

Bootstrap não paramétrico para análise de consistência de método para zoneamento agrícola

Alfredo José Barreto Luiz¹, Aline de Holanda Nunes Maia¹, Adalberto Koiti Miura², Marcos Correa Neves¹, Antonio Roberto Formaggio³, Yosio Edemir Shimabukuro³

Entende-se por área potencial para determinada atividade agrícola aquela que, por características naturais, geográficas, sociais, econômicas ou políticas, é vocacionada ou apresenta aptidão para este tipo de atividade, uso ou fim. A avaliação do potencial das terras é um importante instrumento de apoio na elaboração de políticas públicas e instrumentos de gestão, pois auxilia na conciliação das necessidades sociais e econômicas regionais, salvaguardando os recursos naturais para usos futuros (Miura et al., 2011). Estão disponíveis vários métodos capazes de prever o potencial das terras com base em seus atributos físicos, biológicos e socioeconômicos, cujos resultados são normalmente apresentados na forma de um zoneamento territorial (Andersen et al., 2005). A partir de um estudo de zoneamento pode-se obter um resultado quantitativo que especifique, por exemplo, quantos hectares de um município são considerados aptos para um conjunto específico de culturas agrícolas.

A pergunta que se apresenta é: como avaliar a qualidade desse prognóstico? Neste trabalho, propõe-se o uso da técnica *bootstrap* não paramétrico para obtenção de evidências que permitam responder a esta questão. Também são utilizados diversos recursos de visualização de dados quantitativos no intuito de orientar a discussão dos resultados. Partiu-se de um estudo de caso realizado por Miura et al. (2011) em 40 municípios de três microrregiões (Cerro Largo, Santo Ângelo e Santa Rosa) da Mesorregião Noroeste do Rio Grande do Sul (Figura 1), cujo objetivo foi identificar áreas aptas para o cultivo de espécies vegetais com potencial bioenergético.

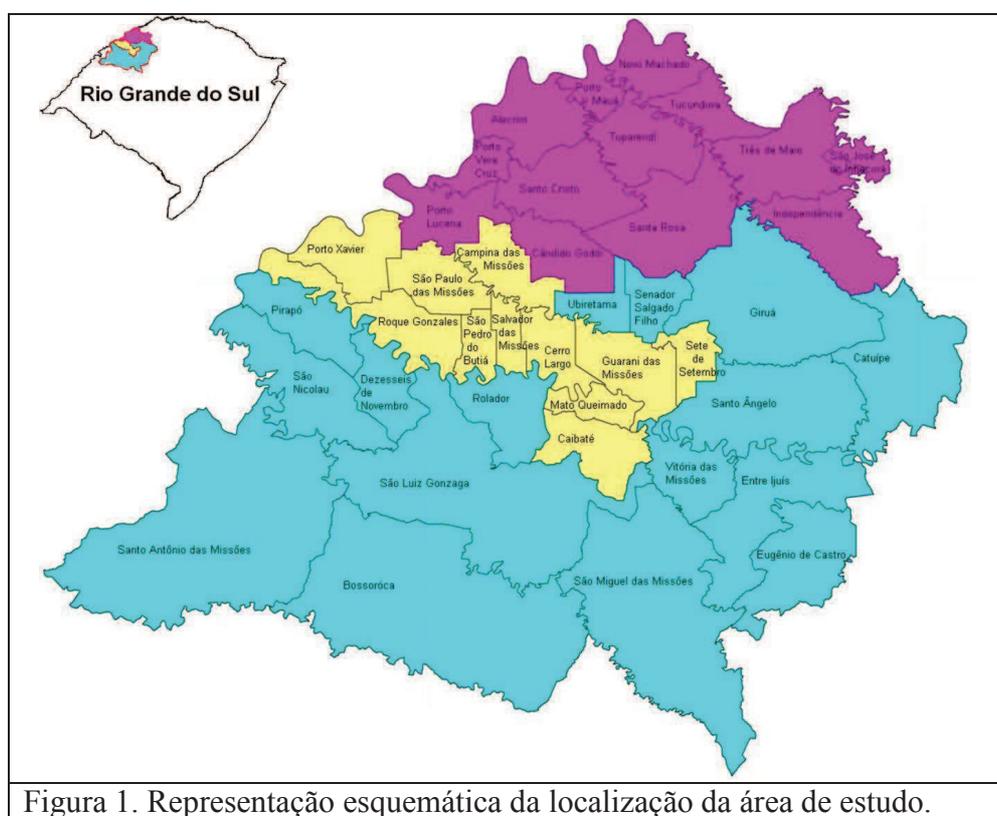


Figura 1. Representação esquemática da localização da área de estudo.

