Multi-Date Divergence Matrices for the Analysis of SAR Image Time Series

Abdourrahmane Atto, Emmanuel Trouvé, Yannick Berthoumieu, Grégoire
Mercier

Multi-Date Divergence Matrices for the Analysis of SAR Image Time Series

Abdourrahmane M. Atto[1], Emmanuel Trouvé[2], Yannick Berthoumieu[3], Grégoire Mercier[4]

Abstract—The paper provides a spatio-temporal change detection framework for the analysis of image time series. In this framework, the detection of *changes in time* is addressed at the image level by using a matrix of cross-dissimilarities computed upon wavelet and curvelet image features. This makes possible identifying the acquisitions-of-interest: the acquisitions that exhibit singular behavior with respect to their neighborhood in the time series and those that are representatives of some stationary behavior. These acquisitions-of-interest are compared at the pixel level in order to detect *spatial changes* characterizing the evolution of the time series. Experiments carried out over ERS and TerraSAR-X time series highlight the relevancy of the approach for analyzing SAR image time series.

Index Terms—Image Time Series ; Wavelets ; Curvelets ; Parametric modeling ; Kullback-Leibler Divergence ; Change Detection.

I. INTRODUCTION

IMAGE analysis from remote sensing time series plays an essential role in many applications involving monitoring of regions that are difficult to access, all over the world.

The analysis of remote sensing time series has been addressed in the literature by exploiting some specific (class dependent) image features and the signal coherency in time. For instance, the *Synthetic Aperture Radar* (SAR) image features proposed in [1] consist of long-term coherence and backscattering temporal variability learned from different land-cover classes, for a classification purpose. For ground deformation monitoring, [2] has proposed estimating some linear and nonlinear deformation features from neighborhood networking of temporally coherent radar targets (permanent-scatterer) in interferometric SAR time series. For the detection of buildings destroyed in an earthquake, [3] has constructed some building features consisting in non-overlapping rectangular footprints

and has proposed a post-event building feature predictor as a benchmark relating building states (destroyed or non-destroyed) over time.

The analysis of remote sensing time series has also been addressed by transforming the observed data in some appropriate representation spaces. As a matter of example, the detection of spatial similarities has been addressed in the spectral domain in [4] and [5]. This detection has been performed upon a compressed representation of the change information in [6]. This detection has also been addressed by focusing on a low frequency representation of the SAR signal from some iterative filtering of the observed data to remove speckle in [7].

In addition, the selection of suitable similarity measures with respect to the remote sensing data distribution has deserved much interest in the literature on multi-temporal change detection. In particular, we have:

- the time warping distance associated with a disk-based suffix tree indexing is a accurate in comparing sequences of different lengths and/or different sampling rates, see [8],
- the Bhattacharyya distance between the probability density functions is efficient for measuring contrast similarity, see [9],
- the mutual information between Wishart processes over time is suitable for characterizing temporal polarimetric and interferometric informations in SAR data, see [10].

Exploiting new generation remote sensing images is actually facing three major challenges: high resolution, large database (long acquisition sequence) and speckle effect. The conjunction of these facts makes the pursuit of a temporal event intricate on a long time series, when no *a priori* is available concerning i) the spatial location of changing areas and ii) the change information. Indeed, local analysis from overlapped sliding windows all over the scene and all along the whole time series is not thinkable due to computational complexity. This analysis is however necessary for the concerns of early detection of some physical phenomenon such as ground displacement or melting cycle of glaciers that can be evaluated through long image time series.

The approach proposed in this work for simplifying the analysis of image time series relates to the description

[1] Université de Savoie - Polytech Annecy-chambéry, LISTIC, *Abdourrahmane.Atto@univ-savoie.fr*
[2] Université de Savoie - Polytech Annecy-Chambéry, LISTIC, *Emmanuel.Trouve@univ-savoie.fr*
[3] Université de Bordeaux, IPB, IMS, CNRS UMR 5218, *Yannick.Berthoumieu@ims-bordeaux.fr*
[4] Institut TELECOM, TELECOM Bretagne, Lab-STICC, CNRS UMR 3192, *Gregoire.Mercier@telecom-bretagne.fr*

Subsequence j Subsequence k Subsequence ℓ

t

$\mathcal{I}_{j-1}$ $\mathcal{I}_j$ $\mathcal{I}_{j+1}$ $\mathcal{I}_{k-1}$ $\mathcal{I}_k$ $\mathcal{I}_{k+1}$ $\mathcal{I}_{\ell-1}$ $\mathcal{I}_\ell$ $\mathcal{I}_{\ell+1}$

Stage 1 / Temporal analysis [stationary subsequence identification] KL$(\mathcal{TI}_k/\mathcal{TI}_{k+m})$

j

k

ℓ

MDDM

j k ℓ

Stage 2 /

Spatial analysis Spatial analysis Spatial analysis

$\mathcal{I}_j$ $\mathcal{I}_k$ $\mathcal{I}_k$ $\mathcal{I}_\ell$ $\mathcal{I}_j$ $\mathcal{I}_\ell$

$\mathcal{I}_j^i$ $\mathcal{I}_k^i$ $\mathcal{I}_k^i$ $\mathcal{I}_\ell^i$ $\mathcal{I}_j^i$ $\mathcal{I}_\ell^i$

KL$(\mathcal{TI}_j^i/\mathcal{TI}_k^i)$ KL$(\mathcal{TI}_k^i/\mathcal{TI}_\ell^i)$ KL$(\mathcal{TI}_j^i/\mathcal{TI}_\ell^i)$
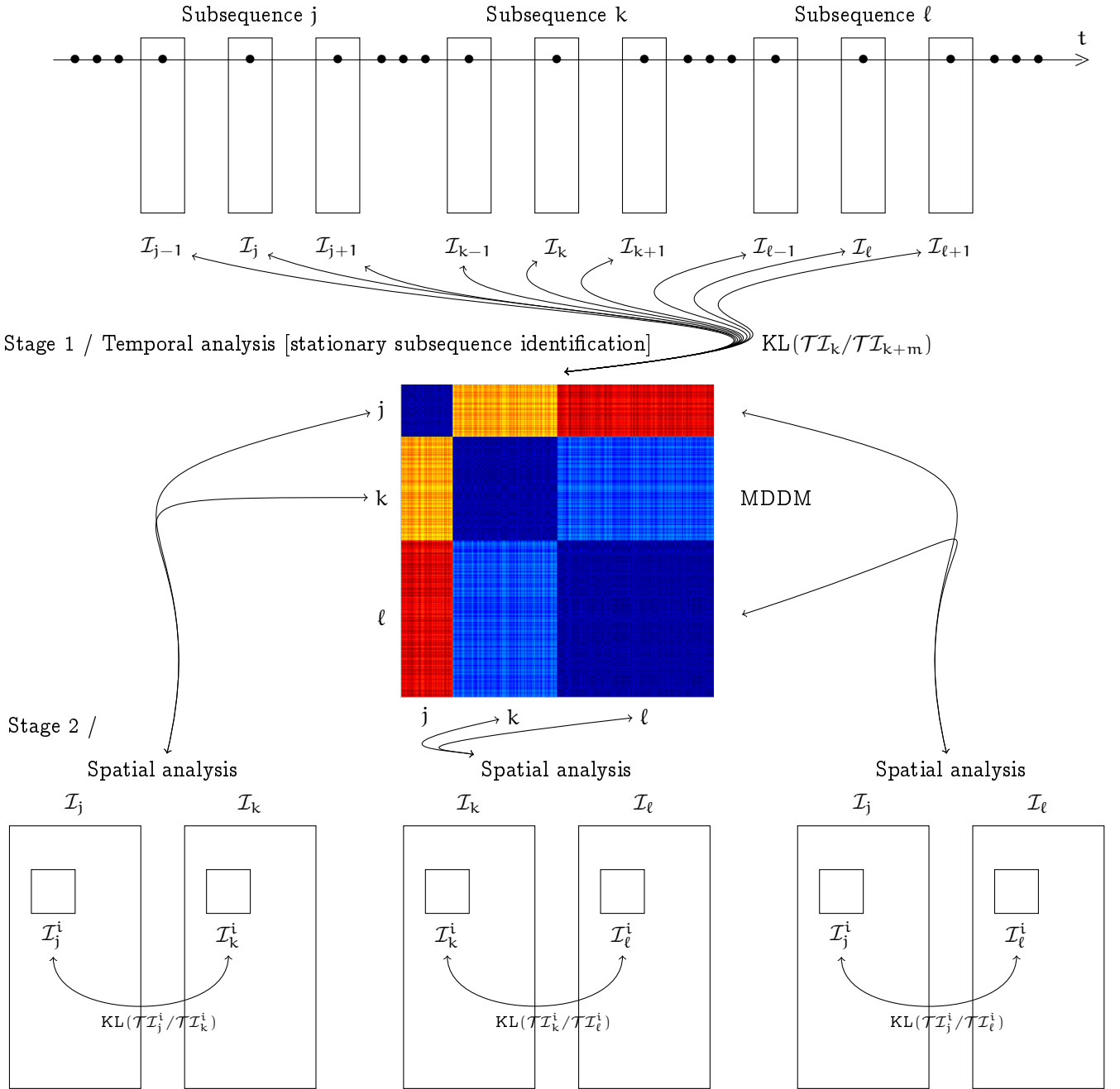
Fig. 1. Block diagram representing temporal and spatial image similarity measurements. In this diagram, transform $\mathcal{T}$ designates either the SWT or the FDCT. Specifically, the term KL$(\mathcal{T}X/\mathcal{T}Y)$ designates KL based similarity measurements from the SWT or the FDCT decompositions of X and Y, see for instance (3) and (4). Temporal analysis involves large size image (or subimage) comparison whereas spatial analysis applies only for specific images at the pixel level. The MDDM shown in the center of the figure has been computed upon a multi-date Gaussian random field composed with 3 stationary subsequences.

of any image/subimage by a set of essential description parameters obtained from representations having suitable statistical and geometrical properties. These representations are the discrete *Stationary Wavelet Transform*, SWT, [11] and the *Fast Discrete Curvelet Transform*, FDCT, [12] [13], [14].

The SWT and FDCT coefficients of the SAR images exhibit stochastic nature because of the presence of speckle in radar images. In this respect, we consider parameter extraction from stochastic modeling of the coefficients of these representations. The modeling is achieved by using 1) Edgeworth expansions for the SWT and FDCT approximations and 2) Generalized Gaussian, Log-Normal and Weibull distributions for the details and orientations of these transforms. These distributions are used in a joint framework requiring similarity evaluation from distribution models pertained to different parametric families: the closed-form expressions for inter-family parameter

Input image $\mathcal{I}_{\mathrm{m}}$

SWT representation $\mathcal{I}_{\mathcal{W}}$ of $\mathcal{I}_{\mathrm{m}}$

FDCT representation $\mathcal{I}_{\mathcal{C}}$ of $\mathcal{I}_{\mathrm{m}}$

Approximation $c_{J,0}$

Details $\left\{c_{j,n}\right\}_{\substack{1\leqslant j\leqslant J \\ n\in\{1,2,3\}}}$

Approximation $\mathcal{I}_{\mathcal{C}}^{\mathrm{app}}$

Orientations $\left\{\mathcal{I}_{\mathcal{C}}^{\mathrm{ang}}[q]\right\}_{q\in\{0,1,\ldots,Q-1\}}$

$\kappa_1[c_{J,0}]$ $\kappa_2[c_{J,0}]$ $\kappa_3[c_{J,0}]$ $\kappa_4[c_{J,0}]$

Model selection for $c_{j,n}$

$\kappa_1[\mathcal{I}_{\mathcal{C}}^{\mathrm{app}}]$ $\kappa_2[\mathcal{I}_{\mathcal{C}}^{\mathrm{app}}]$ $\kappa_3[\mathcal{I}_{\mathcal{C}}^{\mathrm{app}}]$ $\kappa_4[\mathcal{I}_{\mathcal{C}}^{\mathrm{app}}]$

Model selection for $\mathcal{I}_{\mathcal{C}}^{\mathrm{ang}}[q]$

$\rho_1\left[c_{j,n}\right]$ $\rho_2\left[c_{j,n}\right]$

$\rho_1\left[\mathcal{I}_{\mathcal{C}}^{\mathrm{ang}}[q]\right]$ $\rho_2\left[\mathcal{I}_{\mathcal{C}}^{\mathrm{ang}}[q]\right]$
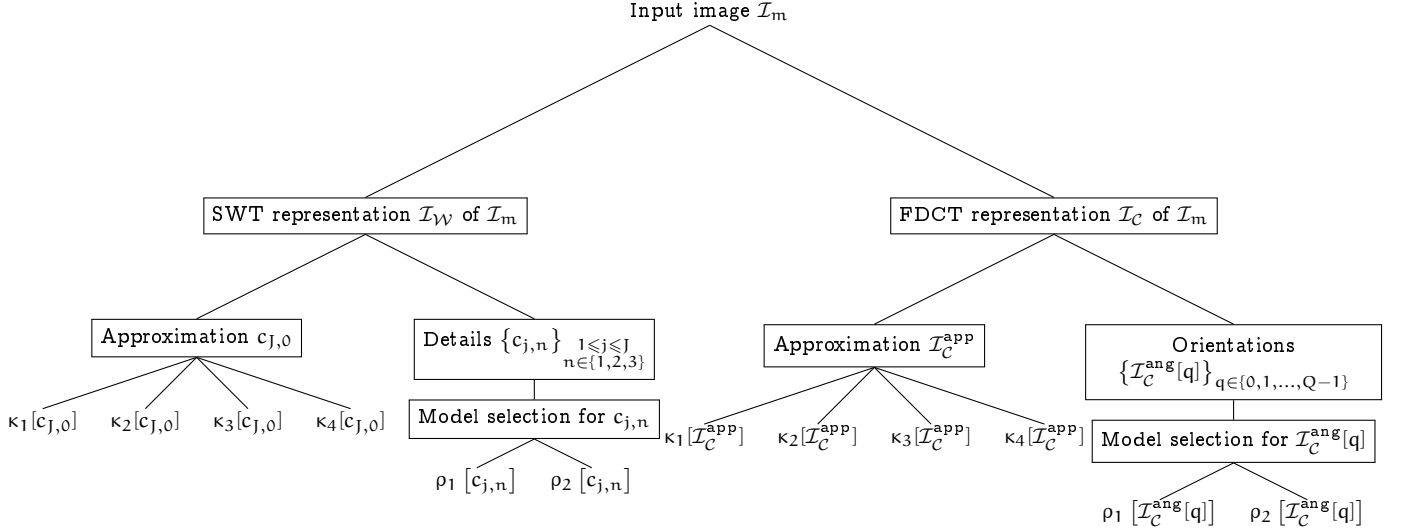
Fig. 2. Block diagram representing image feature extraction. Parameters $(\kappa_\ell[X]), \ell = 1, 2, 3, 4$ designate the first four cumulants of X: κ_1 and κ_2 are the mean and variance of X, κ_3 and κ_4 relate to the skewness and kurtuosis of X, respectively. These cumulants are associated with an Edgeworth expansion of X. Notation $\rho_1[X]$ and $\rho_2[X]$ designate the scale and shape parameters of the best model among the GG, LOGN and WBL distributions associated with |X| and issued from a model selection step.

comparison are then derived in the paper.

