



**Marco Antônio da Cunha Ferreira**

**GPFIS-Forecast: Um sistema fuzzy-genético  
baseado em programação genética multigênica  
para problemas de previsão univariada**

**Dissertação de Mestrado**

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre pelo Programa de Pós-graduação em Engenharia Elétrica do Departamento de Engenharia Elétrica da PUC-Rio

Orientador : Prof. Ricardo Tanscheit  
Co-Orientador: Prof. Marley Maria Bernardes Rebuszi Vellasco

Rio de Janeiro  
Setembro de 2015



**Marco Antônio da Cunha Ferreira**

**GPFIS-Forecast: Um sistema fuzzy-genético  
baseado em programação genética multigênica  
para problemas de previsão univariada**

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre pelo Programa de Pós-graduação em Engenharia Elétrica do Departamento de Engenharia Elétrica do Centro Técnico Científico da PUC-Rio. Aprovada pela comissão examinadora abaixo assinada.

**Prof. Ricardo Tanscheit**

Orientador

Departamento de Engenharia Elétrica — PUC-Rio

**Prof. Marley Maria Bernardes Rebuzzi Vellasco**

Co-Orientador

Departamento de Engenharia Elétrica — PUC-Rio

**Prof. Jorge Luís Machado do Amaral**

Faculdade de Engenharia – UERJ

**Prof. Douglas Mota Dias**

Departamento de Engenharia Elétrica – PUC-Rio

**Prof. Karla Tereza Figueiredo Leite**

Departamento de Engenharia Elétrica – PUC-Rio

**Prof. José Eugenio Leal**

Coordenador do Centro Técnico Científico — PUC-Rio

Rio de Janeiro, 15 de Setembro de 2015

Todos os direitos reservados. Proibida a reprodução total ou parcial do trabalho sem autorização da universidade, do autor e do orientador.

### **Marco Antônio da Cunha Ferreira**

Graduou-se em Engenharia de Produção pela Universidade Federal Fluminense. Durante a graduação trabalhou principalmente com Estatística Aplicada, Administração Financeira e Métodos de Apoio à Decisão para Análise de Impacto Regulatório. No mestrado envolveu-se com a área de Métodos de Apoio à Decisão, em destaque a Inteligência Computacional. Sua área de interesse é métodos quantitativos de apoio à decisão.

#### Ficha Catalográfica

Ferreira, Marco Antônio da Cunha

GPFIS-Forecast: Um sistema fuzzy-genético baseado em programação genética multigênica para problemas de previsão univariada / Marco Antônio da Cunha Ferreira; orientador: Ricardo Tanscheit; co-orientador: Marley Maria Bernardes Rebuszi Vellasco. — Rio de Janeiro : PUC-Rio, Departamento de Engenharia Elétrica, 2015.

v., 55 f: il. ; 29,7 cm

1. Dissertação (Mestrado em Engenharia Elétrica) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Engenharia Elétrica.

Inclui referências bibliográficas.

1. Engenharia Elétrica – Tese. 2. Sistemas Fuzzy-Genéticos. 3. Programação Genética. 4. Previsão. I. Tanscheit, Ricardo. II. Vellasco, Marley Maria Bernardes Rebuszi Vellasco. III. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Engenharia Elétrica. IV. Título.

CDD: 621.3

## **Agradecimentos**

Ao CNPq e a PUC-Rio pelos auxílios e estrutura concedidos, sem os quais este trabalho não poderia ter sido realizado.

A meus orientadores, que me ajudaram a tornar minhas ideias em realidade.

Ao Adriano Koshyiama por ter a paciência e interesse na minha colaboração para aprimorar o modelo muito bacana desenvolvido por ele.

Aos professores Alex e Ana Paula por acreditarem em meu trabalho.

A minha mãe, a meus pais e a minhas irmãs por me apoiarem nesta e em tantas outras jornadas.

Aos poucos, mas verdadeiros, amigos pelos quais compartilho experiências e pensamentos.

## Resumo

Ferreira, Marco Antônio da Cunha; Tanscheit, Ricardo; Vellasco, Marley Maria Bernardes Rebuzzi Vellasco. **GPFIS-Forecast: Um sistema fuzzy-genético baseado em programação genética multigênica para problemas de previsão univariada**. Rio de Janeiro, 2015. 55p. Dissertação de Mestrado — Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

Métodos de previsão são muito importantes para o desenvolvimento de diversas atividades no cotidiano produtivo de nossa sociedade. Vários modelos estatísticos são desenvolvidos até hoje, contudo possuem muitos pressupostos que devem ser seguidos com o intuito de se obter uma resposta aceitável. Modelos não estatísticos para prever séries temporais como os que envolvem Sistemas de Inferência Fuzzy (SIFs) fornecem uma descrição do processo de previsão por meio de regras linguísticas. Explora-se, nesta dissertação, o GPFIS-Forecast: versão do GPFIS – Sistema de Inferência Fuzzy Genético baseado em Programação Genética Multigênica – para previsão de série temporais univariadas. O modelo apresenta, em sua execução, quatro etapas básicas: Fuzzificação, Inferência, Defuzzificação e Avaliação. Em cada uma destas etapas, pode-se fazer uso de diferentes configurações, com implicações evidentes nos resultados finais. Este trabalho propõe o aperfeiçoamento do GPFIS-Forecast em duas vertentes principais: (i) aumentar a quantidade de possibilidades de configurações, avaliando se podem contribuir significativamente para a acurácia das previsões; (ii) adicionar informações complementares como alternativas para a interpretação do resultado do modelo, tendo como compromisso tanto a acurácia e quanto a interpretabilidade. Os estudos de caso demonstram que, em casos de séries temporais com pouca tendência, o GPFIS-Forecast apresenta uma acurácia entre as 10 melhores da competição NN3; quando há forte tendência, faz-se necessário o uso de pré-processamento, prejudicando a interpretabilidade do resultado. Os Limites de Previsão Fuzzy introduzidos neste trabalho agregam mais informação ao resultado da previsão pontual, apontando possíveis ajustes finais nas bases de regras de modelos com maior granularidade.

## Palavras-chave

Sistemas Fuzzy-Genéticos ; Programação Genética ; Previsão.

## Abstract

Ferreira, Marco Antônio da Cunha; Tanscheit, Ricardo (Advisor); Vellasco, Marley Maria Bernardes Rebuzzi Vellasco (Co-Advisor). **GPFIS-Forecast: A Genetic-Fuzzy system based on genetic programming for forecast problems.** Rio de Janeiro, 2015. 55p. MSc. Dissertation — Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

Forecasting methods are very important for the development of various activities in everyday society. Several statistical models have been developed, but many assumptions must be made in order to obtain an acceptable response. Nonstatistical models for time series forecasting such as those involving systems Fuzzy Inference Systems (FIS) provide a description of the process through linguistic rules. This dissertation delves into GPFIS-Forecast: a version of GPFIS – Fuzzy Inference System based on Multigene Genetic Programming – for univariate time series forecasting. This model consists of four basic stages: Fuzzification, Inference, Defuzzification and Evaluation. In each of these steps, different configurations will have distinct impacts on the results. This work proposes the improvement of GPFIS-Forecast along two main lines (i) increase the amount of possible configurations and assess their contribution to a better forecasting accuracy and (ii) add further information to the interpretation of results, keeping in mind both accuracy and interpretability. The case studies show that in the case of time series with small tendency, GPFIS-Forecast provides a good accuracy; when tendency is larger and pre-processing becomes necessary, interpretability is affected. The Fuzzy Forecasting Limits introduced here add more information to the result, pointing to possible adjustments to rule bases of models with greater granularity.

## Keywords

Genetic-Fuzzy Systems ; Genetic Programming ; Forecasting.

## Sumário

1	Introdução	<b>10</b>
1.1	Motivação	11
1.2	Objetivos	12
1.3	Contribuições	13
1.4	Descrição e Organização da Dissertação	13
2	Sistema Fuzzy Genético (SFG)	<b>14</b>
2.1	Meta-heurística Genética	14
2.2	Sistema Fuzzy Genético Baseado em Regras	16
2.3	Previsor Fuzzy Para Séries Temporais	18
3	Modelo GPFIS - Forecast	<b>21</b>
3.1	Visão Geral	21
3.2	Configuração do GPFIS-Forecast	26
4	Experimentos e Resultados	<b>37</b>
4.1	Delineamento Experimental	37
4.2	Experimento Sequencial	37
4.3	Teste de Friedman	38
4.4	Estudo de casos	39
5	Conclusões e Trabalhos Futuros	<b>51</b>
6	Referências Bibliográficas	<b>52</b>

## Lista de figuras

2.1	Exemplo de um indivíduo cuja estrutura é em árvore.	15
2.2	Exemplo de um indivíduo multigênico.	16
2.3	Operação de mutação em um indivíduo multigênico.	16
2.4	Operadores: Cruzamento de baixo nível.	17
2.5	Operadores: Cruzamento de alto nível.	17
2.6	Esquema geral de um Sistema de Inferência Fuzzy	18
3.1	GPFIS-Forecast: Fluxo de processos	21
3.2	Funções de pertinência para as defasagens $y_{t-p} \in Y_p$ .	22
3.3	Exemplo de um conjunto de premissas codificadas em um indivíduo da PGMG.	23
3.4	Funções de Pertinência com granularidade 5.	27
3.5	Suporte do conjunto fuzzy de saída contem o valor observado.	31
3.6	Suporte do conjunto fuzzy de saída não contem o valor observado.	31
3.7	Exemplos de LPF	32
4.1	Previsão da Série 103 da Competição NN3: Granularidade 7	48
4.2	Previsão da Série 102 da Competição NN3	49
4.3	Previsão da Série 103 da Competição NN3: Granularidade 3	50

## Lista de tabelas

3.1	Conjuntos Fuzzy de Entrada e Operadores Fuzzy usados para gerar a solução ilustrada na Figura 3.3.	24
3.2	T-normas e Modificadores	28
4.1	Resultados dos experimento para três t-normas	38
4.2	Ordenação dos resultados de cada bloco da Tabela 4.1	39
4.3	Informações sobre a base de dados da M3.	40
4.4	Configuração Base do GPFIS-Forecast	41
4.5	Influência de Parâmetros: Gerações e Populações	41
4.6	Número de gerações: teste de Friedman	42
4.7	Parâmetros do GPFIS-Forecast.	42
4.8	Etapa de Fuzzificação: Ordenamento pelo Teste de Friedman	43
4.9	Influência de Parâmetros: Fuzzificação	43
4.10	Influência de Parâmetros: Formulação	44
4.11	Influência de Parâmetros: Associação	44
4.12	Influência de Parâmetros: Agregação	44
4.13	Influência de Parâmetros: Defuzzificação	45
4.14	Influência de Parâmetros: Avaliação	45
4.15	Configuração Base do GPFIS-Forecast	45
4.16	Dez primeiros colocados na competição NN3, considerando os resultados referentes às últimas 11 séries temporais.	46
4.17	Melhores resultados da NN3 para 111 séries.	46
4.18	Ranking M3 para base de dados industriais mensais	47
4.19	Exemplo de Base de Regras para granularidade 5	48

# 1

## Introdução

Métodos de previsão são muito importantes para o desenvolvimento de diversas atividades no cotidiano produtivo de nossa sociedade. Os gestores precisam de informações sobre o futuro para estabelecer planos em diversas esferas organizacionais tais como projeto de fluxos de caixa, previsão de demanda, planejamento de contratação mão-de-obra, etc Krajewski et al. (2012). Existem várias metodologias para efetuar previsões e todas elas dependem do tipo de informação disponível. Em processos que possuem um histórico de dados dependentes do tempo, o mesmo pode ser utilizado para prever eventos futuros. Tais metodologias de previsão podem servir de auxílio ao planejamento e à alocação de recursos, proporcionando, em consequência, uma economia de gastos.

Segundo Makridakis (1976) o marco para o desenvolvimento da área de análise e previsão de séries temporais deu-se no século XIX com os trabalhos de Fourier. No século seguinte Yule (1927), Walker (1931) e Slutsky (1937) desenvolveram o conceito de modelos autorregressivos e médias móveis enquanto coube a Wold (1938) desenvolver a validação teórica dos modelos ARMA Kirchgässner et al. (2012). Nas décadas seguintes, vários modelos estatísticos foram desenvolvidos, destacando-se as pesquisas de Holt (1957), Brown (1961), Winters (1960) e Box et al. (2008). Todas essas técnicas estatísticas possuem muitos pressupostos que devem ser seguidos com o intuito de se obter uma resposta aceitável.

Na década de noventa surgiram outras vertentes de pesquisa em modelos não estatísticos para prever séries temporais, com base em Redes Neurais Artificiais Park et al. (1991); Zhang et al. (1998), Lógica Fuzzy Wang e Mendel (1992) e Sistemas Híbridos Kim e Kim (1997). Tais abordagens foram impulsionadas pela diversidade de um grande número de informações armazenadas ao longo do anos em bases de dados e pelo poder computacional que passou a estar disponível. Em oposição aos modelos baseados puramente em Redes Neurais Artificiais, os que envolvem Sistemas de Inferência Fuzzy (SIFs) fornecem uma descrição do processo de previsão por meio de regras linguísticas. Nestes, tem-se também uma preocupação com a interpretabilidade linguística, e não somente com a acurácia das previsões Mangai et al. (2014); Stepnicka et al. (2014); Lu et al. (2014).

## 1.1

### Motivação

Os Sistemas Fuzzy Genéticos são modelos híbridos que se baseiam numa combinação entre uma metaheurística genética (MHG) e um Sistema Fuzzy (SF), sendo que o primeiro é responsável pelo design do segundo Herrera (2008).

As MHGs são otimizações estocásticas baseadas no conceito natural da evolução de espécies (Seleção Natural). Candidatos à solução (também denominados indivíduos, ou cromossomos) são avaliados por meio de algum critério ou função objetivo e uma nova amostra de soluções é gerada a cada iteração com base na amostra da iteração anterior Kim e Kim (1997).

O tipo de SF mais adequado quando se leva em conta a interpretabilidade é o Sistema de Inferência Fuzzy (SIF), ou Sistema Fuzzy Baseado em Regras (SFBR) Herrera (2008), dada a sua característica de proporcionar uma representação linguística da relação entre as variáveis de entrada e de saída de um sistema. Os SIFs têm sido empregados com sucesso nas áreas de Controle Shaw e SIMOES (2007), Classificação Ishibuchi e Nojima (2011) e em muitas outras Herrera e Magdalena (1997).

