UNIVERSIDADE ESTADUAL PAULISTA - UNESP

CÂMPUS DE JABOTICABAL

ANÁLISE ESPACIAL E TEMPORAL DE ATRIBUTOS DO SOLO E DA PRODUTIVIDADE DAS CULTURAS VISANDO MANEJO ESPECÍFICO EM ÁREAS AGRÍCOLAS

Marcos Sales Rodrigues

Engenheiro Agrônomo

2013

UNIVERSIDADE ESTADUAL PAULISTA - UNESP
CÂMPUS DE JABOTICABAL

ANÁLISE ESPACIAL E TEMPORAL DE ATRIBUTOS DO SOLO E DA PRODUTIVIDADE DAS CULTURAS VISANDO MANEJO ESPECÍFICO EM ÁREAS AGRÍCOLAS

Marcos Sales Rodrigues
Orientador: Prof. Dr. José Eduardo Corá
Co-orientador: Prof. Dr. Thomas Mueller

Tese apresentada à Faculdade de Ciências Agrárias e Veterinárias – Unesp, Câmpus de Jaboticabal, como parte das exigências para a obtenção do título de Doutor em Agronomia (Produção Vegetal).

2013

R696a Rodrigues, Marcos Sales
Análise espacial e temporal de atributos do solo e da produtividade das culturas visando manejo específico em áreas agrícolas / Marcos Sales Rodrigues. – – Jaboticabal, 2013
xii, 90 p. ; 28 cm

Tese (doutorado) - Universidade Estadual Paulista, Faculdade de Ciências Agrárias e Veterinárias, 2013
Orientador: José Eduardo Corá
Banca examinadora: Carlos Eduardo Angeli Furlani, Rouverson Pereira da Silva, Luciano Shozo Shiratsuchi, Gláucia de Mello Pelá
Bibliografia

1. algoritmo de grupamento. 2. calagem. 3. modelo espaço-temporal. 4. sensor de pH. 5. zonas de manejo. I. Título. II. Jaboticabal-Faculdade de Ciências Agrárias e Veterinárias.

CDU 631.42:519.23

Ficha catalográfica elaborada pela Seção Técnica de Aquisição e Tratamento da Informação – Serviço Técnico de Biblioteca e Documentação - UNESP, Câmpus de Jaboticabal.

unesp UNIVERSIDADE ESTADUAL PAULISTA
CAMPUS DE JABOTICABAL
FACULDADE DE CIÊNCIAS AGRÁRIAS E VETERINÁRIAS DE JABOTICABAL

CERTIFICADO DE APROVAÇÃO

TÍTULO: ANÁLISE ESPACIAL E TEMPORAL DE ATRIBUTOS DO SOLO E DA PRODUTIVIDADE DAS CULTURAS VISANDO MANEJO ESPECÍFICO EM ÁREAS AGRÍCOLAS

AUTOR: MARCOS SALES RODRIGUES
ORIENTADOR: Prof. Dr. JOSE EDUARDO CORA
CO-ORIENTADOR: Prof. Dr. THOMAS MUELLER

Aprovado como parte das exigências para obtenção do Título de DOUTOR EM AGRONOMIA (PRODUÇÃO VEGETAL), pela Comissão Examinadora:

Prof. Dr. JOSE EDUARDO CORA
Departamento de Solos e Adubos / Faculdade de Ciências Agrárias e Veterinárias de Jaboticabal

Prof. Dr. CARLOS EDUARDO ANGELI FURLANI
Departamento de Engenharia Rural / Faculdade de Ciências Agrárias e Veterinárias de Jaboticabal

Prof. Dr. ROUVERSON PEREIRA DA SILVA
Departamento de Engenharia Rural / Faculdade de Ciências Agrárias e Veterinárias de Jaboticabal

Prof. Dr. LUCIANO SHOZO SHIRATSUCHI
Empresa Brasileira de Pesquisa Agropecuária / Sinop/MT

Profa. Dra. GLÁUCIA DE MELLO PELÁ
Universidade Estadual de Goiás / Ipameri/GO

Data da realização: 21 de março de 2013.

DADOS CURRICULARES DO AUTOR

MARCOS SALES RODRIGUES – Nascido dia 22 de setembro de 1983, em Ipameri, GO. Em fevereiro de 2003, ingressou no Curso de Graduação em Agronomia pela Universidade Estadual de Goiás – Unidade Universitária de Ipameri, e em fevereiro de 2008, obteve o título de Engenheiro Agrônomo. Foi bolsista de iniciação científica pelo programa PIBIC/CNPq/UEG durante o período de agosto de 2006 a julho de 2007. Paralelamente ao curso de Agronomia, em 2003 ingressou no Curso Superior de Tecnologia em Sistemas de Informação pelo Centro Federal de Educação Tecnológica de Urutaí-GO, atual Instituto Federal Goiano, Câmpus de Urutaí, e em dezembro de 2005, obteve o título de Tecnólogo em Sistemas de Informação. Iniciou em março de 2008 o curso de Mestrado em Agronomia (Ciência do Solo) na Universidade Estadual Paulista – Câmpus de Jaboticabal, SP, onde foi bolsista FAPESP. No dia 22 de janeiro de 2010, submeteu-se à banca para a defesa da Dissertação e obteve o título de Mestre em Agronomia (Ciência do Solo). Iniciou em março de 2010 o curso de Doutorado em Agronomia (Produção vegetal) também na UNESP-Jaboticabal, onde foi bolsista CAPES. Participou de doutorado sanduíche no período de março de 2012 a dezembro de 2012, na University of Kentucky, EUA, sendo bolsista CAPES. No dia 21 de março de 2013, submeteu-se à banca para a defesa da Tese e obteve o título de Doutor em Agronomia (Produção Vegetal).

"Pouca ciência afasta o homem de Deus, porém muita ciência a Deus o conduz."

Possível autor: Louis Pasteur (1822-1895)

Cientista francês que realizou descobertas que tiveram enorme importância na história da química e da medicina.

À minha família,

Moacir Amâncio Rodrigues, Maria de Fátima Sales Rodrigues, Priscila Sales Rodrigues Aquino e Marcelo Sales Rodrigues,

pelo amor, paciência e incentivo.

Dedico.

AGRADECIMENTOS

Ao meu **Senhor Jesus Cristo**, por ter me dado vida, salvação e força necessária para alcançar esta vitória.

À minha namorada **Talita Cristina Mamedes**, pelo seu constante apoio e ajuda nos momentos bons e difíceis.

À **Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES)** pela concessão da bolsa de estudo.

Ao meu orientador **Prof. Dr. José Eduardo Corá**, pelos ensinamentos e correções necessárias ao meu crescimento profissional. Agradeço também, por ter sido o meu maior incentivador no estágio nos EUA. Sem o seu apoio esse período no exterior, que foi de fundamental importância para meu crescimento, não teria acontecido.

Ao meu co-orientador estrangeiro **Prof. Dr. Thomas Mueller**, pela sua paciência, ajuda e compreensão que foi fundamental para o período que estive nos EUA.

À **Dra. Annamaria Castrignanó**, pela ajuda na confecção da presente tese, além de sua amizade muito valorosa.

Aos colegas com quem trabalhei no estágio nos EUA **Blazan Mijatovic**, **Surendran Neelakantan** e **Eduardo Rienzi** pelos constantes auxílios.

Aos meus amigos **Daniel Oliveira**, **Wes Nowell**, **Todd Martin**, **James Riker**, **Xu Zhang**, **Lu Manman**, **Steven Shofner**, **Gao Fei**, **Qin Wang**, **Eduardo Mônaco Gianello**, **José Eduardo Guiotti**, **Isabel Cristina Violatti**, **Rafael Mazza Barbieri**, **João Miarelli Neto, Laz Ude Eze** e **Kevin Yu**, pela amizade, agradável convivência e o apoio nos EUA.

À **Prof. Dra. Carolina Fernandes**, pelo constante auxílio em meus trabalhos durante todo este período em Jaboticabal.

Aos colegas de Pós-Graduação **Adolfo Valente Marcelo**, **Roniran Pereira da Silva**, **Liliane Pereira Campos**, **Getulio Seben Junior**, **Llerme Navarro Vasquez**, **Márcio dos Reis Martins** e **Junior Steve Garrido Perez** pelas constantes ajudas e sugestões que me fizeram crescer profissionalmente durante este período.

Ao técnico de laboratório **Tiago Fieno**, pela sua imprescindível ajuda nos trabalhos de campo e laboratório.

A todos os funcionários do Depto. de Solos e Adubos, sem os quais não seria possível a realização deste trabalho.

Aos meus amigos **Leandro Braz**, **Michelle Roberta Santos**, **Alessandra Domingos** e **Mateus Fernandes**, pela grande amizade durante este período em Jaboticabal.

E todos os demais **amigos** que sempre estiveram do meu lado e me apoiaram em todos os momentos.

SUMÁRIO

Página

ANÁLISE ESPACIAL E TEMPORAL DE ATRIBUTOS DO SOLO E DA PRODUTIVIDADE DAS CULTURAS VISANDO MANEJO ESPECÍFICO EM ÁREAS AGRÍCOLAS

RESUMO – Os atributos do solo e a produtividade das culturas podem variar no tempo e no espaço, assim como suas relações de causa/efeito. O presente trabalho objetivou analisar atributos do solo e da produtividade das culturas visando o manejo específico de áreas agrícolas. Para isto, foram utilizadas três ferramentas: uso de modelos de regressão espaço-temporal para entendimento da relação entre atributos do solo e produtividade do milho; uso de algoritmo de agrupamento 'fuzzy c-means' para delineamento de zonas homogêneas de manejo; e uso de um sensor de pH do solo em tempo real. Os dois primeiros estudos foram realizados em Jaboticabal, SP. Foram avaliados a produtividade de milho e os atributos químicos e físicos do solo de seis anos agrícolas entre os anos de 2001 a 2010. O primeiro estudo avaliou quatro modelos de regressão os quais foram: mínimos quadrados ordinários; mínimos quadrados generalizados assumindo heteroscedasticidade; modelo espacial e temporal, assumindo homocedasticidade; e modelo espacial e temporal, assumindo heteroscedasticidade. Verificou-se que o modelo espacial e temporal, assumindo heteroscedasticidade foi superior aos demais e que a acidez do solo (pH) foi o fator que mais influenciou a variabilidade da produtividade de milho em todos os anos agrícolas. No segundo estudo, verificou-se que o uso de algoritmo de agrupamento ''fuzzy c-means' foi eficiente para delinear zonas homogêneas de manejo, quando foi considerada a variabilidade espacial e temporal dos atributos do solo e da produtividade de milho. O terceiro estudo foi conduzido em três campos de produção agrícola nos EUA, e o objetivo foi avaliar o sensor de pH em tempo real Veris (MSP)® em Kentucky, avaliando a acurácia das medidas de pH, 'buffer pH' e a necessidade de calagem usando técnicas analíticas e propondo um novo método de correção de erros de interpolação. Verificou-se que o sensor de pH foi eficiente para a produção de mapas de pH do solo e, consequentemente, de mapas de necessidade de calagem. O método proposto para a correção dos erros de interpolação utilizando um banco de dados de calibração, foi eficiente para a correção de mapas de necessidade de calagem. Verificou-se por meio dos resultados dos três estudos, que as ferramentas de Agricultura de Precisão estudadas são eficientes para o manejo específico de áreas agrícolas.

Palavras-Chave: algoritmo de grupamento, calagem, modelo espaço-temporal, sensor de pH, *Zea mays* L., zonas de manejo.

SPATIAL AND TEMPORAL ANALYSIS OF SOIL ATTRIBUTES AND CROP YIELD TO APPLY SITE-SPECIFIC MANAGEMENT ON CROP FIELDS

ABSTRACT – Soil attributes and crop yield may vary over space and time as well as their cause-effect relationships. Therefore, the objectives of this study were analyze soil attributes and crop yield to apply site-specific management on crop fields, in this regard, it was used three techniques, which were: models that accounted for spatial-temporal autocorrelation to evaluate the relationship between corn yield and soil attributes; identify management zones using fuzzy c-means clustering analysis based on the spatial and temporal variability of soil attributes and corn yield, and use the Veris (MSP)® on-the-go soil sensing system to lime requirement recommendation. The first two studies were carried out in Jaboticabal, SP, Brazil. Soil chemical and physical attributes were measured between 2001 and 2010 year. Ordinary least square, generalized least squares assuming heteroscedasticity, spatial-temporal least squares assuming homoscedasticity, and spatial-temporal model assuming heteroscedasticity analyses were used to estimate corn yield in the first study. The spatial and temporal autocorrelation assuming heteroscedasticity was superior to the other studied model for prediction and soil acidity (pH) was the factor that most influenced corn yield over time in this study. In the second study, which the fuzzy c-means cluster analysis based on the spatial and temporal variability of soil attributes and corn yield was used, it was observed that this technique was efficient to delineate management zones. The third study was carried out in three crop fields in Kentucky, EUA, and it aimed to evaluate the Veris on-the-go pH sensor system for soil pH management by assessing the accuracy of soil pH measurements and associated spatial predictions of buffer pH, and lime requirement proposed a new method for correcting interpolation errors. The pH sensor was efficient to map pH and lime requirement recommendation. The approach that corrects interpretations with the calibration data set was efficient to correct lime requirement recommendation. The results of the three studies showed that studied Precision Agriculture techniques was efficient to apply site-specific management on crop fields.

Keywords: clustering algorithm, lime requirement, spatial-temporal models, pH sensor, *Zea mays* L., management zones

CAPÍTULO 1 - Considerações gerais

1.1. Introdução

A globalização da agricultura tem mudado os conceitos de como conduzir a produção agrícola tornando-se um desafio cada vez maior dado a pressão dos mercados internacionais por produtos competitivos. Adicionalmente, a preocupação com o ambiente e as mudanças climáticas tem pressionado a agricultura para que suas técnicas sejam sócio-ambientalmente corretas. Portanto, torna-se imprescindível o uso de tecnologias que possibilitem o aumento da produtividade das culturas, o uso racional de insumos agrícolas, a minimização dos impactos ambientais e a maximização dos lucros do produtor agrícola. Desta forma, as tecnologias preconizadas na agricultura de precisão (AP) tem se mostrado com grande potencial para alcançar tais objetivos.

O princípio da AP é que os campos de produção agrícola não são homogêneos, portanto, a variabilidade espacial da produtividade das culturas e dos fatores que controlam esta variabilidade devem ser exaustivamente estudados. Deste modo, ferramentas que permitam o entendimento da variabilidade da produtividade das culturas e suas relações no tempo e no espaço, assim como os fatores que a controlam, devem ser desenvolvidas e testadas.

Os atributos do solo e sua relação com a produtividade das culturas tem sido largamente estudados, porém, muitas das ferramentas aplicadas para entender esta relação não levam em consideração a variabilidade espacial e temporal desses fatores, como é o caso das regressões simples, que são amplamente utilizadas neste tipo de estudo. Porém, estudos tem mostrado que a relação entre atributos do solo e produtividade das culturas é um processo dinâmico no espaço e no tempo. Por isso, ferramentas que levam em consideração a variabilidade espaço-temporal devem ser utilizadas para entender as relações de causa/efeito entre atributos do solo e produtividade das culturas. Portanto, este será o tema abordado no capítulo 2 deste trabalho, no qual modelos espaço-temporal serão testados para entender a relação entre a produtividade da cultura de milho e os atributos do solo.

Conhecendo os fatores que controlam a variabilidade das culturas no tempo e no espaço, é, então, necessário que este conhecimento permita a aplicação de técnicas que ajudem o produtor agrícola na tomada de decisão quanto ao manejo do solo e das culturas. Umas das ferramentas que podem ser utilizadas para tal propósito é o delineamento de zonas homogêneas de manejo. Porém, sabendo-se que existe uma relação dinâmica entre a produtividade das culturas e atributos do solo que variam no espaço e no tempo, tais zonas de manejo devem ser delineadas baseadas em dados que permitam a avaliação, não somente espacial, mas também temporal, dos atributos do solo e da produtividade das culturas. É importante também que o processo de definição de zonas de manejo, seja, relativamente, simples e prático. Desta forma, é necessário o uso de técnicas que permitam a sugestão do número de zonas para determinada área agrícola, como é o caso do delineamento de zonas homogêneas de manejo baseadas em algoritmos de grupamento 'fuzzy', o qual será o tema abordado no capítulo 3.

Um dos fatores limitantes da AP é o grande número de amostras necessárias para a confecção de mapas de atributos do solo com adequada acurácia. Tornando tal processo, muitas vezes, inviável devido ao alto custo das análises. Desta forma, a utilização de sensores que possam aumentar o número de amostras a baixo custo, o que, consequentemente, aumentaria a acurácia dos mapas de atributos do solo, devem ser estudados. Dos diversos sensores utilizados na AP, destaca-se o sensor de pH em tempo real Veris 'Mobile Sensor Platform' (MSP)®, que permite, por meio de dois eletrodos de antimônio a leitura em tempo real do pH do solo no campo, o que permite o aumento do número de amostras por hectare, o que, teoricamente, permite melhor acurácia dos mapas de pH. Um estudo realizado com este sensor em três campos agrícolas no estado de Kentucky, EUA, será o tema do capítulo 4 do presente trabalho.

Desta forma, o trabalho objetivou analisar atributos do solo e da produtividade do milho visando o manejo específico de áreas agrícolas, utilizando-se três ferramentas: modelos de regressão espaço-temporal para entendimento da relação entre atributos do solo e produtividade do milho; uso de algoritmo de grupamento 'fuzzy' para delineamento de zonas homogêneas de manejo; e uso de um sensor de pH do solo em tempo real.

1.2. Revisão de Literatura

1.2.1. Variabilidade espacial e temporal de atributos do solo e da produtividade das culturas

A variabilidade espacial, tanto dos atributos do solo como da produtividade das culturas tem sido estudada (KRAVCHENKO e BULLOCK, 2000; SCHEPERS et al., 2004; RODRIGUES, CORÁ e FERNANDES, 2012). A principal função do estudo da variabilidade dos atributos do solo e da produtividade das culturas está conferida a aplicação de adubos e corretivos à taxas variáveis (CORÁ e BERALDO, 2006) e a definição de zonas homogêneas de manejo (CORÁ et al., 2004).

Resultados divergentes tem sido obtidos em trabalhos que estudam a relação entre a produtividade das culturas e os atributos do solo. Em diversos trabalhos é encontrada baixa correlação entre produtividade das culturas e atributos do solo. Santos et al. (2006) verificaram que, praticamente, não foram observadas correlações lineares nem espaciais entre os atributos do solo e a produtividade da cultura e, quando ocorreram, os valores dos coeficientes de correlação foram considerados baixos em um Latossolo Vermelho distroférrico sob semeadura direta. Bourennane et al. (2004) observaram baixa correlação entre a produtividade de trigo e atributos do solo em um 'Xerifluventic haplocambids' e um 'Udorthent'.

Mallarino, Oyarzabal e Hinz (1999) observaram baixa correlação entre atributos químicos e a produtividade de milho em cinco campos de produção agrícola nos EUA. Contrariamente a estes resultados, Rodrigues, Corá e Fernandes (2012) estudando a relação entre produtividade e atributos do solo em intensidade amostral similar, encontraram alta correlação entre a produtividade de milho e atributos do solo em um Latossolo Vermelho distroférrico sob sistema de semeadura direta. Demonstrando que há alta correlação entre produtividade das culturas e atributos do solo quando avaliados em similar intensidade amostral.

McBratney et al. (2005) estudando as diretrizes futuras da AP, ressaltaram que o conhecimento do aspecto temporal necessita ser grandemente incrementado na aplicação da AP. Isto porque os mapas de produtividade podem variar substancialmente de ano para ano como comprovado por Milani et al. (2006), que verificaram que, de forma geral, os mapas de produtividade apresentaram um padrão de variabilidade pouco semelhante quanto à distribuição na área de estudo ao longo dos cinco anos de avaliação, confirmando a existência de variabilidade temporal da produtividade. Blackmore, Godwin e Fountas (2003), ao avaliar mapas de produtividade de seis anos, observaram variabilidade espacial aleatória da produtividade nas áreas de estudo. Portanto, a análise dos dados de um só ano poderia, potencialmente, conduzir a erros de decisão no manejo da lavoura. Esta análise temporal da produtividade tornou-se possível devido ao desenvolvimento de sensores de obtenção de mapas de produtividade, fazendo com que a fase de obtenção de dados de produtividade tornasse relativamente simples e economicamente viável (BAZZI et al., 2008).

A mudança da estrutura espacial da produtividade das culturas ao longo dos anos está estritamente ligada com as condições climatológicas de cada ano. Amado et al. (2007), verificaram que em anos de déficit hídrico, ocorreu aumento da variabilidade da produtividade. Concordando com estes resultados, Guedes Filho et al.(2010) observaram que a variabilidade da produtividade das culturas ao longo de 23 anos, foi mais alta nos anos com déficit hídrico e/ou distribuição pluviométrica irregular.

A alteração na estrutura espacial pode ocorrer porque segundo Diker Heermann e Brodahl (2004), a dominância dos fatores que influenciam a variabilidade da produtividade das culturas podem mudar de ano para ano devido às condições climatológicas. Indicando que as relações entre produtividade das culturas e atributos do solo podem sofrer alterações de ano para ano. Timlin et al. (1998) verificaram que a produtividade de milho foi correlacionada com os atributos P, K^+ e matéria orgânica (MO) somente em anos secos em um 'Typic Fragiochrept. Bakhsh' et al. (2000) afirmaram que é necessário incluir as variáveis climáticas e os fatores de manejo nos modelos de predição de produtividade das culturas para um detalhado diagnóstico dos fatores que limitam a produtividade.

Dessa maneira, estudos que objetivam o manejo especifico de lavouras devem levar em consideração, tanto a variabilidade espacial como a variabilidade temporal dos fatores que influenciam a produtividade das culturas. Assim como, considerar a influência de fatores climáticos que comumente mudam de ano para ano.

1.2.2. Uso de modelos estatísticos para estimação da produtividade das culturas

Um dos métodos mais utilizados para descrever o relacionamento entre produtividade das culturas e atributos do solo são os modelos de regressão. Eles consistem, basicamente, de técnicas que modelam e analisam o relacionamento entre uma variável dependente com uma ou mais variável independente.

Um modelo estatístico de regressão consiste dos componentes: resposta, que é o resultado da variável de interesse estudada, como por exemplo a produtividade de uma determinada cultura (geralmente se usa **Y** como notação na equação); parâmetros, que podem ser quaisquer constantes desconhecidas na função média ou da distribuição de variáveis aleatórias; parte sistemática, que é a função média do modelo; erro do modelo, que é a diferença entre a observação e a função média; predição, que é a avaliação da função média nos valores estimados dos parâmetros; e o ajuste dos resíduos que é a diferença entre valores observados e ajustados. A grande maioria dos modelos usados em variáveis em ciências agrárias são baseados na estrutura resposta = estrutura + erro (SCHABENBERGER e PIERCE, 2001).

Os modelos podem ser classificados como lineares, quando se considera que a relação da resposta às variáveis é uma função linear de alguns parâmetros e não-lineares, quando não são uma função linear dos parâmetros. Os modelos podem ser também classificados como de efeito fixos, efeito aleatórios e mistos. A diferença entre efeitos fixos e aleatórios são aplicados nos componentes não-conhecidos do modelo, ou seja, um efeito fixo é uma constante não conhecida (não varia) e o efeito aleatório é uma variável aleatória. Os efeitos aleatórios são advindos de subamostragens, seleção aleatória de níveis de tratamento e processos aleatórios hierárquicos. Nos modelos de efeito fixo todos os efeitos são fixados (além do erro), enquanto que os modelos de efeito aleatório todos os efeitos são aleatórios (além do intercepto). Os modelos chamados mistos possuem alguns efeitos fixados e outros aleatórios (sem contar o intercepto e o erro do modelo) (SCHABENBERGER e PIERCE, 2001).

Algumas suposições estatísticas devem ser levados em conta para o uso de modelos de regressão, as quais são: distribuição normal e independência dos erros; linearidade; ausência de colinearidade; confiabilidade das medidas; e homocedasticidade da variância (OSBORNE e WATERS, 2002). Infelizmente, essas suposições tem sido frequentemente ignoradas na literatura.

Em estudos que relacionam produtividade das culturas e atributos do solo, a regressão linear que faz uso do método de estimação dos mínimos quadrados ordinários (sigla em inglês: OLS - 'Ordinary least squares') é a mais amplamente utilizada. Os métodos de estimação de mínimos quadrados ordinários buscam minimizar a soma de quadrados das diferenças entre valores observados e estimados (FARACO et al., 2008). Esse método de regressão levam em consideração a normalidade e independência dos erros e a homocedasticidade da variância.

Em trabalho realizado por Santos et al. (2006), foi utilizado a regressão linear múltipla para explicar a relação entre produtividade de milho em um sistema de semeadura direta e os atributos densidade do solo, densidade de partículas e porosidade total. Semelhantemente, Megda et al. (2008) utilizaram a regressão múltipla para relacionar a produtividade de feijão em um sistema de semeadura direta com os atributos macroporosidade, microporosidade e porosidade total do solo. Martins et al. (2009) fizeram uso da mesma técnica para, também, explicar o relacionamento entre a produtividade de feijão e atributos físicos do solo, assim como Rosa Filho et al. (2009) que utilizaram esse tipo de regressão para explicar a produtividade de soja por meio de atributos físicos do solo. A regressão linear também foi utilizada para explicar a produtividade de soja sob sistema convencional de manejo de solo, baseado em atributos físicos e químicos, em estudo conduzido por Reichert et al. (2008). Em todos esses estudos que utilizaram da regressão do mínimos quadrados ordinários, a dependência espacial não foi considerada. Contudo, os resíduos dos dados de produtividade das culturas são geralmente espacialmente correlacionados (LARK, 2000; LOBELL et al., 2005). Quando se ignora a dependência espacial nos parâmetros da regressão OLS, a estimação da variância pode ser significativamente impactada e, então, pode ser fornecida conclusões erradas sobre os fatores que foram estatisticamente significantes.

Os modelos espaciais, ou seja, aqueles que levam em consideração a dependência espacial das amostras, assume que os erros (os elementos do **e** no modelo misto $Y = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}$) possuem correlação. A correlação espacial pode ser refletida no **G**, a matriz de covariância do modelo de efeitos aleatórios, **u**, ou em **R**, a matriz de covariância do erro do modelo (LITTELL et al., 2006). Apesar de ainda escassos na literatura, existem trabalhos que apontam que a incorporação da variabilidade espacial nos modelos de regressão mistos podem melhorar o entendimento dos fatores que afetam a produtividade das culturas (HURLEY, MALZER e KILIAN, 2004; LAMBERT, LOWENBERG-DEBOER e BONGIOVANNI, 2004). Lambert, Lowenberg-Deboer e Bongiovanni (2004) observaram que regressões espaciais apresentaram maior confiabilidade estatística para entender a relação produtividade de milho e nitrogênio do que a regressão OLS.