The proposed similarity measurements are based on the Kullback-Leibler (KL) divergences and applied in a two stage approach illustrated in Fig. 1. In the first stage, a global temporal analysis is performed at the image/subimage level, *i.e* between any pair of image description parameters. These KL based similarity measurements are gathered to form a Multi-Date Divergence Matrix (MDDM). The analysis of the MDDM eases the selection of the acquisitions-of-interest: these acquisitions are either *change-images*, *i.e* images relating the transitions between the stationary subsequences composing the image time series or *image-outliers*, *i.e* acquisitions that exhibit a suspicious behavior with respect to their neighborhood in time.

In the second stage, a local spatial analysis is performed at the pixel level between acquisitions-of-interest derived from the MDDM analysis. Pixelwise similarity measurements are computed within a sliding window in order to build change maps and detect the changing areas.

The presentation of the paper is as follows. Section II addresses the extraction of image description parameters and the similarity measurements in the wavelet and curvelet domains. Section III presents the MDDM and provides an interpretation of this matrix in term of abrupt and progressive changes, as well as multiple change-images in time. Section IV is dedicated to experimental results on multi-temporal change detection in SAR time series from TerraSAR-X (TS-X) images. Section V addresses spatial change detection from MDDM critical acquisitions. Section VI concludes the work.

II. FEATURE EXTRACTION FROM SWT AND FDCT PARAMETRIC MODELING - SIMILARITY MEASUREMENTS

A. Motivation

Let $\mathcal{I} = \{\mathcal{I}_{\mathrm{m}}, \mathrm{m} = 1, 2, \ldots, \mathrm{M}\}$ be a sequence of M images of a given scene acquired at different dates: $\mathcal{I}_{\mathrm{m}} = \mathcal{I}[\mathrm{t_m}]$ where $\mathrm{t_m}$ is the acquisition time.

The method proposed below for analyzing $\mathcal{I}$ relates to a *parsimonious description* of every $\mathcal{I}_{\mathrm{m}}$ by a set of P *essential description parameters* extracted from the image features. This set is required to be parsimonious in the sense that it consists of an accurate modeling with $P \ll N$ elements, where N is the number of pixels of $\mathcal{I}_{\mathrm{m}}$ (images are assumed to have the same sample size). In this respect, the set of parameters used hereafter derives from *parametric modeling* of the statistical distributions of wavelet and curvelet contributions. Indeed, in presence of non-stationarities (the modeling concerns an image/subimage and not a stationary region from this image), wavelet frames are relevant for building parsimonious models by reducing several types of lower and higher order dependencies (some statistical properties of wavelet transforms for analyzing non-stationary random processes are given in [15]).

The main steps of this modeling are described in the block diagram given by Fig. 2 and are developed below.

B. SWT coefficients - FDCT subimages

Among the variants of wavelet frames, we consider the SWT because of its suitable statistical properties [11], [16], [17] and the FDCT for its sensitivity to image geometry: the transform is a quasi-optimal representation for 2 di-

TABLE I
Average retrieval rates of some distribution families over the SWT and FDCT contributions for two TS-X time series over the Argentière glacier. A retrieval rate per family indicates the percentage of occurrence of the distributions pertaining to this family as those with the smallest Kolmogorov stochasticity parameter, among the distribution families considered. The sequences of TS-X images considered are 13 images in ascending orbit and 10 images in descending orbit.

Transform	Acquisition mode	GG	LOGN	WBL	GAM	GEV	EXP
SWT	Ascending	60.90	37.82	1.28	0	0	0
	Descending	62.50	21.17	14.33	0	0	0
FDCT	Ascending	63.25	35.04	1.71	0	0	0
	Descending	63.33	35.5556	0	0	1.11	0

mensional regular signals admitting some singularities like edges and contours for natural images [12], [13], [14].

The J-level SWT and FDCT representations of an image $\mathcal{I}$ have the following form:

$$\mathcal{I} = \sum_{k_1,k_2} \langle \mathcal{I}, \Phi_{J,[k_1,k_2]} \rangle \Phi_{J,[k_1,k_2]}$$
$$+ \sum_{j=1}^{J} \sum_{\substack{n_1,n_2 \\ k_1,k_2}} \langle \mathcal{I}, \Psi_{j,[n_1,n_2],[k_1,k_2]} \rangle \Psi_{j,[n_1,n_2],[k_1,k_2]}. \quad (1)$$

In this expansion, $\Phi_{J,[k_1,k_2]}$ denotes translated versions of a scaling function Φ_J and the wavelet/curvelet waveforms $\Psi_{j,[n_1,n_2],[k_1,k_2]}$ are generated at decomposition level j from the scaled, translated/oriented versions of a waveform Ψ.

When the SWT is concerned, $\mathcal{I}$ will be denoted by $\mathcal{I}_{\mathcal{W}}$ and the SWT approximation coefficients are denoted by

$$c_{J,0}[k_1, k_2] = \langle \mathcal{I}, \Phi_{J,[k_1,k_2]} \rangle$$

whereas the SWT detail coefficients are obtained from

$$c_{j,n}[k_1, k_2] = \langle \mathcal{I}, \Psi_{j,n,[k_1,k_2]} \rangle.$$

In the above notations, $n \in \{0, 1, 2, 3\}$ is a capital parameter associated with indices $n_1, n_2 \in \{0, 1\}$. We will use the standard terminology *subband wavelet coefficients* to designate the subimage of coefficients $c_{j,n} = (c_{j,n}[k_1, k_2])_{(k_1,k_2) \in \mathbb{Z} \times \mathbb{Z}}$.

Parametric modeling of this SWT representation concerns subband coefficients $c_{j,n}$. In the experimental results, we consider an SWT with decomposition level is $J = 4$ and the "Symlet" wavelet of order 8 (a wavelet with high order have more stringent statistical properties with respect to the parametric modeling concerned, see [15], [16]).

When the FDCT is concerned, $\mathcal{I}$ will be denoted by $\mathcal{I}_{\mathcal{C}}$. In FDCT implementation [14], detail subbands located at a given decomposition level j are generated from scaled and oriented versions of the curvelet waveform Ψ, where orientations are associated with a rotation by an angle $\theta_n = 2\pi 2^{-\lfloor j/2 \rfloor} n$, with $n \in \{0, 1, \ldots, 2^{\lfloor j/2 \rfloor} - 1\}$ (with the correspondence $n_1, n_2 \to n$, as above).

FDCT coefficients form a huge number of subbands, many of these subbands having a small number of coefficients when the decomposition level and the number of

orientations are large. Thus, we need to pre-process FDCT in order to make suitable, the curvelet based parametric modeling. In this respect, we consider synthesizing from the curvelet coefficients, an approximation and some orientation *subimages* with the same sizes as the input image.

The curvelet *approximation subimage* is computed by setting to zero all the curvelet coefficients except the approximation ones and by applying the "inverse" FDCT to reconstruct from these coefficients:

$$\mathcal{I}_{\mathcal{C}}^{\mathrm{app}} = \sum_{k_1,k_2} \langle \mathcal{I}, \Phi_{J,[k_1,k_2]} \rangle \Phi_{J,[k_1,k_2]}.$$

A curvelet *orientation subimage* is obtained by aggregating the different curvelet subbands lying in a given angular wedge and reconstructing the corresponding *subimage* by using the inverse FDCT. Assume that the total number of orientations at coarse scale is Q. Then the following steps are used for selecting the FDCT coefficients involved in a subimage synthesis.

Computation of curvelet orientation subimages: For every $q \in \{0, 1, \ldots, Q - 1\}$:

- Identify for every $j = 1, 2, \ldots, J$, the set of polar *wedges*:

$$\mathcal{A}[q] = \{(j, n) : [\theta_n(j), \theta_{n+1}(j)[\subset [\theta_q, \theta_{q+1}[\}.$$

- Then, set to zero all the curvelet coefficients, except those pertaining to the above wedges, and reconstruct the subimage $\mathcal{I}_{\mathcal{C}}^{\mathrm{ang}}[q]$ by using the "inverse" FDCT:

$$\mathcal{I}_{\mathcal{C}}^{\mathrm{ang}}[q] = \sum_{(j,n) \in \mathcal{A}[q]} \sum_{k_1,k_2} \langle \mathcal{I}, \Psi_{j,n,[k_1,k_2]} \rangle \Psi_{j,n,[k_1,k_2]}$$

It follows from the above decomposition and by taking into account (1) that

$$\mathcal{I}_{\mathcal{C}} = \mathcal{I}_{\mathcal{C}}^{\mathrm{app}} + \sum_{q \in \{0, 1, \ldots, Q-1\}} \mathcal{I}_{\mathcal{C}}^{\mathrm{ang}}[q] = \mathcal{I}.$$

In the following experimental results, we consider a number $Q = 8$ orientations.

C. Parametric modeling of the SWT and FDCT contributions

The parametric modeling is addressed by using 1) Edgeworth expansions for describing SWT and FDCT

TABLE II
OVERVIEW OF THE PARAMETRIC MODELING AND THE ASSOCIATED NUMBER OF PARAMETERS.

Parametric modeling	FDCT	SWT
Edgeworth	$\{\kappa_p[\mathcal{I}_C^{\mathrm{app}}]\}_{1\leqslant p\leqslant 4}$	$\{\kappa_p[c_{J,0}]\}_{1\leqslant p\leqslant 4}$
GG / LOGN / WBL	$\{\rho_1\left[\mathcal{I}_C^{\mathrm{ang}}[q]\right],\rho_2\left[\mathcal{I}_C^{\mathrm{ang}}[q]\right]\}_{q=0,1,\ldots,Q-1}$	$\{\rho_1\left[c_{j,n}\right],\rho_2\left[c_{j,n}\right]\}_{\substack{1\leqslant j\leqslant J \\ n\in\{1,2,3\}}}$
Number of parameters	$2\times Q+4$	$6\times J+4$

approximations and 2) a dictionary $\mathcal{D}$ composed with the Generalized Gaussian (GG), Log-Normal (LOGN) and Weibull (WBL) distributions for describing the amplitudes of SWT and FDCT details and orientations. The analytic forms of these distribution functions are given in Appendix A.

The SWT approximation subband $c_{J,0}$ and FDCT approximation subimage $\mathcal{I}_C^{\mathrm{app}}$ have specific behavior [15] because these coefficients result from inner products with scaling functions (a scaling function has no vanishing moments). Finding a class of parametric models for the approximation contributions is very intricate because of this specificity. In this respect, we consider the Edgeworth expansion of order $p = 4$ in order to capture the information conveyed by the first 4 cumulants of SWT $c_{J,0}$ and FDCT $\mathcal{I}_C^{\mathrm{app}}$ approximations.

For high frequency SWT and FDCT contributions (details and orientations), we focus on GG, LOGN and WBL distributions from *model validation* results that are given in Table[1] I. This model validation has been performed upon distribution families with exponential decays so as to comply with the statistical properties of wavelet-based transforms. From this model validation step, it follows that: when compared with Gamma (GAM), Generalized Extreme Values (GEV) and Exponential (EXP) distributions, the GG, LOGN and WBL distributions yield more than 99% of retrieval results when the query concerns the "distribution with the smallest Kolmogorov stochasticity parameter" (see average retrieval results given in Table I). These 3 distribution families are thus aggregated to form the dictionary $\mathcal{D}$ of parametric models used in the sequel.

We then use the following model selection procedure to retrieve the parameters of the best distribution from this dictionary.

D. Model selection

Consider an SWT detail subband $c_{j,n}$ with some $n \neq 0$ or an FDCT orientation subimage $\mathcal{I}_C^{\mathrm{ang}}[q]$ for some $q \in \{0, 1, \ldots, Q-1\}$.

The *model selection* step involved in the diagram of Fig. 2 consists in deriving from dictionary $\mathcal{D}$, the best

---

[1]One can note the results given in Table I are specific to TS-X images considered in the framework of the SWT and the FDCT. Thus, statistics given in this table may differ, depending on the data type and the transform used.

---

distribution family and the parameters of this distribution for modeling the subband under consideration: this distribution follows from 1) a first stage that aims at computing 3 *maximum-likelihood estimates* associated with GG, LOGN and WBL distributions respectively (see Table I for the motivation on selecting these families) and 2) a second stage for selecting the best distribution as the one admitting the smallest Kolmogorov stochasticity parameter among these maximum-likelihood estimates.