Como parte dos resultados (Tabela 1), para cada município foi quantificada de forma absoluta (em hectares) e relativa (em porcentagem da área do município) a “Área de Potencial Ajustado à Agroenergia” (APAA). A APAA ainda foi dividida em três classes, com alta, média e baixa aptidão (Figura 2). Para testar a adequação desta quantificação, os percentuais observados

¹ Embrapa Meio Ambiente; ² Embrapa Clima Temperado; ³ INPE/OBT/DSR

foram utilizados para gerar 10.000 amostras *bootstrap*, nas quais os valores foram atribuídos aleatoriamente, com reposição, aos municípios. Posteriormente os percentuais foram transformados para hectares, ao se multiplicar pela área total do município. De forma independente da utilizada na construção do zoneamento, foram obtidos dados relativos às áreas anualmente utilizadas por algumas culturas agrícolas e florestais (AREA) nos mesmos municípios. As culturas agrícolas consideradas foram: Amendoim; Arroz; Cana-de-açúcar; Girassol; Mamona; Mandioca; Milho; Soja; Sorgo e Trigo. Sabe-se que algumas dessas lavouras são cultivadas em seqüência, na mesma área e no mesmo ano, mas as áreas foram somadas para dar uma idéia da intensidade da atividade de produção agrícola do município e não exatamente para representar a área total cultivada. Devido a este procedimento, é possível que a soma da área cultivada seja maior que a área total do município, significando que grande parte da área com culturas anuais é utilizada por mais de uma cultura durante o ano.

