



# Modelo de regressão para a previsão de produtividade de cafeeiros no Estado de Minas Gerais<sup>1</sup>



Luiz G. de Carvalho<sup>2</sup>, Gilberto C. Sediya<sup>3</sup>, Paulo R. Cecon<sup>4</sup> & Helena M. R. Alves<sup>5</sup>

<sup>1</sup> Parte da Tese de Doutorado do primeiro autor. Projeto financiado pelo CNPq.

<sup>2</sup> Departamento de Engenharia, UFLA, CP 37, CEP 37200-000, Lavra, MG. Fone: (35) 3829-1481. E-mail: [lgonsaga@ufla.br](mailto:lgonsaga@ufla.br) (Foto)

<sup>3</sup> DEA/UFV, CEP 36571-000, Viçosa, MG. Fone: (31) 3899-1905. E-mail: [sediya@ufv.br](mailto:sediya@ufv.br)

<sup>4</sup> DPI/UFV, CEP 36571-000, Viçosa, MG. Fone: (31) 3899-1781. E-mail: [cecon@dpi.ufv.br](mailto:cecon@dpi.ufv.br)

<sup>5</sup> Departamento de Ciências do Solo, UFLA. Fone: (35) 3829-1252. E-mail: [helena@ufla.br](mailto:helena@ufla.br)

Protocolo 100 - 3/6/2003 - Aprovado em 1/3/2004

**Resumo:** Com este trabalho, objetivou-se parametrizar e testar um modelo de regressão linear múltipla aplicado sobre os componentes principais mais significativos obtidos de séries de produtividades da cultura do café, representativas de três municípios da região Sul do Estado de Minas Gerais, tomando-se por base o modelo de Stewart et al. (1976), porém se acrescentando novas variáveis representadas por elementos agrometeorológicos, além das penalizações hídricas para os quatro trimestres do ciclo agrícola (julho a junho) da cultura. Tendo em vista ser o número de observações inferior à quantidade de variáveis, recorreu-se à análise multivariada de componentes principais para reduzir a dimensão do conjunto dessas variáveis. A análise de regressão linear múltipla foi aplicada nos três primeiros componentes principais. Os resultados dos testes apresentaram erros relativos percentuais das estimativas bastante discrepantes, ocorrendo tendência de superestimarem as produtividades; contudo, verificou-se que as estimativas pelo modelo tenderam a apresentar comportamento similar ao dos dados observados.

**Palavras-chave:** rendimentos, modelagem agrometeorológica, componentes principais, café

## A regression model to predict coffee productivity in Southern Minas Gerais, Brazil

**Abstract:** The objective of this work was to set up and test a multiple linear regression model applied to principal components for representative coffee crop yield series for three places in Southern Minas Gerais, based on the model proposed by Stewart et al. (1976), with new variables, represented by agrometeorological elements, besides the soil water depletion for the four quarterly periods in agricultural cycle (July to June). Since the number of observations was lower than the amount number of variables, we resorted to principal component analysis to reduce the dimension of this set of variables. The multiple linear regression analysis was applied to the first three principal components. In agreement with the tests, the model presented relative errors of estimates with high discrepancies and a tendency to overestimate productivity for the three places. However, it was verified that the estimates for the model tended to present behavior similar to observed data.

**Key words:** yield, agrometeorological modeling, principal components, coffee bean

## INTRODUÇÃO

A cultura do café tem grande importância na economia do Brasil, razão por que a previsão da sua produtividade permite também ao agricultor, um planejamento melhor de suas atividades agrícolas, minimizando as oscilações de preços, garantindo sua estabilidade e regularidade do abastecimento

do mercado interno e externo. Contudo, a previsão de safras agrícolas no Brasil tem sido um grande problema. Neste sentido, procurando-se elaborar modelos de previsão, diversos trabalhos existentes na literatura mostram a importância das condições climáticas na produção agrícola. Segundo Hoogenbom (2000), os elementos climáticos têm sido os mais requeridos para a operação da maioria dos modelos agrometeorológicos.

O modelo “empírico-estatístico”, proposto por Robertson (1983), é o mais simples, em que os elementos climáticos correspondem às condições de temperatura, chuva ou disponibilidade hídrica reinantes em todo o ciclo da cultura, ou em parte dele. Tais modelos utilizam uma ou mais variáveis independentes, sendo a variável dependente a produtividade da cultura e as independentes, em geral os elementos do clima, cujos coeficientes de ponderação se obtêm normalmente pela análise de regressão múltipla.

Na cultura do café, um fator significativo que interfere na variação de sua produção, próprio de sua natureza fisiológica, é a alternância bienal, com safras altas e baixas, necessitando vegetar em um ano para produzir bem no ano seguinte (Rena & Maestri, 1985). Dean (1939) estudando a correlação de índices pluviais na produtividade da cultura do café no Havaí, já havia observado este efeito também verificado por Beaumont (1939) na mesma região, em grupos de plantas de café com diferentes idades (7 e 12 anos).

Doorenbos & Kassam (1979) propuseram o modelo desenvolvido por Stewart et al. (1976) como alternativa de modelagem, tendo a produtividade como função de parâmetros hídricos, tendo sido tal modelo trabalhado por Picini (1998) e adaptado para a cultura do café. Mais tarde, conforme apresentam Picini et al. (1999), o referido modelo foi testado e parametrizado, obtendo-se resultados de produtividade do cafeeiro altamente dependente da produtividade do ano anterior e particularmente sensível ao estresse hídrico durante o estágio fenológico do “final da dormência das gemas/florescimento” e do estágio “final do florescimento/início da formação do grão”; além disso, constataram que os períodos trimestrais, agosto/setembro/outubro; novembro/dezembro/janeiro e fevereiro/março/abril, adotados na penalização hídrica do modelo, foi a combinação que apresentou melhor resultado.

