
MEDIDAS KAPPAS MODIFICADAS PARA
A COMPARAÇÃO DE MAPAS TEMÁTICOS

JOÃO MARCELO LOPES TAVARES

Orientação
Profa. Dra. Viviana Giampaoli

Área de concentração
Estatística Matemática

Dissertação submetida como requerimento parcial para obtenção do
grau de Mestre em Estatística pela Universidade Federal de Pernambuco

Recife, fevereiro de 2004

À minha esposa Inês e filhos Rodrigo, Gustavo e
Leonardo, razões do meu viver.

Aos meus pais (in memoriam) Raimundo Lopes e
Maria Cristina, responsáveis pela minha existência.
A minha esposa Inês, amiga e companheira.
Aos meus filhos Rodrigo, Gustavo e Leonardo,
razões da minha vida.

Agradecimentos

Primeiramente a Deus, por me cercar de pessoas maravilhosas e permitir a conclusão de mais esta etapa.

Aos meus pais (in memorian) Raimundo Lopes Tavares e Maria Cristina Tavares, responsáveis pela minha existência.

A minha esposa Inês, companheira e amiga, pelo seu incentivo, apoio, compreensão e disponibilidade, sem os quais este trabalho não seria concretizado.

Aos meus filhos Rodrigo, Gustavo e Leonardo, pela compreensão com minha ausência em momentos importantes de suas vidas, abdicando muitas das vezes de minha atenção.

À professora Viviana Giampaoli, pela dedicação e orientação no direcionamento desse estudo e firmeza de objetivos. Obrigado pela confiança e segurança depositadas em mim e pela paciência e ajuda nos momentos de insegurança.

À professora Cristina Raposo, pelo seu carinho e amizade, além dos ensinamentos e orientações sempre que solicitei.

Ao professor André Toom, pela dedicação e paciência no ensino da Matemática.

À professora Cláudia Regina, pelo carinho e contribuição para o meu aprendizado em Estatística.

Ao professor Francisco Cribari Neto, pela eficiente e exemplar competência na condução do Programa de Mestrado em Estatística.

Ao excelente professor Klaus Vasconcellos, por sua amizade e pela forma competente na construção dos conhecimentos em sala de aula.

Ao professor Yossi Zana, pela boa vontade, disponibilidade em ajudar e suas contribuições.

Aos professores do Programa de Mestrado em Estatística da UFPE, por suas contribuições à minha formação em Estatística.

A Keila Mara, por sua disposição em socializar seus conhecimentos e me ajudar sempre que solicitei.

À minha turma de mestrado, por se preocuparem sempre comigo: Keila Mara, Felipe, Moisés, Tatiane, Tarciana, Patrícia Leal, Sílvia, Bartolomeu, Raydonal e Gilson.

Aos colegas Almir, Antônio, Waldemar, Valdecy, Franklin, Cristiano, Veronica, Cristina Moraes, Patrícia Leone, Micheli, Diana, Carla, Heráclito, pela convivência gratificante e enriquecedora.

A Valéria, pela atenção e pelo apoio incondicional, dispensados ao longo do curso de pós-graduação.

A Adriana e Cícero, funcionários do departamento de estatística, pelo carinho e amizade.

Aos participantes da banca examinadora, pelas sugestões.

Resumo

O estudo de concordância, em relação a variáveis categorizadas, tem sido aplicado a vários contextos e seu uso não é novo. Em 1968, Cohen propôs o coeficiente Kappa Ponderado. Neste trabalho, propomos medidas de Kappas modificadas para a análise de concordância entre mapas temáticos, isto é, provenientes de algum tipo de classificação, a partir da proposta dada por Pontius em 2000. Estudamos o comportamento dessas medidas, através de um conjunto de simulações que fornecem mapas hipotéticos e, realizando as comparações correspondentes, observamos as vantagens delas. Realizamos uma aplicação comparando essas medidas em mapas geográficos referentes ao desflorestamento da região amazônica nos anos de 2000 e 2001. Nossos resultados revelam que as medidas Kappas modificadas propostas são mais estáveis quando se varia a resolução utilizada no experimento, sendo assim, medidas mais confiáveis.

Abstract

The study of concordance in relation to categorized variables has been used in many situations and this use is not new. In 1968, Cohen proposed the Weighted Kappa coefficient. In this dissertation, we are proposing modified Kappas measures for an analysis of concordance among thematic maps, which are from some kind of classification, based on the proposal given by Pontius in 2000. We had studied behaviour of those measures. Through a set of simulations, which provide hypothetic maps, we made correspondent comparisons and we observed the advantages of them. We had realized an application comparing those measures in geographic maps associated to damage in Amazonas region from the years 2000 and 2001. Our results show that the proposed modified Kappas measures are more stable when we vary the resolution used in the experiment, or in better words, they become more reliable measures.

Índice

1. Introdução	1
1.1. Considerações básicas sobre medidas de concordância	1
1.2. Vantagens e desvantagens do Kappa	4
1.3. Organização da dissertação	4
1.4. Suporte computacional	5
2. Matemática para Kappa	6
2.1. O caso de dois observadores	6
2.2. Múltiplas avaliações por sujeito	9
2.3. Kappa ponderado	14
3. Kappas Modificados	16
3.1. Considerações iniciais	16
3.2. Kappa separado em Kappa quantidade e Kappa localização	16
3.3. Kappa separado em Kappa histograma e Kappa localização	22
3.4. Algumas desvantagens da medida Kappa propostas por Pontius e Hagen	25
3.5. Propostas de novas medidas de concordância	25
4. Comparação entre as medidas de concordância	28
4.1. Considerações iniciais	28
4.2. Primeiro estudo de simulação	28
4.3. Resultados obtidos do primeiro grupo de simulações	33
4.4. Segundo estudo de simulação	35
4.5. Resultados obtidos do segundo grupo de simulações	40
5. Estudo de comparação de mapas geográficos	70
5.1. Considerações iniciais	70

5.2. Formato dos dados	70
6. Conclusões	80
6.1. Propostas de trabalhos futuros	81
◦ Apêndice A	82
◦ Apêndice B	84
◦ Referências.....	91

Capítulo 1

Introdução

1.1. Considerações básicas sobre medidas de concordância

Em muitos casos, é bastante comum se estudar concordância entre razões de múltiplos julgamentos, experimentos, testes de diagnose, entre outros. Muitos índices de concordância são usados em taxonomia numérica para descrever o grau de similaridade entre unidades de estudos diferentes (Sneath and Sokal, 1973). Por exemplo, em estudos de correlação ou de determinantes do uso de drogas entre adolescentes, pode ser de interesse determinar como concordam as atitudes sobre o uso de drogas entre os pais ou do melhor amigo de cada sujeito, considerando sempre sujeitos do mesmo sexo. Suponhamos que se monitora, em várias estações de uma certa cidade, o nível de vários poluentes no ar (por exemplo, dióxido de enxofre). Em cada estação se controla se se excede ou não um certo nível para cada poluente oficialmente designado. Outro exemplo é o estudo do papel dos fatores familiares no desenvolvimento de uma condição, como hipertensão de adolescentes. O estudo de concordância também pode ser aplicado para descrever o grau de união na condição familiar, (Fleiss (1981)).

O foco de estudo consiste, justamente, em medidas de concordância em relação às variáveis categorizadas, do tipo *dicotômicas*, isto é, com duas categorias (por exemplo, sim ou não, presença ou ausência etc), *policotômicas*, ou seja, mais de duas categorias, *ordinais* (baixo, médio ou alto) e *nominais* (depressão, esquizofrenia etc). Há poucos consensos sobre quais métodos estatísticos é o melhor para analisar concordância. Para o não estatístico, o número de alternativas e a falta de consistência na literatura são indubitavelmente causas de dificuldades na seleção do método. Apesar das muitas opções aparentes para analisar dados de concordância, os assuntos básicos são mesmo simples. Normalmente, é melhor um ou dois métodos para uma aplicação particular. Mas é necessário identificar o propósito de análise e as perguntas substantivas claramente para serem respondidas. O engano mais comum é quando, analisando dados de concordância, não ter um objetivo explícito. Não é suficiente estarmos medindo concordância “ou” descobrindo se os observadores concordam, deve haver, presumivelmente, alguma razão por que a pessoa quer medir concordância, e a escolha de qual método estatístico é melhor, depende desta razão. Por exemplo, estudos de índices de concordância ou confiabili-

dade são usados freqüentemente para avaliar um novo sistema ou um novo instrumento. Se tal estudo está sendo administrado durante a fase de desenvolvimento do instrumento, a pessoa pode desejar analisar os dados e os métodos usados, para identificar como o instrumento poderia ser mudado para melhorar a concordância. Porém, se um instrumento já está em um formato final, os mesmos métodos poderiam não ser úteis. Enquanto faltar uma variável de critério definida ou um “padrão definido”, a precisão de um instrumento é avaliada, comparando seus resultados através de observadores diferentes. A pessoa pode distinguir amplamente duas razões para estudar concordância de avaliação. Às vezes, o objetivo é estimar a validade (precisão) das avaliações na ausência de um “padrão” definido, isto é, um uso razoável de dados de concordância: se duas avaliações discordarem, então pelo menos uma deve ser incorreta. Na própria análise de dados de concordância, está definido, em certas conclusões, sobre como é provável uma determinada avaliação estar correta. Outra razão é saber a consistência de avaliações feitas por observadores diferentes. Em alguns casos, a avaliação de precisão pode não ter nenhum significado até mesmo, por exemplo, as avaliações podem ter interesse de opiniões, atitudes, ou valores. Em outras situações, a pessoa pode estar considerando as avaliações de dois ou mais observadores para obter resultados de precisão satisfatória. Nesse caso, novamente, métodos específicos satisfatórios para este propósito devem ser usados. Um segundo problema comum de análise de concordância é o fracasso para pensar nos dados do ponto de vista teórico. Quase todos os métodos estatísticos de análise de concordância fazem suposições. Se a pessoa não pensou se dados do ponto de vista teórico satisfazem estas suposições, terá dificuldades na seleção de um método apropriado.

As variáveis podem ser categorizadas e tratadas como discretas?, ou uma característica pode ser considerada contínua por conveniência?, é razoável para assumir que uma variável normalmente é distribuída?, ou alguma outra distribuição é plausível? As respostas a essas perguntas muitas vezes não temos. O ponto principal é pensar dados aproximados para que esses sejam emparelhados com o método. Assim, os dados emparelhados com o método e considerando a teoria, são as chaves principais para uma análise próspera de dados de concordância.

O ato de efetuar uma medição baseia-se na produção de uma observação que descreva, com rigor, determinado fenômeno, de forma a que este possa ser analisado estatisticamente.

Ao realizar um trabalho, um dos propósitos que deve estar subjacente a ele, é obter amostras aleatórias mediante a execução dos mesmos procedimentos. Todavia, sabe-se que, em teoria, o que é desejável não se verifica na prática. Para tal há de se ter em conta

que determinados fatores, como o método utilizado, as características do observador e do observado, podem ter influência no resultado final. Isto obriga o observador a ter, a priori, que definir quais e de que tipo são as variáveis que vai estudar e, consoante esse estudo, saber usar o índice estatístico mais apropriado a cada uma delas. No Quadro 1.1 serão mostrados esses índices, Hulley (2001). Cada um desses índices serão apresentados no Capítulo 2.

Quadro 1.1. Índices a escolher mediante a variável em estudo

Índice	Tipo de variável
Proporção de concordância	Nominal ou Ordinal
Estatística Kappa	Nominal
Kappa ponderado	Ordinal
Coeficiente de correlação intraclasse	Contínua

O coeficiente de estatística *Kappa* (κ), que foi originalmente proposto por Jacob Cohen (1960), é uma medida de concordância utilizada para a análise de variáveis categorizadas em escalas nominais (não precisam de uma organização ordenada) que nos dá uma idéia do quanto as observações se afastam daquelas esperadas, fruto do acaso, indicando-nos assim o quão legítimas são as intepretações. Enquanto o coeficiente de estatística *Kappa Ponderado*, proposto por Jacob Cohen (1968), é usado para análise de variáveis categorizadas em escalas ordinais, admite uma ordenação. *Kappa* tem sido, ao longo dos anos, a medida preferida pelos epidemiologistas, na quantificação do grau de concordância em estudos de variabilidade observacional.

Vários trabalhos mostram que *Kappa* apresenta sérias limitações em determinadas situações que serão discutidas posteriormente. Além disso, ele não é apropriado para identificar padrões de concordância. Recentemente, diferentes abordagens que utilizam modelos estatísticos para avaliar a concordância entre os observadores foram desenvolvidas como alternativa à estatística *Kappa*. Essas abordagens consideram o problema de concordância, para respostas ordinais, de forma isolada sem uma interpretação combinada dos efeitos das distribuições marginal e conjunta sobre estruturas de concordância. Observa-se que a análise dos padrões de concordância torna-se mais rica quando é realizado um estudo combinado dos efeitos das distribuições conjunta e marginal. A magnitude da estatística *Kappa* é uma medida de concordância bastante mais significativa do que a sua própria significância estatística.

1.2. Vantagens e desvantagens do Kappa

Embora as estatísticas de Kappas tenham sido desenvolvidas, originalmente para medir concordância entre observadores, as aplicações de Kappas são estendidas além do problema específico da concordância. Na realidade, são úteis para dar uma medida de semelhança ou agrupamento de dados categorizados (Fleiss, (1981)).

Mesmo restrita para categorias não ordenadas, a medida Kappa deve ser usada não somente como uma componente a mais da estatística descritiva, mas como uma base para a inferência estatística.

O uso do Kappa, em inferência estatística, não depende do tipo de distribuição assumida, no processo subjacente, na geração das classificações.

A medida Kappa pode ser usada como um método para testar independência de observadores, ou seja, como uma estatística de teste, embora se saiba que observadores não são independentes, já que eles estão avaliando os mesmos casos.

Também Kappa pode ser usado como um método para quantificar o nível de concordância, isto é, como uma medida de efeito-tamanho.

Quando usado para medir concordância, ou, mais geralmente, Kappa trata todos os observadores ou as unidades simetricamente (Fleiss, (1981)).

No caso do Kappa ponderado, sua principal vantagem reside na possibilidade de quantificar diferentes graus de discordância.

Como já foi mencionado, Kappa não é apropriado para identificar padrões de concordância.

A medida Kappa usual não faz distinções entre vários tipos e fontes de discordância. Neste trabalho é realizado um estudo de certas medidas Kappas para a análise de concordância entre mapas categorizados ou temáticos. Essas medidas consideram duas possíveis fontes de discordância: a localização e a quantidade.

1.3. Organização da dissertação

Neste capítulo apresentam-se considerações gerais sobre medidas de concordância bem como os índices a escolher mediante a variável em estudo. Também citamos as vantagens e desvantagens do coeficiente de concordância Kappa.

No capítulo 2, faz-se uma revisão dos diversos aspectos de Kappas, o caso de dois observadores, múltiplos observadores, e o Kappa ponderado. No capítulo 3, apresentam-se as medidas Kappas propostas por Pontius (2000) e Hagen (2002) e se propõem modificações destas medidas. No capítulo 4, são discutidos dois estudos de simulações com

mapas hipotéticos com a finalidade de compararmos as medidas de concordância Kappa. No capítulo 5, realiza-se uma aplicação de Kappas na comparação de mapas geográficos para análise da evolução de desflorestamento. Finalmente, apresentam-se as conclusões gerais dos resultados e propostas de trabalhos futuros no capítulo 6.

1.4. Suporte computacional

Para elaboração do texto desta dissertação, foi usada a linguagem de tipografia (Plain) $\text{\TeX}$, dada sua flexibilidade, sendo esta linguagem especialmente útil para a confecção de documentos que contêm notação e símbolos matemáticos, criada por Donald Knuth. Tipografia científica utilizando $\text{\TeX}$ (ou seu variante $\text{\LaTeX}$) está tornando-se padrão no meio científico, onde a utilização de editores de texto como o **Word**, por exemplo, vem caindo em desuso. Uma implementação gratuita e de boa qualidade do sistema $\text{\TeX}$ para o sistema operacional **Windows** pode ser obtida no endereço <http://www.miktex.org>. Para maiores detalhes, (Knuth (1986)).

Um suporte computacional utilizado no desenvolvimento desta dissertação de mestrado é a linguagem de programação interpretada **R**, e é bastante similar à linguagem **S** e sua versão comercial melhorada **S-PLUS**. Ele possui uma documentação detalhada, tanto em forma de manuais como documentos **on-line** disponíveis em <http://www.r-project.org>. Uma de suas características mais atrativas é o fato de que é um **software** gratuito e, portanto, pode ser obtido e distribuído sem custos. Atualmente **R** está na versão 1.8 e pode ser compilado (**R** é escrito em **ANSI C**) para diferentes sistemas operacionais.

Outro suporte utilizado é o **MATLAB**¹, que pode ser usado como uma linguagem de programação ou como uma ferramenta de cálculo interativa. Em ambos os casos, o ambiente **MATLAB** permite realização de cálculos, visualização de resultados e desenvolvimento de algoritmos usando uma sintaxe muito próxima da notação matemática padrão.

Ele é desenvolvido pela **The MathWorks, Inc.** em ambiente **Unix**, mas pode ser utilizado em ambiente **Windows**. Está disponível em <http://www.mathworks.com>. Uma das grandes virtudes do **MATLAB** é que todas suas funções são escritas em **m-files**, sendo que a maioria delas podem ser editadas/modificadas.

¹ O nome **MATLAB** vem de **MATrix LABoratory** e representa bem o seu modo de funcionamento. Todas as variáveis são definidas de forma matricial, o que faz com que este ambiente seja ideal para resolver problemas de cálculo envolvendo matrizes e vetores.

Capítulo 2

Matemática para Kappa

Neste capítulo, será apresentada uma revisão dos vários aspectos da estatística Kappa que nos introduziram nesse tipo de medidas.

2.1. O caso de dois observadores

Suponha que dois observadores distintos classificam, independentemente, a mesma amostra de n itens, segundo uma variável nominal com g categorias. O resultado desta classificação pode ser resumido em uma tabela de contingência como a mostrada no Quadro 2.1, em que cada valor x_{ij} representa o número de itens que tenha sido classificado pelo observador 1 na categoria i e pelo observador 2 na categoria j , (Altman, (1991)).

Quadro 2.1. Tabela de contingência

observador 1	observador 2				
	1	2	...	g	Total
1	x_{11}	x_{12}	...	x_{1g}	$x_{1.}$
2	x_{21}	x_{22}	...	x_{2g}	$x_{2.}$
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
g	x_{g1}	x_{g2}	.	x_{gg}	$x_{g.}$
Total	$x_{.1}$	$x_{.2}$	.	$x_{.g}$	n

Kappa é calculado a partir das frequências observadas e esperadas na diagonal principal da tabela quadrada de frequências. Se houver n observações em g categorias, então a *proporção de concordância observada* (P_o) é dada por

$$P_o = \sum_{i=1}^g x_{ii}/n, \quad (2.1)$$

em que x_{ii} é o número de concordância para a categoria i . Notemos que $0 \leq P_o \leq 1$. Se $P_o = 1$ implica que não existe diferença entre a classificação dos observadores. A *proporção esperada de concordância* (P_e) é determinada por

$$P_e = \sum_{i=1}^g x_{i.} x_{.i} / n^2, \quad (2.2)$$

em que $x_{i.}$ e $x_{.i}$ são, respectivamente, as frequências totais fila e coluna para a i -ésima categoria. O **coeficiente de concordância**, Kappa (κ), é determinado através do quociente da diferença entre a proporção de concordância observada e a proporção de concordância esperada, e a diferença entre a proporção absoluta e a proporção de concordância esperada (a maior diferença possível entre a concordância observada e a esperada),

$$\kappa = \frac{P_o - P_e}{1 - P_e}. \quad (2.3)$$

Quando o valor da estatística (κ) é “0”, representa não haver concordância, e quando é “1” representa haver concordância perfeita. Se todos os elementos da tabela de contingência (ou matriz de confusão) que não pertencem à diagonal principal possuem valor nulo, o coeficiente Kappa possui valor 1.

As diretrizes para a interpretação (sempre subjetiva) de *Kappa* são dadas no Quadro 2.2, (ligeiramente adaptado de *Landis & Koch* , 1977).

Quadro 2.2. Interpretação do índice

Valor de κ	Intensidade da concordância
< 0.20	pobre
0.21 – 0.40	fraca
0.41 – 0.60	moderada
0.61 – 0.80	boa
0.81 – 1.00	muito boa

O erro padrão aproximado de κ , (Altman(1991)) $ep(\hat{\kappa})$, é obtido por

$$ep(\hat{\kappa}) = \sqrt{\frac{P_o(1 - P_o)}{n(1 - P_e)^2}}. \quad (2.4)$$

Por sua vez, o erro padrão da estatística κ permite-nos estimar o intervalo de confiança a nível $(1 - \alpha)$. No caso em que $(1 - \alpha) = 0.95$, o intervalo de confiança é determinado por

$$\kappa - 1.96 \text{ ep}(\hat{\kappa}) \quad \text{para} \quad \kappa + 1.96 \text{ ep}(\hat{\kappa}).$$

No entanto, o valor de κ depende da proporção de sujeitos em cada categoria, que denominaremos de *prevalência*. Uma grande prevalência resulta num alto nível de concordância esperada pelo acaso, o que resultará num valor de κ mais baixo. Por outro lado, uma baixa prevalência dará origem a valores de κ mais altos.

O *Coefficiente de Correlação Intraclass* (Intraclass Correlation Coefficient) ou coeficiente de reprodutibilidade (r) é uma estimativa da fração da variabilidade total das medidas devido à variação entre os sujeitos. O valor de (r) é obtido dividindo-se a variação entre os sujeitos pela variação total, a qual inclui a variação entre os sujeitos e a variação indesejada (“erros”).

A variação indesejada poderá incluir diferentes componentes dependendo do desenho de estudo utilizado. Por exemplo, num estudo de reprodutibilidade relacionado com um método laboratorial, incluiria a variabilidade provocada pelos os erros de método ou procedimento.

O Coeficiente de Correlação Intraclass (Cochran (1965)) é definido pela fórmula,

$$r = \frac{E[(y_{ij} - \bar{y}_{g.})(y_{.g} - \bar{y}_{g.})]}{E(y_{ij} - \bar{y}_{g.})^2}, \quad (2.5)$$

em que o numerador representa a média de todos os $kn(n - 1)/2$ pares diferentes de unidades e o denominador é a média de todos os N valores de y_{ij} . Em que o símbolo y_{ij} representa o valor de ordem j da amostra sistemática de ordem i , de modo que $j = 1, 2, \dots, n$, $i = 1, 2, \dots, k$; $N = nk$, k é o número de amostras sistemáticas possíveis.

O coeficiente de correlação intraclass é equivalente à estatística κ para as variáveis contínuas e tem a mesma amplitude de valores possíveis (de 0 a 1, em caso de concordância perfeita). Tem uma vantagem em relação aos índices anteriores, na medida em que **é uma verdadeira medida de concordância, combinando a informação resultante das diferenças de correlação**. Quando (r) é igual a zero (0), o estudo não é reprodutível (ou seja, há uma grande variabilidade dentro dos sujeitos mas não há variabilidade entre os sujeitos), quando o (r) é igual a 1, o estudo é reprodutível (ou seja, não há variabilidade dentro dos sujeitos mas há uma grande variabilidade entre os sujeitos). Tal como a estatística Kappa, o Coeficiente de Correlação Intraclass pode também ser estendido para o cálculo da reprodutibilidade entre mais de dois observadores, (Szklo, R.(2000)).

Quadro 2.3. Interpretação do índice

Valor de (r)	Intensidade da reprodutibilidade
$(r) < 0.4$	pobre
$0.4 \leq (r) < 0.75$	satisfatória
$(r) \geq 0.75$	excelente

A interpretação dos valores de (r) é mostrada no Quadro 2.3

2.2. Múltiplas avaliações por sujeitos

Suponha que uma amostra de n sujeitos tenha sido estudada com m_i , sendo o número de avaliações do i -ésimo sujeito. Não é assumido que o(s) avaliador(es) responsável(eis) pela classificação de um sujeito não é (são) o(s) mesmo(os) necessariamente responsável pela classificação de outro sujeito. Suponha $g=2$, então a avaliação consistirá na classificação em uma de duas categorias; o caso $g > 2$ será considerado posteriormente. Denotemos x_i o número de avaliações positivas do sujeito i , e $m_i - x_i$, seja o número de avaliações negativas deste (Fleiss, (1981)).

A proporção total para avaliações positivas é definida como

$$\bar{p} = \frac{\sum_{i=1}^n x_i}{n\bar{m}},$$

em que

$$\bar{m} = \frac{\sum_{i=1}^n m_i}{n}$$

é o número médio de avaliações por sujeito. Se o número de sujeitos n , é tal que $n \geq 20$, a média quadrática entre sujeitos (MQES) é aproximadamente igual a

$$\text{MQES} = \frac{1}{n} \sum_{i=1}^n \frac{(x_i - m_i \bar{p})^2}{m_i}, \quad (2.6)$$

A MQES nos dá uma idéia da distância entre o número de avaliações positivas de cada sujeito em relação à média de avaliações positivas do grupo. A média quadrática dentro do sujeito (MQDS) definida por

$$\text{MQDS} = \frac{1}{n(\bar{m} - 1)} \sum_{i=1}^n \frac{x_i(m_i - x_i)}{m_i}, \quad (2.7)$$

dá uma idéia da variação das avaliações de cada sujeito, ou seja, uma espécie de média ponderada das variâncias de todos os sujeitos.

