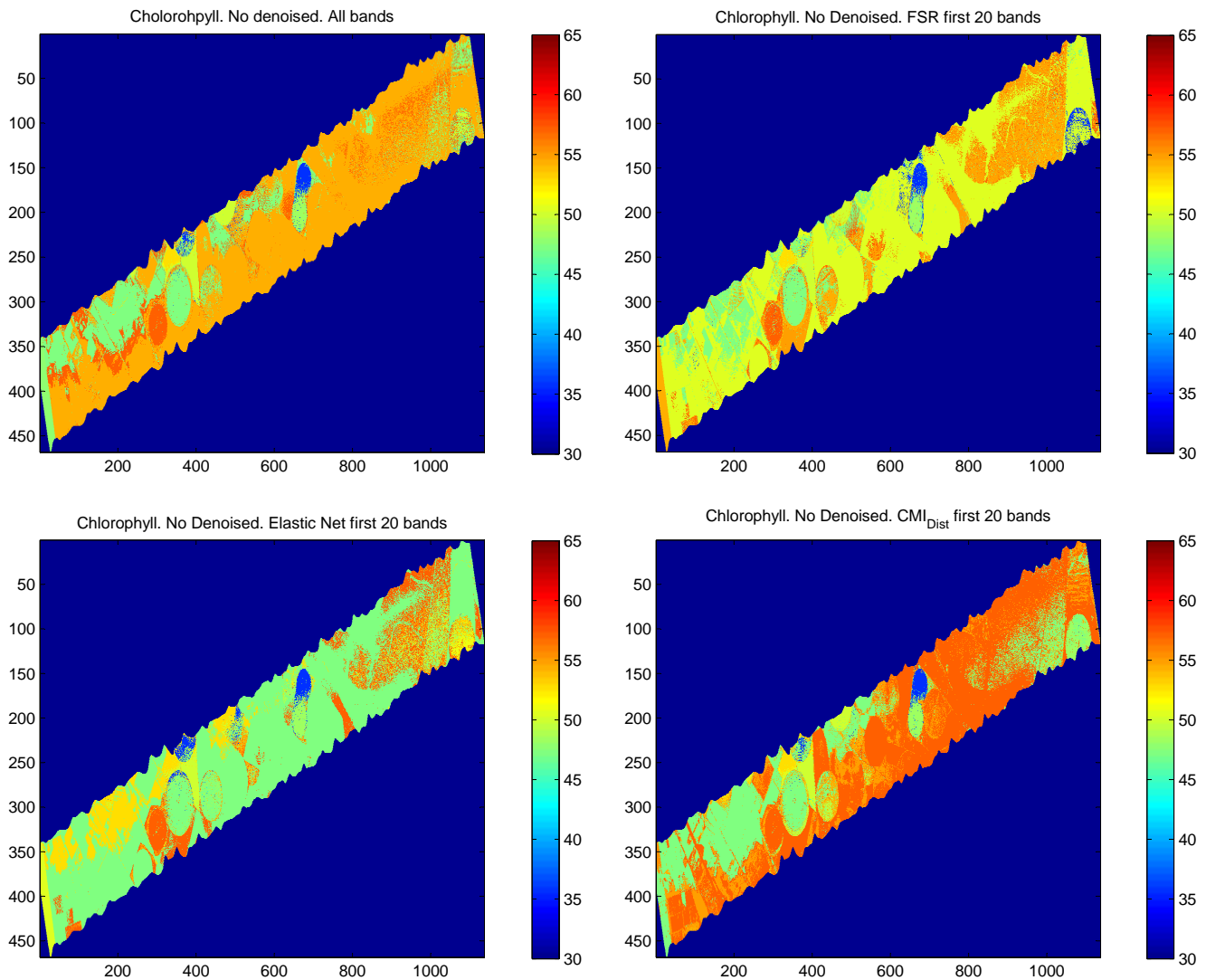
Fig. 4.   Regression Trees (*RT*) results for chlorophyll for the Barrax image acquired by the $CASI$ and AHS sensors. No denoising was applied

5, no. 2, pp. 354-379, 2012.

[8] Yang-Lang Chang; Kun-Shan Chen; Bormin Huang; Wen-Yen Chang; Benediktsson, J.A.; Lena Chang, "A Parallel Simulated Annealing Approach to Band Selection for High-Dimensional Remote Sensing Images", *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 4 , no. 3, pp. 579-590, 2011.