Uma técnica bastante útil em modelagem é a análise estatística multivariada de componentes principais, a qual consiste em transformar um conjunto original de variáveis em outro conjunto, os componentes principais, de dimensões equivalentes, procurando-se reduzir a massa de dados. Em muitos casos, as medidas das variáveis que caracterizam um indivíduo são correlacionadas entre si, cuja correlação indica que algumas informações contidas em uma variável também o são em outra. Então, o objetivo da análise de componentes principais é transformar a quantidade de variáveis originais correlacionadas em um mesmo tanto de variáveis não-correlacionadas, ou em componentes ortogonais, ou seja, nos componentes principais. Cada um desses é uma combinação linear de todas as variáveis originais, independentes entre si e escolhidos por ordem decrescente dos autovalores com o máximo de informação, em termos de variação total, contida nos dados iniciais permitindo, com isso, reduzir a dimensão do conjunto original (Cruz & Regazzi, 1997).

Na análise de componentes principais pode-se definir a variância total existente em um conjunto de dados multivariados pela soma das variâncias de cada uma das variáveis. Numa matriz de variâncias/covariâncias, essas variâncias individuais constituem os elementos da diagonal principal. Basta somá-los, portanto, encontrando-se o traço da matriz (diagonal principal), para se obter a variabilidade total e, em seguida, a contribuição de cada variável.

Os coeficientes das equações lineares de cada componente principal que transformam os dados originais em contagens (escores) são denominados autovetores. Deste modo e se utilizando a multiplicação da matriz de dados originais pela matriz de autovetores, obtêm-se uma matriz de dados transformados, os chamados escores dos componentes principais, em que a principal propriedade é o fato de não serem correlacionados entre si.

A soma dos autovalores da matriz de covariâncias ou correlações dos escores dos componentes principais, igual ao traço dessa matriz, representa também a variabilidade total da mesma e a contribuição de cada autovalor, em termos de variabilidade. Ao primeiro corresponderá a maior variabilidade possível e, ao segundo, a maior variabilidade possível restante, e assim por diante.

A importância relativa de um componente é avaliada pela percentagem da variância total que ele explica. A soma dos primeiros autovalores dividida pela de todos os autovalores, representa a proporção da variância total explicada pelos respectivos primeiros componentes principais (Morrison, 1976).

Ressalta-se que uma análise de componentes principais nem sempre trabalha no sentido de que um grande número de variáveis originais seja reduzido a um pequeno número de variáveis transformadas. Além do mais, se as variáveis originais não são correlacionadas, então a análise é absolutamente nula, ou seja, não há redução alguma no número dessas variáveis. Os melhores resultados são obtidos quando as originais são altamente correlacionadas, positiva ou negativamente. Segundo Manly (1986), neste caso é completamente concebível que 20 ou 30 variáveis originais possam ser adequadamente representadas por dois ou três componentes principais.

A teoria da amostragem de componentes principais não está devidamente desenvolvida, especialmente quando os componentes são extraídos da matriz de correlação, em vez da de covariância. Segundo Mardia et al. (1979) e Dunteman (1984), na maioria dos casos a seleção sobre quantos componentes reter, é baseada em métodos simples e práticos. Morrison (1976) sugeriu que só os primeiros quatro ou cinco componentes devem ser extraídos, desde que os restantes sejam difíceis de interpretar o problema fisicamente. Infelizmente, nenhum teste estatístico pode ser usado para determinar o significado de um componente. De modo geral, a literatura sobre a utilização de componentes principais nas diversas áreas de aplicação mostra que, para interpretar dados com sucesso, basta escolher os primeiros componentes que acumulam uma percentagem de variâncias explicada igual ou superior a 70%. Uma regra prática para desconsiderar os componentes principais de menor importância, é eliminar aqueles com variância inferior à variância média das variáveis originais.

Em certas situações, algumas das variáveis do conjunto original podem ser eliminadas examinando-se as correlações (também conhecidas por fatores de carga) entre cada uma com o respectivo componente principal. Com base no princípio de que a importância ou variância dos componentes principais decresce do segundo para o último, tem-se que os últimos componentes são responsáveis pela explicação de uma fração muito pequena da variância total. Assim, a variável que apresenta a maior correlação com o componente principal de menor importância (componente de menor variância ou menor

autovalor) provavelmente é a menos significativa para explicar a variância total podendo, então, ser eliminada. Tal procedimento segue para o outro componente de menor importância, eliminando-se aí a variável de maior correlação e que não tenha sido descartada anteriormente.

Como foi mostrado, a técnica de componentes principais apresenta tantos componentes principais quanto for o número de variáveis que estão sendo analisadas, porém com a propriedade de que, em ordem decrescente, os primeiros componentes principais retêm a maior parte da variância dos dados originais. Portanto, pode-se, na maioria das vezes e principalmente quando se tem um número de observações inferior à quantidade de variáveis, fazer a regressão linear com os componentes principais que retiverem a maior parte da variância original das variáveis, desde que sejam suficientes para representar as variáveis originais, diminuindo substancialmente as variáveis regressoras dentro do modelo linear de regressão múltipla; sendo assim, o número de parâmetros a ser estimado passa a ser menor que o número de observações; porém, cada componente principal é uma combinação linear das variáveis originais com seus autovetores tendo-se, ainda, estas variáveis como entrada do modelo.

Ao se proceder à regressão em componentes principais, algumas das variáveis originais, as independentes, conforme já referido, podem ser eliminadas da análise antes mesmo de executar qualquer regressão. As variáveis restantes são, então, novamente submetidas a uma análise de componentes principais com a regressão múltipla, que é executada nos novos componentes. Tal procedimento tem a vantagem de reduzir o número de variáveis que devem ser medidas para usar na equação de regressão resultante. Descrições detalhadas da regressão em componentes principais podem ser encontradas, entre outros autores, em Haan (1977) e Montgomery & Peck (1992).

