**Universidade de São Paulo
Escola Superior de Agricultura “Luiz de Queiroz”**

**Modelos de regressão e curvas de produção vegetal: aplicações nas áreas
de nutrição mineral de plantas e adubação**

Iuri Emmanuel de Paula Ferreira

Tese apresentada para a obtenção do título de Doutor em Ciências. Área de concentração: Estatística e Experimentação Agronômica

**Piracicaba
2015**

Iuri Emmanuel de Paula Ferreira
Físico Médico

Modelos de regressão e curvas de produção vegetal: aplicações nas áreas de nutrição mineral de plantas e adubação

versão revisada de acordo com a resolução CoPGr 6018 de 2011

Orientador:
Prof. Dr. **SILVIO SANDOVAL ZOCCHI**

Tese apresentada para a obtenção do título de Doutor em Ciências. Área de concentração: Estatística e Experimentação Agronômica

**Piracicaba
2015**

**Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA - DIBD/ESALQ/USP**

Ferreira, Iuri Emmanuel de Paula

Modelos de regressão e curvas de produção vegetal: aplicações nas áreas de nutrição mineral de plantas e adubação / Iuri Emmanuel de Paula Ferreira. - - versão revisada de acordo com a resolução CoPGr 6018 de 2011. - - Piracicaba, 2015.

133 p. : il.

Tese (Doutorado) - - Escola Superior de Agricultura "Luiz de Queiroz".

1. Níveis críticos 2. Modelos segmentados 3. Funções de produção 4. Retornos decrescentes I. Título

CDD 631.80182
F383m

"Permitida a cópia total ou parcial deste documento, desde que citada a fonte – O autor"

Aos meus pais,

A minha eterna namorada e futura esposa, Ana Carolina,

À Maria Inês (in memoriam), minha avó.

AGRADECIMENTOS

Em primeiro lugar, gostaria de agradecer a Deus. Gostaria de agradecer também a minha família, por todo suporte que tem me dado. Gostaria de agradecer a minha noiva, Ana Carolina, por cuidar de mim com amor e dedicação e, também, por escolher trilhar esse caminho ao meu lado.

Gostaria de agradecer ao meu amigo e orientador, o Prof. Silvio Sandoval Zocchi, pelas nossas discussões durante as reuniões, pela motivação em compreender o problema e detalhar possíveis soluções, por indicar novos caminhos a serem seguidos e corrigir meus erros. Sou grato também pelos seus conselhos sobre abordagem didática, que foram de grande valia durante a minha experiência como docente.

Além disso, gostaria de agradecer às Professoras Luzia Trinca e Claudia Pio Ferreira, da Unesp de Botucatu, que têm sido fundamentais para a minha formação acadêmica, desde os tempos de Iniciação Científica.

Gostaria de dizer, também, que sou imensamente grato às pessoas do Departamento de Ciências Exatas (LCE-ESALQ), que me receberam de braços abertos e foram excelentes companhias durante esse período de 4 anos. Agradeço à Prof.^a Clarice, minha chefe, por todo auxílio durante as disciplinas e no período em que estive atuando como professor substituto. Agradeço à Prof.^a Roseli pela recepção cordial, pela parceria durante o estágio em matemática aplicada e pelos exemplos trocados sobre as aplicações das equações diferenciais. Gostaria de agradecer aos demais professores (Idemauro, Sônia, Taciana, Edwin, César, Renata, Cristian, Gabriel, Tadeu, Rafael) do LCE pelas disciplinas ministradas, pelos ensinamentos e pela parceria no trabalho. Sou muito grato, também, aos funcionários do departamento (Solange, Luciane, Jorge, Rosni, Mayara e Eduardo), os quais sempre estiveram prontamente disponíveis para ajudar no que fosse preciso.

Por fim, gostaria de agradecer aos meus companheiros de curso, da minha turma de Doutorado e, também, da turma de discussões avançadas sobre Epistemologia que se formou na copinha: Pedro Henrique, Rodrigo, Rafael Maia, Rafael Moral, Djar, Edilan, Ricardo, Simone, Maria, César, Thiago Gentil, Thiago Viana, Thiago Oliveira, Ana Júlia, José Nilton, Davi, Marcus, Lucas, Izabela, Everton, Elisângela, Nakamura, Reginaldo, dentre outros.

O presente trabalho foi realizado com apoio do CNPq, Conselho Nacional de Desenvolvimento Científico e Tecnológico - Brasil.

``Howard Wolowitz: [after everyone cheers for him and his team design going to space] It gets better! Someone has to go up with the telescope as a payload specialist, and guess who that someone is!

Sheldon Cooper: Mohammed Lee.
[everyone's looking confused]

Howard Wolowitz: Who's Mohammed Lee?

Sheldon Cooper: Mohammed is the most common first name in the world, and Lee the most common surname. As I didn't know the answer, I thought that gave me a mathematical edge.''

Chuck Lorry, in ``The Big Bang Theory".

SUMÁRIO

RESUMO.....	11
ABSTRACT.....	13
1 INTRODUÇÃO	15
Referências.....	17
2 REGRESSÃO SEGMENTADA COM TRANSIÇÕES DE FASE SUAVES SOB A PRESENÇA DE LIMIARES ALEATÓRIOS	21
Resumo	21
Abstract.....	21
2.1 Introdução	22
2.2 Revisão Bibliográfica	23
2.2.1 A Formulação Sinal: Modelos Transicionais e Hiperbólicos	25
2.2.2 A Formulação Máx-Mín	30
2.3 Metodologia	35
2.3.1 A Regressão Bent-Cable Estendida	35
2.3.2 A Inclusão de Limiares Aleatórios e Novas Funções de Junção Suave.....	38
2.3.3 Estimação por Máxima Verossimilhança.....	43
2.3.4 Simulação de Dados.....	49
2.4 Resultados e Discussão	52
2.4.1 Resultados da Simulação de Dados	53
2.4.2 Resultados Teóricos	58
2.5 Conclusão.....	63
Agradecimentos	63
Referências.....	63
3 CURVA DA PRODUÇÃO VEGETAL MEDIANTE ESTADOS NUTRICIONAIS DIS- TINTOS	67
Resumo	67
Abstract.....	67
3.1 Introdução	67
3.2 Revisão Bibliográfica	69
3.3 Material	77
3.4 Métodos.....	78
3.4.1 Modelos de Regressão.....	78
3.4.2 Estimação por Máxima Verossimilhança e Testes de Hipóteses.....	81
3.4.3 Quantidades de Interesse Agrícola.....	85
3.5 Resultados e Discussão	87
3.6 Conclusão	96
Agradecimentos	96
Referências.....	97
4 AS LEIS DE LIEBIG E MITSCHERLICH E A FRAÇÃO DE APROVEITAMENTO DO NUTRIENTE LIMITANTE NA PLANTA.....	101
Resumo	101
Abstract.....	101
4.1 Introdução	101
4.2 Metodologia	103

4.2.1 Função de Produção Linear.....	107
4.2.2 Função de Produção Linear-platô	107
4.2.3 Função de Produção Quadrática-Platô.....	108
4.2.4 Função de Produção de Mitscherlich	109
4.2.5 Novas Funções de Produção	111
4.3 Uma Aplicação dos Modelos Propostos	119
4.4 Conclusão.....	122
Agradecimentos	122
Referências.....	122
ANEXOS	125

RESUMO

Modelos de regressão e curvas de produção vegetal: aplicações nas áreas de nutrição mineral de plantas e adubação

A avaliação do estado nutricional das culturas requer o uso de padrões para comparação, tais como os níveis críticos e as faixas de suficiência. Esses padrões de interesse agrícola geralmente são estimados a partir da curva de calibração da produção vegetal, que relaciona os rendimentos de uma cultura com a dose do adubo aplicada no solo ou com o teor de algum elemento nos tecidos. A curva de produção vegetal é fundamental para o desenvolvimento de pesquisas nas áreas de nutrição mineral de plantas e adubação, uma vez que o conhecimento a seu respeito possibilita o uso econômico de adubos e a elaboração de estratégias corretivas para o solo. No entanto, a curva de produção divide-se em diferentes regimes nutricionais, que lhe conferem uma forma peculiar e dificultam sua descrição matemática. O objetivo desse trabalho é a construção de modelos de regressão flexíveis, capazes de descrever adequadamente a curva de produção vegetal. Em uma primeira abordagem, modelos lineares segmentados com transições de fase suaves são usados para descrever a produção do trigo de inverno em função da adubação com nitrogênio. Por meio dos ajustes destes modelos, também são estimados os níveis críticos e as faixas de suficiência. Posteriormente, realiza-se a modelagem matemática da resposta vegetativa à adubação, desconsiderando doses tóxicas. Observa-se que cada tipo possível de resposta vegetativa é ligado de forma unívoca a uma função de produção agrícola. A técnica proposta, além de permitir a construção de novas funções de produção, fornece interpretações fisiológicas claras para parâmetros de alguns modelos de uso comum na agricultura. Com base na segunda abordagem, um novo enunciado para a lei de Mitscherlich é apresentado: quanto maior é o conteúdo total do nutriente limitante na planta, menor é o aproveitamento feito (a fim de lhe prover crescimento) de um incremento em sua disponibilidade – como consequência, os retornos da adubação são decrescentes.

Palavras-chave: Níveis críticos; Modelos segmentados; Funções de produção; Retornos decrescentes

ABSTRACT

Regression models and yield curves: with applications in plant mineral nutrition and fertilization

The evaluation of the plant nutritional status requires standards for comparisons, such as the critical levels and sufficiency ranges. These standards are generally estimated by means of the yield curves in response to fertilizers, which relates production with doses applied or nutrient content in the plant tissues. The yield curve is essential for the development of researches on plant mineral nutrition and fertilization, since knowledge of this curve allows economic use of fertilizers and planning strategies for soil correction. However, the production curve is divided into different nutritional regimes, which give it a peculiar shape and hinders its mathematical description. In this work, the objective is to build flexible regression models, able to adequately describe the yield curves. In a first approach, linear piecewise models with smooth transitions between regimes are used to describe the winter-wheat yield in function of the nitrogen doses. By means of the fits of these models, the critical levels and sufficiency ranges are also estimated. In another approach, it is proposed a mathematical model for yield response (or returns) from fertilizers – toxic doses are neglected in this phase. It is verified that each possible yield response type is uniquely related with a production function. This approach allows to construct new production functions and, indeed, provides clear physiological interpretations for the parameters of some common models in agriculture. Based on the second approach, it is presented a new statement for the Mitscherlich law: the greater is the total content of the limiting nutrient in the plant, the lower is the exploited fraction (in order to provide growth) of an increment in its availability – as a consequence, the returns from fertilization are diminishing.

Keywords: Critical levels; Linear piecewise functions; Production functions; Diminishing returns

1 INTRODUÇÃO

Nos ensaios de adubação, o pesquisador provoca deliberadamente uma deficiência mineral e, depois, avalia como aumentos gradativos nos níveis de fertilização afetam o rendimento das culturas. O objetivo destes experimentos é a obtenção da curva de calibração da produção vegetal (HAVLIN et al., 2004; MAIA et al., 2001; MILES, 2010), que relaciona a produção da cultura com a dose de adubo aplicada no solo ou com o conteúdo (ou teor) do nutriente na planta.

A curva de produção vegetal é essencial para os estudos de fertilização e nutrição mineral de plantas. Por meio de sua forma é possível inferir o estado nutricional das lavouras e traçar estratégias para a correção do solo, assim favorecendo a obtenção de colheitas produtivas e o uso racional de adubos (MALAVOLTA, 1967; MALAVOLTA et al., 1989). A importância da curva de produção vegetal está atrelada ao fato de que os padrões usados na recomendação de doses de adubo, tais como os níveis críticos e as faixas de suficiência, são quantidades que derivam-se dela (FAQUIN, 2002; MAIA et al., 2001; NEVES et al., 2008; WADT et al., 2008).

A despeito da sua importância prática, a descrição matemática da curva de produção vegetal não é uma tarefa fácil. O problema concentra-se no fato de que a curva divide-se em diferentes partes, que correspondem aos regimes nutricionais distintos, ou seja, os regimes de deficiência mineral, consumo de luxo e toxidez (FAQUIN, 2002; HAVLIN et al., 2004; JONES JR, 1993; MACY, 1936; OHKI, 1984; SMITH, 1962; TAIZ; ZEIGER, 2004).

Durante o regime de deficiência mineral, a relação entre produção vegetal e dose de adubo é determinada pela lei do mínimo de Liebig. Conforme Justus von Liebig, a produção vegetal é controlada pelo elemento mineral essencial que encontra-se em quantidade relativa mínima na planta (LIEBIG, 1840; PARIS, 1992a). Dessa maneira, aumentos na produção são possíveis apenas quando eleva-se a disponibilidade desse mineral limitante. Entretanto, a planta não responde indefinidamente aos incrementos na dose de adubo. De acordo com a lei do mínimo de Liebig, doses de adubo crescentes levam a planta a sua suficiência pelo mineral e, em certo ponto, outro fator de produção passa a ocupar a posição de mínimo relativo (MALAVOLTA, 1980; MITSCHERLICH, 1909; PARIS, 1992a). A partir desse nível crítico, a planta ingressa no regime nutricional do consumo de luxo.

No regime do consumo de luxo, a planta não responde a incrementos na dose do adubo. Nesta fase, o excesso do mineral aplicado ao solo pode até ser absorvido pela planta, elevando o teor do elemento nos tecidos ou sendo armazenado em órgãos de reserva, mas não reflete na produção vegetal (STERNER; ELSER, 2002). O regime de consumo de luxo requer que a curva de produção apresente um platô (ou uma assíntota), caracterizando uma faixa de valores de dose (ou teores) do elemento que não ocasionam mudanças significativas na produção. Quando as

culturas encontram-se dentro dessa faixa, dizemos que as plantas estão no estado de nutrição adequado.

O regime do consumo de luxo se estende até o ponto em que os excessos do nutriente são capazes de intoxicar a planta ou causar desequilíbrios nutricionais, o que leva a decréscimos na produção (HAVLIN et al., 2004). A partir desse ponto, a planta ingressa no regime nutricional de toxidez. Há um número crescente de situações em que é preciso identificar níveis tóxicos dos elementos minerais, mas pouca informação têm sido obtida numa escala completa de doses, partindo-se da deficiência mineral à toxidez (JONES JR, 1993). Talvez isso ocorra devido à dificuldade de se modelar a curva de produção vegetal ao longo dos três regimes nutricionais existentes. Na bibliografia consultada, verifica-se que apenas dois modelos são usados quando deseja-se incluir o regime de toxidez, o modelo polinomial quadrático e o modelo de 2ª ordem de Mitscherlich (MITSCHERLICH, 1928; PIMENTEL-GOMES; NOGUEIRA, 1980; SCHABENBERGER; PIERCE, 2001).

Acredita-se que estes modelos não são descrições matemáticas fiéis dos processos que governam a resposta à adubação, uma vez que seus gráficos apresentam um ponto de máximo, ao invés do platô característico do consumo de luxo. Além disso, regimes nutricionais diferentes implicam a existência de mudanças estruturais na relação dose-resposta, as quais impossibilitam o uso de funções simples para descrever os conjuntos de dados.

Uma alternativa para descrever a curva de produção vegetal é dividir o espaço de regressão em partes, que correspondem aos diferentes regimes nutricionais. Então, torna-se possível modelar a produção dentro de cada regime, utilizando-se de funções bem simples, como polinômios de 1º e 2º graus. Assim, uma aproximação para o comportamento complexo da curva de produção vegetal é obtida pela união desses submodelos. Essa é a ideia que fundamenta a regressão segmentada (FEDER, 1975; HINKLEY, 1969, 1971; McGEE; CARLETON, 1970; QUANDT, 1958; SPRENT, 1961).

Outra alternativa é ignorar o regime que seria, talvez, o menos importante: o regime de toxidez. Assim, muitos modelos de regressão conhecidos tornam-se aplicáveis ao problema, tais como os modelos linear-platô, quadrático-platô e de Mitscherlich (CERRATO; BLACKMER, 1990; MITSCHERLICH, 1909; SCHABENBERGER; PIERCE, 2001). Em geral, pode-se afirmar que estes modelos são funções com gráficos côncavos, monótonos não-decrescentes e limitados superiormente por uma assíntota horizontal ou um platô. Estas funções de produção são conhecidas por funções de von Liebig, uma vez que seus comportamentos assintóticos são determinados pela lei do mínimo de Liebig (BERCK; HELFAND, 1990; PARIS, 1992a,b).

Neste trabalho estas duas alternativas são exploradas. O objetivo geral é a construção de modelos de regressão flexíveis, que descrevam bem as curvas de produção vegetal e a partir

dos quais seja possível obter os padrões de interesse agrícola requeridos nas áreas de nutrição mineral de plantas e adubação. Esse trabalho divide-se em três artigos, desenvolvidos para atingir o objetivo geral, mas que abordam partes diferentes do mesmo problema. Uma breve descrição de cada um dos artigos que compõem esse trabalho é apresentada a seguir.

No primeiro artigo, disposto no capítulo 2, realiza-se um amplo estudo de revisão sobre os modelos de regressão segmentada com transições suaves entre regimes (BACON; WATTS, 1971; GRIFFITHS; MILLER, 1973, 1975; SEBER; WILD, 1989; TISHLER; ZANG, 1981a,b; WATTS; BACON, 1974). Além disso, são introduzidas novas alternativas para a formulação destes modelos, baseadas na existência de limiares aleatórios que dividem os domínios de atuação de cada regime. O primeiro artigo fornece a base teórica necessária para o entendimento do segundo artigo.

No segundo artigo, disposto no capítulo 3, o objetivo é modelar a curva de produção vegetal mediante os três regimes nutricionais existentes. Para atingir este objetivo, propõe-se o uso de modelos lineares segmentados com transições suaves entre regimes, previamente abordados no capítulo 2. A metodologia apresentada nesse artigo permite a descrição da curva de produção vegetal considerando-se uma escala de doses completa, que vai da deficiência mineral até a toxidez. Além disso, por meio dos ajustes dos modelos estudados, torna-se possível a estimação dos níveis críticos (de deficiência e toxidez) e das faixas de suficiência, que são padrões essenciais para a determinação do estado nutricional das culturas. Como motivação para a prática, os modelos estudados nesse artigo são ajustados aos dados de Schneeberger (2009), sobre a produção do trigo de inverno em função da adubação com nitrogênio.

No terceiro artigo, disposto no capítulo 4, o regime de toxidez é negligenciado e funções de produção vegetal mais simples são estudadas. Neste artigo, a resposta vegetativa das plantas a incrementos na dose de adubo é modelada por meio da lei de Liebig, mas considerando-se uma distribuição heterogênea do elemento limitante (e das exigências nutricionais) dentro da planta. Verifica-se que há uma relação unívoca entre o tipo de resposta vegetativa à adubação e a função de produção vegetal. Como exemplos, diversas funções de produção de von Liebig são obtidas como casos particulares da metodologia proposta, para tipos de resposta vegetativa específicos (BERCK; HELFAND, 1990; PARIS, 1992a,b). A partir da metodologia introduzida nesse artigo, também são propostas novas funções de produção de von Liebig, úteis para descrever dados oriundos de experimentos de adubação.

Referências

BACON, D.W.; WATTS, D.G. Estimating the transition between two intersecting straight lines. **Biometrika**, Oxford, v. 58, n. 3, p. 525-534, 1971.

- BERCK, P.; HELFAND, G. Reconciling the von Liebig and differentiable crop production functions. **American Journal of Agricultural Economics**, Oxford, v. 72, n. 4, p. 985-996, 1990.
- CERRATO, M.E.; BLACKMER, A.M. Comparison of models for describing corn yield response to nitrogen fertilizer. **Agronomy Journal**, Madison, v. 82, p. 138-143, 1990.
- FAQUIN, V. **Diagnose do estado nutricional de plantas**. Lavras: UFLA e FAEPE, 2002. 77p.
- FEDER, P.I. On asymptotic distribution theory in segmented regression problems - identified case. **The Annals of Statistics**, Philadelphia, v. 3, n. 1, p. 49-83, 1975.
- HAVLIN, J.L.; BEATON, J.D.; TISDALE, S.L.; NELSON, W.L. **Soil fertility and fertilizers: an introduction to nutrient management**. 7.ed. New Jersey: Pearson, 2005. 515 p.
- GRIFFITHS, D.A.; MILLER, A.J. Hyperbolic regression - a model based on two-phase piecewise linear regression with a smooth transition between regimens. **Communications in Statistics**, Abingdon, v. 2, n. 6, p. 561-569, 1973.
- GRIFFITHS, D.; MILLER, A.; DRAPER, N.R.; BOX, M.J.; JONES, E.R. Letters to the editor. **Technometrics**, Alexandria, v. 17, n. 2, p. 281-282, 1975.
- HINKLEY, D.V. Inference about the intersection in two-phase regression. **Biometrika**, Oxford, v. 56, n. 3, p. 495-504, 1969.
- HINKLEY, D.V. Inference in two phase regression. **Journal of the American Statistical Association**, Alexandria, v. 66, n. 336, p. 736-743, 1971.
- JONES JR, J.B. Modern interpretation systems for soil and plant analyses in the United States of America. **Australian Journal of Experimental Agriculture**, Clayton South, v. 33, n. 8, p. 1039-1043, 1993.
- LIEBIG, J. von. **Organic chemistry in its applications to agriculture and physiology**. London: Taylor and Walton, 1840. 387 p.
- MACY, P. The quantitative mineral requirements of plants. **Plant Physiology**, Bethesda, v. 11, n. 4, p. 749-764, 1936.
- MAIA, C.E.; MORAIS, E.R.C.; OLIVEIRA, M. Nível crítico pelo critério da distribuição normal reduzida: uma nova proposta para interpretação de análise foliar. **Revista Brasileira de Engenharia Agrícola e Ambiental**, Campina Grande, v. 5, p. 235-238, 2001.
- MALAVOLTA, E. **Manual de química agrícola - adubos e adubação**. 2.ed. São Paulo: Editora Agronômica CERES, 1967. 606 p.
- MALAVOLTA, E. **Elementos de nutrição mineral de plantas**. São Paulo: Editora Agronômica CERES, 1980. 251 p.
- MALAVOLTA, E.; VITTI, G.C.; OLIVEIRA, S.A. **Avaliação do estado nutricional de plantas: princípios e aplicações**. Piracicaba: Associação Brasileira para Pesquisa da Potassa e do Fosfato, 1989. 201 p.
- McGEE, V.E.; CARLETON, W.T. Piecewise regression. **Journal of the American Statistical Association**, Alexandria, v. 65, n. 331, p. 1109-1124, 1970.

MILES, N. Challenges and opportunities in leaf nutrient data interpretation. In: SOUTH AFRICAN SUGAR TECHNOLOGISTS' ASSOCIATION, South Africa. **Proceedings...** South Africa: SASTA, p. 205-215, 2010.

MITSCHERLICH, E. A. Das gesetz des minimums und das gesetz des abnehmenden bodenertrages. **Landwirtschaftliche Jahrbücher**, Berlim, v. 38, p. 537-552, 1909.

MITSCHERLICH, E.A. Die zweite annäherung des wirkungsgesetzes der wachstumsfaktoren. **Zeitschrift für Pflanzenernährung**, Berlim, v. 12, p. 273-282, 1928.

NEVES, S.C.N.; CARVALHO, J.G.; FERREIRA, E.V.O.; ASSIS, R.P. Nutrição mineral, crescimento e níveis críticos foliares de cálcio e magnésio em mudas de umbuzeiro, em função da calagem. **Revista Ceres**, Viçosa, v. 55, n. 6, p. 575-583, 2008.

OHKI, K. Critical nutrient levels related to plant growth and some physiological processes. **Journal of Plant Nutrition**, Athens, v. 10, p. 1583-1673, 1987.

PARIS, Q. The von Liebig hypothesis. **American Journal of Agricultural Economics**, Oxford, v. 74, n. 4, p. 1019-1028, 1992a.

PARIS, Q. The return of von Liebig's law. **Agronomy Journal**, Madison, v. 84, p. 1040-1046, 1992b.

PIMENTEL-GOMES, F.; NOGUEIRA, I.R. **Análise matemática**. 2.ed. Piracicaba: Departamento de Ciências Exatas da ESALQ, 1980. 371 p.

QUANDT, R.E. The estimation of the parameters of a linear regression system obeying two separate regimes. **Journal of the American Statistical Association**, Alexandria, v. 53, n. 284, p. 873-880, 1958.

SCHABENBERGER, O.; PIERCE, F.J. **Contemporary statistical models for the plant and soil sciences**. New York: CRC PRESS, 2001. 730 p.

SCHNEEBERGER, H. **Over-fertilization: an inverse Mitscherlich process**. Institute of Statistics, University of Erlangen-Nürnberg. Disponível em: <<http://soil-statistic.de/overfertilization/overfertilization.html>>. Acesso em: 24 abril 2015.

SEBER, G.A.F.; WILD, C.J. **Nonlinear regression**. United States: John Wiley & Sons, Inc., 1989. 768 p.

SMITH, P.F. Mineral analysis of plant tissues. **Annual Review of Plant Physiology**, Florida, v. 13, p. 81-108, 1962.

SPRENT, P. Some hypotheses concerning two phase regression lines, **Biometrics**, Arlington, v. 17, n. 4, p. 634-645, 1961.

STERNER, R.W.; ELSER, J.J. **Ecological stoichiometry: the biology of elements from molecules to the biosphere**. Princeton, New Jersey: Princeton University Press, 2002. 464 p.

TAIZ, L.; ZEIGER, E. **Fisiologia vegetal**. 3.ed. Porto Alegre: Artmed, 2004. 719 p.

TISHLER, A.; ZANG, I. A new maximum likelihood algorithm for piecewise regression. **Journal of the American Statistical Association**, Alexandria, v. 76, n. 376, p. 980-987, 1981a.

TISHLER, A.; ZANG, I. A maximum likelihood method for piecewise regression models with a continuous dependent variable. **Journal of the Royal Statistical Society. Series C (Applied Statistics)**, Malden, v. 30, n. 2, p. 116-124, 1981b.

WADT, P.G.S.; ALVAREZ, V. H.; NOVAIS, R. F.; FONSECA, S.; BARROS, N. F. O método da chance matemática na interpretação de dados de levantamento nutricional de eucalipto.

Revista Brasileira de Ciência do Solo, Viçosa, v. 22, p. 773-778, 1998.

WATTS, D.G.; BACON, D.W. Using an hyperbola as a transition model to fit two-regime straight-line data. **Technometrics**, Alexandria, v. 16, n. 3, p. 369-373, 1974.

2 REGRESSÃO SEGMENTADA COM TRANSIÇÕES DE FASE SUAVES SOB A PRESENÇA DE LIMIARES ALEATÓRIOS

Resumo

O objetivo deste trabalho é descrição de processos multifase com limiares aleatórios e transições de fase suaves via regressão segmentada. Inicialmente, realiza-se uma revisão das formulações sinal de Bacon e Watts (1971) e Griffiths e Miller (1973) e máx-mín de Tishler e Zang (1981a,b) para a construção de modelos segmentados. Em seguida é apresentada uma nova parametrização da formulação máx-mín, útil para a construção de modelos de regressão simples com inúmeros regimes lineares. Com o auxílio dessa nova parametrização, uma extensão da regressão “bent-cable” de Chiu et al. (2006) é obtida. Além do mais, pela inclusão de limiares aleatórios na estrutura dos modelos, são derivadas duas novas famílias de junções suaves entre regimes, denominadas aqui por junções médias dos tipos I e II. As junções médias são aplicáveis em ambas as formulações (sinal e máx-mín) e, além disso, permitem a estimação da variabilidade dos limiares aleatórios. Por fim, é realizado um estudo de simulação a fim de se avaliar o desempenho dos modelos propostos. Por meio dos resultados, observa-se que esses modelos são flexíveis para descrever o comportamento complexo de processos multifase e são fáceis de se ajustarem, considerando-se o fato de que são não lineares. No entanto, a obtenção de informação sobre os limiares aleatórios depende de um grande esforço amostral e requer delineamentos com muitos pontos de suporte nas zonas transições de fase.

Palavras-chave: Modelos segmentados; Formulação sinal; Formulação máx-mín; Junções suaves; Regressão “bent-cable”

Abstract

The aim of this work is the modeling of multi-phase processes with random thresholds and smooth phase transitions by means of segmented regression. Firstly, a review of two segmented model formulations is presented: the sign (BACON; WATTS, 1971; GRIFFITHS; MILLER, 1973) and max-min (TISHLER; ZANG, 1981a,b) formulations. This work also introduces a new parametrization of the max-min models, which is helpful to obtaining simple regression models in cases in which there are several linear regimes. From the new max-min parametrization, an extension of the bent-cable regression of Chiu et al. (2006) is introduced in this work and also two new families of approximating junction functions between regimes are derived by means of the adoption of random thresholds. These functions are called here as averaged junctions of type I and II. The averaged junctions are applicable in both of the two piecewise formulations studied (sign and max-min) and they have the advantage of providing variance estimates for the random thresholds. Finally, a simulation study (based on a real data set of the plant response to fertilizers) is conducted in order to check the performance of the proposed models. The results indicate that the models are flexible enough to describe the complex behavior shown by multi-phase data sets and, indeed, they are easily fitted to the data by using nonlinear optimization techniques. However, the gathering of information about the random thresholds is a complicated task which depends on a large amostral effort and requires designs with many support points in the transition phase zones.

Keywords: Piecewise models; Sign formulation; Max-min formulation; Smooth phase transitions; Bent-cable regression

2.1 Introdução

Em alguns problemas de regressão, as regras que relacionam as variáveis resposta e explanatória parecem se modificar ao longo do espaço de regressão. Um exemplo disso é encontrado na resposta à adubação: geralmente a produção vegetal é relacionada à adubação direta e proporcionalmente, mas só até certo nível crítico; a partir desse limiar de suficiência, as plantas não são mais capazes de responder a incrementos nos insumos e a curva de resposta torna-se constante; eventualmente as plantas se intoxicam com os excessos de nutriente, o que leva a um decréscimo da produção para doses de adubo que vão além de um ponto de toxidez (FAQUIN, 2002; HAVLIN et al., 2004; MALAVOLTA, 1993; ULRICH, 1952).

Problemas similares são abundantes nas mais variadas áreas da ciência, tais como em biologia (CHAPPEL, 1989), medicina (HOUT et al., 2011; KHAN, 2010; MUGGEO, 2003), meteorologia (RANA, 2012), ecologia (BARROWMAN; MYERS, 2000; QIAN; CUFFNEY, 2012; TOMS; LESPERANCE, 2003), finanças (HAYNE et al., 2012) e agronomia (PRUNTY, 1983; SHUAI et al., 2003). Neles, essas mudanças estruturais são ligadas à existência de diferentes fases em que o fenômeno sob pesquisa se divide, dentro das quais as observações apresentam comportamentos mais parecidos.

Nesses casos, encontrar uma expressão matemática que descreva o comportamento global das variáveis não é uma tarefa fácil. Uma alternativa para simplificar o problema é dividir o espaço de regressão e modelar os dados dentro de cada fase, por meio de funções bem simples, tais como polinômios de 1º e 2º graus. A aproximação para o comportamento complexo dos dados ao longo de todo o domínio de regressão é obtida, então, pela união de submodelos. É sobre esta ideia que se fundamenta a regressão segmentada (FEDER, 1975; HINKLEY, 1969, 1971; McGEE; CARLETON, 1970; QUANDT, 1958; SPRENT, 1961).

A regressão segmentada, em sua versão original, é caracterizada por mudanças de fase abruptas, as quais muitas vezes carecem de interpretação prática, além de causarem descon continuidades nas derivadas do modelo com respeito aos pontos de mudanças, indesejáveis para os processos de estimação e inferência (SEBER; WILD, 1989). A bibliografia especializada apresenta algumas alternativas para descrever transições de fases mais graduais, por meio da inclusão de junções suaves entre os regimes, o que torna os modelos de regressão mais realísticos e tratáveis sob o ponto de vista estatístico (BACON; WATTS, 1971; CHIU et al., 2006; GALLANT; FULLER, 1973; GRIFFITHS; MILLER, 1973; TISHLER; ZANG, 1981a,b; WATTS; BACON, 1974).

O objetivo deste trabalho é a construção de modelos segmentados com transições suaves quando o processo em estudo é caracterizado pela existência de limiares aleatórios. O ponto central do estudo é demonstrar que a inclusão de limiares aleatórios conduz, naturalmente, a

modelos segmentados com transição de fase suave, por meio dos quais a variabilidade desses limiares pode ser estimada. Neste contexto, as zonas de transição de fase são compreendidas como domínios nos quais diferentes fenômenos geradores dos dados operam de maneira simultânea. O presente trabalho restringe-se ao caso da regressão simples e univariada, e sob a pressuposição de regimes com a forma linear.

Este trabalho é dividido da seguinte forma: na seção 2.2 é apresentada uma revisão bibliográfica sobre regressão segmentada e modelos com transições suaves, sendo detalhadas as alternativas mais populares para modelar mudanças de fase graduais: as formulações sinal de Bacon e Watts (1971) e máx-mín de Tishler e Zang (1981a,b). Na seção 2.3.1 é introduzida uma nova parametrização dos modelos máx-mín, a qual nos possibilita estender a regressão “bent-cable” de Chiu et al. (2006). A seção 2.3.2 trata da inclusão de limiares aleatórios na estrutura dos modelos lineares segmentados e nela, como resultados dessa inclusão, são derivadas duas novas famílias de junções suaves entre regimes, as junções médias dos tipos I e II. A seção 2.3.3 aborda os processos de obtenção de valores iniciais e estimação dos parâmetros dos modelos propostos pelo método da máxima verossimilhança. Já a seção de resultados do trabalho é dividida em três partes: em 2.4.2 são apresentados alguns resultados teóricos acerca da inter-relação existente entre as junções médias do tipo I e as aproximações suaves da família transicional de Bacon e Watts (1971) e Griffiths e Miller (1973); em 2.4.1 são apresentados os resultados da simulação de dados conduzida para avaliar o desempenho dos modelos propostos durante os processos de estimação e inferência. Por fim, na seção 2.5 são dispostas as conclusões do trabalho.

2.2 Revisão Bibliográfica

Seja um modelo de regressão simples, que considera apenas uma variável explanatória e usualmente é expresso por

$$y_i = \eta(x_i) + \epsilon_i \quad (i = 1, 2, \dots, n), \quad (1)$$

em que y_i é a i -ésima das n respostas observadas de uma variável aleatória Y , ϵ_i é o i -ésimo erro aleatório e $\eta(x_i)$ é a resposta esperada condicionada ao i -ésimo valor da variável explanatória, denotado por x_i . No contexto de análise de regressão, o valor x_i é fixado pelo pesquisador ou mensurado de forma precisa, com erros desprezíveis (RATKOWSKY, 1990).

Na regressão segmentada, a parte fixa do modelo é desdobrada em

$$\eta(x) = \eta(x|\boldsymbol{\beta}, \boldsymbol{\tau}) = \begin{cases} g_1(x|\boldsymbol{\beta}_1), & A \leq x \leq \tau_1, \\ g_2(x|\boldsymbol{\beta}_2), & \tau_1 < x \leq \tau_2, \\ \vdots & \vdots \\ g_D(x|\boldsymbol{\beta}_D), & \tau_{D-1} < x \leq B, \end{cases}$$

na qual: as funções g_j 's, com $j = 1, 2, \dots, D$, denominadas também por regimes, geralmente são polinômios de 1º e 2º graus ou funções não lineares de x com poucos parâmetros que caracterizam a forma das D fases do processo estudado; $\boldsymbol{\beta}_j$'s são os vetores de parâmetros associados a cada um dos regimes, de maneira que $\boldsymbol{\beta}^T = (\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T, \dots, \boldsymbol{\beta}_D^T)$; A e B são os limites inferior e superior do domínio de regressão e $\boldsymbol{\tau}^T = (\tau_1, \tau_2, \dots, \tau_{D-1})$, com $A < \tau_1 < \tau_2 < \dots < \tau_{D-1} < B$. Os elementos do vetor $\boldsymbol{\tau}$ são os valores das abscissas que delimitam a partição proposta do domínio de regressão e usualmente recebem nomes como pontos de junção, de transição, de mudança ou pontos finais. Por razões fenomenológicas, também os chamaremos de limiares. A modelagem de respostas contínuas exige, ainda, a imposição da seguinte restrição

$$g_l(x = \tau_l|\boldsymbol{\beta}_l) = g_{l+1}(x = \tau_l|\boldsymbol{\beta}_{l+1}), \quad (2)$$

para $l = 1, \dots, D - 1$.

