



HAL
open science

An Interactive Platform for Extracting Gradual Patterns from Multivariate Temporal Data

Jerry Lonlac, Yannick Miras, Vincent Mazenod, Engelbert Mephu Nguifo

► **To cite this version:**

Jerry Lonlac, Yannick Miras, Vincent Mazenod, Engelbert Mephu Nguifo. An Interactive Platform for Extracting Gradual Patterns from Multivariate Temporal Data. 2021. hal-03986063v2

HAL Id: hal-03986063

<https://hal.science/hal-03986063v2>

Preprint submitted on 19 Apr 2023

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.

MobiPaleo: An Interactive Platform for Extracting Gradual Patterns from Multivariate Temporal Data

Jerry Lonlac

IMT Nord Europe, Institut Mines-Télécom, Univ. of Lille
Centre for Digital Systems, F-59000 Lille, France
jerry.lonlac@imt-lille-douai.fr

Vincent Mazenod

Univ. Clermont Auvergne, CNRS, Mines de Saint-Etienne
Clermont Auvergne INP, LIMOS, Clermont-Ferrand
vincent.mazenod@uca.fr

Yannick Miras

CNRS, UMR 7194 HNHP
Institut de Paléontologie Humaine, Paris, France
yannick.miras@mnhn.fr

Engelbert Mephu Nguifo

Univ. Clermont Auvergne, CNRS, Mines de Saint-Etienne
Clermont Auvergne INP, LIMOS, Clermont-Ferrand
engelbert.mephu_nguifo@uca.fr

Abstract

This paper presents a tool namely *Mobipaleo*¹ to automatically extract simultaneous attribute co-variations called gradual patterns from multivariate temporal data. It proposes two types of temporal resolution to integrate in the search process according to user preferences: the high temporal resolution taken into account referenced as the temporal constraint between consecutive objects and the low temporal resolution referenced as the temporal constraint between distant objects. *Mobipaleo* highlights the benefits of using artificial intelligence techniques such as pattern mining in real-world applications.

1 Introduction

Gradual patterns model complex attributes co-variations of the form "the more or less is X_1, \dots , the more or less is X_n " in a numerical database. These last years, gradual patterns attracted a lot of attention and many algorithms have been proposed to automatically extract such patterns from different data models [1–3, 6, 7]. Nowadays, gradual patterns are largely applied for extracting knowledge from data in several domains, including for example in *financial market* for discovering co-evolution between financial indicators, in biology for discovering correlations between genomic expressions. Despite of the contribution of gradual patterns for understanding the data from their ability for interpretation and summary large amounts of data, there is nowadays no available tool for interactive extraction of gradual patterns. Note that, few extracting gradual patterns software are available [5], however, they do not process temporally annotated numerical data. Moreover, they are not user-friendly wrt mining process by specifying some mining constraints and analyzing patterns. The number of those patterns can be very large, making difficult for a user to interpret them.

MobiPaleo is an online tool for extracting gradual patterns, and it deals with temporal data and allows users to specify the mining constraints, visualize and easily navigate through the extracted patterns. It offers a customizable user friendly

interface to easily explore and filter on the extracted patterns, it enables users to identify correlations between attributes most useful for their usecase. Our platform offers a suitable framework to identify correlations from temporal data while leave user free to specify temporal resolution constraints to be integrated into the mining process. It is used by paleoecological experts to analyze biodiversity trajectories, climate and ecosystems dynamics from the extracting ecological correlations of groups of key-proxy multi-variate indicators. Our tool highlights the use of pattern mining based methods for finding patterns that provide valuable knowledge and that are interpretable by humans.

Gradual patterns. A gradual item has form i^* : where i is an attribute and $*$ is a variation such that $*$ \in $\{\uparrow, \downarrow\}$. " \uparrow " denotes an increasing variation and " \downarrow " a decreasing variation. A gradual pattern is a set of gradual items $\{i_j^*, j = 1 \dots k\}$. A gradual pattern captures the simultaneous co-variation of several attributes. For instance $\{i_1 \uparrow, i_2 \downarrow\}$ is interpreted as the more i_1 and the less i_2 . The quality of a gradual pattern is measured by its frequency support [5]. The algorithms within *MobiPaleo* consider the definitions of frequency proposed by [2, 3] as they were designed for temporal data.

