

UNIVERSIDADE ESTADUAL PAULISTA

“JÚLIO DE MESQUITA FILHO”

Edgar Simkevic Martins Dornelas

**Estimativa de rendimento agrícola sobre o estado de SP a partir
de um modelo de regressão linear usando o Índice de
Precipitação-Evapotranspiração Normalizado (SPEI)**

UNIVERSIDADE ESTADUAL PAULISTA

“JÚLIO DE MESQUITA FILHO”

Edgar Simkevic Martins Dornelas

**Estimativa de rendimento agrícola sobre o estado de SP a partir
de um modelo de regressão linear usando o Índice de
Precipitação-Evapotranspiração Normalizado (SPEI)**

Trabalho de conclusão de curso de graduação apresentado à
Universidade Estadual Paulista “Júlio de Mesquita Filho”
como requisito parcial para a obtenção do título de Bacharel
em Meteorologia.

BANCA EXAMINADORA

Prof. Dr. Luiz Felipe Gozzo (orientador(a))
Universidade Estadual Paulista Júlio de Mesquita Filho

Prof. Dr. Demerval Soares Moreira
Universidade Estadual Paulista Júlio de Mesquita Filho

Prof. Dr. Helber Custódio de Freitas
Universidade Estadual Paulista Júlio de Mesquita Filho

D713e

Dornelas, Edgar Simkevic Martins

Estimativa de rendimento agrícola sobre o estado de SP a partir de um modelo de regressão linear usando o Índice de Precipitação-Evapotranspiração Normalizado (SPEI) / Edgar Simkevic Martins Dornelas. -- Bauru, 2024
28 f. : il.

Trabalho de conclusão de curso (Bacharelado - Meteorologia) -
Universidade Estadual Paulista (Unesp), Faculdade de Ciências, Bauru
Orientador: Luiz Felipe Gozzo

1. Meteorologia. 2. Agrometeorologia. 3. Arroz. 4. Estimativa agrícola. 5.
São Paulo. I. Título.

SUMÁRIO

RESUMO.....	4
1 INTRODUÇÃO	5
2 JUSTIFICATIVA	6
3 OBJETIVOS.....	6
4 METODOLOGIA.....	7
5 RESULTADOS E DISCUSSÕES	8
5.1 ARROZ	8
5.2 MILHO	17
5.1 TRIGO	26
6 CONCLUSÕES E PERSPECTIVAS	27
AGRADECIMENTOS.....	28
REFERÊNCIAS	28

RESUMO

A estimativa de rendimento de produções agrícolas tem se tornado cada vez um estudo de mais importância, para desenvolver um futuro sustentável. É de fundamental importância para o desenvolvimento de culturas agrícolas a presença de água, seja de forma direta ou indireta, pois o estresse hídrico impacta a qualidade e a quantidade da produção. Neste trabalho foi avaliado um modelo de regressão linear para relacionar condições de anomalias de disponibilidade hídrica (através da média do índice de Evapotranspiração-Precipitação Padronizado na escala de 12 meses - SPEI-12 - para todo o estado de São Paulo) com o rendimento de algumas culturas importantes no estado de São Paulo - arroz, milho e trigo - utilizando dados do período de 1974 a 2000. Apesar de os coeficientes de correlação linear de Pearson entre as culturas do arroz e do milho terem sido moderados e com significância estatística, o modelo linear para estimativa de rendimento futuro apresentou fraco desempenho, com grande viés positivos e erros. Isto pode estar relacionado com a grande variabilidade espacial do fenômeno de seca, sugerindo que um índice único do SPEI-12 para todo o estado pode não ser representativo da seca nas regiões produtoras mais importantes do estado. Além disso, a análise de qualidade do modelo linear mostrou que seus resíduos apresentaram comportamento não-normal e heterocedástico, indicando que um método linear não é a melhor ferramenta para estudar a relação entre as variáveis estudadas.

