LUÍS EMÍLIO CAVECHIOLLI DALLA VALLE

UM MODELO PARA RECONHECIMENTO DE PADRÕES EM IMAGENS DE SATÉLITES CLIMÁTICOS COM BASE EM LINGUAGENS FORMAIS

Dissertação apresentada à Escola Politécnica da Universidade de São Paulo para obtenção do Título de Mestre em Ciências.

São Paulo 2012

LUÍS EMÍLIO CAVECHIOLLI DALLA VALLE

UM MODELO PARA RECONHECIMENTO DE PADRÕES EM IMAGENS DE SATÉLITES CLIMÁTICOS COM BASE EM LINGUAGENS FORMAIS

Dissertação apresentada à Escola Politécnica da Universidade de São Paulo para obtenção do Título de Mestre em Ciências.

Área de Concentração: Sistemas Digitais

Orientador: Ricardo Luis de Azevedo da Rocha

São Paulo 2012 Este exemplar foi revisado e alterado em relação à versão original, sob responsabilidade única do autor e com a anuência de seu orientador.

São Paulo, 23 de Setembro de 2012.

Assinatura do autor

Assinatura do orientador

FICHA CATALOGRÁFICA

Dalla Valle, Luís Emílio Cavechiolli

Um Modelo para Reconhecimento de Padrões em Imagens de Satélites Climáticos com Base em Linguagens Formais/ L. E. Cavechiolli Dalla Valle. – ed. rev. – São Paulo, 2012.

77 p.

Dissertação (Mestrado) — Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia de Computação e Sistemas Digitais (PCS).

1. Linguagens Formais 2. Reconhecimento de Padrões 3. Imageamento de Satélite I. Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia da Computação e Sistemas Digitais Departamento de Engenharia de Computação e Sistemas Digitais (PCS). II. t.

Este trabalho é dedicado à minha família

AGRADECIMENTOS

Deus, que felizmente, pela ciência não podemos provar sua existência ou inexistência, mostrando que nossa ciência, filosofia e religião não encerram tudo, que não são um fim, mas um meio, que dependemos muito mais de nosso lado humano para agir e se sustentar, agradeço por tudo, mesmo que o tudo seja maior que minha concepção. Agradeço ao Sr. Florisvaldo e Dona Zuleica, meus queridos pais, que sempre estiveram presentes na evolução deste desafio, obrigado pela torcida e confiança. Às mulheres da minha vida, Mariana, Marina, Sônia, Sueli, Eliza, Elza, Rosa e Léia, cada uma, de seu modo, apoiando esta jornada, meu muitíssimo obrigado. Ao nosso herói Agamenon e seu filho, meu querido primo João, aos amigos Celso, Marília, Leandro e Tamara pela torcida e apoio. Ao querido orientador, que sempre se manteve íntegro, firme, e soube indicar com precisão o ponto de equilíbrio durante minhas escorregadas. Ao professor João José Neto, que muito auxiliou direta e indiretamente na construção deste trabalho, especialmente nos momentos entre esfihas e tabules.

Agradecimento especial ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelos recursos fornecidos, e ao Instituto Nacional de Pesquisas Espaciais, pelo fornecimento dos dados, sem o qual esta pesquisa não se realizaria.

RESUMO

Uma sequência de imagens de satélite climático é processada aplicando-se um conjunto de operações de filtros, no intuito de extrair padrões de comportamento das nuvens. Caracteres são criados a partir deste tratamento e suas transições são investigadas, explorando a possibilidade de justificar suas ocorrências através de linguagens formais e linguagens bidimensionais, definindo suas gramáticas. Com esta contagem de transições, uma análise de sua forma fractal é iniciada e um paralelo com outras contagens estabelecida, como uma forma de estruturar um modelo computacionalmente menos complexo de prever o tempo, ou o comportamento de qualquer entidade dinâmica que possa ser discretizada. Com estas investigações e experiências, foi possível diminuir a quantidade de símbolos utilizados para justificar as formas das nuvens, bem como criar classes de equivalências para representar conjuntos de símbolos que compartilham mesmas propriedades, diminuindo ainda mais a complexidade da gramática que se espera encontrar.

ABSTRACT

A sequence of weather satellite images are processed by applying a set of filtering operations in order to extract the behavior patterns of clouds. Characters are created from this treatment and their transitions are investigated by exploring the possibility of justifying their occurrence across formal languages and two-dimensional languages, defining their grammar. With these count transitions an analysis of their fractals starts and counts a parallel with others established as a way to structure a model less computationally complex to predict the weather, or the behavior of any dynamic entity that could be discretized. With these investigations and experiments, it was possible to reduce the number of symbols used to explain the shapes of clouds and create equivalent classes to represent the symbol sets that share the same properties, further reducing the complexity of the grammar expected to be found.

SUMÁRIO

Lista de llustrações

Lista de Tabelas

Lista de Abreviaturas e Siglas

Lista de Símbolos

1	Intro	odução		13
	1.1	Apres	entação	13
	1.2	Objeti	vos	16
	1.3	Organ	ização	16
2	Ferr	rament	al Teórico Pesquisado	18
_				
	2.1	Lingua	agens Formais	18
		2.1.1	Conceitos Básicos	18
		2.1.2	Hierarquia de Linguagens	19
	2.2	Autôm	natos Finitos Determinísticos	21
	2.3	Adapt	atividade	23
		2.3.1	Autômatos Adaptativos de Estado Finito	23
	2.4	Lingua	agens Bidimensionais	24
		2.4.1	Formalização das Linguagens Bidimensionais	25

	2.5	Princípio do Tamanho Mínimo de Descrição	26
	2.6	Redes Bayesianas	28
	2.7	Fluídos Estáveis	28
3	Aqu	isição de Dados	31
	3.1	As Imagens	31
	3.2	Os Filtros	33
	3.3	Os Símbolos	42
4	Aná	lise dos Dados	46
	4.1	Contagem dos Símbolos e suas Transições	46
		4.1.1 Análise da Contagem de Transições	47
	4.2	Clusterização: Ordenação por Ocorrências	48
	4.3	Busca de Padrões em Altura, Largura e Profundidade	51
	4.4	Comentários sobre Gramática	54
		4.4.1 Contextualização	54
	4.5	Busca por Equivalências em Profundidade	55
5	O S	imulador Atmosférico	68
	5.1	Descrição	68
	5.2	Mapa de Elevação	68
6	Con	siderações Finais	72
	6.1	Contribuições	72

6.2	Conclusões	73
6.3	Trabalhos Futuros	74

Referências

LISTA DE ILUSTRAÇÕES

1	Representação em Grafo do Autômato da Tabela 1 22
2	Imagem Captada pelo Satélite MeteoSat-9 (MSG2) no Canal
	Visível 2 (0.8μm)
3	Seção Correspondente à América do Sul
4	Imagem de Satélite. Copyright 2010-2012 EUMETSAT 37
5	Filtro Median: Retirada dos Limites Territoriais
6	Threshold de Cor - Retém Apenas Nuvens
7	Threshold de Tons de Cinza - Transformação para Imagem Binária 40
8	Detector de Bordas de Canny - Apenas Contorno das Nuvens . 41
9	Os Primeiros Dezesseis Símbolos do Alfabeto 43
10	Caracteres Indexados pelo Mapa de Cores
22	Busca Utilizada na Rede Bayesiana
11	Representação da Densidade da Contagem Realizada em No-
	vembro de 2011
12	Representação da Densidade da Contagem para Outro Con-
	junto de Dados Realizada em Abril de 2012 60
13	Total de Diferença das Contagens Realizadas em Períodos Dis-
	tintos
14	Representação da Densidade da Contagem para Conjunto Arti-
	ficial de Dados Utilizando Distribuição Normal dos Símbolos 62

15	5 Clusterização da Contagem de Transições Representada pela		
	Figura 11 Ordenada pela Soma das Ocorrências Verticais e Ho-		
	rizontais	63	
16	Partes da Contagem de Transições com Repetições do Padrão		
	Encontrado	64	
17	Estrutura de Dados para Busca de Padrões	65	
18	Expansão das Transições no Tempo	65	
19	Cobertura de Nuvens no Tempo. Nuvens em Azul e Áreas Des-		
	cobertas em Vermelho	66	
20	Falha na Emissão de Dados em uma Sub-Região da Imagem	67	
21	Sequência Regular de Transição	67	
23	Mapa de Elevação global: Modelo ETOPO1	69	
24	Comparação Visual Entre a Região Simulada em Destaque e o		
	Resto do Relevo Terrestre	70	
25	Estado Atual do Simulador	71	

LISTA DE TABELAS

1	Representação em Tabela de um Autômato Finito 2		
2	Ordenação de Bits para Construção dos Símbolos	43	
3	Primeiras Linhas da Tabela de Contagem	46	
4	Probabilidade de Ocorrência dos Símbolos Equivalentes ao Símbolo 64 (30 imagens)	57	
5	Probabilidade de Ocorrência dos Símbolos Equivalentes ao Símbolo 64 (300 imagens)	58	

LISTA DE ABREVIATURAS E SIGLAS

CPTEC	Centro de Previsão do Tempo e Estudos Climáticos		
Eumetsat	European Organization for the Exploitation of Meteorologi-		
cal Satellit	es		
INPE	Instituto Nacional de Pesquisas Espaciais		
NASA	National Aeronautics and Space Administration		
NOAA	National Oceanic and Atmospheric Administration		
Poli	Escola Politécnica da Universidade de São Paulo		
USP	Universidade de São Paulo		

LISTA DE SÍMBOLOS

- Σ Conjunto Finito de Símbolos
- Σ^{\star} Todas as combinações possíveis de símbolos sobre Σ .
- $\Sigma^{\star\star}$ Todas as combinações possíveis de símbolos sobre Σ formando cadeias bidimensionais.
- $\Gamma^{3\times 3}$ Mesmo que $\Sigma^{\star\star}$, porém limitado a cadeias de tamanho 3×3 .

