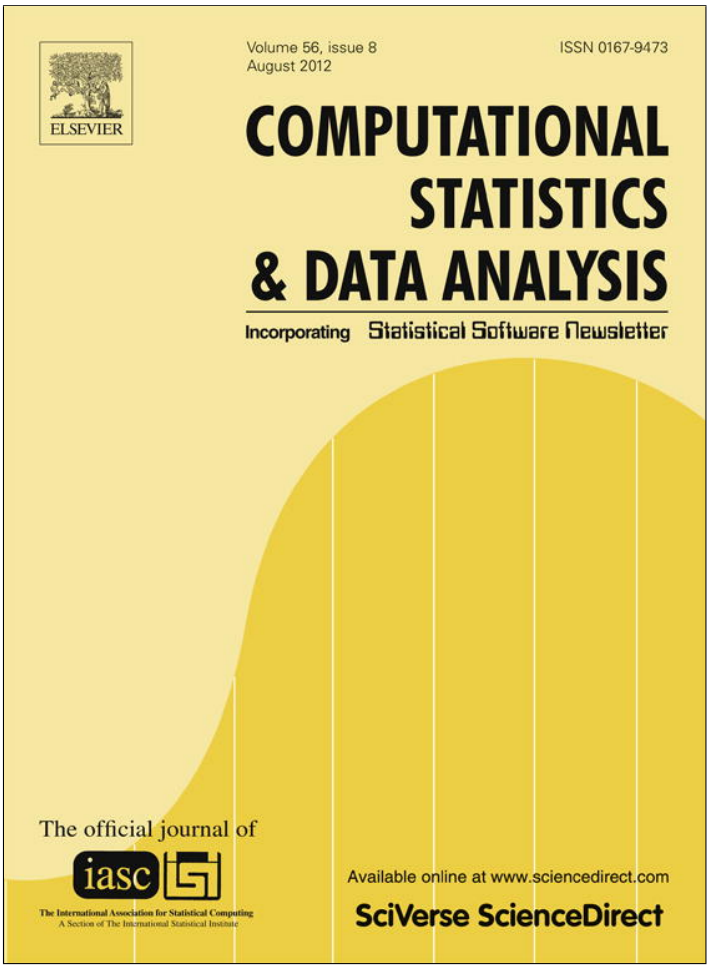

Computational Statistics and Data Analysis 56 (2012) 2388–2403

Contents lists available at SciVerse ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Large gap imputation in remote sensed imagery of the environment

Valeria Rulloni, Oscar Bustos, Ana Georgina Flesia *

CIEM-Conicet and FaMAF-UNC, 5000, Córdoba, Argentina

ARTICLE INFO

Article history:
Received 1 July 2010
Received in revised form 13 February 2012
Accepted 15 February 2012
Available online 24 February 2012

Keywords:
Imputation
Remote sensed imagery
Filling gaps

ABSTRACT

Imputation of missing data in large regions of satellite imagery is necessary when the acquired image has been damaged by shadows due to clouds, or information gaps produced by sensor failure.

The general approach for imputation of missing data, which could not be considered missed at random, suggests the use of other available data. Previous work, like local linear histogram matching, takes advantage of a co-registered older image obtained by the same sensor, yielding good results in filling homogeneous regions, but poor results if the scenes being combined have radical differences in target radiance due, for example, to the presence of sun glint or snow.

This study proposes three different alternatives for filling the data gaps. The first two involves merging radiometric information from a lower resolution image acquired at the same time, in the Fourier domain (Method A), and using linear regression (Method B). The third method considers segmentation as the main target of processing, and proposes a method to fill the gaps in the map of classes, avoiding direct imputation (Method C).

All the methods were compared by means of a large simulation study, evaluating performance with a multivariate response vector with four measures: Q, RMSE, Kappa and Overall Accuracy coefficients. Differences in performance were tested with a MANOVA mixed model design with two main effects, imputation method and type of lower resolution extra data, and a blocking third factor with a nested sub-factor, introduced by the real Landsat image and the sub-images that were used. Method B proved to be the best for all criteria.

1. Introduction

Management of the environment and inventory of natural resources often requires appropriate remote sensing data acquired at specific times on Earth locations. But quite often, good resolution optical images have large damaged areas with lost information due to clouds or shadows produced by clouds. Also, long term projects requiring images acquired regularly have reported problems in data collection due to sporadic sensor failure. A major missing data problem occurred in 2003, when the Landsat 7 satellite had a problem, not with its sensor, but with one of the mirrors that helped image construction, leaving almost 23% of the image as missing data.

The general statistical approaches for imputation usually consider the data loss in one of the three categories: missing at random data, completely missing at random data, (meaning that the missing data is independent of its value), and non-ignorable missingness, Allison (2000). The last case is the most problematic form, which exists when missing values are not randomly distributed across observations, but when the probability of missingness cannot be predicted from the variables in the model. One approach for dealing with non-ignorable missingness is to impute values based on data otherwise external

* Corresponding author. Tel.: +54 351 4334051; fax: +54 351 4334055.
E-mail addresses: vrulloni@famaf.unc.edu.ar (V. Rulloni), bustos@famaf.unc.edu.ar (O. Bustos), flesia@famaf.unc.edu.ar, gflesia@gmail.com (A.G. Flesia).

0167-9473/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2012.02.022

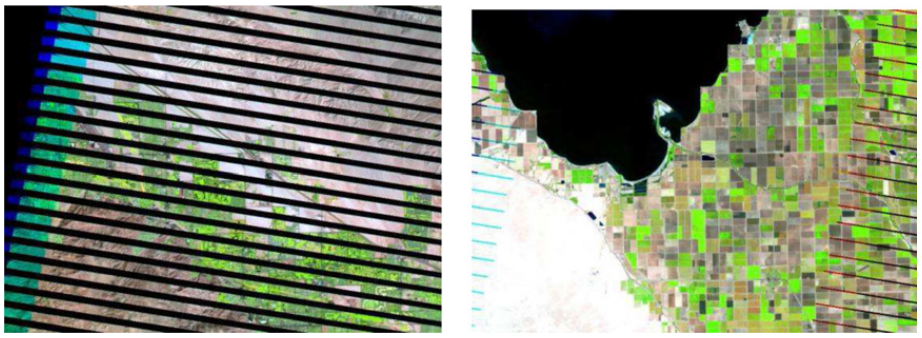

Fig. 1. Left panel, a piece of a Landsat 7 image with lost information. Right panel, a center piece of the same image almost without loss.

to the research design, as older images. Little and Rubin (2002) reported that when a problem in collection of census data occurred, or large regional surveys, older databases could be merged to solve the problem, without deleting the incomplete entries.

In the classical literature of cloud removing, existence of extra information has been a common hypothesis of multitemporal measurement based techniques. These approaches consist of selecting the best measurement (the most cloud-free pixel) among a set of measurements acquired over a limited time period to represent the considered multitemporal pixel over that time period. Examples of these approaches are the recent work of Melgani (2006), Melgani et al. (2007), and Benabdelkader and Melgani (2008). The major limitation of this approach is that the acquisition dates should be close to each other, and the spatial dynamics of the geographical area under analysis should be low compared to the total time interval of the sequence. Extensive fires usually cause major changes in large regions, and dense smoke clouds, so data information cannot be recovered or imputed using old (even recent) imagery.

A second category of cloud-removal techniques are the ones that attempt to fill in the cloud-contaminated region using traditional synthesis and image inpainting techniques. Image inpainting attempts to keep the continuity of the curves of constant intensity on the image, named isophotes. Many existing methods were designed to meet that requirement, but suffered from an over-smoothing effect, especially when the regions to be filled are thick. Most of Partial Differential Equation-based methods use 2D anisotropic diffusion, Bertalmío et al. (2000), Esedoglu and Shen (2002), Bugeau et al. (2010). The Fast Marching method, Telea (2004), uses a local 2D average of pixels. Thus, there exists, though in small quantities, a low-pass filtering effect orthogonal to the isophotes. This effect is negligible for filling thin regions, but for thick regions, the smoothing across isophotes is propagated and enlarged to an unacceptable degree. Texture must be estimated from the surrounded region of the gap, in order to reduce the smoothing effect, Bugeau and Bertalmío (2009).

A third category of cloud removal methods involves data fusion techniques. Merging information acquired at the same time from two or more sensors (with possible different resolutions) is the core of data fusion techniques. Their historical goals are to increase image resolution, sharpening or enhancement of the output, Tsuda et al. (2003), Ling et al. (2007) and Pohl (1998), and their major problem to overcome is the co-registration of the different sources to merge, Blum and Zheng (2005). In Le Hegarat-Mascle et al. (1998), contextual modeling of information was introduced in order to merge data from SAR (Synthetic Aperture Radar) images into optical images, to correct dark pixels due to clouds. Arellano (2003) used Wavelet transforms first for automatic cloud detection, and secondly to correct the information of the located clouds pixels by merging older image information with a wavelet multiresolution fusion technique. Rossi et al. (1994), introduced a spatial technique, kriging interpolation, for correction of shadows due to clouds. Shadows and light clouds do not destroy all information, only distort it, but smoke clouds or dark rain clouds produce non-ignorable missingness, a harder problem for all interpolation techniques if the region is thick and large, Zhou et al. (2006), Maalouf et al. (2009).

