

Universidade Federal de Campina Grande
Centro de Engenharia Elétrica e Informática
Coordenação de Pós-Graduação em Ciência da Computação

Dissertação de Mestrado

Uma abordagem para visualização e análise baseada
em *clustering* de dados espaço-temporais

Maxwell Guimarães de Oliveira

Campina Grande, Paraíba, Brasil

Agosto – 2012

Universidade Federal de Campina Grande
Centro de Engenharia Elétrica e Informática
Coordenação de Pós-Graduação em Ciência da Computação

Uma abordagem para visualização e análise baseada
em *clustering* de dados espaço-temporais

Maxwell Guimarães de Oliveira

Dissertação submetida à Coordenação do Curso de Pós-Graduação em
Ciência da Computação da Universidade Federal de Campina Grande -
Campus I como parte dos requisitos necessários para obtenção do grau
de Mestre em Ciência da Computação.

Área de Concentração: Ciência da Computação
Linha de Pesquisa: Metodologia e Técnicas da Computação

Cláudio de Souza Baptista, Ph.D.
(Orientador)

Campina Grande, Paraíba, Brasil

©Maxwell Guimarães de Oliveira, 20/08/2012

**DIGITALIZAÇÃO:
SISTEMOTECA - UFCG**

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA CENTRAL DA UFCG

O482u Oliveira, Maxwell Guimarães de.
Uma abordagem para visualização e análise baseada em *clustering* de dados espaço-temporais / Maxwell Guimarães de Oliveira. - Campina Grande, 2012.

116f.: il., color.

Dissertação (Mestrado em Ciência da Computação) – Universidade Federal de Campina Grande, Centro de Engenharia Elétrica e Informática.

Orientador: Prof. Ph.D. Cláudio de Souza Baptista.

Referências.

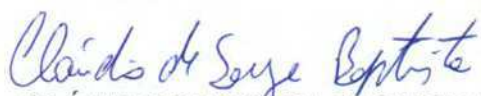
1. Dados Espaço-Temporais. 2. *GIS*. 3. Análise Visual. 4. *Clustering*. 5. Mineração de Dados. I. Título.

CDU 004.22 (043)

**"UMA ABORDAGEM PARA VISUALIZAÇÃO E ANÁLISE BASEADA EM CLUSTERING
DE DADOS ESPAÇO-TEMPORAIS"**

MAXWELL GUIMARAES DE OLIVEIRA

DISSERTAÇÃO APROVADA EM 20/08/2012



CLÁUDIO DE SOUZA BAPTISTA, Ph.D
Orientador(a)



ULRICH SCHIEL, Dr.
Examinador(a)



STANLEY ROBSON DE MEDEIROS OLIVEIRA, Ph.D
Examinador(a)

CAMPINA GRANDE - PB

Resumo

Atualmente, há um volume considerável de dados espaço-temporais disponíveis em vários meios, sobretudo na Internet. A visualização de dados espaço-temporais é uma tarefa complexa, que requer uma série de recursos visuais apropriados para que, em conjunto, possam permitir aos usuários uma correta interpretação das informações analisadas. Além do emprego de técnicas de visualização, a utilização de técnicas de descoberta de conhecimento em bancos de dados tem se mostrado relevante no auxílio à análise exploratória de relacionamentos em dados espaço-temporais. O levantamento do estado da arte em visualização de dados espaço-temporais leva à conclusão de que a área ainda é deficiente em soluções para visualização e análise desses tipos. Muitas abordagens abrangem somente questões espaciais, desprezando as características temporais desses dados. Inserido nesse contexto, o principal objetivo deste trabalho é melhorar a experiência do usuário em visualização e análise espaço-temporal, indo além do universo da visualização dos dados espaço-temporais brutos e considerando, também, a importância em visualização de dados espaço-temporais derivados de um processo de descoberta de conhecimento, mais especificamente algoritmos de *clustering*. Esse objetivo é atingido com a definição de uma abordagem inovadora em visualização de dados espaço-temporais, e de sua implementação, denominada GeoSTAT (*Geographic SpatioTemporal Analysis Tool*), que engloba pontos importantes observados nas principais abordagens existentes e acrescenta, principalmente, técnicas de visualização voltadas à dimensão temporal e à utilização de algoritmos de *clustering*, valorizando características até então pouco exploradas em dados espaço-temporais. A validação deste trabalho ocorre por meio de dois estudos de caso, onde cada um aborda dados espaço-temporais de um domínio específico, para demonstrar a experiência do usuário final diante das técnicas de visualização reunidas na abordagem proposta.

Abstract

Nowadays, there is a considerable amount of spatiotemporal data available in various media, especially on the Internet. The visualization of spatiotemporal data is a complex task that requires a series of visual suitable resources which can enable users to have a correct interpretation of the data. Apart from the use of visualization techniques, the use of techniques of knowledge discovery in databases has proven relevant for the exploratory analysis of relationships in spatiotemporal data. The state of the art in visualization of spatiotemporal data leads to the conclusion that the area is still deficient in solutions for viewing and analysis of those data. Many approaches cover only spatial issues, ignoring the temporal characteristics of such data. Inserted in this context, the main objective of this work is to improve the user experience in spatiotemporal visualization and analysis, going beyond the universe of visualization of spatiotemporal raw data and also considering the importance of visualization of spatiotemporal data derived from a knowledge discovery process, more specifically clustering algorithms. This goal is achieved by defining an innovative approach for the analysis and visualization of spatiotemporal data, and its implementation, called GeoSTAT (Spatiotemporal Geographic Analysis Tool), which includes important points in the main existing approaches and adds especially visualization techniques geared to the temporal dimension and the use of clustering algorithms, enhancing unexplored features in spatiotemporal data. The validation of this work occurs through two case studies, where each one deals with spatiotemporal data of a specific domain to demonstrate the end-user experience on the visualization techniques combined in the proposed approach.

Agradecimentos

Agradeço primeiramente a Deus, pois sem fé, nada faz sentido. Agradeço profundamente aos meus pais, José de Oliveira e Edilene Guimarães, pela força, pelas palavras de encorajamento e de esperança, e pelos ensinamentos que foram fundamentais para que eu estivesse aqui, finalizando este trabalho. Agradeço também ao meu irmão, Everton Guimarães, pelos momentos necessários de descontração.

Agradeço à minha namorada, Selma Barros, pelo apoio, pela presença, pelo carinho e compreensão, sobretudo nos momentos de maior dificuldade.

Agradeço ao meu orientador, professor Cláudio Baptista, pelo apoio dispensado ao longo destes dois anos e meio de pesquisa. Agradeço, também, aos professores da COPIN, em especial a Jacques Sauv , Ulrich Schiel, Carlos Eduardo Pires, Joseana Fechine, Hyggo Almeida e Angelo Perkusich, pela sabedoria, paci ncia, e pelas contribui es fundamentais   minha forma o.

Agradeço aos "parceiros" das horas de descontra o: Leonardo Sampaio, Rodrigo Noel, Marco Rosner, Bruno Dias, Danilo Abreu, Athayde Neto, Elthon Oliveira, Leonardo Soares, Giovanni Calheiros, Gustavo Soares, Daniel Bruno, Paulo Ditarso, Henrique Farias e H lio Moreira. Agradeço tamb m  s "meninas" das resenhas em Campina Grande: Carolina Nogueira, Catuxe Varg o, Lorena Maia, Andreia Oliveira, Kalina Medeiros e Kaline Ribeiro.

Agradeço aos companheiros mais pr ximos do LSI: Magna, Odilon, Vin cius, Thiago (vulgo Baiano), Ana Gabrielle, Adriano, Hugo e Daniel, pelos momentos de discuss o de ideias e, tamb m, pelos momentos de descontra o.

Agradeço tamb m a El as Barros e Tarc sio Ara jo, meus superiores, pela compreens o, amizade, e pelas libera es concedidas no ambiente de trabalho, horas valiosas que contribuíram de forma fundamental na conclus o deste trabalho.

Agradeço   Dona Alba, Dona In s e Sr. Marcos, pelo rubac o, salgadinhos irresist veis, e pelos quilos em excesso que adquiri nos  ltimos anos.

Agradeço   WindFM.com, "*The Classic Rock Station*", pelas v rias horas de companhia, sempre presente nos momentos de implementa o e escrita com m sicas de qualidade.

  CAPES, pelo apoio financeiro concedido aos 45" do segundo tempo.

Conteúdo

1	Introdução	1
1.1	Problemática	2
1.2	Objetivo do Trabalho	4
1.3	Relevância e Contribuições	4
1.3.1	Publicações Realizadas	5
1.4	Estrutura da Dissertação	6
2	Fundamentação Teórica	8
2.1	Dados Espaciais, Temporais e Espaço-Temporais	8
2.1.1	Dados Espaciais	8
2.1.2	Dados Temporais	11
2.1.3	Dados Espaço-Temporais	13
2.2	Mineração de Dados Espaço-Temporal	15
2.2.1	Motivação, Definições e Problemas	15
2.2.2	Fatores a Considerar	17
2.2.3	Tarefas, Técnicas e Metodologias	17
2.3	Padrões de Serviços OGC	20
2.3.1	Definições	20
2.3.2	Principais Serviços	21
2.4	Características da visualização exploratória de dados espaço-temporais	24
2.5	Resumo do Capítulo	27
3	Trabalhos Relacionados	28
3.1	Trabalhos Acadêmicos	28

3.1.1	<i>SocioScape</i>	28
3.1.2	<i>Exploratory spatio-temporal data mining and visualization</i>	31
3.1.3	<i>Spatio-Temporal Visual Query Environment</i>	34
3.1.4	<i>Visual Analytics Tools for Analysis of Movement Data</i>	37
3.1.5	<i>A Framework for Exploring Spatiotemporal Patterns</i>	38
3.1.6	<i>COPLINK: Visualization for Crime Analysis</i>	41
3.1.7	<i>BirdVis: Visualizing and Understanding Bird Populations</i>	43
3.1.8	<i>GeoVISTA CrimeViz</i>	45
3.2	Ferramentas de SIG	46
3.3	Considerações	49
3.4	Resumo do Capítulo	51
4	Uma abordagem para visualização e análise de dados espaço-temporais	52
4.1	Técnicas e Recursos de Visualização reunidos	52
4.1.1	Para a Dimensão Espacial	53
4.1.2	Para a Dimensão Temporal	54
4.2	O uso de <i>Clustering</i> para mineração de dados Espaço-Temporais	56
4.3	O GeoSTAT - <i>Geographic SpatioTemporal Analysis Tool</i>	57
4.3.1	Visão Geral	58
4.3.2	Apresentação do Ambiente	58
4.3.3	Arquitetura	60
4.3.4	Componentes	68
4.4	Comparativo com as principais abordagens existentes	83
4.5	Resumo do Capítulo	85
5	Validação da abordagem proposta com estudos de caso	87
5.1	Estudo 1: Análise de correlação espaço-temporal entre falhas em linhas de transmissão de energia elétrica e focos de incêndio	87
5.1.1	Motivação	87
5.1.2	Dados	88
5.1.3	Condução	90
5.1.4	Resultados	95

5.1.5	Conclusões	98
5.2	Estudo 2: Análise visual espaço-temporal de fotografias georeferenciadas	99
5.2.1	Motivação	99
5.2.2	Dados	100
5.2.3	Condução	100
5.2.4	Resultados e Conclusões	105
5.3	Resumo do Capítulo	105
6	Considerações Finais	107
6.1	Trabalhos Futuros	109
	Referências Bibliográficas	111

Lista de Símbolos

AJAX - *Asynchronous Javascript and XML*

CHESF - Companhia HidroElétrica do São Francisco

CPTEC - Centro de Previsão de Tempo e Estudos Climáticos

GIF - *Graphics Interchange Format*

GML - *Geography Markup Language*

GPS - *Global Positioning System*

HTML - *HyperText Markup Language*

IARIA - *International Academy, Research and Industry Association*

INPE - Instituto Nacional de Pesquisas Espaciais

ISO - *International Organization for Standardization*

JPG - *Joint Photographic Experts Group*

JSP - *Java Server Pages*

KML - *Keyhole Markup Language*

OASIS - *Organization for the Advancement of Structured Information Standards*

OCS - *OpenGIS Catalog Service*

OECD - *Organisation for Economic Co-operation and Development*

OGC - *Open Geospatial Consortium*

OWS - *OGC Web Services*

PNG - *Portable Network Graphics*

SGBD - Sistema de Gerenciamento de Banco de Dados

SIG - Sistema de Informações Geográficas

SOM - *Self-Organizing Map*

SVG - *Scalable Vector Graphics*

W3C - *World Wide Web Consortium*

WFS - *Web Feature Service*

WMS - *Web Map Service*

WRS - *Web Registry Service*

XML - *Extensible Markup Language*

Lista de Figuras

2.1	Representação de dados espaço-temporais com o método " <i>snapshot</i> " (Pessoa e Simões, 2003).	14
2.2	Representação linear e cíclica de uma série temporal de dados relacionados à saúde humana (Andrienko et al., 2010 <i>b</i>).	25
2.3	Agregação de trajetórias a partir da densidade dos movimentos no espaço-tempo (Andrienko et al., 2010 <i>b</i>).	26
3.1	Técnicas de visualização exploradas na ferramenta <i>SocioScope</i> (Reda et al., 2009)	29
3.2	<i>SocioScope</i> exibindo o movimento espaço-temporal das comunidades de zebra " <i>Grevy</i> " (esquerda) e a composição social desses grupos (direita) (Reda et al., 2009).	30
3.3	Arquitetura da solução proposta por Compieta et al. (2007).	31
3.4	Visualização de regras de associação espaço-temporal na implementação baseada em <i>Google Earth</i> (Compieta et al., 2007).	32
3.5	Esquema de interface da implementação baseada em <i>Java-3D</i> (Compieta et al., 2007).	33
3.6	Interface principal do S-TVQE (Cavalcanti et al., 2006).	35
3.7	S-TVQE apresentando resultado visual de consulta espaço-temporal (Cavalcanti et al., 2006).	35
3.8	S-TVQE apresentando gráfico temporal resultante de consulta espaço-temporal (Cavalcanti et al., 2006).	36
3.9	<i>Clusters</i> de paradas realizadas em várias viagens de carro entre o trabalho e a residência de um cidadão (Andrienko et al., 2007).	38

3.10	Possíveis aparências de células em uma matriz gerada a partir da técnica <i>Self-Organizing Map</i> (Andrienko et al., 2010a).	40
3.11	Apresentação da ferramenta <i>COPLINK Spatio-Temporal Visualizer</i> (Chen et al., 2003).	43
3.12	Interface da ferramenta <i>BirdVis</i> (Ferreira et al., 2011).	44
3.13	Ambientes e componentes da ferramenta <i>BirdVis</i> (Ferreira et al., 2011). . .	45
3.14	Interface Interativa da ferramenta <i>GeoVISTA CrimeViz</i> (Roth et al., 2010). .	46
4.1	Interface interativa do GeoSTAT.	59
4.2	Arquitetura do GeoSTAT em três camadas.	61
4.3	Módulos da camada de controle da arquitetura do GeoSTAT.	62
4.4	Diagrama de sequência com as operações realizadas no processamento de uma requisição de mineração de dados.	64
4.5	Esquema de funcionamento do Módulo de Mineração de Dados.	65
4.6	Estrutura interna da camada de persistência da arquitetura do GeoSTAT. . .	66
4.7	Esquema relacional do banco de dados espaço-temporal utilizado pelo GeoSTAT.	68
4.8	Componentes da camada de visualização do GeoSTAT.	69
4.9	Caixa de diálogo para adicionar nova camada de mapa no GeoSTAT.	71
4.10	Utilização do controlador temporal do GeoSTAT. Em a) todos os dados espaço-temporais distribuídos no mapa, e em b) apenas os dados referentes a "04/11/2010" (116 pontos).	74
4.11	Caixa de diálogo para definição de filtro temporal no GeoSTAT.	76
4.12	Caixa de diálogo para definição de filtro espacial no GeoSTAT.	77
4.13	Caixa de diálogo para execução de algoritmo de mineração de dados baseada em <i>clustering</i> no GeoSTAT.	79
4.14	Caixa de diálogo de acompanhamento da execução de mineração de dados no GeoSTAT.	80
4.15	Caixa de diálogo de visualização de resultados de mineração de dados no GeoSTAT.	80

4.16	GeoSTAT exibindo os resultados relevantes de um processo de mineração de dados executado com duas camadas espaço-temporais.	81
4.17	GeoSTAT exibindo todos os <i>clusters</i> da marcação temporal "2010-01". . . .	82
5.1	GeoSTAT exibindo as camadas de dados utilizadas neste estudo de caso. . .	91
5.2	Dois visualizações de mapa em instantes de tempo distintos: a) eventos registrados às 05h49 do dia 01/01/2002, e b) eventos registrados às 20h21 do mesmo dia.	92
5.3	Gráficos de distribuição temporal gerados automaticamente pelo GeoSTAT: a) distribuição temporal dos focos de incêndio, e b) distribuição temporal das falhas em linhas de transmissão.	93
5.4	Resultados obtidos com a execução da mineração de dados a partir do GeoSTAT.	95
5.5	GeoSTAT exibindo todos os <i>clusters</i> relevantes gerados.	96
5.6	GeoSTAT exibindo todos os <i>clusters</i> relevantes gerados para 03/11/2004, das 15 às 17 horas e 59 minutos.	97
5.7	GeoSTAT exibindo, em detalhes, o <i>cluster</i> espaço-temporal nº 97, da marcação temporal "03/11/2004 15:00 às 17:59".	98
5.8	GeoSTAT exibindo a visualização espaço-temporal da camada "FOTOS". .	101
5.9	Sequência de solicitações realizadas por usuário do GeoSTAT para realizar a consulta espacial desejada.	102
5.10	Resultados sequenciais de consulta espacial realizada por meio do GeoSTAT.	102
5.11	Resultado final da consulta espacial que retorna os municípios, que fazem limite com João Pessoa, que contém fotografias capturadas.	103
5.12	Distribuição espacial e temporal do conjunto de fotografias georeferenciadas capturadas nos últimos quatro anos.	104

Lista de Tabelas

3.1	Tabela comparativa entre os trabalhos acadêmicos de visualização espaço-temporal relacionados nesta seção.	50
3.2	Tabela comparativa entre as soluções gratuitas de SIG espacial e espaço-temporal relacionadas nesta seção.	51
4.1	Nova tabela comparativa entre os trabalhos acadêmicos de visualização espaço-temporal relacionados no Capítulo 3 e a nova abordagem proposta neste trabalho.	84

Lista de Códigos Fonte

4.1 Pseudo-código do algoritmo de execução da mineração de dados espaço-temporais no GeoSTAT	66
--	----

Capítulo 1

Introdução

Atualmente, há um volume considerável de dados espaço-temporais disponíveis em vários meios, sobretudo na Internet. Em meio a tanta informação, emerge a necessidade de sistemas de análise e suporte à decisão, que possibilitem aos usuários tomadores de decisão extrair conhecimentos relevantes, como a predição de determinados eventos futuros, por exemplo. Estes sistemas precisam ser intuitivos de tal modo que se gaste pouco tempo na obtenção de conclusões relevantes (predições, padrões de recorrência, agrupamentos, etc.) a partir do processo de análise.

Técnicas de visualização são amplamente reconhecidas por serem poderosas no domínio de suporte à decisão (Johnston, 2001), uma vez que elas se aproveitam das habilidades humanas para, rapidamente, perceber padrões visuais e interpretá-los (Kopanakis e Theodoulidis, 2003; Andrienko et al., 2003). No entanto, sabe-se que os recursos de visualização espacial fornecidos na maioria das aplicações geográficas existentes não são suficientes para sistemas de suporte à decisão, quando utilizados sozinhos (Bédard et al., 2001).

Além disso, dados espaço-temporais estabelecem sérios desafios para os analistas. Em primeiro lugar, devido à complexidade do espaço geográfico, que requer o envolvimento humano e seu senso de determinação de espaço, lugar, e dos relacionamentos espaciais (Andrienko et al., 2008). E, em segundo lugar, devido ao complexo fenômeno denominado tempo. O tempo flui de forma linear, no entanto, alguns eventos que ocorrem ao longo do tempo podem ser periodicamente recorrentes, com múltiplos ciclos, formando estruturas hierárquicas que se sobrepõem e interagem entre si. Os humanos têm um bom entendimento das propriedades do tempo, mas passar todo esse conhecimento para uma máquina é um

desafio. Os dados temporais também exigem o envolvimento humano (Peuquet, 2002).

Além das características complexas presentes em dados espaço-temporais, a existência de fontes heterogêneas recai num problema de interoperabilidade, uma vez que é cada vez mais presente a necessidade de realizar análises sobre dados oriundos, simultaneamente, de fontes de dados heterogêneas, obtidas de fontes distintas. Na tentativa de, ao menos, minimizar tal problema, o *Open Geospatial Consortium* (OGC) (OGC, 2011) desenvolveu padrões para serviços de conectividade com bancos de dados espaciais heterogêneos, a exemplo do *Web Map Service* (WMS) e do *Web Feature Service* (WFS), amplamente utilizados por soluções de Sistemas de Informações Geográficas (SIG).

1.1 Problemática

A visualização de dados espaço-temporais é uma tarefa complexa, que requer uma série de recursos visuais apropriados para que, em conjunto, possam permitir aos usuários uma correta interpretação das informações analisadas. Visualizar e analisar dados espaço-temporais é uma tarefa que vem ganhando destaque em diversas áreas, como biologia, criminalidade, transmissão de energia elétrica, tráfego urbano, construção civil, entre outras. Esse notório destaque se dá, principalmente, em razão da difusão de dispositivos de captura de localização geográfica e de Sistemas de Informações Geográficas, gerando grande quantidade de informações relacionadas ao espaço x tempo (trajetórias de objetos móveis ou ocorrência de focos de incêndio, focos de dengue, descargas atmosféricas, etc).

Além do grande volume de dados espaço-temporais para analisar, o usuário interessado em um domínio específico ainda se depara com a complexidade inerente aos dados espaciais, uma vez que relacionamentos espaciais não são explícitos se visualizados de forma tabular, prática comum na visualização de dados de qualquer natureza, diferente da visualização dos relacionamentos cartesianos, geralmente encontrados em dados convencionais (numérico ou de texto). A simples tarefa de visualizar, de forma tabular, um conjunto de transações no espaço-tempo não é suficiente para visualizar relacionamentos de sobreposição ou agrupamento espaço-temporal, por exemplo. Para auxiliar na observação de relacionamentos em dados espaço-temporais, existem algoritmos de mineração de dados. O uso do algoritmo DBScan (Ester et al., 1996), por exemplo, possibilita a identificação de grupos

espaço-temporais em função da sua densidade, apontando regiões no espaço-tempo em que se concentra grande parte da massa de dados analisada.

Unir os esforços dos algoritmos de mineração de dados espaço-temporais com modernas técnicas de visualização e usabilidade, voltadas à visualização de dados brutos e de dados derivados de um processo de mineração, parece uma estratégia de solução viável para reduzir os problemas de visualização e análise sobre dados espaço-temporais.

Segundo Andrienko et al. (2010b), uma boa solução em análise visual de dados espaço-temporais deveria reunir todos esses fatores e, ainda, oferecer baixo custo de processamento e um ambiente colaborativo, dois desafios para esta área de pesquisa.

O levantamento do estado da arte em visualização de dados espaço-temporais leva à conclusão de que a área ainda é deficiente em soluções para visualização e análise de dados espaço-temporais, deixando a desejar principalmente nos seguintes fatores:

Temporalidade: muitas soluções abordam a dimensão espacial, mas deixam de levar em consideração a temporalidade dos dados, que deveria ser tratada como mais uma dimensão;

Genericidade: muitas soluções resolvem problemas voltados a domínios específicos e pequenos grupos de usuários. Faltam soluções genéricas, funcionais e eficazes em diferentes domínios de aplicação;

Interoperabilidade: muitas soluções não oferecem suporte à interoperabilidade de bancos de dados. Quando se tem uma solução que só funciona para um único e exclusivo conjunto de dados, tem-se uma solução que só se aplica a fontes de dados específicas.

Praticidade: ausência de componentes intuitivos, que permitam sua utilização por qualquer tipo de usuário;

Flexibilidade: mecanismos que deem liberdade ao usuário, de tal forma que este visualize a informação que julgar importante e descarte o que julgar irrelevante, evitando visualizações confusas em função da grande quantidade de informações;

Mecanismos de levantamento de evidências: muitas soluções não passam de visualizadores de dados baseados em mapas. Faltam mecanismos para mineração de dados e visualização dos resultados.

Dessa forma, nos deparamos diante do seguinte problema: soluções de visualização espaço-temporal existentes são deficientes na manipulação da dimensão temporal, tornando-se meras soluções de visualização espacial. Como oferecer recursos de manipulação temporal que, somados aos recursos de manipulação sobre dados espaciais, possam melhorar a experiência dos usuários finais, interessados em realizar análises visuais sobre dados de natureza espaço-temporal?

1.2 Objetivo do Trabalho

Diante do cenário apresentado, este trabalho tem como objetivo propor uma nova abordagem para visualização e análise de dados espaço-temporais, que leve em consideração questões de temporalidade, genericidade, interoperabilidade, praticidade, flexibilidade, e ainda ofereça mecanismos para levantamento de evidências.

Para alcançar esse objetivo, pretende-se alcançar os seguintes objetivos específicos:

- Reunir as principais técnicas de visualização de dados espaço-temporais, abrangendo não somente a dimensão espacial como também a dimensão temporal, e oferecendo um tratamento genérico, independente do domínio de natureza dos dados;
- Incorporar algoritmos de mineração de dados baseados em *clustering* espaço-temporal em um ambiente de visualização de dados espaço-temporais;
- Utilizar os padrões de serviço de compartilhamento de dados espaciais para prover interoperabilidade;
- Melhorar a experiência do usuário em visualização exploratória espaço-temporal, indo além do universo da visualização dos dados espaço-temporais brutos e, considerando também, a importância na visualização de dados espaço-temporais derivados de um processo de descoberta de conhecimento.

1.3 Relevância e Contribuições

A principal contribuição deste trabalho está na definição de uma abordagem inovadora no domínio da visualização de dados espaço-temporais, partindo de pontos fortes observados

nas principais abordagens existentes e incluindo técnicas voltadas, principalmente, à dimensão temporal e a mecanismos de levantamento e visualização de evidências, valorizando características espaço-temporais até então pouco exploradas.

Além de integrar algoritmos de mineração de dados para promover mineração de dados espaço-temporal semiautomatizada, a nova abordagem proposta é implementada, surgindo como uma nova solução para usuários interessados em visualizar e analisar dados dessa natureza, independente do domínio aplicado.

1.3.1 Publicações Realizadas

Durante o desenvolvimento deste trabalho, algumas publicações relacionadas foram realizadas em eventos nacionais e internacionais da área de geoprocessamento. Essas publicações são listadas a seguir:

GeoINFO 2011 (Baptista et al., 2011): No '*XII Brazilian Symposium on GeoInformatics*', organizado pelo INPE (Instituto Nacional de Pesquisas Espaciais) e sediado em Campos do Jordão, no interior de São Paulo, uma versão inicial da implementação da abordagem proposta neste trabalho foi utilizada como ambiente de visualização de dados espaço-temporais fornecidos a partir de um servidor que implementa os padrões de serviço OGC para compartilhamento de bancos de dados espaciais *NoSQL*, apresentado no artigo intitulado "*Using OGC Services To Interoperate Spatial Data Stored In SQL And NoSQL Databases*". Na ocasião, o GeoSTAT possibilitou a realização de testes de funcionalidade do servidor de compartilhamento proposto;

GEOProcessing 2012 (Oliveira et al., 2012b): No '*4th International Conference on Advanced Geographic Information Systems, Applications, and Services*', organizado pela IARIA (*International Academy, Research and Industry Association*) e sediado em Valência, na Espanha, a implementação da abordagem proposta neste trabalho é apresentada como uma nova solução para visualização e análise de dados espaço-temporais, com um artigo intitulado "*A Web-based Environment for Analysis and Visualization of Spatio-temporal Data provided by OGC Services*". Um estudo de caso é realizado para validar a proposta e enfatizar os pontos de melhoria na experiência do usuário final;

ENGEO 2012 (Oliveira et al., 2012a): No *'II Encontro Nacional de Geoprocessamento no Setor Elétrico'*, organizado pela Eletrobrás/Chesf (Companhia HidroElétrica do São Francisco) e sediado em Recife, Pernambuco, o artigo intitulado "Utilizando Geoprocessamento e Mineração de Dados em Busca de Correlação entre Ocorrências de Focos de Incêndio e Desligamentos em Linhas de Transmissão" apresenta um estudo de caso realizado através da implementação da abordagem proposta, com duas pequenas bases de dados espaço-temporais reais obtidas junto à Eletrobrás/Chesf e ao INPE, com o objetivo de visualizar a correlação espaço-temporal entre eventos de falhas em linhas de transmissão de energia elétrica e focos de incêndio detectados no estado de Pernambuco no ano de 2010. O estudo chamou a atenção dos gestores da empresa, que lançaram vários questionamentos sobre as facilidades oferecidas pelo GeoSTAT.

Além dessas três publicações já realizadas, outras duas estão sendo elaboradas para serem submetidas ainda em 2012.

1.4 Estrutura da Dissertação

O restante deste trabalho está organizado da seguinte forma:

- No *Capítulo 2*, é apresentada uma fundamentação teórica abordando a natureza dos dados espaciais, temporais e espaço-temporais, apresentando conceitos e técnicas de mineração de dados espaço-temporais, e os padrões de compartilhamento de dados espaciais desenvolvidos para promover interoperabilidade entre SIGs;
- No *Capítulo 3*, são apresentados os trabalhos relacionados. São relacionados os principais trabalhos científicos e ferramentas gratuitas de SIG, permitindo a realização de uma avaliação comparativa que expõe pontos fortes e fracos de cada solução;
- No *Capítulo 4*, é apresentada a nova abordagem para visualização e análise de dados espaço-temporais, descrevendo as técnicas de visualização e os recursos de manipulação de dados espaço-temporais reunidos. É apresentado o GeoSTAT, uma implementação da nova abordagem proposta, descrevendo detalhadamente sua arquitetura e todos os seus componentes. Por fim, é realizada uma avaliação comparativa com os trabalhos relacionados;

- No *Capítulo 5*, são apresentados dois estudos de caso desenvolvidos para avaliar e validar a nova abordagem proposta;
- No *Capítulo 6*, são apresentadas as considerações finais deste trabalho e as discussões sobre propostas para trabalhos futuros.

Capítulo 2

Fundamentação Teórica

Neste capítulo, são abordados os conteúdos relevantes e indispensáveis ao perfeito entendimento deste documento, sendo composto por uma revisão de literatura abordando dados espaciais, temporais e espaço-temporais, mineração de dados espaço-temporais, e os padrões para serviços de compartilhamento de dados espaciais especificados pelo OGC.

2.1 Dados Espaciais, Temporais e Espaço-Temporais

Esta seção apresenta três tipos de dados: espaciais, temporais e espaço-temporais; e aborda as principais características que diferenciam estes dados dos dados convencionais e que, portanto, requerem um tratamento diferenciado.

2.1.1 Dados Espaciais

Dados espaciais são aqueles que apresentam, além das estruturas e relacionamentos encontrados nos tipos de dados tradicionais (numéricos ou texto), informações relacionadas implicitamente com um espaço pré-determinado. Essas informações são responsáveis pela determinação de localizações e de topologias espaciais.

Um exemplo de espaço pré-determinado é o espaço geográfico. Os dados geográficos constituem uma especialização de dados espaciais. Doravante, nesta dissertação, a utilização do termo "dado espacial" remeterá ao tipo de dado geográfico.

Um banco de dados espacial geralmente é composto por atributos espaciais e não-

espaciais, onde os atributos espaciais especificam a forma (pontos, linhas ou polígonos) e a localização espacial (latitude, longitude e elevação) dos dados, enquanto que os atributos não-espaciais caracterizam essa localização especificada com informações de interesse, como, por exemplo, país, população, município, precipitação média, tipo de vegetação, tipo de objeto (rodovia, edificação, árvore, etc), entre outras. Dessa forma, um banco de dados espacial nada mais é do que um banco de dados tradicional estendido para a dimensão espacial.

Para Shekhar et al. (2008), os conjuntos de dados espaciais são representações discretas de fenômenos contínuos. Tal discretização do espaço é necessária em razão da natureza da representação digital, onde tudo se torna finito. Existem dois modelos básicos de representação dos dados espaciais: *raster* e vetorial. Os dados *raster* são constituídos por imagens georeferenciadas que carregam consigo a descrição de cada um de seus *pixels* e a relação que eles mantêm com o espaço geográfico. Um bom exemplo de dados *raster* são as imagens de satélite georeferenciadas, onde cada *pixel* representa uma região delimitada do espaço geográfico. Os dados vetoriais são constituídos por pontos, linhas ou polígonos, informações geométricas vetoriais, baseadas no sistema de coordenadas geográficas, que usam a álgebra vetorial para uma representação mais precisa do espaço geográfico.

Em se tratando do aspecto de armazenamento, os relacionamentos encontrados em dados não-espaciais são geralmente explícitos, ou seja, basicamente relacionam-se dados não-espaciais através de atributos do tipo numérico, compondo relações aritméticas, ou relacionam-se esses dados através de atributos do tipo alfanumérico (texto), possibilitando relacionamentos categóricos. Por outro lado, também no aspecto de armazenamento, relacionamentos em dados espaciais são geralmente implícitos, a exemplo da sobreposição (ex: um polígono pode estar sobrepondo toda área espacial de outro polígono ou ponto), intersecção (ex: um ponto pode intersectar espacialmente um determinado polígono se estiver na mesma área preenchida por ele) e orientação (ex: um polígono pode estar ao norte ou à noroeste de outro), e, portanto, não são perceptíveis quando visualizados a partir de estruturas tabulares, uma visão que simula a forma de armazenamento dos conjuntos de dados.

Uma característica interessante em dados espaciais é que eles podem ser correlacionados, já que são oriundos de uma estrutura espacial comum. Um dos pressupostos fundamentais da análise estatística, é que as amostras de dados são geradas de maneira independente,

como jogadas sucessivas de uma moeda ou de um dado. Entretanto, com dados espaciais, esse pressuposto que assume a independência das amostras de dados geralmente não é válido (Shekhar et al., 2008). Os dados espaciais tendem a serem correlacionados, e essa correlação é inerente ao espaço compartilhado pelos dados, por exemplo:

1. Pessoas com características sociais semelhantes, como a renda e a ocupação, tendem a se agrupar na mesma vizinhança, ou no mesmo bairro;
2. Mudanças em variáveis do meio ambiente, como o clima e a temperatura, ocorrem gradualmente no espaço.

Essa propriedade dos dados espaciais é conhecida na estatística espacial, uma área destinada à análise de dados espaciais, como **autocorrelação espacial**.

Outra importante característica dos dados espaciais é a heterogeneidade espacial, uma vez que os dados não são distribuídos no espaço de forma estritamente uniforme. Um bom exemplo de heterogeneidade espacial está no domínio de população: é sabido que a distribuição espacial da população ao longo de todo o planeta não é idêntica sob vários parâmetros, como espécie, classe social e até mesmo clima. Regiões com clima de frio extremo costumam agrupar muito menos pessoas que aquelas com climas mais amenos. É possível perceber que as propriedades dos dados espaciais são dependentes da localização (Shekhar et al., 2003).

Para Miller (1999), a heterogeneidade espacial é medida através de medidas locais de autocorrelação espacial. Dessa forma, é possível saber se os grupos têm dimensões semelhantes e quais regiões no espaço registram maior concentração.

Shekhar et al. (2008) levantaram o problema dos dados espaciais apresentarem relacionamentos implícitos e apontaram, como uma possível solução, a materialização desses relacionamentos em atributos tradicionais, de tal modo que eles sejam facilmente localizados e entendidos.

Em razão das estruturas e relacionamentos complexos encontradas em dados espaciais, os Sistemas de Gerenciamento de Bancos de Dados (SGBD) espaciais disponibilizam diversas funções de suporte a essas informações, a exemplo de métodos que permitem calcular a distância espacial entre dois objetos, ou criar um objeto mais complexo a partir da união de

diversos pontos espalhados no espaço, ou ainda, exibir todos os objetos vizinhos espacialmente de um determinado objeto.

Os SGBDs *Oracle*¹ e *PostgreSQL*² são exemplos de Sistemas de Gerenciamento de Bancos de Dados que dão suporte à dados espaciais. Este último provê o suporte espacial através da extensão *PostGIS*³, que além de permitir o armazenamento e a manipulação de dados espaciais, fornece uma série de métodos de suporte para tratamento específico desse tipo de dado.

2.1.2 Dados Temporais

Muitas das aplicações de *software* tem necessidade de manipular, de alguma forma, informações históricas - dados relacionados a estados passados da aplicação (Jensen e Snodgrass, 1999). Dados armazenados por um SGBD tradicional normalmente referem-se à informações válidas em um tempo presente, como o momento em que uma transação aconteceu e foi armazenada num banco de dados, por exemplo. Quando os dados são modificados, removidos ou inseridos, o estado do banco de dados é sobrescrito por um novo estado.

Analisar as características do tempo acarreta a divisão dessa dimensão em três partes essenciais: passado, presente e futuro. Os dados passados consistem em informações que foram armazenadas e permaneceram válidas durante um período de tempo no passado, não sendo mais válidas no presente ou no futuro, embora permaneçam armazenadas. Os dados futuros referem-se à informações armazenadas no presente ou no passado, mas que só serão válidas em um tempo no futuro, ou seja, que só serão verdadeiras em um futuro próximo (Patel, 2003).

Dada a importância de abordar as três partes da dimensão temporal, surgiu o conceito de dado temporal, ou seja, dados que não contêm apenas informações instantâneas, no presente, como também podem conter informações do passado ou do futuro, e, juntamente com esse conceito, o tratamento da temporalidade dos dados.

Os bancos de dados temporais associam aos dados instantes ou períodos de tempo. Eles

¹SGBD *Oracle*. Mais informações em (inglês):

<http://www.oracle.com/us/products/database/options/spatial/index.htm>

²SGBD *PostgreSQL*. Mais informações em (inglês): <http://www.postgresql.org/>

³Extensão espacial *PostGIS*. Mais informações em (inglês): <http://postgis.refrains.net/>

oferecem tipos de dados temporais e podem armazenar informações relacionadas com o passado, presente e futuro. Dessa forma, um banco de dados temporal armazena uma coleção de dados relacionados com o tempo (Patel, 2003). Tal característica requer maior espaço de armazenamento, uma vez que, geralmente, em bancos de dados temporais, os dados não são fisicamente excluídos.

