MINISTÉRIO DA CIÊNCIA, TECNOLOGIA, INOVAÇÕES E COMUNICAÇÕES
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

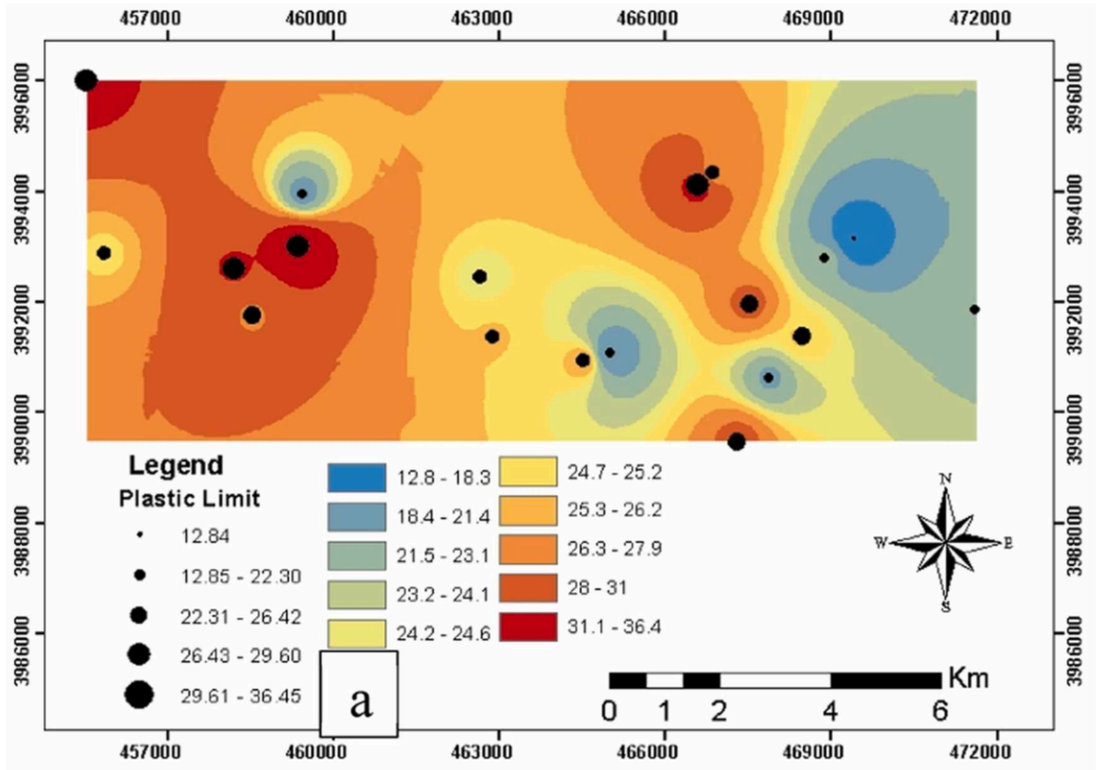
EXERCÍCIO PRÁTICO

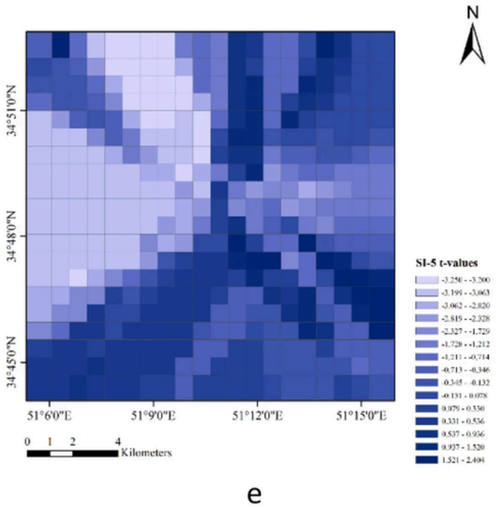
- Isabel Chuizaca
- Ítalo Mira
- Jocilene Barros

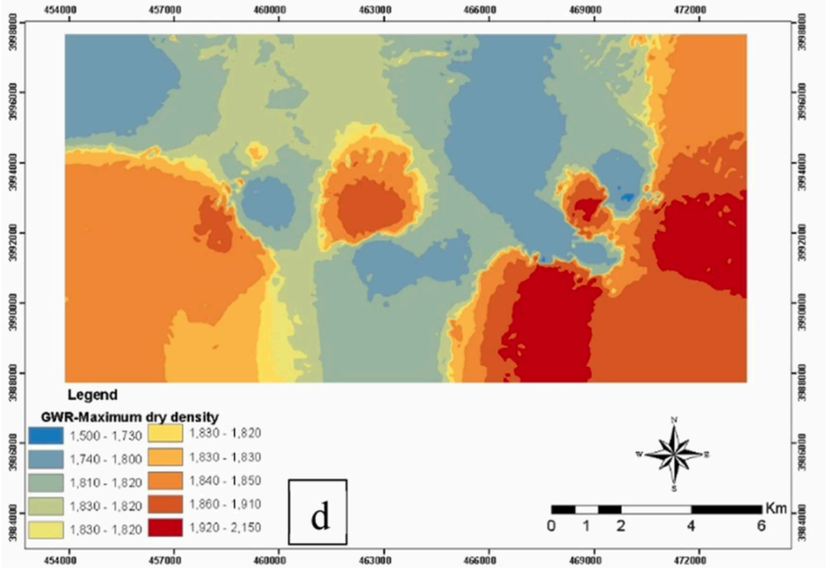
Análise Espacial de Dados Geográficos - SER301

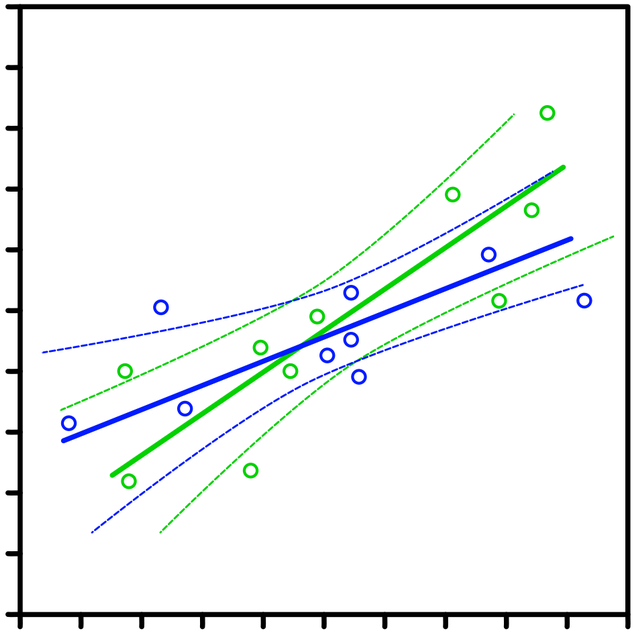
Dr. Miguel Monteiro

São José dos Campos
2024









REGRESSÃO

ESPACIAL

1

Banco de dados e variáveis

FONTE DOS DADOS

Análise Espacial Integrada da Degradação da Terra no Submédio São Francisco, Brasil

Integrated Spatial Analysis of Land Degradation in Submedium São Francisco Basin, Brazil

Ana Catarina das Neves Alves
INPE
https://orcid.org/0000-0002-5395-7151

Jocilene Dantas Barros
Instituto Nacional de Pesquisas Espaciais - INPE
https://orcid.org/0000-0001-7869-7424

Gustavo Felipe Balué Arcoverde
Instituto Nacional de Pesquisas Espaciais - INPE
https://orcid.org/0000-0002-0082-5710

Jean Pierre Henry Balbaud Ometto
Instituto Nacional de Pesquisas Espaciais - INPE
https://orcid.org/0000-0002-4221-1039

DOI: https://doi.org/10.21680/2447-3359.2023v9n2ID33877

PDF | PDF (English)

Publicado
16-12-2023

Como Citar

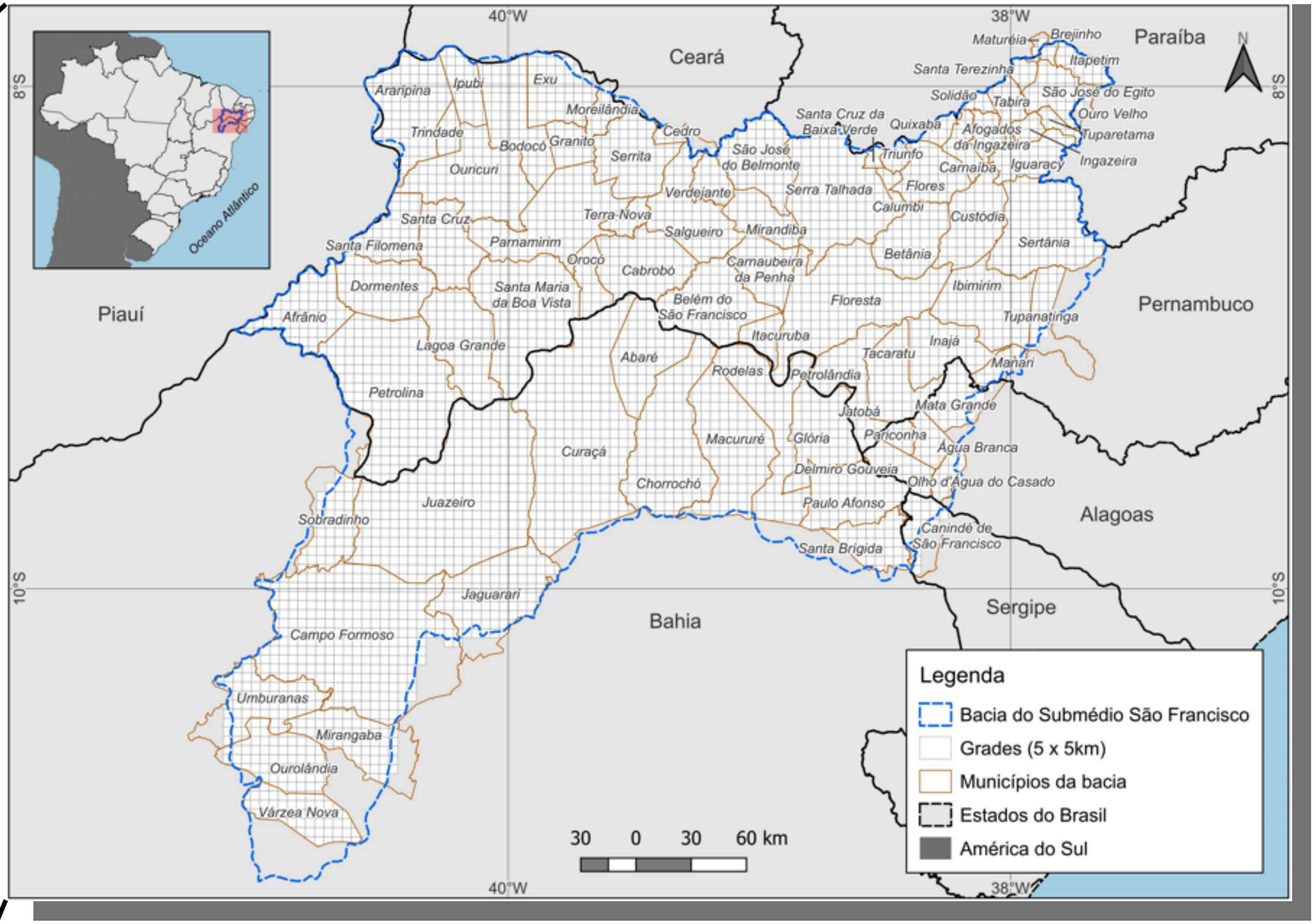
DAS NEVES ALVES, A. C.; BARROS, J. D.; ARCOVERDE, G. F. B.; OMETTO, J. P. H. B. Análise Espacial Integrada da Degradação da Terra no Submédio São Francisco, Brasil: Integrated Spatial Analysis of Land Degradation in Submedium São Francisco Basin, Brazil. **Revista de Geociências do Nordeste**, *[S. l.]*, v. 9, n. 2, p. 183–202, 2023. DOI: 10.21680/2447-3359.2023v9n2ID33877. Disponível em: https://periodicos.ufrn.br/revistadoregne/article/view/33877. Acesso em: 17 nov. 2024.

