



Guilherme Gonçalves Schardong

**Visual interactive support for selecting
scenarios from time-series ensembles**

Tese de Doutorado

Thesis presented to the Programa de Pós-graduação em Informática of PUC-Rio in partial fulfillment of the requirements for the degree of Doutor em Ciências – Informática.

Advisor : Prof. Hélio Côrtes Vieira Lopes
Co-advisor: Prof^a. Simone Diniz Junqueira Barbosa

Rio de Janeiro
September 2018



Guilherme Gonçalves Schardong

**Visual interactive support for selecting
scenarios from time-series ensembles**

Thesis presented to the Programa de Pós-graduação em Informática of PUC-Rio in partial fulfillment of the requirements for the degree of Doutor em Ciências – Informática. Approved by the undersigned Examination Committee.

Prof. Hélio Côrtes Vieira Lopes

Advisor

Departamento de Informática – PUC-Rio

Prof^a. Simone Diniz Junqueira Barbosa

Co-advisor

Departamento de Informática – PUC-Rio

Prof. Fernando Luiz Cyrino Oliveira

Departamento de Engenharia Industrial – PUC-Rio

Prof. Bruno Fânzeres dos Santos

Departamento de Engenharia Industrial – PUC-Rio

Prof. Regis Kruehl Romeu

CENPES

Prof. Sergio Lima Netto

UFRJ

Prof. Abelardo Borges Barreto Junior

Departamento de Matemática – PUC-Rio

Prof. Alex Laier Bordignon

UFF

Prof. Márcio da Silveira Carvalho

Vice Dean of Graduate Studies

Centro Técnico Científico – PUC-Rio

Rio de Janeiro, September 13th, 2018

All rights reserved.

Guilherme Gonçalves Schardong

Bachelor's in Computer Science (2011) and Masters' in Informatics (2014) at the Federal University of Santa Maria (UFSM), with emphasis on Applied Computing. Worked for the Applied Computing Laboratory (LaCA) from 2009 to 2014 and Tecgraf from 2014 to 2016 in the Reservoir visualization group. Since 2016 works at GALGOS, former ATD-Lab, at PUC-Rio.

Bibliographic data

Schardong, Guilherme Gonçalves

Visual interactive support for selecting scenarios from time-series ensembles / Guilherme Gonçalves Schardong; advisor: Hélio Côrtes Vieira Lopes; co-advisor: Simone Diniz Junqueira Barbosa. – Rio de Janeiro: PUC-Rio, Departamento de Informática, 2018.

v., 17 f: il. color. ; 30 cm

Tese (doutorado) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Informática.

Inclui bibliografia

1. Redução de Cenários. 2. Interatividade. 3. Visualização Científica. 4. Tomada de Decisão. 5. Conjuntos de Dados Temporais. I. Lopes, Hélio Côrtes Vieira. II. Barbosa, Simone Diniz Junqueira. III. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Informática. IV. Título.

CDD: 004

Acknowledgments

To Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for partially financing this research under grant XXXXXX/YYYY-X. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001

Abstract

Schardong, Guilherme Gonçalves; Lopes, Hélio Côrtes Vieira (Advisor); Barbosa, Simone Diniz Junqueira (Co-Advisor). **Visual interactive support for selecting scenarios from time-series ensembles**. Rio de Janeiro, 2018. 17p. Tese de doutorado – Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

Stochastic programming and scenario reduction approaches have become invaluable in the analysis and behavior prediction of dynamic systems. However, such techniques often fail to take advantage of the user's own expertise about the problem domain. This work provides visual interactive support to assist users in solving the scenario reduction problem with time-series data. We employ a series of time-based visualization techniques linked together to perform the task. By adapting a multidimensional projection algorithm to handle temporal data, we can graphically present the evolution of the ensemble. We also propose to use cumulative bump charts to visually compare the ranks of distances between the ensemble time series and a baseline series. To evaluate our approach, we developed a prototype application and conducted observation studies with volunteer users of varying backgrounds and levels of expertise. Our results indicate that a graphical approach to scenario reduction may result in a good subset of scenarios and provides a valuable tool for data exploration in this context. The users liked the interaction mechanisms provided and judged the task to be easy to perform with the tools we have developed. We tested the proposed approach against state-of-the-art techniques proposed in the literature and used in the industry and obtained good results, thus indicating that our approach is viable in a real-world scenario.

