A process mining approach to the analysis of the structure of time series

Valdés, Julio J.; Céspedes-González, Yaimara; Tapping, Kenneth; Molero-Castillo, Guillermo

Publisher's version / Version de l'éditeur:
https://doi.org/10.1007/978-3-030-63089-8_25
Proceedings of the Future Technologies Conference (FTC) 2020, Volume 2, Advances in Intelligent Systems and Computing; Volume 1289, pp. 379-392, 2020-11-01

A Process Mining approach to the analysis of the structure of time series

Julio J. Valdés[1], Yaimara Céspedes-González[2], Kenneth Tapping[3], and Guillermo Molero-Castillo[4]

[1] National Research Council Canada
Digital Technologies Research Centre
M50, 1200 Montreal Rd. Ottawa K1A0R6, Canada
julio.valdes@nrc-cnrc.gc.ca
[2] Faculty of Accounting and Administration
Universidad Veracruzana
Veracruz, Mexico
zs16017485@estudiantes.uv.mx
[3] Herzberg Institute for Astrophysics
National Research Council Canada
Victoria, Canada
ken.tapping@nrc-cnrc.gc.ca
[4] Engineering Faculty
Universidad Nacional Autónoma de México
Mexico City, Mexico
guillemolero@comunidad.unam.mx

Abstract. This paper presents a discussion of the potential of Process Mining for the analysis of general processes involving the time variation of magnitudes given by real-valued variables. These scenarios are common in a broad variety of domains, like natural and life sciences, engineering, and many others beyond business processes, where in general complex systems are observed and monitored using sensor data, producing time series information. Two approaches are presented to construct event logs for such types of problems and one of them is applied to a real world case (monitoring the F10.7 cm electromagnetic flux produced by the Sun). The results obtained with the Fuzzy Miner and the Multi-Objective Evolutionary Tree Miner algorithms successfully exposed the differences in the internal structure of the F10.7 cm series between Solar cycles. For this application, Process Mining proved to be a valuable tool for analyzing the rhythm of solar activity and how it is changing. The approach introduces here is general and could be use in the analysis of data from a broad variety of time-dependent information from many domains.

Keywords: process mining, machine learning, graph an trace clustering, fuzzy models, evolutionary multi-objective models

1 Introduction

Process mining (PM) is a set of techniques originally developed within the process management domain, mainly oriented to model, analyze and optimize business processes. It is an already well established research discipline that combines machine learning and data mining with process modeling and analysis [2].

PM methods work with data consisting of event logs. Typically, an event log consists of a set of instances of a process, where each instance consists of ordered events. The event log has properties such as activity and time as well as additional ones like resource or cost. The following is an example of an event log.

Case ID	Event ID	dd-MM-yyyy:HH.mm	Activity	Resource	Costs
1	35654423	30-12-2010:11.02	register request	Pete	50
1	35654424	31-12-2010:10.06	examine thoroughly	Sue	400
1	35654425	05-01-2011:15.12	check ticket	Mike	100
1	35654426	06-01-2011:11.18	decide	Sara	200
...	...	...	...	...	...

In particular, there are certain *activities* that are performed at certain *times*, which are performed by certain subjects (*resources*). Besides these key elements, an event log may provide additional information about the process (e.g. cost). Thus, one of the most interesting possibilities offered by process mining is to discover the process models from event logs, which guarantees that the discovered model describes the actual behavior recorded [2], [3]. Models are suitable representations because they allow communicating complex knowledge in a more intuitive way [11].

Process mining has become a very useful tool for the analysis of systems of events and has been used very successfully in domains like business, management, health and social studies, among others. However, despite its great potential, applications in non-business domains, like natural and life sciences, engineering and many others are comparatively fewer [9]. PM algorithms cover a wide variety of problems and approaches [1], [4], [8], [12], [13].

The purpose of this paper is to start filling this gap by introducing process mining to the analysis of time series from processes in any domain producing continuous magnitudes (e.g. sensor data). The objective is to uncover patterns of behavior within the dynamics of a general system and to characterize the changes associated to these patterns using process mining.

The paper is organized as follows, Section 2 presents process mining, discusses its applications to the study of natural processes and presents approaches for constructing process logs from time series of real-valued magnitudes. Section 3 describes two process mining techniques used in the example application (Fuzzy Miner and the Evolutionary Multi-Objective Tree Miner). Section 4 presents the application example (monitoring the Sun's F10.7 cm flux) and the results obtained with the process mining modeling, and Section 5 summarizes the paper.

2 Process Mining

The aim of process mining is to discover, monitor and improve processes by extracting knowledge from event logs. As described above, these are collections of cases containing sequences of certain events. Typical scenarios described by sequences of events are database systems, transaction logs in a trading system, message logs, among others.