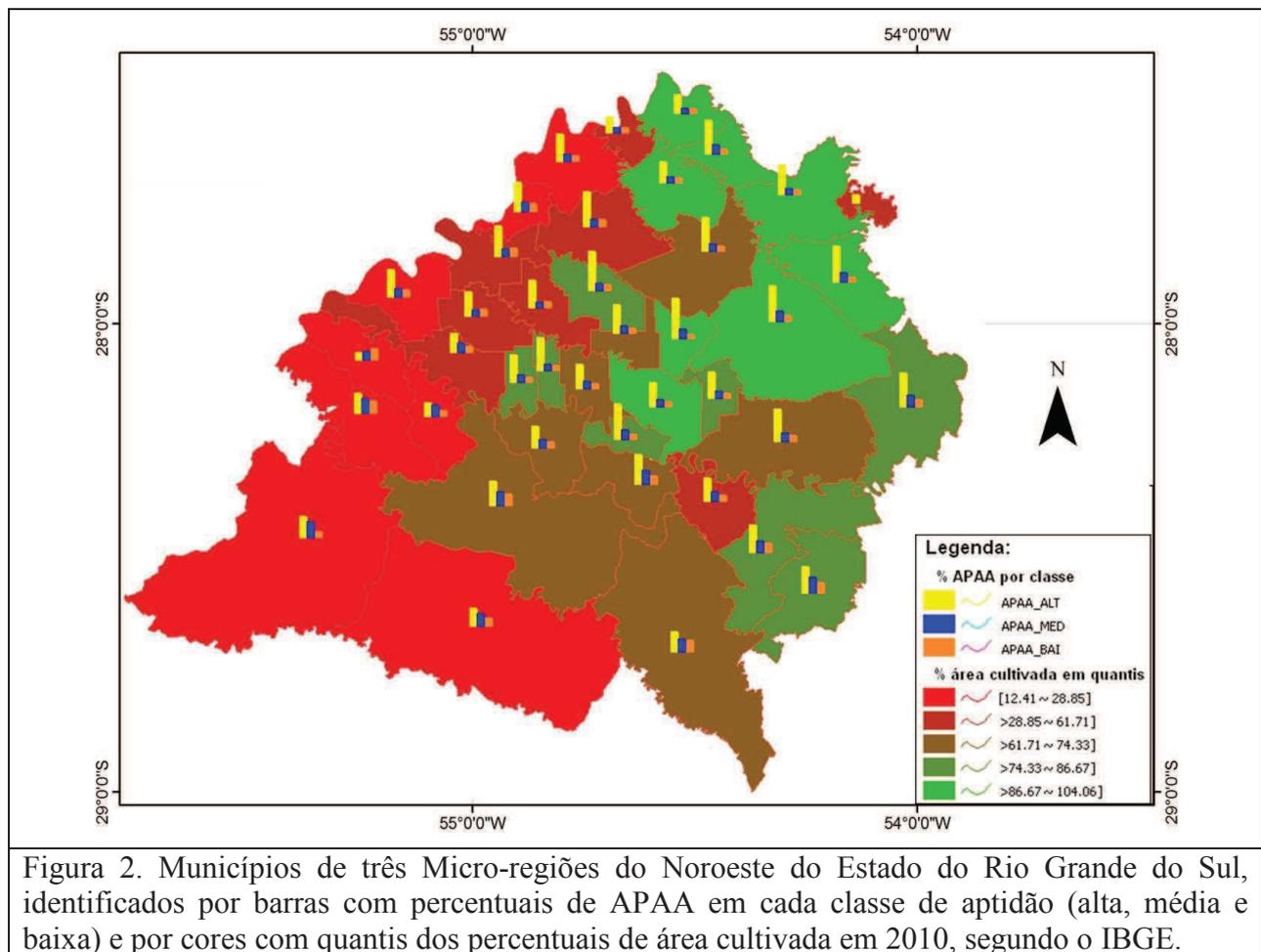


Figura 2. Municípios de três Micro-regiões do Noroeste do Estado do Rio Grande do Sul, identificados por barras com percentuais de APAA em cada classe de aptidão (alta, média e baixa) e por cores com quantis dos percentuais de área cultivada em 2010, segundo o IBGE.

Outro ponto a considerar é que nem todas essas culturas são atualmente utilizadas para geração de energia no Brasil, mas foram escolhidas por que todas elas apresentam, pelo menos teoricamente, potencial para geração de energia. Outras culturas com relatos de presença na região de estudo e com potencial para produção de energia, mas que não são avaliadas nos levantamentos da Produção Agrícola Municipal (PAM) do IBGE, como a canola, por exemplo, não tiveram suas áreas consideradas. Para estimar a área ocupada por espécies florestais, foram utilizados os dados de produção de carvão vegetal, lenha e madeira em toras, obtidos pelo levantamento Produção da Extração Vegetal e da Silvicultura do IBGE.

Para transformar em medida de área os valores da produção de carvão, fornecido em toneladas pelo levantamento do IBGE, foi utilizada uma média de 15 toneladas por hectare. De maneira semelhante, utilizou-se a média de 80 metros cúbicos por hectare na transformação dos

valores da produção de lenha e de 60 metros cúbicos por hectare no caso da madeira em tora. Todos os valores de área, por município, foram somados para constituir a variável denominada AREA. Foi então calculado o coeficiente de correlação entre os valores identificados pelo método proposto (APAA) e os obtidos de forma independente (AREA), cujos percentuais por município podem ser observados na Figura 3.