Em se tratando do aspecto de armazenamento, os relacionamentos temporais podem ser considerados como parcialmente explícitos, uma vez que, assim como acontece com dados tradicionais, alguns relacionamentos temporais podem ser visualizados a partir de relações aritméticas. É possível utilizar operadores de igualdade ou de diferença (como maior ou menor) para saber se um instante ou período de tempo antecede ou sucede outro instante ou período. Da mesma forma, a simples tarefa de visualizar dados temporais em uma estrutura tabular permite a identificação de alguns relacionamentos temporais, como antes, durante e depois. Entretanto, dados temporais também podem apresentar relacionamentos implícitos. Um fato que aconteceu em um período de tempo do ano atual, por exemplo, pode estar relacionado com outros fatos de mesma natureza que aconteceram em períodos de mesmo mês e duração, em anos anteriores.

A diferença entre dados temporais e não-temporais é que um período de tempo é adicionado aos dados com o objetivo de especificar quando esses dados são válidos ou armazenados em um banco de dados (Patel, 2003). Associar tempo com os dados possibilita o armazenamento concomitante de diferentes estados do banco de dados, que podem ser restaurados a qualquer tempo.

No universo dos dados temporais, inúmeros conceitos têm sido definidos. Dois conceitos importantes, na visão de Jensen e Snodgrass (1999), são:

Tempo de Validade: tempo em que um fato do banco de dados é verdadeiro na realidade modelada. Esse tempo é definido, geralmente, pelo próprio usuário.

Tempo de Transação: tempo durante o qual um fato é conhecido no banco de dados e pode ser recuperado.

Existem os bancos de dados temporais de tempo de validade e também os bancos de dados temporais de tempo de transação. Jensen e Snodgrass (1999) definem ainda o con-

ceito de banco de dados bitemporal, como sendo aquele em que seus dados podem possuir, simultaneamente, tempo de transação e tempo de validade.

Os dados temporais obedecem uma ordem no tempo. Essa ordem pode ser linear, ramificada ou circular. A ordem linear é a mais usual, onde cada dado tem, no máximo, um sucessor e um predecessor. Na ordem ramificada, cada dado pode ter vários sucessores ou vários predecessores, como, por exemplo, alternativas para o último capítulo de uma novela, onde o penúltimo capítulo, nesse caso, tem vários sucessores. Já na ordem circular, um conjunto de dados se repete após o decorrer de um determinado período, de forma ordenada. Um exemplo de ordem circular está nos períodos de promoção no comércio, como verão, dia das mães, dia dos namorados e natal, que se repetem anualmente.

A **granularidade temporal** é outro importante aspecto a ser considerado nesse tipo de dado. Também conhecida por resolução temporal ou *Chronon*, é a unidade de tempo não decomponível de um dado (Patel, 2003). São exemplos de granularidade temporal: mês, ano, semestre, dia e hora. Ela é dita não-decomponível em razão de descartar informações necessárias a uma decomposição. Por exemplo, não é possível identificar o mês e o dia de um dado temporal com granularidade ano, simplesmente porque essa informação não está disponível nesse dado.

A principal diferença existente entre dados temporais e outros tipos de dados está no fato dos dados temporais não poderem ser modificados pelo usuário, já que estão sujeitos a uma semântica específica.

2.1.3 Dados Espaço-Temporais

Os dados espaço-temporais podem ser definidos como dados que apresentam, de forma simultânea, as características vistas tanto em dados espaciais como em dados temporais. Geralmente, descrevem um fenômeno geoespacial e, portanto, são dotados de informações sobre a localização espacial desse fenômeno e o momento em que ocorreu, ou o período em que esteve acontecendo.

Erwig (2002) define dados espaço-temporais como um caso particular de dados temporais em geral, onde tal particularidade seria a presença da dimensão espacial. Em Geerts et al. (2002), os autores definem os bancos de dados espaço-temporais como conjuntos de figuras n -dimensionais móveis, descritas por meio de um conjunto de tuplas do tipo $(x_1, x_2, \dots, x_n; t)$

em $R^n \times R$, onde R é o conjunto dos números reais, (x_1, x_2, \dots, x_n) representa as coordenadas espaciais de um ponto em um espaço n -dimensional R^n , e t é a coordenada do tempo em R . Já em Wang et al. (2000), os autores definem bancos de dados espaço-temporais como bases de dados que incorporam conceitos espaciais, temporais e espaço-temporais, e capturam, simultaneamente, os aspectos espaciais e temporais dos dados.

Trabalhar com dados espaço-temporais no nível de armazenamento, ou seja, a partir de representações geométricas (pontos, linhas e polígonos) para dados espaciais com marcações de tempo é, geralmente, indesejável, dada a complexidade e particularidades das estruturas de dados envolvidas. Dessa forma, transformações são necessárias para se descrever as unidades de análise em altos níveis conceituais, onde relações e propriedades facilmente interpretáveis por humanos são expressas (Andrienko et al., 2006).

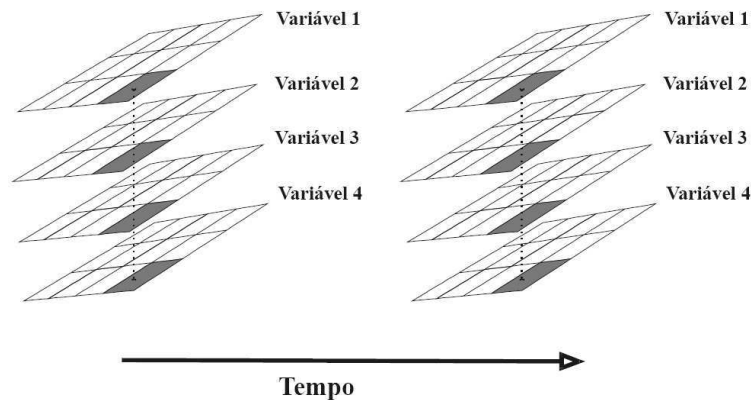


Figura 2.1: Representação de dados espaço-temporais com o método "snapshot" (Pessoa e Simões, 2003).

Uma das formas de visualizar dados espaço-temporais de forma mais "amigável" é através do método "snapshot", exemplificado na Figura 2.1, onde os atributos não-espaciais e não-temporais, tratados na figura como variáveis, são dispostos em mapas, onde cada mapa apresenta um estado válido de um atributo para uma marcação temporal, e então, a partir do valor do atributo para o tempo t e posição espacial x , cada registro de uma tabela de dados espaço-temporais é montado (Pessoa e Simões, 2003). A marcação temporal pode estar representando um instante ou período de tempo. Se for realizada uma animação manual, alternando a visualização dos mapas e obedecendo a ordem sequencial crescente da variação temporal de cada um deles, será possível perceber a evolução de um fenômeno ao longo do

tempo.

Além da necessidade de transformações nos dados, outros problemas merecem atenção quando se trabalha com dados espaço-temporais, como a resolução espacial e a granularidade temporal, fatores que podem ter impactos diretos nos padrões descobertos em bancos de dados espaço-temporais (Andrienko et al., 2006).

A resolução espacial está relacionada com o nível de detalhes que se deseja observar em uma região. Quanto maior a resolução espacial, melhor a visualização de objetos pequenos (ruas, casas, pessoas caminhando, etc.), do ponto de vista espacial.

A granularidade temporal está relacionada com o período de tempo observado, que pode partir de uma marcação ou período de tempo mais preciso, registrando-se data, hora, minutos, segundos e milésimos, até uma marcação ou período de tempo mais genérico, como anos, décadas ou séculos.

Pode ser desejável e mais interessante analisar os dados em níveis de baixa resolução, tanto temporal como espacial, como também pode ser útil e mais interessante analisar os mesmos dados em resoluções mais elevadas. Tais possibilidades recaem na necessidade dos usuários realizarem um estudo mais detalhado sobre o fenômeno analisado a partir de dados espaço-temporais, de tal forma que se determine níveis de resolução apropriados, não desprezando informações importantes e nem utilizando informações em excesso, o que pode vir a dificultar o processo de análise.

2.2 Mineração de Dados Espaço-Temporal

Nesta seção, será abordado o processo de Mineração de Dados Espaço-Temporal, dando ênfase em sua relevância, utilidade, e em suas principais tarefas, técnicas e metodologias.

2.2.1 Motivação, Definições e Problemas

A mineração de dados espaço-temporais é uma fase do processo de descoberta de conhecimento em bancos de dados espaço-temporais. Sua principal diferença, em relação à mineração de dados tradicional, está na necessidade de levar em consideração tanto as características do espaço quanto as características do tempo, associadas aos dados analisados, o que

torna os algoritmos de mineração de dados tradicionais ineficazes quando trabalhados com dados espaço-temporais, sem nenhuma adaptação para tal finalidade.

A mineração de dados espaço-temporal tem utilidade, por exemplo, quando se deseja verificar regularidades ou irregularidades em trajetórias de objetos móveis, como aparelhos celulares ou veículos rastreados por satélite, ou analisar áreas e períodos em que se concentra considerável quantidade de focos de incêndio, ou ainda, analisar ruas ou avenidas de um município e períodos em que são registrados grandes congestionamentos. Tais exemplos deixam clara a necessidade de levar em conta duas dimensões nesse tipo de dado: o espaço e o tempo.

De forma semelhante à definição de mineração de dados espaciais, a mineração de dados espaço-temporal refere-se à extração de conhecimento implícito, seja por meio de propriedades espaciais e temporais, ou por outros padrões não armazenados explicitamente no banco de dados (Koperski et al., 1998). Ambas dimensões acrescentam complexidade à tarefa de mineração de dados.

Um desafio crucial em mineração de dados espaço-temporais é o desenvolvimento de métodos eficientes, devido ao grande volume de dados e de toda a complexidade encontrada tanto na representação desses dados, quanto na estrutura dos dados espaciais e temporais (Yao, 2003). As relações espaciais, tanto métricas (como a distância) quanto não-métricas (como a topologia e a direção), e as relações temporais (como antes e depois), podem conter informações relevantes, e, portanto, precisam ser levadas em consideração pelos métodos de mineração de dados, ao tratarem dados espaço-temporais. Essas relações devem ser extraídas dos dados e para isso existe um *trade-off* entre "pré-computá-las" antes do processo de mineração iniciar (uma abordagem denominada ansiosa), e computá-las *on-the-fly*, ou seja, quando forem realmente necessárias (uma abordagem denominada preguiçosa ou *lazy*) (Andrienko et al., 2006).

Atualmente, a mineração de dados espaço-temporal é uma área de pesquisa emergente, dedicada ao desenvolvimento e aplicação de novas técnicas computacionais para análise de grandes bancos de dados espaço-temporais (Andrienko et al., 2006). Pesquisadores têm focado na mineração de dados espaciais assumindo uma visão estática dos fenômenos geográficos e, dessa forma, considerando apenas essa dimensão. No entanto, como todo fenômeno geográfico sofre alterações com o passar do tempo, tanto a "espacialidade" quanto a "tem-

poralidade" podem ser essenciais para o entendimento de processos e eventos geográficos (Yao, 2003).

A exploração da mineração de dados espaciais (Koperski et al., 1998) (Miller e Han, 2001) e da mineração de dados temporais (Roddick e Spiliopoulou, 2002) tem recebido muita atenção nas comunidades de pesquisa em mineração de dados e descoberta de conhecimento em bancos de dados.

Segundo Kamp et al. (1998), a análise de dados multidimensionais, como espaciais, temporais e estatísticos, ocorre em muitas aplicações científicas e deve ser considerada pelas ferramentas de bancos de dados mais modernas.

2.2.2 Fatores a Considerar

Para que apresente bons resultados, o processo de mineração de dados espaço-temporal precisa levar em consideração os seguintes fatores (Han, 2003):

1. **Escalabilidade**, por trabalhar com grandes volumes de dados;
2. **Eficiência** para oferecer baixo tempo de resposta, pois, em alguns casos, predições podem se confirmar antes de serem descobertas;
3. **Cumprimento das restrições espaciais e temporais**, de modo que as conclusões possam fazer sentido com o mundo real;
4. **Aproximação geográfica e temporal** das respostas, que podem propiciar maior precisão e utilidade das informações obtidas.

2.2.3 Tarefas, Técnicas e Metodologias

As tarefas e técnicas aplicáveis à mineração de dados espaço-temporal assemelham-se às tarefas e técnicas utilizadas para a mineração de dados espacial. A diferença está na necessidade de abordar também a dimensão temporal, cabendo assim uma extensão.

Com esse objetivo, Yao (2003) apontou as necessidades de adaptação temporal às técnicas e tarefas da mineração de dados espacial, de tal forma que tais técnicas possam funcionar também com dados espaço-temporais.

De forma semelhante, porém um pouco mais aprofundada, Han (2003) enumerou as tarefas e técnicas consideradas essenciais para uma boa mineração de dados espaço-temporal, tomando por base a mineração de dados espacial. Essas tarefas e técnicas são brevemente descritas a seguir.

Análise Multidimensional: Quando se trabalha com dados espaço-temporais surge imediatamente a ideia de duas dimensões: uma representando o espaço e a outra representando o tempo. Dessa forma, a melhor maneira de representar esses dados e realizar análise sobre eles pode ser baseada em um modelo de dados multidimensional, um *Data Warehouse*.

Minerando a dinâmica dos *streams* de dados: Uma vez que *streams* de dados representam grandes volumes de informação, por se tratarem de transações ocorridas ao longo do tempo, e onde muitos registros podem ser obtidos em frações de segundos, pode ser necessário realizar uma mineração inicial para consolidar o banco de dados transacional e, conseqüentemente, reduzir a quantidade de registros. Para Han (2003), uma saída seria o desenvolvimento de métodos que permitissem uma mineração rápida e *on-line* de *streams* de dados, ou seja, à medida que as transações acontecem, elas já são mineradas com transações passadas, armazenadas no banco de dados.

Classificação: Semelhante ao que acontece na mineração de dados espaciais, onde um banco de dados com transações passadas é treinado para a construção de um modelo de predição, com as condições em que novas transações poderão ocorrer, a técnica de classificação precisa de extensões que permitam tratar também a temporalidade, de modo que as condições do modelo de predição gerado possam abordar, simultaneamente, **onde** e **quando** novas transações poderão ocorrer.

Descoberta de Tendências: Também semelhante ao que acontece na mineração de dados espaciais, com o acréscimo da mineração de sequências temporais (Yao, 2003).

Agrupamento (*clustering*): Técnica muito popular com dados espaciais, também pode ser utilizada para dados espaço-temporais. Para isso, Han (2003) afirma que é mais interessante agrupar os dados colocando-se a dimensão temporal como um fator principal,

ou seja, agrupando-se inicialmente os dados sob o ponto de vista temporal. Após os agrupamentos temporais, seria realizado o agrupamento espacial.

Análise de *Outliers*: Técnica aplicada de forma semelhante àquela aplicada em mineração de dados espaciais, acrescentando-se apenas a necessidade de observar a dimensão temporal, podendo assim detectar objetos móveis estranhos com base não só na localização espacial como também na direção, na velocidade ou em períodos de tempo, por exemplo.

Padrões de recorrência e análise de correlação: Uma tarefa interessante em mineração de dados espaço-temporal é a tarefa de detecção de padrões de recorrência e correlação em objetos móveis. Com ela é possível descobrir objetos, por exemplo, que compartilharam das mesmas mudanças de direção, velocidade ou localização espaço-temporal.

Descoberta de padrões sequenciais para objetos móveis e ambientes dinâmicos: Esta tarefa preocupa-se com a análise de objetos móveis ou ambientes dinâmicos que apresentam mudanças espaciais semelhantes e em sequência, com o decorrer do tempo. Pode ser adotada como uma tarefa precedente à análise de *outliers* ou descoberta de tendências.

Além dessas tarefas e técnicas aplicáveis à mineração de dados espaço-temporal, Han (2003) propôs três metodologias genéricas que podem ser úteis para uma mineração de dados espaço-temporal mais eficiente e eficaz: **Aprofundamento progressivo**, onde a mineração primeiramente estima padrões em potencial, encontra *hot-spots* interessantes, e, assim, facilita a análise detalhada apenas em porções de dados que despertarem interesse após essa mineração inicial; **Modelagem de janela de tempo inclinada**, onde o objetivo é reduzir a quantidade de dados analisados, gerando registros consolidados de histórico de dados ao longo de um período de tempo, com diferentes níveis de resolução; e **Micro-clustering como técnica de pré-processamento**, diminuindo a quantidade de pontos a serem analisados na mineração de dados, uma vez que eles passarão por um agrupamento inicial, baseado na proximidade espaço-temporal dos dados, e esse agrupamento permitirá descartar dados que não trarão resultados relevantes.

2.3 Padrões de Serviços OGC

Nesta seção, são abordados os padrões de serviços de compartilhamento de dados espaciais e espaço-temporais na Internet, estabelecidos pelo *Open Geospatial Consortium* (OGC).

2.3.1 Definições

O *Open Geospatial Consortium* (Consórcio Geoespacial Aberto - OGC) (OGC, 2011) é uma organização internacional que desenvolve especificações para integrar dados geográficos e recursos de geoprocessamento. Seus esforços estão em prover protocolos padronizados para acessar informações e serviços espaciais na Internet, e prover também padrões para metadados de serviço (Aktas et al., 2005).

O principal objetivo do OGC é viabilizar a interoperabilidade entre sistemas que envolvam dados espaciais. Com o estabelecimento de padrões de acesso às informações, qualquer SIG que implemente tais especificações poderá acessar e/ou disponibilizar qualquer fonte de dados espaciais. Com esse objetivo, o OGC propôs uma arquitetura para distribuição de dados e funcionalidades geográficas na Internet, liderando o processo de padronização de formatos de dados, métodos e especificações de interfaces. Essa arquitetura é chamada de *OpenGIS Services Framework*, ou, de maneira simplificada, *OpenGIS*.

O OGC utiliza esforços existentes de outras organizações de padrões, como o W3C (*World Wide Web Consortium*), a ISO (*International Organization of Standardization*), e a OASIS (*Organization for the Advancement of Structured Information Standards*), para dar sua contribuição no domínio de dados espaciais (Kralidis, 2008).

O OGC propõe o uso de serviços baseados em catálogos, para a implementação das operações de localização, publicação e consulta, detendo-se de uma interface própria para as requisições aos serviços, sem utilização de descritores de serviço, e, conseqüentemente, impondo restrições quanto à indexação e busca desses serviços (OGC, 2011).

O OGC introduz especificações de catálogo (*OpenGIS Catalog Service*, ou OCS), que definem a descoberta e a recuperação de metadados que descrevem dados espaciais e serviços de geoprocessamento, e especificações de registro de serviço *web* (*Web Registry Service*, ou WRS) (Aktas et al., 2005), que definem a implementação das especificações de catálogo. As especificações WRS definem padrões para descoberta e publicação de serviços de dados

espaciais, e apresentam mecanismos que tornam estes serviços capazes de operar com dados espaciais de domínios específicos.

Para especificar os objetos de dados espaciais oferecidos por seus serviços, o OGC desenvolveu a *Geographic Markup Language* (GML), uma linguagem de marcação baseada em XML, voltada exclusivamente para dados geográficos, modelando características espaciais e relacionamentos topológicos entre elas.

Além disso, a versão 2.2 da linguagem KML⁴ (*Keyhole Markup Language*) foi adotada como um padrão OGC em abril de 2008 (Kralidis, 2008). KML é uma linguagem baseada em XML, que ganhou uma grande popularidade e vem sendo utilizada por muitas aplicações da *web*. Ela oferece suporte a dados de estilo (forma de exibição dos objetos espaciais) e à referência espacial (localização dos objetos) em um único documento. Essa linguagem permite a visualização de dados espaciais em aplicações como o *Google Earth*, *Google Maps* e *Google Maps for Mobile*.

2.3.2 Principais Serviços

Alguns dos principais serviços OGC, genericamente conhecidos como *OGC Web Services* (OWS), são apresentados a seguir:

WMS (*Web Map Service*): serviço de mapas para a *web*. Fornece mapas renderizados em diversos formatos de imagem (*raster*) ou texto (vetorial), para que sejam visualizados em aplicações de SIG. É um serviço focado para apresentação visual dos dados espaciais;

WTS (*Web Terrain Service*): serviço de especificação de terreno para a *web*. É um serviço similar ao WMS, com a particularidade de ser destinado à visualização tridimensional de superfícies;

WFS (*Web Feature Service*): serviço de feições para a *web*. Este serviço oferece uma interface que possibilita a realização de consultas aos dados espaciais e, também, operações de inserção, atualização e remoção de feições;

⁴Documentação da linguagem KML. Disponível em (inglês):
<https://developers.google.com/kml/documentation/>

WCS (*Web Coverage Service*): serviço de cobertura para a *web*. O termo "cobertura" normalmente faz referência ao conteúdo, que pode ser imagens de satélite, fotos aéreas digitais, dados digitais de elevação, e outros fenômenos representados por valores observados em cada ponto de medição. É um serviço que, em vez de retornar somente mapas, simples imagens renderizadas, como no caso do WMS, provê os dados espaciais em conjunto com suas descrições detalhadas, em formato de imagens georeferenciadas, que preservam a semântica original dos dados.

Dentre os serviços OGC apresentados, os mais utilizados são o WMS e o WFS. Se utilizados em conjunto, fornecem todas as operações necessárias para manipulação de dados espaciais bidimensionais. A seguir, são detalhados os serviços WMS e WFS, descrevendo suas operações e principais funções.

Web Map Service (WMS)

O serviço de mapa para a *web* surgiu em 1999 como a primeira maior especificação do OGC, com o objetivo de viabilizar o acesso às informações espaciais. A versão 1.3.0, de março de 2006 (de la Beaujardiere, 2006), define três operações: *GetCapabilities*, *GetMap* e *GetFeatureInfo*. Essas operações são apresentadas a seguir:

GetCapabilities: uma operação obrigatória na implementação do WMS. É responsável por fornecer as informações detalhadas de *metadados* do serviço, ou seja, informações sobre formatos de saída aceitáveis pelas demais operações do WMS, sobre os sistemas de referência de coordenadas espaciais suportados, e também sobre as camadas espaciais disponibilizadas pelo serviço;

GetMap: outra operação obrigatória na implementação do WMS. É a operação responsável por retornar mapas de uma ou mais camadas espaciais disponibilizadas, em diversas opções de formatos de saída, baseados tanto em imagens (JPG, PNG, GIF), gerando mapas rasterizados, como em texto (SVG, GML, KML), gerando mapas vetoriais;

GetFeatureInfo: uma operação opcional. Tal operação é responsável por fornecer aos clientes do serviço WMS mais informações sobre as feições dos mapas retornados pela operação *GetMap*. Funciona, basicamente, para retornar os dados não-espaciais disponíveis para uma determinada camada, em uma região específica do mapa.

Web Feature Service (WFS)

A especificação do serviço de feições para a *web* teve sua primeira versão em 2002. Ela define interfaces para descrever operações de manipulação de dados espaciais. Sua versão 2.0.0, de novembro de 2010 (Vretanos, 2010), define seis operações: *GetCapabilities*, *DescribeFeatureType*, *GetFeature*, *GetPropertyValue*, *GetFeatureWithLock* e *Transaction*. Essas operações são apresentadas a seguir:

GetCapabilities: assim como no WMS, uma operação obrigatória para o WFS. De maneira semelhante à operação especificada para o WMS, é a responsável por fornecer informações detalhadas sobre *metadados* do serviço, sobre parâmetros e formatos de saída aceitáveis pelas demais operações do WFS, além da listagem de todas as camadas disponibilizadas, e ainda, os tipos de operadores e funções de manipulação e transformação de dados aceitáveis pelo serviço;

DescribeFeatureType: uma operação obrigatória para o WFS. Retorna uma descrição contendo o esquema de dados (atributos e tipos) de uma camada disponibilizada pelo serviço. O esquema de dados descrito define como o serviço WFS espera receber novas instâncias de dados da camada (texto, numérico, data, etc) a partir de comandos de inserção ou atualização de dados, além de definir também como as instâncias de dados da camada serão codificadas nas operações de consulta ao serviço;

GetFeature: mais uma operação obrigatória para a implementação do WFS. Retorna uma camada ou um conjunto de camadas, e possibilita a realização de consultas espaciais. O serviço WFS processa a requisição da consulta e retorna um documento de resposta para o cliente contendo as instâncias que satisfazem os critérios de consulta especificados na requisição;

GetPropertyValue: uma operação opcional para o WFS. Operação de consulta utilizada para retornar apenas os valores de uma propriedade (ou atributo) de uma camada, em vez de toda a instância contendo todos os atributos;

Transaction: uma operação opcional na implementação do WFS. Ela é utilizada para descrever operações de transformação de dados aplicáveis sobre instâncias de camadas

espaciais disponibilizadas pelo serviço. Utilizando esta operação, aplicações cliente podem criar, modificar, substituir ou excluir feições do servidor;

GetFeatureWithLock: uma operação opcional no WFS. Sua função é similar à operação `GetFeature`, exceto pelo fato de que, na resposta desta operação, além de encaminhar o documento de resposta para o cliente, o serviço bloqueia o acesso de escrita para todas as instâncias envolvidas no resultado, presumindo que a próxima requisição do cliente será uma *Transaction*, de atualização ou remoção de instâncias.

Diante dessas operações especificadas para o serviço WFS, fica evidente que a utilização desse serviço, por parte de aplicações de SIG, possibilita a realização de operações com os dados espaciais, independente da maneira como estes foram armazenados. Se utilizado em conjunto com o WMS, possibilita a construção de ricas aplicações de SIG, por exemplo.

2.4 Características da visualização exploratória de dados espaço-temporais

Andrienko et al. (2010b) apresentam um documento retratando, principalmente, a análise espaço-temporal. Trata-se de um guia para os pesquisadores em visualização espaço-temporal, que traz questionamentos interessantes sobre visualização temporal, visualização colaborativa, manipulação de grandes massas de dados, eficácia das técnicas visuais e integração de diversas fontes de dados.

Na visão dos autores, todos nós estamos, já estivemos, ou um dia estaremos interessados em visualizar alguma informação que envolva localização espacial e variação temporal simultaneamente. Para Andrienko et al. (2010b), é importante valorizar a variação temporal e entender que uma visão qualquer de um mapa representa um estado do espaço no tempo, válido no instante (ou período) em que tal visão foi capturada. Métodos visuais têm-se revelado úteis na análise de dados relacionados com o tempo e, em razão disto, pesquisas em técnicas de visualização temporal têm surgido.

Dada a dificuldade em considerar todos os aspectos da dimensão temporal em uma única visualização, a maioria dos métodos disponíveis foca casos específicos onde a visualização dos dados temporais baseia-se em um eixo de tempo linear (Andrienko et al., 2010b).

Métodos visuais para dados temporais podem ser categorizados baseando-se nas características do tempo. Eles podem tratar o tempo como linear ou cíclico, ou podem tratar como pontos ou intervalos, por exemplo. A Figura 2.2 mostra a diferença entre a representação linear e cíclica, a partir de um exemplo relacionado a padrões temporais em dados de saúde humana. Enquanto o gráfico de barras é útil para destacar tendências gerais e *outliers*, a visualização em espiral foca nos aspectos cíclicos (repetitivos) dos dados ao longo do tempo (Tominski e Schumann, 2008).

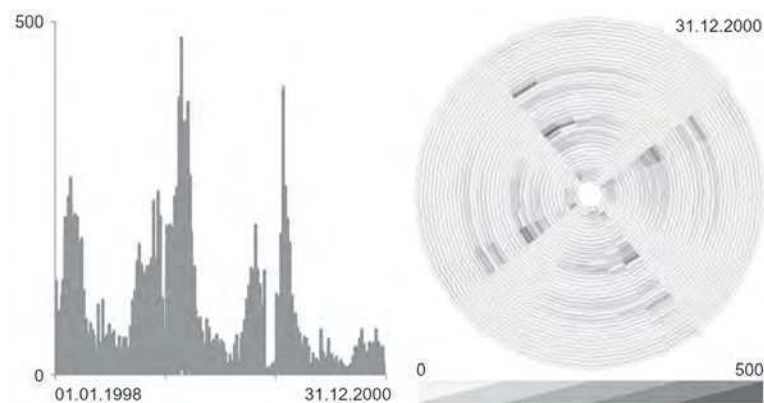


Figura 2.2: Representação linear e cíclica de uma série temporal de dados relacionados à saúde humana (Andrienko et al., 2010b).

No gráfico linear, é muito difícil descobrir padrões de recorrência. Por outro lado, com a visualização em espiral, é possível observar um padrão cíclico implícito e aparente graças à técnica de visualização.

Para Andrienko et al. (2010b), as técnicas mais conhecidas de visualização espaço-temporal são: mapas interativos, animação de mapas, e o cubo espaço-temporal. Um exemplo do cubo espaço-temporal é apresentado na Figura 2.3.

Na Figura 2.3, são apresentadas trajetórias de objetos móveis por meio do cubo espaço-tempo. Nesse cubo, a dimensão espacial está representada em sua base, e a dimensão temporal está representada na altura. Na parte superior da figura, é possível visualizar trajetórias percorridas por petroleiros em um único dia. As trajetórias estão representadas como uma sequência de pontos sobre o espaço-tempo. Já na parte inferior, a representação pontual das trajetórias é substituída pela densidade dos movimentos, obtida por uma técnica de renderização de volume (Andrienko et al., 2010b).

Os autores ainda destacam o uso do método computacional SOM como outra forma

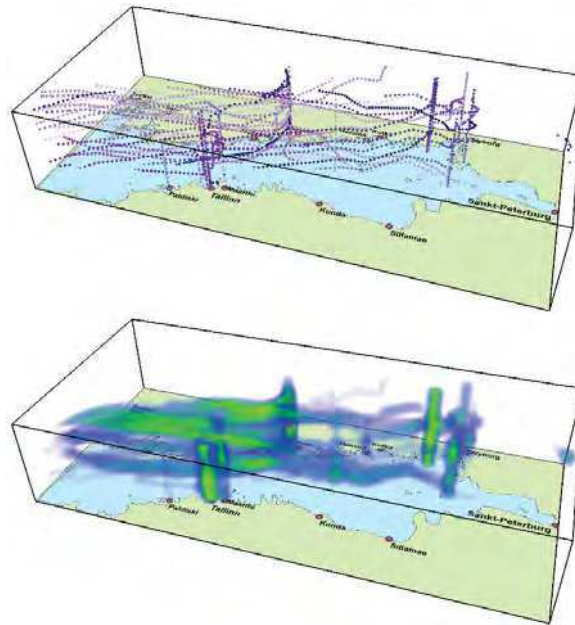


Figura 2.3: Agregação de trajetórias a partir da densidade dos movimentos no espaço-tempo (Andrienko et al., 2010b).

de visualização de dados espaço-temporais. Eles deixam claro que é preciso lidar com o tempo de forma mais eficaz, e enumeram algumas características que seriam ideais para um bom sistema de análise visual de dados espaço-temporais. Com base nessas características, enumeram-se seis características principais. Essas características são:

Espacialidade: oferecer recursos para tratamento e exploração das propriedades da dimensão espacial;

Temporalidade: oferecer recursos para tratamento e exploração das propriedades da dimensão temporal;

Genericidade: oferecer recursos genéricos e independente de qualquer domínio de aplicação;

Flexibilidade: dar liberdade para manipular os dados visualizados, possibilitando realização de consultas e aplicação de filtros;

Interoperabilidade: permitir conexão com diversas fontes de dados de forma prática e eficiente;

Mineração de Dados: possibilita a descoberta de conhecimentos implícitos.

Dar ao usuário recursos para manipular, tanto a dimensão temporal quanto espacial, é indispensável em uma solução para análise sobre dados espaço-temporais. Não se deve descartar as particularidades existentes em uma dessas dimensões, uma vez que, estas juntas, poderão revelar relacionamentos implícitos que condizem com a realidade dos dados analisados.

Ser visual também é outra característica indispensável em razão dos estudos já comprovados sobre a percepção humana e as vantagens obtidas em análise visual de informações.

2.5 Resumo do Capítulo

Este capítulo apresentou a fundamentação teórica deste trabalho. Foram apresentadas as principais características dos dados espaciais, temporais e espaço-temporais.

Foi possível destacar a relevância e a complexidade na manipulação de dados espaço-temporais, uma tarefa que requer esforços para o desenvolvimento de novas soluções que possam agilizar e facilitar o processo de descoberta de conhecimento implícito em bases com dados desse tipo.

Discutiu-se a mineração de dados espaço-temporais, diferenciando-a da mineração de dados espaciais e enumerando os principais problemas, os fatores a serem considerados, e as principais tarefas e técnicas.

Por fim, apresentou-se uma discussão detalhada sobre os padrões de serviço de dados espaciais desenvolvidos pela OGC para prover interoperabilidade entre aplicações de geoprocessamento.

No capítulo seguinte, será apresentado o estado da arte em visualização e análise de dados espaciais e espaço-temporais.

Capítulo 3

Trabalhos Relacionados

Neste capítulo, é apresentado o estado da arte em visualização e análise de dados espaciais e espaço-temporais. São abordados trabalhos dessa área de interesse, além das principais soluções gratuitas de SIG.

3.1 Trabalhos Acadêmicos

Existem muitos trabalhos publicados na comunidade acadêmica voltados à visualização de dados espaciais, entretanto, poucos se voltam à visualização espaço-temporal, ou seja, à inclusão da dimensão temporal na visualização de dados espaciais. Em meio a estes poucos trabalhos, são apresentados, nesta seção, os principais.

3.1.1 *SocioScape*

Reda et al. (2009) desenvolveram uma ferramenta que possibilita a exploração visual de mudanças de grupos de redes sociais dinâmicas espacialmente referenciadas ao longo do tempo. Denominada *SocioScape*, essa ferramenta foca no estudo dos grupos sociais presentes na população e como esses grupos evoluem com o tempo.

A motivação desse trabalho surgiu a partir do estudo das ferramentas de visualização e análise espaço-temporal, voltadas para o domínio de grupos sociais. Os cientistas têm dificuldades para explorar a evolução da estrutura da sociedade e chegar a hipóteses sobre o papel de vários fatores ambientais na formação dessa estrutura (Reda et al., 2009). Para

realizar o estudo, os autores utilizaram uma rede social dinâmica visualizada a partir de séries de grafos estáticos representando indivíduos que foram observados interagindo entre si em diferentes marcações temporais.

A abordagem desenvolvida consiste basicamente em duas técnicas de visualização de dados espaço-temporais: a primeira, denominada *Affiliation Timeline*, que retrata a estrutura da comunidade na população e sua evolução ao longo do tempo; e a segunda, chamada de cubo espaço-temporal, que permite visualizar o movimento das comunidades em um ambiente físico. A Figura 3.1 apresenta essas duas técnicas de visualização espaço-temporal.

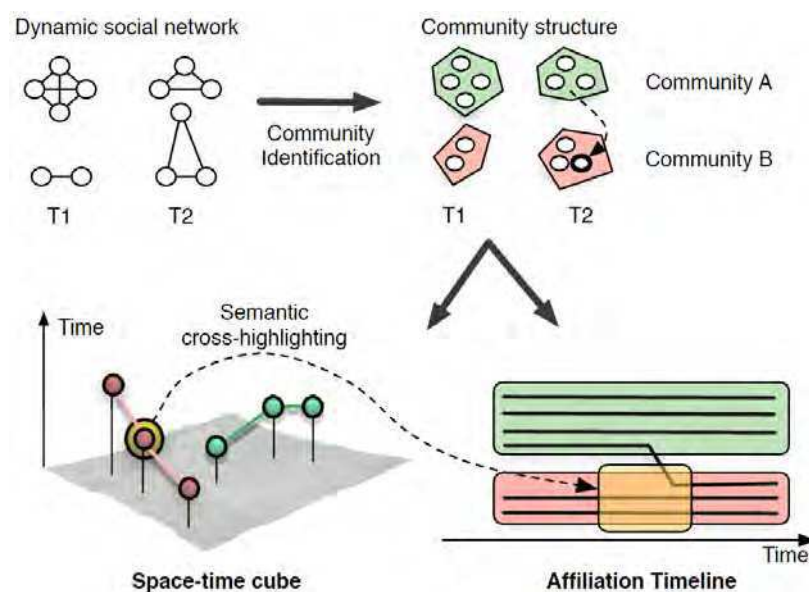


Figura 3.1: Técnicas de visualização exploradas na ferramenta *SocioScape* (Reda et al., 2009)

De acordo com a Figura 3.1, o processo de análise tem início com a identificação das comunidades e de suas estruturas em cada marcação temporal. Após a identificação, o usuário segue com duas opções para visualizar a evolução das comunidades e assim extrair conclusões a partir das interações entre elas.

A técnica do cubo espaço-temporal (à esquerda, na Figura 3.1) consiste em uma visualização 3D das comunidades (em forma de esferas) e de suas mudanças ao longo do tempo. A base do cubo retrata a dimensão espacial representada por um mapa topológico, enquanto que o eixo vertical retrata a dimensão temporal. Uma linha conecta as visões de uma comunidade em marcações temporais distintas. Uma linha vertical é projetada das esferas para o

mapa no sentido de destacar a localização espacial de cada uma.

A técnica denominada *Affiliation Timeline* (à direita, na Figura 3.1) consiste de um layout 2D semelhante a um diagrama de coordenadas paralelas, usado com o objetivo de ilustrar, explicitamente, a evolução das comunidades ao longo do tempo. O eixo X é usado para a representação temporal, e o eixo Y usado para as comunidades e seus indivíduos. As comunidades são representadas, nessa técnica, como retângulos posicionados uns sobre os outros. Os indivíduos são representados como linhas contidas nos retângulos. Quando uma linha cruza dois retângulos, significa que os indivíduos ali representados estão migrando de uma comunidade para outra na marcação temporal em que ocorre esse cruzamento. Essa técnica de visualização permite claramente observar, no nível de indivíduos, as mudanças ocorridas nas comunidades ao longo do tempo.

A ferramenta *SocioScape* explora as duas técnicas de visualização espaço-temporal em uma visão simultânea e interativa, oferecendo um recurso chamado de *semantic cross-highlighting*, conforme ilustrado na Figura 3.2. Esse recurso permite, por exemplo, que um usuário, ao clicar em uma esfera representando uma comunidade no mapa 3D do cubo espaço-temporal, possa ver também os detalhes de evolução da comunidade clicada no *layout* 2D da *Affiliation Timeline*.

