



HAL
open science

A pattern-based mining system for exploring Displacement Field Time Series

Tuan Nguyen, Nicolas Méger, Christophe Rigotti, Catherine Pothier, Noel
Gourmelen, Emmanuel Trouvé

► **To cite this version:**

Tuan Nguyen, Nicolas Méger, Christophe Rigotti, Catherine Pothier, Noel Gourmelen, et al.. A pattern-based mining system for exploring Displacement Field Time Series. 19th IEEE International Conference on Data Mining (ICDM) Demo, Nov 2019, Beijing, China. pp.1110-1113, 10.1109/ICDMW.2019.00165 . hal-02361793

HAL Id: hal-02361793

<https://hal.science/hal-02361793v1>

Submitted on 13 Nov 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A pattern-based mining system for exploring Displacement Field Time Series ^{*}

Tuan Nguyen¹, Nicolas Méger¹, Christophe Rigotti², Catherine Pothier³
Noël Gourmelen⁴, and Emmanuel Trouvé¹

¹ Université Savoie Mont Blanc, Polytech Annecy-Chambéry, LISTIC Laboratory,
F-74944 Annecy-le-Vieux, France,

{hoang-viet-tuan.nguyen, nicolas.meger, emmanuel.trouve}@univ-smb.fr

² Université de Lyon, CNRS, INRIA, INSA-Lyon, LIRIS laboratory (UMR 5205),
20 avenue A. Einstein, F-69621 Villeurbanne Cedex, France,
christophe.rigotti@insa-lyon.fr

³ Université de Lyon, CNRS, INSA-Lyon, LIRIS laboratory (UMR 5205),
20 avenue A. Einstein, F-69621 Villeurbanne Cedex, France,
catherine.pothier@insa-lyon.fr

⁴ University of Edinburgh, Edinburgh EH89XP, United Kingdom,
noel.gourmelen@ed.ac.uk

Abstract. This paper presents the first available system for mining patterns from Displacement Field Time Series (DFTS) along with the confidence measures inherent to these series. It consists of four main modules for data preprocessing, pattern extraction, pattern ranking and pattern visualization. It is based on an efficient extraction of reliable grouped frequent sequential patterns and on swap randomization. It can be for example used to assess climate change impacts on glacier dynamics.

1 Introduction

Displacement Field Time Series (DFTS) analysis is an emerging research field fueled by the development of photogrammetry and remote sensing techniques, and by application needs of primary importance, such as landslide monitoring (e.g., [1, 2]) or glacier dynamics analysis (e.g., [3–6]). Basically, a DFTS covers a geographical zone and gives the surface displacements observed between acquisitions performed at different dates. These spatiotemporal and vectorial data generally come with confidence measures, each measure characterizing a single displacement vector. They express to which extent displacement vectors can be trusted, and are either obtained while computing displacements (e.g., correlation

^{*} Funding for this project was provided by a grant from la Région Auvergne-Rhône-Alpes (Tuan Nguyen’s grant). The work was also supported by a University Savoie Mont Blanc AAP Montagne grant - project BG *Big data pour la surveillance des Glaciers* and benefited from the PHOENIX ANR-15-CE23-0012 grant of the French National Agency of Research. Catherine Pothier and Christophe Rigotti are members of LabEx IMU (ANR-10-LABX-0088) that provided complementary support.

peaks or signal to noise ratio [7, 8]) or afterwards, by assessing the displacements using their spatial and/or temporal distributions (e.g., [9, 5]).

These raw, large and complex data sets are usually processed by filtering out low confidence data points, and then plotting temporal displacement profiles established for a few transects that are chosen manually, before studying these profiles visually [5]. Data mining methods are thus needed to fully exploit such data and complete existing knowledge. A first proposal has been recently published in [10]. It is based on the extraction of Grouped Frequent Sequential pattern (GFS-patterns) as defined in [11], for which a notion of pattern reliability is proposed. Such reliable patterns are the ones for which the confidence measures, at the occurrence level, are sufficiently high on average. Once extracted, patterns are ranked according to the swap randomization method described in [11]. This paper presents the corresponding system, *DFTS-P2miner* (DFTS-Pattern maP miner), that implements the method of [10]. In addition to the pattern extraction and ranking modules, it also comprises dedicated preprocessing and visualization tools. The salient features of the resulting system, with respect to other state-of-the-art methods, are: its ability (1) to automatically explore the whole data set without being limited to a manual study of transects; and (2) to handle the reliability of the patterns based on their occurrence quality.