A estrutura da dependência espacial pode mudar ao longo dos anos de estudo e, por consequência, alterar a relação entre produtividade das culturas e atributos do solo, portanto, a regressão OLS não é a mais apropriada para esse caso. Desta forma, a regressão que usa como método de estimação, o método dos mínimos quadrados generalizados (sigla em inglês: GLS - 'generalized Least Squares') deve ser utilizada (FLINN e DE DATTA, 1984). O método GLS é aplicado quando a variância dos erros não é a mesma (heteroscedasticidade), ou quando há certa correlação entre as observações. Quando se ignora a heteroscedasticidade, erros do Tipo I podem ocorrer (OSBORNE e WATERS, 2002), ou seja, pode aumentar a chance de fatores significativos serem incorretamente excluídos do modelo. Isso poderia levar a críticos erros de interpretação e consequentemente poderia conduzir a incorretas decisões de manejo. Por exemplo, Hurley et al. (2004) verificaram que o melhor modelo para descrever a resposta de N na produtividade de milho foi àquele que levou em consideração a heteroscedasticidade e a correlação espacial.

Baseado nestas informações, verifica-se que é importante testar a viabilidade do uso de modelos estatísticos, para estimar a produtividade das culturas, os quais sejam mais robustos, tais como os que levam em consideração a heteroscedasticidade e a correlação espacial das variáveis.

1.2.3. Algoritmos de agrupamentos 'fuzzy' para determinação de zonas de manejo

Algumas importantes questões precisam ser respondidas para realizar o delineamento de zonas homogêneas de manejo, as quais são: quais as informações devem ser utilizadas para definir zonas em campos de produção agrícola? Como essas informações devem ser processadas para a obtenção das zonas? Em quantas zonas o campo deve ser dividido? (FRIDGEN et al., 2004). Uma técnica que possui potencial para auxiliar nas respostas destas questões são os algoritmos de grupamentos 'fuzzy', que são ferramentas matemáticas utilizadas para detectar similaridade entre membros de uma coleção de objetos (FRAISSE, SUDDUTH e KITCHEN, 2001; SANTOS et al., 2003).

Uma técnica de grupamento que tem sido amplamente utilizada é a classificação algorítmica não supervisionada 'Fuzzy c-means'. Como sendo uma classificação não supervisionada, não há a exigência do uso de uma área de treinamento conhecida como é o caso da classificação supervisionada. Adicionalmente, não é requerido que as variáveis utilizadas na classificação possuam variância similar e distribuição gaussiana (IRVIN, VENTURA e SLATER, 1997). O algoritmo 'c-means' é baseado na minimização de uma função objetiva definida como a soma dos quadrados das distâncias de todos os pontos no domínio de um grupamento ('cluster') até o grupamento central. Este é um processo interativo e o algoritmo termina quando um específico critério de convergência é alcançado (FRIDGEN et al., 2004).

Objetivando determinar zonas de manejo para uso em agricultura de precisão, baseado em atributos topográficos e condutividade elétrica do solo, Fraisse, Sudduth e Kitchen (2001), utilizaram o algoritmo 'fuzzy' não supervisionado 'c-means'. Estes autores verificaram que esta técnica de classificação é rápida e facilmente implementada com os atuais Sistemas de Informações Geográficas (SIG) comercialmente disponíveis. Kitchen et al. (2005) utilizaram o algoritmo 'fuzzy' não supervisionado 'c-means' para delinear zonas homogêneas de manejo utilizando dados de condutividade elétrica (CE) e elevação e compararam com os dados de produtividade de dois campos em Missouri, EUA. Estes autores verificaram que os dados de CE e elevação relacionaram com a produtividade das culturas e, portanto, poderiam ser utilizados para definição de zonas homogêneas de manejo.

Em estudo conduzido por Santos et al. (2003), foi utilizado o algoritmo 'fuzzy' não supervisionado para definir zonas de manejo em um campo de milho no estado de São Paulo, fazendo uso de atributos do solo e planta. Estes autores verificaram que essa técnica mostrou-se adequada para o reconhecimento de padrões de agrupamento nos atributos do solo. Li et al. (2007) observaram que a técnica 'fuzzy' pode ser utilizada para classificar zonas de manejo já que a variabilidade espacial de atributos químicos do solo mostraram-se concordantes com o padrão de distribuição espacial da produtividade das culturas em um campo de produção agrícola na China. Tagarakis et al. (2013) utilizaram do algoritmo 'fuzzy' não supervisionado 'c-means' para delinear zonas homogêneas de manejo, não apenas baseados na produtividade, mas também na qualidade de frutos da uva na Grécia central.

Para facilitar o uso, comercialmente, desta técnica no delineamento de zonas homogêneas de manejo , Fridgen et al. (2004), desenvolveram um programa de computador que utiliza o algoritmo 'fuzzy' não supervisionado chamado de 'Management Zone Analyst' (MZA). Este programa tem sido amplamente utilizado em pesquisas e comercialmente, isto porque o MZA permite a produção de zonas de manejo de forma simples e rápida. Adicionalmente, o MZA sugere o número ótimo de zonas de manejo para a área avaliada por meio de dois índices, o índice de performace fuzzy (sigla em inglês: FPI - 'fuzziness performance index') (ODEH, CHITTLEBOROUGH e MCBRATNEY, 1992) e a classificação de entropia normalizada (sigal em inglês: NCE - 'normalized classification entropy') (BEZDEK, 1981). Jiang, Fu e Wang (2011) afirmaram que o uso deste programa pode reduzir a desvantagem do alto nível teórico necessário para o uso do algoritmo 'fuzzy' não supervisionado, além de reduzir o número de análises para obter os resultados. Portanto, o MZA permite alta precisão e velocidade na obtenção de resultados para o delineamento de zonas de manejo, sem com que se perca a cientificidade do método.

1.2.4. Uso do sensor de pH do solo em tempo real

Um dos mais importantes fatores que pode impactar a produtividade das culturas na maioria dos campos de produção agrícola no Brasil, são àqueles relacionados com a acidez do solo, tais como a saturação por bases (AMADO et al., 2009), saturação por magnésio (NOGARA NETO et al., 2011) e o pH do solo (DALCHIAVON et al., 2011). Isto ocorre devido aos baixos valores de pH e os altos valores de Al^{3+}, comumente associados com Latossolos, os quais são prevalentes em regiões de produção agrícola no Brasil (MUNIZ et al., 2011).

A técnica mais comumente utilizada para se conhecer a variabilidade espacial de atributos do solo, e portanto, o pH do solo, é o uso de malhas amostrais (ADAMCHUK, MORGAN e ESS, 1999). Porém, os mapas de atributos do solo obtidos a partir de malhas amostrais não tem apresentado boa correlação com os padrões de distribuição da variabilidade da produtividade das culturas como mencionado no item 1.1.1. Mesmo amostragens consideradas intensas como as que são praticadas nos EUA, onde são coletadas uma amostra por hectare, não tem se mostrado eficientes para obter mapas de atributos do solo que permitam a aplicação de fertilizantes e corretivos à taxa variável com a acurácia necessária (MUELLER et al., 2001; ADAMCHUK et al., 2007).

Para o pH do solo, por exemplo, pode se observar variações de seus valores de 3,9 a 6,4, com coeficiente de variação de até 12%, em Latossolos, que são considerados taxionomicamente homogênios (RODRIGUES, CORÁ e FERNANDES, 2012). Essa variabilidade dos atributos químicos do solo, tais como o pH, faz com que seja difícil a obtenção de mapas com acurácia quando se usa baixa intensidade amostral. Desta forma, alta variabilidade produz baixa acurácia em locais não amostrados (ADAMCHUK, MORGAN e LOWENBERG-DEBOER, 2004).

Em estudos realizados por Rodrigues, Corá e Fernandes (2012) em um Latossolo Vermelho distroférrico sob cultivo de milho em sistema de semeadura direta, os autores verificaram que a correlação espacial entre produtividade das culturas e atributos do solo foi alta somente quando foram avaliados em intensidade amostral similar. Contudo, o aumento da intensidade amostral se torna impraticável devido aos elevados custos da amostragem e análises laboratoriais (ADAMCHUK et al., 2007). Portanto, já que a amostragem georreferenciada de solos ainda é um desafio não solucionado (LUND et al., 2004), uma alternativa seria o desenvolvimento de sensores que permitam a avaliação dos atributos do solo com maior intensidade amostral e que seja economicamente viável.

Em 2003, um sensor que mede em tempo real o pH do solo, Veris MSP® (VERIS-TECHNOLOLOGIES, 2010), foi lançado comercialmente nos EUA, e já tem sido largamente utilizado em campos de produção agrícola nos EUA (ERICKSON, 2004) e no mundo. Esse sensor tem aumentado significativamente a acurácia dos mapas de pH, pois, é possível aumentar consideravelmente a densidade de amostras (ADAMCHUK, MORGAN e LOWENBERG-DEBOER, 2004).

Esse sensor faz o uso de dois eletrodos de antimônio que mede o pH de amostras de solo, que são coletadas no campo em tempo real. Estudos preliminares, como os realizados por Adamchuk, Morgan e Ess (1999) mostraram alta correlação (R^2 = 0,92) entre os valores de pH do solo obtidos pelos eletrodos e aqueles obtidos pela análise padrão em laboratório.

Concordando com esses resultados, Lund et al. (2004) verificaram alta correlação (R^2 = 0,81) entre os valores de pH do solo obtidos com o sensor e aqueles obtidos por meio de análises de laboratório padrão quando se estudou 15 campos de produção agrícola nos EUA. Em estudos realizados em vários campos de produção agrícola sob diferentes tipos de solo na Alemanha, Olfs, Borchert e Trautz (2010) observaram coeficientes de determinação das regressões entre os valores de pH do solo obtidos com o sensor e aqueles por meio de análises em laboratório variando entre 0,36 a 0,79.

Esse sensor pode ser operado a 8 km h^{-1}, realizando uma amostra a cada 20 m, durante a realização de medições a cada 10 s. Isso resulta em torno de 20 amostras por ha (ADAMCHUK et al., 2007). Contudo, as configurações do sistema podem ser alteradas, assim como a velocidade de operação, desta forma, o número de amostras por hectare pode ser modificado.

Comparando a distribuição de calcário a taxa variável utilizando dados obtidos com sensor de pH em tempo real e amostragem tradicional utilizando-se de malhas com intensidade de uma amostra por hectare, Adamchuk, Morgan e Lowenberg-

Deboer (2004) verificaram um benefício anual de $ 6,13 ha^{-1} quando se utilizou os dados do sensor de pH do solo.

1.2.5. Interpoladores aplicados às ciências agrárias

A interpolação é uma técnica matemática que ajusta uma função à pontos não amostrados, baseando-se em valores obtidos em pontos amostrados. Esta técnica sido largamente utilizada na agricultura de precisão para a confecção de mapas de atributos do solo, recomendação de fertilizantes e corretivos, produtividade das culturas entre outros. Os métodos de interpolação podem ser divididos em determinísticos, o qual é usado funções matemáticas para interpolação e; geoestatísticos, a qual se baseia tanto em métodos estatísticos que consideram a dependência espacial das amostras e podem ser utilizados para criar mapas de superfícies e também permite avaliar a incerteza da predição (JOHNSTON et al., 2001).

Os métodos de interpolação determinísticos podem ser divididos em dois grupos: global e local. As técnicas globais calculam a predição usando todo o banco de dados, enquanto que as técnicas locais calculam a predição com a medida dos pontos no alcance estabelecido para os vizinhos (JOHNSTON et al., 2001). O presente trabalho teve o seu foco em descrever os métodos locais, tais como: o Inverso do quadrado da distância (IQD), Interpolação polinomial local (IPL) e Funções de base radial (FBR).

Um interpolador determinístico pode forçar ou não, o resultado da interpolação a passar pelo valor real. A técnica de interpolação que prediz um valor idêntico com o valor medido na mesma posição de amostragem é conhecido como interpolador exato. Um interpolador inexato prediz um valor que pode ser diferente do valor medido na mesma posição. O IQD e FBR são interpoladores exatos, enquanto que o IPL é inexato (JOHNSTON et al., 2001).

O inverso do quadrado da distância implementa a suposição de que um valor de um atributo em um local não amostrado é a média ponderada de pontos conhecidos dentro dos vizinhos circundantes ao local não amostrado (BURROUGH e MCDONNELL, 1998). A interpolação polinomial local ajusta um polinômio de uma ordem específica (primeira, segunda e terceira ordem), usando somente os valores dos vizinhos previamente definidos. Os vizinhos se sobrepõem e o valor usado para cada predição é o valor do polinômio ajustado no centro do valor do vizinho (JOHNSTON et al., 2001). O método de interpolação Funções de base radial, consistem de polinômios que descrevem partes de uma linha ou superfície, que são, então, encaixados de modo que os pontos se juntem por meio de uma suavização (WEBSTER e OLIVER, 2007). Informações detalhadas sobre cada um desses interpoladores podem ser encontradas em Burrough e McDonnell (1998), Webster e Oliver (2007) e Johnston et al. (2001).

A Krigagem diferencia-se dos outros métodos de interpolação pela forma como os pesos são distribuídos nas diferentes amostras. Na Krigagem, o procedimento é semelhante ao de interpolação por média móvel ponderada, contudo, nela, os pesos são determinados a partir de uma análise espacial, baseada no semivariograma experimental (ISAAKS e SRIVASTAVA, 1989). Para se utilizar a Krigagem é necessário que exista dependência espacial entre as amostras vizinhas, expressa no semivariograma, para estimar valores em qualquer posição dentro do campo, sem tendência e com variância mínima (VIEIRA, 2000). Informações detalhadas sobre o interpolador Krigagem podem ser encontradas em Isaaks e Srivastava (1989) e Vieira (2000).

Os dois interpoladores mais comumente utilizados na agricultura de precisão são o IQD e a Krigagem (MUELLER et al., 2004). Diversos estudos tem comparado o IQD e a Krigagem. Em alguns casos a Krigagem apresentou melhor performace na predição de atributos do solo do que o IQD (KRAVCHENKO e BULLOCK, 1999; CORÁ e BERALDO, 2006; COELHO et al., 2009). Em outros o IQD apresentou melhores resultados do que a Krigagem (WEBER e ENGLUND, 1992; NALDER e WEIN, 1998). Contudo, o que é observado, frequentemente, é uma mistura de resultados entre os dois métodos (MUELLER et al., 2001; SCHLOEDER, ZIMMERMAN e JACOBS, 2001; SOUZA et al., 2010).

Outros métodos de interpolação como o IPL e o FBR poderiam ser mais utilizados, já que eles são facilmente implementados utilizando os atuas SIGs. Porém, estudos que utilizam o IPL e o FBR para manejo específico da fertilidade do solo ainda são escassos. Silva, Lima e Oliveira (2010) encontraram que o IQD e a

Krigagem foram melhores do que a IPL, quando avaliaram a variabilidade espacial do pH do solo em um Latossolo Vermelho-Amarelo distrófico sob um campo de cultivo de café. Zandi, Ghobakhlou e Sallis (2011) verificaram que o FBR foi o mais adequado método de interpolação para predizer e mapear a distribuição espacial do pH do solo em um campo de produção de uvas.

Para determinar qual o melhor interpolador a se utilizar é preciso utilizar algumas medidas de qualidade de mapa. A qualidade de um mapa é tipicamente determinada pela comparação de valores observados nos mapas com aqueles obtidos por várias técnicas analíticas, tanto quantitativas como qualitativas. Gráficos de valores preditos vs. valores medidos devem ser visualmente examinados para avaliar a qualidade de predição de um determinado interpolador. Uma das medidas quantitativas é a precisão do mapa que é calculada como sendo o desvio padrão dos resíduos. A acurácia do mapa é uma outra medida quantitativa de qualidade, que consiste do quadrado do viés (erro sistemático) e o viés da média dos resíduos. O erro quadrático médio é a soma da precisão e da acurácia (MUELLER et al., 2004).

A escolha do interpolador é altamente dependente do método usado para calcular os valores preditos e medidos (MUELLER et al., 2004). A validação cruzada com substituição de valores é um procedimento rápido e economicamente viável para comparar valores preditos e medidos, contudo ela não é adequada para descrever os erros da predição espacial em várias situações (ISAAKS e SRIVASTAVA, 1989; MUELLER et al., 2001). Portanto, o uso de um banco de dados independente para a validação de mapas é, geralmente, superior que a validação cruzada comumente utilizada.

Sabendo-se que a escolha do método de interpolação dos dados de atributos do solo são essenciais para obtenção de mapas com precisão e acurácia, diferentes métodos de interpolação devem ser testados em detrimento dos que já são amplamente utilizados. Adicionalmente, a escolha do método de avaliação da qualidade dos mapas obtidos com os interpoladores devem ser cuidadosamente testados.

1.3. Referências

ADAMCHUK, V.; LUND, E.; REED, T.; FERGUSON, R. Evaluation of an on-the-go technology for soil pH mapping. **Precision Agriculture**, New York, v. 8, n. 3, p. 139-149, 2007.

ADAMCHUK, V. I.; MORGAN, M. T.; LOWENBERG-DEBOER, J. A model for agro-economic analysis of soil pH mapping. **Precision Agriculture**, New York, v. 5, n. 2, p. 111-129, 2004.

ADAMCHUK, V. I.; MORGAN, M. T.; ESS, D. R. An automated sampling system for measuring soil pH. **American Society of Agricultural and Biological Engineers. Transactions**, Saint Joseph, v. 42, n. 4, p. 885-891, 1999.

AMADO, T. J. C.; PES, L. Z.; LEMAINSKI, C. L.; SCHENATO, R. B. Chemical and physical attributes of oxisols and their relation with irrigated corn and common bean yields. **Revista Brasileira de Ciência do Solo**, Viçosa, v. 33, n. 4, p. 831-843, 2009.

AMADO, T. J. C.; PONTELLI, C. B.; SANTI, A. L.; VIANA, J. H. M.; SULZBACH, L. A. D. S. Variabilidade espacial e temporal da produtividade de culturas sob sistema plantio direto. **Pesquisa Agropecuária Brasileira**, Brasília, v. 42, n. 8, p. 1101-1110, 2007.

BAKHSH, A.; JAYNES, D. B.; COLVIN, T. S.; KANWAR, R. S. Spatio-temporal analysis of yield variability for a corn-soybean field in Iowa. **American Society of Agricultural and Biological Engineers. Transactions**, Saint Joseph, v. 43, n. 1, p. 31-38, 2000.

BAZZI, C. L.; SOUZA, E. G.; OPAZO, M. A. U.; NÓBREGA, L. H. P.; PINHEIRO NETO, R. Influência da distância entre passadas de colhedora equipada com monitor de colheita na precisão dos mapas de produtividade na cultura do milho. **Engenharia Agrícola**, Jaboticabal, v. 28, p. 355-363, 2008.

BEZDEK, J. C. **Pattern Recognition with Fuzzy Objective Function Algorithms**. Kluwer Academic Publishers: Norwell, 1981. p. 256.

BLACKMORE, S.; GODWIN, R. J.; FOUNTAS, S. The analysis of spatial and temporal trends in yield map data over six years. **Biosystems Engineering**, Camden, v. 84, n. 4, p. 455-466, 2003.

BOURENNANE, H.; NICOULLAUD, B.; COUTURIER, A.; KING, D. Exploring the spatial relationships between some soil properties and wheat yields in two soil types. **Precision Agriculture**, New York, v. 5, n. 5, p. 521-536, 2004.

BURROUGH, P. A.; MCDONNELL, R. A. **Principles of Geographical Information Systems**. Oxford: Clarendon, 1998. p 34.

COELHO, E. C.; SOUZA, E. G. D.; URIBE-OPAZO, M. A.; PINHEIRO NETO, R. Influência da densidade amostral e do tipo de interpolador na elaboração de mapas temáticos. **Acta Scientiarum. Agronomy**, Maringá, v. 31, n. 1, p. 165-174, 2009.

CORÁ, J. E.; ARAUJO, A. V.; PEREIRA, G. T.; BERALDO, J. M. G. Variabilidade espacial de atributos do solo para adoção do sistema de agricultura de precisão na cultura de cana-de-açúcar. **Revista Brasileira de Ciência do Solo**, Viçosa, v. 28, n. 6, p. 1013-1021, 2004.

CORÁ, J. E.; BERALDO, J. M. G. Variabilidade espacial de atributos do solo antes e após calagem e fosfatagem em doses variadas na cultura de cana-de-açúcar. **Engenharia Agrícola**, Jaboticabal, v. 26, n. 2, p. 374-387, 2006.

DALCHIAVON, F. C.; CARVALHO, M. P.; FREDDI, O. S.; ANDREOTTI, M.; MONTANARI, R. Variabilidade espacial da produtividade do feijoeiro correlacionada com atributos químicos de um Latossolo Vermelho Distroférrico sob sistema de semeadura direta. **Bragantia**, Campinas, v. 70, n. 4, p. 908-916, 2011.

DIKER, K.; HEERMANN, D. F.; BRODAHL, M. K. Frequency analysis of yield for delineating yield response zones. **Precision Agriculture**, New York v. 5, n. 5, p. 435-444, 2004.

ERICKSON, B. **Field experience validates on-the-go soil pH sensor**. West Lafayette: Purdue Top Farmer Crop Workshop Newsletter, 2004. 4 p.

SILVA, F. A.; LIMA, S. S., J.; OLIVEIRA, B, R. Métodos de interpolação para estimar o pH em solo sob dois manejos de café arábica. **Idesia**, Arica, v. 28, n. 2, p. 61-66, 2010.

FARACO, M. A.; URIBE-OPAZO, M. A.; SILVA, E. A. A. D.; JOHANN, J. A.; BORSSOI, J. A. Seleção de modelos de variabilidade espacial para elaboração de mapas temáticos de atributos físicos do solo e produtividade da soja. **Revista Brasileira de Ciência do Solo**, Viçosa, v. 32, n. 2, p. 463-476, 2008.

FLINN, J. C.; DE DATTA, S. K. Trends in irrigated-rice yields under intensive cropping at Philippine research stations. **Field Crops Research**, Amsterdam, v. 9, p. 1-15, 1984.

FRAISSE, C. W.; SUDDUTH, K. A.; KITCHEN, N. R. Delineation of site-specific management zones by unsupervised classification of topographic attributes and soil electrical conductivity. **American Society of Agricultural and Biological Engineers. Transactions**, Saint Joseph, v. 44, n. 1, p. 155-166, 2001.

FRIDGEN, J. J.; KITCHEN, N. R.; SUDDUTH, K. A.; DRUMMOND, S. T.; WIEBOLD, W. J.; FRAISSE, C. W. Management Zone Analyst (MZA): Software for subfield management zone delineation. **Agronomy Journal**, Madison, v. 96, n. 1, p. 100-108, 2004.

GUEDES FILHO, O.; VIEIRA, S. R.; CHIBA, M. K.; NAGUMO, C. H.; DECHEN, S. C. F. Spatial and temporal variability of crop yield and some Rhodic Hapludox properties under no-tillage. **Revista Brasileira de Ciência do Solo**, Viçosa, v. 34, n. 1, p. 1-14, 2010.

HURLEY, T. M.; MALZER, G. L.; KILIAN, B. Estimating site-specific nitrogen crop response functions. **Agronomy Journal**, Madison, v. 96, n. 5, p. 1331-1343, 2004.

ISAAKS, E. H.; SRIVASTAVA, R. M. **An introduction to applied geostatistics**. Oxford University Press: New York, 1989. p. 561.

IRVIN, B. J.; VENTURA, S. J.; SLATER, B. K. Fuzzy and isodata classification of landform elements from digital terrain data in Pleasant Valley. **Geoderma**, Amsterdam, v. 77, n. 2-4, p. 137-154, 1997.

JIANG, Q.; FU, Q.; WANG, Z. Study on delineation of irrigation management zones based on Management Zone Analyst software. In: LI, D., LIU, Y. e CHEN, Y. (Ed.). **Computer and Computing Technologies in Agriculture IV**: 4th IFIP TC 12 Conference, CCTA 2010, Nanchang, China, October 22-25, 2010, Selected Papers, Part III. Boston: Springer, v.346, 2011. p.419-427.

JOHNSTON, K.; HOEF, J. M. V.; KRIVORUCHKO, K.; LUCAS, N. **Using ArcGIS Geostatistical Analyst**. ESRI: Redlands, 2001. p. 300.

KITCHEN, N.; SUDDUTH, K.; MYERS, D.; DRUMMOND, S.; HONG, S. Delineating productivity zones on claypan soil fields using apparent soil electrical conductivity. **Computers and Electronics in Agriculture**, Amsterdam, v. 46, n. 1, p. 285-308, 2005.

KRAVCHENKO, A.; BULLOCK, D. G. A comparative study of interpolation methods for mapping soil properties. **Agronomy Journal**, Madison, v. 91, n. 3, p. 393-400, 1999.

KRAVCHENKO, A. N.; BULLOCK, D. G. Correlation of corn and soybean grain yield with topography and soil properties. **Agronomy Journal**, Madison, v. 92, n. 1, p. 75-83, 2000.

LAMBERT, D. M.; LOWENBERG-DEBOER, J.; BONGIOVANNI, R. A comparison of four spatial regression models for yield monitor data: A case Study from Argentina. **Precision Agriculture**, New York, v. 5, n. 6, p. 579-600, 2004.

LARK, R. M. Regression analysis with spatially autocorrelated error: Simulation studies and application to mapping of soil organic matter. **International Journal of Geographical Information Science**, Abingdon, v. 14, n. 3, p. 247-264, 2000.

LI, Y.; SHI, Z.; LI, F.; LI, H.-Y. Delineation of site-specific management zones using fuzzy clustering analysis in a coastal saline land. **Computers and Electronics in Agriculture,** Amsterdam, v. 56, n. 2, p. 174-186, 2007.

LITTELL, R. C.; MILLIKEN, G. A.; STROUP, W.; WOLFINGER, R. D.; SCHABENBERBER, O. **SAS system for mixed models**. SAS Institute: Cary, 2006.

LOBELL, D. B.; ORTIZ-MONASTERIO, J. I.; ASNER, G. P.; NAYLOR, R. L.; FALCON, W. P. Combining field surveys, remote sensing, and regression trees to understand yield variations in an irrigated wheat landscape. **Agronomy Journal**, Madison, v. 97, n. 1, p. 241-249, 2005.

LUND, E. D.; COLLINGS, K. L.; DRUMMOND, P. E.; CHRISTY, C. D.; ADAMCHUK, V. I. Managing pH variability with on-the-go pH mapping. In: CONFERENCE ON PRECISION AGRICULTURE AND OTHER PRECISION RESOURCES

MANAGEMENT, 7., 2004, Minneapolis. **Anais**... Madison: ASA-CSSA-SSSA, 2004. p.120-132.

