

A sparsity-based variational approach for the restoration of SMOS images from L1A data

Javier Preciozzi, *Student Member, IEEE*, Andrés Almansa, *Member, IEEE*, Pablo Musé, Sylvain Durand, Ali Khazaal and Bernard Rougé

Abstract—The SMOS mission senses ocean salinity and soil moisture by measuring Earth’s brightness temperature using interferometry in the L-band. These interferometry measures known as *visibilities* constitute the SMOS L1A data product. Despite the L-band is reserved for Earth observation, the presence of illegal emitters cause radio frequency interferences (RFI) that mask the energy radiated from the Earth and strongly corrupt the acquired images. Therefore, the recovery of brightness temperature from corrupted data by image restoration techniques is of major interest. In this work we propose a variational model to recover super-resolved, denoised brightness temperature maps by decomposing the images into two components: an image T that models the Earth’s brightness temperature and an image O modeling the RFIs. The approach is totally new to our knowledge, in the sense that it is directly and exclusively based on the visibilities. Experiments with synthetic and real data support the suitability of the proposed approach.

Index Terms—SMOS, MIRAS, RFI, non-differentiable convex optimization, total variation minimization.

I. INTRODUCTION

HAVING the possibility to observe Earth variables such as surface soil moisture (SSM) and sea surface salinity (SSS) is crucial to obtain meteorological and climate predictions [1]. The SMOS satellite [2] carries an instrument called MIRAS (Microwave Imaging Radiometer by Aperture Synthesis) [3], [4], that provides indirect measures of the corresponding brightness temperatures of both SSM and SSS, in the L-band microwave, using interferometry and sensing the so-called *visibility function* [5].

A. From brightness temperatures to visibilities: The MIRAS instrument and the forward problem

The MIRAS instrument [3], [4] is composed of a set of antennas. The interferometry principle used by the MIRAS instrument leads from brightness temperatures to the visibility function [5]. Let (A_k, A_l) be any pair of its antennas; the visibility function $V_{k,l}$ is defined as the complex cross-correlation between the received signals at A_k and A_l :

$$V_{k,l} = \frac{1}{\sqrt{\Omega_k \Omega_l}} \iint_{\|\xi\| \leq 1} U_k(\xi) U_l^*(\xi) (T_b(\xi) - T_r) \tilde{r}_{kl}(t) \frac{e^{-i2\pi \mathbf{u}_{kl}^T \xi}}{\sqrt{1 - \|\xi\|^2}} d\xi \quad (1)$$

Here, U_k and U_l are the corresponding normalized voltage patterns of the antennas and Ω_k, Ω_l their corresponding solid angles; \mathbf{u}_{kl} is the frequency baseline associated to (A_k, A_l) . The Cartesian coordinates $\xi = (\xi_1, \xi_2)$ are the spatial domain coordinates, which are restricted to the unit circle. T_r is the physical temperature of the receivers (assumed the same for all receivers); \tilde{r}_{kl} is the Fringe-Wash function, a function of the spatial delay $t = \frac{\mathbf{u}_{kl}^T \xi}{f_0}$, where $f_0 = \frac{c}{\lambda_0}$ is the central frequency of observation. Note that the brightness temperature T_b is a 2D function defined on the unit circle $\{\xi : \|\xi\| \leq 1\}$. It is worth mentioning that (1) is valid only for the interferometric measurements, not for the zero-spacing ones (provided by the NIRs). Moreover, this equation does not account for dual-polarization nor full-polarization considerations which are the two operating modes of MIRAS [5].

Since the support of the temperatures is the unit circle, it is well known that the best regular sampling grid is a hexagonal one [6], leading to the largest alias-free Field of View for a given spacing between the antennas [7]. The antennas configuration chosen for the MIRAS instrument is a Y-shaped array [7], shown in Figure 1 along with its corresponding hexagonally sampled, star-shaped domain Ω . The star-shaped domain Ω , known in aperture synthesis as the (experimental) frequency coverage [8, Ch. 5], is contained within the hexagon, hence requiring extrapolation to recover the missing parts (see Figure 2). Because of other practical

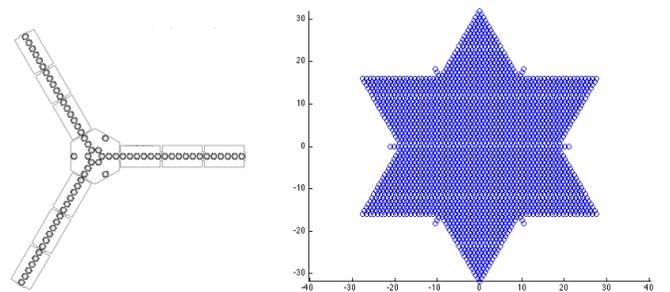


Fig. 1. Diagram of the MIRAS instrument onboard of SMOS satellite (from [9]) and the star-shaped, hexagonally sampled visibilities domain derived from it. The number of antennas in the miras instrument is 69: 22 antennas type Light-Weight Cost-Effective Front-End (LICEF) in each arm, plus three antennas of type Noise Level Injection Radiometers (NIR) located in-between each arm in the center of the instrument.

Javier Preciozzi and Pablo Musé are with the Department of Electrical Engineering, Facultad de Ingeniería, Universidad de la República, Uruguay.

Andrés Almansa is with LTCI - Télécom ParisTech, France.

Sylvain Durand is with MAP5, Université René Descartes, France.

Ali Khazaal and Bernard Rougé are with the Centre d’Etudes Spatiales de la Biosphère (CESBIO), France.

Manuscript received September 29, 2014; revised XXXX XX, 201X.

issues, the final separation between antennas in the MIRAS instrument is $d = 0.875$ wavelengths. This value is larger than the critical sampling distance on a hexagonal grid (which is $d = 1/\sqrt{3}$ wavelengths), and then, some amount of aliasing

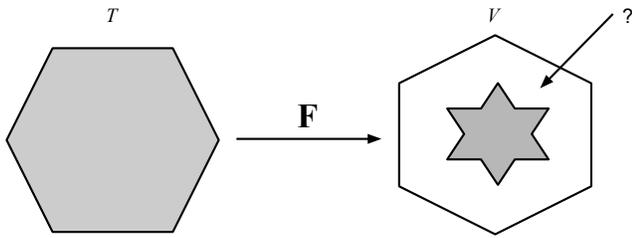


Fig. 2. The absence of information beyond the star-shaped domain Ω makes the recovery of the brightness temperature map an ill-posed problem.

is introduced on the spatial domain. As shown in Figure 3, the Alias-Free Field of View is the small central zone in the hexagon delimited by the blue dashed lines. Nevertheless, because of the tilt of the satellite acquisition system, most of the aliasing comes from the intersection between the Earth disk with the known cold sky, and therefore it is customary to consider the Field of View delimited by the red plain lines, generating the so-called *Extended Alias-Free Field of View* (E-AF-FOV) [7].

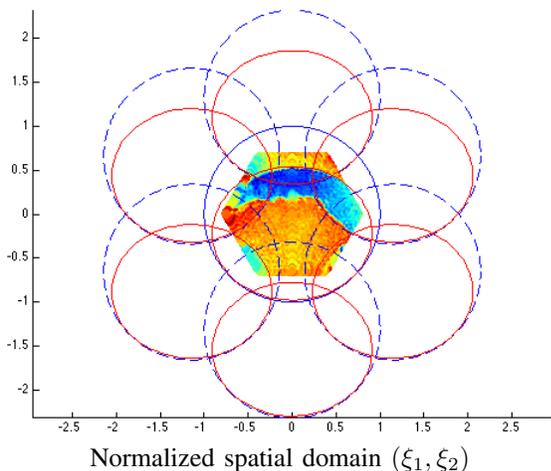


Fig. 3. Alias Free - Field of View (AF-FOV) – zone delimited by the unity circle contours in blue lines – and the corresponding Extended Alias-Free Field of View (E-AF-FOV) in red lines.

B. Restoring brightness temperatures from visibilities: An ill-posed inverse problem

Now, going back to the forward problem (1), if we denote by $T = T_b - T_r$, the samples of T in the hexagonal grid could be obtained from the visibility samples by solving the linear system $\mathbf{G}T = V$, where matrix \mathbf{G} represents the discrete linear operator given by (1). Of course, the inversion of the forward problem is ill-posed since \mathbf{G} is not invertible because the number of visibility measures is significantly lower than the samples of T . The ill-posedness of the problem can also be interpreted as the lack of information beyond Ω , as illustrated in Figure 2. Hence, additional constraints must be added to the model. In [10], the authors propose to solve it as a constrained least square problem, imposing that T has no frequency

components outside Ω . This problem can be formulated as the unconstrained minimisation of $\|V - \mathbf{G}\mathbf{F}^*\mathbf{Z}_\Omega\hat{T}\|_2^2$ on \hat{T} , where \mathbf{F}^* denotes the matrix corresponding to the hexagonal Inverse Fourier Transform, \mathbf{Z}_Ω the zero padding operator and \hat{T} the Fourier coefficients of T for frequencies in Ω .

Let $\mathbf{J} = \mathbf{G}\mathbf{F}^*\mathbf{Z}_\Omega$, then $\hat{T} = \mathbf{J}^+V$ where \mathbf{J}^+ is the pseudo-inverse of \mathbf{J} : $\mathbf{J}^+ = (\mathbf{J}^*\mathbf{J})^{-1}\mathbf{J}^*$. This is the way the L1B product is obtained, and corresponds exactly to \hat{T} . In what follows we shall refer to the L1B data product as D_B .

Using L1B data, T can be recovered from D_B very easily: it is the inverse Fourier transform of D_B : $T = \mathbf{F}^*\mathbf{Z}_\Omega D_B$. As we shall see later, this simple Fourier inversion leads to potentially very strong Gibbs effects which are partially alleviated (as proposed by [10]) by the use of a Blackman window \mathbf{B} : $T = \mathbf{F}^*\mathbf{B}\mathbf{Z}_\Omega\hat{T}$, which can be seen as a kind of linear Tikhonov regularization:

$$T = \arg \min_T \|\mathbf{B}(\mathbf{F}T - D_B)\|_2^2. \quad (2)$$

This approach has been used for five years with some success. However, linear approaches such as the one described above can be improved by the use of more sophisticated image models that nowadays can be solved thanks to recent advances in optimization techniques, such as modern splitting methods and non-smooth, non-convex optimization methods. In SMOS image restoration, the limitation of linear methods like the one described above are twofold. First, zero-padding in the frequency domain leads to blurry images, specially in the edges. Second, measurements may be polluted by outliers or radio frequency interferences (RFI); indeed, despite the fact that according to international radio regulations, the L-band is exclusively allocated to the Earth observation, soon after SMOS was launched the data revealed there were many signals being transmitted within this protected passive band, rendering a great amount of data unusable for scientific purposes. Because these outliers have frequencies beyond Ω and their power is far stronger than Earth radiation, very strong Gibbs effects can be seen on the final brightness temperature images (see for instance the top left image in Figure 11 corresponding to western Europe, obtained by the inverse Fourier transform of the L1B product).