1 INTRODUÇÃO

1.1 Apresentação

Tanto a criação da vida, como sua evolução e manutenção, têm sua engrenagem principal nos diversos ciclos que ocorrem em todo o planeta, além da influências de outros corpos celestes como a Lua e o Sol. A começar pela rotação, translação e a translação lunar, influenciando nos ciclos de dias, das estações e das marés, seguindo para a variação de temperatura dos oceanos, evaporação da água, criação de nuvens, precipitação e raios, passando para as movimentações internas do planeta, através das placas tectônicas, com formação de vulcões, alteração de terreno, maremotos, etc. Toda a vida se molda através destas variações.

O poder transformador destes fenômenos é impressionante. Contudo, assim como a vida foi criada com auxílio destes, eles também podem ser responsáveis por sua extinção parcial ou total.

Nós, seres humanos, como responsáveis por este planeta, influenciamos diretamente as alterações que estes ciclos sofrem, tornando fundamental termos ferramentas que auxiliem no entendimento destes fenômenos tão complexos e caóticos, tanto para prever o seu comportamento, como auxiliar a tomar decisões que podem mudar os rumos do futuro que nos espera.

A variação das condições atmosféricas, em suas infinitas possibilidades,

sempre foi questão de preocupação para a humanidade. Associada ao estado emocional de divindades em várias civilizações primitivas (e modernas), e com muitos recursos incentivando seu estudo nestes últimos séculos, sua importância se deve principalmente por nossa fragilidade diante dos vários tipos de eventos catastróficos que podem ocorrer, a curto e longo prazo, e também para nos tornar conscientes sobre como nossas ações influenciam na intensidade e variações destes eventos.

A computação, como pilar para todas as ciências, também contribui para a análise destes fenômenos, auxiliando em sua previsão, tomada de decisões e emissão de alertas quando necessário, possibilitando evacuações e, dessa forma, salvar vidas.

Na meteorologia, ciência que estuda a atmosfera, seus processos e seus fenômenos, um dos usos da computação é a previsão numérica do tempo. Neste processo, é utilizado um modelo topográfico em alta definição do território a ser analisado. Seu estado inicial é configurado à partir dos dados coletados de fontes diversas como em fotos de satélites, estações meteorológicas e radares meteorológicos. Na sequência, o simulador é acionado, recebendo o modelo numérico como entrada e a previsão do tempo para alguns dias é criada com algumas horas de processamento ((CPTEC-INPE), 2010).

Por se tratar de movimento de fluidos, que é caracteristicamente caótico, quanto mais distante a previsão, menor a precisão (LORENZ, 1965), tornando necessário o reprocessamento constante do modelo numérico com dados mais recentes.

Normalmente, a precisão do modelo é configurada de acordo com a capacidade de processamento do supercomputador utilizado, e do tempo estimado para sua realização, bem como de outros tipos de simulação que precisam ser executadas. Desta forma, o uso do hardware é otimizado no intuito de não desperdiçar seus recursos, especialmente tempo de uso, de forma a enquadrar as simulações necessárias, no tempo disponível para processamento ((CPTEC-INPE), 2010).

No Brasil, o órgão que cuida da previsão do tempo é o Centro de Previsão do Tempo e Estudos Climáticos do Instituto Nacional de Pesquisas Espaciais. Este é responsável por assimilar dados meteorológicos/climáticos de diversas fontes diferentes, inclusive internacionais; criar e avaliar modelos de simulação atmosférica e oceânica; realizar previsões do tempo com boletins e análises técnicas; monitorar ocorrências como geadas, queimadas, ondas, qualidade do ar e radiação ultra-violeta; realizar previsões climáticas (de longo prazo); divulgar os resultados dos diversos modelos regionais e globais, através de meteogramas e disponibilizar as imagens de satélite e mapas correspondentes, entre outras ((CPTEC-INPE), 2010).

O conjunto de modelos dinâmicos que o CPTEC utiliza para realizar as previsões do tempo são o ETA regionais, ETA globais, ensemble regional e global, e oceânico. Nestes são utilizados os modelos ETA (originário da legra grega η que representa a pressão atmosférica em função da altura neste modelo) (ETA, 2010), Brams (BRAMS, 2010), Ensemble (ENSEMBLE, 2010), Global (GLOBAL, 2010) e WWatch (WWATCH, 2010; TOLMAN, 1989), integrados em regiões e resoluções diferentes, que variam entre 5x5Km até 100x100km e de 3 a 15 dias de previsão.

Todos estes modelos utilizam-se de integrações numéricas das equações físicas que regem os fluídos, e a interação com o relevo, vegetação, etc., Consequentemente, necessitam de supercomputadores para realizar estas tarefas em tempo hábil.

O presente trabalho investiga um modelo alternativo para analisar as variações do tempo e os movimentos da atmosfera, utilizando apenas imagens de satélite, no intuito de encontrar um método de previsão que se iguale ao modelo simulado em termos estatísticos, mas que seja computacionalmente mais rápido, sem exigir tantos recursos.

1.2 Objetivos

O objetivo geral deste projeto é explicar a formação e movimentação das nuvens em sequências de imagens de satélite, através da definição de uma linguagem, cujas de regras gramaticais, ao serem aplicadas, possam reproduzir com fidelidade o mesmo comportamento encontrado na natureza.

Esta dissertação cuida de estruturar um método para a produção de modelos computacionais de previsão de tempo cuja complexidade computacional seja inferior aos métodos definidos pela área de meteorologia.

Como objetivo secundário, associar um meio de medir a acuidade do método em comparação aos modelos atualmente utilizados.

Definir um modelo de simulação onde o método criado possa ser testado sob condições controladas.

1.3 Organização

O trabalho está organizado de forma a dar uma sequência nos conhecimentos da mesma maneira como foram adquiridos durante a pesquisa e nos experimentos, na ordem em que foram realizados.

O Capítulo 2 aponta algumas técnicas formais pesquisadas durante o trabalho, como Linguagens Formais, Linguagens Bidimensionais, Autômatos Finitos, Adaptatividade, Princípio do Tamanho Mínimo de Descrição e Fluídos Estáveis. O Capítulo 3 descreve como os dados foram obtidos à partir de uma sequência de imagens de satélite, e como se extraiu os símbolos das mesmas.

O Capítulo 4 descreve os processos de análise dos dados realizados neste trabalho, as contagens e as buscas de padrões, bem como a utilização de métodos probabilísticos para inferir regularidades entre os símbolos, como base para entendimento de alguns fenômenos.

O Capítulo 5 descreve detalhes do simulador atmosférico em construção para este projeto.

Capítulo 6 conclui o trabalho.

2 FERRAMENTAL TEÓRICO PESQUISADO

2.1 Linguagens Formais

As linguagens formais desempenham um papel muito importante na computação. Graças à formalização das linguagens, compiladores, tradutores, e diversos sistemas se beneficiam de suas características e funcionalidades.

2.1.1 Conceitos Básicos

Uma linguagem nada mais é do que um conjunto de cadeias formado sobre um conjunto de símbolos denominado alfabeto. A este conjunto de símbolos, dá-se o nome de Σ , e a linguagem é formada pela combinação destes símbolos. De todas as linguagens possíveis sobre um alfabeto Σ , representado por Σ^* , formando todas as combinações possíveis de cadeias de tamanho $n: 0 \le n \le \infty$, é possível extrair um subconjunto bastante grande de linguagens que são definidas através das regras de produção, ou também denominadas gramáticas.

Uma gramática, como mostrado em (PAPADIMITRIOU; LEWIS, 1998), (RA-MOS; NETO; VEGA, 2009) e (ROSEN, 2007), é uma quádrupla $G = (V, \Sigma, P, S)$, onde *V* representa o vocabulário da gramática, é um conjunto finito e não vazio composto de símbolos terminais e não-terminais, Σ é o conjunto finito e não vazio de símbolos terminais, ou denominado alfabeto da linguagem, $N = V - \Sigma$ é o conjunto de símbolos não-terminais da linguagem, $P = \{(\alpha,\beta) | (\alpha,\beta) \in V^* N V^* \times V^*\}$ forma o conjunto finito de regras de produção de cadeias e $S \in N$ é o símbolo inicial ou raiz da linguagem.

O conjunto *P* composto pelas regras de produção, representa sequências de derivações ou substituições envolvendo os símbolos não-terminais e terminais. Elas se apresentam como duplas, mas representam substituição de cadeias.

Sendo duas cadeias, $\alpha \in V^*NV^*$ e $\beta \in V^*$, representa-se por $\alpha \to \beta$ a substituição da cadeia α por β , cujo intuito é atingir uma cadeia no final composta apenas por símbolos de Σ (terminais).

Como exemplo, uma Linguagem *L*, composta pelos símbolos 0 e 1, de forma a, se começar com 0, deve terminar com 1, apresentando a mesma quantidade de ambos, e os símbolos semelhantes devem permanecer agrupados, pode ser representada pela gramática G = $(\{0,1,S\},\{0,1\},\{S \rightarrow 0S1,S \rightarrow 01\},S)$ e possui cadeias como 01, 0011, 000111, 00001111, ..., $0^n 1^n$.

Conforme a configuração destas regras, definem-se classes ou hierarquias à qual uma determinada linguagem pertence.

2.1.2 Hierarquia de Linguagens

Conforme a maneira que o conjunto de regras de produção forma a linguagem, é possível classificá-la dentro de uma hierarquia de linguagens. Esta hierarquia, foi primeiramente definida por Chomsky em 1956 (CHOMSKY, 1956), que as classificou em 4 tipos diferentes.

As descrições dos tipos pertencentes às hierarquias seguem como mostradas em (RAMOS; NETO; VEGA, 2009) As Gramáticas do Tipo 0, chamadas de gramáticas irrestritas, caracterizam uma classe de linguagens denominada recursivamente enumeráveis. São gramáticas que permitem qualquer tipo de associação entre símbolos, gerando todas as possibilidades de combinações de cadeias. Podem ser reconhecidas/geradas por Máquinas de Turing de fita infinita.