MALLARINO, A. P.; OYARZABAL, E. S.; HINZ, P. N. Interpreting within-field relationships between crop yields and soil and plant variables using factor analysis. **Precision Agriculture**, New York, v. 1, n. 1, p. 15-25, 1999.

MARTINS, M. V.; CARVALHO, M. D. P.; MARCELO, A.; MONTANARI, R. Correlação linear e espacial entre a produtividade do feijoeiro e atributos físicos de um Latossolo Vermelho distroférrico de Selvíria, Estado de Mato Grosso do Sul. **Acta Scientiarum. Agronomy**, Maringá, v. 31, n. 1, p. 147-154, 2009.

MCBRATNEY, A.; WHELAN, B.; ANCEV, T.; BOUMA, J. Future directions of precision agriculture. **Precision Agriculture**, New York, v. 6, n. 1, p. 7-23, 2005.

MEGDA, M. M.; CARVALHO, M. D. P. E.; VIEIRA, M. X.; ANDREOTTI, M.; PEREIRA, E. C. Correlação linear e espacial entre a produtividade de feijão e a porosidade de um Latossolo Vermelho de Selvíria (MS). **Revista Brasileira de Ciência do Solo, Viçosa**, v. 32, n. 2, p. 781-788, 2008.

MILANI, L.; SOUZA, E. G.; URIBE-OPAZO, M. A.; GABRIEL FILHO, A.; JOHANN, J. A.; PEREIRA, J. O. Unidades de manejo a partir de dados de produtividade. **Acta Scientiarum. Agronomy**, Maringá, v. 28, n. 4, p. 591-598, 2006.

MUELLER, T. G.; MIJATOVIC, B.; SEARS, B. G.; PUSULURI, N.; STOMBAUGH, T. S. Soil electrical conductivity map quality. **Soil Science**, Philadelphia, v. 169, n. 12, p. 841-851, 2004.

MUELLER, T. G.; PIERCE, F. J.; SCHABENBERGER, O.; WARNCKE, D. D. Map quality for site-specific fertility management. **Soil Science Society of America Journal**, Madison, v. 65, n. 5, p. 1547-1558, 2001.

MUNIZ, B. M.; CURI, N.; SPAROVEK, G.; CARVALHO FILHO, A.; SILVA, S. R. H. G. Updated Brazilian's georeferenced soil database: An improvement for international scientific information exchanging. In: GÜNGÖR, E. B. Ö. (Ed.). **Principles, Application and Assessment in Soil Science**. Rijeka: InTech, 2011. cap. 16, p.309-332.

NALDER, I. A.; WEIN, R. W. Spatial interpolation of climatic Normals: test of a new method in the Canadian boreal forest. **Agricultural and Forest Meteorology**, Amsterdam, v. 92, n. 4, p. 211-225, 1998.

NOGARA NETO, F.; ROLOFF, G.; DIECKOW, J.; MOTTA, A. C. V. Atributos de solo e cultura espacialmente distribuídos relacionados ao rendimento do milho. **Revista Brasileira de Ciência do Solo**, Viçosa, v. 35, n. 3, p. 1025-1036, 2011.

ODEH, I. O. A.; CHITTLEBOROUGH, D. J.; MCBRATNEY, A. B. Soil pattern recognition with fuzzy-c-means: Application to classification and soil-landform interrelationships. **Soil Science Society of America Journal,** Madison**,** v. 56, n. 2, p. 505-516, 1992.

OLFS, H. W.; BORCHERT, A.; TRAUTZ, D. Validation of on-the-go soil pH measurements: Primary results from Germany. In: INTERNATIONAL CONFERENCE ON PRECISION AGRICULTURE, 10., 2010, Denver. **Anais**... Denver: International Society of Precision Agriculture, 2010. 1-CD-ROM.

OSBORNE, J.; WATERS, E. Four assumptions of multiple regression that researchers should always test. **Practical Assessment, Research and Evaluation**, College Park, v. 8, n. 2, 2002.

REICHERT, J. M.; DARIVA, T. A.; REINERT, D. J.; SILVA, V. R. D. Variabilidade espacial de Planossolo e produtividade de soja em várzea sistematizada: análise geoestatística e análise de regressão. **Ciência Rural**, Santa Maria, v. 38, n. 4, p. 981-988, 2008.

RODRIGUES, M. S.; CORÁ, J. E.; FERNANDES, C. Spatial relationships between soil attributes and corn yield in no-tillage system. **Revista Brasileira de Ciência do Solo**, Viçosa, v. 36, n. 2, p. 599-609, 2012.

ROSA FILHO, G.; CARVALHO, M. D. P. E.; ANDREOTTI, M.; MONTANARI, R.; BINOTTI, F. F. D. S.; GIOIA, M. T. Variabilidade da produtividade da soja em função de atributos físicos de um latossolo vermelho distroférrico sob plantio direto. **Revista Brasileira de Ciência do Solo**, Viçosa, v. 33, n. 2, p. 283-293, 2009.

SANTOS, M. L.; CARVALHO, M. D. P.; RAPASSI, R. M. A.; MURAISHI, C. T.; MALLER, A.; MATOS, F. A. Correlação linear e espacial entre produtividade de milho (Zea mays L.) e atributos físicos de um Latossolo Vermelho distroférrico sob plantio direto do Cerrado Brasileiro. **Acta Scientiarum. Agronomy**, Maringá, v. 28, n. 3, p. 313-321, 2006.

SANTOS, A. O.; PALLONE FILHO, W. J.; UNGARO, M. R. G.; LINO, A. C. L.; RADIN, B. Prospecção de zonas potenciais para manejo diferenciado em agricultura de precisão utilizando-se padrões de solo-planta-clima. **Revista Brasileira de Engenharia Agrícola e Ambiental**, Campina Grande, v. 7, n. 3, p. 463-468, 2003.

SCHABENBERGER, O.; PIERCE, F. J. **Contempoary statistical models for the plant and soil sciences**. CRC Press: Boca Raton, 2001. p. 556-697.

SCHEPERS, A. R.; SHANAHAN, J. F.; LIEBIG, M. A.; SCHEPERS, J. S.; JOHNSON, S. H.; LUCHIARI, A. Appropriateness of management zones for characterizing spatial variability of soil properties and irrigated corn yields across years. **Agronomy Journal**, Madison, v. 96, n. 1, p. 195-203, 2004.

SCHLOEDER, C. A.; ZIMMERMAN, N. E.; JACOBS, M. J. Comparison of methods for interpolating soil properties using limited data. **Soil Science Society of America Journal**, Madison, v. 65, n. 2, p. 470-479, 2001.

SOUZA, G. S.; LIMA, J. S. D. S.; XAVIER, A. C.; ROCHA, W. S. D. Krigagem ordinária e inverso do quadrado da distância aplicados na espacialização de atributos químicos de um argissolo. **Scientia Agraria**, Piracicaba, v. 11, n. 1, p. 73-81, 2010.

TAGARAKIS, A.; LIAKOS, V.; FOUNTAS, S.; KOUNDOURAS, S.; GEMTOS, T. Management zones delineation using fuzzy clustering techniques in grapevines. **Precision Agriculture**, New York, v. 14, n. 1, p. 1-22, 2013.

TIMLIN, D. J.; PACHEPSKY, Y.; SNYDER, V. A.; BRYANT, R. B. Spatial and temporal variability of corn grain yield on a hillslope. **Soil Science Society of America Journal**, Madison, v. 62, n. 3, p. 764-773, 1998.

VERIS-TECHNOLOLOGIES. **Soil pH manager (MSP)**. 2010. Disponível em: < http://www.veristech.com/products/soilph.aspx >. Acesso em: 15 de set. de 2011.

VIEIRA, S. R. Geoestatística em estudos de variabilidade espacial do solo. In: NOVAIS, R. F.;ALVAREZ, V., et al (Ed.). **Tópicos em ciência do solo**. Sociedade Brasileira de Ciência do Solo:Viçosa, v.1, 2000. p.1-54.

WEBER, D.; ENGLUND, E. Evaluation and comparison of spatial interpolators. **Mathematical Geology**, Heidelberg, v. 24, n. 4, p. 381-391, 1992.

WEBSTER, R.; OLIVER, M. A. **Geostatistics for Environmental Scientists**. John Wiley: Chichester, 2007. p. 271.

ZANDI, S.; GHOBAKHLOU, A.; SALLIS, P. Evaluation of spatial interpolation techniques for mapping soil pH. In: INTERNATIONAL CONGRESS ON MODELLING AND SIMULATION, 19., 2011, Perth. **Anais**... Perth: Modelling and Simulation Society of Australia and New Zealand. 2011. p. 1153-1159.

CAPÍTULO 2 - A spatial and temporal prediction model of corn grain yield as a function of soil attributes

ABSTRACT – Effective site-specific management requires an understanding of soil and environmental factors influencing crop yield variability. Moreover, it is necessary assess the techniques used to define these relationships. The objective of this study was to assess whether statistical models that accounted for heteroscedastic and spatial-temporal autocorrelation were superior to ordinary least squares models (OLS) when evaluating the relationship between corn yield and soil attributes in Brazil. The study site (10 by 250-m in size) was located in São Paulo State, Brazil. Corn yield (planted with 0.9-m spacing) was measured in one hundred 4.5x10-m cells along four parallel transects (25 observations per transect) over six growing seasons between 2001 and 2010. Soil chemical and physical attributes were measured. OLS, generalized least squares assuming heteroscedasticity (GLS_{he}), spatial-temporal least squares assuming homoscedasticity (GLS_{sp}), and spatial-temporal model assuming heteroscedasticity ($GLS_{he\text{-}sp}$) analyses were used to estimate corn yield. Soil acidity (pH) was the factor that most influenced corn yield over time in this study. The OLS model suggested that there would be a 0.59 Mg ha^{-1} yield increase for each unit increase in pH, whereas $GLS_{he\text{-}sp}$ increased 0.43 Mg ha^{-1} which means that model choice impacted prediction and regression parameters and this is critical because accurate estimation of yield is necessary for correct management decisions. The spatial and temporal autocorrelation assuming heteroscedasticity was superior to the OLS model for prediction. Farms should use a long enough historical data set of crop yield and soil attributes to understand case-effect relationship between crop yield and soil attributes and predict crop yield.

Keywords: mixed models, soil pH, tropical soils, *Zea mays* L.

2.1. Introduction

Yield mapping is a simple, inexpensive tool for monitoring crop yield at fine spatial resolutions. However, yield maps have little or no value unless they can be used for decisions that will improve crop and soil management (Pierce and Nowak, 1999). Consequently, effective and meaningful spatial analysis of yield variability and yield limiting factors has become a critical issue throughout the world. Moreover, yield maps can vary substantially from year to year, so that analysis of data from only one year might potentially lead to incorrect decisions. Thus, it is critical to assess temporal variability and stability of crop yield (McBratney et al., 2005).

Several factors can affect crop yield variability, such as soil fertility, terrain properties, weeds and diseases. Numerous studies have found that crop yields are frequently spatially correlated to soil fertility attributes, but it is critical that these kinds of studies use similar sampling support (Rodrigues et al., 2012). According to Diker et al. (2004) the dominance of factors that influence crop yield variability can change from year to year because of seasonal weather variation. For example, Timlin et al. (1998) found that corn yields were correlated with soil P, K^+, and OM content only in dry years in a Typic Fragiochrept soil. Bakhsh et al. (2000) pointed out that it may be necessary to also include climate and management factors in yield prediction models for detailed diagnosis of yield limiting factors.

Regression is one of the most common ways to describe the relationship between crop yield and soil attributes. However, it has a number of important assumptions that should be taken into account (e.g., normal distribution and independence of errors, linearity, lack of collinearity, reliability of measurement and homoscedasticity) (Osborne and Waters, 2002). These assumptions have often been ignored in the literature.

Ordinary least squares (OLS) regression, the most common statistical procedure used for yield estimation, assumes normality and independence of errors and homoscedasticity. However, yield residuals generally are often spatially autocorrelated (Lark, 2000; Lobell et al., 2005). Type I errors tend to increase with OLS when spatial dependence (Schabenberger and Gotway, 2005) and heteroscedasticity are ignored (Osborne and Waters, 2002). This could lead to critical misinterpretation errors of yield variation and consequently lead to improper management decisions.

Spatial dependence structure may change from year to year; therefore it is critical to investigate crop yield and soil property interactions over time. In these situations, Generalized Least Squares (GLS) regression with correlated errors should be used (Flinn and De Datta, 1984), because it allows spatial and temporal error correlation components to be assessed and then filtered from the total residual term of the model to improve the power of the statistical tests.

As an alternative to traditional regression analyses, spatial-temporal mixed effect models could be used which assume normally distributed, and spatially-temporally correlated errors, and include both fixed and random effects (Bolker et al., 2009). The study of the spatial–temporal structure of the errors is very important for monitoring and evaluating crop yield. Generally space and time effects are considered separately, however, interactions are common across spatial and temporal scales and modeling efforts should account for both effects simultaneously (Landagan and Barrios, 2007).

Some studies (Hurley et al., 2004; Lambert et al., 2004) have pointed out that mixed effect models, which incorporate spatial variability, will help improve the understanding of the factors that affect crop yield. For example, Hurley et al. (2004) found that the best model to describe the N response in corn yield was the one that took into account heteroscedasticity and spatial correlation. However, there are not many studies that treat spatial-temporal variability and heteroscedasticity of crop yield in the same model.

Therefore, the objective of this study was to assess whether statistical models that accounted for heteroscedasticity and spatial-temporal autocorrelation were superior to ordinary least squares models when evaluating the relationship between corn yield and soil attributes in a field in Brazil.

2.2. Material and Methods

2.2.1. Site description

This experiment was conducted in the city of Jaboticabal, in São Paulo state, Brazil (21º14'05'' S, 48º17'09'' W, and altitude of 613 m a.s.l.). Climatologically the area belongs to tropical/megathermal zone or Köppen Aw (tropical climate with dry winter and average temperature of the coldest month greater than 18º C). The mean annual rainfall (1971–2006) was 1417 mm, with the distribution peaking in the period of October–March and a relatively dry season in the period of April–September. The soil of the experimental area was classified as clayey Rhodic Hapludox (Latossolo Vermelho distroférrico). The experimental area was managed as a corn-fallow rotation under no-tillage system over 12 years.

2.2.2. Yield and soil sampling, and climatic data

The size of the experimental area was 18 by 250-m with the longer dimension oriented in the East-West direction. Each of the 100 experimental plots had a dimension of 10 by 4.5 m and were arranged in a 25 by 4 grid. The experimental scheme is depicted in Figure 1.

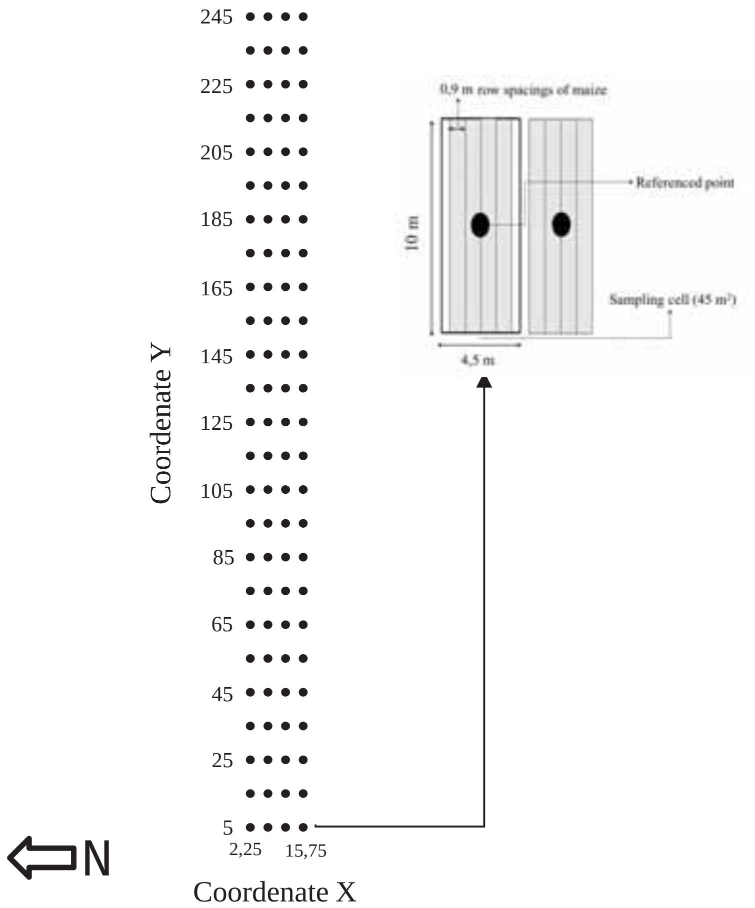

Figure 1. Sampling scheme of soil attributes and corn yield in a Rhodic Hapludox (Latossolo Vermelho distroférrico) under no-tillage system.

Before planting, non-selective herbicides were applied. Corn (*Zea mays* - triple-hybrid Syngenta Master) was planted at 65,000 plants ha^{-1} with 0.9-m row spacing in early December between 2001 and 2010 growing seasons, but the data were collected according to sampling design depicted in Figure 1 only in the 2001/2002, 2002/2003, 2003/2004, 2007/2008, 2008/2009 and 2009/2010 growing seasons. The starter fertilization consisted of 30 kg of N, 70 kg of P_2O_5 and 50 kg of K_2O ha^{-1}. Nitrogen fertilizer (urea) was applied at 100 kg N ha^{-1} when plants had four to six pairs of leaves totally developed. The field was uniformly treated. Corn was harvested about 150 days after planting with a 1-row plot combine that deposited the grain into a burlap bag. Grain weights were obtained for each plot with a manual balance in the field. The grain for each plot was sub-sampled for moisture and grain yields were determined at 13% gravimetric moisture.

Each year, five soil sub-samples were collected within each plot using a Dutch auger (0.1 m depth) and were composited. One of the soil sub-samples was collected in the middle of the plot and the other four samples were collected 2-meters apart from the middle in all four cardinal directions from the centroid. The area associated with soil measurements (support) was assumed similar to the one of crop yield (45 m^2). Each soil composite sample was analyzed for particle size (pipette method) (Gee and Dani, 2002), pH (1:1 soil/water mixture), organic matter (OM) (Walkley-Black method), P (ion-exchange resin**)**, K^+, Ca^{2+} and Mg^{2+} (1M NH_4O Ac. extractable at buffered at pH 7) according to Page et al. (1982). From the analytical determinations, cation exchange capacity (CEC = $K^+ + Ca^{2+} + Mg^{2+} + H + Al^{+3}$) and percentage of soil base saturation (BS = ($K^+ + Ca^{2+} + Mg^{2+}$ / CEC) x 100) were calculated. These attributes were chosen to be measure and analyzed because they are commonly measured by Brazilian farmers to characterize soil fertility.

Monthly cumulative rainfall, growing degree days (base temperature of 10°C), average temperatures, relative humidity and number of days with rainfall were recorded by the climatological station of the São Paulo State University (21º14’05’’ S, 48º17’09’’ W and altitude of 615 m a.s.l.) from December 2001 to April 2010 (Figure 2a and b). The weather station was located 30 m from the experimental site.

a) b)

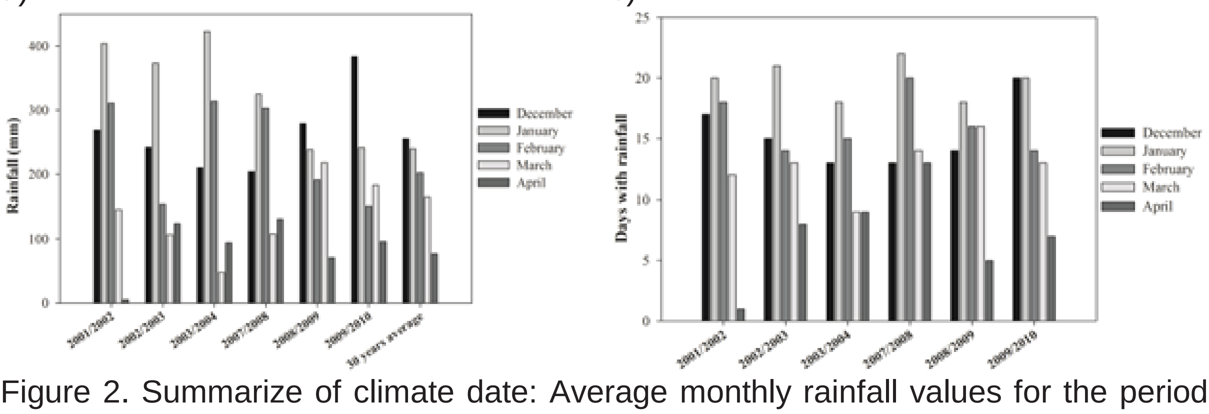

Figure 2. Summarize of climate date: Average monthly rainfall values for the period December–April of the studied years and 30 years average (a); Days with rainfall for the period December-April of the studied years (b).

2.2.3. Preliminary statistics analyses

Yield and soil attributes were tested for normality (Shapiro and Wilk, 1965) and yield was tested for heteroscedasticity with Bartlett's test (Snedecor and Cochran, 1989). Yield was also initially tested for spatial autocorrelation with Moran's I test (Moran, 1950). Moran's I test is a weighted correlation coefficient used to detect departures from spatial randomness, that is, it determines whether neighboring areas were more similar than it would be expected under the null hypothesis.

Exploratory stepwise regression analysis was performed to select the best subset of regressors out of all study variables (clay, sand, silt, pH, OM, K^+, Ca^{2+}, Mg^{2+}, CEC and BS). The final variables selected for the mixed effects model were chosen because they performed well with initial testing, the variables did not show collinearity and were biophysically meaningful. Stepwise selection is similar to the forward method, except that variable already in the model do not necessarily stay there. In a forward selection each of the available predictors is added if it meets the statistical criterion of entry, which is the significance level ($p<0.15$) for the increase in the R^2 produced by addition of the predictor. At each step variables that are already in the model are first evaluated for removal and if any are eligible for removal ($p>0.15$), the one whose removal would cause the least decrease of R^2 is removed. This procedure is repeated until there remain no more predictors that are eligible for entry or removal.

2.2.4. Linear mixed effect model

Four different regression models were compared to assess the impact of soil attributes and growing seasons (temporal variability) on corn yield: ordinary least squares (OLS), generalized least squares assuming heteroscedasticity (GLS_{he}), spatial-temporal model assuming homoscedasticity (GLS_{sp}), and spatial-temporal model assuming heteroscedasticity ($GLS_{he\text{-}sp}$) approaches (Schabenberger and Gotway, 2005) were used.

The standard linear model (OLS) can be written as:

$$y_i = \sum_{j=1}^{\rho} x_{ij}\,\beta_j + e_i \qquad (1)$$

i=1,…N

where y_i are N data points of the response variable (i.e., corn yield), x_{ij} are the observations of p explanatory variables (j), which can be continuous variables (i.e., soil variables) or dummy variables declaring class membership of a categorical variable (i.e., growing season), $\beta_{i\ldots}\beta_p$ are fixed effect coefficients to be estimated and e_i unknown independent and identically distributed normal random variables with mean 0 and variance σ^2.

For convenience and simplicity the previous formula can be written using matrix notation:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}, \qquad (2)$$

where **Y** denotes the vector of the responses, **X** is the matrix of the observations, **β** the vector of the unknown fixed effect coefficients and **e** is the vector of independent and identically distributed normal random errors or in symbol: $\mathbf{e} \sim N(0, \sigma^2\mathbf{I})$ where **I** is identity matrix. If the error variance σ^2 is not constant but varies as a function of a class variable (growing season), the model is called generalized least-squares model (GLS_{he}).

2.2.5. Spatial mixed effect model

Many times the independence distributional assumption about **Y** is too restrictive and the linear mixed model extends the general linear model by allowing elements of **Y** to be correlated. This was performed in two ways: through a specification of the covariance of **e** as a function of the distance between two locations, say $\mathbf{e} \sim N(0, \mathbf{R})$, for spatial variability; and the addition of a random effect and random coefficient in the analysis, giving rise to a **Zu** term in the model, where **u** is normal with mean 0 and variance **G**, for temporal variability. **Z** is a matrix, similar to **X**, that captures the complex covariance structure of the temporal factor.

The spatial-temporal linear mixed effect model (GLS_{sp}) can then be written as:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e} \tag{3}$$

where **e** ~ N(0, **R**), **u** ~ N(0, **G**), Cov [**u**, **e**] = 0**,** which implies the assumption that **u** and **e** are uncorrelated. Differently than **β**, the vector **u** does not contain parameters but random variables. The temporal relationship (growing season) was explored by postulating an autoregressive structure of order 1 for the matrix **G**, which has homogeneous variances and correlations that decline exponentially with the time series. The AR(1) covariance structure has two unknown parameters: the variance (σ_t^2) and the lag-one correlation (ρ_t). However, temporal factor was included also in the model as fixed effect (growing season) and dummy variable in the matrix **X**, in order to assess systematic or trend component in corn yield variation and then evaluate its stability over time.

The spatial relationship was modeled by using three different isotropic covariance functions of the distance such as spherical, exponential, and Gaussian (Littell et al., 2006) and adding an additional parameter (nugget effect) to adequately account for abrupt changes over relatively small distances. The resulting covariance model has the form:

$$\mathrm{Var}[e_i] = \sigma^2 + \sigma_1^2 \tag{4}$$

$$\mathrm{Cov}[e_i, e_j] = \sigma^2 [f(d_{ij})], \tag{5}$$

where: $f(d_{ij})$ is one of geostatistical spatial covariance functions of distance d_{ij} between two observations i and j, using a parameter ρ for spatial scale (range), and σ^2_l, $\sigma^2+\sigma^2_l$ corresponding to the geostatistical parameters nugget and sill, respectively. The fitting process relies on an iterative procedure aimed at maximizing the log likelihood of data by restricted maximum likelihood method (REML) (Littell et al., 2006). The fixed effects estimates are obtained as generalized least squares estimates evaluated at the REML estimate of the covariance parameters. A further complexity in the spatial-temporal model was added by allowing crop yield variance to vary over the growing seasons, because it can cause heterogeneity in the covariance structure. The whole analysis was then repeated using a different set of covariance function parameters for each growing season ($GLS_{he\text{-}sp}$).

2.2.6. Models comparison

To compare the different competing spatial covariance models, their modeling criteria were compared: the best model was selected as the one whose -2 Log Likelihood, Akaike's information (AIC), Akaike's information corrected (AICC), and Schwarz's Bayesian Information (BIC) criteria were the smallest (Littell et al., 2006).