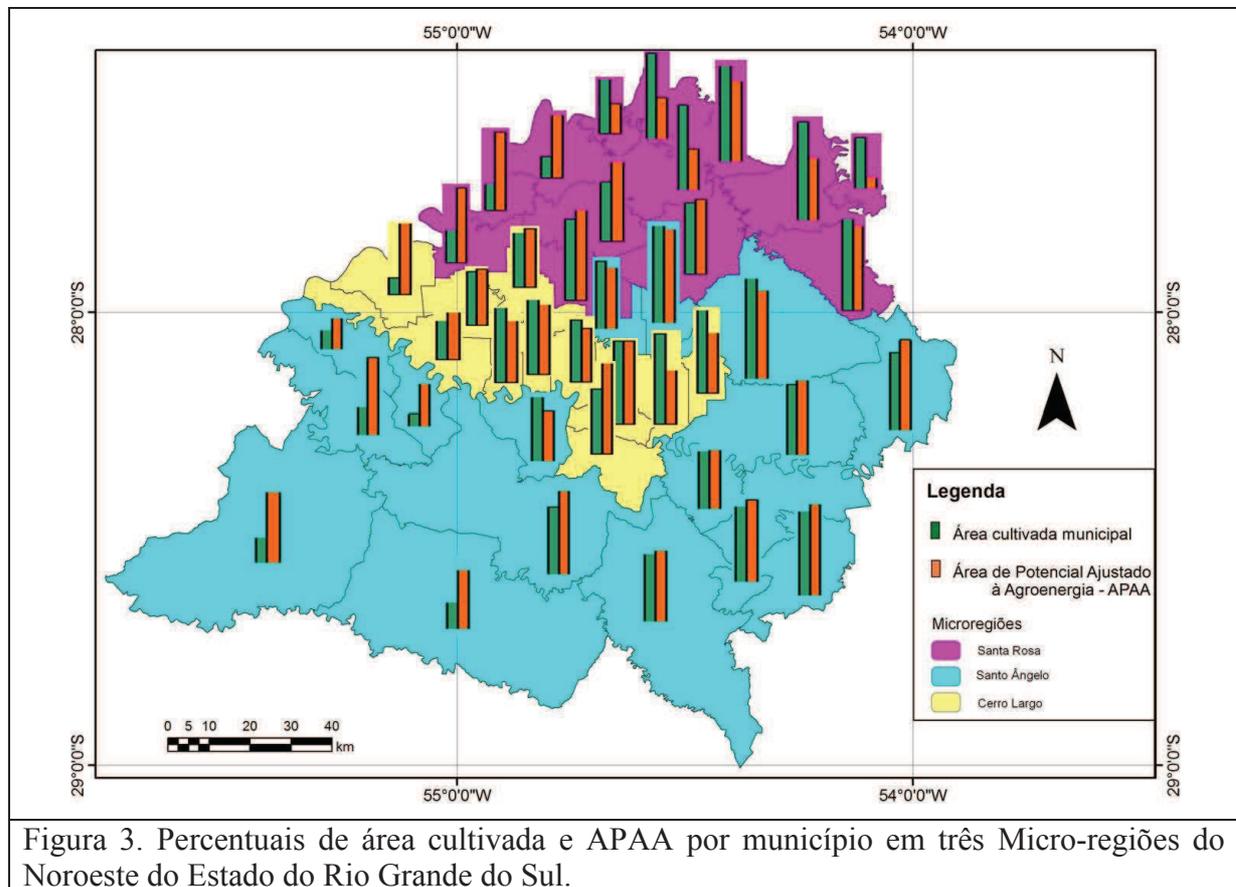


Figura 3. Percentuais de área cultivada e APAA por município em três Micro-regiões do Noroeste do Estado do Rio Grande do Sul.

Como cada município teve a APAA quantificada de forma absoluta (em hectares) e relativa (em porcentagem da área do município), para testar a adequação desta quantificação os percentuais observados foram utilizados para gerar 10.000 amostras *bootstrap*, nas quais os valores foram atribuídos aleatoriamente, com reposição, aos municípios. Posteriormente, os percentuais foram transformados para hectares, ao se multiplicar pela área total do município. Dessa forma, puderam ser calculados também os coeficientes de correlação entre os valores APAA e os valores de AREA nas dez mil amostras *bootstrap* e, a partir deles, foi construída uma distribuição de referência.

O coeficiente de correlação de Pearson observado (0,83602) foi superado pelos valores estimados em apenas 489 amostras *bootstrap*, o que constitui uma forte evidência da consistência do método empregado (nível de significância nominal $p = 0,0489$). Foram calculadas também as estatísticas de correlação de Spearman (0,81144) e Kendall (0,65128) e a medida da dependência D de Hoeffding (0,35355), cujos valores p obtidos pelo método *bootstrap* não paramétrico foram 0,0535, 0,0337 e 0,0394, respectivamente. Também é apresentada a distribuição espacial do percentual de aptidão nas microrregiões estudadas. Tanto a geração das amostras *bootstrap* como os cálculos das correlações foram realizados no software SAS (SAS, 2010). O programa SAS para os cálculos é apresentado na Tabela 2.

O gráfico da Figura 4 é um exemplo de como as técnicas quantitativas, incluindo as de visualização de dados, podem ser úteis na investigação da qualidade dos prognósticos de um sistema que pretende realizar zoneamento agrícola ou semelhante. Bossoroca e Santo Antônio das Missões são os dois maiores municípios em área total e ficam na extremidade sudoeste da

região estudada. Os motivos pelos quais os dois municípios destacados se afastam do comportamento dos demais devem ser profundamente estudados e podem vir a auxiliar no aprimoramento do instrumento de zoneamento ou, por outro lado, levar à melhor compreensão da dinâmica espacial da ocupação agrícola.