O pioneiro na regressão segmentada foi Quandt (1958), que seguido por Sprent (1961), Hinkley (1969, 1971) e Feder (1975), estudou aspectos da estimação dos parâmetros e inferência para sistemas formados por duas retas que se interceptam. No trabalho de McGee e Carleton (1970), por sua vez, a ideia de Quandt (1958) foi estendida para um número indeterminado (e possivelmente desconhecido) de regimes polinomiais, cunhando-se o termo “piecewise regression”, método para a busca da função contínua segmentada que melhor descreve o conjunto de dados.

Os gráficos desses modelos caracterizam-se pela presença de quinas, que ocorrem nas intersecções de regimes adjacentes, o que implica a inexistência da derivada de $\eta(x)$ sob os pontos de mudança. O problema é que sem derivadas contínuas da parte fixa do modelo, os métodos usuais de estimação e inferência em regressão não linear não são aplicáveis (SEBER; WILD, 1989). Além disso, mudanças repentinas são incomuns na natureza; as transições de fase geralmente são fenômenos graduais.

Gallant e Fuller (1973) propuseram, então, uma das primeiras formulações de modelos segmentados com junções suaves entre os regimes. A proposta generaliza a metodologia de

McGee e Carleton (1970) ao impor restrições de continuidade para as derivadas de $\eta(x)$ em modelos segmentados com regimes polinomiais. Como caso particular dos modelos de Gallant e Fuller (1973) podem ser citados os q -“splines”, curvas constituídas por segmentos polinomiais de grau q e com derivadas em relação a x contínuas até a ordem $q - 1$ (SEBER; WILD, 1989).

A bibliografia especializada também apresenta outras alternativas para suavizar as junções e obter modelos mais tratáveis do ponto de vista estatístico. Elas seguem praticamente o mesmo caminho: formulam modelos com o auxílio de operadores abruptos, como máximos, mínimos, módulos ou sinais e, em seguida, os substituem por aproximações que conduzem a funções continuamente diferenciáveis (ou seja, funções da classe C^1). Na bibliografia especializada destacam-se duas abordagens: as formulações sinal de Bacon e Watts (1971) e máx-mín de Tishler e Zang (1981a,b), as quais são descritas a seguir.

2.2.1 A Formulação Sinal: Modelos Transicionais e Hiperbólicos

Como verifica-se no trabalho de Bacon e Watts (1971), a função sinal pode ser convenientemente aplicada para construir modelos segmentados contínuos. Sejam D regimes lineares, expressos por

$$g_j(x|\beta_j) = \alpha_j + \beta_j x \quad (j = 1, 2, \dots, D), \quad (3)$$

em que $\beta_j^T = (\alpha_j, \beta_j)$ é o vetor de parâmetros para o j -ésimo regime, sendo α_j e β_j seus coeficientes linear e angular, respectivamente. Para o caso em que há apenas dois regimes ($D = 2$), Bacon e Watts (1971) propõem a seguinte parametrização:

$$\begin{aligned} \eta(x|\theta, \tau_1) &= \theta_0 + \theta_1(x - \tau_1) + \theta_2|x - \tau_1| \\ &= \theta_0 + \theta_1(x - \tau_1) + \theta_2(x - \tau_1) \text{sinal}(x - \tau_1), \end{aligned} \quad (4)$$

em que τ_1 é o ponto de mudança e $\theta^T = (\theta_0, \theta_1, \theta_2)$, sendo que: $\theta_0 = \alpha_1 + \beta_1\tau_1$ é a resposta esperada para $x = \tau_1$, ou seja, no ponto de mudança; o parâmetro $\theta_1 = (\beta_1 + \beta_2)/2$ é a média das inclinações; e o parâmetro $\theta_2 = (\beta_2 - \beta_1)/2$ é a metade da diferença entre as inclinações dos regimes (Figura 1). Note que, por meio dessa parametrização, $\beta_1 = \theta_1 - \theta_2$ e $\beta_2 = \theta_1 + \theta_2$, de maneira que: se $\theta_1 = -\theta_2$, então β_2 é nulo e, conseqüentemente, a resposta no segundo regime é constante e igual a θ_0 (ver Figura 2, à direita, modelo abrupto); por outro lado, se $\theta_1 = \theta_2$, β_1 é nulo e a resposta no primeiro regime é constante e igual a θ_0 . Dessa forma, o gráfico do modelo é côncavo quando $\beta_1 > \beta_2$ ($\theta_2 < 0$) e convexo caso $\beta_2 > \beta_1$ ($\theta_2 > 0$).

Observe que os domínios nos quais cada regime opera aparecem implícitos em (4) por meio da função sinal: a observação é classificada como pertencente ao primeiro regime se

Figura 1 – Gráficos da junção de dois regimes por meio da formulação sinal de Bacon e Watts (1971) para $\theta_2 > 0$ e $\theta_2 < 0$

$x - \tau_1 \leq 0$ e é classificada como pertencente ao segundo regime se $x - \tau_1 > 0$. A transição entre regimes ocorre abruptamente quando $x = \tau_1$.

De modo a evitar a transição abrupta, Bacon e Watts (1971) sugerem a substituição da função sinal por aproximações continuamente diferenciáveis, introduzindo uma família de funções transicionais suaves. Os modelos que se incluem nesta adaptação são denominados modelos transicionais e são originalmente expressos por

$$\eta(x|\boldsymbol{\theta}, \tau_1, \gamma) = \theta_0 + \theta_1(x - \tau_1) + \theta_2(x - \tau_1) \text{trn}(x - \tau_1; \gamma), \quad (5)$$

em que trn denota a função transicional. O parâmetro γ ($\gamma \in \mathbb{R}_+^*$) controla a curvatura do modelo sob o ponto de junção, partindo de uma transição abrupta quando $\gamma \rightarrow 0^+$ a uma mudança mais gradual conforme seu valor aumenta (Figura 2, à esquerda).

A família de funções transicionais é definida por um conjunto de condições introduzidas por Bacon e Watts (1971) e adaptadas por Griffiths e Miller (1973), dadas por:

- (i) $\lim_{s \rightarrow -\infty} s [\text{trn}(s; \gamma) + 1] = 0$ e $\lim_{s \rightarrow +\infty} s [\text{trn}(s; \gamma) - 1] = 0 \quad \forall s \in \mathbb{R}$,
- (ii) $\lim_{\gamma \rightarrow 0^+} \text{trn}(s; \gamma) = \text{sinal}(s) \quad \forall s \in \mathbb{R}$,
- (iii) $\text{trn}(0; \gamma) = 0$.

No caso, $s = x - \tau_1$. A condição (i) garante o comportamento linearmente assintótico do modelo (5) às retas que lhe dão origem (expressão (3)) à medida que x se afasta do ponto de mudança τ_1 pela esquerda ou pela direita. A condição (ii) garante que o modelo linear abrupto seja um caso particular do modelo transicional em (5) quando $\gamma \rightarrow 0^+$. A condição (iii) determina que a função $\text{trn}(s)$ seja igual a função $\text{sinal}(s)$ quando $s = 0$ e, dessa maneira, que o gráfico do

Figura 2 – Modelos segmentados da formulação sinal com $\theta_0 = 6$, $\theta_1 = 0,3$, $\theta_2 = -0,3$ e $\tau_1 = 10$. À esquerda são apresentados modelos transicionais com diferentes valores para γ . À direita são apresentados os modelos abrupto, transicional e hiperbólico

modelo passe através do ponto de mudança, de coordenadas (τ_1, θ_0) , onde as assíntotas que representam os regimes se interceptam (SEBER; WILD, 1989).

Muitas funções satisfazem às condições da família transicional. Como exemplo, Bacon e Watts (1971) mencionam a tangente hiperbólica, expressa por

$$\tanh\left(\frac{s}{\gamma}\right) = \frac{e^{\frac{s}{\gamma}} - e^{-\frac{s}{\gamma}}}{e^{\frac{s}{\gamma}} + e^{-\frac{s}{\gamma}}}.$$

Os autores ainda afirmam que se pode utilizar a função de distribuição acumulada de variáveis aleatórias com densidades simétricas como função transicional. É preciso tomar cuidado com essa última colocação de Bacon e Watts (1971). Na realidade, nem sempre uma função de distribuição pode ser usada na construção de uma função transicional e, além do mais, a simetria não é sequer uma condição necessária. As condições necessárias e suficientes para se utilizar uma função de distribuição acumulada como base para a obtenção de uma função transicional serão discutidas adiante, na seção 2.4.2.

Conforme Bacon e Watts (1971), a forma da função transicional tem importância secundária, por várias razões: a variabilidade dos dados geralmente mascara as diferenças entre as funções; o processo de estimação a ser empregado é praticamente o mesmo para qualquer função transicional que se queira usar; e a função transicional é disposta no modelo multiplicando o termo $(x - \tau_1)$, de maneira que o valor de $(x - \tau_1) \operatorname{trn}[(x - \tau_1)/\gamma]$ é pouco influenciado pela forma de $\operatorname{trn}[(x - \tau_1)/\gamma]$. A fim de demonstrar que a escolha da função transicional é secundária, Bacon e Watts (1971) apresentam uma comparação dos valores de $s \operatorname{trn}(s; \gamma)$ para as

funções transicionais

$$\text{sinal}(s) [1 - \exp(-|s|/\gamma)], \quad \text{sinal}(s) \left[\frac{(s/\gamma)^2}{1 + (s/\gamma)^2} \right] \text{ e } \tanh(s/\gamma),$$

obtendo diferenças que, conforme os autores, são extremamente pequenas. Como a escolha da função transicional não é influente, Bacon e Watts (1971) usam no decorrer de seu trabalho apenas a tangente hiperbólica como função transicional.

É importante notar que o uso de aproximações transicionais distorce o gráfico do modelo original. Os modelos abruptos e transicionais apresentam respostas iguais apenas sob o ponto de mudança, quando $\text{trn}(x - \tau_1; \gamma) = 0$ e $\eta(x | \boldsymbol{\theta}, \tau_1, \gamma) = \theta_0$, e para todos os outros pontos as curvas são diferentes. Se a curva de resposta é côncava ($\theta_2 < 0$), o gráfico do modelo transicional fica acima do gráfico do modelo abrupto no plano, tocando-o apenas sob $x = \tau_1$ (ver Figura 2, à direita). Por outro lado, se a curva de resposta é convexa ($\theta_2 > 0$), então o gráfico do modelo transicional se dispõe abaixo do gráfico do modelo abrupto, tocando-o apenas em $x = \tau_1$. A distorção causada pela função transicional é maior nas vizinhanças do ponto de mudança, onde caracteriza-se por uma saliência (“bulge”, no Inglês) no gráfico do modelo (BACON; WATTS, 1971). A saliência decorre da condição (iii), que força o gráfico do modelo a atravessar o ponto (τ_1, θ_0) . Observa-se que quanto maior é o valor do parâmetro de suavização γ , mais acentuada é essa saliência (Figura 2).

Um dos problema com os modelos transicionais é o fato de que, em diversas aplicações dos modelos segmentados, a saliência presente na curva de resposta não deveria existir. Como alternativa às funções transicionais, Griffiths e Miller (1973) sugerem substituir a função sinal por

$$\text{hip}(s; \gamma) = \frac{\sqrt{s^2 + \gamma}}{s}, \quad s \neq 0,$$

o que gera gráficos similares aos das hipérboles. Quando $s = 0$, considera-se o fato de que

$$\lim_{s \rightarrow 0} s \text{ hip}(s; \gamma) = \sqrt{\gamma}.$$

Na família de aproximações hiperbólicas de Griffiths e Miller (1973) as condições (i) e (ii) devem ser atendidas, mas a condição (iii) é suprimida, justamente para evitar a saliência presente no gráfico dos modelos transicionais.

A função $\text{hip}(s; \gamma)$ é descontínua em $s = 0$ e, para valores de s próximos de zero, as funções $\text{sinal}(s)$ e $\text{hip}(s; \gamma)$ apresentam comportamentos completamente distintos. No entanto, a função $\text{hip}(s; \gamma)$ se aproxima de -1 e $+1$ assintoticamente, quando s tende ao infinito nega-

tivo e positivo, respectivamente. Dessa forma, apesar da descontinuidade existente nas funções da família hiperbólica, elas geram modelos continuamente diferenciáveis, os quais são expressos por

$$\begin{aligned}\eta(x|\boldsymbol{\theta}, \boldsymbol{\tau}, \gamma) &= \theta_0 + \theta_1(x - \tau_1) + \theta_2(x - \tau_1) \operatorname{hip}(x - \tau_1; \gamma) \\ &= \theta_0 + \theta_1(x - \tau_1) + \theta_2 \sqrt{(x - \tau_1)^2 + \gamma}.\end{aligned}\quad (6)$$

Esses modelos também foram propostos de maneira independente por Watts e Bacon (1974), com a única diferença do uso de $\gamma^2/4$ ao invés de γ (SEBER; WILD, 1989).

O gráfico do modelo hiperbólico une os regimes de forma similar ao que faz o gráfico da hipérbole com respeito as suas assíntotas (ver Figura 2). Se a curva de resposta é côncava ($\theta_2 < 0$), o gráfico do modelo hiperbólico se dispõe sempre abaixo do gráfico do modelo abrupto. Por outro lado, se a curva de resposta é convexa ($\theta_2 > 0$), o gráfico do modelo hiperbólico se dispõe acima do gráfico do modelo abrupto. Em resumo, verifica-se que:

a) Para curvas de resposta côncavas ($\theta_2 < 0$):

$$\text{Hiperbólico} < \text{Abrupto} \leq \text{Transicional};$$

b) Para curvas de resposta convexas ($\theta_2 > 0$):

$$\text{Transicional} \leq \text{Abrupto} < \text{Hiperbólico}.$$

Os modelos apresentados nas expressões (5) e (6) são destinados às situações em que o processo estudado se divide em duas fases, mas muitos fenômenos naturais caracterizam-se pela existência de mais do que dois regimes. Griffiths et al. (1975) propõem uma extensão da formulação sinal de Bacon e Watts (1971) flexível o bastante para descrever processos com quantos regimes forem necessários, sendo que os mesmos não precisam ser lineares. Esses modelos são expressos por

$$\eta(x|\boldsymbol{\beta}, \boldsymbol{\tau}) = \frac{1}{2} \left\{ g_1(x|\boldsymbol{\beta}_1) + g_D(x|\boldsymbol{\beta}_D) + \sum_{k=2}^D \left[g_k(x|\boldsymbol{\beta}_k) - g_{k-1}(x|\boldsymbol{\beta}_{k-1}) \right] \operatorname{sinal}(x - \tau_{k-1}) \right\}. \quad (7)$$

Seber e Wild (1989) obtêm a expressão para o caso particular de regimes lineares (em que cada fase do processo pode ser descrita adequadamente por uma reta), dada por

$$\eta(x|\boldsymbol{\theta}, \boldsymbol{\tau}) = \theta_0 + \theta_1 x + \sum_{k=2}^D \theta_k (x - \tau_{k-1}) \operatorname{sinal}(x - \tau_{k-1}), \quad (8)$$

em que $\beta_1 = \theta_1 - \sum_{k=2}^D \theta_k$ e $\beta_d = \sum_{k=1}^d \theta_k - \sum_{k=d+1}^D \theta_k$ ($d = 2, \dots, D-1$), de maneira que $\theta_k = (\beta_k - \beta_{k-1})/2$. Neste caso os segmentos lineares se interceptam nos pontos $(\tau_l, \theta_0 +$

$\theta_1 \tau_l + \sum_{k=2}^D \theta_k |\tau_l - \tau_{k-1}|$) e, portanto, os coeficientes lineares dos regimes são dados por

$$\alpha_l = \theta_0 + (\theta_1 - \beta_l) \tau_l + \sum_{k=2}^D \theta_k |\tau_l - \tau_{k-1}|$$

e

$$\alpha_D = \theta_0 + (\theta_1 - \beta_D) \tau_{D-1} + \sum_{k=2}^D \theta_k |\tau_{D-1} - \tau_{k-1}|.$$

Já o intercepto do modelo é dado por

$$\eta_0 = \theta_0 + \sum_{k=2}^D \theta_k |\tau_{k-1}|$$

e o parâmetro θ_0 é expresso por

$$\theta_0 = \alpha_1 - (\theta_1 - \beta_1) \tau_1 - \sum_{k=2}^D \theta_k |\tau_1 - \tau_{k-1}|. \quad (9)$$

É importante notar que a interpretação dos parâmetros θ 's no modelo disposto em (8) difere daquela discutida para a expressão (4).

Modelos com junções suaves entre as fases são facilmente obtidos ao substituirmos as funções do tipo sinal em (8) por aproximações transicionais e/ou hiperbólicas.

2.2.2 A Formulação Máx-Mín

A formulação máx-mín de Tishler e Zang (1981a,b) é uma das formas mais flexíveis para lidar com problemas de regressão segmentada. Esta abordagem consiste na utilização de operadores de máximo e mínimo para a descrição sucinta da resposta nos diversos regimes. Por exemplo, quando há apenas dois regimes, o modelo segmentado em questão pode ser descrito como

$$\begin{aligned} \eta(x|\beta) &= \text{máx}\{g_1(x|\beta_1), g_2(x|\beta_2)\} \\ \text{ou} \\ \eta(x|\beta) &= \text{mín}\{g_1(x|\beta_1), g_2(x|\beta_2)\}, \end{aligned} \quad (10)$$

dependendo da concavidade da curva a ser ajustada (veja a Figura 3). Um raciocínio similar pode ser aplicado quando há mais do que dois regimes. Os regimes caracterizados por regiões convexas são unidos pelo operador de máximo, enquanto o mínimo é aplicado às regiões côncavas.

Regimes além dos existentes podem ser adicionados na formulação do modelo sem qualquer ônus. Se a função que representa o regime não for necessária para explicar o processo, ela será desconsiderada após a estimação dos parâmetros. Por exemplo, considere o modelo $\eta(x) = \max\{g_1(x), g_2(x), g_3(x)\}$ e suponha que os regimes obtidos após o processo de estimação são expressos por $g_1(x) = -1 + x$, $g_2(x) = 1 - x$ e $g_3(x) = -3 - x$. Neste caso, o terceiro regime pode ser suprimido do modelo sem qualquer consequência, uma vez que ele é paralelo ao segundo regime e situa-se sempre abaixo do mesmo no plano cartesiano, o que resulta em $\eta(x) = \max\{g_1(x), g_2(x)\}$.

Figura 3 – Gráficos de duas formas de modelos segmentados com dois regimes: côncava, à esquerda e convexa, à direita. Figura adaptada de Tishler e Zang (1981b)

A formulação máx-mín também habilita o pesquisador a trabalhar com diversas variáveis explanatórias ao mesmo tempo. Na regressão múltipla, os regimes são planos ou superfícies que se interceptam em retas e/ou contornos de transição. Sob quaisquer circunstâncias, o mesmo método recursivo pode ser aplicado para a construção do modelo (TISHLER; ZANG, 1981a). Seja

$$u(z) = z_+ = \max\{0, z\} \quad \forall z \in \mathbb{R}. \quad (11)$$

Então, um modelo com dois regimes ($D = 2$) e para o qual a superfície de resposta é convexa é obtido a partir de

$$\eta(\mathbf{x}|\boldsymbol{\beta}) = g_1(\mathbf{x}|\boldsymbol{\beta}_1) + u[g_2(\mathbf{x}|\boldsymbol{\beta}_2) - g_1(\mathbf{x}|\boldsymbol{\beta}_1)], \quad (12)$$

em que $\mathbf{x}^T = (x_1, \dots, x_m)$ é o vetor de valores fixos para as m variáveis explanatórias existentes. Se há três regimes ($D = 3$), o modelo convexo é descrito por

$$\eta(\mathbf{x}|\boldsymbol{\beta}) = g_1(\mathbf{x}|\boldsymbol{\beta}_1) + u[g_2(\mathbf{x}|\boldsymbol{\beta}_2) - g_1(\mathbf{x}|\boldsymbol{\beta}_1)] + u[g_3(\mathbf{x}|\boldsymbol{\beta}_3) - g_2(\mathbf{x}|\boldsymbol{\beta}_2)],$$

Recursivamente, a expressão do modelo convexo para qualquer número de regimes $D \geq 2$ é dada por

$$\eta(\mathbf{x}|\boldsymbol{\beta}) = g_1(\mathbf{x}|\boldsymbol{\beta}_1) + u[g_2(\mathbf{x}|\boldsymbol{\beta}_2) - g_1(\mathbf{x}|\boldsymbol{\beta}_1) + u[\cdots + u[g_D(\mathbf{x}|\boldsymbol{\beta}_D) - g_{D-1}(\mathbf{x}|\boldsymbol{\beta}_{D-1})]\cdots]]. \quad (13)$$

Por outro lado, um modelo com dois regimes ($D = 2$) e superfície de resposta côncava é expresso por

$$\eta(\mathbf{x}|\boldsymbol{\beta}) = g_1(\mathbf{x}|\boldsymbol{\beta}_1) - u[g_1(\mathbf{x}|\boldsymbol{\beta}_1) - g_2(\mathbf{x}|\boldsymbol{\beta}_2)]. \quad (14)$$

Já um modelo com três regimes ($D = 3$) e superfície côncava é obtido por

$$\eta(\mathbf{x}|\boldsymbol{\beta}) = g_1(\mathbf{x}|\boldsymbol{\beta}_1) - u[g_1(\mathbf{x}|\boldsymbol{\beta}_1) - g_2(\mathbf{x}|\boldsymbol{\beta}_2) + u[g_2(\mathbf{x}|\boldsymbol{\beta}_2) - g_3(\mathbf{x}|\boldsymbol{\beta}_3)]]. \quad (15)$$

Recursivamente, a expressão do modelo côncavo para qualquer número de regimes $D \geq 2$ é dada por

$$\eta(\mathbf{x}|\boldsymbol{\beta}) = g_1(\mathbf{x}|\boldsymbol{\beta}_1) - u[g_1(\mathbf{x}|\boldsymbol{\beta}_1) - g_2(\mathbf{x}|\boldsymbol{\beta}_2) + u[\cdots + u[g_{D-1}(\mathbf{x}|\boldsymbol{\beta}_{D-1}) - g_D(\mathbf{x}|\boldsymbol{\beta}_D)]\cdots]]. \quad (16)$$

Para demonstrar as fórmulas dispostas nas expressões de (12) a (16), basta considerar o fato de que $\min\{a, b\} = -\max\{-a, -b\}$ e $\max\{a, b\} = a + \max\{b - a, 0\} = a + u(b - a)$, para quaisquer a e b reais. Consequentemente, regiões côncavas e convexas podem ser unidas por meio da estrutura apresentada em (13) realizando-se pequenas mudanças de sinais.

Os modelos oriundos da formulação máx-mín são diferenciáveis com relação às variáveis independentes em todas as ordens que os regimes também os são, com exceção dos valores de $\mathbf{x}$ que satisfazem as igualdades

$$g_l(\mathbf{x}|\boldsymbol{\beta}_l) = g_{l+1}(\mathbf{x}|\boldsymbol{\beta}_{l+1}) \quad \forall l,$$

ou seja, as derivadas não existem sob os pontos ou contornos de transição de fase. Tishler e Zang (1981a) mostram que esse problema pode ser superado substituindo-se o operador de máximo u nas redondezas dos pontos (ou contornos) de transição por funções continuamente diferenciáveis que aproximam o seu comportamento. A fim de obter tais aproximações, torna-se conveniente definir um novo conjunto de variáveis que é baseado nas diferenças dos valores de regimes adjacentes, sendo essas variáveis expressas por:

$$z_l = z_l(\mathbf{x}) = g_{l+1}(\mathbf{x}|\boldsymbol{\beta}_{l+1}) - g_l(\mathbf{x}|\boldsymbol{\beta}_l) \quad \forall l.$$

A função u é substituída ao longo de pequenas faixas (regiões de transição) que englobam os

Figura 4 – Gráfico da função $u_\gamma(z)$, construída a partir da função $\psi(z; \gamma)$ expressa em (22) e com $\gamma = 20$

pontos (ou contornos) de transição ($z_l = 0$ para algum l) e são delimitadas por valores de $z_l \in [-\gamma_l, \gamma_l]$, sendo que os parâmetros γ_l 's assumem valores positivos e controlam a largura dessas faixas (Figura 4). O operador de máximo u associado a l -ésima transição de fase é substituído pela função suave

$$u_{\gamma_l}(z_l) = \psi_l(z_l, \gamma_l) \mathbf{I}_{\{|z_l| \leq \gamma_l\}}(z_l) + z_l \mathbf{I}_{\{z_l > \gamma_l\}}(z_l), \quad (17)$$

em que:

- ◇ $\mathbf{I}_{\{\mathbb{A}\}}(z_l)$ denota a função indicadora, sendo $\mathbf{I}_{\{\mathbb{A}\}}(z_l) = 1$ se $z_l \in \mathbb{A}$ e $\mathbf{I}_{\{\mathbb{A}\}}(z_l) = 0$ caso contrário;
- ◇ As funções $\psi_l : [-\gamma_l, \gamma_l] \rightarrow \mathbb{R}$ são da Classe C^1 e satisfazem as seguintes condições:

(a) Restrições de continuidade, expressas por

$$\psi_l(z_l = -\gamma_l, \gamma_l) = 0 \quad \text{e} \quad \psi_l(z_l = \gamma_l, \gamma_l) = z_l = \gamma_l; \quad (18)$$

(b) Restrições de diferenciabilidade, expressas por

$$\partial_+ \psi_l(-\gamma_l) = 0 \quad \text{e} \quad \partial_- \psi_l(+\gamma_l) = 1, \quad (19)$$

em que $\partial_+ \psi_l$ e $\partial_- \psi_l$ denotam as derivadas laterais à direita e à esquerda, respectivamente, da função ψ_l com respeito a z_l ;

(c) Restrição de comportamento limite, dada por

$$\lim_{\gamma_l \rightarrow 0^+} u_{\gamma_l}(z_l) = u(z_l) \quad \forall z_l \in \mathbb{R}. \quad (20)$$

A condição (a) garante a construção de modelos contínuos, a condição (b) garante derivadas contínuas ao longo de todo o domínio de regressão e a condição (c) permite que o modelo abrupto seja um caso particular do modelo suave. Observa-se que a condição (c) é satisfeita somente se

$$\lim_{\gamma_l \rightarrow 0^+} \psi_l(0, \gamma_l) = 0. \quad (21)$$

A aproximação suave proposta por Tishler e Zang (1981a) modifica o modelo apenas nas regiões de transição, sendo que o restante da curva se mantém inalterada. A interpretação dos parâmetros associados aos regimes também se mantém a mesma. Os parâmetros γ_l 's controlam a curvatura do modelo nas regiões de transição, de maneira que as transições são mais graduais conforme seus valores aumentam e mais abruptas à medida que esses valores tendem a zero.

Tishler e Zang (1981a) utilizam junções quadráticas do tipo $\psi_l(z_l, \gamma_l) = (z_l + \gamma_l)^2 / 4\gamma_l$ para obter aproximações com derivadas de primeira ordem contínuas. Observe, entretanto, que deste modo as derivadas de segunda ordem da função $u_{\gamma_l}(z_l)$ com respeito a z_l são descontínuas nos pontos $z_l = \pm \gamma_l$. Derivadas contínuas de maior ordem podem ser obtidas a partir de junções com maior grau, como

$$\psi_l(z_l, \gamma_l) = -\frac{z_l^4}{16\gamma_l^3} + \frac{3z_l^2}{8\gamma_l} + \frac{z_l}{2} + \frac{3\gamma_l}{16}, \quad (22)$$

que produz derivadas de segunda ordem contínuas (ZANG, 1980).

Mais recentemente, Chiu et al. (2006) estudaram aspectos da teoria assintótica para os estimadores de máxima verossimilhança e mínimos quadrados dos parâmetros do modelo

$$\eta(x | \beta_1, \delta, \tau_1, \zeta) = \alpha_1 + \beta_1 x + \delta q(x | \tau_1, \zeta), \quad \zeta > 0, \quad (23)$$

em que $\delta = \beta_2 - \beta_1$ e

$$q(x | \tau_1, \zeta) = \frac{(x - \tau_1 + \zeta)^2}{4\zeta} \mathbf{I}_{\{|x - \tau_1| \leq \zeta\}}(x) + (x - \tau_1) \mathbf{I}_{\{|x - \tau_1| > \zeta\}}(x).$$

A proposta de Chiu et al. (2006) é conhecida como regressão “bent-cable” e pode-se demonstrar que ela decorre do uso da aproximação $u_{\gamma_l}(z_l)$, associada à junção quadrática $\psi_l(z_l, \gamma_l) = (z_l + \gamma_l)^2 / 4\gamma_l$ de Tishler e Zang (1981a), para o caso de regressão simples com dois regimes lineares. Nesta formulação, o parâmetro ζ controla a curvatura do modelo na região de transição, e a

largura dessa faixa de transição tem comprimento igual a 2ζ (ao longo do eixo x).

2.3 Metodologia

Primeiramente, na seção 2.3.1 é apresentada uma nova forma de construir modelos segmentados pela aplicação de operadores de máximo e mínimo, especialmente para casos de regressão simples e cujos regimes têm a forma linear. Na mesma seção, as aproximações suaves para os operadores de máximo $u_{\gamma_l}(z_l)$ propostas por Tishler e Zang (1981a) são redefinidas para o caso da regressão simples e, com esse artifício, torna-se possível estender a regressão “bent-cable”. Nesta versão estendida da regressão “bent-cable” é possível incluir inúmeros regimes lineares, ao invés de apenas dois, e ainda pode-se utilizar junções diferentes da quadrática, tal como a proposta por Zang (1980) e apresentada em (22). Já na seção 2.3.2, o foco é a inclusão de limiares aleatórios na estrutura dos modelos segmentados e, como resultado dessa inclusão, duas famílias de junções suaves entre regimes lineares são descobertas. A seção 2.3.3 trata de aspectos relacionados à estimação dos parâmetros de modelos segmentados com regimes lineares e transições suaves por meio do método da máxima verossimilhança. Por fim, a seção 2.3.4 descreve a metodologia usada na condução do estudo de simulação de dados, realizado para caracterizar as distribuições amostrais dos estimadores de máxima verossimilhança para os parâmetros.

2.3.1 A Regressão Bent-Cable Estendida

A disposição hierárquica dos operadores de máximo em (13) é o que possibilita a construção de modelos segmentados no caso de regressão múltipla, mas é completamente desnecessária no contexto da regressão simples. Nesta seção apresenta-se uma forma mais sucinta para aplicar operadores de máximo e mínimo na construção de modelos segmentados. A nova abordagem é especialmente voltada a problemas de regressão simples e sob a pressuposição de regimes lineares. Os modelos propostos nesta seção são estatisticamente mais tratáveis do que os descritos por Tisher e Zang (1981a), pois são aditivos, boa parte de seus parâmetros é linear e, além disso, permitem a demarcação de zonas de transição ao longo dos valores da variável explanatória x (ao invés da variável artificial z_l), o que pode ser de interesse prático. Demonstra-se que a regressão “bent-cable” de Chiu et al. (2006) é um caso particular dessa nova formulação de modelos máx-mín e, por esse motivo, ela será denominada no decorrer do texto por regressão “bent-cable” estendida.

Sejam D regimes lineares, expressos por:

$$g_j(x|\beta_j) = \alpha_j + \beta_j x, \quad j = 1, \dots, D.$$

Pelas restrições de continuidade impostas em (2) para o modelo, tem-se que $\alpha_l + \beta_l \tau_l = \alpha_{l+1} + \beta_{l+1} \tau_l$, para $l = 1, \dots, D-1$, de maneira que os pontos de mudança τ_l 's são dados por

$$\tau_l = \frac{\alpha_l - \alpha_{l+1}}{\beta_{l+1} - \beta_l}. \quad (24)$$

Considere, inicialmente, a união de apenas dois regimes adjacentes, indicados pelos índices l e $l+1$. Uma vez que: (i) $\max\{a, b\} = a + \max\{b - a, 0\}$ e (ii) $\min\{a, b\} = -\max\{-a, -b\}$ para quaisquer a e b reais, então é possível unir os dois regimes por meio dos operadores de máximo ou mínimo, cuja escolha depende da concavidade do modelo na região de transição. Se $\beta_{l+1} > \beta_l$, então a curva a ser modelada é convexa e, portanto, aplicamos o operador de máximo da seguinte forma

$$\begin{aligned} \eta(x|\beta) &= \max\{g_1(x|\beta_1), g_2(x|\beta_2)\} = g_1(x|\beta_1) + \max\{g_2(x|\beta_2) - g_1(x|\beta_1), 0\} \\ &= \alpha_l + \beta_l x + u(z_l). \end{aligned}$$

Por outro lado, se $\beta_{l+1} < \beta_l$, a curva é côncava e a união é possível por meio do operador de mínimo, resultando em

$$\begin{aligned} \eta(x|\beta) &= \min\{g_1(x|\beta_1), g_2(x|\beta_2)\} = -\max\{-g_1(x|\beta_1), -g_2(x|\beta_2)\} \\ &= \alpha_l + \beta_l x - u(-z_l). \end{aligned}$$

A situação em que $\beta_{l+1} = \beta_l$ é trivial, dado que neste caso não há mudança de regime.

Partindo-se da união dos dois regimes adjacentes, os demais podem ser adicionados recursivamente ao modelo corrente pela aplicação dos operadores de máximo e mínimo, até que todos sejam contemplados. Consequentemente, o modelo para um número arbitrário de regimes D ($D \geq 2$) é descrito por

$$\eta(x|\beta) = \alpha_1 + \beta_1 x + \sum_{l=1}^{D-1} \pm u(\pm z_l). \quad (25)$$

Uma parametrização conveniente é obtida ao se definir $\delta_l = \beta_{l+1} - \beta_l$. Dessa maneira, $z_l = \alpha_{l+1} - \alpha_l + \delta_l x$ e, por meio de (24), verifica-se que $\alpha_{l+1} - \alpha_l = -\delta_l \tau_l$ e $z_l = \delta_l(x - \tau_l)$. Logo, se a curva a ser modelada é convexa, temos que $\beta_{l+1} > \beta_l \Leftrightarrow \delta_l > 0$ e, portanto,

$$u(z_l) = u(\delta_l(x - \tau_l)) = \delta_l u(x - \tau_l). \quad (26)$$

Por sua vez, se a curva é côncava, temos que $\beta_{l+1} < \beta_l \Leftrightarrow \delta_l < 0$ e, portanto,

$$-u(-z_l) = -u(-\delta_l(x - \tau_l)) = \delta_l u(x - \tau_l). \quad (27)$$

Substituindo (26) e (27) na expressão (25), obtém-se o modelo

$$\eta(x|\beta_1, \tau, \delta) = \alpha_1 + \beta_1 x + \sum_{l=1}^{D-1} \delta_l u(x - \tau_l), \quad (28)$$

em que $\delta^T = (\delta_1, \dots, \delta_{D-1})$.

A fim de obter modelos continuamente diferenciáveis, os operadores de máximos presentes em (28) devem ser substituídos por aproximações suaves. Anteriormente, as aproximações para o máximo foram definidas com base nos valores das variáveis z_l 's, que são as diferenças entre regimes adjacentes, o que dificulta a interpretação das zonas de transição. No caso da regressão simples é conveniente reescrever tais aproximações como funções da variável explanatória x . Assim, considere $\delta_l \neq 0$, $\zeta_l = \gamma_l/|\delta_l|$ e $\psi_l^*(x, \tau_l, \zeta_l) = \psi_l(z_l', \gamma_l)/|\delta_l|$, com $z_l' = z_l \text{ sinal}(\delta_l) = |\delta_l|(x - \tau_l)$. Então, propõe-se a seguinte família de aproximações para o operador de máximo:

$$u_{\zeta_l}(x - \tau_l) = \psi_l^*(x, \tau_l, \zeta_l) \mathbf{I}_{\{|x - \tau_l| \leq \zeta_l\}}(x) + (x - \tau_l) \mathbf{I}_{\{|x - \tau_l| > \zeta_l\}}(x), \quad \zeta_l > 0 \forall l. \quad (29)$$

De modo análogo ao que foi verificado em (26) e (27), observa-se que

$$\delta_l u_{\zeta_l}(x - \tau_l) = \begin{cases} u_{\gamma_l}(z_l) & \text{se } \beta_{l+1} > \beta_l \Leftrightarrow \delta_l > 0, \\ -u_{\gamma_l}(-z_l) & \text{se } \beta_{l+1} < \beta_l \Leftrightarrow \delta_l < 0. \end{cases} \quad (30)$$

Além disso, a condição de diferenciabilidade imposta em (19) às funções u_{γ_l} 's é prontamente verificada para as u_{ζ_l} 's, visto que:

$$\frac{\partial \psi_l^*(x, \tau_l, \zeta_l)}{\partial x} = \frac{1}{\text{sinal}(\delta_l)} \frac{\partial}{\partial z_l'} \left(\frac{\psi_l(z_l', \gamma_l)}{\delta_l} \right) \frac{\partial z_l'}{\partial z_l} \frac{\partial z_l}{\partial (x - \tau_l)} \frac{\partial (x - \tau_l)}{\partial x} = \frac{\partial \psi_l(z_l', \gamma_l)}{\partial z_l'}.$$

Os modelos continuamente diferenciáveis são obtidos ao substituirmos os operadores de máximo em (28) por suas aproximações suaves apresentadas em (29). Os parâmetros ζ_l 's controlam a curvatura nas regiões de transições de fase, de maneira análoga a verificada para os γ_l 's. As faixas dos valores de x dentro das quais as aproximações u_{ζ_l} 's substituem o modelo abrupto em (28) são as zonas de transição, sendo denotadas por $G_l = \{x \in \mathbb{R} : \tau_l - \zeta_l \leq x \leq \tau_l + \zeta_l\}$. Logo, a quantidade $2\zeta_l$ é interpretável como o comprimento da l -ésima zona de transição (veja a Figura 5).