The following items pinpoint the steps involved in this model selection:

- Generate the set of parameters

$$\mathfrak{S}_{\mathrm{ML}} = \{(\alpha_{\mathrm{ML}}, \beta_{\mathrm{ML}}), (\mu_{\mathrm{ML}}, \sigma_{\mathrm{ML}}), (a_{\mathrm{ML}}, b_{\mathrm{ML}})\}$$

of the "bests" GG, LOGN and WBL distributions from Maximum Likelihood (ML) estimation with respect to the set y of coefficients analyzed,

- the Kolmogorov stochasticity parameter [18] with respect to the above ML based distributions is derived from:

$$\lambda_N(y, F_{\rho_1,\rho_2}) = \sup |F_{y,N}(t) - F_{\rho_1,\rho_2}(t)|, \quad (2)$$

where $F_{y,N}$ is the empirical *cumulative distribution function (cdf)* of the N-sample dataset y and F_{ρ_1,ρ_2} is one among the ML based *cdfs* computed above.

- The relevant distribution is then the distribution $F_{\rho_1^*,\rho_2^*}$ with parameters derived from

$$(\rho_1^*, \rho_2^*) = \arg\min_{(\rho_1,\rho_2)\in\mathfrak{S}_{\mathrm{ML}}} \lambda_N(y, F_{\rho_1,\rho_2}).$$

The distribution type (GG, LOGN or WBL) is the index of the pair (ρ_1^*, ρ_2^*) in $\mathfrak{S}_{\mathrm{ML}}$.

The parsimonious description represented by the block diagram of Fig. 2 thus consists in a concise wavelet-curvelet modeling of any $\mathcal{I}_m$ by using $P = 2Q + 6J + 8$ parameters. Table II summarizes this parametric modeling. By taking into account that $Q = 8$ and $J = 4$ in the following, we have $P = 48$ essential parameters per image description. These parameters encompass multiscale and orientation based image features.

It is worth emphasizing that the P parameters associated with the description of $\mathcal{I}_m$ from the multi-temporal sequence $\mathcal{I}$ can be computed and stored as long as the image $\mathcal{I}_m$ is available, during the acquisition process. Furthermore, as $\mathcal{I}_m$ is available, the similarity measurements between $\mathcal{I}_m$ and the sequence $\{\mathcal{I}_\ell, \ell \leqslant m-1\}$ can

be computed as well, for an operational purpose. These similarity measurements proposed are developped in the following section.

E. Similarity measurements

We use KL divergence as *similarity measure* between the different statistical distributions issued from dictionary $\mathcal{D}$. The similarity measurements between image features (distribution parameters) are performed by using the approach proposed in [16]. This approach is extended so as to make possible comparing features issued from two different parametric families. Furthermore, the approach proposed below also takes into account the modeling of SWT and FDCT approximations by using Edgeworth expansions.

We consider as similarity measure between two random variables X_1 and X_2 distributions f_{X_1} and f_{X_2} (being either Edgeworth expansions or parametric models from $\mathcal{D}$), the symmetric KL similarity measure defined as

$$\mathcal{K}(X_1, X_2) = \mathcal{K}(X_1\|X_2) + \mathcal{K}(X_2\|X_1),$$

where $\mathcal{K}(X_1\|X_2)$ is the divergence

$$\mathcal{K}(X_i\|X_j) = \int_{\mathbb{R}} f_{X_i}(x) \log \frac{f_{X_i}(x)}{f_{X_j}(x)} dx, \quad i, j = 1, 2.$$

The similarity measures are applied to compare adjacent SWT subbands or FDCT subimages, the global similarity measure between $\mathcal{I}_m$ and $\mathcal{I}_\ell$ being:

$$
\mathcal{K}_{\mathcal{W}}(\mathcal{I}_m, \mathcal{I}_\ell) = \mathcal{K}(c_{J,0}[\mathcal{I}_m], c_{J,0}[\mathcal{I}_\ell])
+ \sum_{\substack{j \in \{1,2\ldots,J\} \\ n \in \{1,2,3\}}} \mathcal{K}(c_{j,n}[\mathcal{I}_m], c_{j,n}[\mathcal{I}_\ell]) \quad (3)
$$

for the SWT, where $c_{j,n}[\mathcal{I}_m]$, $c_{j,n}[\mathcal{I}_\ell]$ are the SWT coefficients of $\mathcal{I}_m$ and $\mathcal{I}_\ell$ respectively and

$$
\mathcal{K}_{\mathcal{C}}(\mathcal{I}_m, \mathcal{I}_\ell) = \mathcal{K}(\mathcal{I}_{\mathcal{C}}^{app}[m], \mathcal{I}_{\mathcal{C}}^{app}[\ell])
+ \sum_{q=0}^{Q-1} \mathcal{K}(\mathcal{I}_{\mathcal{C}}^{ang}[m][q], \mathcal{I}_{\mathcal{C}}^{ang}[\ell][q]), \quad (4)
$$

when the FDCT is considered ($\mathcal{I}_{\mathcal{C}}^{ang}[m][q]$ denotes the q-th oriented curvelet subimage of the m-th image of the sequence $\mathcal{I}$).

In the following, $\mathcal{K}(X_1\|X_2)$ is computed from parametric models associated with the distributions of X_1 and X_2. When the detail and orientation coefficients are under consideration, these distributions follow from a dictionary $\mathcal{D}$ composed with the 3 best distribution families given in Table I. Appendices B and C therefore provide the parametric forms of the KL divergences involved in comparing two arbitrary elements of dictionary $\mathcal{D}$. We recall that approximation coefficients are processed separately in the sense that they are modeled by using Edgeworth expansions: the close form of the KL divergence between Edgeworth expansions for $p = 4$ can be found [19], [20].

III. KL BASED MDDM FROM SWT AND FDCT MODELING

A. MDDM

Consider the M size sequence $\mathcal{I}$ given in Section II-A. Section II-E provides $2M^2$ SWT and FDCT similarity measurements (with a redundancy factor of 2 due to symmetry).

We propose constructing SWT and FDCT MDDMs from:

$$\mathbf{K}_{\mathcal{W}} = (\mathcal{K}_{\mathcal{W}}(\mathcal{I}_m, \mathcal{I}_\ell))_{1 \leqslant m \leqslant M, 1 \leqslant \ell \leqslant M}$$

and

$$\mathbf{K}_{\mathcal{C}} = (\mathcal{K}_{\mathcal{C}}(\mathcal{I}_m, \mathcal{I}_\ell))_{1 \leqslant m \leqslant M, 1 \leqslant \ell \leqslant M}.$$

These matrices relate to the different cross similarities between pairwise observations $(\mathcal{I}_m, \mathcal{I}_\ell)_{1 \leqslant m \leqslant M, 1 \leqslant \ell \leqslant M}$.

Matrices $\mathbf{K}_{\mathcal{W}}$ and $\mathbf{K}_{\mathcal{C}}$ are symmetric and have the form of the following matrix:

$$
\mathbf{K} = \begin{pmatrix}
0 & \star & \cdot & \cdot \\
\star & 0 & \star & \cdot \\
\cdot & \star & 0 & \star \\
\cdot & \cdot & \star & 0
\end{pmatrix}
$$

where a star symbol is used to indicate the elements of the *second diagonal*.

The motivation beyond the use of $\mathbf{K}$ is to provide an intuitive and relevant tool for the analysis of image time series in change-image/epoch detection purpose. It is worth emphasizing that matrix $\mathbf{K}$ can be used in an operational context: when a new acquisition is provided, updating MDDM $\mathbf{K}$ only requires:

1) computing the P parameters associated with the parsimonious description of the current acquisition,
2) computing the KL divergences between the above parameters and those of the previous acquisitions and
3) adding a new row and a new column composed with the above KL divergences to $\mathbf{K}$.

Thus the MDDM matrix format suits for long size image time series.

B. Aggregation of SWT and FDCT MDDMs

The $2M^2$ SWT and FDCT cross similarity measurements between elements of sequence $\mathcal{I}$ will hereafter be displayed by using a single MDDM. Indeed, since we have considered the symmetric Kullback-Leibler divergence, then the corresponding SWT and FDCT MDDMs are symmetric, so that only the upper triangular MDDM matrix is necessary and sufficient for information processing, per transform. Aggregating these MDDMs can thus be performed by superimposing information from SWT and FDCT MDDMs: by replacing the redundant information issued from a single SWT MDDM with complementary
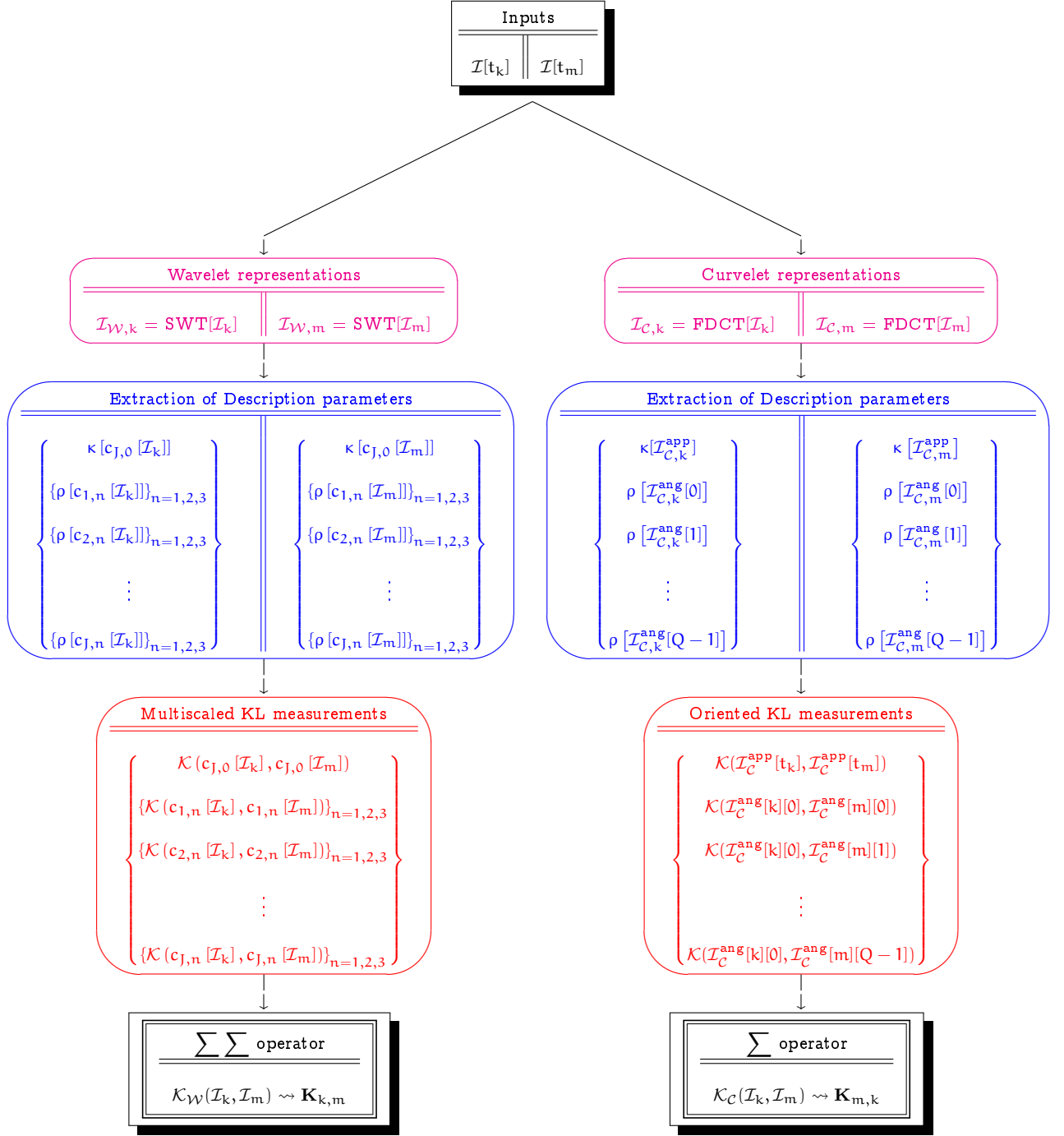
Fig. 3. Flowchart representing the steps involved in MDDM value computation. In this flowchart, vector $\kappa = (\kappa_\ell)_{\ell=1,2,3,4}$ designate the first four cumulants and vector $\rho = (\rho_1, \rho_2)$ represents the scale and shape parameters of the best model obtained from $\mathcal{D}$.

information issued from the FDCT MDDM (redundancy follows from the symmetry of the MDDM), we derive the joint SWT - FDCT MDDM $\mathbf{K}_{\mathcal{W},\mathcal{C}}$, with

$$\mathbf{K}_{\mathcal{W},\mathcal{C}} = \left(\mathcal{K}_{\mathcal{W}}(\mathcal{I}_m, \mathcal{I}_\ell) \mathbb{1}_{\{m \leqslant \ell\}} + \mathcal{K}_{\mathcal{C}}(\mathcal{I}_m, \mathcal{I}_\ell) \mathbb{1}_{\{m \geqslant \ell\}} \right)_{\substack{1 \leqslant m \leqslant M \\ 1 \leqslant \ell \leqslant M}}$$

In MDDM $\mathbf{K}_{\mathcal{W},\mathcal{C}}$, the upper triangular divergence matrix represents SWT similarity measurements as expressed in (3) and the lower triangular divergence matrix provides FDCT similarity measurements from (4). Fig. 3 provides

an overview of the different steps involved in the MDDM $\mathbf{K}_{\mathcal{W},\mathcal{C}}$ computation.