A singular example of incomplete data is the damaged imagery provided by the Landsat 7 ETM+ satellite after its failure on May 2003. A failure of the Scan Line Corrector, a mirror that compensates for the forward motion of the satellite during data acquisition, resulted in overlaps of some parts of the images acquired thereafter, leaving large gaps, as large as 14 pixels, in others. About 22% of the total image is missing data in each scene. In Fig. 1 we can see two parts of the same Landsat 7 image, one with missing information and another almost without loss.

Sensor failure is usually an sporadic cause of missing data, producing random black pixels or an specific event of partial data. But in the Landsat 7 case, all data after the failure was damaged, producing several years of partial information in the historical Earth monitoring projects.

The USGS, NASA and Landsat 7 Team (2003), developed and tested composite products, made with the damaged Landsat 7 images and a database of older Landsat images, using local linear histogram matching, a method that merges pieces of older images that best matches the surroundings of the missing information region, Scaramuzza et al. (2004), USGS/NASA 2004, (United States Geological Survey/National Aeronautics and Space Administration). Commonwealth of Australia (2006) reported that the composite products appear similar in quality to nondamaged Landsat 7 images, but masked environment changes, an usual problem in data fusion when one of the sources is temporally inaccurate.

Zhang et al. (2007) developed a spatial interpolation method based on kriging, the kriging geostatistical technique, for filling the data gaps in Landsat 7 ETM+ imagery, without the need of extra information. They compare their method with

the standard local linear histogram matching technique chosen by USGS/NASA. They show in their case study that kriging is better than histogram matching in targets with radical differences in radiance, since it produces an imputation value closer to the values of the actual neighbors. A drawback of spatial techniques is that they rely on a neighborhood that could be completely empty of real information. The algorithms start imputation in the gap contour, where there are many good pixels, and use imputed and good values to create the next pixel value. In center pixels of large gaps, algorithms based only on damaged images will use previously imputed values to generate the next imputation, degrading visual quality and increasing interpolation error. Modeling with other types of spatial models, like Conditionally Autoregressive (CAR) Models, and Simultaneous Autoregressive (SAR) Models, is a more sophisticated way of generating imputations in small gaps, as model predictions, when the parameters of the model were estimated with the surrounded regions of the gap. Large regions as the ones we are considering in this paper, at first glance, will be imputed with information suffering from the same problems that inpainting methods have, texture will be correctly inferred if the region is homogeneous, but texture changes will not be correctly determined, a situation that will invariably occur when the gap is large. To our knowledge, the influence of the gap size in imputation methods has not been studied in any area of image processing, not only on remote sensing image processing works. Ojeda (2011) is currently studying SAR models robust parameter estimation and modeling for several applications, including imputation, but not in the framework of extra data availability.

In this paper, we propose three methods based on data fusion techniques for imputation of missing values in images with non-ignorable missingness on large gaps. We suppose there is availability of temporally accurate extra information for the gap scenes, produced by a lower resolution sensor. This is not a very restrictive hypothesis, since there are many satellite constellations that can provide temporally accurate data with different sensors at different resolutions.

This study proposes three different alternatives for filling the data gaps. The first method involves merging information from a co-registered older image and a lower resolution image acquired at the same time, in the Fourier domain (Method A). The second used a linear regression model with the lower resolution image as extra data (Method B). The third method considers segmentation as the main target of processing, and proposes a method to fill the gaps in the map of classes, avoiding direct imputation for the segmentation task (Method C). Radiometric imputation is later made assigning a random value from the convex hull made by the neighbor pixels within its class.

All the methods were compared by means of a large simulation study, evaluating performance with a multivariate response vector with four measures: Q, RMSE, Kappa and Overall Accuracy coefficients. Two of these performance measures (Kappa and Overall Accuracy) are designed for the assessment of segmentation accuracy. The other two, Q and RMSE, measure radiometric interpolation accuracy. Difference in performance were tested with a MANOVA mixed model design with two main effects, imputation method and type of lower resolution extra data, and a blocking third factor with a nested sub-factor, introduced by the real Landsat image and the sub-images that were used. Method B proved to be the best for all criteria.

A discussion section will provide a closure to the paper, in which we will elaborate possible extensions of the methods and future work.

2. Methods

We are considering images acquired from the same geographic target, with fixed number of bands K, and 256 levels of gray per band. Damaged and older images have the same support, lower resolution images have different supports, and we suppose that there are $n_z \times n_z$ pixels in the damaged image for each pixel in the lower resolution image.

Let X_D be the damaged image, Z a lower spatial resolution image, and X_{old}, an older image acquired with the same sensor as X_D. Let S_D be the gap to be filled, i.e. the set of pixels of X_D with missing values. The goal is to input values on S_D using available good data from X_D in a small neighborhood of each pixel, the values of Z and eventually, the values of X_{old}.

The first step in processing is the re-sampling of the lower resolution image in order to match the support of the three images. This is done replacing each pixel (c, r) of Z by a matrix of size $n_z \times n_z$ with constant value $Z(c, r)$.

Imputation is done in each band separately with the same algorithm, thus the methods descriptions consider the images as one band only.

2.1. Method A

The image X_{old} has been co-registered with reference to the damaged image X_D, despite they have been acquired by the same sensor. Calibration is also necessary to increase merging accuracy.

We suppose a Gaussian calibration has been made on X_{old} per column, taking the values in the matched column of X_D as pivots. Let X_R^A be the composite image, output from the imputation method A. We define X_R^A values in the gap as a mixture of $(1 - C)100\%$ of the high frequencies of the older image and $C100\%$ of the low frequencies taken from the actual but lower resolution image.

$$X_R^A(c, r) = \begin{cases} |L_C(Z)(c, r) + H_C(X_{old})(c, r)| & (c, r) \in S_D \\ X_D(c, r) & \text{otherwise} \end{cases}$$

with H_C and L_C the Ideal high pass and low pass Fourier filters.

High pass filters are related to details, edges and random noise, and low pass filters to structural information. Therefore, Method A takes structural information from the low resolution image and details from the older one.

The C value is considered a constant that must be fixed by the practitioner. Most fusion methods have this type of constant values that are not tuned up precisely, since they are not parameters to be estimated, but are part of the design of the model. In our experiments, we have chosen three main values for C, "low" 20%, "equal" 50% and "high" 80%, in order to have a feedback of the influence of the choice of the constant in the imputation when real Landsat data is considered. The first one preserves texture at expense of losing actual information; the last one, on the contrary, degrades texture, but preserves possible changes that should appear in the damaged image. Finally, the second one represents a balance between the two possibilities. The users should be able to decide if their applications need better texture in the imputed gap, or to detect a change in the environment, as a line of flood or a fire consumed area. Subtle changes in C produce no influence in the output of the method.

2.2. Method B

Imputation will be made with the only help of Z, a temporally accurate image with lower resolution. We have expanded the lower resolution image to match the support of X_D, replacing each cross grained pixel by a block of $n_z \times n_z$ pixels with the same radiometric value. Each one of the constant blocks of the expanded Z image has a matched block in the damaged image X_D that could be valid (having all the information), or nonvalid, (with some loss). We denote B^v the set of valid blocks.

We will follow a time series approach now. Let us think we have data collected month to month along several years, and we have a missing January data. It is reasonably to impute that value using a regression model that only involves other January data, and extra data collected on Summer that year.

In our case, we have missing information on a location (c, r) inside a block B. We may impute that value using a regression model that only involves data in the same position inside the other blocks of the image (January data), the values of Z (Summer data) as regressors.

$$X_D(c, r, B^v) = \alpha_{(c,r)} * Z(c, r, B^v) + \beta_{(c,r)} + \varepsilon(c, r),$$

with $\varepsilon(c, r) \sim N(0, \sigma(c, r)^2)$. We suppose that valid blocks are independent of each other.

The coefficients $\alpha_{(c,r)}$ and $\beta_{(c,r)}$ are estimated by ordinary least squares using only valid blocks. Then

$$\widehat{\alpha}(c, r) = \frac{\sum_{B \in B^v} (Z(c, r, B) - \overline{Z}(c, r))(X_D(c, r, B) - \overline{X_D(c, r)})}{\sum_{B \in B^v} (Z(c, r, B) - \overline{Z})^2},$$

$$\widehat{\beta}_{(c,r)} = \overline{X_D(c, r)} - \widehat{\alpha}_{(c,r)}\overline{Z},$$

where

$$\overline{Z}(c, r) = \sum_{B \in B^v} \frac{Z(c, r, B)}{|B^v|},$$

and

$$\overline{X_D(c, r)} = \sum_{B \in B^v} \frac{X_D(c, r, B)}{|B^v|}.$$

We define the imputed image X_R^B as X_D in the pixels with no missing information and the value predicted by the regression in each damaged pixel.

$$X_R^B((d, g, B)) = \begin{cases} \widehat{\alpha}(d, g) * Z(d, g, B) + \widehat{\beta}(d, g) & (d, g, B) \in S_D \\ X_D((d, g, B)) & \text{otherwise.} \end{cases}$$

This regression model is very simple, thus, easy to compute in a large sized image, as remote sensed imagery usually are. Data fusion has always the extra complications of image co-registration, so simple data functions minimize errors due to image mismatch. Including spatial correlation in this model, directly or in a hierarchical fashion, will automatically add mismatch issues between the two images bands. Also, independence is a very common assumption in image processing applications, like image segmentation, where Gaussian maximum likelihood methods are still the most applied, given their simplicity and low computational complexity. We will discuss extensions of this model in the Discussion Section.