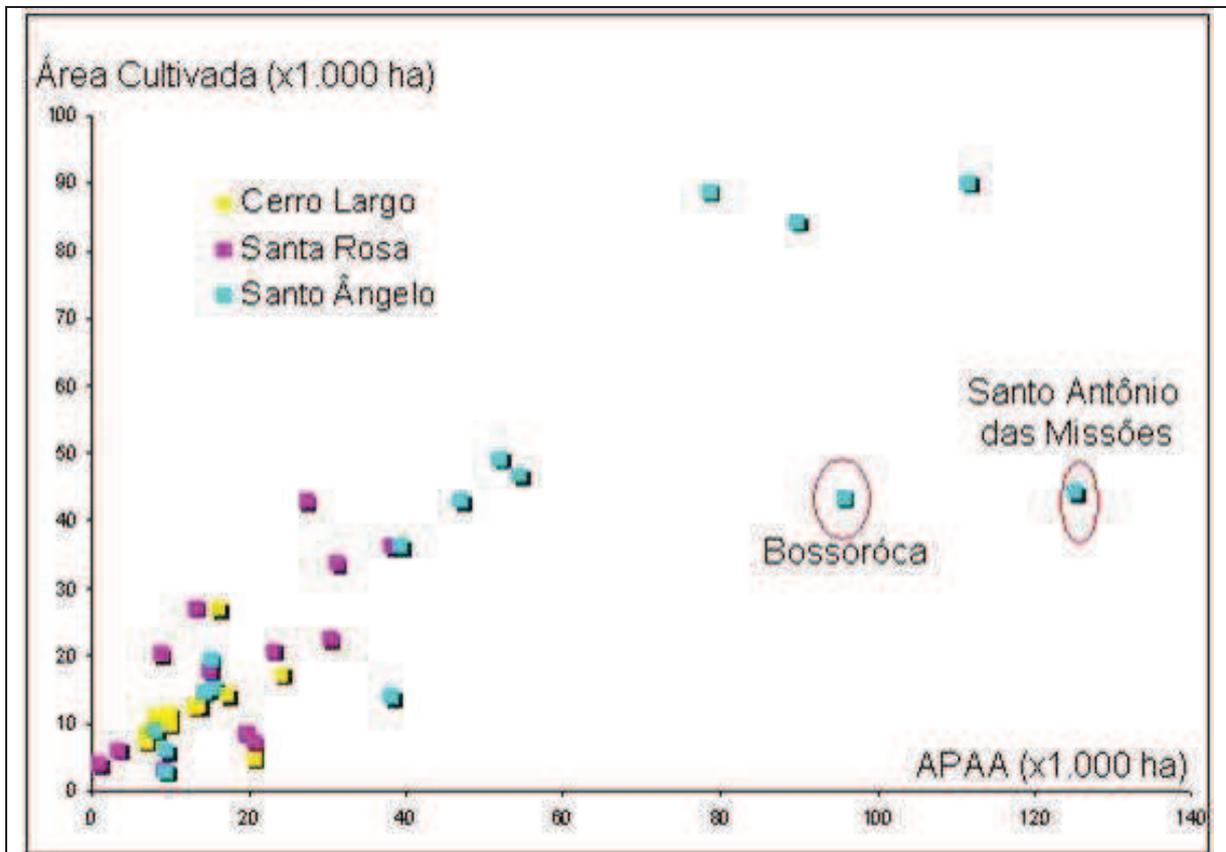


Figura 4. Gráfico da relação entre área cultivada e APAA para os municípios de três Micro-regiões do Noroeste do Estado do Rio Grande do Sul.

As diferenças entre os percentuais de APAA e de área cultivada foram ajustas por um modelo de regressão que considerou como variáveis independentes a latitude, a longitude e a interação (multiplicação) entre elas. Foram adotadas as coordenadas da sede dos municípios. Embora o ajuste tenha sido pouco explicativo (coeficiente de determinação ajustado igual a 0,428), a análise de variância indicou que o ajuste ao modelo foi altamente significativo ($p < 0,0001$). Isso significa que pode existir um viés relacionado à posição geográfica dos municípios no procedimento adotado para o estabelecimento da aptidão das áreas ou que existem grupos de municípios vizinhos com comportamento atual com relação às atividades agrícolas, fortemente determinado por alguma característica local ou regional que não foi levada em consideração na elaboração do sistema de atribuição de aptidão.

A análise de regressão não é a melhor maneira de diagnosticar essa tendência, pois o conjunto dos municípios estudados não se configura uma amostra probabilística e é lícito supor que deve existir alguma estrutura de covariância entre municípios vizinhos, que não é levada em consideração nesse tipo de ajuste linear. Entretanto a simples visualização dos dados quantitativos pode auxiliar em muito na análise dos resultados. Na Figura 5 já é possível perceber que a distribuição espacial dos erros negativos e positivos, e dos seus valores absolutos não parece ser aleatória.