Figura 5 – Gráfico do modelo da regressão “bent-cable”, com aproximação $u_{\zeta_1}(x, \tau_1 = 50, \zeta_1 = 20)$ para unir os regimes $g_1(x) = 10x$ e $g_2(x) = 500$ e ψ^* calculada a partir da junção quadrática $\psi_l(z_l, \gamma_l) = (z_l + \gamma_l)^2/4\gamma_l$ de Tishler e Zang (1981a)

Demonstra-se que a regressão “bent-cable” de Chiu et al. (2006) é um caso particular dos modelos propostos nesta seção, obtido quando são unidos dois regimes lineares por meio da junção quadrática apresentada em Tishler e Zang (1981a). Seja $\psi_1(z'_1, \gamma_1) = (z'_1 + \gamma_1)^2/4\gamma_1$, de maneira que:

$$\psi_1^*(x, \tau_1, \zeta_1) = \frac{\psi_1(z'_1, \gamma_1)}{|\delta_l|} = \frac{\left[|\delta_l|(x - \tau_1) + |\delta_l|\zeta_1\right]^2}{4\zeta_1|\delta_l|^2} = \frac{(x - \tau_1 + \zeta_1)^2}{4\zeta_1}. \quad (31)$$

Então, ao empregar a aproximação suave para o máximo que resulta de (31), a expressão em (23) é obtida e, assim, a relação está provada.

2.3.2 A Inclusão de Limiares Aleatórios e Novas Funções de Junção Suave

Mudanças estruturais ao longo das variáveis explanatórias implicam a existência de diferentes processos geradores dos dados, distintos em suas regras. Zonas de transição suave entre regimes adjacentes podem ser concebidas como domínios dentro dos quais dois ou mais processos operam de maneira simultânea. Tal sobreposição de processos é verificada em muitos fenômenos naturais e pode ser explicada por meio de limiares aleatórios, que possuem valores diferentes para as observações individuais.

Admita que as transições de fase de cada uma das observações individuais são determinadas por uma sequência de variáveis aleatórias contínuas e independentes $\{T_l\}_{l=1}^{D-1}$, com funções de densidade $f_{T_l}(x|\tau_l, \gamma_l)$ e de distribuição $F_{T_l}(x|\tau_l, \gamma_l)$ e para as quais $\tau_1 < \tau_2 < \dots < \tau_{D-1}$ são os valores esperados e γ_l 's são os parâmetros de dispersão, respectivamente. Defini-se

Figura 6 – Ilustração dos diferentes regimes de um processo e das transições de fase ao longo da variável regressora x

que a l -ésima transição de fase ocorre quando o valor observado do limiar aleatório $T_l = t_l$ é menor ou igual ao valor fixado para a variável explanatória x , isso condicionado ao fato de que todas as transições de fase anteriores já ocorreram (veja as Figuras 6 e 7). Nesse momento, a variável resposta deixa de se comportar de acordo com o l -ésimo regime e passa a ser regida pelo regime seguinte. A probabilidade da l -ésima transição de fase ocorrer é $P_{T_l} = P(T_l \leq x \mid T_1 \leq x, T_2 \leq x, \dots, T_{l-1} \leq x)$.

Figura 7 – Ilustração da junção de quatro regimes lineares, dos pontos de mudança (τ_1 , τ_2 e τ_3), dos limiares aleatórios (T_1 , T_2 e T_3) e de suas distribuições

Com a inclusão de limiares aleatórios na estrutura dos modelos, surgem novas alternativas para a construção de modelos segmentados com transições de fase suaves. Nesta seção sugere-se que, ao invés de se utilizarem operadores abruptos como o sinal e o máximo, sejam usados seus valores esperados mediante os limiares aleatórios. Por exemplo, a função $\text{sinal}(x - \tau_l)$ apresenta valor igual a -1 caso a l -ésima transição de fase não ocorra e igual a 1 caso ela

ocorra. Portanto, o valor esperado da função $\text{sinal}(x - \tau_l)$ mediante T_l é:

$$E_{T_l}[\text{sinal}(x - \tau_l)] = -1 \times (1 - P_{T_l}) + 1 \times P_{T_l} = 2 P_{T_l} - 1. \quad (32)$$

Por outro lado, o operador $u(x - \tau_l)$ vale 0 se a l -ésima transição de fase não ocorre e $x - \tau_l$, caso contrário. Logo, seu valor esperado mediante T_l é:

$$E_{T_l}[u(x - \tau_l)] = (x - \tau_l) P_{T_l}. \quad (33)$$

Uma forma mais sucinta das aproximações (32) e (33) é obtida sob a pressuposição de que as regiões de transição de fase não se interceptam ao longo do eixo x , o que se obtém quando os limiares aleatórios para transições de fase sucessivas são bem distantes uns dos outros ($T_1 \ll T_2 \ll \dots \ll T_{D-1}$) com frequência. Neste caso, verifica-se que

$$P_{T_l} = P(T_l \leq x \mid T_1 \leq x, T_2 \leq x, \dots, T_{l-1} \leq x) \approx F_{T_l}(x \mid \tau_l, \gamma_l), \quad (34)$$

de maneira que as transições de fase se tornam separadas ao longo do eixo x .

Para averiguar se as transições de fase podem ser consideradas como separadas, sugere-se utilizar as restrições dispostas a seguir:

$$\begin{cases} F_{T_1}(x = \tau_2 \mid \tau_1, \gamma_1) \approx 1, \\ F_{T_r}(x = \tau_{r-1} \mid \tau_r, \gamma_r) \approx 0 \text{ e } F_{T_r}(x = \tau_{r+1} \mid \tau_r, \gamma_r) \approx 1 \quad (r = 2, \dots, D-2), \\ F_{T_{D-1}}(x = \tau_{D-2} \mid \tau_{D-1}, \gamma_{D-1}) \approx 0, \end{cases} \quad (35)$$

As condições em (35) devem ser verificadas após o ajuste do modelo, considerando-se os valores estimados dos parâmetros τ_l 's e γ_l 's. Caso violações sejam encontradas, a simplificação realizada em (34) não será cabível.

A simplificação proposta é útil, pois, na maioria das aplicações de modelos segmentados, as transições de fase podem ser distinguidas umas das outras com clareza. Não fosse esse o caso, seria impossível identificar os regimes, muito menos reconhecer qual é a forma aproximada de cada fase, e o uso de um modelo segmentado não seria adequado à situação.

A fim de construir modelos com segmentos lineares e transições de fase suaves, basta substituir os operadores abruptos em (8) e (28) pelas funções

$$E_{T_l}[\text{sinal}(x - \tau_l)] = 2 F_{T_l}(x \mid \tau_l, \gamma_l) - 1 \text{ e } E_{T_l}[u(x - \tau_l)] = (x - \tau_l) F_{T_l}(x \mid \tau_l, \gamma_l), \quad (36)$$

respectivamente. Nas seções 2.4.2 demonstra-se que, sob certas condições, essas aproximações são equivalentes às da família transicional de Bacon e Watts (1971) e Griffiths e Miller (1973). As funções em (36) serão chamadas de junções médias do tipo I no decorrer do texto, com a

finalidade de distingui-las das outras aproximações apresentadas.

As duas junções médias do tipo I produzem curvas iguais, ou seja, seu uso leva a parametrizações diferentes do mesmo modelo. A fim de demonstrar a equivalência dessas junções, considere a substituição do operador sinal em (8). O resultado é

$$\eta(x) = \theta_0 + \theta_1 x + \sum_{k=2}^D \theta_k (x - \tau_{k-1}) \left[2 F_{T_{k-1}}(x | \tau_{k-1}, \gamma_{k-1}) - 1 \right],$$

que, por meio da expressão em (9), pode ser reescrito como

$$\eta(x) = \alpha_1 + \beta_1 x + \sum_{k=2}^D \theta_k \left\{ -|\tau_1 - \tau_{k-1}| + (x - \tau_1) + (x - \tau_{k-1}) \left[2 F_{T_{k-1}}(x | \tau_{k-1}, \gamma_{k-1}) - 1 \right] \right\}.$$

Uma vez que $\tau_1 < \tau_2 < \dots < \tau_{D-1}$, tem-se que $\tau_1 - \tau_{k-1} \leq 0 \Rightarrow |\tau_1 - \tau_{k-1}| = \tau_{k-1} - \tau_1$. Além disso, sabe-se que $\theta_k = \delta_{k-1}/2$. Portanto, conclui-se que

$$\eta(x) = \alpha_1 + \beta_1 x + \sum_{l=1}^{D-1} \delta_l (x - \tau_l) F_{T_l}(x | \tau_l, \gamma_l). \quad (37)$$

O modelo em (37) é o que se obtém quando o operador de máximo em (28) é substituído pela junção média do tipo I correspondente. Apesar das junções médias do tipo I produzirem o mesmo resultado quando aplicadas em (8) e (28), a parametrização da formulação máx-mín proposta neste trabalho (expressão (28)) possui uma interpretação mais simples dos parâmetros do que aquela apresentada por (SEBER; WILD, 1989) com base na formulação sinal (expressão (8)) e, por esse motivo, recomenda-se o uso da primeira.

É importante notar que o ponto de mudança τ_l é considerado constante na obtenção das junções médias do tipo I. Nessa abordagem ele representa a abscissa sob a qual, na média, a l -ésima transição de fase ocorre. A única variável aleatória na obtenção das junções médias é o limiar T_l , o qual é usado apenas para decidir se um indivíduo em particular se comporta de acordo com um regime ou outro.

Outra abordagem possível para a construção das junções médias é obtida ao substituímos o parâmetro τ_l dentro dos operadores abruptos pelo próprio limiar aleatório T_l . Antes do cálculo das esperanças matemáticas é conveniente realizar as seguintes translações nas variáveis aleatórias: $S_l = T_l - \tau_l, \forall l$. Dessa maneira, as novas variáveis S_l 's são centradas em zero e distribuem-se de acordo com as funções $F_{S_l}(s_l | \gamma_l) = F_{T_l}(x | \tau_l, \gamma_l)$, em que $s_l = x - \tau_l$. Nesta

Figura 8 – Junções médias do Tipo I (esquerda) e II (direita) como aproximações do operador de máximo $u(x - \tau) = \max\{x - \tau, 0\}$ considerando-se a distribuição Normal para o limiar aleatório T , com $\tau = 1$ e variados valores de γ

situação, a aproximação suave para o operador de máximo é expressa por

$$\begin{aligned} E_{T_l}[u(x - T_l)] &= E_{S_l}[u(x - S_l - \tau_l)] = E_{S_l}[\max\{x - \tau_l, S_l\} - S_l] = \\ &= \int_{-\infty}^{x-\tau_l} (x - \tau_l) f_{S_l}(S_l | \gamma_l) dS_l + \int_{x-\tau_l}^{+\infty} S_l f_{S_l}(S_l | \gamma_l) dS_l - \int_{-\infty}^{+\infty} S_l f_{S_l}(S_l | \gamma_l) dS_l, \end{aligned} \quad (38)$$

em que f_{S_l} é a função de densidade da variável S_l . Caso a variável S_l seja integrável e visto que ela possui média nula, então a esperança do operador de máximo resultará em

$$E_{T_l}[u(x - T_l)] = (x - \tau_l) F_{T_l}(x | \tau_l, \gamma_l) + \int_{x-\tau_l}^{+\infty} S_l f_{S_l}(S_l | \gamma_l) dS_l. \quad (39)$$

No caso da formulação sinal, deve-se lembrar que a função sinal surge por meio da fatoração do operador módulo em $(x - \tau_l) \text{ sinal}(x - \tau_l)$ (ver a expressão (4)). Logo, a junção média é dada por

$$\begin{aligned} E_{T_l}[x - T_l] &= E_{S_l}[(x - S_l - \tau_l) \text{ sinal}(x - S_l - \tau_l)] \\ &= \int_{-\infty}^{x-\tau_l} (x - S_l - \tau_l) f_{S_l}(S_l | \gamma_l) dS_l - \int_{x-\tau_l}^{+\infty} (x - S_l - \tau_l) f_{S_l}(S_l | \gamma_l) dS_l \\ &= (x - \tau_l) \left[2 F_{T_l}(x | \tau_l, \gamma_l) - 1 \right] - \int_{-\infty}^{x-\tau_l} S_l f_{S_l}(S_l | \gamma_l) dS_l + \int_{x-\tau_l}^{+\infty} S_l f_{S_l}(S_l | \gamma_l) dS_l. \end{aligned} \quad (40)$$

As aproximações em (39) e (40) são chamadas de junções médias do tipo II no decorrer do texto.

Observe que as junções médias do tipo I e II são similares, com exceção dos primeiros

momentos incompletos de S_l somados e subtraídos das junções do tipo II. Pode-se verificar que as junções do tipo II apresentam comportamento assintótico igual ao das junções do tipo I, mas nas proximidades dos pontos de mudança seus comportamentos são bem distintos (ver Figura 8). O comportamento das junções do tipo I nas proximidades dos pontos de mudança é similar ao das funções transicionais propostas por Bacon e Watts (1971). Por outro lado, o comportamento das junções do tipo II é mais parecido com o das curvas hiperbólicas de Griffiths e Miller (1973) ou da regressão “bent-cable” de Chiu et al. (2006).

O uso de junções média do tipo II é o mais indicado para a prática, pois neste caso o modelo obtido é a esperança da resposta condicionada a pontos de mudança aleatórios T_l 's. No entanto, além das expressões complicadas para o modelo que as junções médias do tipo II acarretam, elas trazem alguns problemas a mais: se os pontos de mudança são tratados como quantidades aleatórias, então os parâmetros ligados aos regimes devem ser aleatórios, também. A razão para isso é que não se pode fazer com que dois regimes lineares se interceptem em pontos de mudança diferentes, a não ser que pelo menos um dos regimes tenha o valor de algum parâmetro alterado. Dessa forma, a fim de garantir uma interpretação exata dos parâmetros, seria preciso incluir efeitos aleatórios nos pontos de mudança e nos regimes, ao mesmo tempo, o que na inferência clássica requer metodologias específicas para a análise de dados longitudinais.

As junções médias do tipo I podem ser vistas como uma simplificação útil das junções do tipo II, especialmente quando a saliência característica das funções transicionais é observada no conjunto de dados. Afinal, o comportamento assintótico do modelo é o mesmo. Além do mais, qualquer junção que se escolha terá sua forma verdadeira mascarada pela variabilidade ao acaso, tal como foi relatado por Bacon e Watts (1971).

2.3.3 Estimação por Máxima Verossimilhança

A fim de estimar os valores dos parâmetros e realizar inferências, considera-se o modelo não linear clássico

$$y_i = \eta(x_i | \phi) + \epsilon_i, \quad i = 1, \dots, n,$$

em que: y_i é a i -ésima observação da variável resposta Y , condicionada ao valor x_i da variável explanatória; ϕ é o vetor de parâmetros associados a parte fixa do modelo, sendo

$$\phi^T = (\beta_1^T, \delta^T, \tau^T, \gamma^T) \quad \text{ou} \quad (\theta^T, \tau^T, \gamma^T),$$

dependendo do modelo a ser ajustado; e ϵ_i é o erro aleatório associado à i -ésima observação. Assumem-se erros aleatórios independentes, identicamente distribuídos e normais com esperança

nula e variância constante, ou seja, $\epsilon_i \sim N(0, \sigma^2)$, com $\sigma^2 > 0$.

Neste caso, a função de verossimilhança é expressa por

$$L(\phi, \sigma^2) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{[y_i - \eta(x_i | \phi)]^2}{2\sigma^2} \right\}, \quad (41)$$

sendo seu logaritmo equivalente a

$$\ln L(\phi, \sigma^2) = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} S(\phi), \quad (42)$$

em que $S(\phi)$ é a soma de quadrados dos erros, expressa por

$$S(\phi) = \sum_{i=1}^n (y_i - \eta(x_i | \phi))^2. \quad (43)$$

A estimativa de máxima verossimilhança para o vetor de parâmetros ϕ , denotada por $\hat{\phi}$, é obtida ao maximizarmos $\ln L$ sobre o espaço paramétrico Φ , para um valor fixo de σ^2 . A estimativa de máxima verossimilhança para σ^2 é calculada posteriormente à obtenção de $\hat{\phi}$, por meio da seguinte expressão

$$\hat{\sigma}^2 = \frac{S(\hat{\phi})}{n},$$

em que $S(\hat{\phi})$ é a soma de quadrados dos resíduos. Isso é possível porque neste caso as estimativas $\hat{\phi}$ e $\hat{\sigma}^2$ são independentes (SEBER; WILD, 1989). Consequentemente, por meio da expressão (42) pode-se concluir que as estimativas de máxima verossimilhança para ϕ coincidem com as obtidas pelo método dos mínimos quadrados.

Uma vez que os modelos a serem ajustados possuem parâmetros não lineares, inexistem forma fechada para os estimadores. Vários procedimentos podem ser empregados para maximizar a função $\ln L$, tais como métodos heurísticos ou baseados no cálculo de gradientes. Os métodos heurísticos geralmente não requerem o cálculo de gradientes e, dessa forma, são aplicáveis na otimização de funções não diferenciáveis. Comumente eles são baseados em dinâmicas de trocas aleatórias dos elementos de ϕ que incrementam gradualmente o valor da função de verossimilhança, até que se atinja um critério de parada. Para os métodos heurísticos, no entanto, não há garantias da convergência ao ótimo. Os algoritmos heurísticos encontram as estimativas em $\hat{\phi}$ por meio de processos aleatórios e geram resultados diferentes a cada vez que são executados. Essa variabilidade pode inflacionar a variância dos estimadores e é difícil mensurar qual seria seu impacto sobre as inferências, o que faz com que métodos de estimação determinísticos sejam preferíveis.

Os procedimentos mais usados na estimação de parâmetros de modelos não lineares são

Figura 9 – Esquema para a obtenção de estimativas iniciais dos parâmetros na estimação de modelos segmentados com regimes lineares e transições suaves. Dados adaptados de Schneeberger (2009) referentes à produtividade do trigo de inverno (100 kg/ha) em resposta à adubação nitrogenada (kg/ha) mediante dois estados de nutrição distintos, a saber, de deficiência mineral por N e de consumo de luxo

iterativos, consistem na atualização de um vetor de estimativas iniciais ϕ^0 e requerem o cálculo de gradientes. O gradiente é utilizado a cada passo para determinar a direção e o sentido que se deve percorrer no espaço paramétrico Φ , partindo de um ponto, a fim de maximizar mais rapidamente a função $\ln L$. A cada iteração percorre-se um tamanho de passo pré-especificado e, em seguida, o gradiente é recalculado para indicar a direção de subida mais rápida. O processo é repetido inúmeras vezes, até que o gradiente se aproxime suficientemente de zero, garantindo que um ótimo local (ou ponto de sela) tenha sido atingido. Os métodos de busca que empregam o gradiente são determinísticos, no sentido de que se obtém o mesmo resultado quando o vetor de valores iniciais ϕ^0 utilizado é o mesmo. Entretanto, o uso de valores iniciais distintos podem conduzir a ótimos locais diferentes e no processo de estimação o que importa é o ótimo global. Logo, a escolha do vetor de valores iniciais ϕ^0 é crucial para a estimação dos parâmetros de modelos não lineares. Aconselha-se o uso de métodos de estimação determinísticos e heurísticos, em concomitância, e a comparação dos resultados, a fim de se evitar a obtenção de ótimos locais.

Conforme Watts e Bacon (1974), os valores iniciais para o ajuste de modelos segmentados podem ser facilmente obtidos por meio de análise gráfica, já que nestes modelos os parâmetros possuem interpretações geométricas claras. Com o auxílio de um diagrama de dispersão, é fácil introduzir um procedimento para encontrá-los visualmente. O primeiro passo é delimitar as

observações de cada regime por faixas compostas por retas paralelas. As abscissas nas quais as faixas de diferentes regimes encontram-se são os valores iniciais para os pontos de mudança. Por sua vez, o coeficiente angular das retas que as formam determinam os coeficientes angulares de cada um dos regimes. Já a média dos interceptos das retas que formam a faixa para o primeiro regime definem o intercepto do mesmo. Por fim, os valores dos parâmetros de suavização podem ser escolhidos visualmente, ao construirmos diferentes curvas de resposta e verificarmos a concordância entre as observações e a transição desenhada (ver Figura 9). Os valores iniciais para os parâmetros de suavização não são influentes no resultado dos ajustes (WATTS; BACON, 1974).

A análise visual é subjetiva e depende da atenção de um observador. Neste trabalho propõe-se um método objetivo e automático para a obtenção de estimativas iniciais, o qual requer, como conhecimento prévio, somente o número de regimes D . O método consiste em duas etapas:

- ◇ Etapa 1: nesta fase são ajustados diversos modelos abruptos e, a partir do melhor ajuste, são obtidos os valores iniciais para os coeficientes dos regimes e pontos de mudança. Considere a existência de um número conhecido de regimes D ($D \geq 2$) e admita que o delineamento experimental (ou amostral) tenha N pontos de suporte, com $N \geq 2D$. Propõe-se ajustar todos os modelos segmentados abruptos possíveis considerando-se as combinações dos N pontos de suporte, D a D , como sendo os próprios pontos de mudança. Obviamente, o máximo e o mínimo dos pontos de suporte não podem ser considerados como pontos de mudança. Além disso, dois ou mais pontos de mudança não podem se situar sobre o mesmo ponto de suporte. Assim, o número total de regressões a serem realizadas nesta etapa é

$$N_R = \binom{N-2}{D-1}.$$

Durante o processo, os modelos segmentados abruptos são ajustados por meio de regressões lineares com restrições de continuidade, uma vez que os pontos de mudança são considerados conhecidos. Após o ajuste de todos os N_R modelos segmentados possíveis, o modelo que produz o menor quadrado médio do resíduo é escolhido e as estimativas de seus parâmetros são usadas como valores iniciais para os coeficientes associados aos regimes.

Alternativamente, ao invés de se fazer uso dos pontos de suporte do delineamento, pode-se usar uma grade de pontos situados entre os valores mínimo e máximo da variável

explanatória para representar os pontos de mudança. Mas é preciso lembrar que tal grade não poderá incluir muitos pontos, uma vez que o número de regressões a serem realizadas N_R cresce exponencialmente com o número de pontos N .

- ◊ Etapa 2: dados os valores dos coeficientes dos regimes e dos pontos de mudança, são obtidos os melhores valores iniciais para os parâmetros de suavização por meio da minimização da soma quadrática dos erros em (43) ou, analogamente, pela maximização do logaritmo da verossimilhança em (42), para algum $\sigma^2 > 0$.

Se desejar, o pesquisador pode suprimir esta etapa e fixar valores razoavelmente pequenos para os parâmetros de suavização. No entanto, tais valores não podem ser pequenos a ponto de distorcer a forma suave que as junções devem possuir, pois isso pode provocar problemas numéricos no cálculo do gradiente durante a estimação.

Ao final das duas etapas, os valores obtidos são utilizados como ponto de partida ϕ^0 para o método iterativo de estimação dos parâmetros.

Neste trabalho faz-se uso do método quasi-Newton conhecido como BFGS (BROYDEN, 1970; FLETCHER, 1970; GOLDFARB, 1970; SHANNO, 1970), implementado na função *optim* do “software” estatístico R (R CORE TEAM, 2015). O método de Newton para a busca de extremos relativos se baseia na expansão da função objetivo que queremos minimizar $f(\psi)$ em série de Taylor de segunda ordem ao redor de ϕ^k :

$$f(\phi^k + \Delta\phi) \approx f(\phi^k) + \nabla f(\phi^k)' \Delta\phi + \frac{1}{2} \Delta\phi' \mathbf{H} \Delta\phi, \quad (44)$$

em que $\Delta\phi = \phi - \phi^k$, $\nabla f(\phi^k)$ é o vetor gradiente da função $f(\phi)$ avaliado no ponto ϕ^k e $\mathbf{H}$ é a sua matriz Hessiana. Assim, o gradiente da função aproximada por série de Taylor é nulo quando

$$\nabla f(\phi^k + \Delta\phi) \approx \nabla f(\phi^k) + \mathbf{H} \Delta\phi = 0,$$

o que resulta na equação de Newton

$$\Delta\phi = -\mathbf{H}^{-1} \nabla f(\phi^k),$$

isso caso a matriz Hessiana seja inversível.

A equação de Newton indica em quais direção e sentido deve-se variar o vetor ϕ para que seja possível otimizar a aproximação de $f(\phi)$ em (44). O método de Newton para a otimização consiste em um procedimento iterativo em que a cada iteração essa direção é recalculada e o vetor de estimativas é atualizado. Seja ϕ^k o vetor de estimativas dos parâmetros na k -ésima

iteração, então nessa etapa calcula-se

$$\phi^{k+1} = \phi^k - \rho_k \mathbf{H}_k^{-1} \nabla f(\phi^k), \quad \rho_k > 0,$$

em que ϕ^{k+1} é o vetor de estimativas atualizadas na iteração $k + 1$, $\mathbf{H}_k$ é a matriz Hessiana para a função $f(\phi^k)$ e ρ_k é o tamanho do passo estabelecido pelas condições de Wolfe (1969,1971).

Os métodos quasi-Newton são adaptações úteis quando a matriz Hessiana (ou sua inversa) não é disponível ou então requer um custo computacional muito grande. Nos métodos quasi-Newton, ao invés de calcular a matriz Hessiana a cada iteração, é realizada apenas a atualização de seu valor, com o auxílio da função gradiente. Outra vantagem é que nesses métodos não é preciso inverter $\mathbf{H}$, pois com os cálculos de atualização de $\mathbf{H}$ é possível obter diretamente uma estimativa de $\mathbf{H}^{-1}$. No método BFGS, por exemplo, as matrizes $\mathbf{H}$ e $\mathbf{H}^{-1}$ são atualizadas por meio das seguintes fórmulas

$$\mathbf{H}_{k+1} = \mathbf{H}_k + \frac{\mathbf{g}_k \mathbf{g}_k'}{\mathbf{g}_k' \Delta \phi_k} - \frac{\mathbf{H}_k \Delta \phi_k (\mathbf{H}_k \Delta \phi_k)'}{\Delta \phi_k' \mathbf{H}_k \Delta \phi_k}$$

e

$$\mathbf{H}_{k+1}^{-1} = \left(\mathbf{I} - \frac{\mathbf{g}_k \Delta \phi_k'}{\mathbf{g}_k' \Delta \phi_k} \right)' \mathbf{H}_k^{-1} \left(\mathbf{I} - \frac{\mathbf{g}_k \Delta \phi_k'}{\mathbf{g}_k' \Delta \phi_k} \right) + \frac{\Delta \phi_k \Delta \phi_k'}{\mathbf{g}_k' \Delta \phi_k},$$

nas quais $\mathbf{g}_k = \nabla f(\phi^{k+1}) - \nabla f(\phi^k)$ e $\Delta \phi_k = \phi^{k+1} - \phi^k$. Uma propriedade importante do método BFGS é o fato de que uma vez que $\mathbf{H}_k$ é positiva definida e ρ_k atende as condições de Wolfe, então $\mathbf{H}_{k+1}$ também é positiva definida.

Como ponto de partida, pode-se definir $\mathbf{H}_0 = \mathbf{I}$, de maneira que no início da busca o algoritmo se comporte como um método de descida pelo gradiente e, após algumas atualizações, a matriz $\mathbf{H}_k$ se aproxime da Hessiana verdadeira e a curvatura da função $f(\phi^k)$ seja levada em conta. O procedimento iterativo cessa quando o valor absoluto do gradiente $|\nabla f(\phi^k)|$ torna-se menor que um valor de tolerância pré-estipulado.

O método BFGS foi introduzido para otimizar funções com derivadas contínuas de primeira e segunda ordem, mas, devido às aproximações numéricas, pode ser utilizado para encontrar extremos de funções que não são suaves, apresentando inclusive um bom desempenho nesta tarefa. No entanto, a convergência do método só é garantida se a função objetivo $f(\phi)$ tiver uma expansão de Taylor quadrática ao redor do ponto ótimo $\hat{\phi}$.

Para a obtenção de estimativas de máxima verossimilhança, a função critério a se minimizar é definida como

$$f(\phi) = -\ln L(\phi, \sigma^2).$$

Uma vez que a otimização nesse caso é irrestrita, os coeficientes de suavização presentes nos modelos segmentados devem ser reparametrizados para assumirem valores na reta real, o que pode ser feito da seguinte maneira

$$w_l = \ln(\gamma_l) \text{ ou } \ln(\zeta_l), \quad l = 1, \dots, D, \quad (45)$$

seguindo a sugestão de (SEBER; WILD, 1989).

Neste trabalho faz-se uso do algoritmo BFGS implementado na função *optim* do R, em associação com a função *mle2* do pacote *bbmle* (BOLKER, 2015), fixando-se o número máximo de iterações em 10.000. A função *mle2* é mais robusta do que a *mle* presente no pacote *stats4*, garantindo maior percentual de convergência no ajuste de modelos segmentados com transições suaves.

2.3.4 Simulação de Dados

Neste trabalho é conduzido um estudo de simulação de dados com a finalidade de ilustrar o uso dos modelos segmentados com transições suaves abordados e avaliar o desempenho dos mesmos nos processos de estimação e inferência. Como motivação prática, as simulações realizadas baseiam-se no estudo de adubação apresentado por Schneeberger (2009), o qual apresenta dados da produtividade do trigo de inverno (em 100 kg/ha) em função da dose de adubação nitrogenada (em kg/ha). A escala de doses utilizada neste experimento abrange três regimes nutricionais distintos do trigo, os de deficiência mineral, consumo de luxo e toxidez. Nesses regimes, as relações funcionais existentes entre a produtividade e a dose de adubo são diferentes. Os autores da área de nutrição mineral de planta geralmente representam a curva de produtividade por meio de três segmentos de retas, os quais são ligados por junções suaves (FAQUIN, 2002; HAVLIN et al., 2004; MALAVOLTA, 1993).

A simulação de dados é conduzida considerando-se a união dos regimes lineares expressos abaixo:

$$g_1(x) = 64,5 + 0,50x, \quad g_2(x) = 101,75 \quad \text{e} \quad g_3(x) = 153,39 - 0,26x$$

e, conseqüentemente, os pontos de mudança são $\tau_1 = 74,5$ e $\tau_2 = 198,6$. Os conjuntos de dados simulados são gerados por meio de três modelos lineares segmentados com transições suaves, a saber:

1.) Modelo linear segmentado com junções médias do tipo I baseadas na distribuição Logística:

$$\eta(x|\beta_1, \tau, \delta) = \alpha_1 + \beta_1 x + \sum_{l=1}^2 \delta_l (x - \tau_l) F_{T_l}(x|\tau_l, \gamma_l), \quad (46)$$

em que

$$F_{T_l}(x|\tau_l, \gamma_l) = \frac{1}{1 + e^{-\frac{(x-\tau_l)}{\gamma_l}}}, \quad (47)$$

é a função de distribuição acumulada Logística.

2.) Modelo linear segmentado com junções médias do tipo II baseadas na distribuição Logística:

$$\eta(x|\beta_1, \tau, \delta) = \alpha_1 + \beta_1 x + \sum_{l=1}^2 \delta_l \left[(x - \tau_l) F_{T_l}(x|\tau_l, \gamma_l) + \int_{x-\tau_l}^{+\infty} S_l f_{S_l}(S_l|\gamma_l) dS_l \right], \quad (48)$$

em que a função F_{T_l} é a acumulada da Logística expressa em (47) e

$$f_{S_l}(S_l|\gamma_l) = \frac{e^{-\frac{S_l}{\gamma_l}}}{\gamma_l (1 + e^{-\frac{S_l}{\gamma_l}})^2}$$

é a função de densidade da Logística centrada no zero.

3.) Modelo da regressão “bent-cable” estendida com junção polinomial de 4ª grau de Zang (1980):

$$\eta(x|\beta_1, \tau, \delta) = \alpha_1 + \beta_1 x + \sum_{l=1}^2 \delta_l u_{\zeta_l}(x - \tau_l), \quad (49)$$

em que u_{ζ_l} é expresso por (29).

Os modelos dispostos em (46, 48 e 49) são obtidos por meio da nova formulação máx-mín, proposta na seção 2.3.1 para o caso de regressão simples. A formulação máx-mín é escolhida para o estudo porque apresenta parâmetros mais fáceis de se interpretar por meio dos coeficientes associados aos regimes. Entretanto, as junções médias dos tipos I e II também poderiam ser aplicadas aos modelos da formulação sinal em (8). Dos modelos máx-mín apresentados, apenas o modelo da regressão “bent-cable” estendida não possui uma reparametrização na formulação sinal.

A distribuição Logística é escolhida para a construção das junções médias por dois motivos: em primeiro lugar, ela apresenta função de distribuição acumulada com forma fechada, o que simplifica os cálculos; além disso, na seção 2.4.2 demonstra-se que o uso da junção média do tipo I baseada na Logística produz um modelo equivalente ao transicional de Bacon e Watts

(1971) por meio da tangente hiperbólica.

A teoria da nutrição mineral de plantas determina que, na fase de consumo de luxo, a resposta das plantas à adubação é muito pequena ou até mesmo nula (Macy, 1936). Dessa forma, o segundo regime g_2 é representado nesse estudo por uma reta de coeficiente angular nulo, ou seja, uma constante. Consequentemente, têm-se que $\delta_1 = -\beta_1$ e $\delta_2 = \beta_3$. No estudo, assumem-se transições de fase suaves, de maneira que os parâmetros γ_1 e γ_2 são fixados em 10,0 e os parâmetros ζ_1 e ζ_2 são fixados em 15,00. Esses valores foram considerados porque o ajuste prévio dos modelos ao conjunto de dados original de Schneeberger (2009) evidenciou que, para se obter o mesmo grau de suavização nas curvas de resposta, a razão entre os valores de γ e ζ deve ser fixada em $2/3$, aproximadamente.

Para que seja possível obter informação sobre os parâmetros de suavização, faz-se necessária uma coleta de dados que elucide a forma da região de transição. Os parâmetros de suavização são associados à curvatura dos modelos na região de transição e, no caso das junções médias, representam a dispersão da distribuição dos limiares aleatórios. Dessa forma, dificuldades em estimá-los são esperadas. Acredita-se que, a fim de se obter informação relevante com respeito a esses parâmetros, o delineamento deve ter muitos pontos de suporte repetidos ao longo das regiões de transição. Isso porque o uso de muitos pontos de suporte permitirá uma descrição satisfatória da forma das regiões de transição, enquanto as repetições impedirão que a variabilidade ao acaso a mascare.

O experimento apresentado por Schneeberger (2009) apresenta 52 unidades experimentais divididas entre 13 pontos de suporte, em uma escala de doses que varia de 0 a 300 kg/ha de adubo. Nesse estudo de simulação utiliza-se um delineamento com o mesmo esforço experimental do apresentado por Schneeberger (2009), mas, no entanto, o número de pontos de suporte é duplicado: são usados 26 pontos de suportes igualmente espaçados ao longo da região de delineamento, sendo esses repetidos duas vezes.