Each model with no spatial correlation (OLS and GLS_{he}), i.e. with $\mathbf{R}=\sigma^2\mathbf{I}$, was compared to the corresponding homoscedastic (GLS_{sp}) and heteroscedastic ($GLS_{he\text{-}sp}$) spatial model with nugget effect respectively, whose $\mathbf{R}=\sigma^2\mathbf{F}+\sigma^2\mathbf{I}$ where $\mathbf{F}$ is an NxN matrix whose ijth element is $f(d_{ij})$. Since $\mathbf{F}$ reduces to $\mathbf{I}$ if the spatial parameter $\rho=0$ and $\sigma^2_l=0$, to compare the -2 Log Likelihood for spatial and non spatial models, a likelihood ratio test for null hypothesis: $\rho=0$ and $\sigma^2_l=0$ with 2 degrees of freedom was performed. Under the null hypothesis that the spatial model is not different from the non-spatial model, the likelihood ratio statistics is distributed as χ^2 with the number of freedom degrees equal to the difference in the number of parameters between the two models. Therefore, because the fixed part is the same in the spatial and non-spatial models, only the parameters in the variance-covariance structure needs to be considered.

For comparison bias, accuracy and precision of the four regression models were assessed by means of leave-one-out cross-validation using the three cross-validation statistics, suggested by Carroll and Cressie (1996), as follows. Let $Y_{[i]}$ be the observed corn yield value removed at the i^{th} iteration, $\hat{Y}_{[i]}$ be its corresponding prediction obtained by fitting the model to the remaining N - 1 points, $e_{[i]} = Y_{[i]} - \hat{Y}_{[i]}$ be the difference between the observed and estimated values and $\sigma_{[i]}$ be the mean squared prediction error of $\hat{Y}_{[i]}$, then the three cross-validation statistics are, respectively:

$$CV_1=\frac{1}{n}\sum_{i=1}^{n}\frac{e_{[i]}}{\sigma_{[i]}},\ CV_2=\left(\frac{1}{n}\sum_{i=1}^{n}\frac{e^2_{[i]}}{\sigma^2_{[i]}}\right)^{\frac{1}{2}}\text{and, } CV_3=\left(\frac{1}{n}\sum_{i=1}^{n}e^2_{[i]}\right)^{\frac{1}{2}} \quad (6)$$

CV_1 was used to assess the unbiasedness of the predictor and the optimal value should be approximately zero; CV_2 was used to assess the accuracy of mean squared prediction error and should be approximately 1; CV_3 was used to check the goodness of prediction, and models with smaller values of CV_3 should be preferred because this means that fitted values are close to observed values (Carroll and Cressie, 1996).

All statistics analyses in this study were computed with SAS (SAS Institute, Inc.; Cary, North Carolina, release 9.3) and linear spatial mixed effect model was estimated with Proc MIXED procedure.

2.3. Results and Discussion

2.3.1. Descriptive statistics of soil attributes

All soil fertility data was classified as low, medium, or high according to criteria determined by Raij et al. (1997) for the state of Sao Paulo. Percent soil base saturation values (Table 1) were medium (i.e., 51-70%) in the 2001/2002 and 2002/2003 growing seasons, and low (i.e., 26-50%) in the 2003/2004, 2007/2008, 2008/2009, and 2009/2010. The levels of soil K^+ (1.6-3.0 $mmol_c$ dm^{-3}) as well as the soil P levels were medium (16-40 mg dm^{-3}) in all the study years. The levels of soil Ca^{2+} were high (greater than 7 $mmol_c$ dm^{-3}) in all the study years, whereas soil Mg^{2+} were high (greater than 8 $mmol_c$ dm^{-3}) in the 2001/2002, 2002/2003, and 2003/2004, and medium (5-8 $mmol_c$ dm^{-3}) in the 2007/2008, 2008/2009, and 2009/2010.

Medium values of pH (5.1-5.5) were observed in the 2001/2002 and 2002/2003 and low (4.4-5.0) in the 2003/2004, 2007/2008, 2008/2009, and 2009/2010. Medium soil organic matter contents (OM) were observed in all the years. Most variables showed large variability on the basis of their range of variation (Table 1). In general, the soil fertility variables were not normally distributed and coefficients of asymmetry and kurtosis varied over years. However, this did not prevent regression analyses, since normality is a requirement only for the response variable. The clay ($\bar{x}$ = 332 g kg^{-1}, standard deviation = 19 g kg^{-1}) and sand ($\bar{x}$ = 623 g kg^{-1}, standard deviation = = 17) content of the surface samples did not vary substantially throughout the experimental area and were considered normally distributed according to the Shapiro-Wilk test. The soil was classified as a sandy clay loams according to the USDA (United States Department of Agriculture) texture method.

Table 1. Descriptive statistics of soil chemical properties from 0 to 0.1 m depth in a Rhodic Hapludox under no-tillage system over six growing seasons.

Attributes	P ($mg\ dm^{-3}$)	OM ($g\ dm^{-3}$)	pH ($CaCl_2$)	K^+	Ca^{2+}	Mg^{2+}	CEC	BS (%)
					($mmol_c\ dm^{-3}$)			
			2001/2002 growing season					
Mean	30	20	5.16	2.77	21	10	62	53
Std Dev	16	2.13	0.53	0.94	9	5	8	16
Minimum	10	15	4.2	1.2	7	3	51	20
Maximum	75	24	6.6	6.6	50	30	91	84
Skewness	1.06	0.09	0.29	1.11	0.68	1.25	1.20	-0.26
Kurtose	0.42	-0.52	-0.13	2.33	0.64	2.88	2.04	-0.78
Pr < W	0.00	0.04	0.05	0.00	0.00	0.00	0.00	0.01
			2002/2003 growing season					
Mean	25	19	5.08	1.86	22	10	64	53
Std Dev	20	2.53	0.46	0.63	9	5	9	16
Minimum	1	11	3.9	0.8	5	2	38	12
Maximum	103	26	6.0	3.9	46	40	109.4	84
Skewness	1.82	-0.60	-0.43	1.01	0.21	1.97	1.43	-0.54
Kurtose	3.47	1.88	-0.47	1.65	0.07	9.63	7.43	-0.24
Pr < W	0.00	0.00	0.03	0.00	0.17	0.00	0.00	0.01
			2003/2004 growing season					
Mean	23	18	4.94	1.82	21	11	64	50
Variance	17	1.67	0.57	0.42	12	6	12	20
Minimum	7	15	3.9	1.1	4	2	45.2	11
Maximum	86	23	6.1	2.7	58	30	108.7	88
Skewness	2.39	0.06	-0.03	0.30	0.62	1.01	1.40	-0.32
Kurtose	5.56	-0.37	-0.97	-0.97	0.18	1.21	2.59	-0.96
Pr < W	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.00
			2007/2008 growing season					
Mean	33	16	4.67	1.93	14	7	57	39
Std Dev	16	2.20	0.37	0.55	7	4	12	14
Minimum	8	11	3.9	0.9	3	2	28	16
Maximum	67	22	5.6	3.6	36	28	93.3	78
Skewness	0.43	0.03	0.17	0.81	0.92	2.11	0.10	0.58
Kurtose	-0.87	0.30	-0.20	1.26	0.80	8.37	0.22	0.09
Pr < W	0.00	0.08	0.45	0.00	0.00	0.00	0.93	0.02
			2008/2009 growing season					
Mean	38	17	4.41	2.39	13	6	65	33
Std Dev	22	3.08	0.36	0.69	6	3	5	13
Minimum	8	10	3.8	1.4	2	1	46.4	8
Maximum	114	26	5.3	4.7	29	16	75.7	63
Skewness	1.28	-0.20	0.08	0.77	0.48	1.21	-0.10	0.18
Kurtose	1.63	0.17	-0.27	0.20	-0.25	3.40	0.77	-0.41
Pr < W	0.00	0.18	0.00	0.00	0.05	0.02	0.34	0.28
			2009/2010 growing season					
Mean	30	17	4.35	2.24	10	6	63	30
Std Dev	17	2.62	0.35	0.51	5	3	7	12
Minimum	11	2	3.8	1.3	3	2	50.8	10
Maximum	112	27	5.3	4.4	28	16	80.1	60
Skewness	1.96	-1.31	0.36	0.72	0.62	1.19	0.43	0.26
Kurtose	6.15	12.99	-0.22	2.48	0.66	2.16	-0.28	-0.43
Pr < W	0.00	0.00	0.01	0.00	0.00	0.00	0.06	0.05

CEC = Cation exchange capacity; BS = Soil base saturation; Pr < W = result of Shapiro-Wilk normality test (P<W > 0.01).

2.3.2. Preliminary statistics analyses for corn yield

Corn yields ranged from 6.0 to 7.6 Mg ha^{-1} (Table 2), and showed normal data distribution in all the years, excepted in the 2003/2004 and 2008/2009 growing seasons, as indicated by the Shapiro-Wilk test (Table 2). However, their coefficients of skewness and kurtosis were close to zero, indicating no substantial departure from normality and the yield was then assumed to be normal.

Table 2. Descriptive statistics of corn yield of six growing season in a Rhodic Hapludox under no-tillage system.

Corn yield growing season	Mean (Mg ha^{-1})	Standard Deviation	Minimum	Maximum	Skewness	Kurtosis	Pr < W
2001/2002	7.6	0.50	6.3	8.6	-0.35	0.08	0.20
2002/2003	7.0	0.54	5.7	8.3	-0.20	-0.20	0.51
2003/2004	6.0	0.49	5.3	7.3	0.96	0.64	0.00
2007/2008	7.3	0.83	5.4	9.0	-0.14	-0.73	0.11
2008/2009	7.8	0.78	5.6	8.8	-0.91	0.25	0.00
2009/2010	6.9	0.75	5.4	8.6	-0.10	-0.49	0.53

Pr < W = result of Shapiro-Wilk normality test ($P<W > 0.01$).

Bartlett's test was significant at probability level of $p < 0.001$, indicating heteroscedasticity of corn yield over time, which was treated in this study by fitting different spatial models of covariance function differing in sill and range parameters. Moran's I test was also significant ($p < 0.001$) indicating spatial association. The best subset of soil variables selected by stepwise regression included pH, K^+, P and clay content (data not shown), which were used as soil regressors into the mixed effect.

2.3.3. Linear and spatial mixed effect models analyses

The variables K^+ and P were not significant ($p > 0.05$) for all the mixed effect models tested, whereas clay content was not significant in the models that took heteroscedasticity into account (i.e., GLS_{he} and $GLS_{he\text{-}sp}$) (Table 3). This last result indicates that ignoring heteroscedasticity can distort results and increase the possibility of falsely declaring significant effects (i.e., type I errors) (Osborne and Waters, 2002). Soil pH was significant for all the models, indicating that in all the study growing seasons, the spatial variability of corn yield was highly correlated with soil acidity (Table 3).

The covariance function that best described the spatial dependence was the spherical model. The two parameters (σ_t^2 and ρ_t) of AR(1) model were not significant at probability level of $p<0.05$ (data not shown) in the GLS_{sp} and $GLS_{he\text{-}sp}$ model. This indicated that the stochastic component of temporal variability of corn yield was not significantly different than 0, thus temporal observations could be considerated to be uncorrelated. However, the temporal fixed effect (i.e., growing season) was highly significant for all the models (Table 3). The temporal effect was considered to be a deterministic process and yield was not stable over the six-season study period as proved by the changing coefficients (GS1-6), which represent the specific contribution of each season to the overall yield average.

The goodness of fit statistics (i.e., -2 log likelihood, AIC, AICC and BIC) differed among the models and was ordered as follows: OLS > GLS_{he} > GLS_{sp} > $GLS_{he\text{-}sp}$ (smaller is better) (Table 4). This order was established on the basis of the values of at least three out of four criteria resulting consistent. If BIC criterion had been the only one used to compare the models, OLS would have been considered better than GLS_{he}, whereas a different conclusion was stated on the basis of the other criteria. This results confirmed warnings that multiple criteria should be considered when comparing models (Littell et al., 2006).

Table 3. Tests of fixed effects of ordinary least squares (OLS), generalized least squares assuming heteroscedasticity (GLS_{het}), spatial model assuming homoscedasticity (GLS_{sp}), and spatial model assuming heteroscedasticity ($GLS_{he\text{-}sp}$) approaches used to assess the impact of soil properties and growing seasons on corn yield.

Statistics	pH	Clay	P	K^+	†GS 1	†GS 2	†GS 3	†GS 4	†GS 5	†GS 6
					OLS					
Estimate	0.5901	0.00355	0.0008	0.03107	3.2899	2.7033	1.7551	3.2758	3.833	3.0939
Std. err.	0.07302	0.00165	0.00154	0.04459	0.5106	0.5128	0.5143	0.5058	0.5049	0.5022
DF	473	473	473	473	473	473	473	473	473	473
Pr>|t|	0.00	0.03	0.60	0.48	0.00	0.00	0.00	0.00	0.00	0.00
Lower	0.4466	0.0003	-0.00222	-0.05655	2.2867	1.6957	0.7445	2.2819	2.841	2.107
Upper	0.7335	0.0068	0.003822	0.1187	4.2932	3.7109	2.7656	4.2696	4.8251	4.0808
					GLS_{het}					
Estimate	0.5224	0.0007	0.000049	0.03629	4.578	3.9884	3.0377	4.5409	5.0899	4.3349
Std. err.	0.06485	0.00155	0.00146	0.04033	0.4683	0.4704	0.4721	0.4696	0.4693	0.4663
DF	473	473	473	473	473	473	473	473	473	473
Pr>|t|	0.00	0.63	0.97	0.36	0.00	0.00	0.00	0.00	0.00	0.00
Lower	0.395	-0.00231	-0.00282	-0.04296	3.6578	3.0642	2.11	3.6182	4.1677	3.4185
Upper	0.6498	0.003785	0.002918	0.1155	5.4982	4.9127	3.9654	5.4636	6.0121	5.2512
					GLS_{sp}					
Estimate	0.2971	0.004709	0.000943	-0.02375	4.6071	3.942	2.9365	4.3961	4.8854	4.123
Std. err.	0.0852	0.002391	0.00151	0.04441	0.8328	0.8303	0.8288	0.8207	0.8164	0.814
DF	375	91.7	468	470	84.2	84.5	84.8	82.7	82.1	81.9
Pr>|t|	0.00	0.05	0.53	0.59	0.00	0.00	0.00	0.00	0.00	0.00
Lower	0.1295	-0.00004	-0.00202	-0.111	2.9511	2.291	1.2885	2.7636	3.2614	2.5037
Upper	0.4646	0.009458	0.00391	0.06351	6.2631	5.593	4.5845	6.0286	6.5094	5.7423
					$GLS_{he\text{-}sp}$					
Estimate	0.4343	0.00177	-0.0006	0.009109	4.7604	4.1788	3.3036	4.7121	5.1693	4.4624
Std. err.	0.06993	0.001882	0.001375	0.04016	0.6423	0.6364	0.6522	0.6652	0.6354	0.6761
DF	279	118	440	304	96.4	105	69.4	103	114	59.7
Pr>|t|	0.00	0.34	0.66	0.82	0.00	0.00	0.00	0.00	0.00	0.00
Lower	0.2967	-0.00196	-0.0033	-0.06993	3.4856	2.917	2.0026	3.3928	3.9106	3.1098
Upper	0.572	0.005497	0.002102	0.08814	6.0353	5.4406	4.6045	6.0315	6.4279	5.8149

† GS = growing season (1, 2, 3, 4, 5 and 6 = 2001/2002, 2002/2003, 2003/2004, 2007/2008, 2008/2009 and 2009/2010, respectively); Std. err. = standard error; DF = Degrees of freedom; Lower and Upper = 95% confidence limits.

Table 4. Model fit statistics of ordinary least squares (OLS), generalized least squares assuming heteroscedasticity (GLS_{he}), spatial model assuming homoscedasticity (GLS_{sp}), and spatial model assuming heteroscedasticity ($GLS_{he\text{-}sp}$) approaches used to assess the impact of soil attributes and growing seasons on corn yield.

Models	-2 Res log.	AIC	AICC	BIC
OLS	899.3	901.3	901.3	905.5
GLS_{he}	869.6	881.6	881.7	906.6
GLS_{sp}	869.2	875.2	875.2	887.7
$GLS_{he\text{-}sp}$	781.3	811.3	812.4	808.2

-2 Res log. = likelihood ratio test; AIC = Akaike’s information criterion; AIC = Akaike’s information criterion corrected; BIC = Schwarz’s Bayesian Information criterion (smaller is better).

The likelihood ratio test of spatial covariance was significant ($p<0.001$) when the non-spatial homoscedastic (OLS) and heteroscedastic (GLS_{he}) models were compared with the correspondent spatial models (GLS_{sp} and $GLS_{he\text{-}sp}$), respectively, stating that the structured component of spatial dependence was significant at the study site (Table 5). The differences in the goodness of fit indices between GLS_{he} and GLS_{sp} model were quite small (Table 4) and the likelihood ratio test of spatial covariance was not significant (data not shown), which indicated that accounting for heteroscedasticity of variance was critical in this study which is in agreement with the results observed by Lambert et al. (2004) and Hurley et al. (2004). It is important to highlight that heteroscedasticity in this study was treated by fitting different spatial models of covariance differing in sill and range parameters instead of homogenizing the variance as it is usually done with this kind of analyses.

Table 5. Likelihood ratio test of spatial covariance between ordinary least squares (OLS) and spatial model assuming homoscedasticity (GLS_{sp}), and between generalized least squares assuming heteroscedasticity (GLS_{he}) and spatial model assuming heteroscedasticity ($GLS_{he\text{-}sp}$) approaches used to assess the impact of soil attributes and growing seasons on corn yield.

Models	DF	Chi-square	Pr > Chi-square
OLS vs GLS_{sp}	2	30.1	0.001
GLS_{he} vs $GLS_{he\text{-}sp}$	2	87.9	0.001

DF = degree of freedom; Significant at probability level of $p<0.001$.

The results of cross-validation analyses showed that the better model was the $GLS_{he\text{-}sp}$, in terms of unbiasedness (CV_1), accuracy (CV_2) and precision (CV_3) of the prediction (Table 6), confirming the results of goodness of fit statistics (Table 4). Moreover, the results obtained for $GLS_{he\text{-}sp}$ can be considered very satisfactory, since the CV_1 was zero (up to four digits), CV_2 was quite close to 1 (0.92), and CV_3 was less than the standard deviation of corn yield at any season (Table 2). These results demonstrated that spatial-temporal models that account for heteroscedasticity of variance were more reliable than other crop prediction models. Additionally, the results of residual analyses for the best model ($GLS_{he\text{-}sp}$) showed that the residual had mean close to zero and standard deviation close to one (Figure 3). Histogram and QQ plot indicated that the residuals were normal, whereas the scattergram of the residuals vs estimates showed no trend (Figure 3). These results further suggested that this model was not biased and the assumptions (Osborne and Waters, 2002) required for regression analysis were confirmed.

Table 6. Cross validation errors of ordinary least squares (OLS), generalized least squares assuming homoscedasticity (GLS_{he}), spatial model assuming homoscedasticity (GLS_{sp}), and spatial model assuming heteroscedasticity (GLS_{he-sp}) approaches used to assess the impact of soil attributes and growing seasons on corn yield.

Model fitted	CV_1	CV_2	CV_3 (Mg ha^{-1})
OLS	-0.1295	53.76	0.559
GLS_{he}	-0.0348	55.62	0.559
GLS_{sp}	-0.0093	0.91	0.538
GLS_{he-sp}	0.0000	0.92	0.489

CV_1 = unbiasedness of the predictor; CV_2 = accuracy of mean squared prediction error; CV_3 = goodness of prediction.

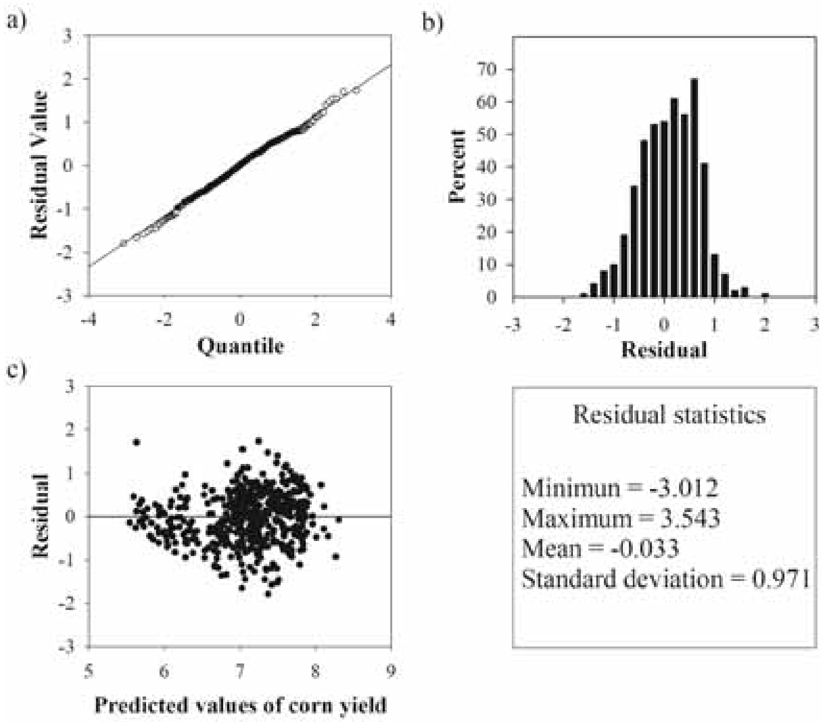

Figure 3. Results of residual statistics for corn yield using spatial-temporal model assuming heteroscedasticity (GLShe-sp).

The estimated coefficients for fixed effects of GLS_{he-sp} model of grain yield are given in Table 3 showing that among the soil attributes, only pH has a significant impact on yield because the main factor that impacts crop yields in most Brazilian agricultural fields is soil acidity (Amado et al., 2009; Dalchiavon et al., 2011; Nogara

Neto et al., 2011). This can be explained because low pH values are often associated with high Al^{3+} contents in Latosols (Oxisol according to the USDA soil Taxonomy) which are prevalent throughout many of the cropped regions of Brazil (Muniz et al., 2011). Crop season had a large impact on yield as indicated by the significant coefficients for year that varied substantially over time. This variability can be ascribed to the total amount of rainfall and distribution of rainfall (days with rain) in March, as confirmed by correlating the average values of corn yield to amount of rainfall in March (r = 0.76) and distribution of rainfall in March (r = 0.84). This can be explained because March is a critical month for corn in São Paulo state, Brazil, since it corresponds to the period in which maize is in its milky/dough stage and water stress may be quite critical.

Growing season (temporal variability) was a significant effect in the $GLS_{he\text{-}sp}$ model as well as in all the other models, causing a different coefficient for each year. The full spatial and temporal nature of the model prevented us from testing the significance of the impact of individual climate variables (e.g., rainfall, growing degree days, humidity) on corn yield because only one datum was recorded for each season. Moreover, the model failed to reach the convergence with these regressors. To explore these relationships, we propose that a study be conducted over a large area and with a longer historical data yield data set.

The results stress also that it is critical to account for temporal variation of the most dynamic soil attributes on a fine spatial scale as they can be influenced by climate factors between the growing seasons. Hence, the relationships between crop yield and these soil attributes may become extremely variable over the seasons, as confirmed by Lambert et al. (2006), who observed spatial and temporal variability of corn yield response (relationship) to P and N in a 5-year corn–soybean rotation in Minnesota. Therefore, it is recommended that farms use a set of data collected over time, not only for yield data, but also for the most temporally variable soil attributes in order to better understand the cause-effect relationships.

2.3.4. Practical implications of model selection

Model choice impacted prediction and regression coefficient estimates and this is critical above all for the soil attributes affected by management. For example, the OLS model suggested that there would be a 0.59 Mg ha^{-1} yield increase for each unit increase in pH (see the parameter estimate in Table 3). This prediction by OLS model was significantly different from the one of GLS_{sp} model (0.30 Mg ha^{-1}) and the one of GLS_{he-sp} model (0.43 Mg ha^{-1}), as confirmed by comparing the confidence limits of pH coefficient estimates of the corresponding models (Table 3). Accounting for spatial autocorrelation in this study had a larger impact on the pH parameter estimate than did accounting for heterogeneity of errors.

The literature suggests that Type I errors in OLS models can result from the violations of the assumptions of independent observations (Schabenberger and Gotway, 2005) and homoscedasticity (Osborne and Waters, 2002). The significance (i.e., OLS) and near significance (i.e., GLS_{sp}) of clay in the models (Table 3) that assumed homoscedasticity were considered to be or nearly be Type-1 errors because clay was clearly not significant in the models that accounted for heteroscedasticity (i.e., GLS_{he} and GLS_{he-sp}) and the GLS_{he-sp} model was assumed to be most accurate. It is important to note however, the coefficient for clay in the OLS and GLS_{sp} models were only 0.004 and 0.005 Mg ha^{-1} corn yield for clay unit, respectively. Experienced agronomists would have correctly concluded that this effect was biologically insignificant.

2.4. Conclusions

The heteroscedastic spatial-temporal autocorrelation model (GLS_{he-sp}) was superior to the other models (OLS, $GLS_{he,}$ and GLS_{sp}) for analyzing yield-soil attributes relationships as determined with cross-validation analysis. This paper supports the modeling of spatial residual errors and accounting for heterogeneity of the variance over time to obtain more accurate estimates of treatment effects for interpreting yield variation. Soil acidity, as assessed by pH, was the soil effect that most influenced corn yield over time in this study. The full spatial and temporal nature of the model prevented testing the significance of the impact of climate variables on yield; however, it was possible to verify that the dynamics of the soil-yield relationship quite likely due to the total amount and distribution of rainfall in March, a period that is critical month for the reproductive stage of maize in São Paulo state. Farms should use a sufficiently long historical data set of crop yield and soil attributes to understand cause-effect relationship between crop yield and soil attributes and predict crop yield accurately.

2.5. References

Amado, T.J.C., L.Z. Pes, C.L. Lemainski and R.B. Schenato. 2009. Chemical and physical attributes of oxisols and their relation with irrigated corn and common bean yields. (In Portuguese, with English abstract.) Rev. Bras. Cienc. Solo 33:831-843.

Bakhsh, A., D.B. Jaynes, T.S. Colvin and R.S. Kanwar. 2000. Spatio-temporal analysis of yield variability for a corn-soybean field in Iowa. Transactions of the Asae 43:31-38.

Bolker, B.M., M.E. Brooks, C.J. Clark, S.W. Geange, J.R. Poulsen, M.H.H. Stevens and J.S.S White. 2009. Generalized linear mixed models: a practical guide for ecology and evolution. Trends in Ecology and Evolution 24:127-135.

Carroll S.S. and N. Cressie. 1996. A comparison of geostatistical methodologies used to estimate snow water equivalent. Water Resource Bulletin 32:267–278.

Dalchiavon, F.C., M.P. Carvalho, O.S. Freddi, M. Andreotti and R. Montanari. 2011. Spatial variability of the bean yield correlated with chemical attributes of a Typic acrustox under no-tillage system. (In Portuguese, with English abstract.) Bragantia 70:908-916.