Fomatos de Citação

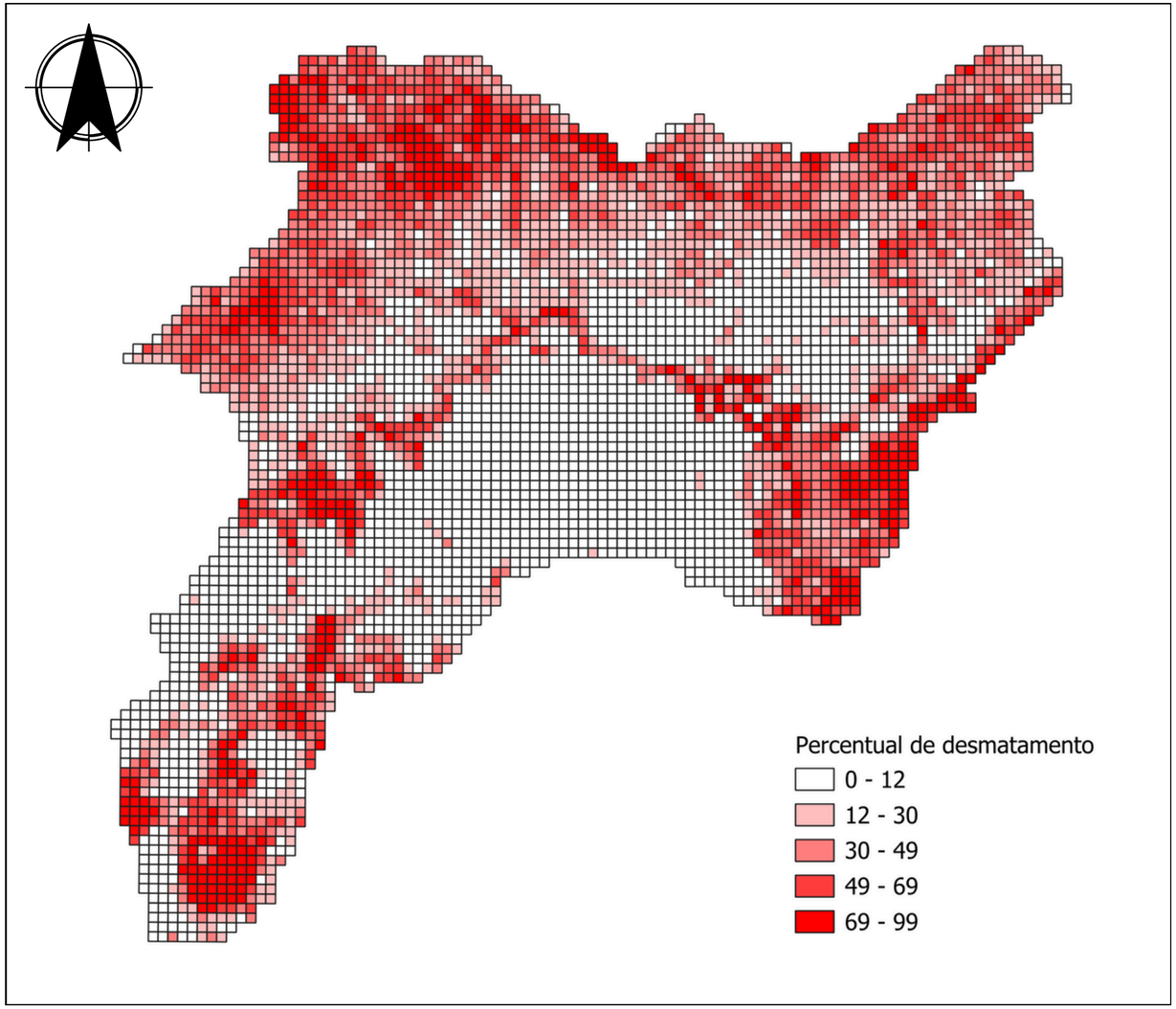
Edição
v. 9 n. 2 (2023)

Resumo

O estudo analisou a degradação da terra na bacia do Submédio São Francisco por meio de índices segundo o modelo FPEIR (Força-Pressão-Estado-Impacto-Resposta). Para seleção dos 17 indicadores foi feita revisão bibliográfica e análise de correlação, os indicadores foram gerados em grades de 5 x 5 km no SIG QGIS através de dados secundários e primários. Foi feito um tratamento para exclusão de outliers, seguido de normalização dos indicadores. Foram gerados índices para cada componente FPEIR por meio da mediana dos indicadores e elaborados mapas dos indicadores e índices. As áreas mais vulneráveis à degradação (maiores valores nos índices FPEI) localizam-se principalmente nas regiões noroeste, centro e sudeste da bacia. Em relação à resposta (maiores valores no índice R), destacam-se as áreas a oeste, noroeste e sudeste da bacia. Os indicadores que mais influenciaram os índices FPEI foram Índice de Desenvolvimento Humano Municipal, estabelecimentos com produtor ocupante, desmatamento, erodibilidade, índice de aridez e pastagens degradadas. E no caso do índice de Resposta, a orientação técnica e programas de crédito. As políticas públicas e iniciativas particulares de restauração da qualidade ambiental são fundamentais para reverter o processo de degradação, contribuindo para o manejo sustentável da Caatinga.



SELEÇÃO DE VARIÁVEIS



6 variáveis das 18
do artigo de referência

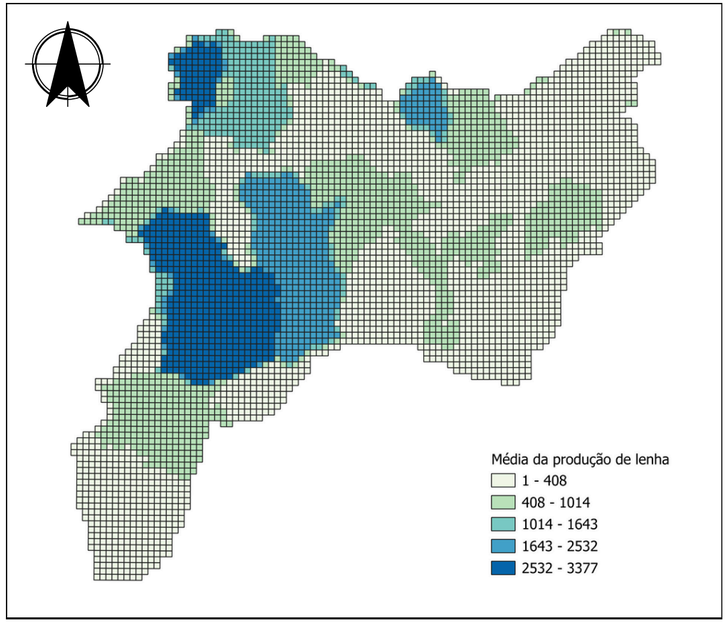
Variável dependente:

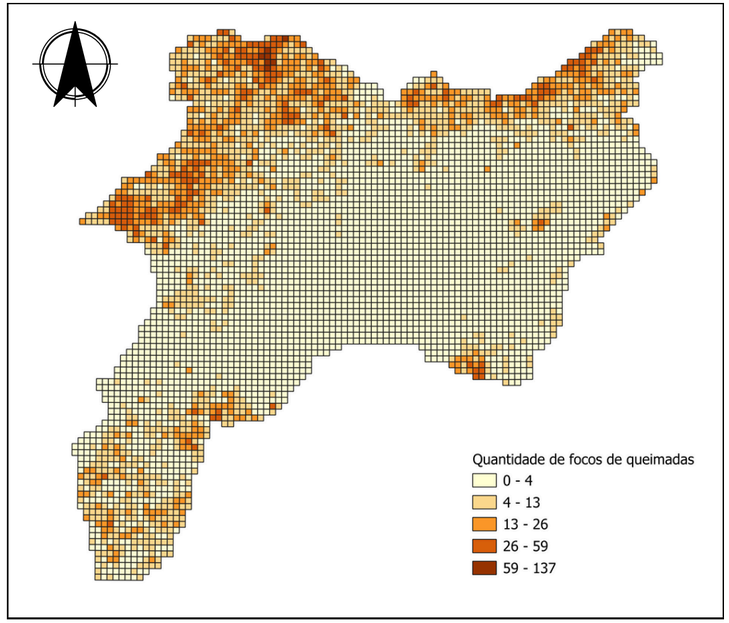
- Desmatamento acumulado (%)

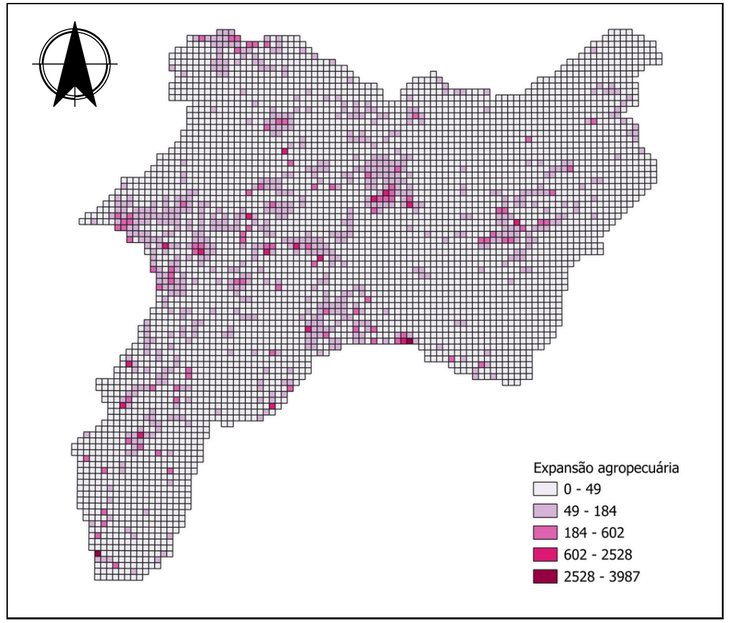
Variáveis independentes:

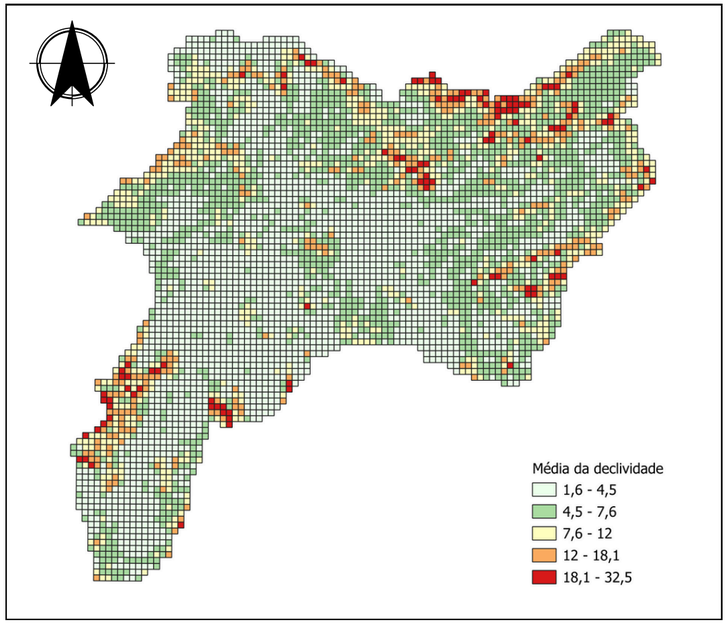
- Média da quantidade produzida de lenha (m^3)
- Quantidade de focos de calor (acumulada)
- Expansão agropecuária (%)
- Média da declividade
- Percentual de Unidades de Conservação (PI e US)

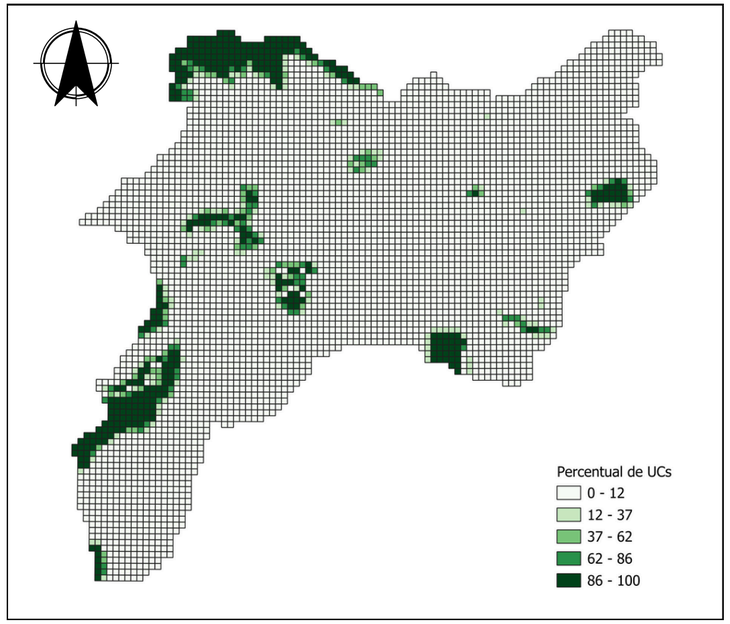
DISTRIBUIÇÃO ESPACIAL DE VARIÁVEIS











2

Resultados Modelo de Regressão

Resíduos

São as diferenças entre os valores observados e os previstos no modelo

Indicam quanto que o modelo se aproxima dos resultados.