Sendo assim, objetivou-se, com este trabalho, parametrizar um modelo de regressão linear múltipla a partir dos componentes principais mais significativos para a previsão de produtividades da cultura do café, utilizando-se um conjunto de variáveis predictoras, além dos componentes hídricos, proposto no modelo de Stewart et al. (1976), para três localidades da região Sul do Estado de Minas Gerais.

## MATERIAL E MÉTODOS

O modelo desenvolvido por Stewart et al. (1976) e adaptado por Picini (1998) para a cultura do café, conforme a Eq. 1, foi a base para a realização do presente trabalho.

$$\frac{Y_e}{Y_p} = ky_0 \left( \frac{Y_a}{Y_p} \right) + \sum_{i=1}^n ky_i \left( \frac{ETR}{ETP} \right) \quad (1)$$

em que  $Y_e$  é a produtividade estimada;  $Y_p$  a produtividade potencial;  $ky_0$  o coeficiente de penalização relativo à produtividade do ano anterior;  $Y_a$  é a produtividade do ano anterior;  $ky_i$  o coeficiente de resposta da cultura ao suprimento de água para os sucessivos estádios fenológicos da cultura; ETR a evapotranspiração real da cultura; ETP a evapotranspiração potencial e ETR/ETP a penalização hídrica.

O modelo representado pela Eq. 1 segue o modelo estatístico tradicional:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + e_i \quad (2)$$

em que  $Y$  é a variável dependente;  $X_1, X_2, \dots, X_k$  são as variáveis independentes;  $\beta_0, \beta_1, \beta_2, \dots, \beta_k$  são os parâmetros da regressão e  $e_i$  representa o erro residual.

Os dados observados de produtividades de café utilizados são representativos dos municípios de Alfenas, Monte Belo e São Sebastião do Paraíso, localizados na região Sul do Estado de Minas Gerais, cuja caracterização da bienalidade de produção é bem definida (Figura 1).

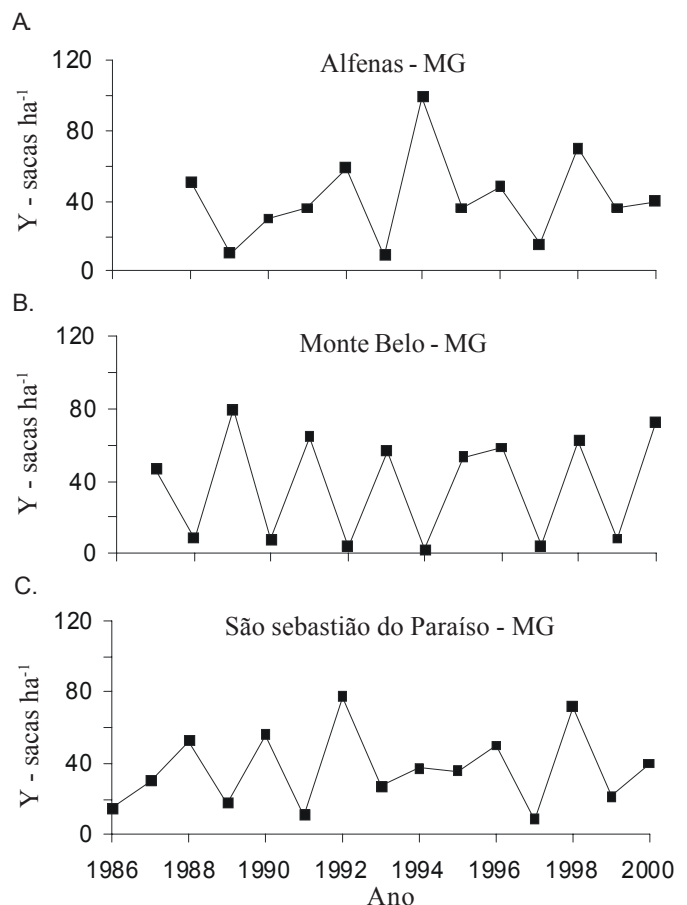


Figura 1. Séries anuais de produtividades (Y) de glebas de lavouras cafeeiras representativas dos municípios da região Sul do Estado de Minas Gerais e utilizadas na parametrização dos modelos.

As produtividades potenciais adotadas foram 99, 81 e 88 sacas  $ha^{-1}$  para os municípios de Alfenas, Monte Belo e São Sebastião do Paraíso, respectivamente. As duas primeiras, por estarem acima das médias das lavouras onde se obtiveram esses dados, e a última surgiu do acréscimo de 10% sobre o valor médio das produtividades máximas de todas as glebas do local, com base na recomendação de Kanemasu (1983).

As estimativas da disponibilidade hídrica do solo, para uma capacidade de água disponível de 100 mm, foram realizadas pelo método proposto por Thornthwaite & Mather, cuja

Tabela 1. Variáveis utilizadas para parametrização dos modelos de regressão linear múltipla em componentes principais, para três localidades do Sul de Minas Gerais