As Gramáticas do Tipo 1, chamadas de gramáticas sensíveis ao contexto, são gramáticas cujas regras de produção associam não-terminais com nãoterminais, de forma que transições do tipo $aSb \rightarrow aAb$ ocorram. São reconhecidas/geradas por Máquinas de Turing com fita limitada, ou por autômatos adaptativos.

As Gramáticas do Tipo 2, chamadas de livres de contexto, são linguagens $L = \{a^n b^n | n > 1\}$. As regras de produção da gramática *G* que gera *L* são da forma $P = \{S \rightarrow aS b | ab\}$ e são reconhecidas por autômatos de pilha.

Expressões algébricas como ((3+2)*5+(2-1)/2) são exemplos de cadeias pertencentes a Linguagens Livres de Contexto

As Gramáticas do Tipo 3, chamadas de gramáticas regulares, são as mais simples, cujas regras e produção são caracterizadas por permitir um crescimento unilateral da cadeia. Esta cresce ou para a direita ou para a esquerda, e são reconhecidas por um autômato finito.

G é um exemplo de gramática regular à direita, onde *G* = $({S,A,a,b,c},{a,b,c},P,S)$, onde *P* segue:

$$P = \{S \to aS, S \to bA, A \to cA, A \to \epsilon\}$$

Além destes modelos de gramáticas, outros modelos, como os sistemas de reescrita (BAADER; NIPKOW, 1999), (BOOK; OTTO, 1993) e (DERSHOWITZ; JOU-ANNAUD, 1990) também foram pesquisados.

2.2 Autômatos Finitos Determinísticos

Automato Finito é um modelo matemático computacional que representa, de forma sintética, um dispositivo mínimo capaz de realizar computação. É constituído de duas partes, uma fita de entrada que contém símbolos, que o autômato está configurado para reconhecer através de uma cabeça de leitura, e o controle finito, uma "caixa preta" que contém todos os estados finais ou não e suas respectivas transições, que verifica o símbolo atual da fita, dada a posição da cabeça de leitura, analisa se o estado atual possui transição para o símbolo lido e transita de estado conforme o caso, repetindo esta operação enquanto houver elementos na fita. Possui três maneiras diferentes de parar. A primeira ocorre quando a cadeia termina e o estado de transição é final, parando em acerto. A segunda, é quando lê um símbolo cujo estado atual não o reconhece, não existindo transição correspondente, parando em erro. A terceira ocorre quando a fita acaba, e o estado de transição não é final, parando em erro novamente. Segundo (HOPCROFT; MOTWANI; ULLMAN, 2001), suas utilizações mais importantes são:

- Software para design e verificação do comportamento de circuitos digitais.
- O "analisador léxico" de um compilador convencional, que quebra o texto em unidades lógicas, como identificadores, palavras reservadas e pontuação.
- Software de escaneamento de textos longos, como coleções de páginas web, para encontrar e contabilizar ocorrências de palavras, frases ou outros padrões.
- 4. Software para verificar sistemas de todos os tipos que possuem número

Estado	Entrada	Transição
δ_0	a	δ_1
δ_1	а	δ_1
δ_1	b	δ_0

Tabela 1: Representação em Tabela de um Autômato Finito

finito de estados, como protocolos de comunicação, ou protocolos para troca segura de informação.

Sua definição matemática segue como apresentado em (PAPADIMITRIOU; LEWIS, 1998):

Um automato finito determinístico é uma quíntupla $M = (K, \Sigma, \delta, s, F)$, onde:

K é um conjunto finito de estados,

 Σ é um alfabeto reconhecido pelo autômato,

 $s \in \Sigma$ é o estado inicial,

 $F \subseteq K$ é o conjunto de estados finais,

 δ é a função de transição, $\delta: K \times \Sigma \to K$

Existem duas formas de representação de autômatos mais comumente utilizadas, através de tabelas de decisão (Tabela 1) e grafos (Figura 1).



Figura 1: Representação em Grafo do Autômato da Tabela 1.

A palavra "determinístico", no autômato finito implica na característica de que cada estado possui no máximo uma transição para um determinado símbolo (não vazio " ϵ ") do alfabeto, de forma a não haver ambiguidades em seu

comportamento.

2.3 Adaptatividade

Segundo (NETO, 2002), adaptatividade é uma característica apresentada por um dispositivo formal quando este possui a capacidade de alterar sua própria estrutura interna de forma autônoma, sem interferência externa, nem mesmo de seus usuários, mudando, em consequência, seu próprio comportamento em resposta direta ao estímulo de entrada.

Pode ser aplicada a virtualmente qualquer dispositivo baseado em regras, como tabelas adaptativas e autômatos finitos. Para isto, basta acrescentar à estrutura original do dispositivo, um conjunto de ações adaptativas que serão executadas quando o dispositivo transitar pelos estados.(NETO, 1993; NETO; BRAVO, 2003).

2.3.1 Autômatos Adaptativos de Estado Finito

Baseado em (NETO, 2002), um Autômato Adaptativo nada mais é do que a aplicação de um mecanismo adaptativo ao autômato finito convencional definido em 2.2

Como visto em (PISTORI; NETO, 2002):

Definição 1. Um autômato Adaptativo de Estado Finito (AAF) pode ser definido como uma óctupla $M = (Q, \Sigma, q_0, F, \delta, Q_{\infty}, \Gamma, \Pi)$, onde:

 $Q \subseteq Q_{\infty}$ é um conjunto finito de estados

 Σ é o alfabeto

 q_0 é o estado inicial

 $F \subseteq Q$ é o conjunto de estados finais

 $\delta \subseteq (Q_{\infty} \times (\Sigma \cup \{\epsilon\}) \times Q_{\infty})$ é a relação de transição

 Q_{∞} conjunto teoricamente infinito de estados

 $\Gamma \subseteq (\{+,-\} \times (Q_{\infty} \times (\Sigma \cup \{\epsilon\}) \times Q_{\infty})) \text{ conjunto de ações primitivas de inclusão}$ (+) ou exclusão (-) de transições ¹

 $\Pi: (Q_{\infty} \times (\Sigma \cup \{\epsilon\}) \times Q_{\infty}) \to \Gamma^* \text{ função que associa a cada transição em } \delta \text{ uma sequência de ações adaptativas primitivas. Pode-se associar uma sequência vazia a uma determinada transição para indicar que é uma transição não adaptativa.$

2.4 Linguagens Bidimensionais

Informalmente, conforme mostrado em (GIAMMARRESI; VENEZIA; RESTIVO, 1997), uma cadeia bidimensional é uma imagem, e é definida como um *array* retangular de símbolos pertencentes a um alfabeto finito.

Em (GIAMMARRESI; VENEZIA; RESTIVO, 1997) e (REGHIZZI; PRADELLA, 2005), os autores mostram como a formalização das linguagens convencionais podem ser transportadas para duas dimensões, de forma a manter a coerência, no caso, para que linguagens bidimensionais de tamanho (n,1) ou (1,m) tenham definições equivalentes com as linguagens de cadeias unidimensionais, e estendendo-as para duas dimensões de qualquer tamanho.

Especificamente, as linguagens tratadas, são as correspondentes ao tipo 3 da hierarquia de Chomsky, apresentada em 2.1.2, pois representa o tipo mais simples de linguagem, e formam a base para os tipos mais complexos.

¹Pelo mesmo motivo apresentado em (PISTORI; NETO, 2002), a operação de consulta (?) é omitida para simplificar a especificação, bem como a ordem em que uma ação adaptativa será executada, se antes ou depois da transição.

Nota-se também que este tipo de linguagem é reconhecida/gerada por dispositivos de estado finito.

2.4.1 Formalização das Linguagens Bidimensionais

Sendo Σ um alfabeto finito.

Definição 2. Uma cadeia bidimensional (ou imagem) sobre Σ é um array bidimensional retangular de elementos de Σ . O conjunto de todas as cadeias bidimensionais sobre Σ é denotada por Σ^{**} . Uma linguagem bidimensional sobre Σ é um subconjunto de Σ^{**} .

Com esta definição, uma imagem é entendida como uma cadeia bidimensional sobre Σ , porém, este trabalho necessita uma definição onde partes da imagem possa ser definida como símbolos que produzem a imagem original.

Consoante com a pesquisa, esta definição também é suprida por (GIAM-MARRESI; VENEZIA; RESTIVO, 1997), onde, para aplicar a noção de reconhecibilidade de uma imagem, símbolos bidimensionais de tamanho dois por dois são descritos como símbolos de um alfabeto Γ . Estes, por sua vez, podem ser projetados como ladrilhos (*tiles*) sobre a imagem formada sobre o conjunto Σ , encaixando suas propriedades locais compartilhadas, de forma a reconstituíla integralmente. Essa projeção é equivalente a transições em um autômato, segundo a referência citada.

Dada uma imagem $p \in \Sigma^{\star\star}$, chama-se $l_1(p)$ ao número de linhas de p e $l_2(p)$ ao número de colunas. O par ordenado $(l_1(p), l_2(p))$ corresponde ao tamanho da imagem p. A imagem vazia ϵ é a imagem com tamanho (0,0). Imagens com tamanho (0,n) ou (n,0) com n > 0 não foram definidas. O conjunto de todas as imagens sobre Σ com o tamanho (m,n), onde m,n > 0 é indicado por $\Sigma^{m\times n}$. Para referenciar um símbolo específico na cadeia bidimensional que forma uma imagem p, se $1 \le i \le l_1(p)$ e $1 \le j \le l_2(p)$, usa-se p(i, j) ou $p_{i,j}$ para o símbolo em p com coordenadas (i, j).

Para percorrer uma imagem e ter controle sobre seus limites horizontais e verticais, o caractere # é utilizado para rodear a imagem p, desta forma, a imagem de tamanho (m,n) passa a ter tamanho (m+2,n+2), e é denominada \hat{p} .

Dada uma imagem *p* de tamanho (*m*,*n*), sendo $h \le m, k \le n$: denota-se por $B_{h,k}(p)$ o conjunto de todos os blocos (ou sub-imagens de p do tamanho (*h*,*k*), chama-se de ladrilho uma imagem quadrada de tamanho (2,2).