Section 2 presents an application where there are temporal data with the challenge of gradual patterns mining which can be addressed using our tool. Section 3 gives an overview of *MobiPaleo*, describes it and the temporal constraint-based gradual patterns mining algorithms running in the background which are proposed by [2, 3]. Note that this demonstration paper highlights the works published in [2, 3] on the extracting of gradual patterns from temporal data, with an emphasis on the software and applications. Therefore, for details on mining gradual patterns in temporal data, we refer the reader to the more detailed papers [2, 3] which include context and formal definitions.

2 Application domain in Paleocology

Paleoecological research enables to rebuild over time the ecological dynamics and the evolution of the biodiversity (for example the evolution of vegetation or the functioning

¹<https://mobipaleo.limos.fr/login>

of a lake ecosystem) under the influence of climatic variations and human activities (e.g. agriculture and pastoralism) [8]. The reconstruction of these ecological trajectories over 7 millennia for Lake Aydat, located in the Chaîne of Puys in the Auvergne Rhône-Alpes region and threatened with eutrophication, makes it possible to better understand its current ecological status and contributes to the establishment of viable governance [4]. This research is based on the abundance of different paleoecological indicators (pollen grains and plant spores, non-pollen micro-fossils: various forms of resistance of phytoplankton and zooplankton, diatoms) conserved in the lacustrine sediment records. All these paleoecological data are then stored in large databases with temporal relationships between components. These data are complex, since they are heterogeneous and evolving, consists variables of different studies.

Indeed, the paleoecological data include a set of numerical attributes for which values correspond to the quantity of each paleoecological indicator contained in a sediment record taken, by coring operations, in a lake ecosystem. The sedimentary sequence obtained is then dated, sampled, and for each sample, at a given depth, a date is calculated. The abundance of each indicator is then recorded for each sample. The database objects correspond to different dates obtained on the considered sedimentary record, and the columns to the different paleoecological recorded.

More formally, let \mathcal{D} be the dates of the different samples of the sedimentary sequence, \mathcal{I} the different paleoecological indicators recorded at these dates \mathcal{D} , then the paleoecological data is defined by $\Delta = \mathcal{D} \times \mathcal{I}$. The following notations characterize the data contained in Δ . Let $d \in \mathcal{D}$ and $i \in \mathcal{I}$:

- $\Delta(d, i)$ indicates the number of instances of the indicator i present at the date d ;
- $\Delta(d, i) = 0$ indicates the absence of the indicator i at the date d ;
- $\Delta(\mathcal{D}, i)$ shows the temporal evolution of i ;
- $\Delta(d, \mathcal{I})$ characterizes the state of palaeoecological conditions at date d

A peculiarity of the paleoecological data contained in Δ is that they are evolutionary. Moreover, Table 1 generally contains few lines (the dates of the samples) in relation to the number of columns (the number of paleoecological indicator) and is very sparse (it contains a large number of zero values).

Table 1 is a sample of paleoecological data, its columns are labeled with scientific names of different indicators.

Table 1. Example of paleoecological database Δ .

Date	Fabaceae	Secale	Rumex	Equisetum	Plantago	Filipendula	Vulgaris
d_1	94	46	11	3	8	4	0
d_2	126	21	9	4	18	4	31
d_3	100	37	7	6	12	4	13
d_4	134	19	6	8	19	3	36
d_5	112	34	5	9	14	2	17
d_6	145	2	5	4	25	2	62

It is then a challenge for the paleoecological research experts, to analyze the database in order to extract specific knowledge, such as the discovery of coevolution groupings of multi-varied paleoecological indicators (eg formed groupings of pollen grains, micro-fossils non pollen and diatoms) necessary to understand the evolution of biodiversity, ecosystem operation over time and to predict future change. All these implicit information are generally sought by the paleoenvironmental experts using conventional statistical analysis methods. These methods are often based on a simple plot of a graph containing the curves of evolution of different indicators from the paleoecological data (for more information, see the website <http://mobipaleo.univ-bpclermont.fr/index.html>), and on an empirical comparison of these curves in order to identify groups of curves that evolve at identical periods.