Process mining provides a set of techniques that allows the analysis of event logs in three main directions [2]: *i)* discovery, *ii)* conformance and *iii)* enhancement. Discovery techniques produce a process model. Conformance techniques compare a process model with the event log of the same process. It verifies if the process model is adjusted to the reality recorded in event logs. Enhancement techniques extend or improve an existing process model by using event log information.

The minimum requirements for process mining are that any event can be related to both an activity and a case and that events within a case are ordered. At the same time, events can be characterized by various attributes (timestamp, resource or performer, activity name and other data). Different techniques use these attributes for specific analyzes [2], [3].

The learnt model can cover different approaches: *i)*control-flow, which describes the order of execution of the activities within the process; *ii)* organizational, which discovers the actors that participate in the process and how they are related; *iii)* case, which focuses on the values of the data elements to characterize the process and *iv)* the time approach, that allows time analysis.

2.1 Extraction of process logs from time series

A time series is a sequence of values of a certain magnitude that is recorded at specific times. If T is an index set (e.g. time), a time series is given by $Y = \{Y_t : t \in T\}$. In the case of real valued magnitudes, $Y_t \in \mathbb{R}$. Thus, from a process mining perspective, where the series is seen as a process, it is necessary to represent it as an event log and therefore, to identify the elements (cases, activities, resources and time). This could be done in several ways. Cases will be defined here as segments of the series of a certain length with or without overlapping (they will be unique).

In this sense, since activities and resources are discrete entities, in time series of continuous magnitudes where $X_t \in \mathbb{R}$, a discretization process is required. Change is an essential aspect in a time series and it could be seen as the *activity* that the series as a process, performs at a given time. Since time series values experience these changes, they could be seen as the entities that perform the changes, that is, as the *resources* of the process.

In a simple approach, resources could be interpreted as the intensity levels of the discretized series. If $\mathcal{C}_p = \{C_1, \cdots C_n\}, n \in \mathbb{N}^+, C_i \in \mathbb{R}$, for all $i \in [1, n]$ with $C_i < C_{i+1}, i \in [1, n)$ is a sequence of cut points, it induces a partition of the range of Y into $p = n - 1$ categories. They could be considered as the state of the series at time t, $L_k = \{Y_t \in (C_k, C_{k+1}]\}$ and collectively, as the resources of the process.

In this simple approach, the activity is the change experienced by a time series value and it is defined by characterizing qualitatively and quantitatively the difference $(Y_{t+1} - Y_t)$. Three classes of change are considered: Increasing, Decreasing, and Constant (no change). If min, max are functions returning the minimum and maximum of the time series respectively, $R = |max(Y_t) - min(Y_t)|$ is the range of the time series and $\alpha \in (0,1), \beta \in (0,1)$ are constants, then the activity performed by a resource at time t can be defined as is $Act(L_k) = C$, iff $|Y_{t+1} - Y_t| \leq \alpha R$. Otherwise, if $(Y_{t+1} > Y_t)$, $Act(L_k) = I_d$, where $d = round(|Y_{t+1} - Y_t|/(\beta R))$ ($round(x)$ returns the closest integer to x). Finally, if $(Y_{t+1} < Y_t)$, $Act(L_k) = D_d$. The scenarios are shown in Fig 1.

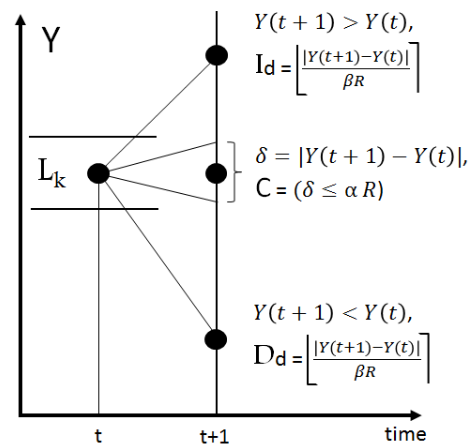

Fig. 1. Simple approach for defining *resources* and *activities* in a time series.

This simple approach is the one used in the paper for illustrating the analysis of a continuous time series (Section 4). Note that other approaches for defining an event log from a time series are possible. For instance, *resources* and *activities* could be defined in terms of category levels with respect to the mean of the series $\overline{T_t}$ by creating the cutpoints $\mathcal{C}_p = \{C_1, \cdots C_n\}$ covering intervals given by a certain fraction γ of the standard deviation $\sigma(Y)$ of Y. In the same way, constants $\alpha \in (0,1), \beta \in (0,1)$ for defining activity levels, could act upon $\sigma(Y)$ instead of on the range.