Definição 3. Seja Γ um alfabeto finito. Uma linguagem $L \subseteq \Gamma^{\star\star}$ é local se existe um conjunto finito Φ de ladrilhos sobre o alfabeto $\Gamma \cup \{\#\}$ de forma que $L = \{p \in \Gamma^{\star\star} | B_{2,2}(\hat{p}) \subseteq \Phi\}$

2.5 Princípio do Tamanho Mínimo de Descrição

A MDL surgiu da ideia que toda regularidade nos dados pode ser utilizada para compactar estes dados, ou seja, descrevê-los utilizando menos símbolos que os necessários para descrever literalmente os dados. Paralelamente ao conceito de compactação, está o de entendimento, ou seja, se foi possível encontrar regularidades que permitam compactar os dados, foi necessário entender pelo menos as partes regulares.

Nesta busca por regularidades, pode-se utilizar modelos, ou hipóteses. A que melhor descreve os dados, ou a que possui maior regularidade é a que a faz no menor tamanho. O MDL funciona como ferramenta algorítmica para seleção da menor hipótese, ou melhor modelo que explica um determinado conjunto de dados.

Conforme pesquisado em (GRÜNWALD, 2007) e reforçado por (LI; VITANYI,

1997), sendo $\mathcal{H}^{(1)}, \mathcal{H}^{(2)}, ...,$ uma lista de modelos candidatos (*e.g.*, $\mathcal{H}^{(k)}$ é o conjunto de polinômios de grau *k*), cada um contendo um conjunto de hipóteses puntuais. A melhor hipótese puntual $H \in \mathcal{H}^{(1)} \cup \mathcal{H}^{(2)} \cup ...$ que explicam os dados *D* é aquele que minimiza a soma L(D) + L(D|H), onde L(H) é o tamanho da descrição da hipótese (em bits) e L(D|H) é o tamanho da descrição dos dados quando codificada com a ajuda da hipótese (também em bits).

Em (GRÜNWALD, 2007) também mostra a necessidade de se refinar as definições, de forma a torná-las mais práticas, não necessariamente mais simples. Para isso, aborda dois tipos de MDL. O descrito acima, denominado de MDL bruto, e o MDL refinado.

No MDL refinado, a definição de L(D|H), assumindo que H defina distribuições de probabilidade. Se H é um polinômio, pode-se transformá-lo em uma distribuição assumindo valores em Y dados por Y = H(X) + Z, onde Z é um termo de ruído com distribuição normal.

Definição de L(H): No MDL refinado, não se usa um código $H \in \mathcal{H}$ para codificar D, mas todo o modelo \mathcal{H} . Assim, um dado modelo \mathcal{H} é usado para codificar os dados, não em duas partes, como na definição do MDL bruto, mas com um único código de tamanho $\overline{L}(D|\mathcal{H})$. Significando que qualquer que seja o membro de \mathcal{H} que sirva bem aos dados, no sentido de L(D|H) ser pequeno, então o tamanho de código $\overline{L}(D|\mathcal{H})$ também será pequeno.

Outro conceito importante no MDL refinado, é a complexidade paramérica de \mathcal{H} , uma medida de riqueza do modelo, indicando sua habilidade de servir dados aleatórios e descrita por $COMP(\mathcal{H})$. Sendo o dado D, denota-se \hat{H} a distribuição em \mathcal{H} que maximiza a probabilidade e, consequentemente, minimiza o tamanho do código $L(D|\hat{H})$. Isso significa que a complexidade estocástica de D dado \mathcal{H} é $L(D|\hat{H}) + COMP(\mathcal{H})$.

2.6 Redes Bayesianas

Segundo (HECKERMAN, 1996), uma rede Bayesiana é um modelo gráfico de relações probabilísticas sobre um conjunto de variáveis. É amplamente utilizada como uma forma de adquirir conhecimento através de dados.

Como (MARQUES; DUTRA, 2003) mostrs, são importantes por conseguirem extrair informações mesmo com conjuntos incompletos de dados, no caso, através de aprendizado supervisionado. Permitem também aprendizado sobre relações causais, onde se analisa como um efeito pode ter sido influenciado ou não por uma causa.

Formalmente, supondo que $\bigcup_{n} A_n = \Omega$ é o espaço amostral com conjuntos disjuntos, então:

$$p(A_n|B) = \frac{p(B|A_n)p(A_n)}{\sum_n p(B|A_n)p(A_n)}$$

p(A|B) é a probabilidade de um evento A (hipótese) condicionada pela ocorrência de um evento B (evidência), sendo que geralmente se obtém o inverso (p(B|A)).

Uma rede Bayesiana cria uma distribuição de probabilidades organizando os diversos eventos através de grafos.

2.7 Fluídos Estáveis

Para a construção do simulador, foi adotado o princípio apresentado em (STAM, 1999), (STAM, 2003), discutido em detalhes em (HARRIS, 2005) e também em (AMADOR; GOMES, 2010), que mostram uma discretização da equação de Navier-Stokes, para fluídos incompressíveis, além de deixá-la estável para variações de tempo maiores. O comportamento do fluído, ou seu estado em um determinado momento é representado por um campo vetorial de velocidades. Esta velocidade é a entidade mais importante a ser representada, pois determina a forma que o fluído se movimenta e como outras entidades também são carregadas por este fluído.

$$\frac{\partial u}{\partial t} = -(u \cdot \nabla) u - \frac{1}{\rho} \nabla p + \nu \nabla^2 u + F$$
(2.1)

$$\nabla \cdot u = 0 \tag{2.2}$$

Neste caso, ρ é a densidade do fluido, ν é a viscosidade cinemática e $F = (f_x, f_y)$ representa forças externas que agem no fluido. A Equação 2.1 é dividida em duas equações, pois representa quantidades vetoriais, e cada uma representa sua projeção no eixo correspondente:

$$\frac{\partial u}{\partial t} = -(u \cdot \nabla)u - \frac{1}{\rho}\nabla p + v\nabla^2 u + f_x$$
$$\frac{\partial v}{\partial t} = -(u \cdot \nabla)v - \frac{1}{\rho}\nabla p + v\nabla^2 v + f_y$$

Deixando três incógnitas $(u, v \in p)$ e três equações.

A Equação 2.1 pode ser dividida em quatro termos, cada um representa uma característica do fluido. O primeiro termo, denominado divergente, $\nabla \cdot u$ representa a advecção, a capacidade que um fluido tem de transportar valores como densidade e outros objetos através de seu fluxo pelo campo vetorial.

O segundo termo, denominado gradiente, ∇_p representa a distribuição de pressão ao longo do campo vetorial, agindo como aceleração entre pontos de maior pressão para os de menor pressão.

O terceiro termo, denominado laplaciano, $\nabla^2 u$ representa a viscosidade do fluido, o quão resistente ele é em fluir, por esse motivo, resulta em difusão do momento e, consequentemente, da velocidade.

O quarto termo engloba qualquer aceleração consequente de forças que possam agir externamente ao fluido. Podem ser locais, aplicadas em regiões específicas do fluido ou globais, agindo no campo vetorial como um todo.

3 AQUISIÇÃO DE DADOS

3.1 As Imagens

Ao pesquisar pela base de imagens do CPTEC, observou-se que possui fontes distintas, recebendo imagens de vários satélites, dentre eles o GOES 10, 12 e 13, sete satélites diferentes da NOAA, os satélites Aqua e Terra da NASA, e o satélite Meteosat-9 da EUMETSAT. Dentre todos, o que fornece imagens com melhor qualidade e constância é o Meteosat-9, também denominado MSG2. Por esta razão, este foi escolhido como provedor das imagens que fundamentam este trabalho.

O Meteosat-9, em sua órbita, possui os seguintes parâmetros e característica (Fonte: (EUMETSAT, 2012)).

» MSG-2 Orbit State:

```
Epoch: 2012/04/30 00:00:00.000 ( 12.121.00.00.00.000 ) UTC
X-Position: X = -33244.1903628911 km,
X-Velocity: XDOT = 1.89122647844 km/s
Y-Position: Y = -25941.4614874085 km,
Y-Velocity: YDOT = -2.42378650209 km/s
Z-Position: Z = -80.24829260437 km,
Z-Velocity: ZDOT = -0.022489899 km/s
```

SRP coefficient: CR = 1.20449645748Semi-Major Axis: A = 42164.9577858 km Eccentricity: E = 0.00008647Inclination: I = 0.43308001281 deg Right Asc. of Asc. Node: RAAN = 23.3838831148 deg Argument of Perigee: AOP = 47.4004789958 deg True Anomaly: TA = 147.182045256 deg Longitude at epoch: LONG = -0.37417335 deg Mean Longitude: MEAN LONG = -0.378541 deg Mean Drift Rate: MEAN DRIFT = 0.00345 deg/day Perigee Height at Epoch: HPERI = 35783.17484304 km Apogee Height at Epoch: HAPO = 35790.46672856 km

» MSG-2 Attitude State:

Epoch: 2012/04/30 00:00:00.000 (12.121.00.00.00.000) UTC Spin Axis Right Asc.: RA = 301.526269383 deg Spin Axis Declination: DEC = 89.6502777959 deg Spin Axis Unit X: UX = 0.00319159 Spin Axis Unit Y: UY = -0.00520285 Spin Axis Unit Z: UZ = 0.99998137

» MSG-2 Physical State:

Epoch: 2012/04/30 04:00:00.000 (12.121.04.00.00.000) UTC
Spin Rate: SPIN = 99.8635038729 rpm
Total Mass: MTOT = 1135.32269222 kg
Fuel Mass: MFUEL = 30.44905722 kg

Oxidizer Mass: MOXID = 51.03863500 kg

Com este posicionamento em 0° de Longitude e 0° de Latitude e a 35.800Km de altitude, caracterizando uma órbita geoestacionária, a captação do globo fica limitado em -79° Oeste e 79° Leste de Longitude, e -81° Sul e 81° Sul de Latitude. O satélite está equipado com sensores que captam a radiação da Terra em 11 diferentes canais espectrais, que vão do infra-vermelho, centralizado em 13.4 μ m até o visível, centralizado em 0.6 μ m, possibilitando captação de imagens como a Figura 2. Há ainda um outro canal em alta resolução no visível, com uma distância focal maior, que aponta dinamicamente para as regiões mais iluminadas pelo sol no globo terrestre.