Tecnicamente, o coeficiente de correlação intraclasse deve ser estimado por

$$r = \frac{\text{MQES} - \text{MQDS}}{\text{MQES} + (m_0 - 1)\text{MQDS}}, \quad (2.8)$$

em que

$$m_0 = \bar{m} - \frac{\sum_{i=1}^n (m_i - \bar{m})^2}{n(n - 1)\bar{m}}.$$

Se n é muito grande, m_0 e $\bar{m}$ estarão muito próximos em magnitude. Se m_0 é substituído por $\bar{m}$ na equação (2.8), o resultado da expressão para o coeficiente de correlação intraclasse, e entretanto para Kappa, é

$$\hat{\kappa} = \frac{\text{MQES} - \text{MQDS}}{\text{MQES} + (\bar{m} - 1)\text{MQDS}}$$

então

$$\hat{\kappa} = 1 - \frac{\sum_{i=1}^n \frac{x_i(m_i - x_i)}{m_i}}{n(\bar{m} - 1)\bar{p}\bar{q}}, \quad (2.9)$$

em que $\bar{q} = 1 - \bar{p}$.

Assim $\hat{\kappa}$ tem as seguintes propriedades:

(P.1) Se não há nenhuma variação de sujeito para sujeito na proporção de avaliações positivas (isto é, se $\frac{x_i}{m_i} = \bar{p}$ para todo i , com $\bar{p} \neq 0$ ou $\bar{p} \neq 1$), então há mais discordância dentro das avaliações de cada sujeito do que propriamente entre os sujeitos. Neste caso, $\hat{\kappa}$ pode ser visto para assumir seu valor mínimo igual a $\frac{-1}{(\bar{m}-1)}$.

Prova: Note que $\frac{x_i}{m_i} = \bar{p} \Rightarrow x_i = \bar{p}m_i$, logo pela equação (2.9), temos

$$\begin{aligned} \hat{\kappa} &= 1 - \frac{\bar{p} \sum_{i=1}^n (m_i - x_i)}{n(\bar{m} - 1)\bar{p}\bar{q}} \\ &= 1 - \frac{\sum_{i=1}^n (m_i - m_i\bar{p})}{n(\bar{m} - 1)\bar{q}} \\ &= 1 - \frac{\sum_{i=1}^n m_i(1 - \bar{p})}{n(\bar{m} - 1)\bar{q}} \\ &= 1 - \frac{\bar{q} \sum_{i=1}^n m_i}{(\bar{m} - 1)\bar{q}n} \\ &= 1 - \frac{\bar{m}}{\bar{m} - 1} \end{aligned}$$

Logo,

$$\hat{\kappa} = \frac{-1}{(\bar{m} - 1)}.$$

Se cada proporção $\frac{x_i}{m_i}$ assume unicamente os valores 0 ou 1, então há concordância perfeita dentro dos sujeitos. Neste caso, $\hat{\kappa}$ pode assumir o valor 1.

(P.2) *Se a maioria das proporções $\frac{x_i}{m_i}$ variam exatamente como a proporção binomial, com m_i observações e com uma probabilidade comum $\bar{p}$. Então há tanta semelhança dentro das avaliações dos sujeitos como entre os sujeitos. Neste caso, o valor de $\hat{\kappa}$ é igual ao número 0.*

Prova: Considere que a variável X_i é o número de avaliações positivas com distribuição Binomial $(n, \bar{p})$, então, as esperanças de X_i e X_i^2 são, respectivamente, dadas por:

$E(X_i) = m_i \bar{p}$ e $E(X_i^2) = m_i \bar{p} \bar{q}$, daí temos que a esperança da média quadrática entre os sujeitos (2.6) é

$$\begin{aligned} E(\text{MQES}) &= \frac{1}{n} \sum_{i=1}^n E[(X_i - m_i \bar{p})^2] \\ &= \frac{1}{n} \sum_{i=1}^n \frac{1}{m_i} m_i \bar{p} \bar{q} \end{aligned}$$

Logo,

$$E(\text{MQES}) = \bar{p} \bar{q}, \tag{2.10}$$

e a esperança da média quadrática dentro dos sujeitos, (2.7) é

$$\begin{aligned}
 E(\text{MQDS}) &= \frac{1}{n(\bar{m} - 1)} \sum_{i=1}^n \frac{E[X_i(m_i - X_i)]}{m_i} \\
 &= \frac{1}{n(\bar{m} - 1)} \sum_{i=1}^n \left[E(X_i) - \frac{E(X_i^2)}{m_i} \right] \\
 &= \frac{1}{m - 1} \left[\sum_{i=1}^n m_i \bar{p} - \sum_{i=1}^n (\bar{p} \bar{q} + \bar{p}^2) \right] \\
 &= \frac{1}{m - 1} \left[\bar{p} \sum_{i=1}^n m_i - \bar{p}^2 - n \bar{p} \bar{q} \right] \\
 &= \frac{1}{n(\bar{m} - 1)} \left[\bar{p} \sum_{i=1}^n m_i - \bar{p}^2 \sum_{i=1}^n m_i - n \bar{p} \bar{q} \right] \\
 &= \frac{1}{n(\bar{m} - 1)} \left[\bar{p} \bar{m} n(1 - \bar{p}) - n \bar{p} \bar{q} \right] \\
 &= \frac{1}{n(\bar{m} - 1)} \left[\bar{m} \bar{p} n \bar{q} - n \bar{p} \bar{q} \right] \\
 &= \frac{1}{n(\bar{m} - 1)} n \bar{p} \bar{q} (\bar{m} - 1)
 \end{aligned}$$

Assim,

$$E(\text{MQDS}) = \bar{p} \bar{q}. \quad (2.11)$$

Portanto (2.10) coincide com (2.11).

Fleiss & Cuzick (1979) obtiveram o erro padrão para $\hat{\kappa}$, apropriado para testar a hipótese que o valor subjacente de Kappa é zero.

Define-se $\bar{m}_h$ para a média harmônica do número de avaliações por sujeito, isto é ;

$$\bar{m}_h = \frac{n}{\sum_{i=1}^n \frac{1}{m_i}}.$$

O erro padrão para $\hat{\kappa}$ é aproximadamente igual a

$$\text{ep}_o(\hat{\kappa}) = \frac{1}{(\bar{m} - 1)\sqrt{n\bar{m}_h}} \sqrt{2(\bar{m}_h - 1) + \frac{(\bar{m} - \bar{m}_h)(1 - 4\bar{p} \bar{q})}{\bar{m} \bar{p} \bar{q}}},$$

e a hipótese de Kappa nulo pode ser testada comparando o valor de

$$z = \frac{\hat{\kappa}}{\text{ep}_o(\hat{\kappa})}$$

em relação ao valor crítico determinado a partir da distribuição normal padrão.

Considere, agora, o número de categorias nas quais são feitas as avaliações em que $g \geq 2$. Denotamos por $\bar{p}_j$ a proporção total de avaliações na categoria j , por $\hat{\kappa}_j$ o erro padrão, que ainda tem que ser calculado, quando os números de avaliações por sujeito variam, para testar a hipótese em que o valor subjacente é igual a zero e por $\hat{\kappa}_j$ o valor de kappa para a categoria j , $j = 1, \dots, g$. Landis & Koch (1977) propuseram considerar a média ponderada

$$\hat{\kappa} = \frac{\sum_{j=1}^g \bar{p}_j \bar{q}_j \hat{\kappa}_j}{\sum_{j=1}^g \bar{p}_j \bar{q}_j}, \quad (2.12)$$

como uma medida geral de concordância ($\bar{q}_j = 1 - \bar{p}_j$). Quando o número de avaliações por sujeito é constante e igual a m , expressões simples para $\hat{\kappa}_j$, $\hat{\kappa}$, e erros padrões respectivos, estão disponíveis. Define-se x_{ij} para o número de avaliações no sujeito i ($i=1, \dots, n$) na categoria j ($j=1, \dots, g$); note que

$$\sum_{j=1}^g x_{ij} = m,$$

para todo i . Então o valor de $\hat{\kappa}_j$, é dado por

$$\hat{\kappa}_j = 1 - \frac{\sum_{i=1}^n x_{ij}(m - x_{ij})}{nm(m-1)\bar{p}_j \bar{q}_j},$$

e o valor para $\hat{\kappa}$ é

$$\hat{\kappa} = 1 - \frac{nm^2 - \sum_{i=1}^n \sum_{j=1}^g x_{ij}^2}{nm(m-1)\sum_{j=1}^g \bar{p}_j \bar{q}_j}.$$

Algebricamente, versões equivalentes dessas fórmulas foram apresentadas primeiramente por Fleiss (1971), que mostrou explicitamente como elas representam medidas de chances corrigidas para concordância.

Quando os números de avaliações por sujeito são iguais, Fleiss, Nee, e Landis (1979) derivaram e confirmaram as fórmulas seguintes para aproximar erros padrões de $\hat{\kappa}$ e $\hat{\kappa}_j$, cada um apropriado para testar a hipótese cujo valor subjacente é zero:

$$ep_o(\hat{\kappa}) = \frac{\sqrt{2}}{\sum_{j=1}^g \bar{p}_j \bar{q}_j \sqrt{nm(m-1)}} \times \sqrt{\left(\sum_{j=1}^g \bar{p}_j \bar{q}_j\right)^2 - \sum_{j=1}^g \bar{p}_j \bar{q}_j (\bar{q}_j - \bar{p}_j)},$$

e

$$ep_o(\hat{\kappa}_j) = \sqrt{\frac{2}{nm(m-1)}}.$$

Note que $ep_o(\hat{\kappa}_j)$ é independente de $\bar{p}_j$ e $\bar{q}_j$

2.3. Kappa ponderado

O *Kappa ponderado* é um índice estatístico de concordância, utilizado quando as variáveis são ordinais, ou seja, que possuem uma ordenação segundo algum tipo de intensidade. É a este nível que o *Kappa Ponderado* difere da Estatística *Kappa*. Ao serem utilizadas três ou mais categorias para se expressar os resultados, certos tipos de discordância podem tomar valores maiores que outros. Os valores e os pesos dependem, em cada caso, da importância que se conceda a cada discordância. Na seção anterior, vimos que os pesos propostos para o cálculo de Kappa (2.12) são determinados pela proporção total de avaliações em cada categoria. Se, em dois estudos diferentes, se calcula o Kappa ponderado com dois sistemas de pesos distintos, é difícil realizar comparação entre eles. Os mais usados, denominados *pesos lineares*, proposto inicialmente por Cohen (1968). Para a cela na linha i e coluna j , um peso, w_{ij} , é calculado como

$$w_{ij} = 1 - \frac{|i - j|}{g - 1},$$

e os denominados por pesos biquadrados, proposto posteriormente por Fleiss & Cohen (1973), calculado como

$$w_{ij} = 1 - \frac{(i - j)^2}{(g - 1)^2}.$$

Este índice confere, assim, uma maior importância à discordância, atribuindo-lhe maior peso quando se procede ao cálculo.

O *Kappa Ponderado* é obtido dando pesos às frequências em cada cela da tabela de contingência com a distância da diagonal principal, que indica a concordância.

Assim que se determina as celas na diagonal principal, com peso 1, enquanto essas em que a diferença está através de uma categoria, adquirem um peso de $1 - \frac{1}{(g-1)}$.

A *proporção observada ponderada* $P_{o(w)}$ e a *proporção de concordância esperada ponderada* $P_{e(w)}$ são obtidas respectivamente por

$$P_{o(w)} = \frac{1}{n} \sum_{i=1}^g \sum_{j=1}^g w_{ij} x_{ij},$$

e

$$P_{e(w)} = \frac{1}{n^2} \sum_{i=1}^g \sum_{j=1}^g w_{ij} x_{i.x.j},$$

e o *Kappa Ponderado* κ_w é determinado por

$$\kappa_w = \frac{P_{o(w)} - P_{e(w)}}{1 - P_{e(w)}}. \quad (2.13)$$

Capítulo 3

Kappas modificados

3.1. Considerações iniciais

Robert Gil Pontius Jr. (2000) apresenta novos métodos estatísticos generalizados para medir concordância entre dois mapas geográficos. Essa metodologia quantifica concordância entre dois mapas categóricos, em várias resoluções. Em cada resolução, o método divide todas as concordâncias devidas à chance, em correções devidas à quantidade, correções devidas à localização, erro devido à quantidade e erro devido à localização. Em síntese, o método calcula três estatísticas que são usadas para interpretar a diferença entre mapas, e mostra como essas estatísticas mudam com a resolução.

Outra proposta da utilização de Kappa para analisar similaridades entre mapas foi dada por Hagen, em (2002), e serão também apresentadas neste capítulo. Nós propomos novas medidas de concordância a partir da proposta de Pontius, para corrigir algumas desvantagens encontradas nas medidas propostas originalmente.

3.2. Kappa separado em Kappa quantidade e Kappa localização

Pontius (2002) verifica que, se se considera a estatística Kappa usual, confunde-se concordância em quantidade com concordância de localização. Neste sentido, quantidade mede a presença total, como uma fração de todas as celas de uma categoria sobre o mapa inteiro. Com “localização” referimos a localização espacial da quantidade sobre o mapa. Pontius introduz duas estatísticas que consideram separadamente **concordância de localização** e **concordância de quantidade**, que serão definidas proximamente.

A resolução de um mapa reflete o poder de discriminação entre dois pontos da realidade. Por exemplo, em mapas de alta resolução duas regiões muito próximas podem ser distinguidas, no entanto em mapas de baixa resolução essas regiões podem não ser distinguidas.

Como exemplo, tomaremos a definição dada por Pontius (2000). Suponhamos que temos duas categorias representadas por 0 e 1 em que 16 celas estão agregadas numa resolução 1×1 (resolução fina),

Resolução 1×1

0	0	0	1
0	1	1	1
1	0	1	1
0	1	1	1

Calculando a média das quatro celas adjacentes na resolução fina acima, obtemos uma cela de resolução 2×2 (resolução grosseira), que mostra o resultado parcial na categoria, ou seja,

Resolução 2×2

0.25	0.75
0.50	1.00

Esta foi a definição de resolução considerada neste trabalho.

Considere dois mapas R e S . Sejam $R_{n,i}$ a proporção da categoria i na cela n do mapa R , $S_{n,i}$ a proporção da categoria i na cela n do mapa S , i a categoria, n a cela, I o número de categorias, N_g o número de celas no mapa de resolução g , logo

$$0 \leq R_{n,i} \leq 1 \quad (3.1)$$

$$0 \leq S_{n,i} \leq 1 \quad (3.2)$$

$$\sum_{i=1}^I R_{n,i} = \sum_{i=1}^I S_{n,i} = 1 \quad (3.3)$$

$$\sum_{n=1}^{N_g} \sum_{i=1}^I R_{n,i} = \sum_{n=1}^{N_g} \sum_{i=1}^I S_{n,i} = N_g \quad (3.4)$$

A estatística para concordância em termos de localização é denotada por κ_{loc} definida por

$$\kappa_{loc} = \frac{M_m - N_m}{P_m - N_m}, \quad (3.5)$$

em que M_m é a **proporção de concordância observada com informações médias de localização e quantidade** (sempre representa a porcentagem de coincidência entre os mapas R e S), é dada por

$$M_m = \frac{\sum_{n=1}^{N_g} [W_n \sum_{i=1}^I \min(R_{n,i}, S_{n,i})]}{\sum_{n=1}^{N_g} W_n}, \quad (3.6)$$

em que W_n são os pesos de cada cela segundo sua importância na análise, N_m é a **proporção de concordância esperada sem informação de localização e com informação média de quantidade**, definida por

$$N_m = \frac{\sum_{n=1}^{N_g} [W_n \sum_{i=1}^I \min(R_{n,i}, S_{.,i})]}{\sum_{n=1}^{N_g} W_n}, \quad (3.7)$$

e P_m é a **proporção de concordância com informações perfeita de localização e média de quantidade**, definida por

$$P_m = \sum_{i=1}^I \min(R_{.,i}, S_{.,i}), \quad (3.8)$$

em que $R_{.,i}$ é a proporção da categoria i no mapa R , definida por

$$R_{.,i} = \frac{\sum_{n=1}^{N_g} [W_n R_{n,i}]}{\sum_{n=1}^{N_g} W_n}, \quad (3.9)$$

e $S_{.,i}$ é a proporção da categoria i no mapa S , definida por

$$S_{.,i} = \frac{\sum_{n=1}^{N_g} [W_n S_{n,i}]}{\sum_{n=1}^{N_g} W_n}. \quad (3.10)$$

Define-se a medida de não informação em relação à localização e quantidade, como

$$N_n = \frac{\sum_{n=1}^{N_g} [W_n \sum_{i=1}^I \min(R_{n,i}, 1/I)]}{\sum_{n=1}^{N_g} W_n}, \quad (3.11)$$

a medida de informação perfeita de quantidade mas de não informação de localização, como

$$N_p = \frac{\sum_{n=1}^{N_g} [W_n \sum_{i=1}^I \min(R_{n,i}, R_{.,i})]}{\sum_{n=1}^{N_g} W_n}, \quad (3.12)$$

Define-se, também, M_n é a **proporção de concordância esperada com informação média de localização e sem informação de quantidade**, definida por

$$M_n = N_n \times (1 - \kappa_{\text{loc}}) + \kappa_{\text{loc}} \times \sum_{i=1}^I \min(R_{.,i}, 1/I), \quad (3.13)$$

e M_p é a **proporção observada de concordância com informações média de localização e perfeita de quantidade**, definida por

$$M_p = N_p \times (1 - \kappa_{\text{loc}}) + \kappa_{\text{loc}}, \quad (3.14)$$

Pontius (2002) define também as estatísticas para concordância em termos de quantidade κ_{quant} e Kappa κ_p , respectivamente, por

$$\kappa_{\text{quant}} = \frac{M_m - M_n}{M_p - M_n}, \quad (3.15)$$

$$\kappa_p = \frac{M_m - N_m}{P_p - N_m}, \quad (3.16)$$

em que M_m está definida em (3.6), N_m é definida em (3.7) e P_p é a **proporção de concordância observada com informações perfeitas de localização e quantidade**, ou seja, $P_p = 1$.

Note a semelhaça das expressões do Kappa para localização (3.5) e para quantidade (3.15) com o Kappa usual (2.3) e o Kappa ponderado (2.13).

Para uma melhor compreensão, considere como exemplo dois mapas hipotéticos, mapa R e mapa S. Em cada mapa, há nove celas e duas categorias. As categorias são denotadas por 0, representando a cor branca, e 1, a cor preta. Assim $N_g = 9$, $i = 0, 1$ e $I = 2$

Mapa R (resolução 1×1)

0	0	0	0	0	0
0	0	1	0	0	0
0	1	0	0	1	0
0	1	0	0	0	0
0	0	1	1	0	0
0	0	0	0	0	0

Assim, temos a proporção da categoria i na cela n no **Mapa R**

$$n = 1 \Rightarrow R_{1,0} = 1, R_{1,1} = 0$$

$$n = 2 \Rightarrow R_{2,0} = 0.75, R_{2,1} = 0.25$$

$$n = 3 \Rightarrow R_{3,0} = 1, R_{3,1} = 0$$

$$n = 4 \Rightarrow R_{4,0} = 0.5, R_{4,1} = 0.5$$

$$n = 5 \Rightarrow R_{5,0} = 1, R_{5,1} = 0$$

$$n = 6 \Rightarrow R_{6,0} = 0.75, R_{6,1} = 0.25$$

$$n = 7 \Rightarrow R_{7,0} = 1, R_{7,1} = 0$$

$$n = 8 \Rightarrow R_{8,0} = 0.5, R_{8,1} = 0.5$$

$$n = 9 \Rightarrow R_{9,0} = 1, R_{8,1} = 0$$

Mapa S (resolução 1×1)

0	0	0	0	0	0
0	0	1	1	0	0
1	1	1	1	0	0
0	0	1	1	1	1
0	0	0	1	0	0
0	0	0	0	0	0

e a proporção da categoria i na cela n no **Mapa S**

$$n = 1 \Rightarrow S_{1,0} = 1, S_{1,1} = 0$$

$$n = 2 \Rightarrow S_{2,0} = 0.5, S_{2,1} = 0.5$$

$$n = 3 \Rightarrow S_{3,0} = 1, S_{3,1} = 0$$

$$n = 4 \Rightarrow S_{4,0} = 0.5, S_{4,1} = 0.5$$

$$n = 5 \Rightarrow S_{5,0} = 0, S_{5,1} = 1$$

$$n = 6 \Rightarrow S_{6,0} = 0.5, S_{6,1} = 0.5$$

$$n = 7 \Rightarrow S_{7,0} = 1, S_{7,1} = 0$$

$$n = 8 \Rightarrow S_{8,0} = 0.75, S_{8,1} = 0.25$$

$$n = 9 \Rightarrow S_{9,0} = 1, S_{8,1} = 0$$

Cálculo dos coeficientes de Kappa segundo Pontius

Calculemos M_m definido em (3.6), considerando todos os pesos iguais a $W_n = 1$,

note que,

$$\sum_{n=1}^9 [1 \sum_{i=0}^1 \min(R_{n,i}, S_{n,i})] = \min(R_{1,0}, S_{1,0}) + \min(R_{1,1}, S_{1,1}) + \dots + \min(R_{9,1}, S_{9,1}) = 7.25,$$

$$\text{logo } M_m = \frac{7.25}{9} = 0.8055.$$

Por (3.10), temos

$$S_{.,i} = \frac{\sum_{n=1}^9 [1 S_{n,i}]}{\sum_{n=1}^9 1} = \sum_{n=1}^9 S_{n,i}/9$$

$$S_{.,0} = 6.25/9 = 0.6944 \text{ e } S_{.,1} = 2.75/9 = 0.3055$$

Então

$$\begin{aligned} \sum_{n=1}^9 [1 \sum_{i=0}^1 \min(R_{n,i}, \sum_{n=1}^9 S_{n,i}/9)] &= \min(R_{1,0}, 0.6944) + \min(R_{1,1}, 0.3055) \\ &+ \dots + \min(R_{9,0}, 0.6944) + \min(R_{9,1}, 0.3055) \\ &= 6.9722 \end{aligned}$$

$$\text{logo } N_m = \frac{6.9722}{9} = 0.7746$$

Finalmente aplicando a equação (3.16) , temos que $\kappa_p = \frac{0.8055-0.7746}{1-0.7746} = 0.1370$.

Cálculo do Kappa para localização

Pelas equações (3.9) e (3.10) temos que,

$$\begin{aligned} R_{.,i} &= \frac{\sum_{n=1}^9 [1 R_{n,i}]}{\sum_{n=1}^9 1} = \sum_{n=1}^9 R_{n,i}/9, \\ S_{.,i} &= \frac{\sum_{n=1}^9 [1 S_{n,i}]}{\sum_{n=1}^9 1} = \sum_{n=1}^9 S_{n,i}/9, \end{aligned}$$

$$R_{.,0} = 7.5/9, R_{.,1} = 1.5/9, S_{.,0} = 6.25/9 \text{ e } S_{.,1} = 2.75/9.$$

Daí,

$$P_m = \min(7.5/9, 6.25/9) + \min(1.5/9, 2.75/9)$$

logo $P_m = \frac{7.75}{9} = 0.8611$, aplicando a equação (3.5) temos que $\kappa_{loc} = 0.3572$

Cálculo do Kappa para quantidade

Por (3.11) e (3.12) temos,

$$N_n = 1/9[\min(R_{1,0}, 0.5) + \dots + \min(R_{9,1}, 0.5)]$$

Daí $N_n = 6/9 = 0.6666$

e

$$N_p = 1/9[\min(R_{1,0}, \sum_{n=1}^9 R_{n,0}/9) + \dots + \min(R_{9,1}, \sum_{n=1}^9 R_{n,1}/9)]$$

Daí $N_p = 1/9[66/9] = 66/81 = 0.8148$

Calculando obtemos que $P_n = 2/3$, $M_n = 0.6666$ e $M_p = 0.8809$,

aplicando a equação (3.15), temos que $\kappa_{quant} = 0.6481$.

3.3. Kappa separado em Kappa Histograma e Kappa localização

No seu trabalho de 2002, Hagen chegou às seguintes conclusões:

- 1) Não é claro porque a formulação do κ_{quant} envolve o κ_{loc} .
- 2) A gama de valores não é a mesma que é utilizada para Kappa padrão; κ_{quant} pode ser maior do que 1.
- 3) κ_{quant} não é estável. Uma diferença mínima nos mapas pode conduzir a uma diferença significativa na estatística.