2.3. Method C

Remote sensed images are used in a wide range of applied sciences. For many of them, as Agricultural and Experimental Biology, or Environmental Science, all useful information is contained on a class map, where the classes are characterized by special features under study. Kind of crop, forested or deforested areas, regions with high, medium and low soil humidity, are examples of such features.

In this section we will change our point of view by thinking on class maps instead of radiometric images. Let us suppose we have our damaged image X_D, and a temporally accurate image Z with lower resolution, and possible different number of radiometric bands. This is an advantage over radiometric imputation, which need the same spectral properties in both images.

We also suppose that a map of classes C_D has been drawn from the damaged image X_D, using a nonsupervised method like K means, with K different classes, or any other method that the practitioner may have confidence on. For example, if there are available accurate training samples drawn for each of the K classes, any supervised classification method could be used to generate the labels of the pixels that are not in the image gap. The class of missing information pixels would be another class, called class 0. The goal is to assign the pixels of class zero to one of the K classes, and generate a radiometric value for them randomly from a subset of the selected class.

Class zero imputation

We have a stable map of K classes C_D, with missing information in an extra class called class zero, and auxiliary information provided by a temporally accurate lower resolution radiometric image Z. If (c, r) is a pixel in class zero, it does not have radiometric information. We will impute a label class on it with the following algorithm

1. Let N_t be the smallest square neighborhood of (c, r) that have a pixel with nonzero label.
2. Let $\overline{Z}_k$ be the arithmetic mean of the Z pixels in class k that have positions in N_t.
3. The label of (c, r) is the label of the class k_0 that makes the smallest Euclidean distance from $Z(c, r)$ to each $\overline{Z}_k$.

$$k_0 = \arg\min_k \|Z(c, r) - \overline{Z}_k(c, r)\|_2.$$

The Euclidean distance has been chosen in this method for simplicity, because any other measure will certainly increase complexity in the final method. Another possibility would be the Mahalanobis distance, using the Correlation matrix computed between the bands of the Z image. Again, none of these methods includes correlation issues, between the different spectral bands or in between the images.

After this process, all pixels will have a class label, but pixels in the gaps will need to be imputed with radiometric information. A pixel's radiometric value will be assigned randomly from the values of the convex hull generated by the pixels of a small neighborhood within its class.

3. Precision assessment by simulation

Impartial imputation assessment is only possible when ground truth is available. Complete simulation of the three types of images involved in the methods (old, damaged and lower resolution) will introduce errors beyond the ones produced by the imputation methods, degrading the quality of the assessment. For this reason, good quality Landsat 7 ETM+ imagery that have older matched imagery available were selected, and strips similar to the gaps in the SLC-off ETM+ imagery were cut manually, guaranteeing ground truth to compare with and co-registration between the images. The four Landsat 7 ETM+ images selected were quite large, having many different textures in them, like crop fields, mountains and cities, which challenge the imputation methods differently.

We have chosen to mimic the damaged Landsat imagery because it is considered by the remote sensing community as extensive missing data. It is very difficult to say when a gap is small or large, or how difficult it is for a given algorithm to produce good imputation results in a large gap, since it depends on the amount of texture information that is lost in the gap, not exactly the size. In the Landsat case, 14 by 14 pixels is more or less 210 m by 210 m of ground truth lost, and the stripes of missing data cover around 20% of the image.

Landsat 7 ETM+ had a lower resolution sensor in its constellation, the MMRS camera from the Argentinean SAC C satellite, whose imagery could be used as extra data. Also, the Landsat's companion, Earth Observing-1, trails one minute following the exact orbital characteristics. But again, to control possible co-registration problems, lower resolution imagery were simulated with three resolution reduction methods (RRM), by block averaging the ground truth, the CONGRID method from ENVI Software (0000) and shifted block averaging. Block averaging is a crude way of reducing resolution, returning a coarse grained image. CONGRID method reduces this blocking effect by smoothing the output, giving the output a better visual appearance. Our implementation of block averaging allows us to shift the blocks in a controlled fashion, simulating lack of co-registration.

In Fig. 2 we see an example of four matched images, good, damaged, older and lower resolution by block averaging. We can see a bright spot in the center of the image, which is still present in the damaged one and the lower resolution one, but it is not present in the older one. Database construction details are given in Section 3.1.

The performance measures (RMSE, Q, Kappa and Overall Accuracy) were modeled as a response vector in a MANOVA mixed model with main effects (imputation method, resolution reduction method), and random nested effects (image and sub-image). Section 3.2 introduces the performance measures and Section 3.4 includes a detailed description of the MANOVA design. Following the MANOVA rejection, simultaneous multivariate comparisons were made by (Bonferroni corrected) two means Hotelling tests, and individual ANOVA results with simultaneous comparisons via Fisher LSD tests were studied to determine the best method, and the influence in performance of co-registration problems.

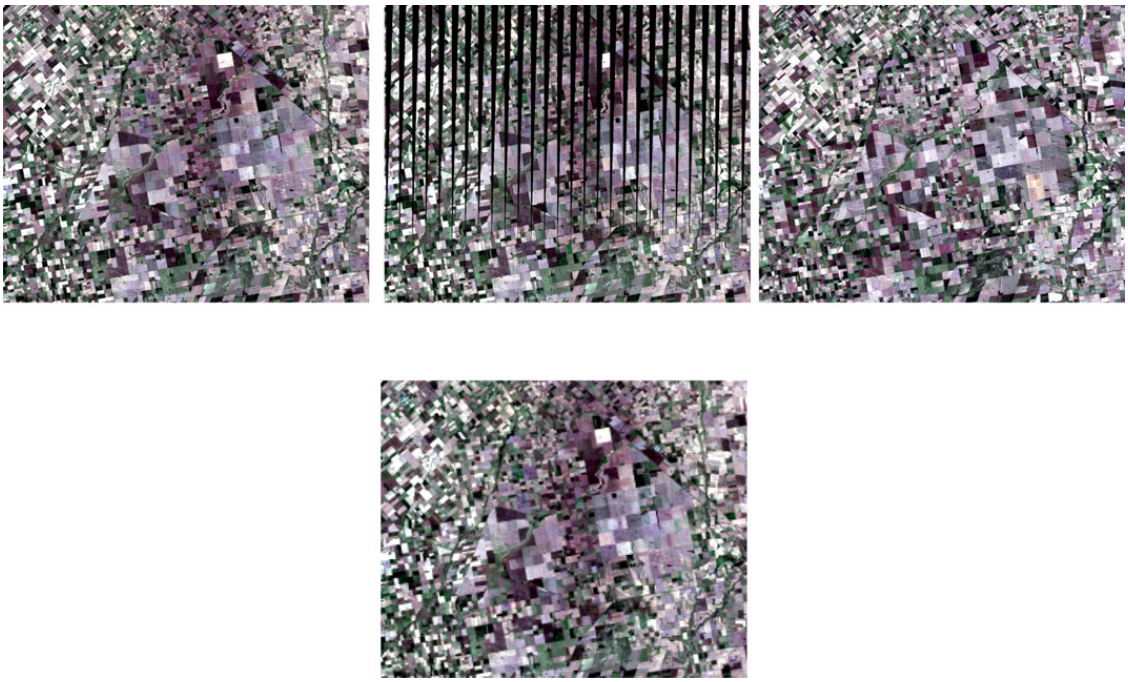

Fig. 2. Examples of images from database. From left to right, top to bottom, sample of Landsat 7 image, damaged image, older image and lower resolution image.

Table 1
Landsat 7 ETM+ images selected for simulation study, courtesy of CONAE-Argentina.

Track	Frame	L7 image date	Old L7 image date
1	73	02-MAY-2003_14:24:38.069	29-APR-2002_14:24:42.108
1	74	02-MAY-2003_14:25:02.021	29-APR-2002_14:25:06.058
229	83	04-NOV-2002_13:56:49.481	17-NOV-2001_13:57:29.017
229	87	04-NOV-2002_13:58:25.329	01-NOV-2001_13:58:52.526

3.1. Database

We selected four Landsat 7 ETM+ images with six bands each, acquired before the failure of the SLC mirror. These images also have companion good quality Landsat images acquired approximately a year before, see Table 1.

