



UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE CIÊNCIAS EXATAS E NATURAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

BENEDITO DE SOUZA RIBEIRO NETO

APLICAÇÃO DE SENSORES VIRTUAIS NA ESTIMAÇÃO DA
CONCENTRAÇÃO DOS PARÂMETROS FÍSICO-QUÍMICOS E
METAIS EM CORPOS D'ÁGUA DE RESERVATÓRIOS DE
HIDRELÉTRICAS: UM ESTUDO DE CASO NA REGIÃO
AMAZÔNICA.

UFPA / ICEN / PPGCC
Campus Universitário do Guamá
Belém-Pará-Brasil
2014

BENEDITO DE SOUZA RIBEIRO NETO

APLICAÇÃO DE SENSORES VIRTUAIS NA ESTIMAÇÃO DA
CONCENTRAÇÃO DOS PARÂMETROS FÍSICO-QUÍMICOS E
METAIS EM CORPOS D'ÁGUA DE RESERVATÓRIOS DE
HIDRELÉTRICAS: UM ESTUDO DE CASO NA REGIÃO
AMAZÔNICA.

Dissertação submetida à Banca
Examinadora do Programa de Pós-
Graduação em Ciência da Computação da
UFPA para a obtenção do Grau de Mestre
em Ciência da Computação.

Orientador: Prof. Dr. Antônio Morais da
Silveira (PPGCC/UFPA)

Co-Orientadora: Prof^ª. Dra. Terezinha
Ferreira de Oliveira (PPME/UFPA)

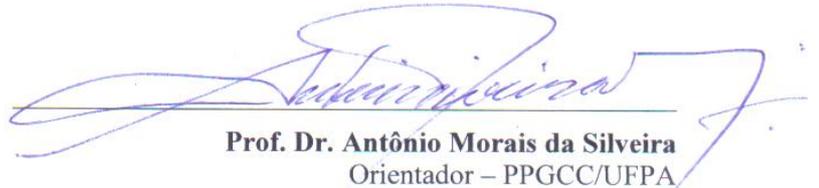
UFPA / ICEN / PPGCC
Campus Universitário do Guamá
Belém-Pará-Brasil
2014

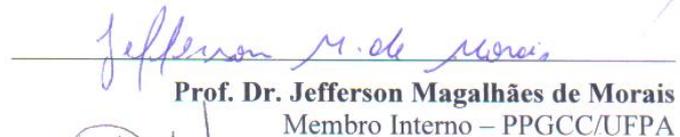
UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE CIÊNCIAS EXATAS E NATURAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

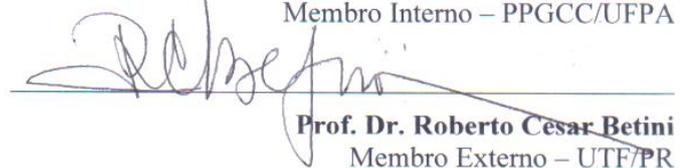
BENEDITO DE SOUZA RIBEIRO NETO

**APLICAÇÕES DE Sensores Virtuais na Estimação da
Concentração dos Parâmetros Físico-Químicos e Metais em
Corpos d' Água do Reservatório de Hidrelétrica: Um Estudo
de Caso na Região Amazônica**

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal do Pará como requisito para obtenção do título de Mestre em Ciência da Computação, defendida e aprovada em 23/10/2014, pela banca examinadora constituída pelos seguintes membros:


Prof. Dr. Antônio Morais da Silveira
Orientador – PPGCC/UFPA


Prof. Dr. Jefferson Magalhães de Morais
Membro Interno – PPGCC/UFPA


Prof. Dr. Roberto Cesar Betini
Membro Externo – UTF/PR

Visto: Nelson Cruz Sampaio Neto.
Prof. Dr. Nelson Cruz Sampaio Neto
Coordenador do PPGCC/UFPA
Prof. Dr. Nelson Cruz Sampaio Neto
Coordenador do PPGCC
Mat. SIAPE: 2659210

AGRADECIMENTOS

Agradeço primeiramente ao bom Deus por ter me dado força e saúde para realizar este trabalho e a todos que me auxiliaram durante o desenvolvimento deste trabalho, em especial à minha mãe Clívia Rodrigues Mendonça, à minha co-orientadora prof^a. Dra. Terezinha Ferreira de Oliveira, ao meu orientador prof. Dr. Antônio Morais da Silveira e a CAPES por ter me beneficiado com Bolsa de estudos.

RESUMO

Esta pesquisa apresenta a utilização de sensores virtuais na estimação da concentração dos parâmetros físico-químicos e metais no monitoramento da qualidade da água de reservatórios da Amazônia, por meio de redes neurais artificiais (RNA's) e imagens de sensoriamento remoto. A análise fatorial das variáveis consideradas no estudo, confirmou o relacionamento da primeira faixa de profundidade do disco de Secchi, Ferro Total, PO₄, P Total, TSS e Turbidez em um único fator, pois estes apresentam grande refletância e boa absorção de energia pelos sensores dos satélites. Estes elementos foram estimados por RNA's, produzindo resultados bem próximos dos valores observados. Os testes de validação também confirmaram a boa aproximação para a primeira faixa de profundidade do disco de Secchi, Fe Total, STS e Turbidez do reservatório de água. No caso específico dos parâmetros PO₄ e P Total, além do problema do pouco número de estações de coleta de dados disponíveis e a variabilidade inerente ao ciclo hidrológico da região, constatou-se, por meio da interpretação das imagens, falta de similaridades entre os dados utilizados no treinamento e na validação da RNA. De modo geral, o estudo demonstrou a eficácia da aplicação de sensores virtuais e imagens de satélite no monitoramento da qualidade da água de reservatórios na Amazônia, proporcionando uma alternativa com bom índice de precisão e menos dispendiosa de recursos no processo de monitoramento desse recurso.

Palavras-chaves: Reservatório de Hidrelétrica, Análise Fatorial, Redes Neurais Artificiais, Parâmetros Físico-químicos e Metais, Sensoriamento remoto.

ABSTRACT

This research introduces the use of virtual sensors to estimate the concentration of physico-chemical parameters and metals in monitoring water quality of reservoirs Amazon through artificial neural networks (ANN) and images of remote sensing. A factor analysis of the variables considered in the study confirmed the relationship of the first factor with Secchi disk, Total Iron, PO₄, Total P, TSS and Turbidity on a single factor, as these have a high reflectance and good energy absorption by satellite sensors. These elements were determined by ANN's, producing satisfactory results approach 100% between observed and estimated. The tests resulted in a good approximation, the first band Secchi disk depth, total Fe, STS, and turbidity of the water reservoir. In the specific case of the parameters PO₄ and Total P, besides the problem of the small number of sampling stations available data and the variability inherent in the hydrological cycle of the region, it was found, through the interpretation of images, lack of similarities between the data used in training and validation of RNA. Overall, the study demonstrated the effectiveness of the application of virtual sensors in monitoring water quality of reservoirs in the Amazon by satellite imagery, providing a precise and less expensive alternative resources in the process of environmental monitoring.

Keywords: Hydroelectric Reservoir. Factor Analysis. Artificial Neural Networks. Physicochemical Parameters and Metals. Remote sensoring.

SUMÁRIO

1. INTRODUÇÃO	1
2. O ESTADO DA ARTE SOBRE O TEMA PESQUISADO	4
3. MATERIAIS E MÉTODOS.....	7
3.1. Área de estudo	7
3.2. Caracterização das estações de amostragem no reservatório de Tucuruí	8
3.3. Análise físico-químicas e metais da água	10
3.4. Estações de amostragem e periodicidade das coletas.....	14
3.5. Aquisição de dados espaciais e sensores virtuais	14
3.6. Métodos estatísticos e computacionais.....	15
3.6.1. Análise Fatorial	15
3.6.2. Estimadores M	21
3.6.3. Métodos de Interpolação.....	21
3.6.4. Redes Neurais Artificiais (RNA's).....	21
3.7. Sensoriamento Remoto e Processamento de Imagens	24
4. APLICAÇÃO DA METODOLOGIA E RESULTADOS	27
4.1. Análise Fatorial	27
4.2. Distribuição espacial dos parâmetros físico-químicos e metais	28
4.3. RNA's	35
4.4. Validação dos sensores	38
5. CONCLUSÕES	46
6. TRABALHOS FUTUROS	48
REFERÊNCIAS BIBLIOGRÁFICAS	49

LISTA DE ABREVIações

AF	Análise Fatorial
Fe	Ferro
MLP	Perceptron Multicamadas
OD	Oxigênio dissolvido
P Total	Fósforo Total
pH	Potencial de Hidrogênio
PO ₄	Fósforos
RNA	Rede Neural Artificial
STS	Sólidos Totais em Suspensão
UHE	Usina Hidrelétrica

LISTA DE FIGURAS

Figura 1- Localização do reservatório da usina hidrelétrica de Tucuruí.....	7
Figura 2 - Arquitetura multicamadas do modelo de RNA perceptron	22
Figura 3 - Funções de ativação de RNA's(a) função degrau; (b) função sinal; (c) função sigmoide; (d) função linear; (e) função tangente hiperbólica.....	24
Figura 4- Imagem digital (a) com área em zoom de grupo de pixels em valores de cinza (b) e correspondentes valores digitais (c) (Menezes e Almeida, 2012).....	26
Figura 5- Distribuição espacial do Secchi nas estações de coleta.....	29
Figura 6 - Distribuição espacial das concentrações de Ferro Total nas estações de coleta.....	30
Figura 7- Distribuição espacial das concentrações de Fósforos nas estações de coleta	31
Figura 8 - Distribuição espacial das concentrações de Fósforo Total nas estações de coleta.	32
Figura 9 - Distribuição espacial das concentrações de STS nas estações de coleta.....	33
Figura 10 - Distribuição espacial da Turbidez nas estações de coleta	34
Figura 11 - Arquitetura da RNA utilizada neste estudo	36
Figura 12 - Desempenho do treinamento das RNA's.....	37
Figura 13 - Validação dos Sensores virtuais para a primeira faixa de profundidade do disco de Secchi (m), considerando o ciclo hidrológico da região e o Erro de aproximação.....	40
Figura 14 - Validação dos Sensores virtuais para o Fe total (mg/L) considerando o ciclo hidrológico da região e o Erro de aproximação.....	41
Figura 15 - Validação dos Sensores virtuais para o PO4 (mg/L) considerando o ciclo hidrológico da região e o Erro de aproximação.....	42
Figura 16 - Validação dos Sensores virtuais para o P total (mg/L) considerando o ciclo hidrológico da região e o Erro de aproximação.....	43
Figura 17 - Validação dos Sensores virtuais para o STS (mg/L) considerando o ciclo hidrológico da região e o Erro de aproximação.....	44

Figura 18 - Validação dos Sensores virtuais para Turbidez (NTU) por ciclo hidrológico e o Erro de aproximação. 45

LISTA DE QUADROS

Quadro 1 - Condições operacionais do ICP-OES	11
Quadro 2 – Procedimentos Estatísticos da Análise fatorial.....	16

LISTA DE TABELAS

Tabela 1 - Análise fatorial com os fatores, comunalidades, variância total e acumulada.	27
Tabela 2 - Erros de aproximação do treinamento das RNA's	36
Tabela 3 - Estimador M para os parâmetros nas estações de coletados para cada ano.	38

1. INTRODUÇÃO

As condições hídricas e do relevo brasileiro, aliados à crescente demanda energética têm proporcionado a construção de várias hidrelétricas no país, nas últimas décadas (TUNDISI, 2003; ATOBATELE e UGWUMBA, 2008; PEREIRA FILHO et.al, 2013).

As alterações do ambiente aquático devido à construção de reservatórios, ainda não foram suficientemente avaliadas, uma vez que os parâmetros físico-químicos e os metais de certos corpos de água podem ser característicos de região para região (TUNDISI e TUNDISI, 2008).

Por outro lado, a presença de metais em corpos de água é importante para o ciclo geoquímico e para a biodiversidade. Contudo, elevadas concentrações geram preocupações sobre sua influência em diversos ecossistemas ambientais mais comuns nas águas e em biotas associadas, podendo indicar a presença de fontes naturais ou antropogênicas destes elementos (KABATA-PENDIAS, 2011; WENNER et. al, 2004).

Diversos estudos foram conduzidos com o intuito de identificar e caracterizar variações sazonais nos parâmetros físico-químicos e nas concentrações dos metais em ambientes aquáticos (Wenner et. al, 2004; Ouyang, 2005; Kazi et.al, 2009; Wang at. al., 2009; Zahraie et.al., 2009; Coletti et. al.,2010; Hauser-davis et.al., 2010; Yenilmez et.al.,2011). Dentre as técnicas de monitoramento utilizadas se destacam as que utilizam sensoriamento remoto (Bazi, et.al.,2012; Zhu e Yu, 2013) e as que utilizam técnicas da inteligência computacional (Wang at. al., 2009; Zahraie et.al., 2009; Hauser-davis et.al., 2010; Hauser-davis et.al., 2012) e, o uso combinado de Redes Neurais Artificiais com sensoriamento remoto, como no estudo de Matthews et.al (2010) e Ribeiro Neto et. al (2014).

A utilização do sensoriamento remoto em sistemas aquáticos é uma técnica de monitoramento dos recursos naturais que faz uso de análises das variações da magnitude

e qualidade da radiação que deixa a água, para se obter informações quantitativas do tipo de substâncias presentes na água e suas concentrações. Assim, a determinação e a quantificação dos fatores ópticos da qualidade da água são alcançadas por meio da assinatura das refletâncias espectrais dos elementos contidos nos corpos d'água (BRANDO e DEKKER, 2003; MICHALOPOULOU et.al., 2010).

A aplicação de sensores virtuais, por meio de RNA's, proporciona uma alternativa de boa precisão e, menos dispendiosa que o uso de sensores físicos em processos de monitoramento ambiental, já que estes são sistemas computacionais desenvolvidos, a partir do uso de dados pretéritos, obtidos em laboratório a partir de um planejamento adequado de coleta de amostras de água, permitindo realizar um processo de descoberta de conhecimento na base de dados construída por meio de análises estatísticas e RNA's. Além disso, as RNA's apresentam outras vantagens, como a capacidade para modelar processos não lineares e complexos, aprendizagem adaptativa, auto-organização, funcionamento em tempo real e facilidade de inserção em tecnologias existentes. Esta flexibilidade, proporciona às RNA's uma vantagem competitiva em relação aos modelos matemáticos e experimentos tradicionais (PADHY, 2005).

Neste contexto, este estudo objetivou desenvolver sensores virtuais para estimação da concentração dos parâmetros físico-químicos e ferro total no monitoramento da qualidade da água no entorno das estações de coleta, à montante do reservatório da Usina Hidrelétrica de Tucuruí (UHE) por meio de RNA's e imagens de sensoriamento remoto, fazendo desta uma pesquisa pioneira na aplicação de sensores virtuais no monitoramento da qualidade da água em reservatórios da Amazônia.

Para atingir o objetivo geral deste estudo, precisou alcançar objetivos específicos a seguir:

- Normalizar dados dos parâmetros físico-químicos e metais com o ciclo hidrológico da região.
- Adquirir e processar imagens de satélite da área de estudo
- Analisar as correlações existentes entre os parâmetros físico- químicos e metais com imagens de satélite.
- Gerar matrizes dos parâmetros físico-químicos.

- Gerar matrizes de pixels das imagens do reservatório
- Criar e Treinar redes neurais artificiais multicamadas
- Validar as saídas geradas pelas RNA's

Através deste estudo torna-se possível a aplicação de sensores virtuais para estimação da concentração de metais no monitoramento da qualidade da água de reservatórios da Amazônia através de redes neurais artificiais e imagens de sensoriamento remoto, proporcionando aos pesquisadores uma nova abordagem de monitoramento da qualidade da água, melhorando o processo de tomada de decisão para uma melhor gestão dos recursos hídricos.

Além deste capítulo introdutório esta dissertação está organizada em 5 capítulos, com os seguintes conteúdos:

Capítulo 2 apresenta o estado da arte sobre o tema.

Capítulo 3 especifica os materiais e métodos utilizados na pesquisa, destacando-se o lócus da pesquisa, a área de estudo, as estações de coleta de amostras e periodicidade das coletas, o processo de análise dos parâmetros físico-químicos e metais, o processo de aquisição de dados espaciais e sensores virtuais, os métodos estatísticos e computacionais utilizados, bem como o processamentos das imagens.

Capítulo 4 descreve a aplicação da metodologia e análise dos resultados.

Capítulo 5 apresenta as conclusões do trabalho e, no capítulo 6 são propostos trabalhos futuros.

2. O ESTADO DA ARTE SOBRE O TEMA PESQUISADO

Pode-se constatar pela literatura que diversos estudos foram desenvolvidos com a aplicação de sensores virtuais voltadas para o monitoramento e estimação de processos industriais por meio das técnicas da inteligência computacional. Dentre esses estudos pode-se destacar:

Rivera et al. (2009) desenvolveram um sensor virtual baseado em redes neurais *perceptron* de múltiplas camadas (MLP) para o monitoramento da concentração de biomassa, etanol e substrato durante a produção de bioetanol. Observou-se que o sistema de monitoramento pode contribuir para a implementação de estratégias avançadas de otimização e de controle para alcançar um maior desempenho operacional.

Gonzaga et al. (2009) desenvolveram um sensor virtual baseado em redes neurais artificiais para a predição da viscosidade do polietileno tereftalato (PET) durante o processo de polimerização. Foi demonstrado que o sensor virtual pode inferir corretamente a viscosidade do PET usando variáveis do processo que podem ser facilmente medidas.

Demori et, al. (2010) propuseram um sistema de sensor para o estudo do fluxo de óleo-água em tubos, estimando a fração de área ocupada por determinado fluido numa dada secção do tubo, aproveitando-se da diferença de constante dielétrica entre os fluidos. A estimativa é feita por meio de medições de capacitância entre os eletrodos de descarga, montados na superfície exterior de uma secção do tubo não condutor. Os resultados obtidos a partir do sistema de sensor desenvolvido para diferentes frações de fluido foram comparados com os obtidos pela válvula de fecho rápido (QCV), técnica utilizada como uma referência.

Morais et.al (2010) propõem uma estratégia para construção de um soft sensor, que envolve desde o processo de aquisição de dados, análise de sensibilidade das variáveis de

entrada, construção de um modelo inferencial caixa-preta que utiliza técnica *neuro-fuzzy*, até a validação do modelo. A estratégia foi desenvolvida utilizando dados de uma planta piloto de destilação, que têm como objetivo separar o metanol da água numa coluna com 10 bandejas. O processo destilação mostrou-se interessante para o presente estudo, uma vez que apresenta dificuldades na obtenção da composição do produto. A estratégia computacional construída mostrou-se adequada e produziu bons resultados na estimativa da composição.

KUSIAK, et. al. (2010) apresentam uma abordagem orientada a dados para a modelagem de sensores de qualidade do ar indoor utilizados em sistemas de aquecimento, ventilação e ar condicionado. Os sensores utilizam três parâmetros básicos, temperatura, CO₂ e umidade relativa. Os modelos construídos com algoritmos de mineração de dados, podem servir como sensores virtuais da qualidade do ar em edifícios e serem usado para monitoramento on-line e calibração dos sensores.