In a DFTS, the observed area is represented as a grid of pixels, each pixel containing the sequence of displacement vectors (magnitudes and directions) computed over time for a given location, and each vector coming along with its respective confidence measure. Our system takes as inputs either the magnitude or the direction of the vectors as well as the associated confidence measures. The latter are left unchanged while the former are quantized in a preprocessing step to replace them by symbols (e.g., ‘1’, ‘2’, ‘3’, ...) denoting magnitude levels or direction quadrants. This symbolic DFTS is then mined to extract GFS-patterns, where a GFS-pattern is a frequent sequential pattern [12] forming homogeneous zones in space: if the pattern occurs in the sequence associated to a pixel, then it must also tend to occur in the spatial neighborhood of this pixel (but possibly with a shift in time). In addition to these frequency and grouping constraints, the patterns are also required to be reliable as defined in [10]. More precisely, for a given pattern, the confidence level of each occurrence is obtained by retaining the minimum confidence value observed for the data points forming the occurrence. Then, a pattern is considered as reliable if the average of the confidence level of its occurrences exceeds a user-defined threshold. Consequently, instead of filtering out low confidence data points in a prior step and extracting patterns from the remaining confident data points only, all data points are considered and patterns are retained as long as their occurrences are confident on average.

2 System Description

DFTS-P2miner takes as input either the magnitude or the orientation of a DFTS as well as the corresponding confidence measures. Each displacement field

is supplied using raw (e.g., 32-bit floating numbers written in a raster mode) or standard (e.g., tiff or jpeg) image formats. Confidence measures are given separately using a raw image format. The DFTS is first quantized by the *preprocessing* module, using one of the different available discretization strategies, to produce a *symbolic DFTS* such that each symbol represents a magnitude level or a direction quadrant, and comes with its respective confidence measure. This symbolic DFTS is then processed by the *pattern extraction* module to extract *reliable GFS-patterns*. Following the strategy proposed in [10], efficiency is achieved by pruning the search space according to the pattern-growth general approach introduced in [13]. Along with the standard anti-monotonic frequency constraint and the grouping constraint, the reliability constraint is exploited using a partial pushing technique based on an upper bound of the reliability measure. More details about this pruning strategy can be found in [10]. In addition, dynamic programming is utilized to locate patterns occurrences and compute reliability measures. Once extracted, the patterns are ranked using a Normalized Mutual Information (NMI) measure indicating whether pattern occurrences are likely to be found or not in a swap randomized data set. Each pattern being characterized by a single NMI measure that is established individually, parallelization is thus performed to lower execution times. Finally, different *maps*, called *SpatioTemporal Localization maps (STL-maps)*, depicting the location in space and time of the occurrences of the top-ranked patterns are computed by the *visualization* module that is also run in parallel mode. These maps can be browsed with a dedicated Graphical User Interface (GUI), illustrated Fig. 1. This interface also gives access to complementary visualization modes and statistics about the pattern occurrence dates (histograms, median values, etc.). The reader is referred to [10] and [11] for the complete definition of the patterns and their maps, the description of the extraction/ranking steps and the guidelines for parameter settings.

The whole DFTS mining process is controlled using a single parameter file gathering information about the input/output directories and the extraction/ranking parameters. This file is interpreted by a Python script that chains and executes all necessary steps. The most resource consuming steps, i.e., the extraction, the swap randomization and the map generation are implemented in C. The results are stored in folders that are structured according to the processing steps, the parameter values and the execution dates. They include patterns, intermediate ranking information, maps as well as monitoring log files that are organized for quick result browsing and easy iterative mining. The system can be run using a standard OS (Linux or Mac OS X) and a standard computing platform (e.g., 2.5 GHz Intel Core i7, 16 GB memory).