A data set of small images for experimenting was constructed in a hierarchical way, in order to account for variability between images acquired at different times and locations. Four pieces of each different Landsat 7 ETM images were selected from regions with few clouds and then subdivided again into four pieces, obtaining 16 sub-images for each of the original set of Landsat Images. The final 64 sub-images have 1250 $\times$ 1250 pixels (and 6 bands), and were stored as ground truth. Their companion damaged images were computed erasing 26% of information of each image, mimicking the strips with missing information in the SLC-off Landsat imagery.

It is well known that resolution in optical images can be degraded by block averaging. The use of a truly lower resolution image in the study, like the one provided by the SAC C satellite camera, would have introduced extra calibration and co-registration errors, without control over them. It was decided then to degrade the ground truth Landsat image by averaging, and simulate lack of co-registration shifting the degraded image a few lines and columns. This procedure will make the assessment of the impact of lack of co-registration in the imputation methods possible.

For each one of the 64 images, three types of lower resolution images with the same support were generated

1. Z_0, averaging radiometric information of 5 $\times$ 5 pixels square regions. Image was divided into 5 $\times$ 5 consecutive blocks, redefining the radiometric value of each pixel within the block as the average value of radiometric information of the original pixels

$$Z_0(ic, ir, B) = \frac{1}{5 * 5} \sum_{c=0}^{5-1} \sum_{r=0}^{5-1} \mathrm{L7}(c, r, B)$$

   $\forall (ic, ir) \in B$, 5 $\times$ 5 block.
2. Z_1 interpolation given by the IDL CONGRID command that reduces the impact of blocking by smoothing the output.
3. Z_2 moving Z_0 3 rows and 2 columns, mimicking a possible co-registration problem.

The companion older Landsat images were also cropped to match the sub-figures design. In Fig. 2 we have shown a sub-image with many agricultural fields and city blocks, and the matched sub-image, taken approximatively one year before. Main structure is the same, roads, ponds, main agricultural fields, but fine scale structure is different, and there is a possible change in crops, since homogeneous patches look different.

3.2. Performance measures

Radiometric evaluation and comparison were made globally and locally by means of RMSE (Root mean square error) and Q coefficients. The differences between the map of classes were assessed with measures related to the confusion matrix, Overall Accuracy and Kappa coefficients. This section gives the definition of all four measures. Further information about radiometric comparisons can be found for instance, in Congalton and Green (1999) and Richards and Jia (1999).

3.2.1. Radiometric measures

3.2.1.1. RMSE. The Root Mean Square Error between two images Y and W in region S_0 is

$$\sqrt{\frac{\sum_{s\in S_0}\|Y_s - W_s\|^2}{|S_0|}}.$$

The values of the RMSE are scale and order dependent, thus there is not universal reference values to compare with. Simulating the damage in the images, the scale and order are preserved, so the RMSE is not contaminated with distortion noise.

Q-coefficient. Wang and Bovik (2002) propose Q as an objective index of image quality that measures distortions with a combination of three factors: correlation loss, illumination distortion and contrast deformation. Let I be

$$I = \frac{4\sigma_{YW}\overline{Y}\,\overline{W}}{(\sigma_Y^2 + \sigma_W^2)[\overline{Y}^2 + \overline{W}^2]}$$

where

$$\overline{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i, \qquad \sigma_Y^2 = \sigma_{YY}, \qquad \sigma_{YW} = \frac{1}{N-1}\sum_{i=1}^{N}(Y_i - \overline{Y})(W_i - \overline{W}).$$

The image is divided in M small regions, and the index I is calculated in each of them. The overall quality index Q is the average over all of them.

$$Q = \frac{1}{M}\sum_{j=1}^{M} I_j.$$

The Q range is $[-1, 1]$, the highest value is obtained when $y_i = w_i$ for all $i = 1, \ldots, N$.

3.2.2. Confusion matrix: measures for class accuracy evaluation

Let M be the standard confusion matrix generated by two map of classes, the original map of classes C_X and the reconstructed C_D. The matrix's $M(i, j)$ entry is n_{ij}, the number of pixels of the true class j assigned to class i in the reconstructed image.

The totals by column and row will be denoted by (n_{+j}) and (n_{i+}), n will be the total of pixels considered and n_{ij}/n, n_{i+}/n, n_{+j}/n the proportions of misclassified pixels.

Overall Accuracy.

$$OA = \frac{\sum_{i=1}^{k} n_{ii}}{n}.$$

This coefficient compute the proportion of coincidences between the reference and reconstructed images (well classified pixels). A classification scheme is considered good if its Overall Accuracy is above 85%.

Cohen's Kappa coefficient Agresti (1996). $\widehat{K}$, the Kappa coefficient, is an estimator of the degree of matching in a classification scheme. It helps to determine if the scheme output is better than random allocation.

$$\widehat{K} = \frac{n\sum_{i=1}^{k} n_{ii} - \sum_{i=1}^{k} n_{i+}n_{+i}}{n^2 - \sum_{i=1}^{k} n_{i+}n_{+i}}.$$

3.3. Imputation methods

The database contains 64 sets with 6 images each. A 6-tuple has one ground truth image X, one matched damaged image X_D, one matched older image X_{old} and three lower resolution images $Z_0,\ Z_1,\ Z_2$, outputs of the resolution reduction methods. The ground truth of each one of the 64 sets is a Landsat image of a different geographic location, with a wide range of textures. Cities, forests, mountains and agricultural fields can be found in each of these images.

Six classes of imputation methods were implemented in IDL: Method A with 3 radios C for the high pass and low pass filters, 20%, 50%, 80% (A1, A2 and A3), method B and Method C with and without classification enhancement (C and C1). All these methods were applied to the damaged image in each set, and the performance measures collected.

Initial classification enhancement for Method C

Automatic classification is a difficult task, and the imputation method under study heavily depends on the accuracy of the initial map of classes. So it may be important to pay special attention to the coherence of the classes. If we do not have the class map as prior knowledge, we suggest the following set of steps to improve class homogeneity:

1. Given X_D, construct a map of classes C_D using K means, and assign the zero label to the missing information regions.
2. Given the class image C_D, detect the pixels with nonhomogeneous neighborhood, i.e, pixels that have no neighbor pixels on the same class. Call this set N_G.
3. For each pixel in N_G, verify its label using a forward mode filter, and a backwards mode filter. The filters will give a label that is consistent with the mode of the labels in a (forward or backward) neighborhood.
   (a) If both filters give the same label, update the label of the pixel to match this one.
   (b) If the labels are not the same, maintain the original label, and put the pixel in a new set called N_M.
4. For each pixel in N_M that do not have label zero, we updated the label as follows.
   (a) Take the one step neighbors of the pixels, and compute the arithmetic mean for each class present in the neighborhood.
   (b) Update the label of the pixel by the label of the class whose arithmetic mean is closest (in Euclidean distance) to the radiometric value of the pixel.

It is important to note that after this process, some of the pixels of the zero class may have been classified into a real class, only using contextual information. The other classes should have more defined borderlines.

The procedure that applies Method C to the enhanced map of classes will be referred to as method C, and the procedure that considered the map of classes as prior knowledge will be referred to as Method C1.

3.4. Mixed model analysis of a randomized complete block design

We want to assess the main effects in the mean of the response vector of four performance measures induced by two factors:

- Imputation method, with 6 levels : A1, A2, A3, B, C and C1.
- Resolution reduction method, with 3 levels: 0, 1 and 2.

Considering our set of Landsat images as a random subset of all possible images, we set image as random factor, and sub-image of each image as a nested random factor.

Our multivariate model design is

$$Y_{ijkl} = \mu + \alpha_i + \gamma_j + m_k + s(m)_{kl} + \varepsilon_{ijkl} \quad i = 1, \ldots, 6,\ j = 1, \ldots, 3,\ k = 1, \ldots, 4,\ l = 1, \ldots, 16$$

where

- Y_{ijkl} is the response vector RMSE, Q, Kappa and Overall Accuracy,
- α_i is the main effect of the imputation method,
- γ_j is the main effect of the resolution reduction method,
- μ is the general mean,
- m_k and $s(m)_{kl}$ are the independent random effects introduced by images and sub-images, $m_k \sim N(0, \Sigma_m)$, $s(m)_{kl} \sim N(0, \Sigma_s)$,
- ε_{ijkl} is the independent experimental error: $\varepsilon_{ijkl} \sim N(0, \Sigma_e)$.

All methods have been applied to all the sub-images, and no interactions have been considered since each image is a block itself and has no interesting interaction with any other methods.