Mohler et al. (2010), desenvolveram sensores virtuais, baseados em redes neurais MLP para estimar o ponto final de destilação do óleo diesel bruto. Foi observado que os sensores virtuais satisfizeram o objetivo proposto e, portanto, podem ser usados para a estimação contínua de propriedades de óleo diesel com densidade, viscosidade, ponto de ignição, tendência à corrosão, entre outras.

Yap e Karri (2011), apresentam o uso de redes neurais artificiais na construção de sensores virtuais para previsão de emissões e controle de um motor a gasolina. Emissões de escape e os parâmetros do motor foram mensurados primeiro experimentalmente para formar um banco de dados abrangente para treinamento da rede e testes. Este trabalho demonstrou uma metodologia viável e precisa de previsão e controle de emissões. Através da aplicação de modelos de sensores virtuais, a necessidade de medição adicional, complicada e onerosa e os dispositivos de monitoramento podem ser eliminadas.

Valdan et al. (2011), desenvolveram um sensor virtual baseado em redes MLP para prever a concentração da emissão de gases resultantes da combustão de misturas de biodiesel/diesel em uma caldeira; observaram que a aplicação do sensor virtual é eficiente no processo de monitoramento e, pode ser implementada nas estratégias de controle e otimização do processo.

Hedegard e Wik (2011), desenvolveram um sensor virtual baseado em Filtro de Kalman Estendido (FKE) para a estimativa online de substratos biodegradáveis em um processo de lodo ativado. O estimador usou padrões de medições como concentração de oxigênio e fluxos de água e ar e se mostrou como uma alternativa robusta contra erros de parâmetro e de medição de ruído.

Gómez e Monteagudo (2012), desenvolveram um Sensor Virtual baseado em RNA aplicado ao estudo de turbo gerador de 4 Mega Watts de uma empresa geradora de energia, estimando a temperatura do rotor do turbo gerador para futuras manutenções .

Yap et.al (2012), apresentam uma ferramenta alternativa para aplicações de ajuste de veículos, incorporando o uso de redes neurais artificiais (RNA's), sensores virtuais para um carro movido a hidrogênio. O desempenho e precisão dos sensores virtuais propostos foram considerados aceitáveis com os erros relativos médios preditivos máximos de 0,65%. Em função da capacidade de previsão obtida, os sensores virtuais foram empregados e simulados como uma ferramenta de medição para o ajuste do veículo e otimização.

Yap e Karri (2013), apresentam uma comparação de modelos preditivos para a estimativa da potência do motor e emissões de escape de uma motocicleta a gasolina *scooter* de 4 kW. Três modelos preditivos de redes neurais artificiais (RNA's) foram investigados neste estudo; modelos diretos, séries redes neurais artificiais e um modelo dinâmico. Este estudo constitui um ponto de referência para o estabelecimento de um controle de emissões de combustão on-line e um sistema de monitoramento para manter as emissões dentro de limites específicos.

Apesar de serem pesquisas que demonstram a aplicação de sensores virtuais, nenhum dos estudos tratam da aplicação desta técnica ao monitoramento da qualidade da água, principalmente em reservatórios da Amazônia.

3. MATERIAIS E MÉTODOS

3.1. Área de estudo

Para este estudo foi escolhido o reservatório da usina hidrelétrica de Tucuruí, um dos maiores do Brasil, sendo considerado um reservatório profundo, com profundidade máxima de 77 m e profundidade média de 19,8 m (ELETRONORTE, 1988).

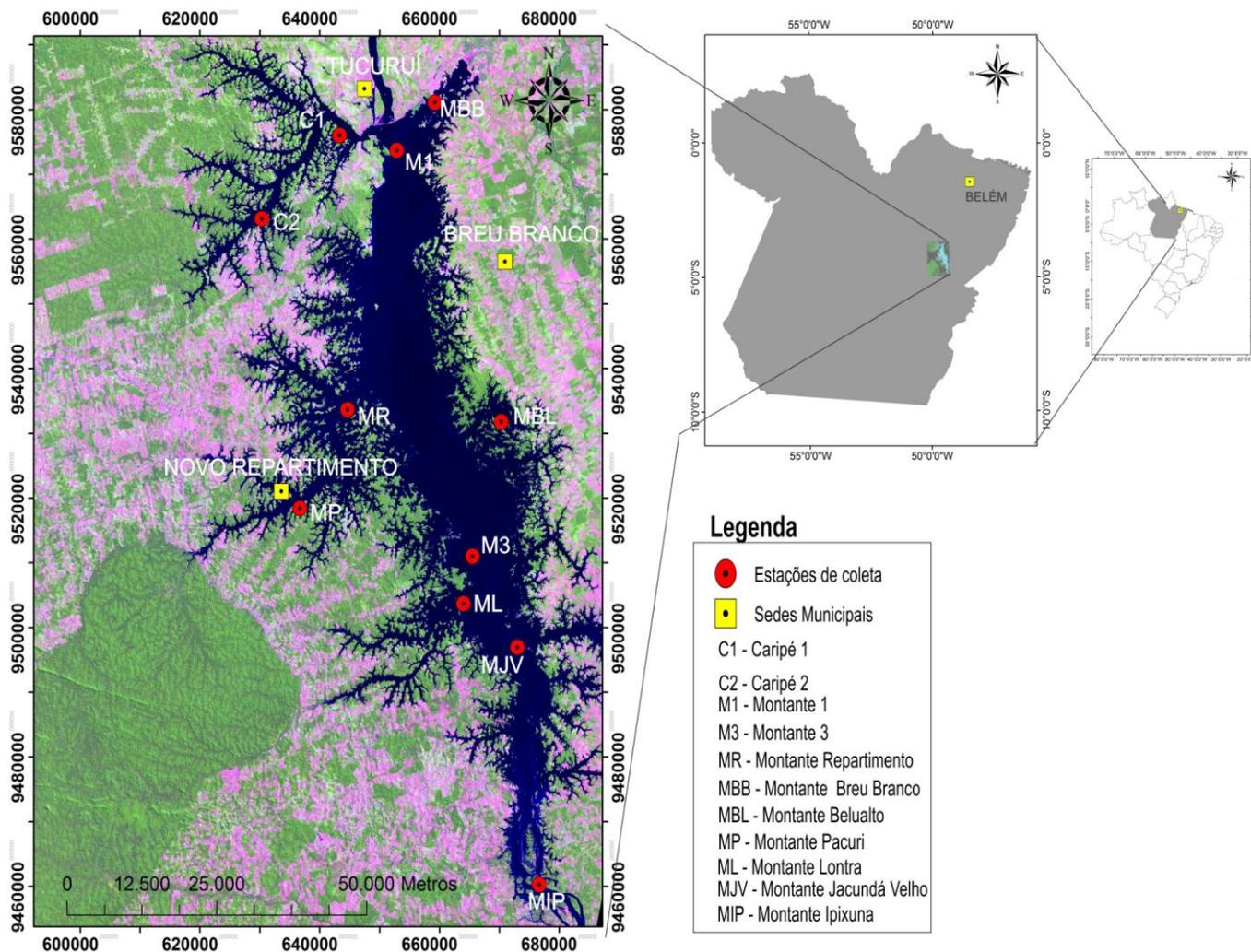


Figura 1- Localização do reservatório da usina hidrelétrica de Tucuruí

A usina hidrelétrica de Tucuruí está localizada no estado do Pará, nas coordenadas: latitude 03° 43' – 05° 15'S, longitude 49° 12' – 50° 00' W (Figura 1). A usina foi construída no rio Tocantins, a cerca de 7 km da cidade de Tucuruí e a 300 km da cidade de Belém. O reservatório possui um total de área inundada de aproximadamente 2.850 Km², com aproximadamente 50,8 milhões de m³ de água (ELETRONORTE, 1988). O período de Dezembro a Maio é considerado chuvoso na região e o período de Junho a Novembro é considerado seco, com precipitação anual entre 2250 e 2500 mm, sendo em Março o mês mais chuvoso e Setembro o mais seco (FISCH et al., 1990).

3.2. Caracterização das estações de amostragem no reservatório de Tucuruí

Estação Montante 1(M1)

Localizada a 2 km da barragem na calha original do rio Tocantins. Esta é importante no monitoramento, por representar a água a ser captada pelas unidades geradoras e também a água a ser lançada a jusante do barramento. Quando o reservatório está na cota máxima de operação (cota de 74m em relação ao mar) chega a ser verificada profundidades de até 70m.

Estações Caraipé 1(C1) e Caraipé 2 (C2)

Localizadas na margem esquerda do reservatório na região denominada atualmente como “Região do Caraipé” onde desde 2002 foi criada a Reserva de Desenvolvimento Sustentável – RDS Alcobaça. Nesta região residem muitos moradores, em torno de 5.000 habitantes, com um nível de antropização elevado, caracterizado pelo elevado nível de desmatamento nesta região. As duas estações localizam-se na antiga calha do Igarapé Caraipé apresentando profundidades de até 28m (quando o reservatório está cheio).

Estação Montante Breu Branco (MBB)

Localizada na margem direita do reservatório, em frente a cidade de Breu Branco, região onde é lançada boa parte da drenagem superficial da cidade. A estação de amostragem apresenta profundidade máxima de 32,0m (quando o reservatório está na cota máxima de operação 74m), nesta região anterior a formação do reservatório não havia nenhum igarapé. Esta região situa-se próximo ao núcleo urbano do município de Breu Branco, apresenta-se muito antropizada, com poucas áreas de floresta primária.

Estação Montante Belauto (MBL)

Localizada na margem direita do reservatório, na área da Zona de Preservação da Vida Silvestre – ZPVS Base 4, área de proteção integral (ou seja, não permitido a presença de moradores). Nesta região, há grandes áreas com floresta primária. Entretanto, esta parte do reservatório caracteriza-se por apresentar um tempo de retenção superior à média do reservatório, vindo a contribuir para que a água nesta região apresente características físico-químicas diferenciadas. Apresenta uma profundidade máxima de 26,0 metros.

Estação Montante Repartimento (MR) e Montante Pucuruí (MP)

Localizada na margem esquerda do reservatório, na região da Reserva de Desenvolvimento Sustentável – RDS Pucuruí – Ararã. Esta região é a mais dendrítica do reservatório, ou seja, ocupou grande área, porém apresenta baixa profundidade. Entretanto, as duas estações apresentam profundidades máximas de 20m (MP) e 32m (MR) devido estarem situadas na antiga calha dos igarapés Pucuruí e Pucuruizinho, respectivamente. Nesta região, há um local de desembarque de peixes “in natura” denominado “Polo Pesqueiro” para seguir ao município de Novo Repartimento. Região bastante antropizada, com significativa redução da vegetação nativa ao longo dos últimos 10 anos.

Estação Montante Jacundá Velho (MJV)

Localizada na margem direita onde havia o antigo núcleo urbano da cidade de Jacundá. Nesta região, há dois locais de desembarque de peixe “in natura” sendo um deles denominado “Porto Novo” – município de Jacundá e o “Porto da Colônia” – município de Goianésia do Pará. Região bastante antropizada com reduzidas áreas de vegetação nativa. A estação de amostragem do monitoramento limnológico localiza-se na antiga calha do rio Jacundá apresentando profundidade de até 22,0m (quando o reservatório está cheio).

Estação Montante Lontra (ML)

Localizada na margem direita, próximo a Terra Indígena dos Índios Parakanãs (hoje a maior extensão continua com vegetação nativa na região do Lago de Tucuruí), localizada

na antiga calha do Igarapé Bacuri, apresentando profundidade máxima de 22,0m (quando o reservatório está cheio).

Estação Montante 3 (M3)

Localizada na parte central do reservatório, distante aproximadamente, 60km em linha reta da barragem, situada na antiga calha do rio Tocantins, apresenta profundidade máxima de até 52,0 metros (quando o reservatório está cheio). Apresenta seu comportamento, em termos de qualidade da água, dependente da vazão afluente ao reservatório.

Estação Montante Ipixuna (MIP)

Localizada na parte central do reservatório, distante aproximadamente, 130 km em linha reta da barragem, situada na antiga calha do rio Tocantins. No período de cheia do reservatório, apresenta profundidades de até 42,0 metros. Esta apresenta a característica de comportar-se como reservatório no período de cheia e como rio no período de seca (setembro a dezembro).

3.3. Análise físico-químicas e metais da água

Cada corpo de água doce tem um padrão individual de características físico-químicas que são largamente influenciadas pelo clima, geomorfologia, e condições geoquímicas que prevalecem na base de drenagem e lençol aquíferos. A qualidade físico-química da água ocorrerá normalmente em condições sem sinais de impactos antropogênicos (CHAPMAN, 1996). Na prática, essa condição é difícil de encontrar devido ao transporte atmosférico de contaminante e subsequente depósito em locais distantes da origem.

As amostras de água do reservatório de Tucuruí utilizadas neste estudo foram coletadas e analisadas nos anos de 2008, 2009, 2010 e 2011 em onze estações de coleta (Figura 1) distribuídas à montante do reservatório, em períodos trimestrais que correspondem ao ciclo hidrológico da região: seco, enchendo, cheio e vazando totalizando 960 amostras.

Algumas análises dos parâmetros físico-químicos foram feitas no local e outras em laboratório. Os parâmetros físico-químicos obtidos foram: temperatura (°C), primeira faixa de profundidade do disco de Secchi (m), condutividade($\mu\text{S}/\text{cm}$), pH, STS(mg /L),

amônia(mg /L), NO₃(mg /L), clorofila-a(mg /L), turbidez (NTU), oxigênio dissolvido (mg/L), fosfatos PO₄(mg/L) e fosforo total P Total(mg /L). A análise de água para determinação dos metais foi feita pela técnica multielementar com espectrometria de emissão óptica em plasma por acoplamento indutivo (*Inductively Coupled Plasma Optical Emission Spectrometry* -ICP OES). O Quadro 1 mostra as condições operacionais do ICP OES para análise dos metais. Estes metais foram: Ca(mg /L), Fe(mg /L), K(mg /L), Mg(mg /L), Na(mg /L). A seguir são descritos a utilização e importância da análise dos parâmetros físico químicos utilizados neste estudo.

Quadro 1 - Condições operacionais do ICP-OES

Potência da radiofrequência (KW)	1,4
Pressão do nebulizador (kPa)	180
Fluxo de argônio do plasma (L min ⁻¹)	15
Fluxo de argônio auxiliar (L min ⁻¹)	0,75
Introdução da amostra (s)	40
Tempo de lavagem (s)	30
Tempo de estabilização (s)	15
Tempo de leitura (s)	1
Velocidade da bomba peristáltica (rpm)	15

Temperatura: O principal interesse do estudo da **temperatura** é seu papel na estruturação térmica dos corpos d'água já que influencia no metabolismo, na reprodução e comportamento migratório de organismos aquáticos (CHAPMAN, 1996; ATOBATELE e UGWUMBA, 2008; DAVIES et.al, 2008; DEEKAE, et. al., 2010).

Primeira faixa de profundidade do Disco de Secchi (Transparência): A medida do **disco de Secchi** vem sendo muito utilizada para inferir sobre a qualidade da água. Desta forma, pronunciadas alterações no padrão sazonal da transparência da água podem servir de sinal de alerta às autoridades locais e ONGs no estabelecimento de propostas e ações visando a recuperação de um dado corpo de água. São aconselháveis medidas periódicas

do disco de Secchi em vários pontos de um lago. Desta forma, será possível estabelecer uma satisfatória análise espacial e temporal do sistema estudado. (POMPÊO, 1999).

Condutividade elétrica (CE): A condutividade elétrica é a medida da capacidade que a água possui de conduzir correntes elétricas. Este parâmetro proporciona uma boa indicação de mudanças na composição da água como partículas poluentes (ATOBATELE e UGWUMBA, 2008) e, depende da concentração iônica e da temperatura, podendo variar se houver alguma mudança nas concentrações de sólidos dissolvidos na água. (DEEKAE, et. al., 2010).

Potencial Hidrogênico (pH) : Segundo DEEKAE et. al. (2010) a concentração do íon hidrogênio na água é usualmente medida em termos de **pH**, que é definido como o logaritmo negativo da concentração do íon hidrogênio, sendo o valor sete, neutro, valores acima de sete tendem para básico, enquanto valores abaixo de sete tendem para ácido. Valores de pH maiores que 7 e menores que 8,5 são ideais para produtividade biológica, enquanto valores de pH inferiores a 4 são prejudiciais a vida aquática. A maioria dos organismos não tolera grandes variações do pH durante o tempo e se tais condições persistirem podem ocorrer mortes.

O pH de águas naturais é fortemente influenciado pela concentração do dióxido de carbono, que é considerado um gás ácido (ATOBATELE e UGWUMBA, 2008; DEEKAE, et. al., 2010).

Turbidez : é a redução da transparência de fluidos devido à presença de materiais em suspensão. De acordo com Atobatele e Ugwumba (2008) a biomassa de fitoplâncton, disponibilidade de nutrientes e aumento de sólidos em suspensão influenciam na transparência da água e conseqüentemente na turbidez. Os autores também observaram que a turbidez reduz a penetração da luz, diminuindo a produção primária e secundária.

Sólidos totais em suspensão (STS): é a porção dos sólidos totais que fica retida em um filtro que propicia a retenção de partículas de diâmetro maior ou igual a 1,2 µm. São caracterizados através de amostras líquidas (mg/L) ou sólidas (%), com a finalidade de se verificar a possibilidade de degradação aeróbia/anaeróbia dos sólidos em suspensão, tendo sido empregado também quando em sólidos totais sedimentados, para estimar o conteúdo orgânico do sedimento, em corpos d'água (CHAPMAN, 1996).

Amônia (NH_3): ocorre naturalmente em corpo de água onde há uma quebra de nitrogênio orgânico e matéria inorgânica no solo e água, excreção pela biota, redução do gás nitrogênio no corpo de água por microrganismo e da troca de gases com a atmosfera, sendo um indicador de descarga de processos industriais, esgotos fertilização e lixos comunitários (CHAPMAN, 1996). A amônia é formada pela decomposição da matéria orgânica, durante o processo de amonificação. Nesse processo, a amônia resulta tanto da decomposição aeróbica quanto anaeróbica da parte nitrogenada da matéria orgânica.

Nitrato (NO_3): é a principal forma do nitrogênio combinado encontrado em águas naturais, e é o resultado da completa oxidação dos compostos de nitrogênio. A nitrificação (conversão de amônia ou nitrito a nitrato) é o principal processo no ciclo do nitrogênio. A maioria das águas de superfície contém nitrato, contudo, a presença de nitrato em concentrações maiores que 5 mg/L pode indicar condições sanitárias inadequadas, desde que resíduos de humanos e de animais são grandes fontes de nitratos. As plantas são capazes de converter nitratos em nitrogênio orgânico. Desta forma o nitrato estimula o crescimento das plantas e organismos aquáticos (como as algas), mas em excesso, ou seja, em concentrações maiores que pode desencadear a proliferação destes organismos (ABNT-NBR 9896/1993).

Oxigênio dissolvido (OD): A quantidade de oxigênio dissolvido na água é importante para organismos aquáticos, pois afeta o crescimento, sobrevivência, distribuição, comportamento e fisiologia dos organismos aquáticos (DEEKAE, et. al., 2010). A medida de sua concentração é usada para indicar o grau de poluição por matéria orgânica e o nível de auto-purificação da água, o valor mínimo é de 5,0 mg/L (ESTEVEZ, 1998).

