# UNIVERSIDADE FEDERAL DE MATO GROSSO INSTITUTO DE FÍSICA PROGRAMA DE PÓS-GRADUAÇÃO EM FÍSICA AMBIENTAL

# ANÁLISE DE DESEMPENHO DE PRODUTOS MODIS PARA MODELAGEM DA DINÂMICA DE INUNDAÇÃO DO PANTANAL MATO-GROSSENSE

# SÉRGIO WAGNER GRIPP DA SILVEIRA

PROF. DR. PETER ZEILHOFER Orientador

> Cuiabá, MT Março de 2015

# UNIVERSIDADE FEDERAL DE MATO GROSSO INSTITUTO DE FÍSICA PROGRAMA DE PÓS-GRADUAÇÃO EM FÍSICA AMBIENTAL

# ANÁLISE DE DESEMPENHO DE PRODUTOS MODIS PARA MODELAGEM DA DINÂMICA DE INUNDAÇÃO DO PANTANAL MATO-GROSSENSE

# SÉRGIO WAGNER GRIPP DA SILVEIRA

Tese apresentada ao Programa de Pós-Graduação em Física Ambiental da Universidade Federal de Mato Grosso, como parte dos requisitos para obtenção do título de Doutor em Física Ambiental.

# PROF. DR. PETER ZEILHOFER Orientador

Cuiabá, MT Março de 2015

### Dados Internacionais de Catalogação na Fonte.



Ficha catalográfica elaborada automaticamente de acordo com os dados fornecidos pelo(a) autor(a).

### Permitida a reprodução parcial ou total, desde que citada a fonte.

# UNIVERSIDADE FEDERAL DE MATO GROSSO INSTITUTO DE FÍSICA Programa de Pós-Graduação em Física Ambiental

### FOLHA DE APROVAÇÃO

### TÍTULO: ANÁLISE DE DESEMPENHO DE PRODUTOS MODIS PARA MODELAGEM DA DINÂMICA DE INUNDAÇÃO DO PANTANAL MATO-GROSSENSE

### AUTOR: SÉRGIO WAGNER GRIPP DA SILVEIRA

Tese de Doutorado defendida e aprovada em 25 de março de 2015, pela comissão julgadora:

**Prof. Dr. Peter Zeilhofer Orientador** Instituto de Ciências Humanas e Sociais/ UFMT

**Prof. Dr. Ibraim Fantin da Cruz Examinador Externo** Faculdade de Arquitetura, Engenharia e Tecnologia UFMT

Profa. Dra. Julia Arieira Couto Examinadora Externa Universidade de Cuiabá - UNIC

Prof. Dr. Marcelo Sacardi Biudes Examinador Interno Instituto de Física – UFMT

Prof. Dr. Osvaldo Borges Pinto Junior

**Examinador Interno** Universidade de Cuiabá - UNIC

Prof. Dr. João dos Santos Vila da Silva Examinador Externo Centro Nacional de Pesquisa Tecnológica em Informática para a Agricultura EMBRAPA – Campinas/SP Dedico este trabalho à minha amada esposa, Deize, e aos meus filhos, Denis, David e Dayanne.

Também à minha mãe, Enir Gripp da Silveira, e ao meu pai, Joel Bastos da Silveira, pois desde pequeno me ensinaram a importância dos estudos.

### AGRADECIMENTOS

A Deus, minha fonte de inspiração, força e esperança.

Ao Prof. Dr. Peter Zeilhofer por toda sua dedicação na orientação deste trabalho, sem a qual este trabalho não teria alcançado este resultado.

Ao coordenador do Programa de Pós-Graduação em Física Ambiental, Prof. Dr. José de Souza Nogueira, o Paraná, por toda ajuda, apoio e incentivo, com palavras certas nos momentos certos.

Ao Prof. Dr. Osvaldo Borges Pinto Junior, por toda ajuda em estatística e no uso da Linguagem e Ambiente para Computação Estatística, R. Também por ter aceitado o convite para participar da minha banca de qualificação e de defesa de tese, dando suas importantes contribuições.

Ao Prof. Dr. Ibraim Fantin da Cruz, que me forneceu os dados de monitoramento da inundação usados nesse trabalho. Também por sua participação na minha banca de qualificação e defesa de tese, dando suas importantes contribuições.

Ao Prof. Dr. Marcelo Sacardi Biudes, que participou da minha banca de qualificação e defesa de doutorado, dando suas importantes contribuições para a melhoria desse trabalho.

Ao Prof. Dr. João dos Santos Vila da Silva, do Centro Nacional de Pesquisa Tecnológica em Informática para a Agricultura, EMBRAPA de Campinas, por ter aceitado o convite para participar da minha banca de defesa de doutorado, dando suas importantes contribuições.

À Prof.<sup>a</sup> Dr.<sup>a</sup> Julia Arieira Couto, da Universidade de Cuiabá - UNIC, por ter aceitado o convite para participar da minha banca de defesa de doutorado, dando suas relevantes contribuições.

Ao Prof. Dr. Thiago Sanna Freire Silva, da UNESP de Rio Claro - SP, que, apesar de não ter participado da minha banca examinadora de defesa, participou da banca de qualificação, com críticas e sugestões que foram muito importantes para melhoria dessa tese.

Ao Prof. Dr. Carlo Ralph De Musis pelas aulas de Estatística e pelas diversas vezes em que ele me ajudou no esclarecimento dessa ciência.

À Prof.<sup>a</sup> Dr.<sup>a</sup> Marta Cristina de Jesus Albuquerque Nogueira pelas aulas de Metodologia da Pesquisa Científica.

A todos os professores do Programa de Pós-graduação em Física Ambiental por compartilharem seus conhecimentos.

À FAPEMAT através do Programa de Apoio a Núcleos de Excelência

(PRONEX) pelo apoio financeiro ao projeto de pesquisa: "Nutrientes e Pesticidas nas Águas Superficiais das principais Bacias do Pantanal Setentrional: Uma Abordagem Integrada". Processo n. 840905/2009.

Ao Governo do Estado de Mato Grosso por ter me concedido a licença para qualificação profissional para fazer o doutorado.

Ao Dr. Pawel Rozenfeld e ao Me. Glauber Paz Miranda, ambos do Instituto Nacional de Pesquisas Espaciais, pelo apoio no tocante à flexibilização do meu horário de trabalho, sem a qual não seria possível a realização dessa tese.

A todos os colegas do Programa de Pós-graduação em Física Ambiental, pela amizade e compartilhamento de seus conhecimentos, em especial a George Sanches Suli, Victor Hugo de Morais Danelichen, Fernando da Silva Sallo, Vinícius Marcio Rodrigues Pereira, Erondina Azevedo de Lima, Paula Valéria de Carvalho, Geison Jader Mello, Suzi Mara Klemp, Gracyeli Santos Souza e Jéssica da Silva Linhares.

Ao secretário Cesário e às secretárias Soilce e Juliana do Programa de Pós-graduação em Física Ambiental, por toda dedicação e presteza referente às atividades de secretaria e por toda a amizade.

"No Pantanal ninguém pode passar régua. Sobremuito quando chove. A régua é existidura de limite. E o Pantanal não tem limites." (Manoel de Barros)

# SUMÁRIO

	LISTA DE FIGURAS	xii
	LISTA DE TABELAS	xvii
	LISTA DE ABREVIATURAS E SIGLAS	xix
	LISTA DE SÍMBOLOS	xxi
	RESUMO	xxii
	ABSTRACT	xxiv
	INTRODUÇÃO	1
1	REVISÃO BIBLIOGRÁFICA	5
1.1	O SENSOR ORBITAL MODIS	5
1.2	ÍNDICES ESPECTRAIS	5
1.2.1	ΝΟΥΙ	6
1.2.2	ΕVΙ	6
1.2.3	LSWI	6
1.2.4	ΝΟΨΙ	7
1.2.5	ΜΝΟΨΙ	8
1.3	QUALIDADE DAS IMAGENS MODIS	8
1.3.1	Influência da Atmosfera	9
1.3.2	Geometria da Aquisição da Imagem	10
1.4	PRODUTOS DO SENSOR MODIS	10
1.4.1	MOD13Q1 - Índice de Vegetação	10
1.4.2	MCD12Q1 - Uso e Ocupação da Terra	15
1.5	REGRESSÃO LOGÍSTICA	18
1.5.1	Teste de Significância dos Coeficientes	18
1.5.2	Métodos de Seleção das Variáveis Independentes	19
1.5.3	Medida da Qualidade de Ajuste do Modelo	19
1.5.3.1	Tabela de Classificação	20
1.5.3.2	Coeficiente Kappa	20
1.5.3.3	Área Sob a Curva ROC	21
1.5.3.4	Coeficiente de Determinação	21

1.5.3.5	Validação Externa	22
1.5.3.6	Validação Cruzada: k-fold	23
1.6	TIMESAT: SUAVIZAÇÃO E PREENCHIMENTO DE FA-	
	LHAS	23
1.7	MAPEAMENTO DA INUNDAÇÃO COM IMAGENS DE	
	SATÉLITE	24
2	MATERIAL E MÉTODOS	28
2.1	MATERIAL	28
2.1.1	Área de estudo	28
2.1.2	Medidas de Campo: Inundação e topografia do terreno	29
2.1.3	Produtos MODIS	32
2.2	ΜΈΤΟDOS	34
2.2.1	Processamento das medidas de inundação	34
2.2.2	Pré-processamento das imagens MODIS	35
2.2.3	Desenvolvimento do Modelo de Inundação	35
2.2.4	Preenchimento de Falhas	36
2.2.5	Análise da Influência da Qualidade dos Pixels Sobre o De-	
	sempenho do Modelo	37
2.2.6	Comparação dos pontos de monitoramento <i>in situ</i> com os	
	pixels MODIS	39
2.2.6.1	Geração do Modelo Numérico de Terreno	39
2.2.6.2	As Réguas Virtuais	41
3	APRESENTAÇÃO DOS RESULTADOS	44
3.1	O MODELO DE INUNDAÇÃO	44
3.1.1	Escolha das Variáveis Independentes	44
3.1.2	Qualidade de Ajuste do Modelo	46
3.1.3	Modelo Simplificado	50
3.2	COMPARAÇÃO PONTO COM PIXEL	50
3.2.1	Interpolação do Modelo Numérico de Terreno	50
3.2.2	Desempenho do Modelo na Utilização das Réguas Virtuais	53
3.3	VARIABILIDADE INTERANUAL DA QUALIDADE DO PRO-	
	DUTO MOD13Q1	54
3.4	DESEMPENHO DO MODELO EM FUNÇÃO DA QUALI-	
	DADE DOS PIXELS	57
3.5	DESEMPENHO DO MODELO EM FUNÇÃO DO CICLO	
	HIDROLÓGICO	60

3.6 3.7	DESEMPENHO EM FUNÇÃO DO TIPO DE COBERTURA VARIABILIDADE ANUAL DA INUNDAÇÃO MEDIDA E	64
	PREDITA	68
4	DISCUSSÃO DOS RESULTADOS	71
4.1	PODER PREDITIVO DAS VARIÁVEIS	71
4.2	O DESEMPENHO DAS RÉGUAS VIRTUAIS	73
4.3	VARIABILIDADE INTERANUAL DA QUALIDADE DO PRO- DUTO MOD13Q1 E A INFLUÊNCIA SOBRE AS CLASSI-	
	FICAÇÕES	75
4.4	A INFLUÊNCIA DO CICLO HIDROLÓGICO	78
4.5	DESEMPENHO DO MODELO EM FUNÇÃO DO USO E	
	OCUPAÇÃO DA TERRA	79
4.6	VARIABILIDADE ANUAL DA INUNDAÇÃO MEDIDA E	
	PREDITA	80
5	CONCLUSÕES	82
	REFERÊNCIAS	84
	APÊNDICES	96
	APÊNDICE A – PROCESSAMENTO DAS MEDIDAS DE CAMPO E PRÉ-PROCESSAMENTO DOS DADOS MODIS	97
	APÊNDICE B – SELEÇÃO DAS VARIÁVEIS E AJUSTE DO MODELO COM O USO DE IMA- GENS MODIS DE QUALIDADE PR_0 .	98
	APÊNDICE C – PREENCHIMENTO DE FALHAS E AJUSTE	E
	DO MODELO COM DADOS PREENCHI- DOS	99
	APÊNDICE D – CRIAÇÃO DOS CONJUNTOS DE NÍVEIS DE QUALIDADE QA_0000 A QA_1100	100

# LISTA DE FIGURAS

Figura 1 –	Efeitos da atmosfera que influenciam as medidas realizadas em	
	sensores passivos	8
Figura 2 –	Ilustração da geometria de aquisição de dados MODIS. A função	
	de distribuição da refletância bidirecional (BRDF) muda com	
	a geometria do ângulo de visada e ângulo de incidência solar $% f(x)=\int dx  dx$ .	9
Figura 3 $-$	Algoritmo de composição do produto MOD13Q1 coleção 5 $\ .$ .	14
Figura 4 $-$	Exemplo de gráfico ROC de três classificadores	22
Figura 5 $-$	Dados do índice de vegetação organizados em imagens (es-	
	querda). Uma imagem $i$ possui o valor de um determinado	
	pixel para cada tempo $t_i.$ Extraindo o valor do pixel (j, k) em	
	intervalos de tempos consecutivos, uma série temporal é obtida	
	para aquele pixel (direita) $\ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots$	24
Figura 6 $-$	Área de estudo apresentando o Sítio de Amostragem de Longa	
	Duração (SALD) Pirizal e o Transecto dentro da Reserva Par-	
	ticular do Patrimônio Natural do SESC Pantanal	29
Figura 7 $-$	Malha com 40 réguas limnimétricas instaladas no SALD Pirizal	
	em uma imagem SPOT, bandas $3/2/1$ de 2007	30
Figura 8 $-$	Transecto formado por 11 réguas na Reserva Particular do	
	Patrimônio Natural SESC Pantanal em uma imagem SPOT,	
	bandas $3/2/1$ de 2007 $\ldots$ $\ldots$ $\ldots$ $\ldots$ $\ldots$ $\ldots$ $\ldots$ $\ldots$	30
Figura 9 $-$	Estimativa da precipitação mensal feita por dados da missão	
	$\operatorname{TRMM}$ (produto 3B42), calculada a partir da média ponderada	
	da grade $(0,25^\circ)$ dentro da bacia de contribuição da régua	
	fluviométrica de Porto Cercado (gráfico de barras). Medida da	
	vazão do rio Cuiabá em Barão de Melgaço (a montante da área	
	de estudo, gráfico de linha)	32
Figura 10 –	Climograma da Fazenda Miranda para os anos de 2011 a 2013.	
	O eixo das abscissas indicam os meses, de julho $(J)$ a junho $(J)$ .	
	O eixo das ordenadas da esquerda corresponde à temperatura	
	do ar, com as médias mensais máxima e mínima em destaque	
	(cor preta). O eixo das ordenadas da direita corresponde à	
	precipitação pluviométrica.	33

Figura 11 –	Exemplo de série temporal (2004 a 2010) do índice LSWI de um pixel da área de estudo. A série original com falhas (pixels	
	nublados removidos) é representada pola linha azul. A série	
	roconstruída é dada pola linha marrom	38
Figure 19	Dispecieño des régues virtueis pa grade de 250 metros de	00
rigura 12 –	MODIS, na região do SALD Pirizal	41
Figura 13 –	Exemplo de extrapolação da inundação das réguas virtuais	42
Figura 14 –	Sensibilidade e especificidade do modelo em função do ponto	
0	de corte. A interseção das duas curvas indica o melhor ponto	
	de corte para o modelo	46
Figura 15 –	Capacidade de discriminação do modelo usando os índices	
0	LSWI, EVI e a variável categórica Land	48
Figura 16 –	Capacidade de discriminação do modelo usando apenas a va-	
	riável LSWI	48
Figura 17 –	Capacidade de discriminação do modelo usando apenas a va-	
	riável NDVI	49
Figura 18 –	Capacidade de discriminação do modelo usando apenas a va-	
	riável EVI	49
Figura 19 –	Interpretação visual de uma imagem SPOT de 2007 para iden-	
	tificar as cotas altimétricas do levantamento topográfico locali-	
	zadas nas cordilheiras do SALD Pirizal	51
Figura 20 –	Gráficos quantil-quantil para os pontos de levantamento topo-	
	gráfico (a) sem e (b) com descarte de medidas em cordilheiras	51
Figura 21 –	Modelo numérico de terreno do SALD Pirizal obtido por kriga-	
	gem dos pontos de levantamento topográfico após a exclusão dos $% \mathcal{A}$	
	pontos localizados nas cordilheiras. São apresentadas, também,	
	as quarenta réguas limnimétricas	52
Figura 22 –	a) Reta de regressão formada com as elevações estimadas e	
	medidas b) Gráfico quantil-quantil dos erros padronizados $\ . \ .$	53
Figura 23 –	Composição da imagem em relação aos valores de Pixel Relia-	
	bility (PR = 0, PR = 1 e PR = 3) $\dots \dots \dots \dots \dots \dots \dots$	54
Figura 24 –	Composição da qualidade dos pixels (barras) e porcentagem	
	de réguas inundadas (linha) para o ano de 2005, na região do	
	transecto do SESC	55
Figura 25 –	Composição da qualidade dos pixels (barras) e porcentagem	
	de réguas inundadas (linha) para o ano de 2006, na região do	
	transecto do SESC	56

Figura 26 –	Composição da qualidade dos pixels (barras) e porcentagem	
	de réguas inundadas (linha) para o período de 2006/2007, na	
	região do SALD Pirizal	56
Figura 27 –	Composição da qualidade dos pixels (barras) e porcentagem	
	de réguas inundadas (linha) para o período de 2007/2008, na	
	região do SALD Pirizal	57
Figura 28 –	Taxas de falso positivo (TFP) e falso negativo (TFN), gráfico	
	de barras, e porcentagem de réguas inundadas para as medidas	
	do transecto SESC RPPN do ano de 2005	60
Figura 29 –	Taxas de falso positivo (TFP) e falso negativo (TFN), gráfico	
	de barras, e porcentagem de réguas inundadas para as medidas	
	do transecto SESC RPPN do ano de 2006	60
Figura 30 –	Taxas de falso positivo (TFP) e falso negativo (TFN), gráfico	
	de barras, e porcentagem de réguas inundadas para as medidas	
	do SALD Pirizal, para o período de medida $2006/2007$	61
Figura 31 –	Taxas de falso positivo (TFP) e falso negativo (TFN), gráfico	
	de barras, e porcentagem de réguas inundadas para as medidas	
	do SALD Pirizal, para o período de medida $2007/2008$	61
Figura 32 –	Mapas de inundação da área de estudo predita pelo modelo	
	usando dados de qualidade PR_0. Ano hidrológico: outubro de	
	2006 a setembro de 2007	62
Figura 33 –	Mapas de inundação da área de estudo predita pelo modelo	
	usando dados reconstruídos pelo TIMESAT. Ano hidrológico:	
	outubro de 2006 a setembro de 2007	63
Figura 34 –	Mapas de inundação da região do transecto do SESC Pantanal	
	de 13 de agosto de 2007 (dia 225) gerados a partir do modelo	
	a) com dados reconstruídos pelo TIMESAT e b) com dados	
	originais de qualidade PR_0	64
Figura 35 –	Taxas de Falso Positivo (TFP) e Falso Negativo (TFN) em	
	função do uso e ocupação da terra e do ciclo de inundação. São	
	apresentadas as TFP e TFN para a série original com qualidade	
	PR_0 e para a série reconstruída pelo TIMESAT. Dados usados	
	no ajuste do modelo para as regiões do transecto do SESC e	
	do SALD Pirizal.	65

Figura 36 –	Mapas com a probabilidade de inundação dada pelo modelo em	
	três períodos de 2007 na região do SALD Pirizal: cheia, vazante	
	(transição) e seca. Três tipos de uso e ocupação da terra são	
	usados como referência: Floresta Inundável, Campo Cerrado e	
	Pastagem Exótica. Imagem do SPOT de 2007 para a área do	
	SALD Pirizal	66
Figura 37 –	Mapas de inundação na região do transecto do SESC em 13	
Ũ	de agosto de 2007 gerados pelo modelos a) simplificado e b)	
	completo. c) Uso e ocupação da terra na região do transecto	
	do SESC fornecidos pela camada <i>Land Cover Type 2</i> (UMD)	
	do produto MCD12Q1	67
Figura 38 –	Distribuição da probabilidade de inundação por pixel ( <i>Boxplot</i> )	
0	e porcentagem de réguas inundadas (linha vermelha). Proba-	
	bilidades obtidas no ajuste do modelo com dados originais de	
	qualidade PR 0. Dados de 2005 e 2006 para o transecto do	
	SESC e de 2006/2007 e 2007/2008 para SALD Pirizal. O pe-	
	ríodo de 16 dias corresponde à periodicidade em que o produto	
	MOD1301 é gerado	68
Figura 39 –	Distribuição da probabilidade de inundação por pixel ( <i>Bornlot</i> )	00
i iguita so	e porcentagem de réguas inundadas (linha vermelha). Probabi-	
	lidades obtidas no ajuste do modelo com dados reconstruídos	
	Dados de 2005 e 2006 para o transecto do SESC e de 2006/2007	
	e 2007/2008 para SALD Pirizal. O período de 16 dias corres-	
	ponde à periodicidade em que o produte MOD13O1 é gerado	60
Figure 40	Probabilidade de inundação por pivel (Boyplet) o porcentagem	09
$r_{1}gura 40 -$	de régues inundades (linha vermelha). Validação externa de	
	reguas mundadas (mina vermeina). Vandação externa do	
	line com dados originais de quandade P.K_0. O periodo de 10	
	dias corresponde a periodicidade em que o produto MODI3QI	70
D. 41		70
Figura 41 –	Probabilidade de inundação por pixel (Boxpiot) e porcentagem	
	de reguas inundadas (linna vermeina). Validação externa do mo-	
	delo com dados reconstruidos. O periodo de 16 dias corresponde	
D: 40	a periodicidade em que o produto MODI3QI e gerado.	70
Figura 42 –	Fluxograma apresentando as etapas de pré-processamento das	
	imagens MODIS e processamento das medidas de inundação .	97
Figura 43 –	Fluxograma apresentando a etapa de seleção de variáveis e	
	ajuste do modelo usando dados MODIS com qualidade PR_0	98

Figura 44 $-$	Fluxograma apresentando a etapa de processamento dos dados	
	MODIS com o uso do software TIMESAT para preenchimento	
	de falhas e suavização. Ajuste do modelo e medida de desempe-	
	nho do modelo $\ldots$	99
Figura 45 –	Fluxograma apresentando a etapa de criação dos conjuntos com	
	os níveis de qualidade QA_0000 a QA_1100 . Ajuste do modelo	
	para cada conjunto de qualidade, com medida de desempenho	
	do modelo para cada conjunto	100

# LISTA DE TABELAS

Tabela 1 –	Características do produto MOD13Q1 – Índice de vegetação .	11
Tabela 2 –	Confiabilidade dos pixels do produto MOD13Q1	11
Tabela 3 $$ –	Descrição do conjunto de dados científicos de avaliação de	
	qualidade dos índices de vegetação ( $V\!I\ Quality)$ do produto	
	MOD13Q1	12
Tabela 4 –	Método de escalonamento do parâmetro ${\it VI}$ ${\it Usefulness}$ para o	
	produto MOD13Q1	13
Tabela 5 $$ –	Descrição das camadas que compõem o produto MODIS de	
	classificação de uso e ocupação da terra MCD12Q1 $\ .$	16
Tabela 6 $\ -$	Descrição das bandas Land Cover Type 1, IGBP (Classificação	
	do Programa Internacional Geosfera-Biosfera) e Land Cover	
	$Type\ 2,\ UMD$ (Classificação da Universidade de Maryland)	16
Tabela 7 $$ –	Descrição das bandas Land Cover Type 3, LAI/fPAR (Sistema	
	de classificação usado pelo algoritmo LAI/FPAR do MODIS) $$	
	e Land Cover Type 4, $BGC$ (Classificação com oito tipos de	
	biomas) $\ldots$	17
Tabela 8 –	Descrição da banda Land Cover Type 5, $PFT$ (Tipo funcional	
	das plantas)	17
Tabela 9 –	Tabela de classificação	20
Tabela 10 –	Período das medidas de inundação obtidas através das réguas	
	do SALD Pirizal e das do Transecto do SESC Pantanal, e a	
	quantidade (n) de medidas em cada período	31
Tabela 11 –	Critério de formação dos conjuntos de dados do produto $\rm MOD13Q1$	
	criados com base nos níveis de qualidade dos pixel s $\ .\ .\ .$ .	39
Tabela 12 –	Resultado do ajuste do modelo de regressão logística com os	
	dados do SALD Pirizal e do transecto do SESC $\ .\ .\ .\ .$	44
Tabela 13 –	Fator de inflação da variância (VIF), para determinação de	
	multicolinearidade das variáveis independentes, e grau de li-	
	berdade (gl). Dados usados no ajuste do modelo para o SALD	
	Pirizal e transecto do SESC	45
Tabela 14 –	Resultado do ajuste do modelo de regressão logística com os	
	dados do SALD Pirizal e do transecto do SESC após excluir a	
	variável NDVI	45

Tabela 15 – Fator de inflação da variância (VIF), para determinação de	
multicolinearidade das variáveis independentes, e grau de li-	
berdade (gl), após excluir a variável NDVI. Dados usados no	
ajuste do modelo para o SALD Pirizal e transecto do SESC $$ .	46
Tabela 16 – Resultado da validação interna da regres são logística $\ .\ .\ .$	47
Tabela 17 – Resultado da validação cruzada tipo $k\mbox{-}fold$ com $k\mbox{ = }10$ da	
regressão logística	47
Tabela 18 – Resultado da validação externa da regressão logí stica $\ .\ .\ .$	47
Tabela 19 – Acurácia, taxas de falso positivo (TFP) e falso negativo (TFN),	
e coeficiente de concordância Kappa para as validações: interna	
(ajuste do modelo), cruzada tipo $k\text{-fold}~(k=10)$ e externa $\ .$ .	47
Tabela 20 – Variáveis na equação logística para o modelo sem a classificação	
do uso e ocupação da terra	50
Tabela 21 – Desempenho do modelo em função da porcenta gem de inunda-	
ção dos pixels dada pelas réguas virtuais	54
Tabela 22 – Desempenho do modelo em função da qualidade dos pixel s $\ .$	58
Tabela 23 – Desempenho do modelo em função da qualidade dos pixel para	
os dados adicionais ao conjunto $PR_0$	58
Tabela 24 – Resultado da classificação para a diferença entre os conjuntos	
PR_3 e QA_0110	59
Tabela 25 – Resultado da classificação para a diferença entre os conjuntos	
TIMESAT e $QA_0110$	59

# LISTA DE ABREVIATURAS E SIGLAS

AIC	Akaike information criterion	
AVHRR	Advanced Very High Resolution Radiometer	
BRDF	Bidirectional Reflectance Distribution Function	
CSV	comma-separated values	
CV-MCV	Constrained View angle - Maximum Value Composite	
DB	Deciduous Broadleaf	
EB	Evergreen Broadleaf	
EVI	Enhanced Vegetation Index	
FPR	False Positive Rate	
HDF	Hierarchical Data Format	
LP DAAC Land Processes Distributed Active Archive Center		
LSWI	Land Surface Water Index	
MIR	Middle Infrared	
MODIS	MODerate resolution Imaging Spectroradiometer	
MRT	MODIS Reprojection Tool	
MCV	Maximum Value Composite	
NASA	National Aeronautics and Space Administration	
NDVI	Normalized Difference Vegetation Index	
NDWI	Normalized Difference Water Index	
MNDWI	Modified Normalized Difference Water Index	
NIR	Near Infrared	
NOAA	National Oceanic and Atmospheric Administration	
PR	Pixel Reliability	

QA	Quality Assessment
ROC	Receiver Operating Characteristic
RPPN	Reserva do Patrimônio Particular Natural
S	Savanna
SALD	Sítio de Amostragem de Longa Duração
SWIR	Short Wave Infrared
TFN	Taxa de Falsos Negativos
TFP	Taxa de Falsos Positivos
TPR	True Positive Rate
TRMM	Tropical Rainfall Measuring Mission
USGS	U.S. Geological Survey
UTM	Universal Transversa de Mercator
WS	Wood Savanna

# LISTA DE SÍMBOLOS

ρ	Refletância espectral da superfície
α	Nível de significância
p	valor-p ou nível descritivo
$R^2$	Coeficiente de determinação

### RESUMO

SILVEIRA, S. W. G. Análise de Desempenho de Produtos MODIS para Modelagem da Dinâmica de Inundação do Pantanal Mato-Grossense,
2015, 100 f. Tese (Doutorado em Física Ambiental) - Instituto de Física,
Universidade Federal de Mato Grosso.