[9] J. J. Fuchs, Sébastien Maria, "A new approach to variable selection using the TLS approach", *IEEE Transactions on Signal Processing*, vol. 55, no. 1, pp. 10-19, 2007.

[10] S. García, A. Fernández, J. Luengo, F. Herrera, "Advanced nonparametric tests for multiple comparisons in the design of experiments in computational intelligence and data mining: Experimental analysis of power", *Information Sciences*, vol. 180, pp. 2044-2064, 2010.

[11] S. Jacquemoud, W. Verhoef, F. Baret, C. Bacour, P. J. Zarco-Tejada, G. P. Asner, C. Franois, S. L. Ustin, "PROSPECT + SAIL models: a review of use for vegetation characterization", *Remote Sensing of Environment*, vol. 113, pp. S56-S66, 2009.

[12] S. Jia, Z. Ji, Y. Qian, L. Shen, "Unsupervised Band Selection for Hyperspectral Imagery Classification Without Manual Band Removal", *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5, no. 2, pp. 531-543, 2012.

[13] R. F. Kokaly, G. P. Asner, S. V. Ollinger, M. E. Martin, C. A. Wessman, "Characterizing canopy biochemistry from imaging spectroscopy and its application to ecosystem studies", *Remote Sensing of Environment*, vol. 113, pp. S78-S91, 2009.

[14] A. Kraskov, H. Stögbauer, P. Grassberger, "Estimating mutual information", *Physical Review E*, vol. 69, 066138, 2004.

[15] E. J. Kwiatkowska, G. S. Fargion, "Application of Machine-Learning Techniques toward the creation of a consistent and calibrated global chlorophyll concentration baseline dataset using remotely sensed ocean color data", *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 12, pp. 2844-2860, 2003.

[16] P. Latorre, J. M. Sotoca , F. Pla, "Filter-Type Variable Selection Based on Information Measures for Regression Tasks", *Entropy*, vol. 14, no. 2, pp. 323-343, 2012.

[17] H. Liang, R. Li, "Variable selection for partially linear models with measurement errors", *Journal of the American Statistical Association*, vol. 104, no. 485, pp. 234-248, 2009.

[18] Y. Ma, R. Li, "Variable selection in measurement error models", *Bernoulli*, vol. 69, no. 1, pp. 274-300, 2010.

[19] J. F. Moreno,"Sen2flex data acquisition report", *Universidad de Valencia*, Technical Report,2005.

[20] S. T. Monteiro, Y. Kosugi, "Particle Swarms for Feature Extraction of Hyperspectral Data", *IEICE Trans. Inf. and Syst.,* vol. E90D, no. 7, pp. 1038-1046, 2007

[21] G. Moser, S. B. Serpico, "Automatic parameter optimization for Support Vector Regression for land and sea surface temperature estimation from remote sensing data", *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 3, pp. 909-921, 2009.