Clorofila-a (Pigmentação): A concentração de **clorofila-a** na água está diretamente relacionada com a quantidade de algas presentes nos corpos d'água. As características da qualidade da água determinam que espécies de algas estão presentes. Lagos com elevados níveis de nutrientes, tendem a suportar um maior número de algas do que aquelas com baixo nível desses elementos. Outros fatores, como: temperatura, profundidade, pH e alcalinidade, também influem nas espécies e no número de algas encontradas nos lagos e reservatórios (EÇA, 2009).

Fósforo Total (P Total) e Fosfatos (PO_4): possuem grande importância em toda a atividade de crescimento de algas e outros organismos biológicos. Dado o efeito nocivo

do crescimento anormal de algas em massas de água. Dessa forma, tem sido dada grande importância no controle dos compostos de fósforo despejados pelos efluentes domésticos e industriais causando superpopulações de algas e a redução do Oxigênio dissolvido na água (KATO, 1983).

Metais: contribuem de forma significativa para a poluição do ar, da água e do solo, interferindo na manutenção da biota terrestre e aquática em decorrência principalmente de fontes naturais ou antropogênicas (PAPAFILIPPAKI et. Al., 2008; LEITE et. al, 2013).

3.4. Estações de amostragem e periodicidade das coletas

As onze estações de coleta deste estudo distribuídas à montante do reservatório foram: C1, C2, M1, M3, MR, MBB, ML, MBL, MIP, MP, MJV. Estas foram escolhidas de modo a se ter uma amostragem representativa de todo o reservatório, considerando as peculiaridades de cada ecossistema existente no reservatório (Figura 1), sendo quatro amostras por ano que correspondem ao ciclo hidrológico da região em cada estação de coleta. No caso específico do PO₄ somente as estações C1, C2, M3 e MIP foram consideradas e para P total foram considerados somente as estações C1, C2, M1, M3, MR, MBB, MBL e MIP, cujos dados estavam disponíveis na base de dados da ELETRONORTE.

3.5. Aquisição de dados espaciais e sensores virtuais

Neste estudo, os dados espaciais utilizados foram obtidos do banco de imagens de satélite do Instituto Nacional de Pesquisas Espaciais – INPE, obtendo-se imagens do satélite LandSat 5 TM do lago de Tucuruí dos anos de 2008, 2009, 2010 e 2011, utilizando a banda espectral 2, que possui sensibilidade à presença de materiais particulados em suspensão de corpos d'água, possibilitando o monitoramento e gestão de recursos hídricos.

As imagens da área de estudo foram classificadas e convertidas para o formato vetorial que serviu de contorno para extração dos corpos d'água da imagem de cada estação de coleta (Lee and Warner, 2006). Estas imagens inicialmente foram produzidas a partir de recortes de cada estação de coleta na escala espacial de 1:24.000, exportadas em

escala cinza invertida para o formato JPG nas dimensões de 794×1123 pixels e ajustadas para 15 × 15 pixels, sem perda de informações significativas.

As matrizes de 15×15pixel de cada estação de coleta foram agrupadas por seus respectivos anos gerando matrizes dos anos de estudo (2008, 2009, 2010) nas dimensões 225×3 para treinamento do modelo de RNA construído para a implementação dos sensores virtuais e 225×1 para validação (2011). Como alvo do treinamento das RNA's, construiu-se matrizes com os resultados das análises dos parâmetros físico-químicos e com as concentrações de metais das estações de coleta de água do reservatório para os anos considerados no estudo.

Os sensores virtuais são aplicativos computacionais cujo objetivo é estimar o comportamento de variáveis de difíceis aferições ou que demandam tempo para serem analisadas em laboratório (FORTUNA et. al.,2007). De modo geral, foram utilizadas 768 amostras de água para treinamento e 192 amostras de água para validação dos sensores, correspondendo a 75% e 25% respectivamente.

A implementação de sensores virtuais pode ser realizada por meio de RNA's em função das suas propriedades. Elas são uma técnica da Inteligência Computacional muito utilizada em soluções para gerenciar problemas complexos e incertos. É uma técnica que permite resolver estes tipos de problemas, possuindo a capacidade de previsão de resultados a partir do treinamento experimental de dados pretéritos (YAP e KARRI, 2013; SILVA, et. al. 2010).

3.6. Métodos estatísticos e computacionais

3.6.1. Análise Fatorial

Dentre as diversas técnicas de análise multivariadas existentes, a Análise Fatorial é uma das que permite realizar redução ou sumarização de dados. Esta técnica de interdependência considera todas as variáveis simultaneamente, assumindo que cada uma delas está relacionada com as demais, a fim de se estudar a inter-relação existente entre as mesmas.

Marcoulides e Hershberger (1997) apresentam a Análise Fatorial, desenvolvida como uma técnica multivariada criada por Hottelling (1933), sendo uma forma de se reduzir a quantidade de dimensões a serem visualizadas em modelos que contemplem um grande

conjunto de variáveis, preservando o máximo possível a estrutura dos relacionamentos e minimizando a perda de informação decorrente da redução de dimensionalidade.

Em termos de pré-requisitos Quadro 2, para aplicação da técnica de análise Fatorial, Morrison (1990) apresenta algumas propriedades restritivas, como por exemplo: (a) presença de correlações superiores a 0,30, em módulo, para as variáveis que situam no fator; (b) presença de baixas correlações parciais ou na matriz anti-imagem; (c) teste de Esfericidade de Barlett significativo, o que significa que a matriz de correlações tem correlações significantes no mínimo em algumas das variáveis (HAIR et.al,1998).

Quadro 2 – Procedimentos Estatísticos da Analise fatorial

Teste de Esfericidade de Bartlett	Estatística de teste usada para examinar a hipótese de que as variáveis não sejam correlacionadas na população, ou seja, a matriz de correlação da população é uma matriz identidade, onde cada variável se correlaciona perfeitamente com ela própria ($r=1$), mas não apresenta correlação com as outras variáveis ($r=0$). A significância para o teste não deve ultrapassar 0,05.
Matriz de Correlação	O triângulo inferior da matriz que exhibe as correlações simples entre todos os pares possíveis de variáveis incluídos na análise. Os elementos da diagonal, que são todos iguais a 1, em geral são omitidos
Comunalidade	Porção da variância que uma variável compartilha com todas as outras variáveis consideradas. É também a proporção de variância explicada pelos fatores comuns
Autovalor (Eigenvalue)	Representa a variância total explicada por cada fator.
Cargas dos fatores	Correlações simples entre as variáveis e os fatores
Gráfico de cargas dos fatores	Gráfico das variáveis originais, utilizando as cargas dos fatores como coordenadas
Matriz de Fatores	Cargas dos fatores de todas as variáveis em todos os fatores extraídos
Escore fatorial	Escore composto estimado para cada entrevistado nos fatores derivados
Medida de Adequação de Kaiser-Meyer-Olkin (KMO)	Índice usado para avaliar a adequação da análise fatorial. Os valores obtidos através deste teste variam entre 0 e 1, onde 0,80 ou acima é admirável; 0,70 ou acima, mediano; 0,60 ou acima, medíocre; 0,50 ou acima, ruim; e abaixo de 0,50, inaceitável
Porcentagem de variância	Porcentagem da variância total atribuída a cada fator
Resíduos	Diferenças entre as correlações observadas na matriz de correlação de entrada e as correlações reproduzidas, conforme estimadas na matriz de fatores
Scree plot	Gráfico dos autovalores versus o número de fatores, por ordem de extração

O método utilizado na Análise de Fatores foi o de Componentes Principais, com rotação ortogonal dos eixos via Varimax, onde o objetivo é encontrar uma rotação dos fatores que maximize a variância da matriz de pesos, para simplificar a interpretação dos fatores. Esta análise é baseada em um modelo explícito, em que as variáveis são

expressas como uma função linear de um número reduzido de fatores latentes. Na análise de componentes principais deseja-se determinar uma transformação linear das variáveis originais, sendo elas correlacionadas entre si, resultando em novas variáveis chamadas componentes principais, que possuem as seguintes características:

- São descorrelacionadas;
- A variância total é igual a variância das variáveis originais;
- $\text{Var}(c_1) \geq \text{Var}(c_2) \geq \dots \geq \text{Var}(c_p)$, onde $\text{Var}(c_1)$ é a variância maior possível.

A principal aplicação da Análise de Componentes Principais é a redução de dimensionalidade através da eliminação das componentes com menor variância. A variância das componentes principais é igual ao autovalor que lhe correspondem. A primeira componente é a de maior variância e as outras aparecem em ordem decrescente de variância.

O principal aspecto que distingue a Análise de Fatores da Análise de componentes Principais é que a Análise de Fatores não só reduz a dimensionalidade de X como também permite a interpretabilidade das novas variáveis por suas correlações.

Descrição das variáveis da base de dados a partir do modelo:

$$\mathbf{X} - \boldsymbol{\mu} = \mathbf{L}\mathbf{F} + \boldsymbol{\varepsilon} \quad (\text{Eq 1})$$

onde:

\mathbf{X} é uma variável;

$\boldsymbol{\mu}$ é o centróide;

\mathbf{L} é matriz de *carga* dos fatores;

\mathbf{F} é o vetor de fatores;

$\boldsymbol{\varepsilon}$ é o vetor de resíduo do modelo (erro).

Para este modelo assume-se que a Variância (E) e Covariância (Cov):

$$E(\mathbf{F}) = 0 \quad \text{e} \quad \text{Cov}(\mathbf{F}) = \mathbf{I}$$

$$E(\boldsymbol{\varepsilon}) = 0 \quad \text{e} \quad \text{Cov}(\boldsymbol{\varepsilon}) = \boldsymbol{\psi}, \quad \text{onde } \boldsymbol{\psi} \text{ é uma matriz diagonal.}$$

\mathbf{F} e $\boldsymbol{\varepsilon}$ são independentes.

É muito frequente traçar um gráfico chamado de *Scree Plot* (porcentagem da variância explicada X componentes encontrados). O seu exame permite que se identifique visualmente, o ponto em que os autovalores se tornaram pequenos ou que a porcentagem

da variância explicada é pequena em relação ao ônus de se trabalhar com mais uma componente.

A variância de cada variável X_i pode ser dividida em duas parcelas, aquelas devido a fatores comuns e aquela devido a fatores individuais, de acordo com o seguinte modelo:

$$\text{Var}(X_i) = s_{ii} = \lambda_{i1}^2 + \lambda_{i2}^2 + \dots + \lambda_{iq}^2 + \Psi_i \quad (\text{Eq. 2})$$

A variância devido aos fatores comuns é chamada de comunalidade. A comunalidade total é dada por:

$$H = h_1^2 + h_2^2 + \dots + h_p^2 = \sum_{i=1}^p \sum_{j=1}^q \lambda_{ij}^2 \quad (\text{Eq. 3})$$

Assim, a matriz de covariância de X será a soma da matriz cheia LL^T e da matriz diagonal ψ , isto é:

$$\text{Cov}(X) = \Sigma = LL^T + \psi \quad (\text{Eq. 4})$$

Onde:

L $p \times m$ é a matriz de carga dos fatores, constituída dos coeficientes L_{ij} chamados de carga da i -ésima variável no j -ésimo fator.

$$\psi \text{ } p \times p \text{ é a matriz de } \text{Cov}(\epsilon) = E[\epsilon\epsilon^T], \text{ com } \Psi = \begin{bmatrix} \psi_1 & 0 \dots & 0 \\ 0 & \psi_2 \dots & 0 \\ \vdots & \vdots & \vdots \\ 0 & 0 \dots & \psi_p \end{bmatrix} \quad (\text{Eq. 5})$$

Logo, no modelo de Análise Fatorial, os fatores comuns são responsáveis pela estrutura de covariância, enquanto os valores individuais só influenciam a diagonal da matriz de covariância de X . Os fatores comuns são responsáveis pela estrutura de dependência entre as variáveis.

Um recurso para melhorar a interpretabilidade de L , é realizar uma rotação conjunta dos eixos. Utilizou-se nesse o método de rotação ortogonal Varimax, o qual gera uma matriz G dos fatores que maximizam a variância do quadrado das colunas de $L^T = GL$.

Pode-se provar que a estimativa de máximo verossimilhança das comunalidades são:

$$\hat{h}_i^2 = \hat{l}_{i1}^2 + \hat{l}_{i2}^2 + \dots + \hat{l}_{im}^2 \quad \text{para } i=(1,2,\dots,p \text{ componentes}). \quad (\text{Eq. 6})$$

com a proporção da variância amostral total para o j-ésimo fator dada por:

$$\frac{\hat{l}_{i1}^2 + \hat{l}_{i2}^2 + \dots + \hat{l}_{im}^2}{s_{11} + s_{22} + \dots + s_{pp}} \quad (\text{Eq. 7})$$

a proporção padronizada da variância amostral total para o j-ésimo fator dada por:

$$\frac{\hat{l}_{i1}^2 + \hat{l}_{i2}^2 + \dots + \hat{l}_{im}^2}{p} \quad (\text{Eq. 8})$$

O algoritmo básico do Varimax define para $j=1,\dots,m$ e $i=1,\dots,p$:

$$\tilde{l}_{ij}^* = \frac{\hat{l}_{ij}^*}{\hat{h}_i} \quad (\text{Eq. 9})$$

para ser os coeficientes finais de rotação escalados pela raiz quadrada da comunalidade. O procedimento utiliza uma transformação ortogonal que faz com que,

$$V = \frac{1}{p} \sum_{j=1}^m \left[\sum_{i=1}^p \tilde{l}_{ij}^{*4} - \left(\sum_{i=1}^p \tilde{l}_{ij}^{*2} \right)^2 / p \right] \quad (\text{Eq. 10})$$

seja a maior possível.

De forma sintética, pode-se dizer que a análise fatorial possui quatro grandes etapas para a sua realização:

- 1) Cálculo da matriz de correlação (ou de covariância conforme o caso) das variáveis que irão compor os fatores, nesta etapa é verificada a adequação da aplicação da técnica a base de dados em questão;
- 2) Extração dos fatores e seleção dos mais significativos que representarão os dados. Nesta etapa, é possível verificar o quão bem o modelo representa os dados;
- 3) Aplicação de rotação, nos fatores mais significativos, para facilitar o entendimento dos mesmos;

4) Salvar os escores fatoriais que podem ser utilizados em outras análises.

A matriz de correlação dos dados de entrada pode ser calculada a partir da computação das correlações entre as variáveis ou entre os respondentes individuais. As variáveis para AF devem ser métricas e deve-se minimizar o número de variáveis, mas manter um número significativo de variáveis por fator (HAIR *et al.*, 1998)

Neste trabalho, os pré-requisitos para adequação da análise fatorial utilizados como critérios para escolher os fatores que melhor representam a estrutura dos dados, foram: o *Kaiser-Meyer-Olkin Measure of Adequacy* (KMO) maior que 0,50 e o teste de Bartlett. O KMO é uma medida que serve para quantificar o grau de correlação entre variáveis. Este índice varia de 0 a 1, alcançando 1 quando cada variável é perfeitamente prevista sem erro pelas outras variáveis, sendo inaceitáveis valores a baixo de 0,50. O teste de Bartlett de esfericidade é um método que examina a matriz de correlação inteira, fornecendo a probabilidade de correlações significantes existem em algumas das variáveis (JOHNSON e WICHERN, 2001; HAIR *et al.*, 1998).

A matriz fatorial não-rotacionada é computada para auxiliar na obtenção de uma indicação preliminar do número de fatores a extrair. Soluções com fatores não-rotacionados atingem o objetivo de redução de dados, porém, podem não fornecer a interpretação mais adequada das variáveis. Em geral, a rotação é desejável porque simplifica a estrutura fatorial. A rotação fatorial pode ser ortogonal, na qual os eixos são mantidos a 90 graus, ou oblíqua, quando não há restrição de os eixos serem ortogonais. A rotação fatorial ortogonal é a mais comum e entre as principais abordagens ortogonais destacam-se a *quadrimax* (simplifica as linhas da matriz fatorial), *varimax* (simplifica as colunas da matriz fatorial) utilizada neste trabalho e *equimax* (simplifica as linhas e colunas da matriz fatorial, porém não apresenta bons resultados) (JOHNSON E WICHERN, 2001; HAIR *et al.*, 1998).

As cargas fatoriais e as comunalidades (valor que representa a quantia de variância explicada pela solução fatorial para cada variável) devem ser analisadas caso não apresentem valores satisfatórios de correlações, é preciso remover alguma variável e executar a análise de componentes ou de fatores comum novamente.

3.6.2. Estimadores M

Estes são estimadores robustos de tendência central que aplicam uma ponderação aos dados de acordo com as distâncias destes ao centro da distribuição. Assim, os dados mais próximos do centro da distribuição recebem os maiores pesos. Quando os dados pertencem a uma distribuição simétrica com caudas longas, ou quando os dados têm valores extremos, os estimadores M fornecem melhores estimativas do que as obtidas por outro estimador de tendência central como a mediana ou a média (HAMPEL et.al., 1986).

3.6.3. Métodos de Interpolação

A classificação dos métodos de interpolação encontrados na literatura é subjetiva, entretanto Hartkamp et al. (1999), classificam os interpoladores em globais, locais ou estocásticos. Os interpoladores globais consideram todos os pontos da região amostrada, permitindo assim interpolar valores da função em qualquer ponto dentro do domínio dos dados originais, ou seja, a remoção ou adição de pontos afetará todo o mapa. Já com relação aos interpoladores locais, cada ponto da superfície é estimado baseando-se nos valores das amostras mais próximas. Assim, a alteração de um valor afeta localmente os pontos próximos ao mesmo. Vale salientar, que tanto os interpoladores globais como os locais não levantam qualquer hipótese estatística sobre a variabilidade espacial.

Os modelos estocásticos levam em consideração que cada ponto da superfície é estimado apenas tendo como fundamento a interpolação dos valores das amostras mais próximas, utilizando um estimador estatístico. Assim, os erros são definidos através das auto-correlações e correlações espaciais dos atributos (DEUSTCH e JOURNAL, 1996).

3.6.4. Redes Neurais Artificiais (RNA's)

Uma rede neural artificial (RNA) é um sistema de processamento de informação que possui algumas características de desempenho em comum com as redes neurais biológicas. Os modelos neurais artificiais têm como principal fonte de inspiração as redes neurais biológicas, inspirados na habilidade apresentada pelos seres humanos e outros animais no desempenho de funções como o processamento de informações sensoriais e a capacidade de interação com ambientes pouco definidos. Logo, a ciência da computação está preocupada em desenvolver sistemas artificiais capazes de desempenhar estas

tarefas, resultando em habilidades como capacidade de processamento de informações incompletas ou imprecisas (NEGNEVITSKY, 2005; RUSSELL e NORVIG, 2010).

A arquitetura da RNA (Figura 2) do tipo *perceptron* de múltiplas camadas (MLP) constituem um dos paradigmas neurais artificiais mais utilizados e difundidos. Tipicamente, esta arquitetura consiste de um conjunto de unidades sensoriais que formam uma camada de entrada, uma ou duas camadas intermediárias (ou escondidas) de unidades computacionais e uma camada de saída. Os sinais de entrada são propagados camada a camada pela rede da camada de entrada para a camada de saída (GIROSI et.al, 1995).