Assim, este autor propôs uma nova estatística para concordância de localização, κ_{locH} que compara a proporção de concordância observada com a proporção de concordância esperada, relativa à proporção máxima de concordância observada, quando o

número total de pixels por categoria não se altera. κ_{locH} é calculado pela equação (3.17) e independe do número total de celas em cada categoria.

$$\kappa_{locH} = \frac{P_o - P_e}{P_{max} - P_e}, \quad (3.17)$$

em que

$$P_{max} = \sum_{i=1}^g \min(x_i, x_{.i}). \quad (3.18)$$

P_{max} denota máxima fração de concordância segundo a distribuição observada, P_e representa a fração de concordância esperada segundo a distribuição observada, p_{ij} denota a porcentagem da área do mapa na categoria i do mapa 1 e na categoria j do mapa 2 e x_{ij} representa o número de coincidências que tenha sido classificada no mapa 1 na categoria i e no mapa 2 pela categoria j ,

$$P_e = \sum_{i=1}^g p_{i.i}.$$

Hagen (2002), então, propôs uma alternativa para a concordância dos resultados em modelos quantitativos, baseada no número total de celas dentro de cada categoria, e que pode ser relacionada diretamente com Kappa. A estatística resultante é chamada de κ_{histo} definida pela equação

$$\kappa_{histo} = \frac{P_{max} - P_e}{1 - P_e} \quad (3.19)$$

A definição de κ_{histo} tem uma propriedade poderosa de que Kappa agora está definido como um produto de dois fatores. O primeiro fator é κ_{locH} , que é uma medida para a **concordância da localização espacial** das categorias dos dois mapas comparados. O segundo fator é κ_{histo} , que é uma medida de **concordância quantitativa** dos dois mapas comparados. Assim, Hagen define a estatística κ_H

$$\begin{aligned} \kappa_H &= \kappa_{locH} \times \kappa_{histo} \\ &= \frac{P_o - P_e}{P_{max} - P_e} \times \frac{P_{max} - P_e}{1 - P_e}, \end{aligned}$$

daí,

$$\kappa_H = \frac{P_o - P_e}{1 - P_e}. \quad (3.20)$$

Além de calcular a estatística κ para todas categorias combinadas, há a opção para calcular a estatística κ por categoria. Para uma estatística κ categórica, os dois mapas

são transformados em um mapa, que consiste em só duas categorias. A primeira categoria nova é a categoria para a qual a estatística κ é derivada; a segunda é a combinação de todas as outras categorias.

Como exemplo, suponhamos os resultados da classificação dos mapas **R** e **S**, definidos na secção anterior, resumidos em uma tabela de contingência mostrada no quadro a seguir, em que cada valor x_{ij} representa o número de coincidências que tenha sido classificado no mapa S na categoria i e no mapa R pela categoria j .

Tabela de contingência
Proporções em cada categoria

	Mapa R		
Mapa S	0	1	Total
0	5.5/9	0.75/9	6.25/9
1	2/9	0.75/9	2.75/9
Total	7.5/9	1.5/9	1

Cálculo dos coeficientes de Kappa, segundo Hagen

Cálculo do **Kappa Padrão**, definido na equação (2.3),

Note que $P_o = \frac{5.5}{9} + \frac{0.75}{9} = \frac{6.25}{9} = 0.6944$ e $P_e = \sum_{i=1}^2 (p_{i.} \times p_{.i})$, então $P_e = \frac{51}{81} = 0.6296$,

$$\text{Daí, } \kappa_H = \frac{0.6944 - 0.6296}{1 - 0.6296} = \frac{0.0648}{0.3704} = 0.1749$$

Este valor é levemente superior ao obtido segundo o método de Pontius (0.1370).

Cálculo do **Kappa para localização**, definido na equação (3.17)

$$P_{max} = \sum_{i=1}^2 \min(x_{i.}, x_{.i}), \text{ então } P_{max} = \frac{7.75}{9} = 0.8611,$$

$$\text{Assim, } \kappa_{locH} = \frac{0.6944 - 0.6296}{0.8611 - 0.6296} = \frac{0.0648}{0.2315} = 0.2799$$

Esta medida é inferior a κ_{loc} de Pontius (0.3572).

Cálculo do **Kappa em termos de histogramas**, definido na equação (3.19)

$$\kappa_{\text{histo}} = \frac{0.8611-0.6296}{1-0.6296} = \frac{0.2315}{0.3704} = 0.6250, \text{ valor próximo a } \kappa_{\text{quant}} (0.6481).$$

Note ainda que $\kappa_H = \kappa_{\text{histo}} \times \kappa_{\text{loc}} = 0.6250 \times 0.2799 = 0.1749$

3.4. Algumas desvantagens da medida Kappa propostas por Pontius e Hagen

Algumas desvantagens das medidas propostas por Pontius já foram citadas por Hagen (2002).

Outro problema que podemos mencionar é que não existe uma prova matemática formal das ordenações destas medidas. Por exemplo, $N_n < N_m < N_p$ ou $M_n < M_m < M_p$, ou que pode levar em alguns casos a obter medidas negativas ou inclusive não estarem definidas. Por exemplo, se P_p está próximo de N_m , κ_p não está definido ou se M_p está muito próximo de M_n , κ_{quant} não definido. Esses problemas foram observados nos estudos de simulação, o que levou a considerar uma série de condições nos cálculos dos Kappas. Foi considerado que se ambos (numerador e denominador) forem próximos de zero, a medida Kappa (padrão, localização ou quantidade) é nula. Se o denominador é próximo de 1, considera-se a medida Kappa igual a 1.

Outra desvantagem é que ele é baseado na suposição de que κ_{loc} é constante através de vários cenários de quantificação. Contudo κ_{loc} pode mudar quando muda a quantificação. Este fato inclusive foi observado em nossos estudos de simulação.

A desvantagem da medida de Hagen é de que ela não foi definida para diferentes resoluções.

3.5. Propostas de novas medidas de concordância

O problema principal para a aplicação da medida proposta por Pontius (2000), em dois mapas quaisquer, é a determinação de qual deve ser considerado como o real (R) e qual como o simulado (S). Com esta idéia surgem novas medidas modificadas das anteriores. Assim definimos agora a medida modificada sem informação de localização

e com informação média de quantidade, como

$$N_{mM} = \frac{\sum_{n=1}^{N_g} W_n \sum_{i=1}^I \min[\min(R_{n,i}, S_{.,i}), \min(S_{n,i}, R_{.,i})]}{\sum_{n=1}^{N_g} W_n}. \quad (3.21)$$

Esta medida é inspirada na medida N_m dada em (3.7) mas considera a informação dos dois mapas, escolhendo o valor mínimo de concordância.

A medida de informação perfeita de quantidade mas de não informação de localização modificada, como

$$N_{pM} = \frac{\sum_{n=1}^{N_g} W_n \sum_{i=1}^I \min[\min(R_{n,i}, R_{.,i}), \min(S_{n,i}, S_{.,i})]}{\sum_{n=1}^{N_g} W_n}. \quad (3.22)$$

É uma modificação da medida (3.12) e leva em conta a informação de quantidade dos dois mapas. Evidentemente as medidas (3.21) e (3.22) não dependem de qual mapa é considerado real e qual mapa é o simulado.

Define-se agora a medida de não informação em relação a localização e quantidade modificada, como

$$N_{nM} = \frac{\sum_{n=1}^{N_g} W_n \sum_{i=1}^I \min[\min(R_{n,i}, 1/I), \min(S_{n,i}, 1/I)]}{\sum_{n=1}^{N_g} W_n}. \quad (3.23)$$

e a proporção não de informação perfeita em relação à localização e quantidade modificada, como sendo

$$P_{nM} = \sum_{n=1}^{N_g} \min[\min(R_{.,i}, 1/I), \min(S_{.,i}, 1/I)] \quad (3.24)$$

Consideramos a proporção de concordância esperada modificada com informação média de localização e sem informação de quantidade, M_{nM} , como

$$M_{nM} = \frac{P_{nM} + N_{nM}}{2}. \quad (3.25)$$

Define-se a medida Kappa modificada denotada por κ_M sendo igual a

$$\kappa_M = \frac{M_m - N_{mM}}{1 - N_{mM}}. \quad (3.26)$$

É natural considerar que $\kappa_{quantM} = M_m$, pois efetivamente esta medida representa a porcentagem de valores coincidentes. Portanto pode-se definir a concordância em termos de localização como sendo

$$\kappa_{locM} = \frac{\kappa_M}{M_m}. \quad (3.27)$$

Logo $\kappa_{locM} \times \kappa_{quantM} = \kappa_M$.

A medida κ_M está bem definida e que assume valores entre 0 e 1. Como provaremos a seguir

Fixado o i -ésimo termo que define N_{mM} , (3.21), isto é, $\min[\min(R_{ni}, S_{.i}), \min(S_{ni}, R_{.i})]$.

Suponhamos que este mínimo seja R_{ni} logo $R_{ni} \leq S_{ni}$, então $R_{ni} = \min(R_{ni}, S_{ni})$ coincide com o termo da soma que define M_m . De igual maneira se o mínimo for S_{ni} .

Agora, se o mínimo for $S_{.i}$, então $S_{.i} < S_{ni}$ e $S_{.i} < R_{ni}$ logo $S_{.i} < \min(R_{ni}, S_{ni})$, portanto o i -ésimo termo da soma que define N_{nM} é menor que o i -ésimo termo da soma que define M_m , logo, efetivamente, temos que

$$M_m \geq N_{nM}.$$

Como os termos das somas (3.3) e (3.21) são menores ou iguais que 1, então $M_m \leq 1$ e $N_{nM} \leq 1$.

Agora por (3.3) existe um valor de índice i tal que R_{ni} ou $S_{ni} < 1$, logo $S_{.i}$ e $R_{.i}$ são menores que 1, logo existe algum i para o qual o i -ésimo termo de N_{nM} é menor que 1 e assim N_{nM} é estritamente menor que 1, logo κ_M está bem definido. Portanto para esta medida não existe a necessidade de definir condições como na medida κ_p .

Notemos que a diferença de κ_{loc} , M_m tem a mesma escala que κ_M , isto é, assume valores entre 0 e 1.

Outra medida proposta para análise é a chamada Kappa correta dada por

$$\kappa_{co} = 1 - \frac{\sum_{n=1}^{N_g} W_n \sum_{i=1}^I (|R_{n,i} - S_{n,i}|)}{\sum_{n=1}^{N_g} W_n}. \quad (3.28)$$

Esta medida considera as diferenças entre os valores observados para cada categoria e em cada cela.

Capítulo 4

Comparação entre as medidas de concordância

4.1. Considerações iniciais

Com o intuito de analisar o desempenho de todas as medidas de concordância apresentadas, foi realizada uma série de simulações em que consideramos amostras obtidas de variáveis binárias, para representar “mapas hipotéticos”. Por exemplo, se o valor da amostra j -ésima é 1, então na cela j o mapa terá a cor branca e, se o valor é zero, a cela do mapa terá a cor preta. Foram realizados dois estudos de simulação em que no primeiro a variação entre as amostras foi feita em relação à quantidade e no segundo uma amostra difere da outra na posição dos valores.

4.2. Primeiro estudo de simulação

Foram obtidas 1000 amostras em cada caso de tamanho 200, $(X_{i1}, \dots, X_{i200})$, de uma variável aleatória X_i com distribuição binomial de parâmetros (n, p) . Considerou-se $n = 200$ e p variando de 0.10, 0.25, 0.50, 0.75 e 0.90, com $i = 1, \dots, 1000$.

Numa primeira etapa, a partir da amostra original, obteve-se uma segunda amostra $(Y_{i1}, \dots, Y_{i200})$ com $i = 1, \dots, 1000$ em que $Y_{ij} = X_{ij}$, $j = 1, \dots, k$ e $k = \text{round}(n \times r)$, para $r=0.10, 0.25, 0.50, 0.75$ e 0.90 e $Y_{ij} = 1$ para $j = k + 1, \dots, n$. Então, por exemplo, para $r = 0.25$ significa que a segunda amostra coincide com a primeira em 25%, isto é, nos primeiros 50 valores e $Y_{ij} = 1$ para os valores restantes. Portanto, espera-se uma mudança nos 75% dos valores restantes.

Para visualizar melhor as diferenças, as amostras foram representadas por meio das Figuras de 4.1. a 4.6., nas quais se o valor da amostra foi 0 a cor do pixel é preta e 1 se a cor é branca. As figuras indicadas por **a)** representam a primeira amostra $(X_{i1}, \dots, X_{in})$ e as indicadas por **b)** representam a segunda amostra $(Y_{i1}, \dots, Y_{in})$.

As figuras foram geradas com uma probabilidade (p) e, obviamente, quando (p) é maior a ocorrência de pixels² brancos é maior, também, quando a porcentagem de coincidência é maior os mapas hipotéticos são mais semelhantes.

² Pixels é a abreviatura de “picture elements”. É a menor unidade de informação em uma imagem ou mapa matricial. Referido como uma célula em uma imagem ou grid.

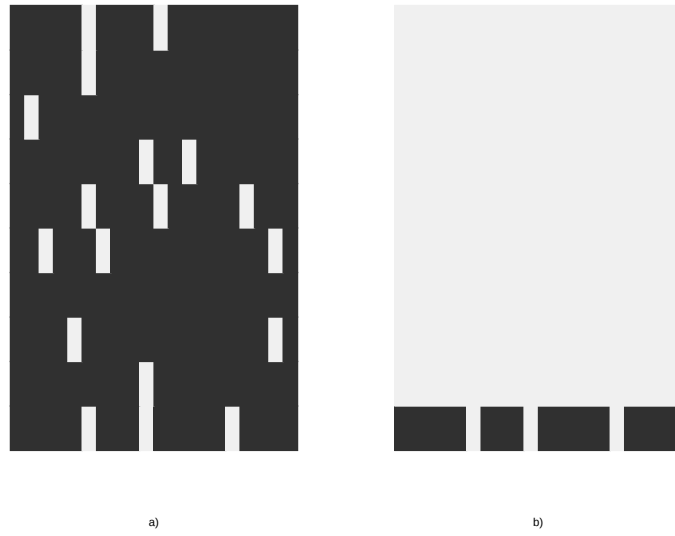


Figura 4.1. Mapas hipotéticos gerados no primeiro estudo para uma coincidência de 10% e $p = 0.10$

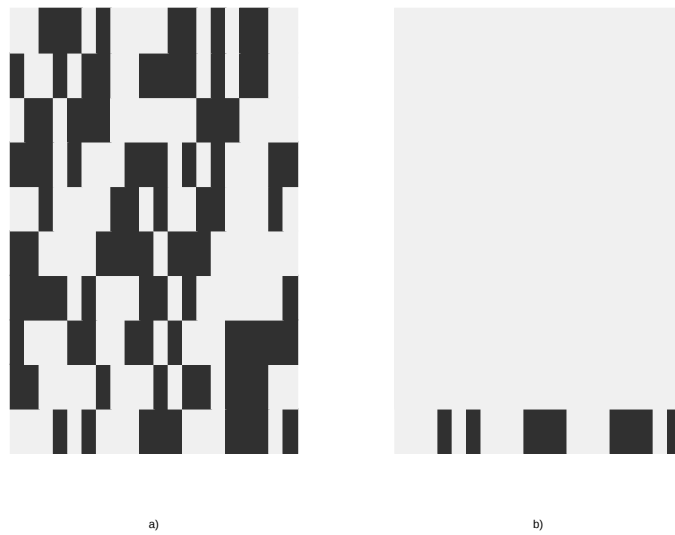


Figura 4.2. Mapas hipotéticos gerados no primeiro estudo para uma coincidência de 10% e $p = 0.50$

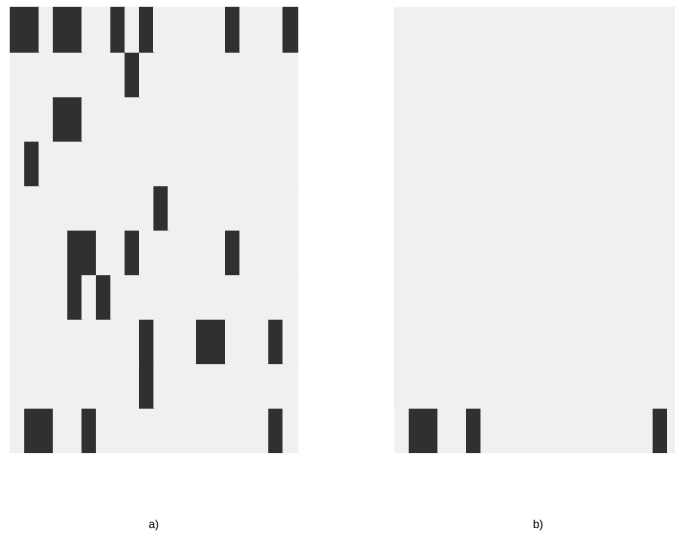


Figura 4.3. Mapas hipotéticos gerados no primeiro estudo para uma coincidência de 10% e $p = 0.90$

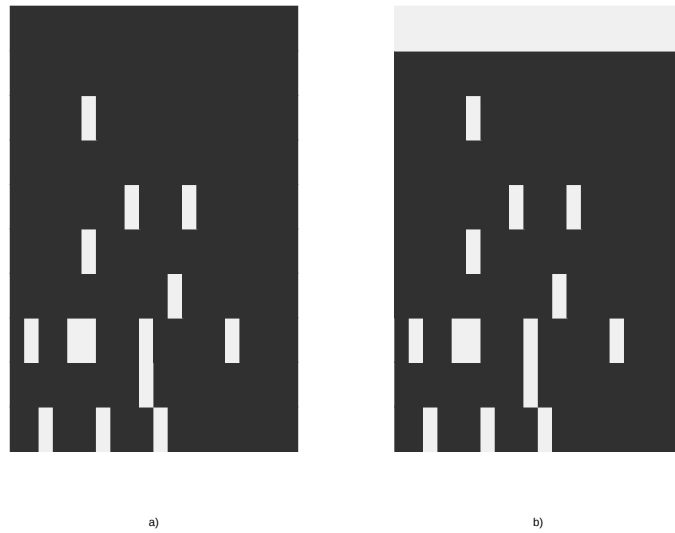


Figura 4.4. Mapas hipotéticos gerados no primeiro estudo para uma coincidência de 90% e $p = 0.10$



Figura 4.5. Mapas hipotéticos gerados no primeiro estudo para uma coincidência de 90% e $p = 0.50$

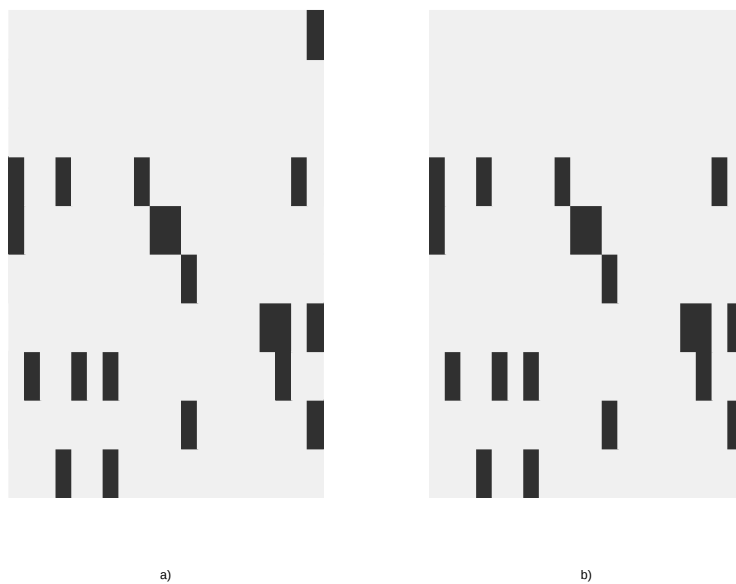


Figura 4.6. Mapas hipotéticos gerados no primeiro estudo para uma coincidência de 90% e $p = 0.90$

Calcularemos, então, a concordância *esperada* em cada situação. Por exemplo, se se mantêm em 25% os valores, ou, equivalentemente, se tem uma mudança de 75% nos valores da amostra. Se consideramos $p = 0.25$ e $n = 200$, então as amostras $(X_{i1}, \dots, X_{i200})$ e $(Y_{i1}, \dots, Y_{i200})$ coincidirão nos 50 valores iniciais, ou seja, $X_{ij} = Y_{ij}$ com $j = 1, \dots, 50$ e $j = 51, \dots, 200$, respectivamente. Como se espera que os X_{ij} , com $j = 51, \dots, 200$, com probabilidade $p = 0.25$, assumirão o valor 1, logo $X_{ij} = Y_{ij}$ nestes casos, isto é, entre 37 e 38 valores coincidirão; ou seja, no total teremos uma concordância esperada nas amostras entre $50 + 37 = 87$ e $50 + 38 = 88$ valores, isto é, uma concordância de 87 ou 88 em 200, isto é esperamos uma coincidência entre 0.435 e 0.440.

Se $p = 0.50$ se espera que os X_{ij} com $j = 51, \dots, 200$ assumirão o valor 1 com probabilidade 0.50, logo $X_{ij} = Y_{ij}$ nestes casos, ou seja, 75 valores coincidirão. No total, teremos uma concordância esperada nas amostras de $50 + 75 = 125$ valores, isto é, uma concordância de 125 em 200, ou seja, esperamos uma concordância de 0.625.

A forma geral deste cálculo pode ser realizada considerando que para $j = k+1, \dots, 200$ $P(X_{ij} = Y_{ij}) = P(X_{ij} = 1) = p$, logo a probabilidade nestes $(200 - k)$ será de $p(200 - k)$, logo a coincidência esperada será $r + p \frac{(200-k)}{200}$. Estes valores são apresentadas no Quadro 4.1.

Quadro 4.1. Valores esperados de concordância C_{esp} (primeiro estudo)

	probabilidade (p)				
coincidência (r)	0.10	0.25	0.50	0.75	0.90
10%	0.19	0.33	0.55	0.78	0.91
25%	0.33	0.44	0.63	0.81	0.93
50%	0.55	0.63	0.75	0.88	0.95
75%	0.78	0.81	0.88	0.94	0.98
90%	0.91	0.93	0.95	0.98	0.99

Para cada par de amostras X_i, Y_i , gerados segundo o caso, foram calculadas as medidas κ_{loc} , M_m , κ_{quant} , κ_p e κ_H definidas em (3.5), (3.6), (3.15), (3.16) e (3.20), respectivamente, e as medidas modificadas κ_{locM} , κ_M e κ_{co} definidas em (3.25), (3.26) e (3.27), respectivamente.

Em cada caso se calculou a quantidade de amostras que se apresenta bem em cada intervalo, isto é apresentado no Quadro 4.5. Por exemplo, para uma coincidência de 25% e $p = 0.10$, do Quadro 4.5, observamos que, para 887 amostras, o valor M_m esteve no intervalo $(0.3; 0.4]$ e 113 no intervalo $(0.2; 0.3]$. Para 442 amostras, o valor de κ_{loc}

pertence ao intervalo $(0.9; 1.0]$ e para 558 foi maior que 1. Para 1000 amostras, κ_{quant} foi menor que zero. Para 998 amostras, κ_p esteve no intervalo $(0.0; 0.1]$, 2 no intervalo $(0.1; 0.2]$.

4.3. Resultados obtidos do primeiro grupo de simulações

Para analisar o desempenho das diferentes medidas de concordância, compararemos a concordância esperada apresentadas no Quadro 4.1. com os intervalos em que se observou a maior frequência de amostras. Em todas as amostras, os valores de M_m (porcentagem de valores corretos) estão no intervalo ao qual pertence o valor de coincidência esperado, o qual é o esperado já que este representa a probabilidade de os valores das celas serem coincidentes. Ao considerar uma mudança de 90% ou uma coincidência de 10%, os valores de Kappa de Pontius, κ_P , foram inferiores aos esperados no Quadro 4.1. Nos últimos casos em que se fixou uma coincidência de 90% para as probabilidades 0.50, 0.75 e 0.90, os valores de κ_p foram próximos dos valores esperados.

Nesse conjunto de simulações, observam-se valores altos para κ_{loc} entre 0.90 e 0.10. Enquanto que para o caso do Kappa para localização modificado, κ_{locM} , diminui quando a mudança foi de 90% à medida que a probabilidade (p) é maior. Um padrão similar foi observado quando a mudança foi de 75%. Para uma mudança de 50%, estes valores estiveram fundamentalmente no intervalo $(0.70; 0.90]$. Para uma mudança de 25%, os valores estiveram no intervalo $(0.70; 1.00]$ e, para uma mudança de 10%, os valores de κ_{loc} estiveram no intervalo $(0.80; 1.00]$.

Em relação ao κ_{quant} , podemos observar do Quadro 4.5 que, para uma mudança de 90%, os valores são negativos para todas as probabilidades (p), o que demonstra, em parte, a ineficiência dessa medida. Para uma mudança de 75%, os valores são negativos para as probabilidades 0.10, 0.25 e 0.50, mudando para o intervalo $(0.20; 0.40]$ para $p = 0.75$ e para o intervalo $(0.80; 1.00]$ para $p = 0.90$. Para uma mudança de 50% é menor que zero para $p=0.10$, 0.25 e 0.50 mudando para o intervalo $(0.50; 0.60]$ para $p = 0.75$. Para $p = 0.90$ muda para o intervalo $(0.80; 0.90]$. Uma mudança de 25% e 10% a exceção do caso é em que $p = 0.50$, que é negativo, temos que κ_{quant} tem valores superiores a 0.20 e 0.60 respectivamente.

