INTER-SENSOR REGRESSION ANALYSIS FOR OPERATIONAL SENTINEL-2 AND SENTINEL-3 DATA PRODUCTS

Juan M. Haut, *Student Member, IEEE,* Ruben Fernandez-Beltran, Mercedes E. Paoletti, *Student Member, IEEE,* Javier Plaza, *Senior Member, IEEE,* Antonio Plaza, *Fellow, IEEE,* and Filiberto Pla

***Abstract*— The relatively recent availability of operational products from Sentinel-2 and Sentinel-3 missions gives widespread opportunities to combine data collected from different sensors in order to provide products of a higher processing level. Nonetheless, the availability of these products may be affected by multiple factors, such as cloud occlusions, band saturation, geolocation errors or even misaligned detectors. All these anomalies affecting remote sensing data may eventually limit the accessibility to fused products because some of the required information may become partially unavailable for specific areas of interest. In this scenario, the work presented here aims at analyzing the effectiveness of several state-of-the-art regression models in order to restore Sentinel-3 products with partial anomalies from Sentinel-2 integral data. In particular this work investigates three regression methods, two linear-regression method and a non-linear artificial neural networks based method. Obtained results prove that the non-linear approach and linear RIDGE method are able to carry out a good estimation of S3 from S2 data.**

***Index Terms*— Remote Sensing, Sentinel-2 (S2), Sentinel-3 (S3), Product Restoration, Data Regression**

I. INTRODUCTION

The European Space Agency (ESA) has developed a new family of Earth observation missions, called Sentinels, specially designed to provide operational products for the Copernicus program [1]. Among all the six planned missions, Sentinel-2 (S2) and Sentinel-3 (S3) have an especial synergy to generate image fusion products due to the fact that both missions are focused on the global monitoring of terrestrial surfaces by means of mid-resolution and high-resolution Multi-Spectral (MS) optical imagery [2], [3]. On the one hand, S2 mission comprises two identical satellites, both currently under operation, that incorporate a Multi-Spectral Instrument (MSI) which provides 13 spectral bands ranging from the visible and near infrared (VNIR) to the shortwave infrared (SWIR). On the other hand, S3 mission, which only has one pre-operational satellite at present, is able to provide data products with 21 VNIR spectral bands captured by the Ocean and Land Colour Instrument (OLCI).

In addition to the existing radiometric differences, both MSI and OLCI instruments offer a remarkable spatial resolution diversity. Whereas the former provides a spatial resolution ranging from 10 mpp (meters per pixel) to 60 mpp, OLCI's bands are acquired at a 300 mpp spatial resolution. Precisely, these differences motivate the constitution of high-level fusion products useful for multiple remote sensing applications, such as topographic mapping, land use classification, agriculture and forestry, flood and ice monitoring, geology, etc.

Regarding the missions' data availability, ESA has recently launched the Copernicus Open Access Hub [4], previously known as Sentinels Scientific Data Hub, in order to provide complete, free and open access to S2 operational data and S3 pre-operational land products. Even though S3 related products are still in a pre-operational phase, this data access point brings an excellent opportunity for the scientific community to start developing new fusion models and methodologies to deal with future challenges. Nonetheless, the early stage of the S3 infrastructure occasionally limits the process of collecting coupled S2 and S3 data over specific areas of interest. For instance, filtering products by cloud coverage is one of the main limitations of the pre-operational hub. That is, S3 products cannot be still filtered by the percentage of cloud coverage what makes difficult to find S3 images without clouds given a cloud free S2 product.

Additionally, there are many other factors affecting pre-operational products that may generate S3 data partially unavailable, such as band saturation, partial data degradation, geolocation errors or misaligned detectors. Even though all these anomalies can occur in S2 products as well, they are more likely to happen in a pre-operational scenario like in the S3 case. In fact, it is easy to check this having a look at the Copernicus Open Access Hub news [4]. During the last month of the year 2017, the number of anomalies reported by the two S2 operational satellites has been only one while the single S3 pre-operational satellite has reported a total of four anomalies.