A distribuição amostral dos estimadores dos parâmetros de cada um dos modelos é estudada por meio de 1.000 conjuntos de dados simulados. Durante o processo de simulação de dados, o modelo em estudo é assumido como verdade e amostras das respostas são obtidas por meio da expressão

$$y_i = \eta(x_i | \phi^*) + \epsilon_i, \quad i = 1, 2, \dots, n,$$

na qual: ϕ^* é o vetor dos valores verdadeiros para os p parâmetros existentes na parte fixa do modelo e ϵ_i 's são valores independentes e identicamente distribuídos de acordo com a distribuição Normal, com esperança nula e desvio padrão $\sigma = 1,5 \times 100$ kg/ha.

Estatísticas descritivas do conjunto de estimativas de máxima verossimilhança para cada parâmetro são usadas para informar sobre o desempenho dos modelos nos processos de estimação e inferência. Sejam ϕ_m o m -ésimo parâmetro do vetor ϕ (com $m = 1, 2, \dots, p$) e $\hat{\phi}_{mt}$ a estimativa de máxima verossimilhança para esse parâmetro na t -ésima das $B = 1000$ amostras simuladas. Então, as estatísticas calculadas são

a) Média:

$$\bar{\hat{\phi}}_m = \frac{1}{B} \sum_{t=1}^B \hat{\phi}_{mt};$$

b) Coeficiente de Assimetria:

$$sk(\hat{\phi}_m) = \frac{\frac{1}{B} \sum_{t=1}^B (\hat{\phi}_{mt} - \bar{\hat{\phi}}_m)^3}{\left[\frac{1}{B-1} \sum_{t=1}^B (\hat{\phi}_{mt} - \bar{\hat{\phi}}_m)^2 \right]^{3/2}};$$

c) Coeficiente de Curtose:

$$k(\hat{\phi}_m) = \frac{\frac{1}{B} \sum_{t=1}^B (\hat{\phi}_{mt} - \bar{\hat{\phi}}_m)^4}{\left[\frac{1}{B} \sum_{t=1}^B (\hat{\phi}_{mt} - \bar{\hat{\phi}}_m)^2 \right]^2} - 3;$$

d) Erro-Padrão:

$$EP(\hat{\phi}_m) = \sqrt{\frac{1}{B-1} \sum_{t=1}^B (\hat{\phi}_{mt} - \bar{\hat{\phi}}_m)^2};$$

e) Viés e Erro Quadrático Médio:

$$b(\hat{\phi}_m) = \bar{\hat{\phi}}_m - \phi_m^* \text{ e } EQM(\hat{\phi}_m) = b^2(\hat{\phi}_m) + EP^2(\hat{\phi}_m);$$

f) Erro Médio (Absoluto e Relativo):

$$RMSE(\hat{\phi}_m) = \sqrt{EQM(\hat{\phi}_m)} \text{ e } RMSE\%(\hat{\phi}_m) = \frac{RMSE(\hat{\phi}_m)}{|\phi_m^*|} \times 100\%.$$

2.4 Resultados e Discussão

Os resultados deste trabalho são divididos em duas partes: na seção 2.4.1, os resultados do estudo de simulação de dados são apresentados e discutidos; já a seção 2.4.2 traz alguns resultados teóricos acerca das junções médias do tipo I e o relacionamento dessas junções com a família de aproximações transicionais de Bacon e Watts (1971).

2.4.1 Resultados da Simulação de Dados

Os dados são simulados neste trabalho considerando-se três regimes nutricionais distintos, de maneira que a forma da relação existente entre a produtividade e a dose de adubo é complicada demais para ser explicada por um modelo de uso frequente na adubação, tais como os descritos por Cerrato e Blackmer (1990). A forma particular que a relação apresenta, a de uma banheira invertida, é esperada pela teoria de nutrição mineral de plantas (FAQUIN, 2002; HAVLIN et al., 2004; MALAVOLTA, 1993), o que indica a adequação de um modelo linear segmentado. As Figuras 10 e 11 ilustram os ajustes dos modelos segmentados suaves propostos neste trabalho aos conjuntos de dados simulados (ver Tabela 1).

Tabela 1 – Conjuntos de dados simulados sobre a produtividade (100 kg/ha) do trigo de inverno em função da dose de nitrogênio (kg/ha), gerados a partir dos modelos lineares segmentados com junções médias do tipo I e II baseadas na distribuição Logística e o da regressão “bent-cable” estendida (BC Estendida)

Doses	Produtividades (Junção I)		Produtividades (Junção II)		Produtividades (BC Estendida)	
0	62,22	64,67	63,71	66,85	64,86	64,18
12	69,82	70,22	69,40	70,11	68,08	70,43
24	75,92	76,61	76,73	71,52	75,67	76,27
36	84,65	82,30	82,32	83,49	82,96	83,96
48	90,95	90,68	88,44	89,72	90,47	88,05
60	96,07	94,04	94,07	91,61	93,80	94,93
72	102,10	101,66	98,07	96,09	99,88	99,75
84	106,55	104,21	98,65	100,99	101,38	100,50
96	103,00	102,48	103,07	100,58	103,44	102,21
108	102,03	100,54	104,98	101,33	102,75	99,02
120	100,58	99,16	101,62	100,38	102,69	105,30
132	101,84	102,68	101,10	99,96	101,37	102,44
144	103,66	99,96	102,18	102,90	99,95	101,80
156	101,10	101,24	103,83	104,61	100,41	104,92
168	102,73	101,25	101,11	100,52	103,51	102,36
180	104,52	99,88	101,74	102,03	100,62	102,01
192	101,43	104,79	101,62	98,67	100,49	100,69
204	103,27	101,40	98,66	101,74	100,37	98,74
216	98,51	96,05	94,50	97,94	96,64	96,99
228	95,66	93,35	93,80	93,17	95,30	94,40
240	93,42	91,02	90,74	93,04	88,62	91,22
252	89,25	85,32	88,86	84,01	86,38	87,75
264	85,71	84,06	86,56	86,90	84,76	84,37
276	78,45	81,03	80,17	82,00	79,53	82,51
288	79,19	78,74	79,91	77,24	80,26	78,43
300	77,29	76,99	76,16	75,07	75,62	76,45

Observa-se que os modelos estudados são flexíveis o bastante para modelar os dados ao longos dos três regimes nutricionais e das duas transições de fase. Eles nos permitem estimar com precisão os parâmetros associados a cada um dos regimes, além dos pontos de mudança τ_1 e τ_2 (Tabelas 2 e 3). Os modelos obtidos a partir das junções médias ainda possibilitam estimar as probabilidades de transição de fase (de acordo com os valores de x), uma vez que são construídos por meio de funções de distribuição acumuladas. Por exemplo, por meio dos

Figura 10 – Em (a) e (b) são apresentados os gráficos dos ajustes dos modelos lineares segmentados (linha contínua) com junções médias do tipo I e II, respectivamente. Em (c) e (d) apresentam-se as probabilidades estimadas para as transições de fase do 1º ao 2º e do 2º ao 3º regimes nos casos das junções médias do tipo I e II, respectivamente

ajustes apresentados na Figura 10, estima-se que as doses de adubo

$$x_{D_1} = 101,05 \text{ kg/ha e } x_{D_2} = 107,18 \text{ kg/ha}$$

são aquelas que garantem 95% de suficiência por nitrogênio, conforme os modelos com junções médias do tipo I e II, respectivamente. Isso quer dizer que, mediante tais doses, 95% das amostras já transitaram do regime de deficiência mineral por nitrogênio ao estado de nutrição adequado (regime do consumo de luxo). Pode-se dizer ainda, que mediante as doses

$$x_{T_1} = 222,30 \text{ kg/ha e } x_{T_2} = 228,10 \text{ kg/ha,}$$

95% das amostras já transitaram do regime do consumo de luxo ao de toxidez, conforme as junções médias do tipo I e II, respectivamente.

As junções médias do tipo I e II são obtidas por meio da pressuposição de transições de fase separadas. Portanto, as interpretações acerca das transições de fase são válidas apenas quando as condições em (35) são satisfeitas. No caso dos ajustes apresentados na Figura 10, observa-se que:

- Para a junção média do tipo I:

$$F_{T_1}(x = \hat{\tau}_1 | \hat{\tau}_1, \hat{\gamma}_1) = 0,999998 \text{ e } F_{T_2}(x = \hat{\tau}_1 | \hat{\tau}_2, \hat{\gamma}_2) = 6,34 \times 10^{-8};$$

- Para a junção média do tipo II:

$$F_{T_1}(x = \hat{\tau}_2 | \hat{\tau}_1, \hat{\gamma}_1) = 0,999989 \text{ e } F_{T_2}(x = \hat{\tau}_1 | \hat{\tau}_2, \hat{\gamma}_2) = 1,2 \times 10^{-5}.$$

Logo, pode-se afirmar que para esses ajustes tais condições são praticamente atendidas.

O modelo da regressão “bent-cable” estendida, entretanto, não permite estimar as probabilidades de transição de fase, mas, em compensação, possibilita a delimitação das faixas de transição de fase. Por meio do ajuste apresentado na Figura 11, obtêm-se os intervalos

$$G_1 = [\hat{\tau}_1 - \hat{\zeta}_1, \hat{\tau}_1 + \hat{\zeta}_1] = [44,39; 101,40] \text{ e } G_2 = [\hat{\tau}_2 - \hat{\zeta}_2, \hat{\tau}_2 + \hat{\zeta}_2] = [159,06; 229,88],$$

que são as faixas, em kg/ha, correspondentes as 1ª e 2ª transições de fase. De acordo com o ajuste, verifica-se que doses entre 101,40 e 159,06 kg/ha são associadas ao estado nutricional adequado.

Os métodos de obtenção de valores iniciais e estimação apresentam bons resultados diante do cenário de simulação proposto. Neste estudo, o critério de convergência do algoritmo de estimação foi satisfeito em 100% das execuções, para ambos os modelos. No entanto, é

Figura 11 – Gráfico do ajuste do modelo de regressão “bent-cable” estendida (linha contínua) com junção polinomial de quarta ordem proposta por Zang (1980), curvas dos regimes lineares estimados (linha tracejada) e faixas de transição de fase (área hachurada) para um dos 1.000 conjuntos de dados simulados

Tabela 2 – Estimativas dos parâmetros (Est.) e seus erros-padrão (EP) aproximados para os ajustes dos modelos lineares segmentados obtidos pela regressão “bent-cable” (BC) estendida e por meio das junções médias do tipo I e II aos dados simulados da Tabela 1

Parâmetros*	Modelos Ajustados					
	Junção I		Junção II		BC Estendida	
	Est.	EP	Est.	EP	Est.	EP
α_1	63,615	0,758	64,249	0,875	63,811	0,659
β_1	0,526	0,018	0,499	0,033	0,527	2,199
δ_2	-0,261	0,012	-0,257	0,019	-0,252	0,119
τ_1	72,494	1,617	75,962	3,769	72,897	2,199
τ_2	199,70	2,789	196,630	5,339	194,468	3,352
$w_1^\dagger$	2,272	0,256	2,361	0,321	3,350	0,307
$w_2^\dagger$	2,038	0,738	2,369	0,599	3,567	0,356

*: $\alpha_1 = 64,50$, $\beta_1 = 0,50$, $\delta_2 = -0,26$, $\tau_1 = 74,50$ e $\tau_2 = 198,60$

†: $w_l = \ln(\gamma_l)$ ou $\ln(\zeta_l)$, $\ln(\gamma_l) = 2,303$ e $\ln(\zeta_l) = 2,708$, $l = 1, 2$

importante esclarecer alguns fatos com respeito à estimação: em primeiro lugar, o método de estimação parte apenas de um ponto ϕ^0 no espaço paramétrico e pode estacionar em um ótimo local; depois, deve-se considerar o fato do delineamento apresentar muitos pontos de suporte e repetições, o que torna a superfície da função $\ln L$ mais suave e, conseqüentemente, facilita a estimação; por fim, observa-se que a implementação do algoritmo de otimização é um fator chave para o sucesso do método. Neste estudo utiliza-se a função *mle2* do pacote *bbmle* do R, a qual é mais robusta do que a função *mle* do pacote *stats4*.

Tabela 3 – Medidas de resumo da distribuição amostral para os estimadores de máxima verossimilhança dos parâmetros dos modelos lineares segmentados propostos: modelos máx-mín com junção média dos tipos I e II baseadas na distribuição Logística e modelo da regressão “bent-cable” estendida com junção polinomial de 4º grau de Zang (1980). Resultados obtidos a partir de 1.000 conjuntos de dados simulados

ϕ_m	$\hat{\phi}_m$	$sk(\hat{\phi}_m)$	$k(\hat{\phi}_m)$	$EP(\hat{\phi}_m)$	$b(\hat{\phi}_m)$	$EQM(\hat{\phi}_m)$	$RMSE(\hat{\phi}_m)$	$RMSE\%(\hat{\phi}_m)$
Junção I								
α_1	64,423	-0,035	0,087	0,764	-0,077	0,589	0,768	1,190
β_1	0,506	0,251	0,197	0,019	0,006	0,000	0,020	4,094
δ_2	-0,264	-0,049	0,250	0,013	-0,004	0,000	0,014	5,274
τ_1	74,020	-0,154	0,047	1,904	-0,480	3,856	1,964	2,636
τ_2	198,947	-0,073	0,455	3,505	0,347	12,404	3,522	1,773
γ_1	7,218	-0,232	-1,023	4,555	-2,782	28,493	5,338	53,379
γ_2	6,450	1,054	1,091	6,052	-3,550	49,227	7,016	70,162
σ^2	1,988	0,607	0,697	0,440	-0,262	0,262	0,512	22,759
Junção II								
α_1	64,503	0,023	-0,210	0,830	0,003	0,689	0,830	1,287
β_1	0,504	0,780	1,365	0,036	0,004	0,001	0,036	7,150
δ_2	-0,259	-0,829	1,358	0,020	0,001	0,000	0,020	7,634
τ_1	74,172	-0,505	0,807	3,840	-0,328	14,853	3,854	5,173
τ_2	198,062	0,324	0,774	5,570	-0,538	31,314	5,596	2,818
γ_1	9,458	-0,296	0,473	3,877	-0,542	15,322	3,914	39,144
γ_2	5,896	0,896	-0,256	7,227	-4,104	69,080	8,311	83,114
σ^2	1,953	0,433	0,134	0,403	-0,297	0,251	0,501	22,250
BC Estendida								
α_1	64,483	-0,034	0,003	0,775	-0,017	0,601	0,775	1,202
β_1	0,502	0,976	3,526	0,023	0,002	0,001	0,023	4,556
δ_2	-0,262	-1,727	10,901	0,017	-0,002	0,000	0,017	6,470
τ_1	74,451	-0,731	2,638	2,330	-0,049	5,431	2,330	3,128
τ_2	198,901	1,016	6,137	4,321	0,301	18,764	4,332	2,181
γ_1	11,230	0,817	-0,494	13,004	-3,770	183,314	13,539	90,262
γ_2	11,066	1,772	2,789	18,452	-3,934	355,961	18,867	125,779
σ^2	1,987	0,612	0,699	0,437	-0,263	0,260	0,510	22,670

As distribuições amostrais de $\hat{\phi}_m$ são unimodais e razoavelmente simétricas, com exceção das obtidas para os parâmetros γ_i 's e ζ_i 's (ver anexo A). As distribuições amostrais dos estimadores dos parâmetros de suavização apresentam uma acumulação de valores próximos de zero, o que indica que para muitos conjuntos de dados os ajustes caracterizaram transições de fase abruptas, ao invés de suaves. Como os dados foram simulados partindo-se de transições suaves, pode-se afirmar que frequentemente as transições de fase suaves são mal interpretadas como

mudanças abruptas, por falta de informação sobre elas nos conjuntos de dados.

Provavelmente isso ocorre porque a variabilidade ao acaso se confunde com a própria forma do gráfico do modelo na região de transição e impossibilita a estimação correta dos parâmetros de suavização. Então, o fato do valor do parâmetro de suavização não diferir significativamente de zero não implica um processo de transição de fase abrupto. As simulações indicam que amostras retiradas de um processo de transição de fase gradual podem não conter informação suficiente a respeito da transição de fase. Em contrapartida, se o ajuste evidenciar uma transição suave, tem-se o caso no qual a forma da relação na região de transição de fase pôde ser capturada pela informação contida nos dados e as inferências sobre os parâmetros de suavização tornam-se possíveis.

2.4.2 Resultados Teóricos

Na Figura 8 é possível observar que as junções médias do tipo I e as funções transicionais apresentam gráficos bem parecidos, sinalizando que há alguma equivalência entre essas duas aproximações suaves para ligar regimes. Diante desse fato, pergunta-se:

- (a) Sob quais condições uma junção média do tipo I é uma função pertencente à família transicional de Bacon e Watts (1971) e Griffiths e Miller (1973)?
- (b) Sob quais condições um modelo transicional de Bacon e Watts (1971) e Griffiths e Miller (1973) é considerado uma junção média do tipo I e, dessa maneira, está associado a alguma distribuição para os limiares aleatórios T_l 's?

A seguir serão demonstradas as seguintes proposições:

◇ **Proposição (a):** a junção média do tipo I para a formulação sinal, expressa por

$$E[\text{sinal}(S_l)] = 2F_{S_l}(s_l | \gamma_l) - 1, \quad (50)$$

é uma função transicional sempre que f_{S_l} é contínua e S_l é integrável ou possui suporte $\{s_l \in \mathbb{R} : f_{S_l}(s_l) > 0\}$ limitado.

◇ **Proposição (b):** toda função transicional de Bacon e Watts (1971) e Griffiths e Miller (1973) não decrescente pode ser reescrita como

$$\text{trn}(s_l; \gamma_l) = 2F_{S_l}(s_l | \gamma_l) - 1, \quad (51)$$

em que F_{S_l} é a função de distribuição de alguma variável aleatória contínua S_l . Ou seja, toda função transicional não decrescente é uma junção média do tipo I.

Demonstração de (a): a fim de que a junção média do tipo I para a formulação sinal em (50) seja uma função transicional, em primeiro lugar é preciso que ela seja da Classe C^1 . Portanto, a função de distribuição F_{S_l} também precisa ser continuamente diferenciável nos reais, o que significa que a densidade f_{S_l} deve ser contínua.¹ Além disso, para que uma junção média do tipo I pertença à família de funções transicionais, é preciso que ela satisfaça as condições impostas por Bacon e Watts (1971) e Griffiths e Miller (1973):

◇ Condição (i)

$$(c.1) \lim_{s_l \rightarrow -\infty} 2s_l F_{S_l}(s_l | \gamma_l) = 0 \quad \text{e} \quad (c.2) \lim_{s_l \rightarrow +\infty} 2s_l [F_{S_l}(s_l | \gamma_l) - 1] = 0 \quad (52)$$

A validade da condição (i) é clara para densidades com suporte compacto (ou apenas limitado), pois neste caso F_{S_l} atinge os valores 0 e 1 nos limites inferior e superior do suporte, respectivamente, sob valores finitos de s_l .² Em geral, pode-se afirmar que a condição (i) é válida se S_l é integrável, isto é, $E|S_l| < +\infty$:

Parte (c.1) - Sabemos que

$$0 \geq \lim_{s_l \rightarrow -\infty} 2s_l F_{S_l}(s_l | \gamma_l) = \lim_{s_l \rightarrow -\infty} 2s_l \int_{-\infty}^{s_l} dF_{S_l}(\mu) \geq \lim_{s_l \rightarrow -\infty} 2 \int_{-\infty}^{s_l} \mu dF_{S_l}(\mu), \quad (53)$$

uma vez que, pela propriedade de dominação das integrais definidas, o fato de $s_l \geq \mu$ para todo μ em $(-\infty, s_l]$ implica a desigualdade

$$\int_{-\infty}^{s_l} s_l dF_{S_l}(\mu) \geq \int_{-\infty}^{s_l} \mu dF_{S_l}(\mu).$$

Visto que

$$\int_{-\infty}^{s_l} \mu dF_{S_l}(\mu) = E[S_l] - \int_{s_l}^{+\infty} \mu dF_{S_l}(\mu), \quad (54)$$

o lado direito da desigualdade em (53) tem valor definido (e igual a zero) somente quando $E|S_l| < +\infty$. Logo, pelo teorema do confronto, a existência de esperança finita para a variável aleatória S_l implica o valor nulo para o limite em (c.1).

¹Por exemplo, o uso da função de distribuição de uma variável aleatória Uniforme $U(a, b)$ em (50) resultaria em uma junção contínua dos regimes adjacentes, dado que essa função de distribuição é contínua. No entanto, as derivadas do modelo com respeito a s_l (ou analogamente a x) não existiriam sob os pontos a e b , uma vez que a função de densidade Uniforme é descontínua nos limites de seu suporte. Por outro lado, as distribuições Gaussiana, T, Laplace, Logística e Cauchy são exemplos de distribuições que poderiam ser utilizadas para obter junções médias do tipo I continuamente diferenciáveis.

²Há um conjunto extenso de distribuições com suportes compactos que atendem a condição (i), como a Triangular, a de Epanechnikov, a Quártica (“Quartic”), Triponderada (“Triweight”), a Tricúbica (“Tricube”) e a Cosseno (“Cosine”), dentre outras.

Parte (c.2) - Temos que

$$0 \leq - \lim_{s_l \rightarrow +\infty} 2s_l [F_{S_l}(s_l | \gamma_l) - 1] = \lim_{s_l \rightarrow +\infty} 2s_l \int_{s_l}^{+\infty} dF_{S_l}(\mu) \leq \lim_{s_l \rightarrow +\infty} 2 \int_{s_l}^{+\infty} \mu dF_{S_l}(\mu). \quad (55)$$

Assim, por (54) temos que $E|S_l| < +\infty$ implica o limite em (c.2) nulo.

Observe que se a variável aleatória S_l não é integrável, nada se pode afirmar.³

◇ Condição (ii)

$$\lim_{\gamma_l \rightarrow 0^+} 2 F_{S_l}(s_l | \gamma_l) - 1 = \text{sinal}(s_l).$$

Imagine que conforme o parâmetro de dispersão γ_l se aproxima de zero pela direita, a função de densidade f_{S_l} se aproxima da função Delta de Dirac, expressa por

$$\Delta(s_l) = \begin{cases} \infty & s_l = 0, \\ 0 & s_l \neq 0, \end{cases}$$

e para a qual a função de distribuição é $I_{\{s_l \geq 0\}}(s_l)$. Dessa maneira,

$$\lim_{\gamma_l \rightarrow 0^+} 2 F_{S_l}(s_l | \gamma_l) - 1 = 2 I_{\{s_l \geq 0\}}(s_l) - 1 = \text{sinal}(s_l) \quad \forall s_l \neq 0. \quad (56)$$

Quando $s_l = 0$, as junções médias do tipo I e a função sinal assumem valores diferentes em (56) e a condição (ii) não é satisfeita. Mas na prática, pouco importa o valor da junção média no ponto $s_l = 0$, visto que esse ponto determina a transição de fase, onde os regimes adjacentes apresentam valores iguais.

◇ Condição (iii)

$$\text{trn}(0, \gamma_l) = 0$$

e, como consequência,

$$2 F_{S_l}(0 | \gamma_l) - 1 = 0 \Leftrightarrow F_{S_l}(0 | \gamma_l) = 0,5.$$

³Um exemplo de distribuição que não satisfaz a condição (i) é a de Cauchy. Considere S_l com distribuição de Cauchy centrada em zero e com parâmetro de dispersão $\gamma_l = 1$. Então, a sua função de distribuição é dada por

$$F_{S_l}(s_l | 1) = \frac{1}{\pi} \tan^{-1}(s_l) + \frac{1}{2}.$$

Dessa maneira, tem-se que:

$$\lim_{s_l \rightarrow -\infty} 2s_l F_{S_l}(s_l | 1) = \lim_{s_l \rightarrow +\infty} 2s_l [F_{S_l}(s_l | 1) - 1] = -\frac{2}{\pi}.$$

A fim de satisfazer a condição (iii), a variável S_l deve ter mediana nula ou, equivalentemente, a mediana de T_l deve ser igual ao seu valor esperado τ_l . Então, para que a condição (iii) seja satisfeita, a distribuição de T_l deve ser simétrica, tal como Bacon e Watts (1971) a especificaram em seu artigo.

Na realidade, a condição (iii) não é necessária para que o gráfico do modelo atravessa o ponto de mudança quando se faz uso de uma junção média do tipo I. Pode-se mostrar que, nesse caso, a região de transição no modelo é resultado de uma combinação linear convexa dos regimes adjacentes:

Sem perda de generalidade, considera-se a união de dois regimes lineares por meio da parametrização descrita em (4). Substituindo (50) em (4), obtém-se

$$\eta(x|\boldsymbol{\theta}, \tau_1, \gamma_1) = \theta_0 + \theta_1(x - \tau_1) + \theta_2(x - \tau_1) \left[2 F_{T_l}(x - \tau_1 | \gamma_1) - 1 \right],$$

ou seja,

$$\eta(x|\boldsymbol{\alpha}, \boldsymbol{\beta}, \tau_1, \gamma_1) = \alpha_1 + \beta_1 \tau_1 + \frac{(x - \tau_1)}{2} \left\{ (\beta_1 + \beta_2) + (\beta_2 - \beta_1) \left[2 F_{T_l}(x - \tau_1 | \gamma_1) - 1 \right] \right\}.$$

Dado que $\alpha_2 = \alpha_1 + (\beta_1 - \beta_2) \tau_1$, verifica-se que

$$\eta(x|\boldsymbol{\alpha}, \boldsymbol{\beta}, \tau_1, \gamma_1) = (\alpha_1 + \beta_1 x) \left[1 - F_{T_l}(x - \tau_1 | \gamma_1) \right] + (\alpha_2 + \beta_2 x) F_{T_l}(x - \tau_1 | \gamma_1).$$

Dessa maneira, o modelo $\eta(x)$ é uma combinação linear dos regimes $g_1(x|\boldsymbol{\beta}_1) = \alpha_1 + \beta_1 x$ e $g_2(x|\boldsymbol{\beta}_2) = \alpha_2 + \beta_2 x$, com pesos iguais a $1 - F_{T_l}$ e F_{T_l} , respectivamente.

Visto que os regimes apresentam o mesmo valor sob o ponto de mudança $(\tau_1, \alpha_1 + \beta_1 \tau_1)$, o modelo sempre o atravessa. Logo, a condição (iii) é suprimida quando se utiliza uma junção média do tipo I como função transicional e, assim, distribuições assimétricas podem ser usadas para garantir maior flexibilidade aos modelos.

O mesmo raciocínio se aplica quando é desejado unir três ou mais regimes, ao invés de dois. Se a pressuposição de que as regiões de transição de fase não se interceptam for respeitada, então a união dos regimes, dois a dois, será caracterizada por uma combinação linear convexa de seus valores.

Demonstração de (b): considera-se a função trn em (51) como não decrescente, de maneira que F_{S_l} atenda às condições a seguir:

(c.1) A função F_{S_l} é não decrescente, uma vez que trata-se de uma transformação monótona de trn ;

(c.2) A função F_{S_l} é contínua à direita, dado que trn pertence a Classe C^1 ;

(c.3) A função F_{S_l} atende às seguintes condições:

$$\lim_{s_l \rightarrow -\infty} F_{S_l}(s_l | \gamma_l) = 0 \quad \text{e} \quad \lim_{s_l \rightarrow +\infty} F_{S_l}(s_l | \gamma_l) = 1,$$

visto que, pela condição (i) de Bacon e Watts (1971) e Griffiths e Miller (1973),

$$\lim_{s_l \rightarrow \pm\infty} \left(\text{trn}(s_l; \gamma_l) - \text{sinal}(s_l) \right) = 0,$$

e, consequentemente,

$$\lim_{s_l \rightarrow -\infty} \text{trn}(s_l; \gamma_l) = -1 \quad \text{e} \quad \lim_{s_l \rightarrow +\infty} \text{trn}(s_l; \gamma_l) = 1.$$

Pelo fato da função transicional ser não decrescente, decorre da condição (i) que sua imagem é o conjunto $[-1, 1]$, de maneira que $F_{S_l} : \mathbb{R} \rightarrow [0, 1]$. Sabe-se que toda função de $\mathbb{R}$ em $[0, 1]$ que satisfaz as condições (c.1), (c.2) e (c.3) é função de distribuição de alguma variável aleatória (JAMES, 2011; MAGALHÃES, 2006). Então, pode-se afirmar que para toda função transicional não decrescente existe uma função F_{S_l} , expressa por

$$F_{S_l}(s_l | \gamma_l) = \frac{\text{trn}(s_l; \gamma_l) + 1}{2},$$

a qual é função de distribuição de alguma variável aleatória S_l .

Uma vez que a função transicional é da Classe C^1 e não decrescente, existe uma função não negativa

$$f_{S_l}(s_l | \gamma_l) = \frac{1}{2} \frac{d \text{trn}(s_l; \gamma_l)}{ds_l}$$

tal que

$$\int_{-\infty}^{s_l} f_{S_l}(w | \gamma_l) dw = \frac{\text{trn}(s_l; \gamma_l) + 1}{2} = F_{S_l}(s_l | \gamma_l), \quad \forall s_l \in \mathbb{R}.$$

Por esse motivo, S_l é uma variável aleatória contínua e f_{S_l} é a sua função de densidade (MAGALHÃES, 2006).⁴

⁴Por exemplo, se $\text{trn}(s_l; \gamma_l) = \tanh(s_l/\gamma_l)$ for considerada como função transicional, então obtém-se

$$F_{S_l}(s_l | \gamma_l) = \frac{1}{1 + \exp(-2s_l/\gamma_l)},$$

a qual é função de distribuição Logística com centro 0 e dispersão igual a $\gamma_l/2$.

2.5 Conclusão

Este trabalho apresenta uma nova parametrização da formulação máx-mín, especializada na construção de modelos lineares segmentados para o caso da regressão simples e univariada de processos multi-fase. A fim de modelar transições de fase suaves, uma extensão da regressão “bent-cable” de Tishler e Zang (1981a) e Chiu et al. (2006) é proposta. Também são apresentadas duas novas famílias de junções suaves entre regimes, que decorrem da inclusão de limiares aleatórios na estrutura do modelo: elas são as junções médias do tipo I e II. O trabalho ainda estabelece uma inter-relação entre as junções médias do tipo I e a família de aproximações transicionais de Bacon e Watts (1971) e Griffiths e Miller (1973), de forma que as aproximações transicionais não decrescentes sejam interpretáveis pela distribuição de limiares aleatórios.

Observa-se que a regressão “bent-cable” estendida é útil quando deseja-se estabelecer faixas onde as transições de fase ocorrem. Por outro lado, o uso de junções médias permite estimar a variabilidade de limiares aleatórios e as probabilidades de transição de fase condicionadas ao valor de uma variável explanatória. No entanto, frequentemente se obtêm modelos ajustados com transições de fase abruptas, mesmo quando o processo gerador dos dados é caracterizado por mudanças de fase graduais.

Agradecimentos

O presente trabalho foi realizado com apoio do CNPq, Conselho Nacional de Desenvolvimento Científico e Tecnológico - Brasil.

Referências

- BACON, D.W.; WATTS, D.G. Estimating the transition between two intersecting straight lines. **Biometrika**, Oxford, v. 58, n. 3, p. 525-534, 1971.
- BOLKER, B; **Tools for general maximum likelihood estimation**. R Development Core Team. Disponível em: <<http://cran.r-project.org/web/packages/bbmle/bbmle.pdf>>. Acesso em: 16 junho 2015.
- BARROWMAN, N.J.; MYERS, R.A. Still more spawner-recruitment curves: the hockey stick and its generalizations. **Canadian Journal of Fisheries and Aquatic Sciences**, Ottawa, v. 57, n. 4, p. 665-676, 2000.
- BROYDEN, C.G. The convergence of a class of double-rank minimization algorithms. **Journal of the Institute of Mathematics and Its Applications**, Essex, v. 6, p. 76-90, 1970.
- CERRATO, M.E.; BLACKMER, A.M. Comparison of models for describing corn yield response to nitrogen fertilizer. **Agronomy Journal**, Madison, v. 82, p. 138-143, 1990.
- CHAPPEL, R. Fitting bent lines to data, with applications to allometry. **Journal of Theoretical Biology**, London, v. 138, n. 2, p. 235-256, 1989.

CHIU, G.; LOCKHART, R.; ROUTLEDGE, R. Bent-cable regression theory and applications. **Journal of the American Statistical Association**, Alexandria, v. 101, n. 474, p. 542-553, 2006.

FAQUIN, V. **Diagnose do estado nutricional de plantas**. Lavras: UFLA e FAEPE, 2002. 77p.

FEDER, P.I. On asymptotic distribution theory in segmented regression problems - identified case. **The Annals of Statistics**, Philadelphia, v. 3, n. 1, p. 49-83, 1975.

FLETCHER, R. A new approach to variable metric algorithms. **Computer Journal**, Oxford, v. 13, n. 3, p. 317-322, 1970.

GALLANT, R.A.; FULLER, W.A. Fitting segmented polynomial regression models whose join points have to be estimated. **Journal of the American Statistical Association**, Alexandria, v. 68, n. 341, p. 144-147, 1973.

GOLDFARB, D. A family of variable metric updates derived by variational means. **Mathematics of Computation**, Providence, v. 24, n. 109, p. 23-26, 1970.

GRIFFITHS, D.A.; MILLER, A.J. Hyperbolic regression - a model based on two-phase piecewise linear regression with a smooth transition between regimens. **Communications in Statistics**, Abingdon, v. 2, n. 6, p. 561-569, 1973.

GRIFFITHS, D.; MILLER, A.; DRAPER, N.R.; BOX, M.J.; JONES, E.R. Letters to the editor. **Technometrics**, Alexandria, v. 17, n. 2, p. 281-282, 1975.

HAVLIN, J.L.; BEATON, J.D.; TISDALE, S.L.; NELSON, W.L. **Soil fertility and fertilizers: an introduction to nutrient management**. 7.ed. New Jersey: Pearson, 2005. 515 p.

HAYNE, S.C.; WANG, H.; MENDONCA, S. eBay as the "Terminator": determining user suspension from feedback ratings. **Journal of Organizational Computing and Electronic Commerce**, London, v. 22, n. 2, p. 160-183, 2012.

HINKLEY, D.V. Inference about the intersection in two-phase regression. **Biometrika**, Oxford, v. 56, n. 3, p. 495-504, 1969.

HINKLEY, D.V. Inference in two phase regression. **Journal of the American Statistical Association**, Alexandria, v. 66, n. 336, p. 736-743, 1971.

HOUT, A. van den; MUNIZ-TERRERA, G.; MATTHEWS, F.E. Change point models for cognitive tests using semi-parametric maximum likelihood. **Computational Statistics and Data Analysis**, The Hague, v. 57, p. 684-698, 2013.

JAMES, B.R. **Probabilidade: um curso em nível intermediário**. 3.ed. Rio de Janeiro: IMPA, 2011. 304 p.

KHAN, S.A. **Flexible bent-cable models for mixture longitudinal data**. 2010. Tese (Doutorado em Estatística e Bioestatística) – University of Waterloo, Ontário, Canadá, 2010. Disponível em:
<https://uwspace.uwaterloo.ca/bitstream/handle/10012/5092/Khan_Shahedul.pdf?sequence=1>. Acesso em: 24 abril 2015.

MACY, P. The quantitative mineral requirements of plants. **Plant Physiology**, Bethesda, v. 11, n. 4, p. 749-764, 1936.