Para resumir a informação, consideramos no Quadro 4.2 os valores de C_{esp} e as médias das medidas κ_p e κ_M em cada caso. Desse Quadro pode-se observar que os valores de κ_p são muito inferiores inclusive a porcentagem de coincidência r fixa com a qual as amostras foram geradas. Isto não acontece com a nova medida κ_M , com exceção de alguns casos em que p é igual a 0.10 ou 0.25.

Os valores Kappa modificados, κ_M , são menos afetados pelos valores da medida de quantidade. Este fato foi observado ao analisar a medida do κ_p que apresenta valores muito inferiores, talvez pelo fato de que, em muitas situações, o κ_{quant} foi muito baixo. Esses resultados nos dão uma alerta sobre a ineficiência dessa medida.

Quadro 4.2. Valores médios encontrados de Kappas (primeiro estudo)

(r)	(p)	C_{esp}	κ_p	κ_M
10%	0.10	0.19	0.02	0.18
	0.25	0.33	0.05	0.28
	0.50	0.55	0.10	0.40
	0.75	0.78	0.14	0.48
	0.90	0.91	0.16	0.53
25%	0.10	0.33	0.06	0.28
	0.25	0.44	0.14	0.38
	0.50	0.63	0.25	0.49
	0.75	0.81	0.33	0.57
	0.90	0.93	0.36	0.60
50%	0.10	0.55	0.17	0.43
	0.25	0.63	0.33	0.53
	0.50	0.75	0.50	0.64
	0.75	0.88	0.60	0.70
	0.90	0.95	0.64	0.73
75%	0.10	0.78	0.37	0.58
	0.25	0.81	0.60	0.70
	0.50	0.88	0.75	0.79
	0.75	0.94	0.82	0.84
	0.90	0.98	0.84	0.86
90%	0.10	0.91	0.64	0.73
	0.25	0.93	0.82	0.84
	0.50	0.95	0.90	0.91
	0.75	0.98	0.93	0.94
	0.90	0.99	0.94	0.95

Analisando, agora, as resoluções 2×2 e 3×3 apresentadas para as coincidências de 10%, 50% e 90%, e $p = 0.10, 0.50$ e 0.90 .

Do Quadro 4.6 observamos que a resolução 1×1 que os valores de κ_p apresentam valores inferiores aos esperados. Com exceção de quando a coincidência foi de 90% e $p = 0.90$. Os valores de κ_{quant} foram também sempre muito baixos.

Do Quadro 4.7 para a resolução 3×3 , conclui-se de maneira similar.

Dos Quadros 4.9 e 4.10, que consideram resoluções 2×2 e 3×3 , respectivamente, obtêm-se as mesmas conclusões que para a resolução 1×1 . Os valores de κ_{co} se apresentam nos mesmos intervalos que M_m , com o qual podemos concluir que esta é uma medida que determina concordância em termos de quantidade. A medida κ_M apresentou-se sempre em valores inferiores, que C_{esp} , porém valores superiores a κ_p .

Como conclusão dessas simulações, temos que κ_M é uma medida que teve desempenho melhor e é preferível a κ_p que é muito afetada pelos valores baixos de κ_{quant} . Além do fato de não requerer a especificação de qual mapa é o real e qual é o simulado.

4.4. Segundo estudo de simulação

Da mesma maneira, que no primeiro estudo, foram obtidas 1000 amostras de cada caso de tamanho 200, $(X_{i1}, \dots, X_{i200})$, de uma variável aleatória X_i com distribuição binomial de parâmetros (n, p) , consideramos $n = 200$ e p variando entre 0.10, 0.25, 0.50, 0.75 e 0.90 com $i = 1, \dots, 1000$ em que $Y_{ij} = X_{ij}$ $j = 1, \dots, k$ e $k = \text{trunc}(n \times r)$ para $r=0.10, 0.25, 0.50, 0.75$ e 0.90 ; com $Y_{i(k+1)} = X_{in}$, $Y_{ij} = X_{i(j-1)}$ para $j = k + 2, \dots, n$. Por exemplo, se $r = 0.25$ significa que a segunda amostra coincide com a primeira em 25% dos valores, então esperamos uma concordância de pelo menos r .

Assim para as celas $j = k + 2, \dots, n$, a probabilidade de concordância é dada por

$$\begin{aligned} P(X_{ij} = Y_{ij}) &= P(X_{ij} = X_{i(j-1)}) \\ &= P(X_{ij} = 0, X_{i(j-1)} = 0) + P(X_{ij} = 1, X_{i(j-1)} = 1) \\ &= (1 - p)^2 + p^2 \\ &= 1 - 2p + 2p^2. \end{aligned} \tag{4.1}$$

Logo para calcular os valores esperados de coincidências, devemos multiplicar pelo número de celas, ou seja, por $(200 - k)$, logo a concordância esperada é dada por

$$(200 - k) \frac{(1 - 2p + 2p^2)}{200} + r. \tag{4.2}$$

Os resultados desses cálculos são apresentados no Quadro 4.3 através das coincidências esperadas.

Quadro 4.3. Valores esperados de concordância C_{esp} (segundo estudo)

	probabilidade (p)				
coincidência (r)	0.10	0.25	0.50	0.75	0.90
10%	0.84	0.66	0.55	0.66	0.84
25%	0.87	0.72	0.63	0.72	0.87
50%	0.91	0.81	0.75	0.81	0.91
75%	0.96	0.91	0.88	0.91	0.96
90%	0.98	0.96	0.95	0.96	0.98

Note que, nesse conjunto de simulações, esperamos valores mais altos de quantidade e valores mais baixos de localização.

De maneira análoga ao primeiro estudo, para visualizar melhor as diferenças, as amostras foram representadas por meio das Figuras de 4.7 a 4.12, nas quais, se o valor da amostra foi 0 a cor do pixel é preta e 1 se a cor é branca. As figuras indicadas por **a)** representam a primeira amostra $(X_{i1}, \dots, X_{in})$ e as indicadas por **b)** representam a segunda amostra $(Y_{i1}, \dots, Y_{in})$.

As figuras foram geradas com uma probabilidade (p) e, obviamente, quando (p) é maior, a ocorrência de pixels brancos é maior, também, quando a porcentagem de coincidência é maior, os mapas hipotéticos são mais semelhantes.

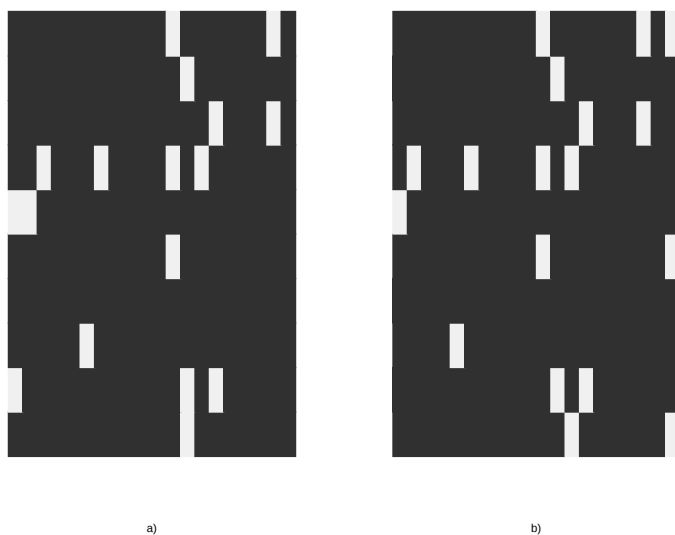


Figura 4.7. Mapas hipotéticos gerados no segundo estudo para uma coincidência de 10% e $p = 0.10$

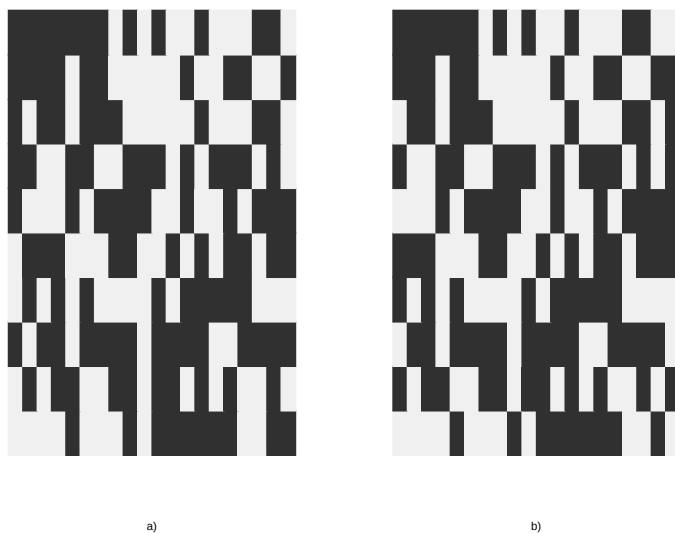


Figura 4.8. Mapas hipotéticos gerados no segundo estudo para uma coincidência de 10% e $p = 0.50$

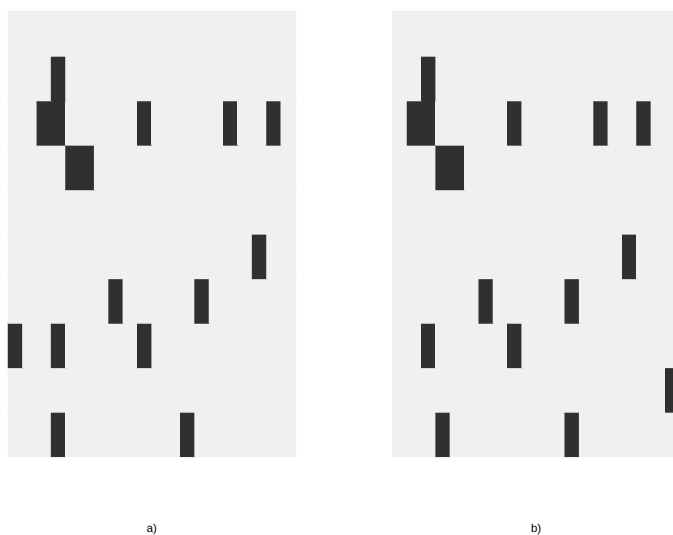


Figura 4.9. Mapas hipotéticos gerados no segundo estudo para uma coincidência de 10% e $p = 0.90$

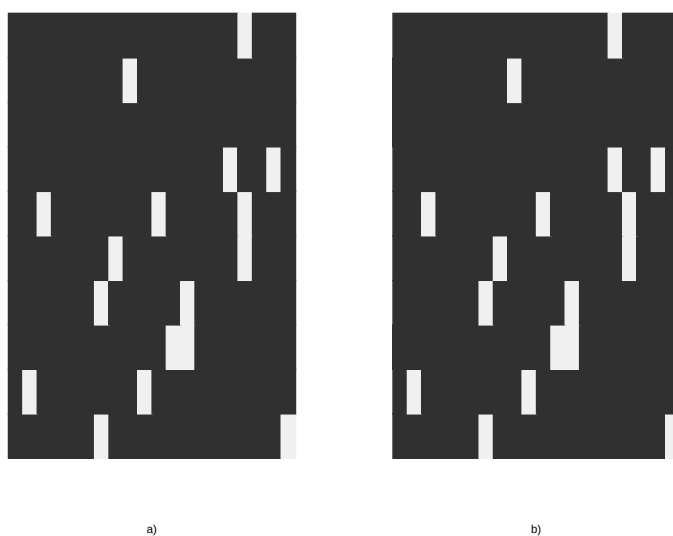


Figura 4.10. Mapas hipotéticos gerados no segundo estudo para uma coincidência de 90% e $p = 0.10$

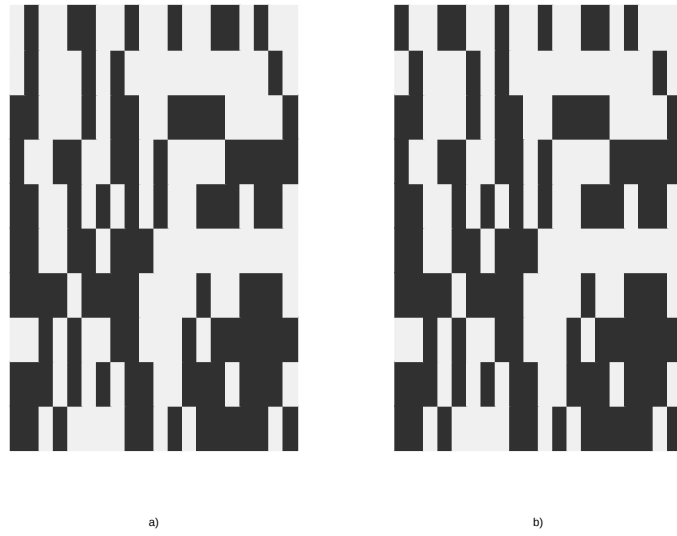


Figura 4.11. Mapas hipotéticos gerados no segundo estudo para uma coincidência de 90% e $p = 0.50$

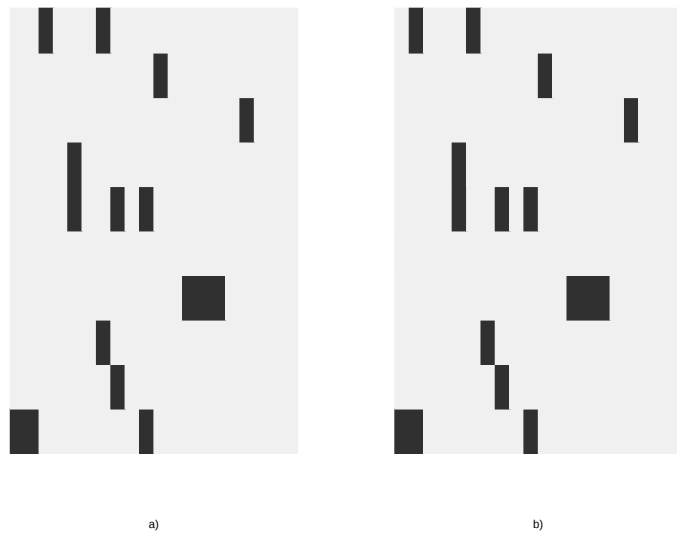


Figura 4.12. Mapas hipotéticos gerados no segundo estudo para uma coincidência de 90% e $p = 0.90$

4.5. Resultados obtidos do segundo grupo de simulações

Para uma coincidência de 10%, a maioria dos valores de κ_p estiveram no intervalo $(0; 0.1]$ e o κ_M no intervalo $(0, 3; 0.4]$.

Para uma coincidência de 25% a maioria dos valores de κ_p estiveram no intervalo $(0.2; 0.3]$ e os de κ_M , no intervalo $(0.4; 0.5]$.

Para uma concordância de 50%, 75% e 90%, os resultados de κ estiveram nos intervalos $(0.4; 0.5]$, $(0.7; 0.8]$, $(0.9; 1, 00]$, respectivamente, e os modificados em $(0.5; 0.6]$, $(0.7; 0.8]$ e $(0.8; 1, 00]$, respectivamente.

Em relação ao κ_{quant} e κ_{quantM} (ou M_m), o κ_{quant} obteve sempre valores no intervalo $(0.9; 1, 00]$ e κ_{quantM} apresentou-se ou no mesmo intervalo, ou em intervalos inferiores. A quantidade total de valores 1 ou zero não se modifica porém os valores das celas sim. Em relação ao κ_{loc} de Pontius, os valores apresentados foram sempre próximos do κ_p .

Ao observar os valores obtidos pelo κ_{locM} constatamos que apresentou em valores superiores ao κ_{loc} e dependem da probabilidade (p). Comparando os valores do κ_p com o κ_M com o Quadro 4.4, observamos que κ_M está mais próximo dos valores de coincidência esperados. Essa medida parece ser mais estável na mudança de localização.

Analisaremos, a seguir, as resoluções 2×2 e 3×3 .

Para a resolução 2×2 , para uma coincidência de 10%, os valores de κ_p estiveram principalmente no intervalo $(0.1; 0.5]$ e κ_M , no intervalo $(0.3; 0.6]$. Para concordâncias de 50% e 90%, os resultados de Pontius estiveram nos intervalos $(0.5; 0.7]$; e $(0.9; 1, 00]$, respectivamente, e para os κ_M , nos intervalos $(0.6; 0.8]$ e $(0.8; 1, 00]$, respectivamente.

Para a resolução 3×3 , para uma coincidência de 10%, os valores de κ_p estiveram principalmente no intervalo $(0.3; 0.5]$ e κ_M , no intervalo $(0.1; 0.4]$. Para concordâncias de 50% e 90%, os resultados de Pontius estiveram nos intervalos $(0.5; 0.7]$; e $(0.9; 1, 00]$, respectivamente, e para os κ_M , nos intervalos $(0.2; 0.6]$ e $(0.5; 0.9]$, respectivamente.

Como era esperado, quando a resolução é menos fina, os valores de Kappa aumentam, “temos menor definição nas imagens”, esperamos maior concordância. Mas os valores do κ_p estão em intervalos mais amplos. Além disso, os valores de κ_M estão no mesmo intervalo da resolução 1×1 o que é uma vantagem.

Nos Quadros 4.11 a 4.16, apresentam-se as quantidades de amostras que tiveram as medidas M_m , κ_{loc} , κ_{quant} , κ_p , κ_H , κ_{locM} , κ_M e κ_{co} em cada intervalo para cada caso de simulação.

Quadro 4.4. Valores médios encontrados de Kappas (segundo estudo)

(r)	(p)	C_{esp}	κ_p	κ_M
10%	0.10	0.84	0.09	0.38
	0.25	0.66	0.09	0.37
	0.50	0.55	0.09	0.38
	0.75	0.66	0.09	0.37
	0.90	0.84	0.09	0.38
25%	0.10	0.87	0.24	0.45
	0.25	0.72	0.24	0.45
	0.50	0.63	0.24	0.45
	0.75	0.72	0.24	0.45
	0.90	0.87	0.24	0.45
50%	0.10	0.91	0.49	0.60
	0.25	0.81	0.49	0.60
	0.50	0.75	0.50	0.60
	0.75	0.81	0.49	0.60
	0.90	0.91	0.50	0.60
75%	0.10	0.96	0.74	0.78
	0.25	0.91	0.74	0.77
	0.50	0.88	0.74	0.77
	0.75	0.91	0.74	0.77
	0.90	0.96	0.74	0.77
90%	0.10	0.98	0.90	0.91
	0.25	0.96	0.89	0.90
	0.50	0.95	0.89	0.90
	0.75	0.96	0.89	0.90
	0.90	0.98	0.89	0.90

Quadro 4.5. Primeiro grupo de simulações
Medidas de concordância de Pontius considerando resolução 1×1

r(%)	(p)	coeficiente	<0	0	(0,;1]	(.1,;2]	(.2,;3]	(.3,;4]	(.4,;5]	(.5,;6]	(.6,;7]	(.7,;8]	(.8,;9]	(.9,;1.0]	>1.0
10	0.10	M_n	0	0	0	716	284	0	0	0	0	0	0	0	0
		κ_{loc}	0	0	0	0	0	0	0	0	0	0	0	192	898
		κ_{quant}	1000	0	0	0	0	0	0	0	0	0	0	0	0
		κ_p	0	0	1000	0	0	0	0	0	0	0	0	0	0
10	0.25	M_n	0	0	0	0	226	773	1	0	0	0	0	0	0
		κ_{loc}	0	0	0	0	0	0	0	0	0	0	0	247	753
		κ_{quant}	1000	0	0	0	0	0	0	0	0	0	0	0	0
		κ_p	0	0	1000	0	0	0	0	0	0	0	0	0	0
10	0.50	M_n	0	0	0	0	0	0	77	872	51	0	0	0	0
		κ_{loc}	0	0	0	0	0	0	0	0	0	0	0	935	65
		κ_{quant}	940	60	0	0	0	0	0	0	0	0	0	0	0
		κ_p	0	0	526	474	0	0	0	0	0	0	0	0	0
10	0.75	M_n	0	0	0	0	0	0	0	0	7	822	171	0	0
		κ_{loc}	0	2	0	0	0	0	0	0	0	0	0	998	0
		κ_{quant}	346	17	127	174	183	98	46	7	0	0	0	0	2
		κ_p	0	2	221	654	119	4	0	0	0	0	0	0	0
10	0.90	M_n	0	0	0	0	0	0	0	0	0	0	369	631	0
		κ_{loc}	0	118	0	0	0	0	0	0	0	0	0	882	0
		κ_{quant}	0	0	0	0	0	0	1	8	121	434	312	6	118
		κ_p	0	118	166	384	250	65	15	0	0	0	0	0	0

Quadro 4.5. (continuação) Primeiro grupo de simulações
Frequências das medidas de concordância de Pontius considerando resolução 1×1

r(%)	(p)	coeficiente	<0	0	(0;1]	(.1;2]	(.2;3]	(.3;4]	(.4;5]	(.5;6]	(.6;7]	(.7;8]	(.8;9]	(.9;1.0]	>1.0
25	0.10	M_m	0	0	0	0	113	887	0	0	0	0	0	0	0
		κ_{loc}	0	0	0	0	0	0	0	0	0	0	0	442	558
		κ_{quant}	1000	0	0	0	0	0	0	0	0	0	0	0	0
		κ_P	0	0	998	2	0	0	0	0	0	0	0	0	0
25	0.25	M_m	0	0	0	0	0	83	909	8	0	0	0	0	0
		κ_{loc}	0	0	0	0	0	0	0	0	0	0	0	536	464
		κ_{quant}	1000	0	0	0	0	0	0	0	0	0	0	0	0
		κ_P	0	0	8	987	5	0	0	0	0	0	0	0	0
25	0.50	M_m	0	0	0	0	0	0	0	233	763	4	0	0	0
		κ_{loc}	0	0	0	0	0	0	0	0	0	0	0	969	31
		κ_{quant}	946	54	0	0	0	0	0	0	0	0	0	0	0
		κ_P	0	0	0	52	872	76	0	0	0	0	0	0	0
25	0.75	M_m	0	0	0	0	0	0	0	0	0	362	638	0	0
		κ_{loc}	0	0	0	0	0	0	0	0	0	0	0	774	226
		κ_{quant}	126	13	112	152	203	207	121	56	10	0	0	0	0
		κ_P	0	0	0	28	305	540	122	5	0	0	0	0	0
25	0.90	M_m	0	0	0	0	0	0	0	0	0	0	105	895	0
		κ_{loc}	0	9	0	0	0	0	0	0	0	0	0	990	1
		κ_{quant}	0	0	0	0	0	0	0	1	41	332	588	29	9
		κ_P	0	9	7	82	197	292	300	96	13	4	0	0	0

Quadro 4.5. (continuação) Primeiro grupo de simulações
Frequências das medidas de concordância de Pontius considerando resolução 1×1

r(%)	(p)	coeficiente	<0	0	(0;1]	(.1;2]	(.2;3]	(.3;4]	(.4;5]	(.5;6]	(.6;7]	(.7;8]	(.8;9]	(.9;1.0]	>1.0
50	0.10	M_m	0	0	0	0	0	0	0	999	1	0	0	0	0
		κ_{loc}	0	0	0	0	0	0	0	0	0	0	0	405	595
		κ_{quant}	1000	0	0	0	0	0	0	0	0	0	0	0	0
		κ_P	0	0	8	73	119	0	0	0	0	0	0	0	0
50	0.25	M_m	0	0	0	0	0	0	0	141	856	3	0	0	0 0
		κ_{loc}	0	0	0	0	0	0	0	0	0	0	0	927	73
		κ_{quant}	1000	0	0	0	0	0	0	0	0	0	0	0	0
		κ_P	0	0	0	0	154	817	29	0	0	0	0	0	0
50	0.50	M_m	0	0	0	0	0	0	0	0	27	954	19	0	0
		κ_{loc}	0	0	0	0	0	0	0	0	0	0	0	958	92
		κ_{quant}	945	54	1	0	0	0	0	0	0	0	0	0	0
		κ_P	0	0	0	0	0	3	522	463	12	0	0	0	0
50	0.75	M_m	0	0	0	0	0	0	0	0	0	1	907	92	0
		κ_{loc}	0	0	0	0	0	0	0	0	0	0	0	980	20
		κ_{quant}	6	0	10	29	67	181	234	274	157	40	2	0	0
		κ_P	0	0	0	0	0	1	52	463	446	38	1	0	0
50	0.90	M_m	0	0	0	0	0	0	0	0	0	0	1	999	0
		κ_{loc}	0	0	0	0	0	0	0	0	0	0	0	960	40
		κ_{quant}	0	0	0	0	0	0	0	0	0	48	654	298	0
		κ_P	0	0	0	0	1	15	76	270	365	234	37	2	0