In this scenario, this paper is concerned about restoring

This work has been supported by Ministerio de Educación (Resolución de 26 de diciembre de 2014 y de 19 de noviembre de 2015, de la Secretaría de Estado de Educación, Formación Profesional y Universidades, por la que se convocan ayudas para la formación de profesorado universitario, de los subprogramas de Formación y de Movilidad incluidos en el Programa Estatal de Promoción del Talento y su Empleabilidad, en el marco del Plan Estatal de Investigación Científica y Técnica y de Innovación 2013-2016. This work has also been supported by Junta de Extremadura (decreto 297/2014, ayudas para la realización de actividades de investigación y desarrollo tecnológico, de divulgación y de transferencia de conocimiento por los Grupos de Investigación de Extremadura, Ref. GR15005). This work has been additionally supported by the Generalitat Valenciana through the contract APOSTD/2017/007 and by the Spanish Ministry of Economy under the project ESP2016-79503-C2-2-P.

J. M. Haut, M. E. Paoletti, J. Plaza and A. Plaza are with the Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, Escuela Politécnica, University of Extremadura, PC-10003 Cáceres, Spain.(e-mail:juanmariohaut@unex.es;mpaolett@unex.es;jplaza@unex.es; aplaza@unex.es).

R. Fernandez-Beltran and F.Pla are with the Institute of New Imaging Technologies, University Jaume I, 12071 Castellón, Spain. (e-mail: (rufernan@uji.es; pla@uji.es).

partially either degraded or unavailable S3 data by means of state-of-the-art regression models. Different sorts of regression algorithms have been widely used in the general purpose remote sensing literature [5]. Cloud removal [6], remote sensing data coding [7], air pollution mapping [8] and remote sensing image classification [9] are some of the applications where different regression models have been successfully used. However, there are few research work done within the specific context of the S2 and S3 missions and this is precisely the gap that motivates this work.

Some quality works in the literature, such as [10], review several state-of-the-art machine learning regression algorithms for S2 and S3, nonetheless the evaluation is exclusively conducted from a biophysical parameter retrieval perspective and besides they only consider simulated data. This work pursues a more general objective where the regression assessment is focused on restoring actual S3 data by integral S2 operational products in order to alleviate some of the current S3 pre-operational limitations. That is, the experimental study presented in this work aims at improving the availability of coupled S2 and S3 real data useful to develop further fusion models.

First, we review in sec. II the three regression models considered in this experimental study, two of them linear and one non-linear, indicating the employed methodology: the widely used Principal Component Analysis (PCA), the RIDGE regression algorithm and the Extreme Learning Machine (ELM). Sec. III introduces and releases a new small-scale test database, called ESA1, which consists of two coupled S2/S3 data products belonging to the Sierra de Andújar Natural Park in southern Spain. Then, we conduct an experimental comparison using the aforementioned state-of-the-art regression algorithms to restore partially unavailable S3 data from clean S2 products of the same area. Experiments reveal that linear RIDGE and non-linear ELM methods perform the best regression task from S2 data to the corresponding S3 pixels.

II. Methodology

This section describes all proposed methods to perform the regression from S2 to S3. In particular, the principal component analysis (PCA), the RIDGE regression algorithm and the extreme learning machine (ELM) artificial neural network. In all cases, the data is scaled in the range $[0, 1]$, being 85% the training percentage. Input data $\mathbf{x}^{(i)} \in \mathbb{R}^{1\times 2925}$ is obtained from S2 images as patches of $15 \times 15 \times 13$, and reshaped to $1 \times (15 \cdot 15 \cdot 13) = 1 \times 2925$ feature vectors, in order to feed the PCA, RIDGE and ELM methods. The desired output is the corresponding pixel in S3 $\mathbf{y}^{(i)} \in \mathbb{R}^{1\times 21}$, whose spectrum is extracted to form the corresponding feature vector, such as $\{\mathbf{x}^{(i)}, \mathbf{y}^{(i)}\}_{i=1}^{N}$.