3 Process Mining Techniques

3.1 Fuzzy Models

Fuzzy miner is a process discovery algorithm capable of handling unstructured processes and numerous activities, providing a simplified process visualization [10], [11]. The algorithm uses correlation and significance metrics to simplify the process model and to build a graph where [11]: *i)* the most significant behavior is conserved, *ii)* the less significant and most correlated behavior is grouped, and *iii)* the less significant and less correlated behavior is not considered in the model.

Measurements of significance and correlation are modifiable in order to get the desired result. In this sense have been developed a set of metrics [10]: *i)* unary significance, *ii)* binary significance, and *iii)* binary correlation.

The algorithm initially creates an early process model where the importance of model nodes (i.e. event class) is determined by the unary significance and the edges are depicted by the binary significance and correlation. Later three transformations are applied to the model to simplify it: conflict resolution, edge filtering and aggregation and abstraction [11].

Conflict resolution. In this first transformations the conflict relation is identified, classified and resolved. There is a conflict relation when two nodes in the model are connected in both directions. The conflict relation can be classified in one of the three situations: length-two-loop, exception or concurrency. For resolving the conflict its relative significance is determined (for a more detailed see [11]). When the relative importance of a conflict relation ($A \rightarrow B$) is known, i.e. rel ($A \rightarrow B$) and rel ($B \rightarrow A$), is possible to resolve the conflict relation as follows: If $rel(A, B)$ or $rel(B, A)$ exceed a threshold value, then it is inferred that A and B build a length-two-loop and both relation remain in the model. When at least one of these two values is below the threshold value, then offset is determined $ofs(A, B) = |rel(A, B) - rel(B, A)|$, whether offset value exceeds a specified ratio threshold then it is assumed that the less significant relation is an exception and it is remove it from the process model. Otherwise, when offset value is inferior to the specified ratio threshold, it is concluding that A and B are concurrent and it is removed from the model.

Edge Filtering. In this transformation, each edge is evaluated by its utility $util(A, B)$, a weighted sum of its significance and correlation (for a more detailed see [11]). Each incoming and outgoing edges is filtered. The edge cutoff parameter $co \in [0, 1]$ allows configuring which edges are preserved. For each node, the utility value is normalized to $[0, 1]$ where is assigned 1 to the strongest edge. All edges whose normalization exceeds utility value are added to the model.

Node Aggregation and Abstraction. In this last transformation the main idea is to preserve highly correlated clusters of less-significant nodes and take away less-significant isolated nodes. Removing nodes is configurable on the node cutoff parameter. Nodes whose unary significance are below parameter can either be aggregated or abstracted.

3.2 Evolutionary Multi-objective Pareto Process Trees

None of the standard techniques for learning process models guarantee the production of syntactically correct models. Moreover, they do not provide insights into the trade-offs between the different quality measures. The Evolutionary Tree Miner algorithm (ETMd) [5], [6] is capable of learning sound process models that balances four established quality measures: *i)* simplicity, *ii)* replay fitness, *iii)* precision and *iv)* generalization.

Simplicity is about reducing the size of a process model by removing nodes that do not improve or compromise behavior in order to give preference to simpler, rather than complex models (Occam's Razor principle). The replay fitness measure quantifies the fraction of the event log supported by the process model, the precision measure quantifies how much of the behavior described by the process model is not observed in the event log and generalization evaluates how the process model explains the behavior of the system, and not only the particular event log describing the observed behavior [7].

ETMd is an implementation of a genetic programming evolutionary algorithm, which evolves trees, each one representing a process model. It works by generating an initial population with candidate solutions that are evaluated (using the aforementioned four quality measures), and processed with evolutionary operators (selection, crossover and mutation), in cycles that produce successive generations (with/without elitism), until a termination criterion is met (number of generations surpassed, lack of improvement of the best solution, performance measures exceeding a given threshold, among others). Common selection mechanisms are roulette-wheel and tournament selection.

In the ETMd algorithm provisos are taken to prevent bloat phenomenon, like prioritizing smaller over larger solutions, common in genetic programming scenarios. The evaluation of candidate solutions is not based on weighted averages of the individual objective functions, which suffers from several disadvantages. Instead, a four dimensional Pareto front is maintained and updated at every generation, ensuring a true multi-objective optimization process, that gradually eliminates the dominated solutions in favor of the non-dominated ones.

At the end of the evolution, the user examines the resulting pairwise Pareto fronts related to the model quality measures and makes his choice. An important element differentiating ELM to other approaches to process mining is that even though models are evaluated using the event log data, they are the result of a generative process involving many candidate solutions, having multiple objectives, where the best solutions balance these objectives.