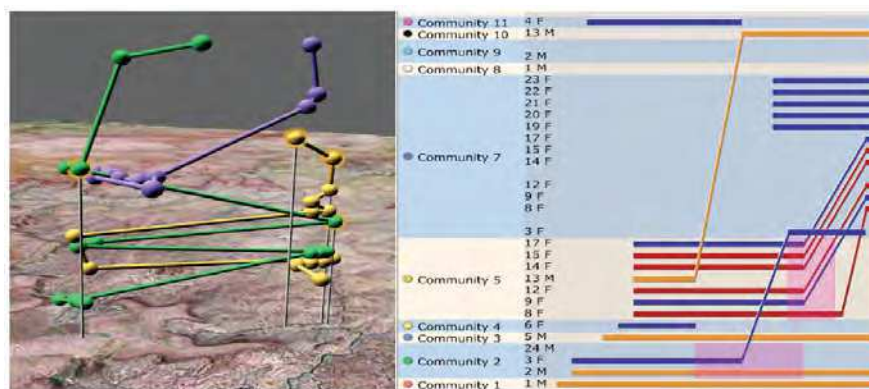


Figura 3.2: *SocioScape* exibindo o movimento espaço-temporal das comunidades de zebra "Grevy" (esquerda) e a composição social desses grupos (direita) (Reda et al., 2009).

Conclui-se que a solução proposta apresenta duas técnicas de visualização espaço-temporal, dessa forma atendendo tanto a **espacialidade** quanto a **temporalidade**. Entretanto, não apresenta **flexibilidade**, pois é restrita a essas duas técnicas de visualização e não é possível aplicar filtros ou adicionar outros dados. Conclui-se também que a solução não

apresenta **genericidade**, por ser válida apenas no domínio dos grupos sociais, e não apresenta **interoperabilidade**, pois não descreve a forma como o usuário fornece os dados para visualização e análise, levando à conclusão de que existem restrições nesse sentido. Por fim, verifica-se que a solução não utiliza nenhuma técnica de **mineração de dados**.

3.1.2 *Exploratory spatio-temporal data mining and visualization*

O trabalho de Compieta et al. (2007) propõe um sistema de mineração de dados espaço-temporais baseado em regras de associação e independente de domínio. Utilizando dois tipos de visualizador espaço-temporal, a solução proposta possibilita a visualização dos resultados obtidos através de mineração de dados sobre dados espaço-temporais.

A solução proposta apresenta uma arquitetura em três camadas: visualização, aplicação e dados, conforme apresentada na Figura 3.3.

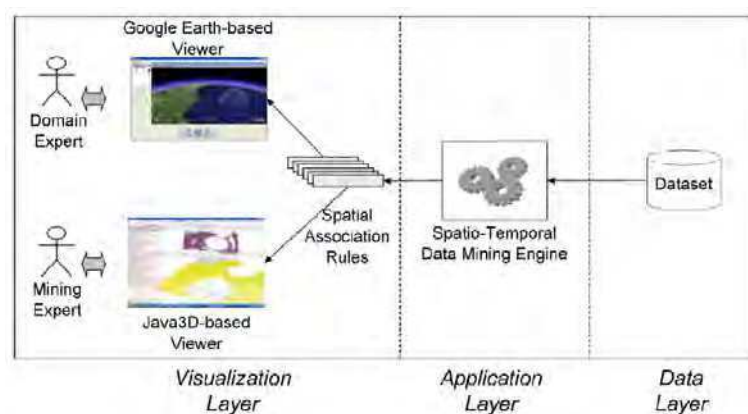


Figura 3.3: Arquitetura da solução proposta por Compieta et al. (2007).

A Figura 3.3 apresenta detalhes das três camadas que compõem a arquitetura dessa solução. Na camada de dados, os autores não especificam o formato em que os dados devem estar disponibilizados para serem corretamente utilizados pela solução proposta, pois sua arquitetura considera apenas um *Dataset*, sem apresentar detalhes sobre SGBDs suportados, por exemplo. Na camada de aplicação, encontra-se o motor de mineração de dados desenvolvido. Este contém mecanismos para tratamento dos dados de entrada e implementa dois módulos: o localizador, que trata as dimensões espaciais e temporais, e retorna os relacionamentos espaço-temporais encontrados; e o minerador, que processa os relacionamentos retornados pelo localizador utilizando uma técnica baseada no algoritmo *Apriori* (Agrawal e

Srikant, 1994). O resultado do minerador consiste em regras de associação espaço-temporal.

Na camada de visualização, a solução propõe duas ferramentas de visualização, que, de acordo com Compieta et al. (2007), oferecem funcionalidades complementares entre elas. A primeira implementação trata-se de um visualizador georeferenciado, baseado no *Google Earth*, que foca no destaque dos relacionamentos espaciais. A segunda implementação é baseada em *Java-3D*, e seu foco é na análise exploratória dos dados, possibilitando visualizar e analisar a estrutura interna, os relacionamentos, e os padrões inferidos pelo mecanismo de mineração.

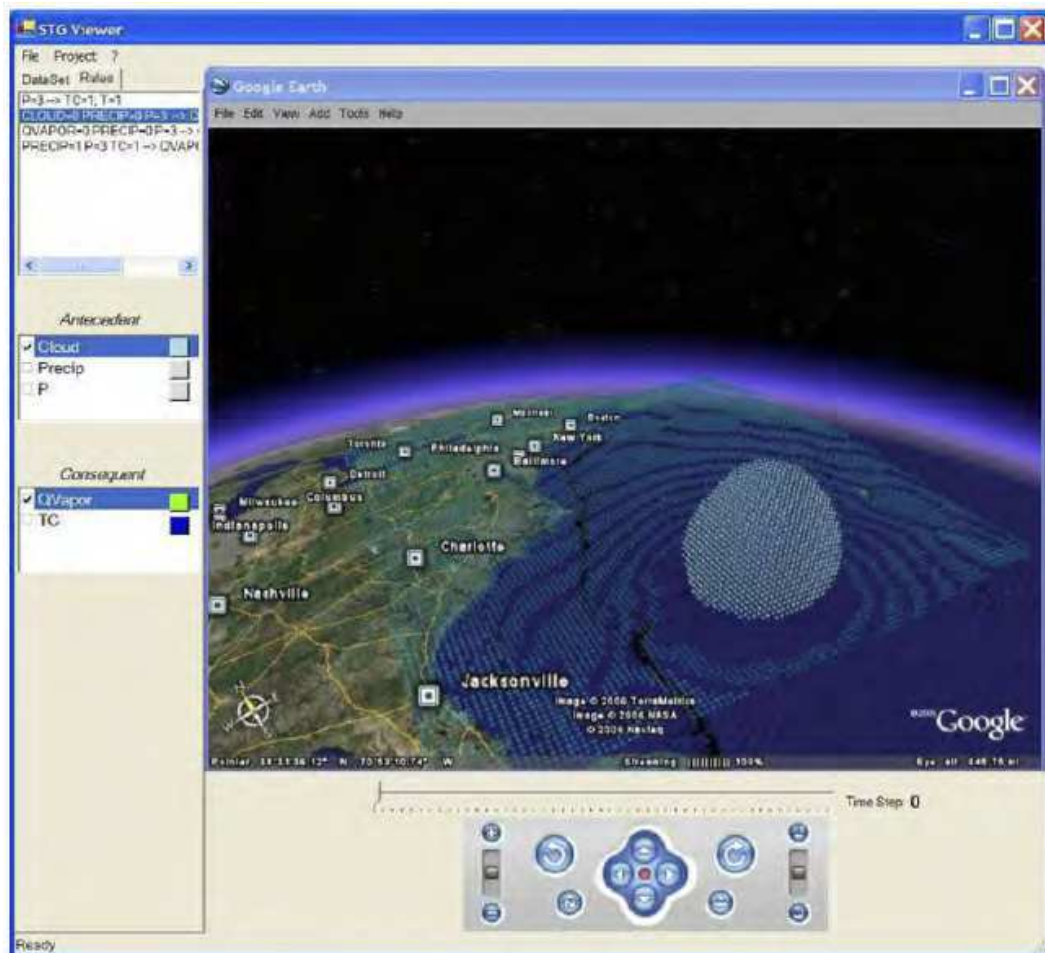


Figura 3.4: Visualização de regras de associação espaço-temporal na implementação baseada em *Google Earth* (Compieta et al., 2007).

A Figura 3.4 apresenta a interface do visualizador baseada no *Google Earth*. No exemplo, é possível observar, no lado esquerdo da figura, três regras de associação e os objetos antecedentes e consequentes dessas regras.

A Figura 3.5 apresenta a interface do visualizador baseada em *Java-3D*. É possível observar a visualização das regras de associação em um espaço tridimensional.

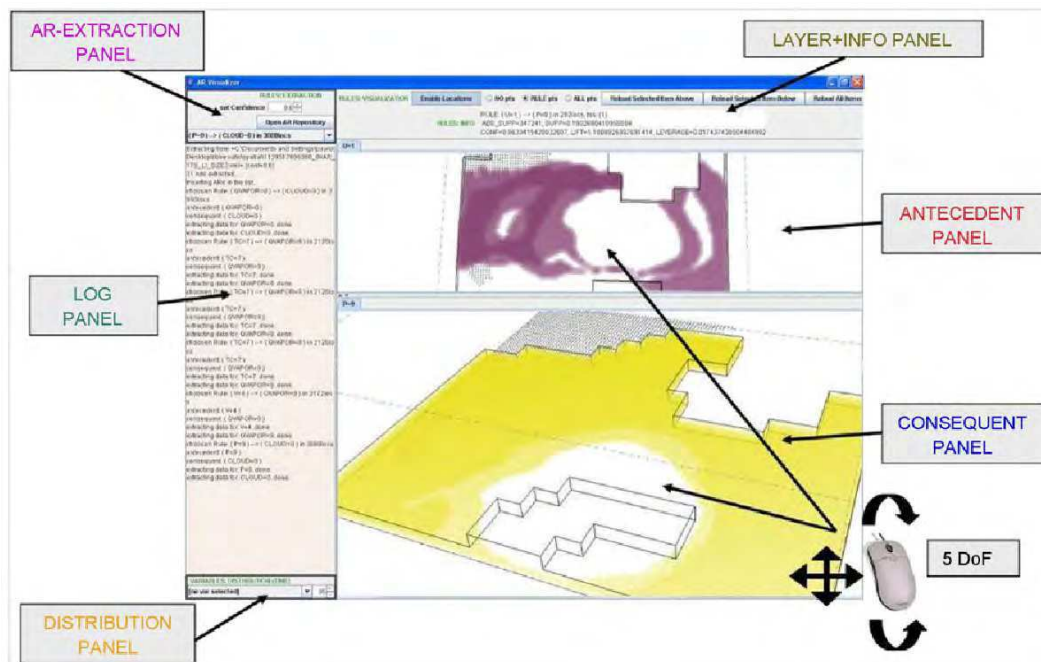


Figura 3.5: Esquema de interface da implementação baseada em *Java-3D* (Compieta et al., 2007).

Compieta et al. (2007) esclarecem que a construção de dois visualizadores objetivou atender tanto os especialistas no domínio, que se sentem familiarizados utilizando a versão georeferenciada, quanto os especialistas em mineração de dados, que veem a versão sem georeferenciamento mais apropriada para explorar os dados e os relacionamentos implícitos.

É importante observar que, mesmo com a promessa de alcançar tanto a **especialidade** quanto a **temporalidade**, os visualizadores apresentados nesse trabalho revelam poucos recursos para a manipulação da dimensão temporal, levando à conclusão de que são, na prática, visualizadores de dados espaciais. Na versão baseada em *Google Earth*, é possível observar um controlador temporal localizado na parte inferior da interface. Já na versão baseada em *Java-3D*, nenhum controle temporal é observado.

O trabalho apresenta **mineração de dados**, através do uso de regras de associação, apresenta **genericidade**, uma vez que os visualizadores e o motor de mineração de dados são funcionais com qualquer domínio aplicado, e também apresenta **flexibilidade**, haja vista que os visualizadores dão liberdade ao usuário na definição do volume de dados visualizado. En-

tretanto, vale ressaltar que não há **interoperabilidade**, em função da não especificação da fonte de dados suportada.

3.1.3 *Spatio-Temporal Visual Query Environment*

O trabalho de Cavalcanti et al. (2006) apresenta o S-TVQE (*Spatio-Temporal Visual Query Environment*), uma ferramenta para realizações de consultas visuais em bancos de dados espaço-temporais. Construído sobre a ferramenta TVQE (*Temporal Visual Query Environment*) (Silva et al., 2002), o S-TVQE acrescenta recursos de consultas visuais à dimensão espacial.

A ferramenta proposta consiste, basicamente, na disponibilização de operadores gráficos para a formulação de consultas espaço-temporais. O usuário pode visualizar os resultados das consultas por meio de tabela de dados, mapas estáticos georeferenciados e gráficos temporais estáticos.

Cada atributo de uma tabela pertencente a um banco de dados espaço-temporal é analisado e pode ser classificado como espacial, temporal ou convencional. A partir de então, as consultas podem ser montadas por meio do editor visual desenvolvido. A Figura 3.6 ilustra a interface principal da ferramenta.

No canto superior esquerdo da Figura 3.6 é possível visualizar a área de seleção espacial, onde o usuário pode fazer restrições espaciais que vão influenciar na abrangência das consultas. No canto inferior esquerdo é possível localizar os atributos de uma tabela pertencente a um banco de dados espaço-temporal e suas classificações. No lado direito da interface estão os operadores, disponibilizados de acordo com a classificação do atributo selecionado. A área central da interface é destinada à formulação da consulta.

A Figura 3.7 apresenta o resultado da seguinte consulta realizada no S-TVQE:

"Recuperar as cidades vizinhas de Campina Grande que pertencem ao III Núcleo Regional de Saúde e tiveram menos que 50 internações hospitalares no ano de 2002".

A imagem da Figura 3.7 apresenta um mapa estático com todos municípios do III Núcleo Regional de Saúde e, em destaque, os quatro que satisfizeram a consulta realizada: Fagundes, Queimadas, Puxinanã e Pocinhos; cada uma registrando apenas 10 internações em 2002. Nos botões localizados na lateral esquerda da interface, é possível alternar a visualização

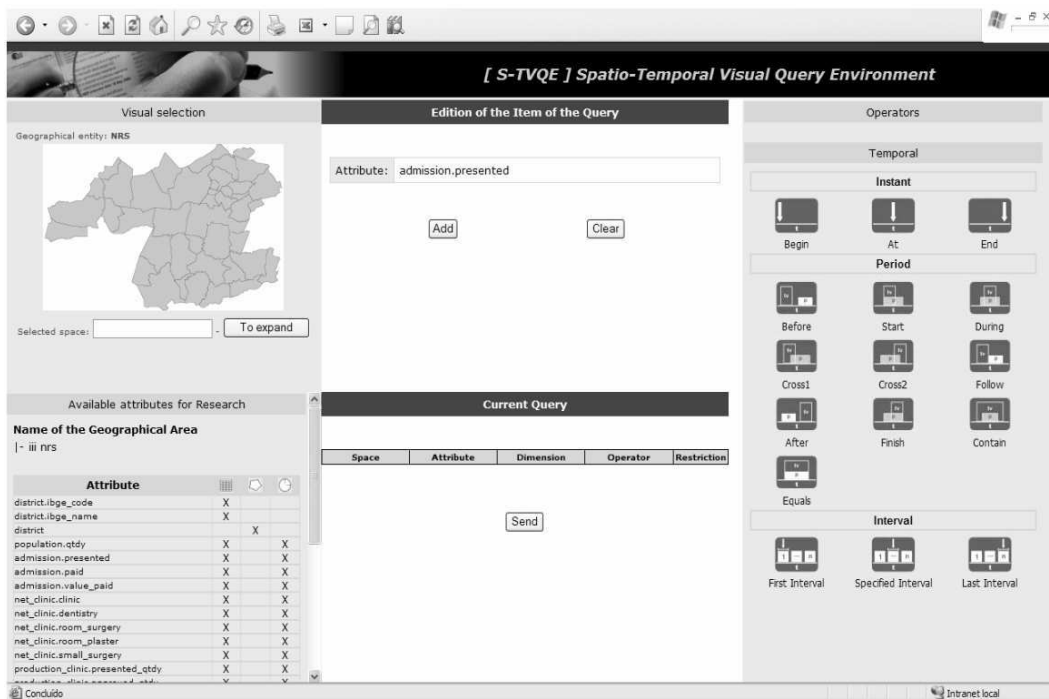


Figura 3.6: Interface principal do S-TVQE (Cavalcanti et al., 2006).

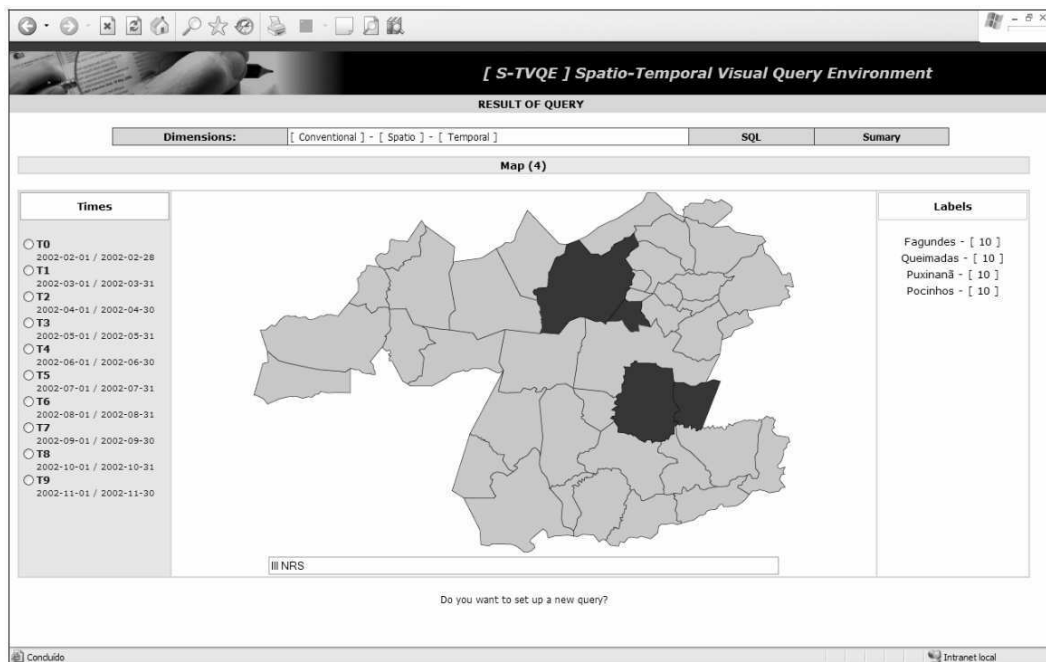


Figura 3.7: S-TVQE apresentando resultado visual de consulta espaço-temporal (Cavalcanti et al., 2006).

dos mapas em função da marcação temporal. Na Figura 3.8, é possível observar o gráfico de distribuição temporal dessa mesma consulta, com os valores temporais observados mês a mês.

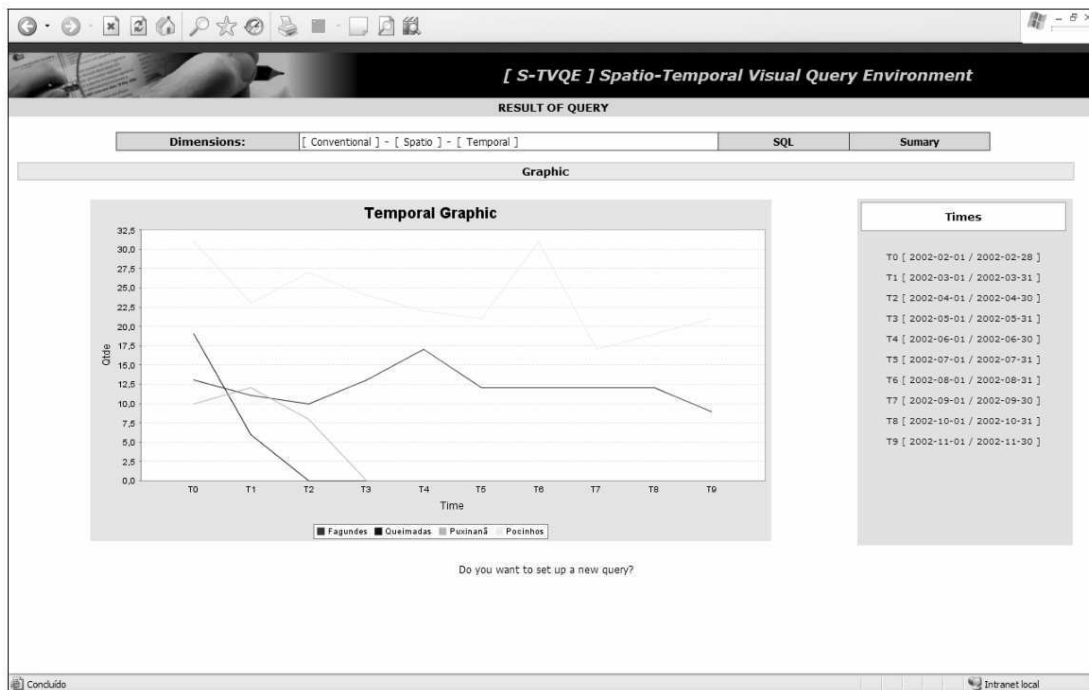


Figura 3.8: S-TVQE apresentando gráfico temporal resultante de consulta espaço-temporal (Cavalcanti et al., 2006).

É possível concluir que o trabalho de Cavalcanti et al. (2006) de fato aborda tanto a **especialidade** quanto a **temporalidade** em dados espaço-temporais, oferecendo um ambiente para realização de consultas espaço-temporais (**flexibilidade**), mas sem considerar **mineração de dados**, e portanto, sem haver mecanismos de descoberta de conhecimento implícito. Apesar disso, as técnicas de visualização adotadas são precárias, fazendo uso de interfaces estáticas tanto para a navegação espacial quanto temporal, o que pode causar um impacto negativo na experiência do usuário.

A solução também apresenta **interoperabilidade** parcial, uma vez que oferece conectividade com três SGBDs relacionais distintos, mas restringe-se somente à SGBDs, e apresenta **genericidade**, por ser independente de domínio.

3.1.4 *Visual Analytics Tools for Analysis of Movement Data*

O trabalho de Andrienko et al. (2007) apresenta um *framework* para análise visual de dados espaço-temporais que representam trajetórias de objetos móveis. Este *framework* combina operações de bancos de dados com processamento computacional, mineração de dados e interfaces visuais interativas.

As operações de bancos de dados são utilizadas na fase de pré-processamento das informações que serão analisadas. Nessa etapa, são extraídas informações relevantes das trajetórias, como as paradas, e incluídos atributos que permitem saber, por exemplo, a aceleração (diferença de velocidade entre um registro e o seguinte), a direção do movimento, a diferença temporal entre um registro e outro, etc. As técnicas de processamento computacional, por sua vez, são utilizadas para agregação e sumarização dos registros, o que permite reduzir grandes volumes de dados antes de visualizá-los. Já as interfaces visuais objetivam facilitar o processo de análise, explorando o poder de cognição e raciocínio dos usuários.

Andrienko et al. (2007) destacam a utilização do algoritmo de *clustering* OPTICS (Ankerst et al., 1999) para detecção de locais visitados com frequência e redução da base de dados, uma vez que o algoritmo é utilizado para suprimir os registros que representam uma mesma situação de parada, podendo oferecer uma redução considerável no conjunto de dados de trajetórias a ser analisado. A Figura 3.9 ilustra a utilidade desse algoritmo de *clustering* na análise de trajetórias.

É possível visualizar cinco pontos de parada relevantes, de acordo com os resultados da mineração de dados. Além dos locais de partida das trajetórias, os resultados apontaram *shoppings centers* e uma quadra de tênis. É importante observar que, sem o auxílio de um mapa georeferenciado como plano de fundo da visualização, a tarefa de análise e identificação dos pontos de parada seria bastante dificultada.

Em se tratando de aspectos de visualização espaço-temporal, a solução oferece histogramas temporais para serem utilizados na análise das trajetórias em função do tempo, e possibilitar a identificação dos padrões temporais. Para a dimensão espacial, a solução não oferece uma interface proprietária, mas sim uma maneira de exportação dos dados para que estes sejam visualizados em visualizadores espaciais georreferenciados.

A solução proposta em Andrienko et al. (2007) permite uma análise visual satisfatória com dados de trajetória (**espacialidade e temporalidade**). É independente de domínio

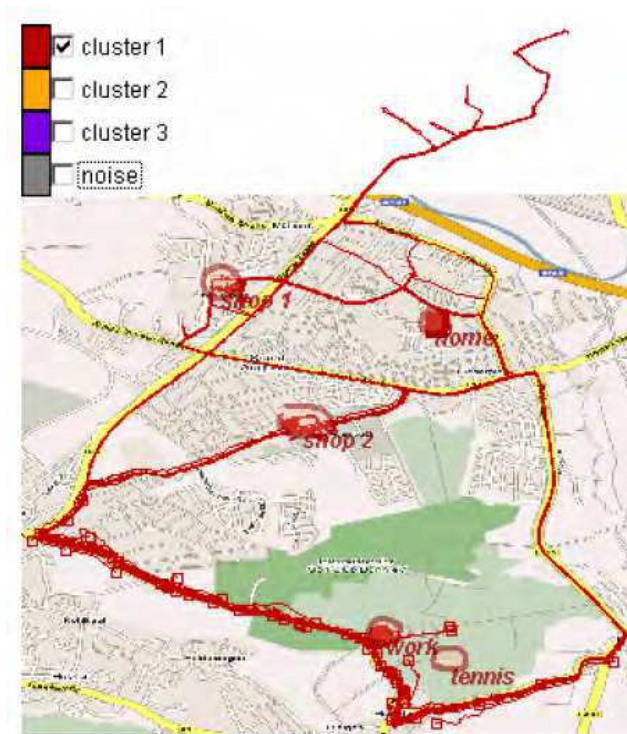


Figura 3.9: *Clusters* de paradas realizadas em várias viagens de carro entre o trabalho e a residência de um cidadão (Andrienko et al., 2007).

(**genericidade**), apesar de restringi-lo a trajetórias de objetos móveis representados apenas por pontos no espaço. Apresenta **flexibilidade**, com recursos de sobreposição de camadas contendo trajetórias ou *clusters*. Utiliza ainda, **mineração de dados** através do algoritmo de *clustering* OPTICS. É importante ressaltar que os autores não deixam claro o formato aceitável para os dados de trajetória, levando à conclusão de que existem restrições e que, portanto, não há **interoperabilidade**.

3.1.5 *A Framework for Exploring Spatiotemporal Patterns*

O trabalho de Andrienko et al. (2010a) sugere um *framework* baseado na técnica *Self-Organizing Map* (SOM), uma combinação de *clustering* com redução de dimensionalidade. Esta técnica segue a ideia de que objetos não são apenas agrupados, mas também arranjados num espaço uni ou bi-dimensional, de acordo com sua similaridade em função de atributos multidimensionais.

A redução da dimensionalidade tem utilidade em registros de dados espaço-temporais

com diversos atributos relevantes (dimensões). Com o uso da técnica, a base de dados visualizada é reduzida, focando em atributos que possam trazer conclusões relevantes para o estudo realizado.

Andrienko et al. (2010a) defendem a ideia de que, para uma análise compreensiva em dados espaço-temporais, esses dados precisam ser considerados sob duas perspectivas: (1) como distribuições espaciais (situações) variando ao longo do tempo e (2) como perfis de variação temporal local distribuídos sobre o espaço. Dessa forma, a solução proposta utiliza a técnica SOM sob a óptica dessas duas perspectivas, podendo ser aplicada em qualquer tipo de dado que possa ser representado por vetores.

A técnica *Self-Organizing Map* (ou, em português, Mapas Auto-Organizáveis) (Kohonen, 2001) consiste de um algoritmo de rede neural para projeção e quantização vetorial. Por meio de um processo de treinamento competitivo e interativo, uma rede de vetores protótipos (ou neurônios, ou células) é treinada para os dados do vetor de entrada. A saída do algoritmo é uma rede vetores que preserva uma topologia aproximada dos dados de entrada. Essa rede pode ser interpretada como um conjunto de *clusters* e, ao mesmo tempo, como um mapa que dispõe os elementos de entrada. A Figura 3.10 apresenta possíveis aparências de células em uma matriz gerada a partir da técnica SOM.

Na Figura 3.10 são apresentadas diversas formas de visualização de uma matriz gerada pela técnica SOM sobre dados espaço-temporais. No lado esquerdo é possível observar agrupamentos temporais de situações espaciais (representados por mapas) enquanto que no lado direito é possível observar agrupamentos espaciais de acordo com a variação temporal dos valores dos atributos (representados por diagramas, também conhecidos por mosaicos temporais).

As Figuras 3.10A e 3.10B retratam um atributo com valores registrados ano a ano, durante 41 anos. Na Figura 3.10A é possível visualizar os anos em que a coloração do mapa se aproxima, ou seja, é possível visualizar a distribuição espacial das médias registradas para o atributo em questão no 30º, 31º e 32º ano da série temporal. Na Figura 3.10B é possível visualizar os estados americanos em que a coloração da linha representando toda a variação temporal se aproxima, ou seja, é possível visualizar a distribuição temporal das médias registradas para o atributo em questão nos estados americanos em destaque no mapa. A partir das imagens, é possível concluir que, por exemplo, entre, aproximadamente, o 13º e o 26º ano

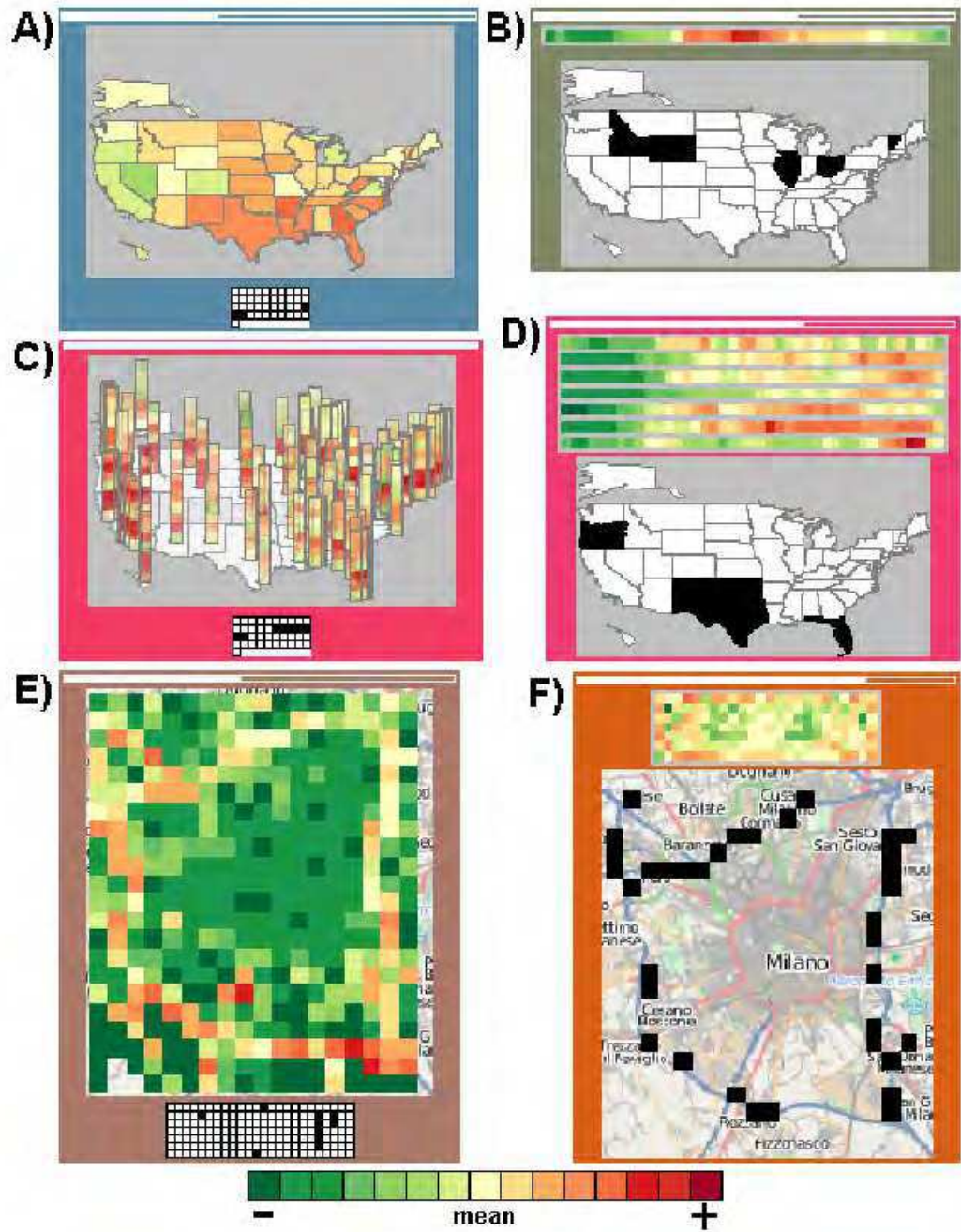


Figura 3.10: Possíveis aparências de células em uma matriz gerada a partir da técnica *Self-Organizing Map* (Andrienko et al., 2010a).

da série temporal, os valores observados no atributo em questão, para cada um dos estados destacados em cor preta, no mapa da figura 3.10B, estão acima da média, de acordo com a legenda, em razão das colorações observadas nesse período da série variarem entre tons de laranja e vermelho.

As Figuras 3.10C e 3.10D retratam sete atributos com valores registrados durante 41 anos. É possível notar, em 3.10C, que a visualização fica mais complexa com o aumento no número de atributos (em 3.10A era apenas um), enquanto que a visualização mostrada em 3.10D parece ser mais apropriada e menos complexa para esse caso.

As Figuras 3.10E e 3.10F retratam um atributo (tráfego) com valores registrados hora-a-hora durante uma semana (7x24). De forma semelhante ao que se vê com as figuras 3.10A e 3.10B, é possível visualizar, na Figura 3.10E, as horas do dia, para cada dia da semana, em que o padrão do mapa se aproxima. Observando a Figura 3.10E, é possível concluir, por exemplo, que o trânsito no centro do mapa apresenta tráfego abaixo da média no período de segunda a sexta-feira, entre 19 e 20 horas.

A capacidade do SOM em organizar os dados de entrada em uma estrutura de rede regular provê boas oportunidades de visualização. Isso torna o método conveniente para exploração de dados multidimensionais (Andrienko et al., 2010a).

Os autores construíram um ambiente de análise visual onde é possível aplicar o SOM para situações espaciais ou variações temporais, e explorar os resultados por meio de técnicas de visualização, entretanto, não há detalhes sobre esse ambiente. O artigo foca apenas a aplicação do algoritmo SOM e os resultados visuais que podem ser obtidos a partir dele.

É possível concluir que o uso do algoritmo SOM trata tanto a **espacialidade** quanto a **temporalidade**, permitindo análises coerentes sobre dados espaço-temporais. A técnica é independente de domínio (**genericidade**) e se mostra funcional com qualquer área de conhecimento, além de trazer consigo a ideia de *clustering* para agregação e redução da base de dados (**mineração de dados**). Entretanto, não há **flexibilidade** nem **interoperabilidade**.

3.1.6 COPLINK: Visualization for Crime Analysis

O trabalho de Chen et al. (2003) apresenta uma ferramenta de visualização de dados espaço-temporais integrando várias técnicas de visualização, a exemplo de SIG, SOM, listas hierárquicas, árvores hiperbólicas, mecanismos de linha do tempo e padrões periódicos.

O principal objetivo dos autores, com a construção dessa ferramenta, era desenvolver um ambiente para análise espaço-temporal de informações sobre ocorrências de crimes. Isso justifica as técnicas de visualização adotadas.

A técnica de árvores hiperbólicas é baseada na geometria hiperbólica. Ela provê um meio conveniente para visualizar exponencialmente árvores hierárquicas crescentes (Chen et al., 2003). Quando utilizada para visualização de relacionamentos entre diferentes entidades do domínio de criminalidade (pessoas, veículos, locais, organização), possibilita localizar todas as entidades que têm uma relação com um determinado termo de busca, e visualizar essas relações em forma de árvores hiperbólicas ou numa estrutura de árvore hierárquica.

A técnica de visualização periódica exhibe padrões relacionados com o tempo, objetivando revelar recorrências em períodos de tempo equivalentes. Já a visualização da linha do tempo mostra as características dos dados temporais em uma disposição linear. O uso de SIG possibilita visualizar os dados em um mapa georeferenciado e, a partir daí, realizar análises espaciais. Os autores combinaram tais técnicas de manipulação espacial e temporal em uma única ferramenta, denominada STV (*Spatio-Temporal Visualizer*), apresentada na Figura 3.11.

A Figura 3.11 ilustra as técnicas de visualização espacial e temporal adotadas por Chen et al. (2003). É possível visualizar a ocorrência de crimes na linha do tempo (lado esquerdo), os locais de ocorrência (lado direito superior), e os períodos de tempo que registram mais eventos (lado direito inferior). Existe ainda um regulador de tempo (*time-slider*) e um controlador de animações temporais. Com o regulador, é possível delimitar o intervalo de tempo visualizado.

Os autores utilizaram ainda a técnica SOM com o objetivo de agrupar os crimes relacionados baseando-se em diferentes atributos, como por exemplo tipo do crime, localização, data e hora, sexo da vítima, etc (Kohonen, 2001).

Apesar de ser uma solução que trata tanto a **espacialidade** quanto a **temporalidade**, utilizando técnicas de visualização que abrangem ambas dimensões, é voltada a um domínio específico, o de criminalidade, e portanto não há **genericidade**. Além disso, por utilizar a técnica SOM, a ferramenta apresenta suporte a agrupamento (*clustering*) espaço-temporal dos dados (**mineração de dados**), embora seja restrito às condições impostas com o uso dessa técnica. Vale ressaltar, também, que não há **flexibilidade** nem **interoperabilidade**.