Os processos ecológicos de áreas úmidas, como o Pantanal Mato-grossense, dependem de sua inundação cíclica, cujo monitoramento *in loco* é inviável devido a sua grande extensão territorial e o difícil acesso. Portanto, o objetivo desse trabalho foi analisar a aplicabilidade de produtos MODIS para monitoramento da dinâmica de inundação, avaliando sistematicamente a interferência do esquema de amostragem em campo, da qualidade dos pixels, da cobertura vegetal, do período dentro do ciclo hidrológico e do uso de um algoritmo de preenchimento de falhas sobre o desempenho dos modelos classificatórios. Para desenvolver as classificações por regressão logística foram utilizadas medidas de inundação *in situ* de duas regiões do Pantanal Norte: o Sítio de Amostragem de Longa Duração do Pirizal, e a Reserva do Patrimônio Particular Natural do SESC Pantanal. As medições, codificadas de forma binária, foram comparadas com os índices espectrais NDVI, EVI e LSWI obtidos do produto MOD13Q1, e com os mapas anuais do uso e ocupação da terra do produto MCD12Q1. O grau de generalização dos modelos foi avaliado através de validações externa e cruzada. Um novo método, denominado réguas virtuais, foi introduzido para averiguar a interferência da comparação de medições pontuais *in situ* com as entidades zonais dos pixels MODIS. Nesse procedimento os níveis de inundação em uma grade virtual ao redor das réguas limnimétricas foram simulados a partir de um modelo numérico de terreno interpolado a partir de um levantamento topográfico existente. Na seleção das variáveis independentes para compor o modelo, a combinação dos índices LSWI e EVI e da classificação do uso e ocupação da terra, obteve o melhor desempenho, com  $R^2 = 0.75$  e Kappa = 0.78. A acurácia geral foi 89.2% no ajuste do modelo e 77.2% na validação externa. O modelo usando réguas virtuais, em termos gerais, não obteve melhor desempenho do que o modelo gerado com as réguas reais, embora a taxa de falso positivo tenha diminuído parcialmente. A exclusão de pixels até o nível de qualidade QA 0110 não melhorou de forma significativa o desempenho do modelo. Um decaimento mais expressivo do desempenho somente foi observado se pixels nublados são incluídos. O modelo usando os dados reconstruídos obteve excelente desempenho quando comparado com modelos usando dados originais em todos os níveis de qualidade dos pixels, tendo como vantagem a classificação de toda área

de estudo em cada passo da resolução temporal. Como esperado, o desempenho da classificação é reduzido em áreas florestadas, porém de forma menos expressiva do que esperado. A qualidade do modelo varia de acordo com os períodos do ciclo de inundação e é principalmente reduzida durante a enchente e vazante. Diferenças superiores a 10% nos indicadores dos acertos dos modelos apontam que somente uma validação externa permite uma avaliação realista sobre o desempenho do monitoramento de inundações por produtos MODIS na planície.

Palavras-chaves: Regressão logística, MOD13Q1, LSWI, EVI, NDVI.

### ABSTRACT

SILVEIRA, S. W. G. **Performance Analysis of MODIS Products for Modeling of North Pantanal Flood Dynamics**, 2015, 100 f. Thesis (Doctorate in Environmental Physic); Institute of Physic, Federal University of Mato Grosso.

The ecological processes of wetlands, such as the Pantanal of Mato Grosso, depend on their cyclic flooding, which field monitoring is unfeasible due to its vast extension and difficult access. Thus this study aimed to evaluate the applicability of MODIS products to monitor the inundation dynamics, assessing systematically the interference of field sampling scheme, pixel quality given by the Quality Assurance band, vegetation cover, period of the hydrological cycle and the use of a gap filling algorithm on the classification model performance. For logistic regression classification training and validation, inundation measurements of two field survey sites were considered: The Pirizal site of the Long Term Ecological Research Program (LTER/PELD), and the Private Natural Heritage Preservation Area (RPPN) of the SESC-Pantanal. Binary coded flood measurements were compared with the NDVI, EVI and LSWI spectral indices obtained from the MOD13Q1 product and annual land use and cover maps of the MCD12Q1 product. Degree of generalizability of the model was measured by external and cross-validation. A new method, called virtual stages, was introduced to test the influence of field point measurements with zonal entities represented by MODIS pixels. In this procedure, flooding levels in regular 25m grid around the limnimetric gauges were simulated by a Digital Elevation Model (DEM) interpolated from an existing topographic survey. The combination of the LSWI, EVI and land use and cover layer resulted in the best flooding predictions with a Nagelkerke  $R^2 = 0.75$  and Kappa = 0.78. The overall accuracy was 89.2% for the adjusted model and 77.2%in the external validation. A model setup using the virtual gauges did not result in a better overall performance of the models. Nevertheless, false positive rates were partially reduced. The omission of pixels down to a QA level of 0110 did not significantly improve the model performance. An expressive performance loss was only observed when pixels flagged as cloudy were included in the models. The filtered and filled in model showed excellent results, if compared with models based on different QA filter levels, having as further advantages the gap-free flood monitoring. Model performance was reduced in forested areas, however less expressive than expected, if compared with the performance in open savannah formations. Model performance varied seasonally, being reduced during rising and falling waters. Differences higher than 10% in the performance indicators showed

that only an external validation allow a realistic assessment of flood mapping performance with MODIS data in the floodplain.

Key-words: Logistic regression, MOD13Q1, LSWI, EVI, NDVI.

## INTRODUÇÃO

As áreas úmidas são regiões caracterizadas pela oscilação periódica entre fases terrestres e aquáticas e cobrem cerca de 4% a 6% da superfície terrestre, ocorrendo em todos os climas e continentes (MITSCH; GOSSELINK, 2000). A sua maioria, porém, é localizada nos trópicos, onde a combinação das condições ambientais com alta incidência de radiação solar, altas precipitações e temperaturas, e o enriquecimento por cargas orgânicas trazidas pelas inundações regulares forma alguns dos ecossistemas de maior produtividade primária do mundo (MITSCH, 1994). São importantes ecossistemas, imprescindíveis para manter o ecossistema local e global, sendo responsáveis pelo armazenamento de água, transformação de nutrientes, crescimento de matéria viva, entre outras funções ecológicas (NO-VITSKI et al., 1996). Possuem ainda contribuição importante no ciclo global do carbono, agindo como recicladores do dióxido de carbono ( $CO_2$ ) atmosférico e como uma das maiores fontes naturais de metano (MELACK et al., 2004).

Todos os processos ecológicos recorrentes, e também os impactos, como degradação ou contaminação, dependem da dinâmica das inundações cíclicas, tornando o seu monitoramento uma atividade crucial para o entendimento das suas funções ecológicas e, assim, para a sua proteção e manejo sustentável. Contudo, essas áreas representam um dos ecossistemas mais frágeis da Terra, pois são altamente susceptíveis às ações humanas (GOPAL; JUNK, 2000). As áreas úmidas e especificamente as planícies de inundação possuem uma alta demanda por um manejo sustentável (TOCKNER; STANFORD, 2002; MERZ et al., 2009; MUELLER et al., 2014).

Com uma área de cerca de 140 mil km<sup>2</sup>, o Pantanal é uma das maiores áreas úmidas do mundo, tendo sido declarado Patrimônio Nacional pela Constituição de 1988 e Patrimônio da Humanidade e Reserva da Biosfera desde 2000 pela UNESCO. Possui uma alternância anual cíclica entre fases aquáticas e terrestres pronunciadas, sendo classificado, portanto, como planície de inundação (NUNES DA CUNHA; JUNK, 2010).

### PROBLEMÁTICA

Devido a sua vasta extensão territorial e o difícil acesso, a realização de medidas *in loco* para qualquer tipo de pesquisa no Pantanal é uma tarefa com alta demanda por recursos humanos e materiais. Neste contexto, e em decorrência das dificuldades da modelagem hidrológica e hidráulica nas grandes planícies, caracterizadas por terreno muito plano e uma expressiva dinâmica geomorfológica (PAZ et al., 2011), técnicas de sensoriamento remoto são fundamentais para o monitoramento e a modelagem da inundação nas grandes planícies, como o Pantanal.

Assim, diversas abordagens de sensoriamento remoto foram avaliadas para esta aplicação, incluindo sensores passivos de micro-ondas (HAMILTON et al., 1996; HAMILTON et al., 2004), radar de abertura sintética (*Synthetic Aperture Radar*, SAR) (TOWNSEND; FOSTER, 2002; KUENZER et al., 2013; EVANS et al., 2010; ARNESEN et al., 2013), radar altimétrico (LEE et al., 2014), bem como sensoriamento remoto óptico de sistemas com diferentes resoluções espaciais e temporais (DARCH, 1979; GUMBRICHT et al., 2004; MACALISTER; MAHAXAY, 2009). De acordo com as limitações específicas dessas abordagens, foram propostas também a utilização de imagens multissensoriais (AUYNIRUNDRONKOOL et al., 2012; LEE et al., 2014).

Contudo, considerando a disponibilidade atual de sistemas sensores, o monitoramento contínuo de grandes planícies através de sensores passivos de microondas é limitado pela sua resolução espacial. Missões orbitais SAR são baseadas na banda C (por exemplo, Radarsat), com capacidade limitada na diferenciação de inundação em áreas cobertas por macrófitas, com arbustos inundáveis e abaixo de vegetação densa entre áreas inundadas (HESS et al., 2003; SILVA et al., 2010), possuem resolução temporal limitada, foram interrompidas (Envisat em 2012) e com novas gerações somente lançadas recentemente, porém com características de suas resoluções alteradas, dificultando comparações temporais (por exemplo, ALOS para ALOS-2, lançado em 2014).

#### JUSTIFICATIVA

Neste contexto, o sistema MODIS, devido a sua resolução espacial e temporal, consiste em um dos principais sistemas de sensoriamento remoto apto para o monitoramento da dinâmica de inundação em planícies de grandes extensões. Estudos realizados no Pantanal (PADOVANI, 2010) e em outras planícies de inundação, como a do rio Amazonas (SILVA et al., 2010) e do rio Mekong (SAKA- MOTO et al., 2007; ARIAS et al., 2012), nos Everglades (ORDOYNE; FRIEDL, 2008), na Austrália (CHEN et al., 2013; CHEN et al., 2014; HUANG et al., 2014) e Namíbia (LANDMANN et al., 2013) comprovaram a sua aplicabilidade para estudos de monitoramento da inundação.

Ressaltam-se, entretanto, diversos aspectos científicos não abordados ou não abordados sistematicamente sobre a aplicabilidade do sistema e seu desempenho, relevantes para um aperfeiçoamento do monitoramento operacional de grandes planícies de inundação.

Também, em decorrência da indisponibilidade de conjuntos de validação *in situ*, estudos anteriores não demonstraram resultados conclusivos sobre aspectos abordados neste estudo, como:

- i quais bandas ou índices espectrais do MODIS possuem o melhor desempenho no monitoramento da inundação, principalmente considerando a heterogeneidade da paisagem pantaneira, característica limitadora da acurácia de classificações da inundação (EVANS et al., 2014);
- ii bandas complementares disponíveis (outros produtos MODIS) podem resultar em melhorias no monitoramento;
- iii qual é a influência da qualidade dos pixels no desempenho das classificações e sua relação com o número de pixels não classificados;
- iv o desempenho de técnicas de preenchimento de falhas sobre o desempenho de classificações versus a porcentagem de não classificações de pixels de baixa qualidade radiométrica
- ${\bf v}$ qual é o melhor procedimento de extrapolação de observações pontuais in situ para os pixels de média resolução do sistema MODIS; e
- vi qual é a influência dos tipos de vegetação e da fase do ciclo hidrológico sobre o desempenho no monitoramento.

Portanto, o objetivo geral desse trabalho foi analisar o desempenho de produtos MODIS para modelagem da dinâmica de inundação do Pantanal Matogrossense.

Através da análise de séries temporais dos produtos Índice de Vegetação (MOD13Q1) e Uso e Ocupação da Terra (MCD12Q1) provenientes do sistema MODIS, de um recorte de cerca de 8400 km<sup>2</sup> na planície setentrional do Pantanal Mato-grossense, foram definidos como objetivos específicos os seguintes itens:

• Validar o desempenho de modelos classificatórios para monitoramento da inundação em função do uso de índices espectrais e produtos derivados;

- Examinar o desempenho do monitoramento da inundação utilizando diferentes abordagens para extrapolação de medições pontuais *in loco* para os pixels do produto MOD13Q1;
- Analisar a influência da qualidade dos pixels, da cobertura vegetal e da fase do ciclo hidrológico na planície sobre a acurácia da detectabilidade da inundação.

# 1 REVISÃO BIBLIOGRÁFICA

#### 1.1 O SENSOR ORBITAL MODIS

O sensor MODIS (do inglês, *MODerate resolution Imaging Spectroradiometer*) foi desenvolvido para fornecer monitoramento aprimorado para pesquisas sobre os continentes, oceanos e a atmosfera. Seu projeto para os componentes da imagem do terreno combina características dos sensores AVHRR, da série NOAA, e TM, da série Landsat-4 e 5 (JUSTICE et al., 1998).

O MODIS possui 36 bandas espectrais, sendo que as bandas 1 e 2 têm 250 metros de resolução espacial. As bandas de 3 a 7 têm 500 metros de resolução. As demais bandas possuem 1 km de resolução espacial. Esse instrumento é um dos sensores a bordo dos satélites Terra e Aqua, lançados em dezembro de 1999 e maio de 2002, respectivamente. Esses satélites fazer parte do Sistema de Observação da Terra (EOS, *Earth Observing System*), da NASA.

Assim como ocorre em qualquer sensor orbital que opera na faixa do visível e do infravermelho, a radiância solar que atinge o MODIS no topo da atmosfera sofre o efeito do espalhamento molecular, da absorção por gases e da influência dos aerossóis. Portanto, é realizado um processamento com base no código de transferência radiativo 6S para calcular a refletância na superfície, com considerações da Função de Distribuição de Refletância Bidirecional (do inglês, *Bidirectional Reflectance Distribution Function*, BRDF) da superfície e da atmosfera. A refletância na superfície é definida como a refletância que seria medida na superfície do terreno se não houvesse atmosfera (VERMOTE; VERMEULEN, 1999; VERMOTE et al., 2002).

#### 1.2 INDICES ESPECTRAIS

Os índices espectrais são combinações matemáticas de determinadas bandas com o objetivo de realçar uma característica que se quer estudar. A seguir são apresentados alguns índices utilizados em detecção de áreas inundadas, incluindo os índices de vegetação, NDVI e EVI.

### 1.2.1 **NDVI**

O Índice de Vegetação Diferença Normalizada (*Normalized Difference Vegetation Index*, NDVI) é a razão normalizada entre as bandas do infravermelho próximo (em inglês, *Near Infrared*, NIR) e do vermelho (*red*, em inglês)

$$NDVI = \frac{\rho_{NIR} - \rho_{red}}{\rho_{NIR} + \rho_{red}} \tag{1}$$

onde  $\rho_{NIR}$  e  $\rho_{red}$  são as refletâncias nas faixas espectrais do infravermelho próximo e do vermelho, respectivamente. Embora o NDVI seja bem sucedido como índice de medida da vegetação, ele apresenta algumas desvantagens, como por exemplo, não linearidade, saturação em condições de grande biomassa, é muito sensitivo a variação do fundo do dossel, e efeitos aditivos de ruídos causados pela radiância externa que ocorre na atmosfera (HUETE et al., 2002).

### 1.2.2 **EVI**

O Índice de Vegetação Realçado (em inglês, *Enhanced Vegetation Index*, EVI) foi desenvolvido para otimizar o sinal de vegetação com maior sensibilidade em regiões de alta biomassa e melhor monitoramento da vegetação através de um desacoplamento entre o sinal de fundo do dossel e uma redução nas influências da atmosfera. A equação toma a forma

$$EVI = G \frac{\rho_{NIR} - \rho_{red}}{\rho_{NIR} + C_1 \times \rho_{red} - C_2 \times \rho_{blue} + L}$$
(2)

onde  $\rho$  é a refletância da superfície para cada banda, corrigida ou parcialmente corrigida para as influências da atmosfera (espalhamento Rayleigh e absorção por ozônio), L é o coeficiente de ajuste do fundo do dossel,  $C_1$ ,  $C_2$  são os coeficientes relativos à resistência dos aerossóis, que usam a banda azul para corrigir a influência destes sobre a banda vermelha, e G é um fator de ganho (HUETE et al., 1997; HUETE et al., 2002).

#### 1.2.3 LSWI

Gao (1996) apresenta uma combinação matemática (3), semelhante à formulação do NDVI, usando neste caso uma banda na faixa espectral do infravermelho próximo (em inglês, *Near Infrared*, NIR), 0,86  $\mu$ m, e outra na faixa do infravermelho médio ou infravermelho de ondas curtas (em inglês, *Short Wave Infrared*, SWIR), 1,24  $\mu$ m, que correspondem, respectivamente, às bandas 2 e 5 do MODIS. A esta combinação ele dá o nome de Índice da Diferença Normalizada da Água (*Normalized Difference Water Index*, NDWI).

$$LSWI = \frac{\rho_{NIR} - \rho_{MIR}}{\rho_{NIR} + \rho_{MIR}} \tag{3}$$

Chandrasekar et al. (2010) chamam a mesma combinação de LSWI (do inglês, *Land Surface Water Index*), mas é usada a banda 7 (2,13  $\mu$ m) no lugar da 5, pois das bandas na região do infravermelho médio, a de número 7 é a que menos é afetada pelo ozônio e pelo espalhamento Rayleigh. Também é menos afetada pelo vapor de água e aerossóis do que a banda 6 (1,64  $\mu$ m).

O LSWI – como esta combinação será chamada daqui em diante –, de acordo com Gao (1996), é útil para medir o estado da água líquida na vegetação. Em geral, este índice é positivo para a vegetação verde devido à fraca absorção da radiação pela água em comprimentos de ondas próximos de 1,24  $\mu$ m. A absorção da radiação pela água líquida na vegetação verde na região de SWIR é significativamente mais forte do que na região de NIR. Os efeitos do vapor d'água na atmosfera sobre o LSWI são muito pequenos. Em relação ao NDVI, o LSWI é muito menos sensitivo aos efeitos de espalhamento pela atmosfera.

Xiao et al. (2002) e Xiao et al. (2005) usam este índice para mapeamento de campos de arrozais na China a partir da detecção da inundação destes campos, usando, respectivamente, imagens do sensor VEGETATION, do satélite SPOT, e imagens multitemporais MODIS. Xiao et al. (2006) estenderam o estudo para o Sudeste Asiático, também com imagens multitemporais MODIS. Já Sakamoto et al. (2007) usam o índice para detecção de mudanças temporais na extensão da inundação anual do Delta do Mekong no Camboja e Vietnã, usando também séries temporais do MODIS.

#### 1.2.4 NDWI

Em McFeeters (1996) é apresentado o índice Normalized Difference Water Index (NDWI), que, apesar de ter o mesmo nome do índice apresentado por Gao (1996), usa as bandas relativas às faixas espectrais do verde e do infravermelho próximo, respectivamente, 4 e 2 do sensor MODIS (4). Este índice, segundo seu autor, é útil no delineamento de corpos d'água. Adicionalmente, pode prover uma ferramenta útil no estudo de qualidade de água.

$$NDWI = \frac{\rho_{green} - \rho_{NIR}}{\rho_{green} + \rho_{NIR}} \tag{4}$$

#### 1.2.5 **MNDWI**

Xu (2006) apresenta o índice "NDWI modificado" (*modified Normalized Water Index*, mNDWI ou MNDWI), onde o mesmo substitui a banda NIR usada em McFeeters (1996) pela banda SWIR (5). Segundo o autor, o MNDWI pode melhorar consideravelmente o aprimoramento de características da água; Pode discriminar com rapidez e precisão entre água e não água; O MNDWI é mais apropriado para discriminar água em terrenos com áreas com construção do que o NDWI.

$$MNDWI = \frac{\rho_{green} - \rho_{MIR}}{\rho_{green} + \rho_{MIR}} \tag{5}$$

### 1.3 QUALIDADE DAS IMAGENS MODIS

A qualidade das imagens MODIS pode, como em qualquer outro sensor passivo, ser influenciada pelas condições atmosféricas (Figura 1) e pela geometria Sol-Alvo-Sensor (Figura 2) na hora da aquisição da imagem. A seguir são descritas essas influências, que são levadas em conta, na hora de compor qualquer produto MODIS.



Figura 1 – Efeitos da atmosfera que influenciam as medidas realizadas em sensores passivos.

Fonte: Adaptada de Lillesand & Kiefer (1994).



Figura 2 – Ilustração da aquisição de dados MODIS. A função de distribuição da refletância bidirecional (BRDF) muda com a geometria do ângulo de visada e ângulo de incidência solar. Note as sombras causadas pelas nuvens e dosséis. As dimensões do pixel (cross-track e along-track) mudam com o ângulo de escaneamento: 0° - 250 m x 250 m; 15° - 270 m x 260 m; 30° - 350 m x 285 m; 45° - 610 m x 380 m. Fonte: Adaptada de van Leeuwen et al. (1999).

#### 1.3.1 Influência da Atmosfera

O sinal na faixa espectral do vermelho que atinge o sensor orbital normalmente aumenta como resultado da contribuição da radiância externa (*path radiance*) espalhada pela atmosfera, enquanto o infravermelho próximo tende a diminuir como resultado da atenuação da atmosfera associado com o espalhamento e absorção pelo vapor d'água. No caso do NDVI, isso provoca uma diminuição no contraste entre estas faixas espectrais, reduzindo, portanto, o valor deste índice. A redução no valor do NDVI leva a uma subestimação da quantidade de vegetação na superfície do terreno (HUETE et al., 1999).

Outro fator causado pela atmosfera sobre o NDVI está relacionado com a presença de aerossóis. Atmosfera túrbida devido a presença de aerossóis provoca

uma atenuação no valor deste índice (GOWARD et al., 1991).

#### 1.3.2 Geometria da Aquisição da Imagem

Os índices de vegetação são afetados por variações nos fatores de refletância bidirecional, que variam muito em função de geometrias Sol-alvo-sensor. A Figura 2 ilustra a posição do Sol e do sensor MODIS para uma passagem do satélite Terra, no período da manhã. O MODIS imageia a superfície da Terra cobrindo uma faixa de 2330 km com o ângulo de visualização do sensor de  $\pm 55^{\circ}$  perpendicular à trajetória, com o ângulo de visada efetiva sobre o terreno sendo ligeiramente maior devido à curvatura da Terra. O tamanho do pixel aumenta proporcionalmente ao ângulo de visão. O ângulo solar zenital através das imagens MODIS pode variar até 20° de uma borda a outra da faixa de 2330 km. Também pode variar espacialmente com a latitude. A reflexão reversa que atinge o sensor tem ângulo zenital solar maior do que a reflexão direta.

Uma vez que o NDVI tende a aumentar proporcionalmente ao ângulo de visada e ao ângulo solar zenital, a variabilidade em ambos os ângulos é importante para intercomparações entre cobertura de vegetação em diferentes latitudes e entre diferentes épocas do ano (VAN LEEUWEN et al., 1999).

### 1.4 PRODUTOS DO SENSOR MODIS

Os muitos produtos de dados provenientes de observações MODIS descrevem os recursos da terra, dos oceanos e da atmosfera que podem ser usados para estudos de processos e tendências em escala local como global. Os produtos MODIS estão disponíveis a partir de várias fontes, dentre elas, o portal NASA Land Processes Distributed Active Archive Center, pertencente ao U.S. Geological Survey (USGS), que fornecem os produtos relacionados aos recursos terrestres. A seguir são listados os produtos utilizados neste trabalho.

#### 1.4.1 MOD13Q1 - Índice de Vegetação

O produto MOD13Q1, gerado em intervalos de 16 dias com resolução espacial de 250 metros, é produzido em unidades retangulares de aproximadamente 1200 por 1200 km, e mapeado em projeção de grade senoidal. Dois índices de vegetação são produzidos em escala global para regiões terrenas. O primeiro é o NDVI, que é referenciado como uma continuidade do NDVI existente produzido a partir do sensor *Advanced Very High Resolution Radiometer* (AVHRR) a bordo dos satélites da série NOAA. O segundo índice é o EVI, cujas melhorias em relação ao NDVI já foram mencionadas na subseção 1.2.2 (SOLANO et al., 2010). A Tabela 1 apresenta as características das bandas que compõem o produto MOD13Q1.

Banda	Unidade	Tipo	Faixa válida
NDVI	NDVI	int16	-2000, 10000
EVI	EVI	int16	-2000, 10000
VI Quality detailed QA	Bits	uint16	0,65534
$red \ reflectance \ (Band \ 1)$	Refletância	int16	0, 10000
$NIR \ reflectance \ (Band \ 2)$	Refletância	int16	0, 10000
blue reflectance (Band 3)	Refletância	int16	0, 10000
$MIR \ reflectance \ (Band \ 7)$	Refletância	int16	0,10000
view zenith angle	Grau	int16	-9000, 9000
sun zenith angle	Grau	int16	-9000, 9000
relative azimuth angle	Grau	int16	-3600, 3600
composite day of the year	Dia do ano	int16	1,366
$pixel\ reliability\ summary\ QA$	Ordinal	int8	0, 3

Tabela 1 – Características do produto MOD13Q1 – Índice de vegetação

A banda *pixel reliability summary QA* apresenta resumidamente a avaliação de qualidade de cada pixel que compõe a imagem. Essa avaliação, chamada de confiabilidade do pixel, é apresentada em escala ordinal, variando de -1 a 3 (Tabela 2).