A. Principal Component Analysis

Traditionally, PCA [11] has been widely used as statistical linear technique to unsupervised data dimensionality reduction, where a set of N observations $X \in \mathbb{R}^{d\times N} = [\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, ..., \mathbf{x}^{(N)}]$, composed by feature vectors with the form $\mathbf{x}^{(i)} \in \mathbb{R}^{d} = [x_1^{(i)}, x_2^{(i)}, ..., x_d^{(i)}]$, is reduce down to its g basic components, being $g < d$, as $X^{(new)} \in \mathbb{R}^{g\times N} = [\mathbf{x}^{(new_1)}, \mathbf{x}^{(new_2)}, ..., \mathbf{x}^{(new_N)}]$ applying an orthogonal transformation $X^{(new)} = WX$, being $W \in \mathbb{R}^{g\times d}$ the projection matrix created by the g selected eigenvectors of the input data covariance matrix. The PCA's reconstruction error is defined by:

$$\min_{W} \parallel W^T(WX) - X \parallel_2 \tag{1}$$

where $W^T(WX)$ is the inverse process to the reduction one. In this work, eq. 1 has been changed in order to feed PCA with patches of S2 $X \in \mathbb{R}^{(1\times 2925)\times N}$, being the corresponding $Y \in \mathbb{R}^{(1\times 21)\times N}$ S3 pixels the target output:

$$\min_{W} \parallel W^T(WX) - X \parallel_2 + \parallel WX - Y \parallel_2 \tag{2}$$

B. RIDGE regression

RIDGE regression is one of the most commonly used linear regression algorithm to approximate the system equation $XZ = Y$. Given N observations $X \in \mathbb{R}^{N\times d} = [\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, ..., \mathbf{x}^{(N)}]$ with the corresponding set of targets $Y \in \mathbb{R}^{N\times g} = [\mathbf{y}^{(1)}, \mathbf{y}^{(2)}, ..., \mathbf{y}^{(N)}]$, RIDGE tries to find Z that minimize the cost function 3:

$$\min_{Z} \parallel XZ - Y \parallel_2 + \parallel \lambda X \parallel_2 \tag{3}$$

where $\lambda = \alpha I$ is the Tikhonov regularization matrix defined as a multiple of the identity matrix that prevents overfitting and underfitting, being α a user-defined penalizing term. Solving eq. 3, Z is defined as $Z = (X^TX + \lambda^T\lambda)^{-1}X^TY$. In our case, X are reshaped patches from S2 and Y corresponding pixels in S3. Cholesky solver has been used in order to obtain a closed-form solution.

C. Extreme Learning Machine

The Extreme Learning Machines (ELM) is a learning algorithm that exploits artificial neural networks (ANNs) to perform data analysis. Suppose a dataset of feature vectors $X \in \mathbb{R}^{N\times d} = [\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, ..., \mathbf{x}^{(N)}]$ and the corresponding targets $Y \in \mathbb{R}^{N\times g} = [\mathbf{y}^{(1)}, \mathbf{y}^{(2)}, ..., \mathbf{y}^{(N)}]$, so $\{\mathbf{x}^{(i)}, \mathbf{y}^{(i)}\}_{i=1}^{N}$. Based on single-hidden layer feed-forward neural networks (SLFNs) with topology $d - L - g$ (see Fig. 1), where L is the number of hidden neurons, the ELM model maps input data X to the L-dimensional hidden layer random feature space as $H(X, W, B) = \phi(XW + B)$, where $W \in \mathbb{R}^{d\times L}$ and $B\mathbb{R}^{L}$ are random weights and biases that have been generated based on a continuous sampling distribution probability and connect the input layer nodes with the hidden layer ones, $H \in \mathbb{R}^{N\times L}$ is the hidden layer's output matrix and $\phi(\cdot)$ is an activation function (e.g. sigmoid, ReLU, tanh...). Knowing that the network output if defined by $Y \simeq f(X) = H\beta$, where $f(X)$ is the network output (desirably similar to Y) and $\beta \in \mathbb{R}^{L\times g}$ is the weights between the hidden and the output nodes, the goal of the ELM is to apply the linear system:

$$H^{\dagger}Y = \beta \tag{4}$$

(being $H^{\dagger} \in \mathbb{R}^{L\times N}$ the Moore-Penrose generalized inverse matrix of H) that calculates the connection weights β

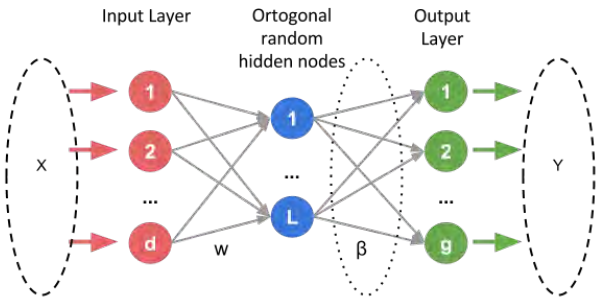

Fig. 1. Extreme learning machine scheme

Fig. 2. Sentinel-2 product.

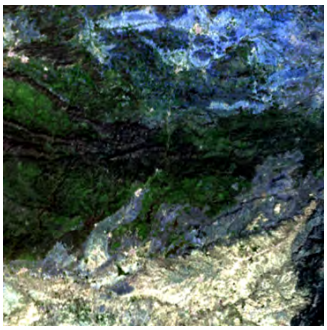

Fig. 3. Sentinel-3 product.

between the hidden and the output nodes of the net that best approximate the output of the hidden layer H to the desired output Y, while exhibiting the minor norm at the same time and minimizing the function cost $\min_{W,B,\beta} \parallel H(W,X,B)\beta - Y \parallel_2$. In our case, the classical ELM model has been adapted in order to accept multispectral patches of S2 pixels as input data $\mathbf{x}^{(i)} \in \mathbb{R}^{1\times 2925}$, being S3 pixels the target output $\mathbf{y}^{(i)} \in \mathbb{R}^{1\times 21}$. Also, eq. 4 has been changed by:

$$\beta = \left(\frac{I}{C} + H^T H\right)^{-1} H^T Y \tag{5}$$

being $\left(H^T H\right)^{-1} H^T = H^{\dagger}$ the Moore-Penrose inverse and C a regularization term [12]. The chosen topology $d - L - g$ contains one input layer, being $d = 1 \times (15 * 15 * 13)$, one hidden layer, with $L = 2000$ hidden nodes, and one output layer with $g = 1 \times 13$ nodes. On the other hand, the training process starts by feeding the entire network with the scaled training sample $\{\mathbf{x}^{(i)}\}_{i=1}^{N}$ and calculating the final weights β. The error incurred by the network during the training phase is calculated by Mean Square Error $MSE = \frac{1}{N}\sum_{i=1}^{N}\left(f\left(\mathbf{x}^{(i)}\right) - \mathbf{y}^{(i)}\right)^2$ where $\mathbf{y}^{(i)}$ is the desired output, $f\left(\mathbf{x}^{(i)}\right)$ is the network output and N the number of training samples. Once the network is trained, the test set is presented to the network. Also, the error incurred by the network during the test phase is calculated by MSE.

III. Results

A. Experimental environment and datasets

Our experiments have been conducted on a hardware environment composed by a 6th Generation Intel® Core™i7-6700K processor with 8M of Cache and up to 4.20GHz (4 cores/8 way multitask processing), 40GB of DDR4 RAM with a serial speed of 2400MHz, a GPU NVIDIA GeForce GTX 1080 with 8GB GDDR5X of video memory and 10Gbps of memory frequency, a Toshiba DT01ACA HDD with 7200RPM and 2TB of capacity, and an ASUS Z170 pro-gaming motherboard. On the other hand, the software environment is composed by Ubuntu 16.04.4 x64 as operating system.