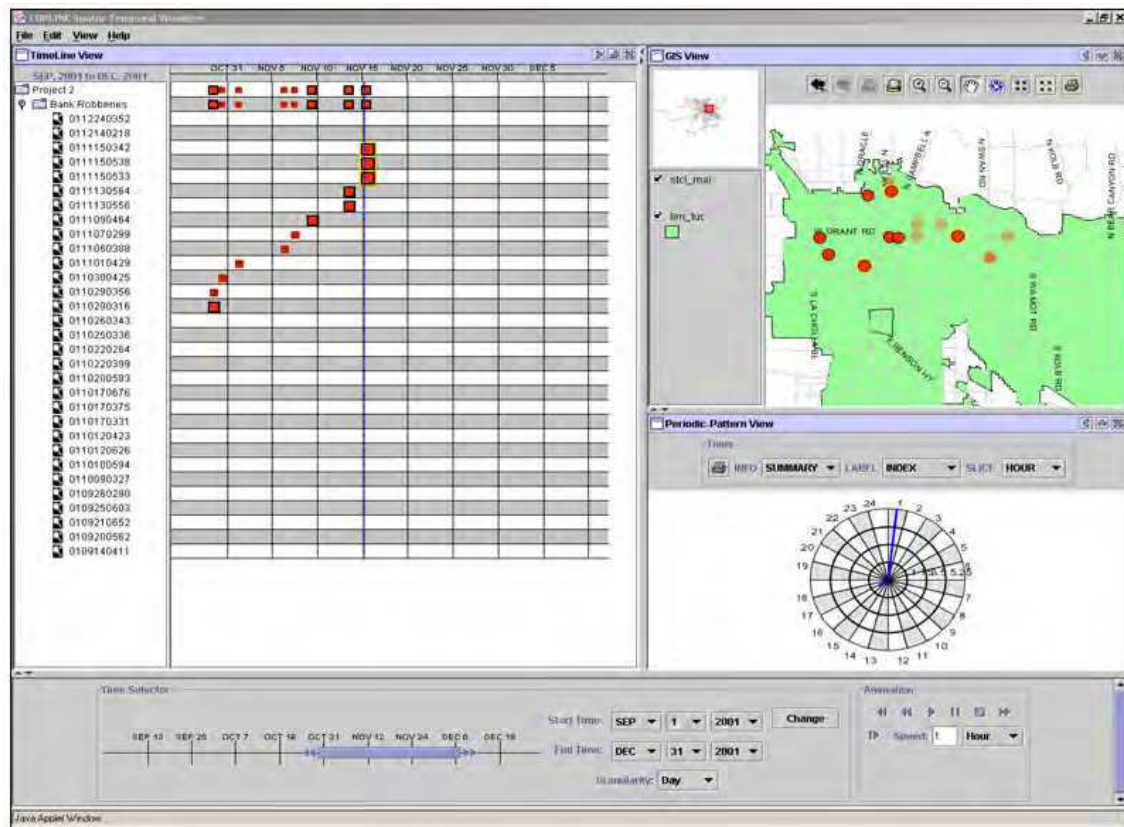


Figura 3.11: Apresentação da ferramenta *COPLINK Spatio-Temporal Visualizer* (Chen et al., 2003).

3.1.7 *BirdVis: Visualizing and Understanding Bird Populations*

O trabalho de Ferreira et al. (2011) consiste de um sistema de visualização interativa, denominado *BirdVis*, que suporta a análise visual de modelos de distribuição espaço-temporal de pássaros. Trata-se de uma abordagem espaço-temporal voltada para o domínio específico de pássaros, fruto de um trabalho de desenvolvimento que envolveu cientistas da computação, estatísticos, biólogos e ornitólogos.

A ideia do sistema surgiu em 2002, a partir do projeto *eBird*, que desde então coleciona registros de observações de pássaros. Essas observações, em conjunto com variáveis ambientais em escala local, como clima, *habitat* e vegetação, compõem um recurso valioso para a comunidade global de educadores, agricultores, ornitólogos e biólogos.

A Figura 3.12 apresenta a interface do *BirdVis*. É possível observar controladores espaciais e temporais para manipulação visual dos dados. Mapas temáticos dispostos lado a lado ilustram a probabilidade de ocorrência de determinadas espécies de pássaros em alguns

instantes de tempo.

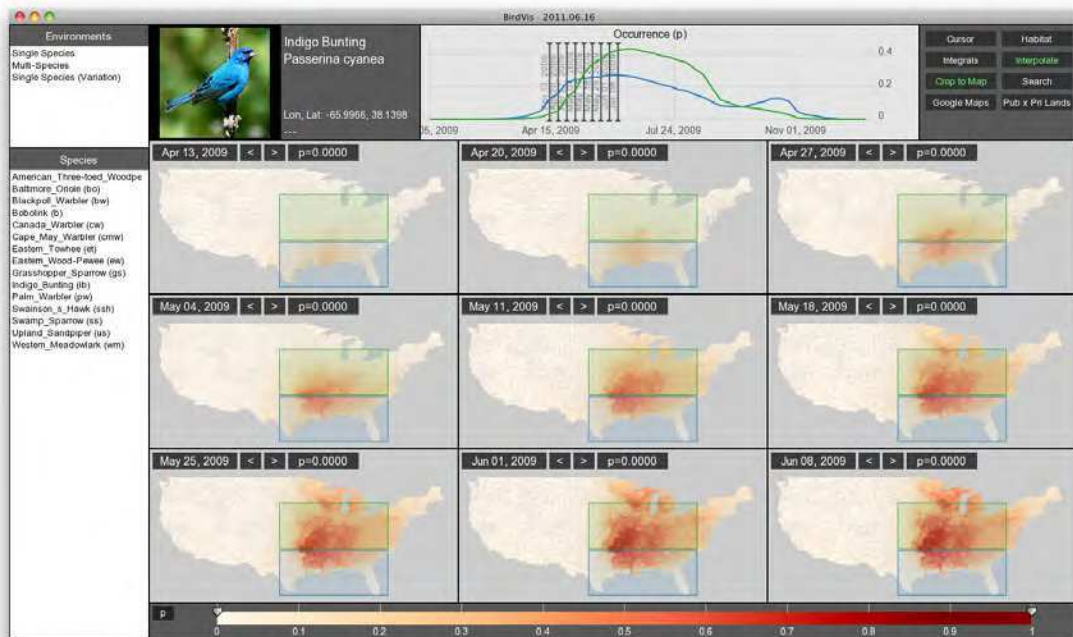


Figura 3.12: Interface da ferramenta *BirdVis* (Ferreira et al., 2011).

Modelos preditivos foram desenvolvidos a partir da associação de dados ambientais com padrões de ocorrência de pássaros observados, oferecendo um *framework* estatístico e aproveitando os dados disponíveis na previsão de distribuição das espécies e na realização de inferências sobre associações espécies-*habitat* (Ferreira et al., 2011).

O *BirdVis* utiliza técnicas de visualização de maneira intuitiva e inovadora, com o objetivo de melhor atender os usuários interessados na exploração de interdependências entre os parâmetros dos modelos desenvolvidos. Os mapas dispostos lado a lado permitem uma visão comparativa dos modelos analisados.

Os componentes visuais são apresentados na Figura 3.13. O *BirdVis* é composto por três ambientes e por sete componentes que tratam tanto a dimensão espacial, a exemplo do visualizador de mapas, como também a dimensão temporal, a exemplo do visualizador temporal de trajetórias.

O trabalho trata tanto a **especialidade** quanto a **temporalidade**, entretanto, vale ressaltar que, além de ser válido apenas para um domínio específico (não há **genericidade**), o *BirdVis* não oferece mecanismos para conexão a bases de dados externas (não há **interoperabilidade**), sendo restrito à base de dados desenvolvida por seus idealizadores, e não utiliza

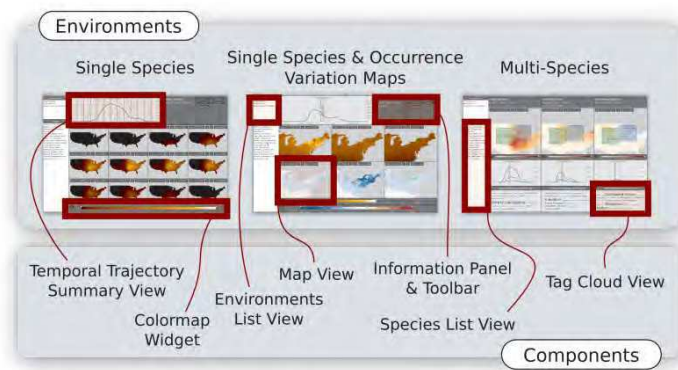


Figura 3.13: Ambientes e componentes da ferramenta *BirdVis* (Ferreira et al., 2011).

mineração de dados. Outro ponto importante é que existe um certo grau de **flexibilidade**, uma vez que os componentes possibilitam várias visualizações dos dados.

3.1.8 *GeoVISTA CrimeViz*

O *GeoVISTA CrimeViz* é uma aplicação de mapas para a *web* que suporta exploração espaço-temporal voltada ao domínio da criminalidade (Roth et al., 2010).

Desenvolvida a partir de uma abordagem de *design* centrada aos interesses do usuário final, a ferramenta provê uma interface de fácil utilização para o combate à criminalidade, permitindo a realização de análise espaço-temporal e mapeamento dos índices de criminalidade em um distrito da cidade de Washington, nos Estados Unidos.

A Figura 3.14 apresenta a interface interativa do *GeoVISTA CrimeViz*. Por meio dela, é possível observar a presença de controladores espaciais e temporais, a exemplo de um visualizador de mapas com camadas de dados pré-definidas, que podem ser utilizadas como sobreposição durante o processo de análise; e de um histograma de frequência, apresentando a variação temporal linear dos dados visualizados no mapa.

A aplicação oferece também o recurso de navegador espaço-temporal, que realiza uma animação simultânea no mapa e no histograma de frequência, possibilitando a visualização das variações dos dados à medida que o tempo evolui, por categoria de crime.

Apesar de ser uma solução que suporta dados espaço-temporais, ou seja, que atende bem a **especialidade** e a **temporalidade**, apresenta limitações no domínio de aplicação (não há **genericidade**) e na fonte de dados utilizada (não há **interoperabilidade**), uma vez que

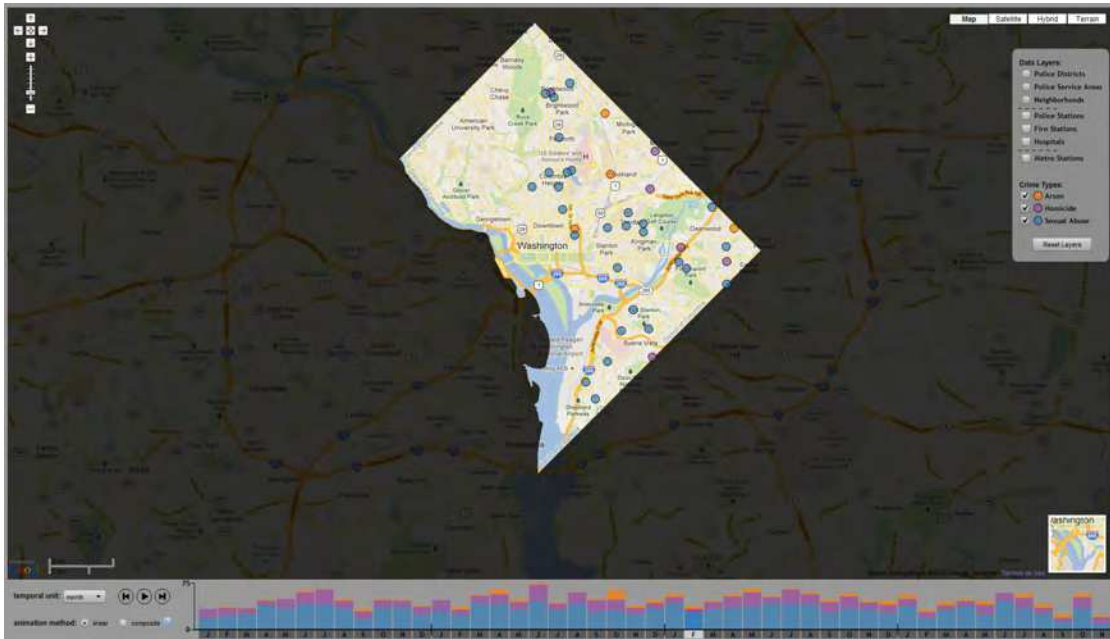


Figura 3.14: Interface Interativa da ferramenta *GeoVISTA CrimeViz* (Roth et al., 2010).

oferece algumas camadas de dados pré-definidas e não oferece recursos para inclusão de camadas contendo outro tipo de informação. A solução também não usa **mineração de dados**, mas oferece um certo nível de **flexibilidade**, permitindo escolher, dentre as camadas disponíveis, as que serão visualizadas.

3.2 Ferramentas de SIG

Existem muitas ferramentas de SIG no mercado. São várias opções tanto para os que procuram soluções gratuitas quanto também para àqueles que preferem as proprietárias. Em meio às soluções gratuitas, esta seção enumera as principais e mais utilizadas.

OpenLayers: é um visualizador de dados espaciais 2D baseado em mapas dinâmicos georeferenciados, desenvolvido e mantido por um grupo de programadores espalhados ao redor do mundo. Implementa recursos básicos de visualização espacial como *pam* (navegação), *zoom* (aproximação), sobreposição de camadas e adição de marcadores. É conectável com os serviços WMS e WFS e suporte para arquivos GeoRSS e KML. É um simples visualizador de dados espaciais, que requer, do usuário interessado em visualizar um conjunto de dados de interesse, conhecimentos em *script*. Não ofe-

rece controladores para tratamento da dimensão temporal em dados espaço-temporais. Mais informações estão disponíveis no endereço eletrônico: <http://openlayers.org/> (último acesso em Janeiro de 2012);

Google Maps: é um visualizador de dados espaciais para a *web* bastante utilizado. Ainda pode ser considerado um visualizador 2D, entretanto, os desenvolvedores da ferramenta já trabalham no desenvolvimento de uma solução 3D e na inclusão de recursos de realidade virtual. Oferece recursos de localização espacial, marcação de locais e determinação de rotas, além dos recursos básicos de um visualizador espacial: *zoom*, *pan*, e sobreposição de camadas pré-determinadas pela aplicação (trânsito, relevo, divisão política, etc). Não oferece recursos que possibilitem a visualização de um conjunto de dados de interesse nem tampouco suporte a dados espaço-temporais. Mais informações estão disponíveis no endereço eletrônico: <http://maps.google.com/> (último acesso em Janeiro de 2012);

Google Latitude: é um serviço que permite localizar amigos e visualizar suas localizações através do *Google Maps*. Usuários com *smartphones* equipados com GPS e conectados à Internet, ou apenas conectados à Internet podem compartilhar sua localização através do serviço. Ele oferece um visualizador espaço-temporal básico, utilizado na exibição do histórico de locais visitados, entretanto, é restrito à base de dados do *Google*, não permitindo a visualização de outros dados espaço-temporais de interesse do usuário. É importante ressaltar, também, que o recurso de linha do tempo confunde o usuário e pouco ajuda na visualização da variação temporal. Mais informações estão disponíveis no endereço eletrônico: <http://www.google.com/mobile/latitude/> (último acesso em Janeiro de 2012);

Google Earth: é um ambiente de visualização espacial 3D, inicialmente desenvolvido para *Desktop*, com funcionalidades similares às encontradas no *Google Maps*, além de permitir a inclusão de camadas sobrepostas ao mapa, a partir de arquivos em formato KML. A versão atual e gratuita apresenta controladores de animação temporal. Por meio de um *slider* temporal, é possível alternar a visualização das imagens de satélite em função do tempo. Apesar de permitir a visualização de dados de interesse do usuário, existe a restrição em só oferecer suporte a fontes de dados dispo-

nibilizadas em arquivos no formato KML. Mais informações no endereço eletrônico: <http://www.google.com/intl/pt-PT/earth/> (último acesso em Janeiro de 2012);

Quantum GIS: é um SIG desenvolvido para ambiente *Desktop*, que possui um visualizador de mapas 2D compatível com diversos formatos de dados espaciais, e um editor de dados espaciais, que possibilita a criação e modificação de geometrias. Diferente do *Google Maps*, não dispõe de um mapa como base. Apesar de apresentar grande quantidade de recursos para manipulação da dimensão espacial, não dá suporte à dimensão temporal nem à utilização de algoritmos de mineração de dados. Mais informações no endereço eletrônico: <http://www.qgis.org/> (último acesso em Janeiro de 2012);

OpenJUMP: é um SIG para ambiente *Desktop*, compatível com diversos formatos de dados espaciais. Possibilita a visualização e edição de dados espaciais, semelhante ao *Quantum GIS*. Também não oferece suporte a dados espaço-temporais. Mais informações no endereço eletrônico: <http://www.openjump.org/> (último acesso em Janeiro de 2012);

OECD eXplorer: é uma ferramenta gráfica interativa de análise de estatísticas regionais divulgadas pela OECD (*Organisation for Economic Co-operation and Development*), uma organização internacional composta por 34 países espalhados em quatro continentes. Trata-se de um ambiente *web* de visualização e análise espaço-temporal com dados estatísticos populacionais referente ao período compreendido entre 1990 e 2009, com diversos recursos visuais para análise: uma combinação de mapas temáticos, gráficos estatísticos e outros recursos visuais, de modo a permitir a visualização flexível dos dados de regiões de interesse do usuário. Um de seus pontos fracos é a não flexibilidade no domínio dos dados analisados, sendo exclusivo do domínio de dados sócio-econômicos. Mais informações no endereço eletrônico: <http://www.oecd.org/gov/regional/statisticsindicators/explorer/> (último acesso em Janeiro de 2012);

ArcGIS Explorer: é um SIG que possibilita a visualização, análise e compartilhamento de informações espaciais. Seu principal diferencial está no fato de oferecer recursos para publicação de dados espaciais proprietários para um grande público. Não

apresenta recursos para tratamento da dimensão temporal, e, portanto, não oferece suporte a dados espaço-temporais. Por também existir uma versão proprietária, sua versão gratuita é bastante limitada. Mais informações no endereço eletrônico: <http://www.esri.com/software/arcgis/explorer/index.htm> (último acesso em Janeiro de 2012).

3.3 Considerações

Em meio aos poucos trabalhos acadêmicos publicados que tratam da visualização e análise de dados espaço-temporais, uma parte aborda soluções voltadas para domínios específicos, sendo, portanto, úteis apenas para um limitado conjunto de usuários. Essas soluções, para serem utilizadas em outros domínios de aplicação, acabam por requerer novas soluções, também específicas, para os inúmeros domínios onde a análise espaço-temporal pode ser aplicada.

A situação é semelhante nas várias ferramentas gratuitas de SIG encontradas no mercado: poucas são, de fato, espaço-temporais e, quando são, geralmente abordam domínios específicos de aplicação. É possível perceber os esforços dos desenvolvedores em prover recursos que privilegiam apenas a dimensão espacial em dados espaço-temporais. As particularidades temporais ainda não recebem o devido tratamento nessas soluções.

Além disso, muitos trabalhos deixam a desejar quanto à flexibilidade na obtenção de dados, muitas vezes requerendo esforço considerável, por parte dos usuários, para adaptar seus conjuntos de dados à aplicação escolhida para realizar a análise, ou então nem permitindo isso, sendo verdadeiras "caixas pretas".

Existem também problemas relacionados à usabilidade, com interfaces que não dão a liberdade necessária ao usuário no sentido de permitir incluir ou remover camadas de informação que julgar relevantes à tarefa que está sendo realizada.

Conforme discussão na seção 2.4, uma boa solução em visualização e análise de dados espaço-temporais deve, portanto, oferecer, no mínimo, seis principais características: (1) Espacialidade, (2) Temporalidade, (3) Genericidade, (4) Flexibilidade, (5) Interoperabilidade, e (6) Mineração de Dados. A partir dessa discussão, realizou-se uma avaliação comparativa entre todos os trabalhos científicos e todas as soluções gratuitas de SIG relacionadas nesta

Tabela 3.1: Tabela comparativa entre os trabalhos acadêmicos de visualização espaço-temporal relacionados nesta seção.

Trabalho / Critério	1	2	3	4	5	6
<i>SocioScope</i>	X	X				
<i>Exploratory spatio-temporal data mining and visualization</i>	X	X	X	X		X
<i>Spatio-Temporal Visual Query Environment</i>	X	X	X	X	X	
<i>Visual Analytics Tools for Analysis of Movement Data</i>	X	X	X	X		X
<i>A Framework for Exploring Spatiotemporal Patterns</i>	X	X	X			X
<i>COPLINK: Visualization for Crime Analysis</i>	X	X				X
<i>BirdVis: Visualizing and Understanding Bird Populations</i>	X	X		X		
<i>GeoVISTA CrimeViz</i>	X	X		X		

seção. Os trabalhos foram avaliados quanto à presença ou ausência dessas características principais. Os resultados são apresentados nas Tabelas 3.1 e 3.2.

Como é possível observar na Tabela 3.1, todas as soluções apresentadas lidam, de alguma forma, com as dimensões espacial e temporal, entretanto, apenas uma leva em consideração a necessidade de permitir conexão de dados entre diversas fontes. Também é importante ressaltar a inclusão de mecanismos de mineração de dados, que já é contemplada em pelo menos metade dessas soluções.

Na Tabela 3.2, fica evidente que a maioria das soluções relacionadas não trata a dimensão temporal, e que nenhuma delas utiliza mineração de dados. Outro ponto importante é que as soluções que são, de fato, espaço-temporais, não apresentam interoperabilidade na comunicação com bases de dados.

A dificuldade em encontrar soluções espaço-temporais genéricas pode ser justificada na complexidade inerente à construção delas.

Diante deste cenário, fica evidente a necessidade de novas soluções que ofereçam recursos de manipulação de dados espaço-temporais que levem em conta as particularidades da dimensão espacial e também da dimensão temporal. Além disso, soluções genéricas e com recursos de visualização intuitivos são importantes para possibilitar o acesso por um maior número de usuários. É necessária uma abordagem que priorize não somente o contexto

Tabela 3.2: Tabela comparativa entre as soluções gratuitas de SIG espacial e espaço-temporal relacionadas nesta seção.

Trabalho / Critério	1	2	3	4	5	6
<i>OpenLayers</i>	X		X	X	X	
<i>Google Maps</i>	X					
<i>Google Latitude</i>	X	X				
<i>Google Earth</i>	X	X	X	X		
<i>Quantum GIS</i>	X		X	X	X	
<i>OpenJUMP</i>	X		X	X	X	
<i>OECD eXplorer</i>	X	X		X		
<i>ArcGIS Explorer</i>	X		X		X	

espaço-temporal, mas também o usuário final.

3.4 Resumo do Capítulo

Este capítulo apresentou os principais trabalhos relacionados com a proposta deste trabalho de dissertação. Relacionando as principais ferramentas gratuitas de SIG e os principais trabalhos acadêmicos da atualidade, este capítulo promoveu uma discussão sobre as abordagens existentes para visualização e análise de dados espaço-temporais.

Em meio aos trabalhos acadêmicos e, com base nas características indispensáveis para uma solução de análise e visualização de dados espaço-temporais, na visão de Andrienko et al. (2010b), enumerou-se seis critérios essenciais a uma solução de visualização espaço-temporal, realizando uma avaliação comparativa entre todos os trabalhos relacionados.

Ficou evidente a necessidade de uma nova abordagem de visualização e análise sobre dados espaço-temporais que seja flexível, prática, e genérica.

No próximo capítulo, será apresentada uma nova abordagem para visualização e análise de dados espaço-temporais, que visa minimizar as lacunas encontradas nas abordagens discutidas neste capítulo.

Capítulo 4

Uma abordagem para visualização e análise de dados espaço-temporais

Neste capítulo, será apresentada uma nova abordagem em visualização de dados espaço-temporais. Essa nova abordagem reúne, basicamente, algumas técnicas para análises visuais e levantamento de evidências em dados espaço-temporais.

O principal objetivo dessa nova abordagem é minimizar as lacunas encontradas nas soluções relacionadas no capítulo anterior, no tocante às questões de espacialidade, temporalidade, genericidade, flexibilidade, interoperabilidade e mineração de dados, e fomentar a área de análise visual espaço-temporal. O resultado é o escopo de uma nova ferramenta de visualização e análise de dados espaço-temporais, cuja implementação será detalhada neste capítulo.

4.1 Técnicas e Recursos de Visualização reunidos

A análise visual é uma tarefa realizada com muita facilidade pelos humanos. Para Andrienko et al. (2010b), a visualização é o meio pelo qual os seres humanos e os computadores cooperam com as suas capacidades distintas para obtenção de resultados mais eficazes. Essa ideia foi o fator gerador de muitos esforços de pesquisa em áreas como visualização da informação, ciência de informações geográficas (*GIScience*), geovisualização e mineração de dados.

Nesta seção, são detalhadas as técnicas de visualização reunidas na nova abordagem pro-

posta. São técnicas que tratam tanto as características inerentes aos dados espaciais, como também àquelas inerentes aos dados temporais, facilitando a visualização dos dados espaço-temporais por parte do usuário final¹. Além de incrementar a experiência em visualização espaço-temporal, a nova abordagem proposta também tem como objetivo permitir a realização de análise visual espaço-temporal. As técnicas reunidas são apresentadas e discutidas nas subseções seguintes.

4.1.1 Para a Dimensão Espacial

A dimensão espacial recebe muita atenção por boa parte das soluções publicadas na comunidade, como se pode observar claramente no Capítulo 3. É evidente que uma abordagem em visualização de dados espaço-temporais deve oferecer recursos que enriqueçam o raciocínio e facilitem os usuários na tarefa de análise espacial visual. Entretanto, é importante levar em consideração o fato de que uma interface de usuário com recursos em demasia pode trazer um efeito contrário àquele desejado por projetistas e desenvolvedores, tornando confusa a interpretação dos dados por parte dos usuários finais.

Dessa forma, evitando incluir recursos que possam prejudicar a experiência do usuário, a nova abordagem reúne os seguintes recursos de visualização de dados espaciais:

- Visualizador baseado em mapas dinâmicos georeferenciados (*Dynamic Map Viewer*);
- Sobreposição de camadas com dados de interesse do usuário (*Map Layers*);
- Consulta espacial (*Spatial Query*).

Um visualizador baseado em mapas georeferenciados é indispensável em uma abordagem para visualização de dados espaço-temporais. Sem ele não é possível visualizar, com facilidade cognitiva, os relacionamentos espaciais, como, por exemplo, vizinhança, distância e intersecção. É importante ressaltar a importância de um mapa georeferenciado como base para o visualizador de mapas. A partir dele, o usuário poderá incluir camadas sobrepostas

¹O termo usuário final é empregado no contexto de analistas de dados espaço-temporais, ou seja, especialistas em determinados domínios de dados espaço-temporais interessados em realizar descobertas sobre uma massa de dados transacionais significativa.

com informações de interesse para a análise, e visualizar relacionamentos geográficos de forma eficaz.

Atualmente, é possível encontrar várias imagens georeferenciadas que servem como bons mapas-base para visualizadores espaciais. São imagens atualizadas que apresentam, por exemplo, dados sobre relevo, hidrografia, divisão política, estradas, rodovias, cidades, entre outros.

Outro recurso indispensável a um visualizador de dados espaciais é o recurso de inclusão de camadas sobrepostas. É por meio de camadas sobrepostas que são adicionadas a uma visualização as informações de domínios específicos que se deseja visualizar e analisar. Alguns exemplos de camadas sobrepostas: pontos de ocorrência de casos de dengue; delimitações territoriais geográficas das mesoregiões brasileiras; rotas das rodovias federais brasileiras; etc. É interessante que este recurso ofereça opções de transparência e estilo. O usuário deve ter a liberdade de selecionar o nível de transparência desejada para cada camada sobreposta adicionada, em uma escala que pode variar entre 0% (totalmente visível) e 100% (invisível). Da mesma forma, é interessante que o usuário possa definir estilos para as camadas, definindo padrões iconográficos ou de variação de coloração e tonalidade, que possibilitem destacar os dados dispostos sobre o mapa, de maneira que sua identificação seja intuitiva.

Por fim, o recurso de consulta espacial. Por meio da consulta espacial, o usuário poderá analisar o relacionamento espacial entre duas camadas sobrepostas. Por exemplo, ao visualizar uma camada contendo uma rota percorrida por um caminhoneiro e, diante de uma camada contendo os estados e as sedes municipais, uma consulta espacial pode ser aplicada sobre essas duas camadas solicitando apenas as sedes municipais e os estados em que tal rota faz intersecção.

A consulta espacial pode ser aplicada, ainda, pelo usuário interessado em visualizar apenas uma partição de uma grande base de dados espaciais, permitindo a definição de filtros baseados nos atributos convencionais desses dados, ou nos resultados obtidos a partir da consulta espacial.

4.1.2 Para a Dimensão Temporal

A ausência de tratamento da dimensão temporal é a principal lacuna percebida nos trabalhos publicados na comunidade científica, como é possível observar também no Capítulo 3. Exis-

tem muitos trabalhos cujas soluções abordam apenas recursos para a visualização de dados e relacionamentos espaciais.

Visando oferecer suporte à dimensão temporal, mas também com cautela para não oferecer recursos em demasia e desnecessários ao usuário, essa abordagem reúne os seguintes recursos de visualização de dados temporais:

- Controlador temporal animado (*Temporal Slider*);
- Gráfico de distribuição temporal (*Distribution Temporal Chart*);
- Consulta temporal (*Temporal Query*).

O controlador temporal é um recurso importante para a visualização de dados temporais. É a partir dele que o usuário pode determinar uma marcação de tempo exata ou um período de tempo de interesse para visualização dos dados. É interessante que, além da função de determinação temporal para a visualização, o controlador temporal tenha o papel de controlador de animação temporal, de forma que permita ao usuário visualizar animações interativas a partir da variação espacial das informações a medida que a linha do tempo evolui. A interatividade é importante, com botões para início, pausa e deslocamento da linha do tempo, dando liberdade de visualização temporal ao usuário.

Adicionar o recurso de controlador temporal a um visualizador espacial possibilita ao usuário a experiência de observar mapas animados em função do tempo, possibilitando então análises espaço-temporais, uma vez que, em uma só visualização, será possível observar tanto relacionamentos espaciais como também relacionamentos temporais em meio aos dados espaço-temporais observados.

Ainda se tratando do controlador temporal, é importante observar a granularidade. O usuário deve ter a liberdade de determinar a granularidade temporal que deseja utilizar no controlador. Pode ser interessante para um usuário visualizar os dados em blocos de periodicidade anual (uma abordagem mais abstrata), entretanto, para outro usuário, e dependendo do domínio da informação que este deseja analisar, pode ser interessante visualizar os dados como blocos diários (uma abordagem mais aprofundada).

O gráfico de distribuição temporal constitui outro recurso importante para a visualização de dados temporais e a extração de conhecimentos implícitos, eventualmente presentes

nesses dados. Por meio da distribuição temporal, o usuário poderá observar periodicidade, ou seja, se existe relacionamento temporal nos eventos visualizados, se eles se repetem em períodos regulares (semanal, mensal, anual, etc.), ou acontecem de forma linear. É possível, também, por meio do gráfico de distribuição temporal, obter a marcação temporal, na granularidade temporal adotada, em que foram registradas o máximo e o mínimo de ocorrências de um dado evento.

Enquanto a visualização espacial foca nas regiões que registraram os eventos, a visualização temporal, por meio do gráfico de distribuição, foca nas marcações ou períodos de tempo que registraram os eventos. Ambas visualizações compartilham o quantitativo e os demais atributos relacionados à ocorrência dos eventos analisados.

O recurso de consulta temporal, por sua vez, visa oferecer ao usuário opções para visualização de relacionamentos temporais entre camadas espaço-temporais sobrepostas. Pode ser interessante, por exemplo, diante de uma camada contendo dados de estados brasileiros atingidos por enchentes no período entre 1990 e 2010, e diante de uma outra camada contendo focos de queimadas registrados em todo o país no período entre 2004 e 2008, realizar uma consulta temporal de forma que se visualize ambas as camadas no período compreendido entre 2005 e 2007. Apesar da primeira camada conter dados do período entre 1990 e 2010, nada impede que no período entre 2005 e 2007 não exista nenhuma ocorrência, por exemplo. O uso desse recurso possibilitaria ao usuário obter uma resposta de forma eficaz e intuitiva.

4.2 O uso de *Clustering* para mineração de dados Espaço-Temporais

Uma vez que o objetivo da nova abordagem proposta não é tão somente prover um visualizador de dados espaço-temporais com recursos necessários a uma boa experiência do usuário final, mas também, permitir a realização de análises visuais sobre dados espaço-temporais, recorreu-se aos algoritmos de mineração de dados com o objetivo de oferecer mecanismos de processamento computacional que vão além da percepção visual dos dados transacionais.

Com o uso dos algoritmos de mineração de dados aplicados a dados espaço-temporais, e os resultados exibidos em um visualizador dotado de recursos para tratamento da dimensão espacial e temporal, será possível promover uma melhor experiência em análise de dados

espaço-temporais.

Para o desenvolvimento desta abordagem e para promover o levantamento de evidências, adotou-se os algoritmos de mineração de dados baseados em agrupamento (*clustering*) espaço-temporal.

Algoritmos de agrupamento (*clustering*) podem revelar padrões de distribuição e agrupamento espaço-temporal em meio aos dados analisados. Por meio de *clustering* é possível promover um estudo de vizinhança de localização no espaço-tempo, por exemplo. É possível encontrar regiões e períodos de grande densidade, ou seja, em que grandes porções de dados se concentram no espaço-tempo. É possível, também, perceber eventos que acontecem raramente em determinadas regiões.

No domínio da criminalidade, por exemplo, algoritmos de *clustering* podem ajudar na percepção dos bairros de uma cidade onde se concentram maior parte dos crimes registrados e os momentos (do dia, da semana, do mês, etc) em que se observa maior atividade criminosa. Essa percepção vai possibilitar esforços proativos das autoridades, no sentido de otimizar a alocação de recursos para regiões e faixas de horário que mais precisam de policiamento, objetivando reduzir o índice de criminalidade nessas regiões (*outliers*).

A concepção dessa abordagem não aponta um ou outro algoritmo de *clustering*, apenas considera as vantagens oferecidas com o uso dessa técnica de mineração de dados, independente do algoritmo que venha ser utilizado.

A nova abordagem proposta reúne, por fim, a funcionalidade de agrupamento espaço-temporal entre duas camadas, objetivando possibilitar a visualização espaço-temporal de dois eventos distintos num espaço-tempo suficientemente próximo, e permitir o estudo acerca da relação espaço-temporal de ocorrência desses eventos. A ocorrência de proximidade espaço-temporal entre dois eventos pode revelar uma relação do tipo causa e consequência, e com isso ajudar usuários interessados em minimizar a ocorrência de um evento em função de outro.

4.3 O GeoSTAT - *Geographic SpatioTemporal Analysis Tool*

Esta seção introduz o GeoSTAT (*Geographic SpatioTemporal Analysis Tool*), uma implementação da nova abordagem de visualização e análise espaço-temporal proposta neste tra-

balho.

4.3.1 Visão Geral

O *Geographic SpatioTemporal Analysis Tool* - GeoSTAT é um ambiente *web* para análise e visualização de dados espaço-temporais. Em razão da arquitetura *web*, o usuário interessado em utilizar o GeoSTAT fica livre de instalações de *softwares* e sistemas de bancos de dados. É necessário apenas dispor de um *browser* (utilizado para navegação na *web*) que ofereça suporte à linguagem *JavaScript*, podendo, inclusive, fazer uso a partir de dispositivos móveis como *Smartphones* e *Tablets*, em razão de não haver necessidade de instalação de *plugins* ou outros complementos.

O ambiente foi construído com o objetivo de validar a nova abordagem para visualização e análise de dados espaço-temporais proposta neste trabalho e de oferecer uma nova solução para tratamento visual de dados espaço-temporais. Dessa forma, o ambiente do GeoSTAT implementa o conjunto de técnicas e recursos apresentados anteriormente.

Além das vantagens de uma aplicação *web*, o GeoSTAT foi concebido sob o ponto de vista da genericidade. Por esta razão, é um ambiente independente de domínio e conectável a qualquer fonte de dados espaço-temporais disponível na *web* por meio da implementação de serviços de compartilhamento de dados espaciais especificados e padronizados pela OGC.

As funcionalidades, arquitetura e as principais questões de implementação são apresentadas nas subseções seguintes.

4.3.2 Apresentação do Ambiente

Por meio do ambiente GeoSTAT, o usuário interessado em analisar um conjunto de dados espaço-temporais poderá visualizá-lo sobre um mapa georeferenciado e terá à sua disposição recursos de visualização que tratam tanto a dimensão espacial como também a dimensão temporal, além de total integração com algoritmos de mineração de dados baseada em *clustering*, de forma que seja possível visualizar os resultados do processo de mineração sobre o mapa e permitir a realização de análises visuais nestes dados. A Figura 4.1 apresenta a interface interativa do GeoSTAT.

Na Figura 4.1, é possível observar a presença dos controladores espaciais e temporais



Figura 4.1: Interface interativa do GeoSTAT.

na interface do ambiente GeoSTAT. Foram implementados os recursos de visualização baseada em mapas, camadas sobrepostas, permitindo a visualização de dados de interesse do usuário, consultas espaciais e temporais, controlador temporal interativo, e gráficos de distribuição temporal.

Além disso, sete algoritmos de mineração de dados baseados em *clustering* espaço-temporal foram integrados ao ambiente, de maneira que fosse possível interagir com o visualizador, no sentido de obter dados pré-processados para o processo de mineração a partir de interações do usuário com o ambiente, e, também, obter os resultados de maneira que pudessem ser manipulados visualmente pelo usuário final. Tais algoritmos foram obtidos a partir da ferramenta de software livre para mineração de dados Weka (Hall et al., 2009) e são apresentados a seguir:

COBWEB: é um algoritmo incremental que agrupa observações em uma árvore de classificação. Cada nó da árvore de classificação representa uma classe (conceito) e é marcado por um conceito probabilístico, que resume as distribuições atributo-valor de objetos classificados sob o nó. Esta árvore de classificação pode ser usada para prever atributos ausentes ou a classe de um novo objeto. Mais informações em Gennari et al. (1990).