Tabela 2 – Confiabilidade dos pixels do produto MOD13Q1

Valor	QA resumida	Descrição
-1	Sem dado	Não processado
0	Dado bom	Utilizável com confiança
1	Dado marginal	Utilizável, mas outras informações
		QA devem ser consultadas
2	Neve/gelo	Alvo coberto por neve ou gelo
3	Nublado	Alvo não visível, coberto por nuvens

A descrição detalhada da avaliação de qualidade dos pixels é fornecida pela banda VI Quality detailed QA (Tabela 3).

De acordo com o manual do usuário do Índice de Vegetação (SOLANO et al., 2010), os índices e bandas espectrais cujos respectivos pixels da banda *pixel reliability summary QA* (PR) têm valor igual a 0, são de boa qualidade, podendo ser usados com confiança. Entretanto, quando o valor do PR para um determinado pixel é igual a 1, é necessário verificar a qualidade deste pixel mediante a análise minuciosa da banda VI Quality detailed QA. Um dos parâmetros a ser verificado nesta banda é o VI Usefulness, dado pelos bits de 2 a 5. Esses bits informam a
utilidade do referente pixel numa escala que vai de 0000, maior qualidade, a 1100, menor qualidade, mas ainda útil. Os valores de *VI Usefulness* são formados a partir da combinação de diversos fatores que influenciam a qualidade do pixel, como condições climáticas e geometria (Sol-alvo-sensor) de aquisição do pixel (Tabela 4).

Bits	Parâmetro	Valor	Descrição
		00	VI produzido com boa qualidade
		01	VI produzido, mas verificar outras
0-1	Qualidade do Índice		informações QA
		10	Pixel produzido, mas provavelmente
			está nublado
		11	Pixel não produzido devido a outras
			razões
		0000	Qualidade mais alta
		0001	Qualidade menor
		0010	Qualidade decrescente
		0100	Qualidade decrescente
		1000	Qualidade decrescente
2-5	VI Usefulness	1001	Qualidade decrescente
		1010	Qualidade decrescente
		1100	Qualidade mais baixa
		1101	Qualidade tão baixa que não é útil
		1110	Dados L1B com falhas
		1111	Não útil por outras razões/não processado
		00	Climatológica
6-7	Quantidade de	01	Baixa
	aerossois	10	Intermediária
		11	Alta
8	Nuvem adjacente	0	Não
	detectada	1	Sim
9	Correção atmosférica	0	Não
		1	Sim
10	Pixel parcialmente	0	Não

Tabela 3 – Descrição do conjunto de dados científicos de avaliação de qualidade dos índices de vegetação (VI Quality) do produto MOD13Q1

Continua na próxima página.

	1		Continuação
Bits	Parâmetro	Valor	Descrição
	nublado	1	Sim
		000	Oceano/Águas rasas
		001	Terreno (nada além de terra)
		010	Litoral ou margens de lago
11-13	Máscara Terra/Água	011	Águas rasas (internas)
		100	Águas efêmeras
		101	Águas profundas (internas)
		110	Oceano continental ou moderado
		111	Oceano profundo
14	Possível neve/gelo	0	Não
		1	Sim
15	Possível sombra	0	Não
		1	Sim

Tabela 3 – Continuação

Tabela 4 – Método de escalonamento do parâmetro  $V\!I$   $U\!sefulness$  para o produto MOD13Q1.

Parâmetro	Condição	Pontuação
Quantidade de aerossois	Se foi utilizado o nível climatológico para a correção atmosférica (00) Se a quantidade de aerossois foi alta (11)	2 3
Correção Atmosférica Adjacente	Se nenhuma correção adjacente foi realizada (0)	1
Correção Atmosférica BRDF	Se nenhuma correção atmosférica foi realizada com a função BRDF de acoplamento atmosfera/solo (0)	2
Mistura de nuvens	Se possivelmente existiam nuvens misturadas (1)	3
Sombra	Se possivelmente existiam sombras $(1)$	2
Ângulo de visada zenital $(q_v)$	Se $q_v > 40^\circ$	1
Ângulo solar zenital $(q_s)$	Se $q_s > 60^{\circ}$	1

Os índices de vegetação do produto MOD13Q1 são corrigidos para os efeitos da atmosfera, uma vez que estes índices são calculados a partir de refletâncias



Figura 3 – Algoritmo de composição do produto MOD13Q1 coleção 5. Fonte: Adaptada de Solano et al. (2010).

da superfície corrigidas para o espalhamento Rayleigh, aerossóis troposféricos e estratosféricos, absorção de gases  $(O_3, O_2, CO_2$  e vapor d'água), além do efeito causado pelas nuvens do tipo cirrus (VERMOTE; VERMEULEN, 1999).

As imagens adquiridas em um período de 16 dias são processadas usando o algoritmo de composição do MODIS VI (Figura 3), que filtra os pixels baseado em sua qualidade, presença de nuvem e geometria de visada. Pixels contaminados por nuvem e ângulo de visada muito fora do nadir<sup>1</sup> são considerados de baixa qualidade. Por outro lado, pixels livres de nuvem, com visada no nadir, sem nenhuma contaminação atmosférica residual representam a melhor qualidade. Para a composição, são mantidos apenas os pixels de maior qualidade e sem presença de nuvem. Assim, o número de pixels aceitáveis no período de 16 dias é tipicamente menor do que 10 (frequentemente varia de 1 a 5).

O objetivo dessa metodologia de composição é extrair um único valor por pixel de todos os dados filtrados acumulados, o qual é representativo de cada pixel ao longo de determinado período de 16 dias. As técnicas de composição, das quais apenas uma é usada, são:

- 1. Principal: Ângulo de visada restrito Valor máximo para composição (*Constrained View angle Maximum Value Composite*, CV-MVC)
- 2. Alternativa: Valor máximo para composição (*Maximum Value Composite*, MVC)

<sup>&</sup>lt;sup>1</sup> O termo nadir é utilizado para definir a reta normal à superfície terrestre a partir do satélite.

O MVC é similar ao que é usado no produto NDVI do sensor AVHRR (série NOAA), em que o pixel com o maior valor de NDVI é selecionado para representar todo o período (16 dias). O CV-MVC é uma melhoria da técnica MVC, em que n observações (neste caso, n = 2) com os maiores valores de NDVI são comparados. A observação com o menor ângulo de visada (mais próxima ao nadir) é escolhida para representar o período.

A técnica empregada depende do número e da qualidade das observações. Se  $n \ge 2$  (ver Figura 3), emprega-se a técnica principal (CV-MVC). Caso n = 1, computa-se o NDVI para esta única observação. Se não existir nenhum pixel bom (n = 0), seleciona-se o maior valor de NDVI para o período (MVC).

#### 1.4.2 MCD12Q1 - Uso e Ocupação da Terra

O produto MCD12Q1 (FRIEDL et al., 2002; FRIEDL et al., 2010) é gerado usando um algoritmo de classificação supervisionado que é estimado usando um banco de dados de áreas de treinamento de cobertura de terreno de alta qualidade. O banco de dados das áreas de treinamento foi desenvolvido usando imagens de alta resolução com dados auxiliares (MUCHONEY et al., 1999). Os dados MODIS usados na classificação incluem um ano completo de observações compostas de 8 dias, incluindo a Refletância Ajustada para o Nadir e a função BRDF (*Nadir BRDF-Adjusted Reflectance*, NBAR) (SCHAAF et al., 2002) e a Temperatura da Superfície do Terreno (*Land Surface Temperature*, LST) (WAN et al., 2002).

Mapas globais de uso e ocupação da terra são produzidos em intervalo de tempo anual e com resolução espacial de 500 metros. Esse produto consiste de cinco tipos de classificação, além das bandas de avaliação da qualidade de cada classificação para cada pixel (Tabela 5).

A camada *Land Cover Type 1 (IGBP)* se refere à classificação do Programa Internacional Geosfera-Biosfera (IGBP), com 17 tipos de uso e ocupação da terra. A segunda camada, *Land Cover Type 2 (UMD)*, é a classificação da Universidade de Maryland (UMD), com 14 tipos de uso e ocupação. A terceira camada, *Land Cover Type 3 (LAI/fPAR)*, é uma classificação com 10 tipos de uso e ocupação, usada pelo algoritmo LAI/FPAR do MODIS. A quarta camada, *Land Cover Type* 4 (*BGC*), apresenta uma classificação com oito tipos de biomas. A quinta camada, *Land Cover Type 5 (PFT)*, apresenta uma classificação de tipo funcional das plantas com 12 classes (FRIEDL et al., 2010). As Tabelas 6, 7 e 8 apresentam as classificações de uso e ocupação da terra do produto MCD12Q1.

Camadas	Unidade	Tipo de dado	Preenchimento	Faixa válida
Land Cover Type 1 (IGBP)	Classe	uint8	255	0 - 254
Land Cover Type 2 (UMD)	Classe	uint8	255	0 - 254
Land Cover Type 3 (LAI/fPAR)	Classe	uint8	255	0 - 254
Land Cover Type 4 (BGC)	Classe	uint8	255	0-254
Land Cover Type 5 (PFT)	Classe	uint8	255	0-254
Land Cover Type 1 Assessment	Inteiro	uint8	255	0-254
Land Cover Type 2 Assessment	$\operatorname{sem} \operatorname{dado}$	uint8	255	0-254
Land Cover Type 3 Assessment	$\operatorname{sem} \operatorname{dado}$	uint8	255	0-254
Land Cover Type 4 Assessment	$\operatorname{sem} \operatorname{dado}$	uint8	255	0-254
Land Cover Type 5 Assessment	$\operatorname{sem} \operatorname{dado}$	uint8	255	0-254
Land Cover QC	Flags	uint8	255	0 - 254
Land Cover Type 1 Secondary	Classe	uint8	255	0 - 254
Land Cover Type 1 Secondary	$\operatorname{sem} \operatorname{dado}$	uint8	255	0 - 254
Land Cover Property 1	$\operatorname{sem} \operatorname{dado}$	uint8	255	0-254
Land Cover Property 2	$\operatorname{sem} \operatorname{dado}$	uint8	255	0-254
Land Cover Property 3	sem dado	uint8	255	0-254

Tabela 5 – Descrição das camadas que compõ<br/>em o produto MODIS de classificação de uso e ocupação da terr<br/>a $\rm MCD12Q1$ 

Tabela 6 – Descrição das bandas *Land Cover Type 1, IGBP* (Classificação do Programa Internacional Geosfera-Biosfera) e *Land Cover Type 2, UMD* (Classificação da Universidade de Maryland)

Classe	IGBP	UMD
0	Water	Water
1	Evergreen Needleleaf forest	Evergreen Needleleaf forest
2	Evergreen Broadleaf forest	Evergreen Broadleaf forest
3	Deciduous Needleleaf forest	Deciduous Needleleaf forest
4	Deciduous Broadleaf forest	Deciduous Broadleaf forest
5	Mixed forest	Mixed forest
6	Closed shrublands	Closed shrublands
7	Open shrublands	Open shrublands
8	Woody savannas	Woody savannas
9	Savannas	Savannas
10	Grasslands	Grasslands
11	Permanent wetlands	
12	Croplands	Croplands
13	ÁUrban and built-up	Urban and built-up
14	Cropland/Natural vegetation mosaic	
15	Snow and ice	
16	Barren or sparsely vegetated	Barren or sparsely vegetated
254	Unclassified	Unclassified
255	Fill Value	Fill Value

Fonte: Adaptada de (FRIEDL et al., 2010)

Tabela 7 – Descrição das bandas *Land Cover Type 3, LAI/fPAR* (Sistema de classificação usado pelo algoritmo LAI/FPAR do MODIS) e *Land Cover Type 4, BGC* (Classificação com oito tipos de biomas)

Classe	LAI/fPAR	BGC
0	Water	Water
1	Grasses/Cereal crops	Evergreen Needleleaf vegetation
2	Shrubs	Evergreen Broadleaf vegetation
3	Broadleaf crops	Deciduous Needleleaf vegetation
4	Savanna	Deciduous Broadleaf vegetation
5	Evergreen Broadleaf forest	Annual Broadleaf vegetation
6	Deciduous Broadleaf forest	Annual grass vegetation
7	Evergreen Needleleaf forest	Non-vegetated land
8	Deciduous Needleleaf forest	Urban
9	Non-vegetated	
10	Urban	
254	Unclassified	Unclassified
255	Fill Value	Fill Value

Fonte: Adaptada de (FRIEDL et al., 2010)

Tabela 8 – Descrição da banda *Land Cover Type 5, PFT* (Tipo funcional das plantas)

Classe	PFT
0	Water
1	Evergreen Needleleaf trees
2	Evergreen Broadleaf trees
3	Deciduous Needleleaf trees
4	Deciduous Broadleaf trees
5	Shrub
6	Grass
7	Cereal crops
8	Broad-leaf crops
9	Urban and built-up
10	Snow and ice
11	Barren or sparse vegetation
254	Unclassified
255	Fill Value

Fonte: Adaptada de (FRIEDL et al., 2010)

# 1.5 **REGRESSÃO LOGÍSTICA**

A regressão logística é um método de regressão que estabelece uma relação entre uma ou mais variáveis independentes (ou preditoras) e uma variável dependente (ou variável resposta), quando esta última é do tipo binário ou dicotômico. Diferentemente da regressão linear, que usa a função identidade para estabelecer essa relação, a regressão logística usa a função transformação logit, g(x),

$$g(x) = \ln\left[\frac{\pi(x)}{1 - \pi(x)}\right] \tag{6}$$

onde  $\pi(x)$  representa o valor esperado de um evento, dado o valor de x. A regressão logística estabelece uma relação linear entre a variável independente, x, e a transformação logit, g(x). Assim, temos

$$g(x) = \beta_0 + \beta_1 x \tag{7}$$

onde  $\beta_0$ , denominado intercepto, e  $\beta_1$ , o coeficiente da variável independente, são estimados pelo método da máxima verossimilhança.

O valor esperado,  $\pi(x)$ , ou a probabilidade de um evento ocorrer, portanto, é escrito da forma

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}},\tag{8}$$

quando há apenas uma variável independente (regressão logística simples), ou

$$\pi(\mathbf{x}) = \frac{e^{g(\mathbf{x})}}{1 + e^{g(\mathbf{x})}},\tag{9}$$

para uma regressão logística múltipla, onde  $g(\mathbf{x})$  é a transformação logit, dada por

$$g(\mathbf{x}) = \ln\left[\frac{\pi(\mathbf{x})}{1 - \pi(\mathbf{x})}\right] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p \tag{10}$$

para uma coleção p de variáveis independentes dado por  $\mathbf{x}' = (x_1, x_2, \dots, x_p)$ (HOSMER JR et al., 2013).

### 1.5.1 Teste de Significância dos Coeficientes

Para inferir sobre a significância de uma variável independente em uma regressão logística simples, usa-se o teste da razão de verossimilhança. Esse teste é obtido multiplicando por -2 a diferença entre o logaritmo da verossimilhança do modelo contendo apenas o intercepto e o logaritmo da verossimilhança do modelo contendo, além do intercepto, a variável independente. O valor-p associado ao teste nos fornece evidência, para um determinado nível de significância escolhido (por exemplo,  $\alpha = 0,05$ ), se a variável independente é significativa ou não na predição da variável resposta.

Para a regressão logística múltipla, o valor-p associado ao teste da razão de verossimilhança nos permite decidir se a hipótese nula deve ou não ser rejeitada. Rejeitar a hipótese nula significa concluir que pelos menos um ou mais coeficientes estimados são diferentes de zero.

Entretanto, antes de concluir que todos os coeficientes são diferentes de zero, é necessário avaliar o valor-p associado ao teste da estatística Wald para cada variável independente.

Maiores detalhes podem ser encontrados em Myers et al. (2010), Logan (2010) e Hosmer Jr et al. (2013).

#### 1.5.2 Métodos de Seleção das Variáveis Independentes

Hosmer Jr et al. (2013, p. 89) apresentam um método de seleção de variáveis, chamado de Seleção Proposital (ou Significativa) de Covariáveis, onde a decisão da inclusão ou exclusão das variáveis independentes (covariáveis) é feita pelo pesquisador mediante a análise dos resultados do teste da razão de verossimilhança e da estatística Wald. Também são apresentados dois métodos automatizados de seleção de variáveis: *Stepwise* e *Best subsets*.

Independentemente do método adotado, o objetivo é o mesmo: encontrar o modelo mais parcimonioso que ainda reflete fielmente os dados de saída.

Uma medida estatística alternativa ao teste da razão de verossimilhança, usada para comparar modelos com diferentes números de covariáveis, é o Critério de Informação de Akaike (AIC). Essa medida é definida como

$$AIC = -2 \times L + 2 \times (p+1), \tag{11}$$

onde L é o logaritmo da verossimilhança do modelo ajustado e p é o número de coeficientes estimados das variáveis independentes.

Embora o AIC seja uma medida estatística, não há testes estatísticos para comparar essa medida obtida de diferentes modelos. Em geral, valores menores são preferíveis a valores maiores. Logan (2010, p. 215) apresenta uma "regra de ouro": quando a diferença entre dois valores de AIC for maior que dois, o modelo com o menor AIC é o mais parcimonioso.

### 1.5.3 Medida da Qualidade de Ajuste do Modelo

As medidas de qualidade de ajuste do modelo fornecem informações que nos permitem inferir se as probabilidades produzidas pelo modelo refletem com exatidão os verdadeiros resultados.

#### 1.5.3.1 Tabela de Classificação

Essa tabela mostra uma classificação cruzada da variável de saída com a variável dicotômica cujos valores são derivados das probabilidades estimadas pelo modelo. Um ponto de corte precisa ser definido para obter a variável dicotômica a partir das probabilidades estimadas. Na forma binária, se a probabilidade estimada excede o ponto de corte, então a variável derivada passa a ser igual a 1; caso contrário, a variável será igual a 0 (HOSMER JR et al., 2013).

A Tabela 9 apresenta um modelo de Tabela de Classificação, em que Y representa a variável resposta binária observada e estimada. A e D representam as concordâncias entre os valores observados e classificados, enquanto  $B \in C$ representam as discordâncias.  $n_0$  é o número de casos observados para Y = 0, e  $n_1$ , para Y = 1. Semelhantemente,  $m_0$  é o número de casos classificados como Y = 0, e  $m_1$ , como Y = 1.  $n = n_0 + n_1 = m_0 + m_1$ .

Tabela 9 – Tabela de classificação

	Obse		
Classificado	Y = 0	Y = 1	Total
Y = 0	А	В	$m_0$
Y = 1	$\mathbf{C}$	D	$m_1$
Total	$n_0$	$n_1$	n

Da tabela de classificação, podemos retirar as seguintes medidas:

- Sensibilidade ou Taxa de Verdadeiro Positivo,  $TVP = D/n_1$
- Especificidade ou Taxa de Verdadeiro Negativo,  $TVN = A/n_0$
- Taxa de Falso Positivo,  $TFP = C/n_0 = 1 Especificidade$
- Taxa de Falso Negativo,  $TFN = B/n_1$
- Acurácia = (A + D)/n

#### 1.5.3.2 Coeficiente Kappa

O coeficiente de concordância Kappa tem sido amplamente usado em sensoriamento remoto como uma medida da qualidade de classificação (CONGALTON, 1991; CONGALTON; GREEN, 2008; FOODY, 2009; TANA et al., 2013; CHEN et al., 2013; HUANG et al., 2014; DRONOVA et al., 2015). Sua equação é dada por

$$\hat{\kappa} = \frac{p_o - p_c}{1 - p_c},\tag{12}$$

onde  $p_o$  é o número de concordância realmente presente (concordância "observada") e  $p_c$  é o número de concordância esperada apenas por acaso (concordância "esperada"). Da Tabela 9, temos

$$p_o = (A+D)/n \tag{13}$$

$$p_c = [(n_1/n) \cdot (m_1/n)] + [(n_0/n) \cdot (m_0/n)]$$
(14)

## 1.5.3.3 Área Sob a Curva ROC

O coeficiente de concordância Kappa e outras medidas de desempenho provenientes da tabela de classificação (sensibilidade, especificidade, acurácia) dependem de um único ponto de corte usado para classificar como verdadeiro o resultado de um teste. Uma descrição mais completa da acurácia da classificação pode ser obtida pelo cálculo da área sobre a curva ROC (*Receiver Operating Characteristic*).

Traçando a sensibilidade (taxa de verdadeiro positivo) em função de 1 especificidade (taxa de falso positivo) para cada ponto de corte, leva à curva ROC (Figura 4). Integrando-se a área sob a curva (*area under the curve*, AUC), obtém-se uma medida de avaliação de desempenho do modelo ou da capacidade preditiva deste. Um modelo com nenhuma capacidade preditiva é aquele com AUC igual a 0,5 (por exemplo, uma linha de 45°), enquanto um modelo perfeito corresponde àquele com AUC igual a 1,0 (BOYCE et al., 2002).

Uma das vantagens da curva ROC é que ela permite visualizar e organizar o desempenho do modelo sem levar em conta as distribuições de classe ou custos de erro (FAWCETT, 2006a).

#### 1.5.3.4 Coeficiente de Determinação

Coeficientes de determinação,  $R^2$  (variando de 0 a 1), intuitivamente conceituados como "variância explicada", têm sido largamente utilizados para quantificar a qualidade de ajuste de modelos. No caso da regressão linear generalizada (a regressão logística é um caso específico de regressão linear generalizada),  $R^2$  pode ser definido usando a máxima verossimilhança dos modelos completo e reduzido (modelo apenas com o intercepto). Uma das definições mais conhecidas é:

$$R^2 = 1 - \left(\frac{L_0}{L_\beta}\right)^{\frac{2}{n}},\tag{15}$$

onde  $L_{\beta}$  é a verossimilhança do modelo ajustado,  $L_0$  é a verossimilhança do modelo reduzido, e n é o tamanho total da amostra (NAKAGAWA; SCHIELZETH, 2013).



Figura 4 – Exemplo de gráfico ROC de três classificadores. Fonte: Adaptada de Fawcett (2006b)

Para a regressão logística, com 50% das observações igual a 1 e 50% igual a 0, o valor máximo do coeficiente de determinação, dado em (15), é de  $R^2 = 0,75$ . Esse máximo ocorre quando todas as observações são preditas com a máxima probabilidade. Assim, para que o coeficiente de determinação pudesse estar entre o intervalo de 0 e 1, Nagelkerke (1991) propôs definir esse coeficiente como:

$$R^{2} = \frac{\left[1 - \left(\frac{L_{0}}{L_{\beta}}\right)^{\frac{2}{n}}\right]}{\left[1 - (L_{0})^{\frac{2}{n}}\right]}.$$
(16)

#### 1.5.3.5 Validação Externa

Na avaliação da qualidade do ajuste de um modelo, não é suficiente avaliar o quão bem ele prediz sobre a amostra usada para ajustar o modelo. É necessário usar uma medida de quão bem o modelo irá predizer sobre uma amostra independente de dados similares, a fim de determinar seu grau de generalização (GIANCRISTOFARO; SALMASO, 2007).

Assim, em alguns estudos, pode ser possível excluir uma sub-amostra das observações, desenvolver o modelo com os dados remanescentes, e avaliar o modelo com os dados originalmente excluídos. Em outras situações, pode ser possível obter novas amostras de dados para avaliar a qualidade de ajuste do modelo. Esse tipo de avaliação é frequentemente chamado de validação do modelo (ou validação externa), e pode ser especialmente importante quando o modelo ajustado será usado para predizer o resultado de dados futuros. A razão para considerar esse tipo de avaliação de desempenho é que o modelo ajustado sempre se desempenha de uma maneira otimista sobre os dados usados no ajuste do modelo (HOSMER JR et al., 2013).

#### 1.5.3.6 Validação Cruzada: k-fold

Muitas vezes não é possível obter uma nova amostra externa independente da mesma população ou de uma similar. Nesse caso, pode ser possível fazer a validação interna do modelo (GIANCRISTOFARO; SALMASO, 2007).

Uma das formas de validar internamente o modelo é através da validação cruzada, como por exemplo, o método k-fold. Nesse método, a amostra original é aleatoriamente particionada em um número k igual de subamostras. Uma subamostra é retida para validação, enquanto o restante (k - 1) é usado para realizar o ajuste do modelo. Esse processo ocorre repetidamente um número k de vezes, com cada uma das k subamostras usada uma vez para validação (KOHAVI, 1995; HARRELL et al., 1996; BOYCE et al., 2002).

# 1.6 TIMESAT: SUAVIZAÇÃO E PREENCHIMENTO DE FALHAS

O software TIMESAT (JÖNSSON; EKLUNDH, 2002; JÖNSSON; EKLU-NDH, 2004; EKLUNDH; JÖNSSON, 2011) foi primariamente desenvolvido para processar séries temporais de índice de vegetação derivados de medidas espectrais de satélites. Esse software executa três métodos de processamento baseado em ajuste de mínimos quadrados: filtragem adaptativa Savitzky-Golay; ajuste com funções Gaussianas assimétricas; e ajuste com função logística dupla. Apenas a técnica de filtragem adaptativa Savitzky-Golay será descrita.

Os índices espectrais usados no processamento do TIMESAT são imagens binárias sem cabeçalho (matriz de dados em duas dimensões). Cada imagem apresenta os valores de índice de vegetação na matriz em um determinado momento. Extraindo os valores do índice de um pixel (j, k) para tempos consecutivos, uma série temporal  $y_1, y_2, ..., y_n$  é obtida (Figura 5).

Dados auxiliares de classificação de nuvens podem ser usados como indicador da incerteza dos valores dos índices espectrais. Essa classificação é traduzida em pesos que por sua vez determinam a incerteza dos valores dos índices.

Uma maneira de suavizar dados e suprimir distúrbios é com a utilização de um filtro, onde cada valor  $y_i, i = 1, ..., N$  é substituído por uma combinação



Figura 5 – Dados do índice de vegetação organizados em imagens (esquerda). Uma imagem i possui o valor de um determinado pixel para cada tempo  $t_i$ . Extraindo o valor do pixel (j, k) em intervalos de tempos consecutivos, uma série temporal é obtida para aquele pixel (direita). Fonte: (EKLUNDH; JÖNSSON, 2011).

linear dos valores próximos em uma janela

$$\sum_{j=-n}^{n} = c_j y_{i+j}.$$
 (17)

No caso mais simples, referido como média móvel, os pesos são  $c_j = 1/(2n+1)$ , e o valor de  $y_i$  é substituído pela média dos valores na janela. No caso do filtro adaptativo Savitzky-Golay, o valor é obtido através de ajuste de mínimos quadrados polinomial. Para cada valor  $y_i$ , i = 1, ..., N é ajustado um polinômio quadrado  $f(t) = c_1 + c_2t + c_3t^2$  para todos os pontos 2n + 1 na janela móvel e substituído o valor  $y_i$  pelo valor do polinômio na posição  $t_i$ .