3.3 Social Networks

Process mining generally focuses on discovering and analyzing the process model [3]. However, when the event log contains information about the resource it is possible to construct and analyze social networks. When the event log contains time information, it is possible to infer causal relations between activities and also social relations between resources [1], [4]. Different metrics allow the identification of relations between resources within the process: *i)* handover of work, *ii)* subcontracting, *iii)* working together, *iv)* similar task, and *v)* reassignment.

From them, *handover of work* and *subcontracting*, are based on causality. Their objective is to identify how work flows between resources within a process and they were the ones used in this paper.

There is *handover of work*, within a process instance from resource i to resource j when there are two successive activities where the first is performed by i and the second by j. When a resource j performed an activity between two activities performed by resource i, it is said that the work was *subcontracted* from i to j. In both metrics it is possible to consider direct and indirect successions using a *causality fall factor*, that specifies the activity number in-between an activity complete by i and other complete by j [1], [4].

4 Application example: The 10.7 cm Solar Radio Flux series

The Sun structure and behavior are largely controlled by magnetic fields. The level of magnetic activity follows an 11 (really a 22) year cycle. This rhythm pervades the Sun and modulates physical processes taking place at different locations throughout the Sun. The result is variations in the Sun's energy output and other emissions, for example, the ultraviolet emissions that heat the Earth's atmosphere and change the ionosphere.

Monitoring the Sun is extremely important because of the impact that it has on Earth, ultimately affecting human activities (both on Earth and in space). Geomagnetic storms caused by solar flares and coronal mass ejections are responsible for distorting communications, satellites, the power grid and many other distortions with economic impact measured in millions. One of the most useful solar activity indices is the 10.7 cm solar radio flux (F10.7), which has been measured by the National Research Council of Canada since 1947 [15].

This index consists of measurements of the total solar radio emission at 10.7 cm wavelength (a frequency of 2800 MHz). It comprises contributions from the solar disc plus emission from all the activity centers on it. At least three emission mechanisms are involved. The most important are thermal free-free emission from plasma concentrations trapped in the chromosphere and lower corona by magnetic fields, and thermal gyroresonance, where those magnetic fields are strong enough for the electron gyrofrequency (fg (MHz) = 2.8 B (Gauss)) to be higher than about a third of the observing frequency.

This requirement is often met in the strong magnetic fields overlying sunspots. The third contribution is gyrosynchrotron (non-thermal) emission, driven by electrons accelerated by flares or other reconnection processes [16]. The main use of this index is to reflect the changes in the general level of magnetic activity-evolutionary changes in active structures, which have characteristic time-scales ranging from hours to weeks. The non-thermal emissions may vary dramatically over seconds to hours [17]. Three flux values are distributed for each measurement: the *Observed Flux*, which is the value as measured, the *Adjusted Flux*, which is the value corrected for the annual variations in the Earth-Sun distance, and the *URSI Series-D Flux*, which is 0.9 times the *Adjusted Flux*.

The flux data were obtained from Canadian Space Weather Centre [14]. For a more detailed discussion of the F10.7 solar radio flux activity index, see [15]. The $F_{10.7}$ solar radio flux data used in this paper are daily local noon values of the Adjusted Flux, smoothed using 27-point adjacent-averaging. This corresponds to the size of the synoptic map ($\approx$ 27 days), covering a single solar rotation. The daily averaged and smoothed $F_{10.7}$ flux for the time period from 2006 to October 2018 is shown in Fig. 2, which includes the discretization levels used for creating the event logs (Section 2.1), as well as the time periods corresponding to the different solar cycles (denoted as C_{xx}, where xx indicates the given year).

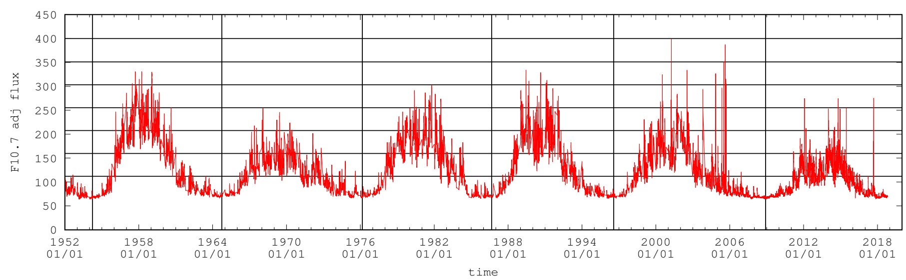

Fig. 2. $F_{10.7}$ solar radio flux series (1947 − 2018). Top labels indicate solar cycles [19 − 24]. Horizontal lines indicate intensity levels categorized into classes (7). Solar cycles covered by the F10.7 series are labeled at the top.

The flux is expressed in solar flux units (1 sfu = 10^{-22} W m^{-2} Hz^{-1}). In the figure we can see declining phase of the solar cycle 23 and the cycle 24, which is nearing its minimum. The flux encompass full solar cycle 23 and 24 more than solar cycle length.