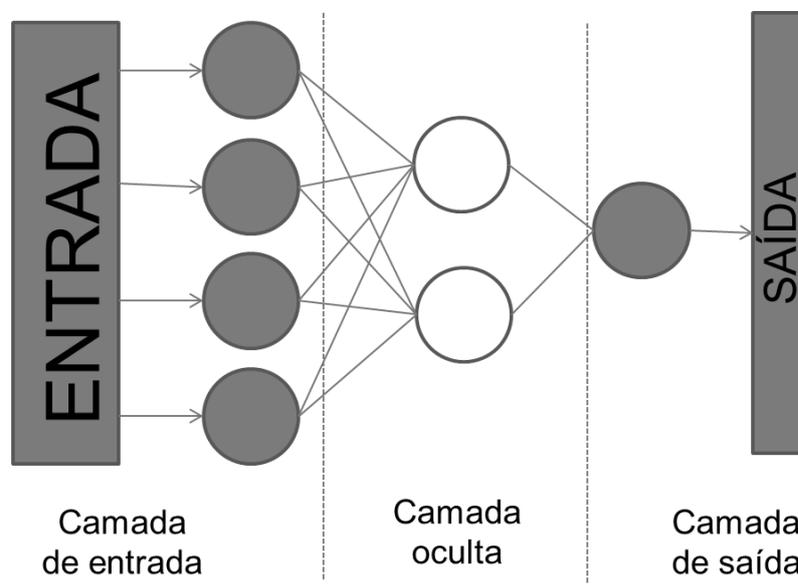


Figura 2 - Arquitetura multicamadas do modelo de RNA perceptron

As redes do tipo MLP têm sido utilizadas com sucesso na solução de vários problemas envolvendo altos graus de não-linearidade. Seu treinamento é do tipo supervisionado e utiliza um algoritmo muito popular chamado retro-propagação do erro (*error backpropagation*). Este algoritmo é baseado numa regra de aprendizagem que minimiza o erro de aproximação da rede durante o treinamento (HAYKIN, 1994).

O processo de treinamento de uma RNA MLP por meio do algoritmo de retro-propagação do erro, é constituído de duas fases: uma fase de propagação do sinal funcional (*feedforward*) e uma de retro-propagação do erro (*backpropagation*). Na primeira fase, os vetores de dados são aplicados às unidades de entrada, e seu efeito se

propaga pela rede, camada a camada até a camada de saída e, um conjunto de saídas é produzido como resposta da rede. Durante essa fase, os pesos das conexões são mantidos inalterados. Na segunda fase, o erro resultante da subtração entre os valores desejados e os valores obtidos pela RNA, são propagados da camada de saída para a camada de entrada da rede (*backpropagation*) e, durante a propagação do erro, os pesos da rede são ajustados de acordo com uma regra de correção do erro. Especificamente, a resposta da rede em um instante de tempo é subtraída da saída desejada (*target*) para produzir um sinal de erro. Este sinal de erro é propagado da saída para a entrada, camada a camada, originando o nome “retro-propagação do erro”. Os pesos são ajustados de forma que a distância entre a resposta da rede e a resposta desejada seja reduzida. Esse processo se repete, até que esta redução alcance uma tolerância estabelecida (SILVA et. al., 2010).

A função de ativação representa o efeito que a entrada interna e o estado atual de ativação exercem na definição do próximo estado de ativação da unidade. Quando propriedades dinâmicas estão envolvidas na definição do estado de ativação, equações diferenciais (caso contínuo) ou a diferenças (caso discreto) são empregadas. Tendo em vista a simplicidade desejada para as unidades processadoras, geralmente define-se seu estado de ativação como uma função algébrica da entrada interna atual, independente de valores passados do estado de ativação ou mesmo da entrada interna. Geralmente, esta função é monotonicamente não decrescente e apresenta um tipo de não-linearidade associada ao efeito da saturação. A Figura 3 a seguir descreve alguns tipos de função de ativação empregados na literatura para as arquiteturas do tipo MLP (HAYKIN, 1994).

A saída de cada neurônio é computada a partir de uma função de ativação. Existem diversas funções de ativação, porém apenas poucas apresentaram aplicação prática (NEGNEVITSKY, 2005). Dentre as escolhas mais comuns, destacam-se as funções degrau, sinal, linear, sigmoide e tangente hiperbólica) (GOMES e LUDERMIR, 2008; NEGNEVITSKY, 2005).

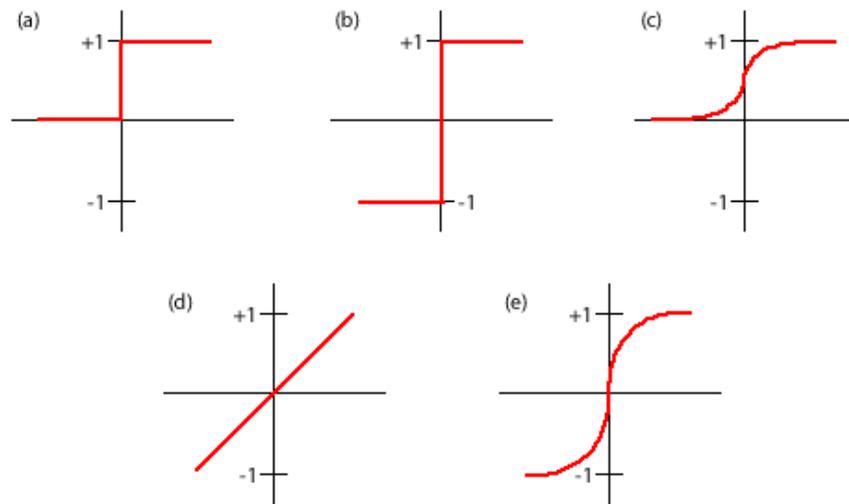


Figura 3 - Funções de ativação de RNA's(a) função degrau; (b) função sinal; (c) função sigmoide; (d) função linear; (e) função tangente hiperbólica

3.7. Sensoriamento Remoto e Processamento de Imagens

Sensoriamento Remoto é a ciência e a arte de obter informações sobre um objeto, área ou fenômeno pela análise de dados adquiridos por um instrumento que não está em contato com o objeto, área ou fenômeno sob investigação. Esses instrumentos recebem o nome de sensores remotos. Esses sensores, são sistemas fotográficos ou ópticoeletrônicos capazes de coletar e registrar a energia proveniente dos alvos sob investigação (objeto, material, área ou fenômeno) de forma a permitir um posterior processamento e interpretação dos registros por um analista. Portanto, a transferência de dados de um alvo para um sensor dá-se por meio da energia. Os sensores mais comuns operam com a energia ou radiação eletromagnética (LILLESAND et. al., 2004).

De acordo com Menezes e Almeida (2012) o processamento de imagens é configurado por algoritmos especializados, que disponibilizam para o usuário a aplicação de uma grande variedade de técnicas de processamento. Os algoritmos tratam matematicamente a imagem como um dado diretamente relacionado ao processo físico que o criou. Porém, a representação e o processamento matemático da imagem não é 100% a representação do mundo real. Por isso, sempre caberá ao usuário e não apenas ao computador, a análise e a interpretação dos dados contidos na imagem.

As imagens de sensoriamento remoto devem ser vistas como uma forma de documentos que representam, em escala e sobre um plano 2D (bidimensional), os

acidentes e as feições naturais e artificiais da superfície terrestre, a partir da medição de um processo físico da radiação eletromagnética. Dessa forma, a energia da radiação eletromagnética conduz de forma analógica a informação sobre os objetos e no sensor um conversor analógico/digital converte essa informação em um valor digital, codificado por uma unidade denominada de pixel. A forma digital do dado é que possibilita o uso de computadores para processar as imagens, com o objetivo principal de representar porções bem definidas do espaço terrestre, utilizando-se de processamentos matemáticos, estatísticos e probabilísticos dos dados. Os limites do espaço são configurados pelos pixels, os quais, como veremos, estão organizados na disposição de linhas e colunas. Embora estejamos falando de valores numéricos (imagem digital), os limites de qualquer porção da imagem não são matemáticos (MATHER, 1987).

A imagem de um sensor é obtida pela varredura do terreno em linhas sequenciais, cada linha sendo amostrada por uma área unitária (pixel). O sensor registra os pixels no formato de uma grade retangular ou quadrada, e o valor do pixel correspondente ao valor de radiância, emitância ou de retroespalhamento dos alvos é convertido no sensor para um valor digital discreto e finito (SCHOWENGERDT, 2007).

O valor digital do pixel é uma média dos valores de radiância, emitância ou retroespalhamento, dos diferentes alvos que podem estar contidos no pixel. Os algoritmos de processamento veem esse formato como uma matriz de dados composta por milhares de linhas e colunas, na qual o valor de cada célula da matriz é definido pelas coordenadas que o pixel ocupa na matriz e pelo seu valor digital. Genericamente, pode se definir qualquer imagem como sendo uma função bidimensional representada por $f(x,y)$ (GONZALEZ e WOODS, 2002).

Quando x,y e os valores de amplitude f são quantidades discretas (números inteiros), a imagem é uma imagem digital. Nas imagens de sensoriamento remoto x,y referem-se às coordenadas espaciais da linha e coluna do pixel e o valor de f em qualquer ponto x,y é a medida da energia radiante refletida, emitida ou retroespalhada. Na forma digital, o valor da função $f(x,y)$, é representado por um número digital quantizado em bits, que assume somente valores inteiros positivos, como mostrado na Figura 4. Uma imagem $f(x,y)$ possível de ser processada por computadores e exposta na tela do monitor, tem de ser digitalizada tanto espacialmente como em amplitude de níveis de brilho (ou cinza).

Digitalização de coordenadas x,y refere-se à “amostragem de imagem”, e digitalização de amplitude refere-se à “quantização de nível de brilho”(MENEZES e ALMEIDA, 2012).

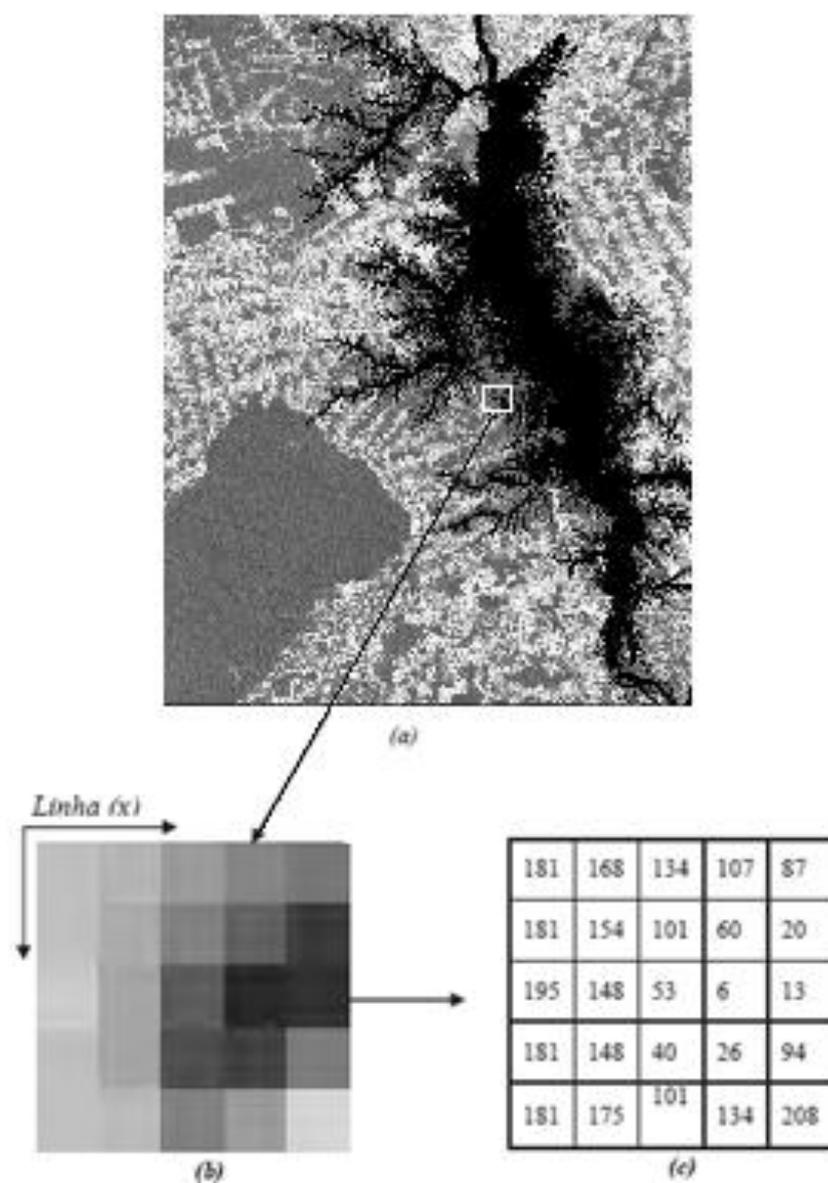


Figura 4- (a) Imagem digital com (b) área em zoom de grupo de pixels em valores de cinza e (c) correspondentes valores digitais (Menezes e Almeida, 2012).

4. APLICAÇÃO DA METODOLOGIA E RESULTADOS

4.1. Análise Fatorial

Na análise fatorial, o valor de KMO foi 0,76, acima de 0,50, para a matriz completa, indicando ser apropriada a aplicação da análise fatorial, e o teste de Bartlett apresentou um valor significância inferior a 0,05. Foram obtidos 6 fatores com variância total de 71,05. A Tabela 1 mostra os fatores, as comunalidades, as variâncias total e acumulada. O método de extração dos fatores foi o de Componentes Principais e a rotação ortogonal varimax.

Tabela 1 - Análise fatorial com os fatores, comunalidades, variância total e acumulada.

Parâmetros	Comunalidade	Fatores					
		F1	F2	F3	F4	F5	F6
Secchi (m)	0,772	-0,766	-0,080	0,163	-0,217	0,064	-0,318
Fe Total (mg /L)	0,725	0,752	0,177	0,339	-0,036	-0,093	-0,057
PO ₄ (mg /L)	0,622	0,740	0,004	0,098	-0,106	-0,181	-0,147
P TOTAL(mg /L)	0,838	0,865	0,199	0,205	-0,037	-0,057	0,062
STS (mg /L)	0,713	0,827	-0,028	-0,094	0,077	0,047	0,111
Turbidez (NTU)	0,761	0,843	-0,051	-0,188	0,109	-0,020	-0,028
Na(mg /L)	0,814	-0,037	0,876	0,067	-0,064	0,095	0,166
K(mg /L)	0,763	0,252	0,793	0,116	0,087	-0,219	-0,032
OD(mg O ₂ /L)	0,645	0,112	-0,053	-0,745	0,076	0,235	0,115
NH ₄ (mg/L)	0,728	0,352	0,162	0,632	0,229	0,152	0,320
pH	0,688	0,149	-0,246	-0,475	0,549	0,069	0,272
Mg(mg /L)	0,726	-0,061	0,094	0,011	0,833	-0,131	-0,036
NO ₃ (mg /L)	0,686	0,378	-0,088	0,357	0,506	0,350	-0,171
Condutividade(μS/cm)	0,660	0,237	-0,004	0,108	-0,060	-0,714	0,281
Ca(mg /L)	0,689	-0,034	-0,082	-0,073	-0,129	0,766	0,269
Temperatura (°C)	0,541	-0,457	-0,260	0,007	-0,212	0,046	0,466
Clorofila-a (mg /L)	0,703	0,096	0,186	-0,046	0,040	-0,005	0,809
Variância total		26,38	9,96	9,31	8,67	8,41	8,32
% Variância acumulada		26,38	36,34	45,65	54,32	62,73	71,05

De acordo com a Tabela 1 é demonstrada a relação dos elementos Transparência, Ferro Total, PO4, P total, STS e Turbidez e estão agrupados no Fator 1 (F1) apresentando forte correlação entre eles com variância Total de 26,38. Estes elementos apresentam grande refletância em corpos d'água e boa absorção de energia pelos sensores do satélite, possibilitando a visualização destes por meio de imagens, pois o F1 possui maior variâncias total que os fatores F2, F3, F4, F5 e F6 (Tabela 1).

4.2. Distribuição espacial dos parâmetros físico-químicos e metais

Para determinar em quais estações de coleta serão aplicados os sensores virtuais, utilizou-se imagens de interpolação das concentrações dos elementos contidos no Fator 1 da análise fatorial (Tabela 1) possuindo forte correlações que afetam diretamente a qualidade da água. O algoritmo de interpolação utilizado foi o Inverso de distância ponderada (IDW) que é considerado o método mais simples de interpolação, que atribui o peso maior ao ponto mais próximo, diminuindo esse peso com o aumento da distância, ou seja, quanto maior for o peso, menor será a influência dos pontos mais distantes do nó. O método tem como característica a geração do chamado efeito de mira ao redor dos pontos observados, retratando a similaridade das estações de coleta como mostram as figuras 5 a 10.

As imagens da transparência das estações de coleta medidas pela primeira faixa de profundidade pelo disco de Secchi dos anos de estudo (Figura 5), demonstra que nas estações próximas da barragem apresentam maior transparência e similaridades em todos os quatro anos.

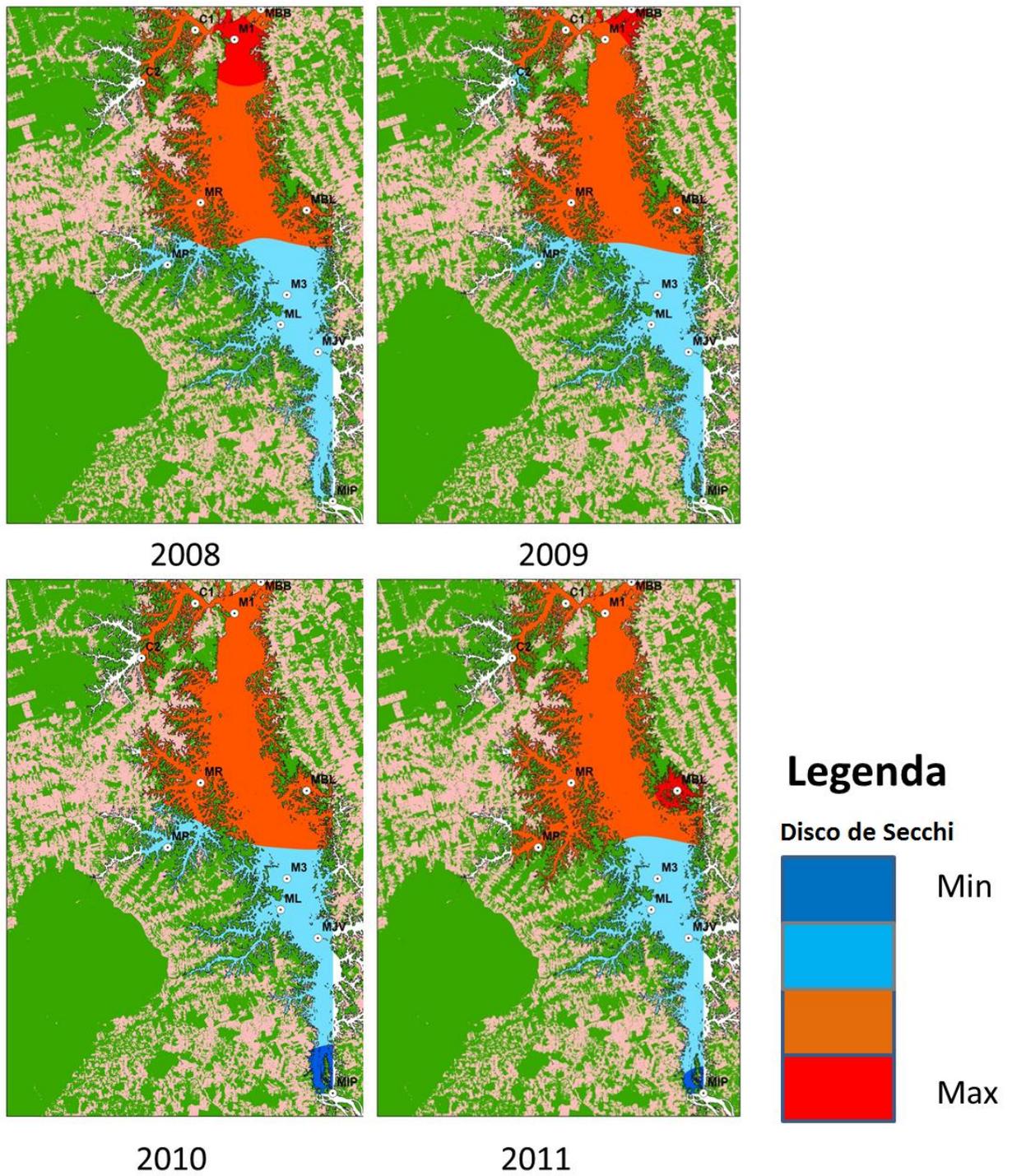


Figura 5- Distribuição espacial do Secchi nas estações de coleta

As imagens da concentração de Fe Total (Figura 6) das estações de coleta, demonstram que nas estações ML, MJV e MP apresentam disparidades de valores, apresentando pouca similaridade em relação aos demais anos.