#### 1.7 MAPEAMENTO DA INUNDAÇÃO COM IMAGENS DE SATÉLITE

Existem diversos métodos utilizados para detectar inundação com o uso de imagens de sensoriamento remoto. Os métodos usando sensores ópticos não mudaram muito desde os anos 1970, mesmo tendo ocorrido melhorias nas resoluções espacial e temporal, disponibilidade e qualidade da imagem (SMITH, 1997; BAKER et al., 2007; BWANGOY et al., 2010).

Um dos métodos mais simples é o fatiamento de histograma, onde um único valor de corte (limiar) é escolhido, de tal modo que todos os valores acima desse limiar são classificados como água (inundado), e abaixo, superfície seca (não inundado) (FRAZIER et al., 2000; FRAZIER et al., 2003; PADOVANI, 2010).

A classificação multibanda também é uma técnica bem conhecida. Esse tipo de classificação integra informações de todas as bandas disponíveis, e classifica a imagem em classes dominantes, incluindo água (MACALISTER; MAHAXAY, 2009).

Os índices espectrais, como o NDVI (ZHU et al., 2011), EVI (SAKAMOTO et al., 2007; PADOVANI, 2010), NDWI (MCFEETERS, 1996; GAO, 1996) e mNDWI (XU, 2006; CHEN et al., 2014; HUANG et al., 2014) são largamente empregados, com a vantagem de combinar diferentes informações da água de bandas multiespectrais, normalizar os dados, bem como reduzir os efeitos atmosféricos.

O índice OWL (do inglês *Open Water Likelihood*) (GUERSCHMAN et al., 2011) tem sido usado para mapeamento da inundação em planícies inundáveis e áreas úmidas na Austrália (CHEN et al., 2013; CHEN et al., 2014; HUANG et al., 2014; TICEHURST et al., 2014).

Embora os sensores que operam nas faixas do visível e do infravermelho próximo tenham sua limitação devido às condições climáticas, seu uso se justifica quando comparado com os sensores ativos, uma vez que esses últimos possuem custos mais elevados (SAKAMOTO et al., 2007). No caso do sensor MODIS, a periodicidade com que os dados são adquiridos o torna promissor nos estudos de dinâmica de inundação em planícies, como o caso do Pantanal.

Alguns estudos avaliaram a aplicabilidade do sensor MODIS para detecção de mudanças espaciais e temporais da extensão da inundação (SAKAMOTO et al., 2007; ORDOYNE; FRIEDL, 2008; KHAN et al., 2011; CHEN et al., 2012; FENG et al., 2012).

Sakamoto et al. (2007), com o objetivo de detectar as variações na extensão da região de inundação no Delta do Mekong, entre o Camboja e Vietnã, criaram uma metodologia que incluía um filtro baseado em *Wavelet* para interpolação de dados faltantes e redução de ruído dentro da série temporal. Os autores usaram os dados de refletância da superfície fornecidos pelo produto MOD09A1, que é gerado em intervalos de 8 dias com resolução espacial de 500 metros. Foram usados dados de março de 2000 a fevereiro de 2005.

Usando os índices EVI e LSWI, e a diferença entre os dois índices (DVEL), filtrados por *Wavelet*, Sakamoto et al. (2007) criaram uma série temporal da inundação do Delta do Mekong. Obtiveram bons resultados quando comparados com medidas de inundação derivados do RADARSAT ( $R^2$  entre 0,89 e 0,92).

Entretanto, regiões cobertas por florestas e pântanos foram subestimadas por seu modelo, provavelmente por causa da extensa cobertura da vegetação nesta área.

Ordoyne & Friedl (2008), com o objetivo de explorar a utilidade de dados MODIS para mapear a dinâmica espacial e temporal de inundação no Everglades, no estado da Flórida, nos Estados Unidos, usaram os produtos MCD43B4 (*Nadir BRDF-Adjusted Reflectance*) e MYD11A2 (*8-day Land Surface Temperature*), do ano de 2004.

Usando dados de elevação de superfície, obtido de modelo numérico de terreno (MNT), os níveis de água em relação à elevação da superfície do terreno local foram obtidos a partir de dados do nível da água, medidos em estações de monitoramento hidrológico do Everglades.

Dos diversos modelos estatísticos testados, a regressão logística obteve o melhor desempenho para resposta binária (inundado/não inundado). Realizando uma validação cruzada *leave-one-out*, obtiveram uma acurácia geral de 83,7%. A maioria dos casos de classificação errada ocorreu onde o nível relativo da água estava dentro de  $\pm 15$  cm em relação à superfície do terreno.

Um dos problemas encontrados, com relação aos dados MODIS, foi a falta de dados causada pela presença de nuvens. Apenas 3,6% dos pixels na região do Everglades continham dados válidos para as 23 imagens usadas. Também obtiveram redução na acurácia na validação externa, demonstrando uma diminuição no grau de generalização do modelo.

Padovani (2010), em sua tese de doutorado, estudou a dinâmica de inundação do Pantanal com o uso de série temporal do índice EVI derivado do produto MOD13Q1, dividindo o Pantanal em sub-regiões e suas respectivas bacias de drenagem. Foram utilizadas imagens de fevereiro de 2000 a dezembro de 2009, num total de 227 imagens.

Usando o Modelo de Mistura Espectral (SHIMABUKURO; SMITH, 1991), Padovani (2010) criou imagens de fração de solo, de água e de vegetação. Em seguida, realizou a suavização e preenchimento de falhas na série temporal dessas imagens de fração e nas imagens formadas pelos índices EVI e NDVI, utilizando o software TIMESAT com filtro adaptativo Savitzky-Golay.

Usando a técnica de fatiamento manual de histograma, as áreas inundadas foram extraídas das imagens de fração de água. O reconhecimento visual dos alvos a serem mapeados foi realizado tendo como referência uma imagem ETM do Landsat-7 de 2000. Um mapa de frequência de inundação, elaborado a partir da soma dos 227 mapas de inundação, foi utilizado para relacionar com um mapa de vegetação e uso e outro de solos, para relacionar a frequência de inundação como o tipo de vegetação e tipo de solo, respectivamente.

Mesmo tendo usado o modelo de mistura espectral, ainda assim verificou-se subestimação da inundação em regiões mais fragmentadas, limitada, portanto, pela resolução espacial moderada que o MODIS possui. Por outro lado, áreas maiores e mais homogêneas tiveram um melhor mapeamento da frequência de inundação.

Huang et al. (2014) estudaram a dinâmica espacial e temporal da inundação da bacia Murray-Darling na Autrália. Para isso, integraram observações hidrológicas e dados do MODIS com mapas de inundação gerado a partir de imagens Landsat TM/ETM+.

Dados hidrológicos consistiram de séries temporais diárias de medidas de vazão de diversos rios da bacia, cujos picos de vazão foram usados para determinar a data das imagens MODIS. Desse último, foi utilizado o produto MOD09A1, de onde foi derivado o índice OWL (GUERSCHMAN et al., 2011).

Os mapas de inundação foram obtidos partir do índice mNDWI (XU, 2006), derivados das imagens Landsat TM/ETM+. Esses foram usados em substituição às medidas de campo ("ground truth") para validar as estimativas de inundação obtida pelo índice OWL.

O índice OWL (modelo obtido por regressão logística múltipla) foi calibrado e otimizado por Guerschman et al. (2011) com base em seu desempenho no continente Australiano. As variáveis preditivas desse índice são as bandas 6 (SWIR) e 7 (SWIR) do produto MOD09A1, os índices NDVI e NDWI (GAO, 1996), derivados, também, do produto MOD09A1, e o índice MrVBF (GALLANT; DOWLING, 2003) derivado do DEM da Missão Topográfica Radar Shuttle (SRTM).

Os resultados da avaliação do índice OWL usando imagens MODIS na detecção da inundação, dados pelo coeficiente de concordância Kappa, demonstraram acurácia aceitável à escala da bacia.

# 2 MATERIAL E MÉTODOS

## 2.1 MATERIAL

#### 2.1.1 Área de estudo

O Pantanal é uma vasta planície inundável na bacia do Alto Rio Paraguai, com a maior parte localizada no Brasil, com áreas menores na Bolívia e no Paraguai. Sua área no Brasil, segundo estudos de Silva & Abdon (1998) usando imagens de satélite Landsat-5, é de 138.183 km<sup>2</sup>. A maior parte se encontra no estado de Mato Grosso do Sul, com uma área de 89.318 km<sup>2</sup> (64,6%). Em Mato Grosso, esse bioma ocupa uma área um pouco menor: 48.865 km<sup>2</sup> (35,4%).

O clima, segundo a classificação de Köppen, é tropical com inverno seco (Aw), durante o qual pode ocorrer um déficit hídrico de até oito meses. As temperaturas médias anuais variam de 24°C a 26°C, e as precipitações anuais, de 1100 mm a 1400 mm (ALVARES et al., 2013), com a maior parte ocorrendo de novembro a março.

O Pantanal tem um ciclo sazonal monomodal de inundação, determinado principalmente pelas chuvas nas suas áreas de contribuição no planalto e nas cabeceiras do rio Paraguai e seus afluentes. Devido à baixa declividade da planície e, assim, à lenta passagem das águas de alagamento, os máximos de inundação correspondem aos máximos da precipitação somente nas partes mais altas da planície perto da sua transição para as depressões interplanálticas e planaltos (HAMILTON et al., 1996). As inundações ocorrem pelo transbordamento dos rios, embora também ocorra parcialmente pelas chuvas locais (HAMILTON et al., 1998). Em função da duração e amplitude da inundação e características pedológicas, o Pantanal é composto por um mosaico de unidades fitofisionômicas, entre elas formações de florestas sempre-verdes, semidecíduas e decíduas, savanas de diferentes densidades de seu dossel, formações herbáceas e palustres (ZEILHOFER; SCHESSL, 2000). Principalmente durante o período de alagamento, os campos inundáveis podem ser parcialmente cobertos por macrófitas, e nos corpos de água ocorrem ilhas flutuantes formadas por ciperáceas e plantas aquáticas, com as raízes entrelaçadas (batume).

### 2.1.2 Medidas de Campo: Inundação e topografia do terreno

Para o estudo da dinâmica de inundação no Pantanal Norte, foram usadas medidas previamente realizadas *in situ* em duas regiões (Figura 6) (FANTIN-CRUZ et al., 2010; GIRARD et al., 2010; FANTIN-CRUZ et al., 2011).



Figura 6 – Área de estudo apresentando o Sítio de Amostragem de Longa Duração (SALD) Pirizal e o Transecto dentro da Reserva Particular do Patrimônio Natural do SESC Pantanal.

Uma das regiões, localizada no município de Nossa Senhora do Livramento, é o Sítio de Amostragem de Longa Duração (SALD) no Pirizal (Figura 7), contendo 40 réguas limnimétricas distribuídas em forma de grade, cobrindo uma área de 25 km<sup>2</sup> (SIGNOR et al., 2010). A outra região, localizada dentro da Reserva Particular do Patrimônio Natural (RPPN) do SESC Pantanal, no município de Barão de Melgaço, é constituída de 11 réguas limnimétricas, formando um transecto de 12 km de extensão (Figura 8).



Figura 7 – Malha com 40 réguas limnimétricas instaladas no SALD Pirizal em uma imagem SPOT, bandas 3/2/1 de 2007



Figura 8 – Transecto formado por 11 réguas na Reserva Particular do Patrimônio Natural SESC Pantanal em uma imagem SPOT, bandas 3/2/1 de 2007

Fantin-Cruz et al. (2010) identificaram quatro unidades fitofisionômicas no SALD Pirizal: (1) Campo inundável (Savana gramíneo-lenhosa sazonalmente inundável); (2) Cordilheira (Savana arbórea densa com *Curatella americana* L. e *Dpteryx alata* Vogel.); (3) Landizal (Floresta Sempre-Verde sazonalmente inundável com *Calophyllum brasiliense* Camb.); e (4) Cambarazal (Floresta monodominante sazonalmente inundável de *Vochysia divergens* Pohl.).

Ao longo do Transecto do SESC Pantanal, foram identificadas, também, quatro unidades fitofisionômicas: (1) Campo Cerrado, caracterizado pela ocorrência de *Byrsonima orbigyana* A. Juss., *Alchornea discolor* Poepp., *Bactris glaucescens* Drube, *Licania parvifolia* Huber, e *Curatella americana* L.; (2) Cambarazal (ARIEIRA; NUNES DA CUNHA, 2006); (3) Espinheiral, consistindo de arbustos espinhosos e trepadeiras; e (4) uma estreita mata ciliar ao longo do dique do rio Cuiabá e de outros canais secundários. Seu perfil topográfico irregular inclui três pontos baixos correspondendo a canais secundários conhecidos localmente como corixos (GIRARD et al., 2010).

Usando as medidas de inundação do SALD Pirizal apresentadas em Fantin-Cruz et al. (2010), e as de inundação ao longo do Transecto do SESC Pantanal apresentadas em Girard et al. (2010) e em Fantin-Cruz et al. (2011), criou-se seis conjuntos de medidas; três conjuntos para cada sítio. A Tabela 10 sumariza as medidas utilizadas neste estudo.

Também foi utilizado o levantamento topográfico do SALD Pirizal apresentado em Fantin-Cruz (2008), para interpolação de um modelo numérico de terreno (MNT). Sua topografia apresenta uma superfície inclinada no sentido NE/SO com a menor cota de 112,52 m e a maior de 116,56 m (FANTIN-CRUZ et al., 2010).

Local	Período	n
	24/01/2005 a 30/04/2005	19
Transecto	10/01/2006 a 01/05/2006	41
	$10/01/2007 \ge 06/05/2007$	10
	$29/12/2006 \ge 01/06/2007$	20
SALD	06/12/2007 = 27/06/2008	25
	$13/12/2008 \ge 04/05/2009$	17

Tabela 10 – Período das medidas de inundação obtidas através das réguas do SALD Pirizal e das do Transecto do SESC Pantanal, e a quantidade (n) de medidas em cada período.

A Figura 9 apresenta uma estimativa da precipitação mensal feita por dados da missão TRMM (produto 3B42), calculada a partir da média ponderada da grade  $(0,25^{\circ})$  dentro da bacia de contribuição da régua fluviométrica de Porto



Cercado (Cód. ANA 66341000) e a medida da vazão do rio Cuiabá em Barão de Melgaço (a montante), durante o período de estudo.

Figura 9 – Estimativa da precipitação mensal feita por dados da missão TRMM (produto 3B42), calculada a partir da média ponderada da grade (0,25°) dentro da bacia de contribuição da régua fluviométrica de Porto Cercado (gráfico de barras). Medida da vazão do rio Cuiabá em Barão de Melgaço (a montante da área de estudo, gráfico de linha).

A Figura 10 apresenta um climograma obtido a partir de medidas de temperatura e precipitação coletados entre os anos de 2011 e 2013 na Fazenda Miranda, cerca de 75 km de distância da área de estudo, onde o Programa de Pós-graduação em Física Ambiental possui uma torre de coleta de dados meteorológicos.

#### 2.1.3 Produtos MODIS

Foram usados os produtos Índice de Vegetação (MOD13Q1, coleção 5) e Tipo de Cobertura do Terreno (MCD12Q1). Esses produtos foram baixados gratuitamente do portal da NASA Land Processes Distributed Active Archive Center (LP DAAC, 2012) usando a ferramenta USGS EarthExplorer. As imagens MODIS são distribuídas no formato HDF (*Hierarchical Data Format*) e projeção senoidal, formando blocos de 1200 km por 1200 km, divididos em linhas (v) e



Figura 10 – Climograma da Fazenda Miranda para os anos de 2011 a 2013. O eixo das abscissas indicam os meses, de julho (J) a junho (J). O eixo das ordenadas da esquerda corresponde à temperatura do ar, com as médias mensais máxima e mínima em destaque (cor preta). O eixo das ordenadas da direita corresponde à precipitação pluviométrica.

colunas (h). Assim, um único bloco (h12v10) foi suficiente para cobrir toda a área de estudo.

Foram baixadas 299 imagens do produto índice de vegetação MOD13Q1, dos anos de 2000 a 2012, e 5 imagens do produto uso e ocupação da terra MCD12Q1, dos anos de 2005 a 2009.

Das bandas do produto MOD13Q1, foram utilizadas: os índices de vegetação NDVI e EVI; a refletância na faixa espectral do infravermelho próximo (NIR); a refletância na faixa espectral do infravermelho de ondas curtas (MIR); a banda *composite day of the year*, que fornece a data (dia do ano: 1 a 366) de aquisição de cada pixel que compõe a imagem; a banda *pixel reliability summary QA* (Tabela 2), que apresenta de forma resumida a qualidade dos pixels; e banda *VI Quality detailed QA* (Tabela 3), que fornece informações detalhadas da qualidade dos pixels.

# 2.2 MÉTODOS

A metodologia descrita nas próximas seções é apresentada em forma de fluxogramas nos anexos:

- **Apêndice A** Processamento das medidas de campo e pré-processamento dos dados MODIS;
- **Apêndice B** Seleção das variáveis e ajuste do modelo com o uso de imagens MODIS de qualidade PR\_0;
- **Apêndice C** Preenchimento de falhas e ajuste do modelo com dados preenchidos;
- **Apêndice D** Criação dos conjuntos de níveis de qualidade QA\_0000 a QA\_1100.

## 2.2.1 Processamento das medidas de inundação

Embora o produto MOD13Q1 seja produzido a cada 16 dias, os pixels que compõem a imagem possuem datas específicas de aquisição, fornecidas pela banda *composite day of the year*. Assim, para associar cada pixel da imagem, localizado na coordenada de cada régua, com as medidas de inundação, obtidas em campo em intervalos de tempo não regulares, as medidas de cada régua foram linearmente interpoladas, obtendo-se medidas diárias.

Para cada régua, a última medida de inundação diferente de zero foi extrapolada com o objetivo de se estimar o fim da inundação, ao invés de se interpolar com a subsequente medida igual a zero. Essa extrapolação foi feita utilizando a maior taxa de esvaziamento da inundação medida por cada régua, para cada ano de medição, conforme a equação 18.

$$T_L = \min\left(\frac{L_{n+1} - L_n}{t_{n+1} - t_n}\right) \tag{18}$$

onde  $T_L$  corresponde à variação da inundação de uma determinada régua ocorrida em um intervalo de tempo,  $t_{n+1} - t_n$ ;  $L_n$  é o nível de inundação de uma medida nregistrada no dia  $t_n$ ;  $L_{n+1}$  é o nível de inundação registrada no dia  $t_{n+1}$ . A função min seleciona o menor valor (a taxa negativa de maior magnitude) para cada régua em cada ano de coleta de dados.

Dados faltantes na planilha de medidas de campo foram examinados individualmente para verificar a possibilidade de preenchimento. Em alguns casos, o valor da régua mais próxima foi ajustado altimetricamente para preencher pequenas falhas. Interpolação e extrapolação, como explicadas acima, em alguns casos também foram usadas. Uma vez interpoladas e/ou extrapoladas as medidas de inundação, obtendose medidas diárias, estas foram convertidas para valores binários. Assim, qualquer valor acima de zero passou a ser representado por 1 (inundado).

As medidas diárias de inundação ("1") ou seca ("0") de cada régua foram armazenadas em uma planilha do tipo csv (*comma-separated values* – valores separados por vírgula, em Português) para posterior associação com as datas de aquisição de cada pixel dos produtos MODIS. Foram usadas apenas as medidas de campo com a mesma data de aquisição dos pixels.

#### 2.2.2 Pré-processamento das imagens MODIS

Utilizando o software *MODIS Reprojection Tool* (MRT) (NASA/USGS, 2011), os produtos Índice de Vegetação (MOD13Q1) e Tipo de Cobertura do Terreno (MCD12Q1) foram reprojetados para a projeção Universal Transversa de Mercator (UTM), Datum WGS1984, zona 21S; recortadas para a área de estudo (latitude entre  $-16, 1^{\circ}$  e  $-16, 8^{\circ}$  e longitude entre  $-55, 7^{\circ}$  e  $-56, 7^{\circ}$ ); separadas as bandas de interesse e salvas em formato Geotiff. As bandas separadas foram: NDVI, EVI, NIR, MIR, *VI Quality detailed QA, pixel reliability summary QA* e *composite day of the year*, do produto MOD13Q1; e a banda *Land Cover Type 2 (UMD)*, do produto MCD12Q1. Dado a grande quantidade de imagens do produto MOD13Q1 (299 imagens; 13 anos de imagens), esse o procedimento foi feito em lote usando o script MRTBatch.jar.

#### 2.2.3 Desenvolvimento do Modelo de Inundação

Com o objetivo de desenvolver o modelo empírico estatístico para classificação da inundação, os pixels nublados e de qualidade duvidosa (*pixel reliability* summary QA > 0) das bandas NDVI, EVI, NIR e MIR foram filtrados. Também foram filtrados os pixels cujos valores estavam fora da faixa válida de cada banda (ver Tabela 1). O índice LSWI foi calculado a partir das bandas NIR e MIR, já filtradas, conforme a equação (3).

Os pixels localizados nos pontos de monitoração da inundação (réguas) foram extraídos e, com o uso da banda *composite day of the year*, foram associados com as medidas diárias de inundação, já convertidas para valores binários (0 e 1). Assim, só foram usados os pixels em que havia medida de campo disponível. Da mesma forma, só foram utilizadas medidas de campo coincidente com a banda *composite day of the year*.

Para o ajuste do modelo, foram usados os dois primeiros períodos de medidas de cada sítio. O período restante de cada sítio foi reservado para validação externa do modelo (ver Tabela 10).

Com o uso do software R (R Core Team, 2014), foi realizado o ajuste do modelo linear generalizado, usando a função de ligação *logit* e família binomial. O ajuste do modelo foi feito com todas as variáveis independentes, para verificar, inicialmente, a significância estatística de cada variável no modelo.

O diagnóstico de multicolinearidade das variáveis independentes foi feito através do cálculo do fator de inflação da variância para cada variável independente. Esse procedimento foi realizado com o uso da função *vif* da biblioteca *car* (FOX; WEISBERG, 2011).

A seleção das variáveis para compor o modelo foi realizada através do método de seleção automático *Stepwise*, usando a função *stepAIC* da biblioteca MASS (VENABLES; RIPLEY, 2002).

Como o objetivo do modelo é a classificação da inundação, a qualidade do ajuste foi verificada com o uso de informações derivadas da tabela de classificação (também chamada de matriz de confusão), como taxa de falso positivo (TFP) taxa de falso negativo (TFN), acurácia e o coeficiente de concordância Kappa. A qualidade de ajuste do modelo também foi medida com o uso do coeficiente de determinação de Nagelkerker ( $R^2$ ).

Um ponto de corte foi estimado para obter a variável dicotômica de saída a partir das probabilidades estimadas pelo modelo. Esse ponto de corte foi utilizado pela tabela de classificação e pelas medidas derivadas desta.

A avaliação da capacidade de discriminação do modelo foi feita através da inspeção visual de quatro gráficos, conforme Royston & Altman (2010), além da verificação do valor calculado para a área sob a curva (AUC) ROC.

Para medir o grau de generalização do modelo, foi realizada uma validação externa com os dados que foram reservados para este fim (último período de medida de cada sítio).

Muitas das vezes, pesquisadores não conseguem uma nova amostra de dados da mesma população ou de uma similar para realizar a validação externa. Portanto, nesse estudo também foi realizada uma validação cruzada do tipo k-fold com k = 10.

## 2.2.4 Preenchimento de Falhas

Embora o uso de pixels com qualidades inferiores aumente a quantidade total de dados a serem utilizados, ainda assim não é possível obter uma imagem com 100% de cobertura da área de estudo para cada ano de estudo; algumas cenas ainda são contaminadas por nuvens (*Pixel Reliability* igual a três). Portanto, para preenchimento dos pixels removidos no processo de filtragem, foi utilizado o software TIMESAT, utilizando a filtragem adaptativa Savitzky-Golay que, embora não tenha tido melhor desempenho do que as outras duas técnicas disponível no software em estudos comparativos de eliminação de ruídos em séries temporais de NDVI (HIRD; MCDERMID, 2009), em outros estudos mostrou-se mais eficiente na redução das contaminações causadas por nuvens e variabilidade atmosférica em séries temporais do índice EVI (CHEN et al., 2004; MINGWEI et al., 2008; WANG et al., 2012; CHAKRABORTY; DAS, 2012).

A aplicação do filtro Savitzky-Golay, para suavização e preenchimento de dados faltantes (dados nublados), foi feita nas imagens EVI, LSWI e NDVI do tipo inteiro de 16 bits, sem cabeçalho. Estas imagens não foram previamente filtradas de acordo com os níveis de qualidade dos pixels, conforme explicado na subseção 2.2.3, contendo, portanto, todos os possíveis níveis de qualidade, incluindo pixels nublados. Entretanto, a banda *Pixel Reliability* foi utilizada como dado auxiliar de entrada. Após alguns ensaios com os diversos níveis de qualidade, resolveu-se atribuir os seguintes pesos: PR\_0: peso = 1,0; PR\_1: peso = 0,8; e PR\_3: peso = 0. O peso igual a zero foi usado para que os dados nublados fossem removidos da série temporal original e, assim, preenchidos pelo software. Também foram usados os seguintes parâmetros de configuração do software: *Window size* = 4; *No. of envelope iterations* = 3; e Adaptation strength = 2.

A Figura 11 ilustra um exemplo de uma série temporal do índice LSWI de um pixel do ano 2004 a 2010. A linha azul é a série original com falhas (os dados nublados foram removidos). A linha marrom apresenta a série reconstruída usando a filtragem adaptativa Savitzky-Golay.

# 2.2.5 Análise da Influência da Qualidade dos Pixels Sobre o Desempenho do Modelo

A fim de se investigar se a qualidade dos pixels do produto MOD13Q1 influencia a capacidade de classificação da inundação, foram criados quinze conjuntos de dados, de acordo com os níveis de qualidade desses pixels, a partir de informações fornecidas pela banda *pixel reliability summary QA* (PR) e pelo parâmetro *VI Usefulness* (bits 2 ao 5) da banda *VI Quality detailed QA*. Na criação dos conjuntos, também foram avaliados os valores dos pixels de acordo a faixa válida dos índices e bandas espectrais usados (ver Tabela 1).

Assim, para formar o conjunto PR\_0, apenas a banda PR foi consultada: pixels com PR diferente de zero são filtrados (excluídos). Por outro lado, o conjunto de dados PR\_3 é formado por todos os pixels, inclusive aqueles marcados como



Figura 11 – Exemplo de série temporal (2004 a 2010) do índice LSWI de um pixel da área de estudo. A série original com falhas (pixels nublados removidos) é representada pela linha azul. A série reconstruída é dada pela linha marrom.

nublados (PR = 3). Para formar os conjuntos QA\_0000 ao QA\_1100, além da banda PR, o parâmetro VI Usefulness da banda VI Quality Assessment também foi avaliado. Assim, para formar o conjunto QA\_0000, o valor do pixel da banda PR deve ter valor igual a 0 ou 1, enquanto o pixel da banda QA<sup>1</sup> deve ter valor igual a 0000. Para criar o conjunto QA\_0001, o valor do pixel da banda PR deve ter valor igual a 0 ou 1, e a banda QA precisa ter valor  $\leq$  0001. Para criar o conjunto QA\_0010, o valor do pixel da banda PR deve ter valor igual a 0 ou 1, e a banda QA precisa ter valor  $\leq$  0010. Dessa forma, sucessivamente, até formar o conjunto QA\_1100. A Tabela 11 sumariza esses conjuntos.