4.1 Fuzzy Models

According to these results, fuzzy models were computed for the first and last two solar cycles contained in the $F_{10.7}$ record (Cycles 19, 23 and 24). They are shown in Fig 3 (Top row), together with the models obtained with the Evolutionary MO techniques and the $F_{10.7}$ series for comparison. Each fuzzy model is a graph that provides a simplified process visualization and describes the precedence relations among event classes. The yellow squares represent significant activities and each node is labeled with the event class name and the significance value. The edges that link nodes express their significance with thickness and darkness proportional to the strength of the connection.

The fuzzy models clearly reveal differences in the structure of the sub processes associated to the cycles. Solar Cycle 19 consists of only six classes of events while cycles 23 and 24 consist of 15 and 10 classes of events respectively, related in a much more complex manner. Taking into account the number of classes of events, the model for Cycle 19 seems much simpler and more balanced, i.e. with less abrupt jumps. The model of Cycle 23 involves the largest number of event types, related to the large number of high spikes characteristic of this cycle.

Finally, Cycle 24 involves 10 classes of events, of which those indicating little changes appear with the highest frequency ($\{D_1, I_1, D_2, I_2$ and $D_3\}$). This situation explains why the cycle is observed flatter than the two previous cycles. However, there is an abrupt peak (D_6) with high frequency, indicating sudden, higher intensity variations, easily identified in the $F_{10.7}$ behavior during Cycle 24.

4.2 Evolutionary Multi-objective Models

The ETMd algorithm was applied to the activities sub-logs of the $F_{10.7}$ for Cycles $\{19, 23, 24\}$ with the following parameters: population size= 20, elite count= 5, nbr. of generations= 1000, cross-over rate= 0.25, random tree creation rate= 0.25, random node addition= 1, random node removal rate= 1, random node mutation= 1 and useless node removal= 1. No solutions were filtered from the Pareto front based on quality measures preset thresholds. Upon termination for each case a trade-off solution was chosen as the one on the Pareto front, closest to the overall optimum given by the vector $< 1, 1, 1, 1 >$ determined by the best values of the individual quality measures (Section 3.2).

The resulting process trees are shown in Fig 3 (left to right for Cycles $\{19, 23, 24\}$). As with the fuzzy models, there are immediate differences in the underlying dynamics of the sub-processes for the starting and the ending solar cycles along the $F_{10.7}$ flux record, exposed by the pareto trade-off models. However, when all quality measures are considered simultaneously, Cycle 19 exhibits a longer sequence of elements (7), with a larger number of constant and $d = 1$ order increasing/decreasing changes and loops, compared to Cycles $23, 24$. On the other hand, Cycle 23 is the one with the deepest tree and with jumps which are either small or more towards the extreme ($d = \{1, 3, 7\}$). Cycle 24 has a simpler activity change schema and slightly shorter sequences. It is structurally more similar to 23 than to 19, a relation that coincides with the one exhibited by their fuzzy model counterparts. These findings provide more insight on the changes in Sun's behavior during the last cycles [17].

4.3 Social Networks

The social networks constructed from the $F_{10.7}$ process log for solar cycles $\{19, 23, 24\}$ are shown in Fig. 4, corresponding (top to bottom) to the handover of work, similar tasks and the subcontracting models respectively. Recall that resources are the radiation intensity levels at which the different types of changes (activities) take place.