The dataset considered in this work, called ESA1, consists of two coupled S2 and S3 reflectance products belonging to the Sierra de Andújar Natural Park in southern Spain. Figures 2-3 show a color visualization of the data.

These two products, which were acquired on March 10 2017, have been selected because of the special interest of this area for the study of some biophysical parameters and besides there are not anomalies present in this data. Both S2 and S3 products, downloaded from the Copernicus Open Access Hub, have been appropriately processed using the Sentinel Application Platform (SNAP) [4] in order generate compatible images. That is, the S2 product has been initially re-sampled to a uniform spatial resolution of 20 mpp. Then, the S3 image has been projected onto the S2 grid and also cropped to the corresponding S2 coverage area. Eventually, each S3 pixel with a size of (1×21) corresponds to a voxel $(15 \times 15 \times 13)$ in S2.

B. Performance evaluation

In order to perform a quantitative evaluation of the results, two reference metrics are used, the Spectral Angle Mapper (SAM) and the Mean Squared Error (MSE). The SAM metric (Eq. (6)) considers each spectral band as a coordinate axis and then it computes the average angle between the reconstructed S3 image (R) and the ground-truth ones (G).

$$\mathrm{SAM}(\mathbf{R},\mathbf{G}) = \frac{1}{N}\sum_{i}^{N}\arccos\frac{\boldsymbol{R_i}\cdot\boldsymbol{G_i}}{||\boldsymbol{R_i}||\ ||\boldsymbol{G_i}||}. \tag{6}$$

Note that the i subindex is used to denote spectral pixels and N represents the total number of pixels per band. Regarding the MSE index, this metric measures absolute differences between both the reconstructed (R) and the ground-truth (G) images as Eq. (7) shows,

$$\mathrm{MSE}(\mathbf{R},\mathbf{G}) = \frac{1}{N\cdot M}\sum_{i}^{N}\sum_{j}^{M}(R_{i,j} - G_{i,j})^2, \tag{7}$$

where M represents the number of spectral bands and the i,j index is used to denote a specific pixel within the j band.

Table I shows the obtained MSE and SAM results for each method, where each experiment has been repeated 10 times in order to measure the robustness and stability of the proposed regression methods. As we can observe, non-linear ELM model and linear RIDGE regression method are able to reach the best MSE and SAM values, i.e. they are the best

method to restore S3 data by integral S2, being the PCA the worst regression method.

Algorithm	MSE	SAM
PCA	0.3168 (0.0006)	1.5898 (0.0066)
ELM	0.0055 (8e-06)	0.0969 (0.0016)
Ridge	0.0053 (5e-06)	0.0918 (0.0006)

TABLE I

REACHED MSE AND SAM BETWEEN ORIGINAL S3 AND PREDICTED S3. EACH EXPERIMENT HAS BEEN EXECUTED 10 TIMES AND THE AVERAGE RESULTS AND THE STANDARD DEVIATION (IN PARENTHESIS) ARE REPORTED.

IV. CONCLUSIONS

In this work, we have presented an experimental study to assess the performance of three different regression algorithms, PCA, ELM and Ridge, in order to restore S3 data from S2 operational products. The conducted experiments reveal that both ELM and Ridge are able to generate an accurate result whereas the PCA approach is unable to successfully restore S3 data.

One of the main conclusions that arises from this work is the feasibility of restoring S3 spectral data by taking advantage of the higher spatial resolution of S2. That is, the spatial patterns captured by the S2 MSI sensor are useful to identify spectral signatures over S3 OLCI instrument which is, precisely, the motivation behind many data fusion algorithms.