Diker, K., D.F. Heermann and M.K. Brodahl. 2004. Frequency analysis of yield for delineating yield response zones. Precision Agric. 5:435-444.

Flinn, J.C. and S.K. De Datta. 1984. Trends in irrigated-rice yields under intensive cropping at Philippine research stations. Field Crops Research 9:1-15.

Gee, G.W. and O. Dani. 2002. Particle-Size Analysis. In: J. H. Dane and G. C. Topp, editors, Methods of Soil Analysis, Part 4 - Physical Methods. Soil Science Society of America, Madison, Wisconsin. p. 255-289.

Hurley, T.M., G.L. Malzer and B. Kilian. 2004. Estimating site-specific nitrogen crop response functions. Agron. J. 96:1331-1343.

Lambert, D.M., J. Lowenberg-Deboer and R. Bongiovanni. 2004. A comparison of four spatial regression models for yield monitor data: A case study from Argentina. Precision Agric. 5:579-600.

Lambert, D.M., J. Lowenberg-DeBoer and G.L. Malzer. 2006. Economic analysis of spatial-temporal patterns in corn and soybean response to nitrogen and phosphorus. Agron. J. 98:43-54.

Landagan O.Z. and E.B. Barrios. 2007. An estimation procedure for a spatial–temporal model. Statistics and Probability Letters 77:401-406.

Lark, R.M. 2000. Regression analysis with spatially autocorrelated error: Simulation studies and application to mapping of soil organic matter. International Journal of Geographical Information Science 14:247-264.

Littell, R.C., G.A. Milliken, W. Stroup, R.D. Wolfinger and O. Schabenberber. 2006. SAS system for mixed models SAS Institute, Cary, NC.

Lobell, D.B., J.I. Ortiz-Monasterio, G.P. Asner, R.L. Naylor and W.P. Falcon. 2005. Combining field surveys, remote sensing, and regression trees to understand yield variations in an irrigated wheat landscape. Agron. J. 97:241-249.

McBratney, A., B. Whelan, T. Ancev and J. Bouma. 2005. Future directions of precision agriculture. Precision Agric. 6:7-23.

Moran, P.A.P. 1950. Notes on Continuous Stochastic Phenomena. Biometrika 37:17-23.

Muniz, B.M., N. Curi, G. Sparovek, A. Carvalho Filho and S.R.H.G. Silva. 2011. Updated Brazilian’s Georeferenced Soil Database: An improvement for international scientific information exchanging. In: E. B. Ö. Güngör, editor Principles, Application and assessment in soil science. InTech, Rijeka, Croatia. p. 309-332.

Nogara Neto, F., G. Roloff, J. Dieckow and A.C.V. Motta. 2011. Spatially varied soil and crop attributes related to corn yield. (In Portuguese, with English abstract.) Rev. Bras. Cienc. Solo 35:1025-1036.

Osborne, J. and E. Waters. 2002. Four assumptions of multiple regression that researchers should always test. Practical Assessment, Research and Evaluation 8.

Page, A.L., R.H. Miller and D.R. Keeney. 1982. Methods of soil analysis: Part 2. Chemical and microbiological properties american society of agronomy, Crop Science Society of America, Soil Science Society of America, Madison.

Pierce, F.J. and P. Nowak. 1999. Aspects of precision agriculture. In: L. S. Donald, editor, Advances in Agronomy. Academic Press. p. 1-85.

Raij, B.V., H. Cantarella, J.A. Quaggio and A.M.C. Furlani. 1997. Recomendações de adubação e calagem para o Estado de São Paulo, Instituto Agronômico, Campinas, São Paulo.

Rodrigues, M.S., J.E. Corá and C. Fernandes. 2012. Spatial relationships between soil attributes and corn yield in no-tillage system. Rev. Bras. Cienc. Solo 36:599-609.

Schabenberger, O. and C. Gotway. 2005. Statistical methods for spatial data analysis. CRC Press, Boca Raton, FL.

Shapiro, S.S. and M.B. Wilk. 1965. An analysis of variance test for normality (complete samples). Biometrika 52:591-611.

Snedecor, G.W. and W.G. Cochran. 1989. Statistical methods. 8 ed. Iowa State University Press.

Timlin, D.J., Y. Pachepsky, V.A. Snyder and R.B. Bryant. 1998. Spatial and temporal variability of corn grain yield on a hillslope. Soil Sci. Soc. Am. J. 62:764-773.

CAPÍTULO 3 - Management zones using fuzzy clustering base on spatial-temporal variability of soil and crop yield

ABSTRACT – A management zone in a crop field expresses a sub-region of a field that has a relatively homogeneous combination of yield-limiting factors. Clustering soil and crop data can be used as a basis for the definition of management zones because the data are grouped into clusters based on the similar interaction of these variables. Therefore, the objective of this study were identify management zones using fuzzy c-means clustering analysis based on the spatial and temporal variability of soil attributes and corn yield. The study site (10 by 250-m in size) was located in São Paulo State, Brazil. Corn yield (planted with 0.9-m spacing) was measured in one hundred 4.5x10-m cells along four parallel transects (25 observations per transect) over six growing seasons between 2001 and 2010. Soil chemical and physical attributes were measured. SAS procedure MIXED was used to identify which variable(s) most influenced the spatial variability of corn yield over the five study years. Basis saturation (BS) was the variable that better related to corn yield, thus, semivariograms models were fitted for BS and corn yield and, then, were kriged. Management Zone Analyst (MZA) software was used to carry out the fuzzy *c-means* clustering algorithm. Spatial dependence was observed for corn yield and BS at all study years. The optimum number of management zones can change over time, therefore, the choice of how many management zones should be carry out base on a temporal data set. The degree of agreement between the BS and corn yield management zone maps can change over time. It is very important take into account the temporal variability of crop yield and soil attributes to delineate management zones accurately. The fuzzy c-means cluster analysis based on the spatial and temporal variability of soil attributes and corn yield was efficient to delineate management zones.

Keywords: MZA, precision agriculture, soil pH, tropical soils, *Zea mays* L.

3.1. Introduction

A management zone in a crop field expresses a sub-region of a field that has a relatively homogeneous combination of yield-limiting factors (VRINDTS et al., 2005; LI et al., 2007). Within-field management zones have many uses, for example, as an alternative to grid soil sampling and to develop nutrient maps for variable rate fertilizer application or relate yield to soil parameters for crop-modeling evaluation (LI et al., 2007).

Several types of information have been used to delineated management zones, such as: terrain attributes, electrical conductivity, soil resistance to penetration, chemical soil attributes, Normalized Difference Vegetation Index (NDVI) image and crop yield (FRAISSE et al., 2001; JIANG et al., 2011; ROSALEN et al., 2011; DELALIBERA et al., 2012). Among these types of information, several studies in Brazil (MOLIN, 2002; MILANI et al., 2006; AMADO et al., 2007) have defined management zones based on yield maps from data collected over time. This is due to yield mapping is a simple, inexpensive tool for monitoring crop yield at fine spatial resolutions.

Some studies (GUEDES FILHO et al., 2010; DIACONO et al., 2012) have correlated temporal crop yield maps to soil attributes maps in order to define correctly management zones. These studies have considered the temporal variability of crop yield only, however, the dominance of factors that influence crop yield variability can change from year to year because of seasonal weather variation (DIKER et al., 2004). Therefore, temporal variability or stability of, both, crop yield and soil attributes should take into account to delineate management zones (LI et al., 2007).

One important question arise when considering managing by zones, which is how many unique zones a field should be divided into. Because, if the user defines a different number of classes, the results of the classification will be different, and consequently, different management zones will be produced (FRAISSE et al., 2001). In order to overcome this problem, cluster analysis procedure, which groups similar individuals into distinct classes through an iterative process called clusters, can be used (FRIDGEN et al., 2004) and the number of management zones can be chosen based on the fuzziness performance index (ODEH et al., 1992) or/and normalized classification entropy (BEZDEK, 1981).

Among the different cluster analysis, fuzzy *c-means* algorithm has been widely used to delineate management zones (FRAISSE et al., 2001; LI et al., 2007; JIANG et al., 2011). In addition, a software developed by FRIDGEN et al. (2004) on the basis of fuzzy c-means clustering algorithm called Management Zone Analyst (MZA) can be used easily for researchers and farms, who just need to input data into the software and will obtain results quickly (JIANG et al., 2011). MZA allows to cluster soil and crop data as a basis for the definition of management zones because the data are grouped into clusters based on the similar interaction between the soil and crop data (FRAISSE et al., 2001).

Therefore, the objective of this study was identify management zones using fuzzy *c-means* clustering analysis based on the spatial and temporal variability of soil attributes and corn yield.

3.2. Material and Methods

3.2.1. Site description

The experiment was conducted in Jaboticabal city, São Paulo State, Brazil (21º 14’ 05’’ S, 48º 17’ 09’’w, 613 m asl). Climatologically, the area belongs to the tropical/megathermal zone or Köppen aw (a tropical climate with dry winter and temperature average of the coldest month above 18 ºC). The mean annual rainfall (1971–2006) is 1417 mm, peaking in the period of October–March and a relatively dry season in the period of April–September. the soil of the experimental area is a clayey Rhodic hapludox ('Latossolo Vermelho distroférrico').

The experimental area has been managed in a corn-fallow rotation under no-tillage for 12 years. Before corn seeding, weeds were eliminated with non-selective herbicides.

3.2.2. Yield and soil sampling, and climatic data

The size of the experimental area was 18 by 250-m with the longer dimension oriented in the East-West direction. Each of the 100 experimental plots had a dimension of 10 by 4.5 m and were arranged in a 25 by 4 grid. The experimental scheme is depicted in Figure 1.

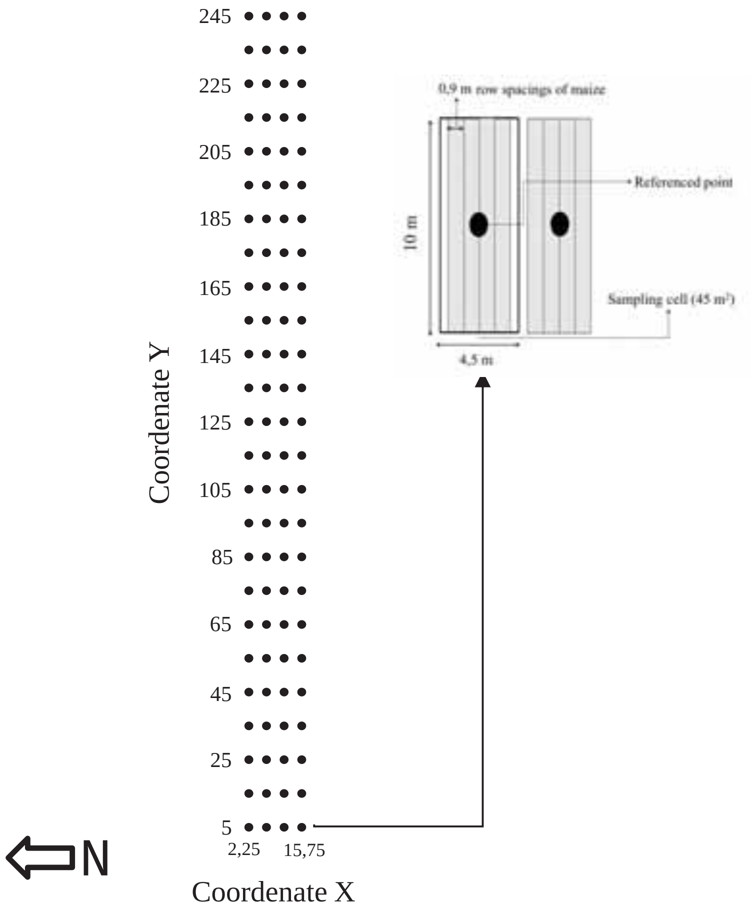

Figure 1. Sampling scheme of soil attributes and corn yield in a Rhodic Hapludox under no-tillage system.

Before planting, non-selective herbicides were applied. Corn (*Zea mays* - triple-hybrid Syngenta Master) was planted at 65,000 plants ha^{-1} with 0.9-m row spacing in early December between 2001 and 2010 growing seasons, but the data were collected only in the 2001/2002, 2002/2003, 2003/2004, 2007/2008, 2008/2009 and 2009/2010 growing seasons. The starter fertilization consisted of 30 kg of N, 70 kg of P_2O_5 and 50 kg of K_2O ha^{-1}. Nitrogen fertilizer (urea) was applied at 100 kg N ha^{-1} when plants had four to six pairs of leaves totally developed. Corn was harvested about 150 days after planting with a 1-row plot combine that deposited the grain into a burlap bag. Grain weights were obtained for each plot with a manual balance in the field. The grain for each plot was sub-sampled for moisture and grain yields were determined at 13% gravimetric moisture.

Each year, five soil sub-samples were collected within each plot using a Dutch auger (0.1 m depth) and were composited. One of the soil sub-samples was collected in the middle of the plot and the other four samples were collected 2-meters apart from the middle in all four cardinal directions from the centroid. The support of soil measurements was assumed similar to the one of crop yield (45 m^2). Each soil composite sample was analyzed for particle size (pipette method) (GEE and DANI, 2002), pH (1:1 soil/water mixture), organic matter (OM) (Walkley-Black method), P (ion-exchange resin**)**, K^+, Ca^{2+} and Mg^{2+} (1M NH_4O Ac. extractable at buffered at pH 7) according to PAGE et al. (1982). From the analytical determinations, cation exchange capacity (CEC = $K^+ + Ca^{2+} + Mg^{2+} + H + Al^{+3}$) and percentage of soil base saturation (BS = ($K^+ + Ca^{2+} + Mg^{2+}$ / CEC) x 100) were calculated.

Monthly cumulative rainfall, growing degree days (base temperature of 10°C), average temperatures, relative humidity and number of days with rainfall were recorded by the climatological station of the São Paulo State University (21º14’05’’ S, 48º17’09’’ W and altitude of 615 m a.s.l.) from December 2001 to April 2010 (Figure 2a and b). The weather station was located 30 m from the experimental site.

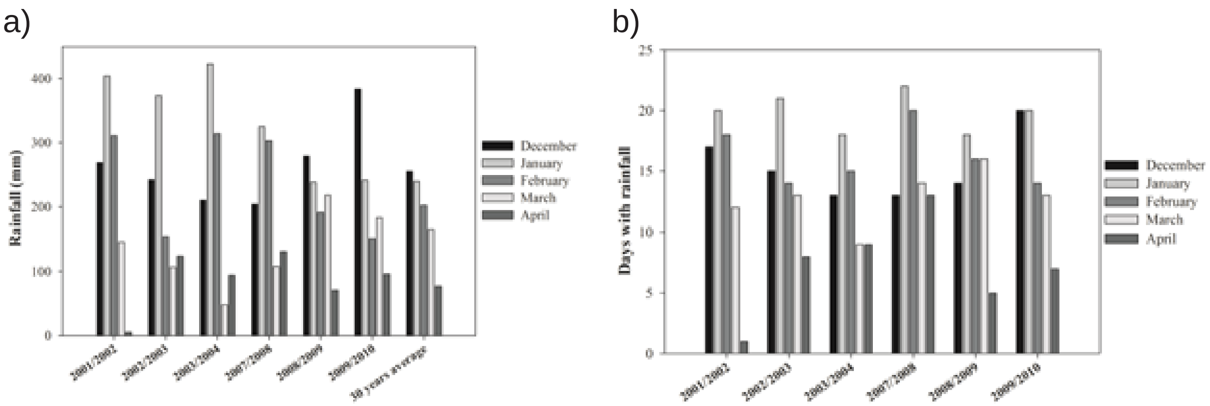

Figure 2. Summarize of climate date: Average monthly rainfall values for the period December–April of the studied years and 30 years average (a); Days with rainfall for the period December-April of the studied years (b).

3.2.3. Preliminary statistics analyses

The heteroscedastic spatial-temporal autocorrelation model was used to identify which variable(s) most influenced the spatial variability of corn yield over the four study years. This procedure was chosen because it takes into account the spatial and temporal variability of crop yield and soil attributes as well as heteroscedastic of variance and it showed better results to determine the yield-limiting factor than the traditional regression analyses using Ordinary least square. A detailed description of this analysis for the present study area was given in Chapter 2.

3.2.4. Data processing

Semivariograms models were fitted to soil attributes (chosen with mixed model) and corn yield data for all study years. The semivariograms were validated with cross-validation analysis, and isotropy in all adjusted models was observed. The soil attributes and corn yield data were interpolated using ordinary kriging, and then gridded to a common 1-m cell. All these analyses were computed using ArcGIS (Redlands, CA) with the Geostatistical Analyst Extension, afterwards data files were imported as a comma delimited text file with the first row as labels for the columns to MZA software.

3.2.5. Data analyses using MZA

Fuzzy clustering algorithms were used to classify data into homogenous zones by MZA software (FRIDGEN et al., 2004). The options settings chosen were measure of similarity as Euclidean, fuzziness exponent = 1.3, maximum number of iterations = 300, convergence criterion = 0.0001, minimum number of zones = 2, and maximum number of zones = 6.

A number of validity statistics were used to determine the best combinations of fuzziness index and number of clusters, as well as the overall clustering performance. The fuzziness performance index (FPI) is a measure of the degree to which different classes share membership (fuzziness) and values are constrained between 0 and 1 (ODEH et al., 1992). The normalized classification entropy (NCE) to aid in deciding how many clusters are most appropriate for creating management zones, which can help the users obtain management zones simply and quickly (BEZDEK, 1981). Minimizing the value for NCE and FPI, the optimum number of clusters can be found.

3.2.6. Overall an specific accuracy analyses

In order to compare the relationship between soil attributes maps and corn yield in every study year two approaches were tested. The first one was the overall accuracy statistic which was calculated using the number of matching cells between maps (i.e., at the same zone) divided by the total number of cells, providing an index of the overall accuracy (i.e. a degree of spatial agreement between the compared data) as suggested by TAGARAKIS et al. (2013). The second one was an adaptation from the first one, which the zones of clustering procedure were classified in high, medium and low for corn yield and soil attributes, then the number of matching cells, i.e., low yield to low soil attribute was divide by the total number of cells, which we called it as specific accuracy.

3.3. Results and Discussion

3.3.1. Preliminary statistics analyses results

The preliminary statistic analyses with mixed models indicated that pH was the variable that better related to corn yield spatially at all study years. High correlation between base saturation (BS) and soil pH is often observed in tropical soils (CATANI and GALLO, 1955) and it also found in our study with Pearson correlation coefficient (r) greater than 0.97. Therefore, we chose BS instead of soil pH, because it is more practical attribute for agronomic management purpose. A complete discussion about mixed model results for this present study area can be found in Chapter 2.

Spatial dependence was observed for corn yield at all study years (Table 1), however, in 2003/2004 growing season corn yield did not show a very good spatial structure, therefore, it was not used in the clustering analyses. Spherical model was fitted for corn yield at the five study years (Table 1). Also spatial dependence was observed for BS at all study years and Gaussian model was fitted for these variables (Table 1). In order to match with corn yield maps, only five years of BS data were used in the clustering analyses. The kriged and gridded to a common 1-m cell maps provided 5061 cells each (Figure 3).

Table 1. Variogram model parameters of corn yield and base saturation of five growing seasons in a Rhodic Hapludox under no-tillage system.

Growing season	Model	Nugget effect	Partial Sill	Range (m)
		Corn Yield		
2001/2002	Spherical	0.07	0.28	12
2002/2003	Spherical	0.15	0.08	47
2007/2008	Spherical	0.10	0.80	27
2008/2009	Spherical	0.65	1.13	51
2009/2010	Spherical	0.35	0.23	49
		Base saturation		
2001/2002	Gaussian	75.59	146.40	25
2002/2003	Gaussian	125	98.07	26
2007/2008	Gaussian	46.13	170.29	20
2008/2009	Gaussian	75	170.42	17
2009/2010	Gaussian	90	150	15

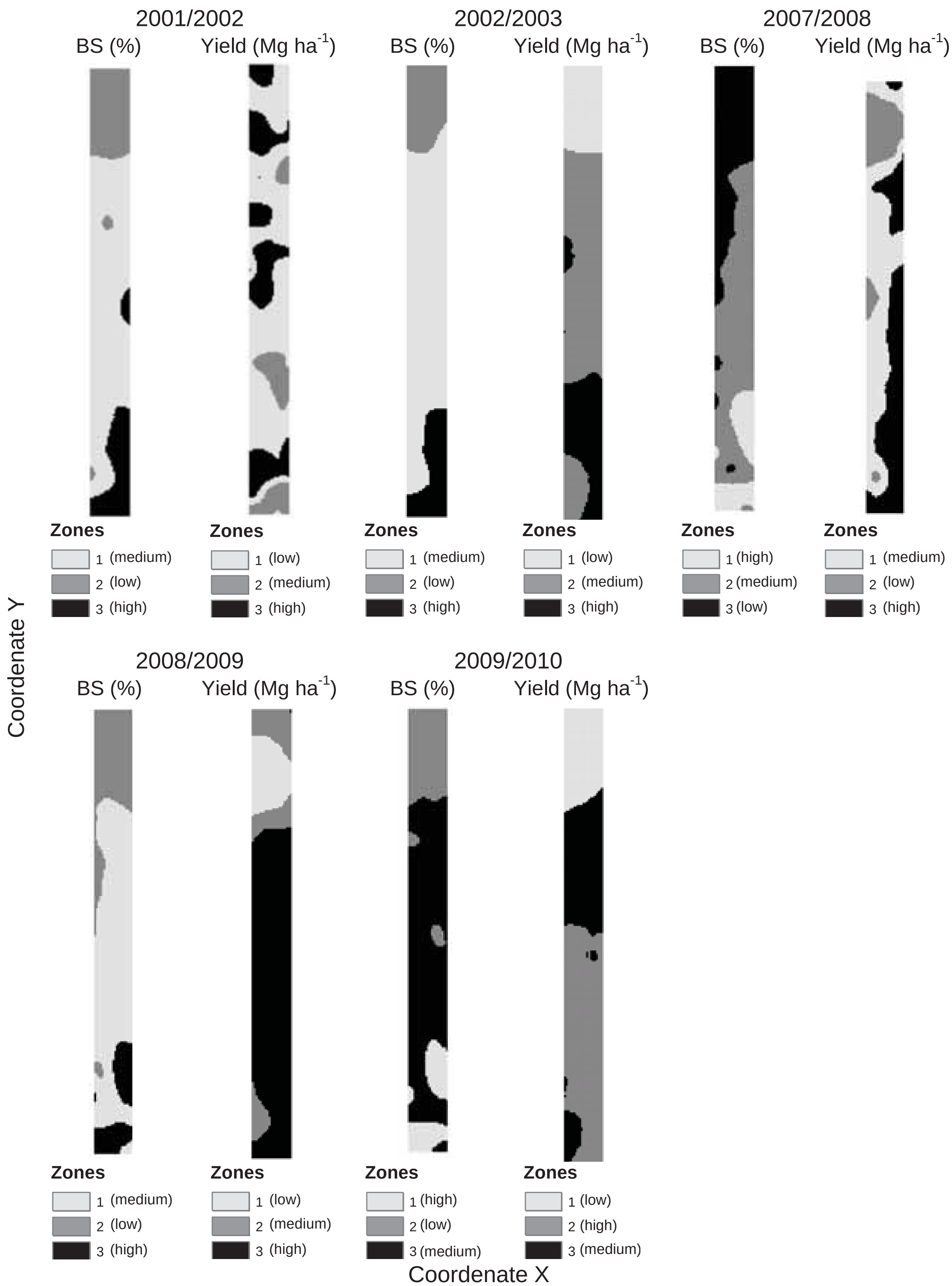

Figure 3. Management zone maps using fuzzy *c-means* clustering analyses of base saturation (BS) and corn yield of five growing seasons in a Rhodic Hapludox (Latossolo Vermelho distroférrico) under no-tillage system.

*low = low values; medium = medium values; high = high values of BS and corn yield.

3.3.2. Fuzzy c-means clustering analyses

The minimum least membership sharing (FPI) and amount of organization (NCE) were similar, that is, the minimum value was obtained at the same zone, for BS and corn yield at all years, excepted, for BS in the 2009/2010 growing season and for corn yield in the 2007/2008 growing season, which were dissimilar (Figure 4). The optimal number of clusters for each computed index is when the index is at the minimum, representing the FPI and NCE as a result of the clustering process. The optimum number of clustering zones, considering both index, varied from 2 to 4. These results indicate that the optimum number of management zones can change over time, thus, the choice of how many management zones should be carry out base on a temporal data set. Fraisse et al. (2001) also pointed out that the optimum number of zones may vary from year to year and is mainly a function of weather and the crop planted.

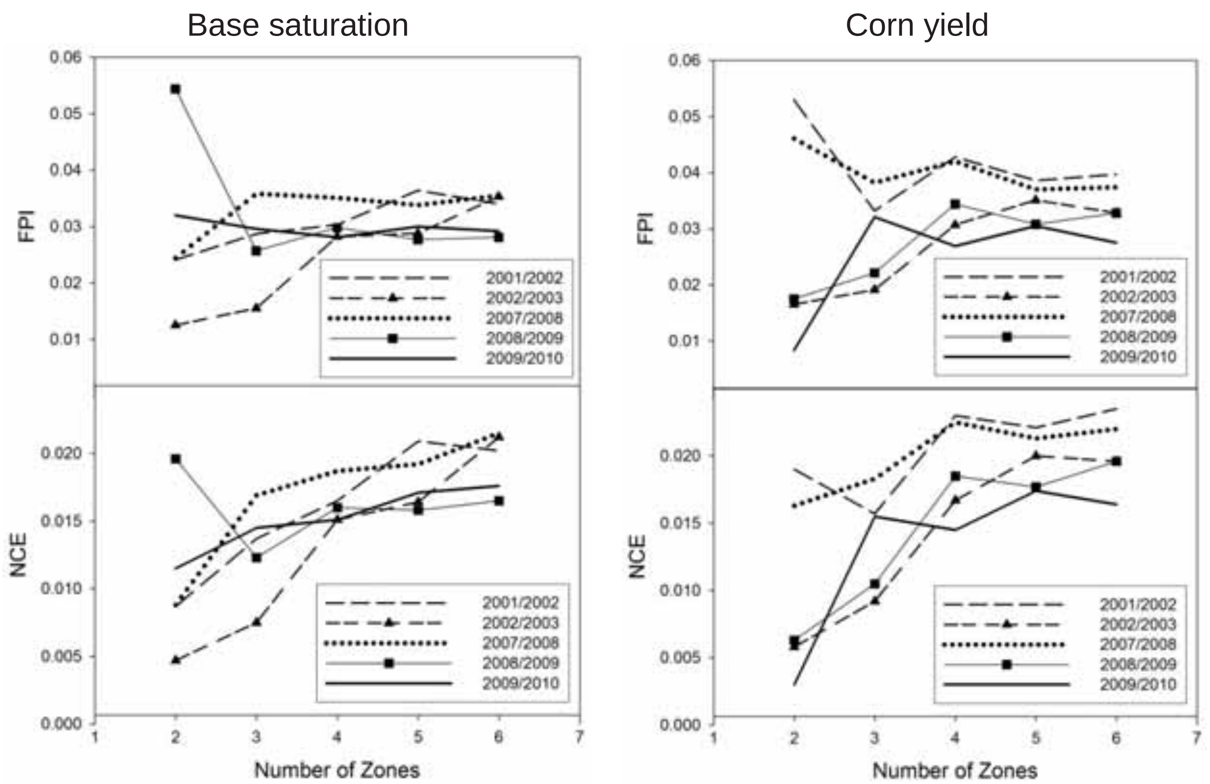

Figure 4. Fuzziness performance index (FPI) and normalized classification entropy (NCE) as calculated by MZA software for base saturation and corn yield of five growing seasons in a Rhodic Hapludox under no-tillage system.