- MAGALHÃES, M.N. **Probabilidade e variáveis aleatórias**. 2.ed. São Paulo: Editora da Universidade de São Paulo, 2006. 428 p.
- MALAVOLTA, E. **Nutrição mineral e adubação do cafeeiro**. São Paulo: Editora Agronômica CERES, 1993. 210 p.
- McGEE, V.E.; CARLETON, W.T. Piecewise regression. **Journal of the American Statistical Association**, Alexandria, v. 65, n. 331, p. 1109-1124, 1970.
- MUGGEO, V.M.R. Estimating regression models with unknown break-points. **Statistics in Medicine**, Malden, v. 22, p. 3055-3071, 2003.
- PRUNTY, L. Curve fitting with smooth functions that are piecewise-linear in the limit, **Biometrics**, Arlington, v. 39, n. 4, p. 857-866, 1983.
- QIAN, S.S.; CUFFNEY, T.F. To threshold or not to threshold? That's the question. **Ecological Indicators**, Kiel, v. 15, p. 1-19, 2012.
- QUANDT, R.E. The estimation of the parameters of a linear regression system obeying two separate regimes. **Journal of the American Statistical Association**, Alexandria, v. 53, n. 284, p. 873-880, 1958.
- RANA, M. **Spatial-longitudinal bent-cable model with an application to atmospheric CFC data**. 2012. Dissertação (Mestrado em Ciência) – University of Saskatchewan, Saskatoon, Canadá, 2012. Disponível em: <http://math.usask.ca/~longhai/students/Thesis_mdr091.pdf>. Acesso em: 24 abril 2015.
- R CORE TEAM. **R: a language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria, 2015. Disponível em: <<http://www.r-project.org>>. Acesso em: 24 abril 2015.
- RATKOWSKY, D.A. **Handbook of nonlinear regression models**. New York: M. Dekker, 1990. 241 p.
- SCHNEEBERGER, H. **Over-fertilization: an inverse Mitscherlich process**. Institute of Statistics, University of Erlangen-Nürnberg. Disponível em: <<http://soil-statistic.de/overfertilization/overfertilization.html>>. Acesso em: 24 abril 2015.
- SEBER, G.A.F.; WILD, C.J. **Nonlinear regression**. United States: John Wiley & Sons, Inc., 1989. 768 p.
- SHANNO, D.F. Conditioning of quasi-Newton methods for function minimization. **Mathematics of Computation**, Providence, v. 24, n. 111, p. 647-656, 1970.
- SHUAI, X.; ZHOU, Z.; YOST, R.S. Using segmented regression models to fit soil nutrient and soybean grain yield changes due to liming. **Journal of Agricultural, Biological, and Environmental Statistics**, Alexandria, v. 8, n. 2, p. 240-252, 2003.
- SPRENT, P. Some hypotheses concerning two phase regression lines, **Biometrics**, Arlington, v. 17, n. 4, p. 634-645, 1961.
- TISHLER, A.; ZANG, I. A new maximum likelihood algorithm for piecewise regression. **Journal of the American Statistical Association**, Alexandria, v. 76, n. 376, p. 980-987, 1981a.
- TISHLER, A.; ZANG, I. A maximum likelihood method for piecewise regression models with a continuous dependent variable. **Journal of the Royal Statistical Society. Series C (Applied Statistics)**, Malden, v. 30, n. 2, p. 116-124, 1981b.

TOMS, J.D.; LESPERANCE, M.L. Piecewise regression: a tool for identifying ecological thresholds. **Ecology**, Ithaca, v. 84, n. 8, 2003.

ULRICH, A. Physiological bases for assessing the nutritional requirements of plants. **Annual Review of Plant Physiology**, Florida, v. 3, p. 207-228, 1952.

WATTS, D.G.; BACON, D.W. Using an hyperbola as a transition model to fit two-regime straight-line data. **Technometrics**, Alexandria, v. 16, n. 3, p. 369-373, 1974.

WOLFE, P. Convergence conditions for ascent methods. **SIAM Review**, Philadelphia, v. 11, n. 2, p. 226-235, 1969.

WOLFE, P. Convergence conditions for ascent methods. II: Some corrections. **SIAM Review**, Philadelphia, v. 13, n. 2, p. 185-188, 1971.

ZANG, I; A smoothing-out technique for min-max optimization. **Mathematical Programming**, Philadelphia, v. 19, n. 1, p. 61-77, 1980.

3 CURVA DA PRODUÇÃO VEGETAL MEDIANTE ESTADOS NUTRICIONAIS DISTINTOS

Resumo

A curva de calibração da produção vegetal apresenta uma forma complicada, parecida com uma banheira invertida, em decorrência dos diferentes regimes nutricionais que as plantas podem apresentar. Os modelos comumente usados na agricultura não são capazes de descrever essa curva em toda sua extensão. Neste trabalho, a curva de produção vegetal foi modelada por meio de modelos lineares segmentados com transições de fase suaves. Como motivação prática, os modelos foram ajustados para descrever os dados de Schneeberger (2009) sobre a resposta do trigo de inverno à adubação nitrogenada. Os modelos mostraram-se flexíveis o bastante para descrever a produção da cultura mediante os diferentes regimes nutricionais existentes. Por meio dos ajustes foi possível estimar quantidades de interesse agrícola, tais como os níveis críticos e as faixas adequada e de suficiência. O uso de modelos com transições de fase suaves também possibilitou interpretações biológicas interessantes acerca das mudanças de estados nutricionais.

Palavras-chave: Nutrição mineral de plantas; Modelos segmentados; Transições suaves; Níveis críticos

Abstract

The yield curve in fertilizer experiments shows a complicated shape due to the existence of different nutritional regimes. This form looks like an upside-down bathtub and models commonly used in agriculture are not able to describe this curve throughout its length. In this work the yield curve was modeled by means of piecewise linear models with smooth phase transitions. The studied models were applied to describe the Schneeberger's data on winter wheat response to different doses of nitrogen (SCHNEEBERGER, 2009). The models shown flexibility, describing adequately the yield curve along the different nutritional regimes. By using these models, it was possible to estimate the critical levels and sufficiency ranges for nitrogen doses. It has been found that the use of smooth transition models brought important biological interpretations concerning changes of the nutritional status.

Keywords: Plant nutrition; Piecewise models; Smooth phase transitions; Critical levels

3.1 Introdução

Em experimentos de adubação, o pesquisador provoca deliberadamente uma deficiência mineral e avalia como aumentos graduais nos níveis de fertilização afetam os rendimentos de uma cultura. A curva resultante, dita curva de calibração da produção vegetal, é subdividida em estados nutricionais distintos (FAQUIN, 2002).

Muitos autores ilustram a curva de produção vegetal como tendo a forma de uma banheira invertida (Figura 12), formada por três segmentos de retas que são ligados uns aos outros por junções suaves (JONES JR, 1993; MALAVOLTA, 1993; MILES, 2010; PREVOT; OLLAGNIER, 1957; TAIZ; ZEIGER, 2004; SMITH, 1962). Nela há o regime de deficiência mineral,

no qual a planta responde ao fertilizante e a produção aumenta. A partir de certo ponto, entretanto, a cultura deixa de responder a incrementos nos níveis de fertilização e curva passa a ter a forma de um platô, ou seja, se estagna em um valor de produção ótimo - a essa região dá-se o nome de regime de consumo de luxo. Por outro lado, níveis altos de fertilização podem ocasionar decréscimos na produção, caracterizando o regime nutricional de toxidez.

Figura 12 – Curva de calibração da produção vegetal adaptada de Malavolta (1993)

A maioria dos autores apresentam a curva de produção vegetal em função da concentração do elemento mineral nos tecidos. No entanto é importante notar que a curva em função da dose de fertilizante aplicada (ou da concentração de elemento na solução do solo) comumente tem forma similar, uma vez que há relação direta entre a quantidade de nutriente assimilável e o seu conteúdo na planta. Além disso, a mesma curva também pode ser usada para modelar dados de produtividade vegetal, visto que a produtividade é, por definição, a produção por unidade de área.

Em muitas circunstâncias, pesquisas em nutrição mineral de plantas e fertilização dependem do conhecimento acerca da forma da curva de produção vegetal. Nessas áreas, geralmente é requerida a estimação de padrões de comparação derivados da curva de produção, como os níveis críticos e as faixas de suficiência (FAQUIN, 2002; GRADES, 2011; MACY, 1936; MAIA et al., 2001; MALAVOLTA et al., 1989; MILES, 2010; NEVES et al., 2008; WADT et al., 1998). Esses padrões auxiliam na avaliação do estado nutricional das culturas, o que possibilita a recomendação de doses de fertilizante e o uso econômico de adubos, favorecendo a obtenção de colheitas mais produtivas (MALAVOLTA, 1967; MALAVOLTA et al., 1989).

A despeito da importância que a curva de produção vegetal tem na nutrição mineral de

plantas e adubação, são raros os estudos devotados a descrevê-la em uma escala completa, indo do regime da deficiência mineral até a fase de toxidez. Talvez isso ocorra porque os modelos mais usados nessas áreas não são flexíveis o bastante para elucidar a forma da relação. Alguns modelos de uso comum, por exemplo, apresentam um máximo de produção: esse é o caso dos polinômios de 2º grau e do modelo de Mitscherlich (1928) de segunda ordem. Eles podem até ser adequados para descrever dados nos quais o regime de consumo de luxo não é pronunciado, mas inúmeros estudos exibem de forma clara um platô ao invés de um máximo de produção (CERRATO; BLACKMER, 1990; LOPES; SOUSA, 2013; MAUST; WILLIAMSON, 1994; RIED, 2002; SCHERER, 1998; SCHNEEBERGER, 2009). Outros modelos de uso frequente são o linear-platô, o quadrático-platô e o de Mitscherlich (1909) (CERRATO; BLACKMER, 1990). Esses modelos possuem gráficos monótonos não-decrescentes e, portanto, podem ser aplicados apenas na porção inicial da curva, ao longo dos regimes de deficiência mineral e consumo de luxo.

O objetivo deste trabalho foi o de modelar a curva de produção vegetal em uma escala de doses (ou teores) completa. Propôs-se o uso de modelos lineares segmentados com transições de fase suaves (previamente estudados no capítulo 2) para descrever o comportamento da produção vegetal mediante os diferentes regimes nutricionais (BACON; WATTS, 1971; CHIU et al., 2006; GRIFFITHS; MILLER, 1973; SEBER; WILD, 1989; TISHLER; ZANG, 1981; WATTS; BACON, 1974). Diversas abordagens para estimar as quantidades de interesse agrícola (como níveis críticos e faixas de suficiência) também foram propostas. A fim de ilustrar a metodologia, os modelos sugeridos neste trabalho foram ajustados aos dados de Schneeberger (2009) sobre a produtividade do trigo de inverno em função da dose de nitrogênio aplicada no solo.

3.2 Revisão Bibliográfica

Os elementos minerais são determinantes para o desenvolvimento e o crescimento dos vegetais (TAIZ; ZEIGER, 2004). Todos eles, com exceção do C, H e O, provêm do solo (LIEBIG, 1840). Por tal motivo, a produtividade no campo está intimamente ligada ao uso eficiente de adubos. A fim de garantir boas colheitas, o agricultor deve estudar a composição mineral do solo e da planta e, fazendo uso das informações levantadas, traçar intervenções corretivas, tal como a aplicação de doses suplementares de fertilizantes (MALAVOLTA, 1967). Neste contexto destaca-se a avaliação do estado nutricional de plantas, que consiste no estudo da composição mineral dos tecidos e o seu relacionamento com o crescimento do vegetal e a produtividade da cultura (MALAVOLTA et al., 1989). A técnica permite a detecção de deficiências ou excessos de minerais na planta e possibilita o acompanhamento do programa de fertilização corrente

(MILES, 2010). O estado nutricional das culturas pode ser inferido por diagnose visual (estudo dos sintomas visuais de deficiência e toxidez) ou por meio de análises químicas de amostras de órgãos específicos das plantas.

A diagnose visual, a despeito de sua simplicidade, apresenta muitas desvantagens: os sintomas relacionados às desordens nutricionais geralmente são visíveis tardiamente, quando parte da produção já está comprometida (MEMON et al., 2005); eles também podem ser confundidos com os decorrentes da ação de pragas e moléstias; além disso, quando a desordem nutricional é causada por mais de um nutriente, eles são muito difíceis de serem interpretados (MALAVOLTA, 1967; MALAVOLTA et al., 1989).

A análise química de amostras, por sua vez, requer o uso de padrões para a interpretação dos resultados. Sem a existência de padrões, os dados obtidos a campo não possuem qualquer significado. Estes padrões podem ser definidos a partir da análise de materiais coletados em áreas altamente produtivas ou por meio de experimentos de adubação (MALAVOLTA, 1967).

A comparação das composições nutricionais observadas no cultivar sob estudo com a de áreas produtivas possibilita a detecção de deficiências e excessos de minerais, bem como o estabelecimento de faixas adequadas para os teores (as concentrações) de cada elemento na planta (MALAVOLTA et al., 1989). Todavia, a técnica, por si só, não habilita o planejamento de estratégias corretivas do solo ou a predição das colheitas futuras. Estas habilidades dependem do conhecimento que o pesquisador tem acerca das relações entre as doses de adubo aplicadas, as frações assimiladas pelas plantas e o seu impacto na produção vegetal. A nutrição mineral de plantas se sustenta sobre a premissa de que estas quantidades têm relações diretas entre si (FAQUIN, 2002), e elas só podem ser determinadas por meio da experimentação (MALAVOLTA, 1967).

A fim de estudá-las, são realizados os ensaios de calibração da produção vegetal (MAIA et al., 2001), experimentos nos quais o pesquisador provoca deliberadamente uma deficiência mineral e avalia como aumentos graduais dos níveis de fertilização afetam os rendimentos de uma cultura. Eles são bem simples e podem ser realizados sob condições controladas, em vasos com areia ou por meio de soluções nutritivas, ou até mesmo no campo, onde deseja-se determinar as doses requeridas de cada nutriente (MALAVOLTA et al., 1989). Tais experimentos resultam em curvas da produção vegetal, que podem ser ajustadas em função das doses aplicadas no solo ou dos teores dos nutrientes na planta.

Ao ajustarmos a produção da cultura em função da dose de fertilizante, podemos calcular a concentração ótima do elemento no solo ou mesmo predizer, até certo ponto, os retornos a serem obtidos em função da quantidade total de adubo aplicada (PIMENTEL-GOMES; NOGUEIRA, 1980). A exatidão das previsões, entretanto, depende das propriedades do solo e do clima, e de

outros fatores que também influenciam a fração do fertilizante assimilável pelas plantas. A fim de traçar recomendações de adubação, é mais razoável ajustar a produção em função dos teores do elemento nos tecidos e, posteriormente, estudar qual é a melhor dose a ser ministrada para cada um dos tipos de solo e clima (MALAVOLTA, 1967).

A obtenção dos teores presentes na planta requer, por sua vez, o emprego de análise química de amostras teciduais. O estudo da composição nutricional dos tecidos e seu relacionamento com a produção vegetal é escopo da diagnose foliar. Na técnica, geralmente são analisadas as folhas maduras, cujo crescimento já terminou, uma vez que essas refletem particularmente bem as relações entre o teor de um elemento no solo e o conteúdo do mesmo na planta (MALAVOLTA, 1967). Ademais, elas são os principais órgãos responsáveis pela produção vegetal, graças à fotossíntese, de forma que os seus teores de nutrientes apresentam melhor correlação com o desempenho (FLOSS, 2006). As amostras são coletadas em um período de vida pré-definido, no qual a composição nutricional das plantas é mais estável (FLOSS, 2006).

A interpretação correta das análises laboratoriais depende, em grande parte, do conhecimento adquirido sobre as curvas de produção (HAVLIN et al., 2004). As curvas podem variar em suas características e formas, dependendo do elemento e da espécie vegetal (SMITH, 1962), mas apresentam muitas similaridades entre si, traços que são repetidamente verificados nos experimentos (GRADES, 2011; HAVLIN et al., 2004; JONES JR, 1993; MACY, 1936; MILES, 2010; OHKI, 1984; PREVOT; OLLAGNIER, 1956, ULRICH, 1952). Tais curvas possuem a forma aproximada de uma banheira invertida e são divididas em diferentes regimes nutricionais: a deficiência mineral, o consumo de luxo e a toxidez (FAQUIN, 2002).

O regime de deficiência mineral é verificado quando a disponibilidade de um elemento está aquém das exigências nutricionais da planta. Neste estágio há a redução do potencial produtivo e, às vezes, sinais visuais da deficiência podem ser observados no campo (MALAVOLTA et al., 1989). A deficiência mineral é caracterizada por concentrações citoplasmáticas relativamente pequenas do nutriente, o que, por questões estequiométricas, pode comprometer a atividade de síntese na planta (STERNER; ELSER, 2002).

Justus von Liebig foi um dos primeiros a avaliar como a escassez de nutrientes afeta os rendimentos de uma cultura. Ele é responsável por introduzir a lei do mínimo, a qual postula que a produção vegetal é limitada pelo fator que encontra-se disponível em quantidade relativa mínima (LIEBIG, 1840). Ao propor a lei do mínimo, Liebig imaginou a produção vegetal de maneira reducionista, sendo similar a uma reação química simples, em que a quantidade final de produto é determinada pelo reagente que é consumido primeiro (STERNER; ELSER, 2002).

A lei do mínimo de Liebig não se aplica apenas aos nutrientes, ela pode ser estendida para qualquer fator de produção (MITSCHERLICH, 1909). A disponibilidade de água e luz, a

incidência de pragas e a temperatura, por exemplo, podem ser fatores limitantes da produção, ou seja, podem ocupar a posição de mínimo. O cenário mais comum de se verificar na prática é, entretanto, aquele em que as colheitas são limitadas pela escassez de algum mineral no solo.

Figura 13 – Ilustração do barril de Liebig adaptada de “Wikipedia, the free encyclopedia: Liebig’s law of the minimum” (WIKIPEDIA, 2015)

Uma analogia clássica para a lei do mínimo é o barril de Liebig (Figura 13). Nele, o nível de água representa a produção a ser atingida, enquanto o tamanho das ripas são indicadores das quantidades relativas de cada um dos nutrientes na planta. Pelo barril de Liebig é possível observar que a ripa que controla o nível de água é aquela que apresenta o menor comprimento. De maneira análoga, o elemento que controla a produção é aquele que se encontra em quantidade relativa mínima. Se desejamos aumentar o nível de água do barril, devemos incrementar o comprimento da menor ripa. Equivalentemente, se é desejado aumentar a produção vegetal, precisamos elevar a disponibilidade do elemento que se encontra no mínimo. Alterar os níveis dos elementos que não limitam a produção é algo ineficaz, não produz retornos. Conforme a lei de Liebig, a cultura responde a incrementos no nutriente limitante até que uma dose crítica (ou nível crítico) seja atingida, quando outro fator de produção passa a ocupar a posição de mínimo. Além deste ponto, dose adicionais do nutriente não produzem retorno algum.

A lei de Liebig também admite que o crescimento das plantas é diretamente proporcional ao suprimento do elemento limitante, o que implica numa resposta linear (FLOSS, 2006; MACY, 1936; TAIZ; ZIEGER, 2004). Esta premissa foi fundamentada por uma série de experimentos realizados por Liebig e seus contemporâneos e resulta da escala de doses de fertilizantes aplicadas na época.

Quando doses pequenas são aplicadas em solos pobres, é bem provável que a cultura se estabeleça no que Macy (1936) chamou de zona dos percentuais mínimos: a este nível a deficiência pelo nutriente é tão severa que os aumentos em seu suprimento não ocasionam incrementos nos teores encontrados na planta, pois praticamente todo o fertilizante assimilado é metabolizado em função do crescimento, o volume da planta aumenta e a concentração cito-

plasmática do elemento permanece aproximadamente constante.

Por outro lado, se doses consideráveis do nutriente são ministradas às plantas, a concentração citoplasmática do elemento tende a se elevar até um valor que determina a sua suficiência (MACY, 1936; STERNER; ELSEER, 2002). A taxa de crescimento do organismo é diretamente relacionada com o teor citoplasmático do elemento limitante (STERNER; ELSEER, 2002) e, portanto, a produção também cresce. A planta, entretanto, não responde indefinidamente à adubação e o comportamento linear descrito por Liebig dá lugar a uma curva mais complexa. Aumentando-se progressivamente a dose de um determinado adubo, os retornos são cada vez menores, até o ponto em que a produção se estabiliza (MACY, 1936; MALAVOLTA et al., 1967; MITSCHERLICH, 1909).

Para explicar o comportamento gradual de saturação da resposta frente à fertilização, que contrasta com as funções lineares de Liebig, Mitscherlich propôs a lei dos retornos decrescentes (MACY, 1936; MALAVOLTA, 1967; MITSCHERLICH, 1909, 1928). Mitscherlich (1909) postulou que, conforme as doses de fertilizantes crescem, menores são os retornos obtidos por um incremento no fertilizante. De acordo com Mitscherlich (1909), os retornos a serem obtidos por um aumento no nível do fator limitante são proporcionais à diferença entre o máximo de produção que se pode atingir e a produção corrente.

A lei de Mitscherlich (1909) é uma aproximação empírica, mas geralmente produz bons ajustes aos dados. Conforme Macy (1936), as lei de Liebig (1840) e Mitscherlich (1909) aplicam-se a regiões distintas da curva de produção: a lei básica é a do mínimo, sendo a lei dos retornos decrescentes uma modificação para a transição de fase.

Assim que outro fator de produção, que não o nutriente estudado, ocupa a posição de mínimo, a planta deixa de responder à adubação e aumentos nas doses do elemento provocam pouco ou nenhum crescimento. Nesta etapa, a concentração do nutriente no citoplasma atinge o nível de suficiência e a sua fração em excesso é armazenada em estruturas de reserva, como os vacúolos, para ser usada em épocas de escassez (STERNER; ELSEER, 2002). Este regime nutricional é chamado consumo de luxo e caracteriza-se aproximadamente por um platô na curva de produção vegetal, região na qual a resposta se mantém próxima ao seu nível máximo.

O consumo de luxo é comum na maioria das plantas (MEMON et al., 2005), mas há culturas em que este regime nutricional é suprimido. Usualmente ele ocupa um domínio mais amplo para macronutrientes, como o potássio, e em outros casos, muito estreitos, como para o boro (MALAVOLTA et al., 1989).

O consumo de luxo, por sua vez, prossegue até o momento em que o excesso de nutriente nos tecidos é capaz de intoxicar a planta e/ou provocar desequilíbrios nutricionais, ocasionando decréscimos na produção (HAVLIN et al., 2004). A esse regime nutricional dá-se o nome de

toxidez. Há um número crescente de instâncias em que se torna importante identificar os níveis de concentração dos nutrientes excessivos ou tóxicos, mas, infelizmente, pouca informação tem sido obtida sobre a resposta à fertilização em uma escala completa, partindo da deficiência à toxidez (JONES JR, 1993).

Figura 14 – Produção relativa em função dos teor do nutriente na planta ou da dose de adubo aplicada no solo. Níveis críticos inferior e superior denotados por NC_D e NC_T , respectivamente; faixas de suficiência inferior e superior denotadas por FS_D e FS_T , respectivamente; e limites inferior e superior da faixa ótima denotados por LI e LS , respectivamente

Na realidade, o interesse agrônomo não se concentra na curva de calibração da produção vegetal, mas sim na estimação de quantidades derivadas que sirvam de padrões para a interpretação das análises, como os níveis críticos e as faixas de suficiência (FAQUIN, 2002; MAIA et al., 2001; NEVES et al., 2008; WADT et al., 1998). Os padrões constituem valores e intervalos de referência para a determinação do estado nutricional de amostras e, portanto, são originalmente expressos em termos das concentrações teciduais do elemento estudado. Entretanto, os teores dos elementos nos tecidos estão relacionados às doses aplicadas, de maneira que os padrões podem ser definidos, inclusive, por meio da quantidade do nutriente aplicado ou presente no solo, como é feito nos estudos de fertilidade (Figura 14).

De acordo com Grades (2011), os padrões nutricionais mais comuns são os níveis críticos. No entanto, para entendermos o seu significado, primeiramente devemos definir a faixa (ou intervalo) adequada. Ela consiste no conjunto de valores do teor foliar ou da dose do adubo que garantem produção próxima à máxima - para muitos autores, isto significa produção que atinja pelo menos 90% do valor máximo (MALAVOLTA, 1993). Enfim, Grades (2011) define os níveis críticos como sendo os valores que se situam nos extremos da faixa adequada, limitando

as regiões afetadas pelos efeitos da deficiência ou da toxidez por minerais. O extremo inferior da faixa adequada é dito nível crítico inferior ou de deficiência (denotado por NC_D), enquanto o extremo superior é referido como nível crítico superior ou de toxidez (denotado por NC_T).

Outra definição para o nível crítico de deficiência é o ponto abaixo do qual a produção é reduzida e acima do qual ela não é econômica. Quando nos referimos às doses de adubo ministradas, este valor é conhecido como dose economicamente aconselhável (FAQUIN, 2002; MALAVOLTA, 1967; PIMENTEL-GOMES; NOGUEIRA, 1980). Alternativamente, Miles (2010) também o interpreta como sendo o valor abaixo do qual o nutriente é o fator limitante do crescimento.

A despeito das várias definições já existentes, neste trabalho é adotado um conceito fisiológico para os níveis críticos, no qual eles são dados pelos teores teciduais ou doses do nutriente associados às mudanças de estado nutricional. Dessa maneira, o nível crítico de deficiência é relacionado com a concentração citoplasmática do mineral que garante sua suficiência na célula, o que coincide com a interpretação apresentada por Miles (2010). O nível crítico de toxidez, por sua vez, é associado ao valor de concentração do nutriente na célula (citoplasma e estruturas de reserva) capaz de causar desequilíbrios nutricionais pela competição com outros nutrientes ou intoxicação.

Dessa maneira, o regime de deficiência mineral é compreendido como a situação na qual as células que compõem a amostra foram afetadas pela falta do nutriente em algum estágio do desenvolvimento do órgão coletado. O regime de consumo de luxo é verificado quando todo material se desenvolveu mediante uma quantidade adequada do nutriente no órgão, sem excessos. Enfim, a toxidez determina um material no qual as células foram, de alguma forma, prejudicadas por uma quantidade excessiva do nutriente durante o desenvolvimento. Situações nas quais apenas parte do material coletado foi afetada pela deficiência mineral ou toxidez caracterizam regiões de transição entre os regimes nutricionais e são concebíveis pela variabilidade dos níveis críticos (no senso fisiológico) existente entre células, tecidos, órgãos ou, até mesmo, indivíduos.

Apesar de não existir uma divisão precisa entre os diferentes regimes nutricionais, neste trabalho assume-se que os seus domínios de atuação são delimitados pelas estimativas dos níveis críticos de deficiência e toxidez (ver Figura 14).

De acordo com Neves et al. (2008), as estimativas dos níveis críticos, independentemente de como eles são definidos, sempre estão associadas a intervalos de confiança e, dessa maneira, elas devem ser apresentadas como faixas estreitas de valores recomendados para as doses e/ou os teores do nutriente.

Outro padrão semelhante na avaliação do estado nutricional é dado pelas faixas de su-

ficiência. Para entendermos o que elas são, primeiramente devemos definir as faixas ótima, infra-ótima e supra-ótima (WADT et al., 1998). A faixa ótima é o conjunto de valores para a dose ou o teor foliar que garantem produção máxima. Por sua vez, a faixa infra-ótima é determinada pelos valores para os quais a produção é reduzida devido à deficiência pelo nutriente. Já a faixa supra-ótima consiste nos valores que ocasionam perdas na produção devido à toxidez por excesso do nutriente. Em resumo, as faixas de suficiência são os intervalos situados entre os níveis críticos e os limites da faixa ótima (NEVES et al., 2008).

Portanto, a faixa de suficiência inferior ou de deficiência (denotada por FS_D) é o intervalo de valores para as doses ou teores foliares compreendidos entre o nível crítico de deficiência e o limite inferior (denotado por LI) da faixa ótima, também conhecido como nível ótimo. De forma análoga, a faixa de suficiência superior ou de toxidez (denotada por FS_T) é o conjunto de valores compreendidos entre o limite superior (denotado por LS) da faixa ótima e o nível crítico de toxidez.

Uma vez que as definições dos níveis críticos e das faixas de suficiência são resultantes da curva de calibração da produção vegetal, o uso de formulações matemáticas que a descreva adequadamente é fundamental para a estimação acurada e precisa dessas quantidades. No entanto, a maioria dos trabalhos encontrados na bibliografia especializada utilizam modelos de regressão inadequados para este fim. Conforme Maia et al. (2001), os modelos geralmente ajustados para estes estudos são o polinomial quadrático, o exponencial e o de Mitscherlich (1909). Tais modelos são capazes de produzir resultados razoáveis apenas na porção inicial da curva, em uma escala de teores ou doses na qual a lei dos rendimentos decrescentes de Mitscherlich (1909) é aplicável. Entretanto, eles não contemplam o comportamento observado nas fases de consumo de luxo e toxidez.

Para modelar o platô ocasionado pelo consumo de luxo, muitos autores têm utilizado modelos segmentados de dois regimes lineares e com transições abruptas, denominados por modelos de resposta linear e platô, que constituem um caso especial dos modelos “broken-stick” (BERCK; HELFAND, 1990; MAUST; WILLIAMSON, 1994; PARIS, 1992a,b; SANCHEZ et al., 1991). Estes modelos, devido à transição abrupta, explicam de maneira irrealista a relação na região de maior interesse agrônômico, onde se situa o nível crítico de deficiência. Outro modelo de uso comum é o de segunda ordem de Mitscherlich (1928), o qual é bem flexível e permite a inclusão da fase de toxidez, mas mostra-se falho ao ajustar o regime de consumo de luxo.

A fim de modelar a curva de produção vegetal, Prunty (1983) apresenta uma família de modelos segmentados constituídos por regimes lineares e que permitem a introdução de transições suaves. Prunty (1983) mostra que seus modelos possuem grande flexibilidade na prática, se

ajustando bem a conjuntos de dados em comparação a outros modelos, como os de Mitscherlich. No entanto, a família proposta por Prunty (1983) é formulada com o auxílio da função módulo e, conseqüentemente, apresenta derivadas descontínuas que dificultam os processos de estimação e inferência (SEBER; WILD, 1989). Neste trabalho são propostos modelos similares aos de Prunty (1983), porém mais tratáveis sob o ponto de vista estatístico.

3.3 Material

Neste trabalho foram analisados os dados de Schneeberger (2009), sobre a produtividade de trigo de inverno (em 100 kg/ha) em função da adubação nitrogenada (kg/ha), obtidos em 2005, em um experimento realizado na Universidade de Tecnologia de München, Weihens-tephan, Alemanha, apresentados na Tabela 4.

Tabela 4 – Produtividade (100 kg/ha) de trigo de inverno segundo a dose de nitrogênio (kg/ha) aplicada no solo

Dose de Nitrogênio	Produtividade			
0	72,7	60,9	60,6	58,3
40	89,6	86,4	91,1	79,8
60	91,6	91,6	96,6	108,8
80	103,0	102,6	100,4	109,2
100	108,6	101,0	103,6	111,3
120	110,2	96,9	102,5	118,3
140	104,1	105,6	105,4	110,8
180	99,9	94,9	104,1	108,4
200	106,2	113,6	111,2	100,6
200	109,5	100,3	108,9	101,5
240	93,5	95,9	98,1	96,6
260	87,0	82,9	89,5	90,5
300	85,7	85,0	86,8	83,1

Fonte: Schneeberger (2009)

O delineamento experimental consistiu em 12 pontos de suporte, cada um deles repetido ao menos quatro vezes, sendo que as doses de nitrogênio foram ministradas em uma ampla escala de valores que vai da deficiência mineral por nitrogênio até a fase de toxidez. O autor não faz menção sobre o fato do delineamento ser inteiramente ao acaso ou realizado em blocos. Apesar disso, em análise preliminar não verificou-se o efeito de blocos.

A análise estatística deste trabalho foi realizada com o auxílio do *software* estatístico R (R CORE TEAM, 2015).

3.4 Métodos

Na seção 3.4.1 são apresentadas diversas versões de modelos lineares segmentados (com transições de fase suaves) úteis para modelar as curvas de produção vegetal mediante os três regimes nutricionais existentes: a deficiência mineral, o consumo de luxo e a toxidez. Já na seção 3.4.2 são descritos os procedimentos usados na estimação dos parâmetros e na realização de testes de hipóteses. Por fim, na seção 3.4.3 é descrito como obter as quantidades de interesse agrícola (tais como níveis críticos e faixas de suficiência) por meio das curvas de produção ajustadas.

3.4.1 Modelos de Regressão

Como mencionado anteriormente, as curvas de produção vegetal podem ser divididas em três regimes nutricionais distintos. Na terminologia da regressão segmentada, regime não é apenas o nome que se dá a uma fase de um processo, mas também é como denomina-se a função que a expressa. Então, para a construção dos modelos segmentados, o primeiro passo é a apresentação dos regimes que o constituem.

A fim de descrever adequadamente a curva de produção, o modelo segmentado deve ser constituído por três regimes lineares (segmentos de reta), denotados aqui por g_1 , g_2 e g_3 .¹ Neste caso, o regime de deficiência mineral é expresso por

$$g_1(x) = \alpha_1 + \beta_1 x, \quad 0 \leq x \leq \tau_1,$$

no qual x é a dose do nutriente aplicada no solo ou o teor do elemento nos tecidos da planta; τ_1 é o nível crítico do nutriente necessário para que a planta atinja a suficiência por esse elemento; α_1 é a produção obtida quando nenhum fertilizante é aplicado ao solo ou quando não há teor detectável do elemento nos tecidos;² e β_1 é o efeito do adubo, ou seja, o incremento que se observa na produção quando a dose (ou o teor do elemento nos tecidos) é aumentada em uma unidade.

A planta não responde indefinidamente aos incrementos no nutriente, pois, a partir do nível crítico de deficiência τ_1 , as plantas ingressam no regime do consumo de luxo. Neste caso,

¹Retas foram escolhidas para simplificar as descrições dos regimes, mas a metodologia abordada neste trabalho pode ser facilmente adaptada para incluir regimes com outras formas, tais como funções quadráticas ou não-lineares.

²Observe que se não há quantidades mínimas de um nutriente essencial na planta, então, pelo próprio conceito de essencialidade, a produção dessa planta será desprezível. Assim, quando trabalha-se com o teor do elemento nas plantas, convencionou-se que $\alpha_1 = 0$.

o regime de consumo pode ser descrito por

$$g_2(x) = \alpha_2 + \beta_2 x, \quad \tau_1 < x \leq \tau_2.$$

Em teoria, os incrementos na quantidade do nutriente durante o regime do consumo de luxo não devem surtir efeito na produção, uma vez que o excedente do nutriente exportado à planta é armazenado em órgãos de reserva ou excretado e não participa dos processos de síntese de compostos orgânicos complexos. Na prática, isso significa dizer que β_2 é igual a zero e, se esse for o caso, então o valor de α_2 determina a produção da cultura na sua faixa ótima (Figura 14).

Para doses muito elevadas do nutriente, a produção da cultura é governada pelo regime de toxidez, expresso por

$$g_3(x) = \alpha_3 + \beta_3 x, \quad x > \tau_2,$$

em que o parâmetro β_3 determina o decréscimo observado na produção a cada unidade em que se aumenta a quantidade do nutriente (dose ou teor) além do nível crítico de toxidez τ_2 .

A fim de produzir curvas de produção vegetal contínuas, os regimes nutricionais g_1 , g_2 e g_3 foram unidos com o auxílio da função sinal (formulação sinal) ou de operadores de máximo (formulação máx-mín). Essas são duas abordagens amplamente usadas para a construção de modelos segmentados abruptos (ver Capítulo 2), também conhecidos por “broken-stick models”. O modelo obtido pela formulação sinal é expresso por

$$\eta(x|\boldsymbol{\theta}, \boldsymbol{\tau}) = \theta_0 + \theta_1 x + \theta_2 (x - \tau_1) \text{ sinal}(x - \tau_1) + \theta_3 (x - \tau_2) \text{ sinal}(x - \tau_2), \quad (57)$$

em que: $\boldsymbol{\tau}' = (\tau_1, \tau_2)$ e $\boldsymbol{\theta}' = (\theta_0, \theta_1, \theta_2, \theta_3)$, de maneira que $\theta_3 = (\beta_3 - \beta_2)/2$, $\theta_2 = (\beta_2 - \beta_1)/2$, $\theta_1 = \beta_1 + \theta_2 + \theta_3$ e $\theta_0 = \alpha_1 - (\theta_1 - \beta_1) \tau_1 - \theta_3 |\tau_2 - \tau_1|$.

Enquanto que, pela formulação máx-mín, o modelo é expresso por

$$\eta(x|\boldsymbol{\beta}_1, \boldsymbol{\tau}, \boldsymbol{\delta}) = \alpha_1 + \beta_1 x + \delta_1 u(x - \tau_1) + \delta_2 u(x - \tau_2), \quad (58)$$

em que: $\boldsymbol{\beta}'_1 = (\alpha_1, \beta_1)$ e $\boldsymbol{\delta}' = (\delta_1, \delta_2)$, de maneira que $\delta_1 = \beta_2 - \beta_1$, $\delta_2 = \beta_3 - \beta_2$ e $u(z) = \max\{z, 0\} \forall z \in \mathbb{R}$.