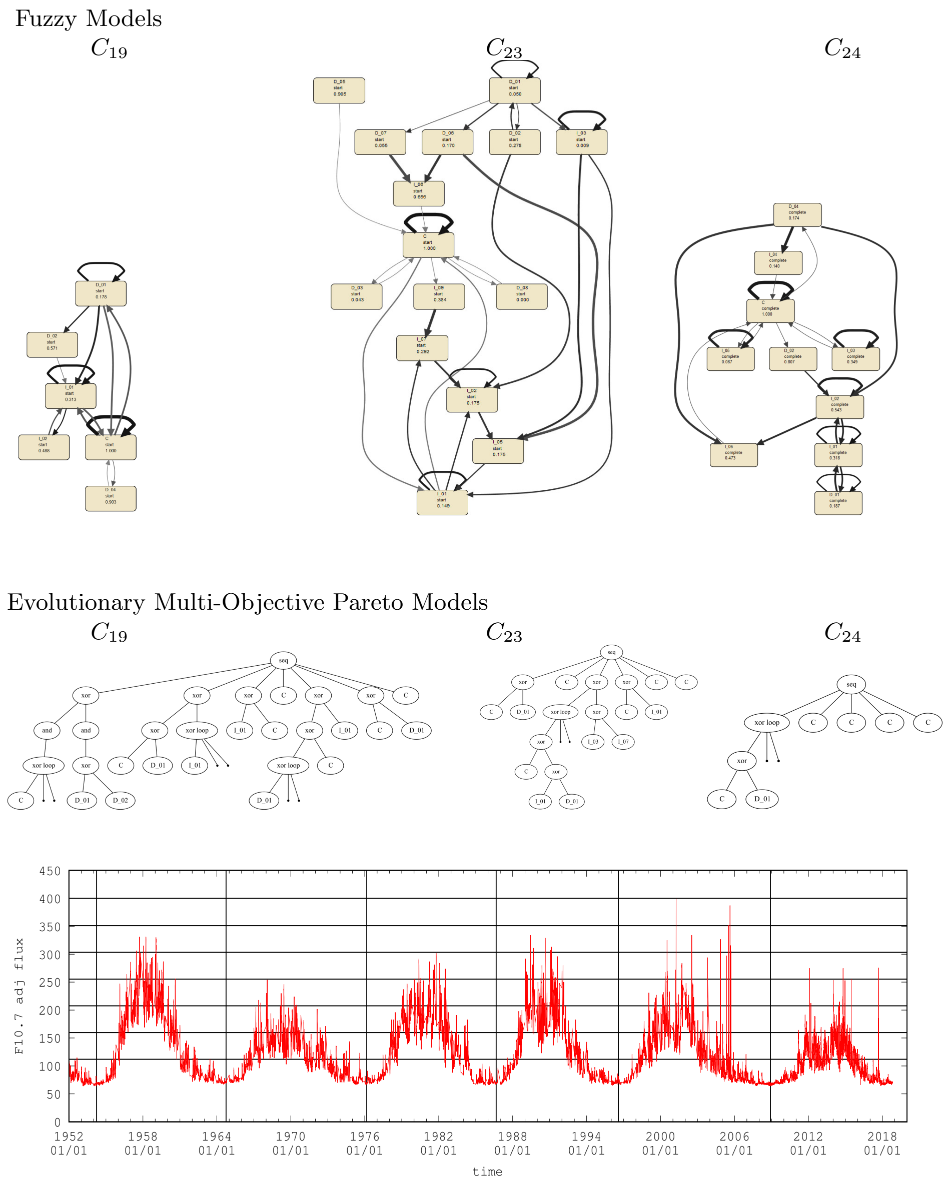

Fig. 3. $F_{10.7}$ solar radio flux series (1947 – 2018). Fuzzy and Evolutionary Multi-Objective Pareto Models corresponding to Solar cycles 19, 23 and 24 (each row contains networks of the same type). Left hand side: Cycle 19. Right hand side: Cycles 23 and 24.

The handover of work analysis indicates how activities are passed from one resource to another, and it depends on two parameters: The first one indicates whether to consider multiple transfers within one instance. The second parameter indicates whether to consider direct succession. In the representation, node sizes indicate the frequency with which resources have executed activities per process instances.

The network for Cycle 19 involves six resources, with R_3 being has one with the greatest participation. It receives work from R_2, R_4 and R_6 and it gives work to R_2 and R_4. The opposite happens with with the R_1 which has less participation with only two edges, one for giving work to R_2 and other for receiving work from R_2. On the other hand, for Cycle 23, seven resources are involved. The central nodes are R_2 and R_3 with 9 edges each. R_2 receives work from $\{R_1, R_3, R_4, R_5, R_6\}$ and it gives work to $\{R_1, R_3, R_5, R_6\}$. R_3 receives work from $\{R_1, R_3, R_4, R_5, R_6\}$ and it gives work to $\{R_1, R_3, R_5, R_6\}$.

The opposite happens with R_5 and R_7, which have less number of edges. R_5 is related to R_1 and R_3; while R_7 is related to R_2 and R_4. Altogether, the dynamics is very different from the exhibited by Cycle 19. The network of Cycle 24, involves five resources. Node R_2 is the one with the highest number of relations. It gives work to four other nodes and also receives work from them. R_3 is the node with fewer edges, as it only relates to R_2 and R_4, from which it gives and receives work. As was seen with other techniques, the structure and behavior is more similar to Cycle 23 than to Cycle 19.