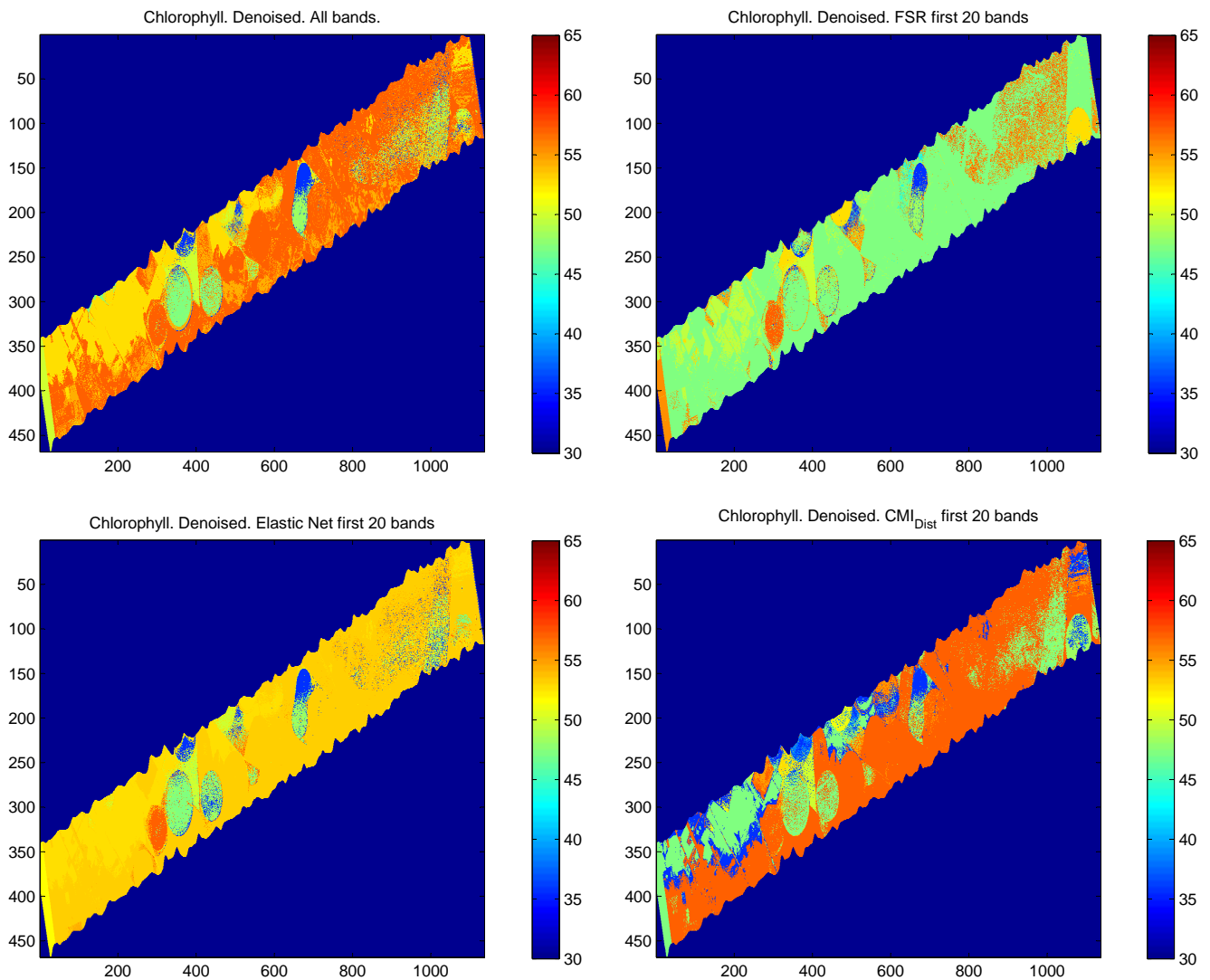[22] F. Rossi, A. Lendasse, D. François, V. Wertz, M. Verleysen, "Mutual in-

Fig. 5. Regression Trees (*RT*) results for chlorophyll for the Barrax image acquired by the $CASI$ and AHS sensors. *Hysime* denoising was applied

formation for the selection of relevant variables in spectrometric nonlinear modelling", *Chemometrics and Intelligent Laboratory Systems,*, vol. 80, pp. 215 226, 2006.

[23] C. Saunders, A. Gammerman and V. Vovk, "Ridge Regression Learning Algorithm in Dual Variables", *in ICML98*, pp. 515-521, 1998.

[24] G. Schaepman-Strub, M.E. Schaepman, T.H. Painter, S. Dangel, J.V. Martonchik, "Reflectance quantities in optical remote sensing definitions and case studies", *Remote Sensing of Environment*, vol. 103, pp. 27 42, 2006.

[25] R. Shah, R. J. Samworth, "Variable selection with error control: another look at stability selection", *J. Royal Stat. Soc., B*, 2012, to appear.

[26] S. Siegel, "Non-parametric statistics for the behavioral sciences", New York: McGraw-Hill. pp. 7583, 1965.

[27] H. Yang, Q. Du, G. Chen, "Unsupervised Hyperspectral Band Selection Using Graphics Processing Units", *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 4, no. 3, pp. 660-668, 2011.

[28] H. Yang, Q. Du, G. Chen, "Particle Swarm Optimization-Based Hyperspectral Dimensionality Reduction for Urban Land Cover Classification", *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5, no. 2, pp. 544-554, 2012.

[29] X. Ye, K. Sakai, A. Sasao, S.-I.-Asada, "Estimation of citrus yield from canopy spectral features determined by airborne hyperspectral imagery", *Int. Journal of Remote Sensing*, vol. 30, no. 18, pp. 4621-4642, 2009.