In matrix $\mathbf{K}_{\mathcal{W},\mathcal{C}}$, the SWT and FDCT MDDMs are fused (lossless fusion) in order to have a complementary viewpoint that can be helpful for decision taking in many situations. In what follows, representation $\mathbf{K}_{\mathcal{W},\mathcal{C}}$ will be used for plotting MDDMs whereas analysis of the time series with respect to the MDDM properties will be addressed with respect to one among MDDMs $\mathbf{K}_{\mathcal{W}}$ and $\mathbf{K}_{\mathcal{C}}$.
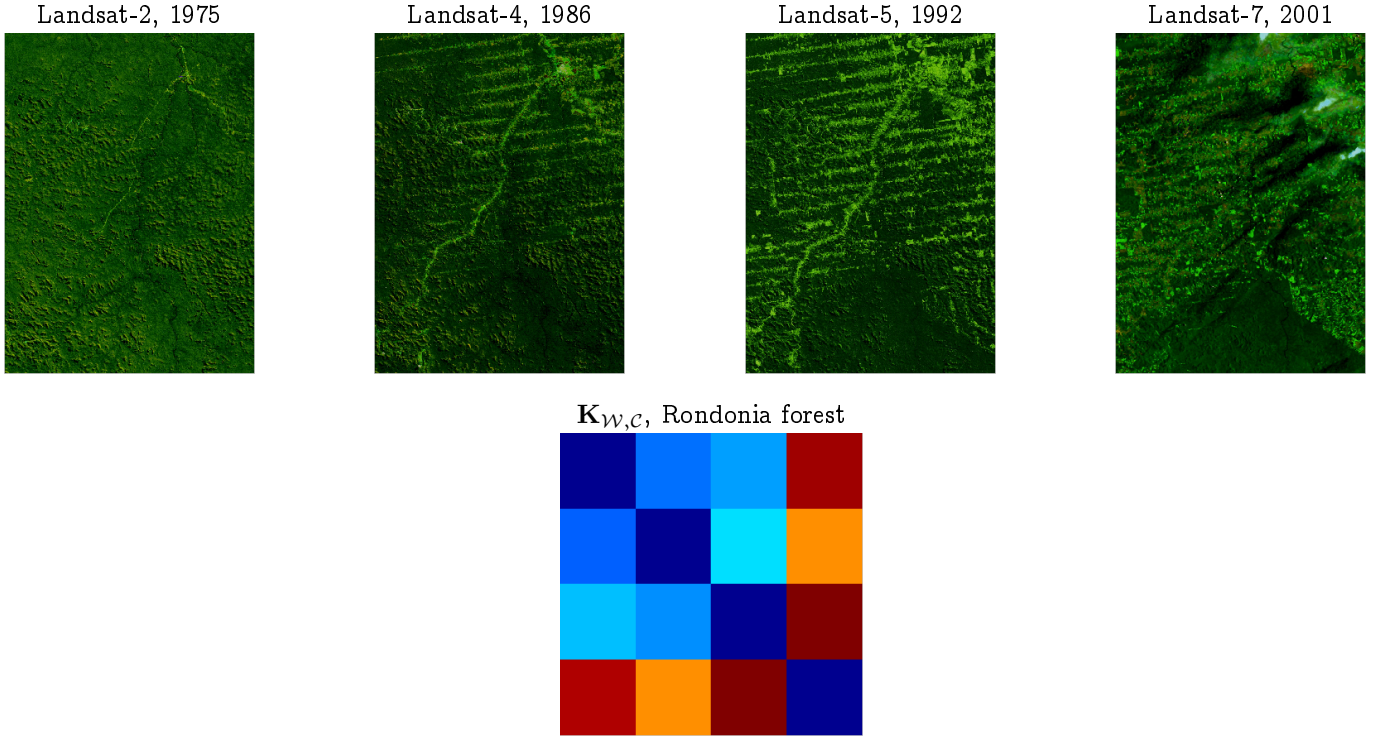
Fig. 4. Multi-date image sequence from Rondonia forest, NASA/courtesy of nasaimages.org, and its MDDM $\mathbf{K}_{\mathcal{W},\mathcal{C}}$. The upper (resp. lower) triangular matrix of the MDDM represents the SWT (resp. FDCT) contributions in $\mathbf{K}_{\mathcal{W},\mathcal{C}}$. The first row and the second diagonal of $\mathbf{K}_{\mathcal{W},\mathcal{C}}$ are increasing sequences: the changes are progressive and maintained. In addition, the changes from the last column of $\mathbf{K}_{\mathcal{W},\mathcal{C}}$ appears abrupt as corroborated by the visual image comparison. Images are with size 1920×2880.

C. Change-image detection from MDDM analysis

Let us focus on an MDDM $\mathbf{K}$ being either $\mathbf{K}_{\mathcal{W}}$ or $\mathbf{K}_{\mathcal{C}}$. The analysis of $\mathcal{I}$ derives from the construction of MDDM $\mathbf{K}$:

- In absence of significant changes from the observed scene and the acquisition system, matrix $\mathbf{K}$ does not exhibit any particular structure: it represents only acquisition noise.
- In presence of a gradually maintained change (Example: deforestation),
  - the second diagonal of $\mathbf{K}$, as well as,
  - each semi-row of $\mathbf{K}$ starting from a second diagonal element (that is $\{\mathbf{K}_{m,\ell} : \ell = m + 1, \ldots, M\}$),

  are increasing sequences, if we assume that the different images are conform, i.e. the acquisitions have been achieved in the same conditions.
- An abrupt change occurs when every semi-row of $\mathbf{K}$ starting from a second diagonal element can be roughly associated with a step-like function in the following sense: the divergence values across the semi-row form two clusters (or stationary subsequences), a first cluster with small divergence values representing only acquisition noise (before changes) and a second cluster with large divergence values. The location in time of the changes lies in-between the two clusters.

- In most cases, the analysis of the neighborhood of a divergence values given in the MDDM characterizes and helps understanding the changes or the acquisition conditions at the corresponding acquisition time.

Fig. 4 presents an illustration of progressive changes from an MDDM $\mathbf{K}_{\mathcal{W},\mathcal{C}}$ computed from a sequence of Landsat images. Images given in Fig. 4 show a part of the Rondonia forest (Brazil) that has been subject to a drastic deforestation (more than 15000km^2 devastated per year). The images clearly highlight through a visual comparison, the progressive changes that have led to the forest destruction. These changes and their progressive nature appear in MDDM[2] $\mathbf{K}_{\mathcal{W},\mathcal{C}}$: the second diagonals and any of the semi-column/row starting with a diagonal element are increasing sequences. The abrupt nature of the global change also appear by noting that the last column (respectively row) has very large $\mathbf{K}_{\mathcal{W},\mathcal{C}}$ values in comparison to the other columns (respectively rows).

[2]The MDDM colors have been obtained from the following convention: We first apply the normalization: $\mathcal{K}_{\mathcal{W}} \equiv \mathcal{K}_{\mathcal{W}}/\max[\mathcal{K}_{\mathcal{W}}]$ and $\mathcal{K}_{\mathcal{C}} \equiv \mathcal{K}_{\mathcal{C}}/\max[\mathcal{K}_{\mathcal{C}}]$ where $\max[M]$ denotes here the maximum over elements of the matrix M. We then construct the map $\mathbf{K}_{\mathcal{W},\mathcal{C}}$ from the above normalized similarity measurements. The color map then ranges from blue (associated with 0) to red (associated with 1) by passing through cyan, yellow, and orange. The same convention is used for all the MDDM color compositions given in the paper.

D. MDDM and the non-conformity of an acquisition with respect to the image time series

ERS 2, June 06, 2003   ENVISAT, August 15, 2003

ENVISAT, June 25, 2004   ERS 2, June 10, 2005

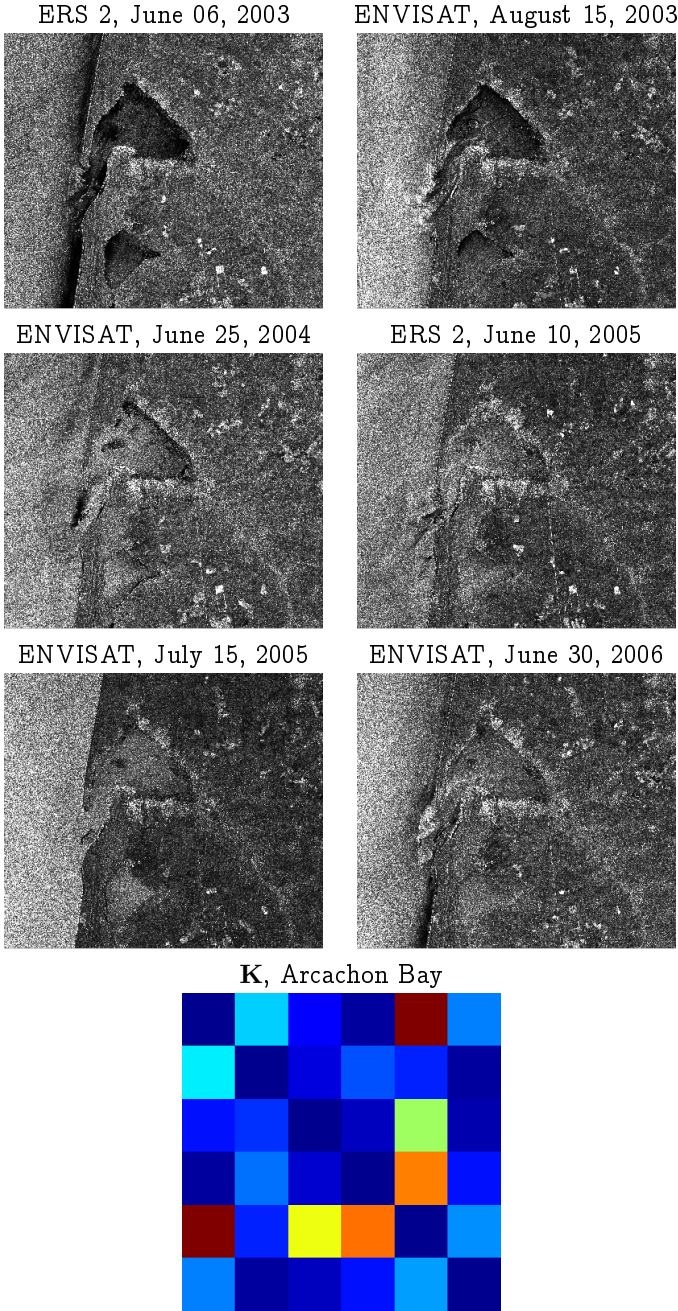ENVISAT, July 15, 2005   ENVISAT, June 30, 2006

$\mathbf{K}$, Arcachon Bay

Fig. 5. Multitemporal ERS 2 and ENVISAT image sequence, CNES/courtesy. The MDDM $\mathbf{K}_{\mathcal{W},\mathcal{C}}$ only reflects acquisition noise and an outlier (ENVISAT acquisition of 2005, July 15. Images are with size 3840×4480.

A non-conform acquisition (image-outlier) is characterized by a column of $\mathbf{K}$ with large divergence values, among the other divergence values representing acquisition noise. Note that the ℓ-th column $\mathbf{K}[\ell]$ of $\mathbf{K}$ represents dissimilarities of acquisitions in $\mathcal{I}$ with respect to the acquisition performed at time $\mathbf{t}_\ell$. As a consequence, an index relating the non-conformity (or inconsistency) of a given acquisition is the cumulative dissimilarities from the columns of $\mathbf{K}$:

$$\mathfrak{D}[\ell] = \|\mathbf{K}[\ell]\|_1 = \sum_{m=1}^{M} \mathbf{K}_{m,\ell}. \qquad (5)$$

TABLE III
Non-conformity of acquisitions in Rondonia and Arcachon time series. Strong deviation from the normal behavior of the sequence occur for acquisitions with large $\mathfrak{D}$ values (red, blue and green colors in the table).

Rondiana	ℓ	1	2	3	4
	$\mathfrak{D}_{\mathcal{W}}[\ell]$	11.54	7.27	13.03	25.32
	$\mathfrak{D}_{\mathcal{C}}[\ell]$	7.41	4.94	8.49	16.31

Arcachon						
ℓ	1	2	3	4	5	6
$\mathfrak{D}_{\mathcal{W}}[\ell]$	5.56	2.89	3.23	3.85	8.82	2.36
$\mathfrak{D}_{\mathcal{C}}[\ell]$	9.97	4.58	4.81	6.68	15.67	3.94

For illustrating the role of index $\mathfrak{D}$, let us consider the Rondonia images of Fig. 4: Table III provides index $\mathfrak{D}$ measurements from the MDDM of the corresponding images. It appears clearly that the image acquired in 2001 differs significantly from the other images of the Rondonia time series. Note that for this trivial example, the above remark can be observed directly from the Rondonia optical image time series or its MDDM $\mathbf{K}$.

The analysis of SAR time series such as the one given in Fig. 5 is more intricate. Images in Fig. 5 are ERS 2 and ENVISAT SAR images showing the Arcachon Bay (France). In addition with the noise induced by acquisition conditions, other disturbances characterizing this scene are swell movement, sand displacement, vegetation variability and the presence of boats depending on the acquisition time. The MDDM $\mathbf{K}$ does not highlight any suspicious behavior, except that acquisition 5 (July 15, 2005) is not conform (the column 5 and row 5 have rather large values), in comparison with the other elements of the Arcachon time series. In particular, Table III provides the 3 acquisitions that are less conform from the analysis with respect to index $\mathfrak{D}$ (colors red, blue and green in Table III).

The following section provides experimental results on the application of $\mathbf{K}$ and $\mathfrak{D}$ for the analysis of image time series involving 13 and 11 acquisition dates respectively.

IV. Analysis of TS-X image time series by using MDDMs

In this section, we consider two high resolution SAR image time series with respectively 13 and 10 images. The problem tackled is glacier monitoring in a change detection context. The sequences considered are TS-X images of Chamonix Mont-Blanc test site in the French Alps. Among these mountains, we focus hereafter on the Argentière glacier.