O modelo recente GPFIS Koshiyama (2014) – um Sistema de Inferência Fuzzy Genético baseado em Programação Genética Multigênica – mostrou-se um modelo muito rico devido ao seu potencial de extração de informações a partir de uma base de dados, conforme demonstrado pelos resultados já obtidos em aplicações em Classificação, Regressão, Previsão e Controle. No entanto restam muitos caminhos a ser explorados neste modelo, envolvendo a avaliação de diferentes configurações de seus componentes internos.

Explora-se, nesta dissertação, o GPFIS-Forecast: versão do GPFIS Koshiyama et al. (2013, 2014); Koshiyama (2014); Koshiyama et al. (2015) para previsão de série temporais univariada. Cada etapa desse modelo é repleta de possibilidades de diferentes configurações. Por exemplo: algumas exigem um esforço computacional maior, outras diminuem o espaço de busca do problema, aumentam a interpretabilidade ou absorvem mais informações que antes eram desconsideradas.

O GPFIS-Forecast apresenta, em sua execução, quatro etapas básicas: Fuzzificação, Inferência, Defuzzificação e Avaliação. Em cada uma destas etapas, pode-se fazer uso de diferentes configurações, com implicações evidentes nos resultados finais. A exploração e a avaliação destas configurações é importante para a definição dos modelos mais promissores em diferentes circunstâncias.

## **1.2** **Objetivos**

Este trabalho propõe o aperfeiçoamento do GPFIS-Forecast em duas vertentes principais:

- aumentar a quantidade de possibilidades de configurações, avaliando se elas podem contribuir significativamente para a acurácia das previsões;
- implementar informações alternativas que possam complementar a interpretação dos resultados do modelo, mesclando compromisso entre a acurácia e a interpretabilidade do modelo.

Explicitando as tarefas de forma mais detalhada, são efetuadas as seguintes avaliações:

- maneira pela qual o modelo representa os dados por meio de variáveis linguísticas, abrangendo a granularidade e também os tipos de funções de pertinência;
- reação do modelo a alterações na representação das conexões entre as proposições fuzzy (t-normas)
- significância da utilização de diferentes modificadores fuzzy para o aumento da variedade de proposições que formarão as premissas das regras;
- impacto de diferentes maneiras de associação dos consequentes às premissas (formação de regras);
- influência de diferentes maneiras de agregação das regras nos resultados;
- efeito do uso de diferentes métodos de defuzzificação;
- desenvolvimento de novas avaliações para potencializar a utilização da informação produzida pelo modelo.

No processo de avaliação das diferentes configurações, utilizam-se os bancos de dados das competições de previsão de séries temporais NN3 (2015) e M3 (2015).

### **1.3**

#### **Contribuições**

A contribuição principal deste trabalho é a avaliação de diferentes configurações do modelo GPFIS-Forecast, explorando granularidade, formas de funções de pertinência, diferentes t-normas, associações entre premissas e consequentes, operadores de agregação e métodos de defuzzificação. Como resultado, tem-se indicações das melhores configurações e um modelo mais rico em termos de variedade de configurações passíveis de utilização.

O estabelecimento de Limites de Previsão Fuzzy (LPF) é outra contribuição importante para a tomada de decisão, pois aproveita a informação antes perdida no processo de defuzzificação, que, como se sabe, produz como resultado um valor preciso.

### **1.4**

#### **Descrição e Organização da Dissertação**

Esta dissertação é estruturada da seguinte maneira: o Capítulo 2 trata de Sistemas Fuzzy Genéticos e apresenta, brevemente, os modelos mais utilizados atualmente; o Capítulo 3 aborda o modelo GPFIS-Forecast e suas principais características; o Capítulo 4 descreve a metodologia, os estudos de caso e apresenta os resultados das investigações; finalmente, o Capítulo 5 apresenta as conclusões e sugestões de trabalhos futuros.

## Sistema Fuzzy Genético (SFG)

Esse capítulo faz um apanhado de geral dos principais conceitos relacionados ao tema da dissertação, assim como um breve contextualização do panorama de modelos de previsão fuzzy presentes na literatura. A primeira seção discorre sobre a Meta-Heurística Genética (MHG), apresentando os principais conceitos de Programação Genética e a sua generalização para a Programação Genética Multigênica. A segunda seção dedica-se a definir os principais aspectos do Sistema Fuzzy Genético Baseado em Regras, elucidando os vários tipos de sistemas fuzzy-genéticos e suas principais características. A terceira seção traz uma breve descrição da importância de técnicas fuzzy de previsão e um apanhado geral dos modelos mais recentes de previsores fuzzy.

### 2.1

#### Meta-heurística Genética

As MHGs, também conhecidas como Algoritmos Evolucionários (AE), são otimizações estocásticas que se fundamentam na teoria da seleção natural. Existem várias áreas de pesquisa de AEs; dois exemplos são Algoritmos Genéticos (AG) e Programação Genética (PG) Bergmeir et al. (2012). Esta seção se inicia pela descrição dos conceitos gerais da PG, para em seguida apresentar a Programação Genética Multigênica (PGMG).

#### 2.1.1

##### Programação Genética

A Programação Genética (Koza, 1992; Poli et al., 2008; Walker, 2013) é o conjunto de métodos evolutivos que operam com populações cujos indivíduos são caracterizados como programas de computador.

A representação do indivíduo é o ponto chave para a PG, pois existem várias possibilidades de estruturá-lo, cada qual com suas particularidades. A estrutura mais comum para o indivíduo é sob a forma de árvore, conforme ilustrado na Figura 2.1, que representa o programa  $\max(X*X, X+Y)$ . Observe, neste exemplo, que o indivíduo possui duas características importantes em sua composição: é formado por terminais (X,Y), que são as folhas da árvore, e pelas operações ( $\max, *, +$ ), que são os nós.

O fluxo de acontecimentos tem início com a criação de uma população cujos indivíduos (programas de computador ou candidatos à solução) são aleatoriamente gerados por combinações entre operadores e terminais.

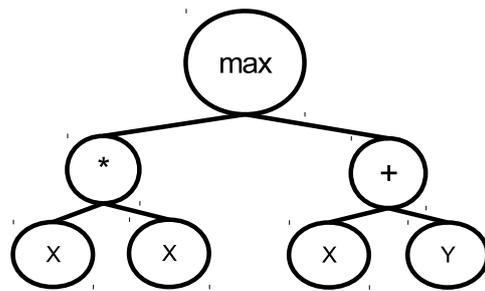


Figura 2.1: Exemplo de um indivíduo cuja estrutura é em árvore.

Utilizando novamente a Figura 2.1, os terminais são as variáveis de entrada do problema a ser otimizado, enquanto as operações, ou funções matemáticas, podem ser do tipo unitárias (logaritmo, exponencial), binárias (soma, subtração), etc.

Gerados os primeiros indivíduos, cada um deles é avaliado por um tipo de critério, ou medida de ajuste. Essa primeira avaliação, também chamada “Busca Cega” (Koza, 1992), fornecerá as diferenças de desempenho entre os indivíduos da geração zero.

A etapa seguinte compreende a seleção dos melhores indivíduos e, a partir destes, a geração de novos indivíduos por meio das operações de recombinação (cruzamento). Os novos candidatos à solução criados nesse processo compõem a nova geração da população, podendo preenchê-la completamente (não aproveitando os indivíduos antigos) ou parcialmente (aproveitando alguns programas da geração anterior). Eventualmente são realizadas alterações aleatórias em um indivíduo, processo este denominado mutação. Esta operação gera mais variabilidade na busca pelo melhor programa e evita que o algoritmo permaneça em um mínimo local. Novas avaliações e operações genéticas são efetuadas em gerações seguintes até que se atinja um critério de parada.

PUC-Rio - Certificação Digital Nº 1322099/CB

### 2.1.2

#### Programação Genética Multigênica

A Programação Genética Multigênica (Hinchliffe et al., 1996; Searson et al., 2007) é uma ampliação do conceito de PG tradicional: um indivíduo é definido como um constructo de genes, os quais são formados por estruturas em árvores (Figura 2.2). Cada gene possui um peso e as medidas de ajuste normalmente se baseiam na combinação linear dos genes. Assim, o algoritmo evolutivo básico e os processos de avaliação e de seleção são efetuados de forma similar à da PG Clássica. Cada árvore desta estrutura é uma solução parcial para o problema.

A operação genética de mutação na PGMG, esquematizada na Figura 2.3,

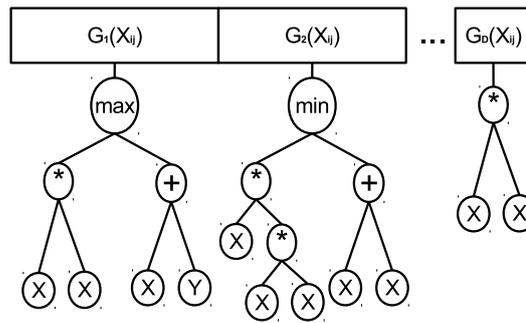


Figura 2.2: Exemplo de um indivíduo multigênico.

é conceitualmente similar à efetuada na PG clássica. Na PGMG, substitui-se uma sub-árvore por uma outra produzida aleatoriamente.

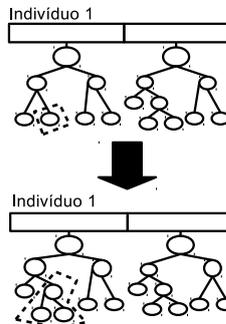


Figura 2.3: Operação de mutação em um indivíduo multigênico.

No caso da operação de cruzamento, é possível aplicar o cruzamento em dois níveis: baixo e alto.

No baixo nível, é possível manipular as estruturas (terminais e operações matemáticas) das equações presentes em um indivíduo. A Figura 2.4 apresenta uma uma operação de cruzamento de baixo nível para um indivíduo multigênico com quatro equações.

O alto nível é o espaço em que se manipulam de forma macro as equações presentes no indivíduo: a partir de dois pontos aleatórios, são permutadas equações de um indivíduo para o outro. Um exemplo de cruzamento de alto nível para um indivíduo com quatro equações é apresentado na Figura 2.5.

## 2.2 Sistema Fuzzy Genético Baseado em Regras

Os Sistemas Fuzzy Genéticos são modelos originados da combinação entre uma metaheurística genética (MHG) e um Sistema Fuzzy (SF) (Herrera, 2008). Um Sistema Fuzzy muito comum nesse tipo de abordagem é o baseado em regras (SFBR) (Herrera e Magdalena, 1997), ilustrado na Figura 2.6. O SFBR

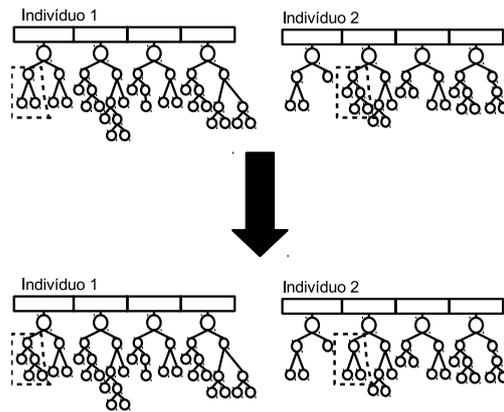


Figura 2.4: Operadores: Cruzamento de baixo nível.

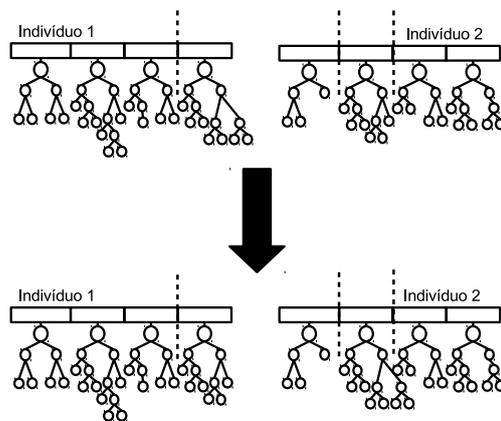


Figura 2.5: Operadores: Cruzamento de alto nível.

possui um núcleo (Base de Conhecimento), em forma de regras fuzzy, que armazena a informação relevante para a solução de problemas complexos. A Base de Conhecimento (BC) tem duas partes, conforme mostrado na Figura 2.6: a Base de Regras (BR), onde se encontram as regras construídas através de conexões entre as variáveis do modelo, e a Base de Dados (BD), onde são armazenados os termos linguísticos e as funções de pertinência, que traduzem as variáveis do modelo, possibilitando que se realizem as operações fuzzy. A evolução de um SFBR proporcionada pela MHG pode se dar em toda a Base de Conhecimento ou apenas em uma de suas partes.

Os problemas a ser resolvidos por meio de um SFG podem ser caracterizados como de ajuste (tuning), quando o objetivo é obter a melhor configuração da Base de Dados, ou de aprendizado, em que o objetivo é obter a melhor Base de Regras ou a melhor Base de Conhecimento (Herrera e Magdalena, 1997).

Observando os tipos de problemas de aprendizado a serem resolvidos pelo SFG – sob o ponto de vista de como o cromossomo da MHG é codificado – e

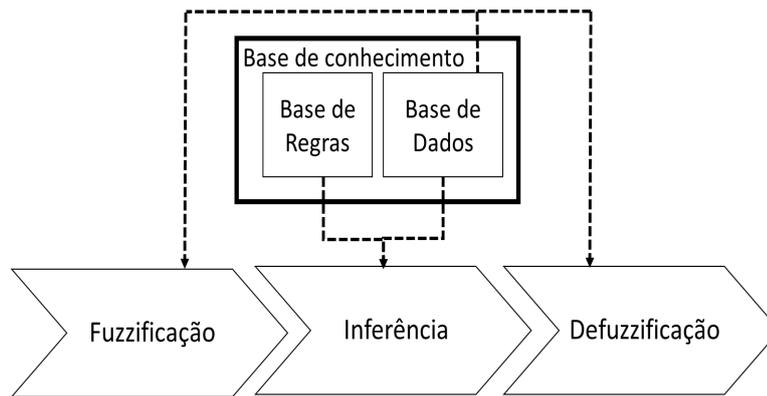


Figura 2.6: Esquema geral de um Sistema de Inferência Fuzzy

considerando que a Base de Dados é fixa, distinguem-se as seguintes categorias (Herrera e Magdalena, 1997; Herrera, 2008):

- SFG do tipo Pittsburg: um cromossomo é uma base de regras ou uma base de conhecimento. A MHG se encarrega de gerar uma população de BRs e de evoluí-la para obter o melhor indivíduo de acordo com o número de gerações estipuladas.
- SFG do tipo Michigan: um cromossomo é uma regra e a Base de Regras é representada pela população inteira. A BR é avaliada como um todo e cada regra possui um indicador de credibilidade (algumas regras podem ser mais significativas individualmente).
- SFG do tipo Iterative Rule Learning: assim como no método Michigan, cada cromossomo é uma regra, mas apenas o melhor indivíduo da população da MHG é considerado para fazer parte da BR final, sendo o resto da população descartado. O processo é repetido várias vezes até que se complete uma base de regras satisfatória.