A influência da qualidade dos pixels e dos dados reconstruídos pelo TI-MESAT sobre o desempenho do modelo foi avaliada de acordo com os valores do coeficiente de determinação de Nagelkerker  $(R^2)$  e coeficiente de concordância Kappa, calculados para cada conjunto de dados. As variáveis independentes do modelo são as mesmas selecionadas conforme subseção 2.2.3.

 $<sup>^{1}</sup>$  Aqui, QA se refere ao parâmetro VI Usefulness como se fosse uma banda propriamente dita.

Tabela 11 – Critério de formação dos conjuntos de dados do produto MOD13Q1 criados com base nos níveis de qualidade dos pixels

Conjunto	Critério utilizado
PR_0	Pixels com valores de $PR = 0$
$QA_{0000}$	Pixels com valores de $0 \le PR \le 1$ e VI Usefulness = 0000
$QA_{0001}$	Pixels com valores de $0 \le PR \le 1$ e VI Usefulness $\le 0001$
$QA_0010$	Pixels com valores de $0 \le PR \le 1$ e VI Usefulness $\le 0010$
$QA_{0011}$	Pixels com valores de $0 \le PR \le 1$ e VI Usefulness $\le 0011$
$QA_0100$	Pixels com valores de $0 \le PR \le 1$ e VI Usefulness $\le 0100$
QA_0101	Pixels com valores de $0 \le PR \le 1$ e VI Usefulness $\le 0101$
$QA_0110$	Pixels com valores de $0 \le PR \le 1$ e VI Usefulness $\le 0110$
QA_0111	Pixels com valores de $0 \le PR \le 1$ e VI Usefulness $\le 0111$
$QA_{1000}$	Pixels com valores de $0 \le PR \le 1$ e VI Usefulness $\le 1000$
QA_1001	Pixels com valores de $0 \le PR \le 1$ e VI Usefulness $\le 1001$
QA_1010	Pixels com valores de $0 \le PR \le 1$ e VI Usefulness $\le 1010$
QA_1011	Pixels com valores de $0 \le PR \le 1$ e VI Usefulness $\le 1011$
QA_1100	Pixels com valores de $0 \le PR \le 1$ e VI Usefulness $\le 1100$
$PR_3$	Pixels com valores de $1 \le PR \le 3$

# 2.2.6 Comparação dos pontos de monitoramento *in situ* com os pixels MO-DIS

O desempenho do modelo usando as medidas pontuais *in situ* de inundação e os pixels de 250 metros do MODIS foi comparado com o desempenho de um novo modelo, em que cada medida pontual foi substituída por um conjunto de réguas virtuais.

Conjecturou-se que a porcentagem de área inundada simulada a partir de um modelo numérico de terreno (MNT) ao redor dos pontos de monitoramento *in situ* pudesse representar melhor o estado de inundação na área de abrangência de um pixel MODIS do que a observação pontual, conforme estudo semelhante apresentado em Fensholt & Sandholt (2005).

## 2.2.6.1 Geração do Modelo Numérico de Terreno

Para estimar a inundação nas réguas virtuais a partir da inundação nas réguas reais, foi utilizado um MNT obtido por krigagem ordinária dos pontos de levantamento topográfico. A qualidade de MNTs obtidos por interpolação de dados pontuais depende: (i) da qualidade individual dos pontos; (ii) da densidade dos pontos; (iii) da distribuição desses pontos dentro da superfície; e (iv) do método de interpolação (LANE, 1998; HERITAGE et al., 2009).

O levantamento topográfico obtido de Fantin-Cruz (2008) foi realizado com

dois Receptores GPS L1/L2, (Leica 900 CS). Todos os pontos foram adquiridos em modo estático para pós-processamento, com precisão nominal horizontal de 5 mm  $\pm$  0,5 ppm e vertical de 10 mm  $\pm$  0,5 ppm, utilizando o WGS 1984 como elipsoide de referência. Os cálculos da altitude geoidal foram realizados a partir do programa MAPGEO2004 (www.ibge.gov.br) com o Datum SAD69 como geoide de referência. Foram adquiridos 191 pontos, o que corresponde a uma densidade de 7,64 pontos por km<sup>2</sup>.

Técnicas de krigagem presumem autocorrelação espacial entre as medidas, comumente avaliada pelo Índice Global de Moran (MORAN, 1950; HUBERT et al., 1981). A krigagem ordinária, na qual as médias locais variam e são reestimadas de acordo com valores da vizinhança, tem como pressupostos adicionais, distribuição normal e estacionariedade de primeira ordem (ausência de tendência ou *drift*). Para avaliação desses pressupostos foram realizadas inspeções visuais de gráficos Quantil-Quantil (*Q-Q plots*) e conduzidos testes de aderência dos dados de entrada e dos resíduos após validação cruzada. Gráficos Q-Q permitem determinar se dois conjuntos de dados pertencem à mesma distribuição de probabilidades, hipótese testável a partir dos testes de aderência de Kolmogorov-Smirnov e Shapiro-Wilk, sendo o segundo comumente considerado mais rigoroso (ZAR, 1999). Foram também avaliados visualmente os semivariogramas omnidirecionais, que representam quantitativamente a variação de um fenômeno regionalizado e evidencia a estrutura espacial desse fenômeno (LANDIM et al., 2002).

Heterogeneidade nas feições morfológicas pode causar a violação dos prérequisitos para ajustar modelos de interpolação por krigagem ordinária. Para subsidiar uma análise da distribuição espacial das formas geomorfológicas, foi utilizada uma imagem multiespectral do SPOT, do ano de 2007, com resolução espacial de 2,5 m.

De acordo com estas análises, foi realizada uma krigagem ordinária, utilizando as opções padrões para ajuste do variograma (modelo esférico) como dado no ArcGIS 10.2 (ESRI, 2013). Heritage et al. (2009) mostram que krigagem baseada em opções padrões de variograma resultam em interpolações satisfatórias.

O MNT foi reamostrado para uma resolução horizontal de 30 m. De acordo com as recomendações de Hengl (2006) para estimativa da resolução da grade gerada a partir da densidade das observações, são recomendadas resoluções entre 18,1 m e 36,2 m.

## 2.2.6.2 As Réguas Virtuais

Para cada régua limnimétrica instalada no SALD Pirizal, foram criadas cem réguas virtuais com espaçamento de 25 m entre elas, em forma de grade quadrada, preenchendo todo o pixel do MODIS (250 m de resolução espacial).

As coordenadas X e Y (projeção UTM) de cada conjunto de cem réguas virtuais foram calculadas a partir das coordenadas da régua real localizada naquele pixel. Esse procedimento foi realizado para as quarenta réguas do SALD Pirizal, criando, assim, um total de 4000 réguas virtuais.

A Figura 12 ilustra um exemplo de um conjunto de 100 réguas virtuais localizadas dentro da grade de 250 metros do MODIS.



Figura 12 – Disposição das réguas virtuais na grade de 250 metros do MODIS, na região do SALD Pirizal

Usando o modelo de terreno previamente criado, as cotas topográficas das réguas virtuais foram extraídas e gravadas em uma planilha. As réguas localizadas nas cordilheiras tiveram suas cotas aumentadas em 2 metros, para que as mesmas não apresentem inundação durante todo o período. A inundação das cotas virtuais foi calculada através da equação (19).

$$I_V = Cota_R + I_R - Cota_V, \tag{19}$$

onde  $I_V$  representa a inundação da régua virtual;  $Cota_R$  é a cota topográfica da réguas real;  $I_R$  é a inundação da régua real; e  $Cota_V$  é a cota topográfica da régua virtual. Como a krigagem altera o valor dos pontos originais, foi utilizado como valor da  $Cota_R$  o valor da cota topográfica de uma régua virtual localizada na mesma coordenada.

Durante o período de esvaziamento das águas do SALD Pirizal, quando as réguas reais atingiam inundação igual a zero, as réguas virtuais, cujas cotas topográficas são mais baixas do que a cota da sua régua real correspondente, permaneciam com valor constante diferente de zero ( $I_V = Cota_R + 0 - Cota_V$ ). Para contornar o problema, os valores de inundação dessas réguas foram extrapolados, utilizando, para isso, a maior taxa de esvaziamento da régua real correspondente. A Figura 13 apresenta um exemplo de extrapolação da inundação de duas réguas virtuais.

![](_page_66_Figure_3.jpeg)

Figura 13 – Exemplo de extrapolação da inundação das réguas virtuais. As réguas com apóstrofe representam a réguas com valores extrapolados. A régua A1\_9\_9 é a régua de referência.

Foram criados dez conjuntos de inundação para o SALD Pirizal. Em cada conjunto, o estado da inundação de cada pixel foi relacionado com uma quantidade mínima de inundação daquele pixel. A porcentagem de inundação de um determinado pixel foi medida pela quantidade de réguas virtuais inundadas naquele pixel. Assim, por exemplo, um pixel é considerado inundado, quando pelo menos 10 réguas virtuais naquele pixel estão inundadas. Os conjuntos foram divididos em intervalos de 10%.

Assim, para cada conjunto foi ajustado um modelo de regressão logística, tendo como variáveis preditivas os índices LSWI e EVI. A medida da qualidade de ajuste do modelo para cada conjunto de dados foi dada pelo coeficiente de determinação de Nagelkerke  $(R^2)$  e pelo coeficiente de concordância Kappa. A qualidade do ajuste do modelo usando as medidas das réguas pontuais foi utilizada como referência.

# **3 APRESENTAÇÃO DOS RESULTADOS**

# 3.1 O MODELO DE INUNDAÇÃO

#### 3.1.1 Escolha das Variáveis Independentes

Para a escolha das variáveis independentes, ou variáveis preditivas, foram utilizados apenas os dados de qualidade PR\_0. Foram usados os índices LSWI, EVI e NDVI, e a variável categórica (Land) com a classificação do uso e ocupação da terra.

Quatro tipos de uso e ocupação da terra foram encontrados nas localidades dos pontos (réguas) de monitoramento da inundação. São eles: DB: Floresta Decídua Latifoliada (*Deciduous broadleaf*); EB: Floresta Sempre-Verde Latifoliada (*Evergreen broadleaf*); S: Savana (*Savanna*); WS: Savana arborizada (*Wood Savanna*).

A Tabela 12 apresenta os coeficientes estimados pela regressão logística. Também são apresentados, para cada um desses coeficientes, o erro padrão (E. P.), o teste da estatística Wald (z) e o valor-p associado a esse teste.

Variável	Coeficiente	E. P.	z	valor- $p$
Intercepto	-7,164	1,488	-4,813	< 0,001
LSWI	$25,\!936$	1,754	14,791	< 0,001
NDVI	-3,550	$3,\!149$	-1,127	0,260
EVI	-13,458	2,323	-5,794	< 0,001
LandEB	-2,192	$0,\!457$	-4,797	< 0,001
LandS	-0,051	$0,\!406$	-0,126	0,900
LandWS	-0,053	$0,\!474$	-0,112	0,911
AIC: 712,84				

Tabela 12 – Resultado do ajuste do modelo de regressão logística com os dados do SALD Pirizal e do transecto do SESC

Verifica-se que, para um nível de significância de 5%, a variável NDVI não é estatisticamente significativa. A Tabela 13 apresenta o diagnóstico de multicolinearidade das variáveis independentes, realizado através do cálculo do fator de inflação da variância (VIF) para cada variável independente.

Tabela 13 – Fator de inflação da variância (VIF), para determinação de multicolinearidade das variáveis independentes, e grau de liberdade (gl). Dados usados no ajuste do modelo para o SALD Pirizal e transecto do SESC

	VIF	gl
LSWI	2,408	1
NDVI	5,218	1
EVI	$3,\!539$	1
Land	$1,\!652$	3

Segundo Logan (2010, p. 211), valores de VIF maiores que 5 indicam colinearidade.

Através do método automatizado *Stepwise* de seleção de variáveis por comparação do AIC, obteve-se o modelo mais parcimonioso. A Tabela 14 apresenta os coeficientes estimados das variáveis independentes para esse novo modelo, juntamente com o erro padrão (E. P.), o teste de estatística Wald (z), e o valor-p associado a esse teste.

Tabela 14 – Resultado do ajuste do modelo de regressão logística com os dados do SALD Pirizal e do transecto do SESC após excluir a variável NDVI

Variável	Coeficiente	E. P.	z	valor- $p$
Intercepto	-8,460	0,961	-8,802	< 0,001
LSWI	$24,\!970$	1,508	$16,\!556$	< 0,001
EVI	-15,358	$1,\!640$	-9,366	< 0,001
LandEB	-2,231	$0,\!457$	-4,882	< 0,001
LandS	-0,013	$0,\!404$	-0,033	$0,\!974$
LandWS	-0,067	$0,\!474$	-0,141	0,888
AIC: 712,10				

Embora as variáveis *dummies* LandS e LandWS não sejam significativas no modelo, sempre que uma variável categórica é incluída (ou excluída) em um modelo, todas as suas variáveis *dummies* devem ser incluídas (ou excluídas) (HOSMER JR et al., 2013).

A Tabela 15 apresenta o fator de inflação da variância (VIF) para verificar se nesse modelo ainda há a presença de colinearidade entre as variáveis independentes.

Tabela 15 – Fator de inflação da variância (VIF), para determinação de multicolinearidade das variáveis independentes, e grau de liberdade (gl), após excluir a variável NDVI. Dados usados no ajuste do modelo para o SALD Pirizal e transecto do SESC

	VIF	gl
LSWI	1,770	1
EVI	1,770	1
Land	$1,\!585$	3

## 3.1.2 Qualidade de Ajuste do Modelo

A Figura 14 apresenta os valores da sensibilidade e especificidade para vários pontos de corte em intervalos de 0,05, para determinação do melhor ponto de corte, com o objetivo de classificar a inundação a partir das probabilidades estimadas. O ponto de corte que maximiza a sensibilidade e a especificidade ocorre na interseção das duas curvas. Nesse caso, o ponto de corte vale aproximadamente 0,5.

![](_page_70_Figure_5.jpeg)

Figura 14 – Sensibilidade e especificidade do modelo em função do ponto de corte. A interseção das duas curvas indica o melhor ponto de corte para o modelo

A Tabela 16 apresenta o resultado da validação interna do modelo, ou seja, usando apenas os dados utilizados no ajuste do modelo.

A Tabela 17 apresenta o resultado da validação cruzada do tipo k-fold com k = 10.

A Tabela 18 apresenta o resultado da validação externa do modelo, ou seja, usando os dados do ano 2007 para o transecto do SESC e do ano 2009 para o SALD Pirizal.

	Observado		
Classificado	Não inundado	Inundado	
Não inundado	618	69	
Inundado	66	496	

Tabela 16 – Resultado da validação interna da regressão logística

Tabela 17 – Resultado da validação cruzada tipo<br/> k-fold com k=10 da regressão logística

	Observado		
Classificado	Não inundado	Inundado	
Não inundado	614	72	
Inundado	70	493	

Tabela 18 – Resultado da validação externa da regressão logística

	Observado		
Classificado	Não inundado	Inundado	
Não inundado	289	63	
Inundado	72	169	

A Tabela 19 apresenta resumo das três validações.

Tabela 19 – Acurácia, taxas de falso positivo (TFP) e falso negativo (TFN), e coeficiente de concordância Kappa para as validações: interna (ajuste do modelo), cruzada tipo k-fold (k = 10) e externa

	Validação		
	Interna	k-fold	Externa
Acurácia	89,2%	$88,\!6\%$	$77,\!2\%$
TFP	$9{,}6\%$	10,2%	19,9%
TFN	$12,\!2\%$	12,7%	$27,\!2\%$
Kappa	0,782	0,770	0,525

Enquanto na validação cruzada k-fold ocorre uma redução de menos de 1% na acurácia do modelo, em relação à validação interna, na validação externa essa redução é de 12%.

O poder de discriminação do modelo é apresentado através de quatro gráficos na Figura 15. Também foram ajustados três modelos, cada um com apenas uma variável independente, os índices LSWI, NDVI e EVI, a fim de verificar o poder de discriminação desses modelos para classificação da inundação usando esses índices, respectivamente (Figuras 16, 17 e 18).


Figura 15 – Capacidade de discriminação do modelo usando os índices LSWI, EVI e a variável categórica Land



Figura 16 – Capacidade de discriminação do modelo usando apenas a variável LSWI



Figura 17 – Capacidade de discriminação do modelo usando apenas a variável NDVI



Figura 18 – Capacidade de discriminação do modelo usando apenas a variável EVI

#### 3.1.3 Modelo Simplificado

Embora a variável categórica com a classificação do uso e ocupação da terra tenha sido significativa no modelo, optou-se por criar um modelo sem essa variável para demonstrar seu efeito em mapas de classificação da inundação.

Após realizar a seleção automática *Stepwise*, as variáveis independentes LSWI, NDVI e EVI foram selecionadas, com AIC = 755,64. No teste de multicolinearidade a variável NDVI obteve um VIF = 4,7. Embora esse valor seja menor do que o máximo sugerido por Logan (2010, p. 211), optou-se por retirar do modelo essa variável. O modelo apenas com as variáveis LSWI e EVI obteve um AIC = 758,39, com redução do fator de inflação da variância em relação ao modelo anterior.

A Tabela 20 apresenta os coeficientes estimados das variáveis independentes, o erro padrão (E. P.), o teste de estatística Wald (z) e o valor-p associado a esse teste.

Tabela 20 – Variáveis na equação logística para o modelo sem a classificação do uso e ocupação da terra

Variável	Coeficiente	E. P.	z	valor- $p$
Intercepto	-6,693	$0,\!657$	-10,180	< 0,001
LSWI	$21,\!935$	1,265	$17,\!340$	< 0,001
EVI	-15,997	$1,\!438$	-11,120	< 0,001

AIC: 758,39

### 3.2 COMPARAÇÃO PONTO COM PIXEL

#### 3.2.1 Interpolação do Modelo Numérico de Terreno

Uma comparação da distribuição das cotas altimétricas do levantamento topográfico com a interpretação visual da imagem SPOT (Figura 19) mostra um desvio expressivo das cotas acima de 116 m dos valores padronizados de uma distribuição normal teórica (Figura 20). Esses locais representam exclusivamente cordilheiras.

Portanto, antes de se realizar esta interpolação do MNT, foi criada uma máscara para as formações do tipo cordilheiras que permanecem não inundadas durante todo ano hidrológico, e gerado um conjunto reduzido, excluindo os pontos do levantamento topográfico localizados neste tipo de formação. Em um teste de aderência de Kolmogorov-Smirnov, a hipótese nula da Normalidade é rejeitada (p < 0, 05) para o conjunto total dos pontos de levantamento topográfico, mas



Figura 19 – Interpretação visual de uma imagem SPOT de 2007 para identificar as cotas altimétricas do levantamento topográfico localizadas nas cordilheiras do SALD Pirizal



Figura 20 – Gráficos quantil-quantil para os pontos de levantamento topográfico (a) sem e (b) com descarte de medidas em cordilheiras

aceita para o conjunto reduzido (p > 0, 05). No teste de Shapiro-Wilks a hipótese da normalidade é ainda aceita para o conjunto reduzido para um intervalo de confiança de 99% (p > 0, 01).

A análise da autocorrelação espacial mostra um aumento do Índice Global de Moran de 0,28 (p < 0,05) do conjunto completo para 0,60 (p < 0,01) para o conjunto reduzido.

Em seguida, foram ajustados os modelos de semivariogramas experimentais para os dois conjuntos de pontos e realizada uma krigagem ordinária, com modelo de semivariograma esférico, raio de busca variável e número de pontos igual a 12. O MNT com exclusão das cotas em cordilheiras mostra a inclinação da planície na área do SALD Pirizal em direção NNE-SSW, com as altitudes cerca de 2 m mais baixas na sua região sul (Figura 21).



Figura 21 – Modelo numérico de terreno do SALD Pirizal obtido por krigagem dos pontos de levantamento topográfico após a exclusão dos pontos localizados nas cordilheiras. São apresentadas, também, as quarenta réguas limnimétricas

As saídas da validação cruzada mostram um bom alinhamento dos valores interpolados próxima à bissetriz e com pequena dispersão dos valores ao longo da reta de regressão formada com as elevações medidas (Figura 22a). O gráfico quantilquantil dos erros padronizados mostra ainda o típico alisamento da krigagem com subestimativas dos valores extremos (GOOVAERTS, 1997), porém com boa concordância (Figura 22b).



Figura 22 – a) Reta de regressão formada com as elevações estimadas e medidas b) Gráfico quantil-quantil dos erros padronizados

Em uma validação cruzada, a Raiz do Erro Quadrático Médio (RMSE) de 0,60 m para o conjunto completo diminuiu para 0,31 m para o conjunto reduzido. Salienta-se que esses erros se referem aos pontos do levantamento da grade que possuem distância média mínima entre si de 249 m, enquanto a distância média das réguas virtuais para os pontos de monitoramento dentro dos pixels MODIS é de 135 m. Assumindo proporcionalidade, o RMSE do MNT nas áreas de abrangência dos pixels MODIS para a altitude das réguas virtuais pode ser estimado em menos de 0,17 m.

#### 3.2.2 Desempenho do Modelo na Utilização das Réguas Virtuais

A Tabela 21 apresenta o desempenho das réguas virtuais nos modelos preditivos da inundação, onde a porcentagem de inundação do pixel, dada pela quantidade de réguas virtuais inundadas, define a variável resposta binária de inundação de cada pixel.

Com base no coeficiente de concordância Kappa, o modelo usando apenas medidas pontuais obteve o melhor desempenho. Por outro lado, houve um desempenho similar entre o modelo com réguas pontuais e o modelo de réguas virtuais com 10% de inundação, com base no coeficiente de determinação de Nagelkerke  $(R^2)$ . Foram usados apenas os dados com qualidade PR\_0.

Porcentagem de Inundação	$R^2$	Kappa
10	0,755	0,802
20	0,718	0,762
30	0,716	0,770
40	$0,\!695$	0,744
50	$0,\!644$	$0,\!679$
60	$0,\!631$	$0,\!619$
70	$0,\!497$	$0,\!493$
80	$0,\!439$	$0,\!378$
90	$0,\!459$	0,317
100	$0,\!198$	-0,002
Medidas pontuais	0,755	$0,\!830$

Tabela 21 – Desempenho do modelo em função da porcenta<br/>gem de inundação dos pixels dada pelas réguas virtuais

# 3.3 VARIABILIDADE INTERANUAL DA QUALIDADE DO PRODUTO MOD13Q1

A Figura 23 apresenta a variabilidade anual da qualidade dos pixels, parametrizados apenas pela banda *Pixel Reliability* (PR\_0, PR\_1 e PR\_3), do produto Índice de Vegetação MOD13Q1, ao longo de 13 anos (2000 a 2012), usando como início do ano hidrológico as imagens do dia 273 (29/30 de setembro) ao dia 257 (13/14 de setembro).



Figura 23 – Composição da imagem em relação aos valores de Pixel Reliability (PR = 0, PR = 1 e PR = 3). Imagens de 29/09/2000 (2000/273) a 14/09/2012 (2012/257)

Para cada dia do ano (DOY) existem até três boxplots, cada um mostrando a variabilidade interanual das qualidades PR\_0, PR\_1 e PR\_3. Os pontos representam os *outliers*.

Do dia 97 ao dia 225, salvo algumas exceções (outliers), todos os pixels possuem qualidade PR\_0, representados pelos traços na marca de 100%.

As Figuras 24 e 25 apresentam a composição da qualidade dos pixels localizados nas coordenadas das réguas do transecto do SESC, em intervalos de 16 dias, dos anos 2005 e 2006, respectivamente. Os pontos, ligados pela linha, representam a porcentagem dessas réguas que estavam inundadas em cada data. A porção vermelha das barras representam os pixels nublados (PR\_3). As outras cores representam os níveis de qualidade dos pixels, variando de QA\_0001 a QA\_0111.



Figura 24 – Composição da qualidade dos pixels (barras) e porcentagem de réguas inundadas (linha) para o ano de 2005, na região do transecto do SESC

Para as réguas do SALD Pirizal, temos de maneira análoga ao já descrito, as Figuras 26 e 27 apresentando as medidas, respectivamente, dos anos 2007 e 2008. As medidas de 2007 se iniciam no dia 353 de 2006 (19 de novembro) e terminam no dia 241 de 2007 (29 de agosto). Para o ano de 2008, as medidas se iniciam no dia 337 de 2007 (3 de dezembro) e terminam no dia 241 de 2008 (28 de agosto).



Figura 25 – Composição da qualidade dos pixels (barras) e porcentagem de réguas inundadas (linha) para o ano de 2006, na região do transecto do SESC



Figura 26 – Composição da qualidade dos pixels (barras) e porcentagem de réguas inundadas (linha) para o período de 2006/2007, na região do SALD Pirizal



Figura 27 – Composição da qualidade dos pixels (barras) e porcentagem de réguas inundadas (linha) para o período de 2007/2008, na região do SALD Pirizal

## 3.4 DESEMPENHO DO MODELO EM FUNÇÃO DA QUALIDADE DOS PIXELS

Verificou-se que para os pontos de monitoramento da inundação não houve pixels com qualidade inferior a QA\_0110. Assim, não foi avaliada a influência da qualidade dos pixels sobre o desempenho do modelo para os níveis de qualidade abaixo desse valor. Verificou-se, também, que para toda a área de estudo, e não apenas para os pontos de monitoramento, não foi encontrado nenhum pixel com a qualidade QA\_0000, que, segundo Solano et al. (2010), representa os pixels de maior qualidade.

O desempenho do modelo em função da qualidade dos pixels, usando como medidor do desempenho os valores de  $R^2$  e Kappa, está sumarizado na Tabela 22.

Os conjuntos de dados PR\_3 e TIMESAT possuem o maior tamanho, n = 1444, uma vez que o primeiro possui todas as qualidades de pixels, incluindo os nublados, e o segundo, devido ao preenchimento de falhas. Por outro lado, o conjunto QA\_0001 possui o menor tamanho, indicando que a filtragem para esse nível de qualidade é mais restritivo do que usando apenas PR = 0.

Com relação à qualidade de ajuste do modelo em função da qualidade dos pixels, o modelo usando dados reconstruídos pelo TIMESAT obteve o maior

Tabela 22 – Desempenho do modelo em função da qualidade dos pixels. Estão incluídos, também, os dados do tipo PR\_3 e reconstruídos (TIME-SAT)

Conjunto	n	$\mathbb{R}^2$	Acurácia (%)	Kappa	TFP $(\%)$	TFN (%)
PR_0	1249	0,746	89,2	0,782	$9,\!6$	12,2
$QA_{0001}$	1193	0,761	89,0	0,777	9,5	12,9
$QA_{0010}$	1296	0,749	89,0	0,777	10,0	12,3
$QA_{0011}$	1326	0,743	88,0	0,759	$11,\!6$	12,5
QA_0100	1343	0,744	88,3	0,765	$11,\!4$	12,0
QA_0101	1355	0,744	88,2	0,763	11,7	12,0
QA_0110	1359	0,744	88,2	0,762	11,7	12,0
$PR_3$	1444	$0,\!699$	86,9	0,738	14,2	12,0
TIMESAT	1444	0,767	88,6	0,773	11,7	$11,\!0$

desempenho, com base no valor de  $R^2$ . Usando Kappa como medidor de desempenho do modelo, o conjunto formado por dados do tipo PR\_0 obteve o melhor resultado. O modelo usando os dados PR\_3 obteve o pior desempenho, com base nos valores de  $R^2$  e Kappa.