Min	1Q	Mediana	3Q	Max
-116,25	-18,143	-6,232	13,880	90,020
Limite Inferior	Dispersão	Concentração dos Valores	Dispersão	Limite Superior

Próximo de Zero - Modelo "Ajustado"

Longe do zero - Alta variabilidade

Presença de Outliers e Assimetria dos dados

Coeficientes

Fator	Estimativa	Std. Error	t-value	Pr(>|t|)
Intercepto	25,9301582	0,7535035	34,413	< 2e-16 ***
M_Lenha_M3	-0,0010958	0,0003482**	-3,147	0,001660 **
Foco_acumu	1,2256932	0,0379433	32,303	< 2e-16 ***
ExpAgr_per	-0,0291742	0,0030255**	-9,643	< 2e-16 ***
Declv_mean	-0,3063125	0,0841901	-3,638	0,000277 ***
Perc_UCs	-0,0877836	0,0123295	-7,120	1,25e-12 ***

Residual standard error: 23,1 com 4659 Graus de liberdade

R^2: 0,1958 | **R^2 Ajustado:** 0,195

F-statistic: 226,9 | **p-value:** <2,2e-16

Modelo é **significativo**, porém **não explica** toda **a variação** dos dados devido **R^2 baixo**

Possui **erros relativamente altos**, mas com **coef. significativos**

3

Moran I e LISA

Dependência espacial - Moran I

Global Moran I for regression residuals

data:
model: lm(formula = Perc_Desm ~ M_Lenha_M3 + Foco_acumu + ExpAgr_per + Declv_mean + Perc_UCs, data = dados, na.action = na.exclude)

weights: rwm

Moran I statistic standard deviate = **85.84**, p-value < **2.2e-16**
alternative hypothesis: two.sided
sample estimates:

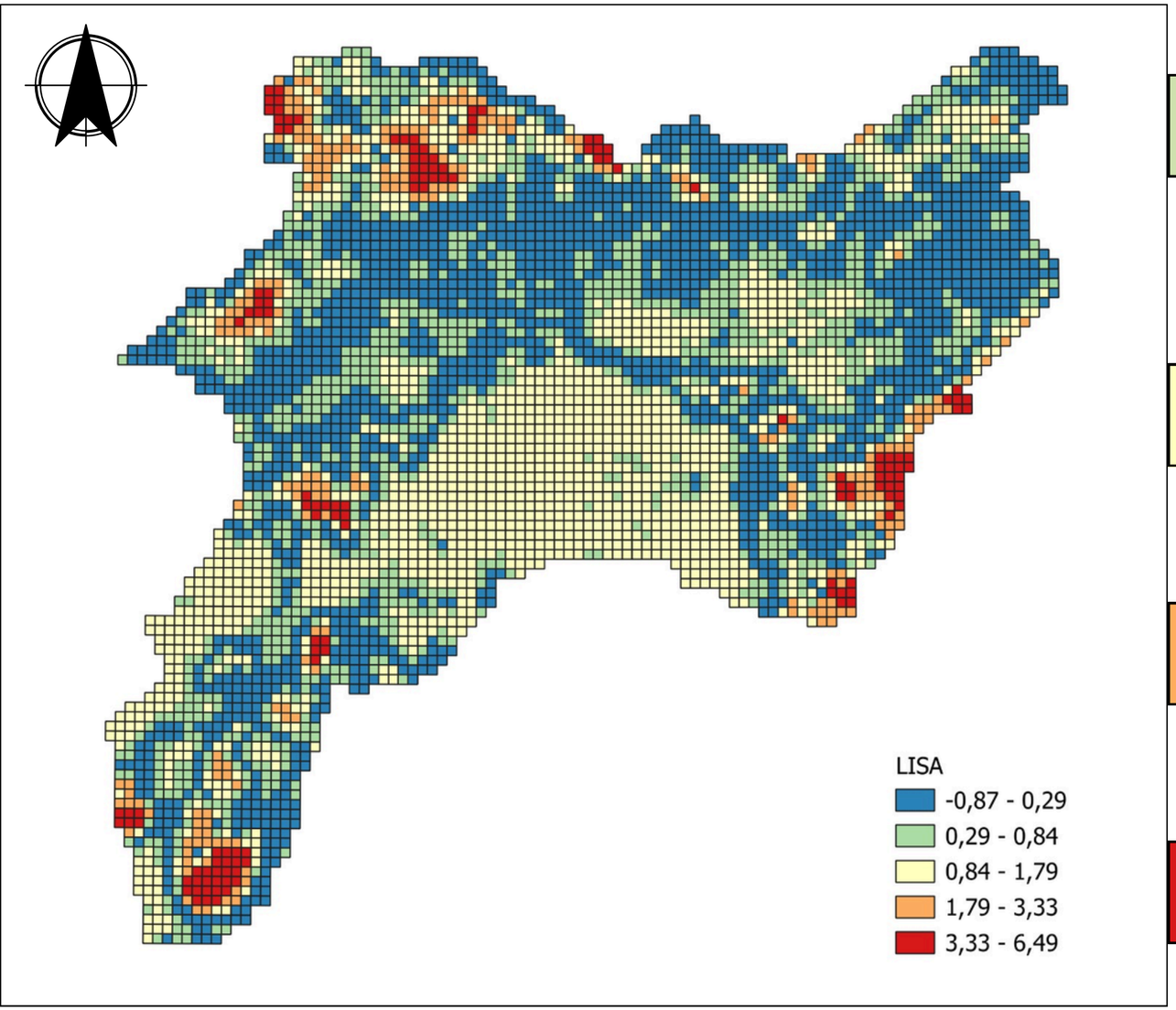
Observed Moran I	Expectation	Variance
6.405108e-01	-8.890516e-04	5.583152e-05

- **Autocorrelação espacial positiva** significativa nos resíduos do modelo
- O **Índice de Moran** observado é **extremamente diferente** do esperado

- Os **resíduos** do modelo **não são espacialmente independentes**

- Fatores espaciais não capturados pelas variáveis independentes podem estar influenciando a variável dependente

LISA



-0.87 - 0.29

Autocorrelação negativa ou positiva muito próxima de zero:
Não mostram um padrão espacial significativo (outliers espaciais negativos)

0.29 - 0.84

Autocorrelação positiva muito baixa:
Agrupamento com menos força, (Zonas de transição entre regiões mais homogêneas ou menos influenciadas por vizinhos)

0.84 - 1.79

Autocorrelação positiva moderada:
Agrupamentos significativos de valores semelhantes (regiões que apresentam características compartilhadas)

1.79 - 3.33

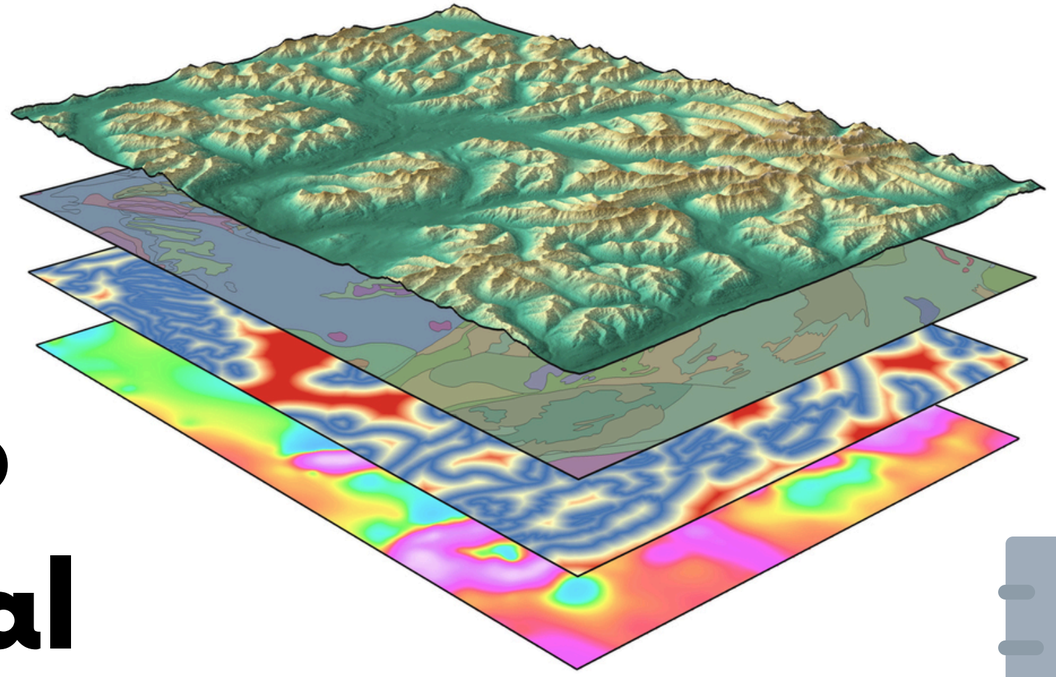
Autocorrelação positiva alta:
Agrupamentos espaciais significativos
A concentração de valores semelhantes é mais evidente aqui

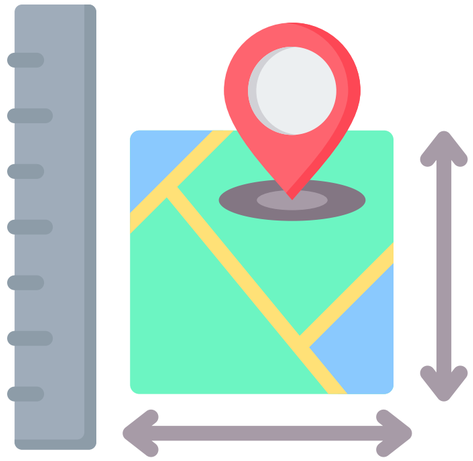
1.79 - 3.33

Autocorrelação positiva muito alta:
Hotspots mais críticos (onde os valores do desmatamento são altamente agrupados)

4

Identificando o Modelo Espacial





Autocorrelação Espacial significativa:

Adoção de modelos de **Regressão Espacial**

Spatial Lag Model
X
Spatial Error Model

Rao's score (Lagrange Multiplier)

- **RSerr:** teste simples para verificar dependência do erro/resíduo;
- **RSlag:** teste simples para verificar omissão de uma variável dependente defasada espacialmente (O valor da variável dependente em um local sendo influenciado pelos valores dela em locais vizinhos)
- **adjRSerr:** teste para dependência do erro/resíduo ajustado para a possibilidade de omissão de uma variável dependente defasada espacialmente
- **adjRSlag:** teste para verificar omissão de uma variável dependente defasada espacialmente, ajustado para a possibilidade de dependência espacial nos resíduos

Autocorrelação Espacial significativa:

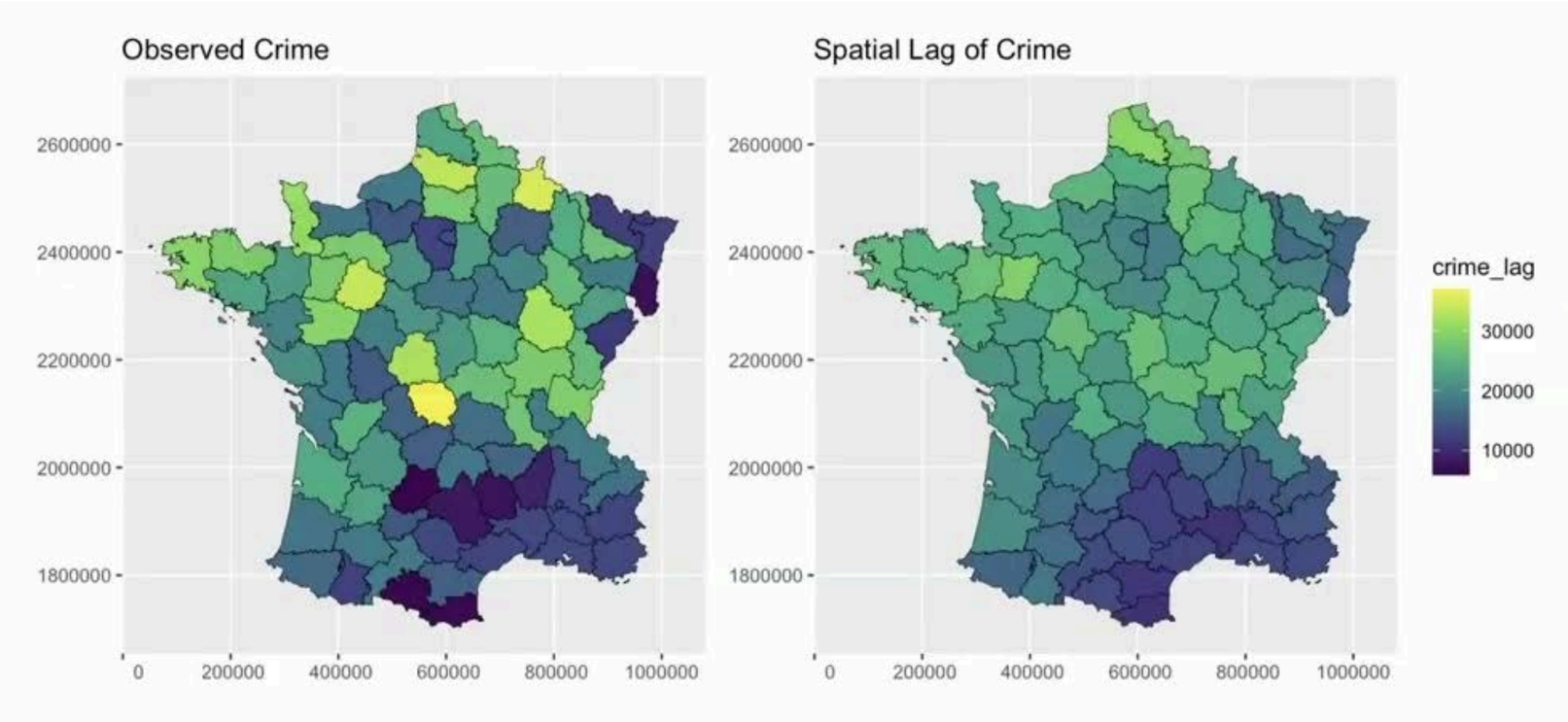
Adoção de modelos de **Regressão Espacial**