### 2.3 Previsor Fuzzy Para Séries Temporais

Uma série temporal é um conjunto de valores dependentes do tempo. Normalmente são dados de um processo físico medido num intervalo de tempo (Kim e Kim, 1997). Matematicamente, uma série temporal pode ser expressa da seguinte forma.

$$\{y_t | t = 1, \dots, T\} \tag{2-1}$$

onde  $y_t$  uma medição registrada num tempo entre  $t$  e  $T$  é o número total de medições.

É comum representar uma série temporal pela Equação 2-2:

$$y_t = T_r + S + C + E \quad (2-2)$$

onde  $T_r$  é a tendência da série, que pode ser crescente ou decrescente,  $S$  é a sazonalidade,  $C$  representa ciclos (eventos que se repetem numa escala mais ampla do que a sazonalidade) e  $E$  é o erro (componente aleatória do processo).

A previsão de séries temporais consiste em fazer uso de dados históricos para prever valores futuros. No caso de previsão univariada, a principal informação usada para descrever e prever o comportamento de uma série temporal  $y_t \in Y$  no instante  $t$  ( $t = 1, \dots, T$ ) consiste nos seus respectivos  $P$  atrasos ou defasagens  $y_{t-1} \in Y_1, \dots, y_{t-p} \in Y_p, \dots, y_{t-P} \in Y_P$  ( $p = 1, \dots, P$ ).

Os Preditores Fuzzy (PF) para séries temporais são muito utilizados atualmente (Wang e Mendel, 1992; Kim e Kim, 1997; Mangai et al., 2014; Stepnicka et al., 2014; Lu et al., 2014; Bergmeir et al., 2012; Hadavandi et al., 2011; Koshiyama, 2014), fornecendo resultados que podem ser explicados, favorecendo a análise e viabilizando planejamentos estratégicos.

A literatura apresenta diversos PFs resultantes de hibridizações. Trabalhos recentes como o de Mangai et al. (2014) desenvolveram um modelo híbrido ARIMA-HyFIS que divide uma série temporal em duas componentes: uma parte linear, modelada por um modelo ARIMA, e o resíduo não linear, ajustado por um modelo HyFIS. Este é muito similar a um sistema do tipo ANFIS, com duas diferenças estruturais importantes: o HyFIS utiliza um sistema fuzzy do tipo Mamdani e as funções de pertinência são todas gaussianas. Os resultados do modelo híbrido ARIMA-HyFIS são melhores que os obtidos com os dos modelos individuais em vários casos; um ajuste por algoritmo genético está previsto.

Stepnicka et al. (2014) desenvolveram uma metodologia para previsão de séries temporais baseada em comitês fuzzy baseados em regras, que é uma combinação linear de resultados de preditores. Os resultados do trabalho foram pouco melhores em comparação com os dos modelos individuais, comprovando o aprimoramento.

Em Lu et al. (2014), o fuzzy c-means efetua a divisão granular da informação para a geração de mapas cognitivos que realizam as previsões.

Os SFG para séries temporais encontrados na literatura são normalmente SFBRs do tipo Takagi-Sugeno (Kim e Kim, 1997; Aznarte et al., 2012;

Bergmeir et al., 2012; Hadavandi et al., 2011), em que interpretabilidade (linguística) das regras é baixa.

O próximo capítulo trata especialmente do GPFIS-Forecast, em que a PGMG é usada como MHG para buscar a melhor BR, definindo-se assim um problema de aprendizado.

### 3

## Modelo GPFIS - Forecast

Uma das variantes do GPFIS (Koshiyama, 2014) contempla a área de previsão de séries temporais univariadas; o modelo para tal é denominado GPFIS-Forecast. Este capítulo se inicia com a descrição básica da estrutura deste modelo. Em seguida são explicitados os componentes internos do modelo.

### 3.1

#### Visão Geral

O GPFIS-Forecast é um típico SFG do tipo Pittsburgh (Herrera, 2008) que busca obter a melhor BR para um determinado problema de previsão de série temporal univariado. Sendo o SFBR desse modelo dotado de diversas configurações possíveis, é importante descrever o fluxo dos processos que o compõem, como ilustrado na Figura 3.1.

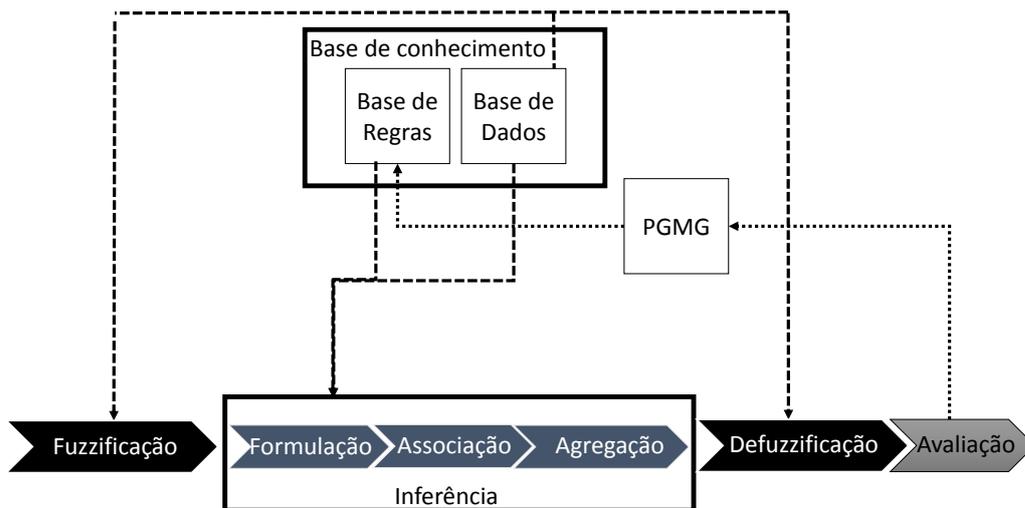


Figura 3.1: GPFIS-Forecast: Fluxo de processos

A primeira etapa executa o mapeamento dos valores precisos de entrada (no caso de séries temporais, as defasagens que formam a janela de previsão), em graus de pertinência a conjuntos fuzzy (Fuzzificação). Em seguida, é realizado o processo de Inferência, que compreende: (i) geração das premissas de regras (Formulação); (ii) união das premissas a termos consequentes (Associação) e (iii) agregação dos graus de ativação de cada regra fuzzy (Agregação). Finalmente, a Defuzzificação, que gera os valores de previsão  $\hat{y}_t$  para, em seguida, o módulo de Avaliação computar os resultados em métricas, possibilitando a evolução da PGMG.

### 3.1.1 Fuzzificação

A Fuzzificação é etapa que estabelece para as entradas do modelo as suas devidas traduções em valores fuzzy, permitindo a execução do processo de inferência e a obtenção das previsões nas etapas posteriores. Esta etapa envolve a definição da granularidade, dos formatos das funções de pertinência e de como estas são distribuídas no universo de discurso.

As entradas do GPFIS-Forecast são as  $p$  defasagens da série temporal  $y_t \in Y$ . Para cada  $y_{t-p}$  são associados  $L$  conjuntos fuzzy  $A_{lp} = \{(y_{t-p}, \mu_{A_{lp}}(y_{t-p})) | y_{t-p} \in Y_p\}$ , cada um munido de uma função de pertinência  $\mu_{A_{lp}} : Y_p \rightarrow [0, 1]$  que assinala a cada observação  $y_{t-p}$  um grau de pertinência  $\mu_{A_{lp}}(y_{t-p})$  ao conjunto fuzzy  $A_{lp}$ . Similarmente, para  $Y$  (variável de saída),  $K$  conjuntos fuzzy  $B_k$  ( $k = 1, \dots, K$ ) são associados.

Em geral, três aspectos norteiam o processo de definição das funções de pertinência: (i) forma (triangular, trapezoidal, etc.); (ii) suporte do conjunto  $\mu_{A_{lp}}(y_{t-p})$ ; e (iii) um rótulo linguístico, qualificando apropriadamente o subespaço compreendido por  $\mu_{A_{lp}}(y_{t-p})$  com um adjetivo correspondente. Idealmente, todas essas tarefas deveriam ser legadas a um especialista, cujo conhecimento sobre o tema ampliaria a acepção linguística sobre a base de regras. Como em muitos casos um especialista está indisponível, é muito comum (Berlanga et al., 2010; Herrera, 2008) definir as funções de pertinência segundo uma partição uniforme ou forte (Figura 3.2).

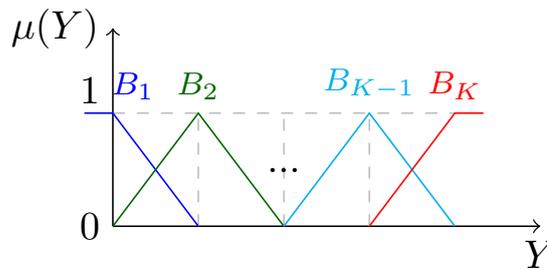


Figura 3.2: Funções de pertinência para as defasagens  $y_{t-p} \in Y_p$ .

### 3.1.2 Inferência Fuzzy

No processo de inferência ocorrem as operações entre conjuntos fuzzy. Nesta etapa definem-se os tipos de conexões entre proposições, os modificadores linguísticos e a associação entre entradas e saídas. Define-se também como as regras ativadas serão combinadas.

**Formulação**

Uma premissa de regra fuzzy é comumente definida por:

$$“Se Y_1 \text{ é } A_{11} \text{ e } \dots \text{ e } Y_p \text{ é } A_{1p} \text{ e } \dots \text{ e } Y_P \text{ é } A_{1P}”$$

ou, em termos matemáticos:

$$\mu_{A_d}(\mathbf{y}_{t,P}) = \mu_{A_{11}}(y_{t-1}) * \dots * \mu_{A_{1P}}(y_{t-P}) \tag{3-1}$$

onde  $\mu_{A_d}(y_{t-1}, \dots, y_{t-P}) = \mu_{A_d}(\mathbf{y}_{t,P})$  é o grau de pertinência conjunto dos  $P$  atrasos de  $y_t$  com respeito a  $d$ -ésima premissa ( $d = 1, \dots, D$ ), computado a partir de uma t-norma  $*$ . Note-se que  $\mathbf{y}_{t,P} = [y_{t-1}, \dots, y_{t-p}, \dots, y_{t-P}]$ . Uma premissa pode ser elaborada combinando cada  $\mu_{A_{1p}}(y_{t-p})$  por meio de t-normas, t-conormas, modificadores linguísticos e operadores de negação. Como consequência, o número de combinações possíveis cresce com o número de defasagens, conjuntos e operadores fuzzy. O GPFIS-Forecast emprega a PGMG para buscar as combinações mais promissoras de premissas  $\mu_{A_{1p}}(y_{t-p})$ . A Figura 3.3 exemplifica uma solução típica proporcionada pela PGMG.

PUC-Rio - Certificação Digital Nº 1322099/CB

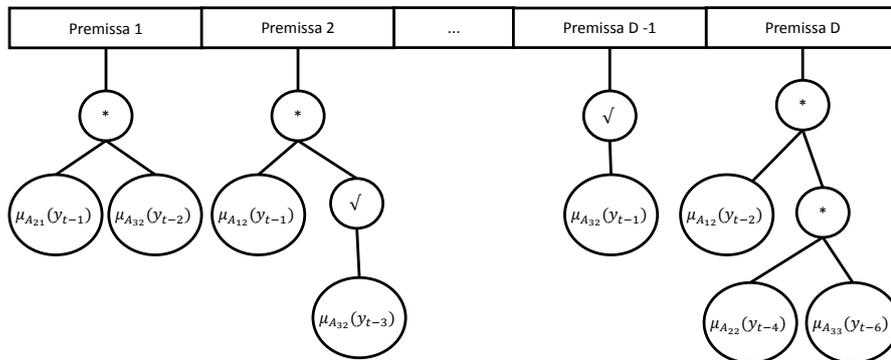


Figura 3.3: Exemplo de um conjunto de premissas codificadas em um indivíduo da PGMG.

Por exemplo, a premissa 1 representa:  $\mu_{A_1}(\mathbf{y}_{t,P}) = \mu_{A_{21}}(y_{t-1}) * \mu_{A_{32}}(y_{t-2})$  que, em termos linguísticos, significa “Se  $Y_1$  é  $A_{21}$  e  $Y_2$  é  $A_{32}$ ”. Seja  $\mu_{A_d}(\mathbf{y}_{t,P})$  a  $d$ -ésima premissa codificada na  $d$ -ésima árvore/função de um indivíduo da PGMG. A Tabela 3.1 apresenta os componentes usados para compor as soluções mostradas na Figura 3.3.

No GPFIS-Forecast, as  $\mu_{A_{1p}}(y_{t-p})$  representam os conjuntos fuzzy de entrada que, na terminologia de PG, seriam os Terminais, enquanto que o Conjunto de Funções é substituído pelos Operadores Fuzzy. A partir disso,

Tabela 3.1: Conjuntos Fuzzy de Entrada e Operadores Fuzzy usados para gerar a solução ilustrada na Figura 3.3.

Conjuntos Fuzzy de Entrada (Terminais)	Operadores Fuzzy (Conjunto de Funções)
$\mu_{A_{11}}(y_{t-1}), \dots, \mu_{A_{L1}}(y_{t-1}), \mu_{A_{lp}}(y_{t-p}), \dots, \mu_{A_{LP}}(y_{t-p})$	t-norma (*), modificadores linguísticos (operador de dilatação - $\surd$ )

a PGMG é empregada para obter um conjunto de premissas de regras fuzzy  $\mu_{A_d}(\mathbf{y}_{t,P})$ . De forma a elaborar uma base de regras plena, torna-se necessário associar cada premissa  $\mu_{A_d}(\mathbf{y}_{t,P})$  a um termo consequente  $B_k$  com o qual ela seja mais compatível. A etapa de Associação é responsável por esse processo.