Os modelos apresentados nas expressões (57) e (58) foram reparametrizados, respectivamente, por

$$\begin{aligned} \eta(x|\boldsymbol{\beta}_1, \beta_3, \boldsymbol{\tau}) &= \alpha_1 + \beta_1 \tau_1 + \frac{\beta_1}{2} (x - \tau_1) \left[1 - \text{sinal}(x - \tau_1) \right] + \\ &\quad \frac{\beta_3}{2} (x - \tau_2) \left[1 + \text{sinal}(x - \tau_2) \right] \end{aligned} \quad (59)$$

e

$$\eta(x|\beta_1, \beta_3, \tau) = \alpha_1 + \beta_1 x - \beta_1 u(x - \tau_1) + \beta_3 u(x - \tau_2), \quad (60)$$

nos casos em que se assumiu $\beta_2 = 0$.

Observe que os parâmetros dos modelos dispostos nas expressões (59) e (60) são os mesmos, sendo que a única diferença entre esses modelos é o operador abrupto (sinal ou máximo) usado para unir os regimes de forma contínua.

De modo que os modelos segmentados tenham transições de fase suaves, os operadores abruptos dos modelos dispostos em (57-60) foram substituídos por funções continuamente diferenciáveis que aproximam seus comportamentos, tais como as aproximações transicionais e hiperbólicas, os “bent-cables” e as junções médias (BACON; WATTS, 1971; CHIU et al., 2006; GRIFFITHS; MILLER, 1973; SEBER; WILD, 1989; TISHLER; ZANG, 1981b; WATTS; BACON, 1974).³ Neste trabalho, foram propostos os modelos suaves descritos a seguir:

Modelo 1.) Modelo de regressão hiperbólico de Griffiths e Miller (1973) e Seber e Wild (1989), dado por

$$\eta(x|\theta, \tau, \gamma) = \theta_0 + \theta_1 x + \theta_2 \sqrt{(x - \tau_1)^2 + \gamma_1} + \theta_3 \sqrt{(x - \tau_2)^2 + \gamma_2}, \quad (61)$$

em que $\gamma' = (\gamma_1, \gamma_2)$.

Modelo 2.) Modelo da formulação sinal com junção média do tipo I baseada na distribuição Logística, expresso por

$$\eta(x|\theta, \tau, \gamma) = \theta_0 + \theta_1 x + \theta_2 (x - \tau_1) [2F(x|\tau_1, \gamma_1) - 1] + \theta_3 (x - \tau_2) [2F(x|\tau_2, \gamma_2) - 1],$$

em que $F(x|\tau, \gamma)$ é a função de distribuição Logística com parâmetro de centro τ e parâmetro de dispersão γ .

Modelo 3.) Modelo da regressão “bent-cable” estendida com junção polinomial de 4º grau de Zang (1980), o qual é dado por

$$\eta(x|\beta_1, \tau, \delta, \zeta) = \alpha_1 + \beta_1 x + \delta_1 u_{\zeta_1}(x - \tau_1) + \delta_2 u_{\zeta_2}(x - \tau_2),$$

em que $\zeta' = (\zeta_1, \zeta_2)$ e

$$u_{\zeta}(x - \tau) = \psi^*(x, \tau, \zeta) I_{\{|x - \tau| \leq \zeta\}}(x) + (x - \tau) I_{\{|x - \tau| > \zeta\}}(x), \quad \zeta > 0,$$

³Mais detalhes sobre essas aproximações são disponibilizados no Capítulo 2

na qual

$$\psi^*(x, \tau, \zeta) = -\frac{(x - \tau)^4}{16 \zeta^3} + \frac{3(x - \tau)^2}{8 \zeta} + \frac{x - \tau}{2} + \frac{3 \zeta}{16}.$$

Modelo 4.) Modelo máx-mín com junção média do tipo I baseada na distribuição Logística, sendo expresso por

$$\eta(x | \beta_1, \tau, \delta, \gamma) = \alpha_1 + \beta_1 x + \delta_1(x - \tau_1)F(x | \tau_1, \gamma_1) + \delta_2(x - \tau_2)F(x | \tau_2, \gamma_2), \quad (62)$$

em que $F(x; \tau, \gamma)$ é a função de distribuição Logística com parâmetro de centro τ e parâmetro de dispersão γ .

Os parâmetros em $\gamma' = (\gamma_1, \gamma_2)$ e $\zeta' = (\zeta_1, \zeta_2)$ controlam a curvatura dos gráficos desses modelos nas regiões de transição de fase. Quando os parâmetros de suavização em γ e ζ tendem a zero, esses modelos tornam-se abruptos. Por outro lado, conforme os valores desses parâmetros aumentam, as transições de fase se tornam mais graduais.

Neste trabalho, os modelos abruptos nas expressões (57-60) não foram ajustados aos dados. O problema é que modelos abruptos possuem derivadas descontínuas nos pontos de mudança e não podem ser ajustados por meio das técnicas de otimização padrão para modelos não lineares. Neste trabalho optou-se por ajustar modelos abruptos com o auxílio dos modelos suaves, seguindo a sugestão de Tishler e Zang (1981). Modelos com transições de fase abruptas são obtidos diretamente pelos modelos 3.4.1, 3.4.1, 3.4.1 e 3.4.1, ao fixarmos os parâmetros de suavização em valores arbitrariamente pequenos, próximos de zero. Como as aproximações suaves garantem a continuidade das derivadas, a estimação não linear por meio de métodos que empregam derivadas torna-se possível. No entanto, se os valores dos parâmetros de suavização forem pequenos demais, as próprias aproximações suaves nas quais os modelos são baseados tornam-se não diferenciáveis e problemas no cálculo numérico das derivadas começam a aparecer. Neste trabalho, observou-se que modelos abruptos são obtidos, sem erros numéricos durante o processo de estimação, ao fixarmos os elementos de γ e ζ em 0,5.

3.4.2 Estimação por Máxima Verossimilhança e Testes de Hipóteses

Nesta seção é abordada a estimação dos parâmetros da curva de produção por meio do método da máxima verossimilhança. As observações da variável resposta Y são consideradas como não correlacionadas e distribuídas de acordo com a distribuição Normal

$$Y_i | x_i \sim N(\eta(x_i | \phi), \sigma^2) \quad (i = 1, \dots, n),$$

em que: $Y_i|x_i$ é a resposta associada a i -ésima observação e condicionada a dose de adubo x_i , $\eta(x_i|\phi)$ é a parte fixa do modelo e representa a produção esperada para a dose x_i e ϕ é o vetor dos parâmetros ligados à ela, sendo $\phi^T = (\beta_1^T, \delta^T, \tau^T, \gamma^T)$ ou $\phi^T = (\theta^T, \tau^T, \gamma^T)$, dependendo do modelo a ser ajustado.

A função de verossimilhança, por sua vez, é dada por

$$L(\phi, \sigma^2) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{[y_i - \eta(x_i|\phi)]^2}{2\sigma^2} \right\}, \quad (63)$$

em que y_i é a i -ésima resposta observada, mediante a dose x_i . Neste caso, o logaritmo da função de verossimilhança é expresso por

$$\ln L(\phi, \sigma^2) = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} S(\phi), \quad (64)$$

em que $S(\phi)$ é a soma de quadrados dos erros, isto é:

$$S(\phi) = \sum_{i=1}^n [y_i - \eta(x_i|\phi)]^2. \quad (65)$$

As estimativas de máxima verossimilhança para os parâmetros do modelo, denotadas por $\hat{\phi}$ e $\hat{\sigma}^2$, são obtidas maximizando o logaritmo da função de verossimilhança em (64) com relação a ϕ e σ^2 .

No caso de modelos não lineares e com erros normais, o estimador de máxima verossimilhança para σ^2 é independente de $\hat{\phi}$ (SEBER; WILD, 1989). Ele pode ser calculado analiticamente, ao solucionar a equação $\partial \ln L(\phi, \sigma^2) / \partial \sigma^2 = 0$ em relação a σ^2 . Esta igualdade implica

$$\hat{\sigma}^2 = \frac{S(\hat{\phi})}{n}.$$

Então os valores do vetor $\hat{\phi}$ correspondem aos que minimizam $S(\phi)$, dado um $\sigma^2 > 0$. Dessa maneira, os estimadores de máxima verossimilhança e mínimos quadrados para ϕ são equivalentes no caso Normal (SEBER; WILD, 1989).

Uma vez que a parte fixa do modelo é não linear, não há forma fechada para os estimadores de máxima verossimilhança e métodos iterativos devem ser empregados, os quais requerem valores iniciais dos parâmetros para o processo de estimação. Neste trabalho utilizou-se o método quasi-Newton de Broyden-Fletcher-Goldfarb-Shanno (BROYDEN, 1970; FLETCHER, 1970; GOLDFARB, 1970; SHANNO, 1970), adotando-se o esquema de dois passos para a obtenção de estimativas iniciais ϕ^0 .⁵ Foi utilizada a versão do algoritmo BFGS (Broyden-Fletcher-

⁵Mais detalhes sobre o procedimento de estimação podem ser encontrados no capítulo 2

Goldfarb-Shanno) implementada na função *optim*, do “software” estatístico R. Em associação ao *optim*, foram usadas as funções *mle* e *mle2*, dos pacotes *stats4* e *bbmle* (BOLKER, 2015), respectivamente. Apesar de as funções apresentarem resultados similares das estimativas, que diferiram apenas a partir da quarta casa decimal, aconselha-se o uso da função *mle2*, por ser uma implementação recente e mais robusta do método.

É importante notar que os parâmetros γ_1 e γ_2 (ou analogamente ζ_1 e ζ_2) são definidos em $\mathbb{R}_+$ e podem causar problemas de estimação quando são próximos de zero, pois encontram-se nos limites do espaço paramétrico Φ . Para evitar problemas de estimação, uma transformação corretiva foi adotada neste trabalho:

$$\gamma \text{ e } \zeta = \exp(w), \quad (66)$$

em que $w \in \mathbb{R}$. Assim, o processo de estimação torna-se irrestrito.

Uma das desvantagens dos modelos não lineares é que não há garantias da existência e/ou unicidade das estimativas por máxima verossimilhança ou mínimos quadrados. Há sempre a possibilidade de que o processo de estimação atinja ótimos locais (SEBER; WILD, 1989). Se os ótimos globais forem encontrados, entretanto, os estimadores de máxima verossimilhança possuirão algumas propriedades importantes. Uma delas é a de invariância: seja $\hat{\phi}$ o estimador de máxima verossimilhança e seja $g(\phi)$ qualquer transformação de ϕ , então o estimador de máxima verossimilhança de $g(\phi)$ é

$$g(\widehat{\phi}) = g(\hat{\phi}).$$

Esta propriedade é útil para estimar quantidades derivadas da curva de produção, como zonas de transição de fase, faixa ótima e faixas de suficiência.

Outras vantagens dos estimadores de máxima verossimilhança são suas propriedades limites: sob certas condições de regularidade, eles são consistentes, eficientes e apresentam normalidade assintótica, isto é,

$$\hat{\pi} \xrightarrow{D} N(\pi^*, \mathbf{I}^{-1}),$$

em que: $\pi^T = (\phi^T, \sigma^2)$, $\hat{\pi}$ é o estimador de máxima verossimilhança para π , π^* é o vetor dos verdadeiros valores dos parâmetros e $\mathbf{I}$ é a matriz de informação (esperada) de Fisher, para a qual as entradas são dadas por meio de

$$I(\pi)_{i,j} = -E\pi \left[\frac{\partial^2 \ln L(\pi)}{\partial \pi_i \partial \pi_j} \right],$$

com $i, j = 1, \dots, p$, sendo p o número de parâmetros em π .

Neste trabalho, a matriz de informação esperada foi aproximada pela matrix de informação observada $\mathbf{J}$, da qual as entradas são dadas por

$$J(\hat{\boldsymbol{\pi}}) = -\frac{\partial^2 \ln L(\boldsymbol{\pi})}{\partial \pi_i \partial \pi_j} \Big|_{\boldsymbol{\pi}=\hat{\boldsymbol{\pi}}}.$$

Observe que a matriz de informação observada $\mathbf{J}$ é a matriz Hessiana da função $\ln L$, com o sinal oposto, calculado com respeito ao vetor $\boldsymbol{\pi}$ e avaliado em $\hat{\boldsymbol{\pi}}$. Dessa forma, o estimador da matriz de variâncias e covariâncias assintóticas de $\hat{\boldsymbol{\pi}}$ foi obtido por meio de

$$\widehat{Var}(\hat{\boldsymbol{\pi}}) = [J(\hat{\boldsymbol{\pi}})]^{-1}$$

e, posteriormente, o erro-padrão de $\hat{\pi}_j$ foi estimado por meio de

$$\widehat{EP}(\hat{\pi}_j) = \sqrt{\widehat{Var}(\hat{\boldsymbol{\pi}})_{j,j}}.$$

Neste trabalho, a significância de cada parâmetro foi avaliada por meio do teste de Wald. No teste de Wald, as hipóteses $H_0: \pi_j = \pi_{0j}$ e $H_A: \pi_j \neq \pi_{0j}$ são investigadas por meio do valor da estatística

$$W = \frac{\hat{\pi}_j - \pi_{0j}}{\widehat{EP}(\hat{\pi}_j)},$$

a qual tem distribuição assintótica normal padrão sob H_0 . O teste de Wald também pode ser usado para testar hipóteses lineares, separada ou conjuntamente. Seja uma matriz $\mathbf{C}_{c \times p}$. Então, para testar as hipóteses lineares $H_0: \mathbf{C} \boldsymbol{\pi} = \mathbf{c}_0$ contra $H_A: \mathbf{C} \boldsymbol{\pi} \neq \mathbf{c}_0$, calcula-se a estatística

$$Q = (\mathbf{C} \hat{\boldsymbol{\pi}} - \mathbf{c}_0)' [\mathbf{C} \widehat{Var}(\hat{\boldsymbol{\pi}}) \mathbf{C}']^{-1} (\mathbf{C} \hat{\boldsymbol{\pi}} - \mathbf{c}_0),$$

a qual tem distribuição χ^2 com c graus de liberdade sob H_0 . Esse teste foi usado para testar se $\beta_1 + \delta_1 = 0$ nos modelos oriundos da formulação máx-mín, o que equivale a testar se o regime de consumo de luxo possui um coeficiente angular $\beta_2 = 0$.

A fim de investigar se há necessidade de estimar os parâmetros de suavização ou se modelos abruptos são mais parcimoniosos, utilizou-se o teste da razão de verossimilhanças para modelos encaixados. Os parâmetros da parte fixa dos modelos originais são elementos do vetor $\boldsymbol{\phi} \in \Phi$. Os modelos abruptos são obtidos ao fixarmos os parâmetros de suavização (presentes em $\boldsymbol{\phi}$) em valores arbitrariamente pequenos. Portanto, os modelos abruptos são casos particulares dos modelos suaves, com o vetor de parâmetros sendo $\boldsymbol{\phi}_0 \subset \boldsymbol{\phi}$. Os modelos abruptos são chamados de nulos, enquanto os modelos suaves são chamados de alternativos. Para testar se o

modelo nulo é mais parcimonioso do que o alternativo, calcula-se a estatística

$$\Lambda = -2 [\ln L(\hat{\phi}_0, \hat{\sigma}_0^2) - \ln L(\hat{\phi}, \hat{\sigma}^2)],$$

em que $\hat{\phi}_0$ e $\hat{\sigma}_0^2$ são as estimativas de máxima verossimilhança para os parâmetros ϕ_0 e σ_0^2 do modelo nulo. Sob o modelo nulo, a estatística razão de verossimilhanças Λ é assintoticamente distribuída de acordo com a distribuição χ^2 , com $(p - p_0)$ graus de liberdade, sendo p_0 a dimensão de ϕ_0 .

3.4.3 Quantidades de Interesse Agrícola

Todos os modelos apresentados na seção 3.4.1 permitem a estimação de quantidades de interesse agrícola, tais como os coeficientes dos regimes nutricionais, os níveis críticos e as faixas de suficiência. No entanto, os modelos suaves distinguem-se entre si pela interpretação biológica das transições de fase. Os modelos 3.4.1 e 3.4.1 são construídos por meio das junções médias do tipo I e, por esse motivo, permitem estimar a variabilidade intrínseca aos níveis críticos (capítulo 2). Mais do que isso, o uso de junções médias do tipo I permite obter as distribuições empíricas dos níveis críticos de suficiência e toxidez, dadas por

$$f(x | \hat{\tau}_1, \hat{\gamma}_1) = \frac{\exp\left(-\frac{x - \hat{\tau}_1}{\hat{\gamma}_1}\right)}{\hat{\gamma}_1 \left[1 + \exp\left(-\frac{x - \hat{\tau}_1}{\hat{\gamma}_1}\right)\right]^2} \quad \text{e} \quad f(x | \hat{\tau}_2, \hat{\gamma}_2) = \frac{\exp\left(-\frac{x - \hat{\tau}_2}{\hat{\gamma}_2}\right)}{\hat{\gamma}_2 \left[1 + \exp\left(-\frac{x - \hat{\tau}_2}{\hat{\gamma}_2}\right)\right]^2},$$

em que f é a função de densidade da Logística e $\hat{\tau}$ e $\hat{\gamma}$ são os vetores de estimativas para os níveis críticos e parâmetros de suavização, respectivamente. Por meio das distribuições empíricas dos níveis críticos, torna-se possível calcular qualquer região de interesse agrícola derivada das curvas de produção vegetal. Podem-se definir tais regiões por meio da função quantílica F^{-1} da distribuição Logística, ao se estipularem dois quantis quaisquer. Para os quantis de 30 e 70%, por exemplo, as zonas de transição de fase entre deficiência mineral e consumo de luxo e entre consumo de luxo e toxidez são expressas por

$$G_1 = \{F^{-1}(0, 30 | \hat{\tau}_1, \hat{\gamma}_1) \leq x \leq F^{-1}(0, 70 | \hat{\tau}_1, \hat{\gamma}_1)\}$$

e

$$G_2 = \{F^{-1}(0, 30 | \hat{\tau}_2, \hat{\gamma}_2) \leq x \leq F^{-1}(0, 70 | \hat{\tau}_2, \hat{\gamma}_2)\},$$

respectivamente. A faixa ótima, por sua vez, é expressa por

$$[LI, LS] = [F^{-1}(0, 30 | \hat{\tau}_1, \hat{\gamma}_1), F^{-1}(0, 70 | \hat{\tau}_2, \hat{\gamma}_2)]$$

e as faixas de suficiência inferior e superior são expressas por

$$FS_D = [NC_D, LI] = [\hat{\tau}_1, F^{-1}(0, 70 | \hat{\tau}_2, \hat{\gamma}_2)]$$

e

$$FS_T = [LS, NC_T] = [F^{-1}(0, 30 | \hat{\tau}_2, \hat{\gamma}_2), \hat{\tau}_2],$$

nas quais $NC_D = \hat{\tau}_1$ e $NC_T = \hat{\tau}_2$.

Observe que a escolha dos quantis depende de como o pesquisador deseja definir as zonas de transição de fase. Considerando-se os quantis de 30 e 70%, por exemplo, tem-se que: a zona de transição de fase entre a deficiência mineral e o consumo de luxo é o intervalo de valores de x para o qual o percentual de amostras deficientes, coletadas ao acaso, irá variar entre 30 e 70%; de modo análogo, a zona de transição de fase entre o consumo de luxo e a toxidez é o intervalo de valores de x para o qual o percentual de amostras, coletadas ao acaso, sob os efeitos de toxidez varia entre 30 e 70%.⁴

Por outro lado, os modelos 3.4.1 e 3.4.1 não permitem estimar a variabilidade dos níveis críticos. Somente os modelos construídos por meio de junções médias são passíveis desta interpretação. O modelo 3.4.1, no entanto, apresenta uma vantagem sobre o modelo 3.4.1: ele permite a delimitação de zonas de transição de fases de modo objetivo. Com base no modelo 3.4.1, as zonas de transição de fase entre deficiência mineral e consumo de luxo e entre consumo de luxo e toxidez são expressas por

$$G_1 = \{\hat{\tau}_1 - \hat{\zeta}_1 \leq x \leq \hat{\tau}_1 + \hat{\zeta}_1\} \text{ e } G_2 = \{\hat{\tau}_2 - \hat{\zeta}_2 \leq x \leq \hat{\tau}_2 + \hat{\zeta}_2\},$$

respectivamente; a faixa ótima, por sua vez, é expressa por

$$[LI, LS] = [\hat{\tau}_1 + \hat{\zeta}_1, \hat{\tau}_2 - \hat{\zeta}_2]$$

e as faixas de suficiência inferior e superior são expressas por

$$FS_D = [NC_D, LI] = [\hat{\tau}_1, \hat{\tau}_1 + \hat{\zeta}_1] \text{ e } FS_T = [LS, NC_T] = [\hat{\tau}_2 - \hat{\zeta}_2, \hat{\tau}_2],$$

respectivamente.

O modelo 3.4.1 não possui interpretações práticas semelhantes acerca das transições de fase, mas possui um gráfico suave nessas regiões, do qual a curvatura pode ser estimada por meio dos parâmetros de suavização.

⁴Os modelos 3.4.1 e 3.4.1 devem apresentar resultados similares para as quantidades de interesse agrícola, pois tratam-se de reparametrizações diferentes do mesmo modelo (ver capítulo 2)

Ao adotarmos o conceito fisiológico de níveis críticos, até mesmo com um modelo abrupto torna-se possível definir as regiões de interesse agrícola. Por exemplo, como zonas de transição de fase podem-se considerar os próprios intervalos de confiança associados aos níveis críticos, ou seja:

$$G = IC(\tau, 100 \times (1 - k)\%) : \hat{\tau} \pm Z_{1-k/2} \widehat{EP}(\hat{\tau}),$$

em que $100 \times (1 - k)\%$ é o nível de confiança, $Z_{1-k/2}$ é o $(1 - k/2)$ -ésimo quantil associado da distribuição normal padrão e $\widehat{EP}(\hat{\tau})$ é uma estimativa do erro padrão de $\hat{\tau}$. Neste caso, a faixa ótima é expressa como

$$[LI, LS] = [\hat{\tau}_1 + Z_{1-k/2} \widehat{EP}(\hat{\tau}_1), \hat{\tau}_2 - Z_{1-k/2} \widehat{EP}(\hat{\tau}_2)],$$

enquanto as faixas de suficiência são expressas por

$$FS_D = [\hat{\tau}_1, \hat{\tau}_1 + Z_{1-k/2} \widehat{EP}(\hat{\tau}_1)] \text{ e } FS_T = [\hat{\tau}_2 - Z_{1-k/2} \widehat{EP}(\hat{\tau}_2), \hat{\tau}_2].$$

3.5 Resultados e Discussão

Nesta seção são apresentados os resultados e discussão referentes aos ajustes dos modelos propostos aos dados de Schneeberger (2009), sobre a produtividade de trigo de inverno (100 kg/ha) em resposta a diferentes doses de nitrogênio (kg/ha) aplicadas ao solo.

A primeira etapa do método de obtenção de estimativas iniciais é comum a todos os modelos. No caso em estudo, utilizou-se o conhecimento prévio da existência de três regimes ($D = 3$) e dois pontos de mudança. O delineamento experimental adotado por Schneeberger (2009) possui $N = 12$ pontos de suporte e, portanto, foram realizadas

$$\begin{pmatrix} N - 2 \\ D - 1 \end{pmatrix} = \begin{pmatrix} 12 - 2 \\ 3 - 1 \end{pmatrix} = 45$$

regressões lineares ao todo. Nelas, foram considerados como pontos de mudança doses indo de 40 a 260 kg/ha. A melhor regressão em termos do quadrado médio do resíduo ($\hat{\sigma}^2 = 5,287$) é apresentada na Figura 15.

Nesta análise, os valores iniciais obtidos para os níveis críticos de deficiência e toxidez foram $\tau_1^0 = 80$ e $\tau_2^0 = 200$, respectivamente; enquanto isso, os parâmetros para os coeficientes dos regimes foram: $\alpha_1^0 = 64,1988$, $\beta_1^0 = 0,5265$, $\alpha_2^0 = 107,3461$, $\beta_2^0 = -0,0129$, $\alpha_3^0 = 148,8640$ e $\beta_3^0 = -0,2204$.

Na etapa 2 do método de obtenção de estimativas iniciais, as melhores estimativas para os

Figura 15 – Melhor regressão obtida na primeira etapa da obtenção de valores iniciais

parâmetros de suavização foram determinadas por meio da minimização da soma de quadrados dos erros, dada pela expressão (65). Os valores iniciais para os parâmetros de suavização são dispostos na Tabela 5.

Tabela 5 – Tabela de valores iniciais para os parâmetros de suavização obtidos por meio da minimização da soma quadrática dos erros

Modelo 3.4.1	Modelo 3.4.1	Modelo 3.4.1	Modelo 3.4.1
$\gamma_1^0 = 35,54$	$\gamma_1^0 = 10,52$	$\zeta_1^0 = 23,70$	$\gamma_1^0 = 10,52$
$\gamma_2^0 = 00,00$	$\gamma_2^0 = 01,39$	$\zeta_2^0 = 00,00$	$\gamma_2^0 = 01,22$

Após a obtenção das estimativas iniciais, cada um dos modelos propostos foi ajustado aos dados de Schneeberger (2009) pelo método BFGS. Os gráficos das curvas ajustadas e de diagnósticos são dispostos no anexo B. Visualmente, pode-se afirmar que todos os modelos ajustaram-se bem aos dados. Em ambos os casos, os valores preditos apresentaram-se associados às observações de maneira linear, o que evidencia a adequação dos modelos na tarefa de descrever a relação existente entre produtividade vegetal e dose de adubo. Em todos os ajustes, as pressuposições de normalidade e homogeneidade de variâncias não foram rejeitadas em nível de 5% de significância, pelos testes de Shapiro-Wilk e Bartlett, respectivamente (SHAPIRO; WILK, 1965; SNEDECOR; COCHRAN, 1989).

Verificou-se, também, que, para nenhum dos modelos estudados, o quadrado médio da falta de ajuste foi significativamente maior do que o do erro puro, por meio do teste F e em

nível de 5% de significância (NIELL1988).

A Figura 16 apresenta os gráficos dos modelos ajustados. Note que as curvas ajustadas são bem similares, sendo que quase não é possível distinguir seus gráficos.

Figura 16 – Gráficos dos ajustes dos modelos 1, 2, 3 e 4 aos dados de Schneeberger (2009) sobre a produtividade do trigo de inverno em função da adubação com nitrogênio

No entanto, a análise cuidadosa da Figura 16 revela que as curvas de resposta separam-se em dois grupos. Os gráficos dos modelos 3.4.1 e 3.4.1 se sobrepõem perfeitamente e isso já era previsto em teoria, visto que o modelo 3.4.1 é uma reparametrização do 3.4.1 (capítulo 2). Os gráficos dos modelos 3.4.1 e 3.4.1 apresentam uma saliência perto do nível crítico τ_1 , característica das junções médias do tipo I (capítulo 2). Por outro lado, os gráficos dos modelos 3.4.1 e 3.4.1 unem os regimes de forma parecida com a qual as funções hiperbólicas unem suas assíntotas.

Dessa maneira, os modelos 3.4.1 e 3.4.1 apresentam respostas sempre maiores do que as observadas pelo modelo abrupto, enquanto que os modelos 3.4.1 e 3.4.1 apresentam respostas menores (ver Capítulo 2). As diferenças entre o modelo abrupto e suas versões suaves só são perceptíveis próximo dos pontos de mudança, nas zonas de transição de fases.

As estimativas dos parâmetros para os modelos ajustados, juntamente com os erros-padrão e testes de Wald, são apresentadas nas Tabelas 6 e 7.

Tabela 6 – Estimativas por máxima verossimilhança dos parâmetros dos modelos 3.4.1 e 3.4.1, erros-padrão $EP(\hat{\phi}_m)$ aproximados e estatísticas obtidas para os testes de Wald

Parâmetro	Modelo 3.4.1				Modelo 3.4.1			
	$\hat{\phi}_m$	$\widehat{EP(\hat{\phi}_m)}$	W_{cal}	Valor-p	$\hat{\phi}_m$	$\widehat{EP(\hat{\phi}_m)}$	W_{cal}	Valor-p
$\phi_1 = \theta_0$	106,53	3,78	28,18	<0,001	103,53	6,32	16,37	<0,001
$\phi_2 = \theta_1$	0,19	0,05	3,51	<0,001	0,17	0,06	2,86	0,004
$\phi_3 = \theta_2$	-0,32	0,07	-4,76	<0,001	-0,27	0,05	-5,97	<0,001
$\phi_4 = \theta_3$	-0,09	0,03	-3,36	<0,001	-0,11	0,02	-4,81	<0,001
$\phi_5 = \tau_1$	74,47	9,30	8,00	<0,001	74,76	11,90	6,28	<0,001
$\phi_6 = \tau_2$	200,03	4,20	47,58	<0,001	200,10	6,31	31,71	<0,001
$\phi_7 = w_1$	5,91	2,18	2,71	0,007	3,29	0,57	5,74	<0,001
$\phi_8 = w_2$	-3,83	21,59	-0,18	0,859	1,04	3,60	0,29	0,773

Tabela 7 – Estimativas por máxima verossimilhança dos parâmetros dos modelos 3.4.1 e 3.4.1, erros-padrão $EP(\hat{\phi}_m)$ aproximados e estatísticas obtidas para os testes de Wald

Parâmetro	Modelo 3.4.1				Modelo 3.4.1			
	$\hat{\phi}_m$	$\widehat{EP(\hat{\phi}_m)}$	W_{cal}	Valor-p	$\hat{\phi}_m$	$\widehat{EP(\hat{\phi}_m)}$	W_{cal}	Valor-p
$\phi_1 = \alpha_1$	63,11	2,49	25,37	<0,001	60,51	5,84	10,37	<0,001
$\phi_2 = \beta_1$	0,60	0,10	6,18	<0,001	0,56	0,11	4,96	<0,001
$\phi_3 = \delta_1$	-0,62	0,11	-5,83	<0,001	-0,54	0,10	-5,54	<0,001
$\phi_4 = \delta_2$	-0,20	0,05	-3,89	<0,001	-0,23	0,05	-4,82	<0,001
$\phi_5 = \tau_1$	74,57	9,59	7,77	<0,001	74,10	12,76	5,81	<0,001
$\phi_6 = \tau_2$	200,03	4,08	48,98	<0,001	200,49	6,32	31,73	<0,001
$\phi_7 = w_1$	3,77	0,77	4,88	<0,001	3,30	0,59	5,57	<0,001
$\phi_8 = w_2$	-1,59	14,49	-0,11	0,913	0,86	6,27	0,14	0,891

Por meio dos coeficientes estimados, observou-se a existência de uma transição de fase suave entre os regimes de deficiência mineral e consumo de luxo, uma vez que o parâmetro de suavização w_1 apresentou valor maior do que zero pelo teste unilateral de Wald, em nível de 5% de significância.

Em contrapartida, as estimativas obtidas para w_2 não foram confiáveis. Os gráficos dos contornos da soma de quadrados dos erros, resultantes da segunda etapa de obtenção dos valores iniciais, evidenciam o mal-condicionamento deste parâmetro (Figura 17). O erro-padrão muito elevado faz com que o valor de w_2 não difira de zero pelo teste unilateral de Wald, em nível de 5% de significância. Enfim, não se pode concluir se a transição entre o consumo

Modelo 3.4.1

Modelo 3.4.1

Modelo 3.4.1

Modelo 3.4.1

Figura 17 – Gráfico dos contornos para a soma dos quadrados dos erros em função dos parâmetros de suavização

de luxo e a toxidez é processo abrupto ou gradual. É importante notar que ajustes de curvas abruptas não significam, necessariamente, transições de fase repentinas. Estudos de simulação de dados prévios indicaram que às vezes, por falta de informação, obtêm-se ajustes abruptos mesmo quando os verdadeiros valores dos parâmetros de suavização (γ e ζ) são consideravelmente maiores do que zero (capítulo 2). Tais resultados podem ser decorrentes do uso de um delineamento experimental inadequado (com poucos pontos de suporte e repetições nas zonas de transição) ou de uma amostra ruim.

Com o auxílio de um teste da razão de verossimilhanças para modelos encaixados foi possível comparar os modelos suaves originais (modelos alternativos) com modelos nos quais a transição de fase entre consumo de luxo e toxidez é abrupta (modelos nulos), os quais foram obtidos ao se fazer $\gamma_2 = 0,5$ e $\zeta_2 = 0,5$. Os resultados dos testes indicaram que não há diferença significativa entre a deviance residual dos modelos originais e nulos em nível de 5%

de significância, o que equivale a não rejeitar a hipótese de nulidade $H_0: \gamma_2 = 0,5$ testada contra a hipótese alternativa $H_A: \gamma_2 \neq 0,5$ (Tabela 8). Dessa forma, foi concluído que os parâmetros de suavização para a segunda transição de fase (γ_2 e ζ_2) não precisam ser estimados.

Tabela 8 – Teste das hipóteses $H_0: \gamma_2 = 0,5$ e $H_A: \gamma_2 \neq 0,5$ por meio da razão de verossimilhanças

Modelo	Λ_{cal}	Valor-p
Modelo 3.4.1	0,1252	0,7234
Modelo 3.4.1	0,0090	0,9244
Modelo 3.4.1	0,0628	0,8021
Modelo 3.4.1	0,0013	0,9708

Adicionalmente, o teste de Wald também foi usado para testar a hipótese de nulidade $H_0: \beta_1 + \delta_1 = 0$ contra a alternativa $H_A: \beta_1 + \delta_1 \neq 0$ nos modelos da formulação máx-mín (modelos 3.4.1 e 3.4.1). A hipótese de nulidade $H_0: \beta_1 + \delta_1 = 0$ equivale a afirmar que $\beta_2 = 0$, uma vez que $\delta_1 = \beta_2 - \beta_1$. As estatísticas de Wald calculadas a partir dos ajustes dos modelos 3.4.1 e 3.4.1 foram $Q_{cal_1} = 0,517$ e $Q_{cal_2} = 0,281$, respectivamente. Comparando-se esses resultados com o quantil crítico $\chi^2_{(0,95,1)} = 3,841$ da distribuição qui-quadrado, não se rejeita a hipótese $H_0: \beta_1 + \delta_1 = 0$ em nível de 5% de significância. Logo, pode-se afirmar que o valor do coeficiente angular do segundo regime β_2 não difere de zero e, conseqüentemente, a planta não responde ao adubo durante o consumo de luxo, tal como é previsto por teoria.

Em face aos resultados obtidos, optou-se por ajustar versões simplificadas dos modelos estudados. Nos modelos simplificados, considerou-se uma transição de fase abrupta entre consumo de luxo e toxidez e o parâmetro β_2 foi fixado em zero. Os modelos simplificados foram obtidos por meio dos modelos abruptos dispostos nas expressões (59) e (60), ao serem substituídos os operadores sinal e u pelas aproximações suaves respectivas, de maneira análoga ao que pode ser verificado nas expressões (61-62).

Nas Tabelas 9 e 10 são apresentadas as estimativas para os parâmetros dos modelos simplificados e seus erros-padrão. Todos os modelos simplificados tiveram seus parâmetros estimados com precisão, excetuando-se o parâmetro γ_1 do modelo 3.4.1. Dessa maneira, por meio da junção hiperbólica não se pôde concluir se a transição de fase entre consumo de luxo e toxidez é abrupta ou gradual. Por outro lado, os ajustes dos modelos 3.4.1, 3.4.1 e 3.4.1 evidenciaram que essa transição é gradual, uma vez que os parâmetros de suavização apresentaram-se significativamente maiores do que zero pelo teste unilateral de Wald, em nível de 5% de significância.