C. Proposed approach and contributions

In the present work we propose to solve the inverse problem defined by the forward problem (1), by introducing a variational formulation that explicitly models the formation of visibilities as a superposition of the Earth's natural brightness temperatures and the RFI emitters. As it will be demonstrated later, the proposed approach automatically removes signal effects generated from illegal emitters (outliers), while at the same time extrapolates the image spectrum in order to minimize Gibbs effects. Another aspect in which our approach differs from the previous SMOS image restoration techniques, is that instead of using the L1B product, we work directly with the L1A data product (the visibilities). As we will show in the following sections, this seemingly simple data change is not easy to implement because of the lack of regularization in the L1A data, and because of several issues that have

to be considered to make the inverse problem numerically tractable. To our knowledge, this is the first work that tackles the problem directly and exclusively from the L1A product. A preliminary, short version of this work was presented in IGARSS 2014 [11].

The article is organized as follows. In Section II we present the model on which our restoration approach is based, and we explain how this model, stated as a variational problem, can be solved with state of the art optimization algorithms. Then, in Section III we describe in detail a few aspects that are fundamental to make the problem numerically tractable and to improve results. The definitive implementation is presented in Section IV, where we discuss major issues like parameter choices and numerical optimization. In Section V we present experiments with synthetic data that validate the proposed approach, we apply our method to real L1A SMOS data and we compare it with a similar approach we proposed in [12], which is based on the L1B data product. Conclusions and future work are presented in Section VI.

II. MODELIZATION AND VARIATIONAL FORMULATION: A FIRST APPROXIMATION TO THE PROBLEM

Visibilities and brightness temperatures are related by the linear operator given by (1). In its discrete form, using matrix notation¹, this is $\mathbf{G}\tilde{T} = V$. As before, the goal is to obtain the original temperatures image \tilde{T} from the given visibilities V , knowing that, as noted before, \mathbf{G} is not invertible. We propose to model the observed brightness temperature image \tilde{T} as

$$\tilde{T} = T + O, \quad (3)$$

where T is the non-polluted Earth's brightness temperature image and O is the outliers' image generated by the RFIs that are assumed to be sparse pointwise sources (delta functions). Because of the physical acquisition system, the visibilities are corrupted with noise, that under realistic assumptions and after a whitening process can be considered to be Gaussian with 0 mean and standard deviation σ_n . The final equation is then:

$$\begin{aligned} \mathbf{G}\tilde{T} &= \mathbf{G}(T + O) = V \\ D_A &= V + n, \end{aligned} \quad (4)$$

where V is the vector of ideal visibilities, and D_A is the vector of measured visibilities which are corrupted by additive Gaussian noise n as described before.

A. Variational formulation

We propose to recover u by solving the following constrained optimization problem:

$$\begin{aligned} \min_{T,O} \{ & \text{TV}(T) + \mu S(O) \} \\ \text{s.t. } & \|\mathbf{G}(T + O) - D_A\|_2^2 \leq |\Omega|\sigma_n^2, \end{aligned} \quad (5)$$

where $\text{TV}(\cdot)$ denotes the total variation semi-norm and $S(\cdot)$ is a *norm* such that when minimized promotes sparsity, for instance the ℓ_1 norm or ℓ_0 counting measure. The total variation is used to super-resolve T beyond the spectral support

Ω while avoiding Gibbs oscillations; the sparsity operator is chosen as the model for the outliers' image O since this image is zero almost everywhere and the RFIs are sparse. Parameter μ controls the trade-off between both terms; its choice can be formally derived from geometric considerations on the outliers, and will be discussed in detail in Section IV-A. The data fit term is derived directly from the data noise model $n = \mathbf{G}(T + O) - V$. The bound $|\Omega|\sigma_n^2$ represents the area of Ω times the visibilities variance σ_n^2 . This variance as explained before, is a realistic approximation of the acquisition system noise (see Section IV for further explanation).

As usual, problem (5) can be reformulated as an unconstrained one:

$$\min_{T,O} \|\mathbf{G}(T + O) - D_A\|_2^2 + \lambda(\text{TV}(T) + \mu S(O)), \quad (6)$$

where the Lagrange multiplier λ will be chosen to take the unique value for that ensures equivalence between both problems. The derivation of this value will also be addressed in Section IV-A.

Assume for now that $S(\cdot) = \ell_1$. Then this functional is convex, and despite not being strictly convex, for the case of $S(\cdot)$ and $\text{TV}(\cdot)$ it has a unique local minimum (see for instance [13], [14]). However, a careful treatment has to be taken since the term multiplied by λ is non-differentiable. Fortunately, there exist optimization methods to solve this kind of problems, for instance the *Forward-Backward splitting* algorithm [15]. We briefly describe it here for the sake of completeness. Let $E(x) = E_1(x) + E_2(x)$, where E_1 and E_2 are convex functions such that E_1 is differentiable with Lipschitz gradient, and E_2 is a *simple* function, in the sense that its associated proximal operator

$$\text{prox}_{\gamma E_2}(x) = \arg \inf_y E_2(y) + \frac{1}{2\gamma} \|x - y\|^2 \quad (7)$$

admits a closed form or a simple algorithm to compute it. If these hypotheses hold, the following generic algorithm can be derived: for each $k \in \mathbb{N}$, the k -th iteration starting from seed $x^0 = (T^0, O^0)$ is given by

$$\begin{cases} x^{k+1/2} &= x^k - \gamma \nabla E_1(x^k) \\ x^{k+1} &= \text{prox}_{\gamma E_2}(x^{k+1/2}). \end{cases}$$

In order to ensure convergence to the minimizer, γ must be smaller than $2/L$, where L is the Lipschitz constant of ∇E_1 . Now, setting

$$\begin{aligned} E_1(T, O) &= \|\mathbf{G}(T + O) - D_A\|_2^2, \\ E_2(T, O) &= \lambda(\text{TV}(T) + \mu S(O)) \end{aligned}$$

yields

$$\begin{aligned} \nabla E_1(T, O) &= \begin{pmatrix} \mathbf{G}^* \mathbf{G}(T + O) - V \\ \mathbf{G}^* \mathbf{G}(T + O) - V \end{pmatrix}, \\ \text{prox}_{\gamma E_2}(T, O) &= \begin{pmatrix} \text{prox}_{\gamma \lambda \text{TV}}(T) \\ \text{prox}_{\gamma \lambda \mu \|\cdot\|_1}(O) \end{pmatrix}. \end{aligned}$$

As for $\text{prox}_{\gamma \lambda \mu \|\cdot\|_1}(O)$, it is straightforward to show from (7) that it corresponds to the *soft thresholding* operator

$$s_\Delta(t) = \begin{cases} \text{sign}(t)(|t| - \Delta) & \text{if } |t| \geq \Delta, \\ 0 & \text{if } |t| < \Delta. \end{cases}$$

¹For the sake of simplicity, we use the same notation to refer to an image and its vectorized form. Disambiguation follows easily from the context.

This algorithm converges to the unique global minimizer that corresponds to the solution of problem (5) with sparsity operator $S(O) = \|O\|_1$. We use this solution as an initialisation for the second step, where the sparsity operator is chosen to be $S(O) = \|O\|_0$, which is non-convex. For this problem, the same Forward-Backward algorithm can be considered and is guaranteed to converge to a local minimizer [16]. Now, instead of the soft thresholding, the proximal operator for $S(O) = \|O\|_0$ becomes the hard thresholding $h_{\sqrt{2\gamma\lambda\mu}}(t) = t \mathbb{1}_{\{|t| \geq \sqrt{2\gamma\lambda\mu}\}}$ (Appendix A).

III. NUMERICAL ISSUES

A. Size of matrix $\mathbf{G}^*\mathbf{G}$

Many papers have already been published on the effects of the assumption of a Fourier like modeling operator, instead of using the actual instrument modeling given by operator \mathbf{G} . It turns out that even if there are some uncertainties on the actual instrument modeling (namely on the antenna patterns), it is however better to consider \mathbf{G} in the inverse problem rather than to assume a Fourier like one (like the restoration method given by (2)), the reconstruction error being much larger with the later one. Of course solving an inverse problem based on \mathbf{G} has an impact on the size of the gradient term and therefore on the computational time and the amount of required memory. More precisely, when we look at the algorithm developed so far, we realize that we perform a multiplication by matrix $\mathbf{G}^*\mathbf{G}$ at each iteration.

Let us give a brief analysis of the size of matrix \mathbf{G} . The dimension of vector T is given by the spatial sampling grid that one chooses to use. A reasonable choice is $128 \times 128 = 16384$. Larger sampling rates lead to larger matrices thus increasing the computational burden. This is the size of vector T is for a single simple polarization, which is the case considered in this work. If we had considered all the polarizations (horizontal, vertical and dual), since in the dual polarization the measurements are complex numbers and each of them is represented with two real entries, the size would have been $16384 \times 4 = 65536$. As for V , The number of antennas determines its dimension. The total number of receptors for the horizontal (or the vertical) polarities is 69, leading to $(69 \times 68)/2 = 2346$ cross-correlations – which are complex numbers – plus 3 supplementary real measurements corresponding to inverse polarization. Since each complex number is represented with two entries (its real and the imaginary parts), the size of V for the horizontal (or vertical) polarization is $2346 \times 2 + 3 = 4695$. This is then the size of the V vector considered in this work. Again, if all the polarizations were considered, the size of V would have been 15996 (4695 (horizontal) + 4695 (vertical) + 3302×2 (dual) = 15996). Consequently, in this work, the dimension of matrix \mathbf{G} – that maps brightness temperatures to visibilities – is 15996×16384 .

It follows that here the dimension of matrix $\mathbf{G}^*\mathbf{G}$ is 16384×16384 . This matrix is a dense matrix, which means that at each iteration, a great number of additions and multiplications are performed. Thus, explicit multiplication by this matrix at each iteration of the algorithm is computationally intractable. However, a change of basis to the Fourier domain yields the

following expression for the gradient term:

$$\nabla E_1(T, O) = \mathcal{F}^*((\mathbf{G}\mathbf{F}^*)^*\mathbf{G}\mathbf{F}^*\mathcal{F}(T + O) - (\mathbf{G}\mathbf{F}^*)^*V).$$

Here, \mathcal{F} denotes the Cooley and Tuckey standard Fast Fourier Transform, that we use instead of matrix multiplication by the hexagonal DFT matrix \mathbf{F} in order to accelerate computations. The use of the standard (rectangular) FFT on a hexagonal lattice can be made possible by re-projecting the samples in a rectangular grid (see [7] or [17, Section III]). The advantage of the previous expression is that it reveals a highly sparse matrix $\mathbf{F}\mathbf{G}^*\mathbf{G}\mathbf{F}^*$: to keep the energy at 99.99%, we only need to keep a fraction of 0.0008 of its entries.