### Associação

Há basicamente duas formas de se definir um termo consequente mais compatível com uma premissa de regra fuzzy: (i) permitir que a PGMG realize essa busca automaticamente (um procedimento comumente aplicado em vários trabalhos); ou (ii) empregar métodos que, a partir de informação da base de dados, defina para uma premissa dada pela PGMG o termo consequente mais apropriado. No GPFIS-Forecast a opção (ii) é a adotada, com o principal objetivo de prevenir que uma premissa com amplo suporte na base de dados, ou que esteja apta a prever corretamente um subespaço da variável de saída, seja associada a um termo consequente pouco apropriado. Com este procedimento, o modelo GPFIS-Forecast reduz o espaço de busca, pois, ao invés de buscar por todos os elementos que compõe uma regra, concentra-se somente na elaboração de cada  $\mu_{A_d}(\mathbf{y}_{t,P})$ .

Neste sentido, deve-se adotar uma métrica que meça a compatibilidade entre  $\mu_{A_d}(\mathbf{y}_{t,P})$  e um termo consequente  $B_k$ .

### Agregação

Uma premissa associada ao k-ésimo termo consequente (ou seja, uma regra fuzzy) é denotada por  $\mu_{A_{d(k)}}(\mathbf{y}_{t,P})$ , que, em termos linguísticos significa: “Se  $Y_1$  é  $A_{11}$ , e ..., e  $Y_P$  é  $A_{lP}$ , então  $Y$  é  $B_k$ ”. Portanto, toda a base de regras pode ser descrita como  $\mu_{A_{1(k)}}(\mathbf{y}_{t,P}), \dots, \mu_{A_{D(k)}}(\mathbf{y}_{t,P}), \forall k = 1, \dots, K$ . Dada uma nova observação temporal  $y_t^*$  e, por conseguinte, seus  $P$  atrasos  $\mathbf{y}_{t,P}^*$  é possível que haja uma ativação de diversas premissas, sendo estas associadas a um

mesmo termo consequente ou não. De maneira a gerar um valor consensual, a etapa de Agregação busca combinar os graus de ativação de todas as regras de um mesmo termo consequente.

Considere-se  $D^{(k)}$  o número de regras fuzzy do k-ésimo termo consequente ( $d^{(k)} = 1^{(k)}, 2^{(k)} \dots, D^{(k)}$ ). Dado um operador de agregação  $g : [0, 1]^{D^{(k)}} \rightarrow [0, 1]$  (ver Calvo et al. (2002)), o grau de pertinência predito de  $\mathbf{y}_{t,P}^*$  ao k-ésimo conjunto fuzzy ( $\hat{\mu}_{B_k}(y_t^*)$ ) é computado por:

$$\hat{\mu}_{B_1}(y_t^*) = g[\mu_{A_{1(1)}}(\mathbf{y}_{t,P}^*), \dots, \mu_{A_{D(1)}}(\mathbf{y}_{t,P}^*)] \quad (3-2)$$

$$\hat{\mu}_{B_2}(y_t^*) = g[\mu_{A_{1(2)}}(\mathbf{y}_{t,P}^*), \dots, \mu_{A_{D(2)}}(\mathbf{y}_{t,P}^*)] \quad (3-3)$$

...

$$\hat{\mu}_{B_K}(y_t^*) = g[\mu_{A_{1(K)}}(\mathbf{y}_{t,P}^*), \dots, \mu_{A_{D(K)}}(\mathbf{y}_{t,P}^*)] \quad (3-4)$$

Há vários operadores de agregação disponíveis na literatura (ver Calvo et al. (2002)), sendo o de Máximo o mais usado entre eles. Não obstante, outros operadores tais como a média aritmética ou a ponderada são passíveis de uso.

Obtidos os valores para cada  $\hat{\mu}_{B_1}(y_t^*), \dots, \hat{\mu}_{B_K}(y_t^*)$ , a etapa de Defuzzificação utiliza essa informação para computar um valor predito  $\hat{y}_t^*$ .

### 3.1.3

#### Defuzzificação

#### Previsão Pontual

O objetivo desta etapa é obter um valor preciso de previsão  $\hat{y}_t$ , que será usado na avaliação do indivíduo do modelo GPFIS-Forecast.

Há diversos métodos de defuzzificação disponíveis, sendo que os mais usados são o Centro de Área, a Média dos Máximos e a Altura (Roychowdhury e Pedrycz, 2001).

### 3.1.4

#### Avaliação

O procedimento de Avaliação no GPFIS-Forecast é definido por um objetivo primário – minimização de erro – e um secundário – redução de complexidade. O objetivo primário é responsável pela ordenação dos indivíduos na população, ao passo que o secundário é usado como critério de desempate.

Uma função objetivo bastante usada em problemas de previsão, e empregada ao longo dos experimentos, é o Erro Médio Simétrico Absoluto Percentual (*SMAPE*):

$$SMAPE = \frac{\sum_{t=1}^T \frac{|y_t - \hat{y}_t|}{(y_t + \hat{y}_t)/2}}{T} \quad (3-5)$$

O melhor indivíduo na população é a solução que minimiza a equação (3-5). Ao utilizar o critério secundário, o GPFIS-Forecast tenta reduzir a complexidade da base de regras por meio de uma simples heurística: a Pressão Lexicográfica Parcimoniosa (Luke e Panait, 2002). Essa técnica somente é usada na fase de seleção: dado dois indivíduos com a mesma aptidão, o melhor entre eles é o que possui o menor número de nós. Menor número de nós indica regras com poucos elementos antecedentes, modificadores linguísticos ou operadores de negação, assim como pouco premissas ( $\mu_{A_d}(\mathbf{y}_{t,P})$ ), e, portanto, uma base de regras mais compacta. Após a etapa de Avaliação, os indivíduos são selecionados (pelo procedimento do Torneio) e recombinaados, aplicando-se os operadores de mutação e cruzamento de baixo e alto níveis. Esse processo é repetido até que um critério de parada seja atendido. Quando isto ocorre, a última população é retornada e o melhor indivíduo providencia as suas previsões.

## 3.2

### Configuração do GPFIS-Forecast

Esta seção discorre sobre as diferentes configurações consideradas no GPFIS-Forecast.

#### 3.2.1

##### Configurações na Fuzzificação

As diferentes configurações consideradas na etapa de fuzzificação podem ser divididas em três categorias:

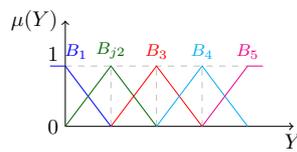
- Forma das funções de pertinência: triangular ou Gaussiana;
- Distribuição da funções de pertinência: uniforme ou segundo o método de Tukey;
- Número de termos linguísticos (granularidade): três, cinco ou sete.

Quanto ao segundo item acima, em uma distribuição uniforme os suportes dos conjuntos fuzzy são uniformemente distribuídos ao longo do universo de discurso, conforme mostrado nas Figuras 3.4a e 3.4c para cinco termos linguísticos e formatos triangular e Gaussiana. Esta distribuição (uniforme) é denominada fortemente particionada.

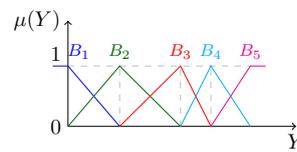
Na partição segundo o método de Tukey (ou simplesmente Tukey), adotam-se as medidas a seguir:

- Para granularidade 3, os pontos do mínimo, da média e do máximo da série temporal são utilizados como referência para distribuir os suportes das funções de pertinência.
- Para granularidade 5, os pontos do mínimo, da mediana (Q2), dos quartis (Q1 e Q3) e do máximo da série temporal são considerados na distribuição de cada suporte para as funções de pertinência.
- Para granularidade 7: o mínimo, a mediana, outros 4 quartis e o máximo da série temporal são considerados na distribuição dos suportes das funções de pertinência.

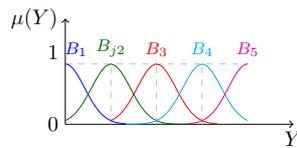
As Figuras 3.4b e 3.4d ilustram distribuições do tipo Tukey.



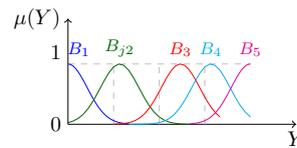
3.4(a): Triangular Fortemente Particionada



3.4(b): Triangular Tukey



3.4(c): Gaussiana Fortemente Particionada



3.4(d): Gaussiana tukey

Figura 3.4: Funções de Pertinência com granularidade 5.

No que diz respeito à granularidade, vale salientar que, quanto maior a granularidade, maior será a quantidade de terminais disponíveis para serem combinados e transformados em regras.

### 3.2.2 Configurações na Inferência

Conforme visto anteriormente, a Inferência compreende as etapas de Formulação, Associação e Agregação, para as quais há várias configurações possíveis.

### Tipos de Formulação

Na etapa de Formulação, diferentes t-normas podem ser empregadas na conexão dos elementos de uma premissa. Além disso, pode-se fazer uso de modificadores, que dimensionam a intensidade dos elementos das premissas e são importantes para um melhor ajuste das sentenças lógicas, embora aumentando o espaço de busca. A Tabela 3.2 mostra as t-normas e modificadores considerados.

Tabela 3.2: T-normas e Modificadores

T-norma Produto	$\mu_{A_{ij}}(y_{t-j1}) * \mu_{A_{ij}}(y_{t-j2})$
T-norma Mínimo	$Minimo(\mu_{A_{ij}}(y_{t-j1}), \mu_{A_{ij}}(y_{t-j2}))$
T-norma Lukasiewicz	$Maximo(0, \mu_{A_{ij}}(y_{t-j1}) + \mu_{A_{ij}}(y_{t-j2}) - 1)$
Modificador Diluidor	$[\mu_{A_{ij}}(y_{t-j1})]^{1/2}$
Modificador Concentrador	$[\mu_{A_{ij}}(y_{t-j1})]^2$

### Tipos de Associação

Os tipos de associação considerados são:

- **Grau de Confiança Fuzzy** ( $FCD_k$ ):

$$FCD_k = \frac{\sum_{t=1}^T \mu_{A_d}(\mathbf{y}_{t,P}) \mu_{B_k}(y_t)}{\sqrt{\sum_{t=1}^T \mu_{A_d}(\mathbf{y}_{t,P})^2} \sqrt{\sum_{t=1}^T \mu_{B_k}(y_t)^2}} \in [0, 1] \quad (3-6)$$

onde  $\sum_{t=1}^T \mu_{A_d}(\mathbf{y}_{t,P}) \mu_{B_k}(y_t)$  é o grau de compatibilidade entre a d-ésima premissa e o k-ésimo termo consequente, enquanto que  $\sqrt{\sum_{t=1}^T \mu_{A_d}(\mathbf{y}_{t,P})^2}$  e  $\sqrt{\sum_{t=1}^T \mu_{B_k}(y_t)^2}$  são fatores de normalização. Quando  $\mu_{A_d}(\mathbf{y}_{t,P}) = \mu_{B_k}(y_t)$  para todo  $t$ , então  $FCD_k = 1$ , ou seja, a premissa e o termo consequente são totalmente compatíveis. O termo consequente de  $\mu_{A_d}(\mathbf{y}_{t,P})$  é definido pelo  $k$  que maximiza a métrica  $FCD_k$ . Uma premissa com  $FCD_k = 0$ , para todo  $k$ , não é associada a nenhum termo consequente (e, portanto, não considerada como uma regra fuzzy).

- **Máximo dos Mínimos:**

$$GM_k = \frac{\sum_{t=1}^T \min(\mu_{A_d}(\mathbf{y}_{t,P}), \mu_{B_k}(y_t))}{\sqrt{\sum_{t=1}^T \mu_{A_d}(\mathbf{y}_{t,P})^2} \sqrt{\sum_{t=1}^T \mu_{B_k}(y_t)^2}} \in [0, 1] \quad (3-7)$$

onde  $\sum_{t=1}^T \min(\mu_{A_d}(\mathbf{y}_{t,P}), \mu_{B_k}(y_t))$  é o grau mínimo de compatibilidade entre a d-ésima premissa e o k-ésimo termo consequente, enquanto

que  $\sqrt{\sum_{t=1}^T \mu_{A_d}(\mathbf{y}_{t,P})^2}$  e  $\sqrt{\sum_{t=1}^T \mu_{B_k}(y_t)^2}$  são fatores de normalização. Quando  $\mu_{A_d}(\mathbf{y}_{t,P}) = \mu_{B_k}(y_t)$  para todo  $t$ , então  $GM_k = 1$ , ou seja, a premissa e o termo consequente são totalmente compatíveis. O termo consequente de  $\mu_{A_d}(\mathbf{y}_{t,P})$  é definido pelo que  $k$  que maximiza a métrica  $GM_k$ . Uma premissa com  $GM_k = 0$ , para todo  $k$ , não é associada a nenhum termo consequente (e, portanto, não considerada como uma regra fuzzy).

- **Mínimos Quadrados Restrito:** o melhor consequente é aquele que apresenta maior peso  $w_{d^{(k)}}$ , segundo a Equação 3-8.

$$\begin{aligned} \min : & \sum_{t=1}^T (\hat{\mu}_{B_k}(y_t) - \sum_{d^{(k)=1}^{D^{(k)}} w_{d^{(k)}} \mu_{A_{d^{(k)}}}(\mathbf{y}_{t,P}))^2 & (3-8) \\ \text{s.a. :} & \sum_{d^{(k)=1}^{D^{(k)}} w_{d^{(k)}} = 1 \text{ e } w_{d^{(k)}} \geq 0 \end{aligned}$$

onde  $w_{d^{(k)}}$  é o peso ou o grau de influência de  $\mu_{A_{d^{(k)}}}(\mathbf{y}_{t,P})$  em prever os valores associados ao  $k$ -ésimo termo consequente. O problema de minimização acima é típico de Programação Quadrática, cuja a solução é computada a partir dos algoritmos descritos em Schölkopf e Smola (2001).