Os modelos baseados nas junções médias (3.4.1 e 3.4.1) apresentaram as menores margens de erro para as estimativas dos parâmetros, com exceção do intercepto α_1 . A estimação mais

Tabela 9 – Estimativas por máxima verossimilhança dos parâmetros dos modelos 3.4.1 e 3.4.1 simplificados, erros-padrão $EP(\hat{\phi}_m)$ aproximados e estatísticas obtidas para os testes de Wald

Parâmetro	Modelo 3.4.1				Modelo 3.4.1			
	$\hat{\phi}_m$	$EP(\hat{\phi}_m)$	W_{cal}	Valor-p	$\hat{\phi}_m$	$EP(\hat{\phi}_m)$	W_{cal}	Valor-p
$\phi_{0_1} = \alpha_1$	63,42	2,44	26,02	<0,001	61,86	4,10	15,09	<0,001
$\phi_{0_2} = \beta_1$	0,58	0,07	8,06	<0,001	0,53	0,06	8,31	<0,001
$\phi_{0_3} = \beta_3$	-0,23	0,03	-8,31	<0,001	-0,21	0,03	-6,43	<0,001
$\phi_{0_4} = \tau_1$	72,78	6,69	10,88	<0,001	78,84	5,92	13,33	<0,001
$\phi_{0_5} = \tau_2$	199,50	5,42	36,84	<0,001	200,23	3,60	55,57	<0,001
$\phi_{0_6} = w_1$	4,32	3,01	1,43	0,152	3,17	0,58	5,46	<0,001

Tabela 10 – Estimativas por máxima verossimilhança dos parâmetros dos modelos 3.4.1 e 3.4.1 simplificados, erros-padrão $EP(\hat{\phi}_m)$ aproximados e estatísticas obtidas para os testes de Wald

Parâmetro	Modelo 3.4.1				Modelo 3.4.1			
	$\hat{\phi}_m$	$EP(\hat{\phi}_m)$	W_{cal}	Valor-p	$\hat{\phi}_m$	$EP(\hat{\phi}_m)$	W_{cal}	Valor-p
$\phi_{0_1} = \alpha_1$	63,16	2,51	25,15	<0,001	61,87	4,08	15,15	<0,001
$\phi_{0_2} = \beta_1$	0,59	0,08	6,94	<0,001	0,53	0,06	8,34	<0,001
$\phi_{0_3} = \beta_3$	-0,23	0,03	-8,54	<0,001	-0,21	0,03	-6,45	<0,001
$\phi_{0_4} = \tau_1$	72,27	7,97	9,07	<0,001	78,86	5,90	13,36	<0,001
$\phi_{0_5} = \tau_2$	199,50	5,17	38,56	<0,001	200,09	3,55	56,45	<0,001
$\phi_{0_6} = w_1$	3,49	0,85	4,11	<0,001	3,17	0,58	5,45	<0,001

Tabela 11 – Intervalos de 95% confiança para os níveis críticos (em kg/ha) de deficiência e toxidez, τ_1 e τ_2 , e critério da informação de Akaike, AIC

Modelos	$IC(\tau_1, 95\%)$	$IC(\tau_2, 95\%)$	AIC
Modelo 3.4.1	[59,68; 85,89]	[188,88; 210,11]	329,571
Modelo 3.4.1	[67,25; 90,44]	[193,17; 207,29]	328,496
Modelo 3.4.1	[56,67; 87,90]	[189,36; 209,64]	329,298
Modelo 3.4.1	[67,29; 90,43]	[193,15; 207,04]	328,501

precisa dos níveis críticos τ_1 e τ_2 por meio dos modelos com junções médias culminou em intervalos de confiança mais estreitos para os mesmos (Tabela 11). Além disso, os modelos 3.4.1 e 3.4.1 apresentaram os menores valores de AIC. Por meio dos níveis críticos e seus intervalos de confiança foi possível estimar algumas das quantidades de interesse agrícola mencionadas anteriormente (Figura 18).

Por exemplo, ao ser considerado o modelo de menor AIC (modelo 3.4.1), têm-se que as

Figura 18 – Ajuste do modelo 3.4.1 aos dados de Schneeberger (2009) e delimitação das zonas de transição de fase por meio dos intervalos de 95% de confiança para os níveis críticos de deficiência e toxidez

zonas de transição de fase são dadas por

$$G_1 = [67,25; 90,44] \text{ e } G_2 = [193,17; 207,29];$$

a faixa ótima é dada por

$$[LI, LS] = [\hat{\tau}_1 + 1,96 \times \widehat{EP}(\hat{\tau}_1), \hat{\tau}_2 - 1,96 \times \widehat{EP}(\hat{\tau}_2)] = [90,44, 193,17];$$

e as faixas de suficiência são dadas por

$$FS_D = [\hat{\tau}_1, \hat{\tau}_1 + 1,96 \times \widehat{EP}(\hat{\tau}_1)] = [78,84; 90,44]$$

e

$$FS_T = [\hat{\tau}_2 - 1,96 \times \widehat{EP}(\hat{\tau}_2), \hat{\tau}_2] = [193,17; 200,23].$$

O modelo 3.4.1, entretanto, é construído por meio de junções médias do tipo I e, dessa forma, as zonas de transição de fase e as faixas de suficiência também podem ser obtidas a

Figura 19 – Ajuste do modelo 3.4.1 aos dados de Schneeberger (2009) e delimitação da zona de transição de fase entre deficiência mineral e toxidez por meio dos quantis 30 e 70% da função de distribuição empírica $f(x; \hat{\tau}_1, \hat{\gamma}_1)$ (à esquerda). Probabilidade de transição de fase estimada entre deficiência mineral e toxidez em função da dose de nitrogênio aplicada (à direita)

partir das distribuições empíricas dos níveis críticos. Neste exemplo, no entanto, o parâmetro de suavização da segunda transição de fase não pôde ser estimado com precisão e, por esse motivo, apenas a transição entre deficiência mineral e consumo de luxo foi considerada (Figura 19). Com o auxílio das estimativas $\hat{\tau}_1 = 78,84$ e $\hat{\gamma}_1 = 23,95$, tem-se que a zona de transição de fase entre deficiência mineral e consumo de luxo, considerando-se os quantis de 30 e 70% da função de distribuição empírica, é dada por

$$G_1 = \{x \in \mathbb{R} \mid F^{-1}(0,30 \mid \hat{\tau}_1, \hat{\gamma}_1) \leq x \leq F^{-1}(0,70 \mid \hat{\tau}_1, \hat{\gamma}_1)\} = [58,15; 99,13];$$

enquanto isso, a faixa de suficiência inferior é expressa por

$$FS_D = [NC_D, LI] = [\hat{\tau}_1, F^{-1}(0,70 \mid \hat{\tau}_1, \hat{\gamma}_1)] = [78,84; 99,13].$$

Os resultados que foram obtidos por meio do modelo 3.4.1 são equivalentes aos do modelo 3.4.1, uma vez que as junções médias do tipo I, quando aplicadas para substituir o operador de máximo e a função sinal, resultam na mesma expressão (capítulo 2).

Alternativamente, as zonas de transição de fase e faixas de suficiência podem ser obtidas com o auxílio do modelo da regressão “bent-cable” estendida (Figura 20). Dadas as estimativas $\hat{\tau}_1 = 72,27$ e $\hat{\zeta}_1 = 32,79$, tem-se que a zona de transição de fase entre deficiência mineral e consumo de luxo é dada por

$$G_1 = [\hat{\tau}_1 - \hat{\zeta}_1, \hat{\tau}_1 + \hat{\zeta}_1] = [39,48; 105,06],$$

Figura 20 – Ajuste do modelo da regressão “bent-cable” estendida (modelo 3.4.1) aos dados de Schneeberger (2009) e delimitação da zona de transição de fase entre deficiência mineral e toxidez por meio da estimativa $\hat{\zeta}_1$

enquanto a faixa de suficiência inferior é expressa por

$$FS_D = [NC_D, LI] = [\hat{\tau}_1, \hat{\tau}_1 + \hat{\zeta}_1] = [72, 27; 105, 06].$$

3.6 Conclusão

Os modelos lineares segmentados com transições suaves foram flexíveis o suficiente para descrever a relação entre produtividade e dose de nitrogênio ao longo dos diferentes estados nutricionais. Os modelos se ajustaram bem aos dados de Schneeberger (2009), fornecendo estimativas precisas dos parâmetros de cada regime linear, bem como dos níveis críticos. Com os modelos também foi possível estimar, de várias maneiras diferentes, quantidades de interesse agrícola que servem como padrões para avaliação do estado nutricional das culturas e recomendação de doses de adubo.

Agradecimentos

Agradecemos ao auxílio do Prof. Francisco Antônio Monteiro, do Departamento de Ciência do Solo (LSO - ESALQ) da Universidade de São Paulo. O presente trabalho foi realizado com apoio do CNPq, Conselho Nacional de Desenvolvimento Científico e Tecnológico - Brasil.

Referências

- BACON, D.W.; WATTS, D.G. Estimating the transition between two intersecting straight lines. **Biometrika**, Oxford, v. 58, n. 3, p. 525-534, 1971.
- BERCK, P.; HELFAND, G. Reconciling the von Liebig and differentiable crop production functions. **American Journal of Agricultural Economics**, Oxford, v. 72, n. 4, p. 985-996, 1990.
- BOLKER, B; **Tools for general maximum likelihood estimation**. R Development Core Team. Disponível em: <<http://cran.r-project.org/web/packages/bbmle/bbmle.pdf>>. Acesso em: 16 junho 2015.
- BROYDEN, C.G. The convergence of a class of double-rank minimization algorithms. **Journal of the Institute of Mathematics and Its Applications**, Essex, v. 6, p. 76-90, 1970.
- CERRATO, M.E.; BLACKMER, A.M. Comparison of models for describing corn yield response to nitrogen fertilizer. **Agronomy Journal**, Madison, v. 82, p. 138-143, 1990.
- CHIU, G.; LOCKHART, R.; ROUTLEDGE, R. Bent-cable regression theory and applications. **Journal of the American Statistical Association**, Alexandria, v. 101, n. 474, p. 542-553, 2006.
- FAQUIN, V. **Diagnose do estado nutricional de plantas**. Lavras: UFLA e FAEPE, 2002. 77p.
- FLETCHER, R. A new approach to variable metric algorithms. **Computer Journal**, Oxford, v. 13, n. 3, p. 317-322, 1970.
- FLOSS, E.L. **Fisiologia das plantas cultivadas: o estudo que está por trás do que se vê**. 3.ed. Passo Fundo: Editora da Universidade de Passo Fundo, 2006. 536 p.
- GOLDFARB, D. A family of variable metric updates derived by variational means. **Mathematics of Computation**, Providence, v. 24, n. 109, p. 23-26, 1970.
- GRADES, E. des. **Comparison of DRIS and critical level approach for evaluating nutrition status of wheat in District Hyderabad, Pakistan**. 2011. Tese (Doutorado em Agronomia) – der Hohen Landwirtschaftlichen Fakultät, der Rheinischen Friedrich-Wilhelms-Universität, Sindh, Paquistão, 2011. Disponível em: <<http://hss.ulb.uni-bonn.de/2012/2638/2638.pdf>>. Acesso em: 24 abril 2015.
- GRIFFITHS, D.A.; MILLER, A.J. Hyperbolic regression - a model based on two-phase piecewise linear regression with a smooth transition between regimens. **Communications in Statistics**, Abingdon, v. 2, n. 6, p. 561-569, 1973.
- HAVLIN, J.L.; BEATON, J.D.; TISDALE, S.L.; NELSON, W.L. **Soil fertility and fertilizers: an introduction to nutrient management**. 7.ed. New Jersey: Pearson, 2005. 515 p.

JONES JR, J.B. Modern interpretation systems for soil and plant analyses in the United States of America. **Australian Journal of Experimental Agriculture**, Clayton South, v. 33, n. 8, p. 1039-1043, 1993.

LIEBIG, J. von. **Organic chemistry in its applications to agriculture and physiology**. London: Taylor and Walton, 1840. 387 p.

LOPES, A.C.; SOUSA, D.M.G. Interpretation of microbial soil indicators as a function of crop yield and organic carbon. **Soil Science Society of America Journal**, Madison, v. 77, p. 461-472, 2013.

MACY, P. The quantitative mineral requirements of plants. **Plant Physiology**, Bethesda, v. 11, n. 4, p. 749-764, 1936.

MAIA, C.E.; MORAIS, E.R.C.; OLIVEIRA, M. Nível crítico pelo critério da distribuição normal reduzida: uma nova proposta para interpretação de análise foliar. **Revista Brasileira de Engenharia Agrícola e Ambiental**, Campina Grande, v. 5, p. 235-238, 2001.

MALAVOLTA, E. **Manual de química agrícola - adubos e adubação**. 2.ed. São Paulo: Editora Agronômica CERES, 1967. 606 p.

MALAVOLTA, E.; VITTI, G.C.; OLIVEIRA, S.A. **Avaliação do estado nutricional de plantas: princípios e aplicações**. Piracicaba: Associação Brasileira para Pesquisa da Potassa e do Fosfato, 1989. 201 p.

MALAVOLTA, E. **Nutrição mineral e adubação do cafeeiro**. São Paulo: Editora Agronômica CERES, 1993. 210 p.

MAUST, B.E.; WILLIAMSON, J.G. Nitrogen nutrition of containerized citrus nursery plants. **Journal of the American Society for Horticultural Science**, Alexandria, v. 119, n. 2, p. 195-201, 1994.

MEMON, N.; MEMON, K.S.; ZIA-UL-HASSAN. Plant analysis as a diagnostic tool for evaluating nutritional requirements of bananas. **International Journal of Agriculture & Biology**, Faisalabad, v. 7, n. 5, p. 824-831, 2005.

MILES, N. Challenges and opportunities in leaf nutrient data interpretation. In: SOUTH AFRICAN SUGAR TECHNOLOGISTS' ASSOCIATION, South Africa. **Proceedings...** South Africa: SASTA, p. 205-215, 2010.

MITSCHERLICH, E. A. Das gesetz des minimums und das gesetz des abnehmenden bodenertrages. **Landwirtschaftliche Jahrbücher**, Berlim, v. 38, p. 537-552, 1909.

MITSCHERLICH, E.A. Die zweite annäherung des wirkungsgesetzes der wachstumsfaktoren. **Zeitschrift für Pflanzenernährung**, Berlim, v. 12, p. 273-282, 1928.

NEVES, S.C.N.; CARVALHO, J.G.; FERREIRA, E.V.O.; ASSIS, R.P. Nutrição mineral, crescimento e níveis críticos foliares de cálcio e magnésio em mudas de umbuzeiro, em função da calagem. **Revista Ceres**, Viçosa, v. 55, n. 6, p. 575-583, 2008.

NIELL, J.W. Testing for lack of fit in nonlinear regression. **The Annals of Statistics**, Beachwood, v. 16, n. 2, p. 733-740, 1988.

OHKI, K. Critical nutrient levels related to plant growth and some physiological processes. **Journal of Plant Nutrition**, Athens, v. 10, p. 1583-1673, 1987.

PARIS, Q. The von Liebig hypothesis. **American Journal of Agricultural Economics**, Oxford, v. 74, n. 4, p. 1019-1028, 1992a.

- PARIS, Q. The return of von Liebig's law. **Agronomy Journal**, Madison, v. 84, p. 1040-1046, 1992b.
- PIMENTEL-GOMES, F.; NOGUEIRA, I.R. **Análise matemática**. 2.ed. Piracicaba: Departamento de Ciências Exatas da ESALQ, 1980. 371 p.
- PREVOT, P.; OLLAGNIER, M. Méthode d'utilisation du diagnostic foliaire. In: **Analyses des plantes et problèmes des fumures minérales**. Paris: Ed. IRHO, 1956, p. 167-192.
- PRUNTY, L. Curve fitting with smooth functions that are piecewise-linear in the limit, **Biometrics**, Arlington, v. 39, n. 4, p. 857-866, 1983.
- REID, J.B. Yield response to nutrient supply across a wide range of conditions 1. Model derivation. **Field Crops Research**, Amsterdã, v. 77, p. 161-171, 2002.
- R CORE TEAM. **R: a language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria, 2015. Disponível em: <<http://www.r-project.org>>. Acesso em: 24 abril 2015.
- SANCHEZ, C.A.; LOCKHART, M.; PORTER, P.S. Response of radish to phosphorus and potassium fertilization on Histosols. **HortScience**, Alexandria, v. 26, n. 1, p. 30-32, 1991.
- SCHERER, E.E. Níveis críticos de potássio para a soja em latossolo húmico de Santa Catarina, **Revista Brasileira de Ciência do Solo**, Viçosa, v. 22, p. 57-62, 1998.
- SCHNEEBERGER, H. **Over-fertilization: an inverse Mitscherlich process**. Institute of Statistics, University of Erlangen-Nürnberg. Disponível em: <<http://soil-statistic.de/overfertilization/overfertilization.html>>. Acesso em: 24 abril 2015.
- SEBER, G.A.F.; WILD, C.J. **Nonlinear regression**. United States: John Wiley & Sons, Inc., 1989. 768 p.
- SHANNO, D.F. Conditioning of quasi-Newton methods for function minimization. **Mathematics of Computation**, Providence, v. 24, n. 111, p. 647-656, 1970.
- SHAPIRO, S.S.; WILK, M.B. An analysis of variance test for normality (complete samples). **Biometrika**, Oxford, v. 52, n. 3-4, p. 591-611, 1965.
- SMITH, P.F. Mineral analysis of plant tissues. **Annual Review of Plant Physiology**, Florida, v. 13, p. 81-108, 1962.
- SNEDECOR, G.W.; COCHRAN, W.G. **Statistical Methods**. 3.ed. Ames: Iowa State University Press, 1989. 487 p.
- STERNER, R.W.; ELSER, J.J. **Ecological stoichiometry: the biology of elements from molecules to the biosphere**. Princeton, New Jersey: Princeton University Press, 2002. 464 p.
- TAIZ, L.; ZEIGER, E. **Fisiologia vegetal**. 3.ed. Porto Alegre: Artmed, 2004. 719 p.
- TISHLER, A.; ZANG, I. A new maximum likelihood algorithm for piecewise regression. **Journal of the American Statistical Association**, Alexandria, v. 76, n. 376, p. 980-987, 1981.
- ULRICH, A. Physiological bases for assessing the nutritional requirements of plants. **Annual Review of Plant Physiology**, Florida, v. 3, p. 207-228, 1952.
- WADT, P.G.S.; ALVAREZ, V. H.; NOVAIS, R. F.; FONSECA, S.; BARROS, N. F. O método da chance matemática na interpretação de dados de levantamento nutricional de eucalipto. **Revista Brasileira de Ciência do Solo**, Viçosa, v. 22, p. 773-778, 1998.

WATTS, D.G.; BACON, D.W. Using an hyperbola as a transition model to fit two-regime straight-line data. **Technometrics**, Alexandria, v. 16, n. 3, p. 369-373, 1974.

WIKIPEDIA, THE FREE ENCYCLOPEDIA. **Liebig's law of the minimum**. Disponível em: <http://en.wikipedia.org/wiki/Liebig%27s_law_of_the_minimum>. Acesso em: 24 abril 2015.

ZANG, I; A smoothing-out technique for min-max optimization. **Mathematical Programming**, Philadelphia, v. 19, n. 1, p. 61-77, 1980.

4 AS LEIS DE LIEBIG E MITSCHERLICH E A FRAÇÃO DE APROVEITAMENTO DO NUTRIENTE LIMITANTE NA PLANTA

Resumo

Neste trabalho, a resposta vegetativa à adubação é modelada por meio da lei de Liebig, sob a pressuposição de uma distribuição desigual do nutriente limitante e de suas exigências nutricionais dentro da planta. É postulado que a resposta vegetativa é proporcional à fração do nutriente limitante existente na planta que é aproveitada nos processos de síntese, durante o crescimento e o desenvolvimento do vegetal. Como resultados da modelagem matemática, obtêm-se diversas funções de produção que, ao mesmo tempo, são interpretáveis por meio das leis do mínimo de Liebig e dos retornos decrescentes de Mitscherlich. Muitas funções de uso comum na agricultura surgem como casos particulares da metodologia adotada e novos modelos também são obtidos. Em face dos resultados verificados neste trabalho, uma nova interpretação para a lei de Mitscherlich emerge: quanto maior é o conteúdo total do nutriente limitante na planta, menor é a fração que se aproveita de um incremento em sua disponibilidade a fim de prover produção vegetal. Como consequência, os retornos obtidos por meio da fertilização são decrescentes.

Palavras-chave: Resposta vegetativa; Fertilização; Funções de produção; Lei do mínimo; Retornos decrescentes

Abstract

In this work, the yield response to fertilizers is modeled by means of Liebig's law, under the premise of a heterogeneous distribution of the limiting nutrient and its nutritional requirements inside the plants. It is stated that the yield response is proportional to the fraction of the limiting nutrient used in synthesis processes, during growth and development of the vegetal. Modeling the yield response to fertilizers allows to construct several production functions, which are interpretable by means of the laws of the minimum and of the diminishing returns, when they are considered jointly. Many production functions that are widely used in agriculture appear as particular cases of the adopted approach and new production models are also derived. Indeed, a new interpretation for the Mitscherlich's law emerges in the context addressed by this work: it is verified that the greater is the amount of the limiting nutrient inside the plant, the lower is the fraction of exploitation of an increment in its availability in order to provide yield. As a result, returns from fertilization are diminishing.

Keywords: Yield response; Fertilization; Production functions; Law of the minimum; Law of diminishing returns

4.1 Introdução

Diversos fatores podem limitar o crescimento dos vegetais, tais como o clima, a falta de água ou luminosidade, a incidência de pragas e doenças, etc. No entanto, o mais comum de se verificar em plantas cultivadas é limitação por falta de um elemento mineral essencial (MALAVOLTA, 1980).

Para estudar os efeitos das deficiências minerais e como as plantas respondem aos adu-

bos, os pesquisadores realizam os experimentos de adubação. A relação entre produção vegetal e dose de adubo é crucial para a agricultura de precisão, pois o conhecimento a seu respeito possibilita o acompanhamento do programa de fertilização corrente, a elaboração de estratégias corretivas para o solo e o uso racional dos adubos, favorecendo a obtenção de colheitas produtivas (MALAVOLTA, 1967; MALAVOLTA et al., 1989).

No entanto, o debate sobre qual função de produção descreve melhor a relação entre produção e dose de adubo já perdura por mais de um século e meio, sem definição (PARIS, 1992). O debate se prolonga porque, geralmente, as funções de produção vegetal são baseadas em duas leis da nutrição mineral de plantas, a lei do mínimo de Liebig e a lei dos retornos decrescentes de Mitscherlich, vistas por muitos autores como teorias conflitantes (BERCK; HELFAND, 1990; LIEBIG, 1840; MITSCHERLICH, 1909).

A lei do mínimo de Liebig postula que, a cada tempo, um único nutriente limita o crescimento da planta e a produção é diretamente relacionada a sua disponibilidade. De acordo com a mesma, o elemento mineral limitante é aquele que se encontra em quantidade relativa mínima na planta e, por esse motivo, controla a sua produção (PARIS, 1992). A lei de Liebig é muitas vezes associada às funções de produção do tipo linear-platô (BERCK; HELFAND, 1990; CERRATO; BLACKMER, 1990). De acordo com essas funções, a produção cresce linearmente até que se atinja uma dose crítica, sendo que, a partir desse ponto, a produção deixa de responder ao adubo e a curva se estagna em um platô. A dose crítica determina a suficiência da planta pelo elemento, determinando o momento em que outro fator de produção torna-se limitante.

Em contraposição a lei do mínimo, Mitscherlich postulou a lei dos retornos decrescentes, a qual afirma que, conforme as doses de adubo aumentam, os retornos obtidos com um incremento na disponibilidade do nutriente limitante diminuem (MITSCHERLICH, 1909). O conflito entre as duas teorias se concentra no fato das funções do tipo linear-platô não apresentar retornos decrescentes, verificados repetidamente nos ensaios de adubação.

Recentemente, Berck e Helfand (1990) apresentaram uma proposta para reconciliar as leis do mínimo e dos retornos decrescentes. Eles observaram que funções de produção do tipo linear-platô, quando associadas a uma distribuição heterogênea dos nutrientes (ou de outros fatores de produção) no campo, produzem curvas de produção esperada côncavas, continuamente diferenciáveis e que podem ser interpretadas pela lei dos retornos decrescentes. Berck e Helfand mostraram, ainda, que funções de produção usadas frequentemente na área de adubação podem ser obtidas por meio das funções do tipo linear-platô, considerando-se distribuições específicas dos insumos no campo. Por exemplo, a função de produção de Cobb-Douglas é relacionada com a distribuição de Pareto, a função de produção quadrática é relacionada com a distribuição Uniforme e o modelo de Mitscherlich é relacionado com a distribuição Exponencial. Assim,

de acordo com Berck e Helfand, não há nenhum conflito entre as funções do tipo linear-platô de von Liebig e a lei dos retornos decrescente, isso quando considera-se uma distribuição heterogênea de pelo menos um dos insumos no campo.

Este trabalho segue uma linha similar a de Berck e Helfand (1990). A lei do mínimo de Liebig é aplicada em nível celular, para explicar a relação existente entre a produção vegetal e o conteúdo total do nutriente limitante na planta. Neste contexto, somente uma fração do nutriente existente dentro da planta é responsável pelo crescimento, que é a fração aproveitada nas suas reações de síntese. Além disso, assume-se uma distribuição desigual do nutriente limitante e das exigências nutricionais dentro da planta.

Observa-se que diferentes distribuições dos níveis críticos de suficiência das células caracterizam tipos distintos de resposta vegetativa à adubação, o que conduz a diferentes funções de produção vegetal. As funções de produção vegetal mais comuns na área de adubação são obtidas como casos particulares da metodologia abordada. Como exemplos, podem ser citadas as funções lineares, lineares-platô, quadráticas-platô e o modelo de Mitscherlich. Novas funções de produção, mais flexíveis do que o modelo de Mitscherlich, também são obtidas pela metodologia proposta.

Apesar da modelagem matemática proposta neste trabalho ser baseada na lei de Liebig e nos conceitos estequiométricos que a fundamentam (STERNER; ELSER, 2002), algumas das funções de produção resultantes são interpretáveis por meio da lei dos retornos decrescentes. Em resumo, esse trabalho introduz uma nova forma de reconciliar as leis de Liebig e Mitscherlich, ao serem incorporados nos modelos matemáticos níveis críticos (de suficiência pelo elemento mineral limitante) variáveis entre diferentes partes da planta.

Como motivação para a prática, os modelos de resposta vegetativa estudados neste trabalho são ajustados aos dados de Hoagland e Martin (1933), a fim de explicar como incrementos no conteúdo total de potássio (K) afetam a produção de tomateiros.

4.2 Metodologia

Como ponto de partida, admite-se que x é a dose do nutriente limitante aplicada no solo; x_0 é a quantidade deste nutriente disponível à planta e pré-existente no solo e a_0 é a quantidade do nutriente pré-existente na planta. Admite-se, também, que a quantidade de nutriente limitante assimilada pela planta é proporcional a sua quantidade disponível no solo, $x + x_0$. Assim, o conteúdo total do nutriente limitante na planta, q , é determinado por

$$q(x) = a_0 + b(x + x_0),$$

em que b ($0 \leq b \leq 1$) é constante e representa a fração do nutriente disponível no solo que é assimilada pela planta.

Dessa forma, o conteúdo total do nutriente limitante na planta pode ser expresso como

$$q(x) = (a_0 + b x_0) + b x = a + b x, \quad (51)$$

em que $a = a_0 + b x_0$ representa a quantidade do nutriente pré-existente na planta, acrescida da fração assimilada pela planta do nutriente pré-existente no solo.

Em nutrição mineral de plantas é comum descrever a produção vegetal como função do conteúdo percentual do nutriente limitante nas plantas ou do teor do elemento (em p.p.m.) em órgãos específicos, como as folhas (MALAVOLTA et al., 1989). O conteúdo percentual, $n(q)$, do nutriente é obtido por meio da razão entre q e a massa da matéria-seca, m , isto é:

$$n(q) = \frac{q}{m} \times 100\%.$$

A produção vegetal y , por sua vez, pode ser medida de muitas formas, sendo uma delas a própria massa da matéria-seca (raiz e parte aérea) e, mesmo quando não é expressa como massa da matéria-seca, indubitavelmente é uma variável correlata ($y = y(m)$). A massa da matéria-seca, no entanto, também é associada ao conteúdo total do nutriente limitante, isto é, $m = m(q)$. Então, para fins de modelagem matemática, não é viável descrever a produção vegetal $y(m)$ como uma função do conteúdo percentual do elemento limitante, pois essa definição de produção é cíclica e só será útil se conhecermos previamente a forma da relação $m(q)$.

Como consequência, a modelagem da produção vegetal y em função do conteúdo total de nutriente limitante na planta, $q(x)$, é uma escolha mais natural. A quantidade do nutriente na planta é a chave para a determinação da produção vegetal, pois trata-se da matéria-prima das reações de síntese que produzem compostos orgânicos associados ao crescimento. Logo, sem incrementos na disponibilidade do nutriente limitante, o organismo simplesmente para de crescer (LIEBIG, 1840; PARIS, 1992).

De maneira reducionista, a produção vegetal é comparável a uma reação química. Nas reações químicas, o total de produto formado é determinado pelo reagente que acaba primeiro, ou seja, o que encontra-se na menor quantidade relativa. Além disso, de acordo com a lei de Proust, os compostos químicos sempre contêm a mesma proporção em massa dos elementos que os constituem. Uma consequência imediata à lei de Proust é o fato de que o total de produto formado é proporcional à quantidade do reagente em escassez.

Na realidade, o que importa para determinar a produção vegetal não é o conteúdo total do nutriente limitante na planta, mas somente a fração que é aproveitada pelo metabolismo

das células e que, por meio das reações de síntese, transforma-se em produtos para a planta. A fração excedente não provém crescimento à planta; serve apenas para elevar o teor do elemento nos tecidos ou, então, é armazenada em órgãos de reserva a fim de que seja utilizada em momentos futuros (STERNER; ELSER, 2002). Por esse motivo, a fração excedente pode ser negligenciada quando deseja-se modelar a produção vegetal.

Neste trabalho é apresentada, primeiramente, a modelagem matemática da resposta vegetativa à adubação, isto é, a taxa de variação da produção em relação à dose de adubo x ou ao conteúdo do nutriente na planta q ($\frac{dy}{dx}$ ou $\frac{dy}{dq}$). Na prática, não há muita diferença entre se utilizar uma definição ou outra, pois, por meio da expressão (51), pode-se notar que as quantidades $\frac{dy}{dx}$ e $\frac{dy}{dq}$ são proporcionais. Nesta etapa, considera-se o fato de que a planta é constituída por uma mistura de células deficientes e bem nutridas, com respeito ao elemento sob estudo. Este fato torna-se evidente quando, por exemplo, estuda-se a progressão dos sinais visuais de deficiências minerais em diferentes partes da planta (MALAVOLTA, 1980). Assim, a resposta vegetativa da planta à adubação é o somatório das respostas vegetativas de cada célula que a constitui.

Considera-se que células no estado de deficiência mineral são reativas ao adubo e aproveitam todo acréscimo no nutriente limitante para prover crescimento à planta. Por outro lado, células no estado de nutrição mineral adequado não são reativas, ou seja, apresentam respostas vegetativas nulas. Com base na lei de Proust, ainda é conveniente supor que as células deficientes apresentam uma resposta vegetativa (em termos de matéria-seca) constante, denotada por k e denominada por eficácia do adubo, o que significa que há um retorno de k unidades de massa da matéria-seca a cada unidade de massa do elemento limitante acrescida às células deficientes.¹

Considere, agora, um incremento Δq no conteúdo total do elemento limitante na planta, $q(x)$. No contexto abordado, faz sentido assumir que a fração de Δq aproveitada nos processos de síntese corresponde à proporção de células no estado de deficiência mineral, aqui denotada por S , $0 \leq S \leq 1$. Dessa forma, o incremento resultante na produção é de $\Delta y = k S \Delta q$. Na forma diferencial, tem-se que

$$\frac{dy}{dq} = k S, \quad (51)$$

em que $k > 0$. Mas, afinal, como determinar S ?

Para responder a questão acima, considera-se o fato de que o conteúdo total do nutriente

¹Observe que isto equivale a aplicar a lei do mínimo de Liebig em nível celular. Na versão original da lei de Liebig, o nutriente limita o crescimento de toda a planta. Aqui, por outro lado, o nutriente é responsável por limitar a produção de cada célula em particular. Observe, ainda, que cada célula transforma o nutriente limitante em produção de acordo com uma função do tipo linear-platô. Assim, se todas as células da planta apresentassem o mesmo estado de nutrição mineral, a resposta da planta como um todo também seria do tipo linear-platô.

na planta, $q(x)$, determina se uma célula, em particular, está bem nutrida ou não. Em outras palavras, para cada célula há um nível crítico q^* para o conteúdo total de nutriente na planta que determina sua suficiência pelo elemento, de maneira que se $q(x) < q^*$, a célula encontra-se deficiente; caso contrário, ela situa-se no estado nutricional adequado com respeito ao elemento sob estudo.

Levando-se em consideração que as exigências nutricionais mudam de acordo com o tecido e que a distribuição do elemento na planta pode ser desigual, então os valores de q^* devem variar entre as células. Logo, algumas partes da planta necessitam de um conteúdo total do nutriente $q(x)$ maior do que outras para atingir a suficiência. Essa variabilidade é incorporada ao modelo por meio de níveis críticos Q^* aleatórios. Assim, o nível crítico q^* associado a uma célula, em particular, é uma realização da variável aleatória Q^* . Por sua vez, a fração de células deficientes na planta é prontamente obtida por meio do cálculo de probabilidades e corresponde à função de sobrevivência da variável aleatória Q^* :

$$S = S_{Q^*}(q) = 1 - P(Q^* \leq q) = 1 - F_{Q^*}(q),$$

em que $F_{Q^*}(\cdot)$ é a função de distribuição acumulada de Q^* .

Diferentes funções de produção podem, portanto, serem obtidas pela escolha de diferentes distribuições para Q^* . Seja a resposta vegetativa expressa por

$$\frac{dy}{dq} = k S_{Q^*}(q | \boldsymbol{\theta}), \quad (51)$$

em que $q \geq 0$ e $\boldsymbol{\theta}$ é o vetor de parâmetros associados à distribuição de Q^* . Então, a função de produção correspondente é obtida pela resolução da equação diferencial ordinária (E.D.O.) separável disposta em (51). A E.D.O. é resolvida por meio de integração, da seguinte maneira:

$$\int_{y_0}^{y(x)} dy = k \int_0^{a+bx} S_{Q^*}(q | \boldsymbol{\theta}) dq = k [IS(a + bx | \boldsymbol{\theta}) - IS(0 | \boldsymbol{\theta})], \quad (52)$$

em que y_0 é a produção inicial e IS é uma primitiva da função de sobrevivência S_{Q^*} com respeito a q . Dado que o elemento mineral em questão é essencial para o desenvolvimento e o crescimento da planta, admite-se que y_0 é desprezível quando o conteúdo total do nutriente limitante na planta $q(x)$ é nulo. Logo, a função de produção é dada por

$$y(x) = k [IS(a + bx | \boldsymbol{\theta}) - IS(0 | \boldsymbol{\theta})], \quad x \geq -a/b. \quad (53)$$

A expressão (53) estabelece uma relação unívoca entre a função de produção $y(x)$, que depende da dose, e a resposta vegetativa dy/dq . Alguns exemplos da relação entre resposta vegetativa e função de produção são apresentados a seguir, considerando-se diferentes distribuições

de Q^* .

4.2.1 Função de Produção Linear

Considere, inicialmente, o caso em que todas as células da planta encontram-se deficientes, independentemente da dose ou do conteúdo total do nutriente limitante na planta. Em outras palavras, $S_{Q^*}(q) = 1 \forall q \geq 0$.

Nesta situação, a resposta vegetativa é constante, sendo dada por:

$$\frac{dy}{dq} = k, \quad k > 0.$$

Por meio da expressão (53), constata-se que

$$y(x) = k(a + bx), \quad x \geq -a/b.$$

Logo, a resposta vegetativa constante é ligada de forma unívoca a uma função de produção linear.

4.2.2 Função de Produção Linear-platô

Considera-se, agora, o caso em que as células da planta transitam do regime da deficiência mineral para o estado de nutrição adequado repentinamente, sob o mesmo valor crítico de conteúdo total do nutriente limitante na planta, denotado por $q^* = q_c$, de maneira que há uma distribuição homogênea do nutriente limitante e das exigências nutricionais dentro da planta.

Nesta situação, as funções de densidade e sobrevivência para a variável aleatória Q^* são dadas, respectivamente, por

$$f_{Q^*}(q^* | q_c) = \Delta(q^*) = \begin{cases} \infty & q^* = q_c \\ 0 & q^* \neq q_c \end{cases}, \quad (52)$$

e

$$S_{Q^*}(q | q_c) = \mathbf{I}_{\{q < q_c\}}(q), \quad (53)$$

sendo que caracterizam a distribuição delta de Dirac para Q^* . Como consequência, a resposta vegetativa é expressa por

$$\frac{dy}{dq} = k \mathbf{I}_{\{q < q_c\}}(q), \quad (54)$$

para $q > 0$. Consequentemente, a função de produção, obtida por meio da expressão (53), é

dada por

$$y(x) = \begin{cases} k(a + bx), & -\frac{a}{b} \leq x < \frac{q_c - a}{b}, \\ k q_c, & x \geq \frac{q_c - a}{b}. \end{cases} \quad (55)$$

Logo, a produção cresce linearmente em resposta ao adubo até que a dose crítica $x_c = \frac{q_c - a}{b}$ seja atingida e, depois, a planta não responde mais a adubação. A partir da dose crítica x_c , a produção se estagna em seu valor máximo $y_{max} = k(a + b x_c) = k q_c$.