Keywords

Scenario Reduction; User Interaction; Scientific Visualization; Decision Making; Time Series Ensembles.

Resumo

Schardong, Guilherme Gonçalves; Lopes, Hélio Côrtes Vieira; Barbosa, Simone Diniz Junqueira. **Uma Abordagem Visual e Interativa para a Seleção de Conjuntos de Cenários Temporais**. Rio de Janeiro, 2018. 17p. Tese de Doutorado – Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

O uso de abordagens de programação estocástica e redução de cenários tem se tornado imprescindível na análise e predição de comportamento de sistemas dinâmicos. Entretanto, tais técnicas não levam em conta o conhecimento prévio sobre domínio que o usuário possui. O presente trabalho tem por objetivo o desenvolvimento de uma abordagem visual e interativa para abordar o problema de redução de cenários com dados temporais. Para tanto, nós propomos a implementação de uma série de visualizações de dados temporais integradas. Também propomos a adaptação de um algoritmo de projeção multidimensional para lidar com dados temporais. Desta forma, podemos representar graficamente a evolução de um conjunto de cenários ao longo do tempo. Outra visualização proposta no presente trabalho é uma adaptação de *Bump chart* para lidar com dados temporais acumulados; através dele, um usuário pode comparar a evolução das distâncias entre os diferentes cenários e um cenário de referência. Para validar a nossa proposta, fizemos uma implementação das técnicas propostas e conduzimos um estudo com usuários de diferentes áreas do conhecimento e níveis de experiência. Os resultados obtidos até então indicam que uma abordagem visual para o problema de redução de cenários é viável, e permite a seleção de um conjunto razoável de cenários. Além disso, constatamos que essa abordagem pode ser útil em um contexto de exploração de dados visando a redução de cenários. O usuário também pode explorar visualmente os resultados de outras técnicas de redução de cenários usando nossa abordagem. Os usuários entrevistados reportaram facilidade em cumprir as tarefas propostas e comentaram positivamente sobre os mecanismos de interação fornecidos pelo nosso protótipo. Também testamos os cenários escolhidos usando nossa proposta contra outras abordagens encontradas tanto na literatura quanto em uso na indústria. Os resultados obtidos foram bons, indicando que nossa proposta é viável em casos de uso reais.

Palavras-chave

Redução de Cenários; Interatividade; Visualização Científica; Tomada de Decisão; Conjuntos de Dados Temporais.

Table of contents

1	Introduction	12
1.1	Thesis Structure	12
	Bibliography	16
A	User profile questionnaire for prototype evaluation sessions	17

List of figures

Figure 1.1 Juxtaposition, Superposition and Explicit Encoding. Image taken from the work of (Szafr2018).	13
Figure 1.2 Comparison between different time encodings on the Time-lapsed LAMP chart of W_p . Superposition (1.2a) and Ex- plicit Encoding (1.2b) from small to large glyphs.	14

List of tables

Table 1.1	Errors for scenarios selected using the industry standard approach.	13
-----------	---	----

List of abbreviations

ANP – Agência Nacional do Petróleo

CENPES – Centro de Pesquisas Leopoldo Américo Miguez de Mello

LAMP – Local Affine Multidimensional Projection

MDS – Multidimensional Scaling

MP – Multidimensional Projections

N_p – Cumulative oil production

P₁₀ – Percentile 10

P₅₀ – Percentile 50

P₉₀ – Percentile 90

PETROBRAS – Petróleo Brasileiro S.A.