3 Demonstration

The magnitudes and the confidence measures of a real DFTS will be processed with *DFTS-P2miner* on a standard laptop for the presentation. This DFTS is

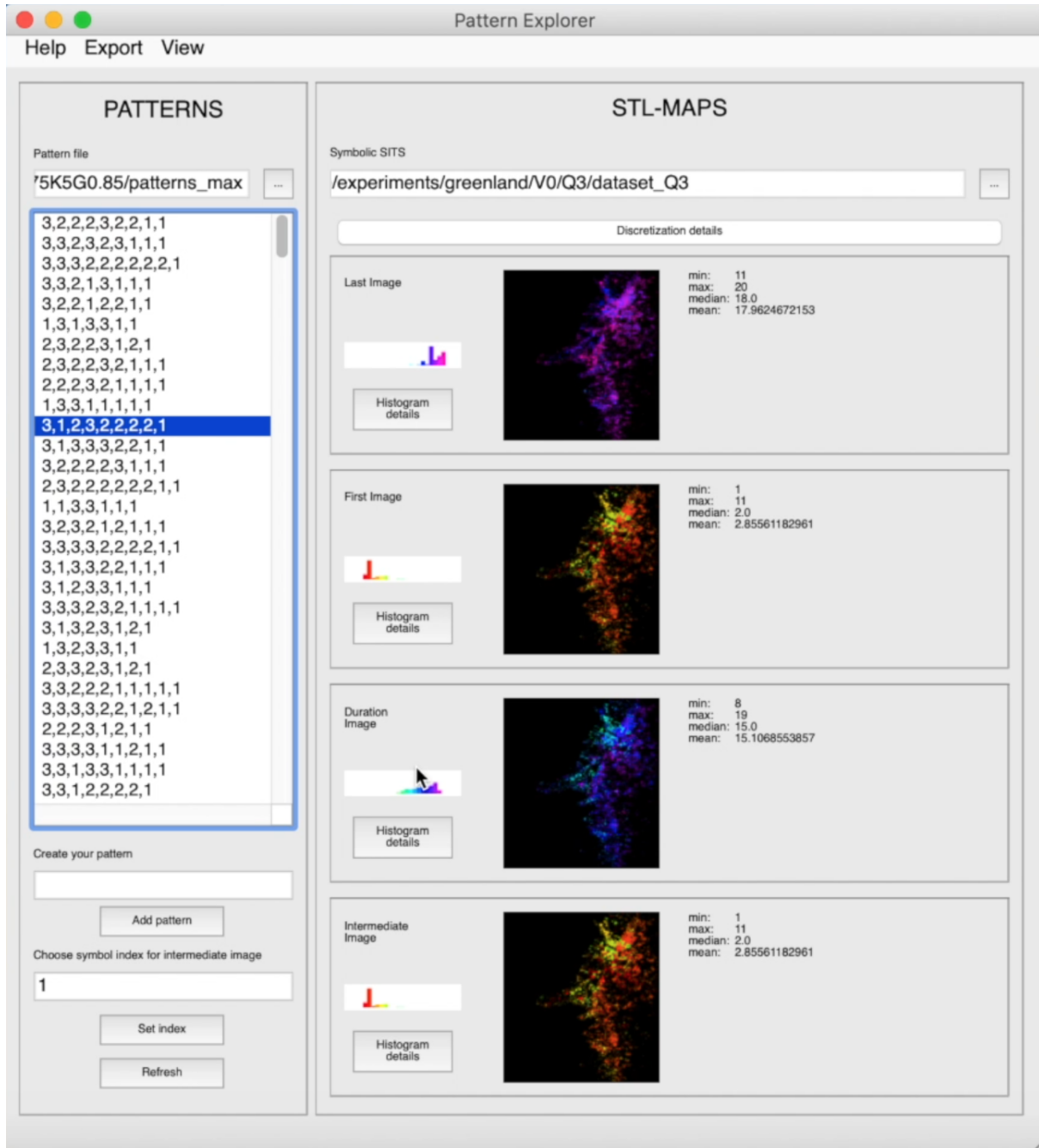


Fig. 1. The graphical user interface for exploring patterns and their maps. The color scale is given by Fig. 2.



Fig. 2. STL-map color scale: 20 timestamps, from 1985 in red to 2013 in magenta.

used in [5] to exhibit a slowdown in the ablation zone of the western Greenland Ice Sheet. It contains twenty displacement fields of size 458×500 , built using Landsat images (optical satellite images) acquired between 1985 and 2013 with a ground resolution of 240×240 m per pixel.

Two of the best ranked maps of reliable GFS-patterns are shown in figures 3 and 4, drawn on a grayscale background image. The blue and pink pixels denote the locations of the occurrences in space and in time (middle of the DFTS in blue and end of the DFTS in pink). The complete color scale is given by Fig. 2.