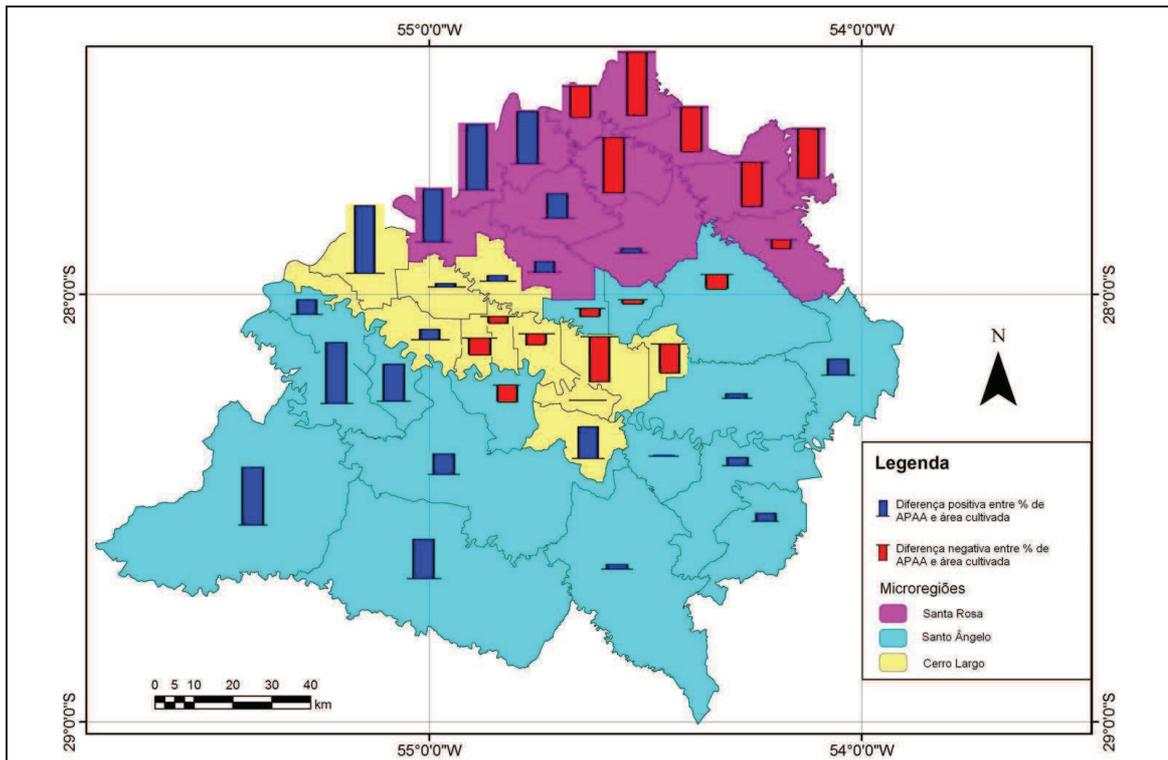


Figura 5. Diferenças entre os percentuais de área cultivada e APAA por município em três Micro-regiões do Noroeste do Estado do Rio Grande do Sul.

Entretanto, se abstrairmos ainda mais as forma de representar os mesmo dados, como feito na Figura 6, fica ainda mais visível a existência de algum tipo padrão espacial na distribuição dos erros observados.