The property of being sparse in the Fourier domain can be easily understood from the image formation point of view: According to (1) visibilities are roughly the Fourier transform of the centered brightness temperatures times a function f which varies smoothly with respect to ξ , k and l . If that function was constant w.r.t. k and l , then \mathbf{G} would be a diagonal matrix. This is not the case, but since f is relatively smooth, the off-diagonal coefficients of \mathbf{G} exhibit fast decay. Consequently, we can modify the Forward-Backward algorithm described above to perform the minimization in the Fourier domain. This algorithm is summarized in Algorithm 1.

Proposed method with Forward-Backward implementation

input : An upper bound $L \geq L(E_1)$ on the Lipschitz constant $L(E_1)$ of $\nabla(E_1)$

output: A brightness temperature image T , an outliers image O

initialization:

Step 0: Set $T^0 = 0$ and $O^0 = 0$;

Step k ($k \geq 0$):

$$\begin{cases} T^{k+1/2} &= T^k - \gamma \mathcal{F}^*(\mathbf{F}\mathbf{G}^*\mathbf{G}\mathbf{F}^*\mathcal{F}(T^k + O^k) - \mathbf{F}\mathbf{G}^*)V \\ O^{k+1/2} &= O^k - \gamma \mathcal{F}^*(\mathbf{F}\mathbf{G}^*\mathbf{G}\mathbf{F}^*\mathcal{F}(T^k + O^k) - \mathbf{F}\mathbf{G}^*)V \\ T^{k+1} &= \text{prox}_{\gamma\lambda \text{TV}}(T^{k+1/2}) \\ O^{k+1} &= s_{\gamma\lambda\mu}(O^{k+1/2}). \end{cases}$$

Algorithm 1: Proposed method with Forward-Backward implementation in the Fourier domain.

B. Spectral TV

In order to reduce the “staircasing” effect inherent to many TV minimization methods [18], [19], we use an approach inspired by the *Spectral TV* method proposed by Moisan [20]. In Moisan’s approach, staircasing reduction is achieved by: (i) Computing image derivatives not by finite differences but analytically on the Fourier series expansion; (ii) Approximating the continuous TV as a Riemann integral over a grid at least two times finer than the critical sampling rate. In our case, instead of doubling the sampling rate of ∇T , we chose to extend the spectral domain of T to an intermediate cell \mathcal{H} , in-between the star domain Ω where measures \hat{O} are taken, and the cell \mathcal{C} corresponding to the (largely overcritical) spatial

sampling rate of T (Figure 4).

There is another reason to use the Spectral TV on this

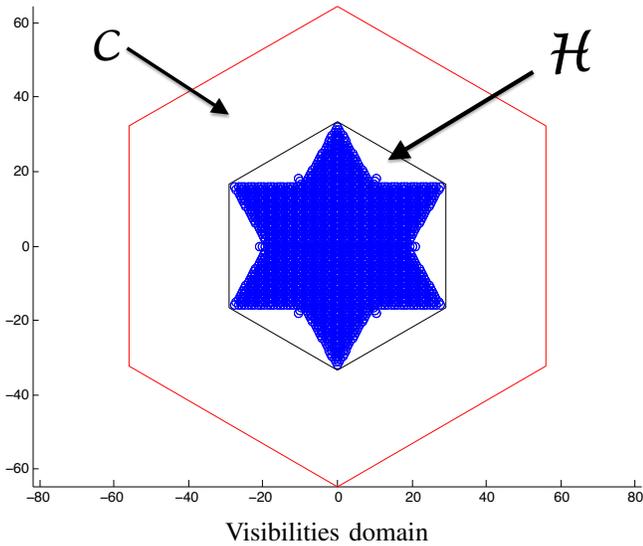


Fig. 4. Intermediate, hexagonally shaped cell \mathcal{H} used to reduce the staircasing effect inherent to discrete TV minimisation methods.

problem. SMOS data is captured on a hexagonal grid, and thus the variables T and O are modeled on a hexagonal grid as well. Most TV algorithms are based on the computation of derivatives using discrete differences, that are not straightforward to adapt to the hexagonal grid. As stated before, the Spectral TV does not use finite differences to compute the image derivatives: they are computed analytically in the Fourier expansion. What is more, because of how the grids are defined on SMOS, we could implement the Fourier derivatives using FFT, as explained before. In the following, we denote the Spectral TV of an image T based on the domain \mathcal{H} as $TV_{\mathcal{H}}(T)$.

IV. FINAL IMPLEMENTATION

A. Parameters

1) *Visibility noise estimation:* The MIRAS instrument has 69 antennas of two different types: 66 antennas are Light-Weight Cost-Effective Front-End (LICEF) and the other three are Noise Level Injection Radiometers (NIR). Each of these receivers has different noise levels. Because each receiver from one arm correlates with all the receivers from the other two arms, we can have three possible combinations: LICEF-LICEF, LICEF-NIR and NIR-NIR (this last one is not used here). The LICEF-LICEF baseline is known as L-baseline while LICEF-NIR is a *mixed* baseline, noted as M-baseline. The SMOS baseline is formed of 2145 L-baselines and 198 M-baselines. The noise in visibilities, while not Gaussian, can be well approximated as Gaussian noise with standard deviation $\sigma_{VL} = 0.098K$ and $\sigma_{VM} = 0.21K$ for the L-Baselines and M-Baselines, respectively. Then, a whitening transform can be applied to the visibilities resulting in a diagonal covariance matrix with fixed diagonal entries. To summarize, we consider the visibilities (after the

forementioned transformation) to be i.i.d. with Gaussian distribution of zero mean and variance $\sigma_n = 0.1$. This explains the squared ℓ_2 norm and the bound used for the data fit term in the model presented so far.

2) *Selection of μ :* We propose to balance the trade-off between the sparsity term $S(O)$ and the regularity term $TV_{\mathcal{H}}(T)$ based on modeling the outliers in the spatial domain as follows. For a cylinder c of radius r and height h , the involved norms or semi-norms are

$$TV(c) = 2\pi r h, \quad \|c\|_0 = \pi r^2 \mathbb{1}_{[h>0]}, \quad \|c\|_1 = \pi r^2 h.$$

The selection of μ determines whether the cylinder is considered to be an outlier or part of the image to be recovered. In other words, the value of μ controls whether a given cylinder is considered to belong to O or to T .

When the sparsity operator is the ℓ_1 norm, c is considered as an outlier if $TV(c) \geq \mu \|c\|_1$, leading to $\mu \leq \frac{2}{r}$. In the examples, we have selected a value $\mu \simeq \frac{2}{10} = 0.2$, which amounts to consider that the radii of the outliers are at most 10 pixels wide.

Note that since the resolution in the brightness temperature map depends on the position in the image, for a fixed value of 10 pixels the support of the outlier in meters depends on its location. Notice moreover that the coefficients neglected in $\mathbf{FG}^* \mathbf{GF}^*$ in order to make the inversion numerically tractable correspond to high frequency components, and therefore some amount of spread is introduced. The amount of spread being signal dependent, for outliers of low intensity the value of 10 is certainly too conservative, but for high intensities the spread may be large and a choice of 10 pixels may be necessary to ensure their complete elimination. A possible additional strategy would consist in measuring the maximum brightness temperature obtained by the zero padding method, and automatically lower the number of outliers' support pixel to 1 or 2 if the maximum brightness temperature falls within regular Earth's brightness temperature. This strategy was followed in the synthetic experiment designed to show the spectral extrapolation capability of the proposed approach, where the outliers' support was set to 1 pixel (Section V-A, Figure 9).

In the case of the ℓ_0 norm, a cylinder is considered as an outlier as soon as $TV(c) \geq \mu \|c\|_0$, i.e. $\mu \leq \frac{2h}{r}$. Hence, μ can then be interpreted as the minimal h/r outlier ratio. Here, contrarily to the ℓ_1 case, the height (here expressed in Kelvin units) plays an important role in distinguishing an outlier from the data, and larger values for μ may be chosen ($\mu \simeq 100 \frac{2}{10} = 20$).

3) *Selection of λ :* Parameter λ has to be chosen carefully in order to make the unconstrained problem equivalent to the constrained one. A wrong choice may drastically affect the results. The correct value of λ depends of course on $|\Omega| \sigma_n^2$, which is given. A classical way to set λ if we know the expected noise variance σ_n^2 is to use Uzawa's algorithm [21]:

$$\begin{cases} (T^{k+1}, O^{k+1}) = \min_{T, O} \|\mathbf{G}(T + O) - D_A\|_2^2 \\ \quad + \lambda^k (\text{TV}(T) + \mu S(O)), \\ \lambda^{k+1} = \max\{\lambda^k + \rho \|\mathbf{G}(T^{k+1} + O^{k+1}) - D_A\|_2^2, 0\} \end{cases} \quad (8)$$

The procedure consists in starting with a fixed λ^0 , then solving the original problem with this λ^0 and adjusting λ^k iteratively, depending on how near or how far the solution is to the original constraint (which is the noise present in the model). The procedure continues with ensured convergence until the data fit term is close enough to the expected noise $|\Omega|\sigma_n^2$.

B. Numerical optimization

Several methods have been proposed in the literature to accelerate the Forward-Backward convergence rate. Two related approaches are FISTA [22] and monotone-FISTA [23].

In FISTA, the main difference with the *Forward-Backward* is that each iteration depends on the two previous ones, whereas in the *Forward-Backward* algorithm it only depends on the last one. This FISTA algorithm does not ensure monotone convergence, which is not a desirable property for an optimization algorithm. In [23], the authors introduce a monotone version of this algorithm known as MFISTA (Algorithm 2), that does not change the convergence rate of the original FISTA algorithm, but guarantees monotonicity. Note

MFISTA

input : An upper bound $L \geq L(E_1)$ on the Lipschitz constant $L(E_1)$ of $\nabla(E_1)$

output: x , which is the solution to the problem

initialization:

Step 0: Set $y^1 = x^0$ where x^0 is some initial value, and $t^1 = 1$;

Step k ($k \geq 1$):

$$\begin{cases} y^{k+1/2} = y^k - \frac{\gamma}{L} \nabla E_1(y^k) \\ z^k = \text{prox}_{\frac{\gamma}{L} E_2}(y^{k+1/2}) \\ t^{k+1} = \frac{1 + \sqrt{1 + 4(t^k)^2}}{2} \\ x^k = \arg \min\{E(x) : x = z^k, x^{k-1}\} \\ y^{k+1} = x^k + \frac{t^k}{t^{k+1}}(z^k - x^k) + \frac{t^k - 1}{t^{k+1}}(x^k - x^{k-1}) \end{cases}$$

Algorithm 2: Monotone-FISTA

that x^k can take two values: z^k or x^{k-1} . Depending on the value it takes, the update of the y^{k+1} may be the same as the one performed on FISTA: $y^{k+1} = x^k + \frac{t^k - 1}{t^{k+1}}(x^k - x^{k-1})$ or it may be given by $y^{k+1} = x^k + \frac{t^k}{t^{k+1}}(z^k - x^k)$.