DBScan: é um algoritmo de agrupamento baseado em densidade, que encontra um certo número de *clusters* a partir da distribuição da densidade estimada de nós correspondentes. Apenas dois parâmetros são requeridos: `minPoints` (Quantidade mínima de pontos para formar um *cluster*) e `radius` (Raio de vizinhança a partir do ponto central de um *cluster*). É um dos algoritmos de agrupamento mais comum e também mais citado na literatura. Mais informações em Ester et al. (1996).

K-Means: é um método de agrupamento que se destina à partição de n observações em k *clusters*, no qual cada observação pertence ao *cluster* com média mais próxima. Mais informações em Arthur e Vassilvitskii (2007).

X-Means: é o algoritmo K-Means melhorado em parte de sua estrutura. Mais informações em Dan Pelleg (2000).

Expectation-Maximization: conhecido como EM, é um algoritmo que atribui uma distribuição de probabilidade para cada instância, indicando a probabilidade de pertencer a cada um dos *clusters*. É capaz de decidir quantos *clusters* serão criados a partir de uma validação cruzada, sem deixar de permitir que o número de *clusters* a serem gerados seja definido *a priori*. Mais informações em Hastie et al. (2001).

Farthest-First: é uma variação do *K-Means* que coloca cada centro de *cluster* no ponto mais afastado dos centros dos *clusters* existentes. Mais informações em Dasgupta e Long (2005).

OPTICS: é um método de agrupamento que pode ser visto como uma generalização do DBScan para vários intervalos, substituindo de forma eficaz o parâmetro de busca com um raio máximo. Mais informações em Ankerst et al. (1999).

4.3.3 Arquitetura

A arquitetura do GeoSTAT foi definida em três camadas interconectadas: visualização, controle e persistência. A Figura 4.2 ilustra essa arquitetura.

A seguir, apresentamos em detalhes cada uma das camadas que compõem a arquitetura do ambiente GeoSTAT.

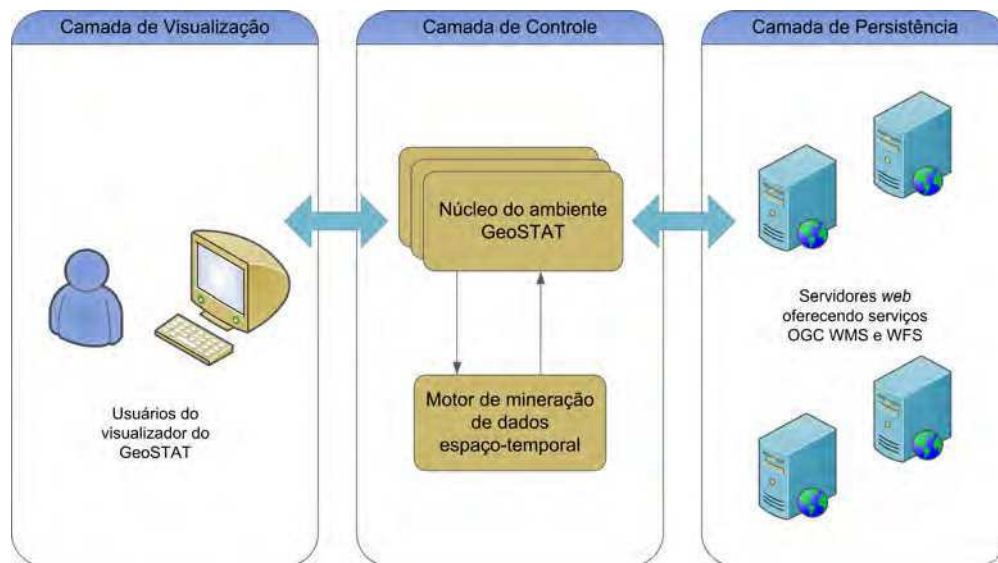


Figura 4.2: Arquitetura do GeoSTAT em três camadas.

Camada de Visualização

A camada de visualização é responsável pela interface com o usuário, oferecendo diversos componentes para carga de dados espaço-temporais e manipulação das dimensões espacial e temporal. Trata-se de um ambiente *web*, desenvolvido com as tecnologias *Java Server Pages*², *AJAX*³, *jQuery*⁴ e *HTML5*⁵.

O visualizador do GeoSTAT é composto por dez componentes intercomunicáveis: mapa interativo, camadas espaço-temporais, controlador temporal dinâmico, filtro temporal, filtro espacial, gráfico de distribuição temporal, mineração de dados, resultados da mineração de dados, menu de ações, e informações dos servidores conectados.

Todos os dez componentes se comunicam com a camada de controle despachando requisições disparadas a partir de eventos do usuário, por meio de comandos realizados na interface interativa do ambiente, e recebendo o retorno dessas requisições. Todos os componentes e suas funcionalidades serão detalhados na subseção seguinte.

²*JavaServer Pages Technology* - Oracle. Disponível em (inglês):

<http://www.oracle.com/technetwork/java/javaee/jsp/index.html>

³*AJAX - Asynchronous Javascript and XML*. Mais informações em:

[http://pt.wikipedia.org/wiki/AJAX_\(programação\)](http://pt.wikipedia.org/wiki/AJAX_(programação))

⁴*jQuery - write less do more*. Disponível em (inglês): <http://jquery.com/>

⁵*HTML5 - Hypertext Markup Language*, versão 5. Mais informações em:

<http://pt.wikipedia.org/wiki/HTML5>

Camada de Controle (Núcleo)

A camada de controle é a responsável pelo processamento de todas as requisições geradas e encaminhadas a partir da camada de visualização, além de também ser responsável pela comunicação com a camada de persistência, sendo, portanto, o núcleo do GeoSTAT.

A Figura 4.3 apresenta os módulos existentes na camada de controle. Estes módulos são acionados de acordo com a natureza da requisição a ser processada por esta camada.

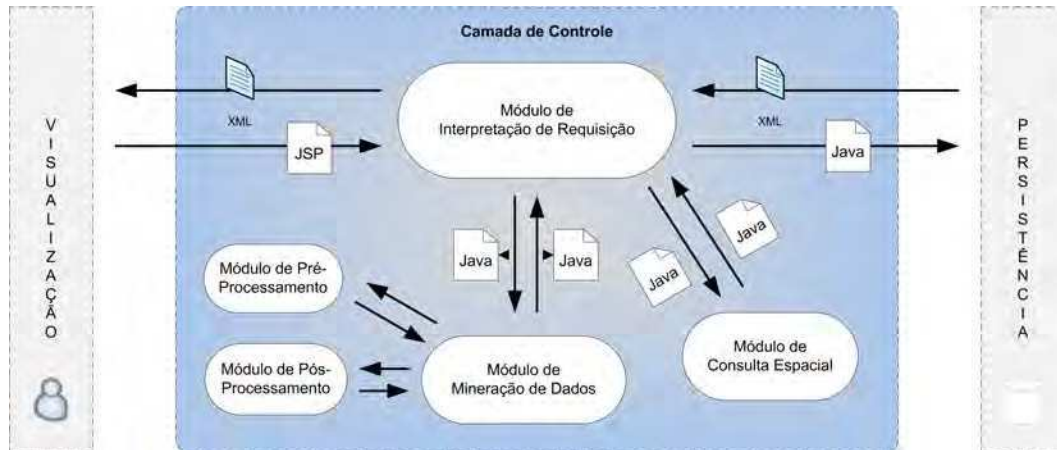


Figura 4.3: Módulos da camada de controle da arquitetura do GeoSTAT.

Como se pode observar na Figura 4.3, a camada de controle do GeoSTAT possui cinco módulos intercomunicáveis: Interpretação de Requisição, Pré-Processamento, Mineração de dados, Pós-Processamento e Consulta Espacial.

O módulo de interpretação de requisição é o módulo principal da camada de controle. Ele é o responsável por receber e tratar todas as requisições vindas da camada de visualização, de estabelecer contato com os demais módulos, quando a natureza da requisição assim exigir, e, também, de realizar o contato com a camada de persistência.

Diversas requisições AJAX (*Javascript* Assíncrono) são disparadas a partir da interação do usuário por meio dos recursos disponibilizados na interface do ambiente do GeoSTAT. Essas requisições são enviadas diretamente à camada de controle, que as recebe por meio de scripts JSP (*Java Server Pages*). Esses scripts são interpretados e, dependendo da natureza da requisição, recebem o devido tratamento.

Basicamente, existem dois tipos de tratamento para as requisições que chegam ao módulo de interpretação de requisição: as requisições de solicitação de dados e as requisições

de solicitação de processamento. As requisições de solicitação de dados são encaminhadas diretamente à camada de persistência, uma vez que é de sua competência a interpretação e processamento delas. Quando a camada de persistência envia a resposta de uma requisição, ela é então encaminhada novamente para a camada de visualização, em formato GML ou em formato XML. Por outro lado, as requisições de solicitação de processamento de dados podem ser encaminhadas ao módulo de mineração de dados ou ao módulo de consulta espacial.

O módulo de consulta espacial é o responsável por realizar o processamento das consultas espaciais entre duas camadas distintas. Sua construção foi necessária em razão dos serviços OGC limitarem as consultas espaciais a uma única camada. O resultado do processamento da consulta (filtro espacial) é encaminhado à camada de visualização para ser exibido ao usuário final.

Como já apresentado na subseção anterior, o módulo de mineração de dados integra sete algoritmos de *clustering* conhecidos. Com isso, o GeoSTAT é capaz de realizar mineração de dados espaço-temporal baseada em *clustering* com qualquer banco de dados espacial ou espaço-temporal, em que a dimensão espacial que ofereça geometrias do tipo ponto bidimensional. A saída retornada pelo módulo de mineração de dados é armazenada em um banco de dados espaço-temporal e disponibilizada para consulta a partir do ambiente, tão logo o processamento esteja completo.

Para tornar possível a integração do GeoSTAT com os sete algoritmos de mineração de dados utilizados, foram desenvolvidos os módulos de pré e pós-processamento de dados. Estes módulos auxiliares são responsáveis por preparar os dados para serem utilizados pelo algoritmo selecionado a partir do usuário, e preparar os resultados obtidos com a execução desse algoritmo para serem tratados pela camada de visualização, respectivamente. A Figura 4.4 apresenta um diagrama de sequência representando as transações realizadas entre alguns módulos da arquitetura do GeoSTAT ao processar uma requisição de mineração de dados.

Conforme ilustrado no diagrama da Figura 4.4, ao receber uma requisição para ser processada, o módulo de mineração de dados solicita os dados da camada espaço-temporal à camada de persistência. De posse dos dados que serão minerados, é a vez do módulo de pré-processamento entrar em ação. Este módulo é o responsável por fazer a leitura do conjunto de dados, obtido em formato GML a partir de uma requisição *WFS GetFeature*, e transformá-

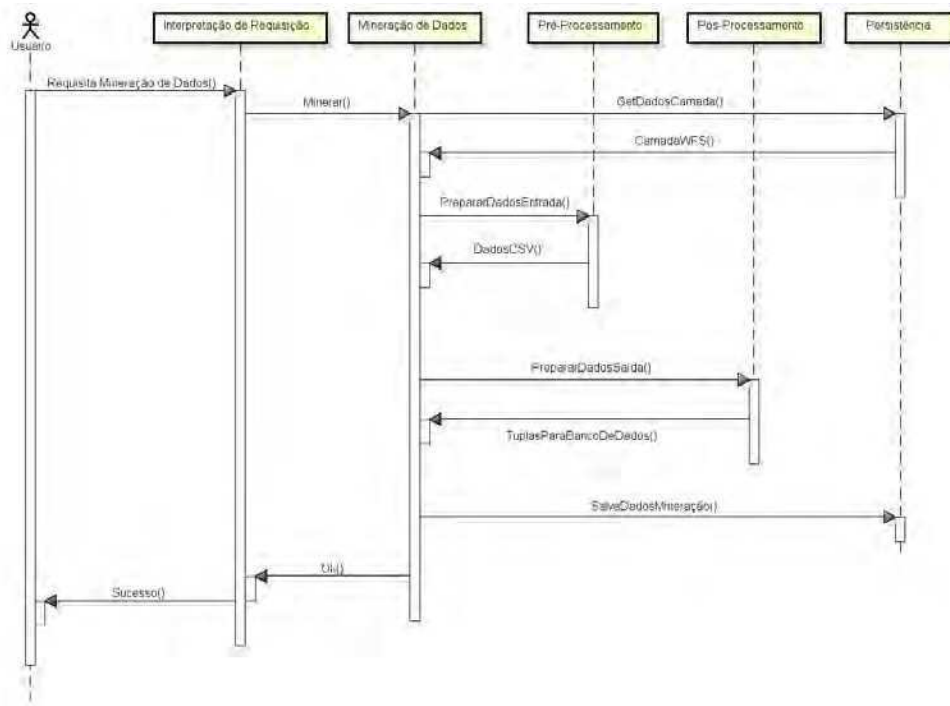


Figura 4.4: Diagrama de sequência com as operações realizadas no processamento de uma requisição de mineração de dados.

lo em um arquivo temporário em formato CSV (*Comma Separated Values*) criado em tempo de execução. O arquivo CSV criado a partir do conjunto de dados de origem é utilizado pelo algoritmo de mineração selecionado como parâmetro de entrada. Este arquivo preserva apenas os atributos referentes à chave primária, dimensão espacial e dimensão temporal do conjunto de dados de origem. Concluída a execução do algoritmo de mineração, que se dá no módulo de mineração de dados, o arquivo CSV temporário é automaticamente destruído.

O módulo de pós-processamento é acionado pelo módulo de mineração de dados assim que um processo de mineração é concluído. O papel do módulo de pós-processamento é preparar os resultados do processo de mineração para serem corretamente disponibilizados e analisados na camada de visualização do GeoSTAT. É nesse módulo que o resultado da mineração de dados é armazenado em um banco de dados espacial *PostgreSQL+PostGIS* e, com isso, disponibilizado para acesso a partir da camada de visualização do ambiente.

O módulo de mineração de dados é o núcleo do processamento da mineração de dados. Ele controla a execução do algoritmo selecionado, separando os dados de entrada em grupos para mineração com base na dimensão temporal e gerenciando a execução de 12 *threads*

concorrentes. Cada *thread* executa o algoritmo com um grupo de dados de entrada, otimizando o processamento e reduzindo o custo da mineração de dados. Ao final de cada *thread*, os resultados (*clusters*) são encaminhados diretamente ao módulo de pós-processamento, e a *thread* recebe então um novo grupo de dados para mineração, repetindo o processo até que a fila de grupos de dados para minerar esteja vazia. A Figura 4.5 detalha o módulo de mineração de dados, ilustrando seus principais componentes.

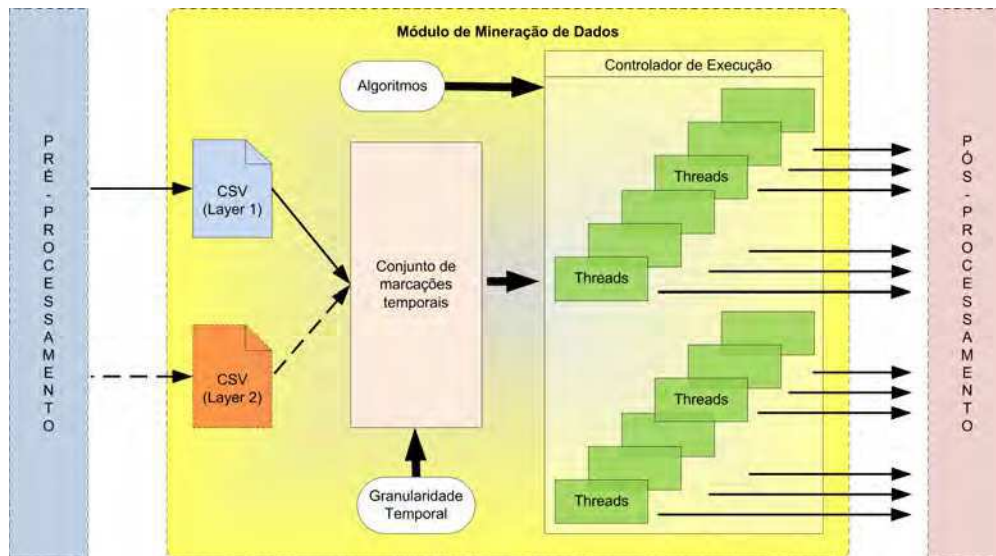


Figura 4.5: Esquema de funcionamento do Módulo de Mineração de Dados.

O código-fonte 4.1 apresenta um pseudo-código do algoritmo que executa a mineração de dados no GeoSTAT. Esse algoritmo representa uma abstração do funcionamento interno do módulo de mineração de dados, detalhado na Figura 4.5. As variáveis representam os dados obtidos do módulo de pré-processamento. O método `AgrupamentoEspacial`, além de utilizar o algoritmo de clustering selecionado, envia os resultados para o módulo de pós-processamento.

Camada de Persistência

A camada de persistência é a responsável pela conexão do GeoSTAT às bases de dados solicitadas pelos usuários a partir dos componentes da camada de visualização. Ao receber uma solicitação de dados da camada de controle, a camada de persistência identifica primeiramente o tipo de acesso que realizará, e em seguida realiza a solicitação junto ao servidor de serviços OGC ou banco de dados espaço-temporal. A Figura 4.6 apresenta os dois tipos de

acesso a dados espaço-temporais implementados na camada de persistência do GeoSTAT.

Código Fonte 4.1: Pseudo-código do algoritmo de execução da mineração de dados espaço-temporais no GeoSTAT

```

VAR DataInicial , DataFinal : datahora ;
VAR CSV1, CSV2, MarcasTemporais : matriz ;
VAR GranTemporal , Algoritmo , i : inteiro ;
INICIO
DataInicial = MenorData(CSV1, CSV2);
DataFinal = MaiorData(CSV1, CSV2);
MarcasTemporais = GerarMarcas(GranTemporal , DataInicial , DataFinal);
AgrupamentoTemporal(MarcasTemporais , CSV1, CSV2);
PARA i DE 1 ATÉ Tamanho(MarcasTemporais) FAÇA
    AgrupamentoEspacial(algoritmo , MarcasTemporais [ i ] );
FIM PARA
FIM
  
```

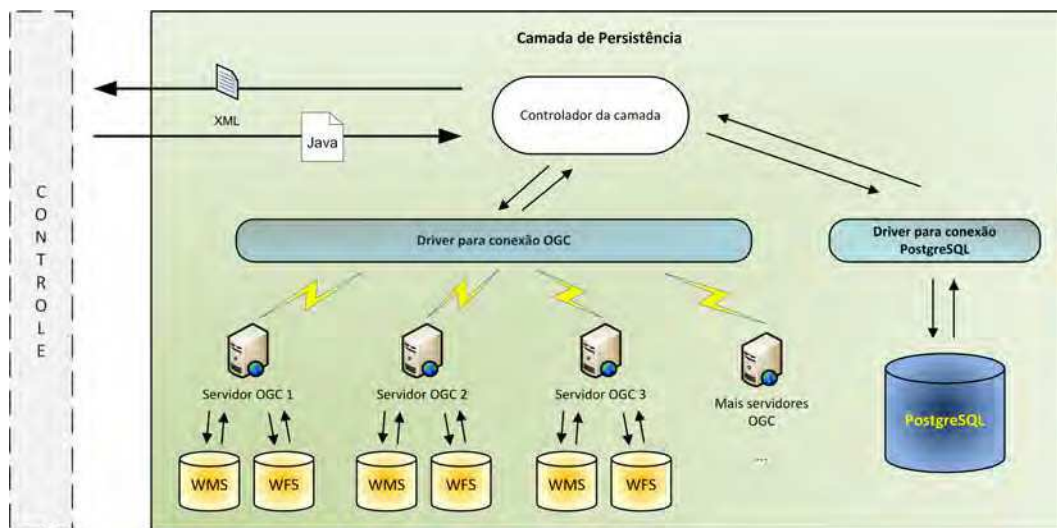


Figura 4.6: Estrutura interna da camada de persistência da arquitetura do GeoSTAT.

A camada de persistência pode realizar conexões aos serviços OGC WMS e WFS ou a um banco de dados espaço-temporal *PostgreSQL+PostGIS* desenvolvido exclusivamente para operar com o GeoSTAT (vide Figura 4.6). Os serviços OGC são acessados a partir de seus servidores *web*. Dois servidores *web* bastante conhecidos que oferecem diversos

serviços OGC são o Geoserver⁶ e o Mapserver (Lime, 2008).

A camada de persistência implementa métodos para conexão e obtenção de dados disponibilizados por meio dos serviços OGC. Com a utilização da biblioteca GeoTools (Turton, 2008), esses métodos realizam operações de obtenção de camadas espaço-temporais e de formulação de consultas espaciais e temporais, acionadas pelo usuário a partir dos recursos de filtro espacial e filtro temporal.

O serviço WMS é requisitado quando o usuário do ambiente deseja, por exemplo: obter a lista de camadas disponibilizadas pelo servidor conectado; visualizar uma camada espacial ou espaço-temporal sobrepondo o mapa base do ambiente; ou obter informações detalhadas de uma camada sobre uma área ou ponto específico do mapa.

O serviço WFS, por sua vez, é requisitado quando o usuário do ambiente realiza operações do tipo: obter atributos de uma camada espacial ou espaço-temporal, seja eles do tipo espacial, temporal ou convencional; realizar consultas espaciais ou temporais e obter apenas os dados que satisfaçam os critérios da consulta (filtro); ou obter os conjuntos de dados espaciais ou espaço-temporais que serão utilizados por um processo de mineração de dados.

No banco de dados *PostgreSQL+PostGIS*, ficam armazenadas as informações utilizadas pelo GeoSTAT para conexão aos serviços OGC, e os resultados completos dos processos de mineração de dados realizados por meio do ambiente e disponíveis para visualização. Os dados de conexão informados pelo usuário no momento de criar uma nova conexão de mapa são armazenados nesse banco de dados e solicitados no momento em que o usuário seleciona a conexão de dados onde estão disponibilizadas as camadas espaciais ou espaço-temporais que utilizará no ambiente.

A Figura 4.7 apresenta o esquema relacional desenvolvido para o banco de dados *PostgreSQL+PostGIS* de suporte ao GeoSTAT. A tabela "*connection*" é a responsável por armazenar os dados das conexões do ambiente aos servidores OGC espalhados na *web*. As tabelas "*mining_result*", "*mining_result_cluster*" e "*mining_result_point_of_cluster*" são responsáveis pelo armazenamento dos resultados dos processos de mineração de dados executados a partir do ambiente. Por fim, as tabelas "*spatial_layer1*" e "*spatial_layer2*" são tabelas de suporte à realização de consultas espaciais. Elas são populadas apenas durante o processa-

⁶*GeoServer - OpenGeo*. Mais informações em (inglês):

<http://geoserver.org/display/GEOS/What+is+Geoserver/>

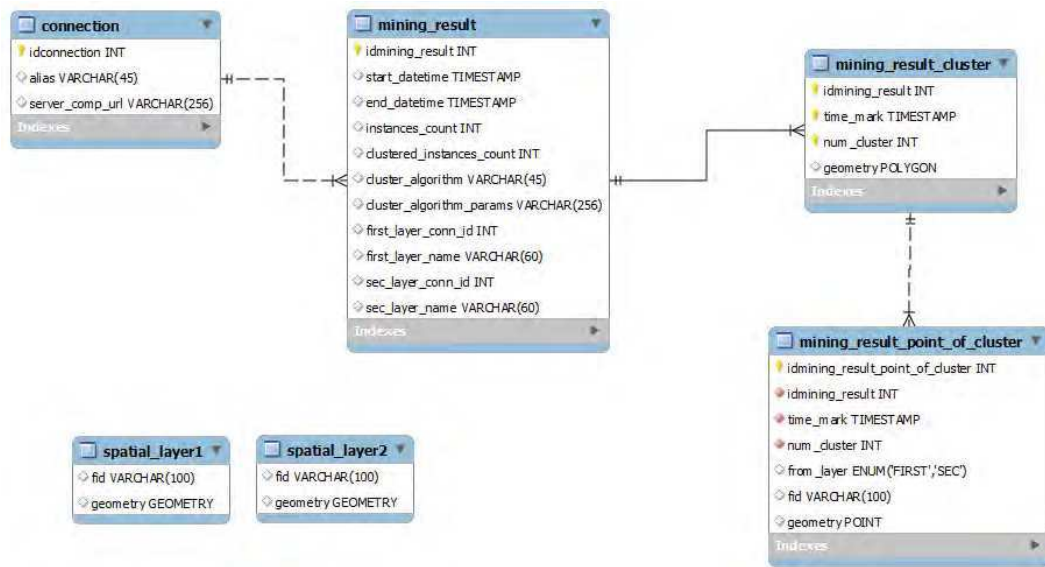


Figura 4.7: Esquema relacional do banco de dados espaço-temporal utilizado pelo GeoSTAT.

mento de uma consulta espacial e são esvaziadas tão logo a consulta esteja concluída.

As respostas da camada de persistência são enviadas à camada de controle já em formato XML interpretável pela camada de visualização. Para as respostas às requisições encaminhadas a serviços WMS e WFS, o retorno é o GML fornecido pelo próprio serviço. Para as respostas às requisições ao banco de dados da aplicação, o retorno é em XML, obtido a partir dos resultados de consultas SQL.

O GeoSTAT possibilita a experiência de visualização e análise de dados espaço-temporais armazenados em bancos de dados heterogêneos disponíveis em qualquer lugar da *web* sem exigir, para isso, nenhum tipo de esforço do usuário final. Além da visualização, o GeoSTAT permite a mineração de dados com até dois bancos de dados espaço-temporais heterogêneos distintos, independente de sua localização e, mais uma vez, sem exigir nenhum tipo de esforço do usuário final.

4.3.4 Componentes

A interface interativa do usuário do GeoSTAT é composta por dez componentes interoperáveis e responsáveis pelas funcionalidades oferecidas pelo ambiente. Estes componentes implementam a abordagem de visualização e análise de dados espaço-temporais proposta neste trabalho.

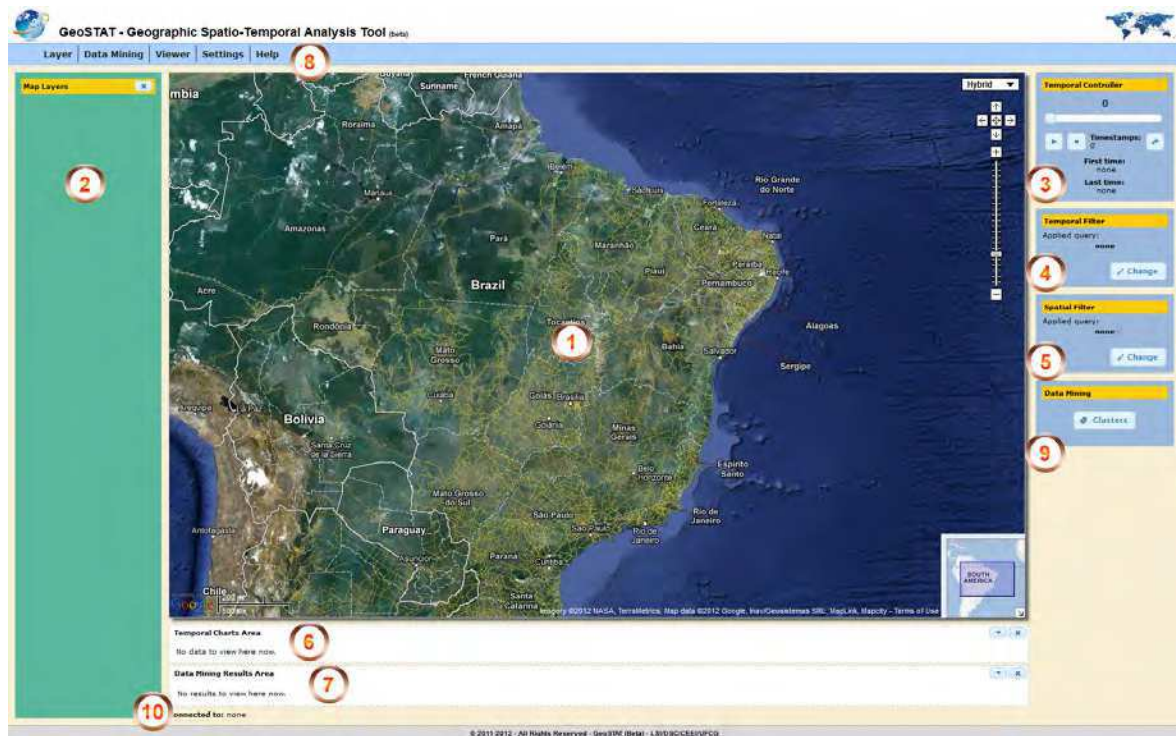


Figura 4.8: Componentes da camada de visualização do GeoSTAT.

A Figura 4.8 apresenta e enumera os componentes: 1) Mapa; 2) Camadas espaço-temporais (sobreposições); 3) Controlador temporal animado; 4) Filtro temporal; 5) Filtro espacial; 6) Gráfico de distribuição temporal; 7) Resultados da mineração de dados; 8) Menu de ações; 9) Mineração de dados; e 10) Informações dos servidores de dados conectados. Cada um deles é detalhado na sequência.

Mapa

O componente de mapa é o componente de maior destaque na interface do GeoSTAT. Trata-se de um visualizador de mapas georeferenciados, implementado com a API do *Google Maps* em linguagem *JavaScript*. Por meio desta API, é possível oferecer um visualizador de mapas com quatro mapas-base padrões oferecidos pelo Google: Cartográfico, apresentando a divisão política, sedes municipais e principais rodovias; Imagem de Satélite, com resoluções em tal nível que é possível visualizar logradouros; Híbrido, a junção dos dois primeiros; e Terreno, enfatizando variações de relevo e vegetação.

Além dos mapas-base, a API do *Google Maps* provê recursos de interatividade de mapa, com funções de *zoom* (aproximação), *pan* (navegação panorâmica), *information* (janela de

informações) e *scale* (escala).

Camadas Espaço-Temporais (Sobreposições)

O GeoSTAT possibilita ao usuário a experiência de adicionar, de forma simplificada e objetiva, camadas sobrepostas com dados espaço-temporais (ou apenas espaciais) publicados em servidores que implementem os serviços OGC WMS e WFS. O primeiro passo para utilizar um conjunto de dados no ambiente é criar uma conexão de dados. A criação de uma conexão de dados é muito simples, bastando apenas que o usuário defina um apelido para a conexão e informe o endereço completo de acesso ao servidor desejado. Essas informações são armazenadas no banco de dados do ambiente, evitando a necessidade de criar uma nova conexão para um servidor de acesso frequente, toda vez que o usuário necessitar utilizá-lo.

Uma vez criada a conexão de dados no GeoSTAT, o usuário poderá utilizá-la para adicionar camadas de dados espaciais ou espaço-temporais no visualizador e, a partir daí, poder realizar análises visuais sobre estes dados, utilizando-se dos demais componentes disponíveis no ambiente.

Na Figura 4.9, é possível visualizar os campos necessários à adição de uma nova camada sobreposta ao mapa. Além de selecionar a conexão de dados dentre aquelas criadas previamente, o usuário deve selecionar uma camada dentre àquelas disponibilizadas pela conexão selecionada, e definir um apelido (*alias*) que melhor defina, de forma sucinta, a camada adicionada no ambiente. Em camadas espaço-temporais, faz-se necessário, ainda, selecionar o atributo responsável por armazenar a informação temporal. Diante dessa informação, o ambiente se encarregará de preparar os componentes temporais para serem devidamente disponibilizados.

O componente de camadas espaço-temporais (ver item 2 da Figura 4.8) exibe as camadas sobrepostas adicionadas pelo usuário e oferece botões de controle de exibição para cada camada adicionada. O usuário tem a liberdade de exibir ou ocultar uma camada adicionada conforme sua necessidade, controlando sua opacidade em uma escala que varia entre 0% (invisível) e 100% (totalmente visível). Este componente implementa, ainda, o conceito de camada ativa, ou seja, a camada que será requisitada quando o usuário desejar exibir uma janela de informações contendo detalhes de uma determinada localização clicada no mapa, ou exibir o gráfico de distribuição temporal da camada, ou ainda, visualizar as informa-

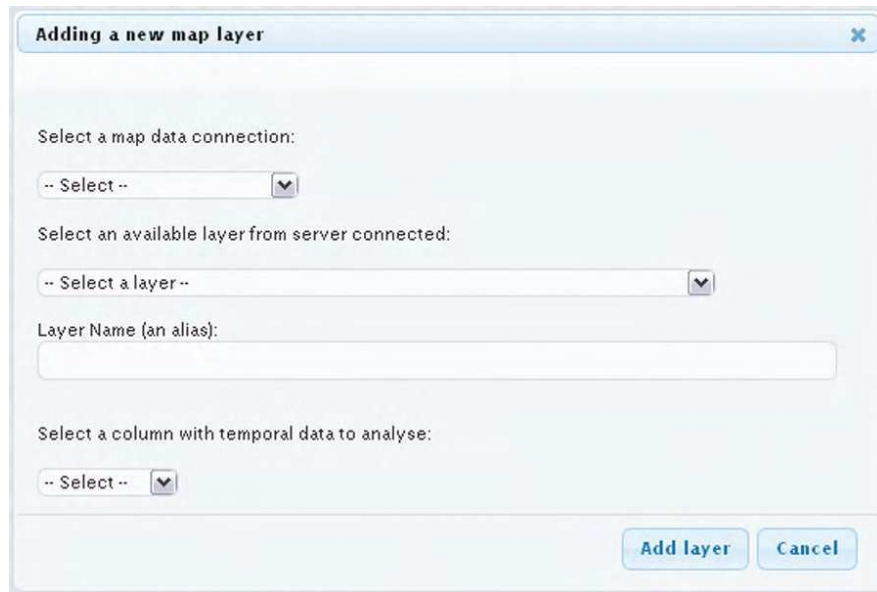


Figura 4.9: Caixa de diálogo para adicionar nova camada de mapa no GeoSTAT.

ções de resultado de um processo de mineração de dados. O usuário tem ainda a opção de ocultar este componente quando julgar necessário, aumentando a área de visualização do componente mapa e melhorando a experiência de visualização espacial na tela do dispositivo utilizado para acesso ao ambiente.

Controlador Temporal

O componente de controlador temporal (ver item 3 da Figura 4.8) é o principal responsável pelo tratamento temporal dos dados espaço-temporais visualizados no GeoSTAT. Por meio do controlador temporal, é possível alterar a visualização do mapa tomando por base um critério temporal. Esse critério temporal pode ser definido como um determinado instante de tempo (*timestamp*), por exemplo, ou em níveis mais abstratos de granularidade temporal, como em meses, por exemplo.

O GeoSTAT oferece vários tipos de granularidade temporal, ou seja, varias formas de fatiar e visualizar os dados espaciais em função do tempo. São 13 tipos no total:

Instante (ou *timestamp*): representa o tempo no formato contendo data completa, hora, minutos e segundos. Ex: "15/12/2009 15:23";

Hora do dia: permite agrupar os dados em função de uma hora do dia. Ex: "15:00" para todas as ocorrências com valores de hora entre 15:00 e 15:59, independente da data;

3-3 Horas: permite agrupar os dados em oito conjuntos diários, cada um representando um período de três horas do dia. Ex: "00:00 às 03:00" para todas ocorrências com valores de hora entre 00:00 e 02:59;

Hora de um dia: permite agrupar os dados em função de uma hora do dia, levando em consideração a data. Ex: "26/12/2010 15:00" para todas as ocorrências do dia 26/12/2010 com valores de hora entre 15:00 e 15:59;

Data: permite agrupar os dados em função da data, independente da hora. Ex: "26/12/2010" para todas as ocorrências registradas nesta data;

Dia da semana: permite agrupar os dados em função de um dia da semana, independente do mês ou do ano, gerando, portanto, sete conjuntos de dados. Ex: "Quarta-feira" para todas as ocorrências com valores de data equivalentes ao quarto dia da semana;

Dia do mês: permite agrupar os dados por dia do mês, independente do mês ou do ano. Ex: "10" para todas as ocorrências em dia 10, como 10/01/2009 ou 10/04/2010;

Semana: permite agrupar os dados por semana em relação à contagem de semanas por ano. Ex: "23" para todas as ocorrências com valores de data referentes à 23ª semana do ano, independente do ano;

Mês do ano: permite agrupar os dados em função do mês, levando em consideração o ano. Ex: "04/2008" para todas as ocorrências com valores de data referentes ao mês de abril de 2008;

Mês: permite agrupar os dados em função do mês, independente do ano. Ex: "04" ou "Abril" para todas as ocorrências com valores de data referentes ao mês de abril;

Quadrimestre do ano: permite agrupar os dados em função do quadrimestre, levando em consideração o ano. Ex: "02/2011" para todas as ocorrências com valores de data referentes aos meses de maio, junho, julho e agosto de 2011;

Semestre: permite agrupar os dados em função do semestre, levando em consideração o ano. Ex: "01/2011" para todas as ocorrências com valores de data referentes ao 1º semestre do ano 2011, ou seja, entre janeiro e junho de 2011;

Ano: permite agrupar os dados em função do ano. Ex: "2011" para todas as ocorrências com valores de data referentes ao ano 2011.

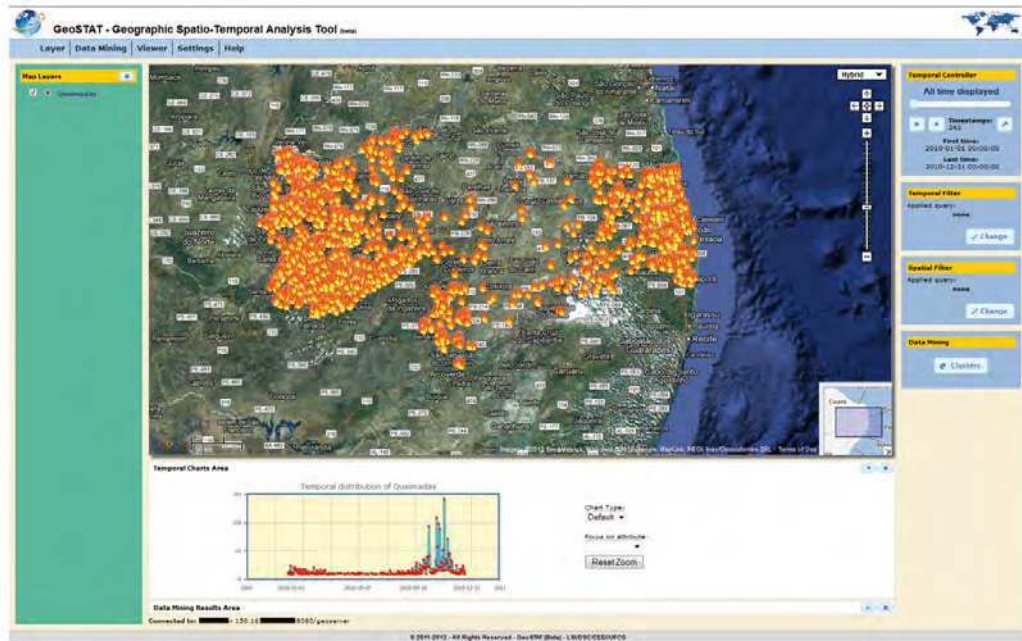
A granularidade temporal deve ser definida pelo usuário do GeoSTAT no momento de visualizar seu conjunto de dados espaço-temporais. Essa definição vai depender dos objetivos do usuário, do domínio e natureza dos dados analisados e, principalmente, do nível menos abstrato de granularidade temporal do conjunto de dados. Por padrão, o GeoSTAT utiliza a granularidade menos abstrata possível, oferecida pelos conjuntos de dados.