Although, in general, the FPI and NCE index minimum values indicated two management zones, it was observed, based on the zone maps and prior knowledge in this area, that only two management zones would not be sufficient to verify all corn yield variability. Therefore, it was chosen three management zones instead of two. Moreover, it appears that there were a little difference in the index values between the number of two and three zones, whereas a much greater difference between the number of three zones and other zones (Figure 4) was observed, thus, the choice of three management zones can be scientifically accepted.

The overall accuracy statistic between BS and corn yield management maps varied from 0.05 to 0.43 (Table 1). According to KITCHEN et al. (2005) and TAGARAKIS et al. (2013), this statistic can indicate the degree of spatial agreement between the yield-limiting factors and crop yield, that is, the higher this value the better crop yield zone map will be explained by the yield-limiting factors maps. However, the zone maps (Figure 3) did not show a good spatial agreement between BS and corn yield when this statistic was used. For example, in the 2007/2008 growing season, the overall agreement value was very poor, however, the visual analysis of BS vs. corn yield maps disagree with the overall agreement result, because it was verified coincident areas of low BS values with low corn yield values as well as high BS values with high corn yield values (Figure 3). This is due to the MZA software did not standardize the way to show the clustering zones, which demonstrate that a zones can provide the same information in two maps, but with a different identification number for the zone. For example, in the 2007/2008 growing season, the higher and lower BS values were displayed as zone number 1and 3, respectively, whereas the higher and lower corn yield values were displayed as zone number 3 and 2, respectively. Therefore, the overall agreement analyses indicated them as not-coincident zones, which explain the poor values of overall agreement between BS and corn yield for this year.

Table 1. Degree of agreement between the base saturation (BS) and corn yield management zone maps of five growing seasons in a Rhodic Hapludox under no-tillage system.

BS vs. Corn yield maps	*Overall agreement	†Specific agreement
2001/2002	0.43	0.22
2002/2003	0.13	0.81
2007/2008	0.05	0.55
2008/2009	0.21	0.32
2009/2010	0.30	0.59

*Overall agreement = the percentage of pixels belonging to the same zones; †Specific agreement = the percentage of pixel belonging to the same level (low, medium and high values of BS and corn yield) zones.

Hence, base on the inconsistent overall agreement results, we decided propose another approach, which we called specific agreement. Firstly, it was identified among the three zones, which zone present higher, medium and low values of BS and corn yield for all study years base on the kriging maps and prior knowledge of this present area. Afterwards, the number of matching cells with the same level values were compared and divided by the total number of cell of the map. The results of the specific agreement (Table 1) had a much better relationship with the visual analyses of the management zones than those results obtained with the overall agreement analyses (Figure 3).

The specific agreement results varied from 0.22 to 0.81 at the five study years (Table 1). These demonstrate that the degree of agreement between the BS and corn yield management zone maps can change over time, which is in agreement with Li et al. (2007), who pointed out that defining management zones should be rely on spatial information that is stable or predictable over time and is related to crop yield. Therefore, a temporal data set of crop yield and yield-limiting factors should be used to delineate management zones.

Based on the specific agreement analyses, BS can be used to delineate management zones in this present study area, since the values of specific agreement were higher than 0.5 in three out of the five study years (0.81; 0.59 and 0.55). This is expected because the main factor that impacts crop yields in most Brazilian agricultural fields is soil acidity (DALCHIAVON et al., 2011; NOGARA NETO et al., 2011).

The results of the specific agreement analyses (Table 1) did not show any direct correlation to the climatic variables assessed. However, this dissimilarity between BS and corn yield zone maps due most likely to related to temporal variability from many other variables such as germination rates, weeds, insects, diseases and climate which are related to crop yield (RODRIGUES et al., 2012), but were not assessed in this study.

For farmers to adopt site-specific management, the development of management zones must be simple, functional, and economically feasible (LI et al., 2007), thus, we proposed to divide the study field into three zones based on the zone maps obtained from the clustering analyses using MZA software for the five study years and it is depicted in Figure 5.

Figure 5. Management zones based on the fuzzy *c-means* clustering analyses of base saturation and corn yield of five growing seasons in a Rhodic Hapludox under no-tillage system.
*Zone 1: high yield potential; Zone 2: medium yield potential; Zone 3: high yield potential.

Zone 1 is the location in the study area with higher BS and corn yield values; zone 2 is the location with medium BS and corn yield values; and zone 3 is the location with lower BS and corn yield values. This zone delineation can allow to use site-specific management in this areas such as alternative to grid soil sampling (LI et al., 2007) and develop lime requirement recommendation maps for variable rate fertilizer application (CORÁ and BERALDO, 2006). When compared to the size of commercial fields, the area in the present study may be considered small, however, the fuzzy *c-means* clustering analysis using temporal data set showed potential to delineate management zones and it may be spread out over larger crop fields. In this study crop yield was related only to soil acidity (i.e., BS), however, the fuzzy *c-means* technique can cluster many yield-limiting factors (i.e., soil attributes) at the same time as showed by FRIDGEN et al. (2004) and TAGARAKIS et al. (2013). Management zone analytics software (MZA) showed be a easy and feasible tool to delineate management zones based on fuzzy *c-means* clustering as found by FRIDGEN et al. (2004), LI et al. (2007), and JIANG et al. (2011).

3.4. Conclusions

The fuzzy *c-means* cluster analysis based on the spatial and temporal variability of soil attributes and corn yield was efficient to delineate management zones. It is very important take into account the temporal variability of crop yield and soil attributes to delineate management zones accurately. Management zone analytics software was a reliable and simple tool to delineate management zones. Soil acidity, as assessed by BS, was the soil effect that most influenced corn yield over time in this study.

3.5. References

AMADO, T. J. C.; PONTELLI, C. B.; SANTI, A. L.; VIANA, J. H. M.; SULZBACH, L. A. D. S. Variabilidade espacial e temporal da produtividade de culturas sob sistema plantio direto. **Pesquisa Agropecuária Brasileira**, Brasília, v. 42, n. 8, p. 1101-1110, 2007.

BEZDEK, J. C. **Pattern Recognition with Fuzzy Objective Function Algorithms**. Kluwer Academic Publishers: Norwell, 1981. p. 256.

CATANI, R. A.; GALLO, J. R. Avaliação da exigência em calcário dos solos do Estado de São Paulo mediante correlação entre o pH e a porcentagem de saturação em bases. **Revista Agricultura**, Piracicaba, v. 30, p. 49-60, 1955.

CORÁ, J. E.; BERALDO, J. M. G. Variabilidade espacial de atributos do solo antes e após calagem e fosfatagem em doses variadas na cultura de cana-de-açúcar. **Engenharia Agrícola**, Jaboticabal, v. 26, n. 2, p. 374-387, 2006.

DALCHIAVON, F. C.; CARVALHO, M. P.; FREDDI, O. S.; ANDREOTTI, M.; MONTANARI, R. Variabilidade espacial da produtividade do feijoeiro correlacionada com atributos químicos de um Latossolo Vermelho Distroférrico sob sistema de semeadura direta. **Bragantia**, Campinas, v. 70, n. 4, p. 908-916, 2011.

DELALIBERA, H. C.; WEIRICH NETO, P. H.; NAGATA, N. Management zones in agriculture acording to the soil and landscape variables. **Engenharia Agrícola,** Jaboticabal, v. 32, n. 6, p. 1197-1204, 2012.

DIACONO, M.; CASTRIGNANÒ, A.; TROCCOLI, A.; DE BENEDETTO, D.; BASSO, B.; RUBINO, P. Spatial and temporal variability of wheat grain yield and quality in a Mediterranean environment: A multivariate geostatistical approach. **Field Crops Research,** Amsterdam, v. 131, p. 49-62, 2012.

DIKER, K.; HEERMANN, D. F.; BRODAHL, M. K. Frequency analysis of yield for delineating yield response zones. **Precision Agriculture**, New York v. 5, n. 5, p. 435-444, 2004.

FRAISSE, C. W.; SUDDUTH, K. A.; KITCHEN, N. R. Delineation of site-specific management zones by unsupervised classification of topographic attributes and soil electrical conductivity. **American Society of Agricultural and Biological Engineers. Transactions**, Saint Joseph, v. 44, n. 1, p. 155-166, 2001.

FRIDGEN, J. J.; KITCHEN, N. R.; SUDDUTH, K. A.; DRUMMOND, S. T.; WIEBOLD, W. J.; FRAISSE, C. W. Management Zone Analyst (MZA): Software for subfield management zone delineation. **Agronomy Journal**, Madison, v. 96, n. 1, p. 100-108, 2004.

GEE, G. W.; DANI, O. Particle-Size Analysis. In: DANE, J. H.; TOPP, G. C. (Ed.). **Methods of Soil Analysis, Part 4 - Physical Methods**. Madison: Soil Science Society of America, 2002. p.255-289.

GUEDES FILHO, O.; VIEIRA, S. R.; CHIBA, M. K.; NAGUMO, C. H.; DECHEN, S. C. F. Spatial and temporal variability of crop yield and some Rhodic Hapludox properties under no-tillage. **Revista Brasileira de Ciência do Solo**, Viçosa, v. 34, n. 1, p. 1-14, 2010.

JIANG, Q.; FU, Q.; WANG, Z. Study on delineation of irrigation management zones based on Management Zone Analyst software. In: LI, D., LIU, Y. e CHEN, Y. (Ed.). **Computer and Computing Technologies in Agriculture IV**: 4th IFIP TC 12

Conference, CCTA 2010, Nanchang, China, October 22-25, 2010, Selected Papers, Part III. Boston: Springer, v.346, 2011. p.419-427.

KITCHEN, N.; SUDDUTH, K.; MYERS, D.; DRUMMOND, S.; HONG, S. Delineating productivity zones on claypan soil fields using apparent soil electrical conductivity. **Computers and Electronics in Agriculture**, Amsterdam, v. 46, n. 1, p. 285-308, 2005.

LI, Y.; SHI, Z.; LI, F.; LI, H.-Y. Delineation of site-specific management zones using fuzzy clustering analysis in a coastal saline land. **Computers and Electronics in Agriculture,** Amsterdam, v. 56, n. 2, p. 174-186, 2007.

MILANI, L.; SOUZA, E. G.; URIBE-OPAZO, M. A.; GABRIEL FILHO, A.; JOHANN, J. A.; PEREIRA, J. O. Unidades de manejo a partir de dados de produtividade. **Acta Scientiarum. Agronomy**, Maringá, v. 28, n. 4, p. 591-598, 2006.

MOLIN, J. P. Definição de unidades de manejo a partir de mapas de produtividade. **Engenharia Agrícola**, Engenharia, v. 22, n. 1, p. 83-92, 2002.

NOGARA NETO, F.; ROLOFF, G.; DIECKOW, J.; MOTTA, A. C. V. Atributos de solo e cultura espacialmente distribuídos relacionados ao rendimento do milho. **Revista Brasileira de Ciência do Solo**, Viçosa, v. 35, n. 3, p. 1025-1036, 2011.

ODEH, I. O. A.; CHITTLEBOROUGH, D. J.; MCBRATNEY, A. B. Soil pattern recognition with fuzzy-c-means: Application to classification and soil-landform interrelationships. **Soil Science Society of America Journal,** Madison**,** v. 56, n. 2, p. 505-516, 1992.

PAGE, A. L.; MILLER, R. H.; KEENEY, D. R. **Methods of Soil Analysis: Part 2. Chemical and Microbiological Properties**. Madison: American Society of Agronomy, Crop Science Society of America, Soil Science Society of America, 1982. 1159.

RODRIGUES, M. S.; CORÁ, J. E.; FERNANDES, C. Spatial relationships between soil attributes and corn yield in no-tillage system. **Revista Brasileira de Ciência do Solo**, Viçosa, v. 36, n. 2, p. 599-609, 2012.

ROSALEN, D. L.; RODRIGUES, M. S.; CHIODEROLI, C. A.; BRANDÃO, F. J. C.; SIQUEIRA, D. S. GPS receivers for georeferencing of spatial variability of soil attributes. **Engenharia Agrícola**, Jaboticabal, v. 31, n. 6, p. 1162-1169, 2011.

TAGARAKIS, A.; LIAKOS, V.; FOUNTAS, S.; KOUNDOURAS, S.; GEMTOS, T. Management zones delineation using fuzzy clustering techniques in grapevines. **Precision Agriculture**, New York, v. 14, n. 1, p. 1-22, 2013.

VRINDTS, E.; MOUAZEN, A. M.; REYNIERS, M.; MAERTENS, K.; MALEKI, M. R.; RAMON, H.; DE BAERDEMAEKER, J. Management Zones based on Correlation between Soil Compaction, Yield and Crop Data. **Biosystems Engineering**, London, v. 92, n. 4, p. 419-428, 2005.

CAPÍTULO 4 - On-the-go soil ph sensor with a proposed new method for correcting interpolation errors

ABSTRACT – The Veris MSP® on-the-go soil sensing system allows rapid mapping of soil pH along with apparent soil electrical conductivity and optional visible/near-infrared soil reflectance. However, for this sensor to be effective for site-specific lime management, maps created with this instrument must be of adequate quality. The goal of this study was to evaluate the Veris on-the-go pH sensor system for soil pH management by assessing the accuracy of soil pH measurements and associated spatial predictions of buffer pH (BpH), and lime requirement proposed a new method for correcting interpolation errors. This research was conducted at three Kentucky study locations where intensive pH sensor data (i.e,. 12-m parallel passes) were collected in each fields and with additional coarser (along 36-m passes) measurements collected in approximately orthogonal directions. Soil samples were also collected and used to relate sensor pH measurements with laboratory measures of water and Buffer pH. Semivariograms and interpolations (ordinary Kriging, inverse distance weighted, local polynomial interpolation, and radial basis functions) were applied to create sensor-based soil pH maps using either the mean electrode pairs (MEP), or the individual electrode (IEV) values. A simple linear regression was used to calibrate sensor-based measurements to measurements obtained according to laboratory procedures. The sensor-based soil pH measurements were significantly correlated with laboratory pH ($0.52 \leq R^2 \leq 0.87$) and Buffer pH ($0.44 \leq R^2 \leq 0.84$). The range of spatial structure was larger for the Henderson (175 m) than Fayette (120-m) and Shelby (80-m) locations. All fields had relative nugget effects (RNE) ranging between 35 and 69%, which may have contributed to the larger than desired interpolation errors for pH (i.e., 0.20 ≤ Root mean square error ≤ 0.44). While the use of the IEV rather than MEP data sets reduced the nugget variance, this did not significantly improve spatial predictions. The choice of interpolation method also had little impact on prediction errors. Spatial interpolation errors were larger than desired and can result in substantial under application of lime. However, these errors can be reduce by simply using the calibration data set to correct for these inaccuracies.

Keywords: lime requirement, buffer pH, spatial variability

4.1. Introduction

Understanding the spatial variability of soil properties is fundamental for the implementation of site-specific crop management. It allows understanding cause-effect relationships between crop yield and soil properties as well as applying fertilizers and lime according to local needs. However, a large number of samples are required to obtain map with sufficient quality.

Sampling based on a pre-defined 1-ha grid pattern has become the most common strategy in many parts of the USA and in other countries (V. Adamchuk, et al., 2007). However, this sampling intensity may not be enough to resolve within-field variability of soil properties (Mueller, et al., 2001; Mueller, et al., 2004a) and increased density of conventional sampling has been shown to be impractical because of high analysis costs. Therefore, one possible way to solve this problem is to develop sensors that monitor soil properties at a high intensity.

On-the-go sensing of soil properties offers a potential solution to increase the sampling density at a relatively low cost. In 2003, an automated on-the-go soil pH sensor based on the technologies developed by Adamchuk et al. (1999), was released commercially by Veris-Technologies (Salina, KS), and it has been widely used in crop fields in the USA (Erickson, 2004) and abroad. It can be operated, for example, at 8 km hr^{-1} along transects with 20 m between passes conducting measurements on average every 10 s, resulting in high mapping densities (e.g., 20 samples per hectare) (Lund, et al., 2004). The on-the-go sensor pH values have been found correlated with conventional laboratory measurements of pH in studies by Adamchuk et al. (1999) (in-field result, R^2 = 0.83), Lund et al. (2004) (R^2 = 0.81), Adamchuk et al. (2007) (R^2 = 0.73-0.98), Staggenborg et al. (2007) (R^2 =0.58-0.78), and Olfs et al. (2010) (R^2 = 0.62).

Soil pH measurements are indicators of soil acidity, but do not measure potential acidity, thus, soil buffer pH is required to measure reserve acidity in soil and, hence, it is used to determine the lime requirement (Havlin, et al., 2005). Therefore, the on-the-go pH sensor or combination of several accessible high-density data layers (e.g., apparent soil electrical conductivity) should provide a good relationship between high-density data and laboratory-based buffer pH (BpH) values.

Staggenborg et al. (2007) observed that buffer pH predictions using only the on-the-go pH sensor were less accurate than water pH predictions (i.e., R^2 = 0.04 and 0.55) on two fields in Kansas. In spite of that, these authors pointed out that if a relationship between lab pH and BpH was developed and used to predict BpH from the pH sensor, the accuracy of BpH maps has improved (R^2 = 0.75 to 0.95), suggesting that the pH sensor could be used to predict BpH and, consecutively, lime requirements.

The spatial variability of soil properties is often assessed with data interpolation. Selecting a proper spatial interpolation method is important, since different methods of interpolation can lead to different results (Zandi, et al., 2011). Two most commonly used interpolation methods for site-specific fertility management are inverse distance weighted (IDW) and Kriging. However, other methods, such as local polynomial interpolation (LPI) and radial basis functions (RBF) have been used as well. All these options can be easily performed using most modern GIS (Geographic Information Systems) software packages.

The LPI deterministic interpolation technique is considered local because the prediction calculation is performed from only the measured neighborhood points. It is an inexact interpolator because it can predict a value different from that measured at a sampled location. RBF are a series of exact interpolation techniques, that is, the surface must go through each measured sample value. It works similar to fitting a rubber membrane through the measured sample values while minimizing the total curvature of the surface. Theoretical detail and methods for using these interpolation procedures have been described by Johnston et al. (2001).

There are still few studies that use LPI and RBF for site-specific soil fertility management. Mueller (2007) observed that ordinary Kriging, IDW, and RBF generally produced soil pH and BpH maps of similar quality, but LPI methods produced maps of inconsistent quality. Silva et al. (2010), found that IDW and Kriging performed better than LPI technique when assessing the spatial variability of soil pH in an Oxisol under a coffee field. Zandi et al. (2011) pointed out that RBF was the most suitable methods for prediction and mapping the spatial distribution of soil pH in a New Zealand vineyard.

Map quality is typically determined by comparing observed mapped values with those obtained by various quantitative and qualitative analytical techniques. Plots of predicted vs. measured values should always be visually examined to assess prediction quality. One of the quantitative measures is map precision, calculated as the standard deviation of the residuals. Map accuracy, another quantitative measure, is the square of bias, and bias is the average of the residuals. The mean square error (MSE) is the sum of precision and accuracy. Frequently, root mean square error (RMSE) is used to compare interpolation methods because it is expressed in the same unit as the study variable (Mueller, et al., 2004b).

The choice of the interpolator is highly dependent on the methods used to calculate predicted and measured values (Mueller, et al., 2004b). Cross-validation with replacement is a rapid, inexpensive procedure for comparing predicted and measured values. Unfortunately, it does not adequately describe spatial prediction errors in many situations (Isaaks and Srivastava, 1989; Mueller, et al., 2001). Validation with an independent data set is a superior and more dependable approach for measuring residuals. Calibration data sets that are recommended for farmers to correcting pH maps (V. Adamchuk, et al., 2007) could also serve as an indicator of quality of predictability. An alternative approach for validating repeatability of interpolation maps created with the sensor is to use an independent data set created with the sensor itself rather than with a separate soil sample data set.

The overall goal of this study was to evaluate the potential use of the Veris on-the-go pH sensor system for soil management in Kentucky. The specific objectives were as follows: 1) study the relationship between sensor pH and laboratory measurements of pH and buffer pH; 2) evaluate and compare methods for spatial estimates of pH, Buffer pH, and lime recommendation, and 3) propose and demonstrate a new procedure for correcting interpolation errors.

4.2. Material and Methods

4.2.1. Site description

The study was conducted across three agricultural fields located in Fayette (38° 6' 1" N, 84° 29' 33"W; 5 ha), Henderson (37° 50' 15" N, 87° 48' 33" W; 8 ha) and Shelby (38° 16' 56" N, 85° 9' 19" W; 25 ha) Counties, Kentucky. Detailed soil taxonomic classification is given in Figure 1. The Fayette county location had been under no-till in a corn-wheat-soybean (Zea mays-Triticum aestivum-Glycine max) rotation for 10 years and continuous maize for five years prior to sampling. The Henderson County location was deep-tilled every fifth year and vertically tilled every other year. It was under continuous soybean for the last eight years with the exception of the final year when Sorghum (Sorghum sp.) was grown. The Shelby county field was no-till under a corn-soybean and corn-wheat rotation for more than 20 years prior to sampling.

4.2.2. Sensor measurements

Soil pH in all fields was measured with a Veris Mobile Sensor Platform (MSP). The instrument was used in the fields for at least half an hour to insure that the electrodes were properly worn. Then the pH electrodes were standardized with pH 4 and 7 buffer solutions. The sensor automatically retrieved a sample with a horizontal coring device (5.9 cm inner- diameter) that was slightly angled in the downward direction. The coring device lifted the soil and pressed it against two ion-selective electrodes as the system was traveling through the fields. The sample depth was set to 9 cm so the sample was considered to be from 4 to 9 cm depth for all study fields. The electrodes remained in contact with the soil until the measurements settled for a maximum time set at 30 s (Veris-Technologies, 2010). A rinsing system provided water for pH measurements and cleaning of the electrodes after each reading. The source of the water was ground water from each farm shop. Sensor measurements were obtained from the Veris pH sensor according to the designs depicted in Figure 1 and the specifications indicated in Table 1.

Maps indicating the location of intensive, coarse, and calibration sensor measurements are provided in Figure 1 with one exception: a coarse dataset was not collected for the Shelby County location. Table 1 provides details about the various datasets.

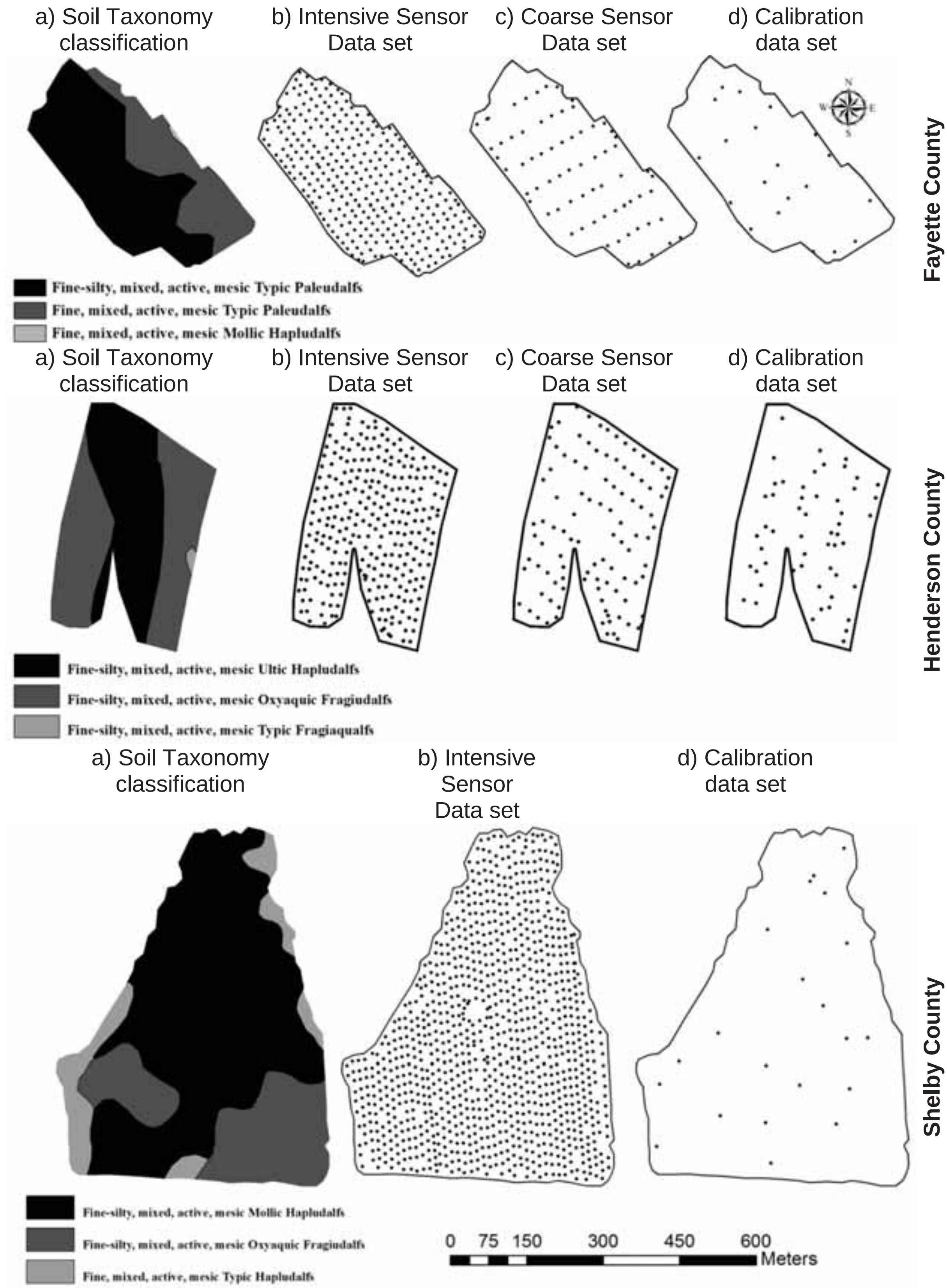

Figure 1. Soil taxonomic classification and sampling design in three crop fields in Kentucky, USA.