The final and complete algorithm is summarized in Algorithm 3. The number of iterations and convergence rate depend on each particular input data: if no outliers are present in the image the method converges much faster. The general process is slow: in the examples where outliers are present it takes more than 100,000 iterations, corresponding to several hours.

The main reason is that in this case, the initial images are initialized with identically null images. A simple improvement consist on initializing the method with the results obtained from the zero padding inversion. In this case, the method converges in the order of thousands of iterations (in the worse case) but in any case it takes several minutes to converge. It is worth mentioning that these reported times were obtained with a code that is not fully optimized, and where no parallelization strategies (which are possible) were performed.

FINAL PROPOSED ALGORITHM

input : L1A SMOS Data - Visibilities: D_A

output: A temperature image T , an outliers image O

initialisation: Set $T^0 = 0$ and $O^0 = 0$;

Initialize Uzawa: Set $\lambda^0 = 1$ and ρ ;

while $s_{uzawa} > uzawa_tol$ **do**

Initialize MFISTA: Set $\tilde{T}^0 = T^0$, $\tilde{O}^0 = O^0$ and $t^0 = 1$

while $s_{fista} > mfista_tol$ **do**

$$\begin{pmatrix} T_{temp}^k \\ O_{temp}^k \end{pmatrix} = \text{prox}_{\frac{\gamma}{L} E_2} \left(\begin{pmatrix} T^k \\ O^k \end{pmatrix} - \frac{\gamma}{L} \nabla E_1(\tilde{T}^k, \tilde{O}^k) \right)$$

$$t^{k+1} = \frac{1 + \sqrt{1 + 4(t^k)^2}}{2}$$

if $E(T_{temp}^k, O_{temp}^k) < E(T^{k-1}, O^{k-1})$ **then**

$$T^k = T_{temp}^k ;$$

$$O^k = O_{temp}^k ;$$

$$\tilde{T}^{k+1} = T^k + \frac{t^k - 1}{t^{k+1}}(T^k - T^{k-1}) ;$$

$$\tilde{O}^{k+1} = O^k + \frac{t^k - 1}{t^{k+1}}(O^k - O^{k-1}) ;$$

else

$$T^k = T^{k-1} ;$$

$$O^k = O^{k-1} ;$$

$$\tilde{T}^{k+1} = T^k + \frac{t^k}{t^{k+1}}(T_{temp}^k - T^k) ;$$

$$\tilde{T}^{k+1} = O^k + \frac{t^k}{t^{k+1}}(O_{temp}^k - O^k) ;$$

end

$$s_{fista} = E(T^{k-1}, O^{k-1}) - E(T^k, O^k)$$

end

$$s_{uzawa} = \lambda^k + \rho \|\mathbf{G}(T^k + O^k) - D_A\|_2^2 ;$$

$$\lambda^{k+1} = \max\{\lambda^k + \rho \|\mathbf{G}(T^k + O^k) - D_A\|_2^2, 0\}$$

end

Algorithm 3: Final proposed algorithm

V. EXPERIMENTAL RESULTS

To show the advantages of the proposed framework, we present two sets of experiments: synthetic simulated data, and real data. We first present results on a set of simulated images in order to evaluate qualitative and quantitative aspects of the proposed method. We compare our approach to previous works: a Fourier inversion of the L1B data, a simple Blackman apodization to smooth the outliers effects, and our previous L1B minimisation framework presented on [12].

A. Simulated Data

Because we are dealing among others with outliers removal, we analyze the performance of the proposed method in the following contexts:

- No outliers are present in the image;
- Outliers with intensities in the same range as the Earth's brightness temperature
- Outliers with intensities on a larger range ($[0,35000]$ K), that corresponds to the range we have observed on real data;
- Outliers located outside the image grid, that is, with sub-pixel accuracy in non-integer positions.
- Only the synthetic Earth's brightness temperature is considered (no instrumental noise nor outliers are added). This scenario is considered to evaluate the extrapolation power of the proposed approach.

To perform this analysis we need to simulate the data. The process to generate the simulated images is the following:

- 1) We generate a temperatures image based on the principles described on [24]. We denote this ground truth image by T_{gt} ;
- 2) With the given visibility variance σ_n^2 , we generate the noise visibility n_v as a random vector with independent, identically distributed Gaussian entries with zero mean and covariance matrix $\sigma_n^2 \mathbf{Id}$;
- 3) With the obtained image T_{gt} we compute a ground truth visibility image: $V_{gt} = \mathbf{G}T_{gt}$;
- 4) We add the noise visibilities to the ground truth visibilities to obtain an outlier-free noisy visibility snapshot: $\tilde{V} = V_{gt} + n_v$;
- 5) We now generate a set of outliers' images, one for each of the cases described above (no outliers, outliers in the image range, outliers between $[0,35000]$ K and outliers outside the grid positions). Let us denote by T_δ any of these synthetic outliers' images. We can now obtain the corresponding induced visibilities: $V_\delta = \mathbf{G}T_\delta$;
- 6) Finally, we obtain the simulated visibilities adding both visibilities: $V_f = \tilde{V} + V_\delta = V_{gt} + V_\delta + n_v$, which are in full agreement with the image formation model considered in this work.

The outliers positions and values are generated randomly, and the number of outliers ranges from 2 to 10 (although this is not a limitation but is based on the observation that rarely a real image has more than 10 outliers).

Figure 5 shows the result of our method when applied to an image where no outliers are present. We recall that no modification to the functional is needed: we still consider the sparsity operator on an image O , whose values at the end of the minimization are all zero. Regarding the Earth's brightness temperature estimation T , because part of the error comes from the attenuation of the image edges due to the bandpass characteristics of the MIRAS instrument and the truncation of $\mathbf{FG}^*\mathbf{GF}^*$ entries, we define a flat zone of 60×60 pixels (in this case in the central part of Spain) and we compute the error in this flat zone. In this flat zone, the error is as expected, smaller than $5K$, which is in agreement with the expected temperature noise model, $\sigma_T = 5.0 K$ [25]. More details are given in the figure's caption.

Figures 6 and 7 show the results obtained when outliers are present in the scene. In the first case, the outliers intensity fall within the Earth's brightness temperature range; in the second

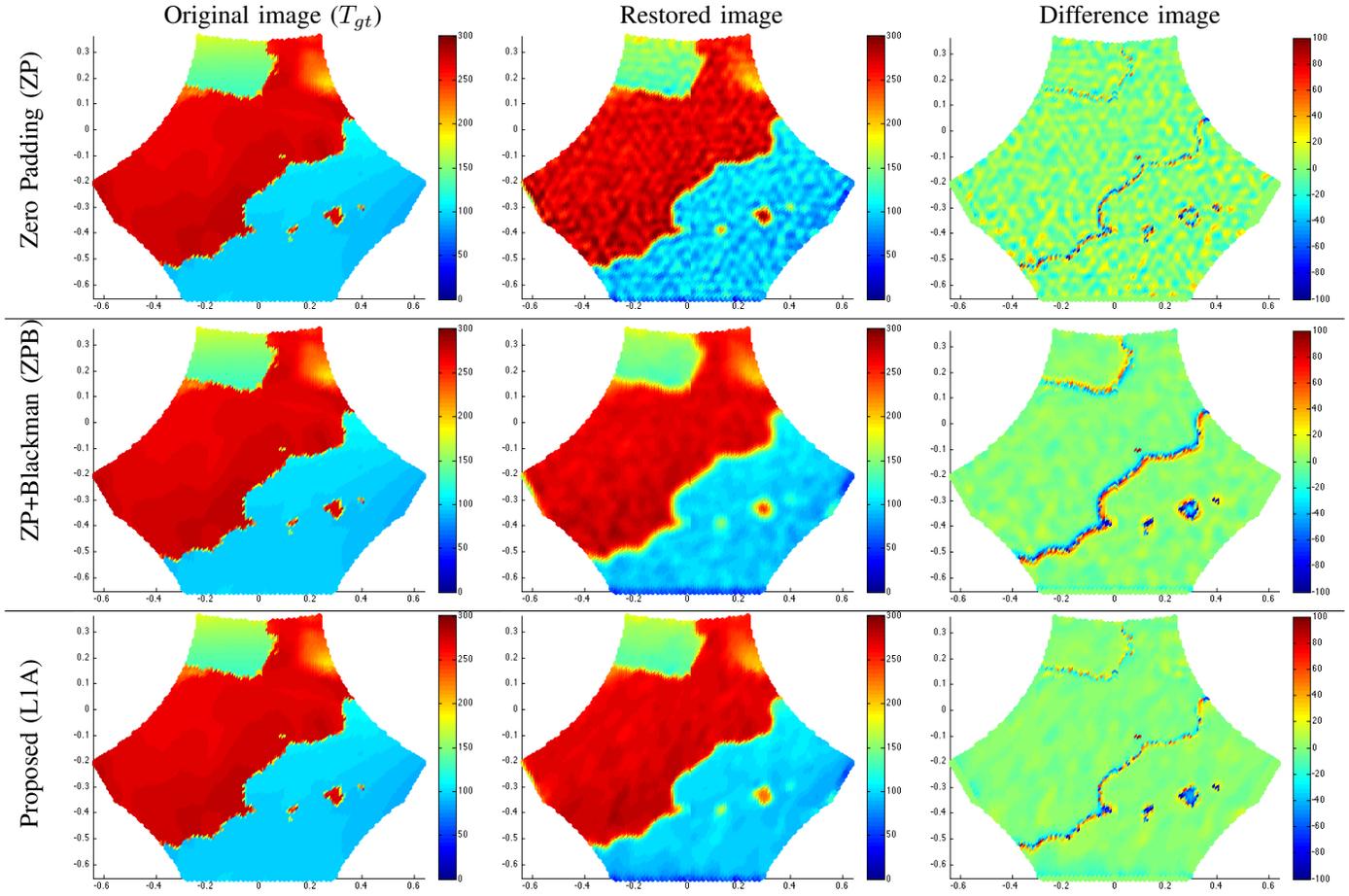
case, the outliers intensities range from 0 to 35000 K. Here again, the proposed method is consistently better than the zero padding approach and even in the presence of very big outliers (35000 K) the ℓ_∞ error is 200 times smaller, leaving an error of near 170 K, which is in the order of the ground truth image values. See the figures' captions for more details. Note that despite the improved accuracy in brightness temperature T , the L1A method shows more artifacts in the image of outliers O . This is due to the approximation of the matrix $\mathbf{FG}^*\mathbf{GF}^*$ by a sparse matrix in the Fourier domain. The approximation level was in fact tuned to smooth temperature images having a relatively fast decay in the Fourier domain. Outliers images, on the other hand, decay much slower and are therefore more sensitive to the Gibbs effects produced by spectral truncation. When accurate detection of RFIs is important, our algorithm should use the full \mathbf{G} matrix at least during the last few iterations. Localization accuracy of RFIs can also be further improved by increasing the spatial resolution of our algorithm or by switching to a continuous outlier model as discussed in Section VI.