ABSTRACT

The estimation of agricultural production yields has increasingly become a study of greater importance for developing a sustainable future. It is of fundamental importance for the development of agricultural crops to have water presence, either directly or indirectly, as water stress impacts both the quality and quantity of production. In this study, a linear regression model was evaluated to relate anomalies in water availability conditions (using the standardized Precipitation-Evapotranspiration Index on a 12-month scale - SPEI-12 - for the entire state of São Paulo) to the yield of some important crops in the state of São Paulo - rice, corn, and wheat - using data from the period 1974 to 2000. Although the Pearson linear correlation coefficients between rice and corn crops were moderate and statistically significant, the linear model for estimating future yields performed poorly, showing significant positive bias and errors. This may be related to the large spatial variability of the drought phenomenon, suggesting that a single SPEI-12 index for the entire state may not adequately represent drought in the most important producing regions of the state. Furthermore, the analysis of the linear model's quality showed that its residuals exhibited non-normal and heteroskedastic behavior, indicating that a linear method is not the best tool for studying the relationship between the variables under consideration.

1 INTRODUÇÃO

A oferta de alimentos tem se tornado, cada vez mais, uma pauta mundialmente discutida. Condições climáticas adversas e eventos extremos decorrentes, possivelmente, das mudanças climáticas, tendem a ter grande impacto na disponibilidade de produtos agrícolas. Dentre estas condições adversas, a seca é um dos principais problemas (Meza, et al., 2020; Batista e Albuquerque, 2022). Jaleel et al. (2009) mostram que a falta de água afeta as plantas de modo a prejudicar seu crescimento, sendo um importante fator limitante na fase inicial da planta, tanto para o seu crescimento quanto para sua estabilização. Portanto, é importante conhecer e quantificar o impacto das secas na produção para que se possa desenvolver um melhor planejamento para mitigação de seus danos.

O estado de São Paulo (SP) é um importante polo agrícola do Brasil. Dentre as principais culturas produzidas no estado está o arroz, que é um dos principais fornecedores de carboidratos para a população brasileira (Boaretto e Moraes, 2010). Apesar de SP ter uma área plantada de arroz relativamente pequena, a produção deste grão vem crescendo e ganhando importância, com o estado tornando-se o estado com maior rendimento de produção do grão na região Sudeste a partir da última década (Moura e Landau, 2020).

A seca afeta o crescimento, maturação e o rendimento do arroz de forma significativa (Lafitte, 2007). Pesquisas recentes indicam probabilidade de diminuição no rendimento agrícola global em períodos de seca – para o Brasil, a probabilidade de queda para o arroz e o milho, durante estes períodos, é estimada em torno de 60% (Leng e Hall, 2019). Eventos de seca são recorrentes sobre o estado de SP, sendo causados por mecanismos climáticos complexos, desde a escala intrasazonal até a multidecadal, decorrentes de processos atmosféricos tanto extratropicais como tropicais (Gozzo et al., 2021; Gozzo et al., 2022). Para identificar tais períodos de seca, foram desenvolvidos diversos índices, como o Índice de Precipitação Padronizado - SPI (WMO, 2012) e o Índice Bhalme-Mooley - BMI (Bhalme e Mooley, 1979). Muitos índices de seca, como os citados, levam em conta apenas a variabilidade da precipitação, porém períodos secos podem estar relacionados tanto à diminuição da precipitação em um local, como também ao aumento da demanda evaporativa da atmosfera (evapotranspiração), associada ao aumento da temperatura do ar. A maior parte de SP já registra tendência de aumento de temperatura nas últimas décadas (Fante e Neto, 2017), e projeta-se que a temperatura siga aumentando, no cenário das mudanças climáticas (de Godoy et al., 2021). Portanto, é recomendável que a identificação de secas no estado ao longo de diferentes décadas seja realizada utilizando-se um índice que leva em conta não só a variabilidade da precipitação, mas também a demanda evaporativa. O SPEI é um índice criado para contabilizar os efeitos da evapotranspiração no balanço hídrico regional e adequado para regiões onde se verifica sensível aumento da temperatura devido às mudanças climáticas (Vicente-Serrano et

al., 2010; Nascimento, 2021).