Q_o – Oil flow

Q_w – Water flow

W_p – Cumulative water production

*ALL THINGS THAT ARE, ARE OURS.
BUT WE MUST CARE. FOR IF WE DO
NOT CARE, WE DO NOT EXIST. IF
WE DO NOT EXIST, THEN THERE IS
NOTHING BUT BLIND OBLIVION. AND
EVEN OBLIVION MUST END SOMEDAY.
LORD, WILL YOU GRANT ME JUST A
LITTLE TIME? FOR THE PROPER BAL-
ANCE OF THINGS. TO RETURN WHAT
WAS GIVEN. FOR THE SAKE OF PRIS-
ONERS AND THE FLIGHT OF BIRDS.
LORD, WHAT CAN THE HARVEST HOPE
FOR, IF NOT FOR THE CARE OF THE
REAPER MAN?*

Terry Pratchett, *Reaper Man*.

1

Introduction

This is a simple, introductory text demonstrating how to use the PUC-Rio's thesis/dissertations L^AT_EX class. Here are some citations (Lee2010), (Hummel2013), (Dupacova2003). In text citations, according to (Armstrong2013) work the same way.

In summary, the contributions of this class are:

- A L^AT_EX template for thesis/dissertations for PUC-Rio's students;
- A way to centralize changes and increase the reach of fixes and adjustments.

Table 1.1 and Figure 1.1 present ways to include tables and figures in the text. Figure 1.2 shows how to include subfigures. Section 1.1 is just a dummy section.

1.1

Thesis Structure

This is just a dummy subsection, don't worry about it.

Table 1.1: Scenarios and errors ($\times 10^{10}$) for the industry standard approach.

Property	Percentile	Scenario	SSE	MSE
N_p	P ₁₀	172	1190.0	30.5
		88	663.0	17.0
		122	1660.0	42.6
		36	312.0	7.9
	P ₅₀	131	162.0	4.1
		4	162.0	4.1
		26	786.0	20.1
		90	236.0	6.0
	P ₉₀	96	975.0	25.0
		100	690.0	17.7
		127	925.0	23.7
		132	526.0	13.5
W_p	P ₁₀	23	7640.0	196.0
		115	42000.0	1080.0
		10	26900.0	690.0
		19	27600.0	707.0
	P ₅₀	29	14200.0	364.0
		153	6260.0	161.0
		84	5840.0	150.0
		88	11200.0	288.0
	P ₉₀	37	14300.0	367.0
		57	22300.0	571.0
		187	14000.0	358.0
		28	2210.0	56.7