Subcontracting Social Network. This type of network provides insight about resources performing an activity in between two other activities performed by other resource. It depends on two parameters: the first one establishes whether to consider multiple subcontracting relations and the second one allows the consideration of indirect subcontracting relationships. Nodes sizes differs, because they are proportional to the amount of contracting and subcontracting relationships. These networks are also shown in Fig. 4 (Subcontracting Models). Cycle 19 has six nodes in total. Nodes $\{R_2, R_3, R_4, R_5\}$ have the same behavior (two incoming and two outgoing edges), whereas nodes R_1 and R_6 have only two edges, one incoming and one outgoing. Specifically, R_1 subcontracts and is subcontracted only by the R_2 node. R_6 node subcontracts and is subcontracted only by the R_5 node. In contrast, Cycle 23 consist of seven nodes and with a very different structure, of which R_2 is the node with the most incoming and outgoing connections. The two incoming edges indicate that R_2 has been subcontracted twice, and the four outgoing edges indicate that R_2 has subcontracted four other nodes $\{R_1, R_3, R_5, R_6\}$. Interestingly, nodes $\{R_5, R_6\}$ only perform subcontracted work, whereas node R_7 is only subcontracted by R_3. Finally, Cycle 24 consists of five nodes all of which have the same behavior, four incoming edges and four outgoing edges, which indicates that all are subcontracted and are subcontracted equally. Although different, its structure is more similar to Cycle 23 than to Cycle 19.

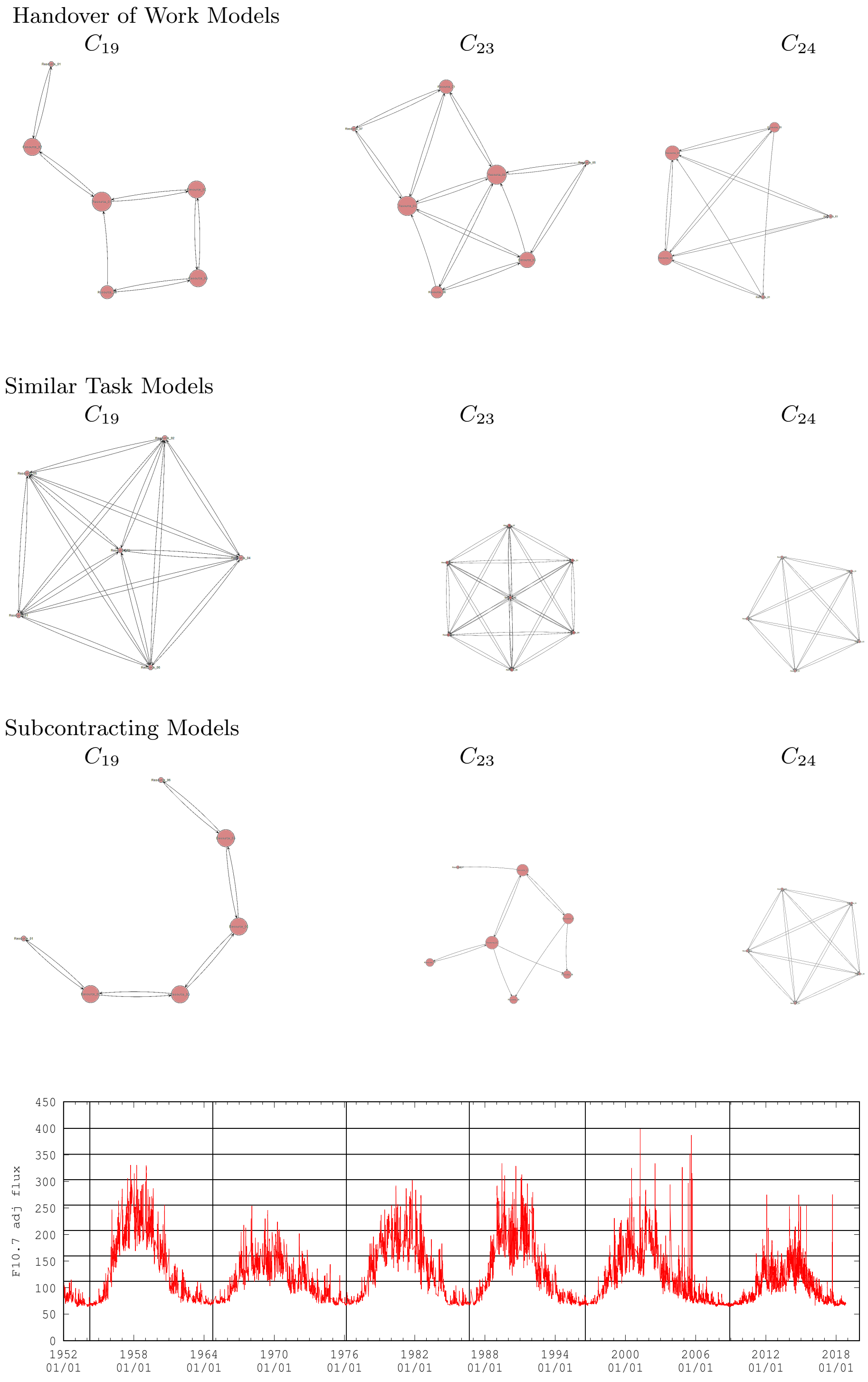

Fig. 4. $F_{10.7}$ solar radio flux series (1947−2018). Social Networks models corresponding to Solar cycles 19, 23 and 24 (each row contains networks of the same type). Left hand side: Cycle 19. Right hand side: Cycles 23 and 24.