It appears obvious that it is very tedious for paleoecology experts to empirically identify the different co-evolutions of these indicators from this plot. Moreover, the risk is to exclude an exhaustive processing of the dates and eventually to miss the small relevant information: groups with low co-evolution but with a strong impact on the biodiversity change. In this context, the gradual patterns are proven suitable to solve the extraction problem of co-evolution groupings of multivariate paleoecological indicators.

In this paper, we propose an interactive platform that makes it possible to extract frequent complex co-variations between numerical attributes in a data table by taking into account the temporal constraints between objects during the mining process. To the best of our knowledge, *MobiPaleo* is the first interactive platform for extracting gradual patterns from temporal data, that offers more flexibility to user for specifying temporal resolution constraint while providing it an interface to explore itself the generated patterns. In terms of applications, this platform provides additional information compared to the traditional techniques used to date in the analysis of paleoecological data, which are essentially statistical methods such as principal component analysis or clustering. It facilitates the analysis of multivariate bio-indicators to an examination of long-term ecosystem dynamics and seeks to detect early anthropogenic impacts.

3 MobiPaleo Platform

This section presents the architecture of *MobiPaleo* and details how use it for exploring multivariate temporal data.

3.1 Architecture Overview

MobiPaleo platform as shown in Figure 1, consists of the three main modules. (1) The Data load module retrieves the data and parameters of gradual patterns mining algorithms from a user interface as shown in Figure 2; (2) The gradual patterns extraction module selects and runs the appropriate

algorithm according to the level of temporal resolution chosen by the user; (3) The Post-processing module shows the gradual patterns found by the selected algorithm and helps to navigate through them as shown in Figure 3.

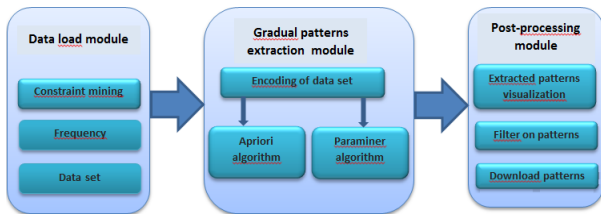


Figure 1. Architecture of *MobiPaleo*

More precisely, the Data load module allows to upload a multivariate temporal data set, to define a minimum support threshold and to specify the temporal resolution level according to which a gradual pattern mining algorithm will be selected. The gradual pattern extraction module is the core part of *MobiPaleo* tool, which runs the selected algorithm and returns the set of gradual patterns found. The Post-processing module displays the generated patterns and allows users to select particular variables to inspect the patterns containing those variables in a user-friendly way.

3.2 User Interface and Demonstration Scenario

This section shows how to perform gradual patterns mining tasks under temporal constraints on *MobiPaleo* platform and use available interface to rapidly navigate through extracted patterns. Figure 2 shows the interface for submitting a gradual patterns mining task on *MobiPaleo*. User should provide some parameters among which the mining constraint that is set according to the temporal resolution preferences of user. There are two mining constraints: 1) high temporal resolution where gradualities are only considered between consecutive objects during the mining process and 2) low temporal resolution where gradualities may be considered between distant objects. The former constraint is particularly suitable to extract paleoecological indicators coevolutions in order to understand paleoecological dynamics over time since palaeoecological research is usually undertaken at high temporal resolution. Hence, *MobiPaleo* helps palaeoecological experts to automatically identify ecological correlations of groups of key-proxies and specific multivariate bio-indicators to explore their potential as early warning signals of tipping points in the lake dynamics. This highlights the application of pattern mining for multi-proxy paleoecological research which usually provides large and heterogeneous data with temporal relationships between data.