Variáveis		Descrição
Y	Yo/Yp	Produtividade observada em relação à potencial (Variável dependente)
X <sub>1</sub>	Ya/Yp	Produtividade observada do ano anterior em relação à potencial
X <sub>2</sub>	(ETR/ETP) <sub>1</sub>	Evapotranspiração relativa para o período de jul./ago./set.
X <sub>3</sub>	(ETR/ETP) <sub>2</sub>	Evapotranspiração relativa para o período de out./nov./dez.
X <sub>4</sub>	(ETR/ETP) <sub>3</sub>	Evapotranspiração relativa para o período de jan./fev./mar.
X <sub>5</sub>	(ETR/ETP) <sub>4</sub>	Evapotranspiração relativa para o período de abr./maio/jun.
X <sub>6</sub>	X <sub>2</sub> .X <sub>3</sub>	(ETR/ETP) <sub>1</sub> multiplicada por (ETR/ETP) <sub>2</sub>
X <sub>7</sub>	X <sub>3</sub> .X <sub>4</sub>	(ETR/ETP) <sub>2</sub> multiplicada por (ETR/ETP) <sub>3</sub>
X <sub>8</sub>	X <sub>4</sub> .X <sub>5</sub>	(ETR/ETP) <sub>3</sub> multiplicada por (ETR/ETP) <sub>4</sub>
X <sub>9</sub>	Tm/Tma	Temperatura média do período de jul./ago./set. em relação à temperatura média anual
X <sub>10</sub>	Tm/Tma	Temperatura média do período de out./nov./dez. em relação à temperatura média anual
X <sub>11</sub>	Tm/Tma	Temperatura média do período de jan./fev./mar. em relação à temperatura média anual
X <sub>12</sub>	Tm/Tma	Temperatura média do período de abr./maio/jun. em relação à temperatura média anual
X <sub>13</sub>	Tn/Tna	Temperatura mínima média do período de jul./ago./set. em relação à temperatura mínima média anual
X <sub>14</sub>	Tn/Tna	Temperatura mínima média do período de out./nov./dez. em relação à temperatura mínima média anual
X <sub>15</sub>	Tn/Tna	Temperatura mínima média do período de jan./fev./mar. em relação à temperatura mínima média anual
X <sub>16</sub>	Tn/Tna	Temperatura mínima média do período de abr./maio/jun. em relação à temperatura mínima média anual
X <sub>17</sub>	Tx/Txa	Temperatura máxima média do período de jul./ago./set. em relação à temperatura máxima média anual
X <sub>18</sub>	Tx/Txa	Temperatura máxima média do período de out./nov./dez. em relação à temperatura máxima média anual
X <sub>19</sub>	Tx/Txa	Temperatura máxima média do período de jan./fev./mar. em relação à temperatura máxima média anual
X <sub>20</sub>	Tx/Txa	Temperatura máxima média do período de abr./maio/jun. em relação à temperatura máxima média anual

descrição foi apresentada por Pereira et al. (1997). O valor do coeficiente da cultura (Kc) utilizado para a estimativa da evapotranspiração da cultura, foi igual a 1, conforme Camargo & Pereira (1994). Para cada município os dados diários dos elementos climáticos (temperaturas e precipitações pluviiais) foram agrupados em escala mensal, obtendo-se as respectivas médias e se efetuando, em seguida, o balanço hídrico seriado (mês a mês) para o mesmo período das produtividades (Figura 1).

Posteriormente, os dados e componentes do balanço hídrico foram separados por uma seqüência de períodos trimestrais, de acordo com Matiello (1991), assim definidos: julho/agosto/setembro; outubro/novembro/dezembro; janeiro/fevereiro/março e abril/ maio/junho. Para cada um desses períodos e para todo o período das produtividades (Figura 1), obtiveram-se as médias dos componentes do balanço hídrico e demais variáveis descritas na Tabela 1, as quais foram submetidas ao modelo de regressão.

Os modelos foram parametrizados, tendo por base o modelo apresentado pela Eq. 1, em que as variáveis dependentes, assim como as independentes, são relativas.

De acordo com a Eq. 2, para cada local representado pela respectiva gleba, os modelos consistiram em submeter à análise de regressão linear múltipla a produtividade da cultura do café, como função de um conjunto de variáveis inerentes à produção, além das apresentadas pela Eq. 1. Com o acréscimo de novas variáveis ao modelo, visou-se encontrar resultados mais satisfatórios aos obtidos por Picini et al. (1999); contudo, observa-se que a quantidade dessas variáveis, no total de 20 (Tabela 1), é superior ao número de observações, por exemplo, 15 no caso de São Sebastião do Paraíso (Figura 1), formando

uma matriz 15 x 20, o que inviabiliza, estatisticamente, a análise de regressão linear múltipla, pois o número de parâmetros a serem estimados é superior ao número de observações. Portanto, adotou-se a técnica de análise multivariada em componentes principais para transformar essas variáveis em um novo conjunto, ou seja, nos chamados escores dos componentes principais.

Os modelos assim realizados foram avaliados por meio da significância dos parâmetros da regressão, utilizando-se o teste “t”, pelo coeficiente de determinação (R<sup>2</sup>) e pelo fenômeno estudado.

Os testes dos modelos foram realizados com dados de produtividades observadas de outras glebas das lavouras cafeeiras de cada município, conforme sugerem Picini (1998) e Hoogenboom et al. (1992). Para a avaliação dos testes foram comparadas as produtividades estimadas e observadas, conhecendo-se o erro relativo percentual (ERP), o qual foi também utilizado por Leal (2000).

Além disso, fez-se a plotagem gráfica das relações entre produtividades estimadas e observadas com a produtividade potencial (Ye/Yp e Yo/Yp). Procedeu-se à análise de regressão linear simples com a reta passando pela origem, sendo os modelos avaliados pelo coeficiente de determinação (r<sup>2</sup>) e pela concordância expressa pelo índice “d”, de Willmott et al. (1985), conforme a Eq. 3. O índice “d”, com variação entre 0 e 1, indica, o grau de exatidão entre valores estimados e observados, sendo que, quanto mais próximo de 1, melhor a exatidão, ao passo que o coeficiente de determinação (r<sup>2</sup>) indica a precisão do modelo, ou seja, quanto a variável dependente é explicada pelas variáveis independentes.

$$d = 1 - \left[ \frac{\sum_{i=1}^n (Y_{e_i} - Y_{o_i})^2}{\sum_{i=1}^n (|Y_{e_i} - \bar{Y}_o| + |Y_{o_i} - \bar{Y}_o|)^2} \right] \quad (3)$$

d = índice de concordância

$Y_{e_i}$  = i-ésimo valor previsto ou estimado

$Y_{o_i}$  = i-ésimo valor observado

$\bar{Y}_o$  = média dos valores observados

## RESULTADOS E DISCUSSÃO

Na Tabela 2 constam os resultados obtidos da aplicação da técnica multivariada de análise em componentes principais onde se observa que os três primeiros componentes principais foram suficientes para interpretar o conjunto das vinte variáveis analisadas, mostrando a eficiência desta técnica na redução de um grande número de variáveis para um conjunto menor. Assim, essas 20 variáveis foram reduzidas para 3 variáveis equivalentes, ou seja, os 3 primeiros componentes principais, passando a dimensão da matriz, que antes era de 15 x 20 a exemplo de São Sebastião do Paraíso, a ser 15 x 3, permitindo a aplicação da regressão linear múltipla sobre os componentes principais.