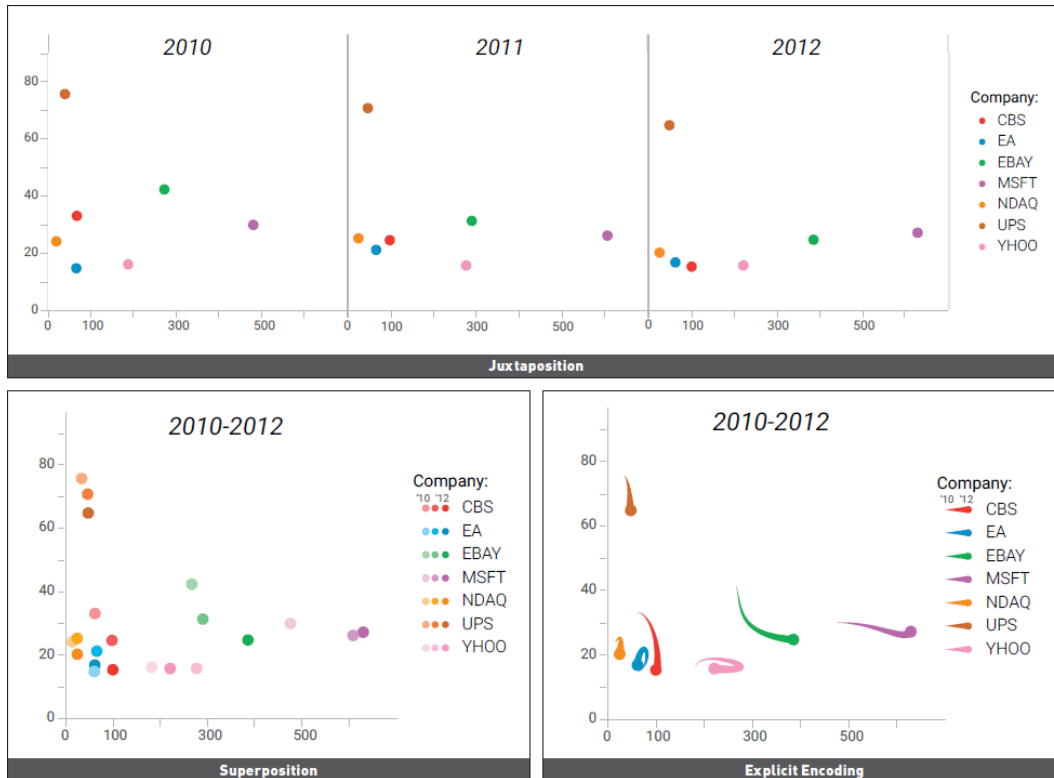
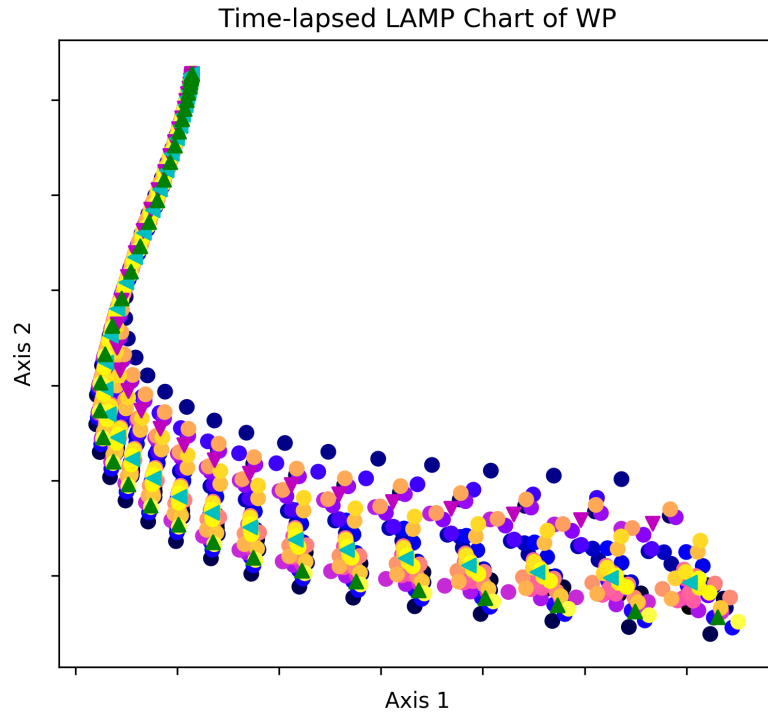
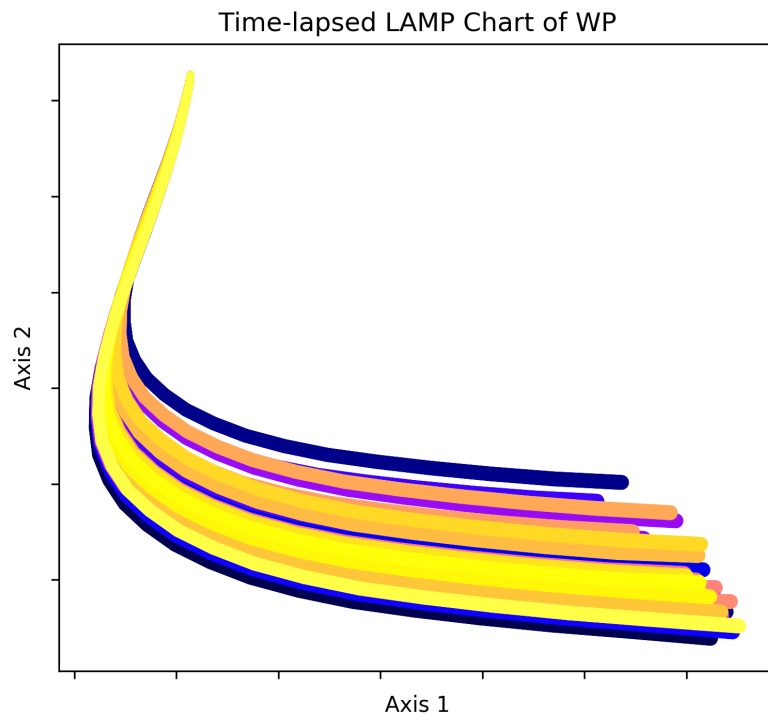