[30] P. Zhao, B. Yu, "On model selection consistency of Lasso", *Journal of Machine Learning Research*, vol. 7, pp. 2541-2563, 2006.

[31] P. Zhao, L. Xue, "Variable selection for varying coefficient models with measurement errors", *Metrika*, vol. 74, pp. 231-245, 2011.

[32] Z. Zhou, R. Jiang, W. Qian, "Variable selection for additive partially linear models with measurement error", *Metrika*, vol. 74, pp. 185-202, 2011.

[33] H. Zou, T. Hastie, "Regularization and variable selection via the elastic net", *J. R. Statist. Soc. B*, vol. 67, part 2, pp.301-320, 2005.

PLACE PHOTO HERE

Pedro Latorre-Carmona Pedro Latorre-Carmona is a Postdoctoral researcher at the Departament de Llenguatges i Sistemes Informàtics of the University Jaume I, Castellón, Spain. He received the B.S. degree in physics from the University of Valencia, Spain, in 1999, and the Ph.D. degree in computer science from the Polytechnical University of Valencia in 2005. His current research interests are feature selection and extraction, pattern recognition, multispectral (including remote sensing) image processing, colorimetry and vision physics.

CASI-AHS-CHLOR dataset

	CMI_{Dist}	CMI_{Kras}	Rossi	FSR	EN	PS$-$FS
K = 5	**14.14**	16.41	15.95	22.36	15.72	15.24
K = 10	**23.58**	26.98	26.68	39.76	26.82	25.40
K = 15	**32.70**	36.93	35.99	55.19	37.09	35.16
K = 20	**41.24**	46.22	44.74	68.62	46.60	44.30

AHS dataset

	CMI_{Dist}	CMI_{Kras}	Rossi	FSR	EN	PS$-$FS
K = 5	**16.25**	18.15	18.40	16.64	16.89	17.19
K = 10	**31.77**	33.72	33.55	32.20	32.46	32.75

FGVC dataset

	CMI_{Dist}	CMI_{Kras}	Rossi	FSR	EN	PS$-$FS
K = 5	**0.33**	0.53	0.50	0.54	0.36	0.34
K = 10	**0.57**	0.89	0.85	0.88	0.59	0.64
K = 15	**0.80**	1.18	1.12	1.18	0.81	0.90
K = 20	**1.01**	1.45	1.35	1.45	1.02	1.14

Orange Juice dataset

	CMI_{Dist}	CMI_{Kras}	Rossi	FSR	EN	PS$-$FS
K = 5	49.64	46.54	**45.25**	49.64	45.88	47.09
K = 10	93.29	91.73	90.61	99.15	**90.16**	94.25
K = 15	137.02	137.25	136.33	147.76	**134.45**	140.64
K = 20	182.58	182.33	181.93	197.24	**178.29**	186.76

TABLE III

AVERAGE $RMSE$ OVER DIFFERENT SUBSETS OF ORIGINAL BANDS OBTAINED BY AVERAGE OVER THE THREE REGRESSION ALGORITHMS.

CASI-AHS-CHLOR dataset

	CMI_{Dist}	CMI_{Kras}	Rossi	FSR	EN	PS$-$FS
K = 5	**12.40**	16.38	16.25	24.82	14.19	13.24
K = 10	**20.66**	26.33	25.33	40.96	23.88	22.73
K = 15	**28.42**	34.89	33.09	54.89	32.60	31.34
K = 20	**35.38**	42.85	40.56	67.18	40.92	39.51

AHS dataset

	CMI_{Dist}	CMI_{Kras}	Rossi	FSR	EN	PS$-$FS
K = 5	**15.23**	15.45	16.63	17.55	18.20	15.97
K = 10	**31.16**	32.00	33.16	35.04	35.52	32.66

FGVC dataset

	CMI_{Dist}	CMI_{Kras}	Rossi	FSR	EN	PS$-$FS
K = 5	**0.23**	0.38	0.37	0.43	0.25	0.26
K = 10	**0.54**	0.85	0.84	0.95	0.56	0.64
K = 15	**0.74**	1.14	1.09	1.29	0.76	0.90
K = 20	**0.92**	1.41	1.32	1.60	0.97	1.14

Orange Juice dataset

	CMI_{Dist}	CMI_{Kras}	Rossi	FSR	EN	PS$-$FS
K = 5	49.88	45.97	**45.26**	49.13	46.12	47.98
K = 10	92.84	91.73	90.96	94.21	**90.27**	94.10
K = 15	133.73	134.68	135.37	137.56	**132.10**	137.00
K = 20	181.29	183.06	184.55	186.67	**178.53**	185.11

TABLE IV

AVERAGE $RMSE$ OVER DIFFERENT SUBSETS OF DENOISED BANDS OBTAINED BY AVERAGE OVER THE THREE REGRESSION ALGORITHMS.