Then

$$Y_{ijkl} \sim N(\mu + \alpha_i + \gamma_j, \Sigma_s + \Sigma_{.m} + \Sigma_e).$$

The MANOVA model can be used to test the hypothesis that the vector of performance measures for each level of the main factors, imputation method and lower resolution method, are equal against the alternative hypothesis that at least one is different. If the null hypothesis is rejected, numerous procedures can be used to determine which of the main effects is significantly different from the others. The comparison-wise Type I error will be fixed at $\alpha = 0.05$, in order to give F-values, but we will also report the p-values in all cases. Following a MANOVA rejection, it has been seen that the overall experiment-wise error rate stays within the limits of the α selected for individual ANOVA tests, helping answering the following research questions

1. Is there at least one method of imputation with a different performance from the others, within each performance measure analysis?
2. Is there at least one method of resolution reduction with a performance different from the others, within each performance measure analysis?
3. Does comparing performance (on any factor level), relative to the different measures of performance, produce the same results?

4. Results

4.1. Empirical findings

4.1.1. Visual inspection

In Fig. 3 it can be seen the original image and its reconstructions using methods B, A1 and A3, C and C1. We should notice that radiometric imputation from the map of classes is the worst of the methods, texture and fine details are not reconstructed, giving a blurred appearance to the imputed stripes.

We should also notice that globally, Method B produces an output with better visual quality than the others. Stripes are missing, but fine lines, small scale structure is less defined, like blurred. The reconstructions made with Methods A1 and A3 shows changes between the zones that are original and the zones imputed, but the imputed stripes have better defined small scale structure. This is particularly noticeable in reconstruction with Method A1, which introduces more information from the older image than Method A3.

4.1.2. Mean profiles

Visual inspection can only be made with few images, we have 64 sub-images twice as large as than the ones just shown. To inspect performance globally, we will study now the mean profiles of all performance measures. We compute the sample means of each performance measure, for each method and each image. Thus, we average the values of a particular performance measure taken over all 16 sub-images from a single image. A Mean Profile is a polygonal that links the four means (one per image) from a particular performance measure.

Methods mean profiles should be parallel, if there is no interaction between the performance measure and the images. Differences are then produced by the methods and not by the images. Flat Profiles indicate that each method produces similar values on any image, but that is not true. Images depend not only on the geographic target observed, but the atmospheric conditions of the moment of acquisition. Variability is huge. If we consider our images as chosen at random, they will be only a random effect that inflates the variance, but does not introduce a main effect that could mask the effect of the imputation methods.

Also, it is important to note that in order to compare all methods using Kappa and Overall Accuracy, a map of classes must be made for each reconstructed image, as well as the original image. Methods A and B generates radiometric imputations, Method C and C1 generate maps of classes first, and then radiometric imputation, choosing at random a value within the class assigned. The automatic K-means method was used to obtain a map of classes for the original images and the ones reconstructed with methods A and B.

The coefficients Kappa, Q and Overall Accuracy are designed to give high scores to good reconstructions. On the contrary, RMSE is an error measure, then reconstruction is best when RMSE is at its lowest.

In Fig. 4 we see two profile plots, Mean Overall Accuracy against Image, and Mean Kappa against Image, where the means are made over all the sub-images of the same image. The highest profile will then be the best method and the lowest the worst, if all profiles are more or less parallel.

In both plots, Method C performed badly, and seems to interact with the images. Its profile does not follow the pattern of the others. Nevertheless, Method C1, a simplification of method C, has good performance with these measures that evaluate classification accuracy. It follows a pattern similar to Method A1, A2, A3 and B, but with scores as good as Method B. Measures Kappa and Overall Accuracy depends heavily on good segmentation of the ground truth. Method C includes an enhancement that may give better defined classes, but separate it from the segmentation of the ground truth, increasing the global error, outside the imputation error.

In Fig. 5 we see other two profile plots, Mean Q against Image, and Mean RMSQ against Image.

The highest profile will then be the best method and the lowest the worst, when considering Q, and the reverse when considering RMSE, if all profiles are more or less parallel. Again, Method C has bad performance, followed by Method C1. In this case, reasons for this behavior may be found in the fact that imputation is made randomly within a convex hull in each class, which could have a very wide range of radiometric values. Method C1 may generate a map of classes very close to the real one, but not the radiometric values of the missing pixels.

We will now explore possible interactions between imputation method and Resolution Reduction Method (RRM), by making Mean Profile Plots of each Performance measure by imputation method and plotting it against RRM. In Fig. 6 we should notice that the plots of three versions of Method A are not parallel, suggesting a possible interaction that should be further studied with an ANOVA test.

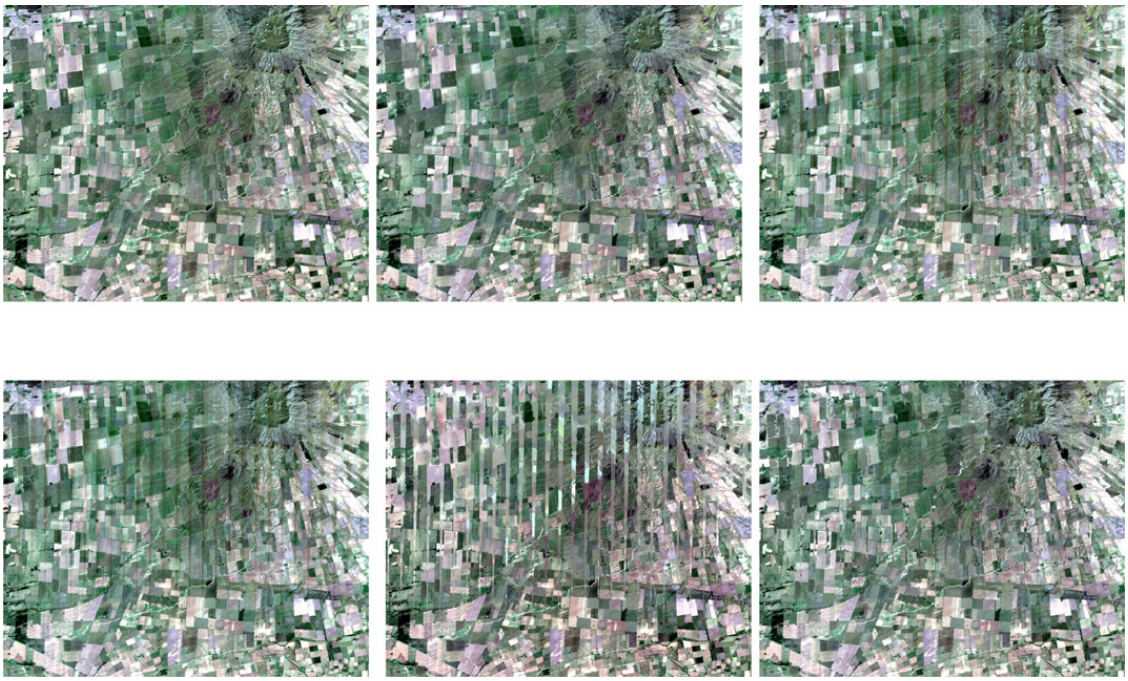

Fig. 3. From left to right. First row: original sub-image, reconstructions from Method B and Method A1. Second row: reconstructions from Method A3, Method C and Method C1.

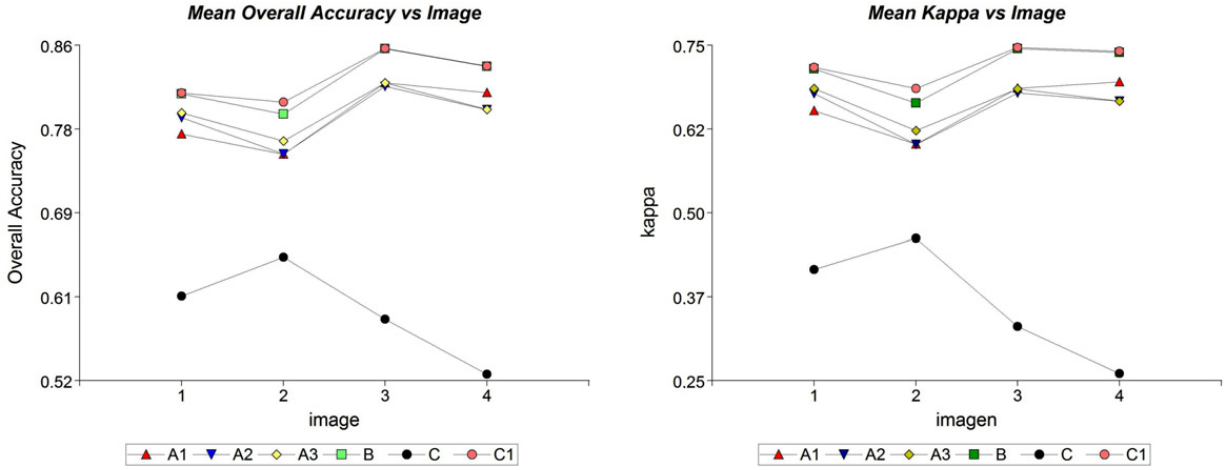

Fig. 4. Left panel: Overall Accuracy against image, right panel: Kappa against image.