We first consider a time series acquired from May 31, 2009 to October 21, 2009, in ascending orbit. The dates of acquisition are given in Fig. 6. The MDDM $\mathbf{K}_{\mathcal{W},\mathcal{C}}$ is provided in the same figure.
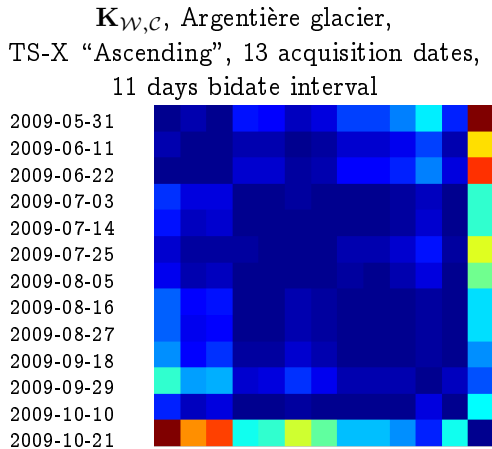
$\mathbf{K}_{\mathcal{W},\mathcal{C}}$, Argentière glacier,
TS-X "Ascending", 13 acquisition dates,
11 days bidate interval

Fig. 6.   MMDM $\mathbf{K}_{\mathcal{W},\mathcal{C}}$ for a multi-date TS-X image sequence in ascending orbit. The changes from the MDDM are progressive from July to September and reflect the glacier surface evolution.

As it can be seen from this MDDM, the upper right wedge of $\mathbf{K}$ has large divergence values whereas the second diagonal reflects only acquisition noise: changes are progressive and follows from climatic seasonal variation that impact the consistency of the primary layer of the glacier. For change evaluation in the spatial domain, acquisitions-of-interest from the above MDDM are the first and the last acquisitions from the image time series (dates 05-31 and 21-10).

The cumulative dissimilarity indices $\mathfrak{D}[\ell], \ell = 1, 2, \ldots, 13$ of this time series are given in Table IV. It appears from this table that neither image $\mathcal{I}_{13}$, nor image $\mathcal{I}_1$ are representative of the glacier state on the acquisition period: $\mathfrak{D}[1]$ and $\mathfrak{D}[13]$ deviate significantly from the rest of the cumulative dissimilarities. Table IV reveals a rather stationary state of the glacier from the period of July 03, 2009 to September 18, 2009. After this stationary period, we observe a significant increase of the cumulative dissimilarity indices that can be explained by recent snowfalls and/or short-time continuous snow deposit on the glacier at the beginning of the autumn.

We now consider a time series of TS-X images of the Argentière glacier acquired in descending orbit (acquisition dates are given in Fig. 7). Images are displayed in the SAR slant range geometry.

This sequence extends to one year, with the specificity that no images are available between 2008-10-22 and 2009-05-28 (period including the winter). As it can be seen from the MDDM $\mathbf{K}_{\mathcal{W},\mathcal{C}}$ of Fig. 7, this "disconnection" have been well detected. Indeed, this "disconnection" involves two different seasons where the glacier state is such that the first layers of the glacier have different backscattering

$\mathbf{K}_{\mathcal{W},\mathcal{C}}$, Argentière glacier,
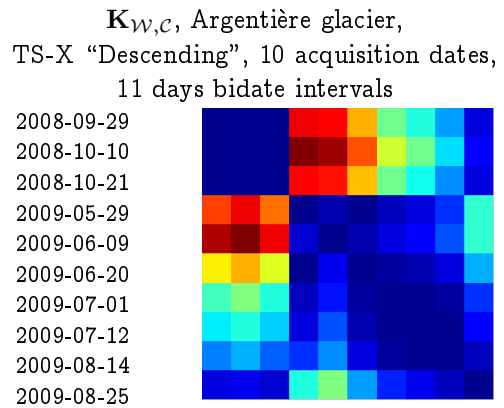TS-X "Descending", 10 acquisition dates,
11 days bidate intervals

Fig. 7.   MDDM $\mathbf{K}_{\mathcal{W},\mathcal{C}}$ for a sequence of TS-X images acquired over Argentière glacier in descending orbit. Changes are abrupt (the first row of the KL MDDM has a step located at the acquisition date 4 whereas the second diagonal has a unique outlier located at date 4) and progressive (decay of semi-row sequences after acquisition date 4). Images are with size 3072×4864.

properties. From the analysis of this MDDM, the acquisitions we will focus on are those performed at 2008: 09-29 ($\mathcal{I}_1$), 05-29 ($\mathcal{I}_4$) and 08-25 ($\mathcal{I}_{10}$).

For this TS-X sequence, SWT and FDCT indices $\mathfrak{D}$ show the same behavior in terms of the stationary sequences (see Table V), but have different ranking of cumulative deviations. The "mean" and the "max" fusion rules applied to SWT and FDCT indices $\mathfrak{D}$ emphasize that FDCT results are more relevant than the SWT ones for analyzing this time series. This can be explained by refering to the suitability of FDCT for representing geometrically regular objects. The above results are instructive in the sense that it reveals a fact that is specific to the acquisition mode and the topography of this Argentière's area: glacier surface and geometry is more visible in descending acquisitions due to the glacier slope and valley main orientation. In this respect, we can recommend analyzing the FDCT based modeling for the TS-X descending mode when observing the Argentière glacier.

Note that, from Table V, one can distinguish mainly 2 changes in time which separate the acquisitions before 2009-05-29 and after 2009-06-20, the corresponding subsequences having rather stationary cumulative divergences.

V. SPATIAL CHANGE DETECTION FROM DISSIMILARITY MAPS BETWEEN STATIONARY SUBSEQUENCES

The analysis performed in Section IV above makes it possible to detect change-images in image time series. From this analysis, certain image time series can be seen, coarsely, as composed of a stationary subsequences in-between the corresponding change-images.

The following addresses spatial change detection between the stationary subsequences identified from a time series. We begin by presenting the basics of the method through an example in Section V-A. We then apply this

11

TABLE IV

NON-CONFORMITY OF ACQUISITIONS IN THE ARGENTIÈRE "ASCENDING" TIME SERIES. STRONG DEVIATION FROM THE NORMAL BEHAVIOR OF THE SEQUENCE OCCUR FOR ACQUISITIONS WITH LARGE $\mathfrak{D}$ VALUES (RED, BLUE AND GREEN COLORS IN THE TABLE).

Dates	Argentière "ascending", 2009												
	05-31	06-11	06-22	07-03	07-14	07-25	08-05	08-16	08-27	09-18	09-29	10-10	10-21
$\mathfrak{D}_\mathcal{W}[\ell]$	6.76	4.03	4.52	2.07	2.02	2.60	2.10	2.23	2.15	2.43	4.28	2.03	14.14
$\mathfrak{D}_\mathcal{C}[\ell]$	6.94	3.50	4.73	2.09	2.06	2.75	2.20	2.21	2.17	2.52	3.84	2.04	15.93

TABLE V

NON-CONFORMITY OF ACQUISITIONS IN THE ARGENTIÈRE "DESCENDING" TIME SERIES. STRONG DEVIATION FROM THE NORMAL BEHAVIOR OF THE SEQUENCE OCCUR FOR ACQUISITIONS WITH LARGE $\mathfrak{D}$ VALUES (RED, BLUE AND GREEN COLORS IN THE TABLE).

	Argentière "descending", 2008-2009									
	2008			2009						
Dates	09-29	10-10	10-21	05-29	06-09	06-20	07-01	07-12	08-14	08-25
$\mathfrak{D}_\mathcal{W}[\ell]$	6.05	6.68	5.61	5.58	7.09	4.22	2.99	2.62	2.33	2.96
$\mathfrak{D}_\mathcal{C}[\ell]$	8.42	9.70	8.21	8.05	8.16	6.06	4.26	3.74	3.21	3.96

method for analyzing subsequences issued from the TS-X time series provided by Fig. 6 and Fig. 7 in Section V-B.

A. Spatial change detection: basics

Specifically, we consider images $\mathcal{I}_2$ and $\mathcal{I}_5$ of the sequence of the Argentière glacier described in Fig. 7. These images are TS-X images of the Argentière glacier (French Alps) acquired at dates 2008-10-10 and 2009-06-09, respectively. Images $\mathcal{I}_2$ and $\mathcal{I}_5$ are with sizes 3072×4864.

Change detection between $\mathcal{I}_2$ and $\mathcal{I}_5$, *i.e.* in the observation space, involves evaluating local dissimilarities between these images. This problem is addressed hereafter by using the parsimonious descriptions and the similarity measures given in Section II. Indeed, for high resolution images such as TS-X images, the (local) information on a region of interest spreads over a large size pixel neighborhood. This makes relevant, the use of the statistical tools presented in Section II for retrieving and comparing local image features.

In particular, when considering the SWT and FDCT approximation modeling, this relevancy can be highlighted by the following remark.

In [21], the information issued from the 4 first cumulants[3] is shown to be suitable for the description of 8-bit RADARSAT images. In contrast, when 16-bit high resolution TS-X images are concerned, Fig. 8 shows that the information derived from the 4 first cumulants[4] is not sufficient to capture the image dynamic. These 4 first cumulants (relating the mean, standard deviation, asymmetry and kurtosis) are rather suitable for modeling SWT and FDCT approximations, due to the fact that statistical distribution regularizes in the wavelet domain (iterative convolution properties).

[3]The information from the 4 first cumulants is expressed in the Pearson system in [21].

[4]The information from the 4 first cumulants is expressed in the Edgeworth expansion in this paper.
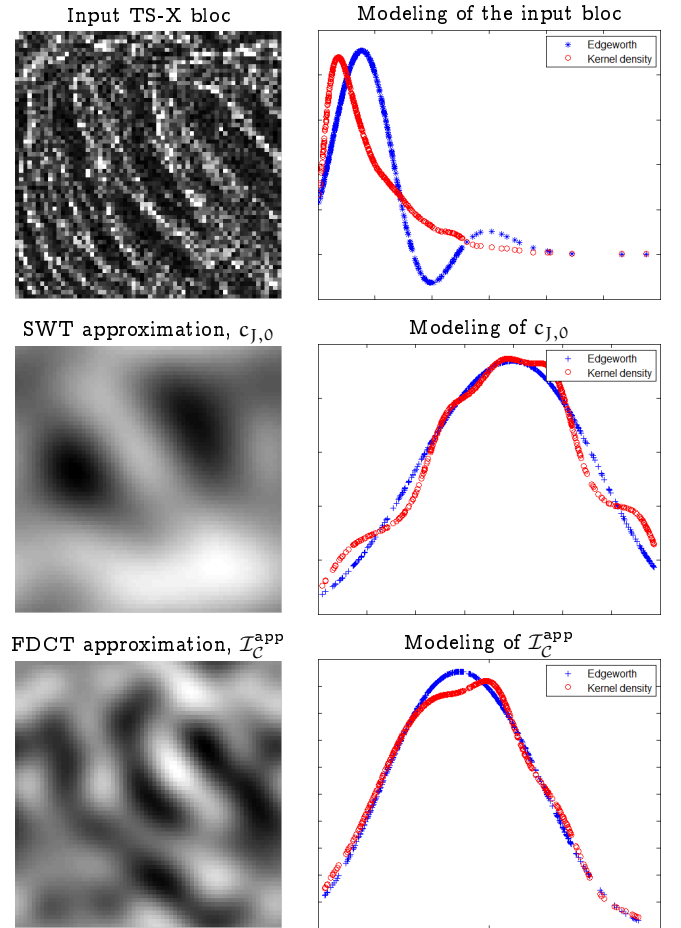

Fig. 8. Experimental setup: Edgeworth (parametric) expansion of order 4 is applied for modeling a 64×64 TS-X bloc (over the Argentière glacier serac area). The modeling performed in the image domain (top) is irrelevant whereas the modeling is relevant for the SWT (middle) and FDCT (down) approximations of the input TS-X bloc. The comparison of the Edgeworth expansion is made with respect to a smoothed histogram obtained by using (non-parametric) kernel density estimation (in red).
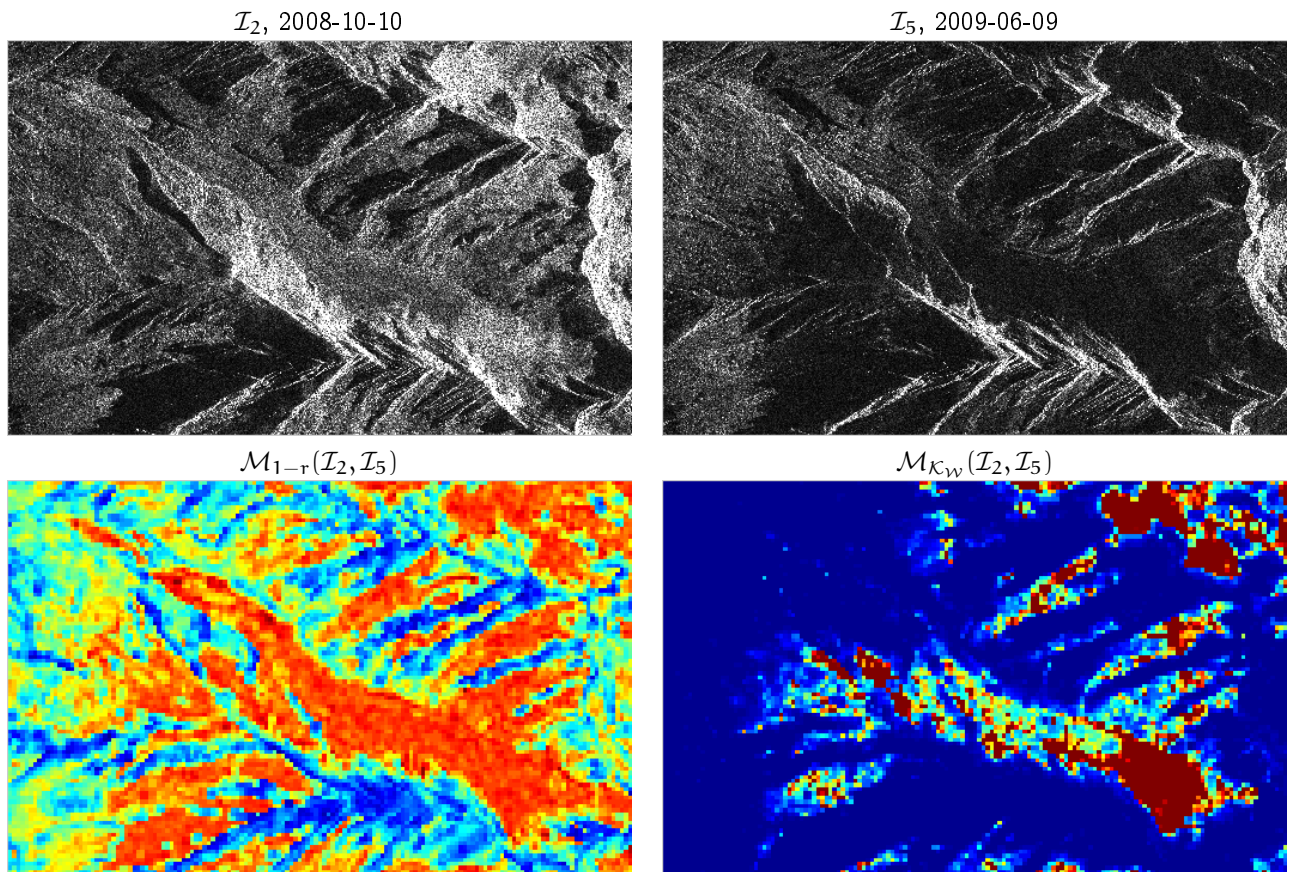
Fig. 9.   TS-X images of the Argentière glacier in descending orbit and their dissimilarity maps $\mathcal{M}_{1-r}$, $\mathcal{M}_{\mathcal{K}_\mathcal{W}}$, where r denotes the use of the correlation coefficient and $\mathcal{K}_\mathcal{W}$ relates to the measure given by (3). For every method, the dissimilarity values are linearly scaled to form a colored image where the map of colors ranges from blue (minimum) to red (maximum) by passing through the colors green, yellow, and orange. Changes appear in red color whereas non-changed areas appear in blue. The bi-date correlation is approximately the same all along the Argentière glacier so that changes seems uniform on the glacier: only second order changes are captured by the correlations coefficients. We need to take into account high order bi-date dependencies from $\mathcal{M}_{\mathcal{K}_\mathcal{W}}$ in order to discriminate the effective changes.