Os conjuntos de dados foram separados através da subtração dos conjuntos, com o objetivo de apresentar a qualidade de ajuste do modelo: apenas para os pixels com qualidade diferente de PR\_0; apenas para os pixels marcados como nublado (PR\_3); e apenas para pixels preenchidos pelo TIMESAT. A Tabela 23 apresenta o desempenho do modelo para esses novos conjuntos de dados.

Conjunto	n	Acurácia (%)	Kappa	TFP $(\%)$	TFN (%)
QA_0010 - PR0	47	89,4	$0,\!693$	0,0	38,5
$QA_0011 - PR0$	77	87,0	0,741	$0,\!0$	$23,\!3$
$QA_0100 - PR0$	94	90,4	$0,\!808$	$^{0,0}$	18,0
$QA_0101 - PR0$	106	$90,\!6$	0,810	$0,\!0$	16,1
$QA_0110 - PR0$	110	90,0	0,798	$^{0,0}$	16,7
$PR_3 - PR0$	195	81,0	$0,\!580$	$^{6,1}$	$23,\!3$
TIMESAT - PR0	195	89,7	0,742	10,2	10,3
$PR_3 - QA_0110$	85	67,1	-0,004	60,0	$31,\!3$
TIMESAT – $QA_{0110}$	85	85,9	-0,172	$100,\!0$	8,8

Tabela 23 – Desempenho do modelo em função da qualidade dos pixel para os dados adicionais ao conjunto  ${\rm PR}\_0$ 

O conjunto PR\_3 – PR\_0 possui todos os pixels diferentes de PR\_0, enquanto o conjunto PR\_3 – QA\_0110 possui apenas os pixels marcados como nublados. O conjunto TIMESAT – PR\_0 possui todos os pixels reconstruídos pelo TIMESAT cujos pixels originais são diferentes de PR\_0, enquanto o conjunto TIMESAT – QA\_0110 possui todos os pixels cujos originais estavam marcados como nublados e foram excluídos da série original.

A classificação para o conjunto PR\_3 – QA\_0110 é apresentada pela Tabela 24.

Tabela 24 – Resultado da classificação para a diferença entre os conjuntos PR\_3 e QA\_0110

	Observado		
Classificado	Não inundado	Inundado	
Não inundado	2	25	
Inundado	3	55	

Verifica-se a ocorrência de 80 casos observados como inundado, e apenas 5 casos observados como não inundado. Há três casos classificados erradamente como inundados, que correspondem a TFP = 60,0%, e 25 casos classificados erradamente como não inundado, que corresponde a TFN = 31,3%.

Embora a concordância observada (acurácia) tenha sido de 67,1%, o valor do coeficiente Kappa apresentou um valor próximo de zero, que indica uma concordância por acaso. Essa discrepância entre esses valores (acurácia e Kappa) se deve à pequena quantidade de casos observados como não inundados em relação aos casos observados como inundados, tornando o Kappa uma medida não confiável nessa situação (VIERA; GARRETT, 2005).

A classificação para o conjunto TIMESAT – QA\_0110 é apresentada pela Tabela 25.

Tabela 25 – Resultado da classificação para a diferença entre os conjuntos TIME-SAT <br/>e $\rm QA\_0110$ 

	Observado		
Classificado	Não inundado	Inundado	
Não inundado	0	7	
Inundado	5	73	

Assim como descrito acima, há 80 casos observados como inundado para apenas 5 casos observados como não inundado. Entretanto, para esse conjunto, há 5 casos classificados erradamente como inundado, que corresponde a TFP = 100,0%, e apenas 7 casos classificados erradamente como não inundado, que corresponde a TFN = 8,8%.

Novamente, há uma discrepância entre a acurácia (85,9%) e o coeficiente de concordância Kappa (-0,172).

## 3.5 DESEMPENHO DO MODELO EM FUNÇÃO DO CICLO HIDROLÓ-GICO

As Figuras 28 e 29 apresentam as taxas de falso positivo (TFP) e falso negativo (TFN), e a porcentagem de réguas inundadas, para os anos de 2005 e 2006, respectivamente, usando imagens MODIS de qualidade PR\_0 e imagens reconstruídas pelo TIMESAT.



Figura 28 – Taxas de falso positivo (TFP) e falso negativo (TFN), gráfico de barras, e porcentagem de réguas inundadas para as medidas do transecto SESC RPPN do ano de 2005



Figura 29 – Taxas de falso positivo (TFP) e falso negativo (TFN), gráfico de barras, e porcentagem de réguas inundadas para as medidas do transecto SESC RPPN do ano de 2006

Semelhantemente, as Figuras 30 e 31 apresentam as taxas de falso positivo (TFP) e falso negativo (TFN) e a porcentagem de réguas inundadas para o SALD Pirizal, para os períodos 2006/2007 e 2007/2008, respectivamente.



Figura 30 – Taxas de falso positivo (TFP) e falso negativo (TFN), gráfico de barras, e porcentagem de réguas inundadas para as medidas do SALD Pirizal, para o período de medida 2006/2007



Figura 31 – Taxas de falso positivo (TFP) e falso negativo (TFN), gráfico de barras, e porcentagem de réguas inundadas para as medidas do SALD Pirizal, para o período de medida 2007/2008

A Figura 32 apresenta vinte e três mapas de inundação da área de estudo para o ano hidrológico de outubro de 2006 a setembro de 2007, preditos pelo modelo usando imagens de qualidade PR\_0, usando ponto de corte igual a 0,5.



Figura 32 – Mapas de inundação da área de estudo predita pelo modelo usando dados de qualidade PR\_0. Ano hidrológico: outubro de 2006 a setembro de 2007

A Figura 33 apresenta vinte e três mapas de inundação da área de estudo para o ano hidrológico de outubro de 2006 a setembro de 2007, preditos pelo modelo usando as imagens reconstruídas pelo TIMESAT.



Figura 33 – Mapas de inundação da área de estudo predita pelo modelo usando dados reconstruídos pelo TIMESAT. Ano hidrológico: outubro de 2006 a setembro de 2007

Em ambas as Figuras (32 e 33), as áreas em bege indicam não inundado, enquanto as áreas em azul marinho, inundado. As áreas em vermelho indicam áreas não classificadas, ou por terem qualidade inferior a PR\_0 (apenas a Figura 32), ou por possuírem máscara de corpos d'água permanente, como é o caso da Baía de Chacororé na parte superior direita do mapa. Os principais rios da região foram incluídos para referência da inundação na área de estudo.

A Figura 34 apresenta em destaque a região do transecto do SESC classificado pelo modelo com dados do TIMESAT (a) e com dados da série original com qualidade PR\_0 (b), para o dia 225/2007 (13 de agosto).



Figura 34 – Mapas de inundação da região do transecto do SESC Pantanal de 13 de agosto de 2007 (dia 225) gerados a partir do modelo a) com dados reconstruídos pelo TIMESAT e b) com dados originais de qualidade PR\_0

### 3.6 DESEMPENHO EM FUNÇÃO DO TIPO DE COBERTURA

A Figura 35 apresenta as taxas de falso positivo (TFP) e falso negativo (TFN) em função do uso e ocupação da terra e do ciclo de inundação, para o modelo usando a série original com qualidade PR\_0 e para o modelo usando os dados reconstruídos pelo TIMESAT. A classificação do uso e cobertura da terra foi feita visualmente usando uma imagem de 2007 do satélite SPOT, com base na localização das réguas.

Verifica-se que no período de transição (períodos de enchente e de vazante), ocorre o maior número de erros de classificação. Nesse período, as TFP para o modelo com dados reconstruídos e para o modelo com dados originais são, respectivamente, 53% e 47% para a formação de Florestas, 34% e 32% para o Cerrado, e 27% e 18% para a Pastagem. Para esse mesmo período, as maiores



Figura 35 – Taxas de Falso Positivo (TFP) e Falso Negativo (TFN) em função do uso e ocupação da terra e do ciclo de inundação. São apresentadas as TFP e TFN para a série original com qualidade PR\_0 e para a série reconstruída pelo TIMESAT. Dados usados no ajuste do modelo para as regiões do transecto do SESC e do SALD Pirizal.

TFN ocorrem para o Cerrado (41% e 26%), seguido das Florestas (19% para ambas as séries) e, por último, da Pastagem (14% e 16%).

O período seco foi o que menos apresentou erros de classificação, com a formação de Cerrado não tendo nenhum erro, tanto na série original quanto na série reconstruída. As áreas de Florestas apresentaram 5% de TFP para o modelo com dados reconstruídos e cerca de 3% para o modelo com dados originais. A Pastagem apresentou menos de 3% em ambas as séries.

Durante o período inundado, as formações de Florestas apresentaram as maiores TFP (12% e 14%), seguido da formação de Cerrado (6% e 11%) e, por último, da Pastagem (7% para ambos os modelos).

A Figura 36 apresenta cinco mapas de probabilidade de inundação estimada pelo modelo usando dados reconstruídos pelo TIMESAT, para três períodos do ano de 2007: cheia, vazante (transição) e seca. O gráfico na parte superior da figura apresenta a medida de inundação em três réguas localizadas em áreas distintas de uso e ocupação da terra: Floresta Inundável, Campo Cerrado e Pastagem Exótica. A parte inferior à direita apresenta uma imagem do SALD Pirizal obtida pelo satélite SPOT em 2007.



Figura 36 – Mapas com a probabilidade de inundação dada pelo modelo em três períodos de 2007 na região do SALD Pirizal: cheia, vazante (transição) e seca. Três tipos de uso e ocupação da terra são usados como referência: Floresta Inundável, Campo Cerrado e Pastagem Exótica. Imagem do SPOT de 2007 para a área do SALD Pirizal.

Usando o modelo simplificado, sem a variável categórica com a classificação do uso e ocupação da terra, foi gerado um mapa de classificação da inundação para o dia 13 de agosto de 2007. A Figura 37 apresenta dois mapas de inundação para uma área que abrange o SESC RPPN. O mapa da Figura 37a) usa o modelo simplificado (variáveis LSWI e EVI), enquando o da Figura 37b) usa o modelo completo (LSWI, EVI e Land). A Figura 37c) apresenta um mapa, da mesma região, com a classificação do uso e ocupação da terra do produto MCD12Q1.



Figura 37 – Mapas de inundação na região do transecto do SESC em 13 de agosto de 2007 gerados pelo modelos a) simplificado e b) completo. c) Uso e ocupação da terra na região do transecto do SESC fornecidos pela camada *Land Cover Type 2* (UMD) do produto MCD12Q1

O mapa gerado pelo modelo com a variável categórica de uso e ocupação da terra (Figura 37.b) possui menos erros de classificação em áreas florestadas quando comparado com o mapa gerado pelo modelo simplificado (Figura 37.a)

### 3.7 VARIABILIDADE ANUAL DA INUNDAÇÃO MEDIDA E PREDITA

A probabilidade de inundação por pixel em função dos índices LSWI e EVI, e da classificação do uso e ocupação da terra (variável categórica Land), usando dados de ajuste do modelo (dados de 2005 e 2006 para o transecto do SESC, e de 2006/2007 e 2007/2008 para o SALD Pirizal) é dada pelas Figuras 38 (dados com qualidade PR\_0) e 39 (dados reconstruídos pelo TIMESAT). Nessas figuras, os *Boxplots* mostram a distribuição da probabilidade de inundação por pixel para cada data do produto MOD13Q1, e as linhas vermelhas, a porcentagem de réguas inundadas medidas em campo, nessas mesmas datas.



Figura 38 – Distribuição da probabilidade de inundação por pixel (Boxplot) e porcentagem de réguas inundadas (linha vermelha). Probabilidades obtidas no ajuste do modelo com dados originais de qualidade PR\_0. Dados de 2005 e 2006 para o transecto do SESC e de 2006/2007 e 2007/2008 para SALD Pirizal. O período de 16 dias corresponde à periodicidade em que o produto MOD13Q1 é gerado.



Figura 39 – Distribuição da probabilidade de inundação por pixel (Boxplot) e porcentagem de réguas inundadas (linha vermelha). Probabilidades obtidas no ajuste do modelo com dados reconstruídos. Dados de 2005 e 2006 para o transecto do SESC e de 2006/2007 e 2007/2008 para SALD Pirizal. O período de 16 dias corresponde à periodicidade em que o produto MOD13Q1 é gerado.

As Figuras 40 (dados com qualidade PR\_0) e 41 (dados reconstruídos pelo TIMESAT) apresentam a probabilidade de inundação por pixel em função dos índices LSWI e EVI, e da classificação do uso e ocupação da terra (variável categórica Land), produzido pelo modelo com o uso dos dados de validação externa (dados de 2007 para o transecto do SESC e de 2008/2009 para o SALD Pirizal). Novamente, os *Boxplots* mostram a distribuição da probabilidade de inundação por pixel, e as linhas vermelhas, a porcentagem de réguas inundadas.



Figura 40 – Probabilidade de inundação por pixel (Boxplot) e porcentagem de réguas inundadas (linha vermelha). Validação externa do modelo com dados originais de qualidade PR\_0. O período de 16 dias corresponde à periodicidade em que o produto MOD13Q1 é gerado.



Figura 41 – Probabilidade de inundação por pixel (Boxplot) e porcentagem de réguas inundadas (linha vermelha). Validação externa do modelo com dados reconstruídos. O período de 16 dias corresponde à periodicidade em que o produto MOD13Q1 é gerado.

# 4 DISCUSSÃO DOS RESULTADOS

### 4.1 PODER PREDITIVO DAS VARIÁVEIS

Analisando a Tabela 14, verifica-se, pela estatística Wald e seu valor-*p*, a importância de cada variável no modelo (HOSMER JR et al., 2013). Dessa forma, verifica-se que o índice LSWI possui maior importância no modelo, seguido pelo índice EVI e pela classificação do uso e ocupação da terra, derivada do produto MCD12Q1.

De fato, o infravermelho de ondas médias (MIR), que compõe o índice LSWI, é altamente sensível ao teor de umidade no solo e no dossel da vegetação, como relatado em Gao (1996), McFeeters (1996), Xiao et al. (2002) e Rogers & Kearney (2004), justificando, assim, a importância desse índice em modelos classificatórios de inundação.

Do mesmo modo, o modelo ajustado apenas com o índice LSWI como variável preditiva, conforme se verifica pala análise gráfica através da Figura 16, ainda tem grande poder de discriminação entre os casos inundados e não inundados.

Por outro lado, modelos usando apenas os índices NDVI ou EVI (Figuras 17 e 18, respectivamente) possuem capacidade de discriminação reduzida. No caso do índice NDVI, a capacidade de discriminação pode ser classificada como apenas "aceitável", enquanto que para o EVI, uma "pobre" discriminação, somente ligeiramente superior do que um modelo nulo, de acordo com a interpretação da medida da área sob a curva (AUC) apresentada em Hosmer Jr et al. (2013, p. 177).

Embora o modelo usando apenas o índice EVI como variável preditiva seja similar a um modelo nulo, usando apenas o intercepto (p-valor = 0,142 em um teste da razão de verossimilhança), esse índice quando usado em conjunto com o índice LSWI passa a ser significativo no modelo (ver Tabela 14). A combinação desses índices para estimar áreas inundadas é usada em Xiao et al. (2005) e Sakamoto et al. (2007).

Resultados superiores do mapeamento da inundação a partir do índice

EVI (PADOVANI, 2010; ALMEIDA et al., 2015), quando comparados com o do presente trabalho, se explicam pelo fato de que as partes centrais da planície, principalmente ao longo do rio Paraguai, possuírem amplas áreas de alagamento, que são comumente bem detectadas pelo uso de qualquer banda espectral do infravermelho ou de índice incluindo esta faixa espectral. Ao contrário, na área de estudo do presente trabalho não ocorre grandes planícies sem cobertura vegetal.

Um estudo prévio no Pantanal usando imagens multitemporais do Landsat TM mostrou que índices de umidade do solo e o estado da inundação extraídos de sensores óticos variam fortemente em função da cobertura e densidade da vegetação verde (ZEILHOFER, 2006).

Portanto, utilizando a variável de classificação de uso e ocupação da terra, proveniente do produto MCD12Q1 como indicador da cobertura vegetal, verificouse que a mesma é significativa no modelo (com a variável: AIC = 712,1; sem a variável: AIC = 758,4), melhorando a qualidade de ajuste. Essa melhoria também foi constatada através da comparação de mapas de inundação criados a partir desses dois modelos (Figura 37). A importância dessa variável no modelo se evidenciou em regiões de florestas inundáveis como o Cambarazal (formação monodominante de *Vochysia divergens*) e a mata ciliar, em que o modelo simplificado superestimou a inundação nesse tipo de formação quando comparado com o modelo completo, principalmente em período de seca.

Também, analisando as medidas de inundação das réguas C, D e E do transecto do SESC, que foram classificadas como EB (*Evergreen Broadleaf*), verificou-se que o número de falsos positivos é menor no modelo completo do que no modelo simplificado, mostrando mais uma vez a importância dessa variável no modelo.

Embora o produto MCD12Q1 seja produzido em escala global de 500 m de resolução espacial (FRIEDL et al., 2010) e que pouco reproduz a heterogeneidade fitofisionômica do Pantanal, relacionada às variações do relevo e topografia (DAMASCENO-JUNIOR et al., 2005), à estrutura do solo (ZEILHO-FER; SCHESSL, 2000) e aos padrões espaço-temporal da inundação (ARIEIRA; NUNES DA CUNHA, 2006), sua utilização melhorou a qualidade de ajuste do modelo, conforme já descrito.

Vale ressaltar que, embora a acurácia geral do produto MCD12Q1 obtida em uma validação cruzada seja de 75%, com piores estimativas para algumas classes, como, por exemplo, *Savanna* e *Woody Savanna* (23% e 45%, respectivamente), a classe *Evergreen Broadleaf* apresentou uma acurácia de 93% (FRIEDL et al., 2010). Essa classe está presente em 24% da área de estudo. Na validação cruzada do tipo k-fold, com k = 10, o modelo obteve uma acurácia de 88,6%, próximo ao valor obtido na validação interna (89,2%).

Enquanto a validação cruzada apresentou uma redução de menos de 1% na acurácia do modelo, a validação externa apresentou uma redução de 12%, em relação à validação interna.

Em estudo similar, realizado na área úmida do Everglades, Flórida, Ordoyne & Friedl (2008) obtiveram, através de validação cruzada, uma acurácia de 83,7% na classificação da inundação. O modelo de regressão logística apresentado por esses autores é constituído de seis variáveis independentes, das quais quatro são derivadas de produtos MODIS: os índices *Tasseled Cap wetness*, *Tasseled Cap greenness* e NDVI, calculados a partir do produto MCD43B4, e a temperatura de superfície da terra (LST), fornecida pelo produto MYD11A2.

Usando o índice OWL (regressão logística com os coeficientes ajustados para o continente Australiano), Huang et al. (2014) obtiveram acurácia geral de 97,0% e 94,8% em uma validação interna para as duas regiões estudas, Narran Lakes e Chowilla, respectivamente, localizadas na Austrália. Entretanto, como a acurácia geral é uma simples porcentagem entre o número de pixels estimados corretamente (incluindo tanto pixels inundados como não-inundados) e o número total de pixels, as acurácias elevadas nessas duas regiões são devidas ao elevado número de pixels localizados em grandes corpos de água. Por outro lado, o coeficiente de concordância Kappa, que é uma medida mais adequada para avaliar a acurácia da detecção, foi de 0,6 para ambas as regiões. Esse valor representa uma acurácia aceitável, de acordo com Viera & Garrett (2005).

Embora no presente trabalho os dois sítios de estudo não sejam caracterizados por conter grandes corpos de água, e a acurácia geral obtida pelo modelo tenha sido ligeiramente inferior ao obtido pelo índice OWL no trabalho de Huang et al. (2014), ainda assim, o coeficiente de concordância Kappa apresentou resultado superior, cerca de 0,78.

### 4.2 O DESEMPENHO DAS RÉGUAS VIRTUAIS

O conceito de réguas virtuais foi introduzido para tentar melhorar a representatividade limitada que uma medida pontual tem em relação a uma entidade zonal de tamanho de 6,25 ha de pixel do MOD13Q1. No lugar de uma medida apenas, várias medidas distribuídas regularmente em uma grade de 25 m pelo pixel foram usadas, totalizando 100 réguas virtuais com nível de inundação simulada a partir do MNT de alta resolução espacial. Entretanto, o modelo ajustado com as medidas das réguas reais obteve melhor desempenho do que quando ajustado com as medidas provenientes das réguas virtuais. O modelo usando 10% das réguas virtuais inundadas para cada pixel como valor de corte obteve um desempenho mais próximo ao do modelo usando réguas reais. Isso foi verificado pelos coeficientes  $R^2$  de Nagelkerke e Kappa (ver Tabela 21). O ajuste do modelo com mais de 10% das réguas virtuais inundadas como valor de corte obteve resultado inferior ao modelo com réguas reais.

Examinando os casos classificados erradamente, para o modelo com valor de corte de 10% de réguas virtuais inundadas, verificou-se que o número de falsos positivos – dados classificados como inundado, porém não inundados – diminuiu em relação ao modelo usando apenas as réguas reais. Isso mostra a melhoria na representatividade das réguas virtuais quando comparadas com medidas pontuais. Por outro lado, o número de falsos negativos – dados classificados como não inundado, porém inundado – aumentou em relação ao modelo pontual.

Portanto, apesar de ter havido uma ligeira melhoria na classificação dos casos não inundados, a piora na classificação dos casos inundados levou a uma diminuição na acurácia geral, quando comparada à acurácia do modelo com réguas reais.

Uma explicação para o baixo desempenho das réguas virtuais pode estar relacionado à baixa magnitude de inundação no SALD Pirizal. De acordo com os resultados apresentados em Fantin-Cruz (2008), a utilização de cota topográfica não é considerada um bom indicativo da inundação na região, sendo a morfologia e o gradiente de escoamento de água fatores que atuam mais fortemente na determinação espacial da amplitude e duração da inundação.

Fensholt & Sandholt (2005), para avaliar o índice de vegetação de 500 m do MODIS quando comparado com medidas *in situ*, propuseram o uso de imagens de 30 metros do LANDSAT para minimizar o problema da comparação ponto-com-pixel. Embora o estudo desses autores não seja sobre inundação, foi uma abordagem similar usada como referência no presente estudo, visto que não foi encontrado na literatura uma abordagem similar para mapeamento da inundação. Os autores entretanto não apresentam um comparativo do desempenho com e sem o uso dessa abordagem.

Ordoyne & Friedl (2008), ciente da incompatibilidade de escala entre medidas pontuais e os pixels de 1 km do MODIS, usaram apenas locais que eram representativos do seu entorno (áreas com anomalias, tais como canais, estradas ou grandes ilhas de árvores foram excluídos). Ainda assim, cientes da variabilidade topográfica dentro do pixel, para converter as medidas de inundação para valores binários, usaram apenas as medidas acima de 15 cm da superfície, que foram codificadas em "1", e apenas as medidas abaixo de 15 cm da superfície, que foram codificadas em "0". Porém, o estudo não apresenta comparativo quantitativo de desempenho dos modelos com e sem a aplicação desse filtro ( $\pm 15$  cm).

## 4.3 VARIABILIDADE INTERANUAL DA QUALIDADE DO PRODUTO MOD13Q1 E A INFLUÊNCIA SOBRE AS CLASSIFICAÇÕES

Um dos principais objetivos que motivaram esse estudo foi investigar se a qualidade dos pixels do produto MODIS MOD13Q1 influencia o desempenho de modelos classificatórios de inundação.

Assim, analisando a Figura 23, verificou-se que a variabilidade da qualidade dos pixels do produto MOD13Q1 entre os anos aumenta principalmente entre o final de agosto e o final de março, período que corresponde ao fim do período de estiagem até o final do período chuvoso.

Os pixels nublados (PR\_3) ocorrem nas imagens do dia 273 (final de setembro) ao dia 81 (segunda quinzena de março), com menor frequência nos dias 49 e 65. As maiores variabilidades interanuais de pixels nublados ocorreram nas imagens do dia 321 ao dia 001, que correspondem à segunda quinzena de novembro, às duas quinzenas do mês de dezembro, e à primeira quinzena de janeiro.

Pixels com qualidade PR\_1 ocorrem do dia 241 (segunda quinzena de agosto) ao dia 81, com maior variabilidade interanual ocorrendo nas imagens dos dias 241, 257 e 273, compreendendo o final de agosto até o final de setembro.

Os pixels de boa qualidade (PR\_0) possuem variabilidade interanual do dia 241 ao dia 81, diminuindo a partir do dia 49. Nas imagens do dia 97 (início de abril) ao dia 225 (primeira quinzena de agosto), não houve variabilidade, salvo algumas exceções (*outliers*), com as imagens contendo 100% de pixels com PR\_0.

Contudo, isso não significa que não há formação de nuvens e nem grande concentrações de aerossóis nesse período. O real significado é que, no período de 16 dias de composição do produto, em pelo menos um dia foi capturada uma imagem em boas condições (SOLANO et al., 2010).

As Figuras 24 e 25 mostram, respectivamente, a composição da qualidade dos pixels para as réguas do transecto do SESC, para os anos de 2005 e 2006. De maneira análoga, as Figuras 26 e 27 mostram a composição para as réguas do SALD Pirizal, para os períodos 2006/2007 e 2007/2008. Verifica-se que os pixels de qualidade degradada, incluindo pixels contaminados por nuvens, ocorrem principalmente no período de inundação. A porcentagem de réguas inundadas serve de indicativo da inundação em cada área de estudo.

Em 2005 (Figura 24) as medidas em campo começaram na segunda quinzena de janeiro, portanto, a primeira imagem útil é a do dia 017/2005. Neste dia, verificase que menos de 20% dos pixels tem qualidade útil; o restante está coberto por nuvens (PR\_3). A partir do dia 33, todos os pixels (localizados nas réguas) tiveram qualidade máxima (QA\_0001), com exceção dos dias 65 e 241, cujos pixels tiveram qualidade QA\_0010 E QA\_0100, respectivamente. Para o período de estudo de 2006 (Figura 25), praticamente todos os pixels tiveram qualidade máxima (QA\_0001). Apenas para o dia 001/2006 ocorrem alguns pixels de qualidade QA\_0010 e QA\_0011.