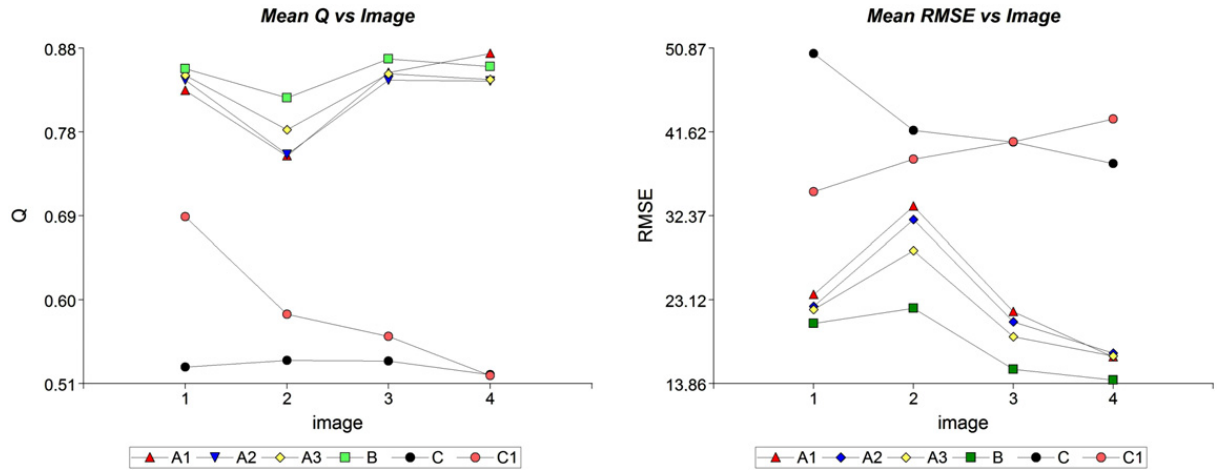

Fig. 5. Left: Q against image, and right: RMSE against image.

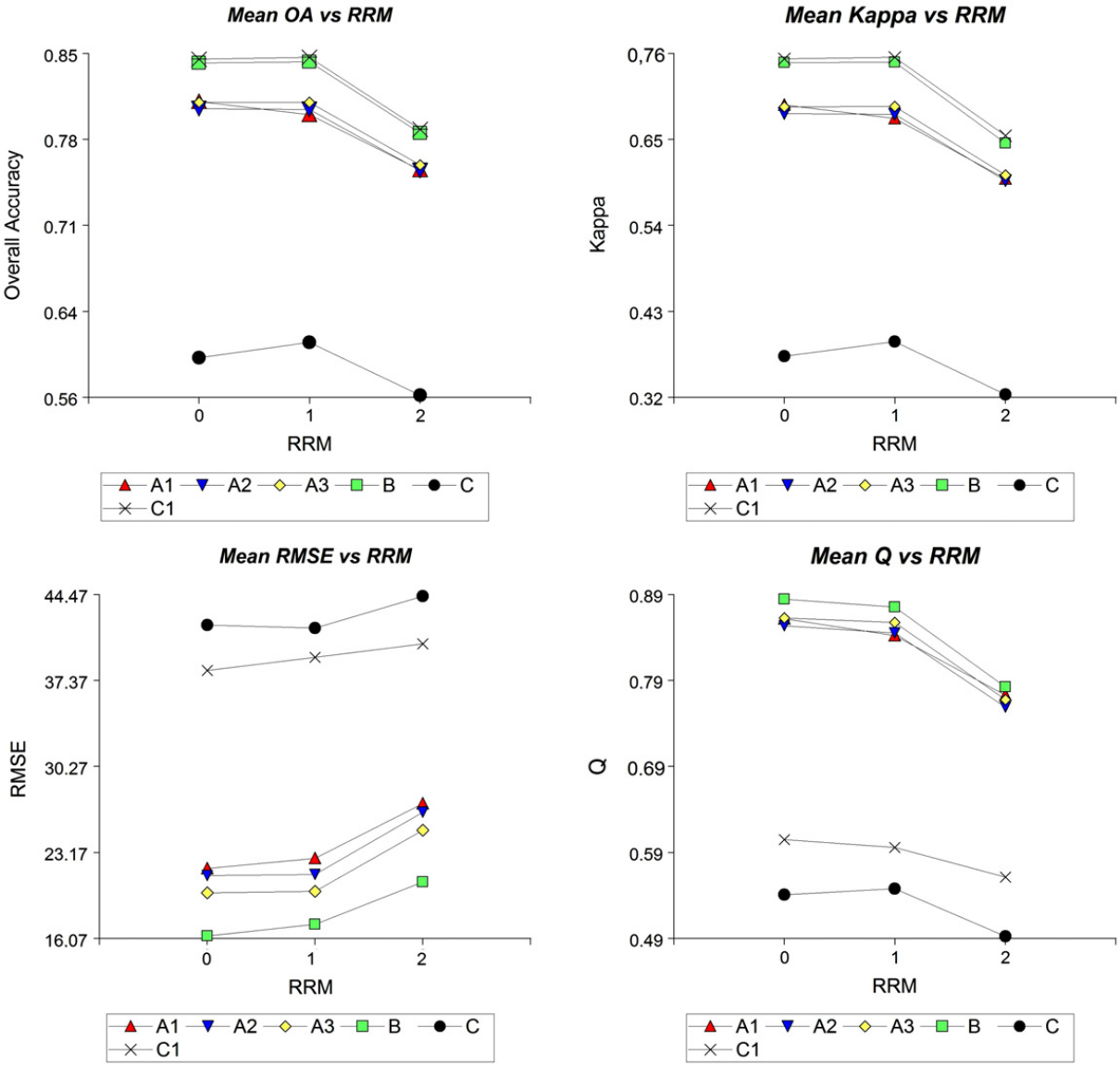

Fig. 6. From left to right. First row: Overall Accuracy against RRM by method and Kappa against RRM by method. Second row: RMSE against RRM by method and Q against RRM by method.

4.1.3. *Preliminary conclusions*

In all cases, the mean profile of Method B seems to be the best, followed by A3, fusion using a high value of threshold, that gives Z more influence than X_{old}. That corroborates the facts pointed out by visual inspection on only one sub-image. But we should test if the differences are statistically significant with level $\alpha = 0.05$. All A Methods are in pack, their differences may not be significant.

When studying interactions between the main factors, the A Methods suggested a change between Mean performance using RRM0 and RRM1. The most important change would be a significant difference on Mean Performance between Methods when using RRM2, which is slightly out of sync.

In the next section, we will corroborate these empirical findings with MANOVA test of significance, and individual ANOVA tests. We will be considering Image as a blocking factor, with a nested sub-factor, Sub-Image, discarding interaction as weak, Kuehl (1999). In both set of plots, we have reasons to believe that the profiles from Method C and C1 do not show interaction with the image factor, but other errors beside the ones measured by the particular performance index.

4.2. *MANOVA test outputs*

The vector of response variables contains the four measures considered, Overall Accuracy, Kappa, RMSE and Q. The MANOVA test searches for significant evidence to reject the following null hypothesis:

$$\mathrm{H1}_0 : \alpha_i = 0, \quad \forall i : i = 1, \ldots, 6$$
$$\mathrm{H2}_0 : \gamma_j = 0, \quad \forall j = 1, \ldots, 3.$$

Table 2
Multivariate analysis of variance table (Wilks).

	S.V.	F statistic	df (num)	df (den)	p	Error
Method	0.08	200.56	20	3543	< 0.0001	
RRM	0.64	67.08	8	2136	< 0.0001	
Image	0.61	2.57	12	151	0.0040	(im>subim)
Image>subimage	0.01	36.78	240	4268	< 0.0001	
Method*RRM	0.95	1.52	40	4052	0.0192	

Table 3
Hotelling test (Bonferroni adjustment) Alpha:= 0.05. Different letters indicate significant differences among location parameters ($p <= 0.05$) Error: Pooled covariance matrix df: 1071.

Method	Overall Accuracy	Kappa	RMSE	Q	n					
C	0.59	0.36	42.67	0.53	192	A				
C1	0.83	0.72	39.30	0.59	192		B			
B	0.83	0.71	18.09	0.85	192			C		
A2	0.79	0.65	23.05	0.82	192				D	E
A1	0.79	0.66	23.93	0.83	192				D	
A3	0.80	0.66	21.62	0.83	192					E

Table 4
Hotelling test (Bonferroni adjustment) Alpha:= 0.05. Different letters indicate significant differences between location parameters ($p <= 0.05$) Error: Pooled covariance matrix df: 1069.

RRM	Overall Accuracy	Kappa	RMSE	Q	n		
2	0.74	0.57	30.71	0.69	384	A	
1	0.79	0.66	27.05	0.76	384		B
0	0.79	0.66	26.58	0.77	384		B

with α_i the vector of main effects introduced by the i-imputation method, and γ_j is the vector of main effects introduced by the j-resolution reduction method. Rejection of each hypothesis means

$H1_0$. There is an imputation method effect, there is at least one method that has a different vector of performances than the others.

$H2_0$. There is a resolution reduction method effect, there is at least one method of resolution reduction with a vector of performance different to the others.