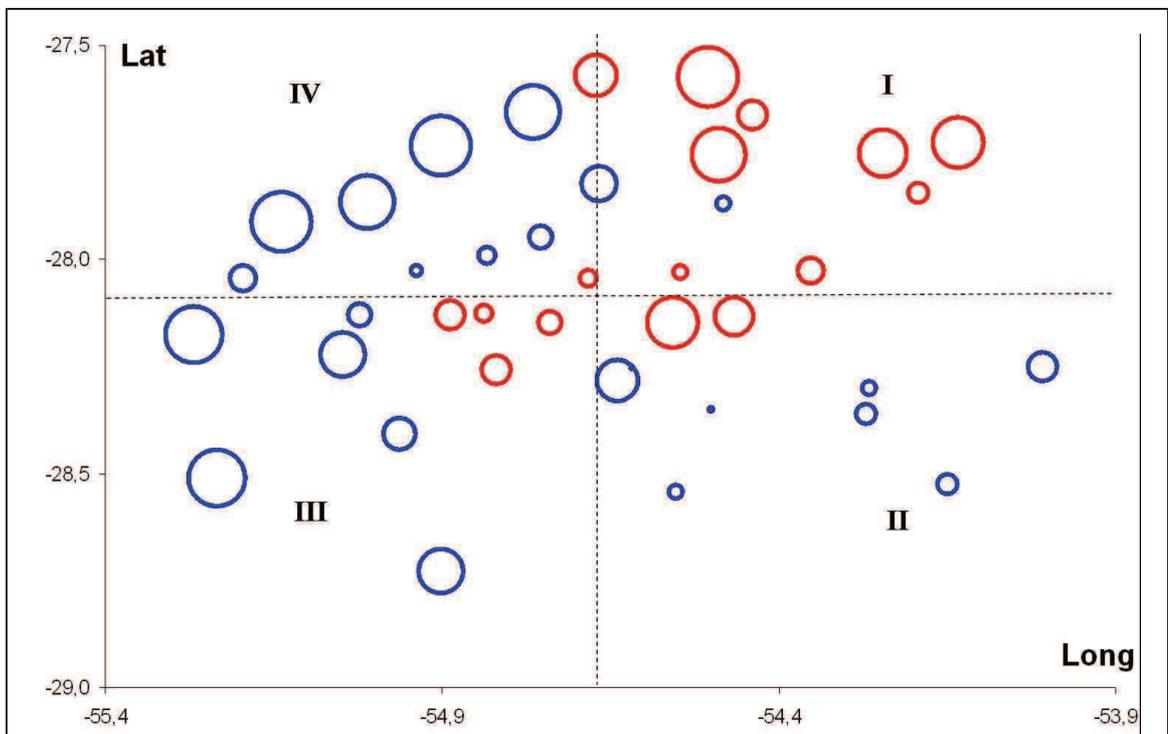


Figura 6. Espacialização esquemática das diferenças entre os percentuais de área cultivada e APAA para 40 municípios do Noroeste do Estado do Rio Grande do Sul.

Na Figura 6, cada quadrante contém as coordenadas das sedes de 10 municípios. Pode ser observado que o quadrante I é predominantemente composto por municípios que já atingiram ou

superaram o potencial para cultivos com possível utilização para a geração de bioenergia. No quadrante II o que chama a atenção é a proximidade entre os valores preditos pelo método (APAA) e os efetivamente encontrados (AREA), que ficam evidenciados por círculos de menor diâmetro, ou seja, baixos erros. Os quadrantes III e IV são predominantemente ocupados por municípios que ainda não atingiram o potencial estimado pelo método, exceto talvez por aqueles situados em posição central na região estudada, onde os erros são pequenos e às vezes positivos, às vezes negativos. A partir dessas análises, tanto o método pode ser revisto como interpretações para os motivos que expliquem a situação atual podem ser elaboradas.

Avaliar a qualidade de um prognóstico, como foi feito no presente caso, é um desafio, pois, mesmo que consistente, ele só se realizará no futuro. Neste trabalho, visando contornar esta dificuldade de avaliação da consistência do método proposto, empregou-se a técnica *bootstrap* para obtenção de evidências que permitisse de alguma forma avaliar a qualidade das estimativas da APAA para os municípios. Também foram utilizadas técnicas de visualização de dados quantitativos e espacialmente distribuídos, de forma a auxiliar nas análises e interpretações.

Referências

ANDERSEN, R. S.; TOWERS, W.; SMITH, P. Assessing the potential for biomass energy to contribute to Scotland's renewable energy needs. **Biomass and Bioenergy**, Oxford, v.29, n.2, p.73-82, 2005.

MIURA, A. K. ; FORMAGGIO, A. R.; SHIMABUKURO, Y. E.; ANJOS, S. D. dos; LUIZ, A. J. B. Avaliação de áreas potenciais ao cultivo de biomassa para produção de energia e uma contribuição de sensoriamento remoto e sistemas de informações geográficas. **Engenharia Agrícola**, Jaboticabal, v. 31, n. 3, p. 607-620, 2011.

SAS Institute Inc. **Base SAS® 9.2 Procedures Guide: Statistical Procedures**. Third Edition. Cary, NC: SAS Institute Inc. 2010.

Tabela 1 - Áreas de Potencial Ajustado à Agroenergia (APAA) por município de três microrregiões do Noroeste do Rio Grande do Sul e área cultivada com espécies potencialmente energéticas (IBGE, 2010).