A imagem de interesse deste estudo é gerada por uma fatia correspondente à porção da América do Sul da imagem, como mostra a Figura 3. Nesta, é realizada uma projeção retangular, e o resultado é a imagem exposta na Figura 4

3.2 Os Filtros

Com a captação de uma amostra de duas mil imagens do satélite METE-OSAT 9, como na Figura 4 um conjunto de filtros foi escolhido e aplicado a todas elas para retirar informações à princípio desnecessárias, e para manter o mínimo de informações necessárias para a análise do comportamento atmosférico.

Primeiramente, as bordas territoriais foram extraídas (Figura 5). No domínio da frequência, as bordas de um pixel de espessura representam informações de frequência com valor elevado. Qualquer filtro passa-baixa (*lowpass*) é suficiente para removê-las total ou parcialmente, conforme cada caso.


Figura 2: Imagem Captada pelo Satélite MeteoSat-9 (MSG2) no Canal Visível 2 (0.8 μm)



Figura 3: Seção Correspondente à América do Sul

Neste, foi utilizado o filtro *median*¹ com raio 3.

Em seguida, tudo o que não é nuvem foi retirado, utilizando-se *threshold*² de cores, como colocado na Figura 6.

O próximo passo foi transformar a imagem em tons de cinza (8 bits) e utilizar novamente o *threshold* para eliminar os tons intermediários de cinza, deixando a imagem apenas em preto e branco (Imagem binária), de forma que a presença de nuvem é indicada pela cor branca, e sua ausência pela cor preta (Figura 7).

Esta imagem, por ter se tornado chapada, apresenta informações interessantes apenas nas bordas, e ocasionalmente, no meio de uma região monocromática devido ao aparecimento de uma nuvem no meio da região preta ou surgimento de um buraco no meio da região branca. Quem intensifica estas informações são os detectores de borda, onde apenas as bordas destes fenômenos serão mantidas. A Figura 8 mostra a aplicação do detector de bordas utilizando o algoritmo de Canny (CANNY, 1986) (HEATH et al., 1997), escolhido por ser o único que mantém as bordas com um pixel de espessura.

Sendo $\Sigma = \{0, 1\}$, 0 representando a cor preta e 1 a cor branca, chamou--se de \mathcal{L} à linguagem formada pelo conjunto de cadeias bidimensionais, $\mathcal{L} \subseteq \Sigma^{1150 \times 1300}$ adquiridas pelo processo descrito acima. Embora exista uma gramática $G(\mathcal{L})$, o conjunto *P* de regras de produção não é conhecido.

¹Filtro que calcula o valor do meio de um grupo de pixeis levando em conta o pixel analisado e seus vizinhos

²Filtro que elimina informações da imagem que estejam contidas acima ou abaixo de um valor limiar pré-definido



Figura 4: Imagem de Satélite. Copyright 2010-2012 EUMETSAT



Figura 5: Filtro Median: Retirada dos Limites Territoriais



Figura 6: Threshold de Cor - Retém Apenas Nuvens



Figura 7: Threshold de Tons de Cinza - Transformação para Imagem Binária



Figura 8: Detector de Bordas de Canny - Apenas Contorno das Nuvens

3.3 Os Símbolos

Com esta imagem contendo apenas as bordas das nuvens, foram extraídos pequenos padrões, *tiles* de tamanho 3x3 pixeis, de forma a criar uma linguagem \mathcal{T} . Sendo o alfabeto $\Gamma = \{0,1\}$, nos mesmos moldes que Σ , $\mathcal{T} = \{\forall \gamma \in \Gamma^{3 \times 3}\}$, ou seja, \mathcal{T} é formado por todas as cadeias de tamanho 3x3 contendo 0s e 1s. Dessa forma, realizou-se uma projeção $\pi : \mathcal{T} \to \mathcal{L}$ de modo que as cadeias de \mathcal{T} ao serem projetadas sobre cadeias individuais de \mathcal{L} pudessem ser compreendidas como símbolos equivalentes que também formam cadeias de \mathcal{L} .

É válido notar pela Figura 9, que a formação destes símbolos é igual à formação dos números binários, cujo bit de menor ordem se encontra na posição (x,y) = (0,0), e o crescimento se dá para a direita e para baixo, como mostra a Tabela 2.

Outra característica digna de nota é que um *tile* 3x3, escolhido para representar os símbolos do alfabeto, contém a definição de um item central e de seus vizinhos em todas as direções, como na rosa-dos-rumos, mantendo, de certa forma, informações de direção e rotação dos itens, bem como informações que podem limitar a possibilidade de caracteres que o seguem tanto na vertical como na horizontal. Também é suficientemente pequeno para manter uma quantidade administrável de possibilidades, pois o formato em *tile* 3x3 formam $2^9 = 512$ caracteres diferentes. O próximo *tile* com as mesmas características seria em 5x5, o que criaria $2^{25} = 33.554.432$ caracteres diferentes, o que não limitaria o trabalho, mas optou-se por utilizar as formas menores por serem as mais simples possíveis.

A Figura 9 mostra alguns exemplos destes caracteres.

Para facilitar a leitura de uma imagem preenchida com símbolos, auxili-

Tabela 2: Ordenação de Bits para Construção dos Símbolos

(0,0) = bit0	(0,1) = bit1	(0,2) = bit2
(1,0) = bit3	(1,1) = bit4	(1,2) = bit5
(2,0) = bit6	(2,1) = bit7	(2,2) = bit8



Figura 9: Os Primeiros Dezesseis Símbolos do Alfabeto

ando na busca por padrões, além de comprimir o tamanho da imagem em três vezes, cada caractere foi transformado em uma cor única indexada por um mapa de cores de 512 posições.

Na Figura 10a, temos um exemplo de uma distribuição típica de símbolos, onde cada cor representa o índice de um determinado símbolo.



(a) Caracteres em suas Posições. Cores Conforme Mapa de Cores. Detalhe da Ampliação de uma Região Arbitrária para Melhor Visualização



Figura 10: Caracteres Indexados pelo Mapa de Cores

4 ANÁLISE DOS DADOS

4.1 Contagem dos Símbolos e suas Transições

Com os símbolos definidos, e como para cada símbolo encontrado em uma imagem, existe uma transição para outro caractere na imagem seguinte, uma contagem de transições foi realizada. Esta contagem, por sua vez, mostra a quantidade de transições que os 512 símbolos transitam para 512 símbolos ao longo da amostra de duas mil imagens.

A contagem foi realizada, a princípio, com o intuito de saber quais transições acontecem, quais nunca ocorrem, quais símbolos mais aparecem e quais não existem, para, com isso, limitar mais ainda o alfabeto.

A Tabela 3 mostra as transições dos primeiros seis caracteres. Importante notar que a transição do símbolo 0 (■) para ele mesmo foi desprezada, pois como a maior parte da imagem é preta, ocorrem muito mais transições deste símbolo para ele mesmo, que qualquer outra.

	297000000	503828	15975	146820	519398	11207	
	496189	94110	614	21229	6804	704	•••
	16457	509	159	278	572	24	•••
	152133	18923	385	16194	2103	122	
	596568	6940	575	2273	97194	767	
	10723	871	15	160	730	367	•••
:		•		•	•	•	·

Tabela 3: Primeiras Linhas da Tabela de Contagem

Por se tratar de uma tabela muito densa, ela foi mapeada em uma imagem para mostrar, através de uma escala de cores que segue o espectro da luz visível, em escala logarítmica (Figura 10), como se distribui a contagem. Esta é exibida na Figura 11. Esta figura, ao ser gerada, surpreendeu pela sua forma, que numa análise mais detalhada, caracteriza-se por um padrão que se repete por toda a contagem. Um valor numérico alto que se dilui para a direita e para baixo na imagem como um todo, e em porções da mesma. Num determinado ponto, este padrão é sobreposto a si mesmo, além de se notar uma aproximação de um espelhamento na diagonal principal.

4.1.1 Análise da Contagem de Transições

Na Figura 11 percebe-se um padrão de crescimento de valores. Em aproximadamente três quartos de qualquer linha horizontal da imagem, há um vazio, uma diminuição de ocorrências de transições que volta a crescer no último quarto da imagem, embora não com a mesma intensidade que as apresentadas nos dois primeiros quartos.

Numa análise mais atenta, percebe-se que este padrão se repete em toda a imagem, não necessariamente com a mesma nitidez, apresentando um entrelaçamento entre padrões, mas ainda assim uma repetição. Tanto que ao se aproximar o primeiro quarto da imagem (128 x 128), o mesmo comportamento é observado. No espaço de um oitavo (32x32), o padrão está entrelaçado, mas em 8x8 e 4x4 o padrão é nítido.

No padrão 4x4 (Figura 16d), observou-se nitidamente o padrão principal no menor tamanho apresentado na imagem. Este padrão se observa também em 8x8 na Figura 16c, que é o 4x4 mais uma repetição deste com valores reduzidos e espelhados na horizontal, vertical e diagonal. O padrão em 16x16, podendo ser observado na primeira metade da Figura 16b possui uma identidade mais próxima do padrão original, unido ao 8x8 em seu início. O padrão seguinte, 32x32 (mesma Figura 16b) parece uma sobreposição de todos, pois não possui alguma característica que possa ser estabelecida como unidade de identificação. O mais próximo que se pode afirmar é que o padrão 16x16 é repetido na vertical e horizontal subtraído de algum outro padrão que não se conhece a origem para formar o padrão 32x32. A Figura 16a, recorte 128x128 da imagem original volta a apresentar o padrão detectado com diversas sobreposições.