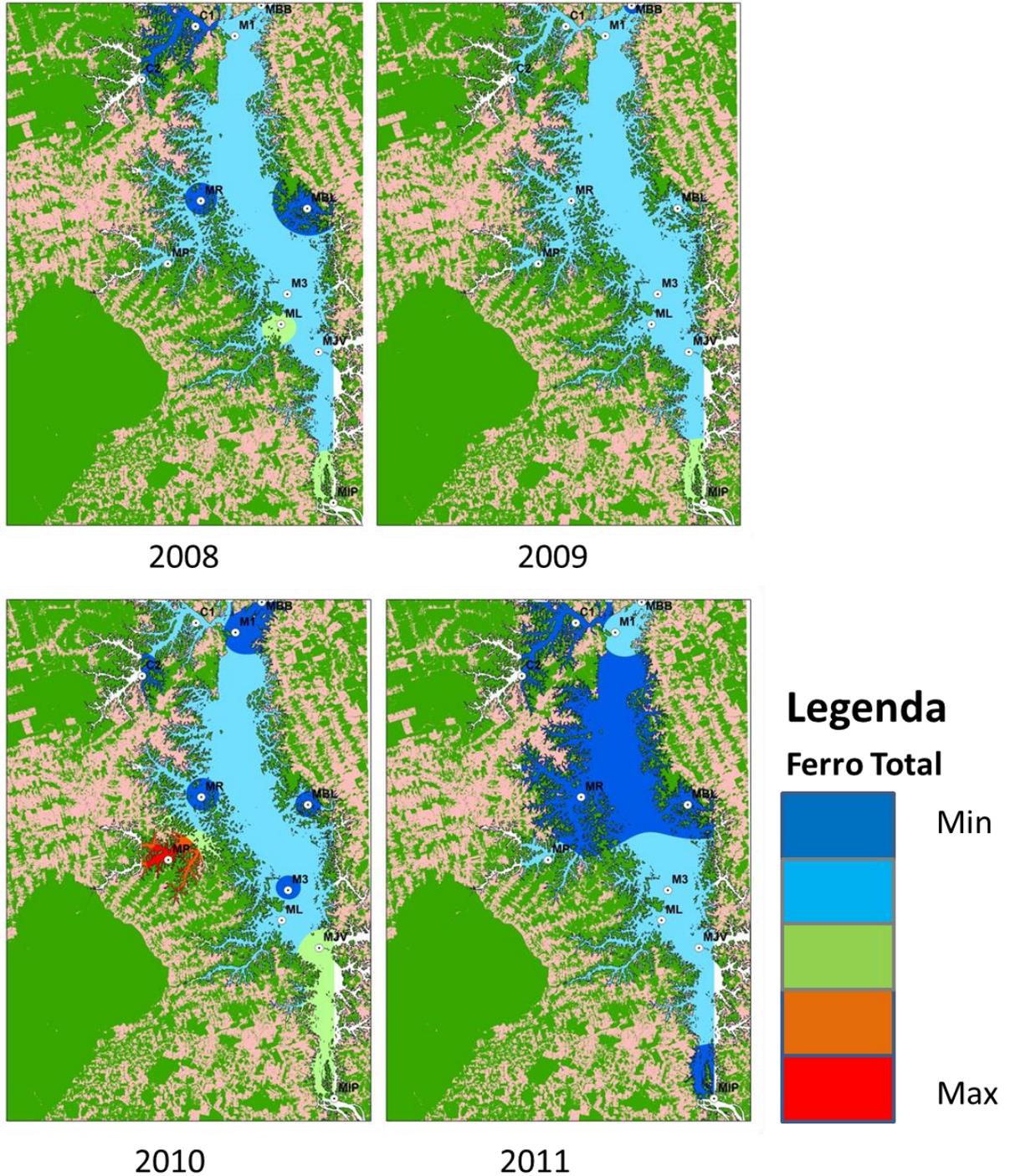


Figura 6 - Distribuição espacial das concentrações de Ferro Total nas estações de coleta

Já para as concentrações de fosfatos (Figura 7) somente as estações C1, C2, M3 e MIP apresentaram similaridades significantes e novamente as estações ML e MJV apresentam disparidades de concentrações nos quatro anos de estudo.

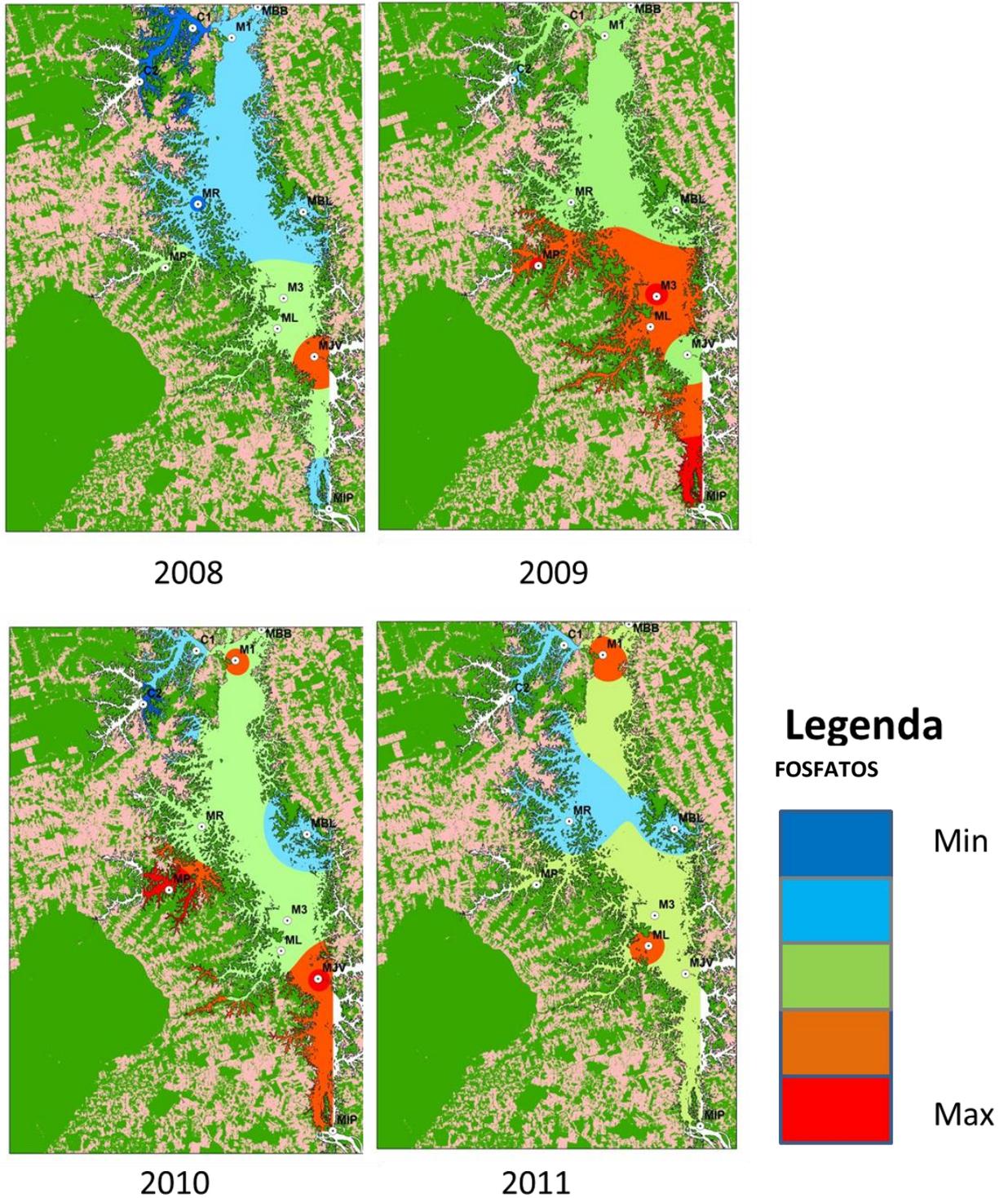


Figura 7- Distribuição espacial das concentrações de Fósphatos nas estações de coleta

Para o Fósforo Total (Figura 8) as estações C1, C2, M1, M3, MR, MBB, MBL e MIP apresentam pouca similaridade, contudo as estações ML e MJV continuam a apresentar disparidades nos anos de estudo.

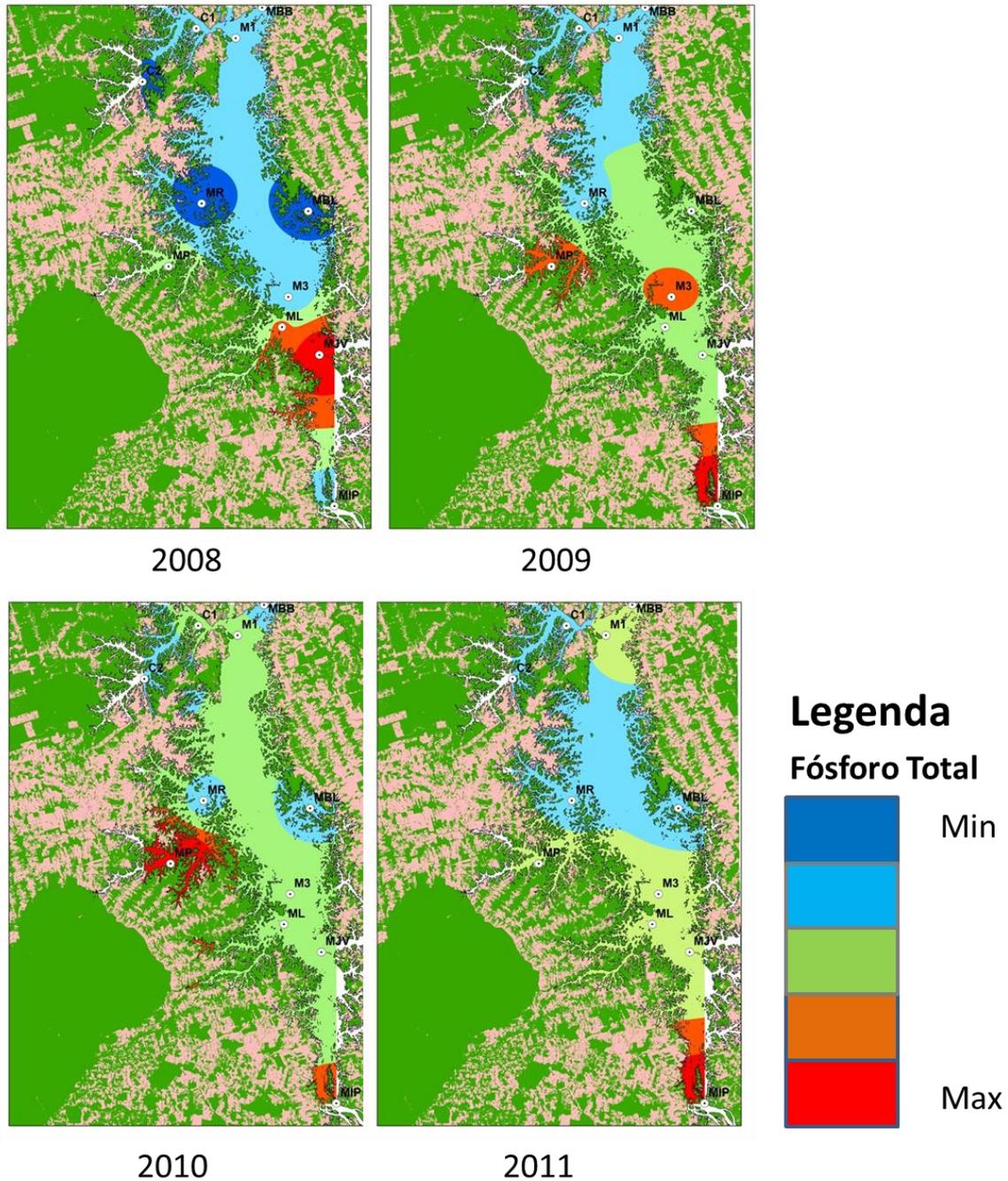


Figura 8 - Distribuição espacial das concentrações de Fósforo Total nas estações de coleta.

Os sólidos totais em suspensão STS apresentam similaridade em todas as estações de coleta nos quatro anos, apresentando baixas concentrações nas proximidades da barragem.

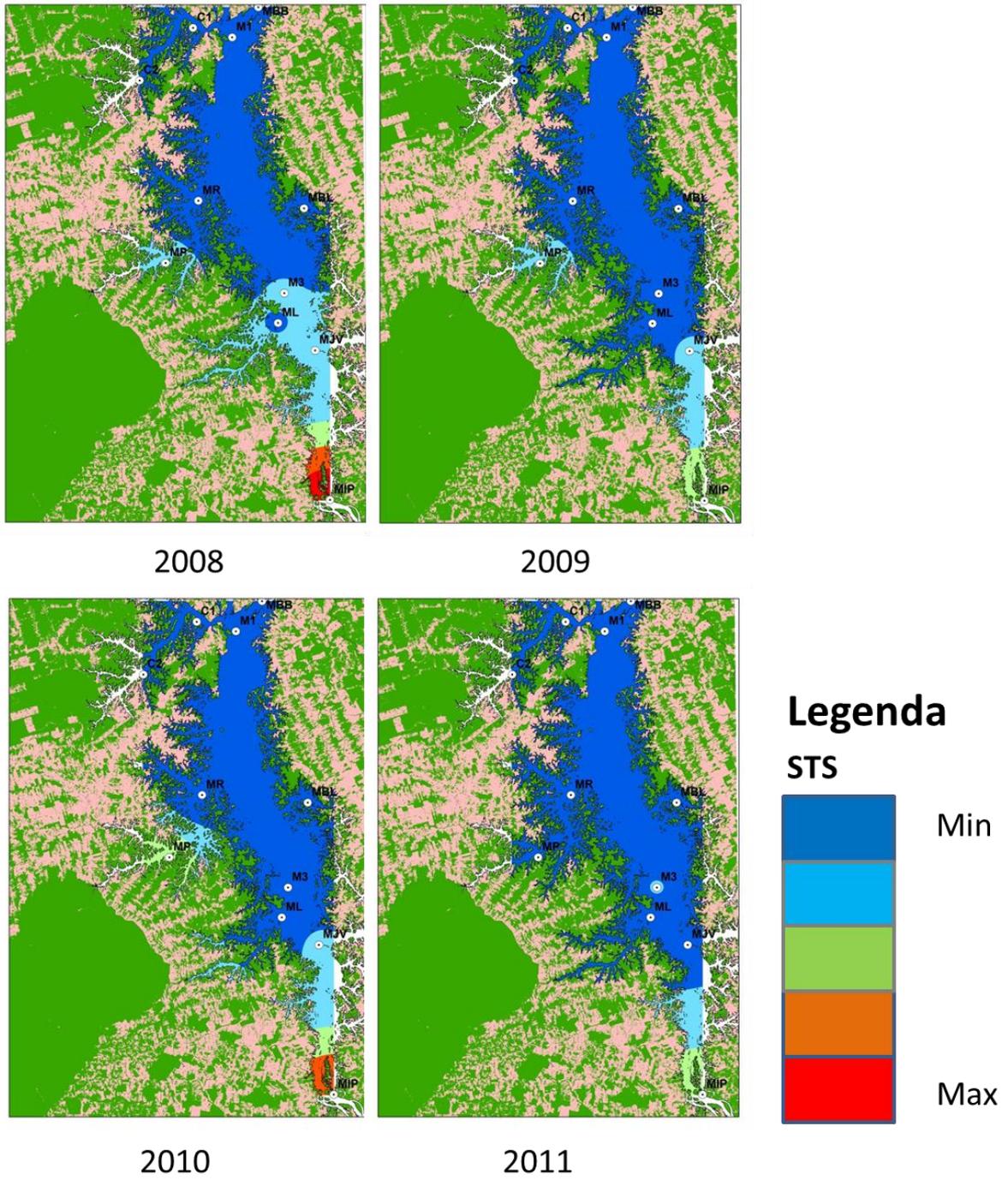


Figura 9 - Distribuição espacial das concentrações de STS nas estações de coleta.

A Turbidez (Figura 10) semelhante ao STS apresentaram similaridades entre os anos de estudo para as estações C1, C2, M1, M3, MR, MBB, MBL, MIP, MP e apresentando baixas concentrações nas proximidades da barragem.