Portanto, utilizaram-se esses três componentes principais para parametrização dos modelos de regressão, adotando-se o

critério, visto apresentarem variâncias acumuladas acima de 80%, as quais corresponderam a 84,04, 84,68 e 90,99% para Alfenas, Monte Belo e São Sebastião do Paraíso, respectivamente.

A tentativa de se eliminar algumas das variáveis originais pelo critério da baixa correlação (baixo valor do autovetor) da variável com seu respectivo componente principal, não se mostrou eficiente, razão pela qual se mantiveram todas as variáveis para os componentes principais retidos.

Na Tabela 3 tem-se o resumo dos resultados estatísticos dos modelos. Verifica-se, para Alfenas e de acordo com o teste “t”, que todas as estimativas dos parâmetros são significativas ( $P \leq 0,0787$ ), indicando que as variáveis transformadas em componentes principais contribuíram satisfatoriamente para a construção do modelo, fato reforçado pelo coeficiente de determinação ( $R^2 = 0,88$ ), mostrando quanto da variação dos valores da variável dependente, que no caso é expressa por  $Y_e/Y_p$  (Eq. 1), é explicada pela regressão sobre as variáveis independentes representadas pelos componentes principais.

Para Monte Belo, a estimativa do parâmetro “ $\beta_3$ ” foi de baixa significância ( $P \leq 0,4987$ ); contudo, manteve-se o componente principal correspondente ( $Z_3$ ) na regressão linear múltipla por praticamente não alterar os resultados com sua retirada e por coerência ao modelo anterior. Os demais parâmetros mostraram-se com significância relevante ( $P \leq 0,1131$ ), e coeficiente de determinação ( $R^2 = 0,75$ ) é também satisfatório.

Em São Sebastião do Paraíso a estimativa do parâmetro “ $\beta_1$ ” possui um nível de significância relativamente baixo

Tabela 2. Coeficientes de ponderação e variâncias dos componentes principais ( $Z_i$ ) gerados a partir da matriz de covariância dos dados originais para três localidades do Sul de Minas Gerais

Variáveis	Alfenas			Monte Belo			São Sebastião do Paraíso		
	$Z_1$	$Z_2$	$Z_3$	$Z_1$	$Z_2$	$Z_3$	$Z_1$	$Z_2$	$Z_3$
$X_1$	0,4282	0,7544	-0,3726	0,8201	-0,5483	-0,0651	0,5722	0,7513	-0,2223
$X_2$	-0,5546	0,2631	0,0817	0,3145	0,4386	0,2095	-0,5754	0,3313	-0,2432
$X_3$	-0,1236	0,1737	-0,1215	0,0303	0,0564	-0,0344	0,0023	0,1314	0,2174
$X_4$	-0,0070	0,0458	-0,0498	0,0070	0,0005	-0,0182	0,0159	-0,0020	0,0001
$X_5$	-0,0956	0,0737	0,3142	0,1709	0,1326	0,3462	-0,0776	0,2384	0,5928
$X_6$	-0,5914	0,3433	-0,0142	0,3176	0,4695	0,1654	-0,5502	0,3909	-0,1506
$X_7$	-0,1285	0,2055	-0,1588	0,0368	0,0568	-0,0515	0,0180	0,1295	0,2176
$X_8$	-0,1008	0,1132	0,2663	0,1770	0,1330	0,3303	-0,0621	0,2365	0,5930
$X_9$	0,0522	-0,0529	0,0017	-0,0789	-0,0939	0,1528	0,0504	-0,0530	0,0592
$X_{10}$	0,0851	0,0336	0,1700	-0,0622	-0,0590	0,1671	-0,0340	-0,0274	0,0725
$X_{11}$	0,0245	-0,0228	0,0023	-0,0549	-0,1062	0,1082	-0,0106	-0,0424	0,0897
$X_{12}$	0,0340	-0,0477	-0,0268	-0,0269	-0,0496	-0,0229	-0,0506	-0,0381	0,0051
$X_{13}$	0,0127	-0,0582	0,1825	-0,1051	-0,0652	0,5790	0,0312	-0,0114	0,0437
$X_{14}$	0,2899	0,3382	0,7278	-0,1208	-0,2741	0,4359	-0,0210	0,0250	0,0586
$X_{15}$	0,0424	0,0122	0,0455	-0,1285	-0,3135	0,2765	-0,0224	-0,0693	0,0768
$X_{16}$	-0,0159	-0,0434	0,0997	-0,0514	-0,1646	0,1138	-0,1204	-0,0578	0,1350
$X_{17}$	0,0741	-0,0502	-0,0974	-0,0666	-0,1095	-0,0543	0,0610	-0,0761	0,0678
$X_{18}$	-0,0269	-0,1333	-0,1353	-0,0346	0,0461	0,0360	-0,0412	-0,0564	0,0801
$X_{19}$	0,0148	-0,0421	-0,0214	-0,0171	-0,0035	0,0289	-0,0040	-0,0275	0,0968
$X_{20}$	0,0614	-0,0501	-0,0962	-0,0141	0,0053	-0,0898	-0,0119	-0,0272	-0,0669
Variância (autovalores)	0,1387	0,0673	0,0314	0,1481	0,0963	0,0281	0,0728	0,0677	0,0226
Variância Acumulada (%)	49,09	72,91	84,04	46,03	75,96	84,68	40,63	78,39	90,99