Teste	Valor	df	p-value
RSerr	7305.1	1	2.2e-16
RSlag	7242.3	1	2.2e-16
adjRSerr	356.04	1	2.2e-16
adjRSlag	293.26	1	2.2e-16

Todos apresentaram significância estatística, há dependência espacial em todos os casos e os dois modelos são adequados

5

Modelo Spatial Lag



Capturar a influência da variável dependente as localidades vizinhas

Min	1Q	Mediana	3Q	Max
-53,4226	-7,0585	-1,8801	6,2871	60,2398

Fator	Estimativa	Std. Error	Z-value	Pr(>|t|)
Intercepto	5,36442699	0,48445421	11,0731	< 2,2e-16
M_Lenha_M3	-0,00047739	0,00019201	-2,4863	0,0129089
Foco_acumu	0,27058162	0,02287393	11,8293	< 2,2e-16
ExpAgr_per	-0,00854675	0,00166531	-5,1322	2,863e-07
Declv_mean	-0,42705373	0,04714909	-9,0575	< 2,2e-16
Perc_UCs	-0,02569379	0,00683566	-3,7588	0,0001707

Rho: 0,88118 Coeficiente de autocorrelação espacial --> Forte dependência espacial

LR test value: 4793,2

p-value: <2,22e-16 Modelo SAR é melhor que o modelo sem autocorrelação espacial

Asymptotic standard error: 0,0079036 Variabilidade esperada --> Quanto menor, mais preciso

Z-value: 111,49 Desvios em relação a média

Wald statistic: 12430 Significância do parâmetro do modelo --> Altamente significativo

Log likelihood: -18866,34 Adequação do modelo --> Quanto maior, melhor ajuste

ML residual variance: 161,22 Variância residual do modelo --> Quanto menor, melhor ajuste

Number of observations: 4665

Number os parameters estimated: 8

AIC: 37749 Qualidade do modelo comparado a outro

LM test for residual autocorrelation

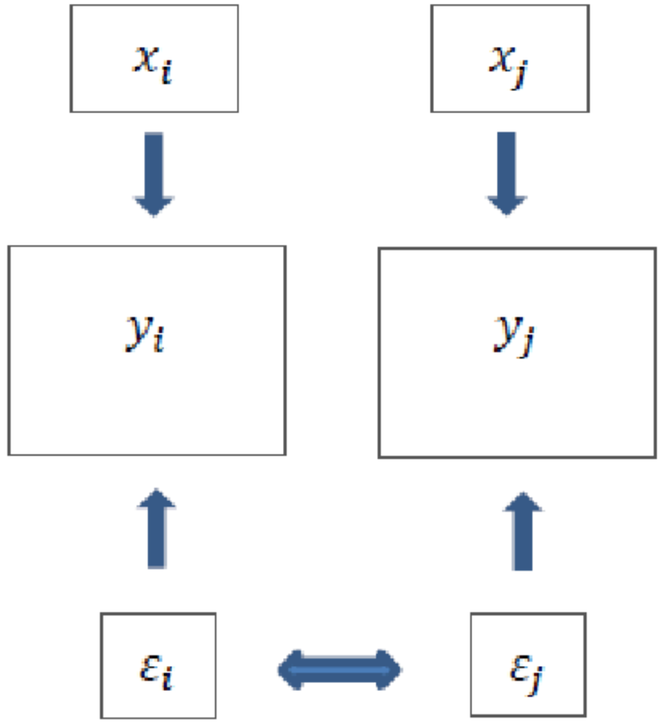
test value: 6,5705

p-value: 0,010368 Autocorrelação espacial dos resíduos --> Há correlação

- O **modelo SAR** revela que a dependência espacial é uma característica **importante** do **desmatamento**
- Isso deve-se ao fato de que as regiões e as variáveis tem uma **alta correlação espacial**

6

Modelo Spatial Error



Call: errorsarlm(formula = Perc_Desm ~ M_Lenha_M3 + Foco_acumu + ExpAgr_per + Declv_mean + Perc_UCs, data = dados, listw = rwm, zero.policy = TRUE)

Residuals:

Min	1Q	Median	3Q	Max
-52.7375	-6.8896	-1.9638	6.6029	59.7838

Type: error
Coefficients:
(asymptotic standard errors)

	Estimate	Std. Error	z value	Pr(>|z|)
Intercept	34.62579786	2.16128101	16.0210	< 2.2e-16
M_Lenha_M3	-0.00126796	0.00077633	-1.6333	0.102411
Foco_acumu	0.33688736	0.03632466	9.2743	< 2.2e-16
ExpAgr_per	-0.00566491	0.00173096	-3.2727	0.001065
Declv_mean	-0.93211853	0.07157553	-13.0229	< 2.2e-16
Perc_UCs	-0.08447262	0.01678414	-5.0329	4.832e-07

Lambda: **0.90816**,

Forte dependência espacial nos resíduos do modelo

LR test value: **4853.8**, p-value: **< 2.22e-16**

p-valor extremamente baixo indica que é importante considerar a dependência espacial nos resíduos

Asymptotic standard error: 0.0074031

z-value: 122.67, p-value: < 2.22e-16

Wald statistic: 15049, p-value: < 2.22e-16

Log likelihood: -18836.07 for error model

Indica uma melhoria na qualidade do ajuste em comparação com o modelo tradicional

ML residual variance (sigma squared): 156.29, (sigma: 12.502)

Em média, os valores previstos pelo modelo têm um desvio de aproximadamente 12,5% na porcentagem de desmatamento em relação aos valores observados

Number of observations: 4665

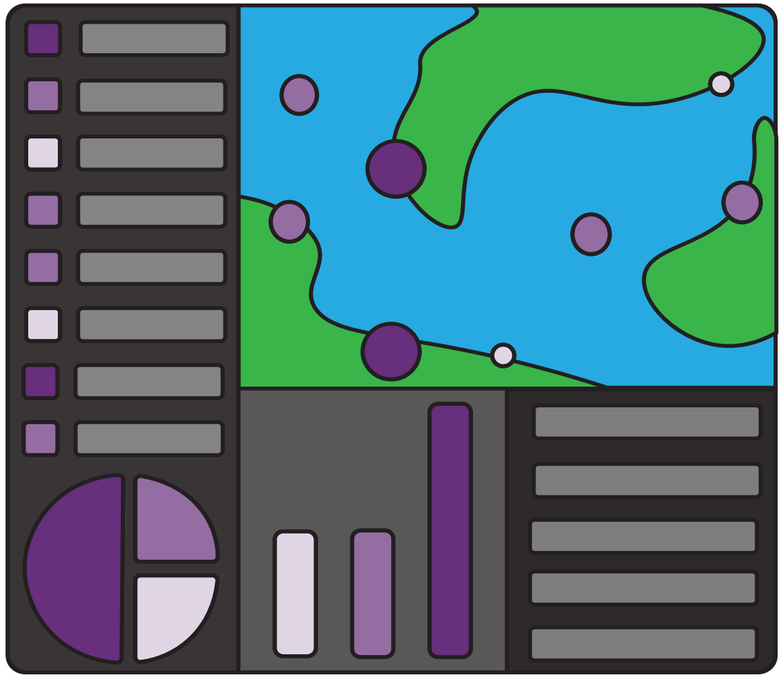
Number of parameters estimated: 8

AIC: 37688, (AIC for lm: 42540)

Significativamente menor do que o do modelo linear tradicional (42,540)

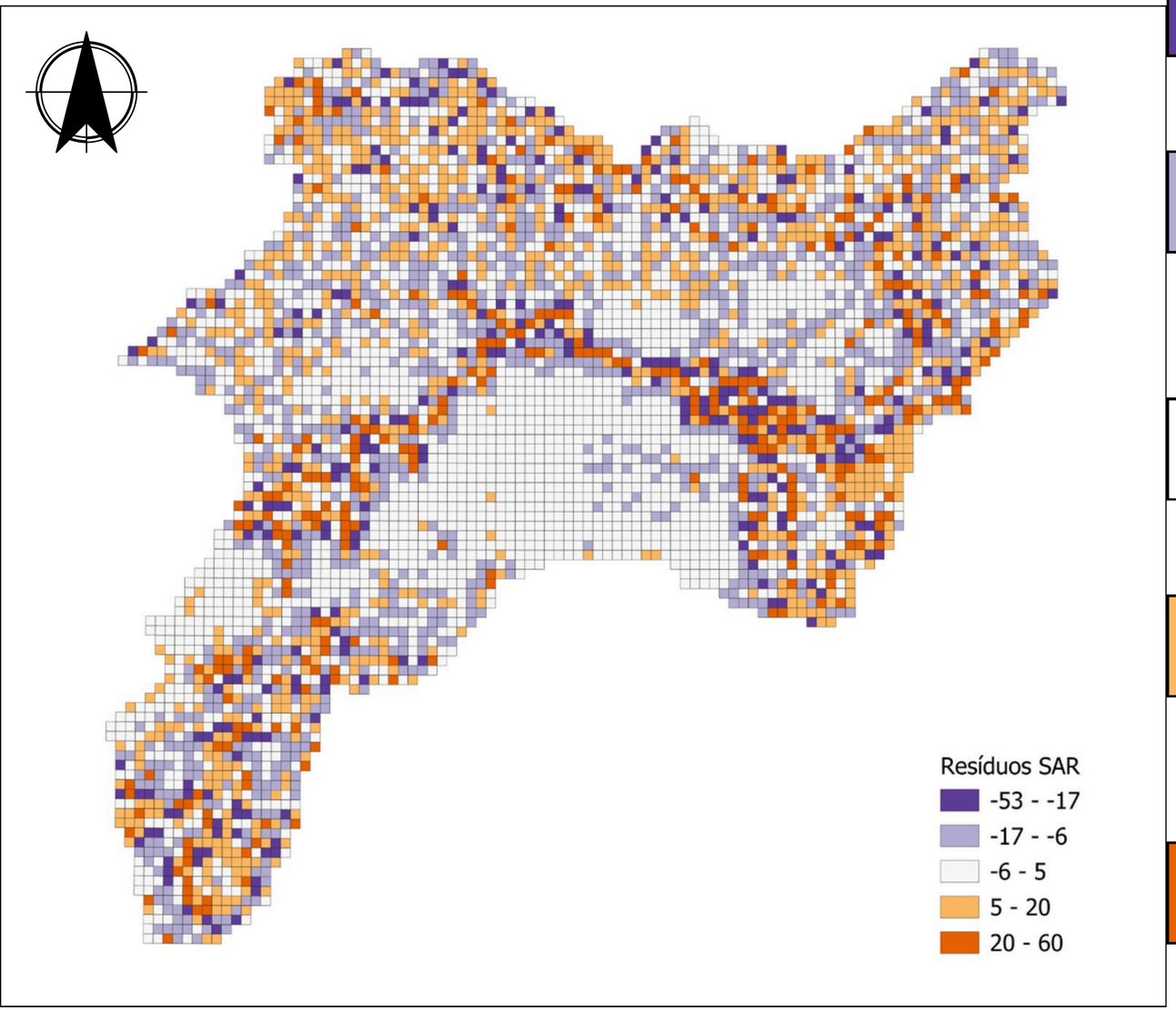
7

Mapas SAR e CAR



$y = mx + b$

Modelo SAR



-53 - -17

RESÍDUOS MAIS NEGATIVOS:
O modelo **subestimou significativamente** os valores observados nessas áreas