### Tipos de Agregação

Nesta etapa, consideram-se as seguintes possibilidades para efetuar a agregação:

- **Média Aritmética Ponderada:** a definição dos pesos advém da resolução do problema de Mínimos Quadrados Restrito (MQR) expresso na Equação 3-8.
- **Mínimos Quadrados Iterativo:** geram-se combinações de regras, conforme expresso abaixo.

$$\begin{aligned} \hat{\mu}_{B_1}(y_t^*) &= g[\mu_{A_{1(1)}}(\mathbf{y}_{t,P}^*), \dots, \mu_{A_{D(1)}}(\mathbf{y}_{t,P}^*)] \\ \hat{\mu}_{B_2}(y_t^*) &= g[\mu_{A_{1(2)}}(\mathbf{y}_{t,P}^*), \dots, \mu_{A_{D(2)}}(\mathbf{y}_{t,P}^*)] \\ &\dots \\ \hat{\mu}_{B_K}(y_t^*) &= g[\mu_{A_{1(K)}}(\mathbf{y}_{t,P}^*), \dots, \mu_{A_{D(K)}}(\mathbf{y}_{t,P}^*)] \\ \mu_{B_{K+1}}(y_t^*) &= \hat{\mu}_{B_1}(y_t^*) * \hat{\mu}_{B_1}(y_t^*), \\ \mu_{B_{K+2}}(y_t^*) &= \hat{\mu}_{B_1}(y_t^*) * \hat{\mu}_{B_2}(y_t^*), \\ &\dots, \\ \mu_{B_{K+\binom{K}{2}}}(y_t^*) &= \hat{\mu}_{B_K}(y_t^*) * \hat{\mu}_{B_K}(y_t^*) \end{aligned}$$

Este método faz com que certas premissas ocorram em conjunto e no momento em que o cálculo dos mínimos quadrados é efetuado pela equação 3-8.

– **Máximo:**

$$\begin{aligned} \hat{\mu}_{B_1}(y_t^*) &= \max[\mu_{A_{1(1)}}(\mathbf{y}_{t,P}^*), \dots, \mu_{A_{D(1)}}(\mathbf{y}_{t,P}^*)] \\ \hat{\mu}_{B_2}(y_t^*) &= \max[\mu_{A_{1(2)}}(\mathbf{y}_{t,P}^*), \dots, \mu_{A_{D(2)}}(\mathbf{y}_{t,P}^*)] \\ &\dots \\ \hat{\mu}_{B_K}(y_t^*) &= \max[\mu_{A_{1(K)}}(\mathbf{y}_{t,P}^*), \dots, \mu_{A_{D(K)}}(\mathbf{y}_{t,P}^*)] \end{aligned}$$

### 3.2.3

#### Configurações na Defuzzificação

Consideram-se os seguintes métodos de defuzzificação nesta etapa:

– **Método da Altura (3-9)**, dado por:

$$\hat{y}_t = \frac{c_1 \hat{\mu}_{B_1}(y_t) + \dots + c_K \hat{\mu}_{B_K}(y_t)}{\hat{\mu}_{B_1}(y_t) + \dots + \hat{\mu}_{B_K}(y_t)} \quad (3-9)$$

sendo  $c_k$  o ponto médio do suporte do termo k.

– **Média dos Máximos (MoM)**: o valor preciso é dado pela média entre o menor e o maior valor no universo que apresentam o maior grau de pertinência ao conjunto fuzzy de saída.

– **Centro de Área (CoA)**, em que o valor defuzzificado é dado por

$$\hat{y}_t = \frac{\sum_{z=\min(Y)}^{\max(Y)} z * \hat{\mu}_{B_1}(y_t) + \dots + \sum_{z=\min(Y)}^{\max(Y)} z * \hat{\mu}_{B_K}(y_t)}{\hat{\mu}_{B_1}(y_t) + \dots + \hat{\mu}_{B_K}(y_t)} \quad (3-10)$$

### 3.2.4 Formas de Avaliação

As previsões são geradas a partir da defuzzificação do conjunto fuzzy de saída. O valor previsto pode estar contido ou não no suporte deste conjunto, como visto nas Figuras 3.5 e 3.6, respectivamente.

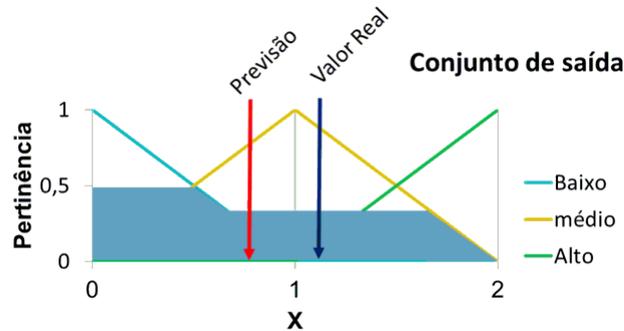


Figura 3.5: Suporte do conjunto fuzzy de saída contém o valor observado.

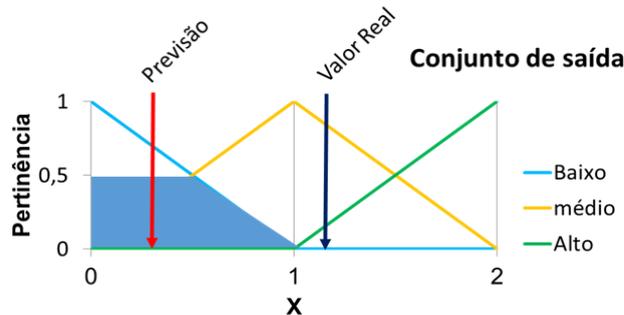


Figura 3.6: Suporte do conjunto fuzzy de saída não contém o valor observado.

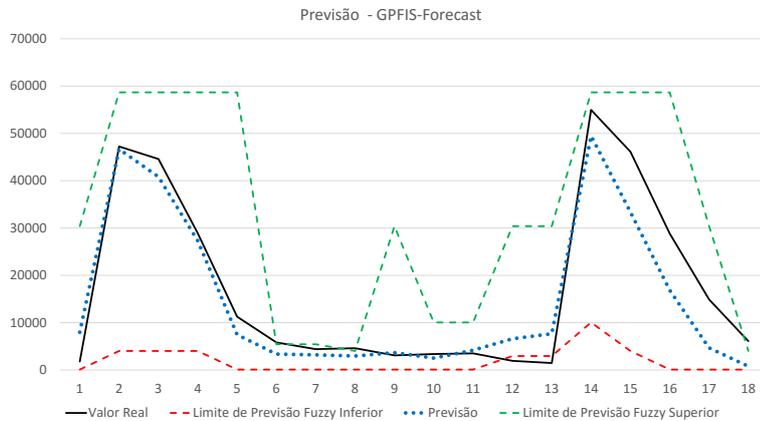
O valor preciso resultante do processo de defuzzificação não expressa de maneira completa toda a informação contida no conjunto fuzzy de saída. Com o objetivo de enriquecer a informação fornecida por tal conjunto, estabelecem-se aqui os Limites de Previsão Fuzzy (LPF), abordados nas seções subsequentes.

#### Limite de Previsão Intervalar

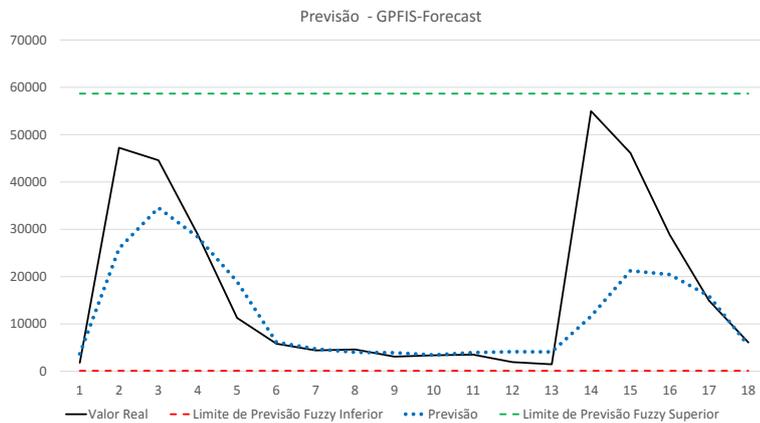
De maneira geral, o objetivo de considerar os LPF numa métrica é:

(1) buscar, no treinamento, o modelo que contempla o maior número de pontos de valores desejados contidos no suporte do conjunto fuzzy de saída. A Figura 3.7(a) ilustra um intervalo de previsão fuzzy que capta tais pontos na fase in-sample.

(2) buscar um modelo que apresente conjunto fuzzy de saída com suporte pequeno e contemple o máximo de valores desejados no treinamento. A Figura 3.7(b) mostra um intervalo fuzzy bem largo, contendo todos os pontos de valores desejados, mas não tão bem ajustado quanto o da Figura 3.7(a).



3.7(a): LPF que contempla muitos pontos



3.7(b): LPF muito largo

Figura 3.7: Exemplos de LPF

A função de ajuste, ou função objetivo, do GPFIS-Forecast é normalmente a minimização do SMAPE. Logo os melhores indivíduos, ou as melhores base de regras, possuem os menores SMAPEs.

Levando em consideração também o fato de que o valor observado pode estar ou não incluso no intervalo entre os limites de previsão fuzzy, faz-se uso de uma nova métrica criada no âmbito desse trabalho, o PISMAPE – Prediction Interval and Symmetric mean absolute percentage errors. Calcula-se inicialmente a frequência de pontos alvo fora dos limites de previsão fuzzy (PIe):

$$PIe = \frac{\sum T_i^{out}}{h} \quad (3-11)$$

sendo  $T_i^{out}$  os tais pontos e  $h$  o horizonte de previsão. Nota-se que o limite de  $PIe$  está entre 0 e 1:  $PIe = 0$  quando todos os pontos são contemplados pelos LPF e  $PIe = 1$  quando nenhum ponto está contido nos LPF.

Calcula-se a seguir a largura do intervalo:

$$PIIe = \frac{\sum_{i=1}^h \frac{U_i - L_i}{\max(Y) - \min(Y)}}{h} \quad (3-12)$$

Sendo  $U_i$  e  $L_i$  os Limites de Previsão Fuzzy do ponto  $i$ . Se, para todo  $i$ ,  $U_i - L_i = \max(Y) - \min(Y)$ ,  $PIIe = 1$ . Logo, minimizar esse indicador significa minimizar a largura do LPF.

O PISMAPE é então computado:

$$PISMAPE = p * SMAPE + (1 - p) * \frac{(PIe + PIIe)}{2} \quad (3-13)$$

onde  $p$  é um peso (atualmente  $p=0,7$ ).

Pode-se notar que o desempenho da métrica PISMAPE depende de três fatores:

- O valor do SMAPE da previsão
- O tipo de cálculo efetuado para os valores dos LPF. Eles são calculados de acordo com a filosofia de aproveitamento da informação originada do próprio GPFIS.
- A largura dos LPF.

### Limite de Previsão Fuzzy Suporte

Os LPF Suporte são baseados no intervalo de suporte do conjunto de saída do Sistema de Inferência Fuzzy.

Considere-se uma variável linguística de saída  $S$ , com  $K$  termos; cada termo  $k$  é representado por um vetor  $p_k$  de parâmetros do tipo  $(p_{1,k}, p_{2,k}, p_{3,k})$ , onde o intervalo  $[p_{1,k}, p_{3,k}]$  é o suporte do termo e  $p_{2,k} \in [p_{1,k}, p_{3,k}]$ . Considere-se ainda o peso  $w_k \in [0, 1]$ ,  $k = 1, 2 \dots K$ , resultante da agregação para o termo  $k$  de  $S$ , e constantes  $M$  e  $Mm$ , que expressam, respectivamente, um valor muito grande e um valor muito pequeno.

Os limites inferior e superior são dados por:

$$\text{Limite Inferior} = \min(L_1, \dots, L_K) \quad (3-14)$$

onde

$$L_k = \begin{cases} p_{2,1}, & \text{Se } k = 1 \text{ e } w_1 > 0 \\ p_{1,k-1}, & \text{Se } k \neq 1 \text{ e } w_k > 0 \\ M, & \text{Se } w_k = 0 \end{cases} \quad (3-15)$$

$$\text{Limite Superior} = \max(U_1, \dots, U_K) \quad (3-16)$$

onde

$$U_k = \begin{cases} p_{2,K}, & \text{Se } k = K \text{ e } w_k > 0 \\ p_{3,k+1}, & \text{Se } k \neq K \text{ e } w_k > 0, \\ Mm, & \text{Se } w_k > 0 \end{cases} \quad (3-17)$$

Seja um exemplo de uma variável linguística de saída com granularidade 5, espaço de busca  $E \in [0, 10]$ , termos fortemente distribuídos,  $p_1 \in (0.0, 0.0, 2.5)$ ,  $p_2 \in (0.0, 2.5, 5.0)$ ,  $p_3 \in (2, 5, 5.0, 7.5)$ ,  $p_4 \in (5.0, 7.5, 10.0)$ ,  $p_5 \in (7.5, 10.0, 10.0)$  e os seguintes pesos resultantes da agregação para o ponto de previsão:  $w = \{0.4, 0.6, 0.0, 0.0, 0.0\}$ .

Neste caso, são ativados os termos  $B_1$  e  $B_2$ , cujos parâmetros são, respectivamente,  $p_1 = (0.0, 0.0, 2.5)$  e  $p_2 = (0, 2.5, 5.0)$ . Os limites de previsão resultantes são:

$$\text{Limite inferior} = \min((0.0, 0.0, 2.5), (0, 2.5, 5.0), M, M, M) = 0$$

$$\text{Limite superior} = \max((0.0, 0.0, 2.5), (0, 2.5, 5.0), Mm, Mm, Mm) = 5$$

Considerando que o alvo está fora do intervalo  $[0, 5]$ , ou seja,  $PIe = 1$ , o  $PHe$  será:

$$PHe = \frac{5 - 0}{10 - 0} = 0.5$$

Para uma previsão de SMAPE = 0.4, o PISMAPE será:

$$PISMAPE = 0.7 * 0.4 + 0.3 * \frac{(1 + 0.5)}{2} = 0.505$$

### Limite de Previsão Fuzzy Sigma

Estes limites baseiam-se na medida de dispersão dos pesos oriundos do método de agregação.