5 Conclusions

Process mining was discussed in the context of the analysis of continuous, real-valued time-varying magnitudes like time series and the monitoring with sensor data. They are important in a broad variety of domains, like natural and life sciences, engineering, and many others.

Approaches were presented that describe the variations of continuous magnitudes as event logs where the intensity levels of the time series are interpreted as the resources involved and the type and magnitude of their variation are mapped to activities. This representation allows the application of process mining techniques to problems like monitoring with sensor data and other types of time-varying phenomena.

In particular, the Fuzzy Miner (FM) and the Multi-Objective Evolutionary Tree Miner (ETMd) process mining algorithms were applied to a time series of the F10.7 flux index of Solar activity. These techniques successfully constructed models for the process that exposed the differences in the internal structure of the time series between Solar cycles, and provided a better understanding of the changing dynamics of the physical system (the Sun). In this application Process Mining proved to be a valuable tool for analyzing the rhythm of solar activity and how it is changing.

The results obtained are promising and further studies should extend the range of application domains, the dimensionality of the time series, the data types of the variables describing the time-dependent processes, as well as comparison with other data mining procedures.

Acknowledgment

The $F_{10.7}$ data are provided by the National Research Council of Canada and Natural Resources Canada. Y. Céspedes-González acknowledges the support the University of Veracruz/Faculty of Accounting and Administration/Ph.D program in Administrative Sciences.

References

1. W.M. van der Aalst, H.R., Song, M.: Discovering social networks from event logs. Computer Supported Cooperative Work (CSCW) **14**(6), 549–593 (2005)
2. van der Aalst, W.M.: Process Mining Data Science in Action. Springer (2016)
3. van der Aalst, W.M., Adriansyah, A., de Medeiros, A.K.A., Arcieri, F., Baier, T., Blickle, T., et al.: Process mining manifesto. Business process management workshops **99**, 169–194 (2012)
4. van der Aalst, W., Song, M.: Mining social networks: Uncovering interaction patterns in business processes. In: Springer (ed.) Proc. of the International conference on business process management. pp. 244–260 (2004)
5. Buijs, J.: Flexible Evolutionary Algorithms for Mining Structured Process Models. Ph.D. thesis, Technical University of Eindhoven (2014)

6. Buijs, J., van Dongen, B., van der Aalst, W.: A genetic algorithm for discovering process trees. In: Proc. of the 2012 IEEE Congress on Evolutionary Computation. pp. 1—-8 (2012)
7. Buijs, J., van Dongen, B., van der Aalst, W.: Quality dimensions in process discovery: The importance of fitness, precision, generalization and simplicity. International Journal of Cooperative Information Systems (2014)
8. Dongen, S.V.: A cluster algorithm for graphs. Tech. rep., National Research Institute for Mathematics and Computer Science in the Netherlands (2000)
9. Enright, A.J., Dongen1, S.V., Ouzounis, C.A.: An efficient algorithm for large-scale detection of protein families. Nucleic Acids Research **30**(7), 1575—-1584 (2002)
10. Günther, C.W.: Process Mining in Flexible Environments. Ph.D. thesis, School for Operations Management and Logistics. Eindhoven University of Technology (2009)
11. Günther, C.W., van der Aalst, W.M.: Fuzzy mining–adaptive process simplification based on multi-perspective metrics. Springer **4714**, 328–343 (2007)
12. Hompes, B., Buijs, J., van der Aalst, W., Dixit, P., Buurman, J.: Detecting change in processes using comparative trace clustering. In: Proc. of the 5th International Symposium on Data-driven Process Discovery and Analysis (SIMPDA 2015). pp. 95—-108 (2015)
13. Hompes, B., Buijs, J., van der Aalst, W., Dixit, P., Buurman, J.: Discovering deviating cases and process variants using trace clustering. In: Proc. of the 27th Benelux Conference on Artificial Intelligence (BNAIC). Hasselt, Belgium (November 5-6 2015)
14. NRCan: Space weather canada (2019), \url{https://spaceweather.gc.ca/index-en.php}
15. Tapping, K.: The 10.7 cm solar radio flux (f10.7). Space Weather **11**, 394—-406 (2013)
16. Tapping, K., DeTracey, B.: The origin of the 10.7 cm solar flux. Solar Physics **127**, 321—-332 (1990)
17. Tapping, K., Valdés, J.: Did the sun change its behaviour during the decline of cycle 23 and into cycle 24? Solar Phys. **272**, 337–347 (2011)