Table 1. Specification of the Veris MSP® on-the-go sensor measurement carried out in three fields in Kentucky.

	Fayette County	Henderson County	Shelby County
Sampling dates	July, 12, 15 and 18 of 2011	August, 17 and 18 of 2011	January, 11 and 12 of 2007
pH Electrodes	Antimony	Antimony	Glass
Average tractor speed	4.2 km h^{-1}	6.8 km h^{-1}	6.3 km h^{-1}
Intensive Sensor Data sets			
Dist. betw. transects (m)	12	12	12
Number of observations	309	284	1046
Coarse Sensor Data sets			
Dist. betw. transects (m)	36	36	not collected
Number of observations	69	102	
Calibration Data sets			
Number of observations	20	54	21

Dist. betw. transects: Distance between transects.

For the Fayette and Shelby fields, the locations of calibration points were selected by the pH sensor operators after visual inspection of the intensive sampling dataset to insure that the calibration points were evenly distributed spatially throughout the fields and across the range of pH sensor variability (e.g., similar numbers of low, medium, and high pH values). At the Henderson location, the calibration data set was collected simultaneously with the intensive data and only a subset of the intensive sampling locations were used also for calibration. These calibration points were selected to allow small scale variability and anisotropy to be resolved efficiently.

The Mean of the Electrode Pairs (MEP) was the mean of the two electrodes. This data was included in the Veris sensor output provided by the software that came with the instrument. Values were removed if pH measurements 1) did not settle within 30 seconds, 2) were not between 3.5 and 9.5, and 3) if the absolute value of the difference between the two electrodes was greater than 0.50 pH units.

We calculated the Individual Electrode Values (IEV) from the raw Veris data. Specifically, the slopes and offsets for the calibration curve created using the pH 4 and 7 buffers, were back calculated from the processed pH and raw data files. These coefficients were then used to calculate predicted pH values from the raw electrode readings. The same criteria that were used to filter the MEP data set were also used to filter the raw data set. The separate electrode values from the raw data file were combined to create the IEV.

A spatial adjustment of the coordinates of one of electrode values of each IEV pair was necessary because Kriging (to be described in a subsection below) could not be performed with multiple measurements that have identical coordinates. Specifically, the coordinates for one of the electrodes was offset from the other by 0.02 meter in the East-West direction (the distance corresponded to the separation distance of the electrodes). This adjustment was done with the full knowledge that it introduced an inaccuracy into the IEV data set; however, the goal with this analysis was to determine whether that any prediction errors that occurred because of the adjustment would be more than offset by improved predictions through better resolution of micro-scale variability (i.e., accurately determining the nugget variance). The number of observations in the IEV data sets were 618, 568 and 2092 observations for the in the Fayette, Henderson, and Shelby fields data sets, respectively. These values were twice the number of those reported in Table 1.

4.2.3. Soil Sampling and laboratory analysis

Next soil samples were collected for analysis at the calibration points indicated in Figure 1. At all study sites, the precise locations of the calibration sensor points (i.e., the locations where the Veris MSP® pH sensor lifted out of the ground) were flagged by an individual following on foot behind the tractor. Then six (2.1-cm diameter) soil sub-samples were collected with a hand probe to a depth of 10-cm and within a 1-m^2 circle centered at the flagged point and composited into one bag.

The soil samples for the calibration dataset were forced air dried and sent to the soil testing laboratory located within the Division of Regulatory Services at the University of Kentucky. The samples were ground to pass a 2 mm sieve and soil water pH (1:1 soil:water mixture) and Sikora buffer pH (Sikora, 2006) were measured.

4.2.4. Correction of the on-the-go sensor measurements

Regression analyses for the calibration dataset were conducted to relate point sensor pH measurements with the laboratory measurements of soil pH and Buffer pH. The regression equations were used to correct the sensor values for the coarse and intensive and calibration datasets.

4.2.5. Geostatistical, interpolation, and validation analyses

The corrected intensive sensor Buffer pH dataset (i.e., IEV and MEP) were tested for normality (Shapiro & Wilk, 1965) using SAS software (SAS Institute, Inc.; Cary, North Carolina) and fit with semivariogram models using ArcGIS (Redlands, CA) with the Geostatistical Analyst Extension.

Next, spatial prediction techniques (i.e., ordinary Kriging, inverse distance weighted, local polynomial interpolation, and radial basis functions) were used to interpolate the corrected intensive IEV and MEP dataset values using ArcGIS (Redlands, CA) with the Geostatistical Analyst Extension. The semivariogram models were used for Kriging. This allowed the creation of interpolated maps for the individual and mean electrode values.

Next the corrected point calibration and coarse datasets were used as validation datasets to test the interpolated quality of the for the IEV and MEP interpolations from the intensive datasets using ArcGIS (Model building). Prediction errors (bias and RMSE) and an indicator of map goodness R^2 were calculated for these validation analyses and reported. A detailed description of these measures of error and map goodness have been described in other papers (e.g., Mueller et al., 2004a).

4.2.6. Calculation of lime recommendations

The 2012-2013 Lime and Nutrient Recommendations for Kentucky (i.e., AGR-1) (Kentucky, 2012) were not used to directly estimate lime requirement from sensor values. This was because the lime recommendations in this document are based on both pH and buffer pH and the Veris sensor only provides one value. Therefore, lime requirement recommendations were based on the SMP buffer publication by Shoemaker, et al. (1961). The recommendations were presented in table format a simple linear regression was used to calculate the lime recommendation from buffer pH as follows:

$$LR_{SMP} = 83.6 + -12.0 \times \text{buffer pH} \tag{1}$$

where LR_{SMP} is the lime recommendation based on the SMP buffer according to the Shoemaker, et al. (1961) publication. A target pH 6.4 was used because the crops in this field fields had historically been used for corn, soybeans, and small grains. AGR-1 recommends using a target pH of 6.4 for these crops.

Since the LR_{SMP} was not calibrated for Kentucky soils, sampling depths, and the Sikora Buffer, the AGR-1 publication was used to adjust LR_{SMP} Specifically, lime recommendations were calculated for our calibration datasets with using Equation 1 (i.e., LR_{SMP}) and with the AGR-1 table for a target pH of 6.4 and the following linear relationship was determined with regression:

$$LR = 0.734 + 1.70 \times LR_{SMP} \tag{2}$$

where LR indicates the lime recommendation based on the SMP buffer (i.e., LR_{SMP}) adjusted for Kentucky soil with AGR-1. Adjusted lime recommendation calculated with eq. 1 and 2 were used to calculate the predicted lime requirement maps from the corrected interpolated Buffer pH maps.

4.2.7. Correcting for lime recommendations interpolation errors

The purpose of the analyses was to determine whether interpolation errors could be removed from prediction maps. A simple approach corrected interpretations by forcing predicted and measured values to adhere to the 1:1 line. Specifically, regression analyses were used to relate laboratory measures of buffer pH with interpolated mean electrode values for the calibration dataset. The resultant regression equations were used to adjust the lime recommendation maps to correct for interpolation errors.

4.2.8. The potential for EC to improve estimates of pH

Next, the potential for apparent soil electrical conductivity (EC_a) to be used for enhancing predictions of lime recommendations was assessed but only for the Henderson location. This was the only field location where pH_{Sensor}, $EC_{shallow}$, and soil test pH and Buffer pH were assessed at the same points. Specifically, multiple regression model was developed with pH_{Sensor}, $EC_{shallow}$, and the interaction between pH_{sensor} and $EC_{shallow}$ (pH_{Sensor}*$EC_{shallow}$) as independent variables for the Henderson field. LR was used as a dependent variable. For the Henderson location, variables with the greatest p-value that was greater than 0.05 were sequentially removed from the regression analyses until all values were significant. These regression analyses were computed with SAS (SAS Institute, Inc.; Cary, North Carolina) using the REG procedure.

4.3. Results and Discussion

4.3.1. Relationship sensor pH and laboratory pH

Sensor pH values and laboratory soil pH and buffer pH for the calibration data sets (Table 2) were significantly correlated (P > 0.01) for all three fields. The strongest relationships were found in the Fayette location (R^2 = 0.87). The coefficients of determination were lower for the Henderson (R^2 = 0.52) and Shelby (R^2 = 0.65). Regression relationships (Table 2) all had slope values less than 1.0 and intercept values greater than zero indicating that the sensor overestimated pH values consistent with the findings of Adamchuk et al. (1999), Lund et al. (2004), and Olfs et al. (2010). This reinforces the importance of the need for a calibration data set as recommended by Adamchuk et al. (2007).

Table 2. Regression between Lab H_2O pH and Sikora buffer pH, and Sensor pH values from calibration data set in three crop field in Kentucky.

Equations	Fayette County	Henderson County	Shelby County
Lab H_2O pH eq.	0.364 + (0.910 x Sensor pH)	1.602 + (0.671 x Sensor pH)	2.382 + (0.561 x Sensor pH)
R^2	0.87	0.52	0.65
buffer pH eq.	3.457 + (0.500 x Sensor pH)	4.834 + (0.295 x Sensor pH)	4.224 + (0.362 x Sensor pH)
R^2	0.84	0.44	0.72

The relationship between the Sikora buffer pH (bpH) and proximal soil pH sensor values (Table 2) were fairly strong for the Fayette (R^2 = 0.84) and Shelby (R^2 = 0.72) locations and smaller, but still significant, for the Henderson (R^2 = 0.44) field. High correlation found between buffer pH and pH sensor indicated that the on-the-go pH sensor may be used to recommend lime requirement without apparent need for additional high-density data representing spatially different soil buffering characteristics.

4.3.2. Sensor data and lime recommendation

The multiple regression analyses for the calibration data sets indicated that $Sensor_{pH}$ could be used to predict lime recommendation (Table 3). The RMSE values ranged from 0.90 to 1.34 Mg ha^{-1} for the simple (Fayette and Shelby) and multivariate (Henderson) analyses.

Veris recommends the use of both sensor EC and pH data (Veris-Technologies, 2010) for predicting lime recommendations. Since sensor $EC_{Shallow}$ and the interaction pH_{Sensor}*$EC_{shallow}$ were not significant factors in the model for the Henderson (Table 3), these factors would not have improved our predictions at this location. These results are in not in agreement with findings of Lund et al. (2004) and Christy et al. (2004) who found improvements in lime prediction accuracy when EC and near infrared reflectance were utilized. We do not know how EC would have impacted predictions at the Fayette and Shelby locations because EC was not collected at the calibration locations.

Table 3. Results of regression analyses for predicting lime recommendation (Mg ha^{-1}) as a function of Veris MSP sensor readings.

Parameters	Parameter Estimate	Pr>ltl	R^2	RMSE
	Fayette County (20 samples)			
Intercept	28.080	>0.001		
pH_{Sensor}	-3.921	>0.001		
			0.84	1.20
	Shelby Country (21 samples)			
Intercept	22.286	>0.001		
pH_{Sensor}	-2.879	>0.001		
			0.71	1.34
	Henderson Country (54 samples)			
Intercept	32.947	>0.001		
Sensor pH	-4.713	0.001		
$EC_{shallow}$	-4.434	0.073		
pH_{Sensor}*$EC_{shallow}$	0.676	0.078		
			0.48	0.90

RMSE = root mean squared error.

4.3.3. Whole field versus within field management

The calibration data set was used to compare whole field and within field lime-recommendations. The whole-field lime recommendations were greater for Fayette County field (4.3 Mg ha^{-1}) than for the Henderson (2.1 Mg ha^{-1}) and Shelby Country (3.1 Mg ha^{-1}) locations. If the farmer were to have used these field average recommendations, many regions of the fields would have received substantial under-applications of lime. For example, 50% of the LR values for individual observations in the calibration data set were between 5.1 and 9.5 Mg ha^{-1} for the Fayette county location. Approximately, 14% of the Henderson County observations required between 3.1 and 4.5 Mg ha^{-1}. For the Shelby County field, 29% of lime recommendations were between 4.3 and 11 Mg ha^{-1}. The data suggested that all three fields would have been good candidates for variable rate lime applications at the times of soil sampling.

4.3.4. Spatial structure

The values showed spatial dependency at all the study fields and isotropic exponential semivariogram models were fitted (Figure 2).

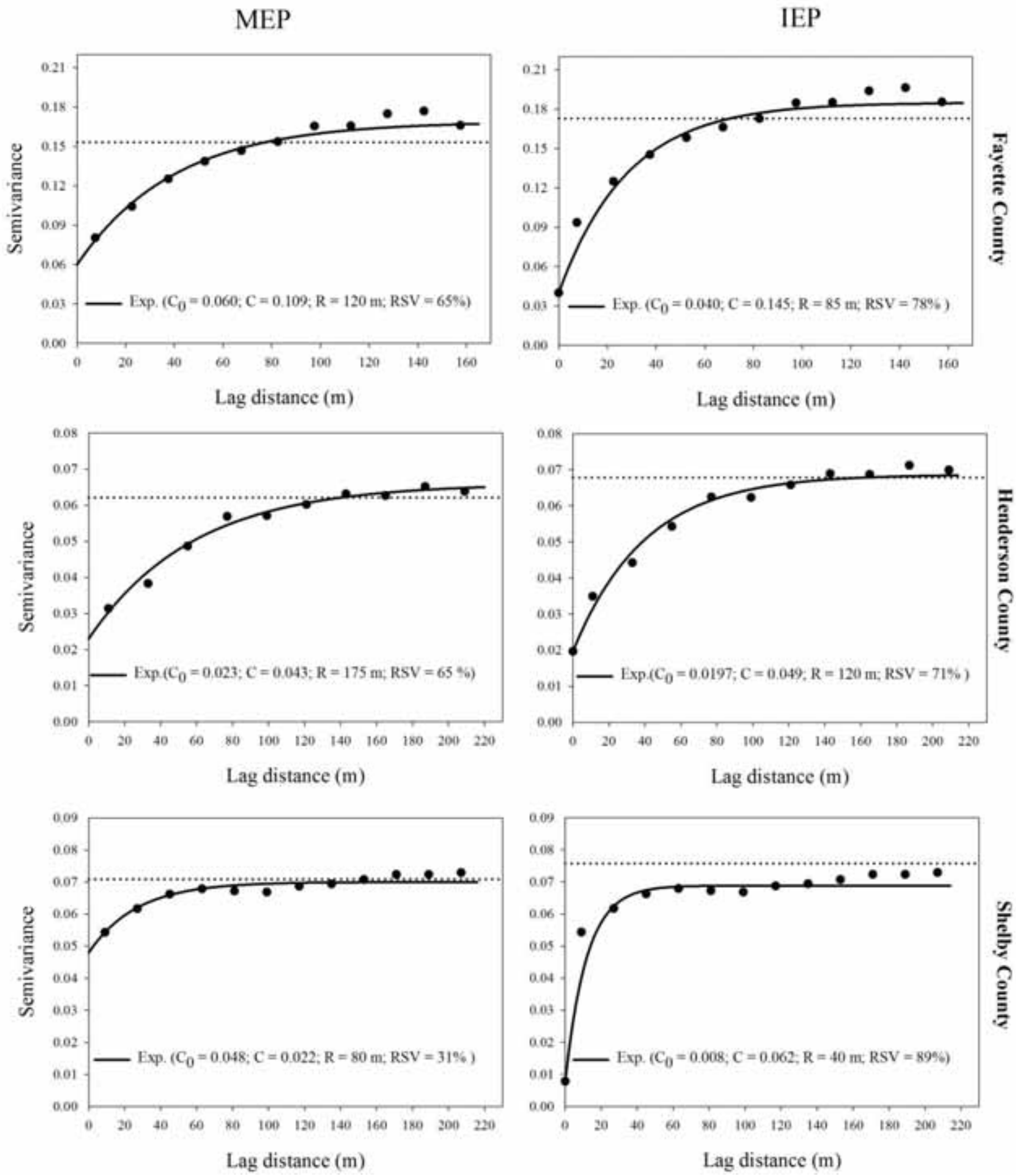

Figure 2. Semivariograms from mean electrode intensive sensor data set (MEP) and individual electrode intensive sensor data set (IEP) in three crop field in Kentucky.

Exp. = Exponential model; Sph. = Spherical model; C_0 = Nugget effect; C = partial sill; RSV = relative structural variability (partial sill) sill^{-1}.

The range of spatial structure was larger for the Henderson (175-m) than the Fayette (120-m) and Shelby (80-m) fields (Figure 2). All fields significant nugget effect (C_0) was found (i.e., $0.023 \leq C_0 \leq 0.060$; $31 \leq RSV \leq 65\%$) (Figure 2). Similar results were found for pH on three Kentucky crop fields (Mueller et al, 2004a) including the Calloway (i.e., $C_0 = 0.028$; RSV = 35%), Hardin (i.e., $C_0 = 0.032$; RSV = 83%), and Shelby (i.e., $C_0 = 0.020$; RSV = 79%) County which were based on 30.5-m grid soil sampling with additional samples collected on a random coarser grid.

The use of the IEV rather than MEP data sets reduced the nugget variance as expected. The Shelby field IEV data set had the greatest reduction of the nugget variance (Figure 2). This may have occurred either because 1) there was less small micro variability in pH at the Shelby county field or 2) the glass electrode was used instead of antimony electrode in this field and possibly the glass electrode produces values with a smaller measurement error. Regardless of the cause, the glass electrodes are much more fragile than the solid-state antimony electrode and thus may still be the better choice. More work is needed to better understand the cause.

4.3.5. Evaluation and comparison of prediction errors for interpolator methods

Bias, RMSE, R^2, and plots of predicted-versus-measured were similar for all interpolation methods at the three locations (Figure 3, 4 and 5). These results indicated that the choice of interpolation technique (i.e., Kriging, IDW, LPI, and RBF) did not substantially impact on map quality at the three locations consistent with the findings of others (Mueller, 2007; Robinson & Metternicht, 2006; Schloeder, et al., 2001). Given these findings, precision agriculture practitioners should not be greatly concerned about which interpolation procedure they use for mapping with the pH sensor. Kriging general exhibited slightly better results than the deterministic interpolation techniques, in agreement with the results found by Kravchenko (2003) for pH field in central Illinois.

Overall, map quality was less than expected or desired for all of the fields and interpolation procedures (Figure 3, 4 and 5). This may have occurred because nugget variances were fairly large (Figure 3). These large nugget variances were likely the result of management in these fields over the last 100 years that that inflated small-scale variability. However, these pH nugget variances were in-line with those of other studies (Mueller, et al., 2004b; Rodrigues, et al., 2012; Weindorf & Zhu, 2010).

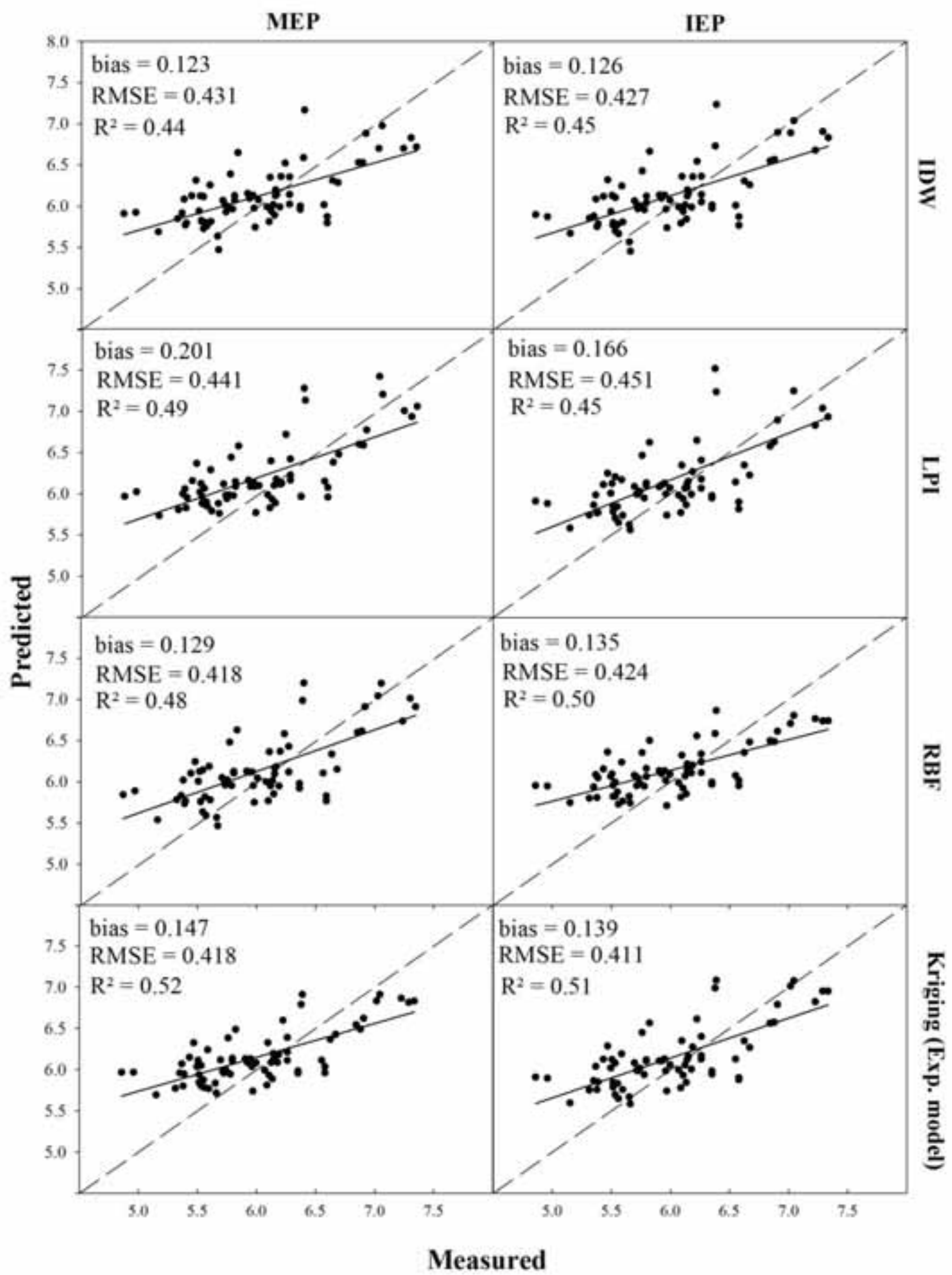

Figure 3. Predicted vs. measured for pH sensor values from mean electrode intensive sensor data set (MEP) and individual electrode intensive sensor data set (IEP) in a crop field in Fayette County-Kentucky.

RMSE = root mean squared error; IDW = Inverse distance weighted; LPI = Local polynomial interpolation; RBF = Radial basis functions; Exp. = Exponential model

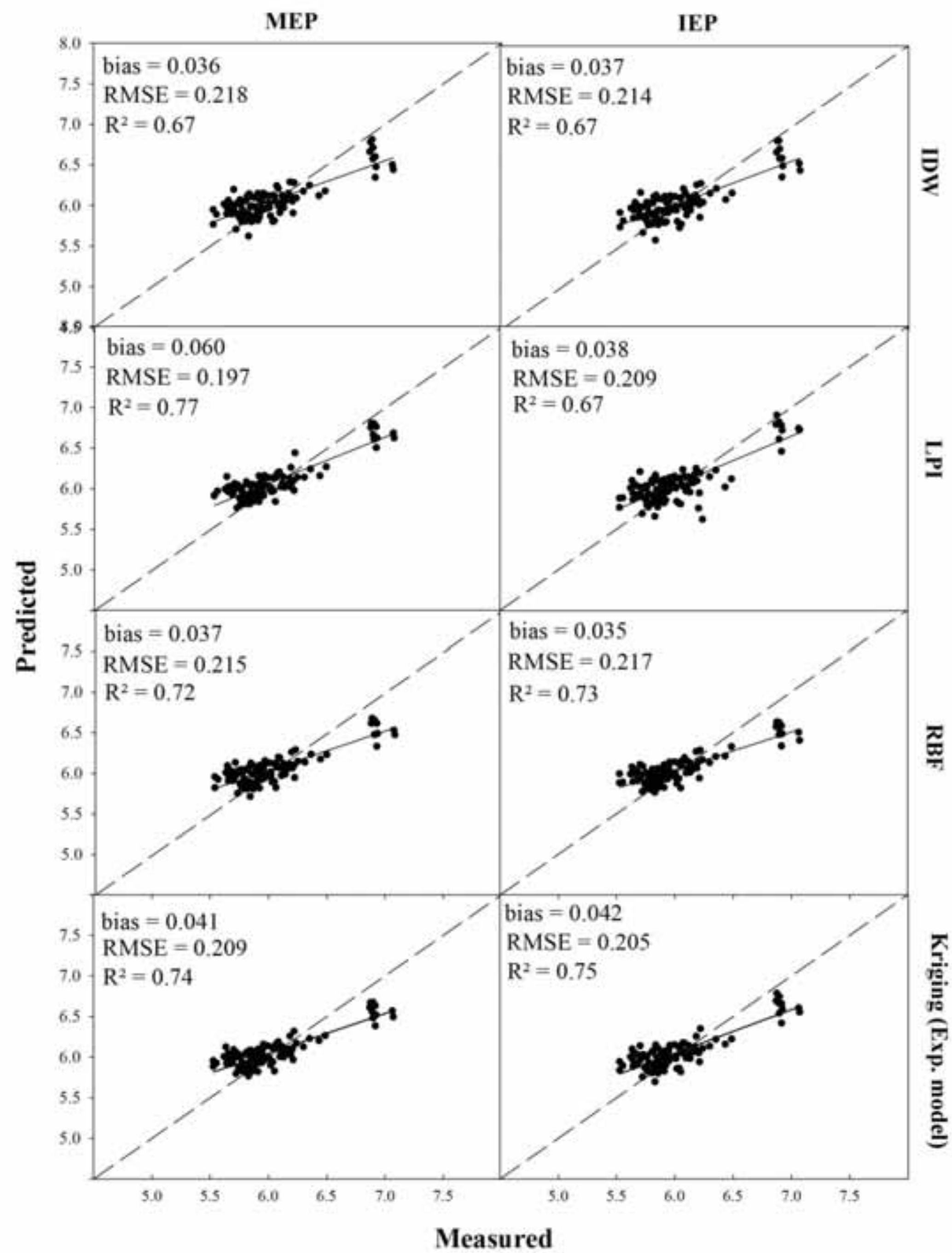

Figure 4. Predicted vs. measured for pH sensor values from mean electrode intensive sensor data set (MEP) and individual electrode intensive sensor data set (IEP) in a crop field in Henderson County-Kentucky.

RMSE = root mean squared error; IDW = Inverse distance weighted; LPI = Local polynomial interpolation; RBF = Radial basis functions; Exp. = Exponential model.