Dada uma variável linguística de saída  $S$ , com  $K$  termos, considerem-se os parâmetros já especificados no caso dos limites de previsão do tipo Suporte.

$$\text{Limite Inferior} = \min(L_1, \dots, L_K) \quad (3-18)$$

onde

$$L_k = \begin{cases} p_{2,1} * (w_k - sd(w)), & \text{Se } k = 1 \text{ e } w_1 > 0 \\ p_{1,k-1} * (w_k - sd(w)), & \text{Se } k \neq 1 \text{ e } w_k > 0 \\ M, & \text{Se } w_k = 0 \end{cases} \quad (3-19)$$

sendo

$$sd(w) = \sqrt{var(w)} \quad (3-20)$$

$$var(w) = \sum_{k=1}^K (k^2 * w_k - \bar{w}^2) \quad (3-21)$$

$$(3-22)$$

$$\text{Limite Superior} = \max(U_1, \dots, U_K) \quad (3-23)$$

onde

$$U_k = \begin{cases} p_{3,K} * (w_k + sd(w)), & \text{Se } k = K \text{ e } w_n > 0 \\ p_{3,k+1} * (w_k + sd(w)), & \text{Se } k \neq K \text{ e } w_k > 0, \\ Mm, & \text{Se } w_k > 0 \end{cases} \quad (3-24)$$

Seja um exemplo de uma variável linguística de saída com granularidade 5, espaço de busca  $E \in [0, 10]$ , com termos fortemente distribuídos,  $p_1 \in (0.0, 0.0, 2.5)$ ,  $p_2 \in (0.0, 2.5, 5.0)$ ,  $p_3 \in (2, 5, 5.0, 7.5)$ ,  $p_4 \in (5.0, 7.5, 10.0)$ ,  $p_5 \in (7.5, 10.0, 10.0)$  e pesos resultantes da agregação para o ponto de previsão dados por  $w = \{0.4, 0.6, 0.0, 0.0, 0.0\}$ .

São ativados os termos  $B_1$  e  $B_2$ , para os quais  $p1 = (0.0, 0.0, 2.5)$  e  $p2 = (0, 2.5, 5.0)$ , respectivamente. Os limites de previsão, resultantes dos cálculos a seguir, são:  $\bar{w} = 1.6$ ;  $var(w) = \sum_{k=1}^5 (k^2 * w_k) - 2.56 = ((1^2) * 0.4 + (2^2) * 0.6 + (3^2) * 0 + (4^2) * 0 + (5^2) * 0) - 2.56 = 0.24$ ; e  $sd(w) = \sqrt{0.24} = 0.48$ .

$$\text{Limite inferior} = \min(p1 * (0.6 - 0.48), p2 * (0.4 - 0.48), M, M, M) = 0$$

$$\text{Limite superior} = \max(p1 * (0.6 + 0.48), p2 * (0.4 + 0.48), Mm, Mm, Mm) = 4.4$$

Considerando que o alvo está fora do intervalo  $[0, 4.4]$ , ou seja,  $PIe = 1$ ,

o *PIle* será:

$$PIle = \frac{4.4 - 0}{10 - 0} = 0.44$$

Para uma previsão de  $SMAPE = 0.4$ , o  $PISMAPE$  será:

$$PISMAPE = 0.7 * 0.4 + 0.3 * \frac{(1 + 0.44)}{2} = 0.496$$

## 4

### Experimentos e Resultados

Este capítulo aborda a metodologia utilizada para obtenção dos resultados, as principais métricas para avaliação de desempenho do modelo, os estudos de caso e as investigações realizadas.

#### 4.1

##### Delineamento Experimental

Um delineamento experimental é definido como um processo em que o planejamento e a condução de um experimento são executados para obter dados que possam ser analisados por meio de técnicas estatísticas adequadas, proporcionando com isso conclusões confiáveis e objetivas. Dois tipos de delineamento são frequentes na literatura. O Delineamento Inteiramente Causalizado (DIC) não utiliza o princípio do controle local pois há certeza de que há homogeneidade das condições experimentais; os dados são arranjados de maneira aleatória, diminuindo significativamente o ruído experimental. No Delineamento em Blocos Causalizados (DBC), em que existe heterogeneidade entre as parcelas experimentais, o controle local é realizado pelo agrupamento de parcelas homogêneas em blocos. Os tratamentos são colocados nos blocos de maneira aleatória, de forma que cada um deles tenha todos os tipos de tratamentos disponíveis para análise.

O delineamento do experimento utilizado nessa pesquisa é do tipo DBC, onde cada bloco é composto por uma amostra de cada configuração do modelo GPFIS.

#### 4.2

##### Experimento Sequencial

Devido a grande quantidade de parâmetros a testar (t-normas, métodos de agregação e de defuzzificação, etc.), é inviável avaliar todas as combinações possíveis para 111 séries temporais. Assim, opta-se por realizar experimentos sequenciais, em que cada um deles é do tipo DBC, composto por vários níveis.

Por exemplo, pode-se avaliar o desempenho do modelo para diversas t-normas, mantendo fixos outros parâmetros, e, em seguida, testar vários modificadores, com os outros parâmetros fixos – inclusive o melhor parâmetro do experimento anterior.

No final de cada experimento realiza-se o teste não-paramétrico de Friedman, com o intuito de identificar se alguma das configurações se destaca

das demais e qual a melhor configuração atual. Esse teste é explicado a seguir.

### 4.3

#### Teste de Friedman

O teste não paramétrico de Friedman é utilizado para comparação de dois ou mais experimentos em blocos ao acaso. O teste de hipótese a ser verificado é o seguinte:

- $H_0$ : as distribuições das  $k$  amostras são idênticas.
- $H_1$ : as distribuições das  $k$  amostras diferem na localização.

Seja um experimento com  $k$  parâmetros a testar e  $b$  blocos. Em cada um dos destes, as observações são ordenadas de maneira crescente e atribuídas números de ordem. Em seguida, somam-se os números de ordem (atribuídos por bloco) em cada um dos tratamentos ( $R_i$ ).

A estatística de teste ( $X^2$ ) é calculada pela seguinte equação:

$$X^2 = \frac{12}{b.k.(k+1)} \sum_{i=1}^k R_i^2 - 3.b.(k+1) \quad (4-1)$$

Como exemplo, pode-se utilizar a Tabela 4.1, que mostra o resultado de dez blocos causalizados e três t-normas (min, prod e Lukasiewicz). A avaliação destes experimentos é dado pela métrica SMAPE (ver equação 3-5). Deseja-se, portanto, saber se existe uma t-norma que se destaca na performance do SMAPE.

Tabela 4.1: Resultados dos experimento para três t-normas

MIN	PROD	LUKA
0,101289607	0,112480411	0,105341341
0,063301308	0,062017334	0,056916278
0,108771433	0,106228711	0,113134553
0,08028098	0,071415085	0,094430834
0,122677387	0,119540229	0,115063401
0,092429609	0,077357699	0,072563856
0,262178105	0,236046577	0,260362754
0,255975831	0,217090451	0,26902413
0,145881268	0,148671092	0,247105416
0,121565084	0,145879562	0,243110842

Os dados da tabela 4.1 são então ordenados, gerando-se a tabela 4.2.

Tabela 4.2: Ordenação dos resultados de cada bloco da Tabela 4.1

	MIN	PROD	LUKA
	1	3	2
	3	2	1
	2	1	3
	2	1	3
	3	2	1
	3	2	1
	3	1	2
	2	1	3
	1	2	3
	1	2	3
$R_i$	21	17	22

A partir dos dados da tabelas 4.1 e 4.2, calcula-se a estatística de teste:

$$X^2 = \frac{12}{10 * 3 * (3 + 1)} \sum_{i=1}^3 (21^2 + 17^2 + 22^2) - 3 * 10 * (3 + 1)$$

então,

$$X^2 = 121,4 - 120 = 1,4$$

Considerando um nível de significância  $\alpha = 0,05$ , o valor crítico da distribuição é  $\chi_{0,05;3}^2 = 7,8147$ ; como  $X^2 = 1,4 < \chi_{0,05;3}^2 = 7,8147$ , não há indícios para rejeitar a hipótese nula, ou seja, não há diferenças estatísticas significativas para as três t-normas.

Como este resultado ainda não indica qual dos três métodos testados é o melhor, faz-se a análise post-hoc Iman-Davenport (Trawinski et al., 2012), definindo-se assim o método que será escolhido para próxima sequência de experimentos.

#### 4.4

##### Estudo de casos

Foram escolhidas bases de dados referentes às competições NN3 e M3 para o estudo de casos. No primeiro estudo, avalia-se o desempenho do GPFIS-Forecast frente a variações em seus parâmetros internos. O segundo serve para validar a melhor configuração obtida no primeiro.

#### 4.4.1

##### Competição NN3

A competição para redes neurais artificiais e inteligência computacional (NN3) (NN3, 2015) foi destinada à avaliação do desempenho de métodos provenientes de Inteligencia Computacional. Os requisitos eram:

- previsão de 11 e 111 series temporais mensais.
- validação cruzada 1-fold.
- sem intervenção do usuário na metodologia de previsão.
- avaliação out-of-sample
- horizonte fixo de 18 meses.
- avaliação pelo SMAPE

#### 4.4.2

##### Competição - M3

As competições M (M3, 2015) têm por objetivo expandir e incentivar a pesquisa na área de previsão. A M3 possui uma grande coleção de séries temporais divididas em 6 categorias, com 4 tipos de períodos, totalizando 3003 séries, cf. Tabela 4.3. Optou-se por utilizar a categoria Industrial mensal da base de dados, totalizando 334 séries temporais, com horizonte de previsão de 18 períodos e janela de previsão de 12 períodos.

Tabela 4.3: Informações sobre a base de dados da M3.

INTERVAL	MICRO	INDUSTRY	MACRO	FINANCE	DEMOG	OTHER	TOTAL
Yearly	146	102	83	58	245	11	645
Quarterly	204	83	336	76	57	0	756
Monthly	474	334	312	145	111	52	1428
Other	4	0	0	29	0	141	174
Total	828	519	731	308	413	204	3003

Tabela 4.4: Configuração Base do GPFIS-Forecast

Configuração Base	
Formato	Triangular
Distribuição	Fort. Parcit.
Granularidade	5
T-norma	Produto
Modificadores	Nenhum
Associação	FCD
Agregação	MQR
Defuzzyficação	Altura
Avaliação	SMAPE

### 4.4.3 Experimentos

A medida de desempenho utilizada no teste de Friedman é o SMAPE out-of-sample médio para dez experimentos: obtém-se a previsão da série dez vezes e a média das dez previsões de cada ponto é utilizada para o cálculo do SMAPE.

Tomando por base a configuração do melhor experimento de Koshiyama (2014), cujos parâmetros são apresentados na Tabela 4.4, avaliou-se o desempenho do modelo para 500, 250, 100, 50 gerações da MHG na previsão das 11 últimas séries da NN3.

A Tabela 4.5 apresenta o SMAPE obtido para cada número de gerações considerado.

Tabela 4.5: Influência de Parâmetros: Gerações e Populações

População	Repetições	nº de gerações	Configuração	SMAPE 11
100	10	500	Base	22,07%
100	10	250	Base	22,87%
100	10	100	Base	22,40%
100	10	50	Base	23,35%
200	10	50	Base	22,95%

Apesar de os resultados apresentados na Tabela 4.6 sugerirem uma pequena vantagem da configuração com 500 gerações(t500) em relação às outras, o Teste de Friedman para os cinco experimentos produziu p-valor = 0.533786, que não permite afirmar que existe uma diferença estatística entre os resultados.

Dada a quantidade de parâmetros a ajustar, realizaram-se testes preliminares com o objetivo de fixar alguns deles. Os que permanecerão fixos

Tabela 4.6: Número de gerações: teste de Friedman

Algorithm	Ranking
t500	2.3636
t250	3.3636
t100	2.7727
t50	3.2727
t50-200	3.2273

ao longo dos experimentos são mostrados na Tabela 4.7.

Tabela 4.7: Parâmetros do GPFIS-Forecast.

Parâmetro	Valor
Tamanho da população	100
Quantidade de gerações	50
Altura máxima da árvore	5
Tamanho do torneio	2
Taxa de cruzamento de alto nível	50%
Taxa de cruzamento de baixo nível	85%
Taxa de mutação	10%
Taxa de clonagem	5%
Taxa de elitismo	1%
Pressão lexicográfica	Sim

Em uma primeira investigação sobre configurações na etapa de fuzzificação – formato e número de funções de pertinência), obteve-se um p-valor = 0.020588, que evidencia uma diferença estatística nos resultados proporcionados pelas configurações mostradas na Tabela 4.8. O ordenamento favorece levemente o formato Triangular, granularidade 3 e uma distribuição fortemente particionada no universo de discurso (tri3). A função de pertinência triangular, além de ter se mostrado mais eficaz nas aplicações consideradas neste trabalho, é também menos custosa computacionalmente do que a função de pertinência gaussiana.

Em uma análise par a par da configuração tri3 com as demais, ocorrem sete empates. Entre estes, a combinação granularidade 3, formato triangular e universo de discurso dividido pelo método Tukey(tukey3) apresenta o menor SMAPE médio (destaque em negrito na Tabela 4.9). Assim, esta configuração foi mantida fixa nos experimentos subsequentes.

Em uma segunda investigação, testaram-se três t-normas e dois modificadores, em um total de doze experimentos (Tabela 4.10). O p-valor = 0.48941 indica que não se pode afirmar que existam diferenças significativas nos resultados. O ordenamento favorece a combinação da t-norma Lukasiewicz com os modificadores de Concentração e de Diluição. Contudo a combinação

Tabela 4.8: Etapa de Fuzzificação: Ordenamento pelo Teste de Friedman

Algorithm	Ranking
tri3	4.1364
tri5	6.1364
tri7	6.2727
Gauss3	8.3636
Gauss5	7.6364
Gauss7	9.2727
Tukey3	4.4091
Tukey5	4.6818
Tukey7	7.3182
TG3	6.5455
TG5	6.7273
TG7	6.5

Tabela 4.9: Influência de Parâmetros: Fuzzificação

Formato	Distribuição	Granularidade	SMAPE 11
Triangular	Fort. Parcit.	3	20,51%
Triangular	Fort. Parcit.	5	23,35%
Triangular	Fort. Parcit.	7	22,50%
Gaussiana	Fort. Parcit.	3	28,42%
Gaussiana	Fort. Parcit.	5	24,34%
Gaussiana	Fort. Parcit.	7	24,66%
<b>Triangular</b>	<b>Tukey</b>	<b>3</b>	<b>17,81%</b>
Triangular	Tukey	5	21,05%
Triangular	Tukey	7	22,74%
Gaussiana	Tukey	3	22,80%
Gaussiana	Tukey	5	23,53%
Gaussiana	Tukey	7	23,77%

Lukasiewicz com o modificador de Diluição mostra-se mais acurado e na comparação par a par estabelece-se um empate. Esta última configuração foi a escolhida para a etapa de Formulação.