A equação de diferenças que se aproxima desta distribuição é:

$$Z_0 = S \operatorname{eed}(4 \times 4)$$
$$Z_{n+1} = 2Z_n - \frac{3Z_{n-1}}{2} + c_n$$

4.2 Clusterização: Ordenação por Ocorrências

Também foi realizada a clusterização dos valores da Figura 11. Os valores foram classificados conforme a quantidade total de ocorrências na vertical e horizontal, onde se observa nitidamente o espelhamento na diagonal principal, observados na Figura 15. Notou-se que a ordem dos símbolos pela quantidade de transições contabilizadas foi a mesma tanto na horizontal como na vertical, embora houvesse divergência nos valores.

 Foi notado que os caracteres 170 ((2), 171 ((2), 174 ((2), 175 ((2), 234 ((2), 235 ((2), 238 ((2), 239 ((2), 341 ((2), 343 ((2), 349 ((2), 351 ((2), 373 ((2), 375 ((2), 381 ((2), 383 ((2), 426 ((2), 427 ((2), 430 ((2), 431 ((2), 443 ((2), 447 ((2), 455 ((2), 469 ((2), 471 ((2), 477 ((2), 479 ((2), 487 ((2), 490 ((2), 491 ((2), 493 ((2), 494 ((2), 495 ((2), 501 ((2), 503 ((2), 506 ((2), 509 ((2), 510 ((2), 511 ((2), 503 ocorreram em duas mil imagens, consequentemente, foram retirados do alfabeto.

Para verificar se esta imagem descreve as transições para este tipo de análise, e não apenas um acidente para este evento específico, uma nova coleta de imagens foi realizada, para um novo período, e uma nova contagem calculada, exibida na Figura 12. É facilmente notada a semelhança na distribuição de valores entre ambas, embora existam pequenas diferenças, apresentadas na Figura 13. Para testar também se este tipo de resultado pode ser obtido através de qualquer distribuição aleatória de caracteres, um novo experimento foi realizado. Duas mil imagens do tamanho exato das imagens coletadas do satélite foram geradas com os caracteres selecionados através de uma distribuição normal, seguindo a região positiva da gaussiana, onde os primeiros símbolos, especialmente o símbolo 0 foram mais selecionados, e os últimos, menos selecionados, como uma forma artificial de imitar o padrão observado. Porém como se trata de contagem de transições, o resultado da contagem apresentou forma muito mais regular. Este resultado pode ser observado na Figura 14 e o algoritmo utilizado para a criação desta distribuição pode ser observado abaixo.

```
Entrada: Imagem de tamanho w \times l e um fator de corte de seleção
           sliceFactor
Resultado: Distribuição Gaussiana de símbolos para preencher a
              imagem
v \leftarrow Conjunto de valores de tamanho w \times l;
max \leftarrow 0;
para cada e \in v faça
   n \leftarrow random.nextGaussian();
   se max < n então
       \max \leftarrow n;
   fim
    se n < sliceFactor então
       n \leftarrow 0;
   fim
   e \leftarrow n;
fim
para cada e \in v faça
   e \leftarrow \frac{e}{\max} \times 511.0;
fim
```

Neste código, *sliceFactor* é uma variável que define um limiar para a seleção do símbolo 0. Quanto maior este fator, maior a presença do símbolo 0
(■). Para o experimento apresentado, foi utilizado um *sliceFactor* de 1,5.

4.3 Busca de Padrões em Altura, Largura e Profundidade

Neste ponto, foi criada uma estrutura de dados como sendo o equivalente numérico à sequência de imagens, com a vantagem de armazenar a posição (x, y, z) de cada ocorrência de símbolo, tornando possível a contagem de ocorrências de uma determinada transição em posições específicas, sendo x e y as posições correspondentes da imagem, e z o índice da imagem na sequência. Desta forma, temos:

- 1. Tamanho horizontal = Quantidade de Caracteres na Horizontal;
- 2. Tamanho vertical = Quantidade de Caracteres na Vertical;
- 3. Profundidade = Quantidade de imagens analisadas

A Figura 17 representa graficamente esta estrutura de dados.

Realizou-se a busca de um determinado caractere em profundidade, para investigar se existe alguma relação entre a região que a imagem representa com a ocorrência de um determinado símbolo. Embora existam muitas ocorrências, não foi detectada regularidade, exceto, claramente, onde o filtro *median* não pôde retirar completamente as bordas territoriais, criando a ocorrência, por exemplo, do símbolo 176 (**F**) na posição (x,y) = (267,233) em todas as imagens da sequência nesta mesma posição. O mesmo acontece para o símbolo 51 (**L**) em (150,399).

A busca retorna com a seguinte formatação:

 (s, x, y, z_1, z_2)

Onde *s* é o índice do símbolo encontrado, *x* e *y* é a posição horizontal e vertical do símbolo, z_1 é o índice da imagem na sequência e z_2 é o índice

da imagem na sequência onde o símbolo também ocorreu. Um outro dado que também aparece no resultado é a quantidade de ocorrências, dado pela forma (s,n) onde s é o símbolo e n é a quantidade de ocorrências. Segue uma pequena amostra dos resultados desta busca:

(116,0)

- (7, 331, 0, 0, 47)
- (7,331,0,0,159)
- (7,331,0,0,171)
- (7,331,0,0,196)
- (7,331,0,0,872)
- (7,331,0,0,1791)
- (7,331,0,0,1799)
- (7,7)
- (64,332,0,0,76)
- (64,332,0,0,159)
- (64, 332, 0, 0, 196)
- (64,332,0,0,847)
- (64,332,0,0,910)
- (64,332,0,0,956)
- (64,332,0,0,1113)
- (64,332,0,0,1178)
- (64,8)
- (9,352,0,0,137)
- (9,352,0,0,1907)
- (9,2)
- (73, 353, 0, 0, 121)
- (73,353,0,0,1834)

(73,2)(146,354,0,0,899)(146,354,0,0,1319)(146,2)(148,355,0,0,900)(148,1)(256,356,0,0,120)(256,356,0,0,287)(256,356,0,0,288)(256,356,0,0,342)(256,356,0,0,1837)(256,356,0,0,1898)(256,6)

Uma análise que também foi realizada, em busca de ciclos de ocorrência de nuvens, foi o cálculo da área de cobertura de nuvens ao longo do tempo. O resultado pode ser visualizado na Figura 19. Esta análise deixa nítido o fato de existir um ciclo de 21 dias ao longo das duas mil imagens observadas. Este ciclo também pode ser observado, porém sem a mesma nitidez, na Figura 18. Com a Figura 19 pôde se verificar uma falha na coleta de dados, representadas pelos picos circulados em vermelho. Esta falha localizou-se na imagem de índice 941, exposta na Figura 20. Este dado demonstra a boa qualidade do serviço do satélite, se comparado aos outros, que até pelo fato de não possuí-rem órbita geoestacionária, não podem produzir imagens com a regularidade necessária, pois parte do território analisado se encontra oculto no outro lado do globo terrestre.

Um outro aspecto do trabalho é, com as análises das imagens de satélite, dadas as formas das nuvens, se chegar aos valores de temperatura, pressão e velocidade do vento que originaram seu comportamento. Porém, isto é muito difícil sem os dados das várias estações meteorológicas sincronizados com a imagem analisada, e com resolução adequada para saber se a predição foi correta, ou o quanto de erro ela cometeu.

Este problema foi solucionado com a criação de um simulador, que gera uma atmosfera artificial, cujos resultados são exibidos como imagens com as características aproximadas das imagens de satélite originais. Consequentemente, os mesmos processos de tratamento da imagem e extração dos padrões são realizados como se estas imagens geradas pelo simulador fossem as imagens de satélite, com a vantagem de se conhecer a pressão, temperatura e velocidade de todos os pontos simulados, tendo uma grande base de dados para comparar os resultados do modelo de predição.

4.4 Comentários sobre Gramática

Foi realizado uma busca de regularidades no espaço de busca, mas não foi encontrado uma forma como as transições apresentadas pela Figura 21

4.4.1 Contextualização

As linguagens bidimensionais são abordadas de duas formas neste trabalho. São utilizadas para pesquisar regularidades em uma única imagem, e são utilizadas para pesquisar regularidades em profundidade.

Na busca por semelhanças em uma única imagem, percebeu-se que alguns caracteres aparecem regularmente próximos a outros. Como é o caso do caractere ■ que muitas vezes aparece do lado esquerdo do caractere ■. Isto se dá pelo fato dos vizinhos se encontrarem pela existência de dado nas bordas do caractere. Se um caractere possui pixeis ligados nas bordas, há uma chance muito maior de que seus oito símbolos vizinhos estejam ligados por estes pixeis, afinal, estes se originaram pelos contornos das nuvens.

A implicação disso é que pode existir um subconjunto das regras de produção que podem agrupar estes casos, algo como { $S \rightarrow A, A \rightarrow \blacksquare B, B \rightarrow \blacksquare$ }, criando a sequência $S \Rightarrow_G \blacksquare B \Rightarrow_G \blacksquare \Box$ dentre inúmeras outras possibilidades.

4.5 Busca por Equivalências em Profundidade

Para auxiliar na criação de regras gramaticais, foi necessário escolher um método que realize busca de equivalências dos símbolos nas imagens, sendo que sua forma pode ser alterada pela posição, pela rotação ou mesmo esticamento. As formas nas imagens de satélite se movimentam continuamente, a velocidades variáveis, de modo que um determinado caractere muito provavelmente não será encontrado na mesma posição na próxima imagem, dessa forma, foi definido um espaço de busca de 9x9 pixeis na próxima imagem para encontrar possíveis equivalências do caractere apontado.

Foi adotada a inferência através de redes Bayesianas para realizar este trabalho. Ela busca dentro do contexto associado a um caractere (*e.g.*caracteres vizinhos) uma relação de equivalência na região 9x9 da imagem seguinte.

A relação de equivalência foi adotada como sendo qualquer símbolo que possui diferença de no máximo um pixel do caractere procurado, podendo estar em qualquer divisão 3x3 do espaço 9x9 de busca. Como observado na Figura 22, o símbolo 64 (**T**) ao ser procurado na próxima imagem da sequência, pode ter transitado para o símbolo 192 (**T**) tanto no canto superior esquerdo como no canto superior direito. Esta transição é contada pois o símbolo 192 é o simbolo 64 acrescido de mais um pixel, enquadrando-se no critério de erro.