Figure 1.1: Juxtaposition, Superposition and Explicit Encoding. Image taken from the work of (Szafir2018).



(a) Superposition of times in the Time-lapsed LAMP chart.



(b) Explicit encoding of time in the Time-lapsed LAMP chart. Starting times encoded with smaller glyphs.

Figure 1.2: Comparison between different time encodings on the Time-lapsed LAMP chart of W_p . Superposition (1.2a) and Explicit Encoding (1.2b) from small to large glyphs.

Lemma 1.1 (Unimodularity preserving matrix operations) *The fol-*

lowing matrix operations, called unimodular operations, preserve the unimodularity property from an already unimodular matrix:

1. Interchange two columns
2. Add an integer multiple of a column to another column
3. Multiply a column by -1
4. Transpose

Bibliography

- [Armstrong2013] ARMSTRONG, M.; NDIAYE, A.; RAZANATSIMBA, R. ; GALLI, A.. **Scenario Reduction Applied to Geostatistical Simulations**. Mathematical Geosciences, 45(2):165–182, feb 2013.
- [Dupacova2003] DUPAČOVÁ, J.; GRÖWE-KUSKA, N. ; RÖMISCH, W.. **Scenario reduction in stochastic programming**. Mathematical Programming, 95(3):493–511, 2003.
- [Hummel2013] HUMMEL, M.; OBERMAIER, H.; GARTH, C. ; JOY, K. I.. **Comparative visual analysis of lagrangian transport in cfd ensembles**. Visualization and Computer Graphics, IEEE Transactions on, 19(12):2743–2752, 2013.
- [Lee2010] LEE, H. K. H.; TADDY, M. ; GRAY, G. A.. **Selection of a Representative Sample**. Journal of Classification, 27(1):41–53, mar 2010.
- [Szafir2018] SZAFIR, D. A.. **The good, the bad, and the biased: Five ways visualizations can mislead (and how to fix them)**. Interactions, 25(4):26–33, June 2018.

A

User profile questionnaire for prototype evaluation sessions

1. Personal information

(a) Name

(b) E-mail

(c) Education degree

(d) Course semester (just for students)

(e) Occupation

2. Please mark below your knowledge about the following subjects:

(a) Division of a sample by percentiles: P10, P50, P80 ...

- ☐ I do not know
- ☐ I know little (I have learned these concepts at some point, but may have to learn again if I have to apply them)
- ☐ I have average knowledge (I may have to revise one concept or another if I have to apply it)
- ☐ I know well (I do not apply often, but I would not need to revise the concepts if I had to apply them)
- ☐ I am a specialist (I apply these concepts frequently)

(b) Analysis of trends and patterns in time series

- ☐ I do not know
- ☐ I know little (I have learned these concepts at some point, but may have to learn again if I have to apply them)
- ☐ I have average knowledge (I may have to revise one concept or another if I have to apply it)