Quadro 4.5. (continuação) Primeiro grupo de simulações
Frequências das medidas de concordância de Pontius considerando resolução 1×1

r(%)	(p)	coeficiente	<0	0	(0,;1]	(.1,;2]	(.2,;3]	(.3,;4]	(.4,;5]	(.5,;6]	(.6,;7]	(.7,;8]	(.8,;9]	(.9,;1.0]	>1.0
75	0.10	M_m	0	0	0	0	0	0	0	0	0	992	198	0	0
		κ_{loc}	0	0	0	0	0	0	0	0	0	0	0	562	438
		κ_{quant}	0	0	0	0	0	99	892	9	0	0	0	0	0
		κ_P	0	0	0	2	89	607	299	13	0	0	0	0	0
75	0.25	M_m	0	0	0	0	0	0	0	0	0	259	741	0	0
		κ_{loc}	0	0	0	0	0	0	0	0	0	0	0	736	264
		κ_{quant}	9	8	49	222	463	239	10	0	0	0	0	0	0
		κ_P	0	0	0	0	0	0	3	507	484	6	0	0	0
75	0.50	M_m	0	0	0	0	0	0	0	0	0	0	935	65	0
		κ_{loc}	0	0	0	0	0	0	0	0	0	0	0	970	30
		κ_{quant}	957	39	1	1	2	0	0	0	0	0	0	0	0
		κ_P	0	0	0	0	0	0	0	0	71	857	72	0	0
75	0.75	M_m	0	0	0	0	0	0	0	0	0	0	19	981	0
		κ_{loc}	0	0	0	0	0	0	0	0	0	0	0	767	233
		κ_{quant}	0	0	0	0	0	4	15	53	220	447	252	9	0
		κ_P	0	0	0	0	0	0	0	0	8	351	617	24	0
75	0.90	M_m	0	0	0	0	0	0	0	0	0	0	0	1000	0
		κ_{loc}	0	0	0	0	0	0	0	0	0	0	0	1000	0
		κ_{quant}	0	0	0	0	0	0	0	0	0	0	108	892	0
		κ_P	0	0	0	0	0	0	1	1	33	227	572	166	0

Quadro 4.5. (continuação) Primeiro grupo de simulações
 Frequências das medidas de concordância de Pontius considerando resolução 1×1

r(%)	(p)	coeficiente	<0	0	(0,;1]	(.1,;2]	(.2,;3]	(.3,;4]	(.4,;5]	(.5,;6]	(.6,;7]	(.7,;8]	(.8,;9]	(.9,;1.0]	>1.0
90	0.10	M_m	0	0	0	0	0	0	0	0	0	0	128	872	0
		κ_{loc}	0	0	0	0	0	0	0	0	0	0	0	585	415
		κ_{quant}	0	0	0	0	0	0	0	0	0	916	84	0	0
		κ_p	0	0	0	0	0	0	12	209	688	90	1	0	0
90	0.25	M_m	0	0	0	0	0	0	0	0	0	0	3	997	0
		κ_{loc}	0	0	0	0	0	0	0	0	0	0	0	625	375
		κ_{quant}	0	0	0	0	0	0	1	24	490	476	9	0	0
		κ_p	0	0	0	0	0	0	0	0	0	272	727	1	0
90	0.50	M_m	0	0	0	0	0	0	0	0	0	0	0	1000	0
		κ_{loc}	0	0	0	0	0	0	0	0	0	0	0	947	1
		κ_{quant}	766	80	30	30	30	22	19	10	8	3	2	0	0
		κ_p	0	0	0	0	0	0	0	0	0	0	514	486	0
90	0.75	M_m	0	0	0	0	0	0	0	0	0	0	0	1000	0
		κ_{loc}	0	0	0	0	0	0	0	0	0	0	0	999	1
		κ_{quant}	0	0	0	0	0	0	0	0	3	33	451	513	0
		κ_p	0	0	0	0	0	0	0	0	0	0	121	8798	0
90	0.90	M_m	0	0	0	0	0	0	0	0	0	0	0	1000	0
		κ_{loc}	0	0	0	0	0	0	0	0	0	0	0	1000	0
		κ_{quant}	0	0	0	0	0	0	0	0	0	0	0	1000	0
		κ_p	0	0	0	0	0	0	0	0	0	1	142	857	0

Quadro 4.6. Primeiro grupo de simulações
Frequências das medidas de concordância de Pontius considerando resolução 2×2

r(%)	(p)	coeficiente	<0	0	(0,;1]	(.1,;2]	(.2,;3]	(.3,;4]	(.4,;5]	(.5,;6]	(.6,;7]	(.7,;8]	(.8,;9]	(.9,;1.0]	>1.0
10	0.10	M_m	0	0	0	716	284	0	0	0	0	0	0	0	0
		κ_{loc}	0	995	0	0	0	0	0	0	0	0	0	1	4
		κ_{quant}	1000	0	0	0	0	0	0	0	0	0	0	0	0
		κ_p	126	0	874	0	0	0	0	0	0	0	0	0	0
10	0.50	M_m	0	0	0	0	0	0	77	872	51	0	0	0	0
		κ_{loc}	0	33	0	0	0	0	0	0	0	0	0	893	55
		κ_{quant}	890	110	0	0	0	0	0	0	0	0	0	0	0
		κ_p	15	28	957	0	0	0	0	0	0	0	0	0	0
10	0.90	M_m	0	0	0	0	0	0	0	0	0	0	369	631	0
		κ_{loc}	0	118	0	0	0	0	0	0	0	0	0	882	0
		κ_{quant}	0	0	0	0	0	0	1	8	121	434	312	6	118
		κ_p	0	118	278	440	138	21	5	0	0	0	0	0	0
50	0.10	M_m	0	0	0	0	0	0	0	999	1	0	0	0	0
		κ_{loc}	0	821	0	0	0	0	0	0	0	0	0	99	80
		κ_{quant}	1000	0	0	0	0	0	0	0	0	0	0	0	0
		κ_p	260	95	645	0	0	0	0	0	0	0	0	0	0
50	0.50	M_m	0	0	0	0	0	0	0	0	27	954	19	0	0
		κ_{loc}	0	10	0	0	0	0	0	0	0	0	0	949	41
		κ_{quant}	935	64	1	0	0	0	0	0	0	0	0	0	0
		κ_p	3	1	349	565	79	3	0	0	0	0	0	0	0

Quadro 4.6.(continuação)Primeiro grupo de simulações
Frequências das medidas de concordância de Pontius considerando resolução 2×2

r(%)	(p)	coeficiente	<0	0	(0;1]	(.1;2]	(.2;3]	(.3;4]	(.4;5]	(.5;6]	(.6;7]	(.7;8]	(.8;9]	(.9;1.0]	>1.0
50	0.90	M_m	0	0	0	0	0	0	0	0	0	0	1	999	0
		κ_{loc}	0	0	0	0	0	0	0	0	0	0	0	960	40
		κ_{quant}	0	0	0	0	0	0	0	0	0	48	654	298	0
		κ_p	0	0	0	0	12	50	223	340	276	86	12	1	0
90	0.10	M_m	0	0	0	0	0	0	0	0	0	0	128	872	0
		κ_{loc}	0	0	0	0	0	0	0	0	0	0	0	585	415
		κ_{quant}	0	0	0	0	0	0	0	0	0	916	84	0	0
		κ_p	0	0	0	0	0	7	365	529	97	2	0	0	0
90	0.50	M_m	0	0	0	0	0	0	0	0	0	0	0	1000	0
		κ_{loc}	0	0	0	0	0	0	0	0	0	0	0	947	53
		κ_{quant}	766	80	30	30	30	22	19	10	18	3	2	0	0
		κ_p	0	0	0	0	0	0	0	8	164	614	207	7	0
90	0.90	M_m	0	0	0	0	0	0	0	0	0	0	0	1000	0
		κ_{loc}	0	0	0	0	0	0	0	0	0	0	0	1000	0
		κ_{quant}	0	0	0	0	0	0	0	0	0	0	0	1000	0
		κ_p	0	0	0	0	0	0	0	0	0	25	310	665	0

Quadro 4.7. Primeiro grupo de simulações
Frequências das medidas de concordância de Pontius considerando resolução 3×3

r(%)	(p)	coeficiente	<0	0	(0,;1]	(.1,;2]	(.2,;3]	(.3,;4]	(.4,;5]	(.5,;6]	(.6,;7]	(.7,;8]	(.8,;9]	(.9,;1.0]	>1.0
10	0.10	M_n	0	0	0	716	284	0	0	0	0	0	0	0	0
		κ_{loc}	0	1000	0	0	0	0	0	0	0	0	0	0	0
		κ_{quant}	1000	0	0	0	0	0	0	0	0	0	0	0	0
		κ_p	9	846	145	0	0	0	0	0	0	0	0	0	0
10	0.50	M_n	0	0	0	0	0	0	77	872	51	0	0	0	0
		κ_{loc}	0	973	0	0	0	0	0	0	0	0	0	27	0
		κ_{quant}	690	310	0	0	0	0	0	0	0	0	0	0	0
		κ_p	139	753	108	0	0	0	0	0	0	0	0	0	0
10	0.90	M_n	0	0	0	0	0	0	0	0	0	0	369	631	0
		κ_{loc}	0	118	0	0	0	0	0	0	0	0	0	882	0
		κ_{quant}	0	0	0	0	0	0	1	8	122	439	343	71	16
		κ_p	0	118	648	206	25	3	0	0	0	0	0	0	0
50	0.10	M_n	0	0	0	0	0	0	0	999	1	0	0	0	0
		κ_{loc}	0	994	0	0	0	0	0	0	0	0	0	3	3
		κ_{quant}	1000	0	0	0	0	0	0	0	0	0	0	0	0
		κ_p	60	547	393	0	0	0	0	0	0	0	0	0	0
50	0.50	M_n	0	0	0	0	0	0	0	0	27	954	19	0	0
		κ_{loc}	0	250	0	0	0	0	0	0	0	0	0	727	23
		κ_{quant}	883	116	1	0	0	0	0	0	0	0	0	0	0
		κ_p	0	223	739	37	1	0	0	0	0	0	0	0	0

Quadro 4.7. (continuação) Primeiro grupo de simulações
Frequências das medidas de concordância de Pontius considerando resolução 3×3

r(%)	(p)	coeficiente	<0	0	(0;1]	(.1;2]	(.2;3]	(.3;4]	(.4;5]	(.5;6]	(.6;7]	(.7;8]	(.8;9]	(.9;1.0]	>1.0
50	0.90	M_m	0	0	0	0	0	0	0	0	0	0	1	999	0
		κ_{loc}	0	0	0	0	0	0	0	0	0	0	0	960	40
		κ_{quant}	0	0	0	0	0	0	0	0	0	48	654	298	0
		κ_p	0	0	1	40	183	296	269	149	40	20	2	0	0
90	0.10	M_m	0	0	0	0	0	0	0	0	0	0	128	872	0
		κ_{loc}	0	4	0	0	0	0	0	0	0	0	0	582	414
		κ_{quant}	0	0	0	0	0	0	0	0	0	912	88	0	0
		κ_p	0	0	110	447	341	93	9	0	0	0	0	0	0
90	0.50	M_m	0	0	0	0	0	0	0	0	0	0	0	1000	0
		κ_{loc}	0	0	0	0	0	0	0	0	0	0	0	999	1
		κ_{quant}	766	180	30	30	30	22	19	10	8	3	2	0	0
		κ_p	0	0	0	1	2	14	120	336	369	133	22	3	0
90	0.90	M_m	0	0	0	0	0	0	0	0	0	0	0	1000	0
		κ_{loc}	0	0	0	0	0	0	0	0	0	0	0	1000	0
		κ_{quant}	0	0	0	0	0	0	0	0	0	0	0	1000	0
		κ_p	0	0	0	0	0	0	0	5	36	196	390	373	0

Quadro 4.8. Primeiro grupo de simulações

Frequências das medidas de concordância modificadas considerando resolução 1×1

r(%)	(p)	coeficiente	<0	(0;1]	(.1;2]	(.2;3]	(.3;4]	(.4;5]	(.5;6]	(.6;7]	(.7;8]	(.8;9]	(.9;1.0]	>1.0
10	0.10	κ_{locM}	0	0	0	0	0	0	0	0	0	11	989	0
		κ_M	0	0	914	86	0	0	0	0	0	0	0	0
		κ_{co}	0	0	716	284	0	0	0	0	0	0	0	0
10	0.25	κ_{locM}	0	0	0	0	0	0	0	0	2	998	0	0
		κ_M	0	0	0	894	106	0	0	0	0	0	0	0
		κ_{co}	0	0	0	220	779	1	0	0	0	0	0	0
10	0.50	κ_{locM}	0	0	0	0	0	0	0	111	889	0	0	0
		κ_M	0	0	0	0	565	435	0	0	0	0	0	0
		κ_{co}	0	0	0	0	0	65	868	67	0	0	0	0
10	0.75	κ_{locM}	0	0	0	0	0	0	128	872	0	0	0	0
		κ_M	0	0	0	0	0	770	230	0	0	0	0	0
		κ_{co}	0	0	0	0	0	0	0	3	826	171	0	0
10	0.90	κ_{locM}	0	0	0	0	0	0	755	245	0	0	0	0
		κ_M	0	0	0	0	0	249	726	25	0	0	0	0
		κ_{co}	0	0	0	0	0	0	0	0	0	369	631	0
25	0.10	κ_{locM}	0	0	0	0	0	0	0	0	0	1000	0	0
		κ_M	0	0	0	876	124	0	0	0	0	0	0	0
		κ_{co}	0	0	0	83	917	0	0	0	0	0	0	0
25	0.25	κ_{locM}	0	0	0	0	0	0	0	0	0	1000	0	0
		κ_M	0	0	0	0	8787	122	0	0	0	0	0	0
		κ_{co}	0	0	0	0	83	908	9	0	0	0	0	0
25	0.50	κ_{locM}	0	0	0	0	0	0	0	0	864	136	0	0
		κ_M	0	0	0	0	0	681	319	0	0	0	0	0
		κ_{co}	0	0	0	0	0	0	220	775	5	0	0	0

Quadro 4.8. (continuação) Primeiro grupo de simulações
Frequências medidas de concordância modificadas considerando resolução 1×1

r(%)	(p)	coeficiente	<0	(0;1]	(.1;2]	(.2;3]	(.3;4]	(.4;5]	(.5;6]	(.6;7]	(.7;8]	(.8;9]	(.9;1.0]	>1.0
25	0.75	κ_{locM}	0	0	0	0	0	0	2	517	481	0	0	0
		κ_M	0	0	0	0	0	16	858	126	0	0	0	0
		κ_{co}	0	0	0	0	0	0	0	0	362	638	0	0
25	0.90	κ_{locM}	0	0	0	0	0	0	142	729	127	2	0	0
		κ_M	0	0	0	0	0	11	496	467	26	0	0	0
		κ_{co}	0	0	0	0	0	0	0	0	0	105	895	0
50	0.10	κ_{locM}	0	0	0	0	0	0	0	0	917	83	0	0
		κ_M	0	0	0	0	67	933	0	0	0	0	0	0
		κ_{co}	0	0	0	0	0	0	999	1	0	0	0	0
50	0.25	κ_{locM}	0	0	0	0	0	0	0	0	0	1000	0	0
		κ_M	0	0	0	0	0	61	928	11	0	0	0	0
		κ_{co}	0	0	0	0	0	0	23	958	19	0	0	0
50	0.50	κ_{locM}	0	0	0	0	0	0	0	0	0	1000	0	0
		κ_M	0	0	0	0	0	0	61	928	11	0	0	0
		κ_{co}	0	0	0	0	0	0	0	23	958	19	0	0
50	0.75	κ_{locM}	0	0	0	0	0	0	0	0	425	575	0	0
		κ_M	0	0	0	0	0	0	2	473	519	6	0	0
		κ_{co}	0	0	0	0	0	0	0	0	1	907	92	0
50	0.90	κ_{locM}	0	0	0	0	0	0	0	89	628	280	3	0
		κ_M	0	0	0	0	0	0	11	267	599	121	2	0
		κ_{co}	0	0	0	0	0	0	0	0	0	1	999	0
75	0.10	κ_{locM}	0	0	0	0	0	0	0	47	945	8	0	0
		κ_M	0	0	0	0	0	0	794	206	0	0	0	0
		κ_{co}	0	0	0	0	0	0	0	0	992	8	0	0

Quadro 4.8. (continuação) Primeiro grupo de simulações

Frequências das medidas de concordância modificadas considerando resolução 1×1

r(%)	(p)	coeficiente	<0	(0;1]	(.1;2]	(.2;3]	(.3;4]	(.4;5]	(.5;6]	(.6;7]	(.7;8]	(.8;9]	(.9;1.0]	>1.0
75	0.25	κ_{locM}	0	0	0	0	0	0	0	0	1	999	0	0
		κ_M	0	0	0	0	0	0	0	551	449	0	0	0
		κ_{co}	0	0	0	0	0	0	0	0	259	741	0	0
75	0.50	κ_{locM}	0	0	0	0	0	0	0	0	0	192	808	0
		κ_M	0	0	0	0	0	0	0	0	592	408	0	0
		κ_{co}	0	0	0	0	0	0	0	0	0	934	66	0
75	0.75	κ_{locM}	0	0	0	0	0	0	0	0	0	489	511	0
		κ_M	0	0	0	0	0	0	0	0	77	880	43	0
		κ_{co}	0	0	0	0	0	0	0	0	0	19	981	0
75	0.90	κ_{locM}	0	0	0	0	0	0	0	1	38	603	358	0
		κ_M	0	0	0	0	0	0	0	1	110	695	194	0
		κ_{co}	0	0	0	0	0	0	0	0	0	0	1000	0
90	0.10	κ_{locM}	0	0	0	0	0	0	0	5	469	526	0	0
		κ_M	0	0	0	0	0	0	0	168	829	3	0	0
		κ_{co}	0	0	0	0	0	0	0	0	0	1287	872	0
90	0.25	κ_{locM}	0	0	0	0	0	0	0	0	0	191	809	0
		κ_M	0	0	0	0	0	0	0	0	19	976	5	0
		κ_{co}	0	0	0	0	0	0	0	0	0	2	998	0
90	0.50	κ_{locM}	0	0	0	0	0	0	0	0	0	0	1000	0
		κ_M	0	0	0	0	0	0	0	0	0	300	700	0
		κ_{co}	0	0	0	0	0	0	0	0	0	0	1000	0
90	0.75	κ_{locM}	0	0	0	0	0	0	0	0	0	1	999	0
		κ_M	0	0	0	0	0	0	0	0	0	63	937	0
		κ_{co}	0	0	0	0	0	0	0	0	0	0	1000	0
90	0.90	κ_{locM}	0	0	0	0	0	0	0	0	0	46	954	0
		κ_M	0	0	0	0	0	0	0	0	0	111	889	0
		κ_{co}	0	0	0	0	0	0	0	0	0	0	1000	0

Quadro 4.9. Primeiro grupo de simulações
Frequências das medidas de concordância modificadas considerando resolução 2×2

r(%)	(p)	coeficiente	<0	(0;1]	(.1;2]	(.2;3]	(.3;4]	(.4;5]	(.5;6]	(.6;7]	(.7;8]	(.8;9]	(.9;1.0]	>1.0
10	0.10	κ_{locM}	0	0	0	0	0	0	0	15	433	552	0	0
		κ_M	0	0	1000	0	0	0	0	0	0	0	0	0
		κ_{co}	0	0	716	384	0	0	0	0	0	0	0	0
10	0.50	κ_{locM}	0	0	0	0	130	835	35	0	0	0	0	0
		κ_M	0	0	35	958	10	0	0	0	0	0	0	0
		κ_{co}	0	0	0	0	0	65	868	67	0	0	0	0
10	0.90	κ_{locM}	0	0	0	0	2	534	447	17	0	0	0	0
		κ_M	0	0	0	0	116	746	134	4	0	0	0	0
		κ_{co}	0	0	0	0	0	0	0	0	0	369	631	0
50	0.10	κ_{locM}	0	0	1	155	784	60	0	0	0	0	0	0
		κ_M	0	0	673	327	0	0	0	0	0	0	0	0
		κ_{co}	0	0	0	0	0	0	999	1	0	0	0	0
50	0.50	κ_{locM}	0	0	0	0	0	209	779	12	0	0	0	0
		κ_M	0	0	0	0	564	432	4	0	0	0	0	0
		κ_{co}	0	0	0	0	0	0	0	23	958	19	0	0
50	0.90	κ_{locM}	0	0	0	0	0	0	19	376	536	68	1	0
		κ_M	0	0	0	0	0	0	94	545	327	33	1	0
		κ_{co}	0	0	0	0	0	0	0	0	0	1	999	0
90	0.10	κ_{locM}	0	0	0	0	0	0	0	108	890	2	0	0
		κ_M	0	0	0	0	0	0	7	933	60	0	0	0
		κ_{co}	0	0	0	0	0	0	0	0	0	128	872	0
90	0.50	κ_{locM}	0	0	0	0	0	0	0	0	47	887	66	0
		κ_M	0	0	0	0	0	0	0	1	416	569	14	0
		κ_{co}	0	0	0	0	0	0	0	0	0	0	1000	0
90	0.90	κ_{locM}	0	0	0	0	0	0	0	0	0	174	826	0
		κ_M	0	0	0	0	0	0	0	0	2	265	733	0
		κ_{co}	0	0	0	0	0	0	0	0	0	0	1000	0

Quadro 4.10. Primeiro grupo de simulações
Frequências das medidas de concordância modificadas considerando resolução 3×3

r(%)	(p)	coeficiente	<0	(0;1]	(.1;2]	(.2;3]	(.3;4]	(.4;5]	(.5;6]	(.6;7]	(.7;8]	(.8;9]	(.9;1.0]	>1.0
10	0.10	κ_{locM}	0	0	0	0	0	18	402	521	59	0	0	0
		κ_M	0	34	966	0	0	0	0	0	0	0	0	0
		κ_{co}	0	0	716	284	0	0	0	0	0	0	0	0
10	0.50	κ_{locM}	0	0	0	178	790	32	0	0	0	0	0	0
		κ_M	0	0	773	227	0	0	0	0	0	0	0	0
		κ_{co}	0	0	0	0	0	65	868	67	0	0	0	0
10	0.90	κ_{locM}	0	0	3	114	597	258	28	0	0	0	0	0
		κ_M	0	0	10	266	586	131	17	0	0	0	0	0
		κ_{co}	0	0	0	0	0	0	0	0	0	369	631	0
50	0.10	κ_{locM}	0	0	0	0	656	344	0	0	0	0	0	0
		κ_M	0	0	191	809	0	0	0	0	0	0	0	0
		κ_{co}	0	0	0	0	0	0	999	1	0	0	0	0
50	0.50	κ_{locM}	0	0	0	0	197	735	68	0	0	0	0	0
		κ_M	0	0	0	235	734	31	0	0	0	0	0	0
		κ_{co}	0	0	0	0	0	0	0	23	958	19	0	0
50	0.90	κ_{locM}	0	0	0	0	1	71	451	386	84	7	0	0
		κ_M	0	0	0	0	3	180	486	275	49	7	0	0
		κ_{co}	0	0	0	0	0	0	0	0	0	1	999	0
90	0.10	κ_{locM}	0	0	0	0	0	0	130	854	16	0	0	0
		κ_M	0	0	0	0	0	0	813	187	0	0	0	0
		κ_{co}	0	0	0	0	0	0	0	0	0	128	872	0
90	0.50	κ_{locM}	0	0	0	0	0	0	0	39	649	306	6	0
		κ_M	0	0	0	0	0	0	4	232	650	113	3	0
		κ_{co}	0	0	0	0	0	0	0	0	0	0	1000	0
90	0.90	κ_{locM}	0	0	0	0	0	0	0	2	74	486	438	0
		κ_M	0	0	0	0	0	0	0	3	108	490	399	0
		κ_{co}	0	0	0	0	0	0	0	0	0	0	1000	0

Quadro 4.11. Segundo grupo de simulações
Frequências das medidas de concordância de Pontius considerando resolução 1×1

r(%)	(p)	coeficiente	<0	0	(0;1]	(.1;2]	(.2;3]	(.3;4]	(.4;5]	(.5;6]	(.6;7]	(.7;8]	(.8;9]	(.9;1.0]	>1.0
10	0.10	M_m	0	0	0	0	0	0	0	0	0	186	798	16	0
		κ_{loc}	168	2	402	306	102	20	0	0	0	0	0	0	0
		κ_{quant}	0	0	0	0	0	0	0	0	0	0	0	906	94
		κ_p	168	2	402	307	102	19	0	0	0	0	0	0	0
10	0.25	M_m	0	0	0	0	0	0	0	69	808	123	0	0	0
		κ_{loc}	112	1	466	363	55	4	0	0	0	0	0	0	0
		κ_{quant}	0	0	0	0	0	0	0	0	0	0	0	805	195
		κ_p	112	0	467	365	52	4	0	0	0	0	0	0	0
10	0.50	M_m	0	0	0	0	0	0	104	850	46	0	0	0	0
		κ_{loc}	108	0	450	398	44	0	0	0	0	0	0	0	0
		κ_{quant}	0	109	2	10	5	2	2	16	34	29	104	617	70
		κ_p	108	0	451	397	44	0	0	0	0	0	0	0	0
10	0.75	M_m	0	0	0	0	0	0	0	66	813	121	0	0	0
		κ_{loc}	128	1	449	358	63	2	0	0	0	0	0	0	0
		κ_{quant}	0	0	0	0	0	0	0	0	0	0	0	809	191
		κ_p	128	1	452	355	63	2	0	0	0	0	0	0	0
10	0.90	M_m	0	0	0	0	0	0	0	0	0	162	820	18	0
		κ_{loc}	180	1	377	317	110	12	3	0	0	0	0	0	0
		κ_{quant}	0	0	0	0	0	0	0	0	0	0	0	910	90
		κ_p	180	1	377	319	108	12	3	0	0	0	0	0	0