Microrregião	Município Nome	Área (km ²)	APPA (%)			Área cultivada (ha)
			Alto	Médio	Baixo	
Cerro Largo	Caibaté	257,33	60,10	23,05	10,33	17.273,95
	Campina das Missões	224,58	55,02	1,61	4,36	12.585,30
	Cerro Largo	177,30	47,88	6,44	1,19	11.407,57
	Guarani das Missões	289,95	48,17	5,97	0,88	26.956,67
	Mato Queimado	114,81	71,83	12,08	1,73	9.815,25
	Porto Xavier	281,51	55,32	10,58	7,63	4.902,05
	Roque Gonzales	348,89	33,30	12,41	3,25	14.093,38
	Salvador das Missões	93,83	66,41	4,58	1,40	7.246,36
	São Paulo das Missões	224,43	47,16	4,98	5,91	12.442,18
	São Pedro do Butiá	106,64	54,82	7,42	1,70	8.271,25
	Sete de Setembro	130,27	52,50	7,36	2,82	11.283,17
Alecrim	315,83	53,95	6,12	5,10	6.998,12	
Santa Rosa	Cândido Godói	245,83	82,16	7,16	4,35	20.706,36
	Independência	356,93	73,43	11,80	2,14	33.871,89
	Novo Machado	220,69	35,65	3,01	1,22	20.326,97
	Porto Lucena	249,54	61,83	7,01	9,58	8.591,44
	Porto Mauá	105,29	28,11	2,12	0,72	5.971,18
	Porto Vera Cruz	111,89	59,98	12,30	9,59	3.094,52
	Santa Rosa	488,28	68,75	7,85	1,54	36.293,55
	Santo Cristo	366,69	70,78	5,76	6,26	22.628,30
	São José do Inhacora	77,55	10,52	0,00	0,53	4.073,67
	Três de Maio	421,36	58,71	4,24	1,56	42.814,52
	Tucunduva	180,52	70,63	10,88	1,51	17.797,68
Tuparendi	308,24	38,24	3,35	0,85	27.243,08	
Santo Ângelo	Bossoroca	1.592,74	32,62	18,14	9,35	43.327,48
	Catuípe	584,46	69,67	16,42	7,02	46.614,00
	Dezesseis de Novembro	217,62	22,52	16,01	4,53	2.701,67
	Entre-Ijuís	552,32	54,76	17,19	12,87	43.019,67
	Eugenio de Castro	418,26	52,77	25,68	15,42	36.251,33
	Giruá	854,72	73,36	14,96	3,53	88.945,83
	Pirapó	293,54	7,90	8,25	16,05	5.839,99
	Rolador	293,30	39,87	8,18	3,26	19.173,47
	Santo Ângelo	679,18	64,02	9,30	3,28	49.158,33
	Santo Antônio das Missões	1.710,05	40,36	27,70	5,08	44.051,71
	São Luiz Gonzaga	1.296,84	48,09	22,14	15,72	90.083,37
	São Miguel das Missões	1.226,89	36,46	19,40	17,24	84.378,00
	São Nicolau	481,85	36,64	24,33	18,46	13.903,23
	Senador Salgado Filho	147,25	84,89	11,28	0,20	14.723,20
	Ubiretama	126,82	55,98	6,61	0,67	8.814,15
Vitória das Missões	258,45	43,81	11,74	4,40	15.240,83	

Tabela 2. Programa SAS para aplicação da técnica bootstrap e cálculo do coeficiente de correlação de Kendal. Os cálculos para os coeficientes de Pearson, Spearman e Hoeffding seguem o mesmo modelo.

```
Data APP_data;
  input municip appa; cards;
193.4713943 ...
;
Data agrosilv;
  input municip area agro; cards;
125733 17273.9500 ...
;
Data boot;
  do index=1 to 10000;
    do i=1 to 40;
      x=round(ranuni(35791)*40);
      set APP_data
        point=x;
      output; end;
    end;
  drop municip;
stop;
Proc sort data=boot;
  by i index;
Data boot;
  set boot;
  municip=i;
Data dois;
  merge agrosilv boot;
  by municip;
  appa=appa*area/100;
Proc sort data=dois;
  by index municip;
Proc corr data=dois noprint outk=ken;
  var appa;
  with agro;
  by index;
Data kendal; set ken;
  if appa>=1 then delete;
  appak=appa;
  keep appak;
Proc sort data=kendal;
  by appak;
Data kendal;
  set kendal;
  if appak< 0.65128 then corrk=0; else corrk=1;
Proc means data=kendal sum;
  var corrk
run;
```