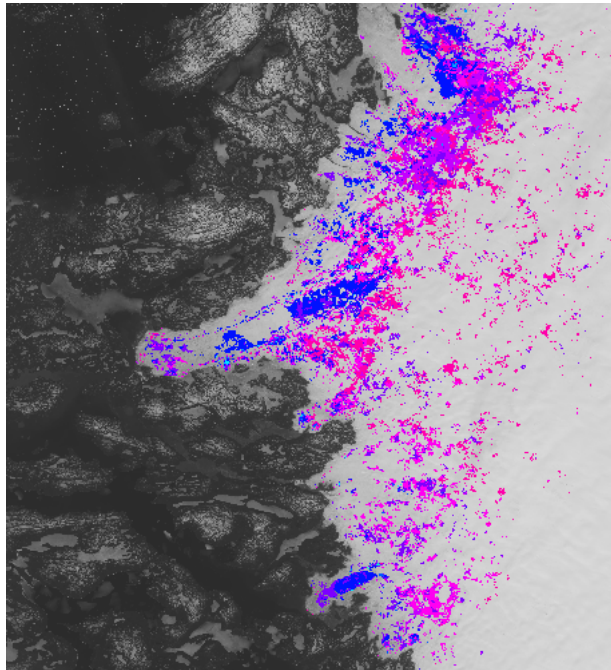


Fig. 3. Map of pattern 3,3,3,3,2,2,2,2,2,1.

The map of Fig. 3 shows the locations of pattern 3,3,3,3,2,2,2,2,2,1. It expresses displacement magnitudes falling from the highest magnitude level (3) to the lowest one (1). It corresponds to the decadal slowdown reported in [5] and also complete the current knowledge available for the area, by giving the whole spatial extent of the phenomenon, while previous studies have been made along

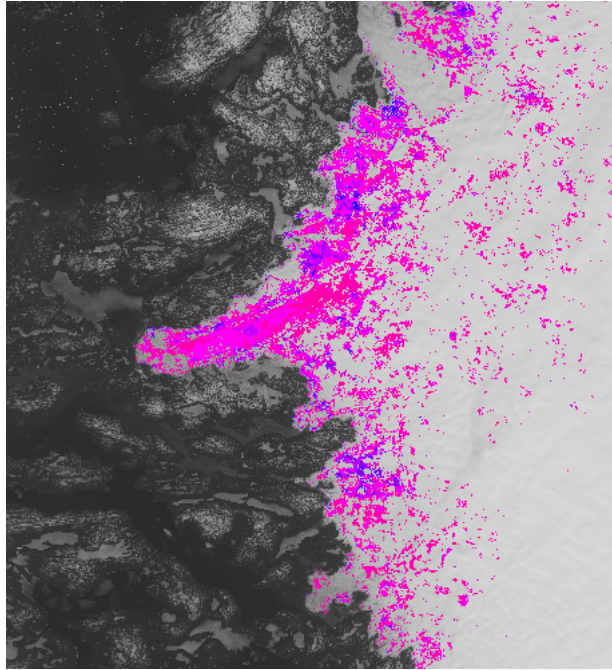


Fig. 4. Map of pattern 3,3,2,3,2,1,1,1,1.

transects only. The map of another top-ranked pattern is given Fig. 4. It corresponds to pattern 3,3,2,3,2,1,1,1,1,1, that unveils a moderate slowdown (from level 3 to level 2) followed by a stronger one (from level 3 to level 2 and finally 1). As it can be observed, these two patterns tend to hold in rather complementary parts of the area.

Another application dealing with short time displacement evolutions over Alpine glaciers (Argentière, Mer de Glace, Bossons and Tacconnaz glaciers) in the Mont Blanc massif will also be demoed during the conference. It is based on a DFTS containing twenty-five 3494×3186 displacement fields built using TerraSAR-X images (radar satellite images) acquired between 2009 and 2011. Its ground resolution is about $6 \text{ m} \times 6 \text{ m}$ per pixel.

The demonstration features a real mining scenario showing the interactive capability of the GUI. This includes: 1) refining the exploration by locating in time the different elements forming the patterns and by computing statistics for sub-areas, 2) trying out user-defined variant patterns and 3) exporting exploration materials (STL-maps, statistics).