Quadro 4.11. (continuação) Segundo grupo de simulações
Frequências das medidas de concordância de Pontius considerando resolução 1×1

r(%)	(p)	coeficiente	<0	0	(0,;1]	(.1,;2]	(.2,;3]	(.3,;4]	(.4,;5]	(.5,;6]	(.6,;7]	(.7,;8]	(.8,;9]	(.9,;1.0]	>1.0
25	0.10	M_m	0	0	0	0	0	0	0	0	0	24	875	101	0
		κ_{loc}	5	0	82	251	372	227	54	7	1	1	0	0	0
		κ_{quant}	0	0	0	0	0	0	0	0	0	0	0	927	73
		κ_p	5	0	82	256	372	224	53	6	1	1	0	0	0
25	0.25	M_m	0	0	0	0	0	0	0	2	335	637	6	0	0
		κ_{loc}	0	0	30	240	507	209	14	0	0	0	0	0	0
		κ_{quant}	0	0	0	0	0	0	0	0	0	0	0	823	177
		κ_p	0	0	30	246	514	196	14	0	0	0	0	0	0
25	0.50	M_m	0	0	0	0	0	0	0	247	749	4	0	0	0
		κ_{loc}	0	0	10	244	568	174	4	0	0	0	0	0	0
		κ_{quant}	0	92	28	2	0	0	0	52	33	72	125	593	3
		κ_p	0	0	10	244	568	174	4	0	0	0	0	0	0
25	0.75	M_m	0	0	0	0	0	0	0	0	330	660	10	0	0
		κ_{loc}	0	0	17	258	509	196	20	0	0	0	0	0	0
		κ_{quant}	0	0	0	0	0	0	0	0	0	0	0	814	186
		κ_p	0	0	18	261	512	190	19	0	0	0	0	0	0
25	0.90	M_m	0	0	0	0	0	0	0	0	0	27	878	95	0
		κ_{loc}	8	0	74	261	357	232	61	3	4	0	0	0	0
		κ_{quant}	0	0	0	0	0	0	0	0	0	0	0	913	87
		κ_p	5	0	74	266	3612	225	59	3	4	0	0	0	0

Quadro 4.11. (continuação) Segundo grupo de simulações
 Frequências das medidas de concordância de Pontius considerando resolução 1×1

r(%)	(p)	coeficiente	<0	0	(0,;1]	(.1,;2]	(.2,;3]	(.3,;4]	(.4,;5]	(.5,;6]	(.6,;7]	(.7,;8]	(.8,;9]	(.9,;1.0]	>1.0
50	0.10	M_m	0	0	0	0	0	0	0	0	0	0	405	595	0
		κ_{loc}	0	0	0	7	34	165	352	276	141	22	3	0	0
		κ_{quant}	0	0	0	0	0	0	0	0	0	0	0	973	27
		κ_p	0	0	0	7	34	166	355	277	137	21	3	0	0
50	0.25	M_m	0	0	0	0	0	0	0	0	0	382	618	0	0
		κ_{loc}	0	0	0	0	2	61	467	420	50	0	0	0	0
		κ_{quant}	0	0	0	0	0	0	0	0	0	0	0	999	1
		κ_p	0	0	0	0	2	71	471	409	47	0	0	0	0
50	0.50	M_m	0	0	0	0	0	0	0	0	30	949	21	0	0
		κ_{loc}	0	0	0	0	0	31	535	411	23	0	0	0	0
		κ_{quant}	0	88	35	0	0	0	12	34	67	52	141	571	0
		κ_p	0	0	0	0	0	31	537	409	23	1	0	0	0
50	0.75	M_m	0	0	0	0	0	0	0	0	0	395	605	0	0
		κ_{loc}	0	0	0	0	2	78	450	413	56	1	0	0	0
		κ_{quant}	0	0	0	0	0	0	0	0	0	0	0	995	5
		κ_p	0	0	0	0	2	83	459	404	51	1	0	0	0
50	0.90	M_m	0	0	0	0	0	0	0	0	0	0	404	596	0
		κ_{loc}	0	0	1	2	34	147	330	312	136	34	4	0	0
		κ_{quant}	0	0	0	0	0	0	0	0	0	0	0	981	19
		κ_p	0	0	1	3	34	148	337	307	134	32	4	0	0

Quadro 4.11. (continuação) Segundo grupo de simulações
 Frequências das medidas de concordância de Pontius considerando resolução 1×1

r(%)	(p)	coeficiente	<0	0	(0,;1]	(.1,;2]	(.2,;3]	(.3,;4]	(.4,;5]	(.5,;6]	(.6,;7]	(.7,;8]	(.8,;9]	(.9,;1.0]	>1.0
75	0.10	M_m	0	0	0	0	0	0	0	0	0	0	11	989	0
		κ_{loc}	0	0	0	0	0	1	6	65	226	410	257	35	0
		κ_{quant}	0	0	0	0	0	0	0	0	0	0	0	1000	0
		κ_p	0	0	0	0	0	0	7	68	230	414	249	31	0
75	0.25	M_m	0	0	0	0	0	0	0	0	0	0	503	497	0
		κ_{loc}	0	0	0	0	0	0	0	3	179	691	126	1	0
		κ_{quant}	0	0	0	0	0	0	0	0	0	0	0	1000	0
		κ_p	0	0	0	0	0	0	0	3	204	678	114	1	0
75	0.50	M_m	0	0	0	0	0	0	0	0	0	1	952	147	0
		κ_{loc}	0	0	0	0	0	0	0	1	122	830	47	0	0
		κ_{quant}	0	79	32	0	0	0	19	40	53	40	173	564	0
		κ_p	0	0	0	0	0	0	0	1	122	830	47	0	0
75	0.75	M_m	0	0	0	0	0	0	0	0	0	0	457	543	0
		κ_{loc}	0	0	0	0	0	0	0	1	183	663	152	1	0
		κ_{quant}	0	0	0	0	0	0	0	0	0	0	0	1000	0
		κ_p	0	0	0	0	0	0	0	2	206	650	141	1	0
75	0.90	M_m	0	0	0	0	0	0	0	0	0	0	3	997	0
		κ_{loc}	0	0	0	0	0	0	9	62	241	425	230	33	0
		κ_{quant}	0	0	0	0	0	0	0	0	0	0	0	1000	0
		κ_p	0	0	0	0	0	0	10	67	247	426	218	32	0

Quadro 4.11. (continuação) Segundo grupo de simulações
Frequências das medidas de concordância de Pontius considerando resolução 1×1

r(%)	(p)	coeficiente	<0	0	(0,;1]	(.1,;2]	(.2,;3]	(.3,;4]	(.4,;5]	(.5,;6]	(.6,;7]	(.7,;8]	(.8,;9]	(.9,;1.0]	>1.0
90	0.10	M_m	0	0	0	0	0	0	0	0	0	0	0	1000	0
		κ_{loc}	0	0	0	0	0	0	0	0	4	59	407	530	0
		κ_{quant}	0	0	0	0	0	0	0	0	0	0	0	1000	0
		κ_p	0	0	0	0	0	0	0	0	5	68	425	502	0
90	0.25	M_m	0	0	0	0	0	0	0	0	0	0	0	1000	0
		κ_{loc}	0	0	0	0	0	0	0	0	0	0	557	443	0
		κ_{quant}	0	0	0	0	0	0	0	0	0	0	0	1000	0
		κ_p	0	0	0	0	0	0	0	0	0	1	3599	400	0
90	0.50	M_m	0	0	0	0	0	0	0	0	0	0	0	1000	0
		κ_{loc}	0	0	0	0	0	0	0	0	0	0	675	325	0
		κ_{quant}	0	70	26	0	0	0	29	31	55	57	202	530	0
		κ_p	0	0	0	0	0	0	0	0	0	0	675	325	0
90	0.75	M_m	0	0	0	0	0	0	0	0	0	0	0	1000	0
		κ_{loc}	0	0	0	0	0	0	0	0	0	2	537	461	0
		κ_{quant}	0	0	0	0	0	0	0	0	0	0	0	1000	0
		κ_p	0	0	0	0	0	0	0	0	0	0	566	429	0
90	0.90	M_m	0	0	0	0	0	0	0	0	0	0	0	1000	0
		κ_{loc}	0	0	0	0	0	0	0	0	3	70	426	501	16
		κ_{quant}	0	0	0	0	0	0	0	0	0	0	0	1000	0
		κ_p	0	0	0	0	0	0	0	0	4	82	435	479	0

Quadro 4.12. Segundo grupo de simulações
Frequências das medidas de concordância de Pontius considerando resolução 2×2

r(%)	(p)	coeficiente	<0	0	(0;1]	(.1;2]	(.2;3]	(.3;4]	(.4;5]	(.5;6]	(.6;7]	(.7;8]	(.8;9]	(.9;1.0]	>1.0
10	0.10	M_m	0	0	0	0	0	0	0	0	0	0	174	826	0
		κ_{loc}	3	0	14	47	142	237	271	181	79	23	2	1	0
		κ_{quant}	0	0	0	0	0	0	0	0	0	0	0	988	12
		κ_p	3	0	15	47	143	247	271	171	77	24	1	1	0
10	0.50	M_m	0	0	0	0	0	0	0	0	0	126	872	2	0
		κ_{loc}	214	9	214	279	201	67	14	2	0	0	0	0	0
		κ_{quant}	19	738	4	2	4	1	16	4	5	7	37	172	1
		κ_p	214	9	218	284	194	66	13	2	0	0	0	0	0
10	0.90	M_m	0	0	0	0	0	0	0	0	0	0	149	851	0
		κ_{loc}	0	0	5	44	120	238	268	201	83	27	4	0	0
		κ_{quant}	0	0	0	0	0	0	0	0	0	0	0	996	4
		κ_p	0	0	5	47	135	235	274	195	78	27	4	0	0
50	0.10	M_m	0	0	0	0	0	0	0	0	0	0	0	1000	0
		κ_{loc}	0	0	0	0	1	10	53	168	321	307	107	34	0
		κ_{quant}	0	0	0	0	0	0	0	0	0	0	0	1000	0
		κ_p	0	0	0	0	0	12	57	174	325	296	106	30	0
50	0.50	M_m	0	0	0	0	0	0	0	0	0	0	442	558	0
		κ_{loc}	0	0	1	4	29	116	293	358	177	21	1	0	0
		κ_{quant}	58	107	5	7	6	10	31	20	60	64	108	524	0
		κ_p	0	0	1	4	32	131	303	346	167	15	1	0	0

Quadro 4.12. (continuação) Segundo grupo de simulações
 Frequências das medidas de concordância de Pontius considerando resolução 2×2

r(%)	(p)	coeficiente	<0	0	(0,;1]	(.1,;2]	(.2,;3]	(.3,;4]	(.4,;5]	(.5,;6]	(.6,;7]	(.7,;8]	(.8,;9]	(.9,;1.0]	>1.0
50	0.90	M_m	0	0	0	0	0	0	0	0	0	0	2	998	0
		κ_{loc}	0	0	0	0	6	51	175	311	296	129	32	1000	0
		κ_{quant}	0	0	0	0	0	0	0	0	0	0	0	1000	0
		κ_p	0	0	0	0	0	7	58	177	292	124	26	1000	0
90	0.10	M_m	0	0	0	0	0	0	0	0	0	0	0	1000	0
		κ_{loc}	0	0	0	0	0	0	0	0	9	152	837	2	0
		κ_{quant}	0	0	0	0	0	0	0	0	0	0	0	1000	0
		κ_p	0	0	0	0	0	0	0	0	15	219	766	0	0
90	0.50	M_m	0	0	0	0	0	0	0	0	0	0	0	1000	0
		κ_{loc}	0	0	0	0	0	0	0	0	15	471	511	3	0
		κ_{quant}	25	71	0	0	0	0	59	2	54	90	180	519	0
		κ_p	0	0	0	0	0	0	0	0	34	538	428	0	0
90	0.90	M_m	0	0	0	0	0	0	0	0	0	0	0	1000	0
		κ_{loc}	0	0	0	0	0	0	0	0	17	161	822	0	0
		κ_{quant}	0	0	0	0	0	0	0	0	0	0	0	1000	0
		κ_p	0	0	0	0	0	0	0	0	1	27	230	742	0

Quadro 4.13. Segundo grupo de simulações
Frequências das medidas de concordância de Pontius considerando resolução 3×3

r(%)	(p)	coeficiente	<0	0	(0;1]	(.1;2]	(.2;3]	(.3;4]	(.4;5]	(.5;6]	(.6;7]	(.7;8]	(.8;9]	(.9;1.0]	>1.0
10	0.10	M_m	0	0	0	0	0	0	0	0	0	0	3	997	0
		κ_{loc}	76	3	68	91	139	176	173	135	85	42	10	2	0
		κ_{quant}	0	0	0	0	0	0	0	0	0	0	0	974	26
		κ_p	76	3	68	96	144	173	176	129	86	39	8	2	0
10	0.50	M_m	0	0	0	0	0	0	0	0	0	0	408	592	0
		κ_{loc}	149	1	98	170	195	199	126	50	12	0	0	0	0
		κ_{quant}	15	155	2	1	1	2	17	12	37	33	118	563	44
		κ_p	149	1	103	174	201	197	118	47	10	0	0	0	0
10	0.90	M_m	0	0	0	0	0	0	0	0	0	0	0	1000	0
		κ_{loc}	69	1	56	88	140	170	161	154	103	44	14	0	0
		κ_{quant}	0	0	0	0	0	0	0	0	0	0	0	978	22
		κ_p	69	1	58	88	140	181	156	157	96	41	13	0	0
50	0.10	M_m	0	0	0	0	0	0	0	0	0	0	0	1000	0
		κ_{loc}	2	0	4	13	28	59	88	160	243	204	164	35	0
		κ_{quant}	0	0	0	0	0	0	0	0	0	0	0	997	3
		κ_p	2	0	4	14	28	62	90	170	243	198	154	35	0
50	0.50	M_m	0	0	0	0	0	0	0	0	0	0	5	995	0
		κ_{loc}	3	0	5	8	29	86	164	259	258	140	46	2	0
		κ_{quant}	24	89	4	5	3	6	24	12	56	69	140	566	2
		κ_p	3	0	5	12	29	87	182	271	249	132	29	1	0

Quadro 4.13. (continuação) Segundo grupo de simulações
Frequências das medidas de concordância de Pontius considerando resolução 3×3

r(%)	(p)	coeficiente	<0	0	(0;1]	(.1;2]	(.2;3]	(.3;4]	(.4;5]	(.5;6]	(.6;7]	(.7;8]	(.8;9]	(.9;1.0]	>1.0
50	0.90	M_m	0	0	0	0	0	0	0	0	0	0	0	1000	0
		κ_{loc}	3	0	5	12	21	58	93	163	227	227	165	26	0
		κ_{quant}	0	0	0	0	0	0	0	0	0	0	0	997	3
		κ_p	3	0	6	12	20	63	104	162	229	227	148	26	0
90	0.10	M_m	0	0	0	0	0	0	0	0	0	0	0	1000	0
		κ_{loc}	0	0	0	0	0	0	0	0	3	20	238	794	5
		κ_{quant}	0	0	0	0	0	0	0	0	0	0	0	1000	0
		κ_p	0	0	0	0	0	0	0	1	4	54	208	733	0
90	0.50	M_m	0	0	0	0	0	0	0	0	0	0	0	1000	0
		κ_{loc}	0	0	0	0	0	0	0	0	1	13	168	798	20
		κ_{quant}	5	86	5	0	0	0	48	12	55	91	174	524	0
		κ_p	0	0	0	0	0	0	0	0	2	39	278	681	0
90	0.90	M_m	0	0	0	0	0	0	0	0	0	0	0	1000	0
		κ_{loc}	0	0	0	0	0	0	0	1	3	18	279	699	0
		κ_{quant}	0	0	0	0	0	0	0	0	0	0	0	1000	0
		κ_p	0	0	0	0	0	0	0	1	8	55	243	693	0

Quadro 4.14. Segundo grupo de simulações
Frequências das medidas de concordância modificadas considerando resolução 1×1

r(%)	(p)	coeficiente	<0	(0;1]	(.1;2]	(.2;3]	(.3;4]	(.4;5]	(.5;6]	(.6;7]	(.7;8]	(.8;9]	(.9;1.0]	>1.0
10	0.10	κ_{locM}	0	0	0	0	117	748	135	0	0	0	0	0
		κ_M	0	0	0	21	688	287	4	0	0	0	0	0
		κ_{co}	0	0	0	0	0	0	0	0	0	0	1000	0
10	0.25	κ_{locM}	0	0	0	0	0	36	772	192	0	0	0	0
		κ_M	0	0	0	13	770	216	1	0	0	0	0	0
		κ_{co}	0	0	0	0	0	0	0	0	0	0	1000	0
10	0.50	κ_{locM}	0	0	0	0	0	0	0	827	173	0	0	0
		κ_M	0	0	0	8	781	211	0	0	0	0	0	0
		κ_{co}	0	0	0	0	0	0	0	0	0	0	1000	0
10	0.75	κ_{locM}	0	0	0	0	0	39	757	204	0	0	0	0
		κ_M	0	0	0	7	773	220	0	0	0	0	0	0
		κ_{co}	0	0	0	0	0	0	0	0	0	0	1000	0
10	0.90	κ_{locM}	0	0	0	0	144	722	130	4	0	0	0	0
		κ_M	0	0	0	18	685	290	7	0	0	0	0	0
		κ_{co}	0	0	0	0	0	0	0	0	0	0	1000	0
25	0.10	κ_{locM}	0	0	0	0	6	343	573	75	3	0	0	0
		κ_M	0	0	0	0	152	660	179	8	1	0	0	0
		κ_{co}	0	0	0	0	0	0	0	0	0	0	1000	0
25	0.25	κ_{locM}	0	0	0	0	0	0	199	787	14	0	0	0
		κ_M	0	0	0	0	88	814	98	0	0	0	0	0
		κ_{co}	0	0	0	0	0	0	0	0	0	0	1000	0
25	0.50	κ_{locM}	0	0	0	0	0	0	0	64	936	0	0	0
		κ_M	0	0	0	0	52	872	76	0	0	0	0	0
		κ_{co}	0	0	0	0	0	0	0	0	0	0	1000	0

Quadro 4.14. (continuação) Segundo grupo de simulações
Frequências das medidas de concordância modificadas considerando resolução 1×1

r(%)	(p)	coeficiente	<0	(0;1]	(.1;2]	(.2;3]	(.3;4]	(.4;5]	(.5;6]	(.6;7]	(.7;8]	(.8;9]	(.9;1.0]	>1.0
25	0.75	κ_{locM}	0	0	0	0	0	1	204	777	18	0	0	0
		κ_M	0	0	0	0	78	812	110	0	0	0	0	0
		κ_{co}	0	0	0	0	0	0	0	0	0	0	1000	0
25	0.90	κ_{locM}	0	0	0	0	9	356	550	80	5	0	0	0
		κ_M	0	0	0	0	155	657	181	7	0	0	0	0
		κ_{co}	0	0	0	0	0	0	0	0	0	0	1000	0
50	0.10	κ_{locM}	0	0	0	0	0	12	185	557	232	14	0	0
		κ_M	0	0	0	0	2	82	475	368	68	5	0	0
		κ_{co}	0	0	0	0	0	0	0	0	0	0	1000	0
50	0.25	κ_{locM}	0	0	0	0	0	0	0	149	834	17	0	0
		κ_M	0	0	0	0	0	6	535	454	5	0	0	0
		κ_{co}	0	0	0	0	0	0	0	0	0	0	1000	0
50	0.50	κ_{locM}	0	0	0	0	0	0	0	0	580	420	0	0
		κ_M	0	0	0	0	0	0	568	431	1	0	0	0
		κ_{co}	0	0	0	0	0	0	0	0	0	0	1000	0
50	0.75	κ_{locM}	0	0	0	0	0	0	0	150	833	17	0	0
		κ_M	0	0	0	0	0	12	529	450	9	0	0	0
		κ_{co}	0	0	0	0	0	0	0	0	0	0	1000	0
50	0.90	κ_{locM}	0	0	0	0	1	6	163	569	243	18	0	0
		κ_M	0	0	0	0	2	71	444	394	80	9	0	0
		κ_{co}	0	0	0	0	0	0	0	0	0	0	1000	0
75	0.10	κ_{locM}	0	0	0	0	0	0	1	41	358	518	82	0
		κ_M	0	0	0	0	0	0	7	143	498	317	36	0
		κ_{co}	0	0	0	0	0	0	0	0	0	0	1000	0

Quadro 4.14. (continuação) Segundo grupo de simulações

Frequências das medidas de concordância modificadas considerando resolução 1×1

r(%)	(p)	coeficiente	<0	0	(0;1]	(.1;2]	(.2;3]	(.3;4]	(.4;5]	(.5;6]	(.6;7]	(.7;8]	(.8;9]	(.9;1.0]	>1.0
75	0.25	κ_{locM}	0	0	0	0	0	0	0	0	0	32	923	45	0
		κ_M	0	0	0	0	0	0	0	0	33	749	217	1	0
		κ_{co}	0	0	0	0	0	0	0	0	0	0	0	1000	0
75	0.50	κ_{locM}	0	0	0	0	0	0	0	0	0	0	811	189	0
		κ_M	0	0	0	0	0	0	0	0	6	792	202	0	0
		κ_{co}	0	0	0	0	0	0	0	0	0	0	0	1000	0
75	0.75	κ_{locM}	0	0	0	0	0	0	0	0	0	22	926	52	0
		κ_M	0	0	0	0	0	0	0	0	27	724	247	2	0
		κ_{co}	0	0	0	0	0	0	0	0	0	0	0	1000	0
75	0.90	κ_{locM}	0	0	0	0	0	0	0	2	48	390	489	71	0
		κ_M	0	0	0	0	0	0	0	10	147	521	285	37	0
		κ_{co}	0	0	0	0	0	0	0	0	0	0	0	1000	0
90	0.10	κ_{locM}	0	0	0	0	0	0	0	0	0	9	311	680	0
		κ_M	0	0	0	0	0	0	0	0	2	39	434	525	0
		κ_{co}	0	0	0	0	0	0	0	0	0	0	0	1000	0
90	0.25	κ_{locM}	0	0	0	0	0	0	0	0	0	0	43	957	0
		κ_M	0	0	0	0	0	0	0	0	0	0	545	455	0
		κ_{co}	0	0	0	0	0	0	0	0	0	0	0	1000	0
90	0.50	κ_{locM}	0	0	0	0	0	0	0	0	0	0	0	1000	0
		κ_M	0	0	0	0	0	0	0	0	0	0	514	486	0
		κ_{co}	0	0	0	0	0	0	0	0	0	0	0	1000	0
90	0.75	κ_{locM}	0	0	0	0	0	0	0	0	0	0	29	971	0
		κ_M	0	0	0	0	0	0	0	0	0	0	519	481	0
		κ_{co}	0	0	0	0	0	0	0	0	0	0	0	1000	0
90	0.90	κ_{locM}	0	0	0	0	0	0	0	0	0	9	332	659	0
		κ_M	0	0	0	0	0	0	0	0	1	50	446	503	0
		κ_{co}	0	0	0	0	0	0	0	0	0	0	0	1000	0

Quadro 4.15. Segundo grupo de simulações
 Frequências das medidas de concordância modificadas considerando resolução 2×2

r(%)	(p)	coeficiente	<0	(0;1]	(.1;2]	(.2;3]	(.3;4]	(.4;5]	(.5;6]	(.6;7]	(.7;8]	(.8;9]	(.9;1.0]	>1.0
10	0.10	κ_{locM}	0	0	0	0	1	112	415	372	87	12	1	0
		κ_M	0	0	0	0	33	244	442	230	47	3	1	0
		κ_{co}	0	0	0	0	0	0	0	0	0	0	1000	0
10	0.50	κ_{locM}	0	0	0	4	132	534	305	24	1	0	0	0
		κ_M	0	0	1	77	493	379	48	2	0	0	0	0
		κ_{co}	0	0	0	0	0	0	0	0	0	0	1000	0
10	0.90	κ_{locM}	0	0	0	0	2	78	410	396	105	8	1	0
		κ_M	0	0	0	0	15	236	443	252	48	6	0	0
		κ_{co}	0	0	0	0	0	0	0	0	0	0	1000	0
50	0.10	κ_{locM}	0	0	0	0	0	0	19	206	485	250	40	0
		κ_M	0	0	0	0	0	1	65	387	424	150	33	0
		κ_{co}	0	0	0	0	0	0	0	0	0	0	1000	0
50	0.50	κ_{locM}	0	0	0	0	0	5	118	557	311	9	0	0
		κ_M	0	0	0	0	0	51	393	487	68	1	0	0
		κ_{co}	0	0	0	0	0	0	0	0	0	0	1000	0
50	0.90	κ_{locM}	0	0	0	0	0	0	11	219	470	205	35	0
		κ_M	0	0	0	0	0	0	57	307	436	170	30	0
		κ_{co}	0	0	0	0	0	0	0	0	0	0	1000	0
90	0.10	κ_{locM}	0	0	0	0	0	0	0	0	1	161	836	0
		κ_M	0	0	0	0	0	0	0	0	7	204	789	0
		κ_{co}	0	0	0	0	0	0	0	0	0	0	1000	0
90	0.50	κ_{locM}	0	0	0	0	0	0	0	0	0	281	719	0
		κ_M	0	0	0	0	0	0	0	0	8	519	473	0
		κ_{co}	0	0	0	0	0	0	0	0	0	0	1000	0
90	0.90	κ_{locM}	0	0	0	0	0	0	0	0	6	182	812	0
		κ_M	0	0	0	0	0	0	0	0	10	224	766	0
		κ_{co}	0	0	0	0	0	0	0	0	0	0	1000	0