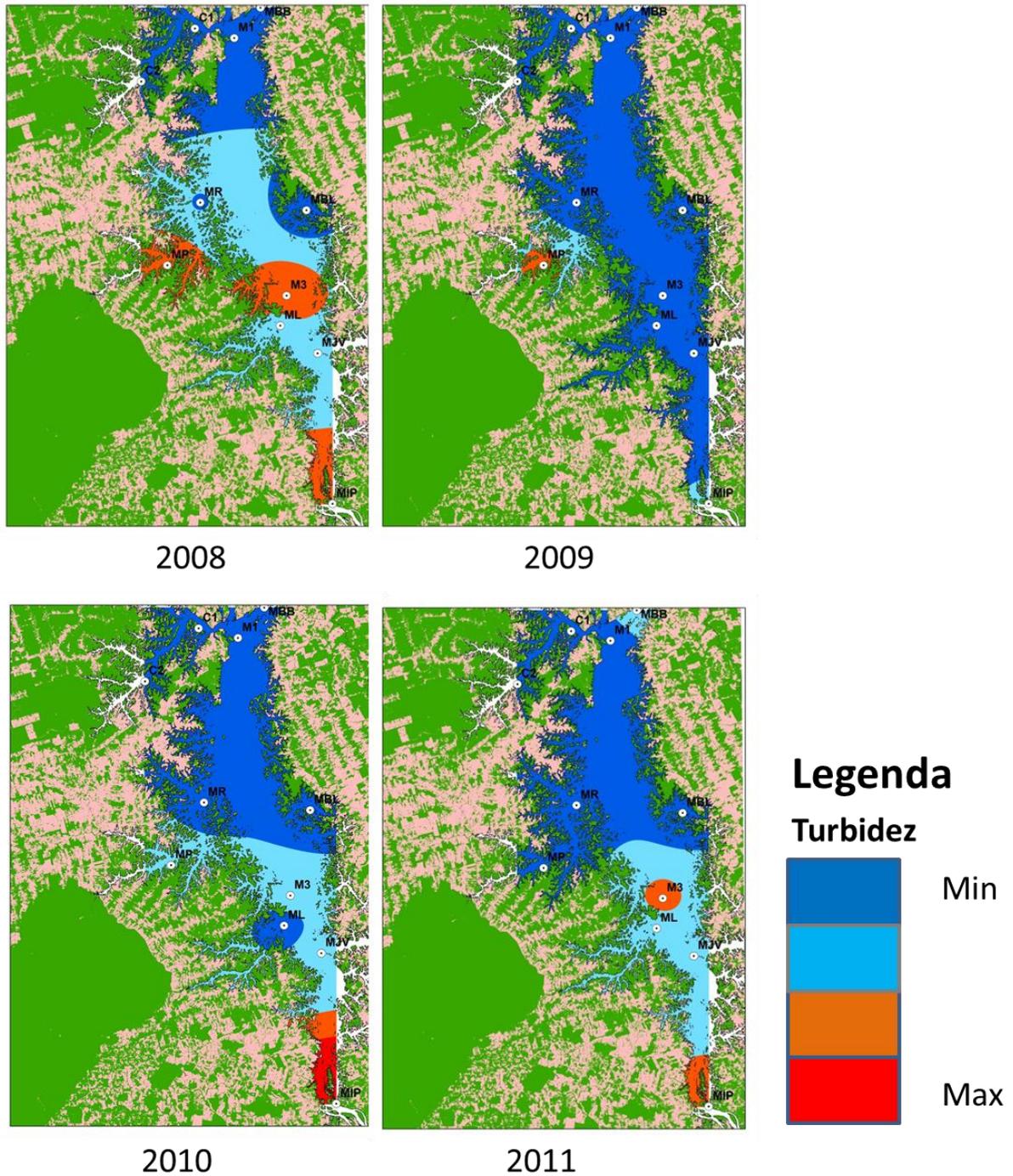


Figura 10 - Distribuição espacial da Turbidez nas estações de coleta

Portanto, após a análise espacial das concentrações dos elementos contidos no Primeiro Fator F1(Tabela 1) demonstram que das 11 estações de coleta serão utilizadas 9 neste estudos pela representatividades nos anos de estudo.

4.3.RNA's

O paradigma de RNA utilizado no estudo foi o *Perceptron* Multicamadas Direta, desenvolvido com o auxílio do programa computacional *Matlab* (MATHWORKS, 2009). Sua arquitetura consistiu de três camadas: camada de entrada, camada escondida e a camada de saída, conforme especificado na Figura 11.

Para treinamento da RNA, foram utilizadas como variáveis de entrada as matrizes de pixels das estações: C1, C2, M1, M3, MR, MBB, MBL, MIP, MP que apresentavam dados amostrais consistentes para processo de treinamento como ausência de nuvens. Como variáveis de saída utilizou as matrizes dos parâmetros físico-químicos e o ferro total de cada estação de coleta, obtidos do primeiro fator da Análise Fatorial (Tabela 1) e as similaridades das estações de amostragens identificadas pelas interpolações (Figuras 5 a 10) dos elementos presentes no primeiro fator. A RNA foi treinada com os seguintes parâmetros:

- RNA: *Perceptron* Multicamada Direta, com uma camada escondida contendo cinco neurônios.
- Algoritmo de treinamento: *Feed-forward back-propagation*.
- Taxa de aprendizagem : 0,001.
- Função de transferência: Tangente hiperbólica em todos os neurônios da rede.

O Erro médio quadrático de aproximação das 9 RNA's utilizadas (uma para cada ponto de coleta de amostra) estão especificados na Tabela 2.

Tabela 2 - Erros de aproximação do treinamento das RNA's .

Nº	RNA	Erro Médio quadrático
1	C1	4.53×10^{-08}
2	C2	2.34×10^{-14}
3	M1	5.21×10^{-09}
4	M3	4.69×10^{-08}
5	MBB	7.05×10^{-15}
6	MBL	9.92×10^{-08}
7	MIP	2.76×10^{-08}
8	MP	6.73×10^{-09}
9	MR	7.95×10^{-08}

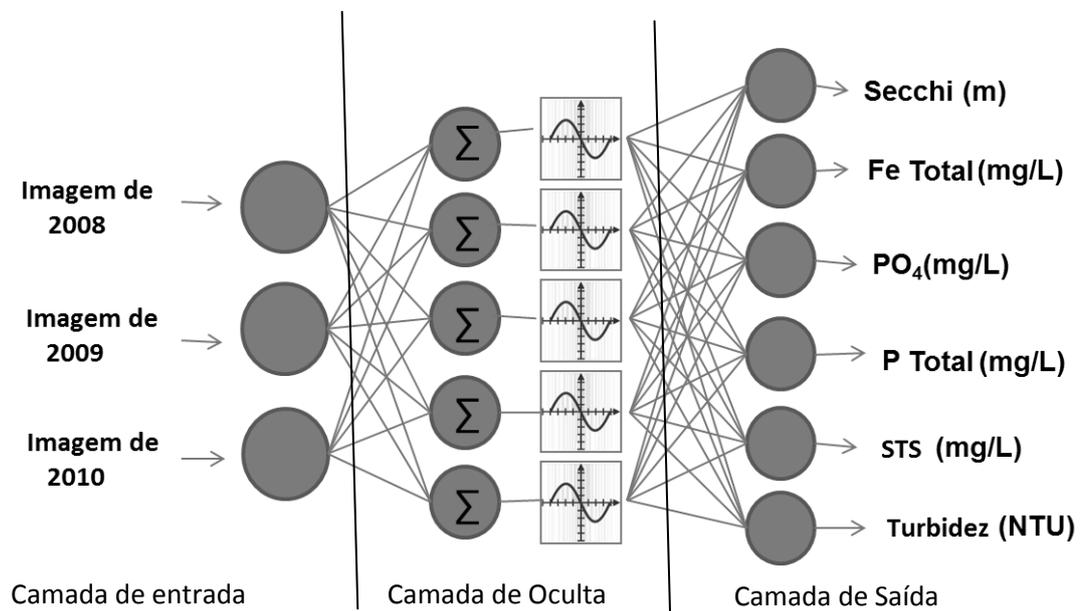


Figura 11 - Arquitetura da RNA utilizada neste estudo

A Figura 12 apresenta o desempenho do treinamento para algumas estações de coleta das RNA's aplicadas aos parâmetros físico-químicos e metais existentes no primeiro fator da análise fatorial Secchi, Fe Total, PO_4 , P Total, STS e Turbidez (Tabela 1), para os pontos de

coleta C1, C2 e M3, comparando os resultados esperados com os processados. Pelos gráficos foi possível ver que o treinamento das RNA's obteve aproximação de 100% entre os valores desejados e os processados pelas RNA's, sinalizando que obteve-se uma generalização por parte das RNA's utilizadas, o mesmo ocorreu para as demais estações de coletas juntamente com os respectivos parâmetros físico-químicos e Ferro total.

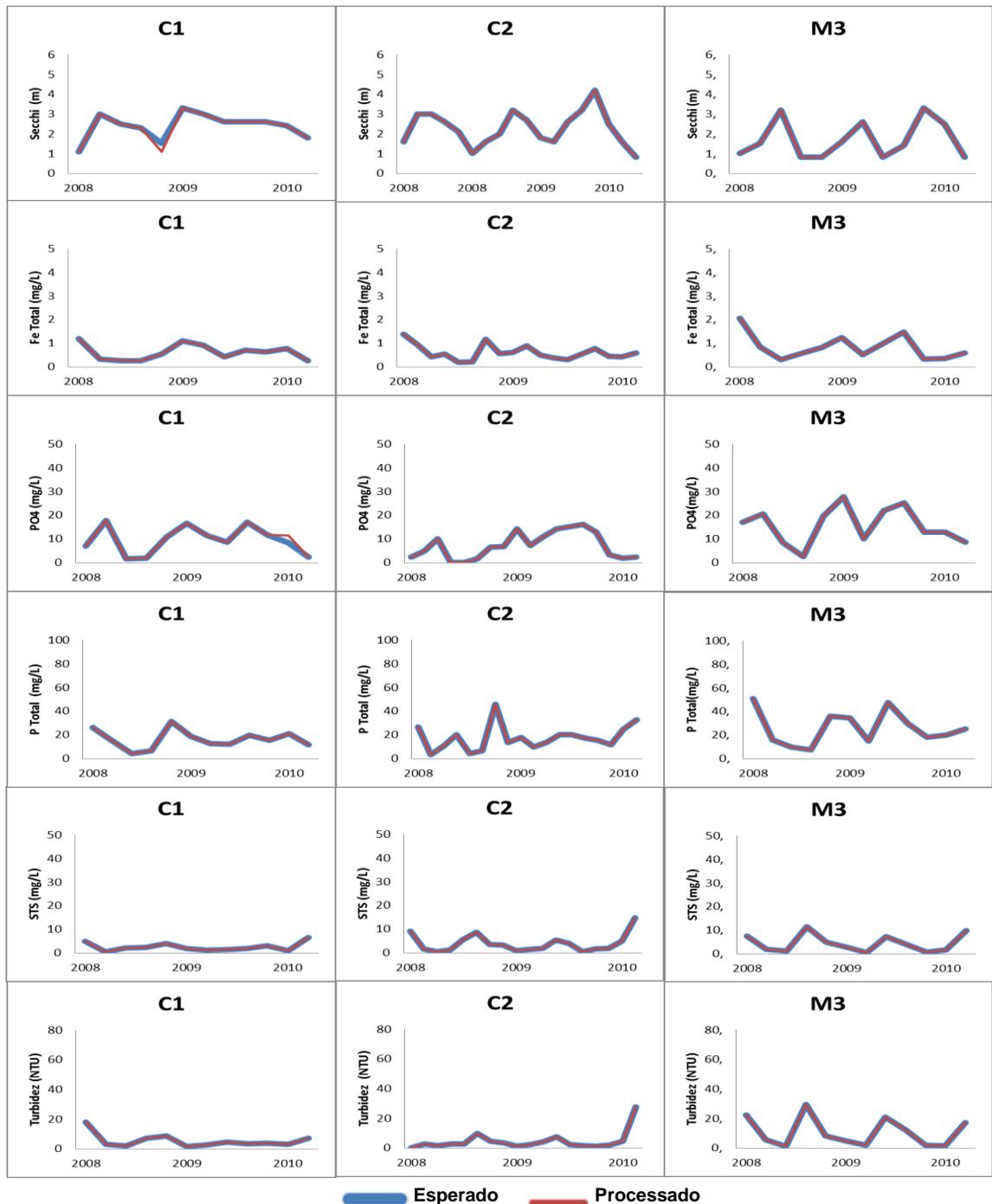


Figura 12 - Desempenho do treinamento das RNA's

4.4. Validação dos sensores

O teste de validação dos sensores foi realizado por meio de dados dos parâmetros físico-químicos e as concentrações de Fe total analisadas no ano de 2011, juntamente com matrizes de pixels das imagens de satélite do mesmo ano, considerando o ciclo hidrológico da região em (1)-seco, (2)- enchendo, (3)-cheio e (4)-vazando.

Para avaliar a validação da RNA foram utilizados os estimadores M. Estas análises foram obtidas pelo programa computacional STATISTICA (STATISTICS, 2005).

A Tabela 3 mostra os resultados obtidos dos estimadores M nas estações de coleta em cada ano estudado retratando as ausências de dados do Fosfato (PO₄) nas estações M1,MR, MBB, MBL e MP. Já para o Fosforo Total (P Total) a ocorrência de ausência foi somente na estação MP. Estes resultados demonstram as faltas de similaridades na validação dos resultados para os parâmetros PO₄ e P Total.

Tabela 3 - Estimador M para os parâmetros nas estações de coletados para cada ano.

Amostras	Ano	Fe Total	Secchi	PO4	P Total	STS	Turbidez
C1	2008	0,28	2,56	3,33	10,86	2,15	3,83
	2009	0,73	2,94	10,29	14,30	1,46	3,20
	2010	0,70	2,54	9,98	17,24	1,92	3,42
	2011	0,23	2,86	7,41	14,87	1,31	1,70
C2	2008	0,27	2,60	3,35	5,16	3,87	2,27
	2009	0,52	1,80	12,91	16,08	2,64	3,60
	2010	0,57	2,43	8,25	21,67	2,26	2,27
	2011	0,31	2,43	2,73	11,64	2,75	4,37
M1	2008	0,65	3,45	-	18,01	1,52	2,97
	2009	0,56	2,76	-	15,13	2,26	4,96
	2010	0,38	3,19	-	17,82	1,09	1,68
	2011	0,48	2,77	-	22,50	1,44	2,20
M3	2008	0,58	1,09	12,31	10,86	5,10	14,51
	2009	0,89	1,14	20,93	36,24	3,80	4,95
	2010	0,41	1,98	11,70	23,13	2,08	8,03
	2011	0,34	2,04	13,01	26,02	3,05	7,25

MR	2008	0,44	2,53	-	5,05	1,68	4,26
	2009	0,88	2,50	-	20,01	2,48	3,30
	2010	0,42	2,54	-	18,62	2,66	3,84
	2011	0,22	2,67	-	17,21	2,52	3,09
MBB	2008	0,59	2,65	-	15,68	2,08	4,35
	2009	0,57	3,03	-	15,03	1,86	2,21
	2010	0,37	2,62	-	16,49	0,97	2,94
	2011	0,42	2,92	-	19,40	2,00	4,89
MBL	2008	0,39	2,50	-	3,95	1,92	2,32
	2009	0,62	2,91	-	27,93	1,60	2,17
	2010	0,42	2,79	-	14,99	1,90	1,99
	2011	0,22	3,41	-	16,03	1,06	1,56
MIP	2008	1,16	1,01	8,19	16,38	23,16	14,17
	2009	0,94	0,81	20,10	57,67	17,37	14,57
	2010	1,06	1,28	16,52	30,49	14,34	16,50
	2011	0,41	0,90	11,20	50,14	13,50	12,83
MP	2008	0,57	1,67	-	-	4,20	9,51
	2009	0,91	1,65	-	-	4,72	9,71
	2010	0,50	1,58	-	-	4,00	6,49
	2011	0,49	2,25	-	-	3,33	2,28

Os resultados do teste de validação demonstraram a eficiência dos sensores nas estimações da primeira faixa de profundidade do disco de Secchi, Fe Total, STS e Turbidez.

O Erro de aproximação ou erro residual revelam o comportamento não modelado, retratando divergências sistemáticas de valores medidos com os valores estimados, pois elevada densidade de valores indicam que o modelo tende a subestimar os valores corretos (BENNETT et al., 2013). As Figuras de 13 a 18 ilustram a comparação dos valores dos parâmetros físico-químicos e das concentrações de ferro observadas com os valores estimados pelos sensores virtuais, por ciclo hidrológico, juntamente com os erros de aproximação.

Na validação do parâmetro primeira faixa de profundidade do disco de Secchi (Figura 13), na estação C1 a diferença entre os valores observados e estimados pode ter sido ocasionado pelos baixos valores no ano de 2008 e 2010, já na estação C2 o baixo valor no ano de 2009.

Nas estações MP e MBL, embora as curvas sejam muito similares, a pouca diferença entre os valores observados e estimados pode ter sido decorrente do alto valor no ano da validação. O MIP pelas diferenças entre os valores nos anos de 2009 e 2010. Nas estações M3 e MIP localizados mais à Montante, primeira faixa de profundidade do disco de Secchi (m), obteve menores valores.

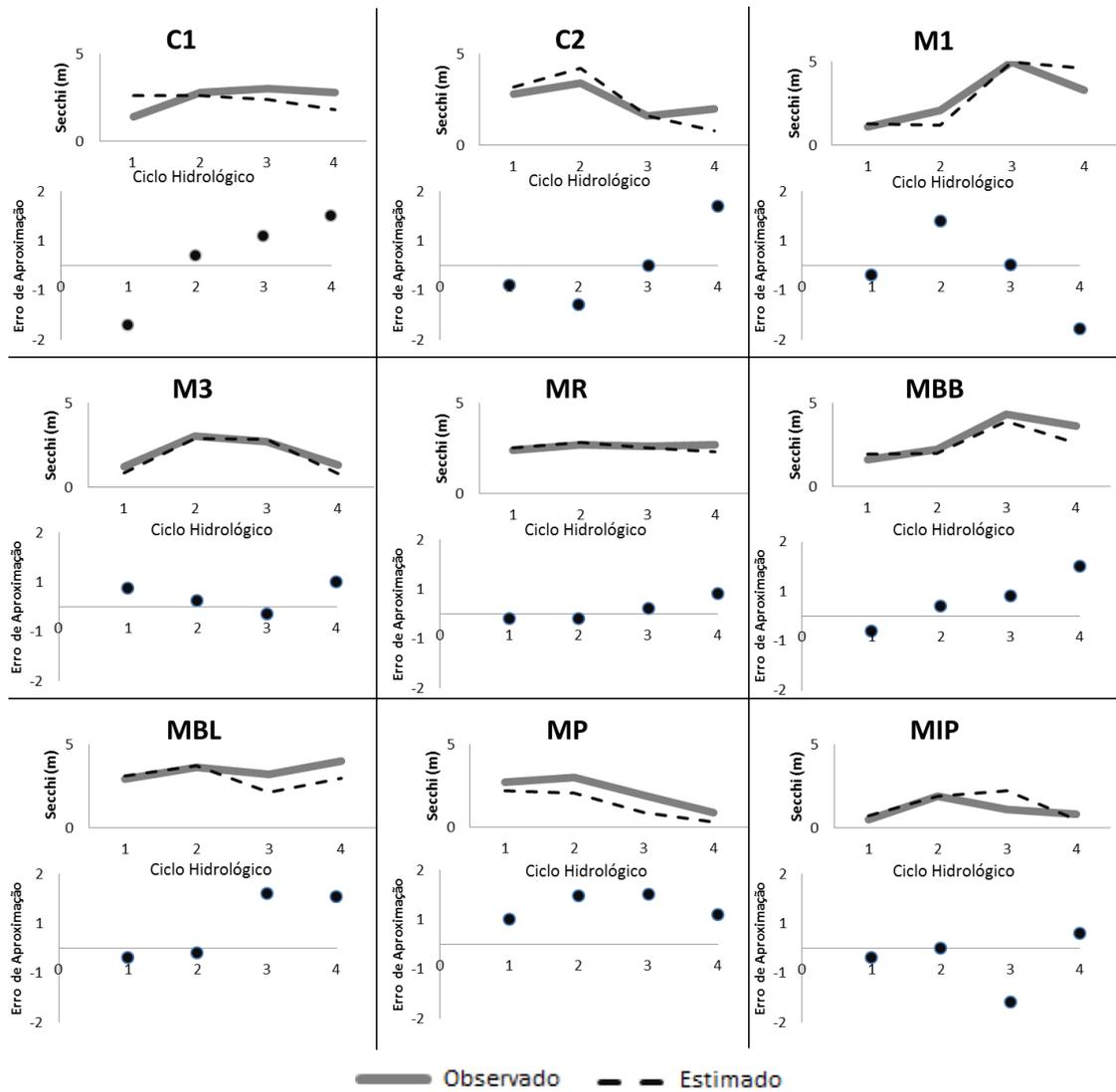


Figura 13 - Validação dos Sensores virtuais para a primeira faixa de profundidade do disco de Secchi (m), considerando o ciclo hidrológico da região e o Erro de aproximação.