Regarding the performance of the tested methods, PCA has obtained an utterly non-competitive result due to the fact that the spatial variability captured by the eigenvectors uncovered from S2 data is not focused on replicating the spectral patterns which are observable through the S3 OLCI sensor. Nonetheless, the ELM and Ridge regression algorithms have shown to restore the S3 spectra with a high level of accuracy in terms of spatial and spectral distortions.

As future research lines, we will perform an evaluation of regression methods that employ both hyperspectral characteristics, spectral and spatial feature information at the same time, in order to minimize the MSE and SAM measurements by adding spatial information. In addition, we will also study the estimation of different biophysical parameters, such as Leaf Area Index or Chlorophyll content, over the restored S3 products.

REFERENCES

[1] Josef Aschbacher and Maria Pilar Milagro-Perez, "The european earth monitoring (gmes) programme: Status and perspectives," *Remote Sensing of Environment*, vol. 120, no. Supplement C, pp. 3–8, 2012.

[2] M. Drusch, U. Del Bello, S. Carlier, O. Colin, V. Fernandez, F. Gascon, B. Hoersch, C. Isola, P. Laberinti, P. Martimort, A. Meygret, F. Spoto, O. Sy, F. Marchese, and P. Bargellini, "Sentinel-2: Esa's optical high-resolution mission for gmes operational services," *Remote Sensing of Environment*, vol. 120, no. Supplement C, pp. 25 – 36, 2012, The Sentinel Missions - New Opportunities for Science.

[3] C. Donlon, B. Berruti, A. Buongiorno, M.-H. Ferreira, P. Fmnias, J. Frerick, P. Goryl, U. Klein, H. Laur, C. Mavrocordatos, J. Nieke, H. Rebhan, B. Seitz, J. Stroede, and R. Sciarra, "The global monitoring for environment and security (gmes) sentinel-3 mission," *Remote Sensing of Environment*, vol. 120, no. Supplement C, pp. 37 – 57, 2012.

[4] "Copernicus open access hub," 2017, Accessed: 2017-12-29.

[5] Gustavo Camps-Valls, Devis Tuia, Luis Gomez-Chova, and Sandra Jimenez, *Remote Sensing Image Processing*, Morgan & Claypool, 2011.

[6] G. Hu, X. Sun, D. Liang, and Y. Sun, "Cloud removal of remote sensing image based on multi-output support vector regression," *Journal of Systems Engineering and Electronics*, vol. 25, no. 6, pp. 1082–1088, 2014.

[7] N. Amrani, J. Serra-Sagrista, V. Laparra, M. W. Marcellin, and J. Malo, "Regression wavelet analysis for lossless coding of remote-sensing data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 9, pp. 5616–5627, 2016.

[8] B. Zou, Q. Pu, M. Bilal, Q. Weng, L. Zhai, and J. E. Nichol, "High-resolution satellite mapping of fine particulates based on geographically weighted regression," *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 4, pp. 495–499, 2016.

[9] R. Hang, Q. Liu, H. Song, Y. Sun, F. Zhu, and H. Pei, "Graph regularized nonlinear ridge regression for remote sensing data analysis," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 1, pp. 277–285, 2017.

[10] Jochem Verrelst, Jordi Muoz, Luis Alonso, Jess Delegido, Juan Pablo Rivera, Gustavo Camps-Valls, and Jos Moreno, "Machine learning regression algorithms for biophysical parameter retrieval: Opportunities for sentinel-2 and -3," *Remote Sensing of Environment*, vol. 118, no. Supplement C, pp. 127 – 139, 2012.

[11] Svante Wold, Kim Esbensen, and Paul Geladi, "Principal Component Analysis," *Chemometrics and Intelligent Laboratory System*, vol. 2, no. 1, pp. 37–52, 1987.

[12] Guang-Bin Huang, Hongming Zhou, Xiaojian Ding, and Rui Zhang, "Extreme Learning Machine for Regression and Multiclass Classification," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 42, no. 2, pp. 513–529, 2012.