Ao visualizar uma camada de dados espaço-temporais, além da distribuição espacial exibida no componente mapa, o controlador temporal habilitará a animação temporal, que permitirá ao usuário visualizar no mapa as eventuais mudanças de distribuição espacial dos dados em função da variação temporal. A velocidade de animação pode ser definida pelo usuário em uma faixa que varia entre 0,5 e 10,0 segundos. O intervalo de animação temporal é determinado através de um prévio processamento do conjunto de dados visualizado, onde o GeoSTAT obtém o registro mais antigo (*First Time*), o registro mais recente (*Last Time*), e a quantidade de marcações de tempo distintas no conjunto de dados, satisfazendo a granularidade temporal escolhida.

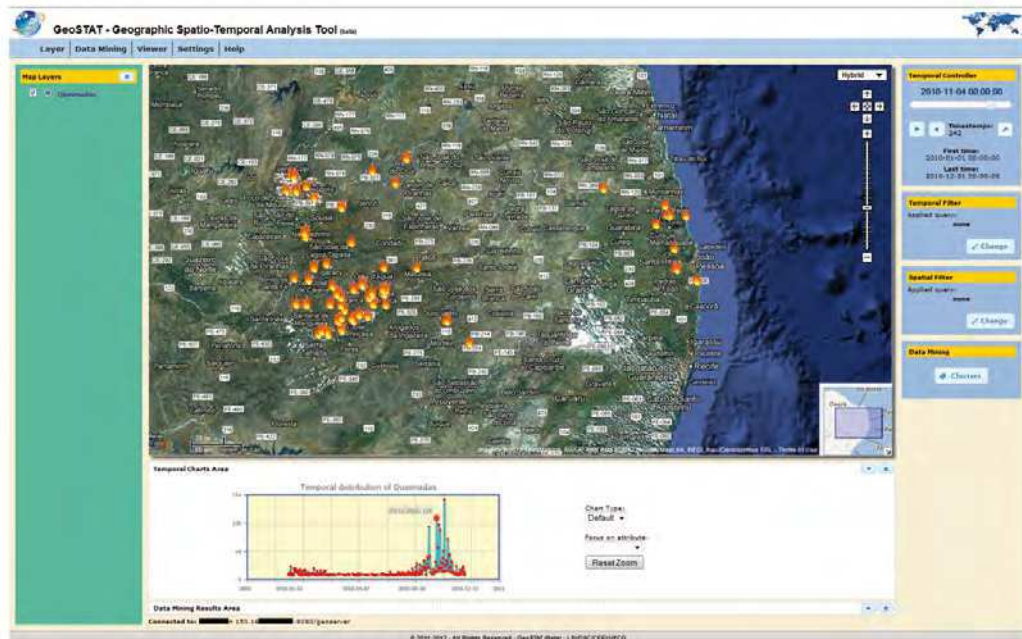
A Figura 4.10 ilustra a utilização do controlador temporal por meio de duas situações. Na situação (a), é possível visualizar no mapa todos os registros do conjunto de dados espaço-temporal, independente da variável temporal e do valor observado para cada registro. Já na situação (b), é possível visualizar apenas os registros correspondentes ao instante "04/11/2010". O valor "*Timestamps*" (242) significa dizer que, no conjunto de dados observado, existem 242 instantes de tempo distintos, onde o primeiro é em "01/01/2010" e o último em "31/12/2010". Se a granularidade fosse alterada para "Ano" o valor de "*Timestamps*" seria modificado para 1, uma vez que todos os registros são do ano de 2010. Consequentemente, o controlador temporal não teria funcionalidade para esse caso específico.

Gráfico de Distribuição Temporal

Além do controlador temporal, outro recurso de visualização temporal oferecido pelo GeoSTAT é o gráfico de distribuição temporal (ver item 6 da Figura 4.8 e exemplo na Figura 4.10). Ele é o responsável por auxiliar o usuário na visualização da variação dos dados



a)



b)

Figura 4.10: Utilização do controlador temporal do GeoSTAT. Em a) todos os dados espaço-temporais distribuídos no mapa, e em b) apenas os dados referentes a "04/11/2010" (116 pontos).

espaço-temporais em função do tempo, somando-se ao recurso de mapa, que possibilita visualizar a distribuição dos dados em função do espaço.

No componente de gráfico de distribuição temporal, o usuário pode visualizar a distribuição temporal de todo o conjunto de dados (*default*), como também pode visualizar distribuições temporais com base em um atributo convencional do conjunto.

Se o atributo for do tipo numérico, o gráfico exibirá a distribuição consolidada do somatório dos valores observados para cada marcação de tempo. Um exemplo pode ser observado em um conjunto de dados onde um atributo convencional é o quantitativo de assassinatos. Neste exemplo o gráfico mostrará a distribuição temporal desse quantitativo consolidado para cada marcação temporal, permitindo, dessa forma, identificar, através do gráfico, a marcação temporal que registrou os maiores valores.

Se o atributo não for numérico, será tratado como categórico, onde o gráfico exibirá a distribuição consolidada das ocorrências de cada categoria para cada marcação de tempo. Um exemplo deste caso pode ser observado em um conjunto de dados onde um atributo convencional é responsável por descrever a causa de uma ocorrência. O gráfico mostrará os quantitativos consolidados de cada ocorrência, para cada marcação de tempo, permitindo identificar qual "categoria" obteve maior índice e em qual marcação temporal este índice foi atingido.

Filtro Temporal

O componente de Filtro Temporal (ver item 4 da Figura 4.8 e Figura 4.11) é o responsável pela consulta e seleção temporal dos dados visualizados a partir das camadas espaço-temporais. O usuário pode estar interessado em visualizar apenas uma porção dos dados espaciais disponibilizados a partir de um servidor de dados OGC, de forma a simplificar o processo de análise.

Através do componente de filtro temporal, o usuário pode, por meio de quatro opções de filtro e observando a granularidade temporal utilizada, reduzir o conjunto de dados espaço-temporais para a visualização e análise, com base na dimensão temporal. As quatro opções disponíveis para o filtro temporal são:

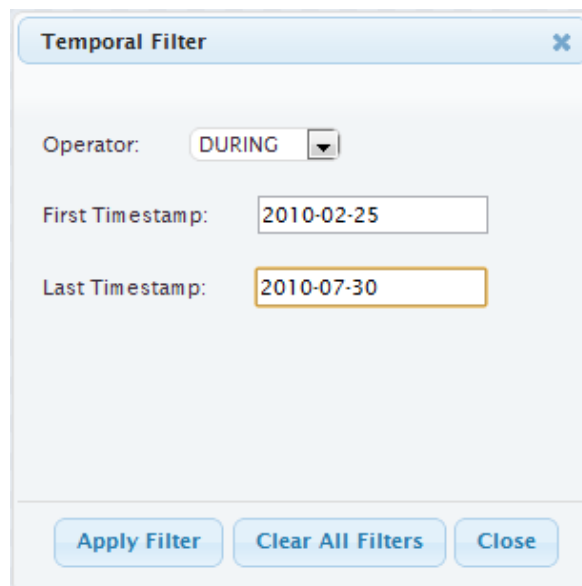
A partir de: permite que o usuário defina uma data e hora inicial e que, portanto, só sejam

apresentados dados de ocorrências com marcação temporal a partir da informada na definição do filtro;

Até: permite que o usuário informe uma data e hora final e que, portanto, só sejam apresentados dados de ocorrências com marcação temporal até a informada na definição do filtro;

Em: permite que o usuário informe uma data e hora específica para visualização dos dados;

Durante: permite que o usuário informe um período específico, com data e hora tanto inicial como também final (inclusive). Dessa forma, só serão apresentados os dados de ocorrências com data pertencente ao período fechado definido.



A caixa de diálogo intitulada "Temporal Filter" possui um botão de fechar no canto superior direito. O campo "Operator" contém o texto "DURING" com uma seta para baixo. Abaixo dele, há dois campos de texto: "First Timestamp" com o valor "2010-02-25" e "Last Timestamp" com o valor "2010-07-30". Na base da caixa, há três botões: "Apply Filter", "Clear All Filters" e "Close".

Figura 4.11: Caixa de diálogo para definição de filtro temporal no GeoSTAT.

Ao aplicar um filtro temporal, todas as camadas espaço-temporais já adicionadas e também aquelas que vierem a ser adicionadas posteriormente ao ambiente, exibirão apenas os dados que satisfaçam os critérios estabelecidos pelo filtro aplicado. É importante ressaltar que o resultado do filtro pode ser um conjunto de dados vazio, apontando, neste caso, que não há informação para o filtro selecionado, ou seja, o intervalo temporal dos dados da camada em questão está fora do intervalo especificado no filtro.

Filtro Espacial

O componente de Filtro Espacial (ver item 5 da Figura 4.8) é o responsável pela consulta e seleção espacial dos dados visualizados em uma camada espacial ou espaço-temporal adicionada. Por meio do filtro espacial, o usuário do GeoSTAT pode estar interessado em:

- visualizar apenas uma região do mapa, reduzindo dessa forma a base de dados de visualização em função da localização geográfica;
- visualizar o resultado de uma relação espacial entre duas camadas espaciais ou espaço-temporais.

Para visualizar apenas uma região do mapa, o usuário pode criar uma consulta baseando-se nos atributos convencionais da camada, de modo a estabelecer critérios de filtro que reduzirão o conjunto de dados visualizado. A Figura 4.12 exhibe um exemplo da caixa de diálogo que possibilita a criação do filtro espacial baseado em consultas (*queries*).

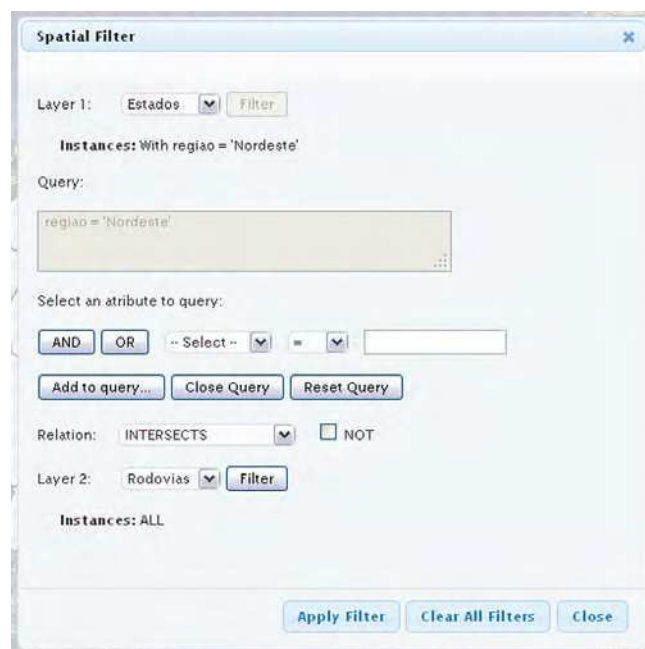


Figura 4.12: Caixa de diálogo para definição de filtro espacial no GeoSTAT.

Para visualizar uma relação espacial entre duas camadas espaciais ou espaço-temporais previamente adicionadas ao GeoSTAT, o usuário faz uso da mesma caixa de diálogo. É possível realizar as seguintes relações espaciais topológicas entre duas camadas: *INTERSECTS*

(Intersecta), *CONTAINS* (Contém), *CROSSES* (Cruza), *TOUCHES* (Toca), *COVERS* (Cobre - Sobreposição total) e *OVERLAPS* (Sobrepõe - Sobreposição parcial). É possível ainda, para cada uma dessas relações espaciais topológicas, aplicar a negação (*NOT*), caso esta seja relevante para a análise realizada pelo usuário.

Mineração de Dados

No componente de atalhos para mineração de dados (ver item 9 da Figura 4.8) é possível realizar a mineração de dados baseada em *clustering* nas camadas previamente adicionadas ao ambiente, visualizar o resultado de um processo de mineração de dados realizado anteriormente, e ainda, visualizar o *status* detalhado de processos de mineração de dados em andamento. Os processos de mineração de dados são executados em segundo plano, dispensando a necessidade de prender o usuário ao aguardar o término para poder fechar a aplicação ou realizar qualquer outra tarefa.

A mineração de dados baseada em *clustering* pode ser realizada somente com camadas espaço-temporais que apresentem a geometria de pontos bidimensionais. A Figura 4.13 apresenta a caixa de diálogo exibida ao usuário interessado em realizar este procedimento com apenas uma camada de dados espaço-temporais.

O GeoSTAT oferece a opção de mineração de dados espaço-temporais com uma ou duas camadas visíveis. A utilização deste recurso com duas camadas é interessante quando o usuário está interessado em realizar agrupamentos entre dois eventos descritos por duas camadas distintas, e verificar a ocorrência de correlação espaço-temporal entre esses eventos.

Além de informar as camadas utilizadas no processo de mineração, o usuário deve informar a granularidade temporal, ou seja, a forma como os dados devem ser agrupados em função do tempo, no processo de mineração, e a aplicação de mineração de dados que deseja utilizar. No momento o GeoSTAT oferece apenas o Weka como opção de aplicação de mineração de dados.

Como já apresentado na subseção anterior, o GeoSTAT integra sete algoritmos de *clustering* disponibilizados pelo software Weka (Hall et al., 2009). Para selecionar o algoritmo de *clustering* desejado, o usuário deve primeiramente selecionar o tipo do algoritmo dentre os disponíveis (Particionamento, Hierárquico, Densidade e Grade) e então, a partir do tipo selecionado, escolher o algoritmo e definir os valores para os parâmetros de entrada desse al-

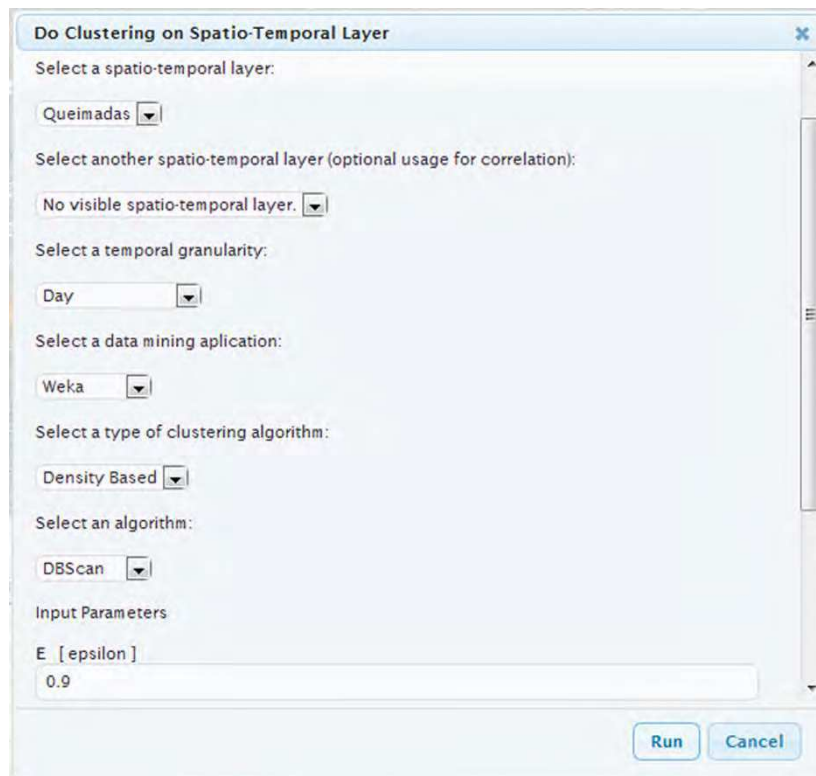


Figura 4.13: Caixa de diálogo para execução de algoritmo de mineração de dados baseada em *clustering* no GeoSTAT.

goritmo. Os resultados (*clusters*) são exibidos em uma nova camada espaço-temporal criada pelo GeoSTAT após finalizado o processo de mineração.

Dependendo do volume de dados utilizado, o processo de mineração de dados pode ser lento. O usuário não precisa manter a aplicação aberta ou deixar de operá-la enquanto um processo de mineração de dados está em execução, uma vez que todo o processamento é realizado em segundo plano no servidor de aplicação do GeoSTAT. O ambiente oferece a opção do usuário acompanhar os processos de mineração de dados ainda em execução.

A Figura 4.14 apresenta a caixa de diálogo responsável por exibir o status dos processos de mineração de dados em execução no servidor. Por meio dessa tela o usuário pode, além de acompanhar o processo, abortar sua execução, se assim desejar. Ao abortar uma execução, o GeoSTAT remove automaticamente todas as informações já armazenadas para o processo abortado.

A Figura 4.15, por sua vez, apresenta a caixa de diálogo exibindo os resultados dos processos de mineração de dados que foram executados a partir do GeoSTAT e já estão con-

cluídos. Além de visualizar os detalhes do resultado, como o tempo total de processamento, o usuário tem a opção de abrir o resultado a partir dessa tela. Ao abrir um resultado de mineração de dados, o GeoSTAT cria uma nova sobreposição de camada espaço-temporal, contendo o resultado do processo de mineração aberto, e habilita o componente de resultados da mineração de dados para que o usuário possa navegar no resultado tanto na dimensão temporal como também na dimensão espacial.

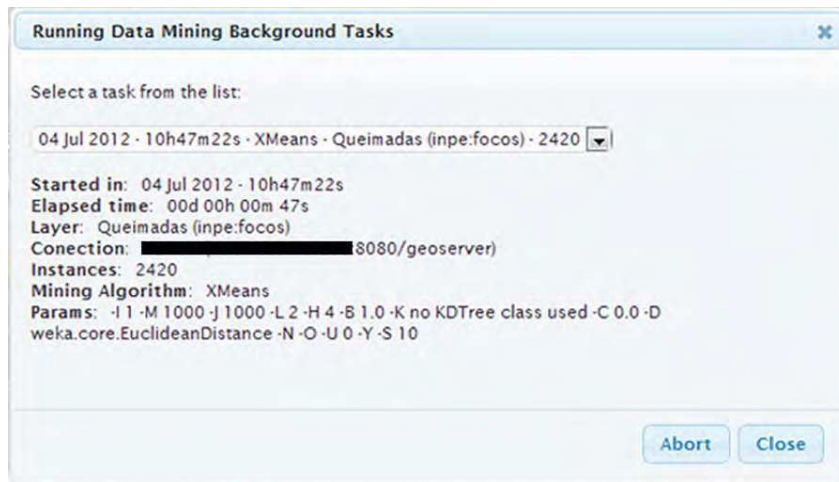


Figura 4.14: Caixa de diálogo de acompanhamento da execução de mineração de dados no GeoSTAT.

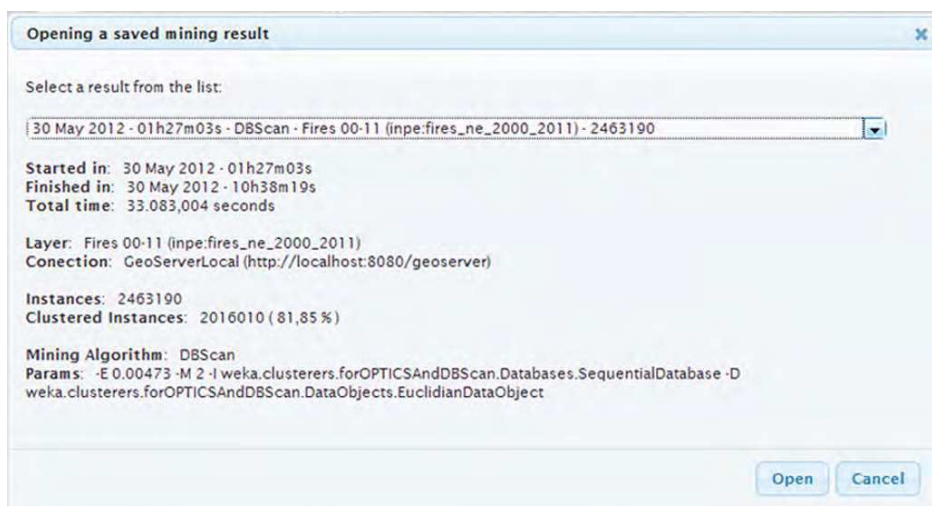


Figura 4.15: Caixa de diálogo de visualização de resultados de mineração de dados no GeoSTAT.

Resultados da Mineração de Dados

O componente de resultados da mineração de dados (ver item 7 da Figura 4.8) é o responsável por oferecer os comandos necessários à visualização e navegação espaço-temporal de uma camada contendo os resultados de um processo de mineração de dados realizada pelo GeoSTAT.

A Figura 4.16 exemplifica a exibição de um resultado de mineração de dados executada a partir do GeoSTAT, e ilustra os comandos de navegação oferecidos pelo componente. Neste exemplo, o mapa exibe o *cluster* de número 10 na marcação temporal "2010-01" (Janeiro de 2010), com ponto central nas coordenadas geográficas (-35.100, -7.607), e agrupando, no total, 12 pontos. É um *cluster* considerado "relevante" pelo ambiente em razão deste agrupar pontos das duas camadas utilizadas no processo de mineração.

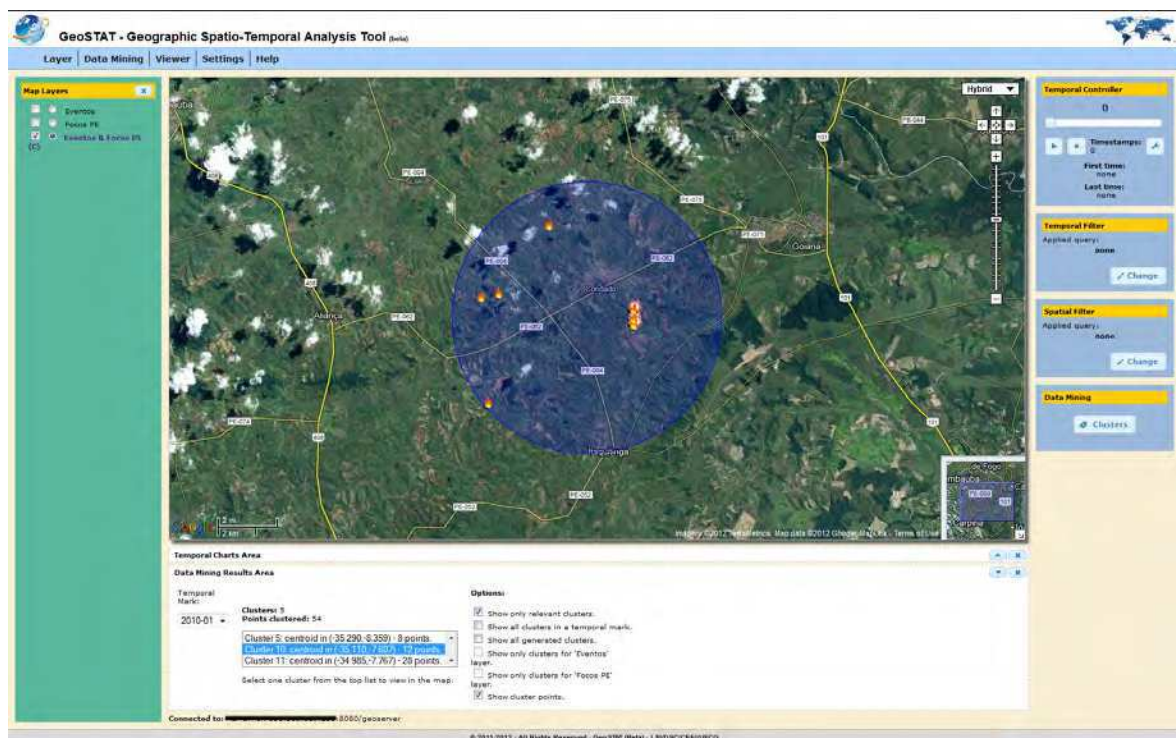


Figura 4.16: GeoSTAT exibindo os resultados relevantes de um processo de mineração de dados executado com duas camadas espaço-temporais.

O usuário do GeoSTAT pode navegar entre as marcações de tempo que tiveram ocorrência de *clusters* e visualizar no mapa cada *cluster* por vez. Se a mineração de dados for realizada com duas camadas, o usuário terá a opção de visualizar somente os *clusters* "re-

levantes", ou seja, aqueles que contêm, pelo menos, um ponto de cada camada, e também, as opções para visualizar apenas *clusters* que agrupam somente pontos da primeira ou da segunda camada.

É possível, também, visualizar todos os *clusters* de uma marcação temporal ou todos os *clusters* gerados independentemente da marcação temporal. A Figura 4.17 exibe um exemplo com todos os *clusters* (relevantes e não relevantes) da marcação temporal "2010-01" (Janeiro de 2010).

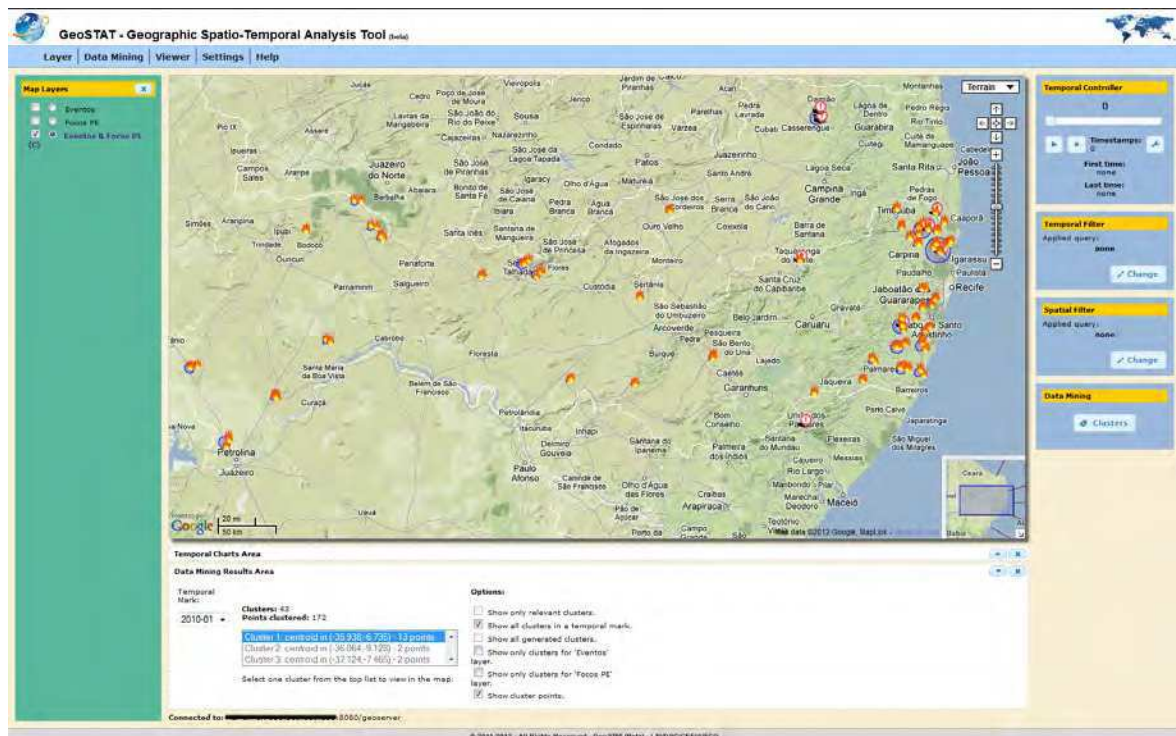


Figura 4.17: GeoSTAT exibindo todos os *clusters* da marcação temporal "2010-01".

Menu de Ações

O componente menu de ações (ver item 8 da Figura 4.8) oferece atalhos para acesso aos demais componentes da interface gráfica interativa do GeoSTAT. Por meio do menu de ações, o usuário pode, ainda, esconder (ou reexibir) os componentes que não são fixos na interface, de modo a ampliar a área de visualização do mapa na tela. Apenas três componentes não são fixos na interface, como a barra com as camadas sobrepostas, a área do gráfico de distribuição temporal e a área dos resultados de mineração de dados. Todos eles são dotados de um botão "fechar", que permite sua ocultação imediata.

Informações dos servidores de dados conectados

O componente de informações dos servidores de dados conectados (ver item 10 da Figura 4.8) é o responsável por exibir as informações sobre todos os servidores de dados que estão conectados em uma sessão de usuário do GeoSTAT. Como já mencionado anteriormente, o GeoSTAT permite que o usuário adicione camadas sobrepostas de diversos servidores de dados em uma mesma sessão, permitindo uma análise espaço-temporal simultânea entre todas essas camadas, independente de sua localização na *web*.

A medida que o usuário adiciona novas camadas ao ambiente, o componente é atualizado, exibindo, portanto, as URLs de todos os servidores conectados.

4.4 Comparativo com as principais abordagens existentes

A seção 3.3 discutiu as principais abordagens existentes para visualização e análise de dados espaço-temporais. Essa discussão foi fundamental para se chegar à proposta apresentada neste capítulo, que tem como objetivo minimizar as lacunas identificadas nas abordagens relacionadas.

Diante da concepção dessa nova abordagem e de sua implementação, o GeoSTAT, realizou-se uma nova avaliação comparativa a partir da avaliação realizada na seção 3.3. A Tabela 4.1, derivada da Tabela 3.1, apresenta o resultado dessa nova avaliação.

Conforme recomendado em "*Space, time and visual analytics*", de Andrienko et al. (2010b), a nova abordagem proposta conseguiu alcançar as seis características necessárias a uma abordagem de visualização e análise espaço-temporal, embora com algumas limitações em alguns pontos:

Espacialidade: através do recurso de mapa dinâmico, com controles para navegação espacial, e do recurso de filtro espacial, que possibilita a execução de consultas espaciais e a visualização dos resultados como um filtro espacial. Entretanto, a solução proposta não aborda a terceira dimensão espacial (3D) e as consultas espaciais podem ter custo muito elevado a depender do tamanho dos conjuntos de dados;

Temporalidade: através do recurso de controlador temporal com animação espaço-temporal, variando as imagens do mapa à medida que a linha do tempo também varia;

Tabela 4.1: Nova tabela comparativa entre os trabalhos acadêmicos de visualização espaço-temporal relacionados no Capítulo 3 e a nova abordagem proposta neste trabalho.

Trabalho / Critério	1	2	3	4	5	6
<i>SocioScope</i>	X	X				
<i>Exploratory spatio-temporal data mining and visualization</i>	X	X	X	X		X
<i>Spatio-Temporal Visual Query Environment</i>	X	X	X	X	X	
<i>Visual Analytics Tools for Analysis of Movement Data</i>	X	X	X	X		X
<i>A Framework for Exploring Spatiotemporal Patterns</i>	X	X	X			X
<i>COPLINK: Visualization for Crime Analysis</i>	X	X				X
<i>BirdVis: Visualizing and Understanding Bird Populations</i>	X	X		X		
<i>GeoVISTA CrimeViz</i>	X	X		X		
<i>GeoSTAT: Geographic SpatioTemporal Analysis Tool</i>	X	X	X	X	X	X

do recurso de gráfico dinâmico de distribuição temporal; e do recurso de filtro temporal, que possibilita a execução de consultas temporais. Entretanto, a solução proposta é limitada no sentido de permitir a visualização de um único intervalo de tempo por vez, e deixa de oferecer outros tipos de gráficos temporais que podem ser interessantes ao usuário;

Genericidade: por oferecer recursos e técnicas de visualização e navegação espaço-temporal funcionais com qualquer domínio de aplicação;

Flexibilidade: através do recurso de camadas sobrepostas e dos controles de exibição e opacidade, que deixam o usuário livre para visualizar as camadas na ordem de sobreposição que melhor lhe convier, e dos recursos de filtro espacial e temporal. Entretanto, ainda existem limitações no sentido de permitir alteração nos estilos das camadas, ficando essa tarefa para o serviço que está disponibilizando os dados na *web*;

Interoperabilidade: através da adoção dos padrões de compartilhamento de dados da OGC. Entretanto, a solução proposta limita-se à utilização dos serviços WMS e WFS, não incorporando ainda outros padrões que tem surgido, como o WCS e o WFS-T, por exemplo. Além disso, é importante observar a necessidade da disponibilização dos

dados por meio de um serviço que implemente tais padrões; e

Mineração de Dados: através da integração de algoritmos de mineração de dados baseados em *clustering* espaço-temporal, e do suporte para execução, visualização e navegação espaço-temporal dos resultados. Entretanto, vale ressaltar que apenas os algoritmos de *clustering* do Weka foram integrados e disponibilizados, e que a solução não oferece outro tipo de técnica para tratamento dos dados, como classificação, por exemplo. Além disso, a mineração de dados funciona apenas com geometrias do tipo ponto bidimensional, não oferecendo, portanto, tratamento a outros tipos conhecidos, como linhas e polígonos.

Os pontos relevantes da abordagem proposta são: utilização de mineração de dados, oferecendo diversos algoritmos para opção do usuário final; um ambiente de visualização independente de domínio, flexível e interoperável, permitindo a visualização de camadas de dados espaço-temporais disponibilizadas a partir de diferentes servidores espalhados na Internet; e inclusão de recursos para tratamento da dimensão temporal.

Os pontos críticos da implementação dessa nova abordagem são: custo considerável na execução de consultas espaciais e custo elevado na visualização simultânea de muitos *clusters*.

4.5 Resumo do Capítulo

Este capítulo apresentou uma nova abordagem para visualização e análise visual em dados espaço-temporais, a principal contribuição desta Dissertação de Mestrado.

Reunindo um conjunto de técnicas, recursos de visualização e algoritmos de mineração de dados espaço-temporais baseados em *clustering*, foi possível desenvolver o escopo de uma nova solução em visualização e análise de dados espaço-temporais, com o objetivo de minimizar as lacunas encontradas na comunidade. A implementação desse escopo, denominada GeoSTAT (*Geographic SpatioTemporal Analysis Tool*), um ambiente *web* para análise e visualização de dados espaço-temporais, foi apresentada e detalhada neste capítulo. Trata-se de um ambiente espaço-temporal independente de domínio e conectável a qualquer fonte de dados disponível na *web* por meio da implementação de serviços de compartilhamento de

dados espaciais especificados pela OGC, com diversos recursos visuais para tratamento das dimensões espacial e temporal.

Além de recursos para tratamento das particularidades inerentes às dimensões espacial e temporal, a implementação apresentada destacou, como principal diferencial, a inclusão de algoritmos de mineração de dados baseados em *clustering* espaço-temporal, no sentido de possibilitar a execução de mineração de dados semiautomatizada e a manipulação visual dos resultados.

Foi realizada uma avaliação comparativa entre a nova abordagem proposta e as abordagens relacionadas e discutidas no capítulo anterior, onde foi possível mostrar que o GeoSTAT cumpriu o objetivo de alcançar as seis características consideradas essenciais em uma boa solução de visualização e análise espaço-temporal, se sobressaindo entre as demais.

No próximo capítulo, serão apresentados dois estudos de caso realizados com a implementação da nova abordagem proposta. A partir de duas situações de análise espaço-temporal envolvendo domínios e fontes de dados distintas, será possível demonstrar a utilidade do GeoSTAT e suas vantagens em relação às demais soluções.

Capítulo 5

Validação da abordagem proposta com estudos de caso

Este capítulo trata da validação da abordagem proposta neste trabalho. São realizados dois estudos de caso, cada um abordando um domínio de aplicação específico, para demonstrar a experiência do usuário final diante das técnicas de visualização espaço-temporal adotadas.

5.1 Estudo 1: Análise de correlação espaço-temporal entre falhas em linhas de transmissão de energia elétrica e focos de incêndio

Este estudo consiste na análise de dois conjuntos de dados espaço-temporais. Cada conjunto é composto por registros de um evento no espaço-tempo. O primeiro evento é a ocorrência de focos de incêndios, enquanto que o segundo é a ocorrência de falhas em linhas de transmissão de energia elétrica.

5.1.1 Motivação

Empresas geradoras de energia elétrica sofrem constantemente com centenas de oscilações em diversos pontos de suas linhas de transmissão. Em meio às oscilações, ocorrem eventos de falta no sistema, ou seja, eventos em que as linhas apresentam falha e geralmente sofrem

algum dano, ficando desligadas por um período. Tais falhas acarretam prejuízos financeiros, perdas de receita, uma vez que existem custos de reparo quando o problema acontece devido a linhas danificadas, e custos indenizatórios aos seus clientes, as empresas fornecedoras de energia elétrica (Amorim et al., 2012).

Um dos caminhos para aumentar a confiabilidade da transmissão de energia elétrica, ou seja, diminuir as ocorrências das oscilações e, conseqüentemente, reduzir os prejuízos financeiros, é identificar os principais elementos causadores de faltas e traçar planos para combatê-los. A identificação desses elementos envolve a análise dos registros de ocorrências passadas, o que significa analisar grande volume de dados espaço-temporais.

A análise histórica das faltas no sistema elétrico, em conjunto com dados meteorológicos históricos, ou dados históricos sobre detecção de focos de incêndio, ou ainda, dados históricos sobre detecção de descargas elétricas, por exemplo, pode revelar informações importantes às empresas geradoras e transmissoras de energia elétrica, que poderão atuar, de forma preventiva, em locais críticos e em determinados momentos do ano, por exemplo, aumentando a qualidade do serviço e reduzindo as perdas de receita.