In a MANOVA test, answers are global, the combination of all performance measures gives power to the rejection, Rencher (2002). Interaction between the main factors is important, since they could mask main effects. Following a MANOVA rejection, individual ANOVA test on each performance measure could give more information about relationships between the main factors, and individual comparisons may show possible clustering in Methods space. Table 2 shows the results of the MANOVA test of significance with the Wilks statistics. Pillai, Roy and Lawley–Hotelling statistics were computed but leaded to the same conclusion and were omitted. Interaction between imputation method and resolution reduction method seems to be significant. Main effects are also significant.

In Table 3 we can see the output of a series of two means — Hotelling tests (with a Bonferroni adjustment for experiment-wise level $\alpha = 0.05$), only with one factor, imputation method. Different letters indicate significant differences among location parameters. Using all measures globally, Method A2 does not seem to be significantly different from Methods A1 and A3, but the last two are significantly different from each other. All other Methods are significantly different from one another.

In Table 4 we have written the output of a series of two means — Hotelling tests (with a Bonferroni adjustment for experiment-wise level $\alpha = 0.05$), with the other main factor, resolution reduction method. Different letters indicate significant differences among location parameters. Using all measures globally, there is a difference in performance when lack of co-registration is present, while there are no differences between block averaging and CONGRID methods.

We have globally studied the main effects of imputation and resolution reduction methods. But interaction between both factors is not advisable. We would like to see in detail which one of the measures (if not all of them) shows changes in performance when both factors are considered at the same time.

4.3. Individual ANOVA tests

Knowing the results of the MANOVA test, we are confident that the level of the test of the individual ANOVA test are not far from its nominal value, $\alpha = 0.05$.

In Table 5, we have the output of an ANOVA test for Q as response variable. Results are similar to MANOVA, and it shows indeed significant interaction between resolution reduction method and imputation method. The other three ANOVA

Table 5
Analysis of variance table (partial SS) Variable: Q measure and Overall Accuracy.

Q	SS	df	MS	F	p-value	(Error)
Model	26.33	80	0.33	57.76	< 0.0001	
Method	19.85	5	3.97	696.52	< 0.0001	
RRM	1.48	2	0.74	129.58	< 0.0001	
Image	0.51	3	0.17	2.36	0.0803	(Image>subimage)
Image>subimage	4.36	60	0.07	12.75	< 0.0001	
Method*RRM	0.14	10	0.01	2.40	0.0080	
Error	6.10	1071	0.01			
Total	32.44	1151				
Overall Accuracy	SS	df	MS	F	p-value	(Error)
Model	10.79	80	0.13	43.07	< 0.0001	
Method	7.72	5	1.54	493.21	< 0.0001	
RRM	0.70	2	0.35	112.28	< 0.0001	
Image	0.29	3	0.10	2.77	0.0493	(Image>subimage)
Image>subimage	2.06	60	0.03	10.95	< 0.0001	
Method*RRM	0.02	10	2.3E−03	0.72	0.7064	
Error	3.35	1071	3.1E−03			
Total	14.15	1151				

Tests, one for each response measure, Kappa, RMSE and Overall Accuracy, also indicate significant main effects, but none of them shows significant interaction between the two main factors, being p-values 0.54, 0.37 and 0.70 respectively for the interaction test. We display only Overall Accuracy output for the sake of completeness in Table 5.

4.4. Grouping

We would like now to study how the location parameters cluster to identify the methods that are significantly different from one another. We have some information from the multivariate Bonferroni corrected set of Hotelling tests we made, but now we will see how the different variables cluster the Methods.

We will use Fisher LSD multiple comparison test as clustering criteria. Other multiple comparison tests such as Tukey LSD and Duncan tests gave the same type of clustering output, and are not reported here.

4.4.1. Grouping by imputation methods

We will now study the information contained in Table 6. When considering imputation method, Method B and C1 are grouped together by the two measures of good classification, Kappa and Overall Accuracy. The other two measures, Q and RMSE divide them into single clusters. The three versions of Method A have been grouped together by all the measures but RMSE, which considered Method A1 and A3 significantly different. The two versions of Method C, C and C1, have been distinguished by all the measures.

The multivariate simultaneous comparison distinguished all methods but the versions of Method A. And even though, Method A1 was considered significantly different from Method A3. It is interesting to see that this is the way of grouping of the two radiometric measures. The classification based measures do not distinguish between Method B and C1, or A1, A2 and A3.

To the naked eye, Method B reconstruction of Fig. 3 is very different from Method C1 reconstruction, and Method A1 is sharper than Method A3. The multivariate simultaneous comparison agreed with this statement, with an experiment-wise Type I error $\alpha = 0.05$.

Now we can come back to the profile analysis of Section 4.1.2. In that section, the performance measures means, computed over the sub-images of each image, were plotted as profiles, and the highest value of Q, Kappa and Overall Accuracy indicated the best imputation method, and the lowest value of RMSE back up the same statement. Method B seemed to be the best of all them. But the analysis did not have any statistical confidence. The simultaneous comparisons made reported Method B as significantly different in mean performance from all the others, and Table 6 shows the mean value of measures Q, Overall Accuracy and Kappa the highest of all, and the lowest of the error measure RMSE. These results give confidence to the previous profile analysis.

4.4.2. Grouping by resolution reduction methods

Now we are concerned with the fact that imputation may have reduced performance when the lower resolution image used as extra information was not accurately co-registered. When simulating the lower resolution image, block averaging and CONGRID methods were chosen as basic resolution reduction methods, and as a third method, the block averaged image was moved slightly (Method RRM2), simulating lack of co-registration. Simultaneous comparisons using Fisher LSD (see Table 7) report that the two versions of block averaging are indistinguishable, but RRM2 produced a significantly different mean in all measures, making them worse. In the case of Kappa, Q and Overall Accuracy, the means are smaller, and in the case of RMSE the mean is higher.

Table 6
Test: Fisher LSD, $\alpha = 0.05$. Different letters indicate significant difference between location parameters.

Method	Mean OA	n						
C	0.59	192	A					
A2	0.79	192		B				
A1	0.79	192		B				
A3	0.80	192		B				
B	0.83	192			C			
C1	0.83	192			C			
Method	Mean Kappa	n						
C	0.36	192	A					
A2	0.65	192		B				
A1	0.66	192		B				
A3	0.66	192		B				
B	0.71	192			C			
C1	0.72	192			C			
Method	Mean RMSE	n						
B	18.09	192	A					
A3	21.62	192		B				
A2	23.05	192		B	C			
A1	23.93	192			C			
C1	39.30	192				D		
C	42.67	192						E
Method	Mean Q	n						
C	0.53	192	A					
C1	0.59	192		B				
A2	0.82	192			C			
A1	0.83	192			C			
A3	0.83	192			C			
B	0.85	192				D		

Table 7
Test: Fisher LSD. $\alpha := 0.05$. Different letters indicate significant differences between location parameters ($p <= 0.05$).

RRM	Mean OA	Mean Kappa	Mean RMSE	Mean Q	n		
2	0.74	0.57	30.71	0.69	384	A	
0	0.79	0.66	27.05	0.76	384		B
1	0.79	0.66	26.58	0.77	384		B

5. Conclusions

Regression models are considered successful models for imputation in a wide range of situations in all Applied Sciences. In this paper, we introduced a Simple Regression Model for imputation of spatial data in large regions of a Remote Sensed image, Method B, which had a statistically significant better performance than the two other main methods, also adapted from the literature, in the frame of a careful simulation study.

Multivariate simultaneous comparisons of all imputation methods agreed with the visual inspection of the reconstructed images. Method B reconstruction shows less contrast between the imputed stripes and the nonimputed regions, but loses sharpness, and appears slightly blurred. Method A1 and Method A3 reconstructions are different, A1 produced finer details, but also a lot more contrast between imputed and nonimputed regions, and A3 has a smoother appearance, closer to Method B reconstruction, also with less detail.

Method C and C1 were designed to give good segmentations, despite the large regions without information, and Method C1 successfully accomplishes the task. Nevertheless, pixels radiometric imputation is made by randomly choosing a value from the convex hull of pixels from its class in a small neighborhood. Without the help of radiometric extra data, radiometric imputation becomes quite poor in both implementations. The first computational implementation of Method C has a preprocessing step that makes the initial segmentation sharper, but it increases the difference with the original image, not only in the imputed regions but in all regions. This preprocessing step seems to be the responsible for the method's strange behavior.

One of the hypothesis of all methods was the existence of good, co-registered, temporally accurate imagery with possible lower resolution. To test the dependence of the methods of co-registration, the block averaged image that acts as lower resolution extra data was shifted slightly and the performance of all methods diminished. This reduction was observed statistically significant with Fisher LSD test of simultaneous comparisons, for each performance measure, and globally with a series of Hotelling tests (Bonferroni adjusted).

6. Discussion

There are many methods in the literature attempting to impute remote sensed data without extra knowledge, this means, only using information that is present in the nondamaged part of the image. A very important set of examples is the so called inpainting methods, which generate good results filling in small gaps, but fail when the gaps are large enough to include textured areas.