Tabela 3. Coeficientes de determinação, testes F, coeficientes de regressão e respectivos testes estatísticos (teste t) dos modelos de regressão linear múltipla em componentes principais, para três localidades do Sul de Minas Gerais

Coeficientes	Alfenas		Monte Belo		São Sebastião do Paraíso	
	Modelo	Valor-P (*)	Modelo	Valor-P (*)	Modelo	Valor-P (*)
R <sup>2</sup>	0,8825		0,7513		0,6649	
Nível Crítico do Teste F	0,0004		0,0044		0,0097	
Interseção (β <sub>0</sub> )	0,6798	0,0191	1,6594	0,1131	-0,1109	0,8551
β <sub>1</sub>	0,1719	0,0787	-0,7688	0,0011	-0,2308	0,1889
β <sub>2</sub>	-0,7805	0,0002	0,4439	0,0571	-0,6448	0,0035
β <sub>3</sub>	0,7034	0,0044	-0,2659	0,4987	0,5447	0,0935

(\*) Valor do nível de significância dos parâmetros da equação de regressão, pelo teste t

(P ≤ 0,1889), porém o componente principal correspondente (Z<sub>1</sub>), por ser o de maior importância, foi mantido no modelo. Assim também o parâmetro “β<sub>0</sub>” não é estatisticamente significativo (P ≤ 0,8551), mas foi mantido no modelo para ser semelhante aos modelos anteriores e pouco alterar o resultado final com sua retirada. Por suas significâncias (P ≤ 0,0935), os demais parâmetros indicam seus níveis de importância dentro do modelo. Por sua vez, o coeficiente de determinação (R<sup>2</sup> = 0,66) ajuda a evidenciar, de forma satisfatória, a “performance” do modelo.

Da Tabela 3, os modelos ficam assim representados:

Para Alfenas:

$$\hat{Y}_e = Y_p[0,6798 + 0,1719Z_1 - 0,7805Z_2 + 0,7034Z_3] \quad (4)$$

sendo

$$Z_1 = 0,4282X_1 - 0,5546X_2 + \dots + 0,0614X_{20} \quad (5)$$

$$Z_2 = 0,7544X_1 + 0,2631X_2 + \dots - 0,0501X_{20} \quad (6)$$

$$Z_3 = -0,3726X_1 + 0,0817X_2 + \dots - 0,0962X_{20} \quad (7)$$

Para Monte Belo:

$$\hat{Y}_e = Y_p[1,6594 - 0,7688Z_1 + 0,4439Z_2 - 0,2659Z_3] \quad (8)$$

sendo

$$Z_1 = 0,8201X_1 + 0,3145X_2 + \dots - 0,0141X_{20} \quad (9)$$

$$Z_2 = -0,5483X_1 + 0,4386X_2 + \dots + 0,0053X_{20} \quad (10)$$

$$Z_3 = -0,0651X_1 + 0,2095X_2 + \dots - 0,0898X_{20} \quad (11)$$

Para São Sebastião do Paraíso:

$$\hat{Y}_e = Y_p[-0,1109 - 0,2308Z_1 - 0,6448Z_2 + 0,5447Z_3] \quad (12)$$

sendo

$$Z_1 = 0,5722X_1 - 0,5754X_2 + \dots - 0,0119X_{20} \quad (13)$$

$$Z_2 = 0,7513X_1 + 0,3313X_2 + \dots - 0,0272X_{20} \quad (14)$$

$$Z_3 = -0,2223X_1 - 0,2432X_2 + \dots - 0,0669X_{20} \quad (15)$$

em que  $\hat{Y}_e$  é a produtividade estimada (sacas ha<sup>-1</sup>);  $Y_p$  a produtividade potencial (sacas ha<sup>-1</sup>);  $Z_1$ ,  $Z_2$  e  $Z_3$  são os componentes principais (funções lineares) cujos coeficientes de ponderação estão apresentados na Tabela 2.

Para os testes, empregaram-se dados de produtividades independentes, os quais não foram utilizados na parametrização dos modelos. Juntamente com esses dados observados, os respectivos valores estimados são apresentados nas figuras. Nota-se, nos três modelos, que as estimativas tenderam a apresentar comportamento similar ao dos dados observados. Os respectivos desempenhos estatísticos estão apresentados nas figuras. Em Alfenas, pela plotagem no plano bidimensional dos valores relativos, observados e estimados (produtividades observadas e estimadas em relação à produtividade potencial) e pela análise de regressão linear simples, com o ajuste da reta passando pela origem, verifica-se o desempenho desse modelo apresentando  $r^2 = 0,58$  e o índice de Willmott et al. (1985),  $d = 0,78$ , indicando a “performance” da concordância entre valores observados e estimados (Fig. 2).

Idêntico aos resultados de Alfenas, para Monte Belo a regressão linear simples, com o ajuste da reta passando pela origem (Figura 2), apresenta  $r^2 = 0,75$ , indicando uma precisão relativamente satisfatória das estimativas. Por sua vez, o índice de Willmott et al. (1985) ( $d = 0,80$ ) mostra boa concordância entre produtividades observadas e estimadas.

Em São Sebastião do Paraíso os resultados mostraram-se com maiores discrepâncias, apresentando  $r^2 = 0,15$  e  $d = 0,58$  (Figura 2), e baixa concordância na comparação das estimativas com as produtividades observadas.