Observa-se que a resposta vegetativa disposta em (54) é ligada de forma unívoca a uma função de produção do tipo linear-platô, muitas vezes referida como função de von Liebig (BERCK; HELFAND, 1990; CERRATO; BLACKMER, 1990; PARIS, 1992).

De acordo com o modelo matemático introduzido nesse trabalho, verifica-se que a função linear-platô é um caso especial das funções de von Liebig quando (i) não há variabilidade dos níveis críticos q^* e (ii) todo o acréscimo no conteúdo do nutriente limitante na planta é aproveitado em suas reações de síntese durante o estado nutricional da deficiência mineral.

4.2.3 Função de Produção Quadrática-Platô

Considera-se o caso em a variável aleatória Q^* tem distribuição Uniforme, com funções de densidade e sobrevivências dadas por,

$$f_{Q^*}(q^* | k, j) = \frac{j}{k} I_{[0, k/j]}(q^*)$$

e

$$S_{Q^*}(q^* | k, j) = \left(1 - \frac{j q^*}{k}\right) I_{[0, k/j]}(q^*)$$

respectivamente, nas quais j é uma constante positiva. Nesta situação, a resposta vegetativa decresce linearmente até que uma quantidade crítica máxima do conteúdo total de nutriente limitante na planta $q_c = k/j$ seja atingida e, depois, torna-se nula.

Dessa forma, a resposta vegetativa é expressa por

$$\frac{dy}{dq} = k \left(1 - \frac{j q}{k}\right) I_{[0, k/j]}(q).$$

Consequentemente, a função de produção é dada por

$$y(x) = k(a + bx) - \frac{j}{2}(a + bx)^2, \text{ se } -\frac{a}{b} \leq x < \frac{k}{j b} - \frac{a}{b}, \quad (53)$$

e $y_{max} = k^2/2j$ se $x \geq \frac{k}{j b} - \frac{a}{b}$, em que y_{max} é o valor máximo de produção (ver Figura 21).

Figura 21 – Curvas da resposta vegetativa linear (à esquerda) e da função de produção quadrática-platô (à direita) para valores diferentes do parâmetro j . Os outros parâmetros foram fixados em: $a = 0,2$, $b = 0,8$ e $k = 100$

A função de produção apresentada na expressão (53) é o modelo quadrático-platô de Galant e Fuller (1973), outro modelo que é de grande uso nos estudos de adubação (CERRATO; BLACKMER, 1990).

4.2.4 Função de Produção de Mitscherlich

Considera-se que a variável aleatória Q^* tem distribuição Exponencial, com funções de densidade e sobrevivência dadas por

$$f_{Q^*}(q^* | \lambda) = \lambda e^{-\lambda q^*}$$

e

$$S_{Q^*}(q^* | \lambda) = e^{-\lambda q^*},$$

respectivamente, nas quais $\lambda > 0$ e $q^* \geq 0$.

Dessa forma, a resposta vegetativa é exponencial, sendo expressa por

$$\frac{dy}{dq} = k e^{-\lambda q}. \quad (52)$$

Por meio da expressão (53), obtém-se a seguinte função de produção:

$$y(x) = \frac{k}{\lambda} [1 - e^{-\lambda(a+bx)}], \quad x \geq -a/b. \quad (53)$$

A função de produção acima é uma reparametrização do modelo de Mitscherlich, na qual k/λ

é a produção máxima e λb é o efeito de Mitscherlich (CERRATO; BLACKMER, 1990; MITSCHERLICH, 1909). Logo, há uma relação unívoca entre a resposta vegetativa Exponencial, a distribuição exponencial para Q^* e o modelo de Mitscherlich (ver Figura 22).

Com isso, no contexto deste estudo, surge uma nova interpretação para a lei de Mitscherlich: quanto maior é o conteúdo total do nutriente limitante na planta, menor é o aproveitamento que se faz de um incremento em sua disponibilidade a fim de prover produção, o que culmina em retornos decrescentes.

Figura 22 – Curvas da resposta vegetativa exponencial (à esquerda) e da função de produção de Mitscherlich (à direita) para valores diferentes do parâmetro λ . Os outros parâmetros foram fixados em: $a = 0,2$, $b = 0,8$ e $k = 100$

O modelo original de Mitscherlich (1909) baseia-se na seguinte taxa de variação:

$$\frac{dy}{dx} = \rho(A - y), \quad \rho > 0, \quad (54)$$

em que o parâmetro ρ é o efeito de Mitscherlich e A é a produção máxima. O modelo de Mitscherlich é formulado a partir da premissa de que os retornos obtidos com a adubação são decrescentes, proporcionais à diferença entre a produção máxima A e a produção corrente $y(x)$. Na versão original deste modelo, o parâmetro ρ é apenas uma constante de proporcionalidade.

O modelo de Mitscherlich é obtido ao resolvermos a equação diferencial separável disposta na expressão (54), o que resulta em:

$$\int \frac{dy}{A - y} = \rho \int dx \Rightarrow y(x) = A - e^{-\rho x + C},$$

em que C é a constante de integração.

A fim de determinar a constante de integração, considera-se o fato de que sem o nutriente

mineral essencial a planta não cresce e se desenvolve. Pela expressão (51), isso implica que $y(-a/b) = 0$. Logo, a constante de integração é $C = \ln A - \rho a/b$ e, conseqüentemente, o modelo de Mitscherlich é expresso por

$$y(x) = A \left[1 - e^{-\rho(x+\frac{a}{b})} \right], \quad (54)$$

em que a fração a/b representa a parcela do nutriente assimilado que encontrava-se presente no solo sem adubação (CERRATO; BLACKMER, 1990; PIMENTEL-GOMES; NOGUEIRA, 1980). Ao serem feitas as substituições $\rho = \lambda b$ e $A = k/\lambda$ na expressão (54), obtém-se a reparametrização do modelo de Mitscherlich disposta na expressão (53).

Algumas distinções entre o modelo original de Mitscherlich e a reparametrização obtida nessa seção devem ser mencionadas: Mitscherlich obteve o modelo que recebe seu nome ao relacionar a resposta vegetativa à dose de adubo de forma empírica. É importante notar que no seu trabalho não há menção sobre como o elemento mineral essencial causa o crescimento da planta. A taxa de variação proposta por Mitscherlich depende de parâmetros que possuem interpretações práticas, mas que não estão ligados diretamente aos aspectos fisiológicos que regem a resposta à adubação. Já a versão do modelo de Mitscherlich proposta nesta seção apresenta interpretações fisiológicas claras para todos os parâmetros. No entanto, deve-se notar que o modelo original de Mitscherlich é mais geral, no sentido de que descreve a produção em função de qualquer fator que possa limitá-la. Enquanto isso, o modelo proposto nessa seção aplica-se somente quando o fator limitante for um mineral essencial para a planta.

4.2.5 Novas Funções de Produção

Anteriormente, foi verificado que para cada distribuição dos níveis críticos Q^* existe uma função de produção relacionada. Nesta seção são apresentadas duas novas funções de produção, obtidas por meio do uso de distribuições mais flexíveis do que a exponencial para modelagem da resposta vegetativa.

4.2.5.1 Funções de produção e a distribuição de Weibull

Considera-se que a variável aleatória Q^* tem distribuição de Weibull, com funções de densidade e sobrevivência dadas por

$$f_{Q^*}(q^* | \lambda, r) = r \lambda (\lambda q^*)^{r-1} e^{-(\lambda q^*)^r}$$

$$S_{Q^*}(q^* | \lambda, r) = e^{-(\lambda q^*)^r},$$

com $\lambda, r > 0$ e $q^* \geq 0$. Desse modo, a resposta vegetativa é expressa por

$$\frac{dy}{dq} = k e^{-(\lambda q)^r}.$$

A partir da expressão (52), verifica-se que

$$y(x) = \int_0^{a+bx} k e^{-(\lambda q)^r} dq.$$

A integral é resolvida por meio da substituição $t = (\lambda q)^r$, de maneira que $dt = \lambda r t^{1-\frac{1}{r}} dq$, sendo a função de produção expressa por

$$y(x) = \frac{k}{\lambda r} \int_0^{\lambda^r (a+bx)^r} t^{\frac{1}{r}-1} e^{-t} dt = \frac{k \gamma(1/r, \lambda^r (a+bx)^r)}{\lambda r}, \quad x \geq -a/b,$$

em que

$$\gamma(s, z) = \int_0^z t^{s-1} e^{-t} dt$$

é a função gama incompleta (inferior).

Neste caso,

$$\lim_{x \rightarrow +\infty} y(x) = \frac{k \Gamma(\frac{1}{r})}{\lambda r},$$

em que Γ denota a função gama. Assim, tem-se que a reta $y = \frac{k \Gamma(\frac{1}{r})}{\lambda r}$ é a assíntota horizontal do gráfico da função de produção $y(x)$, representando a produção assintótica máxima (ver Figura 23).

4.2.5.2 Funções de produção e a distribuição de Erlang

Considera-se que a variável aleatória Q^* tem distribuição de Erlang, com funções de densidade e sobrevivência dadas por

$$f_{Q^*}(q^* | \lambda, r) = \frac{\lambda^r q^{*r-1} e^{-\lambda q^*}}{(r-1)!}$$

e

$$S_{Q^*}(q^* | \lambda, r) = 1 - \frac{\gamma(r, \lambda q^*)}{(r-1)!} = \sum_{n=0}^{r-1} \frac{1}{n!} e^{-\lambda q^*} (\lambda q^*)^n,$$

Figura 23 – Curvas da resposta vegetativa (à esquerda) e da função de produção (à direita) obtidas pela distribuição de Weibull, com valores diferentes dos parâmetros λ e r . Os outros parâmetros foram fixados em: $a = 0,2$, $b = 0,8$ e $k = 100$

em que: $r \in \mathbb{N} - \{0\}$, $\lambda > 0$, $q^* \geq 0$. Então, a resposta vegetativa é expressa por

$$\frac{dy}{dq} = k \left[1 - \frac{\gamma(r, \lambda q)}{(r-1)!} \right] = k \sum_{n=0}^{r-1} \frac{1}{n!} e^{-\lambda q} (q \lambda)^n.$$

Consequentemente, a função de produção é expressa por

$$y(x) = k \sum_{n=0}^{r-1} \frac{1}{n!} \int_0^{a+bx} e^{-\lambda q} (q \lambda)^n dq.$$

Por meio da substituição $t = \lambda q$, obtêm-se que $dt = \lambda dq$ e

$$y(x) = \frac{k}{\lambda} \sum_{n=0}^{r-1} \frac{1}{n!} \int_0^{\lambda(a+bx)} t^n e^{-t} dt = \frac{k}{\lambda} \sum_{n=0}^{r-1} \frac{\gamma(n+1, \lambda(a+bx))}{n!}, \quad x \geq -a/b.$$

Neste caso,

$$\lim_{x \rightarrow +\infty} y(x) = \frac{k}{\lambda} \sum_{n=0}^{r-1} \frac{\Gamma(n+1)}{n!} = \frac{k r}{\lambda},$$

de maneira que a reta $y = \frac{k r}{\lambda}$ é assíntota horizontal do gráfico de $y(x)$ e determina a produção a produção assintótica máxima.

Observa-se que o modelo de Erlang apresenta como caso particular o modelo de Mitscherlich quando $r = 1$, o que decorre do fato da distribuição exponencial ser um caso particular da distribuição de Erlang (ver Figura 24).

Figura 24 – Curvas da resposta vegetativa (à esquerda) e da função de produção (à direita) obtidas pela distribuição de Erlang, com valores diferentes dos parâmetros λ e r . Os outros parâmetros foram fixados em: $a = 0,2$, $b = 0,8$ e $k = 100$

4.2.5.3 Funções de produção e a distribuição Gama

Considera-se que a variável aleatória Q^* tem distribuição de Gama, com funções de densidade e sobrevivência dadas por

$$f_{Q^*}(q^* | \lambda, r) = \frac{\lambda^r q^{*r-1} e^{-\lambda q^*}}{\Gamma(r)} \text{ e } S_{Q^*}(q^* | \lambda, r) = 1 - \frac{\gamma(r, \lambda q^*)}{\Gamma(r)}, \quad (42)$$

em que: $r, \lambda > 0$ e $q^* \geq 0$. Se o parâmetro r é inteiro e positivo, tem-se o caso particular da distribuição Erlang, abordada previamente.

A expansão em séries de potências para a função gama incompleta inferior é dada por

$$\gamma(s, z) = \Gamma(s) \sum_{n=0}^{\infty} \frac{z^{s+n} e^{-z}}{\Gamma(s+n+1)}, \quad z \geq 0, \quad (43)$$

de maneira que a resposta vegetativa à adubação passa a ser expressa por

$$\frac{dy}{dq} = k \left[1 - \frac{\gamma(r, \lambda q)}{\Gamma(r)} \right] = k \left[1 - \sum_{n=0}^{\infty} \frac{(\lambda q)^{r+n} e^{-\lambda q}}{\Gamma(r+n+1)} \right]. \quad (44)$$

Logo, a função de produção é expressa por

$$y(x) = k \left[(a + bx) - \sum_{n=0}^{\infty} \int_0^{a+bx} \frac{(\lambda q)^{r+n} e^{-\lambda q}}{\Gamma(r+n+1)} dq \right], \quad x \geq -\frac{a}{b}, \quad (45)$$

se a soma no integrando é convergente. Por meio da substituição $t = \lambda q$, tem-se que $dt = \lambda dq$

e

$$y(x) = k(a + bx) - \frac{k}{\lambda} \sum_{n=0}^{\infty} \frac{\gamma(r + n + 1, \lambda(a + bx))}{\Gamma(r + n + 1)}, \quad x \geq -\frac{a}{b}. \quad (46)$$

A função de produção (46) é difícil de se implementar nos algoritmos de estimação, pois seu valor depende de uma soma infinita, a qual deve ser truncada. Além disso, seu limite diverge quando a dose x tende ao infinito, o que não faz sentido prático. Portanto, neste trabalho não se recomenda o uso da distribuição Gama para a variável Q^* .

4.2.5.4 Funções de produção e a distribuição Logística

Considera-se que a variável aleatória Q^* tem distribuição Logística, com funções de densidade e sobrevivência dadas por

$$f_{Q^*}(q^* | \mu, s) = \frac{e^{-\frac{(q^* - \mu)}{s}}}{s \left(1 + e^{-\frac{(q^* - \mu)}{s}}\right)^2}$$

e

$$S_{Q^*}(q^* | \mu, s) = \frac{e^{-\frac{(q^* - \mu)}{s}}}{1 + e^{-\frac{(q^* - \mu)}{s}}},$$

em que $s > 0$ e $q^*, \mu \in \mathbb{R}$. Neste caso, a resposta vegetativa é expressa por

$$\frac{dy}{dq} = \frac{k e^{-\frac{(q - \mu)}{s}}}{1 + e^{-\frac{(q - \mu)}{s}}}, \quad (45)$$

o que resulta na função de produção

$$y(x) = k s \left[\ln \left(1 + e^{\frac{\mu}{s}}\right) - \ln \left(1 + e^{-\frac{(a + bx - \mu)}{s}}\right) \right], \quad x \geq -\frac{a}{b}. \quad (46)$$

Neste caso, a função de produção pode ser considerada como um modelo de transição suave constituído por dois regimes lineares, uma vez que as retas $y = k s \ln(1 + e^{\frac{\mu}{s}})$ e $y = k [s \ln(1 + e^{\frac{\mu}{s}}) - \mu + a] + k b x$ são as assíntotas horizontal e inclinada do gráfico de $y(x)$ (ver Figura 25), conforme demonstra-se abaixo:

i) Assíntota horizontal:

$$\lim_{x \rightarrow +\infty} y(x) = k s \left[\ln \left(1 + e^{\frac{\mu}{s}}\right) - \ln \left(1 + e^{-\infty}\right) \right] = k s \ln \left(1 + e^{\frac{\mu}{s}}\right).$$

ii) Assíntota inclinada:

A reta $y = \beta_0 + \beta_1 x$ é assíntota inclinada do gráfico de $y(x)$, em que β_0 e β_1 são coeficientes

Figura 25 – Curvas da resposta vegetativa (à esquerda) e da função de produção (à direita) obtidas pela distribuição Logística, com valores diferentes dos parâmetros μ e s . Os outros parâmetros foram fixados em: $a = 0,2$, $b = 0,8$ e $k = 100$

reais, somente se uma das condições a seguir for satisfeita:

$$(a) \quad \lim_{x \rightarrow -\infty} [y(x) - (\beta_0 + \beta_1 x)] = 0 \quad (46)$$

ou

$$(b) \quad \lim_{x \rightarrow +\infty} [y(x) - (\beta_0 + \beta_1 x)] = 0.$$

No cálculo adiante considera-se apenas o infinito negativo, visto que já fora encontrada uma assíntota horizontal ao se analisar o comportamento da função quando $x \rightarrow +\infty$. A fim de que seja atendida a condição (a) disposta em (46), a igualdade a seguir deve valer:

$$\lim_{x \rightarrow -\infty} \left[\beta_1 x + k s \ln \left(1 + e^{-\frac{(a+bx-\mu)}{s}} \right) \right] = k s \ln \left(1 + e^{\frac{\mu}{s}} \right) - \beta_0.$$

Uma vez que $x \rightarrow -\infty$, a aproximação

$$\ln \left(1 + e^{-\frac{(a+bx-\mu)}{s}} \right) \approx -\frac{(a+bx-\mu)}{s}$$

pode ser aplicada para resolver o problema, o que resulta em

$$\lim_{x \rightarrow -\infty} \left[\beta_1 x + k s \ln \left(1 + e^{-\frac{(a+bx-\mu)}{s}} \right) \right] \approx \lim_{x \rightarrow -\infty} (\beta_1 - k b) x + k (\mu - a). \quad (44)$$

O limite em (44) terá valor finito se e somente se $\beta_1 = k b$ e, portanto, esse é o único valor

possível para o coeficiente angular da assíntota inclinada. Dado β_1 , tem-se que

$$k s \ln \left(1 + e^{\frac{\mu}{s}} \right) - \beta_0 \approx k (\mu - a),$$

de maneira que o único valor admissível para β_0 seja

$$\beta_0 \approx k \left[s \ln \left(1 + e^{\frac{\mu}{s}} \right) - \mu + a \right].$$

4.2.5.5 Nova família de distribuições e funções de produção flexíveis

Nesta seção é apresentada uma nova família de distribuições, útil para a construção de funções de produção vegetal flexíveis e com poucos parâmetros (ver Figura 26).

Considera-se a seguinte distribuição para Q^*

$$f_{Q^*}(q^* | q_c, t) = \frac{t}{q_c} \left(1 - \frac{q^*}{q_c} \right)^{t-1},$$

em que: $t > 0$ e $0 \leq q^* \leq q_c$, sendo que o parâmetro q_c representa o valor máximo para o nível crítico q^* .

Figura 26 – Curvas das funções de densidade (à esquerda) e sobrevivência (à direita) da nova família de distribuições proposta para os níveis críticos Q^* , considerando-se $q_c = 2,0$ e diversos valores para o parâmetro t

Neste caso, a função de sobrevivência de Q^* é dada por

$$S_{Q^*}(q^* | q_c, t) = \left(1 - \frac{q^*}{q_c} \right)^t \mathbf{I}_{[0, q_c]}(q^*),$$

Então, a resposta vegetativa é expressa por

$$\frac{dy}{dq} = k \left(1 - \frac{q}{q_c}\right)^t I_{[0, q_c]}(q).$$

Consequentemente, a função de produção é dada por

$$y(x) = \frac{k q_c}{t+1} - \frac{k (q_c - a - bx)^{t+1}}{q_c^t (t+1)}, \quad -\frac{a}{b} \leq x \leq \frac{q_c - a}{b}, \quad (40)$$

e $y(x) = \frac{k q_c}{t+1}$ se $x > \frac{q_c - a}{b}$ (ver Figura 27). Neste caso, $x_c = \frac{q_c - a}{b}$ é a dose crítica, a partir da qual a planta não responde mais à adubação.

Nota-se que as funções de produção do tipo linear-platô e quadrática-platô são casos particulares da expressão (40):

i) Linear-platô ($t \rightarrow 0$)

$$y(x) = k(a + bx), \quad -\frac{a}{b} \leq x \leq \frac{q_c - a}{b},$$

e $y(x) = k q_c$ se $x \geq \frac{q_c - a}{b}$.

ii) Quadrático-platô ($t = 1$)

$$y(x) = k(a + bx) - \frac{k}{2q_c} (a + bx)^2, \quad -\frac{a}{b} \leq x \leq \frac{q_c - a}{b}, \quad (40)$$

e $y(x) = \frac{k q_c}{2}$ se $x \geq \frac{q_c - a}{b}$.

Observe que o resultado obtido na expressão (40) é equivalente ao que se verifica na fórmula (53) quando toma-se $q_c = \frac{j}{k}$.

Figura 27 – Curvas da resposta vegetativa (à esquerda) e da função de produção (à direita) obtidas pela nova família de distribuições, obtidas com $q_c = 2,0$ e variando-se t . Os outros parâmetros foram fixados em: $a = 0,2$, $b = 0,8$ e $k = 100$

4.3 Uma Aplicação dos Modelos Propostos

Neste trabalho, os modelos propostos para a resposta vegetativa são ajustados aos dados de Hoagland e Martin (1933) sobre a produção de tomateiros em função da adubação com potássio (K). Os dados de Hoagland e Martin apresentam a produção de tomate (em gramas de matéria-seca) obtida em diferentes solos, com e sem a aplicação de potássio, durante variadas estações e em um período total de 4 anos. Os valores dos conteúdos percentuais do potássio são disponibilizados por Macy (1936), que obteve essa informação via comunicação pessoal do próprio Prof. D. R. Hoagland.

A partir do conteúdo percentual de potássio (denotado por K%) é possível calcular o conteúdo total de potássio presente na matéria-seca de cada amostra (K Total). A comparação entre parcelas que, nos mesmos solos e safras, receberam e não receberam o adubo com potássio permite estimar a resposta vegetativa. Adota-se a quantidade

$$\frac{\Delta y}{\Delta q} = \frac{\text{Prod}_2 - \text{Prod}_1}{\text{K Total}_2 - \text{K Total}_1}, \quad (41)$$

como uma aproximação da verdadeira resposta vegetativa (dy/dq) em cada uma das combinações de solo e safra. Na expressão (41), os valores Prod_2 e Prod_1 são as produções médias, em gramas de matéria-seca, observadas nas parcelas adubadas e não adubadas, respectivamente. Já os termos K Total_2 e K Total_1 são os conteúdos totais de potássio na planta, obtidos nas parcelas adubadas e não adubadas, respectivamente (ver Tabela 12).

Neste trabalho, as respostas vegetativas são ajustadas em função do conteúdo médio de potássio nas plantas adubadas e não adubadas, calculado como

$$q_m = \frac{\text{K Total}_2 + \text{K Total}_1}{2}.$$

Para os ajustes, utilizam-se modelos não lineares clássicos, expressos por

$$\left(\frac{\Delta y}{\Delta q}\right)_i = k S_{Q^*}(q_{m_i} | \boldsymbol{\theta}) + \epsilon_i,$$

com $i = 1, 2, \dots, N$, em que $N = 19$ é o tamanho amostral. Assumem-se erros aleatórios ϵ_i normalmente distribuídos e não correlacionados, com média nula e variância comum igual a σ^2 .

Os modelos são ajustados por meio do método de mínimos quadrados não lineares, no qual a soma quadrática dos erros

$$SQ(k, \boldsymbol{\theta}) = \sum_{i=1}^N \left[\left(\frac{\Delta y}{\Delta q}\right)_i - k S_{Q^*}(q_{m_i} | \boldsymbol{\theta}) \right]^2$$

é minimizada com respeito a $\phi \in \Phi$, sendo $\phi' = (k, \boldsymbol{\theta}')$ o vetor de parâmetros do modelo e Φ

Tabela 12 – Dados sobre a produção vegetal (Prod.), em gramas de matéria-seca, do tomate em função do conteúdo de potássio (K), referente a um estudo realizado em 15 solos e mediante variadas estações

Solo	Safrá	Sem K			Com K			Conteúdo Médio	Resposta Vegetativa
		Prod.	K %	K Total	Prod.	K %	K Total		
29	2	86	2,92	2,51	90	4,40	3,96	3,24	2,76
21	2	82	2,12	1,74	86	3,32	2,86	2,30	3,58
101	2	103	1,96	2,02	91	3,72	3,39	2,70	-8,78
95	2	66	1,80	1,19	79	2,64	2,07	1,64	14,48
80	2	69	1,62	1,12	81	2,29	1,85	1,49	16,28
100	2	81	1,36	1,10	98	2,30	2,25	1,68	14,75
38	8	94	1,33	1,25	94	2,15	2,02	1,64	0,00
90	2	98	1,27	1,24	120	3,66	4,39	2,82	6,99
53	5	52	0,88	0,46	70	0,94	0,66	0,56	89,82
30	8	64	0,84	0,54	67	1,38	0,92	0,73	7,75
36	5	45	0,75	0,34	67	0,87	0,58	0,46	89,65
78	2	41	0,74	0,30	86	3,70	3,18	1,74	15,63
37	5	51	0,69	0,35	64	0,84	0,54	0,44	70,01
102	2	45	0,57	0,26	73	2,23	1,63	0,94	20,42
64	5	62	0,56	0,35	71	0,73	0,52	0,43	52,60
36	8	38	0,56	0,21	69	0,62	0,43	0,32	144,19
37	8	43	0,52	0,22	65	0,58	0,38	0,30	143,42
36	6	61	0,49	0,30	70	0,53	0,37	0,33	124,83
37	6	49	0,36	0,18	76	0,48	0,36	0,27	143,31

Fonte: Hoagland e Martin (1993)

o espaço paramétrico.

Uma vez que a parte fixa do modelo, denotada por $k S_{Q^*}(q_{m_i} | \theta)$, possui parâmetros dispostos de maneira não linear, não existe solução analítica para o problema de mínimos quadrados. Neste caso, a estimação dos parâmetros requer métodos de programação não linear, como procedimentos iterativos que exigem valores iniciais para os parâmetros (SEBER; WILD, 1989).

Neste trabalho, a estimação dos parâmetros é realizada por meio do algoritmo de projeção de variáveis de Golub e Pereyra (1973), o qual consiste no desmembramento da soma de quadrados $SQ(k, \theta)$ em partes que dependem somente de parâmetros lineares ou não lineares (O'LEARY; RUST, 2013). A vantagem deste algoritmo é o fato do método não requerer valores iniciais para os parâmetros lineares, como é o caso do parâmetro k na modelagem da resposta vegetativa.

A fim de selecionar o modelo que melhor se aproxima da verdadeira resposta vegetativa, utiliza-se o critério da informação de Akaike corrigido (BURNHAM; ANDERSON, 2002,

2004), expresso por

$$AIC_c = AIC + \frac{2p(p+1)}{N-p-1},$$

em que AIC é o critério de informação de Akaike, p é o número total de parâmetros a serem estimados e N é o tamanho da amostra. O uso do AIC_c é plausível neste caso, pois os estimadores dos parâmetros $\hat{\phi}$ pelo método dos mínimos quadrados coincidem com os de máxima verossimilhança. Dessa forma, torna-se possível calcular o valor do logaritmo da função de verossimilhança para, subsequentemente, obter AIC e AIC_c . Neste trabalho, os cálculos são realizados por meio da função *logLik* do pacote *stats* (R CORE TEAM, 2015).

O modelo que apresenta o menor valor do AIC é considerado como a melhor aproximação para a verdadeira relação existente entre as variáveis resposta e explanatória. O AIC_c é usado para se comparar modelos com diferentes números de parâmetros, levando-se em conta a parcimônia dos mesmos (BURNHAM; ANDERSON, 2004).

Dos modelos apresentados neste trabalho, o que melhor se ajusta aos dados de Hoagland e Martin (1933) é o modelo exponencial (Tabela 13 e Figura 28). Além disso, a estimativa obtida para parâmetro r do modelo de Erlang é igual a unidade, tornando seu ajuste equivalente ao do modelo exponencial.

O fato da resposta exponencial descrever os dados de forma mais parcimoniosa indica que o modelo de Mitscherlich é a função de produção mais plausível para a situação estudada. Esse resultado já era esperado, uma vez que adequação do modelo de Mitscherlich na descrição de dados de adubação vem sendo verificada por mais de um século (CERRATO; BLACKMER, 1990; MACY, 1936; MALAVOLTA, 1993; PIMENTEL-GOMES; NOGUEIRA, 1993; SCHABENBERGER; PIERCE, 2001).

Tabela 13 – Valores do critério da informação de Akaike corrigido para os modelos ajustados aos dados de Hoagland e Martin (1933), e publicados por Macy (1936), sobre a resposta vegetativa à adubação de tomateiros em função do conteúdo total de potássio nas plantas

Modelo Ajustado	AIC_c
Resposta Linear	168,20
Resposta Exponencial	165,19
Resposta Weibull	167,17
Resposta Erlang	166,45

No entanto, esse resultado não conflita com a teoria de Liebig, uma vez que o modelo de Mitscherlich surge naturalmente da lei do mínimo, quando considera-se uma distribuição Exponencial para os níveis críticos de suficiência das células que compõem o vegetal.

Figura 28 – Ajustes de diferentes modelos para descrever os dados sobre a resposta vegetativa do tomate em função do conteúdo total de potássio nas plantas. Dados referentes ao estudo de Hoagland e Martin (1993), publicados pela primeira vez por Macy (1936)

4.4 Conclusão

Conclui-se que a lei do mínimo de Liebig, quando associada à distribuição desigual do nutriente limitante e/ou das exigências nutricionais na planta, conduz a funções de produção côncavas e continuamente diferenciáveis, compatíveis com a lei dos retornos decrescentes de Mitscherlich. Não se encontra neste trabalho nenhuma contradição entre essas duas leis básicas da nutrição mineral de plantas. A modelagem da resposta vegetativa apresentada neste trabalho é fundamentada na lei de Liebig, mas fornece uma nova interpretação à lei dos retornos decrescentes de Mitscherlich: quanto maior é o conteúdo total do nutriente limitante na planta, menor é a fração que se aproveitada de um incremento em sua disponibilidade a fim de prover produção vegetal.

Agradecimentos

O presente trabalho foi realizado com apoio do CNPq, Conselho Nacional de Desenvolvimento Científico e Tecnológico - Brasil.

Referências

BERCK, P.; HELFAND, G. Reconciling the von Liebig and differentiable crop production functions. **American Journal of Agricultural Economics**, Oxford, v. 72, n. 4, p. 985-996, 1990.

- BURNHAM, K.P.; ANDERSON, D.R. **Model selection and multimodel inference: a practical information-theoretic approach**. 2.ed. New York: Springer-Verlag, 2002. 554 p.
- BURNHAM, K.P.; ANDERSON, D.R. Multimodel inference: understanding AIC and BIC in model selection. **Sociological Methods and Research**, Thousand Oaks, v. 33, n. 2, p. 261-304, 2004
- CERRATO, M.E.; BLACKMER, A.M. Comparison of models for describing corn yield response to nitrogen fertilizer. **Agronomy Journal**, Madison, v. 82, p. 138-143, 1990.
- GALLANT, R.A.; FULLER, W.A. Fitting segmented polynomial regression models whose join points have to be estimated. **Journal of the American Statistical Association**, Alexandria, v. 68, n. 341, p. 144-147, 1973.
- GOLUB, G.H.; PEREYRA, V. The differentiation of pseudoinverses and nonlinear least squares problems whose variables separate. **SIAM Journal on Numerical Analysis**, Philadelphia, v. 10, p. 413-432, 1973.
- HOAGLAND, D.R.; MARTIN, J.C. Absorption of potassium by plants in relation to replaceable, non-replaceable, and soil solution potassium. **Soil Science**, Philadelphia, v. 36, p. 1-33, 1933.
- LIEBIG, J. von. **Organic chemistry in its applications to agriculture and physiology**. London: Taylor and Walton, 1840. 387 p.
- MACY, P. The quantitative mineral requirements of plants. **Plant Physiology**, Bethesda, v. 11, n. 4, p. 749-764, 1936.
- MALAVOLTA, E. **Manual de química agrícola - adubos e adubação**. 2.ed. São Paulo: Editora Agronômica CERES, 1967. 606 p.
- MALAVOLTA, E. **Elementos de nutrição mineral de plantas**. São Paulo: Editora Agronômica CERES, 1980. 251 p.
- MALAVOLTA, E.; VITTI, G.C.; OLIVEIRA, S.A. **Avaliação do estado nutricional de plantas: princípios e aplicações**. Piracicaba: Associação Brasileira para Pesquisa da Potassa e do Fosfato, 1989. 201 p.
- MALAVOLTA, E. **Nutrição mineral e adubação do cafeeiro**. São Paulo: Editora Agronômica CERES, 1993. 210 p.
- MITSCHERLICH, E. A. Das gesetz des minimums und das gesetz des abnehmenden bodenertrages. **Landwirtschaftliche Jahrbücher**, Berlin, v. 38, p. 537-552, 1909.
- O'LEARY, D.P.; RUST, B.W. Variable projection for nonlinear least squares problems. **Computational Optimization and Applications**, Amsterdã, v. 54, n. 3, p. 579-593, 2013.
- PARIS, Q. The von Liebig hypothesis. **American Journal of Agricultural Economics**, Oxford, v. 74, n. 4, p. 1019-1028, 1992a.
- PIMENTEL-GOMES, F.; NOGUEIRA, I.R. **Análise matemática**. 2.ed. Piracicaba: Departamento de Ciências Exatas da ESALQ, 1980. 371 p.
- R CORE TEAM. **R: a language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria, 2015. Disponível em: <<http://www.r-project.org>>. Acesso em: 24 abril 2015.
- SCHABENBERGER, O.; PIERCE, F.J. **Contemporary statistical models for the plant and soil sciences**. New York: CRC PRESS, 2001. 730 p.

SEBER, G.A.F.; WILD, C.J. **Nonlinear regression**. United States: John Wiley & Sons, Inc., 1989. 768 p.

STERNER, R.W.; ELSER, J.J. **Ecological stoichiometry**: the biology of elements from molecules to the biosphere. Princeton, New Jersey: Princeton University Press, 2002. 464 p.

ANEXOS

Anexo A

Neste anexo são apresentados os gráficos das distribuições amostrais dos estimadores de máxima verossimilhança dos parâmetros dos modelos lineares segmentados com transições suaves, que foram estudados no capítulo 2. Os resultados foram obtidos com base em 1.000 conjuntos de dados simulados para cada modelo suave. Veja mais detalhes na seção 2.3.4.

Figura 29 – Distribuição amostral empírica dos estimadores de máxima verossimilhança dos parâmetros do modelo linear segmentado, com junção média do tipo I baseada na distribuição Logística. Retas verticais em vermelho indicam os valores verdadeiros de cada parâmetro.

Figura 30 – Distribuição amostral empírica dos estimadores de máxima verossimilhança dos parâmetros do modelo linear segmentado, com junção média do tipo II baseada na distribuição Logística. Retas verticais em vermelho indicam os valores verdadeiros de cada parâmetro.

Figura 31 – Distribuição amostral empírica dos estimadores de máxima verossimilhança dos parâmetros do modelo de regressão “bent-cable” estendido, com junção polinomial de quarta ordem proposta por Zang (1980). Retas verticais em vermelho indicam os valores verdadeiros de cada parâmetro.

Anexo B

Nesta seção são apresentados os gráficos dos ajustes dos modelos lineares segmentados estudados no capítulo 3 aos dados de Schneeberger (2009).

Figura 32 – Gráficos do ajuste do modelo 3.4.1, dos seus resíduos padronizados e dos valores preditos em função dos observados

Figura 33 – Gráficos do ajuste do modelo 3.4.1, dos seus resíduos padronizados e dos valores preditos em função dos observados

Figura 34 – Gráficos do ajuste do modelo 3.4.1, dos seus resíduos padronizados e dos valores preditos em função dos observados

Figura 35 – Gráficos do ajuste do modelo 3.4.1, dos seus resíduos padronizados e dos valores preditos em função dos observados