Com esse objetivo, este estudo de caso utilizará a implementação da abordagem proposta para realizar uma análise espaço-temporal em busca de indícios que reforcem a hipótese de que queimadas ocorridas em uma distância espaço-temporal suficientemente próxima de uma linha de transmissão, são causadoras de eventos de falha no sistema e, conseqüentemente, desenergização temporária em pontos da rede.

5.1.2 Dados

Para a realização deste estudo, utilizou-se dados espaço-temporais georeferenciados de focos de incêndio detectados na Região Nordeste do Brasil, fornecidos pelo Instituto Nacional de Pesquisas Espaciais¹ (INPE), através do Centro de Previsão de Tempo e Estudos Climáticos (CPTEC), que divulga esse tipo de informação diariamente, através do seu Portal de Monitoramento de Queimadas e Incêndios².

Foram obtidos, no total, 2.361.040 registros de focos de incêndio detectados na região, no

¹INPE - Instituto Nacional de Pesquisas Espaciais. Mais informações em: <http://www.inpe.br/>

²INPE/CPTEC - Portal de Monitoramento de Queimadas e Incêndios. Disponível em: <http://www.inpe.br/queimadas/>

período compreendido entre 01/01/2002 e 31/12/2011, ou seja, nos últimos 10 anos. Os dados espaço-temporais foram obtidos em formato *ESRI Shapefile*, com sistema de referência geográfica WGS84³, e informações temporais seguindo o horário mundial GMT⁴.

Algumas informações sobre os focos de incêndio detectados pelo INPE devem ser levadas em consideração. Segundo informações do próprio Instituto, seu sistema detecta a existência de fogo na vegetação sem ter condições de avaliar o tamanho da área que está queimando ou o tipo de vegetação afetada. Outro ponto importante é que trabalhos de validação realizados pelo Instituto indicam que o erro médio na localização espacial dos focos é de, aproximadamente, 400 metros, com desvio padrão de cerca de 3 quilômetros, e com cerca de 80% dos focos detectados ocorrendo em um raio de um quilômetro das coordenadas indicadas. Na validade temporal, os satélites oferecem uma resolução temporal média de 3 horas, ou seja, o tempo médio entre a passagem de dois satélites capturando informações de uma mesma região.

Uma outra base de dados espaço-temporal, também georeferenciada, foi utilizada neste estudo. Trata-se de eventos de falhas em linhas de transmissão, registrados pela Companhia HidroElétrica do São Francisco (Eletrobrás/Chesf)⁵, que atua em toda a região Nordeste do país. Como não foi possível obter informações oficiais junto à Eletrobrás/Chesf devido a questões de ordem técnica e confidencial, foi desenvolvido um algoritmo gerador de eventos de falhas aleatórias no espaço-tempo, obedecendo a restrição espacial imposta pela malha de transmissão da Eletrobrás/Chesf e a restrição temporal imposta pela outra base de dados utilizada neste estudo.

Foram gerados, no total, 131.834 registros de falhas em linhas de transmissão da Eletrobrás/Chesf, no período compreendido entre 01/01/2002 e 31/12/2011, ou seja, também nos últimos 10 anos. Estes registros foram armazenados em um banco de dados espacial *PostgreSQL+PostGIS*, também no sistema de referência geográfica WGS84, e com informações temporais seguindo o horário mundial GMT.

Os dois conjuntos de dados utilizados neste estudo compartilham da mesma geometria

³WGS84 - Sistema Geodésico Mundial. Mais informações em (inglês):

http://en.wikipedia.org/wiki/World_Geodetic_System

⁴Greenwich Mean Time - GMT. Mais informações em (inglês): <http://wwp.greenwichmeantime.com/>

⁵Eletrobrás/Chesf - Companhia HidroElétrica do São Francisco. Mais informações em:

<http://www.chesf.gov.br/>

espacial e também da mesma granularidade temporal. Os registros são formados por geometrias espaciais do tipo PONTO bidimensional, e por registros temporais com granularidade de instante de tempo (*timestamp*) com resolução em segundos.

Para utilização dos dados no GeoSTAT, foi necessária a instalação do *Geoserver*, de forma que ele pudesse ser o servidor dos dados conectado ao ambiente, fornecendo as informações a partir da implementação dos serviços WMS e WFS. Uma camada de dados do *Geoserver* foi criada para cada conjunto de dados.

Com o objetivo de auxiliar na análise visual dos eventos de falhas em linhas de transmissão, utilizou-se, ainda, um conjunto de dados espaciais contendo a malha de transmissão da Eletrobrás/Chesf. Uma terceira camada de dados do *Geoserver* foi criada para oferecer esses dados.

Para condução deste estudo, o usuário do GeoSTAT será denominado analista, um usuário especialista no domínio abordado, em busca de informações relevantes implícitas em uma grande massa de dados espaço-temporais.

5.1.3 Condução

Diante do servidor *Geoserver* em execução, oferecendo as duas camadas espaço-temporais e a camada espacial contendo as Linhas de Transmissão da Eletrobrás/Chesf, utilizou-se o GeoSTAT para, inicialmente, fazer a carga dessas três camadas.

A Figura 5.1 exibe a tela do GeoSTAT com as três camadas espaço-temporais carregadas no ambiente a partir da conexão de dados com o *GeoServer*. O que se observa é o resultado de cerca de dois milhões e meio de pontos plotados no mapa, suficiente para preencher toda a região Nordeste.

Diante dessa visualização, o analista poderia ter a falsa impressão de que toda a extensão de todas as linhas de transmissão tiveram eventos de falha, e todas as falhas em linhas de transmissão aconteceram em regiões onde também foram detectados focos de incêndio. Essa impressão poderia ser verdadeira, se não existisse a dimensão temporal, que é ignorada na visualização de mapa apresentada na Figura 5.1.

Utilizando o controlador temporal do ambiente, é possível visualizar os dados levando em consideração a dimensão temporal. A Figura 5.2 exibe duas imagens ilustrando a utilização do controlador temporal do GeoSTAT. Na primeira imagem (a) é possível observar os eventos

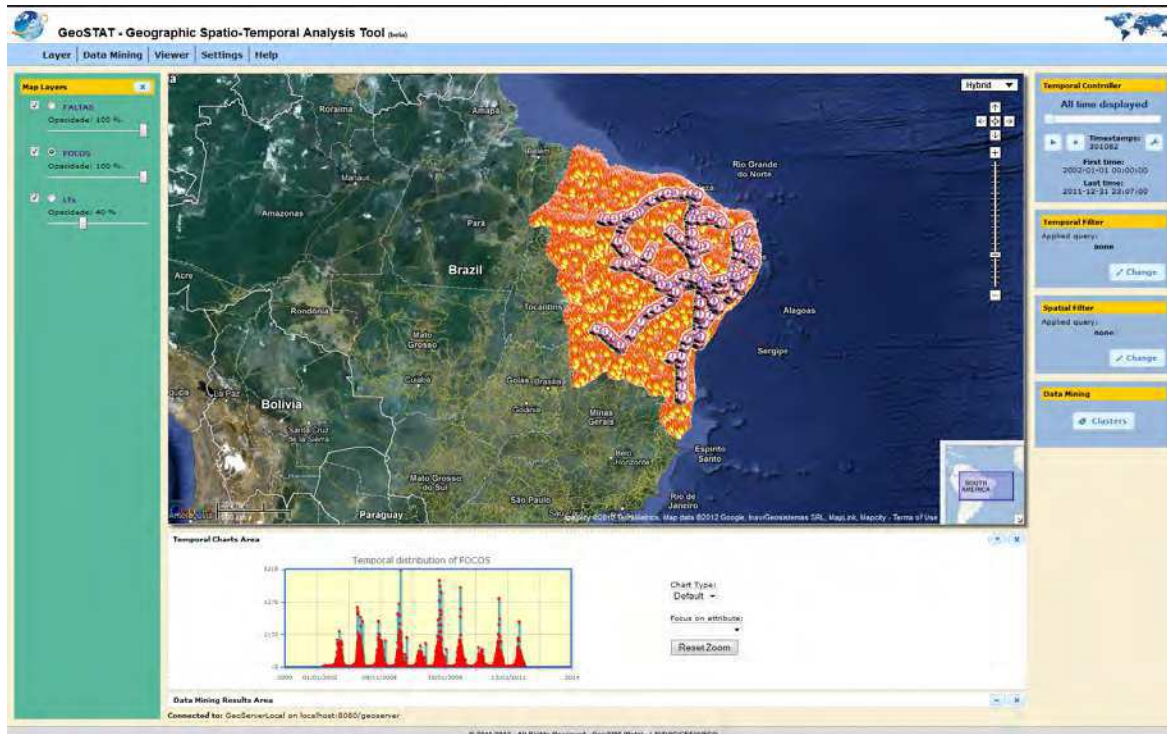


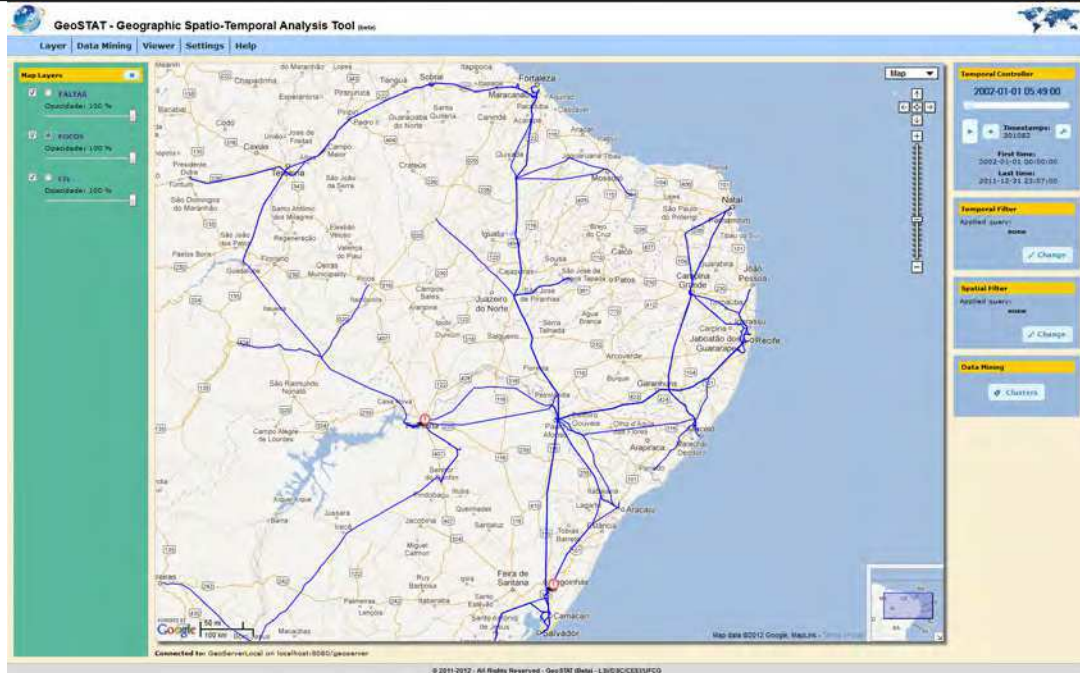
Figura 5.1: GeoSTAT exibindo as camadas de dados utilizadas neste estudo de caso.

que ocorreram às 05:49 hs do dia 01/01/2002: apenas dois eventos de falha, de acordo com o mapa. Na segunda imagem (b) é possível observar os eventos que ocorreram às 20:21 hs do dia 01/01/2002, apenas dois focos de incêndio isolados em dois pontos do estado do Maranhão.

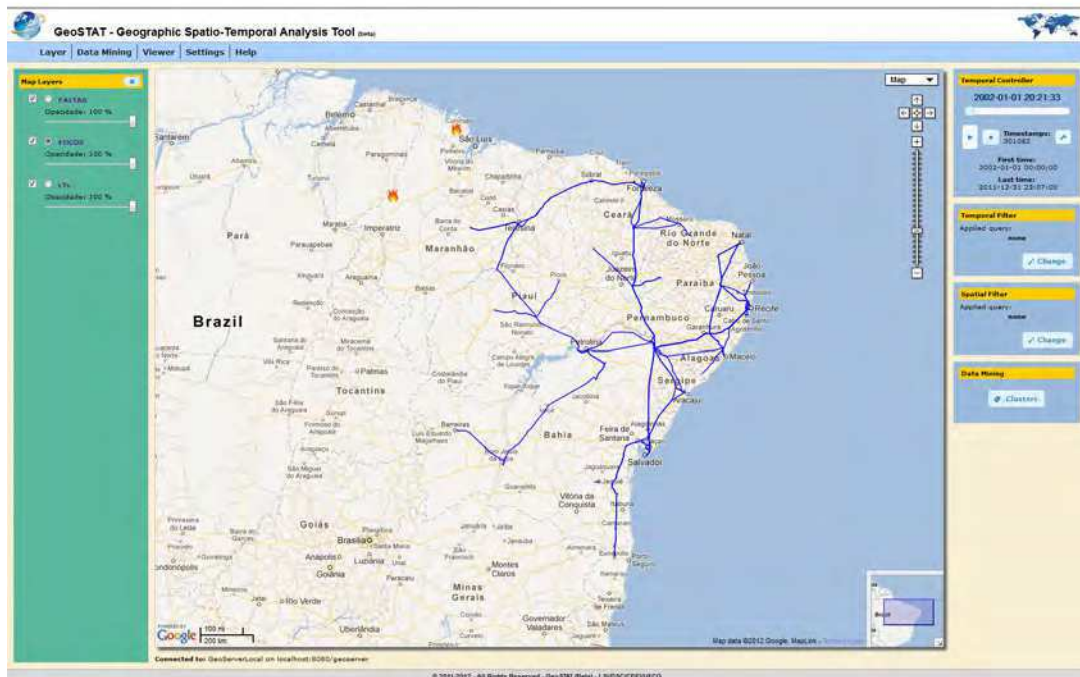
Os gráficos de distribuição temporal, gerados e exibidos automaticamente ao carregar e selecionar uma camada espaço-temporal no GeoSTAT, possibilitam ao analista verificar o comportamento temporal de toda a massa de dados. A Figura 5.3 exibe os gráficos gerados para as camadas de focos de incêndio (a) e de falhas em linhas de transmissão (b).

Observando o gráfico correspondente à camada de focos de incêndio, na Figura 5.3, nota-se que existe uma repetição anual na distribuição da quantidade de ocorrências de focos detectados, onde os máximos se concentram nos últimos e primeiros meses de cada ano, período do ano em que a região Nordeste registra as maiores temperaturas, o que contribui no surgimento de novos focos. Através do gráfico, é possível observar ainda que a quantidade máxima de focos registrada em um dia, durante o período de 10 anos analisado, foi de 6.418 focos. Essa marca foi atingida em 07/11/2005.

Ao observar o gráfico correspondente à camada de falhas em linhas de transmissão, nota-



a)



b)

Figura 5.2: Duas visualizações de mapa em instantes de tempo distintos: a) eventos registrados às 05h49 do dia 01/01/2002, e b) eventos registrados às 20h21 do mesmo dia.

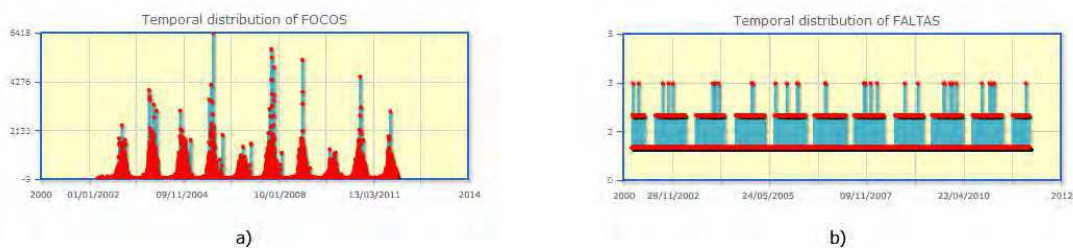


Figura 5.3: Gráficos de distribuição temporal gerados automaticamente pelo GeoSTAT: a) distribuição temporal dos focos de incêndio, e b) distribuição temporal das falhas em linhas de transmissão.

se um comportamento temporal praticamente contínuo. Uma vez que os dados foram gerados de forma aleatória, a partir de um algoritmo, a distribuição temporal das ocorrências foi uniforme, registrando o máximo de três ocorrências em um único dia.

A experiência de visualização espacial dos eventos, levando em consideração a variação temporal, leva à conclusão de que, apesar de ambos eventos acontecerem em dias iguais, seus registros temporais sofrem variações sensíveis em minutos e segundos, o que é um comportamento esperado, pois trata-se de eventos distintos. Para uma melhor visualização das falhas em linhas de transmissão e dos focos de queimadas detectados, talvez fosse melhor usar uma granularidade temporal mais abstrata, como "Hora do dia", por exemplo. Essa estratégia proporcionaria várias visualizações simultâneas de falhas e focos ao longo dos 10 anos de dados, entretanto, o custo para o analista em visualizar imagem por imagem, manualmente, em busca de comportamentos interessantes nos dados espaço-temporais, seria muito elevado.

A utilização da técnica de *clustering* surge como uma boa opção para diminuir o custo do analista, uma vez que um algoritmo de mineração de dados fará o agrupamento espaço-temporal dos eventos e o GeoSTAT oferecerá suporte para visualização e navegação dos resultados gerados pelo algoritmo, ou seja, *clusters* espaço-temporais contendo registros de focos e faltas.

Com as camadas "FALTAS" e "FOCOS" adicionadas ao ambiente do GeoSTAT, aciona-se a opção de *clustering* espaço-temporal, oferecida pelo ambiente, para realização de mineração de dados com as duas camadas. Essa abordagem permitirá que o analista visualize

tanto os agrupamentos espaço-temporais de cada evento em separado, como também os *clusters* relevantes⁶.

Para a execução da mineração de dados, além das camadas de entrada, o usuário deve informar os seguintes parâmetros requeridos:

Granularidade Temporal: dentre as opções disponíveis, adotou-se a opção "3-3 Horas".

Esta decisão foi tomada em razão da resolução temporal do conjunto de dados relacionado a focos de incêndio detectados, que é de, aproximadamente, três horas;

Algoritmo de Mineração Dados: dentre as opções disponíveis, adotou-se o DBScan. Esta escolha se deu pelo fato do algoritmo DBScan não exigir, como parâmetro de entrada, a quantidade total de *clusters* a serem gerados, haja vista que é difícil prever tal informação. Outro ponto importante que justifica essa escolha é em razão da existência de parâmetros que permitem aplicar restrições espaciais de área e distância na formação dos *clusters*;

Parâmetros do Algoritmo: os valores utilizados para os parâmetros de entrada do algoritmo DBScan foram: *MinPoints* = 2, uma vez que um *cluster* com, pelo menos, um objeto de cada conjunto de dados, pode revelar uma possível autocorrelação; *Épsilon* = 0,013472, como o raio dos *clusters* gerados.

A escolha do valor 0,013472 para o parâmetro *Épsilon* do DBScan partiu do conhecimento de que um segundo (unidade de medida angular) é aproximadamente equivalente a 30,9 metros⁷. Como cerca de 80% dos focos de incêndio detectados pelo INPE ocorrem em um raio de um quilômetro das coordenadas indicadas, e que o erro médio na localização espacial dos registros é de 400 metros, achou-se razoável que o raio de um *cluster* gerado fosse um valor entre 1 e 1,5 quilômetros. Logo, sabendo que 48,5 segundos equivale a, aproximadamente, 1.498,65 metros (1,5 quilômetros), e que 1 grau decimal é equivalente a 60 minutos e 60 segundos, tem-se que:

⁶*Cluster Relevante:* um *cluster* é considerado "relevante" neste estudo quando apresenta, em sua formação, pelo menos um registro de ocorrência de cada evento estudado, ou seja, pelo menos um "FOCO" e uma "FALTA".

⁷Cálculo de distância entre coordenadas geográficas. Disponível, em inglês, em: <http://www.movable-type.co.uk/scripts/latlong.html>

$$48,5 \text{ segundos} = \frac{48,5}{(60 \times 60)} = 0,013472 \text{ graus decimais, logo,}$$

$$0,013472 \text{ graus decimais} \approx 1.498,65 \text{ metros.}$$

Com os parâmetros necessários informados pelo analista, o GeoSTAT se encarregou pelo processamento da mineração de dados baseada em *clustering*.

5.1.4 Resultados

O processamento da mineração de dados deste estudo custou 7 horas, 37 minutos e 5 segundos. A execução se deu a partir de um servidor de aplicações *web* rodando em uma máquina com sistema operacional Microsoft Windows 7 Professional (64-bit), com processador Intel Core i7 e 16 GB de memória RAM.

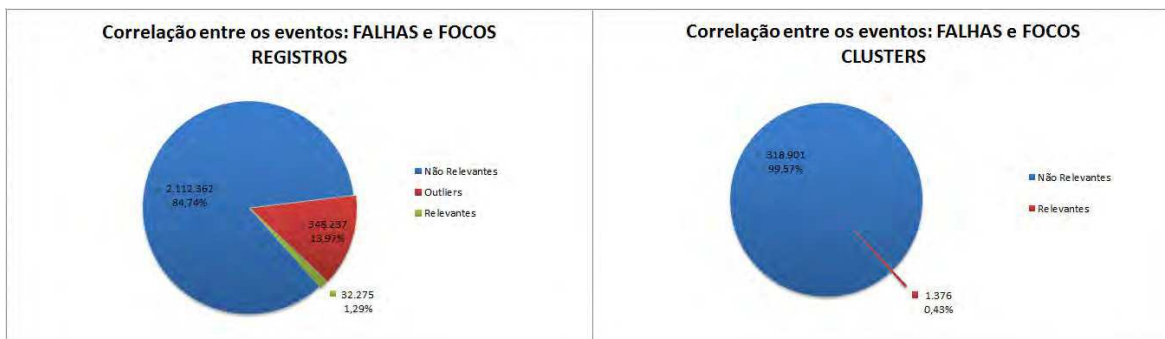


Figura 5.4: Resultados obtidos com a execução da mineração de dados a partir do GeoSTAT.

A Figura 5.4 apresenta os resultados estatísticos obtidos com o processamento da mineração de dados a partir do GeoSTAT. O gráfico de pizza da esquerda apresenta a classificação dos registros após a execução do algoritmo. Apenas 32.275 registros, 1,29% de todo o conjunto de dados, foi considerado relevante pelo GeoSTAT. Isso significa que apenas esses registros estão contidos em *clusters* espaço-temporais relevantes, àqueles que contêm registros de ambos os eventos estudados. Cerca de 86,03% dos registros foram associados

a um *cluster* espaço-temporal. O restante dos registros, 13,97% do total, foram considerados "outliers" por não fazerem parte de nenhum *cluster* espaço-temporal, ou seja, retratam ocorrências pontuais no espaço-tempo.

O gráfico de pizza da direita, por sua vez, apresenta uma relação no quantitativo de *clusters* gerados. Dos 318.901 *clusters* espaço-temporais gerados, apenas 1.376 (0,43%) foram considerados relevantes, sob o ponto de vista dos parâmetros de medida utilizados na execução do algoritmo de mineração de dados. Cada *cluster* não-relevante agrupou, em média, 6.623 registros, enquanto que cada *cluster* relevante agrupou, em média, apenas 23 registros.

A Figura 5.5 apresenta uma imagem capturada do GeoSTAT exibindo no mapa todos os *clusters* espaço-temporais relevantes, gerados nesse processo de mineração de dados. A primeira informação que pode ser percebida pelo analista, diante dessa visualização, é de que a região que concentrou mais *clusters*, foi a região localizada no sudeste do estado do Ceará, mais precisamente na divisa com os estados da Paraíba e do Rio Grande do Norte. A região metropolitana de Maceió (AL) e do Recife (PE), além da região do município de Sobral, no interior do Ceará, são outras regiões que apresentaram muitos *clusters*.

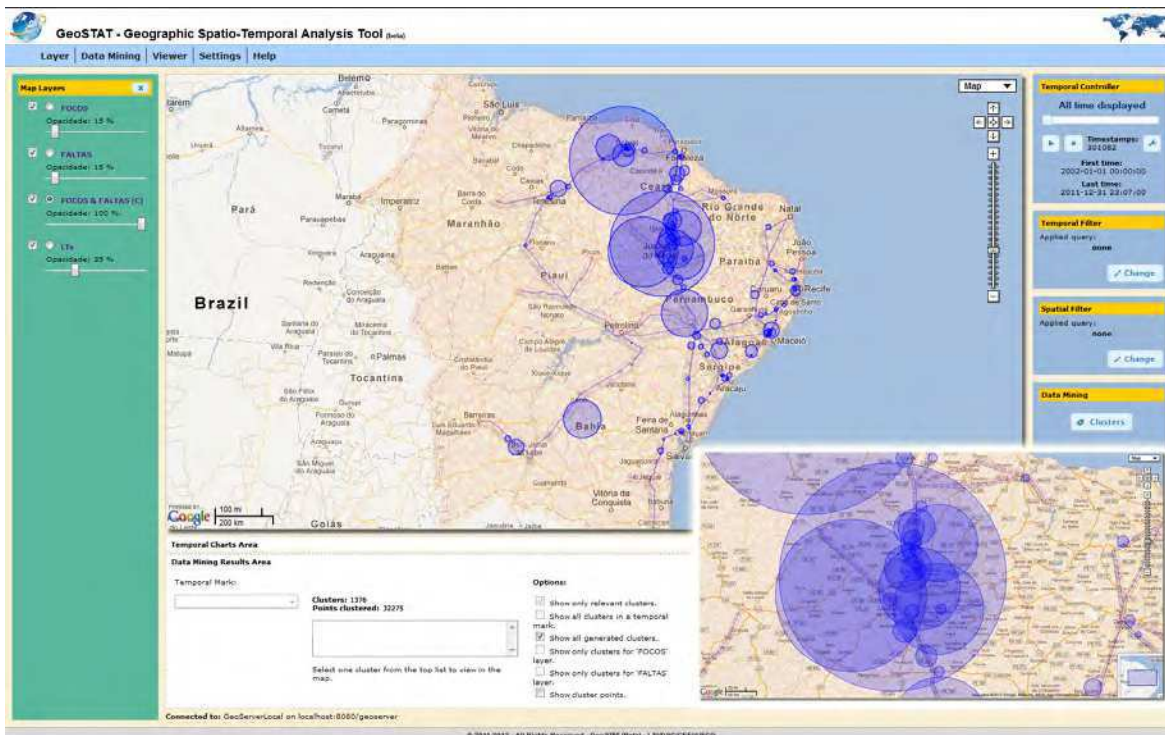


Figura 5.5: GeoSTAT exibindo todos os *clusters* relevantes gerados.

A navegação pelos *clusters* espaço-temporais gerados por um processo de mineração de dados executado a partir do GeoSTAT é realizada a partir dos componentes para seleção da marcação temporal e, a partir dessa definição, a seleção individual de cada *cluster* gerado individualmente. O analista pode optar pela visualização exclusiva dos *clusters* relevantes, ou por visualizar todos os *clusters* gerados. Ele pode, também, visualizar cada *cluster* individualmente, ou ainda, visualizar todos os *clusters* de uma marcação temporal.

A Figura 5.6 apresenta uma imagem capturada do GeoSTAT, exibindo no mapa todos os *clusters* espaço-temporais relevantes, gerados para eventos registrados no dia 03/11/2004 entre 15 e 18 horas. A imagem exhibe cinco *clusters* agrupando, no total, 14 registros.

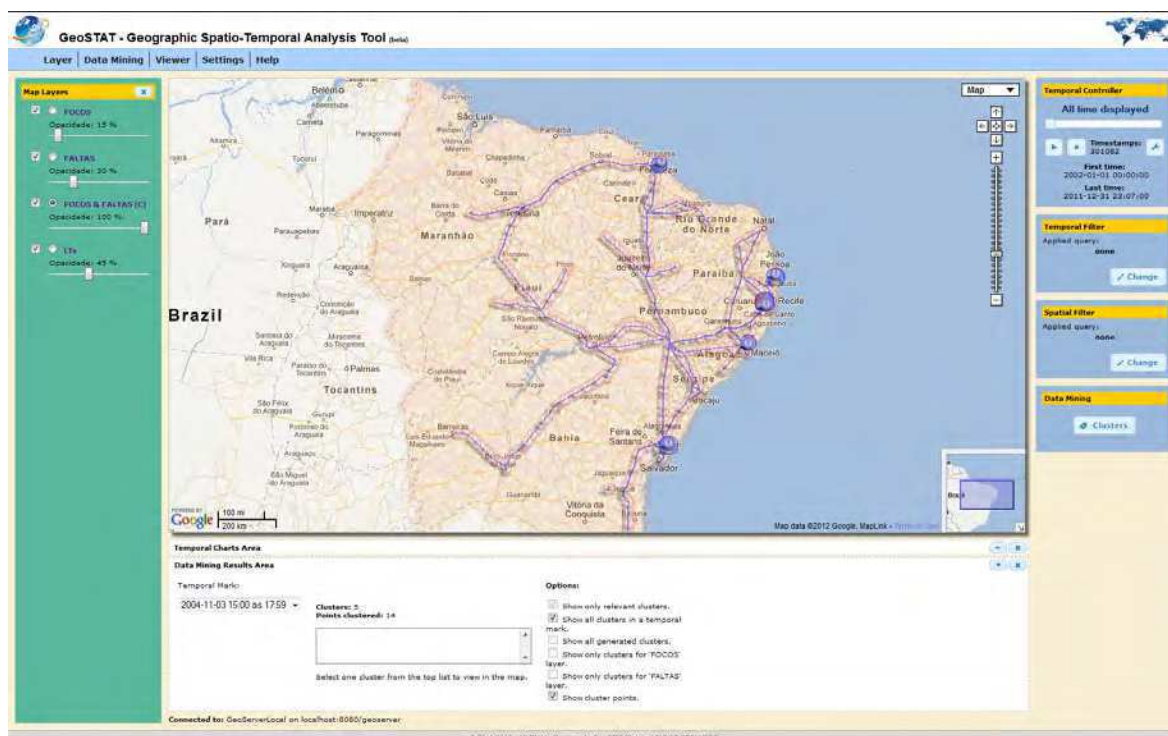


Figura 5.6: GeoSTAT exibindo todos os *clusters* relevantes gerados para 03/11/2004, das 15 às 17 horas e 59 minutos.

Utilizando a opção de navegação "*cluster por cluster*", é possível observar um *cluster* espaço-temporal específico. A Figura 5.7 exemplifica essa opção exibindo, com detalhes, um dos *clusters* registrados na marcação temporal "03/11/2004 15:00 às 17:59".

Para o analista, interessado em confirmar sua hipótese de que alguns focos de incêndio são causadores de falhas em linhas de transmissão de energia elétrica, a Figura 5.7 exemplifica um caso em que tal hipótese se confirma. Uma falha ocorrida na linha "FORTALEZA

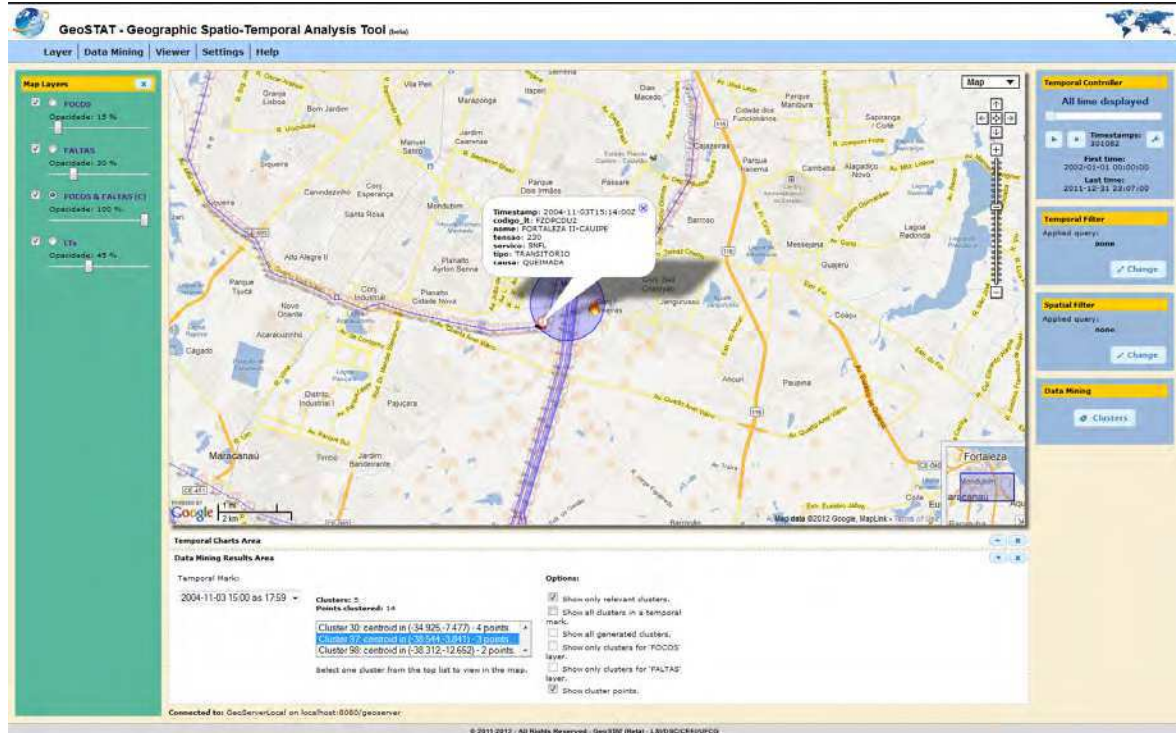


Figura 5.7: GeoSTAT exibindo, em detalhes, o *cluster* espaço-temporal nº 97, da marcação temporal "03/11/2004 15:00 às 17:59".

II - CAUIPE" às 15h14 do dia 03/11/2004, teve como causa apontada "QUEIMADA" e, além disso, em virtude da mineração de dados realizada em conjunto com dados de registro de focos de incêndio detectados na região, no mesmo período da falha, apontou um agrupamento espaço-temporal entre esta falha e dois focos de incêndio: um detectado às 16h08 a, aproximadamente, 1 km ao leste da falha; e outro, detectado às 16h01 a, aproximadamente, 1,5 km ao norte da falha. Se, levados em consideração os erros de precisão espacial e a resolução temporal das informações, o analista poderia julgar esses dois focos como causadores da falha registrada.

5.1.5 Conclusões

Os resultados obtidos com a utilização do ambiente GeoSTAT e, em função disso, da aplicação da nova abordagem em visualização e análise de dados espaço-temporais, foram satisfatórios para o domínio de aplicação explorado neste estudo. Os recursos de visualização explorados permitiram a descoberta de informações interessantes implícitas, em meio a duas

grandes massas de dados.

É importante observar que os resultados estatísticos da mineração de dados apontaram um índice de *clusters* relevantes abaixo do esperado por qualquer especialista nesse tipo de evento. Isso se deve, principalmente, ao fato da utilização de registros simulados de falhas em linhas de transmissão. A utilização de dados reais, capturados e estruturados pela Eletrobrás/Chesf, proporcionarão, certamente, melhores resultados no sentido de maior ocorrência *clusters* relevantes.

Além da utilização de dados reais, os especialistas têm à disposição, através da implementação da nova abordagem proposta, diversos algoritmos de *clustering* espaço-temporal e diversos *setups* para os parâmetros desses algoritmos. Seus resultados podem ser comparados e analisados em busca da descoberta de novas informações relevantes.

5.2 Estudo 2: Análise visual espaço-temporal de fotografias georeferenciadas

Este estudo consiste na análise visual de um conjunto de dados contendo fotografias capturadas no espaço-tempo, utilizando a implementação da nova abordagem proposta neste trabalho.

5.2.1 Motivação

Atualmente, milhares de dados espaço-temporais são gerados ao redor do mundo. Isso se dá, principalmente, com a difusão dos aparelhos portáteis, providos de dispositivos GPS e de conectividade à Internet. Por meio desses dispositivos, seus usuários enviam, frequentemente, para outros dispositivos móveis ou para as redes sociais, mensagens de texto e/ou multimídia, compartilhando suas informações de localização geográfica, além da informação temporal implícita (data e hora de envio).

Em meio às milhares de mensagens multimídia postadas na rede diariamente, surge um interesse na visualização espaço-temporal de fotografias. Um curioso em fotografia, ou uma pessoa interessada em conhecer as características visuais de um lugar que vai visitar, antes de realizar tal viagem, pode estar interessado em visualizar as imagens disponíveis para deter-

minada região do globo terrestre e, também, saber se pode confiar na realidade exposta por essas imagens ou se, em função da informação temporal, deve desconfiar de que a realidade exposta na imagem já não condiz com a realidade atual.

Pelo menos duas marcações espaciais e temporais podem ser encontradas em uma fotografia: a data e hora em que a mesma foi fotografada, e a localização geográfica em que isso aconteceu; e a data e hora em que a mesma foi enviada à rede, juntamente com a localização geográfica de onde esse envio partiu. Diante desse tipo de informação, surge também a necessidade de uma abordagem que possibilite uma visualização eficaz. Partindo dessa necessidade, este estudo de caso utilizará o GeoSTAT para uma análise visual de um conjunto de fotografias contendo dados espaço-temporais de sua captura.

5.2.2 Dados

Para a realização deste estudo, foi utilizado um banco de dados contendo dados georeferenciados de fotografias capturadas em algumas regiões da América do Sul. Esses dados foram obtidos a partir do projeto *PhotoGeo* (de Figueirêdo et al., 2012).

O conjunto de dados consiste de 5.134 imagens, fotografadas entre outubro de 2005 e junho de 2012, em algumas regiões da América do Sul. Os dados espaço-temporais foram obtidos a partir de um banco de dados *PostgreSQL+PostGIS*, com sistema de referência geográfica WGS84, e informações temporais seguindo o fuso horário oficial de Brasília (GMT-3).

Utilizou-se a mesma instalação do *Geoserver* utilizada no estudo de caso anterior, para o compartilhamento dos dados em forma de camada espacial, acessível pelos serviços WMS e WFS.

Utilizou-se, também, uma camada de dados espaciais compartilhada na *web* a partir de um outro servidor *Geoserver*, contendo a geografia de todos os municípios brasileiros obtida junto ao Instituto Brasileiro de Geografia e Estatística (IBGE).

5.2.3 Condução

O ambiente GeoSTAT foi inicializado para carregar as camadas deste estudo de caso. A imagem exibida na Figura 5.8 apresenta a visualização da camada "FOTOS", contendo os

dados espaço-temporais das fotografias.