Observa-se, ainda, que os modelos tendem a superestimar as produtividades. Apesar da inclusão de novas variáveis, aquelas apresentadas na Tabela 1, os desempenhos são ainda relativamente baixos, indicando grande variação entre os valores observados e os estimados. Os erros relativos percentuais (ERP), por sua vez, mostraram-se de, modo geral, bastante discrepantes com relação às produtividades observadas. Os valores de ERP encontraram-se entre -18 a 454%; -18 a 1969% e entre -61 e 1022% para os municípios de Alfenas, Monte Belo e São Sebastião do Paraíso, respectivamente. Constatou-se que as maiores variações das produtividades estimadas ocorreram nos anos consecutivos, com grandes variações nas produtividades observadas. Essas discrepâncias também foram observadas por Picini (1998), ao

trabalhar com a cultura do café, verificando grande variação entre as produtividades observadas e estimadas de -67% a 153% e  $d = 0,98$ , porém as produtividades com as quais foram originados os modelos apresentaram o efeito da bialidade de produção claramente visíveis. Esta característica dos dados se mostra importante para o ajuste do modelo. Leal (2000), por sua vez, também utilizou o modelo de Stewart et al. (1976) para estimar a produtividade das culturas de milho, soja, sorgo e trigo, para as localidades de Capinópolis e Montes Claros, no Estado de Minas Gerais, cujos resultados dos erros relativos percentuais,

de modo geral variaram de -82 a 382%, mas esses autores não aplicaram a análise de componentes principais. Os modelos originados de regressão a partir dos componentes principais, aqui apresentados, visaram justamente avaliar o comportamento do modelo de Stewart et al. (1976) mediante esta técnica.

De acordo com os resultados estatísticos, tanto na parametrização quanto nos testes, o modelo realizado para Alfenas mostrou-se com melhor "performance", e o realizado para São Sebastião do Paraíso, com desempenho inferior em relação aos demais.

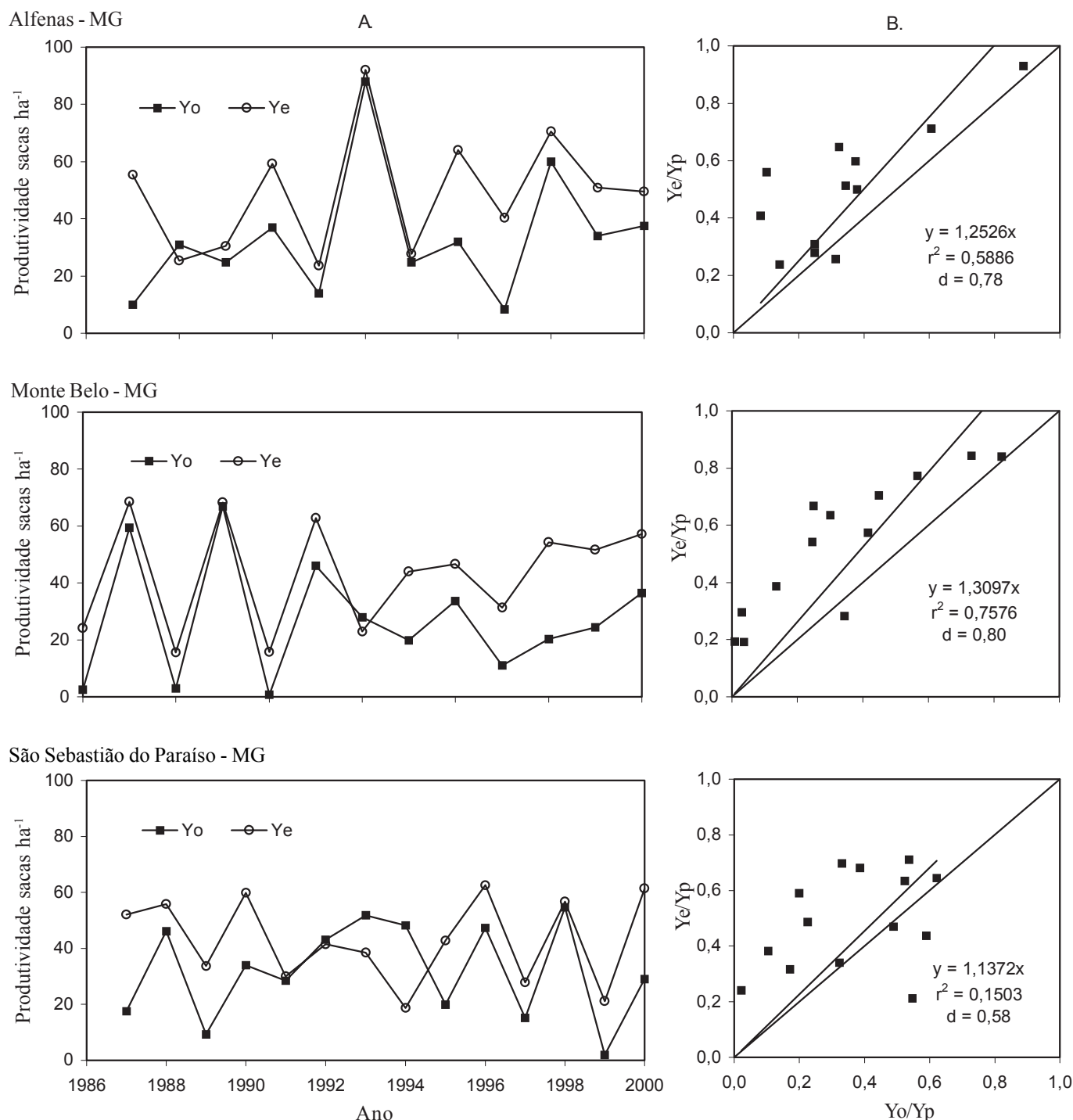


Figura 2. Produtividades observadas  $Y_o$  e estimadas  $Y_e$  pelo modelo de regressão linear múltipla a partir dos componentes principais mais significativos, para diferentes municípios (A) e respectivos índices de desempenho (B)

Os resultados não se mostraram plenamente satisfatórios para o modelo prognóstico proposto. É evidente que, além dos elementos climáticos, outros fatores podem afetar a produção cafeeira, os quais não são contemplados pelo modelo. Dentre esses fatores, o comportamento do mercado, condições adversas do clima, o surgimento de novas tecnologias de manejo e variedades afetam, consideravelmente, a produção, tornando-a bastante instável ao longo dos anos; eles indicam, também, o nível de complexidade para elaborar modelos de previsão de safras cafeeiras. Assim, deduz-se que esse tipo de modelagem se aplica bem somente a dados de produtividades, em que estão inerentes apenas o fenômeno fisiológico da bienalidade e as condições climáticas normais à cafeicultura, razão pela qual tem sua aplicação limitada.