Gozzo et al. (2019) documentaram que eventos de seca severa vêm aumentando em grande parte de SP ao longo do último século. Projeções climáticas indicam que a frequência das secas sobre SP deve aumentar no futuro (IPCC, 2021). Neste cenário, é cada vez mais relevante investigar e quantificar o impacto dos períodos secos na produção agrícola. Muitos estudos têm se dedicado a determinar essa relação. Kamble et al. (2019), por exemplo, relacionaram o SPI com a produção de arroz na Índia, encontrando correlações altas e estatisticamente significativas, e desenvolvendo modelos de regressão capazes de explicar até 70% da variabilidade da produção em termos do SPI. No Brasil, Fernandes e Heinemann (2011) constataram que o índice SPI aplicado em uma escala de 12 meses (SPI-12) é o mais eficiente em identificar variações no rendimento de arroz na região de Goiás, pois é capaz de assimilar secas de maior período e impacto.

Dada a relevância sócio-econômica deste tema, e a escassez de trabalhos focados no estado paulista, o objetivo deste trabalho é avaliar o impacto da seca na produção anual de arroz, milho e trigo em SP. Pretende-se obter e validar um modelo linear que relacione o rendimento com o índice SPEI.

2 JUSTIFICATIVA

O impacto dos períodos secos no rendimento agrícola é uma questão fundamental para a sustentabilidade futura. Pesquisas recentes indicam probabilidade de diminuição no rendimento agrícola global em períodos de seca – para o Brasil, a probabilidade de queda para o arroz e o milho, durante estes períodos, é estimada em torno de 60% (Leng e Hall, 2019). Logo, estudos para estabelecer a relação entre secas e rendimento em escala regional são de grande relevância no cenário atual, e ainda são escassos no Brasil.

A escolha do estado de SP para este estudo baseia-se em sua importância no cenário agrícola brasileiro, e no fato de que projeções e estudos atuais indicam que, no cenário das mudanças climáticas, os extremos secos devem ficar mais frequentes e longos sobre SP, no futuro (IPCC, 2021).

3 OBJETIVOS

O primeiro objetivo deste projeto é avaliar estatisticamente a relação entre o rendimento agrícola do trigo, milho e arroz sobre o estado de SP e o volume de chuva registrado, a partir do SPEI. O segundo objetivo é construir, a partir da relação obtida, um modelo linear de previsão de rendimento a partir do SPEI.

4 METODOLOGIA

A região de estudo compreende o estado de SP, no Sudeste do Brasil (Fig. 1a). Dados de rendimento anual de arroz (em toneladas/ha), assim como o rendimento do milho e do trigo, por município produtor foram obtidos junto ao Instituto Brasileiro de Geografia e Estatística (IBGE), que os disponibiliza em <https://sidra.ibge.gov.br/pesquisa/pam/tabelas>. O período de estudo foi de 1974 a 2010.

Os dados brutos de rendimento apresentaram evidente tendência de aumento ao longo das décadas, e este padrão deve ser removido da série antes da comparação com o SPEI. Para isso, utilizou-se a metodologia da regressão polinomial com ponderação local, seguindo Lu et al. (2017). O cálculo foi realizado utilizando a função *loess* do software estatístico R, com intervalo de ponderação (α) de 0,5.

Para o cálculo do SPEI, foram utilizadas a precipitação acumulada mensal e temperatura média mensal da análise *Climate Research Unit* (CRU) TS4.05 (Harris et al., 2020), com resolução horizontal de 0.5° de longitude e latitude. Na região sob investigação, a confiabilidade dos dados provenientes do *Climate Research Unit* (CRU) pode ser limitada devido à baixa densidade de informações disponíveis na mencionada localidade. Foram calculadas as médias destas variáveis na área 53°W-44°W; 25°S-20°S (Retângulo na Fig. 1a - esquerda), e a partir delas foi calculada a evapotranspiração potencial (PET) pelo método de Thornthwaite (1948). A diferença entre a precipitação e a PET forneceu o balanço hídrico mensal sobre o estado, como expresso na equação 1.

$$D = P - PET, \quad (1)$$

onde P é a precipitação mensal (mm) e PET (mm) é calculado segundo o método de Thornthwaite (1948), que requer apenas dados de temperatura média mensal e localização geográfica da região de interesse.