Before submitting a mining task, the user uploads a dataset, defines the constraints (Constraint mining) on the desired patterns and specifies a minimum support threshold (frequency). Once running terminates, the user can explore and

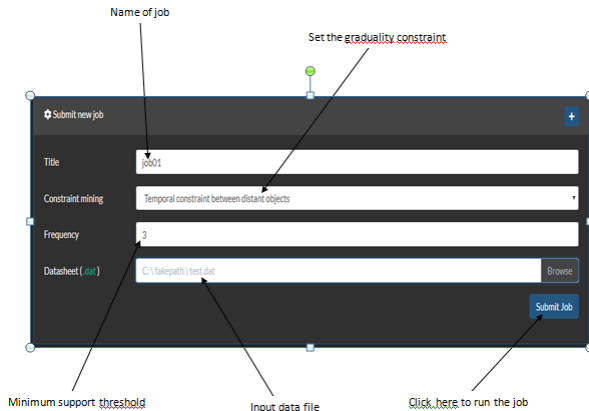


Figure 2. Interface for submitting a task on *MobiPaleo*

post-process extracted patterns like illustrated by Figure 3. On this user interface, user can manually choose a set of variables on the right side to view how they evolve over time. This allows users thru their domain-specific knowledge to filter on desired patterns. Users can also download extracted patterns in a csv file by clicking on *save as csv* button. To use *MobiPaleo* platform, user must first register.

4 Apply *MobiPaleo* on Paleoecological Data

This section shows how *MobiPaleo* can be useful to find interesting hidden patterns describing the co-evolution relationships between paleoecological indicators. We applied *MobiPaleo* on a paleoecological database set taken from Lake Aydat located in the southern part of the Chane des Puys volcanic range (French Massif Central). This database contains 111 objects (the temporal component, be in depth or estimated radiocarbon age) described by a set of 87 attributes (pollen, non-pollen palynomorphs and diatom-types that are paleoecological indicators). It gives the abundance (in percentages) of each attribute for each object (taken at the same high resolution throughout the cores).

It is noteworthy that significant cross-correlations of different indicators and their repeated co-evolutions through time are not easy to characterize empirically by the palaeoecologist. When running *MobiPaleo* on the paleoecological data, we extract some interesting gradual patterns correspond to the groups of co-evolutions of different palaeoecological indicators with their depths of co-evolution (Tids in Figure 3). The analysis of the recurrence of these patterns through time traduces specific patterns, allowing ecological modelling of systems dynamics. Some interesting gradual patterns were characterized (see Figure 3) comprising different types of palaeoecological indicators and consistent co-evolution of abundance of taxa (expressed in percentages).

Evaluation of the quality of extracted patterns. Some co-evolutions are principally associated with human impact, and especially arable agriculture, as the associated attributes