-17 - -6

RESÍDUOS NEGATIVOS:
Áreas onde o modelo **subestimou** os valores observados, mas de uma forma mais moderada

-6 - 5

RESÍDUOS PRÓXIMOS A ZERO:
Áreas em que as previsões estão muito próximas dos valores reais

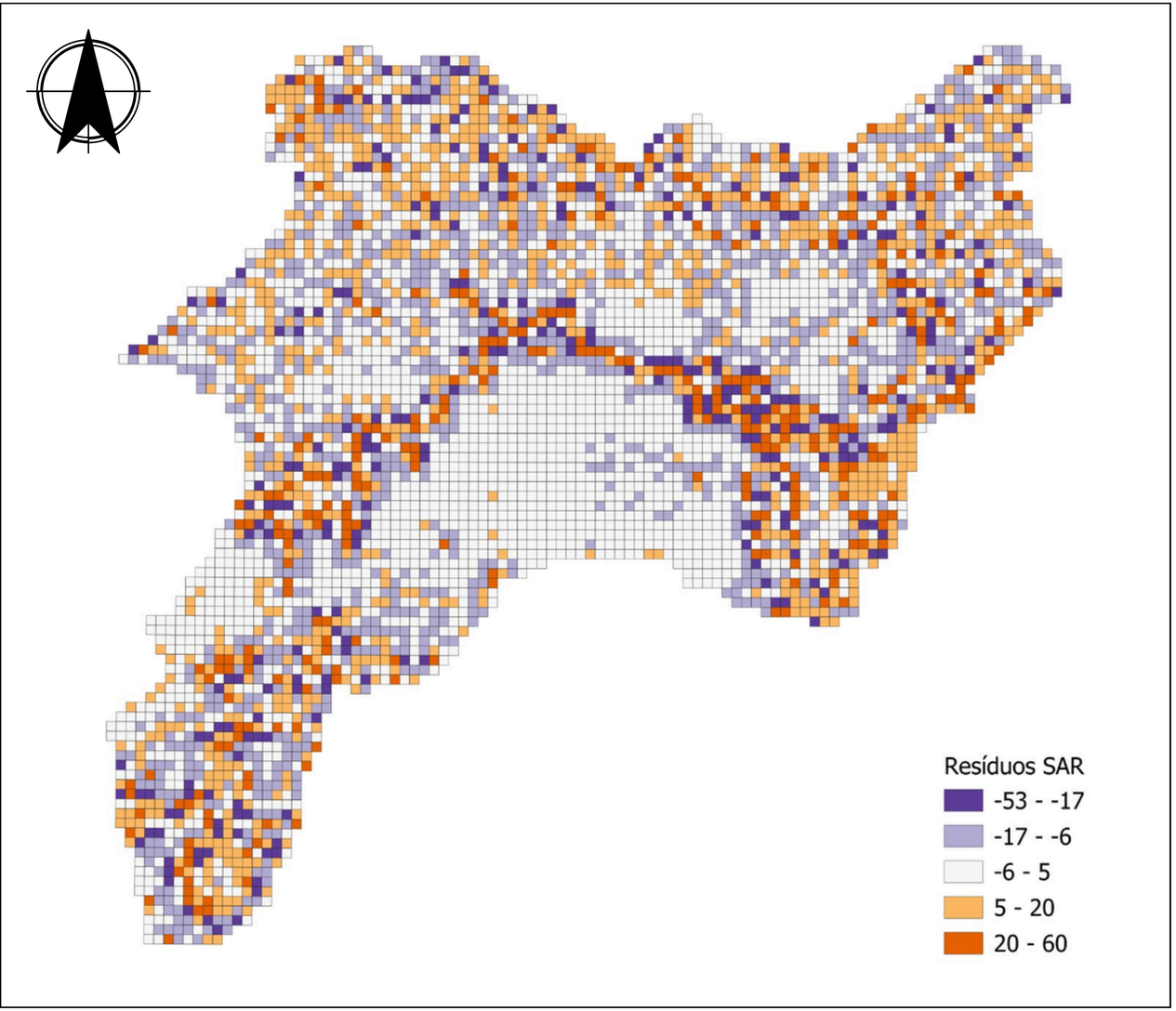
5 - 20

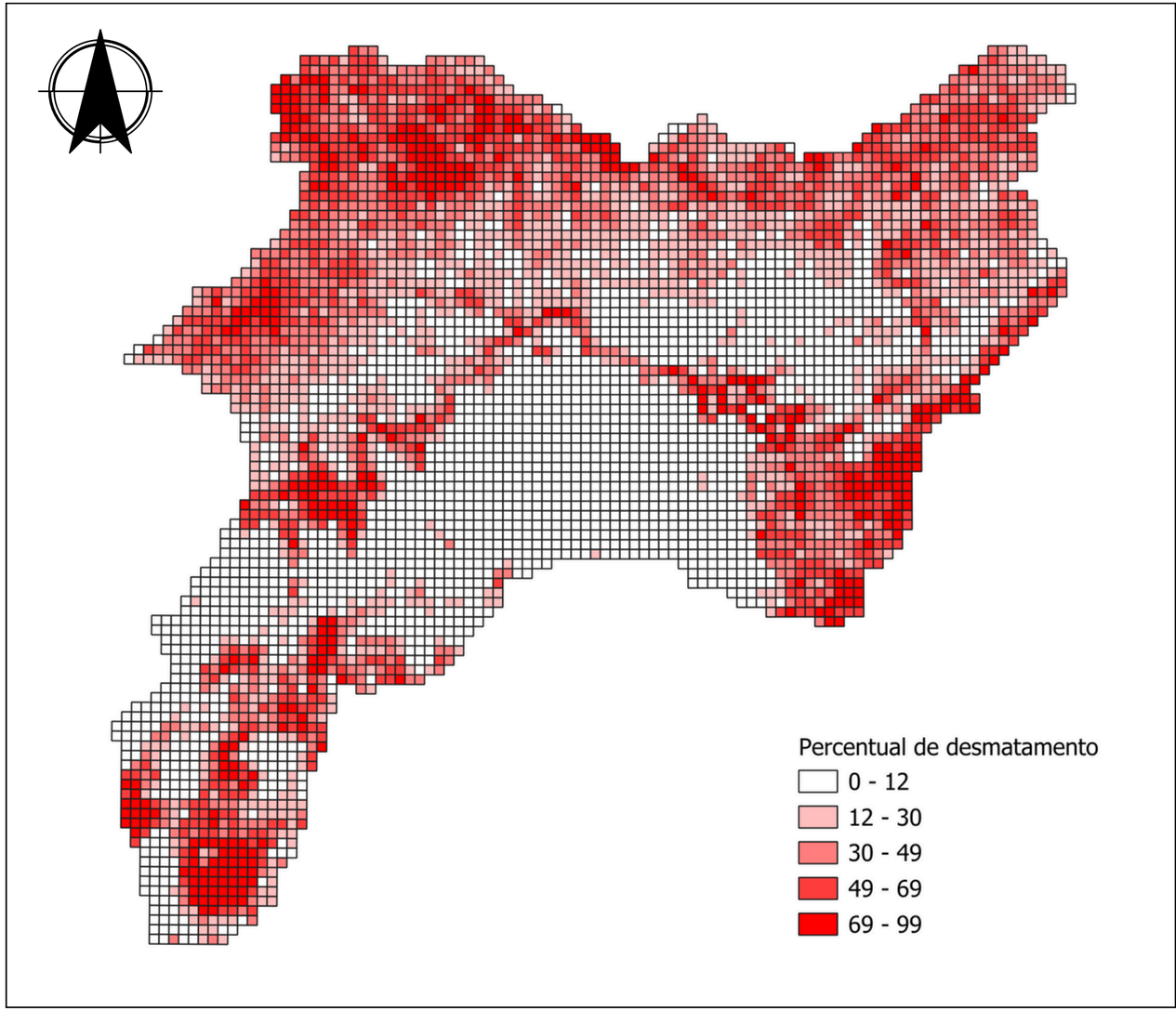
RESÍDUOS POSITIVOS:
O modelo **superestimou** os valores observados nessas áreas, mas de forma moderada

20 - 60

RESÍDUOS MAIS POSITIVOS:
O modelo **superestimou significativamente** os valores observados nessas áreas

Modelo SAR


Resíduos SAR
-53 - -17
-17 - -6
-6 - 5
5 - 20
20 - 60


Percentual de desmatamento
0 - 12
12 - 30
30 - 49
49 - 69
69 - 99

Modelo SAR

Variável dependente:

- Desmatamento acumulado (%) → **Y**

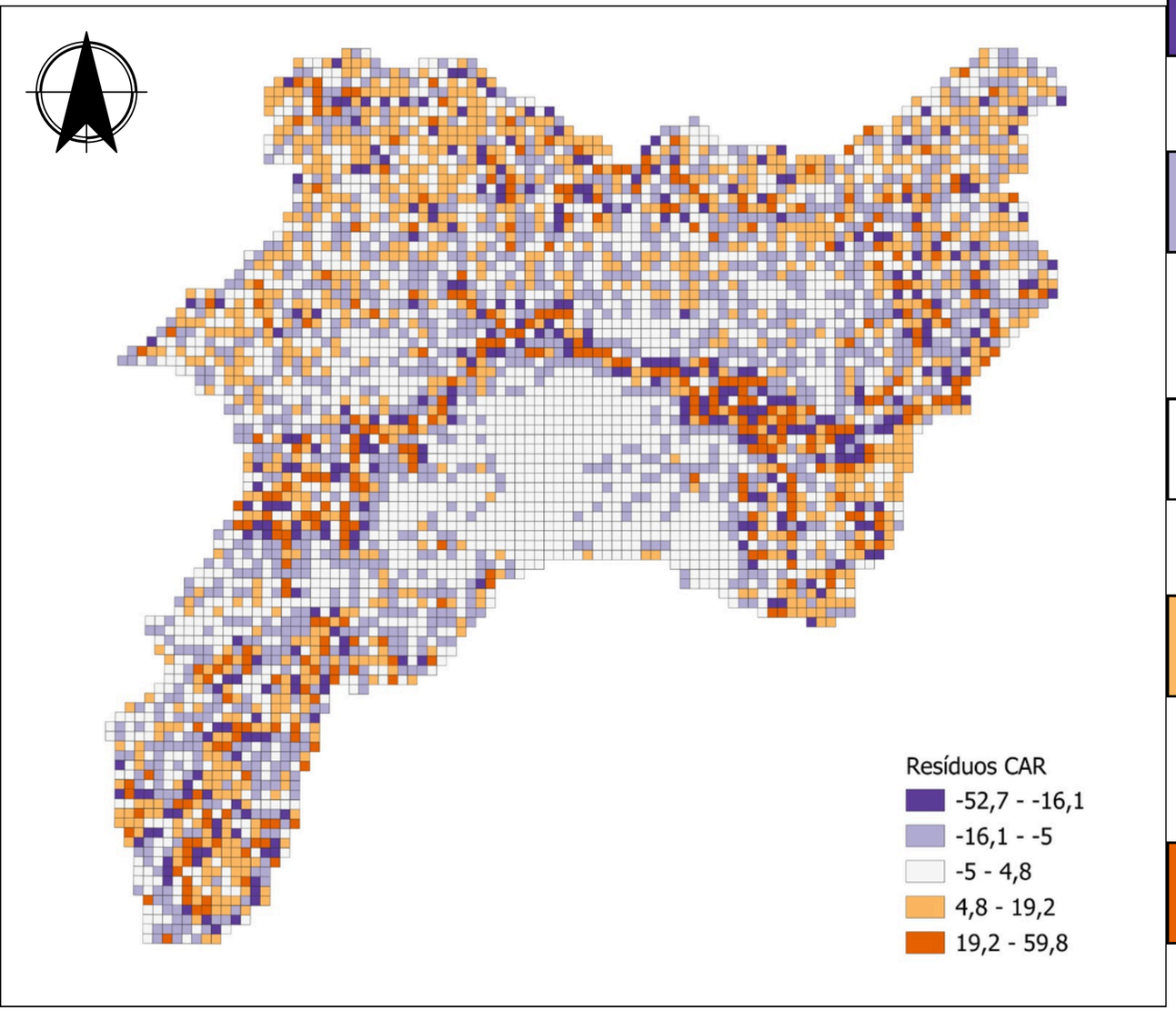
Variáveis independentes:

- Média da quantidade produzida de lenha (m^3) → **A**
- Quantidade de focos de calor (acumulada) → **B**
- Expansão agropecuária (%) → **C**
- Média da declividade → **D**
- Percentual de Unidades de Conservação (PI e US) → **E**

- **Intercepto:** 5.364
- **Erro:** 12.7

$$Y = 5.364 \underbrace{- 0.001\,A}_{\text{Estatísticamente não significativo}} + 0.271\,B \underbrace{- 0.009\,C}_{\text{Contribui negativamente?}} - 0.427\,D \underbrace{- 0.026\,E}_{\text{Baixa contribuição?}} \pm \underbrace{12.7}_{\text{Erro alto}}$$

Modelo CAR



-52.7 - -16.1

RESÍDUOS MAIS NEGATIVOS:
O modelo **subestimou significativamente** os valores observados nessas áreas

-16.1 - -5

RESÍDUOS NEGATIVOS:
Áreas onde o modelo **subestimou** os valores observados, mas de uma forma mais moderada

-5 - 4.8

RESÍDUOS PRÓXIMOS A ZERO:
Áreas em que as previsões estão muito próximas dos valores reais

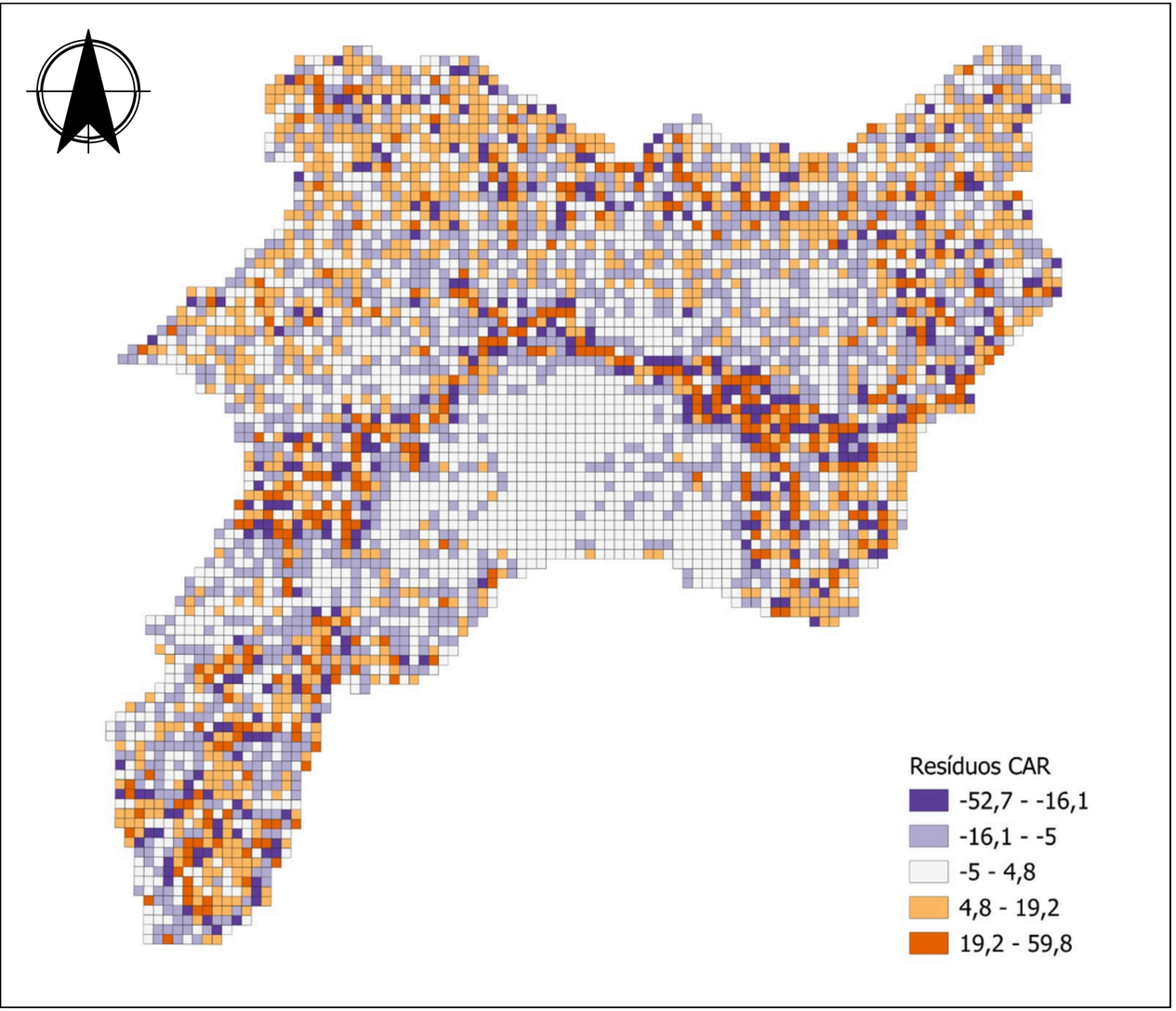
4.8 - 19.2

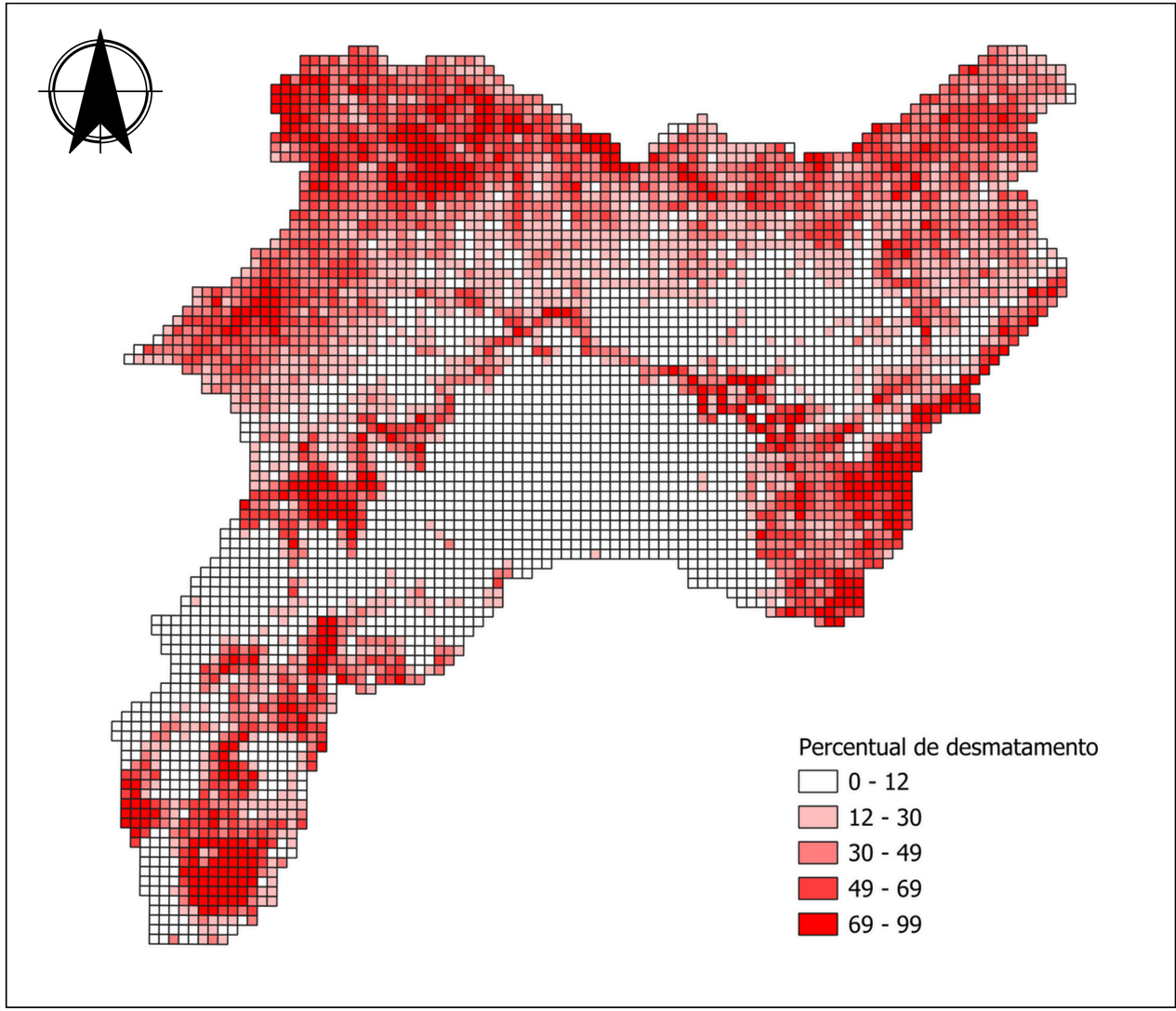
RESÍDUOS POSITIVOS:
O modelo **superestimou** os valores observados nessas áreas, mas de forma moderada

19.2 - 59.8

RESÍDUOS MAIS POSITIVOS:
O modelo **superestimou significativamente** os valores observados nessas áreas

Modelo CAR


Resíduos CAR
-52,7 - -16,1
-16,1 - -5
-5 - 4,8
4,8 - 19,2
19,2 - 59,8


Percentual de desmatamento
0 - 12
12 - 30
30 - 49
49 - 69
69 - 99

Modelo CAR

Variável dependente:

- Desmatamento acumulado (%) → **Y**

Variáveis independentes:

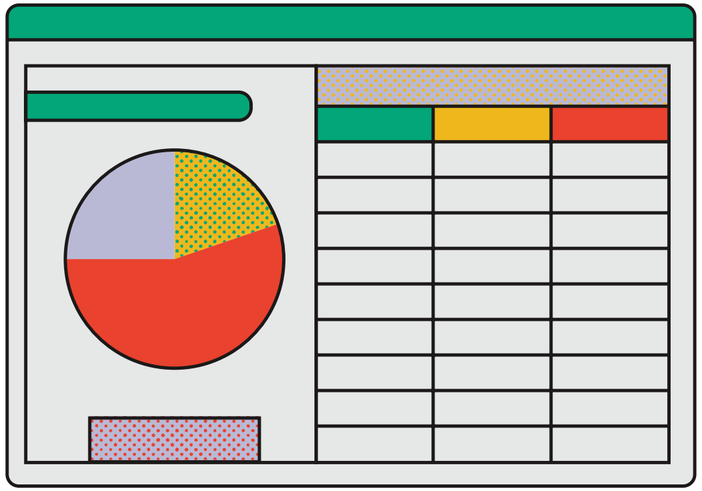
- Média da quantidade produzida de lenha (m^3) → **A**
- Quantidade de focos de calor (acumulada) → **B**
- Expansão agropecuária (%) → **C**
- Média da declividade → **D**
- Percentual de Unidades de Conservação (PI e US) → **E**

- **Intercepto:** 34.626
- **Erro:** 12.5

$$Y = 34.626 \underbrace{- 0.001\,A}_{\text{Estatísticamente não significativo}} + 0.337\,B \underbrace{- 0.006\,C}_{\text{Contribui negativamente?}} - 0.932\,D \underbrace{- 0.085\,E}_{\text{Baixa contribuição?}} \pm \underbrace{12.5}_{\text{Erro alto}}$$

8

Tabelas do Modelo GWR



Largura da Banda Kernel

Define o número de vizinhos considerados para estimar os coeficientes localmente.

Quanto menor, mais o modelo se concentra em capturar a variação local, o que pode ser áreas com alta densidade de dados.

Erro de validação cruzada

Mede o desempenho do modelo ao prever valores em um subconjunto de dados não usados no treinamento.