Furthermore, this relevancy is justified by noting that the parsimonious description under consideration encompasses both low and high order cumulants of the input random field. In this respect, the characterization induced by the parameters of this description is more relevant than using a single variable such as the correlation coefficient. Indeed, the latter parameter mainly reveals second order bidate dependencies, when applied locally on any pair $(\mathcal{I}_m, \mathcal{I}_\ell)$ of observations, whereas second and higher order bidate dependencies are considered when using the parsimonious descriptions of local neighborhoods of $(\mathcal{I}_m, \mathcal{I}_\ell)$.

For illustration, Fig. 9 provides dissimilarities between $\mathcal{I}_1$ and $\mathcal{I}_2$ by using measurements from $\mathcal{M}_{\mathcal{K}_\mathcal{W}}$ and from $1 - r$, where $\mathcal{M}_{\mathcal{K}_\mathcal{W}}(\mathcal{I}_m, \mathcal{I}_\ell)$ (resp. $r(\mathcal{I}_m, \mathcal{I}_\ell)$) denotes the MDDM obtained from measurements by using (3) and (resp. correlation coefficient) between $\mathcal{I}_m$ and $\mathcal{I}_\ell$. These measurements are computed from sliding and overlapping windows with sizes 64×64, the overlap induced by the sliding window beginning at the center of the initial window.

As it can be seen in Fig. 9, the correlation between $\mathcal{I}_1$ and $\mathcal{I}_2$ is too weak all over the Argentière glacier, due to the speckle decorrelation. It follows that correlation is not relevant for assessing change significance in this context. In contrast with the correlation coefficient, $\mathcal{K}_\mathcal{W}$ highlights change amplitudes depending on the mountain's geometry: accumulation areas, seracs, ... This relevancy follows from the capability of the multiscaled and oriented SWT and FDCT similarity measurements in $\mathcal{K}_\mathcal{W}$ for discriminating different change levels which affect both correlation and higher order statistics.

B. Spatial change detection: similarities from acquisitions-of-interest identified in TS-X time series

From the MDDM of Fig. 6, we can derive mainly two acquisitions-of-interest; those performed at dates 2009-05-31 and 2009-10-21, respectively. The corresponding images are given in Fig. 10.

These acquisitions-of-interest reflect the state of the glacier over two "extremal" behaviors, as it can be seen

$\mathcal{I}_1$, 2009-05-31 $\mathcal{I}_{13}$, 2009-10-21 $\mathcal{M}(\mathcal{I}_1,\mathcal{I}_{13})$
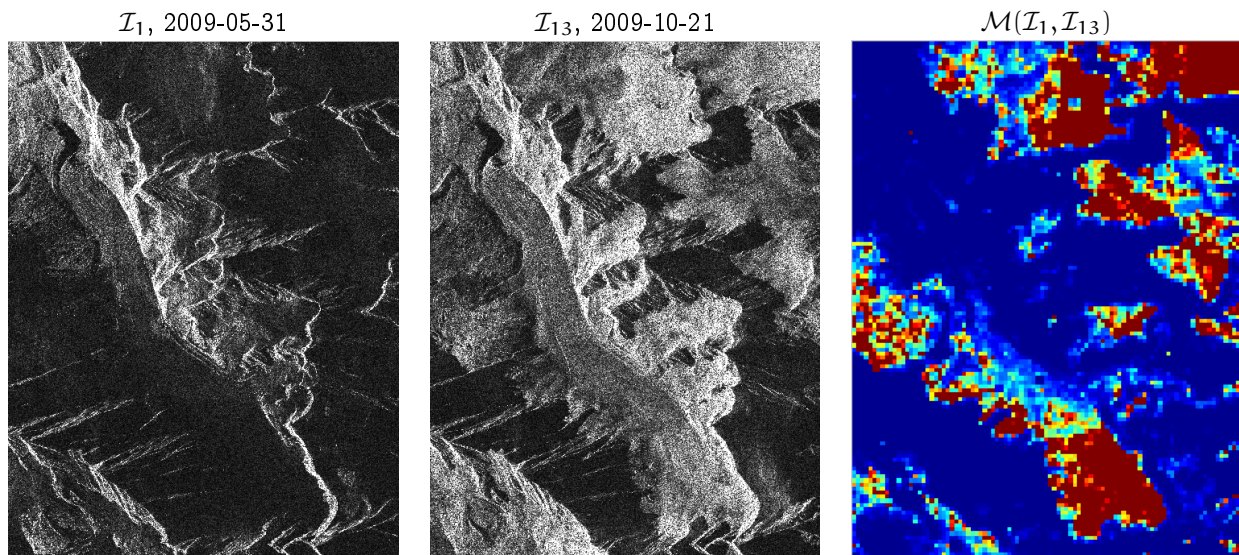

Fig. 10.   Dissimilarity map between two images issued from the TS-X sequence given in Fig. 6. The acquisitions-of-interest are the first and last acquisitions from the sequence. Images are with size 4096×3072.

by comparing image $\mathcal{I}_1$ (dusky) and image $\mathcal{I}_{13}$ (with high luminance) provided in Fig. 10. In particular,

- the "dusky" state characterizing the glacier in May is confirmed by comparing image $\mathcal{I}_1$ of Fig. 10 (acquired in ascending orbit at May 31) with $\mathcal{I}_5$ given in Fig. 9 (acquired at 9 days interval, in descending orbit).
- the "high luminance" characterizing the glacier state in October is assessed by comparing image $\mathcal{I}_{13}$ of Fig. 10 and image $\mathcal{I}_2$ given in Fig. 9: these images are acquired at the same time period (one year period and 11 days interval).

Thus, the state of the glacier during short-time stationary sequences is characterized by images that look similar in terms of intensity variations.

Spatial inter-seasonal dissimilarities can then be evaluated from measurements $\mathcal{M}(\cdot,\cdot)$ computed over overlapping windows extracted from the acquisitions-of-interest $\mathcal{I}_1,\mathcal{I}_{13}$: these dissimilarities are given in Fig. 10. It follows from $\mathcal{M}(\mathcal{I}_1,\mathcal{I}_{13})$ that the main difference between the glacier layer composition at May and October is concentrated at the upper part of the glacier in the *acumulation area*, where fresh snowfall modifies the surface scattering (see [22]).

From the MDDM given in Fig. 7, we can derive 3 acquisitions-of-interest, those performed at dates 2008-09-29, 2009-05-29 and 2009-08-25. These acquisitions and their bidate dissimilarity maps are given in Fig. 11. One can conclude, by comparing the 3 dissimilarity maps $\mathcal{M}(\mathcal{I}_1,\mathcal{I}_4),\mathcal{M}(\mathcal{I}_4,\mathcal{I}_{10})$ and $\mathcal{M}(\mathcal{I}_{10},\mathcal{I}_1)$ that

- changes that do not appear in the 3 dissimilarity maps are occasional or transient. They follow (mainly) climatic conditions which are intrinsic to the different seasons (consequences of the winter and the summer

in a high mountainous glacier area) and depend on 1) the glacier elevation and 2) the glacier slope orientation: in the change maps given in Fig. 11, the change distributions in amplitude relate to the form and the geometry of the glacier.
- Permanent changes may refer to unstable zones like moraines or to continuous displacements of some glacier blocks, for instance in crevasses or serac fall areas.

VI. Conclusion

In this paper, we have proposed a new approach for analyzing SAR image time series. This approach relies on the computation of a multi-date divergence matrix (MDDM) with respect to some cross divergences issued from the sequences of SAR image description parameters.

Description parameters are issued from SWT and FDCT subband modeling. The modeling has been addressed by using Edgeworth expansions for approximation subbands and GG, LOGN and WBL statistical distribution families for detail and orientation subbands. The approach is multi-model (parametric modeling is not restricted to a single distribution family) and the parametric computation of Kullback-Leibler divergence has been extended to measure inter-family dissimilarities (between distribution families GG, LOGN and WBL).

This approach is computationally workable for long size image time series and makes their analysis straightforward from an MDDM display (that can support series with more than 1000 images). This analysis have shown relevancy in monitoring the temporal evolution of Alps glaciers at two different levels: firstly at the image (or sub-image) level, by analyzing the proposed MDDM and cumulative dissimilarity indices that reveal non-conform images
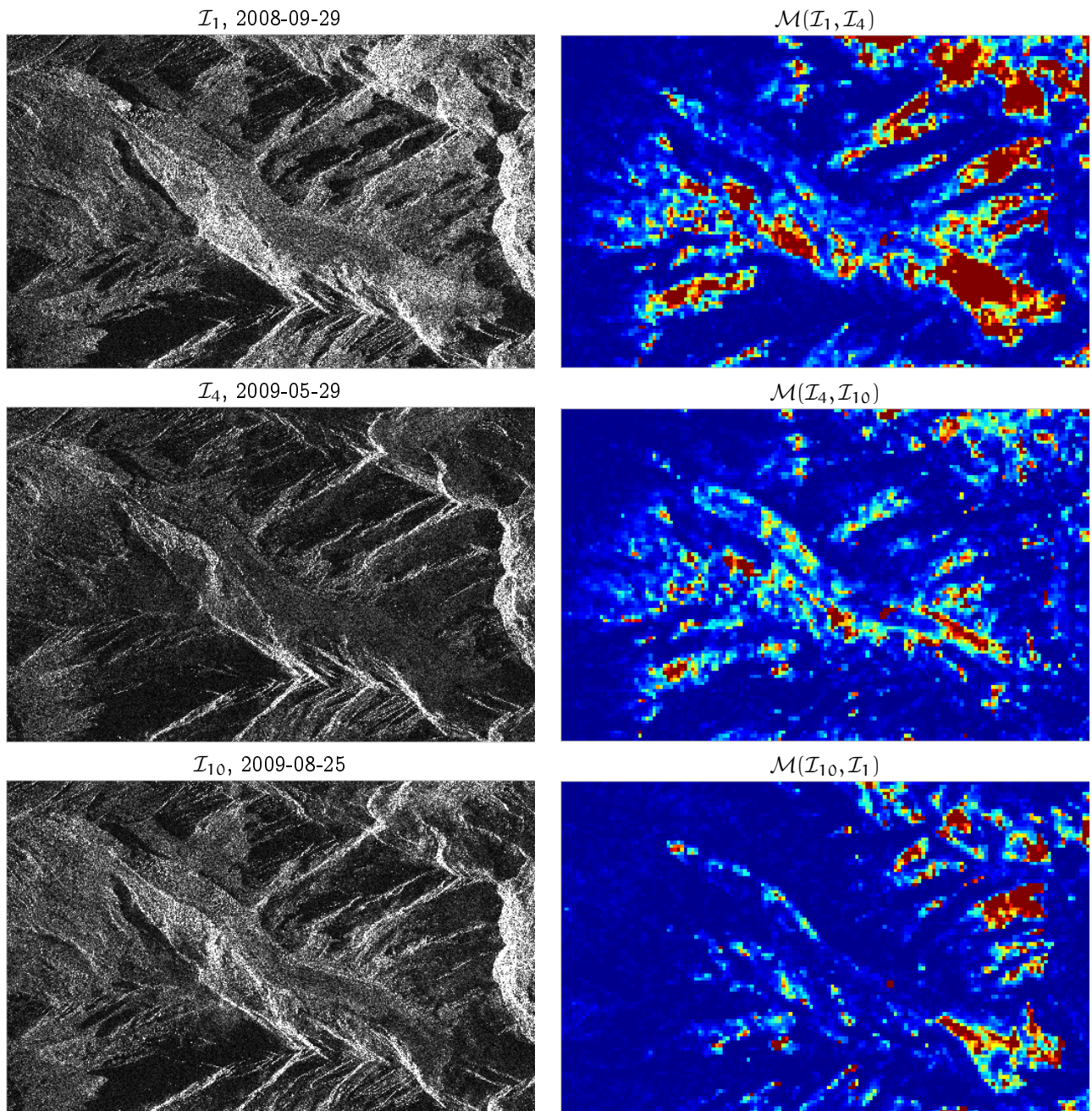
Fig. 11. Change detection for some acquisitions-of-interest issued from the Argentière glacier TS-X sequence of Fig. 7. Images are with size 3072 × 4864.

and stationary subsequences; then at the pixel level, by computing dissimilarity maps between images of interest selected from the MDDM, for a spatial analysis of changing areas.