Information about such texture needs to be obtained from somewhere else, and old information from the same sensor is almost always available; but contemporary information, if available, would come from another source, images mostly from another sensor. Many satellites carry more than one sensor nowadays, and the oldest ones are also part of constellations of satellites, which are able to provide temporally accurate imagery but with different spectral or resolution characteristics.

To our knowledge, there is no paper in the literature of image processing assessing the size or thickness of the gap that could be interpolated or imputed successfully. Nevertheless, most of the method's proposals acknowledge the texture problem, meaning that an over smoothing of the interpolated region is produced, degrading the texture appearance of the imputed region. When extra information is added, better reconstruction of texture is observed in the filled gap.

In this paper, we address the problem of defining and comparing simple methods of imputation based on data fusion techniques, in order to detect the best approach to the problem. We can conclude that the regression based method is the best approach of all tested here, producing imagery that is visually and analytically more accurate than the others.

Nevertheless, the design of all the methods may be refined. For example, Method A is based on a simple mixture of spectral information from two images, provided by ideal filters, and despite its simplicity, this function has given good results in the problem of merging regular data from sensors with different resolutions. A more complex mixture function, based on Butterworth filters, for example, or the merging of wavelets decompositions by band fusion, could be interesting extensions of the method. In the case of Method B, comparisons have been made with the Euclidean measure since remote sensing data is usually massive, and we wanted to maintain the computational complexity of all our imputation methods as low as possible. But the exchange of the Mahalanobis distance for the Euclidean one may include the correlations between bands in the design of the method. Also, we have only considered a simple regression model, while generalized additive models were strong possibilities for increasing complexity. Fusion methods are difficult to handle, and simplicity has been the strongest characteristic of our proposals. Particularly, introducing correlation structure in image fusion techniques is very difficult, and it is beyond the scope of the present paper.

Spatial models like SAR and CAR models are good choices for Image Modeling based imputation, without considering extra knowledge in the design. In that case, imputation is made by pixel value prediction, estimating the parameters of the model from the regions of good data. In the case of small gaps, spatial models for inpainting optical images have been proposed by Ojeda et al. (2010) and Bustos et al. (2009), but its extensions to Landsat data problems are still in the developing stage.

In relation to Method C, we acknowledge that some accurate previous knowledge about the map of classes of the nondamaged part of the image is mandatory. Like method A, which needs a parameter to be declared in advance, method C is not a fully automatic method. Only Method B estimates the values of the coefficients needed in an automatic fashion, and produces an automatic imputation.

As a formal conclusion, we would like to point out that there are very few papers in the image processing literature that make thorough comparisons of imputation methods with real data. Method B results are strikingly good for such a simple method, given the order of the data loss, and it is consistently good through the 16 images considered, even with extra data out of sync.

Acknowledgments

The authors gratefully acknowledge the helpful comments of the anonymous referee and associate editor for improving the clarity of the paper, and for giving useful suggestions. The results introduced in this paper are part of the Master Thesis in Applied Statistics of Valeria Rulloni, at Universidad Nacional de Cordoba. This work was partially supported by Secyt under grant PID Secyt-UNC 214/10 and VR Master Thesis Fellowship. We would like to thank S. Ojeda, J. Izarraulde, M. Lamfri, and M. Scavuzzo for interesting conversations leading to the design of the methods. The imagery used in the simulation section was kindly provided by CONAE, Argentina. Imputation methods and performance measures were computed with ENVI Software. Statistical Analysis was made with INFOSTAT, UNC, provided by Prof. J. Di Rienzo.

References

Agresti, A., 1996. Introduction to Categorical Data Analysis. Wiley-Interscience, 246.
Allison, P., 2000. Multiple imputation for missing data: a cautionary tale. Sociol. Methods Res. 28 (3), 301–309.
Arellano, P., 2003. Missing information in Remote Sensing: Wavelet Approach to Detect and Remove Clouds and their Shadows. M.Sc. Thesis. ITC Library.
Benabdelkader, S., Melgani, F., 2008. Contextual spatio spectral post reconstruction of cloud-contaminated images. IEEE Geoscience Remote Sensing Letters 5 (2), 204–208.
Bertalmío, M., Sapiro, G., Caselles, V., Ballester, B., 2000. Image inpainting. In: Proc. of SIGGRAPH 2000. New Orleans, USA. pp. 417–424.
Blum, R., Zheng, L., 2005. Multi-Sensor Image Fusion and its Applications (Signal Processing and Communications). CRC Press.
Bugeau, A., Bertalmío, M., 2009. Combining texture synthesis and diffusion for image inpainting. VISAPP 1, 26–33.

Bugeau, A., Bertalmío, M., Caselles, V., Sapiro, G., 2010. A comprehensive framework for image inpainting. IEEE Transactions on Image Processing 19 (10), 2634–2645.
Bustos, O., Ojeda, S., Vallejos, R., 2009. Spatial ARMA models and its applications to image filtering. Brazilian Journal of Probability and Statistics 23, 141–165.
Commonwealth of Australia (2006). Landsat 7 ETM+ SLC-off composite products. Available online at http://www.ga.gov.au/acres/referenc/slcoff_composite.jsp.
Congalton, R., Green, K., 1999. Assessing the Accuracy of Remotely Sensed Data: Principles and Practices. Lewis Publishers.
ENVI Software. ITT Visual Information Solutions. http://www.ittvis.com/portals/0/tutorials/envi/ENVI_Quick_Start.pdf.
Esedoglu, S., Shen, J., 2002. Digital inpainting based on the Mumford–Shah–Euler image model. European Journal of Applied Mathematics 13, 353–370.
INFOSTAT Software. http://www.infostat.com.ar/.
Kuehl, R., 1999. Design of Experiments: Statistical Principles of Research Design and Analysis, second ed. Duxbury Press.
USGS, NASA and Landsat 7 Team, 2003. Preliminary assessment of the value of Landsat 7 ETM+ data following scan line corrector malfunction. Available online at http://landsat.usgs.gov/documents/SLC_off_Scientific_Usability.pdf.
Le Hegarat-Mascle, S., Bloch, I., Vidal Madjar, D., 1998. Introduction of neighborhood information in evidence theory and application to data fusion of radar and optical images with partial cloud cover. Pattern Recognition 31 (11), 1811–1823.
Ling, Y., Ehlers, M., Usery, L., Madden, M., 2007. FFT-enhanced IHS transform method for fusing high-resolution satellite images. ISPRS Journal of Photogrammetry and Remote Sensing 61 (6), 381–392.
Little, R., Rubin, D., 2002. Statistical Analysis with Missing Data. John Wiley, New York.
Maalouf, A., Carre, P., Augereau, B., Fernandez-Maloigne, C., 2009. A bandelet-based inpainting technique for clouds removal from remotely sensed images. IEEE Transactions on Geosciences and Remote Sensing 47 (7), 2363–2371.
Melgani, F., 2006. Contextual reconstruction of cloud-contaminated multitemporal multispectral images. IEEE Transactions on Geoscience and Remote Sensing 44 (2), 442–455.
Melgani, F., Benabdelkader, S., Boulemden, M., 2007. Cloud-contaminated image reconstruction with contextual spatio-spectral information. In: Proc. IGARSS, 2007. pp. 373–376.
Ojeda, S., 2011. Personal communication.
Ojeda, S., Vallejos, R., Bustos, O., 2010. A new image segmentation algorithm with applications to image inpainting. Computational Statistics and Data Analysis 54, 2082–2093.
Pohl, C., 1998. Multi-sensor image fusion in remote sensing: concepts, methods and applications. International Journal of Remote Sensing 19 (5), 823–854.
Rencher, A., 2002. Methods of Multivariate Analysis. second ed. Wiley Series in Probability and Statistics.
Richards, J., Jia, X., 1999. Remote Sensing Digital Image Analysis An Introduction. Springer.
Rossi, R., Dungan, J., Beck, L., 1994. Kriging in the shadows: geostatistical interpolation for remote sensing. Remote Sensing of the Environment 49, 32–40.
Scaramuzza, P., Micijevic, E., Chander, G., 2004. SLC gap filled products: phase one methodology. Available online at http://landsat.usgs.gov/documents/L7SLCGapFilledMethod.pdf.
Telea, A., 2004. An image inpainting technique based on the fast marching method. Journal of Graphics Tools 9 (1), 23–34.
Tsuda, K., Akaho, S., Asai, K., 2003. The EM algorithm for kernel matrix completeness with auxiliary data. Journal of Machine Learning Research 4, 67–81.
Wang, Z., Bovik, A., 2002. A universal image quality index. IEEE Signal Processing Letters 9 (3), 81–84.
Zhang, C., Li, W., Travis, D., 2007. Gaps-fill of SLC-off landsat ETM+ satellite image using a geostatistical approach. International Journal of Remote Sensing 28 (22), 5103–5122.
Zhou, F., Wang, Z., Qi, F., 2006. Inpainting thick image regions using isophote propagation. In: Proc. IEEE ICIP, pp. 689–692.