Quanto menor o valor melhor o ajuste do modelo e adequação as variações espaciais

Adaptative q	CV score
0.381966	2329493
0.618034	2399770
0.236068	2227345
0.145898	2076548
0.09016994	1886703
...	
0.03388681	1646795

Fator	Min	1Q	Median	3Q	Max	Global
Intercepto	2.0640507	16.0622409	27.1902688	38.9535665	69.6214577	25.9302
M_Lenha_M3	-0.0534886	-0.0032752	0.0007618	0.0033227	0.0201513	-0.0011
Foco_acumu	-0.6334543	0.5832549	1.0422401	1.6678358	4.1464452	1.2257
ExpAgr_per	-0.0952885	-0.0613443	-0.0403778	-0.0245380	-0.0029615	-0.0292
Declv_mean	-4.1134956	-1.2226860	-0.6061048	-0.0130761	2.4664671	-0.3063
Perc_UCs	-0.2804239	-0.1558406	-0.1075697	-0.0533653	0.0904531	-0.0878

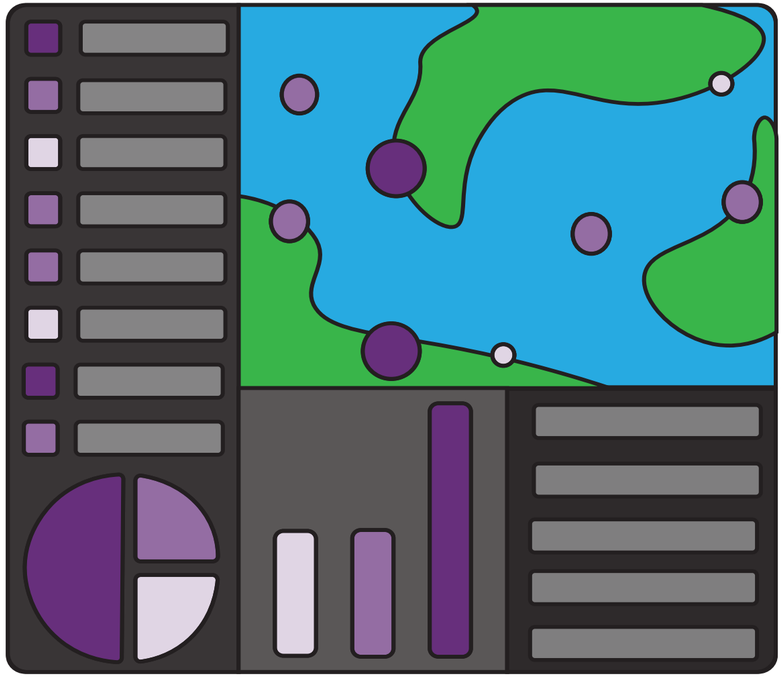
Effective number of parameters (Residuals): 99.17328
Effective degrees of freedom (Residuals): 4565.827
Sigma (Residuals): 17.89825
Effective number of parameters (Model): 72.70988
Effective degrees of freedom (Model): 4592.29
Sigma (Model): 17.84661

Sigma (ML): 17.70698
AICc (GWR p. 61, eq 2.33; p. 96, eq. 4.21): 40202.55
AIC (GWR p. 96, eq. 4.22): 40125.44
Residual sum of squares: 1462650
Quasi-global R2: 0.5267437 Explicação do Desmatamento

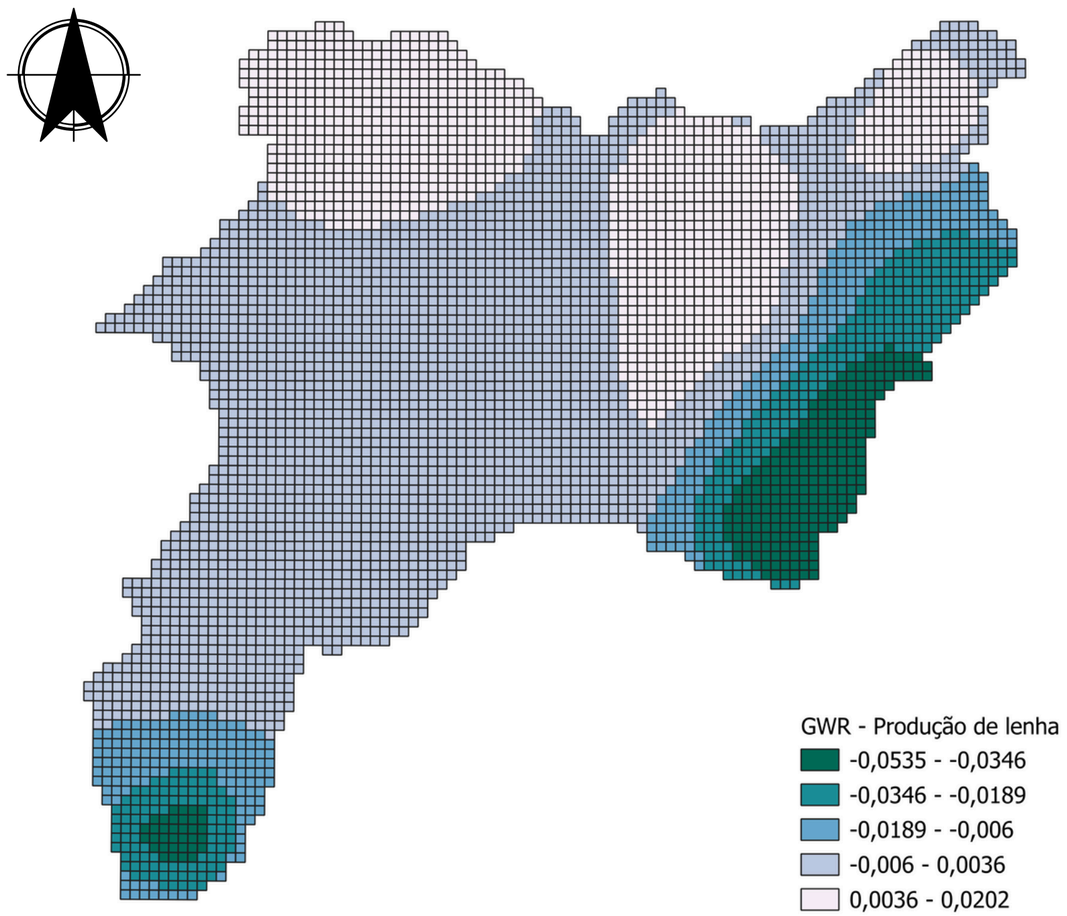
Modelo captura bem a variação e a relação espacial das variáveis.

9

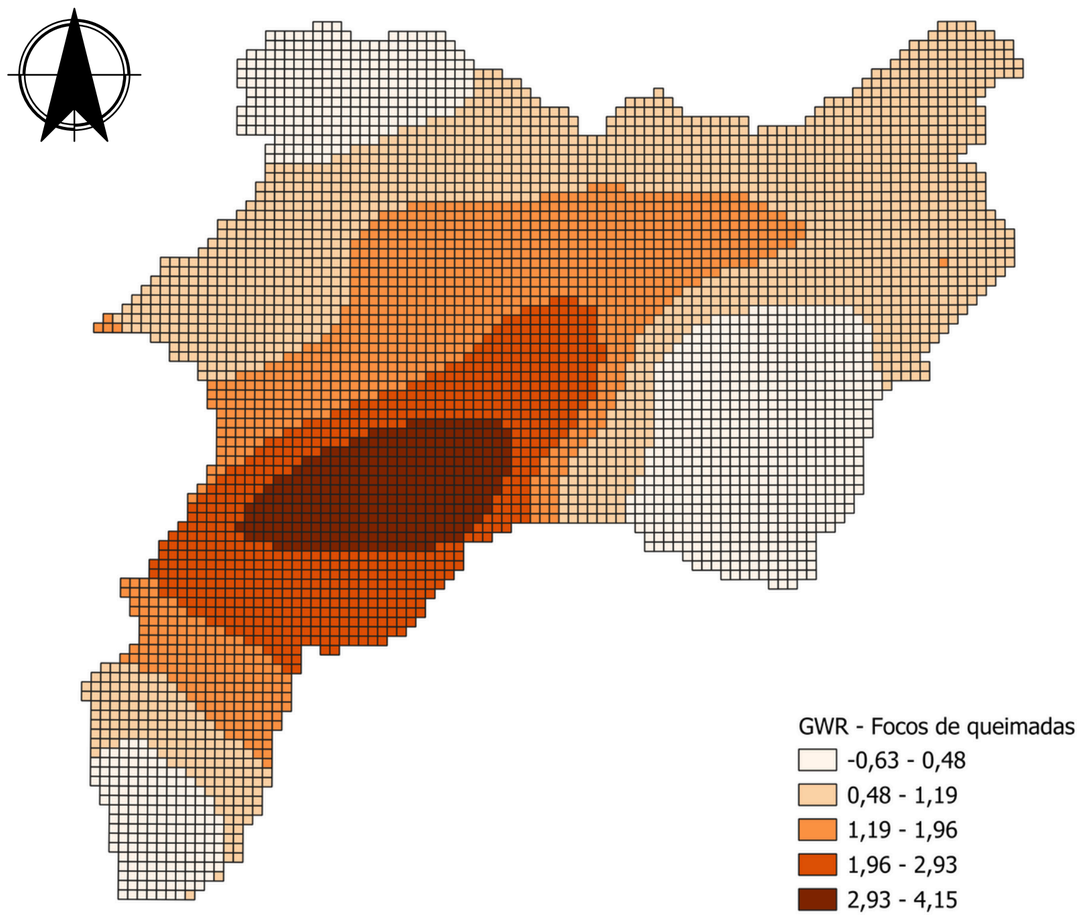
Mapas do Modelo GWR



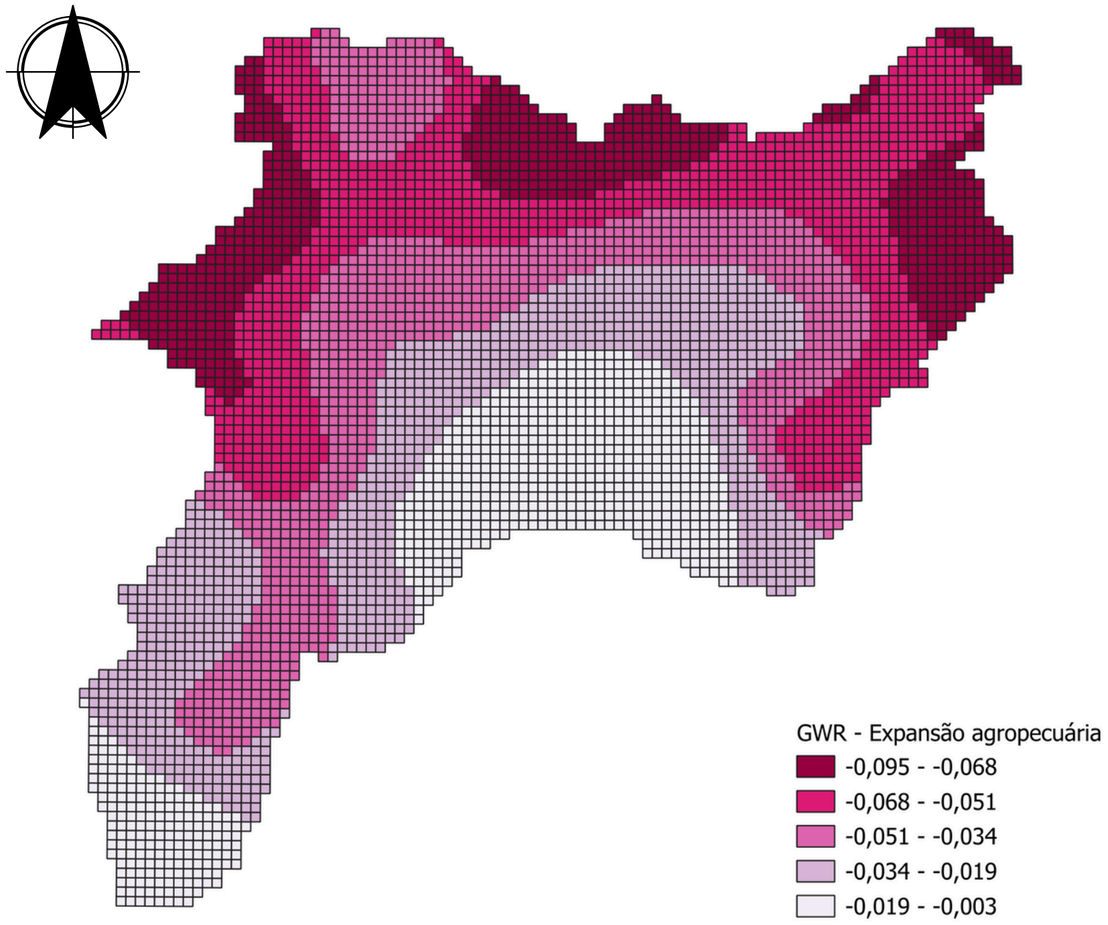
Como as relações entre variáveis mudam espacialmente?
Qual o efeito de cada variável no desmatamento?