À série temporal climatológica do balanço mensal, calculada com o período de referência 1974-2010, foi ajustada uma distribuição de probabilidade log-logística normalizada, e a comparação entre o balanço observado em um certo mês e a distribuição de probabilidade climatológica forneceu o valor do índice SPEI para tal mês.

Maiores detalhes do cálculo do SPEI podem ser encontrados em Vicente-Serrano et al. (2010). Neste trabalho, o índice foi calculado na escala de acumulação de 12 meses (SPEI-12), e o rendimento anual foi correlacionada a uma série temporal

composta pelos SPEI-12 de dezembro de cada ano (correspondendo, portanto, ao desvio do balanço hídrico acumulado anual de janeiro a dezembro daquele ano).

O coeficiente de correlação entre as séries foi obtido pela correlação linear de Pearson, e sua significância estatística foi avaliada a partir do p-valor obtido pelo teste de hipótese da correlação.

O cálculo do SPEI foi realizado utilizando o pacote SPEI, do software estatístico R (<https://cran.r-project.org/web/packages/SPEI/SPEI.pdf>). Mapas de distribuição espacial do índice foram gerados com software de visualização *Grid Analysis and Display* (GrADS).

Com o modelo linear obtido, foi aplicado um ajuste estatístico visando melhores resultados de previsão do modelo, o método de ajuste estatístico utilizado foi o MOS (*Model Output Statistics*) (Glahn e Lowry, 1972) que consiste em determinar uma relação estatística entre a predição e o ocorrido de uma variável, realizando uma regressão linear entre o predito e o ocorrido, encontrando a equação da reta desta regressão e aplicando os coeficientes da reta aos dados preditos, com o objetivo de melhorar a saída original do modelo. Neste estudo, o índice SPEI refere-se ao SPEI-12, calculado para um período de 12 meses.

5 RESULTADOS E DISCUSSÃO

5.1 ARROZ

O arroz é cultivado em praticamente todo o estado, com muitos municípios produtores no leste, oeste e norte de SP. Os principais produtores localizam-se na região do Vale do Paraíba (Fig. 1a), com rendimento anual em torno de 4000 toneladas/ha.

A série temporal de rendimento bruto anual total do estado (Fig 1b) apresenta um crescimento sistemático relacionado ao aperfeiçoamento tecnológico, melhoramento de fertilizantes e de técnicas de produção. A curva de regressão polinomial com melhor ajuste aos dados (em azul) mostra o crescimento ainda mais acentuado a partir da década de 1990. Após a remoção da tendência (por meio de decomposição multiplicativa utilizando a curva de regressão), e a normalização (média subtraída de cada dado anual, e divisão pelo desvio padrão do período de estudo), os dados estão preparados para serem comparados à série do SPEI.