A medição da inundação no Pirizal para o período 2006/2007 (Figura 26) começou em 29/12/2006, portanto, a imagem associada a esta medida é a 353/2006. Nesta imagem, todos os pixels tiveram qualidade QA\_0010. O Pirizal estava no período de enchente, com menos de 80% de réguas inundadas. Na data seguinte, 001/2007, todos os pixels estavam contaminados por nuvens. No dia 017/2007 alguns pixels também estavam contaminados por nuvens. No dia 033/2007 não houve pixels contaminados por nuvens, mas a qualidade estava abaixo da máxima. Apenas a partir do dia 049/2007 os pixels passaram a ter qualidade máxima, mas então, o Pirizal já começava a esvaziar-se.

No período 2007/2008 (Figura 27), cujas medidas de inundação no SALD Pirizal começaram em 06/12/2007, ocorreram pixels nublados em três ocasiões: 337/2007, 017/2008 e 049/2008. Para esses dias, além dos pixels nublados, vários outros com qualidade degradada ocorreram. Os pixels dos dias 353/2007 e 001/2008estavam com qualidade máxima.

Comparando a composição dos pixels das imagens dos períodos 2006/2007 e 2007/2008 (ambas do Pirizal), verifica-se que a última imagem contaminada por nuvens em 2007 foi a do dia 017, com o esvaziamento começando no dia 065, levando-se em conta apenas a data do produto. Portanto, 48 dias de diferença. Para o ano de 2008, a última imagem contaminada por nuvens foi a do dia 049, com o esvaziamento começando na imagem do dia 097. Também 48 dias de diferença.

A variabilidade interanual da qualidade dos pixels é percebida nesses quatro anos de estudo. Contudo, os pixels de pior qualidade ocorrem, principalmente, no início da inundação.

Diante dessa característica de variação da qualidade dos pixels MODIS na área de estudo, verificou-se que o conjunto QA\_0001 é o menor de todos (Tabela 22), tendo, portanto, um filtro mais restritivo no que se refere à qualidade dos pixel, enquanto o conjunto PR\_0 tem tamanho intermediário entre QA\_0001 e QA\_0010. Analisando minunciosamente o parâmetro *VI Usefulness* (Bits 2–5 da Tabela 3) do conjunto PR\_0, verificou-se a presença dos valores 0001 e 0010 dentro desse conjunto; a mesma qualidade dos conjuntos QA\_0001 e QA\_0010, respectivamente.

O conjunto de dados do tipo QA\_0110 tem 94% do total de dados (comparando com o conjunto PR\_3) e um desempenho dado por  $R^2$  de Nagelkerke e pelo coeficiente Kappa de 0,744 e 0,762, respectivamente. Esse desempenho é ligeiramente menor do obtido pelo conjunto PR\_0, que possui apenas 87% do total de dados.

O desempenho do modelo usando dados não filtrados (PR\_3) caiu em cerca de 6%, de acordo com os valores de  $R^2$  e Kappa, o que poderia levar à conclusão equivocada, que a filtragem de pixels nublados do produto MOD13Q1 para a utilização em modelos classificatórios de inundação não é necessária.

Entretanto, analisando a acurácia da classificação apenas dos dados marcados como nublado (Tabela 23), verificou-se que essa medida caiu para 67,1%.

O baixo desempenho do modelo usando esse conjunto de dados também pode ser percebido pelo aumento nas taxas de falso positivo e falso negativo, principalmente para os pixels nublados.

Por outro lado, o desempenho do modelo usando os dados reconstruídos pelo filtro Savitzky-Golay obteve resultado similar ao do modelo com dados PR\_0, com a vantagem de possuir a série completa; sem as lacunas causadas pelas nuvens.

Embora o TIMESAT reconstrua toda a série durante seu processamento, quando analisados apenas os pixels preenchidos, ou seja, apenas os pixels que na série original estavam marcados como nublados, verificou-se uma pequena redução na acurácia do modelo, de 88,6% (série completa de dados) para 85,9% (série reduzida). Esse resultado, quando comparado com o resultado do modelo com o uso de pixels nublados (PR\_3), mostra uma melhoria de quase 20% no desempenho do modelo, justificando, assim, o uso dessa técnica de preenchimento de falhas para obtenção da série completa, ao invés de se utilizar a série original sem filtragem de pixels nublados.

Essas falhas na série original, causadas pelas nuvens, ocorrem principalmente no período de enchente e cheia do Pantanal Norte, quando ocorrem importantes processos ecológicos, como, por exemplo, a popularmente conhecida "Decoada", em que ocorre a diminuição de oxigênio dissolvido, causada pela decomposição da matéria orgânica submersa, levando à mortandade de peixes nos rios e lagoas (CALHEIROS; HAMILTON, 1998).

#### 4.4 A INFLUÊNCIA DO CICLO HIDROLÓGICO

Analisando as Figuras 28, 29, 30 e 31, verificou-se que as taxas de falsos positivos (TFP) e falsos negativos (TFN), tanto na série original, com qualidade PR\_0, quanto na série reconstruída pelo software TIMESAT, ocorrem, principalmente, no período de transição (enchente ou vazante), indicado pela porcentagem de réguas inundadas.

Para os anos de 2005 e 2006, na região do transecto do SESC, e para o período 2007/2008, na região do SALD Pirizal, as taxas de falsos negativos surgem no início do período de medida, primeiro e em maior quantidade para a série do TIMESAT, seguido da série original em menor quantidade. Parece haver um atraso entre as medidas observadas (réguas inundadas) e a predição feita pelo modelo (réguas não inundadas). Os falsos negativos para o período 2006/2007 ocorreram durante a vazante. Assim, as estimativas feitas pelo modelo indicavam que algumas réguas já estavam secas, embora as medidas observadas indicassem o contrário.

Comparando a inundação predita pelo modelo com a medida de inundação obtida pelas réguas do transecto do SESC, na vazante de 2005, observou-se as maiores TFP: 60% para os dados reconstruídos e 50% para a série original. Parece, novamente, haver um atraso entre os valores preditos e os observados. Surgem, ainda, falsos positivos para a série original em 2005 e 2006. Para esse último ano, também ocorrem falsos positivos, em menor quantidade, para os dados reconstruídos.

TFP e TFN ocorreram em menor quantidade para a região do SALD Pirizal (períodos 2006/2007 e 2007/2008) quando comparadas com os resultados do transecto do SESC (anos de 2005 e 2006), possivelmente relacionadas com o uso e ocupação da terra, como será discutida mais adiante (seção 4.5).

Vale ressaltar que nem todas as réguas do SALD Pirizal indicaram inundação, pois algumas delas estão instaladas em cordilheiras, que raramente inundam. Por esse motivo, nas Figuras 30 e 31, a porcentagem de réguas inundadas não atinge o valor de 100%.

O ciclo hidrológico, de outubro de 2006 a setembro de 2007, é apresentado através de mapas de inundação dados pelas Figuras 32 e 33. Como o produto MOD13Q1 é gerado a cada 16 dias, em um ano tem-se um total de 23 imagens.

No início do ciclo, dia 273 de 2006, o mapa de inundação está quase todo incompleto pela cobertura de nuvens. Somente a partir do dia 49 a área de estudo fica sem essa interferência. Dos 23 mapas gerados para esse ciclo, apenas 13 estão sem interferência das nuvens. Portanto, o uso de preenchimento de falhas, nesse estudo realizado pelo software TIMESAT, se torna uma técnica indispensável para o estudo do ciclo completo. Analisando os mapas gerados pelo modelo usando os dados reconstruídos, verificou-se uma grande similaridade com os gerados pelo modelo usando apenas qualidade PR\_0, com a vantagem de possuir o ciclo completo.

Entretanto, em uma inspeção mais detalhada (ver Figura 34), verificou-se que no período de seca os mapas gerados pelo modelo com dados reconstruídos apresentam uma superestimação da inundação em algumas áreas, como por exemplo, em regiões de florestas inundáveis (Cambarazal), quando comparados com mapas de mesma data gerados pelo modelo com dados de qualidade PR\_0.

Embora a área de estudo tenha baixa declividade (FANTIN-CRUZ et al., 2010), ainda assim não é plana o suficiente para garantir que uma medida em uma régua possa representar toda a área coberta pelo pixel de 250 m do MODIS. Em outras palavras, a área pode estar parcialmente inundada (ORDOYNE; FRIEDL, 2008). Assim, durante o período de enchente e vazante, as áreas parcialmente inundadas ocasionam o maior número de erros de classificação.

# 4.5 DESEMPENHO DO MODELO EM FUNÇÃO DO USO E OCUPAÇÃO DA TERRA

Como exposto, a maior quantidade de erros de classificação (falsos positivos e falsos negativos) ocorre no período de transição (enchente ou vazante), seguido do período em que a área de estudo está inundada (Figura 35).

No período de transição, as maiores taxas de falso positivo ocorreram para a formação de Florestas inundáveis, seguida do Cerrado e, por último, da Pastagem. As taxas de falsos negativos também ocorrem em maior número nesse período, com a maior contribuição dada pela formação do tipo Cerrado, seguido pelas Florestas inundáveis e de Pastagem.

No período inundado, as maiores taxas de falsos negativos ocorreram para as Florestas inundáveis, depois para o Cerrado, e, por último, para a Pastagem.

Os erros de classificação são menores no período seco, com menos de 5% de taxas de falsos positivos para as Florestas inundáveis e para a Pastagem. Não houve nenhum erro de classificação para o Cerrado, para esse período.

As Florestas inundáveis, como, por exemplo, o Cambarazal (formação monodominante de *Vochysia divergens*), foram as maiores responsáveis por erros de classificação do tipo falso positivo, como pode ser observado pelas Figuras 32 e 33. Uma explicação para isso é que o dossel fechado (ou parcialmente fechado) das florestas impede o imageamento do solo. Assim, o que se tem é uma medida do teor de água presente nas folhas das árvores.

Erros de classificação em áreas de florestas inundáveis são encontrados não apenas em trabalhos envolvendo o uso de sensores ópticos (passivos), como, por exemplo, o sensor MODIS (SAKAMOTO et al., 2007; ORDOYNE; FRIEDL, 2008; PADOVANI, 2010), mas também em sensores ativos de Radar utilizando a banda C (HESS et al., 2003; FUJII, 2004).

A Floresta Inundável, representada pelo círculo na Figura 36, permaneceu inundada até o dia 113 de 2007 (23 de abril). Entretanto, analisando a medida dessa régua para a mesma data do pixel, verificou-se que não havia inundação, portanto, um falso positivo.

As regiões de Pastagem Exótica e Campo Cerrado, representados pelo quadrado e triângulo, respectivamente, atingiram o período da seca no dia 097 (sete de abril). Apesar da data do produto MODIS indicar o dia 097, os pixels que compõem a imagem podem ter datas de aquisição que vão do dia 097 (sete de abril) ao dia 112 (22 de abril). Assim, verificando a data de aquisição do pixel correspondente à régua aqui representando o Campo Cerrado, obteve a data 109 (19 de abril). Para esse dia, a régua apresentava um centímetro de inundação. Portanto, o valor predito (nesse caso, zero) é um falso negativo.

### 4.6 VARIABILIDADE ANUAL DA INUNDAÇÃO MEDIDA E PREDITA

A probabilidade de inundação das réguas do transecto do SESC e do SALD Pirizal, representada pelos *boxplots* nas Figuras 38, 39, 40 e 41, é comparada com a porcentagem de réguas inundadas (linha vermelha), nas mesmas Figuras.

Para o transecto SESC em 2005 (pixels com qualidade PR\_0), quando todas as réguas estão inundadas, os *boxplots* apresentam sua mediana acima de 0,5 de probabilidade de inundação. Entretanto, para o dia 049/2005, o quartil inferior está abaixo de 0,5, indicando que, embora todas as medidas observadas indiquem inundação, alguns valores preditos têm probabilidade de inundação menor do que 0,5 (falso negativo). À medida que a região do transecto vai se esvaziando, indicado pela diminuição da porcentagem de réguas inundadas (linha vermelha), as probabilidades de inundação também diminuem. Entretanto, as distribuições da inundação predita dos dias 145, 161 e 177 estão todas abaixo de 0,5, indicando alguns casos de falso negativo.

Verificou-se ainda, para os dados de qualidade PR\_0 do transecto SESC 2005, uma variação na largura das caixas e no valor da mediana. Para o mesmo ano,

os dados reconstruídos (Figura 39) apresentam uma variação mais comportada, possivelmente devido à suavização imposta pelo filtro Savitzky-Golay.

Para o transecto SESC em 2006 (pixels com qualidade PR\_0), verificou-se também, que a inundação predita acompanha a observada, com alguns falsos negativos no início da inundação e no esvaziamento. O último dia de medida (225/2006) apresenta um valor predito acima de 0,5, embora todas as réguas indiquem zero (falso positivo). Novamente, os dados reconstruídos apresentam uma variação mais comportada.

Para o SALD Pirizal, períodos 2006/2007 e 2007/2008, a probabilidade de inundação acompanha a inundação observada, com algumas variações nas predições para o início da inundação no período 2007/2008. No período de seca, os *boxplots* se estreitam, com a mediana próxima de zero. Esse comportamento é observado tanto para os dados de qualidade PR\_0 (Figura 38) quanto para os dados do TIMESAT (Figura 39), com esse último apresentando menor variação na distribuição da inundação predita, principalmente para o início das medidas do período 2007/2008.

Para a validação do modelo, transecto SESC 2007 e SALD Pirizal 2008/2009, verificou-se que a inundação predita não tem um comportamento semelhante à inundação medida, tanto para os dados de qualidade PR\_0 (Figura 40) quanto para os dados reconstruídos (Figura 41). Para o transecto do SESC verificou-se uma grande quantidade de falsos positivos e falsos negativos. Para o SALD Pirizal, os *boxplots* dos dados reconstruídos indicam variância inferior, comparados com os de qualidade PR\_0.

# 5 CONCLUSÕES

O modelo de classificação da inundação usando regressão logística com os índices LSWI e EVI, mais a classificação do uso e ocupação da terra, como variáveis preditivas, foi ajustado e validado para duas áreas no Pantanal Norte. A acurácia da classificação obtida no ajuste do modelo (validação interna) foi de 89,2%, enquanto na validação externa foi de 77,2%.

A acurácia na validação cruzada k-fold, com k = 10, foi de 88,6%, próxima ao da validação interna, mostrando que apenas na validação externa o grau de generalização do modelo pode ser estimado com maior confiança.

Foi introduzido o conceito de réguas virtuais com o objetivo de extrapolar as medidas pontuais e, assim, obter maior representatividade da área ocupada pelo pixel do MODIS. Embora não tenha obtido melhor resultado geral quando comparado com as medidas pontuais, em alguns pontos de monitoramento da inundação o número de falsos positivos diminuiu.

A qualidade dos pixels do produto MODIS até o nível QA\_0110 não influenciou significativamente a acurácia na detectabilidade da inundação, devendo filtrar apenas os pixel nublados (*Pixel Reliability* igual a três).

Uma vez que a filtragem de pixels nublados torna a série incompleta, técnicas de preenchimento de falhas se justificam sempre que o ano hidrológico completo for importante para estudos ecológicos.

Assim, nesse estudo foi utilizado o software TIMESAT com filtragem adaptativa Savitzky-Golay para obtenção das séries temporais completas das imagens MODIS, com preenchimento de falhas e suavização de dados espúrios. O modelo de classificação da inundação usando dados reconstruídos por esse software obteve desempenho similar quando comparado com o modelo usando dados da série original.

O uso e ocupação da terra mostrou-se um fator determinante na acurácia do modelo, levando a classificações erradas do tipo falso positivo e falso negativo. A formação de Floresta, como, por exemplo, o Cambarazal da RPPN do SESC Pantanal, foi a que mais apresentou erros de classificação, inclusive no período de seca.

O período hidrológico também influenciou o desempenho do modelo. Os erros de classificação ocorrem mais frequentemente durante os períodos de transição (enchente e vazante), possivelmente devido à irregularidade da superfície do terreno compreendida dentro de cada pixel, associada à imprecisão dos valores de inundação obtidos pela interpolação/extrapolação das medidas obtidas através das réguas.
#### REFERÊNCIAS

ALMEIDA, T. I. R.; PENATTI, N. C.; FERREIRA, L. G.; ARANTES, A. E.; AMARAL, C. H. do. Principal component analysis applied to a time series of MODIS images: the spatio-temporal variability of the Pantanal wetland, Brazil. **Wetlands Ecology and Management**, Springer, p. 1–12, 2015. Citado na página 72.

ALVARES, C. A.; STAPE, J. L.; SENTELHAS, P. C.; MORAES, G. de; LEONARDO, J.; SPAROVEK, G. Köppen's climate classification map for Brazil. **Meteorologische Zeitschrift**, v. 22, n. 6, p. 711–728, 2013. Citado na página 28.

ARIAS, M. E.; COCHRANE, T. A.; PIMAN, T.; KUMMU, M.; CARUSO, B. S.; KILLEEN, T. J. Quantifying changes in flooding and habitats in the Tonle Sap Lake (Cambodia) caused by water infrastructure development and climate change in the Mekong Basin. Journal of Environmental Management, Elsevier, v. 112, p. 53–66, 2012. Citado na página 3.

ARIEIRA, J.; NUNES DA CUNHA, C. Fitossociologia de uma floresta inundável monodominante de *Vochysia divergens* Pohl (Vochysiaceae), no Pantanal Norte, MT, Brasil. Acta Botanica Brasilica, SciELO Brasil, v. 20, n. 3, p. 569–580, 2006. Citado 2 vezes nas páginas 31 e 72.

ARNESEN, A. S.; SILVA, T. S. F.; HESS, L. L.; NOVO, E. M. L. M.; RUDORFF, C. M.; CHAPMAN, B. D.; MCDONALD, K. C. Monitoring flood extent in the lower Amazon River floodplain using ALOS/PALSAR ScanSAR images. **Remote Sensing of Environment**, Elsevier, v. 130, p. 51–61, 2013. Citado na página 2.

AUYNIRUNDRONKOOL, K.; CHEN, N.; PENG, C.; YANG, C.; GONG, J.; SILAPATHONG, C. Flood detection and mapping of the Thailand Central plain using RADARSAT and MODIS under a sensor web environment. **International Journal of Applied Earth Observation and Geoinformation**, Elsevier, v. 14, n. 1, p. 245–255, 2012. Citado na página 2.

BAKER, C.; LAWRENCE, R. L.; MONTAGNE, C.; PATTEN, D. Change detection of wetland ecosystems using Landsat imagery and change vector analysis. Wetlands, Springer, v. 27, n. 3, p. 610–619, 2007. Citado na página 24.

BOYCE, M. S.; VERNIER, P. R.; NIELSEN, S. E.; SCHMIEGELOW, F. K. A. Evaluating resource selection functions. **Ecological Modelling**, Elsevier, v. 157, n. 2, p. 281–300, 2002. Citado 2 vezes nas páginas 21 e 23.

BWANGOY, J.-R. B.; HANSEN, M. C.; ROY, D. P.; GRANDI, G. D.; JUSTICE, C. O. Wetland mapping in the Congo Basin using optical and radar remotely sensed data and derived topographical indices. **Remote Sensing of Environment**, Elsevier, v. 114, n. 1, p. 73–86, 2010. Citado na página 24.

CALHEIROS, D. F.; HAMILTON, S. K. Limnological conditions associated with natural fish kills in the Pantanal wetland of Brazil. **Verhandlungen des Internationalen Verein Limnologie**, 1998. Citado na página 77.

CHAKRABORTY, M.; DAS, S. Determination of Signal to Noise Ratio of Electrocardiograms Filtered by Band Pass and Savitzky-Golay Filters. **Procedia Technology**, Elsevier, v. 4, p. 830–833, 2012. Citado na página 37.

CHANDRASEKAR, K.; SAI, M. V. R. S.; ROY, P. S.; DWEVEDI, R. S. Land Surface Water Index (LSWI) response to rainfall and NDVI using the MODIS Vegetation Index product. **International Journal of Remote Sensing**, v. 31, n. 15, p. 3987–4005, 2010. Citado na página 7.

CHEN, J.; JÖNSSON, P.; TAMURA, M.; GU, Z.; MATSUSHITA, B.; EKLUNDH, L. A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky-Golay filter. **Remote sensing of Environment**, v. 91, n. 3, p. 332–344, 2004. Citado na página 37.

CHEN, Y.; BARRETT, D.; LIU, R.; GAO, L.; ZHOU, M.; RENZULLO, L.; CUDDY, S.; EMELYANOVA, I. A spatial framework for regional-scale flooding risk assessment. In: AMES, D. P.; QUINN, N. W.; RIZZOLI, A. E. (Ed.). **7th International Congress on Environmental Modelling and Software**. San Diego, USA: iEMSs, 2014. Citado 2 vezes nas páginas 3 e 25.

CHEN, Y.; HUANG, C.; TICEHURST, C.; MERRIN, L.; THEW, P. An evaluation of MODIS daily and 8-day composite products for floodplain and wetland inundation mapping. **Wetlands**, Springer, v. 33, n. 5, p. 823–835, 2013. Citado 3 vezes nas páginas 3, 20 e 25.

CHEN, Y.; WANG, B.; POLLINO, C. A.; CUDDY, S. M.; MERRIN, L. E.; HUANG, C. Estimate of flood inundation and retention on wetlands using remote sensing and GIS. **Ecohydrology**, Wiley Online Library, v. 7, n. 5, p. 1412–1420, 2014. Citado na página 25.

CHEN, Y.; WANG, W.; CUDDY, S.; POLLINO, C.; MERRIN, L. E. Spatial modelling of potential water retention under floodplain inundation using remote sensing and GIS. In: SEPPELT, R.; VOINOV, A. A.; LANGE, S.; BANKAMP, D. (Ed.). 6th International Congress on Environmental Modelling and Software. Leipzig, Germany: iEMSs, 2012. Citado na página 25.

CONGALTON, R. G. A review of assessing the accuracy of classifications of remotely sensed data. **Remote Sensing of Environment**, v. 37, n. 1, p. 35 – 46, 1991. Citado na página 20.

CONGALTON, R. G.; GREEN, K. Assessing the accuracy of remotely sensed data: principles and practices. Boca Raton, Florida: Taylor & Francis Group, 2008. Citado na página 20.

DAMASCENO-JUNIOR, G. A.; SEMIR, J.; SANTOS, F. A. M. D.; LEITÃO-FILHO, H. de F. Structure, distribution of species and inundation in a riparian forest of Rio Paraguai, Pantanal, Brazil. **Flora-Morphology, Distribution, Functional Ecology of Plants**, Elsevier, v. 200, n. 2, p. 119–135, 2005. Citado na página 72.

DARCH, J. P. A study of moisture conditions in the Pantanal of Brazil using satellite imagery. **Remote Sensing of Environment**, Elsevier, v. 8, n. 4, p. 331–348, 1979. Citado na página 2.

DRONOVA, I.; GONG, P.; WANG, L.; ZHONG, L. Mapping dynamic cover types in a large seasonally flooded wetland using extended principal component analysis and object-based classification. **Remote Sensing of Environment**, Elsevier, v. 158, p. 193–206, 2015. Citado na página 20.

EKLUNDH, L.; JÖNSSON, P. **Timesat 3.1 Software Manual**. Sweden, 2011. Citado 2 vezes nas páginas 23 e 24.

ESRI. ArcGIS Desktop: Release 10. 2013. Redlands, CA: Environmental Systems Research Institute. Citado na página 40.

EVANS, T.; COSTA, M.; TOMAS, W.; CAMILO, A. Large-scale habitat mapping of the Brazilian Pantanal wetland: A synthetic aperture radar approach. **Remote Sensing of Environment**, Elsevier, v. 155, p. 89–108, 2014. Citado na página 3.

EVANS, T. L.; COSTA, M.; TELMER, K.; SILVA, T. S. F. Using ALOS/PALSAR and RADARSAT-2 to map land cover and seasonal inundation in the Brazilian Pantanal. Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of, IEEE, v. 3, n. 4, p. 560–575, 2010. Citado na página 2.

FANTIN-CRUZ, I. **Dinâmica da inundação em mesoescala na planície de inundação do rio Cuiabá**. Dissertação (Mestrado) — Universidade Federal de Mato Grosso, 2008. Citado 3 vezes nas páginas 31, 39 e 74.

FANTIN-CRUZ, I.; GIRARD, P.; ZEILHOFER, P.; COLLISCHONN, W.; CUNHA, C. N. da. Unidades fitofisionômicas em mesoescala no Pantanal Norte e suas relações com a geomorfologia. **Biota Neotropica**, v. 10, n. 2, p. 31–38, 2010. Citado 3 vezes nas páginas 29, 31 e 79.

FANTIN-CRUZ, I.; PEDROLLO, O.; CASTRO, N. M.; GIRARD, P.; ZEILHOFER, P.; HAMILTON, S. K. Historical reconstruction of floodplain inundation in the Pantanal (Brazil) using neural networks. Journal of Hydrology, v. 399, n. 3–4, p. 376–384, 2011. Citado 2 vezes nas páginas 29 e 31. FAWCETT, T. An introduction to ROC analysis. **Pattern recognition letters**, Elsevier, v. 27, n. 8, p. 861–874, 2006. Citado na página 21.

FAWCETT, T. ROC graphs with instance-varying costs. **Pattern Recognition** Letters, Elsevier, v. 27, n. 8, p. 882–891, 2006. Citado na página 22.

FENG, L.; HU, C.; CHEN, X.; CAI, X.; TIAN, L.; GAN, W. Assessment of inundation changes of Poyang Lake using MODIS observations between 2000 and 2010. **Remote Sensing of Environment**, Elsevier, v. 121, p. 80–92, 2012. Citado na página 25.

FENSHOLT, R.; SANDHOLT, I. Evaluation of MODIS and NOAA AVHRR vegetation indices with *in situ* measurements in a semi-arid environment. International Journal of Remote Sensing, Taylor & Francis, v. 26, n. 12, p. 2561–2594, 2005. Citado 2 vezes nas páginas 39 e 74.

FOODY, G. M. On training and evaluation of svm for remote sensing applications. In: \_\_\_\_\_. Kernel Methods for Remote Sensing Data Analysis. West Sussex, United Kingdom: John Wiley & Sons, Ltd, 2009. p. 85–109. Citado na página 20.

FOX, J.; WEISBERG, S. An R Companion to Applied Regression. Second. Thousand Oaks CA: Sage, 2011. Disponível em: <a href="http://socserv.socsci.mcmaster">http://socserv.socsci.mcmaster</a>. ca/jfox/Books/Companion>. Citado na página 36.

FRAZIER, P.; PAGE, K.; LOUIS, J.; BRIGGS, S.; ROBERTSON, A. Relating wetland inundation to river flow using Landsat TM data. International Journal of Remote Sensing, Taylor & Francis, v. 24, n. 19, p. 3755–3770, 2003. Citado na página 24.

FRAZIER, P. S.; PAGE, K. J. et al. Water body detection and delineation with Landsat Tm data. **Photogrammetric Engineering and Remote Sensing**, ASPRS AMERICAN SOCIETY FOR PHOTOGRAMMETRY AND, v. 66, n. 12, p. 1461–1468, 2000. Citado na página 24.

FRIEDL, M. A.; MCIVER, D. K.; HODGES, J. C. F.; ZHANG, X. Y.; MUCHONEY, D.; STRAHLER, A. H.; WOODCOCK, C. E.; GOPAL, S.; SCHNEIDER, A.; COOPER, A.; BACCINI, A.; GAO, F.; SCHAAF, C. Global land cover mapping from MODIS: algorithms and early results. **Remote Sensing of Environment**, v. 83, n. 1, p. 287–302, 2002. Citado na página 15.