Assim também ocorre com o símbolo 0 (**II**), encontrado na posição superior central, que é o 64 com seu único pixel apagado, apresentando erro de um também.

Na região central do espaço de busca, ainda na Figura 22, o símbolo 64 é encontrado sem modificação (erro zero) e sem deslocamento. Este também é contabilizado.



Figura 22: Busca Utilizada na Rede Bayesiana

Para uma amostra de trinta imagens, as probabilidades obtidas são mostradas na Tabela 4. A Tabela 5 mostra as mesmas variáveis, mas para uma verificação de trezentas imagens.

	•	0,0114	0,0057	0,0133	0,0086	0,0114	0,0057	0,0038	0,0048	0,0048
ins)	F	0,2005	0,1197	0,0589	0,1254	0,3772	0,2166	0,0827	0,1112	0,2394
(30 image	۲	0,0010	0	0	0,0010	0,0010	0	0	0	0
nbolo 64 (P	0	0,0010	0	0	0,0010	0	0	0	0
es ao Sín	P	0,3601	0,2736	0,0513	0,1406	0,5178	0,0865	0,1026	0,3098	0,1767
quivalent	2	0,0010	0	0,0010	0,0029	0	0,0019	0,0019	0,0019	0
mbolos E	7	0,0010	0	0	0,0010	0,0010	0	0	0,0010	0,0029
ia dos Sí	-	0,0019	0,0057	0,0124	0,0019	0,0095	0,0029	0,0019	0,0048	0,0067
Ocorrênc		0,0067	0,0076	0,0067	0,0086	0,0114	0,0133	0,0076	0,0095	0,0086
idade de		1,0575	0,4067	0,2062	0,6433	1,6314	0,3687	0,3288	0,8694	0,6708
: Probabil		0,0048	0,0019	0,0010	0,0019	0,0086	0,0038	0,0019	0,0076	0,0038
Tabela 4		0,0095	0,0143	0,0057	0,0057	0,0067	0,0067	0,0057	0,0067	0,0029
		9.8770	6,4668	11,0095	12,8605	7,7163	7,7885	13,6890	11,5655	8,9771
		0	-	2	ო	4	Ŋ	9	2	ω

T

ens
ag
<u></u>
80
じ +
ð
8
dr
ŝ
go
ŝ
nte
ale
.≧
<u> </u>
ы
90
qu
SíT
S
ğ
<u>Cia</u>
ên
or
ő
é
<u>e</u>
lad
ili
ab
<u>o</u>
٦
4
ela
ap
Ĕ

	₽	0,0106	0,0056	0,0106	0,0077	0,0081	0,0050	0,0037	0,0048	0,0039
		0,1914	0,1307	0,0531	0,1172	0,3382	0,2619	0,0911	0,1302	0,2185
)	¥	0,0007	0,0008	0,0007	0,0010	0,0012	0,0006	0,0002	0,0007	0,0006
	P	0,0002	0,0004	0,0004	0,0004	0,0006	0,0003	0,0001	0,0003	0,0006
	P	0,3106	0,3330	0,0599	0,1551	0,4849	0,1049	0,0977	0,3008	0,2134
-	•	0,0019	0,0009	0,0019	0,0010	0,0015	0,0015	0,0006	0,0009	0,0010
	3	0,0006	0,0004	0,0002	0,0005	0,0015	0,0006	0,0003	0,0010	0,0012
	-	0,0022	0,0039	0,0105	0,0030	0,0097	0,0058	0,0030	0,0064	0,0080
		0,0069	0,0061	0,0066	0,0041	0,0092	0,0136	0,0056	0,0081	0,0093
		0,9521	0,5258	0,1822	0,5556	1,5869	0,4402	0,3095	0,7825	0,7108
		0,0047	0,0046	0,0038	0,0048	0,0049	0,0040	0,0030	0,0045	0,0046
	R	0,0122	0,0114	0,0061	0,0076	0,0074	0,0065	0,0055	0,0069	0,0057
		10,2270	6,3964	10,3617	12,8908	8,2102	7,4739	13,8075	11,6604	9,0169
		0	-	2	ო	4	വ	ဖ	2	∞

Tabela 5: Probabilidade de Ocorrência dos Símbolos Equivalentes ao Símbolo 64 (300 imagens)

									ļ			1			H				
1 65					i				H										
																	1.4. V		rd
																		1	
									Į.				per per						1 = -,
				1111 1415 1415 1415 1415 1415						*					1. A				i konte 1 konte 1 f. dage
		*												linder Linder Linder				1-h-h	
		**** •***									7								
		rer -									€(** ₽€);					263			
														la y Pak	5 X X X X		44 - 24 - 2 44 - 24 - 2	-	
EN LOU	19 A														line and				i e
	1926° 18																		
			șincă.	-citie regis			÷												1112 - 1 1

Figura 11: Representação da Densidade da Contagem Realizada em Novembro de 2011

11.1 COLUMN TODAY 1 CT. 81.1 C.				
	enter bei bei den Beller bei	n Britten fan Line, mar 2 Britten jack fillen Ang 10 Britten ter blan Ang		n - Constanting (n. 1997) (n. 1997) 1 Henri Hallenbelle, Mitterstein (n. 1997) (n. 1997) 19 Henri Hallenbelle, Mitterstein (n. 1997) (n. 1997)
	A PERSON AND A PROPERTY A			
		WER	TERM	e d h ren the s
	sabalanda albanja - porsi reference albanja - porsi	n þeinga staða feldi í staði Staðarísku skilda staði		
	· · · · · · · · · · · · · · · · · · ·			
ANS COLUMN TO PERSON	the state of the			
Aller and the second				A set of the set of

Figura 12: Representação da Densidade da Contagem para Outro Conjunto de Dados Realizada em Abril de 2012



Figura 13: Total de Diferença das Contagens Realizadas em Períodos Distintos



Figura 14: Representação da Densidade da Contagem para Conjunto Artificial de Dados Utilizando Distribuição Normal dos Símbolos



Figura 15: Clusterização da Contagem de Transições Representada pela Figura 11 Ordenada pela Soma das Ocorrências Verticais e Horizontais



Figura 16: Partes da Contagem de Transições com Repetições do Padrão Encontrado



Figura 17: Estrutura de Dados para Busca de Padrões



Figura 18: Expansão das Transições no Tempo



Figura 19: Cobertura de Nuvens no Tempo. Nuvens em Azul e Áreas Descobertas em Vermelho



Figura 20: Falha na Emissão de Dados em uma Sub-Região da Imagem



Figura 21: Sequência Regular de Transição

5 O SIMULADOR ATMOSFÉRICO

5.1 Descrição

Para auxiliar na busca de regularidades nas imagens de satélite, no intuito de criar um mapeamento das formas representadas pelos símbolos para as forças que os geraram, o desenvolvimento de um ambiente controlado, sob a forma de um simulador de dinâmica dos fluidos simplificado foi iniciado.

Neste, todas as movimentações geradas por forças e variáveis são calculadas, como temperatura, pressão e velocidade do fluido em todos os pontos da região estudada, utilizando a proposta de (STAM, 1999) e (STAM, 2003) e implementada em (AMADOR; GOMES, 2010) entre outros para gerar imagens semelhantes às fornecidas pelo satélite Meteosat-9.

O equivalente em termos reais ao simulador seria o território representado pelas imagens do satélite, mas com sensores atmosféricos distribuídos regularmente a cada cinco quilômetros quadrados, incluindo o oceano.

5.2 Mapa de Elevação

Primeiramente, buscou-se nos diversos órgãos internacionais que realizam mapeamento de superfície do planeta, quais modelos de elevação seriam mais adequados para serem implementados. Dentre todos, o que pareceu mais completo é o modelo ETOPO1 da *NOAA - National Oceanic and* *Atmospheric Administration*. Este mapeamento cobre 1 arco-minuto de latitude. Se representado como imagem, exibe um mapeamento em torno de 1,87Km por pixel numa imagem de 21600 por 10800 (NOAA, 2012).

Como a imagem original retirada do satélite está em torno de 5Km por pixel, há espaço suficiente para realizar adaptações futuramente, aumentando a resolução do simulador conforme a necessidade.

Este mapa é apresentado na Figura 23, que o exibe em sua forma original (23a) e o mapa de elevação aplicado a uma superfície tridimensional (23b).



(a) Elevação Tons de Cinza



(b) Elevação em Superfície

Figura 23: Mapa de Elevação global: Modelo ETOPO1


Figura 24: Comparação Visual Entre a Região Simulada em Destaque e o Resto do Relevo Terrestre

Como a região de interesse é apenas uma fração do mapa global, como mostrado na Figura 24, apenas a área que permeia a região de interesse é extraída e aplicada como elemento topográfico no simulador.

O resultado atual pode ser observado na Figura 25



Figura 25: Estado Atual do Simulador

6 CONSIDERAÇÕES FINAIS

6.1 Contribuições

Com a aplicação das técnicas aqui expostas, pôde-se investigar vários aspectos que uma análise de uma sequência de imagens de satélite proporcionou. A análise das transições permitiu criar formas únicas que a caracterizam, pode-se encontrar pequenas regiões de regularidade e criou-se um modelo para análise de imagens de satélite baseado em linguagens formais.

Embora ainda não se tenha definido uma gramática que explique as transições de símbolos na sequência de imagens, conseguiu-se diminuir a quantidade de símbolos de 512 para 473, consequentemente, diminuindo a quantidade de regras que esta gramática pode ter.

Com a ordenação das ocorrências, denominada clusterização, mostrada pela Figura 15, foi possível agrupar símbolos com propriedades compartilhadas em classes de equivalências, de forma que qualquer símbolo que representa sua classe pode ser substituído por qualquer símbolo desta classe, diminuindo mais ainda a quantidade de regras gramaticais.

No presente trabalho, criou-se uma base de dados que pode servir de palco para aplicações dos algoritmos de busca desenvolvidos em (CHAER, 2010) e (GUIRALDELLI, 2012), bem como utilizar a formalização apresentada em (INOJOSA, 2012).