In Figure 8 we present the results obtained by the considered methods when the outliers are not located on the sampling grid points. It can be observed that the performance of the proposed method continues to be consistent with the case where the outliers were located on the grid.

Table I gives quantitative measures for the results obtained with the proposed method in all the simulated scenarios. It is clear that the method behaves consistently in all scenarios. We can conclude that the method is well adapted for several contexts that may occur on real data, and that no context-dependent strategies are needed in order to apply it. In other words, the method and its parameters can be set once for all, independently on the observed data. We recall that in all cases the standard deviation of the residual measured on a flat zone of the spatial domain was smaller than $5K$, the expected temperature noise model. We can see from this table a consistent improvement on the use of the proposed method compared to the previous one.

We end this set of experiments on synthetic data with an analysis of the "spectral extrapolation" capability of the proposed method.² A comparison with the zero padding approach is presented. In order to evaluate spectral extrapolation, the input image is the simulated Earth's brightness temperature: no instrumental noise or outliers are added. For this experiment, the outlier's support in the proposed model was set to 1 pixel, following the strategy described in Section IV-A2. Results are displayed on Figure 9. See caption for details. The Fourier Transform of the obtained images, not presented here, clearly shows for the proposed approach the hexagonal spectral support standing out from the image background; for the zero padding, what stands out is the star-shaped visibilities sampling domain Ω . A similar experiment showing these results patterns can be seen in [12, Figure 3].

²By this we do not pretend to create new information. TV and ℓ_1 "spectral extrapolation" only recreate the high frequency coefficients that are necessary to avoid Gibbs effects that would otherwise be created by a sharp discontinuity in the spatial domain.



	ℓ_1 (K)	ℓ_2 (K)	ℓ_∞ (K)	ℓ_1 flat zone (K)	ℓ_2 flat zone (K)	Relative error ℓ_1	Relative error ℓ_2
$T_{ZP} - T_{gt}$	9.103	13.160	108.523	7.331	9.126	4.75%	6.31%
$T_{ZPB} - T_{gt}$	6.474	13.612	143.147	2.885	3.639	3.38%	6.52%
$T_{L1A} - T_{gt}$	4.500	10.285	146.068	2.149	2.768	2.35%	4.93%

Fig. 5. Comparison with previous approaches when no outliers are present in the image. Notice the improvement both in the denoising power (the error on the flat zone drops down to $2.8K$) as well as on the spectral extrapolation (that can be easily seen on the edges of the difference images). This experiment confirms that the proposed method is generic and that it performs well even when no outliers are present. Finally, the image O obtained was a constant null image, in agreement with the model. Nevertheless, it is important to note that small zones are removed: for instance Ibiza and Formentera islands are completely removed and Menorca is smoothed. Note that precision loss also occurs on the zero padding approach where Formentera is also removed and Menorca is very smooth. Because we could not obtain an original restored image, we don't know if this problem is related to the reconstruction process or is the bandpass effect of the MIRAS instrument. We have tested our method without the outliers term in order to understand if these small islands were considered as outliers. This was not the case (O is zero everywhere) which leads us to conjecture that the problem is more related to the bandpass nature of the acquisition system.

	ℓ_1 (K)	ℓ_2 (K)	ℓ_∞ (K)	ℓ_1 flat zone (K)	ℓ_2 flat zone (K)
No outliers present in the image	6.376	12.982	146.068	2.149	2.768
Outliers with values on the image range	6.723	12.729	147.356	2.692	3.329
Outliers in the range [0,35000] K	6.366	14.245	170.410	1.657	2.050
Outliers outside grid position (25000 K)	6.552	14.129	166.184	1.533	2.032

TABLE I

QUANTITATIVE RESULTS OBTAINED WITH THE PROPOSED METHOD FROM SIMULATED DATA, IN DIFFERENT SCENARIOS: WITHOUT OUTLIERS, WITH OUTLIERS INTENSITIES WITHIN THE RANGE OF THE IMAGE, WITH OUTLIERS LOCATED ON THE SAMPLING GRID POSITIONS AND OUTSIDE OF IT.

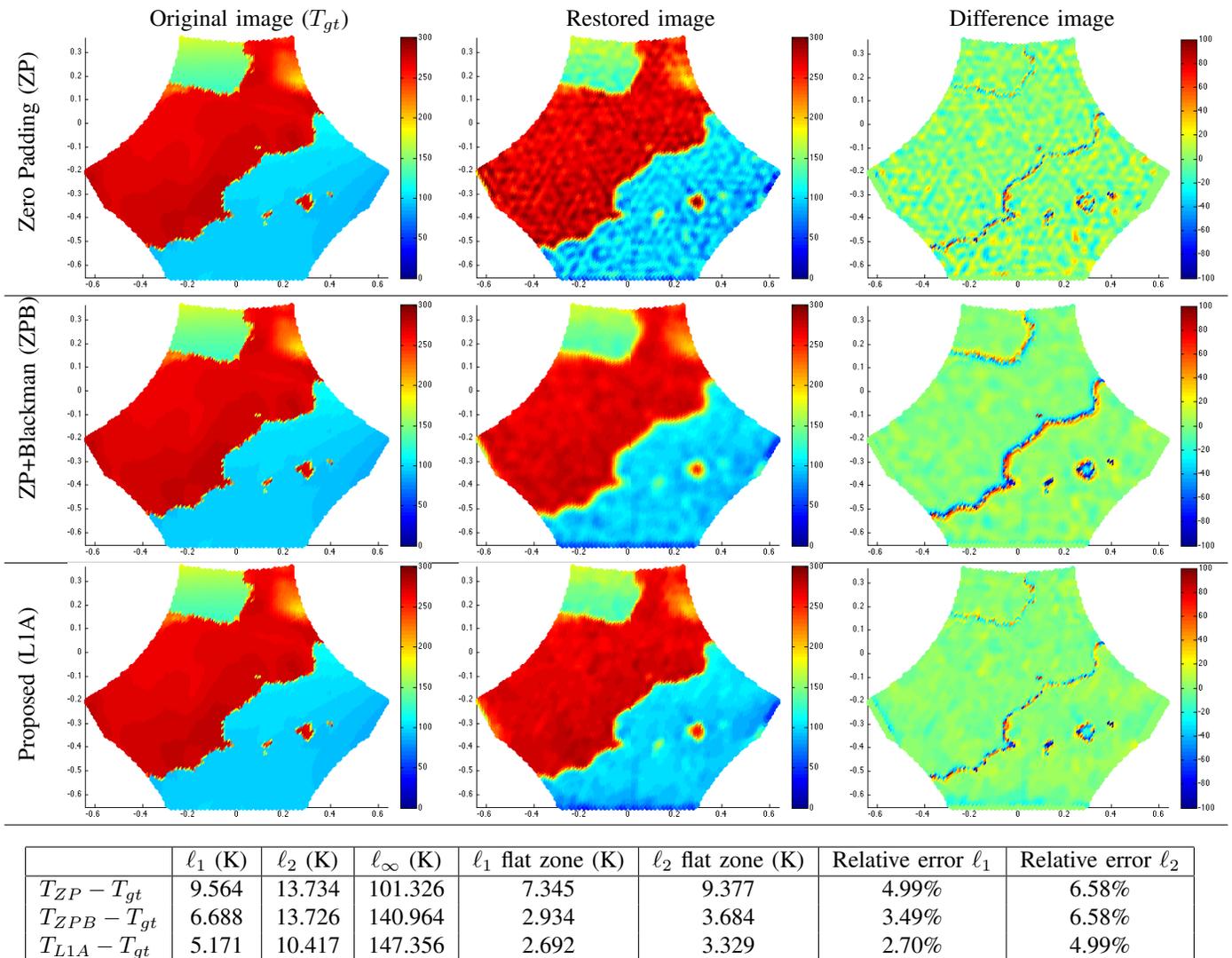


Fig. 6. Analysis of the different methods when the outliers intensities fall within the range of the Earth's brightness temperatures.

B. Experiments on real data

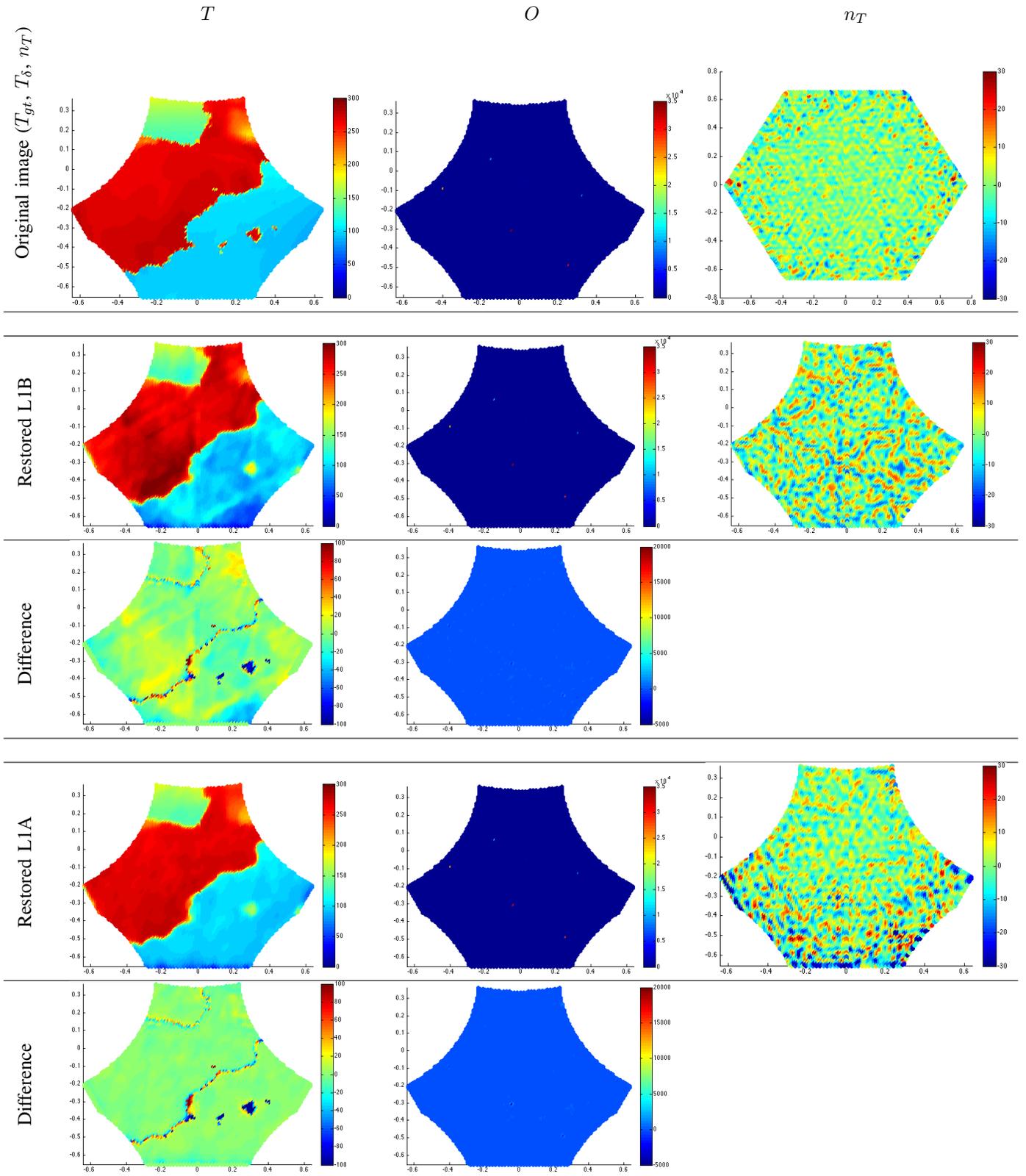
The second set of experiments was run on real data. We compare results from our approach to those obtained by previous works: a Fourier inversion of the L1B data, a simple Blackman apodization to smooth the outliers effects, and our previous L1B minimisation framework presented on [12]. Experiments were performed on several snapshots from the SMOS dataset of march 2010. Here, for the sake of brevity, we illustrate typical results presenting two snapshots. We have set σ_n equal to 0.1, which is the measurement error reported by the SMOS mission.