É importante notar que a inclusão de modificadores fornece ao modelo a possibilidade de corrigir o formato das funções de pertinência, aumentando o número de possibilidades de proposições fuzzy.

Na análise de diferentes configurações para a etapa de Associação, obteve-se um p-valor = 0,004201 – existe diferença significativa. De acordo com o ordenamento, o método MQR é o mais indicado. A análise par a par produz um empate do MQR com a Divisão Uniforme e com o FCD. Sendo o MQR o mais acurado (Tabela 4.11), ele é mantido na próxima investigação.

Na quarta investigação, os métodos de agregação resultam em um p-valor = 0,852919: não há indícios de superioridade de algum método. A análise par

Tabela 4.10: Influência de Parâmetros: Formulação

T-normas	Modificadores	SMAPE 11
Produto	Nenhum	17,81%
Mínimo	Nenhum	20,38%
Lukasiewicz	Nenhum	18,06%
Produto	Concentrador	17,56%
Mínimo	Concentrador	17,73%
Lukasiewicz	Concentrador	17,90%
Produto	Diluidor	19,30%
Mínimo	Diluidor	21,73%
<b>Lukasiewicz</b>	<b>Diluidor</b>	<b>17,17%</b>
Produto	Concentrador / Diluidor	17,61%
Mínimo	Concentrador / Diluidor	17,82%
Lukasiewicz	Concentrador / Diluidor	17,27%

Tabela 4.11: Influência de Parâmetros: Associação

Associação	SMAPE 11
FCD	17,02%
Divisão Uniforme	17,00%
<b>MQR</b>	<b>16,70%</b>
PFCF	18,06%
Máximo dos Mínimos	18,76%

a par evidencia que a alternativa MQR iterativa é a mais adequada, mas há empate com o MQR, que possui a melhor acurácia (Tabela 4.12). Este é então mantido na etapa seguinte.

Tabela 4.12: Influência de Parâmetros: Agregação

Agregação	SMAPE 11
<b>MQR</b>	<b>16,19%</b>
Máximo MQR	18,38%
MQR iterativo	16,29%

Na quinta investigação, os tipos de configuração para defuzzificação produzem um  $p$ -valor = 0,002772, evidenciando diferenças estatísticas significativas. Pelo ordenamento, o método da Altura é o mais adequado. Quanto a acurácia (Tabela 4.13), este mesmo método permanece o mais adequado, sendo então mantido na investigação subsequente.

Finalmente, na etapa de Avaliação, o  $p$ -valor é 0.035405, indicando uma diferença significativa. O ordenamento indica que a avaliação pelo SMAPE é a mais adequada, o que é confirmado pelos resultados apresentados na Tabela 4.14. Vale ressaltar que a pequena granularidade impede que os LPFs sejam mais estreitos, o que torna o fator Pile, que computa a largura do LPF, pouco

Tabela 4.13: Influência de Parâmetros: Defuzzificação

Defuzzificação	SMAPE 11
<b>Altura</b>	<b>16,19%</b>
Altura Corrigida	23,80%
MoM	17,93%
CoA	29,15%

significativo para otimização do PISMAPE. Quanto mais dividido é o espaço de discurso (granularidade grande), melhores são as chances de composições de conjuntos de saída gerarem intervalos estreitos, isso porque a agregação ocorre entre conjuntos cujos suportes são menores.

Tabela 4.14: Influência de Parâmetros: Avaliação

Avaliação	SMAPE 11
<b>SMAPE</b>	<b>16,19%</b>
PISMAPE LPF Suporte	18,38%
PISMAPE LPF Sigma	20,63%

Estabelecida a configuração final (Tabela 4.15), a validação será realizada com 334 séries temporais da categoria Industrial da competição M3.

Tabela 4.15: Configuração Base do GPFIS-Forecast

Configuração Base	
Formato	Triangular
Distribuição	Tukey
Granularidade	3
T-norma	Lukasiewicz
Modificadores	Diluidor
Associação	MQR
Agregação	MQR
Defuzzyficação	Altura
Avaliação	SMAPE

#### 4.4.4

##### Validação da Configuração

O GPFIS-Forecast alcança a nona posição nas duas categorias da competição NN3, conforme apresentado nas Tabelas 4.16 e 4.17.

Considerando as séries da competição M3, o GPFIS-Forecast atingiu o vigésimo segundo lugar (Tabela 4.18).

Os resultados obtidos para as séries da competição M3 demonstram a dificuldade do GPFIS-Forecast de realizar previsões com alta acurácia em séries

Tabela 4.16: Dez primeiros colocados na competição NN3, considerando os resultados referentes às últimas 11 séries temporais.

Rank	Participante	SMAPE 11	Média de regras	Número médio de elementos nas premissas
-	<i>CI Benchmark</i> - Theta AI (Nikolopoulos)	13,07%		
-	<i>Stat. Benchmark</i> - Autobox (Reily)	13,49%		
-	<i>Stat. Benchmark</i> - ForecastPro (Stellwagen)	13,52%		
1	Yan	13,68%		
-	<i>Stat. Benchmark</i> - Theta (Nikolopoulos)	13,70%		
2	Ilies, Jager, Kosuchinas, Rincon, Sakenas e Vaskevcius	14,26%		
3	Chen e Yao	14,46%		
4	Yousefi, Miromeni e Lucas	14,49%		
5	Ahmed, Atiya, Gayar e El-Shishiny	14,52%		
6	Flores, Anaya, Ramirez e Morales	15,00%		
7	Adeodato, Vasconcelos, Arnaud, Chunha e Monteiro	15,10%		
	<i>Stat. Contender - Wildi</i>	15,32%		
8	Luna, Soares e Ballini	15,35%		
9	Theodosiou e Swamy	16,19%		
<b>9</b>	<b>GPFIS-Forecast</b>	<b>16,19%</b>	15,4	1,4

Tabela 4.17: Melhores resultados da NN3 para 111 séries.

Rank	Participante	SMAPE 111	Média de regras	Número médio de elementos nas premissas
-	<i>Stat. Contender - Wildi</i>	14,84%	-	-
-	<i>Stat. Benchmark</i> - Theta Method (Nikolopoulos)	14,89%	-	-
1	Ilies, Jäger, Kosuchinas, Rincon, Sakenas, Vaskevcius	15,18%	-	-
-	<i>Stat. Benchmark</i> - ForecastPro (Stellwagen)	15,44%	-	-
-	<i>CI Benchmark</i> - Theta AI (Nikolopoulos)	15,66%	-	-
-	<i>Stat. Benchmark</i> - Autobox (Reilly)	15,95%	-	-
2	Adeodato, Vasconcelos, Arnaud, Chunha, Monteiro	16,17%	-	-
3	Flores, Anaya, Ramirez, Morales	16,31%	-	-
4	Chen, Yao	16,55%	-	-
5	D'yakonov	16,57%	-	-
6	Kamel, Atiya, Gayar, El-Shishiny	16,92%	-	-
7	Abou-Nasr	17,54%	-	-
8	Theodosiou, Swamy	17,55%	-	-
-	<i>CI Benchmark</i> - Naive MLP (Crone)	17,84%	-	-
9	<b>GPFIS-Forecast</b>	<b>17,96%</b>	<b>15,45</b>	<b>1,5</b>
10	de Vos	18,24%	-	-

Fonte: Retirado do site da NN3.

temporais com fortes tendências. A NN3 apresenta séries com tendências mais suaves em comparação com as da M3, o que justifica o melhor desempenho do modelo no primeiro estudo de caso. A opção por não realizar um pré-processamento, de forma a não comprometer a interpretabilidade do modelo resultante, tem, como seria de se esperar, um impacto na acurácia.

Os experimentos sequenciais realizados não servem para descartar outras configurações não testadas do GPFIS-Forecast. Por exemplo, na segunda investigação, a t-norma Lukasiewicz sem modificadores, apresenta resultados inferiores aos de outras t-normas, mas, em contrapartida, sua utilização em conjunto com o modificador de Diluição proporciona resultados melhores que os da as demais t-normas com e sem modificadores.

Tabela 4.18: Ranking M3 para base de dados industriais mensais

Ranking	Métodos de Previsão	SMAPE	Média de regras	Número médio de elementos nas premissas
1	Forecastpro	11,61 %	-	-
2	ForecastX	11,73 %	-	-
3	B-J automatic	11,94 %	-	-
4	RBF	12,12 %	-	-
5	Dampen	12,13 %	-	-
6	THETA	12,22 %	-	-
7	Single	12,32 %	-	-
8	Comb S-H-D	12,35 %	-	-
9	Autobox1	12,82 %	-	-
10	THETA <sub>sm</sub>	12,87 %	-	-
11	Autobox2	12,98 %	-	-
12	ARARMA	13,01 %	-	-
13	Naive2	13,20 %	-	-
14	SMARTFCS	13,25 %	-	-
15	Auto-ann	13,30 %	-	-
16	Flors-Pearc2	13,32 %	-	-
17	Winter	13,37 %	-	-
18	Holt	13,55 %	-	-
19	Autobox3	13,58 %	-	-
20	PP-Autocast	13,78 %	-	-
21	Robust Trend	13,92 %	-	-
22	<b>GPFIS-Forecast</b>	<b>14,47 %</b>	<b>16</b>	<b>1,45</b>
23	Flores/Pearce1	14,71 %	-	-

De uma maneira geral, a Fuzzificação e a Formulação mostraram-se pontos estratégico na obtenção de bons resultados. Essas investigações demonstraram como a aquisição das novas configurações podem trazer ganhos consideráveis, mesmo sem testar todas as possibilidades de configurações conjuntas com outras etapas do modelo.

#### 4.4.5

##### Interpretabilidade do Modelo

O GPFIS-Forecast fornece ao usuário uma base de regras a partir da série temporal disponibilizada para previsão. A Tabela 4.19 mostra um exemplo de base de regras para a série temporal 102 da NN3. As regras linguísticas apresentam poucos elementos nas premissas (vide Tabela 4.18, por exemplo), são de fácil entendimento e podem servir de guia para análises do comportamento do processo gerador da série.

Os nomes dos termos utilizados para granularidade 3 são: Alto, Médio e Baixo. Para granularidade 5: Bem Baixo, Baixo, Médio, Alto, Bem Alto. Por fim para granularidade 7: Baixíssimo, Bem Baixo, Baixo, Médio, Alto,

Tabela 4.19: Exemplo de Base de Regras para granularidade 5

ID	Regra
1	Se o Período t-10 for Muito Baixo <b>Então</b> a Previsão é Média.
2	Se o Período t-10 for Bastante Alto <b>Então</b> a Previsão é Alto.
3	Se o Período t-11 for médio <b>Então</b> a Previsão é Baixa.
4	Se o Período t-11 for Muito Muito Bem Baixo <b>Então</b> a Previsão é Bem Baixa.
5	Se o Período t-1 for Pouco Bem Alto <b>Então</b> a Previsão é Bem Alto.

Bem Alto, Altíssimo. Os modificadores (Muito ou Pouco), como visto na Tabela 4.19, podem se repetir para fornecer a ênfase necessária para gerar os resultados.

#### 4.4.6 Papel dos LPFs

Os LPFs fornecem informação (visual) sobre o intervalo de previsão: região possível para valor real em um determinado período. Ilustram, também, o comportamento das regras no decorrer do processo de previsão e identificam se o método de defuzzificação é o mais adequado para traduzir o resultado da base de regras em questão.

Um espaço de discurso mais dividido, proporcionando pela granularidade alta, fornece uma variedade maior de opções para composição de conjuntos fuzzy de saída. Essa condição facilita, na etapa de Agregação, a geração de um conjunto cujo suporte é uma parte do universo de discurso; os limites deste suporte podem servir de parâmetros para computar LPFs mais informativos, como observado nas curvas verde e vermelha da Figura 4.1.

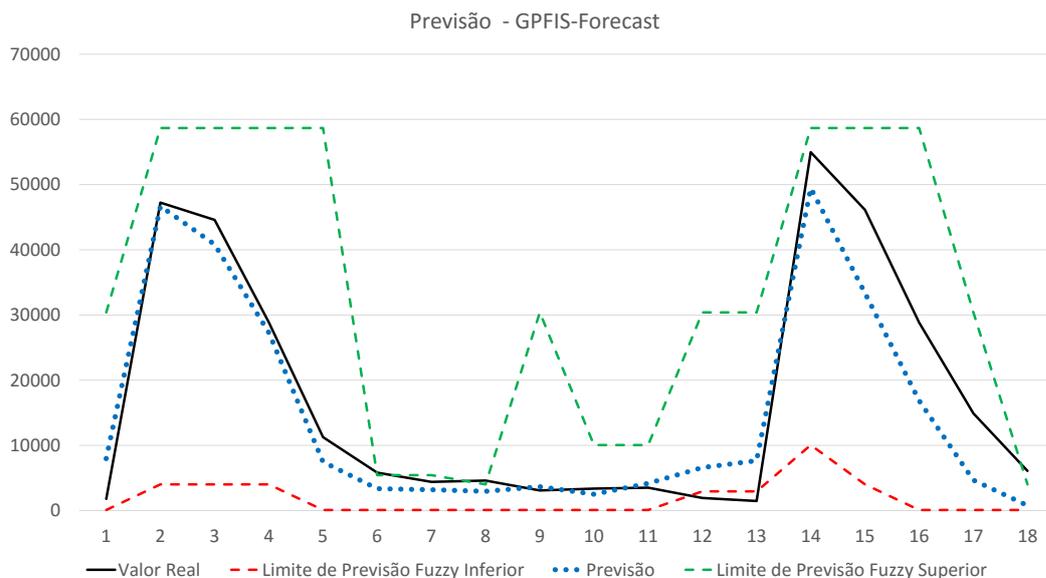


Figura 4.1: Previsão da Série 103 da Competição NN3: Granularidade 7

A largura dos LPFs mostra como as regras estão interagindo entre si. Se a granularidade é alta e os intervalos são muito largos, há regras acionadas paralelamente e a consequente geração de um suporte grande. A Figura 4.1 ilustra um caso de granularidade é 7 e 23 regras. Em  $t = 2$ , a largura dos LPFs é grande, evidenciando que há mais de uma regra acionada para aquela situação. A defuzzificação gera uma previsão boa, evidenciando que as regras acionadas possuem uma boa sinergia.