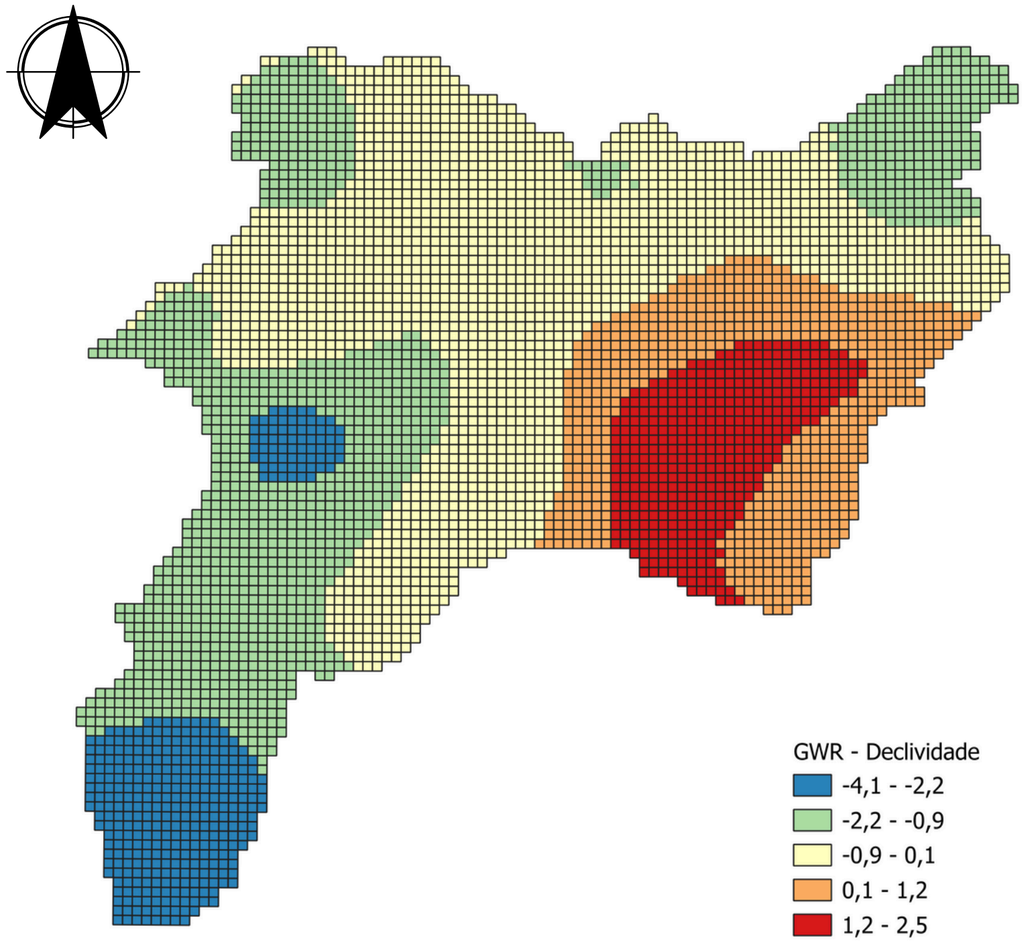
Relação entre desmatamento e produção de lenha pequena no geral. Negativa no extremo sul e leste da bacia.



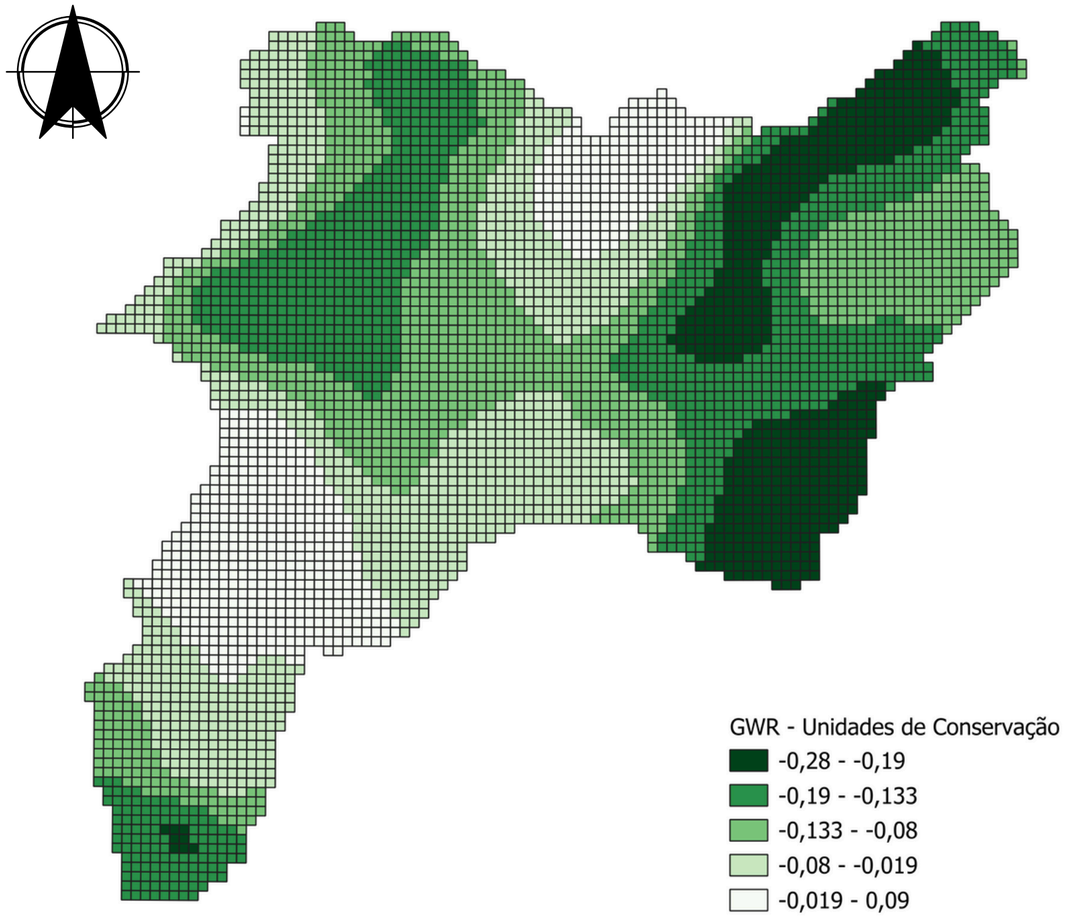
Variação espacial significativa. Relação positiva entre desmatamento e focos de queimadas principalmente no sudoeste da bacia.



Relação negativa entre desmatamento e expansão agropecuária, sobretudo na porção norte.

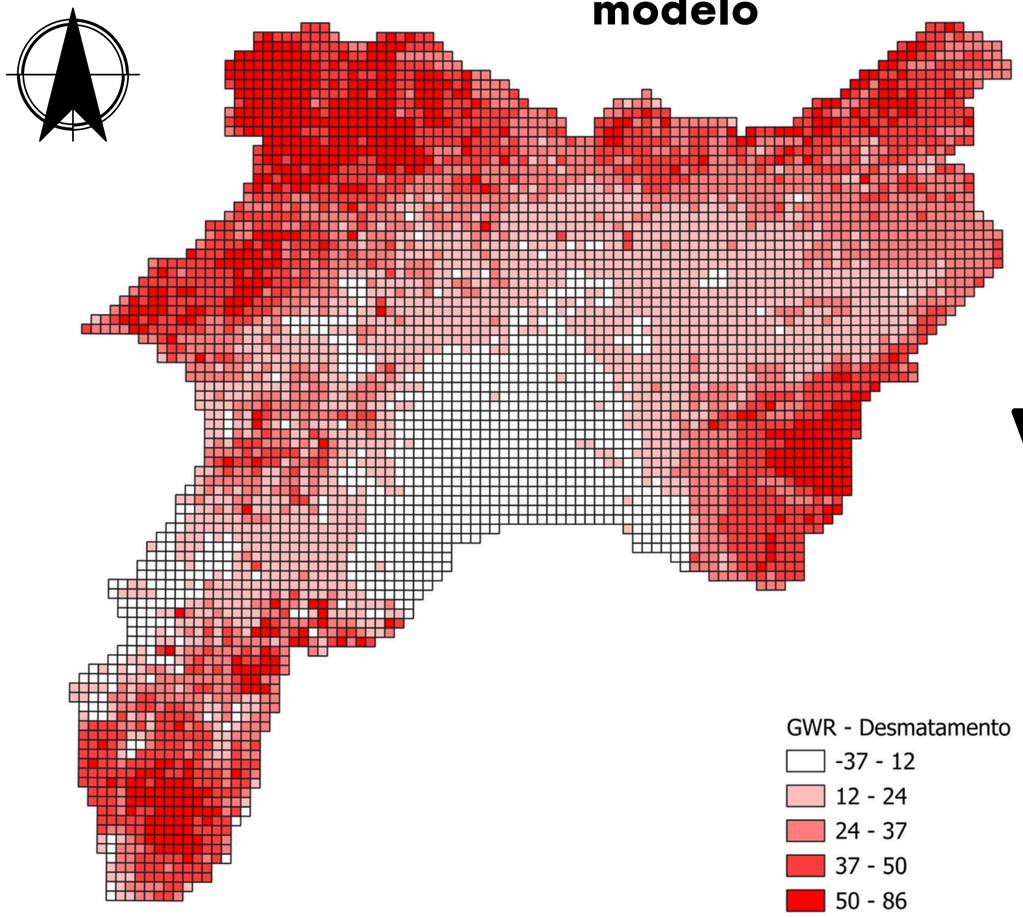


Variação espacial significativa. Relações negativas e positivas entre desmatamento e declividade a depender da área.



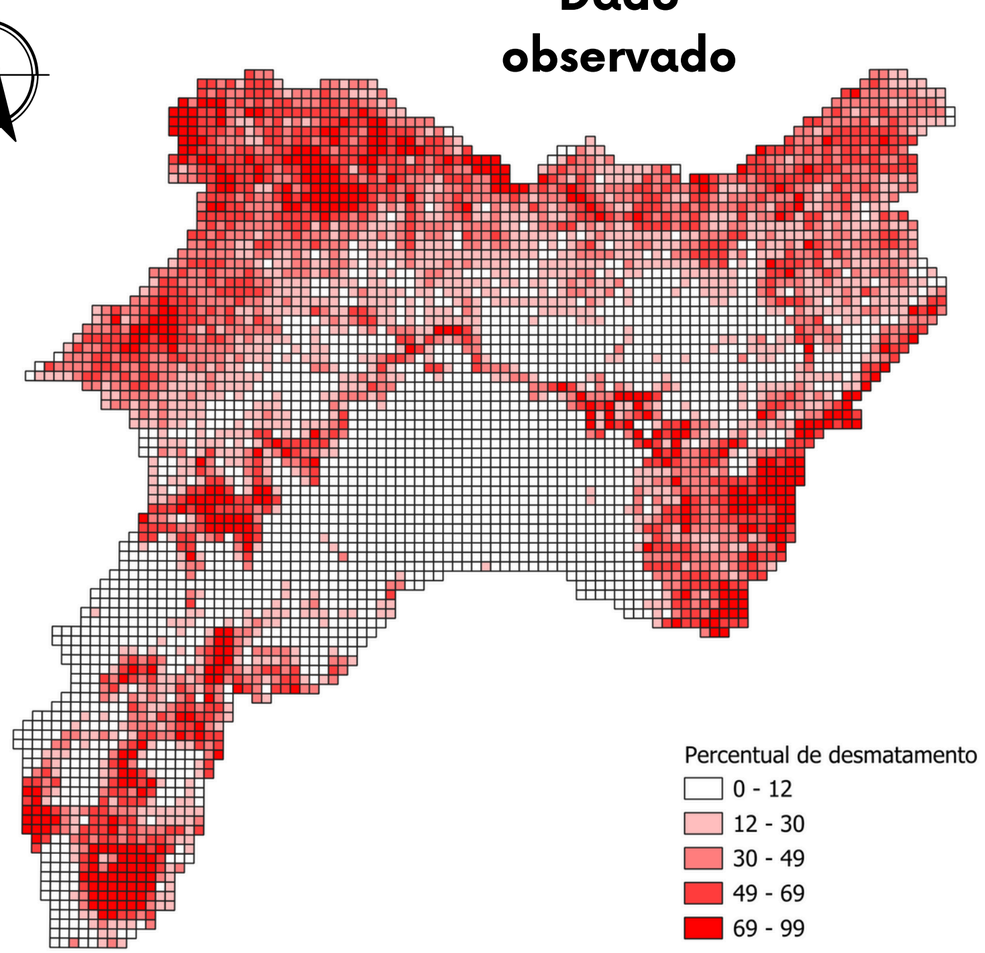
Relação entre desmatamento e Unidades de Conservação negativa, principalmente na porção leste e extremo sul da bacia.

Predito pelo modelo



VS

Dado observado



Estimativas locais baseadas nos coeficientes das variáveis explicativas.
> valor = modelo previu > desmatamento

MINISTÉRIO DA CIÊNCIA, TECNOLOGIA, INOVAÇÕES E COMUNICAÇÕES
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

Obrigado(a)!

"Aprender é a única coisa de que a mente nunca se cansa, nunca tem medo e nunca se arrepende" - Leonardo Da Vinci