Figures 11 and 12 show the results obtained for snapshots 996 and 1050. For geographic reference, Figure 10 shows how these regions look in Google Earth from approximately the same viewing angles as the SMOS acquisitions. Note that the acquired images are corrupted by several outliers that considerably degrade the data. It is clear that the present method outperforms both the direct inverse Fourier transform and the Blackman apodization. It also improves the results from our previous work based on L1B data, which resulted in more regularised images and thus, valuable details where



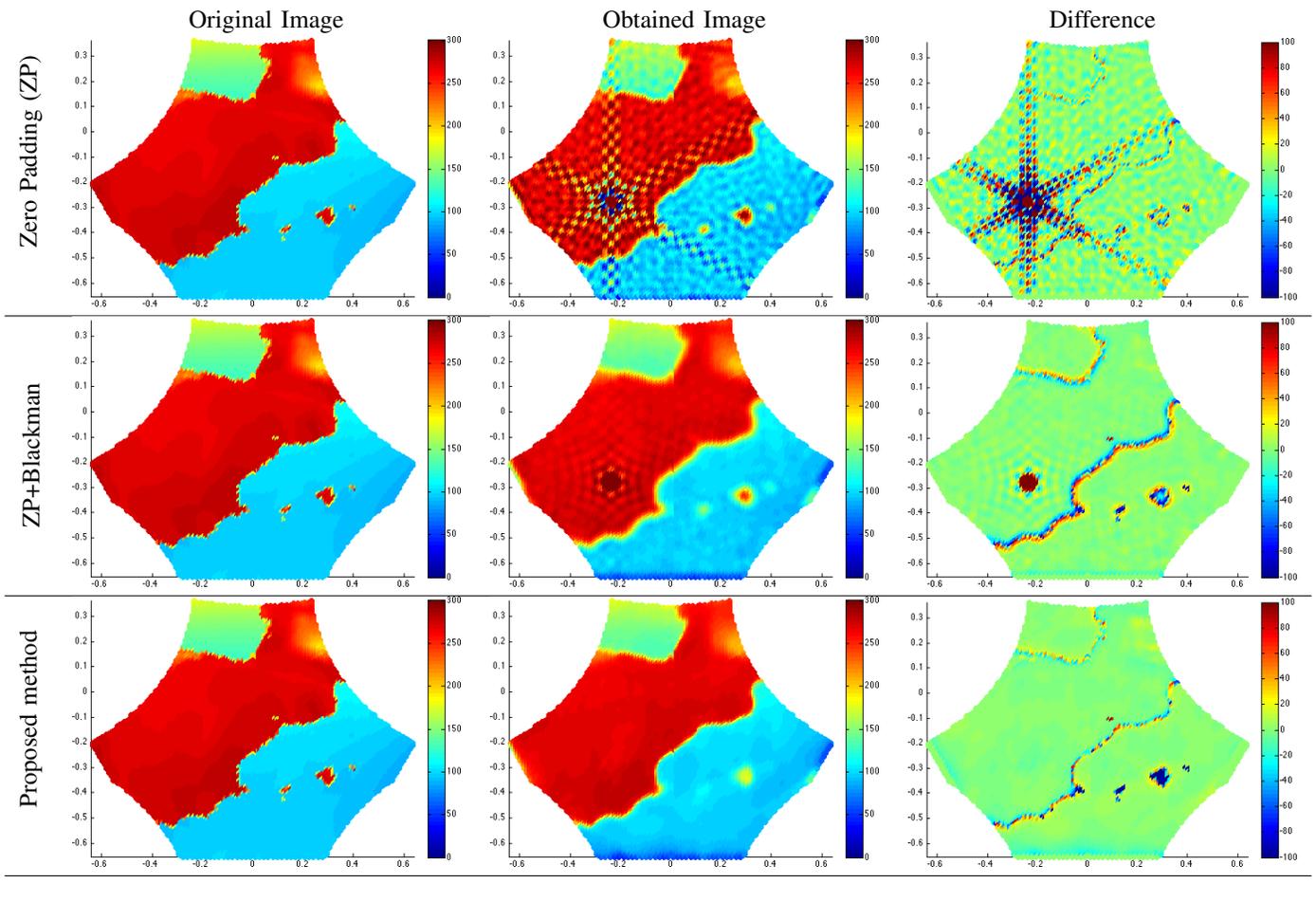
Fig. 10. Google Earth view of two of the regions used on the experiments. The left one corresponds to snapshot 996 (middle and northern Italy, Switzerland, Austria, Slovenia and Croatia) and the right one to snapshot 1050 (western UK, northern France, Belgium, Netherlands, northern Germany, eastern Poland, Denmark, Norway and Sweden). Central European time zone, UTC/GMT+01:00.

lost. An effect that can be seen, in particular in the L1A-based restored brightness temperature image of central Europe, is the correlation between land humidity and topography. Note, for instance, the consistency of the humidity flow coming from the Adriatic Sea and entering inlands in northern Italy, stopped by the barrier imposed by the Alps.



	l_1 (K)	l_2 (K)	l_∞ (K)	l_1 flat zone (K)	l_2 flat zone (K)	Relative error l_1	Relative error l_2
$T_{L1B} - T_{gt}$	9.875	16.107	173.525	8.769	10.588	5.15%	7.72%
$T_{L1A} - T_{gt}$	4.674	13.083	170.410	1.657	2.050	2.44%	6.27%

Fig. 7. Results obtained with simulated data when several outliers with different intensities are present on the image. In this Figure, images obtained by zero padding and Blackman are omitted since they are severely degraded by the presence of strong outliers (similar results to those presented in Figure 8). In this example, outliers were generated on the following grid positions with its corresponding values: (-0.0357,-0.3093), 35.000 K; (-0.1429,0.0619), 10000 K; (-0.3929,-0.0928), 25000 K; (0.5714, -0.1753), 800 K; (0.3304,-0.1289), 8000 K; (0.2589, 0.2629), 35000 K; (0.2589,-0.4897), 30000 K; (0.0357,-0.0515), 2000 K.



	ℓ_1	ℓ_2	ℓ_∞	ℓ_1 flat zone	ℓ_2 flat zone	Relative error ℓ_1	Relative error ℓ_2
$T_{ZP} - T_{gt}$	21.118	87.283	3335.699	61.675	155.806	11.00%	41.82%
$T_{ZPB} - T_{gt}$	8.240	41.006	1185.825	16.971	67.185	4.30%	19.65%
$T_{L1A} - T_{gt}$	4.406	11.946	166.184	1.533	2.032	2.30%	5.72%

Fig. 8. Analysis of the different methods when the outliers are not located on the sampling grid. Note that the method performs well, with a performance similar to the one obtained when outliers are on grid positions. In this experiment, the outlier introduced in the image has an intensity of 20000 K.

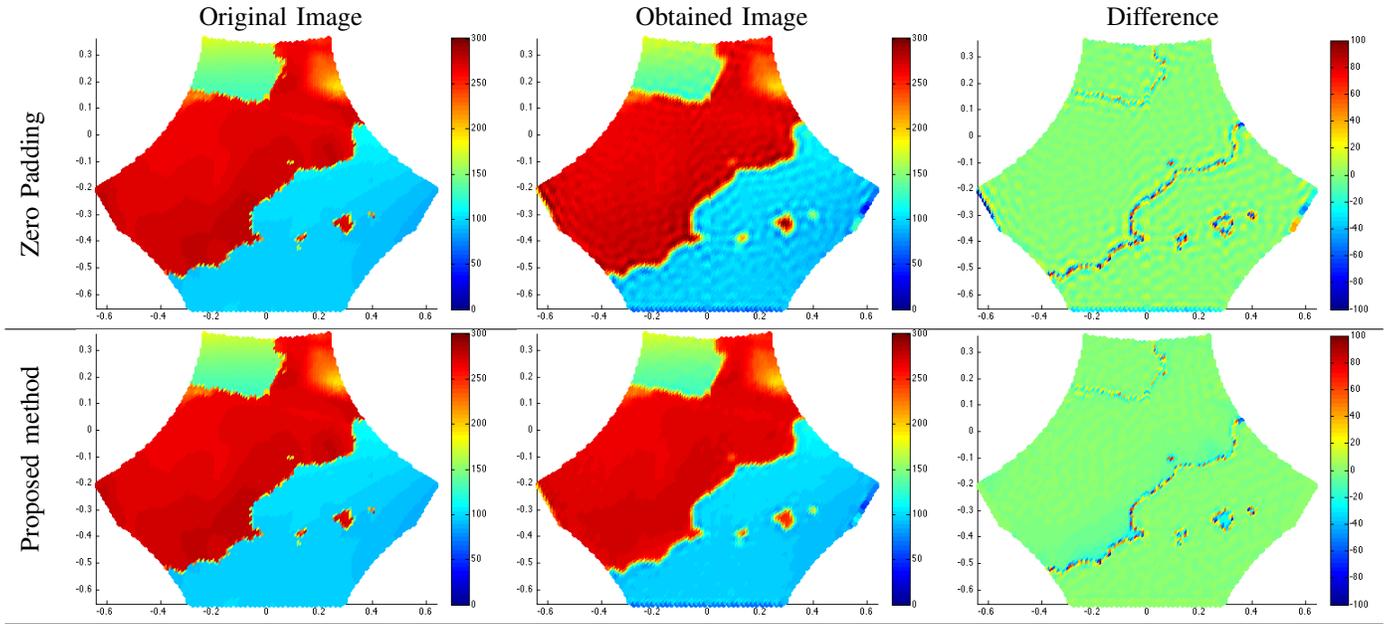
VI. CONCLUSIONS AND FUTURE WORK

In this work we have proposed a novel approach for the restoration of images acquired by the SMOS mission. Two fundamental contributions are presented: First, a variational approach that seeks to restore the image of the Earth's brightness temperature on one side, and the image of outliers or RFIs on the other side. The second contribution is the use of the visibilities or L1A product as the data for the restoration problem; up to our knowledge, this is the first time this product is used for the restoration of SMOS images. We use also a realistic noise model for the data, provided by CESBIO (Centre d'Etudes Spatiales de la Biosphère, France), the agency responsible for the SMOS mission.