The paper opens some prospects relating the statistics of MDDMs for image time series analysis. Indeed, an MDDM makes possible both global and local analysis of a time series: encapsulating global and local MDDMs in the neighborhood of a critical scene will ease a multilevel analysis of the scene under consideration. In addition, beyond the multi-date visual scene observation provided by MDDMs, an MDDM representation makes also possible

variance analysis in the time axis, searching for the *median* of the image time series, etc. Temporal segmentation tools could also be developed to analyze MDDMs and derive sub-matrices associated to seasonal or annual evolutions. This can be derived by applying matrix norms to the series of MDDMs obtained. This issue is challenging with respect to future SAR missions like *Sentinel 1* to be launched in 2013, with more systematic acquisition strategies (every 12 days, then every 6 days with two satellites) and with free data access.

ACKNOWLEDGMENTS

References

[1] L. Bruzzone, M. Marconcini, U. Wegmuller, and A. Wiesmann, "An advanced system for the automatic classification of multi-temporal SAR images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 6, pp. 1321 – 1334, June 2004.

[2] L. Guoxiang, S. M. Buckley, D. Xiaoli, C. Qiang, and L. Xiao-jun, "Estimating spatiotemporal ground deformation with im-proved permanent-scatterer radar interferometry," *IEEE Trans-actions on Geoscience and Remote Sensing*, vol. 47, no. 8, pp. 2762 – 2772, Aug. 2009.

[3] D. Brunner, G. Lemoine, and L. Bruzzone, "Earthquake damage assessment of buildings using VHR optical and SAR imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 5, pp. 2403 – 2420, May 2010.

[4] S. Lhermitte, J. Verbesselt, I. Jonckheere, K. Nackaerts, J. A. van Aardt, W. W. Verstraeten, and P. Coppin, "Hierarchical image segmentation based on similarity of NDVI time series," *Remote Sensing of Environment*, vol. 112, no. 2, pp. 506 – 521, 2008.

[5] A. Bastarrika, E. Chuvieco, and M. P. Martin, "Automatic burned land mapping from MODIS time series images: As-sessment in mediterranean ecosystems," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 9, pp. 3401 – 3413, sept. 2011.

[6] F. Bovolo, S. Marchesi, and L. Bruzzone, "A framework for automatic and unsupervised detection of multiple changes in multitemporal images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 6, pp. 2196 – 2212, June 2012.

[7] Y. Bazi, L. Bruzzone, and F. Melgani, "An unsupervised ap-proach based on the generalized gaussian model to automatic change detection in multitemporal SAR images," *IEEE Trans-actions on Geoscience and Remote Sensing*, vol. 43, no. 4, pp. 874 – 887, Apr. 2005.

[8] S. Park and W. W. Chu, "Similarity-based subsequence search in image sequence databases," *International Journal of Image and Graphics*, vol. 3, no. 1, pp. 31 – 53, 2003.

[9] J. Morio, P. Refregier, F. Goudail, P. C. Dubois-Fernandez, and X. Dupuis, "Information theory-based approach for contrast analysis in polarimetric and/or interferometric sar images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 8, pp. 2185 – 2196, Aug. 2008.

[10] E. Erten, A. Reigber, L. Ferro-Famil, and O. Hellwich, "A new coherent similarity measure for temporal multichannel scene characterization," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 7, pp. 2839 – 2851, July 2012.

[11] R. R. Coifman and D. L. Donoho, *Translation invariant de-noising*. Lecture Notes in Statistics, 1995, no. 103, pp. 125 – 150.

[12] E. J. Candès and D. L. Donoho, "Curvelets - a surprisingly effective nonadaptive representation for objects with edges," *Curves and Surfaces, L. L. Schumaker et al. (eds), Vanderbilt University Press, Nashville, TN*, 2000.

[13] E. J. Candès and D. L. Donoho, "New tight frames of curvelets and optimal representations of objects with piecewise C^2 singu-larities," *Communications on Pure and Applied Mathematics, Wiley Periodicals, Inc., A Wiley Company*, vol. 57, no. 2, pp. 219 – 266, 2004.

[14] E. Candès, L. Demanet, D. Donoho, and L. Ying, "Fast dis-crete curvelet transforms," *Multiscale Modeling and Simulation*, vol. 5, no. 3, pp. 861 – 899, 2006.

[15] A. M. Atto and Y. Berthoumieu, "Wavelet packets of nonstationary random processes: Contributing factors for stationarity and decorrelation," *IEEE Transactions on Information Theory*, vol. 58, no. 1, Jan. 2012. [Online]. Available: http://dx.doi.org/10.1109/TIT.2011.2167496

[16] A. M. Atto and Y. Berthoumieu, "How to perform texture recognition from stochastic modeling in the wavelet domain," *IEEE International Conference on Acoustics, Speech, and Sig-nal Processing, ICASSP*, Prague, Czech Republic, May 22 - 27, 2011.

[17] A. M. Atto and Y. Berthoumieu, "Structuring of large and het-erogeneous texture databases," *IEEE Workshop on Statistical Signal Processing, SSP*, Nice, France, June 28 - 30, 2011.

[18] A. N. Kolmogorov, "Sulla determinazione empirica di una legge di distribuzione," *G. Ist. Ital. Attuari*, vol. 4, pp. 83 – 91, 1933.

[19] J. Lin, N. Saito, and R. Levine, "Edgeworth approximation of the Kullback-Leibler distance towards problems in image anal-ysis," *Univ. California, Davis. Tech. Rep. [Online]. Available: http://www.math.ucdavis.edu/~saito*, 1999.

[20] J. Inglada and G. Mercier, "A new statistical similarity measure for change detection in multitemporal SAR images and its extension to multiscale change analysis," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 5, pp. 1432 – 1445, May 2007.

[21] J. Inglada, "Change detection on SAR images by using a parametric estimation of the kullback-leibler divergence," in *IEEE International Geoscience and Remote Sensing Sympo-sium, IGARSS, Proceedings*, vol. 6, July 2003, pp. 4104 – 4106.

[22] R. Fallourd, O. Harant, E. Trouvé, J.-M. Nicolas, M. Gay, A. Walpersdorf, J.-L. Mugnier, J. Serafini, D. Rosu, L. Bom-brun, G. Vasile, N. Cotte, F. Vernier, F. Tupin, L. Moreau, and P. Bolon, "Monitoring temperate glacier displacement by multi-temporal TerraSAR-X images and continuous GPS mea-surements," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 4, no. 2, pp. 372 – 386, June 2011.

[23] P. McCullagh, *Tensor Methods in Statistics*. London, U.K.: Chapman & Hall, 1987.

[24] A. Stuart and J. K. Ord, *Kendall's Advanced Theory of Statis-tics*. 5th ed. London, U.K.: Arnold, 1991.

[25] M. N. Do and M. Vetterli, "Wavelet-based texture retrieval us-ing generalized gaussian density and kullback-leibler distance," *IEEE Transactions on Image Processing*, vol. 11, no. 2, pp. 146 – 158, Feb. 2002.

[26] R. Kwitt and A. Uhl, "Image similarity measurement by Kullback-Leibler divergences between complex wavelet subband statistics for texture retrieval," *IEEE International Conference on Image Processing, ICIP*, San Diego, California, USA, 12 - 15 October, pp. 933 – 936, 2008.

Appendix A
Analytic forms of the parametric distribution models

The Edgeworth expansion of order p of a random vari-able X, absolutely continuous, with mean μ and standard deviation σ is

$$f_p(x) = \frac{1}{\sqrt{2\pi}\sigma} \left(\sum_{r=0}^{p} \eta_r H_r(x) \right) e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \qquad (6)$$

where H_r is the Chebyshev-Hermite polynomial of order r [23] and the coefficient η_r is a function of the r firsts cumulants of X. Note that when $p = \infty$, the right hand side expression of the above equation is exactly the proba-bility density function of X, under regularity assumptions on this function [24].

The GG, LOGN and the WBL distributions are used in a joint framework: this includes the case where a GG distribution has to be compared with a LOGN or a WBL distribution, and *vice versa*. In this respect (a LOGN and a WBL distribution operating on positive real values), we consider for comparison purpose, the magnitude of a random variable following the GG distribution: the magnitude of a GG distribution with *scale* α, *shape* β is given by:

$$f_{\alpha,\beta}(x) = \frac{\beta}{\alpha\Gamma(1/\beta)}e^{-(x/\alpha)^\beta}\mathbb{1}_{\mathbb{R}^+}(x), \quad (7)$$

where Γ is the standard Gamma function:

$$\Gamma(u) = \int_{\mathbb{R}^+} x^{u-1}e^{-x}dx$$

Note that this expression differs from the standard GG density by a factor 2.

The LOGN distribution with *log-scale* parameter $\mu > 0$ and *shape* σ^2 is defined by:

$$f_{\mu,\sigma}(x) = \frac{1}{x\sigma\sqrt{2\pi}}e^{-\left(\frac{(\log x - \mu)^2}{2\sigma^2}\right)}\mathbb{1}_{\mathbb{R}^+}(x). \quad (8)$$

The WBL distribution with *scale* $a > 0$ and *shape* $b > 0$ is defined by:

$$f_{a,b}(x) = \frac{b}{a}\left(\frac{x}{a}\right)^{b-1}e^{-\left(\frac{x}{a}\right)^b}\mathbb{1}_{\mathbb{R}^+}(x). \quad (9)$$

Appendix B
Intra-family KL divergences for GG, LOGN and WBL distributions

The expression of the KL measure for distributions pertaining to the the same parametric family is available for many distribution families from the literature.

For GG distributions, the KL divergence is given in [25]. The symmetric version of this divergence has the following form:

$$\mathcal{K}(X_1, X_2) = \left(\frac{\alpha_1}{\alpha_2}\right)^{\beta_2}\frac{\Gamma\left(\frac{1+\beta_2}{\beta_1}\right)}{\Gamma(1/\beta_1)}$$
$$+ \left(\frac{\alpha_2}{\alpha_1}\right)^{\beta_1}\frac{\Gamma\left(\frac{1+\beta_1}{\beta_2}\right)}{\Gamma(1/\beta_2)} - \frac{\beta_1 + \beta_2}{\beta_1\beta_2}. \quad (10)$$

For Weibull distributions, the KL divergence is given in [26]. We have that the symmetric version of this divergence is

$$\mathcal{K}(X_1, X_2) = \left(\frac{\lambda_1}{\lambda_2}\right)^{k_2}\Gamma\left(1 + \frac{k_2}{k_1}\right) + \left(\frac{\lambda_2}{\lambda_1}\right)^{k_1}\Gamma\left(1 + \frac{k_1}{k_2}\right)$$
$$+ \mathfrak{e}\left(\frac{k_1}{k_2} + \frac{k_2}{k_1} - 2\right) - 2$$
$$+ (k_1 - k_2)\log\frac{\lambda_1}{\lambda_2}. \quad (11)$$

where $\mathfrak{e}$ is the *Euler-Mascheroni constant.*

Concerning LOGN distributions, the explicite form of the parameter-dependent KL divergence derives from:

Proposition 1: The symmetric KL divergence between LOGN random variables X_1 and X_2 is given by:

$$\mathcal{K}(X_1, X_2) = \frac{1}{2}(\mu_1 - \mu_2)^2\left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}\right)$$
$$+ \frac{1}{2}\left(\frac{\sigma_1^2}{\sigma_2^2} + \frac{\sigma_2^2}{\sigma_1^2}\right) - 1. \quad (12)$$

Proof: Proposition 1 derives from straightforward calculus, by expanding the KL integral form as a sum of Gaussian moments. ∎

The following Appendix C derives close forms of the KL divergence for the joint framework involved in the use of a dictionary $\mathcal{D}$. This joint framework requires comparing distributions issued from different parametric families (Inter-family similarity measurements).

Appendix C
Inter-family KL divergences between GG, LOGN and WBL distributions

Consider two random variables X_1, X_2 with probability density functions f_{X_1}, f_{X_2} respectively. The symmetric KL similarity measure between X_1 and X_2 is:

$$\mathcal{K}(X_1, X_2) = \mathcal{H}(X_1\|X_2) - \mathcal{H}(X_1) + \mathcal{H}(X_2\|X_1) - \mathcal{H}(X_2) \quad (13)$$

where the entropy $\mathcal{H}$ is defined for a random variable X with probability density function f_X by

$$\mathcal{H}(X) = -\int_{\mathbb{R}} f_X(x)\log f_X(x)dx \quad (14)$$

and $\mathcal{H}(X_i\|X_j)$ denotes the entropy of random variable X_i conditional on the random variable X_j. This conditional entropy is

$$\mathcal{H}(X_i\|X_j) = -\int_{\mathbb{R}} f_{X_i}(x)\log f_{X_j}(x)dx \quad (15)$$

when X_i, X_j admit probability density functions f_{X_i}, f_{X_j} respectively.