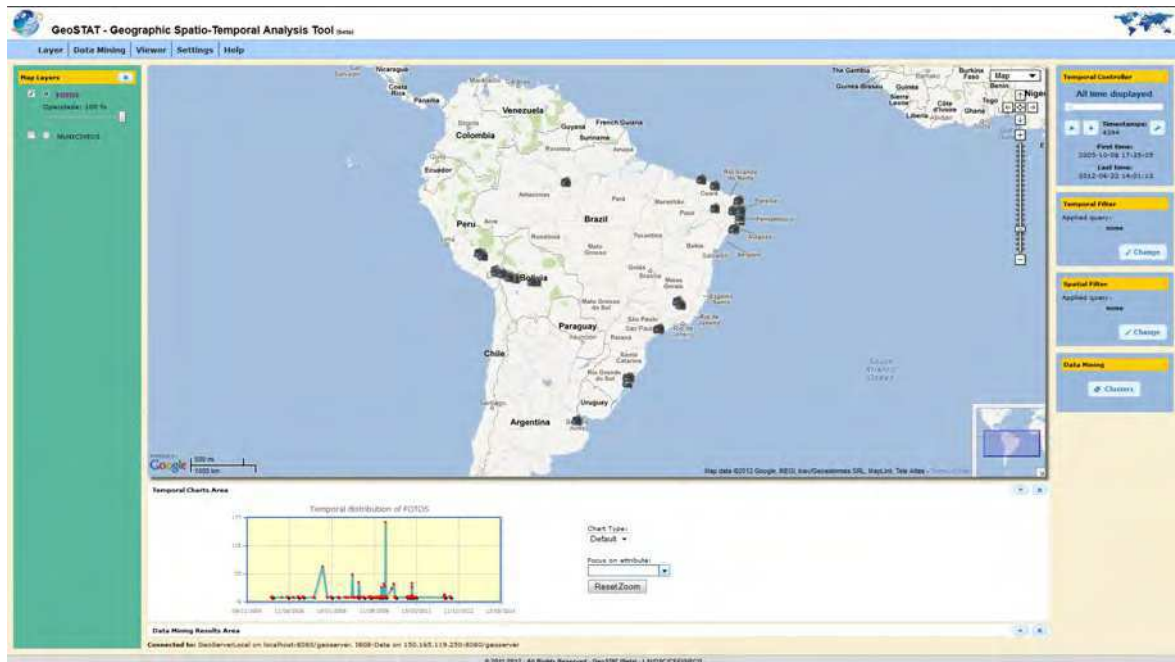


Figura 5.8: GeoSTAT exibindo a visualização espaço-temporal da camada "FOTOS".

É possível observar a distribuição espacial dos dados no mapa. São registros em quatro regiões brasileiras, além de registros também em outros países da América do Sul, como Peru, Bolívia e Argentina. O controlador temporal exibe as marcações temporais que delimitam o conjunto de dados em análise. A marcação inicial data em "08/10/2005 às 17:35", e a final em "22/06/2012 às 14:01", totalizando num período de, aproximadamente, 7 anos, com 4.394 marcações temporais distintas. O gráfico de distribuição temporal permite a identificação imediata do ponto máximo de fotografias capturadas em um dia. Esse ponto foi atingido em "15/01/2010", registrando 167 fotografias somente nessa data.

Diante da visualização dos registros de fotografias, adicionou-se a camada contendo a geografia dos municípios brasileiros. Com essa informação visualizada no ambiente, é possível realizar consultas e aplicar filtros espaciais com o GeoSTAT. O usuário da aplicação, interessado em detalhes sobre as fotografias registradas no estado da Paraíba, gostaria de executar a seguinte consulta espacial:

Quais municípios, que fazem limite territorial com a capital João Pessoa, contêm fotografias capturadas?

Para a realização dessa consulta e visualização dos resultados, o usuário do GeoSTAT precisou executar apenas três solicitações de consulta espacial em sequência. Essas solicitações são apresentadas na Figura 5.9.

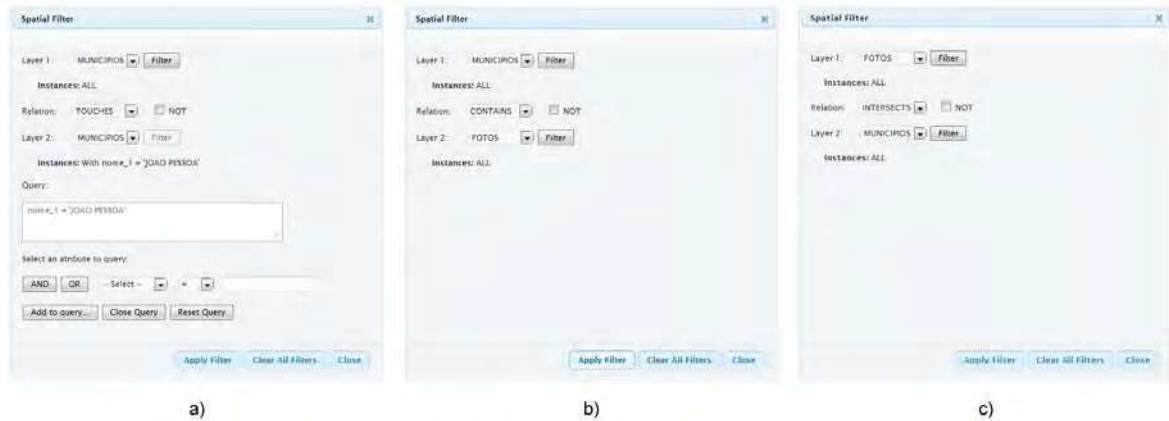


Figura 5.9: Sequência de solicitações realizadas por usuário do GeoSTAT para realizar a consulta espacial desejada.



Figura 5.10: Resultados sequenciais de consulta espacial realizada por meio do GeoSTAT.

O primeiro passo para realização da consulta espacial desejada é solicitar os municípios que fazem fronteira com o município "João Pessoa". Essa solicitação é ilustrada na imagem (a) da Figura 5.9. A primeira consulta retorna então os cinco municípios que fazem fronteira com a capital da Paraíba (vide Figura 5.10a). Utiliza-se então o resultado da primeira consulta para formular a segunda consulta, solicitando, entre esses municípios, aqueles que contêm, pelo menos, uma fotografia capturada. Essa solicitação é apresentada na imagem (b) da Figura 5.9. Dos cinco municípios que fazem fronteira com a capital da Paraíba, apenas

dois deles (Alhandra e Cabedelo) contêm fotografias capturadas (vide Figura 5.10b). Por fim, a terceira e última consulta filtra a camada "FOTOS" de tal forma que só sejam exibidos os registros de fotografias que foram capturados nesses dois municípios. O resultado final é exibido na Figura 5.11.

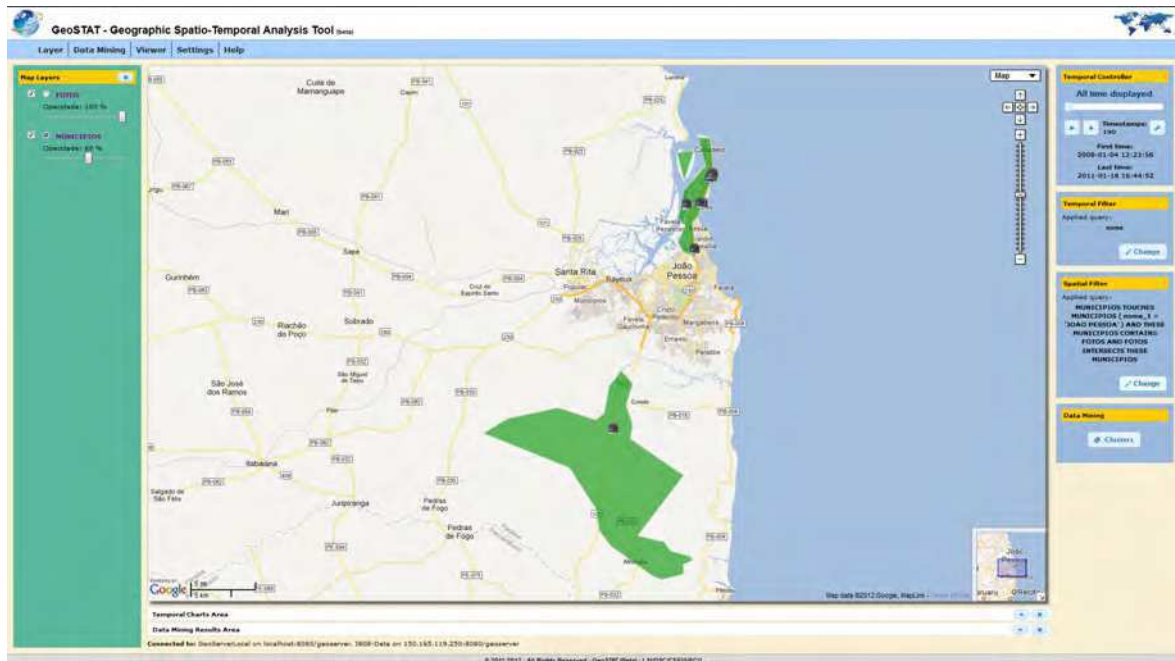


Figura 5.11: Resultado final da consulta espacial que retorna os municípios, que fazem limite com João Pessoa, que contém fotografias capturadas.

O resultado da consulta espacial aplicada retornou dois municípios: Alhandra e Cabedelo; e 190 fotografias capturadas, sendo dessas, 7 capturadas em Alhandra e 183 capturadas em Cabedelo, em meses de dezembro e janeiro. Diante dessas informações, o usuário do GeoSTAT pode concluir que o responsável pelas fotografias estava aproveitando o verão no litoral paraibano.

Além de contemplar o recurso de consultas espaciais, a nova abordagem proposta também contempla a utilização de consultas temporais. Com a base de dados de fotografias utilizada nesse estudo de caso, o usuário utilizou o recurso de consulta temporal (ou filtro temporal) do GeoSTAT para analisar visualmente a variação temporal da base de dados. O filtro temporal foi aplicado para o ano de 2012, até o mês de junho, e para os últimos três anos completos.

A Figura 5.12 apresenta as visualizações de mapa no GeoSTAT, válidas para cada ano

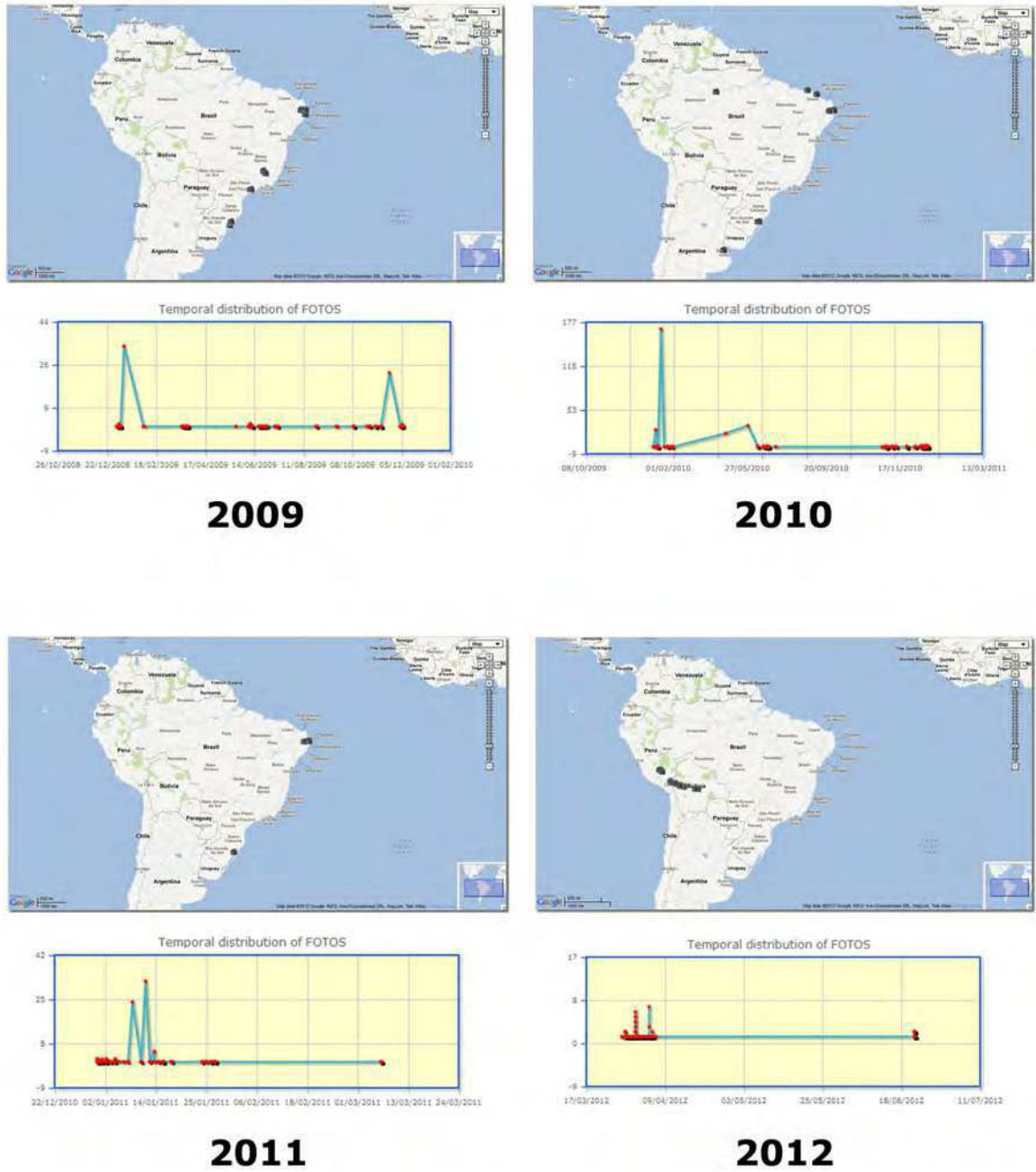


Figura 5.12: Distribuição espacial e temporal do conjunto de fotografias georeferenciadas capturadas nos últimos quatro anos.

solicitado. É possível observar as mudanças nas distribuições espacial e temporal do conjunto de dados ao longo dos últimos três anos e meio. O usuário pode concluir, diante dessas visualizações:

- que 2011 foi o ano em que o fotógrafo realizou menos deslocamento espacial, uma vez que as imagens capturadas se concentram em apenas dois estados brasileiros;
- que, em 2012, o fotógrafo visitou a região de fronteira entre os países do Peru e da Bolívia, e fotografou em duas ocasiões: num período entre os últimos dias de março e primeiros dias de abril, e nos últimos dias de junho; sendo 05 de abril o dia em que mais fotografou, capturando o total de 7 imagens somente neste dia;
- que, em 2010, o fotógrafo registrou o maior número de fotografias capturadas em um único dia, sendo 167 só no dia 15 de janeiro, realizou uma viagem internacional para a Argentina, e uma viagem até Manaus, sendo o ano que realizou o maior deslocamento espacial.

5.2.4 Resultados e Conclusões

Este estudo de caso teve como objetivo aplicar os recursos de visualização espaço-temporal incluídos na abordagem proposta neste trabalho. A utilização da implementação desses recursos, através do GeoSTAT, trouxe resultados satisfatórios, mostrando a experiência do usuário ao visualizar e manipular um conjunto razoável de dados espaço-temporais.

Foi possível demonstrar, também, a facilidade em conectar-se a camadas de dados localizadas em fontes distintas espalhadas na *web*, de maneira que fosse possível realizar uma análise visual a partir do cruzamento de diversas camadas de dados espaço-temporais, realizando operações espaciais e temporais sobre elas.

5.3 Resumo do Capítulo

Este capítulo apresentou dois estudos de caso realizados com a implementação da abordagem de visualização e análise de dados espaço-temporais proposta neste trabalho de dissertação, e teve, como principal objetivo, validar o conjunto de ideias propostas na nova abordagem.

No primeiro estudo de caso, utilizou-se uma grande massa de dados para analisar a correlação espaço-temporal entre focos de incêndio reais e eventos simulados de falha em linhas de transmissão de energia elétrica. Utilizando o GeoSTAT, foi possível visualizar particularidades nas dimensões espacial e temporal dos dados espaço-temporais, a partir de componentes como o mapa dinâmico georeferenciado e o gráfico de distribuição temporal. A mineração de dados baseada em *clustering* espaço-temporal revelou agrupamentos espaço-temporais relevantes, entretanto, em um quantitativo abaixo do esperado na visão do usuário especialista no domínio.

O segundo estudo de caso teve como objetivo demonstrar o funcionamento dos recursos de consulta espacial e consulta temporal oferecidos pelo GeoSTAT, e as vantagens percebidas pelo usuário que analisou um conjunto de dados espaço-temporais de fotografias georeferenciadas. O estudo ainda provou a possibilidade de utilização de vários servidores de compartilhamento de dados espaço-temporais em um único processo de análise visual, e mostrou que o esforço do usuário para realização dessa tarefa é mínimo.

No próximo capítulo, serão apresentadas as considerações finais deste trabalho e as possibilidades de trabalhos futuros visando sua extensão.

Capítulo 6

Considerações Finais

Neste trabalho, foi apresentada uma nova abordagem para visualização e análise de dados espaço-temporais. Dada a necessidade e motivação em visualizar e analisar dados espaço-temporais, realizou-se uma pesquisa sobre as características específicas referentes à dimensão espacial e temporal, e a complexidade em manipular, simultaneamente, essas características em dados espaço-temporais.

A partir do levantamento do estado da arte, foi possível relacionar os pontos fortes e os pontos fracos das principais abordagens existentes. Além da constatação de que ainda existem poucas abordagens que tratam, de fato, as dimensões espacial e temporal de forma simultânea em dados espaço-temporais, as principais lacunas encontradas em visualização e análise espaço-temporal foram em questões como flexibilidade e genericidade das soluções.

Nas poucas abordagens que abrangem, de fato, dados espaço-temporais, ou seja, que oferece o tratamento simultâneo em ambas dimensões, percebeu-se soluções voltadas a domínios específicos, atingindo um grupo específico de usuários. Outro ponto importante no estado da arte em visualização de dados espaço-temporais diz respeito à interoperabilidade. As soluções são construídas para operar com bancos de dados específicos, sem opções de conectividade com outros bancos de dados. Por fim, questões de usabilidade também precisam ser melhor trabalhadas, de forma que a visualização possibilite a obtenção de interpretações imediatas por parte do usuário final.

O uso da mineração de dados se mostrou como uma característica inovadora, que vem sendo incorporada nas soluções mais recentes. O uso do agrupamento (*clustering*) espaço-temporal, além de possibilitar uma considerável redução na base de dados espaço-temporais

analisada e a detecção de *outliers*, ou seja, pontos discretos no espaço-tempo, possibilita a descoberta de padrões de correlação espaço-temporal e, dessa forma, exerce forte contribuição no processo de análise espaço-temporal.

O uso dos padrões de compartilhamento e interoperabilidade de dados espaciais, desenvolvidos pela OGC, abriu o leque de possibilidades de análise espaço-temporal, uma vez que tornou possível a visualização combinada de dados espaciais e espaço-temporais oriundos de diferentes fontes de dados.

A realização de uma avaliação comparativa entre a nova abordagem proposta e as abordagens relacionadas mostrou que a nova abordagem proposta conseguiu endereçar as seis características necessárias a uma abordagem de visualização e análise espaço-temporal: recursos para a dimensão espacial, recursos para a dimensão temporal, independência de domínio, flexibilidade, interoperabilidade e mineração de dados baseada em *clustering* espaço-temporal. Trata-se de uma abordagem que prioriza o usuário final, oferecendo um conjunto de funcionalidades que permitem a realização de um trabalho eficaz.

O ambiente GeoSTAT também foi apresentado neste trabalho. Trata-se de uma nova solução em visualização e análise espaço-temporal que contempla, como principais diferenciais: a incorporação de diversos algoritmos de *clustering*, dando ao usuário várias opções para mineração de dados e a manipulação visual dos resultados; a visualização de gráficos dinâmicos de distribuição temporal; e um controlador temporal para visualização espacial em função do tempo.

O ambiente foi implementado a partir da nova abordagem apresentada, o que permitiu a realização de avaliações concretas de funcionalidade por meio de estudos de caso baseados em situações e problemas reais do cotidiano.

O primeiro estudo de caso mostrou que é possível realizar, com eficiência e praticidade, uma análise espaço-temporal em busca de correlação entre dois grandes conjuntos de dados distintos, obtendo informações relevantes para solução de problemas reais em empresas do ramo de geração e transmissão de energia elétrica, como o caso da Eletrobrás/Chesf.

O segundo estudo de caso, utilizando um banco de dados espaço-temporal de fotografias georeferenciadas, mostrou que o GeoSTAT e a nova abordagem proposta permitem uma análise exploratória sobre conjuntos de dados espaço-temporais, independente de domínio e de modo de armazenamento, dando a liberdade necessária ao usuário final, que tem à

disposição recursos que permitem filtrar as informações no espaço e no tempo, a exemplo dos recursos de consulta espacial e consulta temporal, cujo funcionamento foi demonstrado nesse estudo.

Por fim, chega-se à conclusão de que a nova abordagem proposta atendeu seus objetivos, mostrando-se implementável, satisfatória e eficiente. Chega-se à conclusão, também, de que vários pontos de expansão podem ser explorados em trabalhos futuros, o que certamente contribuirá para a construção de um ambiente mais robusto.

6.1 Trabalhos Futuros

Durante o desenvolvimento deste trabalho, foram notados pontos de expansão que poderiam trazer benefícios à infraestrutura do GeoSTAT.

À medida que os componentes para visualização espaço-temporal têm recebido novas funcionalidades e, também, diante do surgimento de novas técnicas para visualização de dados espaço-temporais, surgem novas ideias que podem ser incorporadas ao ambiente de forma a enriquecê-lo.

O primeiro ponto de expansão é na utilização de outras técnicas para a mineração de dados espaço-temporais: como regras de associação e classificação. A inclusão de algoritmos de mineração de dados baseados em classificação e regras de associação pode, assim como *clustering*, auxiliar os usuários na descoberta de conhecimentos implícitos em meio a grandes volumes de dados espaço-temporais, se configurando como outras opções de descoberta de conhecimento semiautomatizada.

Ainda na questão da mineração de dados, existe a ideia de desenvolver uma interface para inclusão de novos algoritmos de mineração de dados baseados em *clustering* por parte dos usuários, aumentando a gama de opções disponíveis pelo ambiente e dando ao usuário a liberdade de implementar novos algoritmos de agrupamento espaço-temporal e testar o funcionamento e a performance através do GeoSTAT.

A adoção da técnica de redução de dimensionalidade é outra possibilidade de trabalho futuro, visando melhorar o desempenho dos algoritmos de mineração de dados e, consequentemente, reduzir o custo dessa operação.

No recurso de gráficos dinâmicos para a dimensão temporal, existe um ponto de expan-

são com a implementação de outros tipos de gráfico, além do já implementado gráfico de distribuição, a exemplo de histogramas, gráficos de pizza e em espiral, dando ao usuário final mais opções para visualização dos comportamentos temporais.

Outro ponto de expansão consiste na criação de ambientes exclusivos para cada usuário. Dessa forma, cada usuário seria identificado e, a partir daí poderia, por exemplo, ter uma lista personalizada de servidores de dados preferenciais, armazenar esquemas de visualização e sobreposição de camadas para retomar análises em outro momento, e ainda, realizar mineração de dados confidenciais e visualizar os resultados em um ambiente de mesma característica e com a segurança necessária. Outro ponto, a partir dessa expansão, é promover um ambiente colaborativo entre os usuários do ambiente, com componentes de comunicação, anotação e compartilhamento de análises e conclusões entre os usuários.

Por último, um ponto de expansão que consiste no desenvolvimento de uma interface para publicação e exportação dos dados gerados pelo ambiente, para que possam ser utilizados por servidores de compartilhamento de dados espaço-temporais ou por outras ferramentas de interesse do usuário final.

Referências Bibliográficas

- AGRAWAL, R.; SRIKANT, R., Fast algorithms for mining association rules in large databases. In: *Proceedings of the 20th International Conference on Very Large Data Bases*, VLDB. **Anais...** San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1994, p. 487–499.
- AKTAS, M. S.; AYDIN, G.; FOX, G. C.; PIERCE, M.; SAYAR, A., *Information Services for Grid/Web Service Oriented Architecture (SOA) Based Geospatial Applications*, **Technical Report**, Indiana University, USA, 2005.
- AMORIM, E. G.; BAPTISTA, C. S.; OLIVEIRA, M. G.; DE FIGUEIRÊDO, H. F., Uso de Geoprocessamento na Análise das Ocorrências em Linhas de Transmissão. In: *II Encontro Nacional de Geoprocessamento do Setor Elétrico (ENGEEO)*. **Anais...** Recife, PE, Brasil: Eletrobrás/Chesf, 2012.
- ANDRIENKO, G.; ANDRIENKO, N.; BREMM, S.; SCHRECK, T.; LANDESBERGER, T. V.; BAK, P.; KEIM, D., *Space-in-Time and Time-in-Space Self-Organizing Maps for Exploring Spatiotemporal Patterns*, **Computer Graphics Forum**, v. 29, p. 913–922, 2010a.
- ANDRIENKO, G.; ANDRIENKO, N.; DEMSARB, U.; DRANSCHC, D.; DYKES, J.; FABRIKANT, S. I.; JERNF, M.; KRAAKG, M.-H.; SCHUMANNH, H.; TOMINSKIH, C., *Space, time and visual analytics*, **International Journal of Geographical Information Science**, v. 24, n. 10, p. 1577–1600, 2010b.
- ANDRIENKO, G.; ANDRIENKO, N.; DYKES, J.; FRABRIKANT, S. I.; WACHOWICZ, M., *Geovisualization of dynamics, movement and change: key issues and developing approaches in visualization research*, **Information Visualization**, v. 7, n. 3-4, p. 173–180, 2008.
- ANDRIENKO, G.; ANDRIENKO, N.; WROBEL, S., *Visual Analytics Tools for Analysis of Movement Data*, **SIGKDD Explor. Newsl**, v. 9, p. 38–46, 2007.
- ANDRIENKO, G.; MALERBA, D.; MAY, M.; TEISSEIRE, M., *Mining spatio-temporal data*, **Journal of Intelligent Information Systems**, v. 27, p. 187–190, 2006.

- ANDRIENKO, N.; ANDRIENKO, G.; GATALSKY, P., *Exploratory spatio-temporal visualization: an analytical review*, **Journal of Visual Languages & Computing**, v. 14, n. 6, p. 503–541, 2003.
- ANKERST, M.; BREUNIG, M. M.; KRIEGEL, H.-P; SANDER, J., OPTICS: Ordering Points To Identify the Clustering Structure. In: *Proceedings of ACM SIGMOD international conference on Management of data*. **Anais...** ACM Press, 1999, p. 49–60.
- ARTHUR, D.; VASSILVITSKII, S., k-means++: the advantages of careful seeding. In: *Proceedings of the 18th annual ACM-SIAM symposium on Discrete algorithms*. **Anais...** Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2007, p. 1027–1035.
- BAPTISTA, C. S.; JUNIOR, O. F. D. L.; OLIVEIRA, M. G.; ANDRADE, F. G. D.; SILVA, T. E.; PIRES, C. E. S., Using OGC Services To Interoperate Spatial Data Stored In SQL And NoSQL Databases. In: *Proceedings of The XII Brazilian Symposium on GeoInformatics*. **Anais...** Campos do Jordão, SP, Brasil: MCT/INPE, 2011, p. 61–72.
- BÉDARD, Y.; MERRETT, T.; HAN, J., *Fundamentals of Spatial Data Warehousing for Geographic Knowledge Discovery*, **Vol. Research Monographs in GIS**, Taylor & Francis, chapter 3, p. 53–73, 2001.
- CAVALCANTI, V. M. B.; SCHIEL, U.; BAPTISTA, C. S., Querying Spatio-Temporal Databases Using a Visual Environment. In: *Proceedings of the working conference on advanced visual interfaces, AVI '06*. **Anais...** New York, NY, USA: ACM, 2006, p. 412–419.
- CHEN, H.; ATABAKHSH, H.; PETERSEN, T.; SCHROEDER, J.; BUETOW, T.; CHABOYA, L.; O'TOOLE, C.; CHAU, M.; CUSHNA, T.; CASEY, D.; HUANG, Z., COPLINK: Visualization for Crime Analysis. In *Proceedings of The National Conference on Digital Government Research*. **Anais...** Boston, MA, USA: Information Sciences Institute, 2003, p. 1–6.
- COMPIETA, P.; MARTINO, S. D.; BERTOLOTTA, M.; FERRUCCI, F.; KECHADI, T., *Exploratory spatiotemporal data mining and visualization*, **Journal of Visual Languages & Computing**, v. 18, n. 3, p. 255–279, 2007.
- DAN PELLEGG, A. M., X-means: Extending K-means with Efficient Estimation of the Number of Clusters. In: *Proceedings of the Seventeenth International Conference on Machine Learning*. **Anais...** San Francisco, CA, USA: Morgan Kaufmann, 2000, p. 727–734.
- DASGUPTA, S.; LONG, P. M., Performance Guarantees for Hierarchical Clustering, **Journal of Computer and System Sciences - Special issue on COLT 2002**, v. 70, n. 4, p. 555–569, 2005.

- DE FIGUEIRÊDO, H. F.; DA SILVA, J. P. R.; PAIVA, A. C.; BAPTISTA, C. S., People Photo Annotation Using Context and Content. In: *The 6th International Workshop on Multimedia and Signal Processing*. **Anais...** Vienna, Austria, 2012, p. 35–38.
- DE LA BEAUJARDIERE, J., *OpenGIS Web Map Service 1.3.0 Implementation Specification*, **OpenGIS Implementation Specification OGC 06-042**, Open Geospatial Consortium Inc., 2006.
- ERWIG, M., Design of Spatio-Temporal Query Languages. In: *Proceedings of the NSF/BDEI Workshop on Spatio-Temporal Data Models of Biogeophysical Fields for Ecological Forecasting*. **Anais...** San Diego, CA, USA, 2002, p. 11–23.
- ESTER, M.; KRIEGEL, H.-P.; SANDER, J.; XU, X., A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In: *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining (KDD-96)*. **Anais...** Portland, OR, EUA, 1996, p. 226–231.
- FERREIRA, N.; LINS, L.; FINK, D.; KELLING, S.; WOOD, C.; FREIRE, J.; SILVA, C., *BirdVis: Visualizing and Understanding Bird Populations*, **IEEE Transactions on Visualization and Computer Graphics**, v. 17, n. 12, p. 2374–2383, 2011.
- GEERTS, F.; HAESEVOETS, S.; KUIJPERS, B., A Theory of Spatio-Temporal Database Queries. In: *Proceeding DBPL '01 Revised Papers from the 8th International Workshop on Database Programming Languages*. **Anais...** London, UK: Springer-Verlag, 2002, p. 198–212.
- GENNARI, J. H.; LANGLEY, P.; FISHER, D., Models of incremental concept formation, **Artificial Intelligence**, v. 40, p. 11–61, 1990.
- HALL, M.; FRANK, E.; HOLMES, G.; PFAHRINGER, B.; REUTEMANN, P.; WITTEN, I. H., *The WEKA data mining software: an update*, **SIGKDD Explorations**, v. 11, n. 1, p. 10–18, 2009.
- HAN, J., Mining Spatiotemporal Knowledge: Methodologies and Research Issues. In: *Proceedings of The Geospatial Visualization and Knowledge Discovery Workshop*. **Anais...** Lansdowne, PA, USA: University Consortium for GIS, 2003, p. 25–27.
- HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J., **The Elements of Statistical Learning**, New York: Springer, *chapter 8.5. The EM algorithm*, p. 236–243, 2001.
- JENSEN, C. S.; SNODGRASS, R. T., *Temporal Data Management*, **IEEE Transactions on Knowledge and Data Engineering**, v. 11, p. 36–44, 1999.
- JOHNSTON, W. L., **Information visualization in data mining and knowledge discovery**, Morgan Kaufmann Publishers, San Francisco, CA, USA, *chapter 16. Model Visualization*, p. 223–227, 2001.

- KAMP, V.; GRAWUNDER, M.; CRUPE, R.; HINRICHS, M.; OLDENETTEL, F.; WEIDLICH, S.; ZACHEWITZ, L., Database System Support for Spatio-Temporal Aspects in Scientific Applications. In: *Proceedings of International Database Engineering and Applications Symposium*. **Anais...** Cardiff, UK, 1998, p. 48–49.
- KOHONEN, T., *Self-Organizing Maps*, **Information Sciences**, v. 30, 3rd ed., Springer-Verlag, 2001.
- KOPANAKIS, I.; THEODOULIDIS, B., *Visual data mining modeling techniques for the visualization of mining outcomes*, **Journal of Visual Languages and Computing**, v. 14, n. 6, p. 543–589, 2003.
- KOPERSKI, K.; HAN, J.; ADHIKARY, J., *Mining Knowledge in Geographical Data*, **Communications of the ACM**, v. 41, n. 12, p. 47–66, 1998.
- KRALIDIS, A. T., **Open Source Approaches in Spatial Data Handling**, Springer-Verlag Berlin Heidelberg, *chapter 1. Geospatial Open Source and Open Standards Convergences*, p. 1–20, 2008.
- LIME, S., **Open Source Approaches in Spatial Data Handling**, Springer-Verlag Berlin Heidelberg, *chapter 4. MapServer*, p. 65–86, 2008.
- MILLER, H. J., *Potential contributions of spatial analysis to geographic information systems for transportation*, **Geographical Analysis**, v. 1, n. 31, p. 373–399, 1999.
- MILLER, H. J.; HAN, J., *Geographic data mining and knowledge discovery, An overview*. In: **Geographic data mining and knowledge discovery**, Taylor & Francis., London, UK, p. 3–32, 2001.
- OGC, *OGC - Making Location count*, **Open Geospatial Consortium**, 2011. Disponível em: <http://www.opengeospatial.org/>. Consultado em 05/07/2012.
- OLIVEIRA, M. G.; BAPTISTA, C. S.; AMORIM, E. G., Utilizando Geoprocessamento e Mineração de Dados em Busca de Correlação entre Ocorrências de Focos de Incêndio e Desligamentos em Linhas de Transmissão. In: *II Encontro Nacional de Geoprocessamento do Setor Elétrico (ENGE0)*. **Anais...** Recife, PE, Brasil: Eletrobrás/Chesf, 2012a, p. 1–14.
- OLIVEIRA, M. G.; BAPTISTA, C. S.; FALCÃO, A. G. R., A Web-based Environment for Analysis and Visualization of Spatio-temporal Data provided by OGC Services. In: *Proceedings of GEOProcessing 2012: The Fourth International Conference on Advanced Geographic Information Systems, Applications, and Services*. **Anais...** Valencia, Spain: IARIA, 2012b, p. 183–189.
- PATEL, J., *Temporal Database System*, **Technical Report**, Department of Computing, Imperial College, University of London, London, UK, 98 p., 2003.

- PESSOA, A. S. A.; SIMÕES, J. D., Mineração de Dados Espaço-Temporal Aplicada a Previsão Climática Utilizando a Teoria dos Conjuntos Aproximativos. In: *III Workshop dos Cursos de Computação Aplicada do INPE – WORCAP*. **Anais...** São José dos Campos, SP, Brasil, 2003, p. 91–96.
- PEUQUET, D. J., **Representations of Space and Time**, The Guilford Press, 2002.
- REDA, K.; TANTIPATHANANANDH, C.; BERGER-WOLF, T.; LEIGH, J.; JOHNSON, A. E., SocioScape - a Tool for Interactive Exploration of Spatiotemporal Group Dynamics in Social Networks. In: *Proceedings of the IEEE Information Visualization Conference (INFOVIS'09)*. **Anais...** Atlantic City, NJ, USA, 2009, p. 1–2.
- RODDICK, J. F.; SPILIOPOULOU, M., *A survey of temporal knowledge discovery paradigms and methods*, **IEEE Transactions on Knowledge and Data Engineering**, v. 14, p. 750–767, 2002.
- ROTH, R. E.; ROSS, K. S.; FINCH, B. G.; LUO, W.; MACEACHREN, A. M., A user-centered approach for designing and developing spatiotemporal crime analysis tools. In: *Proceedings of GIScience 2010*. **Anais...** Zurich, Suíça, 2010, p. 66–71.
- SHEKHAR, S.; VATSAVAI, R. R.; CELIK, M., **Data Mining: Next Generation Challenges and Future Directions**, AAAI/MIT Press, *chapter 1. Spatial and Spatiotemporal Data Mining: Recent Advances*, p. 1–34, 2008.
- SHEKHAR, S.; ZHANG, P.; HUANG, Y.; VATSAVAI, R., **Data Mining: Next Generation Challenges and Future Directions**, AAAI/MIT Press, *chapter 3. Trends in Spatial Data Mining*, p. 357–380, 2003.
- SILVA, S. L. F.; SCHIEL, U.; CATARCI, T., *Formalizing Visual Interaction with Historical Databases*, **Information Systems**, v. 27, n. 7, p. 487–521, 2002.
- TOMINSKI, C.; SCHUMANN, H., Enhanced interactive spiral display. In: *Proceedings of the annual SIGRAD conference - special theme: interactivity*. **Anais...** Stockholm, Sweden, 2008, p. 1–4.
- TURTON, I., **Open Source Approaches in Spatial Data Handling**, Springer-Verlag Berlin Heidelberg, *chapter 8. GeoTools*, p. 153–170, 2008.
- VRETANOS, P. P. A., *OpenGIS Web Feature Service 2.0 Interface Standard*, **OpenGIS Implementation Specification OGC 09-025r1 e ISO/DIS 19142**, Open Geospatial Consortium Inc., 2010.
- WANG, X.; ZHOU, X.; LU, S., Spatiotemporal Data Modeling and Management: A Survey. In: *Proceedings of the 36th International Conference on Technology of Object-Oriented Languages and Systems*. **Anais...** Xi An, Shaanxi, China: IEEE Computer Society, 2000, p. 202–221.

YAO, X., Research Issues in Spatio-temporal Data Mining. In: *Workshop on Geospatial Visualization and Knowledge Discovery*. **Anais...** Lansdowne, PA, USA: University Consortium for GIS, 2003, p. 60–65.