The parameters in the model are fixed once for all or automatically adjusted by the optimization procedure, since they are either derived from physical modeling or obtained by formal procedures. Experiments on synthetic data show the ability of our method to recover the Earth's brightness temperature with high precision. In summary, both experi-

ments on synthetic and real data confirm the suitability of the proposed method, and show that results are of very high quality, outperforming previous approaches proposed for SMOS images restoration.

A side effect of the approach is the possibility to do subpixel localization of the RFIs. The localization can be improved even more by extrapolating on a finer sampling grid, but this implies a computational cost that increases with the target sub pixel precision. Alternatively one could use finite rate of innovation models (originally introduced in [26] and further developed in [27], [28]) where the outliers are modeled as continuous Dirac deltas that can be localized with arbitrary precision, even when the acquired image has limited resolution. This, as well as the exploration of other optimization techniques in order to improve convergence speed (e.g. [14], [29]), will be addressed in the near future.



	$\ell_1(K)$	$\ell_2(K)$	$\ell_\infty(K)$	ℓ_1 flat zone (K)	ℓ_2 flat zone (K)
$T_{ZP} - T_{gt}$	4.706	9.671	110.410	1.707	2.371
$T_{L1A} - T_{gt}$	2.753	7.859	89.453	0.952	1.484

Fig. 9. Comparison of extrapolation results. In this test, we only analyse the results of the bandpass effect of the MIRAS instrument. No instrumental noise or outliers are added to the input ground truth image. In the first row, we show the result of computing $T_{L1B} = \mathbf{J}^+ \mathbf{G} T_{gt}$. The error reported on the corresponding table is only related to the bandpass matrix \mathbf{G} and zero padding solution with its corresponding matrix \mathbf{J}^+ . It is clear from this example that our method truly performs a spectral extrapolation, whose effects can be easily observed both on the image edges and on the reduction of Gibbs oscillations in the hole image. These two aspects can also be observed in the associated table, on the reported values of ℓ_1 and ℓ_2 error norms and on the more localized error on the flat zone, respectively.

APPENDIX PROXIMAL OPERATORS

For the sake of completeness, we include the derivation of each of the proximal operators involved in this work. We first recall the proximal operator definition of a functional E :

$$\text{prox}_{\gamma E}(x) = \arg \inf_y E(y) + \frac{1}{2\gamma} \|x - y\|^2.$$

Proposition 1 (Proximal operators of ℓ_0 and ℓ_1 norms). $\text{prox}_{\gamma \|\cdot\|_1}(o)[i] = \tau_\gamma(o[i])$, $\text{prox}_{\gamma \|\cdot\|_0}(o)[i] = s_\gamma(o[i])$, where

$$\tau_\gamma(t) = \begin{cases} \text{sign}(t)(|t| - \gamma) & \text{if } |t| \geq \gamma \\ 0 & \text{if } |t| < \gamma \end{cases}$$

$$s_\gamma(t) = \begin{cases} |t| & \text{if } |t| \geq \gamma \\ 0 & \text{if } |t| < \gamma \end{cases}$$

are the hard and soft thresholding operators, respectively.

A. Derivation of $\text{prox}_{\gamma \text{TV}_{\mathcal{H}}}$ proximal operator

We recall first the definition of the total variation semi-norm of an image $u : \Omega \subset \mathbb{R}^2 \rightarrow \mathbb{R}$:

$$\text{TV}(T) = \sup_{\psi \in C_c^1(\Omega, \mathbb{R}^2), \|\psi\|_{L^\infty(\Omega)} \leq 1} \left\{ \int_{\Omega} u(x) \text{div } \psi(x) dx \right\}.$$

With this notation, $\text{prox}_{\gamma \text{TV}}$ can be expressed:

$$\text{prox}_{\gamma \text{TV}}(u) = \arg \inf_y \text{TV}(y) + \frac{1}{2\gamma} \|u - y\|^2.$$

Chambolle [13] proves that the solution to this minimization problem can be expressed as the projection into a suitable convex set:

Proposition 2. The unique solution for $\text{prox}_{\gamma \text{TV}}(u)$ is given by $y = u - P_{\gamma G}(y)$, where $P_{\gamma G}(y)$ is the projection of y into the set γG , and $G = \{v \in X; \exists p \in X, |p| \leq 1 \text{ s.t. } v = \text{div } p\}$. The set $X = \mathbb{R}^{N^2}$ is the set of all discrete images of size N^2 .

The projection $P_{\gamma G}(y)$, defined by

$$\arg \min_{p \in X \times X, |p_{i,j}| \leq 1 \forall i,j=1,\dots,N} \{ |\gamma \text{div } p - u|_{X \times X}^2 \},$$

can be computed iteratively as

$$p_{i,j}^{n+1} = \frac{p_{i,j}^n + \tau(\nabla(\text{div } p^n - u/\gamma))_{i,j}}{1 + \tau|\nabla(\text{div } p^n - u/\gamma)|_{i,j}}.$$

Chambolle [13] has proven that, if $0 < \tau \leq \frac{1}{8}$ then $\gamma \text{div } p^n$ converges to $P_{\gamma G}(y)$ as $n \rightarrow \infty$. Finally,

$$\text{prox}_{\gamma \text{TV}}(u) = u - \lim_{n \rightarrow \infty} \gamma \text{div } p^n.$$

In our case, the spectral TV minimization consists in restricting the frequency domain to the cell \mathcal{H} , i.e.

$$\min_{u \in \text{BL}(\mathcal{H})} \text{TV}(u) + \text{DataFit}(u),$$

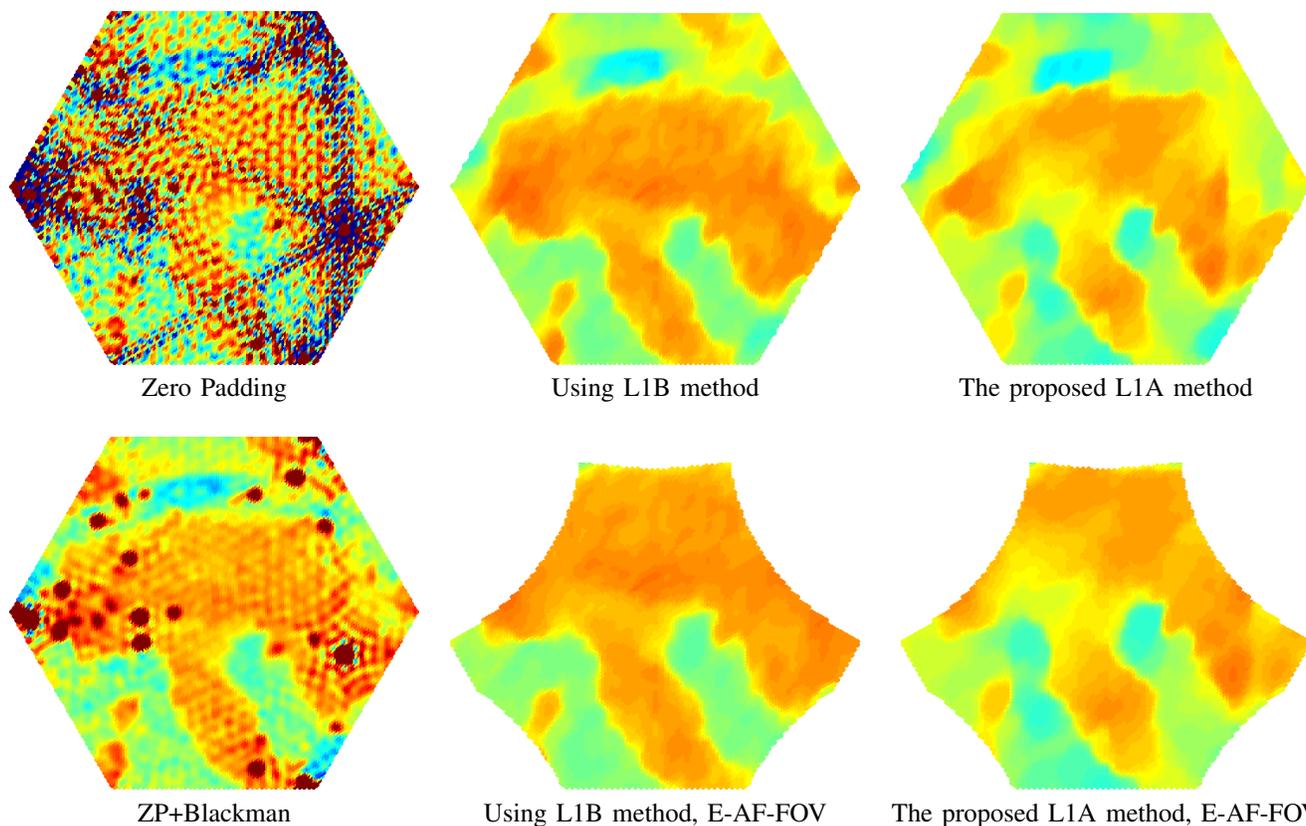


Fig. 11. Comparison between previous works and our method. This snapshot corresponds to Central Europe, with Italy clearly visible, and was acquired on march 2010. Color scale ranges from 0 to 300 Kelvin.