Quadro 4.16. Segundo grupo de simulações
 Frequências das medidas de concordância modificadas considerando resolução 3×3

r(%)	(p)	coeficiente	<0	(0;1]	(.1;2]	(.2;3]	(.3;4]	(.4;5]	(.5;6]	(.6;7]	(.7;8]	(.8;9]	(.9;1.0]	>1.0
10	0.10	κ_{locM}	0	0	1	18	98	234	322	208	96	21	2	0
		κ_M	0	0	1	29	129	267	309	172	76	15	2	0
		κ_{co}	0	0	0	0	0	0	0	0	0	0	1000	0
10	0.50	κ_{locM}	0	0	4	19	151	339	355	122	10	0	0	0
		κ_M	0	1	4	49	246	392	252	53	3	0	0	0
		κ_{co}	0	0	0	0	0	0	0	0	0	0	1000	0
10	0.90	κ_{locM}	0	0	0	12	90	224	311	241	103	18	1	0
		κ_M	0	0	2	22	129	257	285	209	52	16	0	0
		κ_{co}	0	0	0	0	0	0	0	0	0	0	1000	0
50	0.10	κ_{locM}	0	0	0	0	2	30	121	250	344	211	42	0
		κ_M	0	0	0	0	3	51	138	287	302	183	34	0
		κ_{co}	0	0	0	0	0	0	0	0	0	0	1000	0
50	0.50	κ_{locM}	0	0	0	0	1	21	142	413	332	88	3	0
		κ_M	0	0	0	0	2	55	242	436	209	55	1	0
		κ_{co}	0	0	0	0	0	0	0	0	0	0	1000	0
50	0.90	κ_{locM}	0	0	0	0	5	28	119	261	354	195	38	0
		κ_M	0	0	0	0	8	39	143	269	336	175	30	0
		κ_{co}	0	0	0	0	0	0	0	0	0	0	1000	0
90	0.10	κ_{locM}	0	0	0	0	0	0	0	0	15	232	753	0
		κ_M	0	0	0	0	0	0	0	1	23	236	740	0
		κ_{co}	0	0	0	0	0	0	0	0	0	0	1000	0
90	0.50	κ_{locM}	0	0	0	0	0	0	0	0	8	211	781	0
		κ_M	0	0	0	0	0	0	0	0	17	227	706	0
		κ_{co}	0	0	0	0	0	0	0	0	0	0	1000	0
90	0.90	κ_{locM}	0	0	0	0	0	0	0	1	23	262	714	0
		κ_M	0	0	0	0	0	0	0	1	32	263	704	0
		κ_{co}	0	0	0	0	0	0	0	0	0	0	1000	0

Capítulo 5

Estudo de comparação de mapas geográficos

5.1. Considerações iniciais

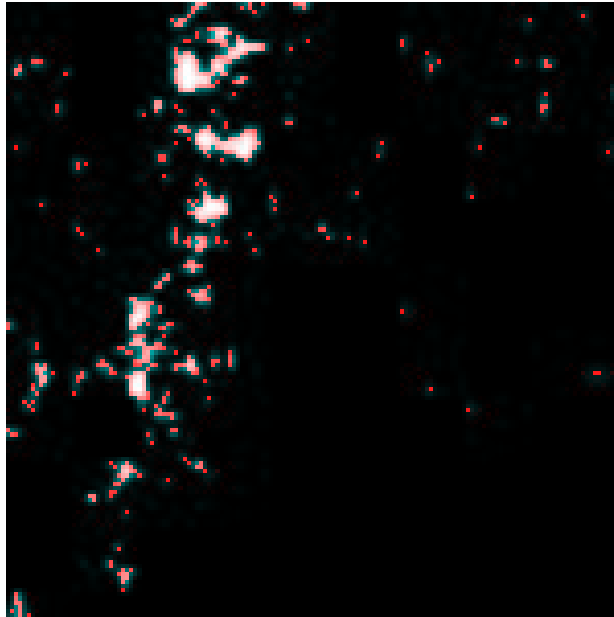
Neste capítulo, com o objetivo de avaliar o desempenho das diferentes medidas de concordância, num conjunto de dados reais, realizamos a comparação de mapas geográficos de desflorestamento da região amazônica dos anos de 2000 e 2001. A medida Kappa fornece um resultado rápido e de fácil obtenção, que poderia ser implementado para o controle de regiões de desflorestamento de maneira periódica.

As regiões em estudo são: Santa Rosa do Purus, no estado do Acre, de coordenadas 09 graus Sul e 70 graus 40 minutos Oeste, ao sul de Acerolândia (Acre), na Bolívia, de coordenadas 09 graus Sul 67 graus Oeste, Santana do Araguaia, no estado do Amazonas, de coordenadas 09 graus Sul e 51 graus 40 minutos Oeste e ao norte de Guarantã do Norte (Mato Grosso), no Pará, de coordenadas 09 graus Sul 55 graus Oeste. A base de dados está disponível gratuitamente em <http://www.dpi.inpe.br/mosaico/mapas>.

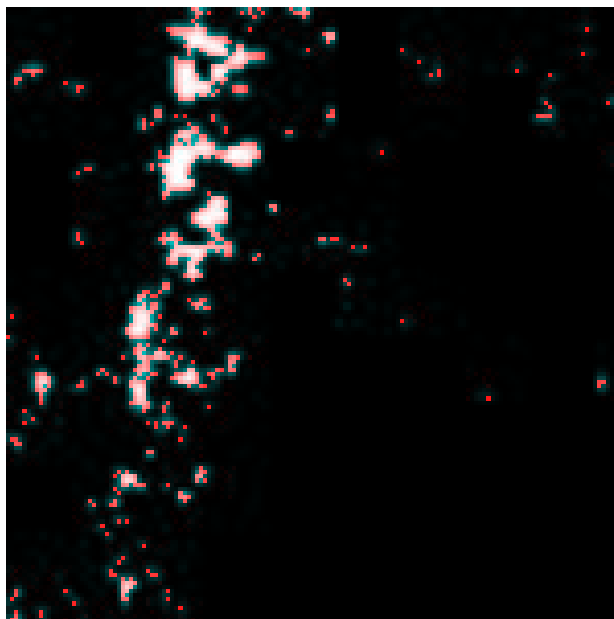
5.2. Formato dos dados

O mapa 1 representa a região de Santa Rosa do Purus, o mapa 2, a região ao sul de Acerolândia, o mapa 3 a região de Santana do Araguaia e o mapa 4, a região ao norte de Guarantã do Norte. Todos os mapas em estudo são de tamanho 20×20 minutos. Os mapas são representados por **a** para o ano 2000 e **b** para 2001. As cores pretas representam regiões com mata e as brancas regiões desflorestadas.

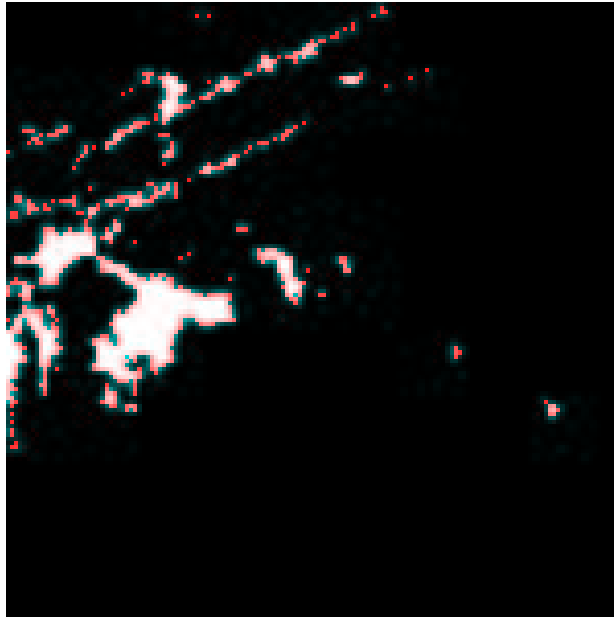
As medidas de concordância entre mapas fornecem valores que permitem avaliar a dimensão do desflorestamento.



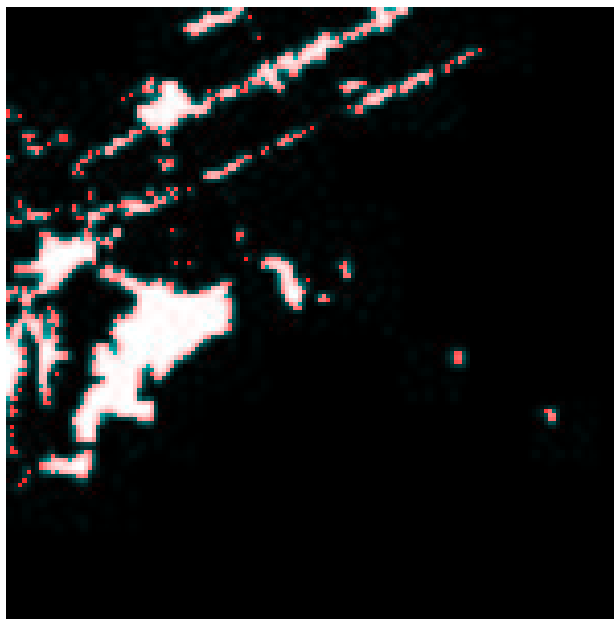
Mapa 1a. Mapa de desflorestamento 2000



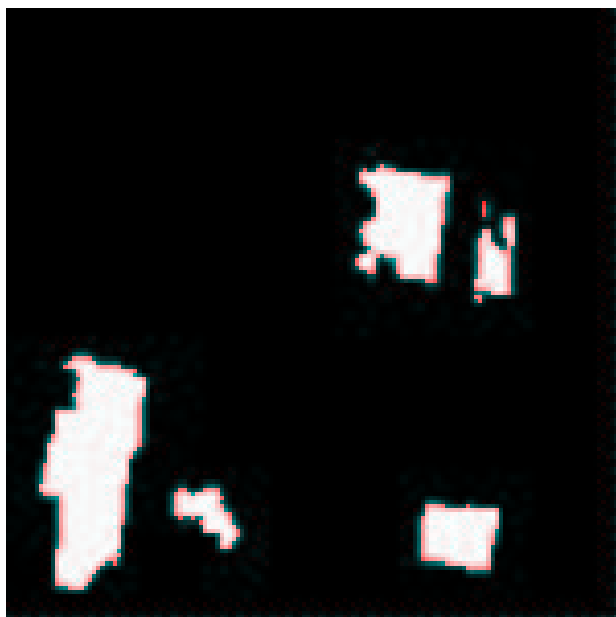
Mapa 1b. Mapa de desflorestamento 2001



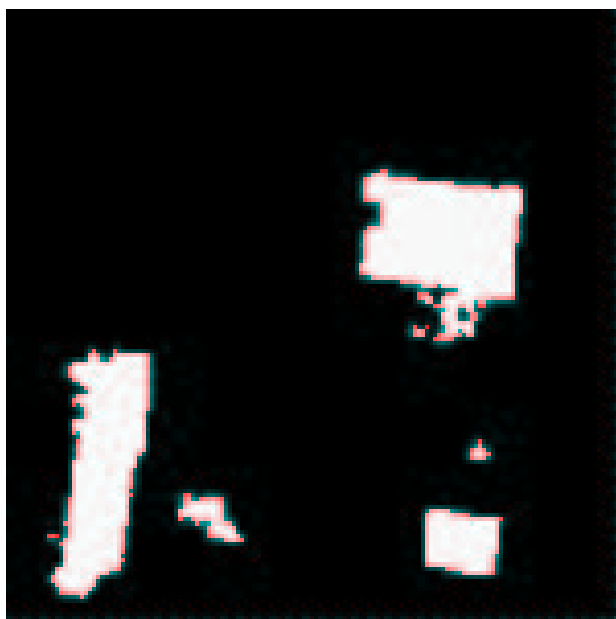
Mapa 2a. Mapa de desflorestamento 2000



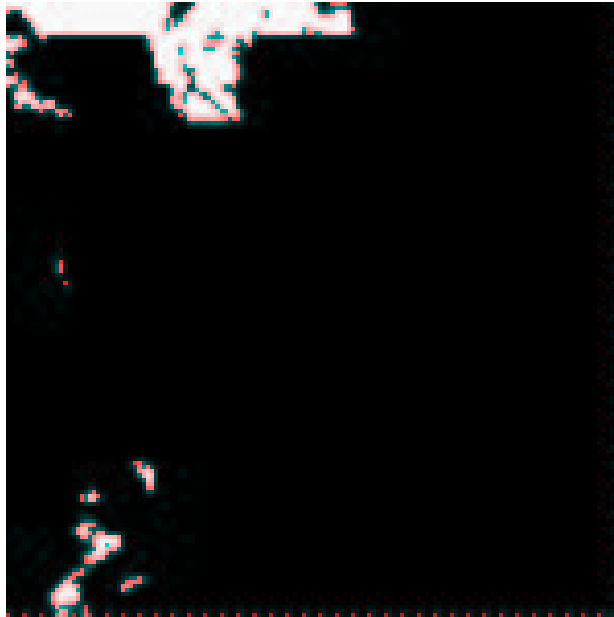
Mapa 2b. Mapa de desflorestamento 2001



Mapa 3a. Mapa de desflorestamento 2000



Mapa 3b. Mapa de desflorestamento 2001



Mapa 4a. Mapa de desflorestamento 2000



Mapa 4b. Mapa de desflorestamento 2001

O quadro 5.1 mostra a porcentagem de concordância na comparação dos mapas geográficos de desflorestamento entre 2000 e 2001. Note que, em todos os casos, o κ_M foi levemente superior ao κ_p . Não existe praticamente diferença entre κ_{quant} e κ_{quantM} . As diferenças maiores são entre κ_{loc} e κ_{locM} .

Quadro 5.1. Valores encontrados de concordância no estudo comparativo dos mapas geográficos de desflorestamento entre 2000 e 2001 para resolução 1×1 .

	valores encontrados			
coeficiente	mapa 1	mapa 2	mapa 3	mapa 4
κ_p	0.48	0.70	0.81	0.77
κ_{quant}	0.98	0.95	1.00	0.97
κ_{loc}	0.57	0.86	0.81	0.87
κ_M	0.61	0.75	0.82	0.81
κ_{quantM}	0.97	0.96	0.97	0.98
κ_{locM}	0.63	0.78	0.85	0.83
κ_H	0.47	0.69	0.80	0.77

A Figura 5.1 mostra o gráfico dos valores dos coeficientes Kappa em função da resolução utilizada para o Mapa 1. Os valores obtidos fornecem, assim, o grau de coincidência entre os Mapas 1a e 1b. Nossos resultados revelam discrepância entre os valores dos diversos coeficientes. Notamos que o coeficiente κ_p é muito volátil e exibe comportamento instável à medida que a resolução varia. Temos, por exemplo, para resolução 1×1 , $\kappa_p = 0.48$ fornecendo, assim, baixo grau de concordância e para resolução 35×35 , $\kappa_p = 0.71$. Ou seja, uma diferença de 0.23 entre as medidas.

Outro coeficiente que apresenta notável variabilidade é o κ_{loc} . Vemos que, para resolução 1×1 , este coeficiente apresenta valor igual a 0.47, contudo apresenta uma tendência crescente à medida que a resolução aumenta, fornecendo, a partir da resolução 13×13 , graus de concordância sempre maiores que 0.95, ou seja, concordância quase perfeita.

Para estes dois mapas, os kappas modificados apresentam comportamento mais estável. Por exemplo, para resolução 1×1 , $\kappa_M = 0.61$ e para resolução 35×35 , $\kappa_M = 0.78$, logo há uma diferença de 0.17, menor que a obtida pelas diferenças entre os κ_p .

Este coeficiente não fornece graus de concordância perfeita entre os mapas, uma vez que seus valores não excedem 0.8, mas caracterizam-se em apresentar menor variabilidade. Os coeficientes M_m e κ_{quant} apresentam-se quase constantes, exibindo taxas de concordância bem próximas a 1 para todas as resoluções, note que $M_m > \kappa_{quant}$. O gráfico mostra também que o coeficiente κ_H fornece grau de concordância constante,

0.47.

Quantitativamente, para os mapas 1a e 1b, podemos afirmar que

$$\kappa_M > \kappa_p > \kappa_H.$$

Como M_m dá uma medida de concordância total no caso de 0.97, se considerarmos que $\kappa_{quant} = 0.98$, ou seja, é muito próxima, com uma concordância de localização de 0.57 ou de 0.63 não faz sentido falar de uma concordância de 48% (κ_p) ou de 47% (κ_H) e sim de uma concordância maior de 61% como a proposta por κ_M .

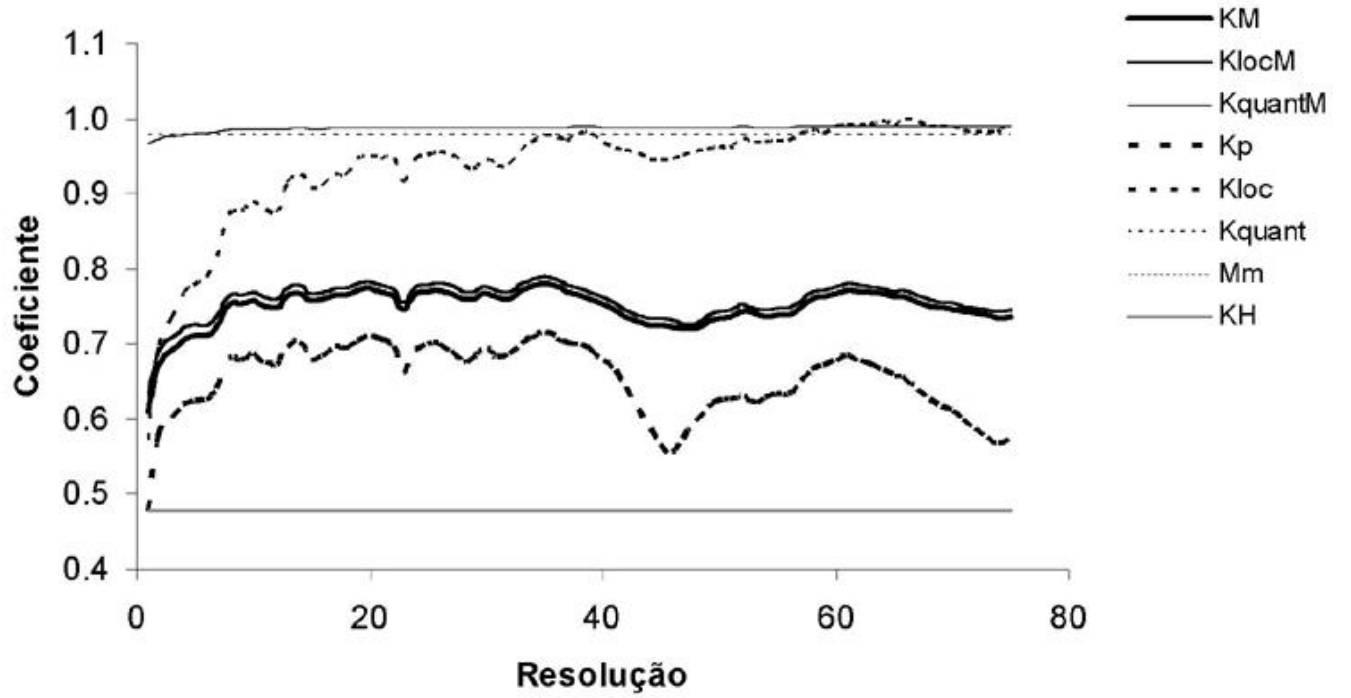


Figura 5.1. Gráfico dos coeficientes de Kappa referente ao mapa 1

Na Figura 5.2, que traz análise dos mapas 2a e 2b, observamos novamente uma forte variabilidade entre os valores para o coeficiente κ_p . Esta característica compromete a confiabilidade deste coeficiente, pois impossibilita ao pesquisador saber exatamente qual resolução escolher. Em contraste, os valores dos coeficientes modificados κ_{locM} e κ_M não variam muito com diferentes resoluções.

O coeficiente κ_{loc} apresenta, como anteriormente, uma tendência crescente com valores quase constantes a partir da resolução 40×40 , os valores deste coeficiente são quase que constantes em 1, representando, assim, uma concordância de 100% entre os mapas, o que não é real.

Os valores dos demais coeficientes kappa são praticamente constantes, sendo que κ_{quantM} têm valores próximos à unidade. Já κ_H , é aproximadamente igual a 0.7 para todas as resoluções. Para os mapas 2a e 2b vemos que também vale

$$\kappa_{loc} > \kappa_{locM}$$

e por outro lado

$$\kappa_M > \kappa_p.$$

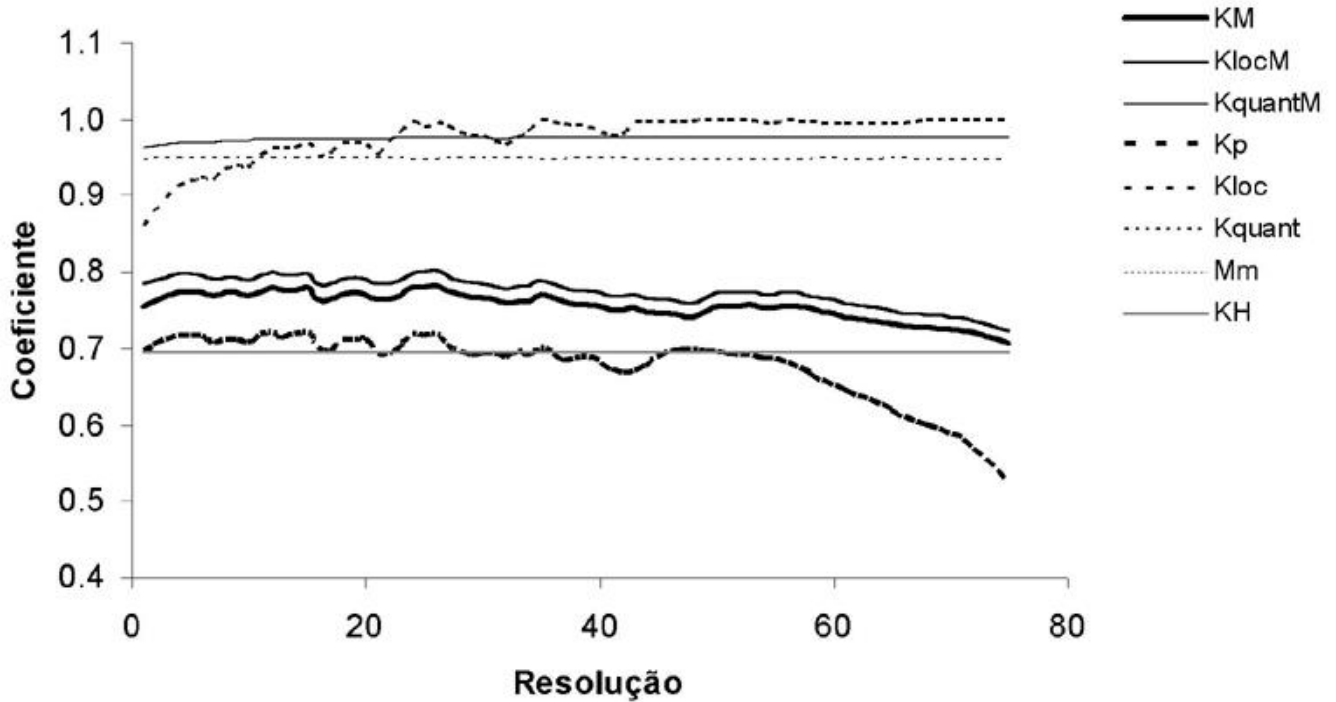


Figura 5.2. Gráfico dos coeficientes de Kappa referente ao mapa 2

Para os mapas 3a e 3b (Figura 5.3), os coeficientes κ_M , κ_p apresentam valores muito próximos, fato não ocorrido para os outros mapas analisados anteriormente. Contudo, estes coeficientes exibem forte variabilidade para resoluções distintas.