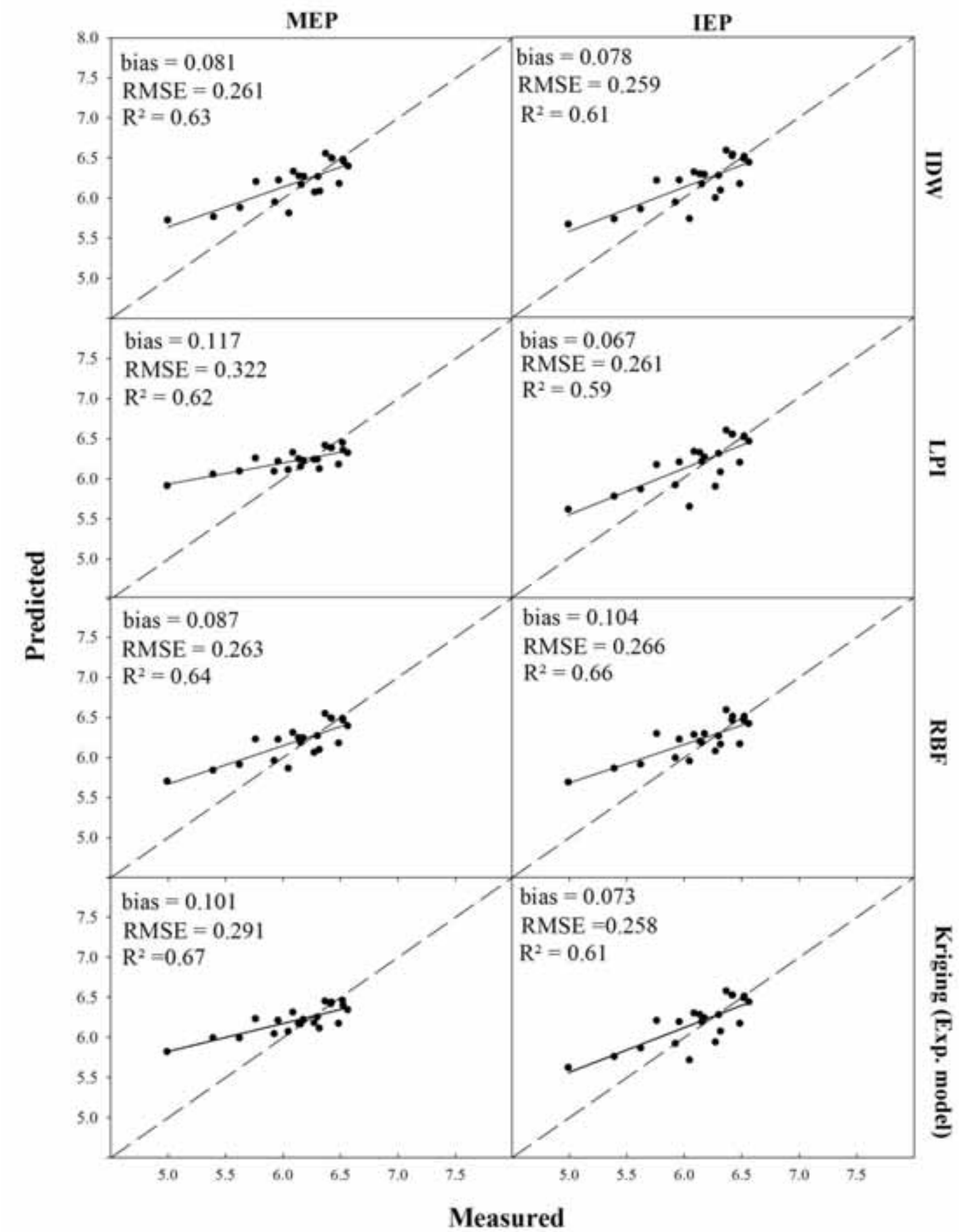

Figure 5. Predicted vs. measured for pH sensor values from mean electrode intensive sensor data set (MEP) and individual electrode intensive sensor data set (IEP) in a crop field in Shelby County-Kentucky.

RMSE = root mean squared error; IDW = Inverse distance weighted; LPI = Local polynomial interpolation; RBF = Radial basis functions; Exp. = Exponential model.

4.3.6. Using the individual electrode values in interpolation analyses to resolve small scale variability and improve prediction quality

The use of the IEV values rather than the mean electrode values did not substantially improve spatial predictions of soil pH (Figure 3, 4 and 5), even though the individual values rather than the mean electrode values did substantially improve the semivariograms (Figure 2). Clearly, improving the resolution of small-scale variability did not necessarily improve map quality.

4.3.7. New procedure for correcting interpolation errors

We considered whether it was possible to use the calibration data to adjust the lime recommendation in order to reduce interpolation errors. The impact of this transformation can be observed by comparing the first and second columns in Figure 6. The largest improvements were observed for Fayette and Shelby County and there was almost no improvement for the Henderson. The Henderson did not change because the calibration points were also belonged to the intensive sensor data set. This transformation analysis suggested that interpolation errors may be correctable in some situations if a calibration data set is collected independently from the intensive sensor data set.

Figure 7 demonstrates how the interpolated pH maps (first column) were used to calculate lime recommendations based on SMP buffer corrected for Kentucky fields (second column). The third column demonstrates how the calculated lime requirement maps would look after they were adjusted for interpolation errors. Clearly some large areas shift from one category to another if the interpolations are not adjusted in the Fayette and Shelby locations.

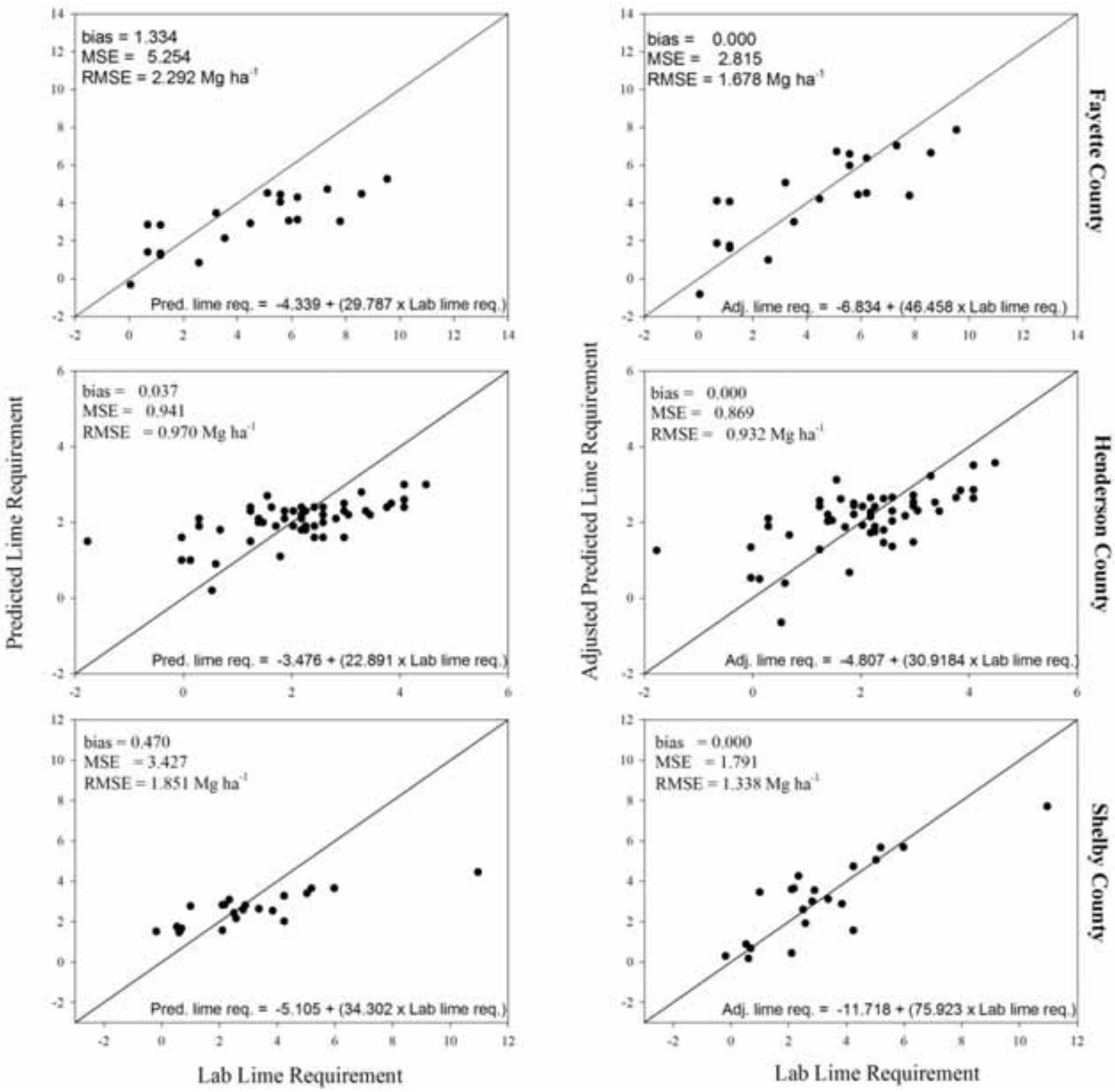

Figure 6. Predicted vs. Lab analysis for Lime requirement interpolated by Kriging method using a calibration data set as a coarse sensor data set and forcing the predicted and measured values to adhere to the 1:1 line.

These results demonstrate that the adjustment could improve lime requirement maps, and this could have a substantial impact on nutrient availability, herbicide efficacy, and crop yield. This approach should be studied more thoroughly on a variety of other fields before applied in commercial agricultural settings.

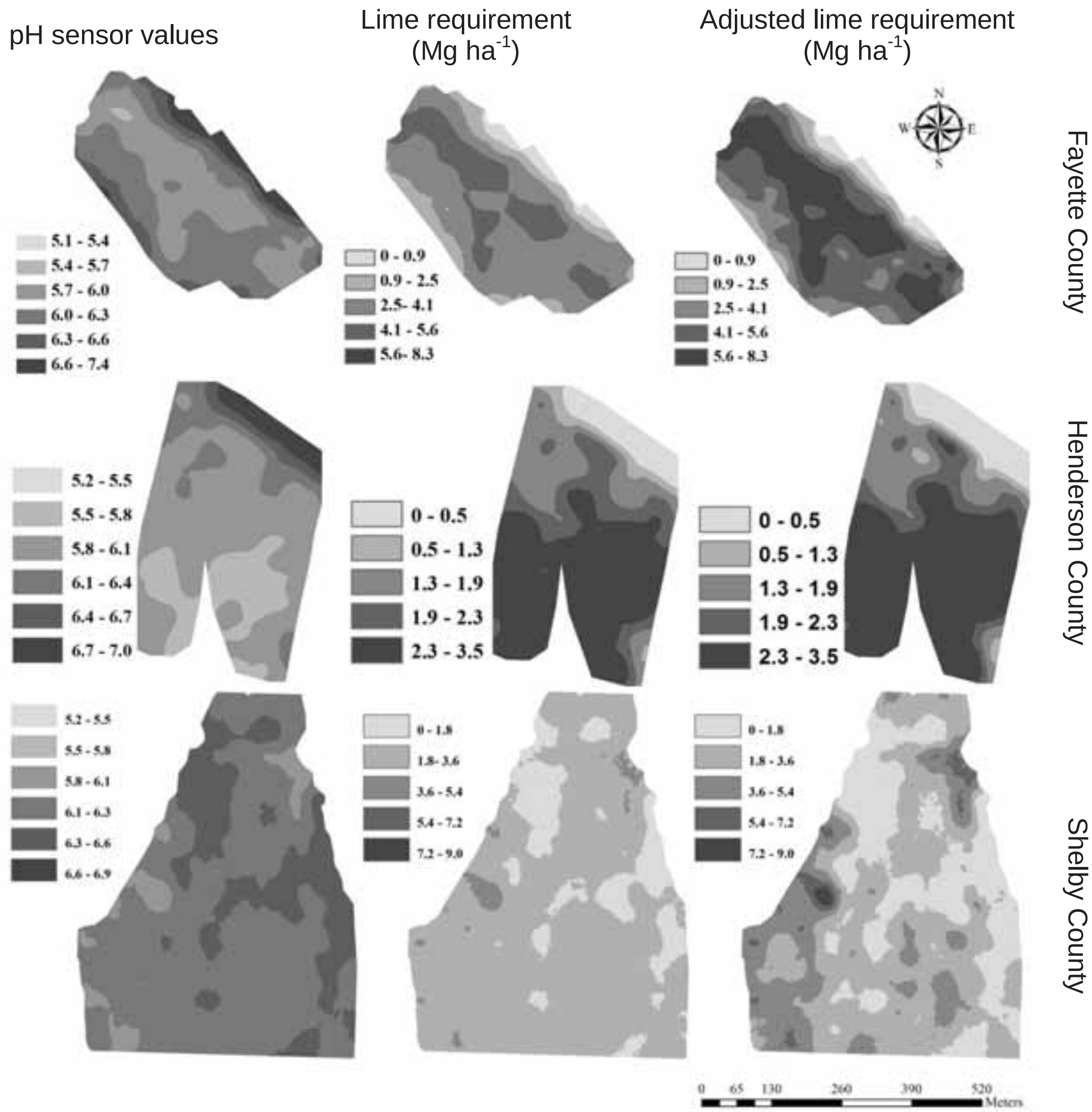

Figure 7. Maps of corrected pH sensor values of intensive sensor data set interpolated by Kriging method, lime requirement, and adjusted lime requirement base on forcing the predicted and measured values to adhere to the 1:1 line with the calibration data set.

4.4. Conclusions

This study indicated that the pH sensor measures can be useful for estimating both pH and Buffer pH in Kentucky soils. Calibration data set can be used to validate pH maps obtained from the sensor pH as external data set. Overall, the interpolation errors were substantial yet lower than 1-ha grid mapping and the choice of interpolation procedure (i.e., Kriging, IDW, LPI, and RBF) did not substantially impact map quality. This research suggests that while using the data from both electrodes separately can help improve estimates of the nugget variances, it will not substantially improve spatial predictions. We found that for the Fayette and Shelby data sets, the RMSE values could be reduced by applying a simple linear transformation to the sensor data. This transformation was obtained with an independently sampled calibration data set. This approach was not successful for the Henderson field because the calibration data set was not independently sampled.

4.5. References

Adamchuk, V., Lund, E., Reed, T., & Ferguson, R. (2007). Evaluation of an on-the-go technology for soil pH mapping. *Precision Agriculture, 8*(3), 139-149.

Adamchuk, V. I., Morgan, M. T., & Ess, D. R. (1999). An automated sampling system for measuring soil pH. *Transactions of the Asae, 42*(4), 885-891.

Christy, C. D., Collings, K., Drummond, P., & Lund, E. (2004). *A mobile sensor platform for measurement of soil pH and buffering.* Paper presented at the ASAE/CSAE Annual International Meeting, St. Joseph, Michigan, USA.

ERICKSON, B. (2004). Field experience validates on-the-go soil pH sensor. West Lafayette: Purdue Top Farmer Crop Workshop Newsletter, 4 p.

Havlin, J. L., Beaton, J. D., Tisdale, S. L., & Nelson, W. L. (2005). *Soil fertility and fertilizers: An introduction to nutrient management*. New Jersey: Pearson Prentice Hall, Upper Saddle.

Isaaks, E. H., & Srivastava, R. M. (1989). *An introduction to applied geostatistics*. New York: Oxford University Press.

Johnston, K., Hoef, J. M. V., Krivoruchko, K., & Lucas, N. (2001). *Using ArcGIS Geostatistical Analyst*. Redlands, CA: ESRI.

Kravchenko, A. N. (2003). Influence of spatial structure on accuracy of interpolation methods. *Soil Science Society of America Journal, 67*(5), 1564-1571.

Lund, E. D., Collings, K. L., Drummond, P. E., Christy, C. D., & Adamchuk, V. I. (2004). *Managing pH variability with on-the-go pH mapping.* Paper presented at the Conference on Precision Agriculture and other Precision Resources Management, Minneapolis, USA.

Mueller, T. G. (2007). Map quality assessment for site-specific fertility management *GIS Applications in Agriculture* (pp. 103-120). Boca Raton, FL: Taylor & Francis.

Mueller, T. G., Pierce, F. J., Schabenberger, O., & Warncke, D. D. (2001). Map quality for site-specific fertility management. *Soil Science Society of America Journal, 65*(5), 1547-1558.

Mueller, T. G., Pusuluri, N. B., Mathias, K. K., Cornelius, P. L., & Barnhisel, R. I. (2004a). Site-specific soil fertility management: A model for map quality. *Soil Science Society of America Journal, 68*(6), 2031-2041.

Mueller, T. G., Pusuluri, N. B., Mathias, K. K., Cornelius, P. L., Barnhisel, R. I., & Shearer, S. A. (2004b). Map quality for ordinary kriging and inverse distance weighted interpolation. *Soil Science Society of America Journal, 68*(6), 2042-2047.

Olfs, H. W., Borchert, A., & Trautz, D. (2010). *Validation of on-the-go soil pH measurements - Primary results from Germany.* Paper presented at the International Conference on Precision Agriculture, Denver.

Robinson, T. P., & Metternicht, G. (2006). Testing the performance of spatial interpolation techniques for mapping soil properties. *Computers and Electronics in Agriculture, 50*(2), 97-108.

Rodrigues, M. S., Corá, J. E., & Fernandes, C. (2012). Spatial relationships between soil attributes and corn yield in no-tillage system. *Revista Brasileira de Ciência do Solo, 36*(2), 599-609.

Schloeder, C. A., Zimmerman, N. E., & Jacobs, M. J. (2001). Comparison Of Methods For Interpolating Soil Properties Using Limited Data. *Soil Science Society of America Journal, 65*(2), 470-479.

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika, 52*(3-4), 591-611.

Shoemaker, H. E., McLean, E. O., & Pratt, P. F. (1961). Buffer Methods for Determining Lime Requirement of Soils With Appreciable Amounts of Extractable Aluminum1. *Soil Science Society of America Journal, 25*(4), 274-277.

Sikora, F. J. (2006). A Buffer that Mimics the SMP Buffer for Determining Lime Requirement of Soil. *Soil Science Society of America Journal, 70*(2), 474-486.

Silva, A. F., Soares de Souza Lima, J., & Batista de Oliveira, R. (2010). Métodos de interpolação para estimar o pH em solo sob dois manejos de café arábica. *Idesia (Arica), 28*, 61-66.

Staggenborg, S. A., Carignano, M., & Haag, L. (2007). Predicting Soil Ph And Buffer Ph In Situ With A Real-time Sensor. *Agronomy Journal, 99*(3), 854-861.

Veris-Technologies. (2010). Mobile Sensor Platform: pH Manager. In V. Techologies (Ed.). Salinas, KS.

Weindorf, D. C., & Zhu, Y. (2010). Spatial Variability of Soil Properties at Capulin Volcano, New Mexico, USA: Implications for Sampling Strategy. [Article]. *Pedosphere, 20*(2), 185-197.

Zandi, S., Ghobakhlou, A., & Sallis, P. (2011). *Evaluation of Spatial Interpolation Techniques for Mapping Soil pH*. Paper presented at the International Congress on Modelling and Simulation.

IMPLICAÇÕES

Tecnologias que possam maximizar os lucros do produtor agrícola, usar de forma racional e sustentável os diversos insumos e permitir uma agricultura sociol-ambiental correta são os objetivos da agricultura moderna. Para que tais objetivos sejam alcançados, o entendimento da variabilidade espaço-temporal dos campos de produção agrícola devem ser estudados. Portanto, este trabalho objetivou analisar atributos do solo e da produtividade das culturas visando o manejo específico de áreas agrícolas baseado em três tecnologias de agricultura de precisão.

Para um efetivo manejo localizado é requerido um entendimento da influencia dos atributos do solo e dos fatores climáticos no padrão de variabilidade da produtividade das culturas. Adicionalmente, é necessário avaliar técnicas estatísticas que determinam as relações de causa/efeito entre atributos do solo e produtividade das culturas.

O objetivo do primeiro estudo foi avaliar se os modelos estatísticos, que levam em consideração a heteroscedasticidade e a autocorrelação espaço-temporal são superiores ao método dos mínimos quadrados (OLS) para avaliar a relação entre a produtividade de milho e os atributos do solo. A área de estudo (10 por 250-m) está localizada em Jaboticabal-SP. A produtividade de milho (espaçamento entre linhas de 0,9-m) foi medida em 100 células amostrais de 4,5x10-m, dispostas em quatro transectas (25 pontos por transecta), em seis anos agrícolas, no período de 2001 a 2010. Atributos químicos e físicos do solo foram obtidos em cada ponto amostral. Para estimar a produtividade de milho foram utilizados quatro modelos: mínimos quadrados ordinários (OLS); mínimos quadrados generalizados assumindo heteroscedasticidade (GLS_{he}); modelo espacial e temporal, assumindo homocedasticidade (GLS_{sp}); e modelo espacial e temporal, assumindo heteroscedasticidade ($GLS_{he\text{-}sp}$).

A acidez do solo (pH) foi o atributo que mais influenciou a produtividade de milho ao longo do tempo. Verificou-se que com o modelo OLS, a estimativa da produtividade de milho aumentou 0,59 Mg ha^{-1} por unidade de pH, enquanto que com o $GLS_{he\text{-}sp}$ o aumento foi de 0,43 Mg ha^{-1}, indicando que a escolha do modelo afeta a predição e os parâmetros da regressão, sendo este fato crítico, pois, modifica a decisão de manejo do solo em relação à produtividade. Verificou-se que o modelo espacial e temporal, assumindo heteroscedasticidade foi superior aos demais. Os resultados demonstram que as relações de causa e efeito entre produtividade da cultura e atributos do solo devem ser estudas utilizando-se dados de longos períodos de tempo.

Uma zona de manejo em um campo agrícola é a expressão de uma sub-região deste campo que é relativamente homogênea, formado pela combinação de dados de produtividade das culturas e seus fatores limitantes.

O agrupamento de dados de solos e planta podem ser utilizados para definir zonas de manejo, isto porque, os dados são agrupados dentro de uma classe chamada de 'cluster' que é baseado na interação similar das variáveis. O objetivo do segundo estudo foi identificar zonas de manejo usando o algoritmo de grupamento 'fuzzy c-means' baseado na variabilidade espacial e temporal dos atributos do solo e da produtividade de milho.

A área de estudo é a mesma descrita para o primeiro estudo e os dados de cinco anos agrícolas de produtividade e atributos químicos e físicos no período entre 2001 a 2010 foram utilizados para as análises. O modelo espacial e temporal que leva em consideração a heteroscedasticidade da variância foi utilizado para identificar qual(ais) variáveis mais influenciaram a variabilidade espacial da produtividade de milho ao longo dos cinco anos de estudo.

A saturação por bases (V) foi a variável que melhor se relacionou com a produtividade de milho, portanto, modelos de semivariogramas foram ajustados aos dados de V e produtividade de milho e posteriormente interpolados pelo método da krigagem. O programa de computador 'Management Zone Analyst' (MZA) foi utilizado para realizar o agrupamento com o algoritmo 'fuzzy c-means'.

Observou-se dependência espacial para produtividade de milho e V para todos os anos estudados. O número ótimo de zonas de manejo alteraram-se ao longo do tempo, portanto, a escolha do número de zonas de manejo devem ser realizados baseados em dados temporais. O grau de similaridade espacial entre as zonas de manejo de V e da produtividade de milho podem mudar ao longo do tempo. Portanto, é de fundamental importância levar em consideração a variabilidade temporal da produtividade das culturas e dos atributos do solo para obter zonas de manejo com maior acurácia. O algoritmo de agrupamento 'fuzzy c-means' baseado na variabilidade espacial e temporal dos atributos do solo e da produtividade das culturas foi eficiente para definir zonas homogêneas de manejo.

O número de amostras de solo é um fator determinante na qualidade dos mapas de atributos do solo. Contudo, número elevado de amostras é um prática onerosa e muitas vezes inviável. Portanto, o desenvolvimento de sensores que realizam a amostragem de atributos do solo com relativo baixo são ferramentas que possuem potencial para resolver o problema do número de amostras.

O sensor Veris MSP® de atributos do solo em tempo real permite avaliar de forma rápida o pH do solo, a condutividade elétrica e opcionalmente a refletância no espectro do visível e do infravermelho. Porém, para este sensor ser utilizado no manejo de calagem em áreas específicas de forma eficiente, os mapas criados com este instrumento devem ter qualidade adequada.

O objetivo do terceiro estudo foi avaliar se o sensor de pH do solo em tempo real Veris MSP® possui acurácia para estimar o pH do solo e associar suas predições com o pH tampão, também chamado de 'Buffer pH' e propor um novo método para correção dos erros de interpolação para a produção de mapas de necessidade de calagem.

Amostras de solo foram coletadas e enviadas ao laboratório para relacionar as medidas em laboratório de pH do solo e 'Buffer pH' com as medidas obtidas com o sensor de pH e este banco de dados foi, então chamado de banco de dados de calibração. Este estudo foi conduzido em três áreas agrícolas em Kentucky, EUA, onde foram coletadas em cada campo com o sensor dois tipos de banco de dados. Utilizando-se da distância entre passadas de 12 m, obteve os bancos de dados chamados de "banco de dados intensivo" e em passadas de 36 m aproximadamente ortogonal a direção das passadas de 12 m, os banco de dados chamados de banco de dados grosseiros.

Os dados foram ajustados aos semivariogramas e interpolados (Kriging, Inverso do Quadrado da Distância, Interpolação Polinomial e Funções de Base Radial) utilizando o banco de dados com a média dos dois eletrodos ('Mean Electrode Prediction' - MEP) e com os valores individuais de cada eletrodo

(Individual eletrode values' - IEV). Uma regressão simples foi utilizada para calibrar os valores obtidos com o eletrodo baseado nas medidas de laboratório.

As medidas obtidas com o sensor de pH apresentaram correlação com as medidas de pH em laboratório (0,52 ≤ R^2 ≤0,87) e com o 'Buffer pH' (0,44≤ R^2 ≤0,84). O alcance da estrutura espacial foi maior para o campo agrícola em Henderson (175 m) do que para o campo agrícola em Fayette (120 m) e Shelby (80 m). Os semivariogramas dos três campos agrícolas estudados apresentaram efeito pepita relativo variando entre 35 e 69%, que possivelmente contribui para o maior erro de interpolação para o pH do solo do que o desejado (0,20 ≤ Erro médio quadrático ≤ 0,44).

O uso do banco de dados IEV reduziu o efeito pepita em relação ao banco de dados MEP, contudo não foi observada melhoras na predição espacial dos valores de pH do solo. Os erros espaciais de interpolação foram maiores do que os desejados e podem substancialmente causar erros na aplicação do calcário. Porém, tais erros podem ser reduzidos utilizando o banco de dados de calibração.