Na Figura 4.2, os Limites de Previsão Fuzzy mostram que as regras estão formando conjuntos fuzzy de saída com suporte largo em alguns casos. O acionamento das regras em paralelo pode ter gerado conflitos no resultado, ou o método de defuzzificação não é o mais adequado para gerar previsões pontuais neste caso. É possível que a eliminação de algumas regras (22, neste caso), torne o modelo mais acurado.

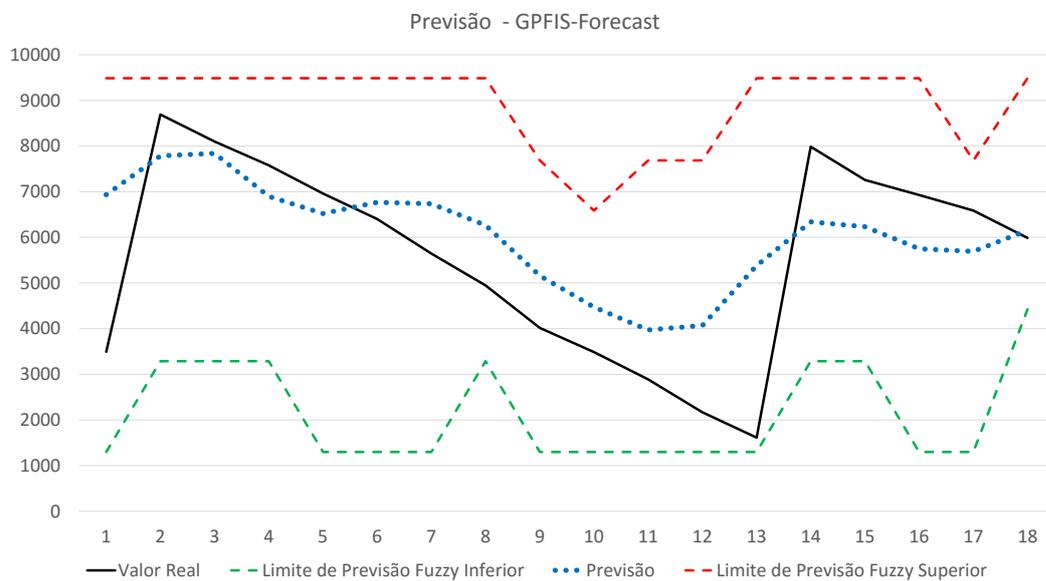


Figura 4.2: Previsão da Série 102 da Competição NN3

Quando a granularidade é baixa, há poucas opções de combinação e a Agregação gera em quase todas as previsões um conjunto fuzzy de saída cujo suporte é o universo de discurso inteiro. Logo a informação relevante que se pode obter com tais LPFs é geralmente inexistente, como observado na situação de granularidade 3 exposta na Figura 4.3. Neste caso, há 11 regras e em todos os 18 passos de previsão os LPFs ocupam todo o universo de discurso. Ou seja, o modelo considera possível que o valor real esteja em qualquer lugar.

Como visto nas subseções anteriores, os melhores resultados para as séries consideradas foram obtidos para granularidades baixas, em que o uso de LPFs têm pouca relevância. No entanto, para situações em que granularidades mais altas proporcionem melhores resultados, os LPFs podem ser de utilidade.

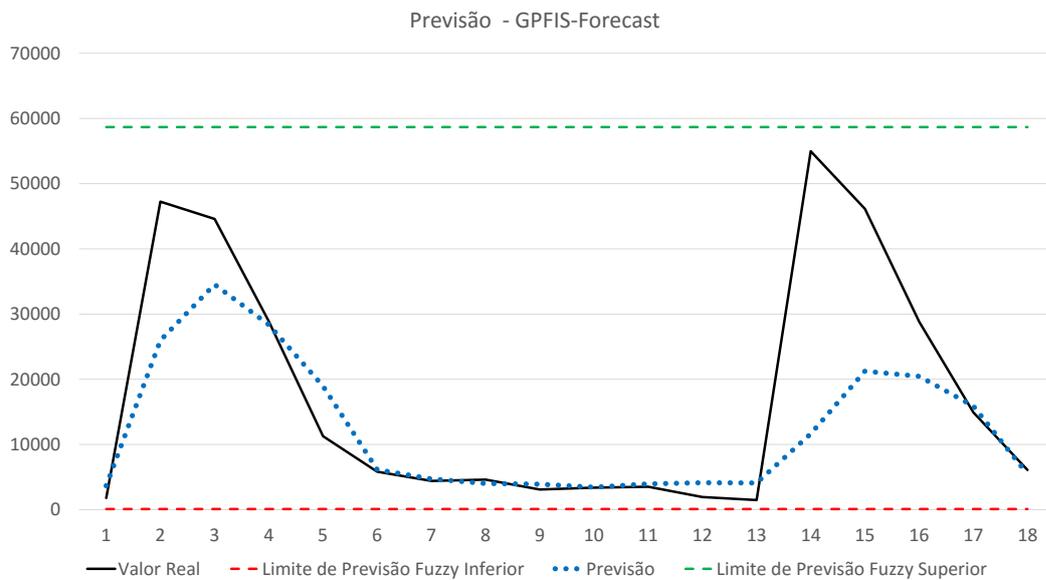


Figura 4.3: Previsão da Série 103 da Competição NN3: Granularidade 3

Quando a granularidade é alta, maior é o espaço de busca da PGMG, e a base de regras tende a ser mais extensa e a apresentar conflitos. Os LPFs podem servir de auxílio para eliminar regras e conflitos.

Na etapa de Avaliação, o PISMAPE leva em consideração: (1) o valor do SMAPE da previsão; (2) se o valor real está contido no intervalo entre os LPFs e; (3) a largura dos LPFs. Estes três termos são ponderados pelos valores 70%, 15% e 15%, respectivamente. Uma sintonia destes pesos pode servir para refinar o processo de evolução.

Em resumo, os LPFs representam um primeiro passo para apontar possibilidades de melhoria dos resultados para casos de granularidade alta.

## 5

### Conclusões e Trabalhos Futuros

Esta dissertação buscou avaliar diversas configurações possíveis do modelo GPFIS aplicado à previsão de séries. Implementaram-se informações visuais sempre tendo em perspectiva um compromisso entre a acurácia e a interpretabilidade das regras geradas.

Os estudos de caso demonstraram que, para séries com pouca tendência e sem a realização de um pré-processamento, o GPFIS-Forecast é capaz de proporcionar resultados com boa acurácia. No caso de séries com forte tendência, um pré-processamento faz-se necessário, o que implica em perda de interpretabilidade linguística das regras geradas.

Os Limites de Previsão Fuzzy introduzidos neste trabalho complementam o resultado da previsão pontual, apontando possíveis ajustes finais na base de regras para situação de alta granularidade. Sua inclusão na etapa de Avaliação – através do PISMAPE – foi uma primeira tentativa de refinamento do processo evolutivo.

No que tange a trabalhos futuros, pode-se:

- desenvolver uma forma de calibração automática dos parâmetros do GPFIS-Forecast por meio de um algoritmo genético.
- investigar o uso de diferentes tipos de métricas para avaliação, tais como o RMSE.
- investigar outras formas de tradução dos resultados fuzzy gerados pelo GPFIS-Forecast, a exemplo do realizado com os LPFs.
- testar os Limites de Previsão Fuzzy em outros tipos de SFG( Michigan e Iterative Learning) para seleção de regras.
- utilizar uma rede neural para aprender padrões de configurações de parâmetros do GPFIS-Forecast de acordo com as características de uma série temporal.
- verificar a frequência de determinados tipos de regras para certas variedades de séries temporais. Poder-se-ia então gerar uma base de regras inicial que funcionasse como *feed* para alguns indivíduos da geração zero da PGMG. Tal procedimento poderia diminuir ainda mais a quantidade de gerações necessárias para a PGMG convergir a valores satisfatórios.

## Referências Bibliográficas

- Aznarte, J. L., Alcalá-Fdez, J., Arauzo-Azofra, A., e Benítez, J. M. (2012). *Financial time series forecasting with a bio-inspired fuzzy model*, volume 39.
- Bergmeir, C., Triguero, I., Molina, D., Aznarte, J. L., e Benitez, J. M. (2012). *Time Series Modeling and Forecasting Using Memetic Algorithms for Regime-Switching Models*, volume 23.
- Berlanga, F., Rivera, A., del Jesus, M., e Herrera, F. (2010). *GP-COACH: Genetic Programming-based learning of COmpact and ACcurate fuzzy rule-based classification systems for High-dimensional problems*, volume 180.
- Box, G. E., Jenkins, G. M., e Reinsel, G. C. (2008). *Time series analysis: forecasting and control*. John Wiley & Sons, New Jersey.
- Brown, R. G. ; Meyer, R. F. (1961). *The fundamental theorem of exponential smoothing*.
- Calvo, T., Kolesárová, A., Komorníková, M., e Mesiar, R. (2002). *Aggregation Operators: Properties, Classes and Construction Methods*, volume 97 of *Studies in Fuzziness and Soft Computing*. Physica-Verlag HD.
- Hadavandi, E., Shavandi, H., e Ghanbari, A. (2011). *An improved sales forecasting approach by the integration of genetic fuzzy systems and data clustering: Case study of printed circuit board*, volume 38.
- Herrera, F. (2008). *Genetic fuzzy systems: taxonomy, current research trends and prospects*, volume 1.
- Herrera, F. e Magdalena, L. (1997). *Genetic fuzzy systems: A tutorial*, volume 13.
- Hinchliffe, M., Hiden, H., McKay, B., Willis, M., Tham, M., e Barton, G. (1996). *Modelling Chemical Process Systems Using a Multi-Gene Genetic Programming Algorithm*. Stanford University, Stanford University, CA, USA.
- Holt, C. C. (1957). *Forecasting seasonal and trends by exponentially weighted moving averages*. Carnegie Institute of Technology, Pittsburgh, Pennsylvania.

- Ishibuchi, H. e Nojima, Y. (2011). Toward quantitative definition of explanation ability of fuzzy rule-based classifiers. Em *2011 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, p 549–556.
- Kim, D. e Kim, C. (1997). *Forecasting time series with genetic fuzzy predictor ensemble*, volume 5. IEEE.
- Kirchgässner, G., Wolters, J., e Hassler, U. (2012). *Introduction to Modern Time Series Analysis*. Springer Texts in Business and Economics. Springer Berlin Heidelberg.
- Koshiyama, A. S. (2014). *GPFIS: Um Sistema Fuzzy-Genético Genérico baseado em Programação Genética*. Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro, 228p., Dissertação de Mestrado, Rio de Janeiro.
- Koshiyama, A. S., Escovedo, T., Dias, D. M., Vellasco, M. M., e Tanscheit, R. (2013). *GPF-CLASS: A genetic fuzzy model for classification*.
- Koshiyama, A. S., Vellasco, M. M., e Tanscheit, R. (2014). *GPFIS-control: a genetic fuzzy system for control tasks*, volume 4.
- Koshiyama, A. S., Vellasco, M. M. B. R., e Tanscheit, R. (2015). *Development of a fuzzy rule-based system using Genetic Programming for Forecasting problems*. IEEE.
- Koza, J. R. (1992). *Genetic programming: on the programming of computers by means of natural selection*. The MIT Press, Massachusetts.
- Krajewski, L., Ritzman, L., e Malhotra, M. (2012). *Administração da Produção e Operações*. Pearson Education.
- Lu, W., Yang, J., e Liu, X. (2014). *Numerical Prediction of Time Series Based on FCMs with Information Granules*, volume 9.
- Luke, S. e Panait, L. (2002). *Lexicographic Parsimony Pressure*. Morgan Kaufmann Publishers, New York.
- M3 (2015). *M Competition*.
- Makridakis, S. (1976). *A survey of time series*. JSTOR.
- Mangai, S. A., Kasinathan, S., Alagarsamy, K., e Sankar, B. R. (2014). *Hybrid ARIMA-HyFIS Model for Forecasting Univariate Time Series*, volume 91. Foundation of Computer Science (FCS).

- NN3 (2015). *Artificial Neural Network & Computational Intelligence Forecasting Competition*.
- Park, D. C., El-Sharkawi, M., Marks, R., Atlas, L., Damborg, M., e others (1991). *Electric load forecasting using an artificial neural network*, volume 6. IEEE.
- Poli, R., Langdon, W. B., e McPhee, N. F. (2008). *A Field Guide to Genetic Programming*. Lulu.com, Raleigh.
- Roychowdhury, S. e Pedrycz, W. (2001). *A survey of defuzzification strategies*, volume 16. John Wiley & Sons, Inc.
- Schölkopf, B. e Smola, A. J. (2001). *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press, Massachussets.
- Searson, D., Willis, M., e Montague, G. (2007). *Co-evolution of non-linear PLS model components*, volume 21.
- Shaw, I. e SIMOES, M. (2007). *Controle e modelagem fuzzy*. Edgard Blucher.
- Slutsky, E. (1937). *The summation of random causes as the source of cyclic processes*. *Econometrica*, 5.
- Stepnicka, M., Stepnickova, L., e Burda, M. (2014). *Fuzzy rule-based ensemble for time series prediction: The application of linguistic associations mining*.
- Trawinski, B., Smetek, M., Telec, Z., e Lasota, T. (2012). *Nonparametric statistical analysis for multiple comparison of machine learning regression algorithms*, volume 22.
- Walker, A. M. (1931). *On the periodicity in series of related terms*. *Proceedings of the Royal Society of London*, A, 131.
- Walker, M. (2013). *Introduction to Genetic Programming*.
- Wang, L.-X. e Mendel, J. M. (1992). *Generating fuzzy rules by learning from examples*, volume 22. IEEE.
- Winters, P. R. (1960). *Forecasting sales by exponentially weighted moving averages*. *Management Science*.
- Wold, H. (1938). *A Study in the Analysis of Stationary Time Series*. Almqvist and Wiksell.

Yule, G. U. (1927). *On the time-correlation problem with special reference to the variate difference correlation method*. Journal of the Royal Statistical Society, 84.

Zhang, G., Patuwo, B. E., e Hu, M. Y. (1998). *Forecasting with artificial neural networks: The state of the art*, volume 14. Elsevier.