PLACE
PHOTO
HERE

José M. Sotoca José Martínez Sotoca received a BSc degree in physics from the Universidad Nacional de Educación a Distancia (UNED, Spain) in 1996 and the MSc degree in physics from the University of Valencia, Spain, in 1999. He received the PhD degree in physics also from the University of Valencia in 2001. Currently, he is Assistant Lecturer at the Departament de Llenguatges i Sistemes Informàtics of the University Jaume I, Castellón, Spain. He has collaborated in different projects, most of which are in the medical application of computer science. He has published more than 70 scientific papers in national and international conference proceedings, books and journals. His research interests include pattern recognition and biomedical applications, including image pattern recognition, computer graphics, hyperspectral data and feature extraction and selection.

Filiberto Pla Prof. Filiberto Pla is a Full Professor at the Departament de Llenguatges i Sistemes Informàtics of the University Jaume I, Castellón, Spain. He has been visiting scientist at Silsoe Research Institute, the University of Surrey, the University of Bristol in UK, CEMAGREF in France, the University of Genoa in Italy, Instituto Superior Técnico of Lisbon in Portugal, the Swiss Federal Institute of Technology ETH-Zurich and the idiap Research Institute in Switzerland. At present, he is the Director of the Institute of New Imaging Technologies, at the University Jaume I. His current research interests are colour and spectral image analysis, visual motion analysis, active vision and pattern recognition techniques applied to image processing. He is member of the Spanish Association for Pattern Recognition and Image Analysis, AERFAI, which is a member of the International Association for Pattern Recognition.

Joseé Bioucas-Dias José Bioucas-Dias (S'87, M'95) received the EE, MSc, PhD, and "Agregado" degrees in electrical and computer engineering from Instituto Superior Técnico (IST), the engineering school of the Technical University of Lisbon (TULisbon), Portugal, in 1985, 1991, 1995, and 2007, respectively. Since 1995, he has been with the Department of Electrical and Computer Engineering, IST, where he was an Assistant Professor from 1995 to 2007 and an Associate Professor since 2007. Since 1993, he is also a Senior Researcher with the Pattern and Image Analysis group of the Instituto de Telecomunicações, which is a private non-profit research institution. His research interests include inverse problems, signal and image processing, pattern recognition, optimization, and remote sensing. Dr. Bioucas-Dias was an Associate Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS (1997-2000) and he is an Associate Editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING and IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. He was a Guest Editor of IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING for the Special Issue on Spectral Unmixing of Remotely Sensed Data and of IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING for the Special Issue on Hyperspectral Image and Signal Processing and he a Guest Editor of IEEE SIGNAL PROCESSING MAGAZINE for the Special Issue on Signal and Image Processing in Hyperspectral Remote Sensing, He was the General Co-Chair of the 3rd IEEE GRSS Workshop on Hyperspectral Image and Signal Processing, Evolution in Remote sensing (WHISPERS'2011) and has been a member of program/technical committees of several international conferences.

Carme Julià Ferré Carme Julià Ferré received the BSc degree in Mathematics in 2002 from the Polytechnic University of Catalonia, Barcelona, Spain. In 2008, she got the PhD degree in Computer Science in the Computer Vision Center (CVC), at the Universitat Autnoma de Barcelona. She is currently collaborating with the Advanced Driver Assistance Systems group (ADAS), at the CVC. Since September 2008, she is working as a member of the Intelligent Robotics and Computer Vision group, at the Universitat Rovira i Virgili (URV), Tarragona. She is a lecturer at the Computer Science and Mathematics Department, at the URV. Her research interest is focused on applying matrix factorization techniques to different low-rank problems, such as rigid and non-rigid tructure from motion (SfM), photometric stereo and recommender systems.