Na validação do Fe total, a Figura 14 mostra os valores observados e estimados da concentração, por ciclo hidrológico com as curvas muito similares. De modo geral, estas diferenças foram muito pequenas. As maiores diferenças ocorreram na estação MBB que teve baixas concentrações no ano de 2010 e no ano da validação.

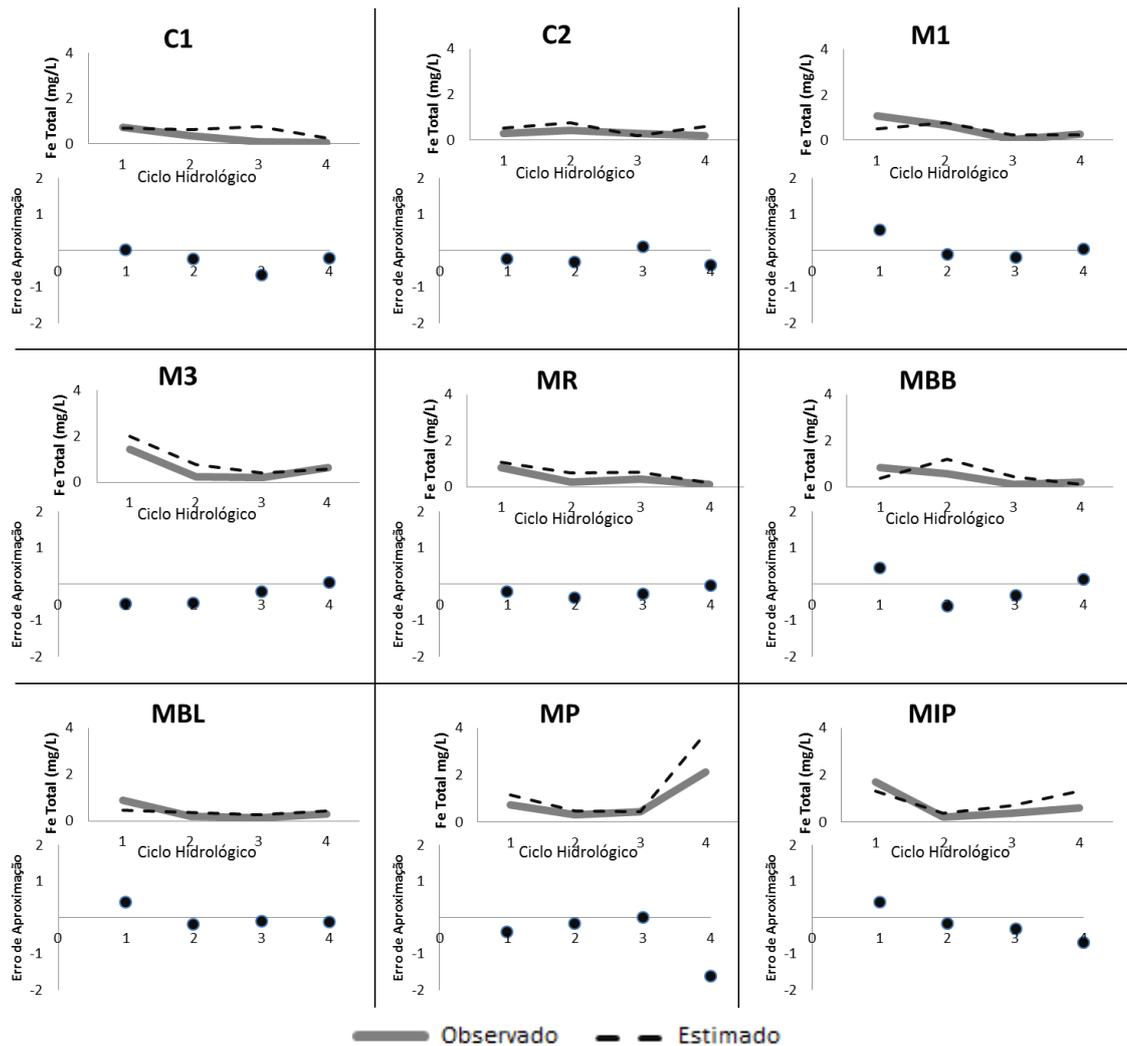


Figura 14 - Validação dos Sensores virtuais para o Fe total (mg/L) considerando o ciclo hidrológico da região e o Erro de aproximação.

Na validação dos Fosfatos de PO_4 , a Figura 15 mostra os valores observados e estimados da concentração, por ciclo hidrológico com as curvas não muito similares, em decorrência das grande variações entre os anos, com maiores concentrações no ano de 2009. O parâmetro PO_4 não obteve boas estimações nas estações pelo fato dos estimadores M nos revelarem altas concentrações no ano de 2009 e baixas concentrações no ano da validação (Tabela 3).

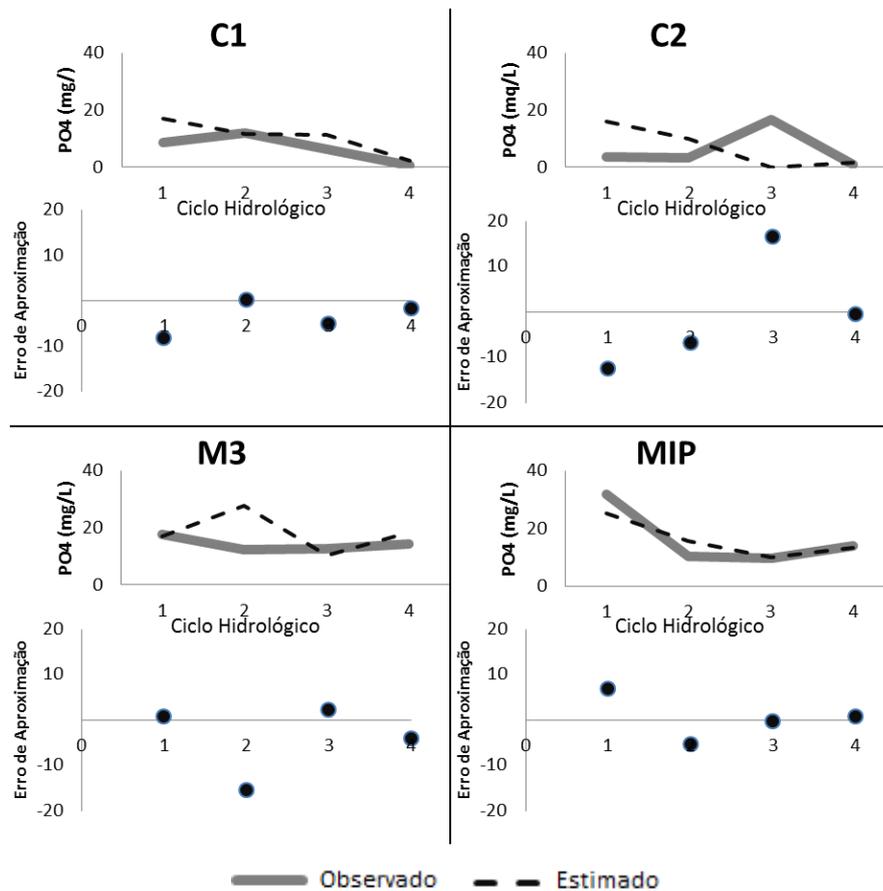


Figura 15 - Validação dos Sensores virtuais para o PO_4 (mg/L) considerando o ciclo hidrológico da região e o Erro de aproximação.

Na validação do Fosforo Total (Figura 16), as concentrações do P total foram muito similares nas estações de coleta C1, C2, MR e MBL. Houve discrepâncias deste parâmetro nas estações M1 e MBB teve alta concentração no ano da validação prejudicando sua estimação. Na estação M3 existem discrepâncias entre as concentrações nos anos de 2008 e 2009.

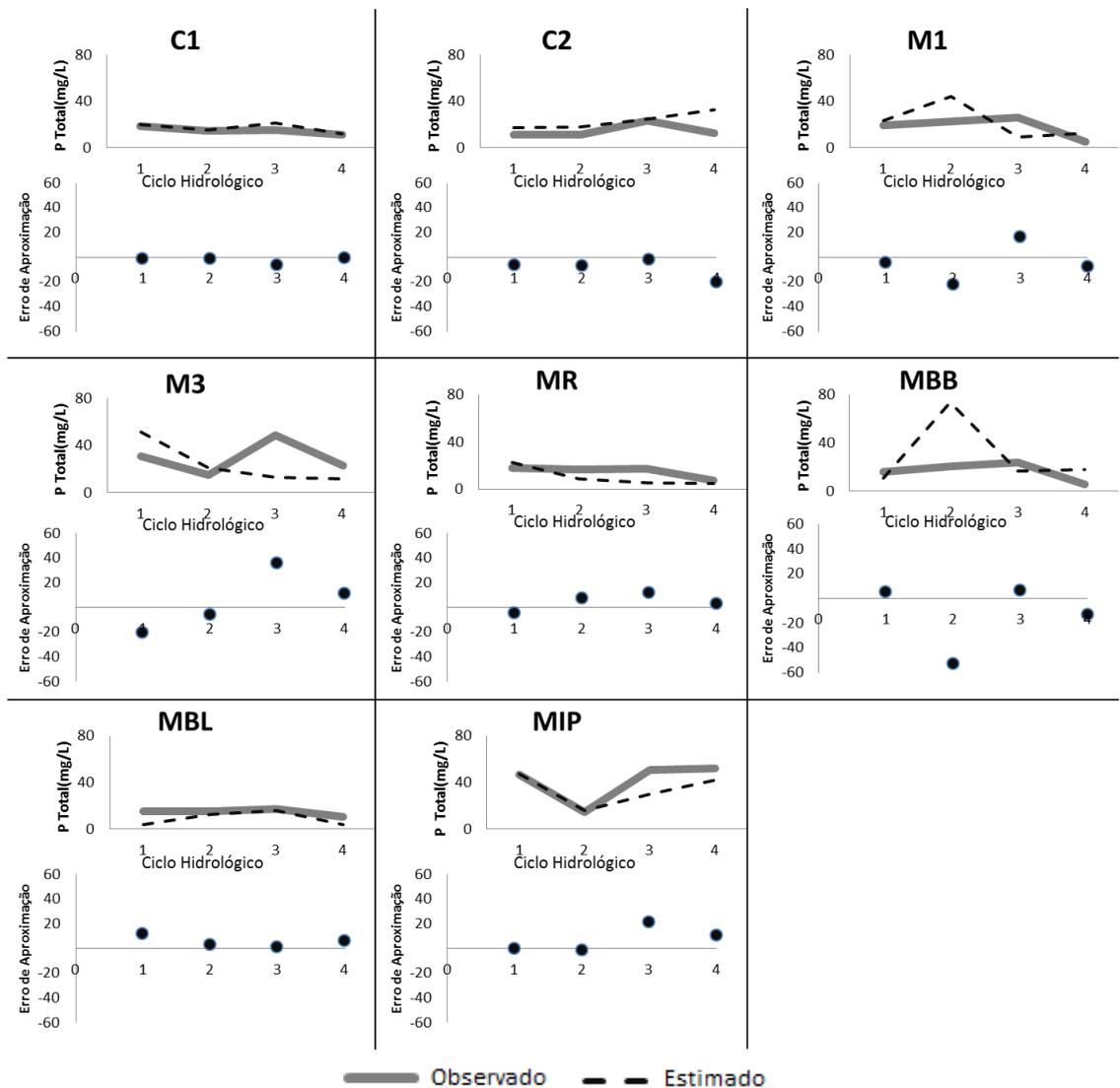


Figura 16 - Validação dos Sensores virtuais para o P total (mg/L) considerando o ciclo hidrológico da região e o Erro de aproximação.

Na validação do de STS, a Figura 17 mostra os valores observados e estimados da concentração por ciclo hidrológico com as curvas muito similares, excetuando a estação MP, em que o estimador M do ano de validação 2011(Tabela 3) é mais baixo do que os demais anos, o erro de aproximação demonstra essa discrepância.

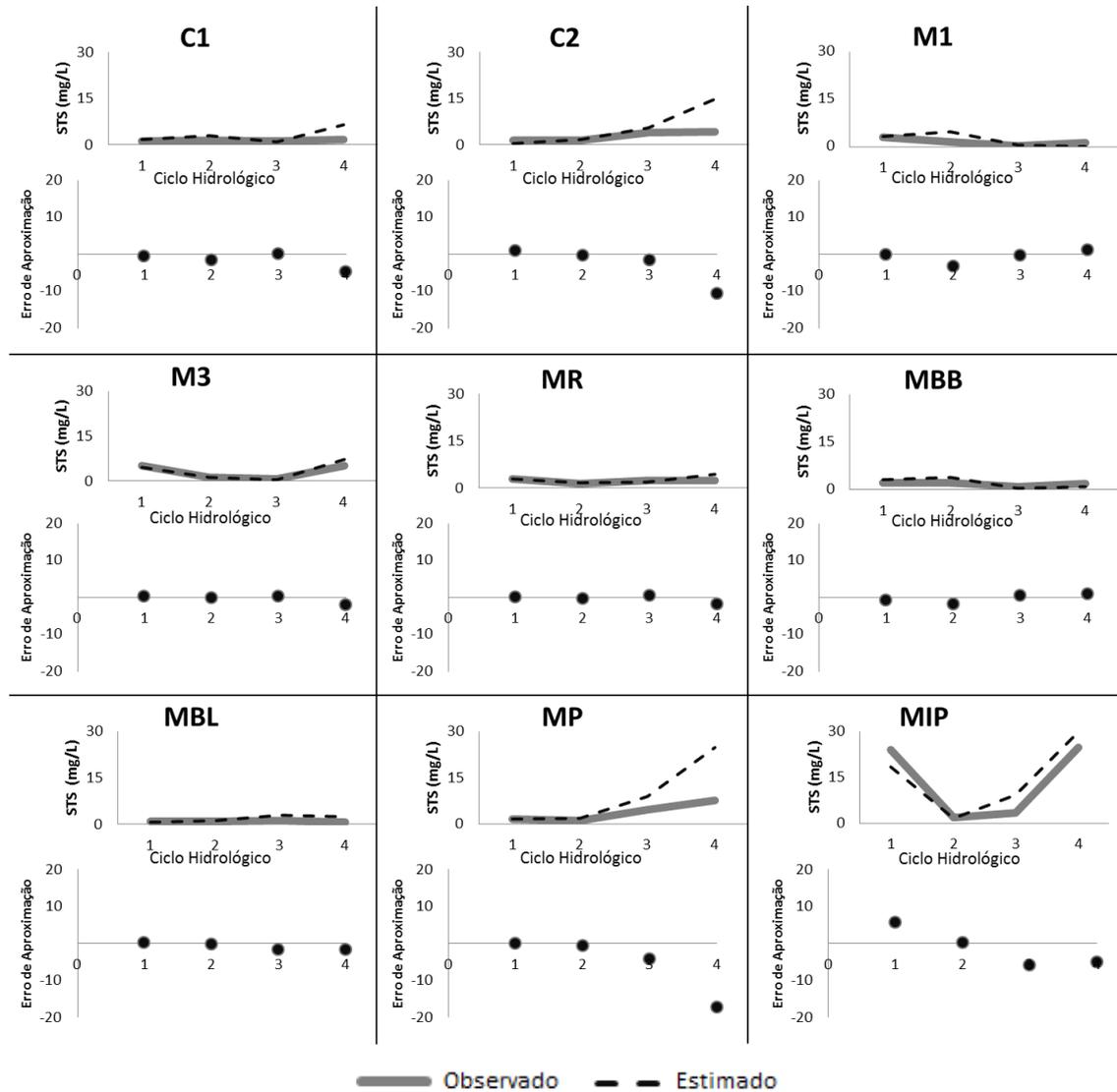


Figura 17 - Validação dos Sensores virtuais para o STS (mg/L) considerando o ciclo hidrológico da região e o Erro de aproximação.

Na validação da Turbidez (NTU), a Figura 18 mostra os valores observados e estimados da por ciclo hidrológico com as curvas similares, excetuando as estações C2, MBL, MIP e MP em que o estimador M do ano de validação 2011(Tabela 3) é mais baixo do que os demais anos, o erro de aproximação retrata essas discrepâncias.

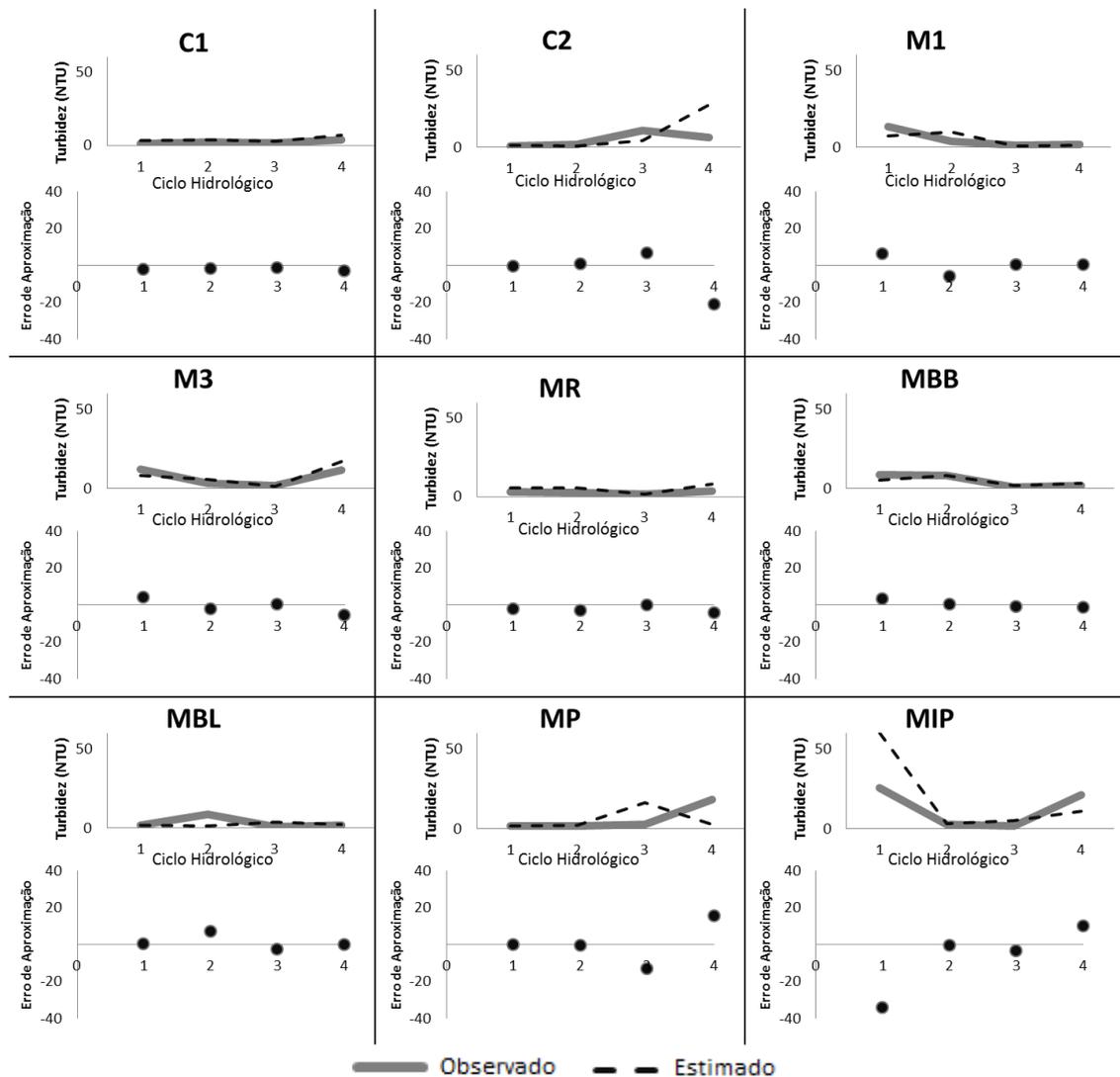


Figura 18 - Validação dos Sensores virtuais para Turbidez (NTU) por ciclo hidrológico e o Erro de aproximação.