O mapa de contagem de transições mostrou-se útil como característica de movimento de ambientes dinâmicos, e sua análise permitiu aplicar regressão bidimensional para encontrar a melhor superfície que descreve estas ocorrências, funcionando como critério de acuidade para o modelo.

O simulador criado permite vários estudos que envolvam dinâmica dos fluidos, pois, além de exibir o comportamento da atmosfera, armazena todas as variáveis calculadas nos instantes de tempo desejados.

6.2 Conclusões

O método para criação de modelos computacionais de análise de imagens de satélite e previsão do tempo foi estruturado, alguns modelos foram gerados e confrontados com dados observados. A análise de acuidade foi realizada utilizando-se da diferença entre os dados obtidos experimentalmente com os observados. Para este teste, a contagem de transições serviu como exemplo. Também foi observada a diferença da contagem de transições coletadas entre períodos distintos de tempo.

O simulador foi desenvolvido utilizando-se critérios arbitrários para formação de nuvens, a ageração de imagens em paralelo com os valores calculados permitiu a realização de experimentos sob condições controladas, para serem comparadas com os dados coletados a partir das imagens de satélite.

Embora a linguagem não tenha sido ainda definida, passos importantes foram dados na direção da redução da complexidade de suas regras gramaticais, ao diminuir o número total de símbolos e a criação de classes de equivalências entre símbolos.

6.3 Trabalhos Futuros

Durante a construção do simulador, notou-se que o surgimento de nuvens não é uma questão simples, ela depende de um contexto de variáveis para ocorrer, consequentemente, abre a possibilidade para a busca de símbolos utilizando outros canais que demonstrem com maior nitidez a movimentação dos fluidos, como é o caso do canal de vapor de água, na faixa de $6,2\mu m$. Com isso, todos os experimentos podem ser refeitos utilizando estas novas imagens.

Outra tarefa importante é realizar o mesmo experimento com uma quantidade maior de imagens. Como as duas mil amostras de imagens compreendem um período de aproximadamente vinte e um dias, coletar uma quantidade equivalente a dois e três anos, setenta mil e cento e cinco mil imagens respectivamente. Traçar paralelo entre o mesmo período com imagens geradas pelo simulador.

Com o banco de dados de profundidade, é necessário criar um *software* que facilite a manipulação e a busca destes dados, de forma a permitir análises mais detalhadas das regularidades encontradas, tanto em altura/largura como em profundidade, bem como exibir graficamente os resultados obtidos, além de permitir a inserção modular de algoritmos de busca e aprendizado.

REFERÊNCIAS

AMADOR, G.; GOMES, A. A cuda-based implementation of stable fluids in 3d with internal and moving boundaries. In: *Proceedings of the 2010 International Conference on Computational Science and Its Applications*. Washington, DC, USA: IEEE Computer Society, 2010. (ICCSA '10), p. 118–128. ISBN 978-0-7695-3999-7. Disponível em: http://dx.doi.org/10.1109/ICCSA.2010.43>.

BAADER, F.; NIPKOW, T. Term rewriting and all that. In: _____. [S.I.]: Cambridge University Press, 1999. cap. 1, p. 316.

BOOK, R. V.; OTTO, F. String-rewriting systems. LNCS, Springer, 1993.

BRAMS. *BRAMS - Brazilian developments on the Regional Atmospheric Modelling System*. São Paulo, Agosto 2010. Acesso em 19 de agosto de 2010. Disponível em: http://stams.cptec.inpe.br/stams

CANNY, J. A computational approach to edge detection. *IEEE Transactions* on *Pattern Analysis and Machine Intelligence*, v. 8, p. 679 – 698, 1986.

CHAER, I. *Um estudo sobre a Teoria da Predição aplicada à análise semântica de Linguagens Naturais*. Dissertação (Mestrado) — Escola Politécnica da Universidade de São Paulo, 2010.

CHOMSKY, N. Three models for the description of language. *Information Theory, IRE Transactions on*, v. 2, n. 3, p. 113–124, september 1956. ISSN 0096-1000.

(CPTEC-INPE), C. de Previsão do Tempo e E. C. *CPTEC website*. Junho 2010. Www.cptec.inpe.br. Acesso em 8 de junho de 2010.

DERSHOWITZ, N.; JOUANNAUD, J.-P. *Rewrite Systems*. [S.I.]: Elsevier and MIT Press, 1990.

ENSEMBLE. *Modelo Ensemble*. São Paulo, Setembro 2010. Explica como o modelo ensemble busca reduzir o erro nas condições iniciais da simulação, gerados por erros dos diversos sensores, utilizando o próprio erro para contornar o problema. Disponível em: http://previsaonumerica.cptec.inpe.br/mod_ens.shtml.

ETA. *Modelo ETA utilizado pelo CPTEC/INPE*. São Paulo, Setembro 2010. Disponível em: br/>.

EUMETSAT. *Meteosat Orbital Parameters*. Europe, Abril 2012. Disponível em: http://www.eumetsat.int/Home/Main/Satellites/orbital_parameters>.

GIAMMARRESI, D.; VENEZIA, F.; RESTIVO, A. Handbook of formal languages - vol 3. In: _____. [S.I.]: Springer-Verlag New York, Inc., 1997. cap. Two Dimensional Languages, p. 215 – 267.

GLOBAL. *Modelo Global*. São Paulo, Setembro 2010. Disponível em: http://previsaonumerica.cptec.inpe.br/modglb.shtml.

GRÜNWALD, P. *The Minimum Description Length Principle*. Mit Press, 2007. (Adaptive Computation and Machine Learning). ISBN 9780262072816. Disponível em: http://books.google.com.br/books?id=mbU6T7oUrBgC.

GUIRALDELLI, R. H. G. *Algorithmic Graph Unification and Simulations for Haplotype Networks*. Dissertação (Mestrado) — Escola Politécnica da Universiadade de São Paulo, 2012.

HARRIS, M. Fast fluid dynamics simulation on the gpu. In: *ACM SIGGRAPH 2005 Courses*. New York, NY, USA: ACM, 2005. (SIGGRAPH '05). Disponível em: http://doi.acm.org/10.1145/1198555.1198790>.

HEATH, M. et al. A robust visual method for assessing the relative performance of edge-detection algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 19, p. 1338–1359, 1997.

HECKERMAN, D. A Tutorial on Learning With Bayesian Networks. [S.I.], 1996.

HOPCROFT, J. E.; MOTWANI, R.; ULLMAN, J. D. *Finite Automata*. [S.I.]: Pearson Education and Addison-Wesley, 2001.

INOJOSA, R. *Uma Nova Formulação Algébrica para o Autômato Finito Adaotativo de Segunda Ordem Aplicada a um Modelo de Inferência Indutiva*. Tese (Doutorado) — Escola Politécnica da Universidade de São Paulo, 2012.

LI, M.; VITANYI, P. An introduction to Kolmogorov Complexity and its Applications: Preface to the First Edition. 1997.

LORENZ, E. N. A study of the predictability of a 28-variable atmospheric model. *Tellus*, v. 17, p. 321333, 1965. Doi: 10.1111/j.2153-3490.1965.tb01424.x.

MARQUES, R. L.; DUTRA, I. Redes bayesianas: o que são, para que servem, algoritmos e exemplos de aplicações. Curso de Inês Dutra. 2003.

NETO, J. J. *Contribuição à Metodologia de Construção de Compiladores*. Tese (Doutorado) — Universidade de São Paulo, 1993.

NETO, J. J. Adaptive rule-driven devices - general formulation and case study. *Lecture Notes in Computer Science*, v. 2494/2002, p. 466–470, 2002.

NETO, J. J.; BRAVO, C. Adaptive automata - a reduced complexity proposal. *Lecture Notes in Computer Science*, v. 2608/2003, p. 1–46, 2003.

NOAA. *ETOPO1 - Global Relief Model*. São Paulo, Fevereiro 2012. Acesso em 27 de fevereiro de 2012. Disponível em: http://www.ngdc.noaa.gov/mgg/global/global.html.

PAPADIMITRIOU, C. H.; LEWIS, H. R. *Elements of the Theory of Computation*. [S.I.]: Prentice-Hall, 1998.

PISTORI, H.; NETO, J. J. Adaptree - proposta de um algoritmo para indução de Árvores de decisão baseado em técnicas adaptativas. *Anais Conferência Latino Americana de Informática*, Novembro/2002, p. 1–10, 2002.

RAMOS, M. V. M.; NETO, J. J.; VEGA Ítalo S. *Linguagens Formais: Teoria, Modelagem e Simulação*. Porto Alegre: Bookman, 2009.

REGHIZZI, S. C.; PRADELLA, M. Tile rewriting grammars and picture languages. *Theoretical Computer Science*, v. 340, p. 2005, 2005.

ROSEN, K. H. Computational Modeling. 6a. ed. [S.I.]: McGraw-Hill, 2007.

STAM, J. Stable fluids. In: *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*. New York, NY, USA: ACM Press/Addison-Wesley Publishing Co., 1999. (SIGGRAPH '99), p. 121–128. ISBN 0-201-48560-5. Disponível em: http://dx.doi.org/10.1145/311535.311548>.

_____. Flows on surfaces of arbitrary topology. *ACM Trans. Graph.*, ACM, New York, NY, USA, v. 22, n. 3, p. 724–731, jul. 2003. ISSN 0730-0301. Disponível em: http://doi.acm.org/10.1145/882262.882338>.

TOLMAN, H. L. The numerical model wavewatch: a third generation model for the hindcasting of wind waves on tides in shelf seas. *Communications on Hydraulic and Geotechnical Engineering, Delft Univ. of Techn., ISSN 0169-6548*, v. 89-2, p. 72 pp., 1989.

WWATCH. *Modelo de Ondas WWATCH III 2.22 no CPTEC/INPE*. São Paulo, Setembro 2010. Disponível em: http://previsaonumerica.cptec.inpe.br/mod_wwatch.shtml.