FRIEDL, M. A.; SULLA-MENASHE, D.; TAN, B.; SCHNEIDER, A.; RAMANKUTTY, N.; SIBLEY, A.; HUANG, X. MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets. **Remote Sensing of Environment**, v. 114, n. 1, p. 168–182, 2010. Citado 4 vezes nas páginas 15, 16, 17 e 72.

FUJII, H. Hydrological survey and water balance of the Cambodian floodplain in the Mekong River. **Technical Report of the National Research Institute of Agricultural Engineering (Japan)**, 2004. Citado na página 80. GALLANT, J. C.; DOWLING, T. I. A multiresolution index of valley bottom flatness for mapping depositional areas. **Water Resources Research**, Wiley Online Library, v. 39, n. 12, 2003. Citado na página 27.

GAO, B.-C. NDWI - A Normalized Difference Water Index for Remote Sensing of Vegetation Liquid Water From Space. **Remote Sensing of Environment**, v. 58, p. 257–266, 1996. Citado 5 vezes nas páginas 6, 7, 25, 27 e 71.

GIANCRISTOFARO, R. A.; SALMASO, L. Model performance analysis and model validation in logistic regression. **Statistica**, v. 63, n. 2, p. 375–396, 2007. Citado 2 vezes nas páginas 22 e 23.

GIRARD, P.; FANTIN-CRUZ, I.; OLIVEIRA de S. M. L.; HAMILTON, S. K. Small-scale spatial variation of inundation dynamics in a floodplain of the Pantanal (Brazil). **Hydrobiologia**, n. 638, p. 223–233, 2010. Citado 2 vezes nas páginas 29 e 31.

GOOVAERTS, P. Geostatistics for natural resources evaluation. New York, Oxford: Oxford Uuniversity Press, 1997. (Applied Geostatistics Series, xiv). Citado na página 53.

GOPAL, B.; JUNK, W. J. Biodiversity in wetlands: an introduction. In: GOPAL, B.; JUNK, W. J.; DAVIS, J. A. (Ed.). **Biodiversity in Wetlands:** Assessment, Function and Conservation. Leiden: Backhuys Publishers, 2000. v. 1, p. 1–10. Citado na página 1.

GOWARD, S. N.; MARKHAM, B.; DYE, D. G.; DULANEY, W.; YANG, J. Normalized Difference Vegetation Index Measurements from the Advanced Very High Resolution Radiomete. **Remote Sensing of Environment**, v. 35, p. 257–277, 1991. Citado na página 10.

GUERSCHMAN, J. P.; WARREN, G.; BYRNE, G.; LYMBURNER, L.; MUELLER, N.; DIJK, A. V. **MODIS-based standing water detection** for flood and large reservoir mapping: algorithm development and applications for the Australian continent. Canberra, Australia: CSIRO, 2011. Citado 2 vezes nas páginas 25 e 27.

GUMBRICHT, T.; WOLSKI, P.; FROST, P.; MCCARTHY, T. Forecasting the spatial extent of the annual flood in the Okavango Delta, Botswana. Journal of Hydrology, v. 290, n. 3, p. 178–191, 2004. Citado na página 2.

HAMILTON, S. K.; SIPPEL, S. J.; MELACK, J. M. Inundation patterns in the Pantanal wetland of South America determined from passive microwave remote sensing. **Archiv für Hydrobiologie**, v. 137, p. 1–23, 1996. Citado 2 vezes nas páginas 2 e 28.

HAMILTON, S. K.; SIPPEL, S. J.; MELACK, J. M. Seasonal inundation patterns in two large savanna floodplains of South America: the Llanos de Moxos (Bolivia) and the Llanos del Orinoco (Venezuela and Colombia). **Hydrological Processes**, v. 18, n. 11, p. 2103–2116, 2004. Citado na página 2. HAMILTON, S. K.; SOUZA, O. C. d.; COUTINHO, M. E. Dynamics of floodplain inundation in the alluvial fan of the Taquari River (Pantanal, Brazil). Verhandlungen International Vereiniging Limnology, v. 26, p. 916–922, 1998. Citado na página 28.

HARRELL, F. E.; LEE, K. L.; MARK, D. B. Tutorial in biostatistics multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. **Statistics in Medicine**, v. 15, n. 4, p. 361–387, 1996. Citado na página 23.

HENGL, T. Finding the right pixel size. Computers & Geosciences, Elsevier, v. 32, n. 9, p. 1283–1298, 2006. Citado na página 40.

HERITAGE, G. L.; MILAN, D. J.; LARGE, A. R.; FULLER, I. C. Influence of survey strategy and interpolation model on DEM quality. **Geomorphology**, Elsevier, v. 112, n. 3, p. 334–344, 2009. Citado 2 vezes nas páginas 39 e 40.

HESS, L. L.; MELACK, J. M.; NOVO, E. M.; BARBOSA, C. C.; GASTIL, M. Dual-season mapping of wetland inundation and vegetation for the central Amazon basin. **Remote Sensing of Environment**, Elsevier, v. 87, n. 4, p. 404–428, 2003. Citado 2 vezes nas páginas 2 e 80.

HIRD, J. N.; MCDERMID, G. J. Noise reduction of NDVI time series: An empirical comparison of selected techniques. **Remote Sensing of Environment**, v. 113, n. 1, p. 248–258, 2009. Citado na página 37.

HOSMER JR, D. W.; LEMESHOW, S.; STURDIVANT, R. X. Applied Logistic Regression. 3rd. ed. New Jersey: John Wiley & Sons, 2013. Citado 6 vezes nas páginas 18, 19, 20, 23, 45 e 71.

HUANG, C.; CHEN, Y.; WU, J. Mapping spatio-temporal flood inundation dynamics at large river basin scale using time-series flow data and MODIS imagery. International Journal of Applied Earth Observation and Geoinformation, Elsevier, v. 26, p. 350–362, 2014. Citado 5 vezes nas páginas 3, 20, 25, 27 e 73.

HUBERT, L. J.; GOLLEDGE, R. G.; COSTANZO, C. M. Generalized procedures for evaluating spatial autocorrelation. **Geographical Analysis**, Wiley Online Library, v. 13, n. 3, p. 224–233, 1981. Citado na página 40.

HUETE, A.; DIDAN, K.; LEEUWEN, W. V. MODIS VEGETATION INDEX (MOD 13) ALGORITHM THEORETICAL BASIS DOCUMENT. Arizona, 1999. Citado na página 9.

HUETE, A.; DIDAN, K.; MIURA, T.; RODRIGUEZ, E.; GAO, X.; FERREIRA, L. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. **Remote Sensing of Environment**, v. 83, n. 1–2, p. 195–213, 2002. The Moderate Resolution Imaging Spectroradiometer (MODIS): a new generation of Land Surface Monitoring. Citado na página 6.

HUETE, A. R.; LIU, H. Q.; BATCHILY, K.; LEEUWEN, W. van. A Comparison of Vegetation Indices over a Global Set of TM Images for EOS-MODIS. **Remote sensing of environment**, v. 59, n. 3, p. 440–451, 1997. Citado na página 6.

JÖNSSON, P.; EKLUNDH, L. Seasonality extraction by function fitting to time-series of satellite sensor data. Geoscience and Remote Sensing, IEEE Transactions on, v. 40, n. 8, p. 1824–1832, Aug 2002. Citado na página 23.

JÖNSSON, P.; EKLUNDH, L. TIMESAT –a program for analyzing time-series of satellite sensor data. **Computers & Geosciences**, v. 30, n. 8, p. 833–845, 2004. Citado na página 23.

JUSTICE, C.; VERMOTE, E.; TOWNSHEND, J. R. G.; DEFRIES, R.; ROY, D.; HALL, D.; SALOMONSON, V.; PRIVETTE, J.; RIGGS, G.; STRAHLER, A.; LUCHT, W.; MYNENI, R.; KNYAZIKHIN, Y.; RUNNING, S.; NEMANI, R.; WAN, Z.; HUETE, A.; LEEUWEN, W. V.; WOLFE, R.; GIGLIO, L.; MULLER, J.-P.; LEWIS, P.; BARNSLEY, M. The Moderate Resolution Imaging Spectroradiometer (MODIS): land remote sensing for global change research. **Geoscience and Remote Sensing, IEEE Transactions on**, v. 36, n. 4, p. 1228–1249, Jul 1998. Citado na página 5.

KHAN, S. I.; HONG, Y.; WANG, J.; YILMAZ, K. K.; GOURLEY, J. J.; ADLER, R. F.; BRAKENRIDGE, G. R.; POLICELL, F.; HABIB, S.; IRWIN, D. Satellite remote sensing and hydrologic modeling for flood inundation mapping in Lake Victoria basin: Implications for hydrologic prediction in ungauged basins. **Geoscience and Remote Sensing, IEEE Transactions on**, IEEE, v. 49, n. 1, p. 85–95, 2011. Citado na página 25.

KOHAVI, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: MELLISH, C. S. (Ed.). **Proceedings of the 14th International Joint Conference on Artificial Intelligence**. San Francisco, CA, USA: Morgan Kaufmann Publishers, 1995. v. 14, n. 2, p. 1137–1145. Citado na página 23.

KUENZER, C.; GUO, H.; HUTH, J.; LEINENKUGEL, P.; LI, X.; DECH, S. Flood mapping and flood dynamics of the Mekong Delta: ENVISAT-ASAR-WSM based time series analyses. **Remote Sensing**, Multidisciplinary Digital Publishing Institute, v. 5, n. 2, p. 687–715, 2013. Citado na página 2.

LANDIM, P. M. B.; STURARO, J. R.; MONTEIRO, R. C. Krigagem ordinária para situações com tendência regionalizada. Rio Claro, SP, 2002. Disponível em: <a href="http://www.rc.unesp.br/igce/aplicada/DIDATICOS/LANDIM/tkrigagem.pdf">http://www.rc.unesp.br/igce/aplicada/DIDATICOS/ LANDIM/tkrigagem.pdf</a>>. Citado na página 40.

LANDMANN, T.; SCHRAMM, M.; HUETTICH, C.; DECH, S. MODIS-based change vector analysis for assessing wetland dynamics in Southern Africa. **Remote Sensing Letters**, Taylor & Francis, v. 4, n. 2, p. 104–113, 2013. Citado na página 3.

LANE, S. The use of digital terrain modelling in the understanding of dynamic river systems. In: LANE, S.; RICHARDS, K. S.; CHANDLER, J. H. (Ed.). Landform Monitoring, Modelling and Analysis. Chichester, UK: John Wiley and Son Ltd., 1998. p. 311–342. Citado na página 39.

LEE, H.; YUAN, T.; JUNG, H. C.; BEIGHLEY, E. Mapping wetland water depths over the central Congo Basin using PALSAR ScanSAR, envisat altimetry, and MODIS VCF data. **Remote Sensing of Environment**, Elsevier, 2014. Citado na página 2.

LILLESAND, T. M.; KIEFER, R. W. **Remote Sensing and Image Interpretation**. 3rd. ed. New York: John Wiley & Sons, Inc., 1994. Citado na página 8.

LOGAN, M. Biostatistical Design and Analysis Using R: A Practical Guide. [S.l.]: John Wiley & Sons, 2010. Citado 3 vezes nas páginas 19, 45 e 50.

LP DAAC. **MOD13Q1**. Sioux Falls, South Dakota: USGS/Earth Resources Observation and Science (EROS) Center, 2012. Citado na página 32.

MACALISTER, C.; MAHAXAY, M. Mapping wetlands in the Lower Mekong Basin for wetland resource and conservation management using Landsat ETM images and field survey data. **Journal of Environmental Management**, Elsevier, v. 90, n. 7, p. 2130–2137, 2009. Citado 2 vezes nas páginas 2 e 25.

MCFEETERS, S. K. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. **International Journal of Remote Sensing**, v. 17, n. 7, p. 1425–1432, 1996. Citado 4 vezes nas páginas 7, 8, 25 e 71.

MELACK, J. M.; HESS, L. L.; GASTIL, M.; FORSBERG, B. R.; HAMILTON, S. K.; LIMA, I. B.; NOVO, E. M. Regionalization of methane emissions in the Amazon Basin with microwave remote sensing. **Global Change Biology**, Wiley Online Library, v. 10, n. 5, p. 530–544, 2004. Citado na página 1.

MERZ, C.; WINKLER, A.; PEKDEGER, A. Trace elements in streambed sediments of floodplains: consequences for water management measures. **Environmental Earth Sciences**, Springer, v. 59, n. 1, p. 25–38, 2009. Citado na página 1.

MINGWEI, Z.; QINGBO, Z.; ZHONGXIN, C.; JIA, L.; YONG, Z.; CHONGFA, C. Crop discrimination in Northern China with double cropping systems using Fourier analysis of time-series MODIS data. International Journal of Applied Earth Observation and Geoinformation, Elsevier, v. 10, n. 4, p. 476–485, 2008. Citado na página 37.

MITSCH, W. J. A comparison of the nonpoint source pollution control function of natural and constructed riparian wetlands. In: MITSCH, W. J. (Ed.). **Global Wetlands: Old World and New**. Amsterdam: Elsevier, 1994. p. 351–161. Citado na página 1. MITSCH, W. J.; GOSSELINK, J. G. The value of wetlands: importance of scale and landscape setting. **Ecological economics**, Elsevier, v. 35, n. 1, p. 25–33, 2000. Citado na página 1.

MORAN, P. A. Notes on continuous stochastic phenomena. Biometrika, JSTOR, p. 17–23, 1950. Citado na página 40.

MUCHONEY, D.; STRAHLER, A.; HODGES, J.; LOCASTRO, J. The IGBP DISCover confidence sites and the system for terrestrial ecosystem parameterization: tools for validating global land-cover data. **Photogrammetric Engineering and Remote Sensing**, American Society for Photogrammetry and Remote Sensing, v. 65, n. 9, p. 1061–1067, 1999. Citado na página 15.

MUELLER, M.; PANDER, J.; GEIST, J. The ecological value of stream restoration measures: an evaluation on ecosystem and target species scales. **Ecological Engineering**, Elsevier, v. 62, p. 129–139, 2014. Citado na página 1.

MYERS, R. H.; MONTGOMERY, D. C.; VINING, G. G.; ROBINSON, T. J. Generalized Linear Models: with Applications in Engineering and the Sciences. 2. ed. New Jersey: John Wiley & Sons, 2010. Citado na página 19.

NAGELKERKE, N. J. A note on a general definition of the coefficient of determination. **Biometrika**, Oxford University Press, v. 78, n. 3, p. 691–692, 1991. Citado na página 22.

NAKAGAWA, S.; SCHIELZETH, H. A general and simple method for obtaining  $R^2$  from generalized linear mixed-effects models. **Methods in Ecology and Evolution**, Wiley Online Library, v. 4, n. 2, p. 133–142, 2013. Citado na página 21.

NASA/USGS. **MODIS Reprojection Tool User's Manual**. 4.1. ed. Sioux Falls, South Dakota, 2011. Citado na página 35.

NOVITSKI, R. P.; SMITH, R. D.; FRETWELL, J. D. Wetland functions, values, and assessment. In: FRETWELL, J. D.; WILLIAMS, J. S.; REDMAN, P. J. (Ed.). National water summary on wetland resources. USGS Water Supply Paper. Washington, D.C.: U.S. Government Printing Office, 1996. v. 2425, p. 79–86. Citado na página 1.

NUNES DA CUNHA, C.; JUNK, W. J. A preliminary classification of habitats of the Pantanal of Mato Grosso and Mato Grosso do Sul, and its relation to national and international wetland classification systems. In: JUNK, W. J.; SILVA, C. J.; NUNES DA CUNHA, C.; WANTZEN, K. M. (Ed.). The Pantanal: Ecology, biodiversity and sustainable management of a large neotropical seasonal wetland. Sofia–Moscow: Pensoft, 2010. p. 127–141. Citado na página 1.

ORDOYNE, C.; FRIEDL, M. A. Using MODIS data to characterize seasonal inundation patterns in the Florida Everglades. **Remote Sensing of Environment**, v. 112, n. 11, p. 4107–4119, 2008. Citado 6 vezes nas páginas 3, 25, 73, 74, 79 e 80.

PADOVANI, C. R. **Dinâmica Espaço-Temporal das Inundações do Pantanal**. 175 p. Tese (Doutorado) — Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, 2010. Citado 6 vezes nas páginas 2, 24, 25, 26, 72 e 80.

PAZ, A. R.; COLLISCHONN, W.; TUCCI, C. E. M.; PADOVANI, C. R. Large-scale modelling of channel flow and floodplain inundation dynamics and its application to the Pantanal (Brazil). **Hydrological processes**, John Wiley & Sons, Ltd., v. 25, n. 9, p. 1498–1516, 2011. Citado na página 2.

R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2014. Disponível em: <a href="http://www.R-project.org/">http://www.R-project.org/</a>. Citado na página 36.

ROGERS, A.; KEARNEY, M. Reducing signature variability in unmixing coastal marsh Thematic Mapper scenes using spectral indices. **International Journal of Remote Sensing**, v. 25, n. 12, p. 2317–2335, 2004. Citado na página 71.

ROYSTON, P.; ALTMAN, D. G. Visualizing and assessing discrimination in the logistic regression model. **Statistics in medicine**, Wiley Online Library, v. 29, n. 24, p. 2508–2520, 2010. Citado na página 36.

SAKAMOTO, T.; NGUYEN, N. V.; KOTERA, A.; OHNO, H.; ISHITSUKA, N.; YOKOZAWA, M. Detecting temporal changes in the extent of annual flooding within the Cambodia and the Vietnamese Mekong Delta from MODIS time-series imagery. **Remote Sensing of Environment**, v. 109, n. 3, p. 295–313, 2007. Citado 5 vezes nas páginas 3, 7, 25, 71 e 80.

SCHAAF, C. B.; GAO, F.; STRAHLER, A. H.; LUCHT, W.; LI, X.; TSANG, T.; STRUGNELL, N. C.; ZHANG, X.; JIN, Y.; MULLER, J.-P.; LEWIS, P.; BARNSLEY, M.; HOBSON, P.; DISNEY, M.; ROBERTS, G.; DUNDERDALE, M.; DOLL, C.; d'ENTREMONT, R. P.; HU, B.; LIANG, S.; PRIVETTE, J. L.; ROY, D. First operational BRDF, albedo nadir reflectance products from MODIS. **Remote sensing of Environment**, Elsevier, v. 83, n. 1, p. 135–148, 2002. Citado na página 15.

SHIMABUKURO, Y. E.; SMITH, A. The least-squares mixing models to generate fraction images derived from remote sensing multispectral data. **Geoscience and Remote Sensing, IEEE Transactions on**, IEEE, v. 29, n. 1, p. 16–20, 1991. Citado na página 26.

SIGNOR, C. A.; FERNANDES, I. M.; PENHA, J. O Pantanal e o Sistema de Pesquisa. In: SIGNOR, C. A.; FERNANDES, I. M.; PENHA, J. (Ed.). **Biodiversidade no Pantanal de Poconé**. Cuiabá: Centro de Pesquisa do Pantanal, 2010. Citado na página 29.

SILVA, J. S. V.; ABDON, M. M. Delimitação do Pantanal brasileiro e suas sub-regiões. **Pesquisa agropecuária brasileira**, v. 33, n. 13, p. 1703–1711, 1998. Citado na página 28.

SILVA, T. S. F.; COSTA, M. P. F.; MELACK, J. M. Spatial and temporal variability of macrophyte cover and productivity in the eastern Amazon floodplain: A remote sensing approach. **Remote Sensing of Environment**, Elsevier, v. 114, n. 9, p. 1998–2010, 2010. Citado na página 2.

SMITH, L. C. Satellite remote sensing of river inundation area, stage, and discharge: A review. **Hydrological processes**, v. 11, n. 10, p. 1427–1439, 1997. Citado na página 24.

SOLANO, R.; DIDAN, K.; JACOBSON, A.; HUETE, A. MODIS Vegetation Index User's Guide (MOD13 Series): Version 2.00, May 2010 (Collection 5). Arizona, 2010. Disponível em: <a href="http://vip.arizona.edu/documents/MODIS/MODIS\_VI\_UsersGuide\_01\_2012.pdf">http://vip.arizona.edu/ documents/MODIS/MODIS\_VI\_UsersGuide\_01\_2012.pdf</a>>. Citado 4 vezes nas páginas 11, 14, 57 e 75.

TANA, G.; LETU, H.; CHENG, Z.; TATEISHI, R. Wetlands Mapping in North America by Decision Rule Classification Using MODIS and Ancillary Data. Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of, IEEE, v. 6, n. 6, p. 2391–2401, 2013. Citado na página 20.

TICEHURST, C.; GUERSCHMAN, J. P.; CHEN, Y. The Strengths and Limitations in Using the Daily MODIS Open Water Likelihood Algorithm for Identifying Flood Events. **Remote Sensing**, Multidisciplinary Digital Publishing Institute, v. 6, n. 12, p. 11791–11809, 2014. Citado na página 25.

TOCKNER, K.; STANFORD, J. A. Riverine flood plains: present state and future trends. **Environmental conservation**, Cambridge Univ Press, v. 29, n. 03, p. 308–330, 2002. Citado na página 1.

TOWNSEND, P. A.; FOSTER, J. R. A synthetic aperture radar-based model to assess historical changes in lowland floodplain hydroperiod. Water Resources **Research**, Wiley Online Library, v. 38, n. 7, p. 20–1, 2002. Citado na página 2.

VAN LEEUWEN, W. J. D.; HUETE, A. R.; LAING, T. W. MODIS vegetation index compositing approach: A prototype with AVHRR data. **Remote Sensing** of Environment, v. 69, n. 3, p. 264–280, 1999. Citado 2 vezes nas páginas 9 e 10.

VENABLES, W. N.; RIPLEY, B. D. Modern Applied Statistics with S. Fourth. New York: Springer, 2002. ISBN 0-387-95457-0. Disponível em: <a href="http://www.stats.ox.ac.uk/pub/MASS4">http://www.stats.ox.ac.uk/pub/MASS4</a>. Citado na página 36.

VERMOTE, E. F.; SALEOUS, N. Z. E.; JUSTICE, C. O. Atmospheric correction of MODIS data in the visible to middle infrared: first results. **Remote Sensing of Environment**, v. 83, n. 1, p. 97–111, 2002. Citado na página 5.

VERMOTE, E. F.; VERMEULEN, A. Atmospheric correction algorithm: spectral reflectances (MOD09). Algorithm Technical Background Document. Maryland, 1999. Citado 2 vezes nas páginas 5 e 14. VIERA, A. J.; GARRETT, J. M. Understanding Interobserver Agreement: The Kappa Statistic. **Fam Med**, v. 37, n. 5, p. 360–363, 2005. Citado 2 vezes nas páginas 59 e 73.

WAN, Z.; ZHANG, Y.; ZHANG, Q.; LI, Z.-l. Validation of the land-surface temperature products retrieved from Terra Moderate Resolution Imaging Spectroradiometer data. **Remote Sensing of Environment**, v. 83, n. 1, p. 163–180, 2002. Citado na página 15.

WANG, H.; CHEN, J.; WU, Z.; LIN, H. Rice heading date retrieval based on multi-temporal MODIS data and polynomial fitting. **International Journal of Remote Sensing**, Taylor & Francis, v. 33, n. 6, p. 1905–1916, 2012. Citado na página 37.

XIAO, X.; BOLES, S.; FROLKING, S.; SALAS, W.; MOORE, B.; LI, C.; HE, L.; ZHAO, R. Observation of flooding and rice transplanting of paddy rice fields at the site to landscape scales in China using VEGETATION sensor data. **International Journal of Remote Sensing**, v. 23, n. 15, p. 3009–3022, 2002. Citado 2 vezes nas páginas 7 e 71.

XIAO, X.; BOLES, S.; FROLKING, S.; LI, C.; BABU, J. Y.; SALAS, W.; III, B. M. Mapping paddy rice agriculture in South and Southeast Asia using multi-temporal MODIS images. **Remote Sensing of Environment**, v. 100, n. 1, p. 95–113, 2006. Citado na página 7.

XIAO, X.; BOLES, S.; LIU, J.; ZHUANG, D.; FROLKING, S.; LI, C.; SALAS, W.; III, B. M. Mapping paddy rice agriculture in southern China using multi-temporal MODIS images. **Remote Sensing of Environment**, v. 95, n. 4, p. 480–492, 2005. Citado 2 vezes nas páginas 7 e 71.

XU, H. Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. **International Journal of Remote Sensing**, v. 27, n. 14, p. 3025–3033, 2006. Citado 3 vezes nas páginas 8, 25 e 27.

ZAR, J. H. **Biostatistical Analysis**. 4. ed. Upper Saddle River, NJ: Prentice Hall, 1999. Citado na página 40.

ZEILHOFER, P. Soil mapping in the Pantanal of Mato Grosso, Brazil, using multitemporal Landsat TM data. Wetlands Ecology and Management, Springer, v. 14, n. 5, p. 445–461, 2006. Citado na página 72.

ZEILHOFER, P.; SCHESSL, M. Relationship between vegetation and environmental conditions in the northern Pantanal of Mato Grosso, Brazil. **Journal of Biogeography**, Wiley Online Library, v. 27, n. 1, p. 159–168, 2000. Citado 2 vezes nas páginas 28 e 72.

ZHU, L.; CHEN, Y.; GONG, H.; JIANG, W.; ZHAO, W.; XIAO, Y. Economic value evaluation of wetland service in Yeyahu Wetland Nature Reserve, Beijing. Chinese Geographical Science, Springer, v. 21, n. 6, p. 744–752, 2011. Citado na página 25.

Apêndices

# APÊNDICE A – PROCESSAMENTO DAS MEDIDAS DE CAMPO E PRÉ-PROCESSAMENTO DOS DADOS MODIS



Figura 42 – Fluxograma apresentando as etapas de pré-processamento das imagens MODIS e processamento das medidas de inundação

# APÊNDICE B – SELEÇÃO DAS VARIÁVEIS E AJUSTE DO MODELO COM O USO DE IMAGENS MODIS DE QUALIDADE PR\_0



Figura 43 – Fluxograma apresentando a etapa de seleção de variáveis e ajuste do modelo usando dados MODIS com qualidade  $PR_0$ 

#### APÊNDICE C – **PREENCHIMENTO DE FALHAS E AJUSTE DO MODELO COM DADOS PREENCHIDOS**



Figura 44 – Fluxograma apresentando a etapa de processamento dos dados MO-DIS com o uso do software TIMESAT para preenchimento de falhas e suavização. Ajuste do modelo e medida de desempenho do modelo

### APÊNDICE D – CRIAÇÃO DOS CONJUNTOS DE NÍVEIS DE QUALIDADE QA\_0000 A QA\_1100



Figura 45 – Fluxograma apresentando a etapa de criação dos conjuntos com os níveis de qualidade QA\_0000 a QA\_1100 . Ajuste do modelo para cada conjunto de qualidade, com medida de desempenho do modelo para cada conjunto