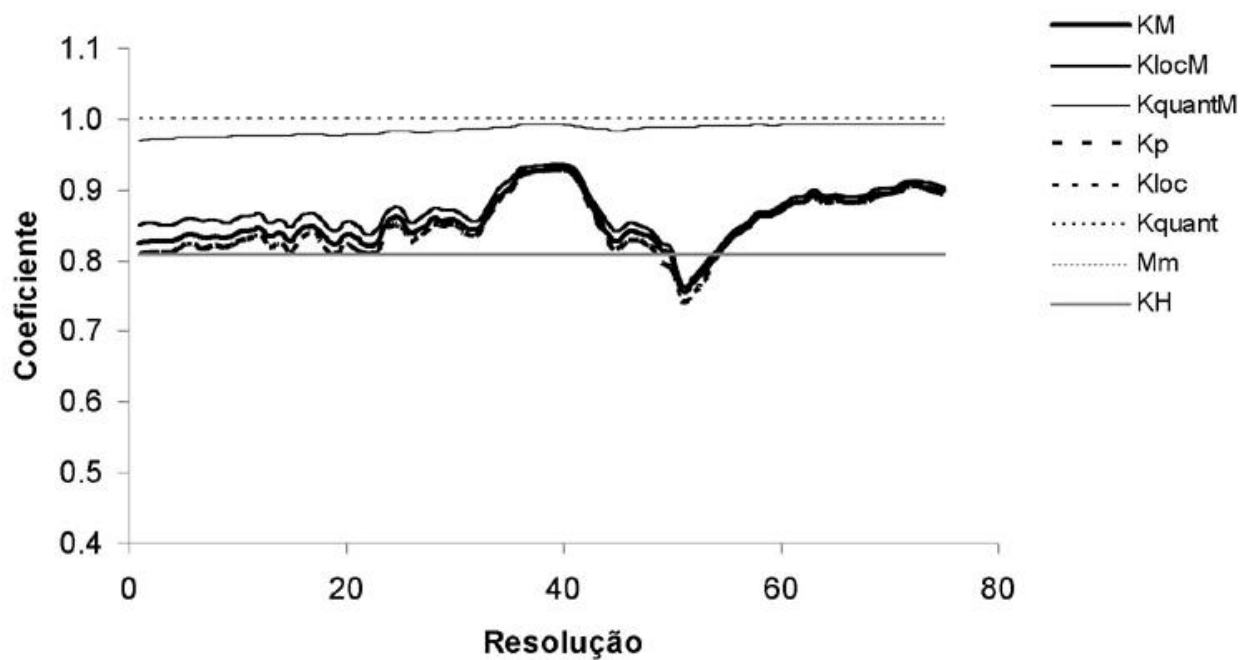


Figura 5.3. Gráfico dos coeficientes de Kappa referente ao mapa 3

Como ocorreu nos mapas 3a e 3b, na comparação entre os mapas 4a e 4b (Figura 5.4), os coeficientes κ_{locM} , κ_M , κ_p exibem valores próximos e fornecem graus de concordância moderada a alta, entre 0.75 e 0.90. Temos para todas as resoluções $\kappa_M > \kappa_p$. Para estes gráficos κ_{loc} fornece maiores graus de concordância, próximos a 1. Os demais coeficientes permanecem constantes, κ_{quant} e M_m fornecendo concordância quase perfeita em termos de quantidade.

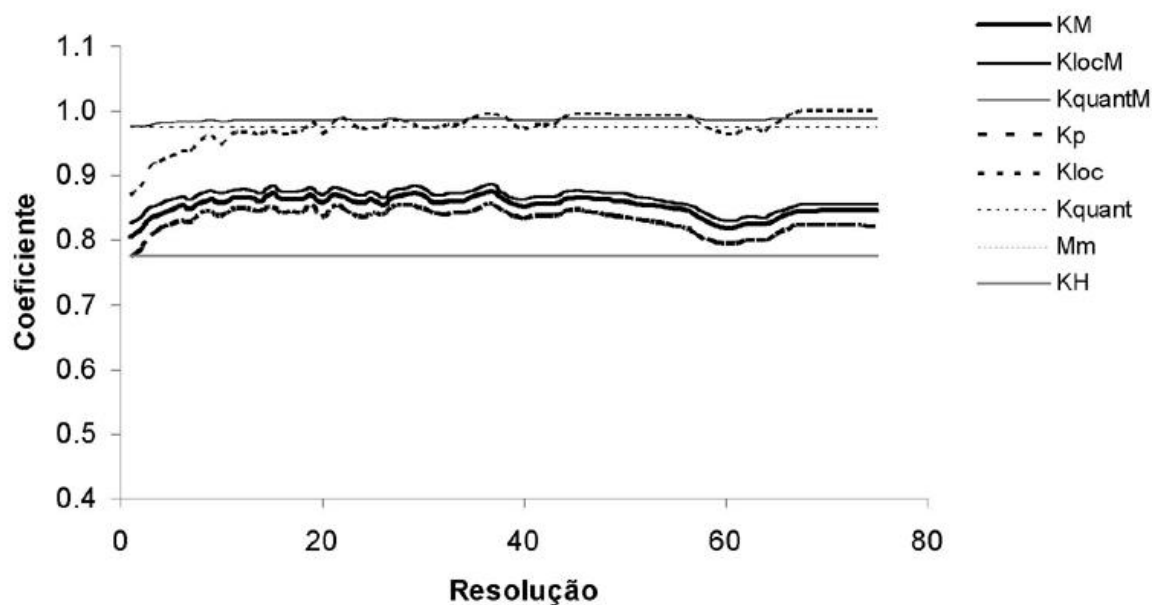


Figura 5.4. Gráfico dos coeficientes de Kappa referente ao mapa 4

Capítulo 6

Conclusões

As medidas de comparação entre mapas temáticos, obtidos a partir de classificações, têm muita utilidade tanto para a comparação direta de mapas de uma determinada região em tempos diferentes, quanto para a avaliação de modelos de predição.

Neste trabalho, apresentaram-se medidas de comparação do tipo kappa, modificadas a partir das propostas dadas por Pontius, em 2000.

Essas novas medidas têm a vantagem de que não requerem a escolha de um mapa real e um mapa simulado, ambos os mapas são tratados de maneira similar. Além disso, não precisam da especificação de condições ou restrições para alguns casos como as requeridas para as medidas de Pontius. Estas condições surgem pela falta de definição em algumas situações, porém, ainda acrescentado estas restrições às medidas de Pontius, apresentam alguns problemas.

Esses problemas foram observados nos estudos de simulação apresentados. A medida proposta para medir concordância em termos de quantidade apresentou, em alguns casos, valores negativos.

Esse tipo de problema surge, acreditamos, porque não é possível obter uma prova formal da ordem das medidas propostas por Pontius e, portanto, de sua correta definição.

Foi apresentada uma prova da boa definição da estatística kappa aqui proposta.

Observamos, a partir dos estudos de simulação, que a medida kappa modificada permanece mais próxima da medida de coincidência total esperada. Consideramos como medida de concordância de localização simplesmente o quociente entre o valor de kappa modificado e a porcentagem de valores corretos, com o qual se logrou dar um “equilíbrio” entre a concordância por localização e por quantidade. Acredita-se que a medida de concordância de localização proposta por Pontius tem uma influência exagerada sobre a medida de concordância geral como foi observado nos estudos de simulação.

Outra vantagem interessante da medida kappa modificada é que sofre menores alterações, quando se consideram resoluções maiores ou mais grosseiras. Em definitivo, tem uma maior estabilidade.

Esse tipo de medida nos fornece um método de rápida implementação para a comparação de mapas temáticos. Isto é de muita utilidade para o estudo da evolução, por exemplo, de áreas de desflorestamento.

6.1. Propostas de trabalhos futuros

Pretende-se realizar a comparação das medidas considerando um maior número de categorias de classificação e dando pesos diferentes às celas.

É necessária também a obtenção de uma medida mais exata para a comparação dos resultados, isto é, alguma medida tipo desvio. Como assim também uma medida de erro da própria medida kappa, que é difícil porque se desconhece a distribuição exata dela. Isto permitiria formular posteriormente algum tipo de estatística de teste.

Pretende-se, também, realizar algum tipo de extensão para a comparação de mais de dois mapas com múltiplas classificações.

Apêndice A

Este apêndice apresenta os comandos do **R** utilizados na elaboração desta dissertação.

O programa a seguir é uma função para obtenção do primeiro grupo de simulações com $p=0.10$

```
>simy10
function(n, tam, proba, ...)
{
  k <- trunc(n * 0.1)
  mx1 <- matrix(1, n, tam)
  my1 <- matrix(1, n, tam)
  for(j in 1:tam) {
    for(i in 1:n) {
      mx1[i, j] <- rbinom(1, 1, proba)
    }
  }
  for(j in 1:tam) {
    for(i in 1:k) {
      my1[i, j] <- mx1[i, j]
    }
  }
  mres <- cbind(mx1[, ], my1[, ])
  return(mres)
}
```

O programa a seguir é uma função para obtenção do segundo grupo de simulações com $p=0.10$

```
> simyr10
function(n, tam, proba, ...)
{
  k <- trunc(n * 0.1)
  mx1 <- matrix(1, n, tam)
  my1 <- matrix(1, n, tam)
  for(j in 1:tam) {
    for(i in 1:n) {
      mx1[i, j] <- rbinom(1, 1, proba)
    }
  }
  for(j in 1:tam) {
    for(i in 1:k) {
      my1[i, j] <- mx1[i, j]
    }
    my1[n, j] <- mx1[k + 1, j]
  }
}
```

```

      for(i in k:(n - 1)) {
        my1[i, j] <- mx1[i + 1, j]
      }
    }
    mres <- cbind(mx1[, ], my1[, ])
    return(mres)
  }

```

O programa a seguir é uma função para obtenção dos mapas hipotéticos

```

exe1<-simy10(200,1,0.1)
cordx<-c(1:20)
cordy<-c(1:10)
par(mfrow=c(1,2))

  m1exe1<-matrix(exe1[,1],ncol=10)
  image(cordx,cordy,m1exe1,axe=F)
  title(sub="a")

  m1exe2<-matrix(exe1[,2],ncol=10)
  image(cordx,cordy,m1exe2,axe=F)
  title(sub="b")
  par(mfrow=c(1,2))

```

Apêndice B

Este apêndice apresenta os comandos do **MATLAB** utilizados na elaboração desta dissertação.

O programa a seguir calcula os coeficientes de Kappa, segundo Pontius, acrescentando-se condições que limitam a faixa de valores de alguns parâmetros.

```
function [KP, KPW, AgreeSpace] = Kpontius01(A,B, res1, res2)
%KPONTIUS kappa Pontius + conditions
% Kappa = agreement between model and test

% A = reality (model) matrix
% B = simulated (test) matrix

% Check possible error
%
% Ri,j and Si,j must be in the 0-1 range.
if any(any(A>1)) any(any(A<0)) any(any(B>1)) any(any(B<0))
    fprintf('Error: R and S must be in the 0-1 range. \n');
end

quase0 = 1e-6;
J = 2; % number of categories. Do not change
[nrow ncol] = size(A); % matrix size

Rsum1 = sum(sum(A))/(nrow*ncol); % R.,1
Rsum2 = 1-Rsum1; % R.,2
Ssum1 = sum(sum(B))/(nrow*ncol); % S.,1
Ssum2 = 1-Ssum1; % S.,2

Pp = 1;

for g=res1:res2

    square = floor([nrow ncol]/g) * g;

    % Work on reality(model) matrix
    %
    A1=[]; A2=[]; A3=[]; Rj=[]; Wn=[]; IJ = []; % clear from last loop
    A1 = A(1:square(1),1:square(2)); % the square part of the matrix
    A2 = A(square(1)+1:nrow,:); % the lower border; considered as a unit
    A3 = A(1:square(1),square(2)+1:ncol); % the right border; considered as a unit
    % calculate mean of each unit
    R = mean(im2col(A1,[g g],'distinct'), 2);
        % aline and calculate R(i,j) = fraction of j in cell i
    n = size(R,1); % number of units
```

```

Wn(1:n,1) = g*g; % units weight according to number of cells
if ~isempty(A2)
    R = [R; mean(mean(A2))];
    n = n+1;
    Wn = [Wn; prod(size(A2))];
end
if ~isempty(A3)
    R = [R; mean(mean(A3))];
    n = n+1;
    Wn = [Wn; prod(size(A3))];

R(:,2) = 1-R; % the second category

Rj(1:n,1) = Rsum1;
Rj(1:n,2) = Rsum2;

% Work on simulated (test) matrix
%
B1=[]; B2=[]; B3=[]; Sj=[]; % clear from last loop
B1 = B(1:square(1),1:square(2));
B2 = B(square(1)+1:nrow,:);
B3 = B(1:square(1),square(2)+1:ncol);

S = mean(im2col(B1,[g g],'distinct'),2); % aline & mean
if ~isempty(B2)
    S = [S; mean(mean(B2))];
end
if ~isempty(B3)
    S = [S; mean(mean(B3))];
end
S(:,2) = 1-S;

Sj(1:n,1) = Ssum1;
Sj(1:n,2) = Ssum2;

% Calculate formulas from Fig. 4 and coefficients
%
IJ(1:n, 1:J) = 1/J;
Wn(:,2) = Wn(:,1); % needed for matrix operations
sumWn = sum(Wn(:,1));

Pn(g) = sum(min(Rj(1,1:J),IJ(1,1:J)));
Pm(g) = sum(min(Sj(1,1:J),Rj(1,1:J)));
Mm(g) = sum(sum(min(R,S).*Wn))/sumWn;
Nn(g) = sum(sum(min(R,IJ).*Wn))/sumWn;

Np(g) = sum(sum(min(R,Rj).*Wn))/sumWn;

Nm(g) = sum(sum(min(R,Sj).*Wn))/sumWn;

if ((Mm(g)-Nn(g)) < quase0) & ((Pp-Nn(g)) < quase0)
    Kno(g) = 0;
elseif ((Pp-Nn(g)) < quase0)

```

```

        Kno(g) = 1;
    else
        Kno(g) = (Mm(g)-Nm(g))/(Pp-Nn(g));
    end

    if ((Mm(g)-Nm(g)) < quase0) & ((Pm(g)-Nm(g)) < quase0)
        Klocation(g) = 0;
    elseif ((Pp-Nn(g)) < quase0)
        Klocation(g) = 1;
    else
        Klocation(g) = (Mm(g)-Nm(g))/(Pm(g)-Nm(g));
    end

    if ((Mm(g)-Nm(g)) < quase0) & ((Pp-Nn(g)) < quase0)
        Kstandard(g) = 0;
    elseif ((Pp-Nn(g)) < quase0)
        Kstandard(g) = 1;
    else
        Kstandard(g) = (Mm(g)-Nm(g))/(Pp-Nn(g));
    end

    Mn(g) = Nn(g) + Klocation(g)*(Pn(g)-Nn(g));

    Mp(g) = Np(g) + Klocation(g)*(Pp-Np(g));

    if ((Mm(g)-Mn(g)) < quase0) & ((Mp(g)-Mn(g)) < quase0)
        Kquantity(g) = 0;
    elseif ((Mp(g)-Mn(g)) < quase0)
        Kquantity(g) = 1;
    else
        Kquantity(g) = (Mm(g)-Mn(g))/(Mp(g)-Mn(g));
    end

    V(g) = n; % number of units at each resolution

end

correct = Mm;
Pp(1:size(V,2)) = 1;
VPIL = Pm-Mm;
VPIQ = Mp-Mm;

% Coefficients weighted by number of cells
sumV = sum(V);
correctW = sum(correct.*V)/sumV;
KnoW = sum(Kno.*V)/sumV;
KlocationW = sum(Klocation.*V)/sumV;
KquantityW = sum(Kquantity.*V)/sumV;
KstandardW = sum(Kstandard.*V)/sumV;
VPIQW = sum(VPIQ.*V)/sumV;
VPILW = sum(VPIL.*V)/sumV;

% Construct output variables

```

```

KP = struct('correct', {correct}, 'no', {Kno}, 'location', {Klocation},...
            'quantity', {Kquantity}, 'standard', {Kstandard}, 'VPIQ', {VPIQ},...
            'VPIL', {VPIL}, 'V', {V});

KPW = struct('correct', {correctW}, 'no', {KnoW}, 'location', {KlocationW},...
            'quantity', {KquantityW}, 'standard', {KstandardW}, 'VPIQ', {VPIQW},...
            'VPIL', {VPILW});

AgreeSpace = struct('Pn', {Pn}, 'Pm', {Pm}, 'Pp', {Pp}, 'Mn', {Mn}, 'Mm', {Mm},...
                   'Mp', {Mp}, 'Nn', {Nn}, 'Nm', {Nm}, 'Np', {Np});

```

O programa a seguir calcula os coeficientes de Kappa, segundo Pontius, com algumas modificações.

```

function [KP, KPW, AgreeSpace] = Kpontius05(A,B, res1, res2)
%KPONTIUS kappa Potnius + 3nd modification + New Correct
% Kappa = agreement between model and test

% res1=1; res2=1;

% A = reality (model) matrix
% B = simulated (test) matrix

% Check possible error
%
% Ri,j and Si,j must be in the 0-1 range.
if any(any(A>1)) any(any(A<0)) any(any(B>1)) any(any(B<0))
    fprintf('Error: R and S must be in the 0-1 range. \n');
end

quase0 = 1e-6;
J = 2; % number of categories. Do not change
[nrow ncol] = size(A); % matrix size

Rsum1 = sum(sum(A))/(nrow*ncol); % R.,1
Rsum2 = 1-Rsum1; % R.,2
Ssum1 = sum(sum(B))/(nrow*ncol); % S.,1
Ssum2 = 1-Ssum1; % S.,2

Pp = 1;

for g=res1:res2

    square = floor([nrow ncol]/g) * g;

    % Work on reality(model) matrix
    %
    A1=[]; A2=[]; A3=[]; Rj=[]; Wn=[]; IJ = []; % clear from last loop
    A1 = A(1:square(1),1:square(2)); % the square part of the matrix

```

```

A2 = A(square(1)+1:nrow,:); % the lower border; considered a as unit
A3 = A(1:square(1),square(2)+1:ncol); % the right border; considered as a unit
% calculate mean of each unit
R = mean(im2col(A1,[g g],'distinct'),'', 2);
      % aline and calculate R(i,j) = fraction of j in cell i
n = size(R,1); % number of units

Wn(1:n,1) = g*g; % units weight according to number of cells
if ~isempty(A2)
    R = [R; mean(mean(A2))];
    n = n+1;
    Wn = [Wn; prod(size(A2))];
end
if ~isempty(A3)
    R = [R; mean(mean(A3))];
    n = n+1;
    Wn = [Wn; prod(size(A3))];
end
R(:,2) = 1-R; % the second category

Rj(1:n,1) = Rsum1;
Rj(1:n,2) = Rsum2;

% Work on simulated (test) matrix
%
B1=[]; B2=[]; B3=[]; Sj=[]; % clear from last loop
B1 = B(1:square(1),1:square(2));
B2 = B(square(1)+1:nrow,:);
B3 = B(1:square(1),square(2)+1:ncol);

S = mean(im2col(B1,[g g],'distinct'),'',2); % aline & mean
if ~isempty(B2)
    S = [S; mean(mean(B2))];
end
if ~isempty(B3)
    S = [S; mean(mean(B3))];
end
S(:,2) = 1-S;

Sj(1:n,1) = Ssum1;
Sj(1:n,2) = Ssum2;

% Calculate formulas from Fig. 4 and coefficients
%
IJ(1:n, 1:J) = 1/J;
Wn(:,2) = Wn(:,1); % needed for matrix operations
sumWn = sum(Wn(:,1));

Pn(g) = sum(min(min(Rj(1,1:J),IJ(1,1:J)),min(Sj(1,1:J),IJ(1,1:J)))); % NOVO
Pm(g) = sum(min(Sj(1,1:J),Rj(1,1:J)));
Mm(g) = sum(sum(min(R,S).*Wn))/sumWn;
Nn(g) = sum(sum(min(min(R,IJ),min(S,IJ)).*Wn))/sumWn; % NOVO
Mn(g) = (Pn(g)+Nn(g))/2; % NOVO

```

```

Nm(g) = sum(sum(min(min(R,Sj),min(S,Rj)).*Wn))/sumWn; % NOVO
Np(g) = sum(sum(min(min(R,Rj),min(S,Sj)).*Wn))/sumWn; % NOVO
Kno(g) = (Mm(g)-Nn(g))/(Pp-Nn(g));
Kstandard(g) = (Mm(g)-Nm(g))/(Pp-Nm(g));
Klocation(g) = Kstandard(g)/Mm(g); % NOVO
Mp(g) = (((Pp-Nm(g))/(Pm(g)-Nm(g)))*(2*Mm(g)-(Pn(g)+Nn(g)))+(Pn(g)+Nn(g)))/2;% NOVO
Kquantity(g) = Mm(g); % NOVO
V(g) = n; % number of units at each resolution

correct_new(g) = 1 - (abs(Rj(1,1)-Sj(1,1)) + abs(Rj(1,2)-Sj(1,2)))/2;

end

correct = Mm;
Pp(1:size(V,2)) = 1;
VPIL = Pm-Mm;
VPIQ = Mp-Mm;

% Coefficients weighted by number of cells
sumV = sum(V);
correctW = sum(correct.*V)/sumV;
KnoW = sum(Kno.*V)/sumV;
KlocationW = sum(Klocation.*V)/sumV;
KquantityW = sum(Kquantity.*V)/sumV;
KstandardW = sum(Kstandard.*V)/sumV;
VPIQW = sum(VPIQ.*V)/sumV;
VPILW = sum(VPIL.*V)/sumV;

% Construct output variables
KP = struct('correct_new',{correct_new},'correct',{correct},'no',{Kno},'location',
    {Klocation},...'quantity',{Kquantity}, 'standard',{Kstandard},'VPIQ',
    {VPIQ},...'VPIL',{VPIL},'V',{V});

KPW = struct('correct',{correctW}, 'no',{KnoW}, 'location',{KlocationW},...
    'quantity',{KquantityW}, 'standard',{KstandardW},'VPIQ',{VPIQW},...
    'VPIL',{VPILW});

AgreeSpace = struct('Pn',{Pn}, 'Pm',{Pm}, 'Pp',{Pp}, 'Mn',{Mn}, 'Mm',{Mm},...
    'Mp',{Mp}, 'Nn',{Nn}, 'Nm',{Nm}, 'Np',{Np});

```

O programa a seguir calcula os coeficientes de Kappa, segundo Hagen

```

function [Khagen, Klocation, Khist, PA, PE, Pmax] = Khagen(R,S)
%KAPPA kappa Hagen
% Kappa = agreement between model and test
% Traditional Method (Hagen 2002)

% R = Reality (model)
% S = Simulated (test)

```

```
% Label all values
[R, nlabels] = label(R);
S = label(S);

m = errormat(R,S); % Error matrix

% Make the error matrix values relative to the total count
Rhyst = hist(R, nlabels);
t1=m;
t2=Rhyst;
for i=1:size(m,2)
    m(i,:) = m(i,:)*Rhyst(i);
end
t3=m;
m = m/sum(Rhyst); % errox matrix

sum1 = sum(m,1); % rows sum
sum2 = sum(m,2); % cols sum
sum2=reshape(sum2,1,size(sum2,1));

PA = sum(diag(m)); % Fraction of agreement
PE = sum(sum1.*sum2); % Expected fraction of agreement
Pmax = sum(min(sum1,sum2)); % Maximum fraction of agreement

Klocation = (PA-PE)/(Pmax-PE); % Similarity of location
Khyst = (Pmax-PE)/(1-PE); % Similarity of quantity
Khagen = (PA - PE)/(1-PE); % = Khyst*Klocation;
```

Referências bibliográficas

- [1] Altman, D. G. (1991). *Practical Statistics for Medical Research*, first edition. New York: Chapman and Hall.
- [2] Cochran, William G. (1965). *Técnicas de Amostragem*. Rio de Janeiro: Fundo de Cultura S.A.
- [3] Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educ. Psychol. Meas*, **20**, 37–46.
- [4] Cohen, J. (1968). Weighted Kappa: Nominal scale agreement with provision for scaled disagreement or partial sets of raters. *Psychol. Bull*, **86**, 974–977.
- [5] Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychol. Bull*, **76**, 378–382.
- [6] Fleiss, J. L. (1981). *Statistical Methods for Rates and Proportions, second edition*. New York: John Wiley.
- [7] Fleiss, J. L. e Cohen, J. (1973). The equivalence of weighted Kappa and the intraclass correlation coefficient as measures of reliability. *Educ. Psychol. Meas.*, **33**, 813–619.
- [8] Fleiss, J.L. e Cuzick, J. (1979). The reliability of dichotomous judgments: Unequal numbers of judges per subject. *Appl. Psychol. Meas*, **3**, 537–542.
- [9] Fleiss, J. L, Nee, J. C. M., and Landis, J. R. (1979). The large sample variance of Kappa in the case of different sets of raters. *Psychol. Bull*, **86**, 974–977.
- [10] Hagen, A. (2002). Multi-method assessment of map similarity. 5 *AGILE, Conference on Geographic Information Science, Palma (Mallorca, Spain)*, April, 25–27.
- [11] Hulley S.B. and Cummings S.R. (2001). *Designing Clinical Research*, second edition. Philadelphia: Lippincott Williams e Wilkins.
- [12] Knuth, D.E. (1986). *The T_EXbook*. New York: Addison - Wesley.
- [13] Landis, J. e Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, **33**, 159–174.
- [14] Pontius, R. G. Jr. (2000). Quantification error versus location errors in comparison of categorical maps. *Photogrammetric Engineering and Remote Sensing*, **66(8)**, 1011–1016.

- [15] Pontius, R. G. Jr. (2002). Statistical Methods to Partition Effects of Quantity and Location During Comparison of Categorical Maps at Multiple Resolutions. *Photogrammetric Engineering and Remote Sensing*, **68**(10), 1041–1049.
- [16] Sneath, P. H. A. e Sokal, R. R. (1973). *Numerical Taxonomy*. San Francisco: W. H. Freeman.
- [17] Szklo, R. Nieto F.J. (2000). Epidemiology, Beyond the Basis. *Aspen Publications*, p. 343–404.