A. KL Divergence between a Generalized Gaussian Magnitude and a Weibull distribution

Let X_w, X_g be random variables distributed as WBL and GG with probability density functions $f_{a,b}, f_{\alpha,\beta}$ respectively. Proposition 4 below provides the symmetric KL similarity measure between X_w and X_g. This proposition requires the following Propositions 2 and 3 relating the cross entropies between X_w and X_g. We have

Proposition 2: The conditional entropy $\mathcal{H}(X_w\|X_g)$ is given by

$$\mathcal{H}(X_w\|X_g) = \frac{\beta}{b}\left(\frac{a}{\alpha}\right)^\beta\Gamma\left(\frac{\beta}{b}\right) + \log\Gamma\left(\frac{1}{\beta}\right) + \log\frac{\alpha}{\beta}. \quad (16)$$

Proof: See Appendix D. ∎

Proposition 3: The conditional entropy $\mathcal{H}(X_g\|X_w)$ is given by

$$\mathcal{H}(X_g\|X_w) = \left(\frac{\alpha}{a}\right)^b \frac{\Gamma((b+1)/\beta)}{\Gamma(1/\beta)} - b\log\left(\frac{\alpha}{a}\right)$$
$$+ \log\frac{\alpha}{b} - \frac{b-1}{\beta}\psi\left(\frac{1}{\beta}\right), \qquad (17)$$

where ψ is the *digamma function* satisfying

$$\frac{d}{dx}\Gamma(x) = \psi(x)\Gamma(x). \qquad (18)$$

Proof: See Appendix E. ∎

From Propositions 2, 3, and by taking into account that the entropy of the magnitude of a GG random variable is

$$\mathcal{H}(X_g) = \frac{1}{\beta} + \log\left(\frac{1}{\beta}\right) + \log\left(\alpha\Gamma\left(\frac{1}{\beta}\right)\right) \qquad (19)$$

and the entropy of a WBL random variable is

$$\mathcal{H}(X_w) = \mathfrak{e}\left(1 - \frac{1}{b}\right) + \log\left(\frac{a}{b}\right) + 1, \qquad (20)$$

we have:

Proposition 4: The symmetric KL divergence between a WBL and the magnitude of a GG random variables is given by:

$$\mathcal{K}(X_w, X_g) = \left(\frac{\alpha}{a}\right)^b \frac{\Gamma((1+b)/\beta)}{\Gamma(1/\beta)} + \frac{\beta}{b}\left(\frac{a}{\alpha}\right)^\beta \Gamma\left(\frac{\beta}{b}\right) - \frac{1}{\beta}$$
$$+ (1-b)\left(\frac{\mathfrak{e}}{b} + \log\frac{\alpha}{a} + \frac{1}{\beta}\psi\left(\frac{1}{\beta}\right)\right) - 1 \qquad (21)$$

B. KL Divergence between a Log-normal and a Weibull distribution

To conclude this section, the following provides cross entropies and KL divergences between a LOGN and a WBL random variables X_ℓ, X_w with probability density functions $f_{\mu,\sigma}, f_{a,b}$ respectively.

Proposition 5: The conditional entropy $\mathcal{H}(X_w\|X_\ell)$ is given by

$$\mathcal{H}(X_w\|X_\ell) = \frac{1}{2}\log(2\pi\sigma^2) + \left(\log a + \frac{\mathfrak{e}}{b}\right) + \frac{\mathfrak{e}^2 + \pi^2/6}{2\sigma^2 b^2}$$
$$+ \frac{1}{\sigma^2}\left(\frac{\mu^2}{2} - (\mu - \log a)\left(\frac{\mathfrak{e}}{b} - \frac{1}{2}(\mu - \log b)\right)\right). \qquad (22)$$

Proof: See Appendix F. ∎

Proposition 6: The conditional entropy $\mathcal{H}(X_\ell\|X_w)$ is given by

$$\mathcal{H}(X_\ell\|X_w) = b\log a - \log b + \mu(1-b) + \frac{e^{b\mu+b^2\sigma^2/2}}{a^2}. \qquad (23)$$

Proof: See Appendix G. ∎

From Propositions 5, 6, Equation (20) and by taking into account that the entropy of a LOGN random variable X_ℓ is

$$\mathcal{H}(X_\ell) = \frac{1}{2} + \mu + \frac{1}{2}\log(2\pi\sigma^2), \qquad (24)$$

we obtain:

Proposition 7: The symmetric KL divergence between an WBL and a LOGN random variables is given by:

$$\mathcal{K}(X_w, X_\ell) = \frac{e^{b\mu+b^2\sigma^2/2}}{a^2} + \frac{1}{2\sigma^2} \times$$
$$\left[\frac{\mathfrak{e}^2 + \frac{\pi^2}{6}}{b^2} + \mu^2 + (\mu - \log a)\left(\mu - \log b - \frac{2\mathfrak{e}}{b}\right)\right]$$
$$- b(\mu - \log a) + \frac{2\mathfrak{e}}{b} - \mathfrak{e} - \frac{3}{2} \qquad (25)$$

C. KL Divergence between a Generalized Gaussian Magnitude and a Log-normal distribution

Consider now a GG magnitude and a LOGN random variables X_g, X_ℓ with probability density functions $f_{\alpha,\beta}, f_{\mu,\sigma}$ respectively. The following provides cross entropies and KL divergences between these random variables.

Proposition 8: The conditional entropy $\mathcal{H}(X_g\|X_\ell)$ is given by

$$\mathcal{H}(X_g\|X_\ell) = \frac{1}{2}\left(\log(2\pi\sigma^2) + \frac{\mu^2}{\sigma^2}\right) + \left(1 - \frac{\mu}{\sigma^2}\right)\log\alpha$$
$$+ \frac{(\log\alpha)^2}{2\sigma^2} + \frac{1}{\beta}\psi\left(\frac{1}{\beta}\right)\left(1 - \frac{\mu}{\sigma^2} + \frac{\log\alpha}{\sigma^2}\right)$$
$$+ \frac{1}{\beta}\frac{\psi^{(2)}\left(\frac{1}{\beta}\right) + \left[\psi\left(\frac{1}{\beta}\right)\right]^2}{2\sigma^2\beta^2} \qquad (26)$$

where $\psi^{(2)}$ denotes the standard *Trigamma function* satysfying

$$\frac{d}{dz^2}\Gamma(z) = \Gamma(z)\left(\psi^{(2)}(z) + (\psi(z))^2\right). \qquad (27)$$

Proof: See Appendix H. ∎

Proposition 9: The conditional entropy $\mathcal{H}(X_\ell\|X_g)$ is given by

$$\mathcal{H}(X_\ell\|X_g) = \log\alpha - \log\beta + \Lambda\left(\frac{1}{\beta}\right) + \frac{1}{\alpha^\beta}e^{\mu\beta+\beta^2\sigma^2/2} \qquad (28)$$

where function Λ denotes the natural logarithm of the Gamma function.

Proof: See Appendix I. ∎

From Propositions 8, 9, Equations (19) and (24), we have:

Proposition 10: The symmetric KL divergence between LOGN and GG magnitude random variables is given by:

$$\mathcal{K}(X_g, X_\ell) = \frac{1}{\beta}\psi\left(\frac{1}{\beta}\right)\left(1 - \frac{\mu}{\sigma^2} + \frac{\log\alpha}{\sigma^2}\right) - \mu$$
$$+ \frac{\psi^{(2)}\left(\frac{1}{\beta}\right) + \left[\psi\left(\frac{1}{\beta}\right)\right]^2}{2\sigma^2\beta^2} + \frac{\mu^2}{2\sigma^2}$$
$$+ \log\alpha\left(1 - \frac{\mu}{\sigma^2} + \frac{\log\alpha}{2\sigma^2}\right) - \frac{1}{\beta}$$
$$+ \frac{1}{\alpha\beta}e^{\mu\beta + \beta^2\sigma^2/2} - \frac{1}{2}. \tag{29}$$

APPENDIX D
PROOF OF PROPOSITION 2

We have:

$$\mathcal{H}(X_w\|X_g) = -\int_{\mathbb{R}} f_{a,b}(x)\log f_{\alpha,\beta}(x)dx, \tag{30}$$

where $f_{\alpha,\beta}(x), f_{a,b}(x)$ are given by (7), (9) respectively.

From (7), (9) and (30), it follows that

$$\mathcal{H}(X_w\|X_g) = -\int_{\mathbb{R}^+} f_{a,b}(x)\left(\log\frac{\beta}{\alpha\Gamma(1/\beta)} - (|x|/\alpha)^\beta\right)dx$$
$$= \log\frac{\alpha\Gamma(1/\beta)}{\beta} + \int_{\mathbb{R}^+} f_{a,b}(x)(|x|/\alpha)^\beta dx. \tag{31}$$

By using integration by parts on the left hand side term of (31), we have

$$\int_{\mathbb{R}^+}\left(\frac{|x|}{\alpha}\right)^\beta f_{a,b}(x)dx = \int_{\mathbb{R}^+}\frac{b}{a}\left(\frac{x}{a}\right)^{b-1}\left(\frac{x}{\alpha}\right)^\beta e^{-(x/a)^b}dx$$
$$= \frac{\beta}{\alpha}\int_{\mathbb{R}^+}\left(\frac{x}{\alpha}\right)^{\beta-1}e^{-(x/a)^b}dx. \tag{32}$$

By applying a change of variable on the latter equality, we get

$$\int_{\mathbb{R}^+}\left(\frac{|x|}{\alpha}\right)^\beta f_{a,b}(x)dx = \frac{\beta}{b}\left(\frac{a}{\alpha}\right)^\beta \int_{\mathbb{R}^+} x^{\frac{\beta}{b}-1}e^{-x}dx. \tag{33}$$

Proposition 2 then follows from (31), (33) and by taking into account the definition of the Γ function (see (8)).

APPENDIX E
PROOF OF PROPOSITION 3

We have:

$$\mathcal{H}(X_g\|X_w) = -\int_{\mathbb{R}} f_{\alpha,\beta}(x)\log f_{a,b}(x)dx, \tag{34}$$

where $f_{\alpha,\beta}(x), f_{a,b}(x)$ are given by (7), (9) respectively.

By taking into account (7), (9) and (34), the conditional entropy $\mathcal{H}(X_g\|X_w)$ can be expressed in the following form

$$\mathcal{H}(X_g\|X_w) = \log\frac{a}{b} + (b-1)\log a$$
$$- \frac{\beta}{\alpha\Gamma(1/\beta)}\left((b-1)A_{\alpha,\beta} + \left(\frac{1}{a}\right)^b B_{\alpha,\beta,b}\right) \tag{35}$$

with

$$A_{\alpha,\beta} = \int_{\mathbb{R}^+} e^{-(x/\alpha)^\beta}\log x\, dx \tag{36}$$

and

$$B_{\alpha,\beta,b} = \int_{\mathbb{R}^+} x^b e^{-(x/\alpha)^\beta}dx. \tag{37}$$

We have:

$$A_{\alpha,\beta} = \frac{\alpha}{\beta}\int_{\mathbb{R}^+} e^{-x}x^{\frac{1}{\beta}-1}\log\left(\alpha x^{\frac{1}{\beta}}\right)dx,$$
$$= \frac{\alpha\log(\alpha)}{\beta}\int_{\mathbb{R}^+} e^{-x}x^{\frac{1}{\beta}-1}dx$$
$$+ \frac{\alpha}{\beta^2}\int_{\mathbb{R}^+} e^{-x}x^{\frac{1}{\beta}-1}\log x\, dx,$$
$$= \frac{\alpha}{\beta^2}\left(\Gamma\left(\frac{1}{\beta}\right)\log\alpha^\beta + \left(\frac{d}{dx}\Gamma(x)\right)_{/x=\frac{1}{\beta}}\right). \tag{38}$$

By noting taking into account (18), we obtain

$$A_{\alpha,\beta} = \frac{\alpha}{\beta^2}\left(\log\alpha^\beta + \psi(\frac{1}{\beta})\right)\Gamma\left(\frac{1}{\beta}\right). \tag{39}$$

In addition, from a change of variable $x \longrightarrow (x/\alpha)^\beta$, we derive:

$$B_{\alpha,\beta,b} = \frac{\alpha^{b+1}}{\beta}\Gamma\left(\frac{b+1}{\beta}\right). \tag{40}$$

Proposition 3 is then a consequence of (35), (36), (37), (39) and (40).

APPENDIX F
PROOF OF PROPOSITION 5

We have:

$$\mathcal{H}(X_w\|X_\ell) = -\int_{\mathbb{R}} f_{a,b}(x)\log f_{\mu,\sigma}(x)dx, \tag{41}$$

where $f_{\mu,\sigma}, f_{a,b}$ are given by (8), (9), respectively. The sketch of the proof is the following: first, decompose $\log f_{\mu,\sigma}$ in (41) as a logarithmic polynomial. Then, identify the combination of derivatives of the gamma function by expanding the integrand of (41).

Expressions (8), (18) involving the Gamma and Digamma functions, as well as the Expression (27) involving the Trigamma function are useful in deriving (22) of Proposition 5.

APPENDIX G
PROOF OF PROPOSITION 6

We have:

$$\mathcal{H}(X_\ell\|X_w) = -\int_{\mathbb{R}} f_{\mu,\sigma}(x)\log f_{a,b}(x)dx, \tag{42}$$

where $f_{\mu,\sigma}, f_{a,b}$ are given by (8), (9), respectively. The integrand in (42) can be expanded as a combination of Gaussian moments. This is performed *via* a change of variable $t = \log x$. Proposition 6 then follows after some straightforward calculus.

Appendix H
Proof of Proposition 8

We have:

$$\mathcal{H}(X_g \| X_\ell) = -\int_{\mathbb{R}} f_{\alpha,\beta}(x) \log f_{\mu,\sigma}(x) dx, \qquad (43)$$

where $f_{\alpha,\beta}, f_{\mu,\sigma}$ are given by (7), (8), respectively. The steps of the proof are similar to those given in Appendix F (proof of Proposition 5).

Appendix I
Proof of Proposition 9

We have:

$$\mathcal{H}(X_\ell \| X_g) = -\int_{\mathbb{R}} f_{\mu,\sigma}(x) \log f_{\alpha,\beta}(x) dx, \qquad (44)$$

where $f_{\alpha,\beta}, f_{\mu,\sigma}$ are given by (7), (8), respectively. The steps of the proof are those given in Appendix 6 relating the proof of Proposition 6.