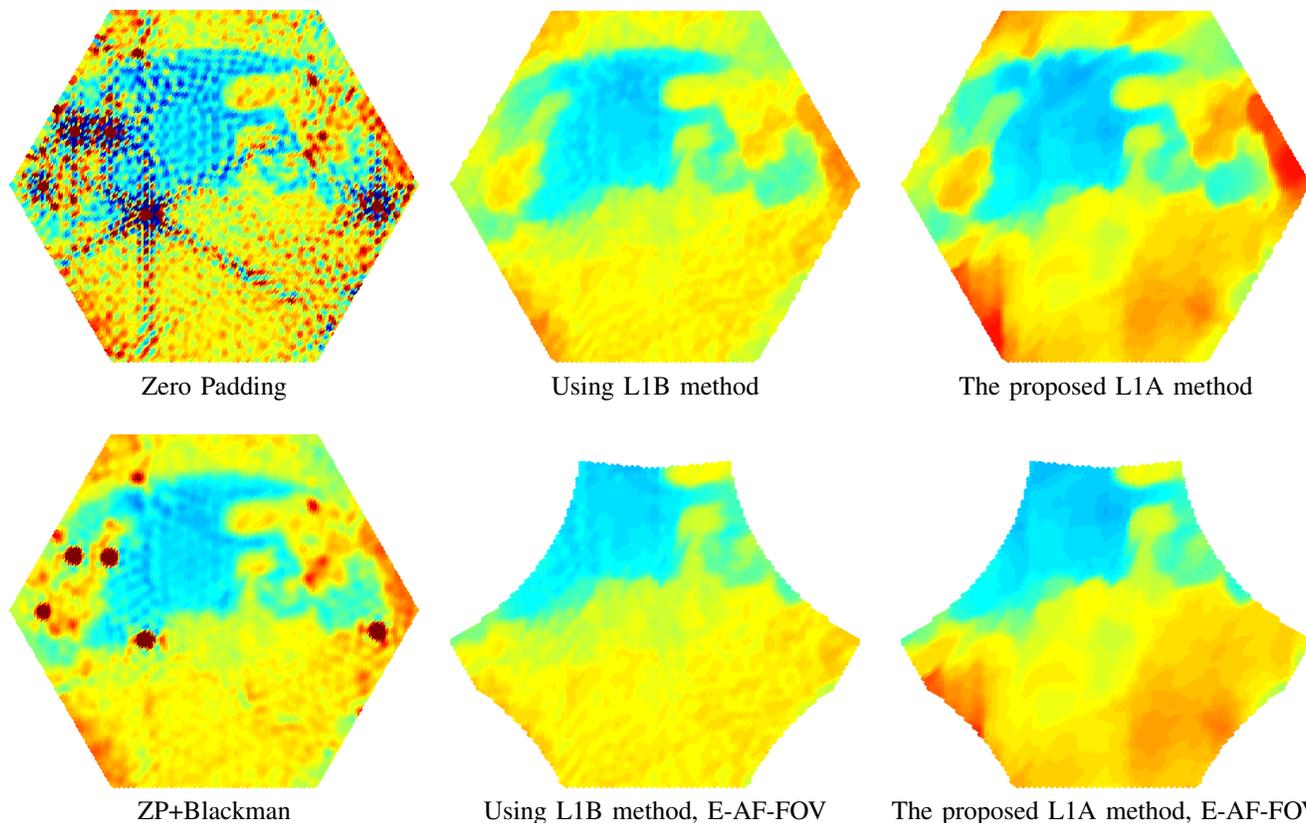


Fig. 12. Comparison between previous works and our method. This snapshot corresponds to Northern Europe and was acquired on march 2010. Color scale ranges from 0 to 300 Kelvin.

where $BL(\mathcal{H})$ denotes the space of band limited functions with spectral support within \mathcal{H} . This constraint can be integrated into the TV operator by means of the indicator function:

$$\iota_A(x) = \begin{cases} 0 & \text{if } x \in A \\ +\infty & \text{if } x \notin A \end{cases}$$

Then the proximal operator for the spectral TV becomes

$$\text{prox}_{\gamma \text{TV}_{\mathcal{H}}}(u) := \text{prox}_{\gamma \text{TV} + \iota_{BL(\mathcal{H})}}(u).$$

Since the sub differential of the indicator function ι_A is the projection operator P_A , we can easily show from the previous expression that a slight modification in Chambolle's algorithm yields the proximal operator for the spectral TV:

Proposition 3. *Let $u \in BL(\mathcal{H})$, and $0 < \tau \leq \frac{1}{8}$. Then the following algorithm with $u^n = -\gamma v^n$ converges to $\text{prox}_{\gamma \text{TV}_{\mathcal{H}}}(u)$*

- 1) $p^0 = 0, v^0 = -u/\gamma, n = 0$
- 2) $p^{n+1} = P_{BL(\mathcal{H})} \left(\frac{p^n + \tau \nabla v^n}{1 + \tau \|\nabla v^n\|} \right)$
- 3) $v^{n+1} = \text{div } p^n + v^0$
- 4) *If not converged go to step 2*

Note that the only modification w.r.t. Chambolle's algorithm is the spectral projection $P_{BL(\mathcal{H})}$ at each iteration p^n .

REFERENCES

- [1] Y. Kerr, P. Waldteufel, J.-P. Wigneron, S. Delwart, F. Cabot, J. Boutin, M.-J. Escorihuela, J. Font, N. Reul, C. Gruhier, S. Juglea, M. Drinkwater, A. Hahne, M. Martin-Neira, and S. Mecklenburg, "The SMOS mission: New tool for monitoring key elements of the global water cycle," *Proceedings of the IEEE*, vol. 98, no. 5, pp. 666–687, May 2010.
- [2] K. D. McMullan, M. Brown, M. Martin-Neira, W. Rits, S. Ekholm, J. Marti, and J. Lemanczyk, "SMOS: The payload," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 46, no. 3, pp. 594–605, March 2008.
- [3] M. Martin-Neira, Y. Menard, J. Goutoule, and U. Kraft, "MIRAS, a two-dimensional aperture synthesis radiometer," in *IEEE International Geoscience and Remote Sensing Symposium 1994. Surface and Atmospheric Remote Sensing: Technologies, Data Analysis and Interpretation*, vol. 3, Aug 1994, pp. 1323–1325.
- [4] Y. Kerr, J. Font, P. Waldteufel, A. Camps, J. Bara, I. Corbella, F. Torres, N. Duffo, M. Vallillossera, and G. Caudal, "New radiometers: SMOS - a dual pol L-band 2D aperture synthesis radiometer," in *Aerospace Conference Proceedings, 2000 IEEE*, vol. 5, 2000, pp. 119–128.
- [5] I. Corbella, N. Duffo, M. Vallillossera, A. Camps, and F. Torres, "The visibility function in interferometric aperture synthesis radiometry," *IEEE Trans. Geosci. Remote Sensing*, vol. 42, no. 8, pp. 1677–1682, 2004.
- [6] D. Petersen and D. Middleton, "Sampling and reconstruction of wave-number-limited functions in N-dimensional euclidean space," *Information and Control*, vol. 5, pp. 279–323, 1962.
- [7] A. Camps, J. Bara, I. Sanahuja, and F. Torres, "The processing of hexagonally sampled signals with standard rectangular techniques: application to 2-D large aperture synthesis interferometric radiometers," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 35, no. 1, pp. 183–190, Jan 1997.
- [8] A. R. Thompson, J. M. Moran, and G. W. Swenson, *Interferometry and Synthesis in Radio Astronomy; 2nd ed.* Weinheim: Wiley-VCH, 2001.
- [9] A. Camps, M. Vallillossera, I. Corbella, N. Duffo, and F. Torres, "Improved image reconstruction algorithms for aperture synthesis radiometers," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 46, no. 1, pp. 146–158, Jan 2008.
- [10] E. Anterrieu, "A resolving matrix approach for synthetic aperture imaging radiometers," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 8, pp. 1649–1656, Aug 2004.
- [11] J. Preciozzi, P. Musé, A. Almansa, S. Durand, A. Khazaal, and B. Rougé, "SMOS images restoration: A sparsity-based variational approach," in *IEEE International Geoscience and Remote Sensing Symposium*, 2014.
- [12] J. Preciozzi, P. Musé, A. Almansa, S. Durand, F. Cabot, Y. Kerr, A. Khazaal, and B. Rougé, "Sparsity-based restoration of SMOS images in the presence of outliers," in *IEEE International Geoscience and Remote Sensing Symposium*, 2012, pp. 3501–3504.
- [13] A. Chambolle, "An algorithm for total variation minimization and application," *Journal of Mathematical Imaging and Vision*, vol. 20, no. 1, 2004.
- [14] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *Journal of Mathematical Imaging and Vision*, vol. 40, no. 1, pp. 120–145, 2011.
- [15] P. L. Combettes and V. R. Wajs, "Signal recovery by proximal forward-backward splitting," *SIAM Journal on Multiscale Modeling and Simulation*, vol. 4, no. 4, pp. 1168–1200, 2005.
- [16] T. Blumensath and M. E. Davies, "Iterative thresholding for sparse approximations," *The Journal of Fourier Analysis and Applications*, vol. 62, pp. 291–294, January 2008.
- [17] E. Anterrieu, P. Waldteufel, and A. Lannes, "Apodization functions for 2-D hexagonally sampled synthetic aperture imaging radiometers," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 12, pp. 2531–2542, Dec 2002.
- [18] M. Nikolova, "Local strong homogeneity of a regularized estimator," *SIAM Journal on Applied Mathematics*, vol. 61, no. 2, pp. 633–658, 2000.
- [19] K. Jalalzai, "Some remarks on the staircasing phenomenon in total-variation based image denoising," *ArXiv e-prints*, Feb. 2014.
- [20] L. Moisan, "How to discretize the total variation of an image?" *ICIAM07 Minisymposia*, 2007.
- [21] H. Uzawa, "Iterative methods for concave programming," *Studies in linear and nonlinear programming*, vol. 6, 1958.
- [22] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.
- [23] —, "Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems," *IEEE Transactions on Image Processing*, vol. 18, no. 11, pp. 2419–2434, 2009.
- [24] A. Camps, I. Corbella, M. Vallillossera, N. Duffo, F. Marcos, F. Martínez-Fadrique, and M. Greiner, "The SMOS end-to-end performance simulator: Description and scientific applications," in *IEEE International Geoscience and Remote Sensing Symposium*, vol. 1, 2003, pp. 13–15.
- [25] F. Cabot, Y. Kerr, P. Waldteufel, and H. Barré, "SMOS overall system performance synthesis report," Centre d'Etudes Spatiales de la Biosphère (CESBIO), Toulouse, France, Tech. Rep. SO-TN-CESBIO-SYS-1184-3.0, January 2007. [Online]. Available: http://www.cesbio.ups-tlse.fr/data_all/SMOS-doc/SMOS-perfos-systeme-V3.pdf
- [26] M. Vetterli, P. Marziliano, and T. Blu, "Sampling signals with finite rate of innovation," *IEEE Trans. Signal Process.*, vol. 50, no. 6, pp. 1417–1428, Jun. 2002.
- [27] E. J. Candès and C. Fernandez-Granda, "Towards a Mathematical Theory of Super-resolution," *Commun. Pure Appl. Math.*, vol. 67, no. 6, pp. 906–956, Jun. 2014.
- [28] V. Duval and G. Peyré, "Exact Support Recovery for Sparse Spikes Deconvolution," CEREMADE - Univ. de Paris Dauphine, Tech. Rep., Jun. 2013. [Online]. Available: <http://arxiv.org/abs/1306.6909>
- [29] I. Daubechies, R. DeVore, M. Fornasier, and C. S. Güntürk, "Iteratively reweighted least squares minimization for sparse recovery," *Comm. Pure Appl. Math.*, vol. 63, no. 1, pp. 1–38, Jan. 2010.