The web page <https://sites.google.com/view/dfts-p2miner> gives public access to the whole system and a link to download the Mont Blanc DFTS.

Acknowledgments

We thank the German Aerospace Agency (DLR) for the TerraSAR-X images (project MTH0232) as well as the National Aeronautics and Space Administration (NASA) and the United States Geological Survey (USGS) for the Landsat images.

References

1. J. Travelletti, C. Delacourt, P. Allemand, J.-P. Malet, J. Schmittbuhl, R. Toussaint, and M. Bastard, "Correlation of multi-temporal ground-based optical images for landslide monitoring: Application, potential and limitations," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 70, pp. 39 – 55, 2012.
2. S. Liu, H. Wang, J. Huang, and L. Wu, "High-resolution remote sensing image-based extensive deformation-induced landslide displacement field monitoring method," *International Journal of Coal Science & Technology*, vol. 2, no. 3, pp. 170–177, Sep 2015.
3. V. Akbari, A. P. Doulgeris, and T. Eltoft, "Monitoring glacier changes using multitemporal multipolarization SAR images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 6, pp. 3729–3741, June 2014.
4. F. Ponton, E. Trouvé, M. Gay, A. Walpersdorf, R. Fallourd, J.-M. Nicolas, F. Vernier, and J.-L. Mugnier, "Observation of the Argentire Glacier Flow Variability from 2009 to 2011 by TerraSAR-X and GPS Displacement Measurements," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 8, pp. 3274–3284, Aug. 2014.

5. A. J. Tedstone, P. W. Nienow, N. Gourmelen, A. Dehecq, D. Goldberg, and E. Hanna, “Decadal slowdown of a land-terminating sector of the Greenland Ice Sheet despite warming,” *Nature*, vol. 526, no. 7575, pp. 692–695, Oct. 2015.
6. M. Fahnestock, T. Scambos, T. Moon, A. Gardner, T. Haran, and M. Klinger, “Rapid large-area mapping of ice flow using Landsat 8,” *Remote Sensing of Environment*, vol. 185, pp. 84 – 94, 2016.
7. T. Strozzi, A. Kouraev, A. Wiesmann, U. Wegmller, A. Sharov, and C. Werner, “Estimation of Arctic glacier motion with satellite L-band SAR data,” *Remote Sensing of Environment*, vol. 112, no. 3, pp. 636–645, Mar. 2008. [Online]. Available: <http://linkinghub.elsevier.com/retrieve/pii/S0034425707002271>
8. M. Debella-Gilo and A. Kb, “Locally adaptive template sizes for matching repeat images of mass movements,” in *Geoscience and Remote Sensing Symposium (IGARSS), 2011 IEEE International*. IEEE, 2011, pp. 4281–4284. [Online]. Available: <http://ieeexplore.ieee.org/abstract/document/6050177/>
9. A. Dehecq, N. Gourmelen, and E. Trouvé, “Deriving large-scale glacier velocities from a complete satellite archive: Application to the Pamir-Karakoram-Himalaya,” *Remote Sensing of Environment*, vol. 162, pp. 55–66, 2015. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0034425715000553>
10. T. Nguyen, N. Méger, C. Rigotti, C. Pothier, E. Trouvé, N. Gourmelen, and J.-L. Mugnier, “A pattern-based method for handling confidence measures while mining satellite displacement field time series. Application to Greenland ice sheet and Alpine glaciers,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 11, pp. 4390 – 4402, Oct. 2018.
11. N. Méger, C. Rigotti, C. Pothier, T. Nguyen, F. Lodge, L. Gueguen, R. Andréoli, M.-P. Doin, and M. Datcu, “Ranking evolution maps for Satellite Image Time Series exploration: application to crustal deformation and environmental monitoring,” *Data Mining and Knowledge Discovery*, vol. 33, no. 1, pp. 131–167, Jan. 2019.
12. R. Agrawal and R. Srikant, “Mining sequential patterns,” in *Proc. of Conf. ICDE*, 1995, pp. 3–14.
13. J. Pei, J. Han, and W. Wang, “Constraint-based sequential pattern mining: the pattern-growth methods,” *Journal of Intelligent Information Systems*, vol. 28, no. 2, pp. 133–160, 2007. [Online]. Available: <http://dx.doi.org/10.1007/s10844-006-0006-z>