5. CONCLUSÕES

Este estudo demonstrou a aplicação de sensores virtuais na estimação da concentração de parâmetros físico-químicos e metais em corpos d'água a partir de imagens de satélite da banda espectral 2 que possui sensibilidade à presença de materiais particulados em suspensão de corpos d'água. Através da análise fatorial foi confirmada a correlação entre os parâmetros físico-químicos: na faixa de primeira profundidade do disco de Secchi, PO₄, P total, STS e Turbidez e o ferro total. Os resultados das estimativas, em vista do reduzido número de distorções entre os resultados estimados pelos sensores virtuais e os observados pela análise das amostras em laboratório no teste de validação, resultaram numa boa aproximação conforme as figuras 13, 14, 17 e 18, comprovando respectivamente que a primeira faixa de profundidade do disco de Secchi, Fe Total, STS, e Turbidez do reservatório de água, podem ser monitorados por sensores virtuais. No caso específico dos parâmetros PO₄ e P Total, figuras 15 e 16, além do problema do pouco número de estações de coleta de dados disponíveis e a variabilidade inerente ao ciclo hidrológico da região demonstrada na Tabela 3, constatou-se, por meio da interpretação das imagens, falta de similaridades entre os dados utilizados no treinamento e na validação da RNA implementada, o que foi comprovado pelos resultados obtidos pelos sensores treinados, levando a uma baixa eficácia na predição desses parâmetros via imagens de satélite.

Portanto, demonstrou-se a eficácia da aplicação de sensores virtuais no monitoramento de parâmetros que afetam a qualidade da água de reservatórios na Amazônia por meio de imagens de satélite, proporcionando uma alternativa de boa aproximação em predições com menos dispêndios de recursos no processo de monitoramento ambiental. A solução proposta neste estudo é específica para o estudo das

águas do reservatório da UHE de Tucuruí, nada impedindo que a metodologia utilizada possa ser utilizada no monitoramento da qualidade da água de outros reservatórios.

6. TRABALHOS FUTUROS

Neste estudo foi abordado a aplicação de sensores virtuais na estimação dos parâmetros físico-químicos: Secchi (m), STS(mg/L), Turbidez (NTU), Fósforos e Fosforo total (PO_4 e P total) e do Ferro total (Fe total) na primeira faixa de profundidade. Como trabalho futuro é proposto aplicar a mesma metodologia para outros tipos de metais e outros parâmetros físico-químicos em diferentes faixas de profundidades, verificando e estimando as concentrações destes elementos em diferentes profundidades.

Outra possibilidade de trabalho futuro é o desenvolvimento de um software que possa alocar as estações de coletas de acordo com a representatividade das concentrações dos parâmetros físico-químicos e metais por meio de imagens de satélite e técnicas de clusterização, abordando técnicas da estatística e da inteligência computacional.

REFERÊNCIAS BIBLIOGRÁFICAS

ABNT . Associação Brasileira de Normas Técnicas – ABNT/NBR 9896. Glossário de poluição das águas. 1993.

ATOBATELE, O. E.; UGWUMBA, O. A. Seasonal variation in the physicochemistry of a small tropical reservoir (Aiba Reservoir, Iwo, Osun, Nigeria). **African Journal of Biotechnology**, 7, Junho 2008. 1962-1971.

BAZI, Y., ALAJLAN, N. and MELGANI, F. “Improved Estimation of Chlorophyll Concentration with Semisupervised Gaussian Process Regression”. **IEEE Transactions on Geoscience and Remote Sensing**, vol 50, 7, pp.2733-2743, 2012.

BENNETT, N. D., CROKE, B.F.W., GUARISO, G., GUILLAUME, J. H.A., HAMILTON, S. H., JAKEMAN, A. J., MARSILI-Libelli, S., NEWHAMA, L.T.H., NORTON, J. P., PERRIN, C., PIERCE, S. A., ROBSON, B., SEPPELT, R., VOINOV, A. A., FATH, B.D. and ANDREASSIAN, V. “Characterising performance of environmental models”. **Environmental Modelling & Software**, vol 40, pp.1-20, 2013.

BRANDO, V.E. and DEKKER, A.G. “Satellite hyperspectral remote sensing for estimating estuarine and coastal water quality”. **IEEE Transactions on Geoscience and Remote Sensing**, vol 41, 6, pp.1378-1387, 2003.

CHAPMAN, D. (Ed.). **Water Quality Assessments: A guide to the use of biota, sediments and water in environmental monitoring**. 2. ed. London: UNESCO/WHO/UNEP, 1996. 651 p.

COLETTI, C.; TESTEZLAF, R.; RIBEIRO, T. A. P.; SOUZA, R. T. G.; PEREIRA, D.A. Water quality index using multivariate factorial analysis. **Revista Brasileira de Engenharia Agrícola e Ambiental**, Campina Grande, 14, Março 2010. 517–522.

DAVIES, O. A.; ABOLUDE, D. S.; UGWUMBA, A. A. A. Phytoplankton of the lower reaches of Okpoka creek, Port Harcourt, Nigeria. **Journal of Fisheries International**, 3, 2008. 83-90.

DEMORI, M., FERRARI, V., STRAZZA, D., POESIO, P. A capacitive sensor system for the analysis of two-phase flows of oil and conductive water . **Int. J. Sensors and Actuators A: Physical** 163 (2010) 172–179.

DEEKAE, S. N.; ABOWEI, J. F. N.; ALFRED-OCKIYA, J. F. Seasonal Variation of Some Physical and Chemical Parameters of Luubara Creek, Ogoni Land, Niger Delta, Nigeria. **Research Journal of Environmental and Earth Sciences**, 2, Outubro 2010. 208-215.

DEUSTCH, C.V. and JOURNAL, A.G, **GSLIB: Geostatistical software Library and User's Guide**. Beta Version. Oxford University press, New York, 1996. 316p.

EÇA, Gilmara Fernandes. Biogeoquímica de nutrientes e Clorofila-a na plataforma continental rasa entre Itacaré e Canavieiras – Bahia. Dissertação de Mestrado.UESC/PPGSAT, 2009.

ELETRONORTE. **Manual do sistema de gestão ambiental - UHE Tucuruí**. ELETRONORTE. Tocantins. 1988.

ESTEVES, F. A. **Fundamentos de Limnologia**. 2. ed. Rio de Janeiro: Interciência, 1998. 602 p.

FISCH, G. F.; JANUÁRIO, M.; SENNA, R. C. Impacto ecológico em Tucuruí (PA): Climatologia. **Acta Amazonica**, 20, 1990. 49-60.

FORTUNA , L., GRAZIANI, S., RIZZO, A., XIBIALIA , M.G.. **Soft Sensors for Monitoring and Control of Industrial Process**, Springer. 2007.

GIROSI F., JONES, M. and POGGIO T., “Regularization Theory and Neural Networks Architectures”, **Neural Computation**, vol. 7, pp.219-269, 1995.

GOMES, G. S. S.; LUDERMIR, T. B. Redes neurais artificiais com funções de ativação complemento loglog e probit para aproximar funções na presença de observações extremas. Learning and nonlinear models. **Revista da Sociedade Brasileira de Redes Neurais (SBRN)**, 6, 2008. 142-153.

GÓMEZ, J. R., MONTEAGUDO, J. P. **A virtual sensor for monitoring excited thermal turbogenerator rotors**. Ingeniería e Investigación. Vol. 32, No. 3. p. 5-9, 2012.

GONZAGA, J. C. B.; MELEIRO, L. A. C.; KIANG, C.; MACIL-FILHO R. ANN-Based softsensor for real-time process monitoring and control industrial polymerization process. **Computers and Chemical Engineering**. v. 33, p. 43 – 49, 2009.

GONZALEZ, R. C., WOODS, R.E. **Digital Image Processing**. Second Edition. Upper Saddle River NJ: Prentice Hall, 2002.

HAIR, J. F.; TATHAM, R. L.; ANDERSON, R. E.; BLACK, W. **Multivariate data analysis**. 5. ed. New Jersey: Prentice Hall, 1998.

HAMPEL, F.R., RONCHETTI, E.M., ROUSSEUW, P.J. and STAHEL, W. A.. **Robust Statistics. The approach based on influence functions**. New York, USA: John Wiley and Sons, 1986.

HARTKAMP, A.D.; De BEURS, K.; STEIN, A. and WHITE, J.W.. Interpolation Techniques for Climate Variables. **NRG-GIS Series** 99-01. Mexico, D.F.: CIMMYT, 1999. 34p.

- HAUSER-DAVIS, R. A.; OLIVEIRA, T. F.; SILVEIRA, A. M.; PROTÁZIO, J. M. B.; ZIOLLI, R. L. Logistic regression and fuzzy logic as a classification method for feral fish sampling sites. **Environ Ecol Stat**, 19, 2012. 473–483.
- HAUSER-DAVIS, R. A.; OLIVEIRA, T. F.; SILVEIRA, A. M.; SILVA, T. B.; ZIOLLI, R. L. Case study: comparing the use of nonlinear discriminating analysis and Artificial Neural Networks in the classification of three fish species: acarás (*Geophagus brasiliensis*), tilapias (*Tilapia rendalli*) and mullets (*Mugil liza*). **Ecological Informatics**, 5, 2010. 474–478.
- HAYKIN S. “**Neural Networks – A Comprehensive Foundation**”, Macmillan College Publishing Inc., 1994.
- HEDEGARD, M.; WIK, T. An online method for estimation of degradable substrate and biomass in an aerated activated sludge process. **Water Research**, v. 45, p. 6308-6320, 2011.
- HOTTELLING, H., 1933. Analysis of a complex of statistical variables into principal components. **J. Educ. Psy.** 24: 417-441.
- JOHNSON, R. A.; WICHERN, D. W. **Applied multivariate statistical analysis**. 5. ed. [S.l.]: Prentice Hall, 2001. 767 p.
- KABATA-PENDIAS, A. **Trace elements in soils and plants**. 4. ed. Boca Raton, Florida: Taylor and Francis Group, 2011.
- KATO, M.T., “Fósforo”. Roteiro de aula da disciplina Qualidade da Água, do Ar e do Solo. Escola de Engenharia Mauá, 1983.
- KAZI, T. G.; ARAIN, M. B.; JAMALI, M. K.; JALBANI, N.; AFRIDI, H. I.; SARFRAZ, R. A.; BAIG, J. A.; SHAH, A. Q. Assessment of water quality of polluted lake using multivariate statistical techniques: A case study. **Ecotoxicology and Environmental Safety**, 72, 2009. 301– 309.
- KUSIAK, A.; LI, M.; ZHENG, H. Virtual models of indoor-air-quality sensors. **Applied Energy** V.87 p. 2087–2094, 2010.
- LEE, J. Y. and WARNER, T. A., 2006, Segment based image classification. **International Journal of Remote Sensing**, 27, pp. 3403–3412
- LEITE, M. A.; DORNFELD, C. B.; ESPÍNDOLA, E. L. G. Análise da concentração de metais na água do reservatório de Salto Grande: diagnóstico ambiental e de saúde pública. Disponível em: <[HTTP://www.bvsde.paho.org/bvsAIDIS/PuertoRico29/busso.pdf](http://www.bvsde.paho.org/bvsAIDIS/PuertoRico29/busso.pdf)>. Acesso em: 23 Abril 2013.
- LILLESAND, T. M.; KIEFER, R.W.; CHIPMAN, J.W. **Remote sensing and image interpretation**. 5. ed. New York: J. Wiley & Sons, 2004. 763 p.
- MARCOULIDES, G.A., HERSHBERGER, S.L. (1997). **Multivariate statistical methods**. Mahwah, NJ: Lawrence Erlbaum Associates.

MATHER, P.M. *Computer Processing of Remotely Sensed Images: an Introduction*. StEdmundsburi Press Ltd, 1987.

MATHWORKS. Neural Network Toolbox User's Guide: for use MATLAB, Matlab 6 p1, 2009.

MATTHEWS, M. W.; BERNARD, S.; WINTER, K. Remote sensing of cyanobacteria-dominant algal blooms and water quality parameters in Zeekoevlei, a small hypertrophic lake, using MERIS. **Remote Sensing of Environment**, 114, 2010. 2070–2087.

MENESES, P. R. e ALMEIDA ,T. Introdução ao processamento de Imagens de sensoriamento remoto. Universidade de Brasilia: Brasilia, 2012. 266p.

MICHALOPOULOU , Z., BAGHERI, S. and AXE, L. “Bayesian Estimation of Optica Properties of Nearshore Estuarine Waters: A Gibbs Sampling Approach”. *IEEE Transactions on Geoscience and Remote Sensing*, vol 48, 3, pp.1579-1587, 2010.

MOHLER, I.; GALINEC, G.; HÖLBLING, N.; BOLF, N.; UJEVIC, Z. Soft sensors for diesel fuel property estimation. **Chemical Engineering Transactions**. v. 21, p. 1423–1428, 2010.

MORRISON, D.F. **Multivariate statistical methods**. 3rd. Ed. McGraw-Hill, New York, 1990.

MORAIS, A. A., SANTOS, R.R.C., SODRÉ, C. H. , CARVALHO, F.O.,VIEIRA, W.G. Uma estratégia de Construção de um “Soft Sensor” Para Inferenciar a composição do produto de um processo de destilação utilizando técnica neuro-fuzzy. **Mecânica Computacional** Vol XXIX, págs. 9345-9365, Buenos Aires, Argentina, 15-18 Noviembre 2010.

NEGNEVITSKY, M. **Artificial Intelligence: A guide to intelligent systems**. 2. ed. [S.l.]: Pearson Education, 2005. 415 p.

OUYANG, Y. Evaluation of river water quality monitoring stations by principal component analysis. **Water Research**, n. 39, 2005. 2621–2635.

PADHY ,N.P. *Artificial Intelligence and Intelligent Systems*. New Delhi: Oxford University Press, 2005.

PAPAFILIPPAKI, A. K.; KOTTI, M. E.; STAVROULAKIS, G. G. Seasonal variations in dissolved heavy metals in the Keritis River, Chania, Greece. **Global NEST Journal**, 10, 2008. 320-325.

PEREIRA FILHO, W., dos SANTOS. F.C., CASSOL, A.P.V., DOMINGUES, A.L., PRADO, D.A. “Influência de componentes oticamente ativos relacionados a reservatórios em cascata – Rio Jacuí-RS”, in **Simpósio Brasileiro de Sensoriamento Remoto**, Foz do de Igauçu, PR, Brasil, 2013.

POMPÊO, M.L.M. **O disco de Secchi**. *Bioikos*, Campinas,v. 13, n. 1/2. p. 40-45, 1999.

RIBEIRO NETO, B. S. ; DAVIS, R. A. H. ; OLIVEIRA, T. F. ; SILVERA, A. M. . Estimating Physicochemical Parameters and Metal Concentrations in Hydroelectric Reservoirs by Virtual Sensors: A Case Study in the Amazon Region. **Computer Science and Engineering**, v. 4, p. 43-53, 2014.

RIVERA, E.A.C.; FARIAS, F.J.; ATALA, D.J.P; ANDRADE, R. R.; COSTA, A.C; MACIEL FILHO, R.. Development and implementation of an automated monitoring system for improved bioethanol production. **Chemical Engineering Transactions**, v. 18, p. 451-456, 2009.

RUSSELL, S.; NORVIG, P. **Artificial Intelligence: A modern approach**. 3. ed. [S.l.]: Pearson Education, 2010. 1132 p.

SCHOWENGERDT, R. A. **Remote Sensing Models and Methods for Image Processing**. 3a ed.,Elsevier, 2007.

SILVA, I.N., SPATTI D. e FLAUZINO, R. **Redes neurais artificiais para engenharia e ciências aplicadas: curso prático**, Artliber Editora Ltda, São Paulo, SP, Brasil, 2010 ISBN978-85-88098-53-4.

STATISTICS. **Methods and Applications**. Pawel Lewicki and Thomas Hill, Statsoft, 2005.

TUNDISI, J. G.; SANTON, M. A.; MENEZES, C. F. S. **Tucuruí reservoir and hydroelectric Power plant**. [S.l.]. 2003.

TUNDISI, J. G.; TUNDISI, T. M. **Limnologia**. São Paulo: Oficina de textos, 2008.

VALDAN, A.; SOUZA, M. B. Jr.; FOLLY, R.; VALDMAN, B. Online monitoring of exhaust gas emissions of a boiler with diesel/biodiesel fuel blends. **Chemical Engineering Transactions**, v. 24, p. 277-282, 2011.

WANG, X. L.; YIN, Z. J.; LV, Y. B.; LI, S. F. Operating rules classification system of water supply reservoir based on Learning Classifier System. **Expert Systems with Applications**, 36, Abril 2009. 5654–5659.

WENNER, E.; SANGER, D.; ARENDT, M.; HOLLAND, A. F.; CHEN, Y. Variability in dissolved oxygen and other water-quality variables within the national estuarine research reserve system. **Journal of Coastal Research**, 45, 2004. 17-38.

YAP,W.K. and KARRI, V. ANN virtual sensors for emissions prediction and control. **Applied Energy** v.88 p.4505–4516, 2011.

YAP, W. K., HO, T., KARRI, V. “Exhaust emissions control and engine parameters optimization using artificial neural network virtual sensors for a hydrogen-powered vehicle”. **International journal of hydrogen energy**, vol.37, pp.8704 - 8715, 2012.

YAP ,W. K. and KARRI, V. Comparative analysis of artificial neural networks and dynamic models as virtual sensors. **Applied Soft Computing** v.13 p.181–188 , 2013.

YENILMEZ, F.; KESKIN, F.; AKSOY, A. Water quality trend analysis in Eymir Lake, Ankara. **Physics and Chemistry of the Earth, Parts A/B/C**, 36, 2011. 135-140.

ZAHRAIE, B.; HOSSEINI, S. M. Development of reservoir operation policies considering variable agricultural water demands. **Expert Systems with Applications**, 36, Abril 2009. 4980–4987.

ZHU, W. and YU, Q. “Inversion of Chromophoric Dissolved Organic Matter”. ”. **IEEE Transactions on Geoscience and Remote Sensing**, vol 51, 6, pp.3286-3298, 2013.