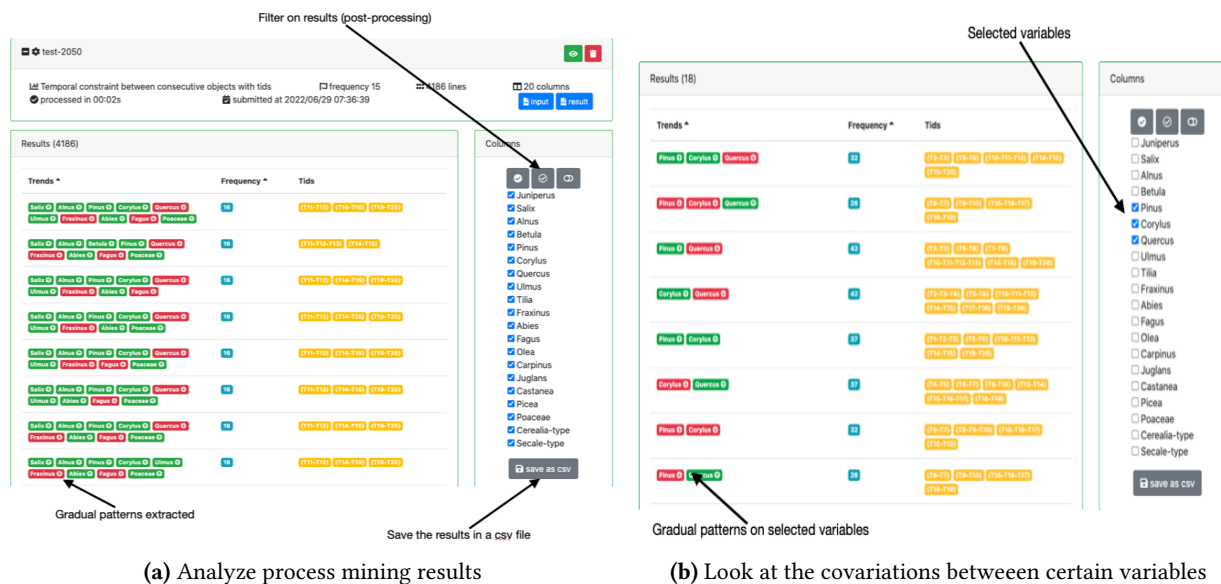


Figure 3. Interface for post-processing of extracted gradual patterns

(pollen taxa) are considered as traditional anthropogenic pollen indicators. Other co-evolutions are associated with human activities, most likely pastoral agriculture. A detailed analysis of the distribution through time (Tids in Figure 3) of the gradual patterns indicative of water trophic status shows an accentuated persistence of them between the earliest ages. *MobiPaleo* highlights ecological associations of groups of multivariate indicators which allowing to obtain a coherent narrative of cause and effect and identify feedbacks over time in ecosystem dynamics, better discriminate between the drivers involved.

5 Conclusion and Future Work

MobiPaleo is to the best of our knowledge, the first interactive tool for automatically extracting attribute co-variations under temporal constraints which furthermore, offers a suitable framework to user to visualize and navigate through the extracted patterns. This platform is very helpful to domain expert in paleoecological research for identifying ecological correlations of groups of key-proxies and specific multivariate bio-indicators to explore their potential. We plan to extend *MobiPaleo* with more parameters to address more complex post-processing of extracted patterns such as, for example, the visualization of the curves of variables.

References

- [1] Jerry Lonlac, Arnaud Doniec, Marin Lujak, and Stéphane Lecoecueche. 2020. Mining Frequent Seasonal Gradual Patterns. In *DaWaK*, Vol. 12393. 197–207.
- [2] Jerry Lonlac, Yannick Miras, Aude Beauger, Vincent Mazenod, Jean-Luc Peiry, and Engelbert Mephu Nguifo. 2018. An Approach for Extracting Frequent (Closed) Gradual Patterns Under Temporal Constraint. In *FUZZ-IEEE*. 878–885.
- [3] Jerry Lonlac and Engelbert Mephu Nguifo. 2020. A novel algorithm for searching frequent gradual patterns from an ordered data set. *Intell. Data Anal.* 24, 5 (2020), 1029–1042.
- [4] Y. Miras, A. Beauger, M. Lavrieux, V. Berthon, K. Serieysson, V. Andrieu-Ponel, and P. Ledger. 2015. Tracking long-term human impacts on landscape, vegetal biodiversity and water quality in the lake Aydat (France) using pollen, non-pollen palynomorphs and diatom assemblages. In *Palaeogeography, Palaeoclimatology, Palaeoecology*. 76–90.
- [5] Benjamin Nègrevergne, Alexandre Termier, Marie-Christine Rousset, and Jean-François Méhaut. 2014. ParaMiner: a generic pattern mining algorithm for multi-core architectures. *DMKD* 28, 3 (2014), 593–633.
- [6] Dickson Owuor, Anne Laurent, and Joseph Orero. 2019. Mining Fuzzy-Temporal Gradual Patterns. In *FUZZ-IEEE*. 1–6.
- [7] NhatHai Phan, Dino Ienco, Donato Malerba, Pascal Poncelet, and Maguelonne Teisseire. 2015. Mining Multi-Relational Gradual Patterns. In *SDM*. 846–854.
- [8] H.J.B. Smol, H.J.B. Birks, and W.M. Last. 2001. Tracking Environmental Change Using Lake Sediments. In *Terrestrial, Algal and Silicaceous Indicators III*. Kluwer Academic Publishers, Dordrecht. 319–349.