## CONCLUSÕES

A tentativa de se acrescentar novas variáveis ao modelo de Stewart et al. (1976), para prever produtividades de cafeeiros, não se mostrou satisfatória apresentando, ainda, grandes discrepâncias entre produtividades estimadas e observadas, além de ocorrer tendência do modelo em superestimar as produtividades; no entanto, notou-se que o comportamento de variação das produtividades estimadas pelo modelo acompanhou a variação das produtividades observadas.

## LITERATURA CITADA

- Beaumont, J.H. An analysis of growth and yield relationships of coffee trees in the Kona District, Hawaii. *Journal of Agricultural Research*, Washington, D.C., v.59, n.3, p.223-235, 1939.
- Camargo, A.P.; Pereira, A.R. *Agrometeorology of the coffee crop*. Geneva: World Meteorological Organization, 1994. 99p. *Agricultural Meteorological CAgM Report*, 58
- Cruz, C.D.; Regazzi, A.J. *Modelos biométricos aplicados ao melhoramento genético*. 2. ed. Revisada. Viçosa: Editora UFV, 1997. 390p.
- Dean, L.A. Relationships between rainfall and coffee yields in the Kona District, Hawaii. *Journal of Agriculture Research*, Washington, D.C., v.59, n.3, p.217-222, 1939.
- Doorenbos, J.; Kassam, A.H. *Yield response to water*. Rome: FAO, 1979. 197p. *FAO Irrigation and Drainage Paper*, 33
- Dunteman, G. H. *Introduction to multivariate analysis*. Beverly Hills: Sage Publications, 1984. 237 p.
- Haan, C.T. *Statistical methods in hydrology*. 5.ed. Ames: The Iowa State University Press, 1977. 378p.
- Hoogenboom, G. Contribution of agrometeorology to the simulation of crop production and its applications. *Agricultural and Forest Meteorology*, Amsterdam, n.103, p.137-157, 2000.
- Hoogenboom, G.; Jones, J.W.; Boote, K.J. Modeling growth, development, and yield of grain legumes using SOYGRO, PNUTGRO and BEANGRO: A review. *Transactions of the ASAE*, St. Joseph, v.35, n.6, p.2043-2056, 1992.
- Kanemasu, E.T. Yield and water-use relationships: Some problems of relating grain yield to transpiration. In: Taylor, H.M.; Jordan, W.R.; Sinclair, T.R. (ed.). *Limitations to efficient water use in crop production*. Madison: American Society of Agronomy, 1983. cap. 9B, p.413-417.
- Leal, B.G. *Sistema computacional para avaliação e estimativa das produtividades potencial e real de culturas agrícolas*. Viçosa: UFV, 2000. 96p. Tese Doutorado
- Manly, B.F.J. *Multivariate statistical methods: A primer*. London: Chapman and Hall, 1986. 159 p.
- Mardia, K.V.; Kent, J.T.; Bibby, J.M. *Multivariate analysis*. London: Academic Press, 1979. 521 p.
- Matiello, J.B. *O café: do cultivo ao consumo*. São Paulo: Editora Globo, 1991. 319p. Coleção do Agricultor - Grãos
- Montgomery, D.C.; Peck, E.A. *Introduction to linear regression analysis*. 2.ed. New York: John Wiley & Sons, Inc., 1992. 527p.
- Morrison, D.F. *Multivariate statistics methods*. 2. ed. New York: McGraw Hill, 1976. 415 p.
- Pereira, A.R.; Villa Nova, N.A.; Sedyama, G.C. *Evapo(transpi)ração*. Piracicaba: FEALQ, 1997. 183p.
- Picini, A.G. *Desenvolvimento e teste de modelos agrometeorológicos para a estimativa de produtividade do cafeeiro (Coffea arabica L.) a partir do monitoramento da disponibilidade hídrica do solo*. Piracicaba: ESALQ-USP, 1998. 132p. Dissertação Mestrado
- Picini, A.G.; Camargo, M.B.P de; Fazuoli, L.C.; Gallo, P.B.; Ortoloni, A.A. Desempenho de modelo matemático agrometeorológico de estimativa de produtividade para a cultura do café no Estado de São Paulo. In: Congresso Brasileiro de Agrometeorologia, 11, e Reunião Latino-Americana de Agrometeorologia, 2, 1999, Florianópolis. *Anais...* Florianópolis: SBA, 1999. p.542-548.
- Rena, A.B.; Maestri, M. *Fisiologia do cafeeiro*. Informe Agropecuário, Belo Horizonte, v.11, n.126, p.26-40, 1985.
- Robertson, G.W. *Guidelines on crop-weather models*. Geneva: World Meteorological Organization, 1983. 115p. *World Climate Application Programme*, 50
- Stewart, J.I.; Hagan, R.M.; Pruitt, W. *Production functions and predicted irrigation programmes for principal crops as required for water resources planning and increased water use efficiency: Final report*. Washington, U.S. Department of Interior, 1976.
- Willmott, C.J.; Ackleson, S.G.; Davis, R.E.; Feddema, J.J.; Klink, K.M.; Legates, D.R.; O'Donnell, J.; Rowe, C.M. *Statistics for the evaluation and comparison of models*. *Journal of Geophysical Research*, Washington, D.C., v.90, n.C5, p.8995-9005, 1985.