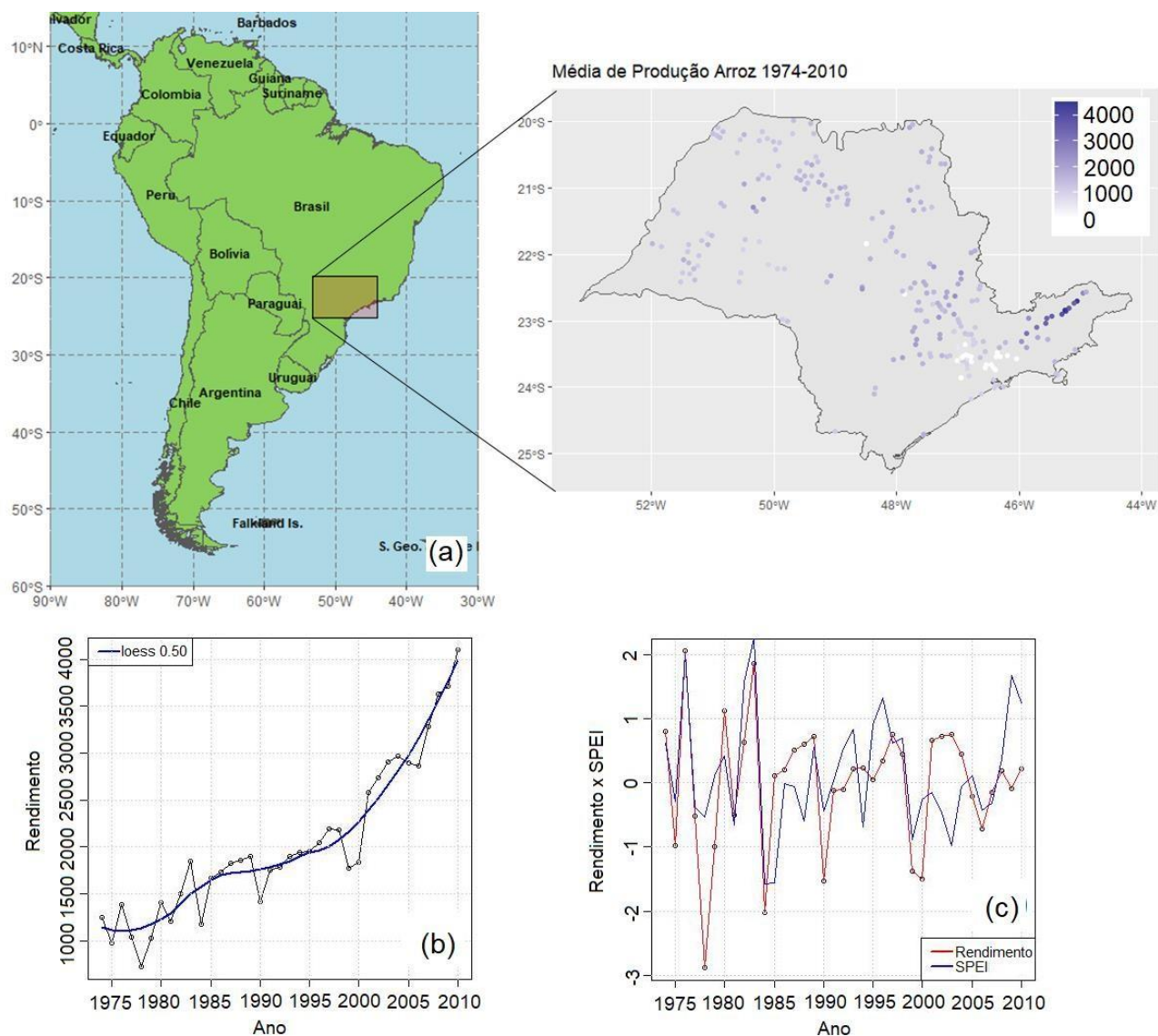


Figura 1: (a) Área de estudo e média anual do rendimento (toneladas/ha) dos municípios produtores de arroz em SP no período 1975-2010; (b) rendimento anual total bruto de arroz no estado e curva de regressão polinomial; (c) séries temporais de rendimento normalizado e SPEI. (Fonte: De autoria Própria).

A Fig. 1c apresenta as séries temporais de rendimento normalizada e do índice SPEI no período de estudo. A correlação linear de Pearson forneceu um coeficiente de +0,54 com significância estatística superior a 99% (p -valor = 0,0004), indicando que o balanço hídrico e a rendimento são moderadamente proporcionais. Pode-se observar na série que anos com forte anomalia negativa de rendimento são anos em que o SPEI está negativo, ou seja, há déficit hídrico (1975, 1978, 1984, 1990, 1999/2000 e 2006). Porém, nem sempre observa-se este comportamento - muitas vezes, quando o SPEI é negativo, o rendimento normalizado é positivo (por exemplo: 1988, 1994, 2003...). Isto pode ocorrer devido a diferentes distribuições espaciais do fenômeno de seca, como pode ser exemplificado pelos anos de 1975 e 1994.

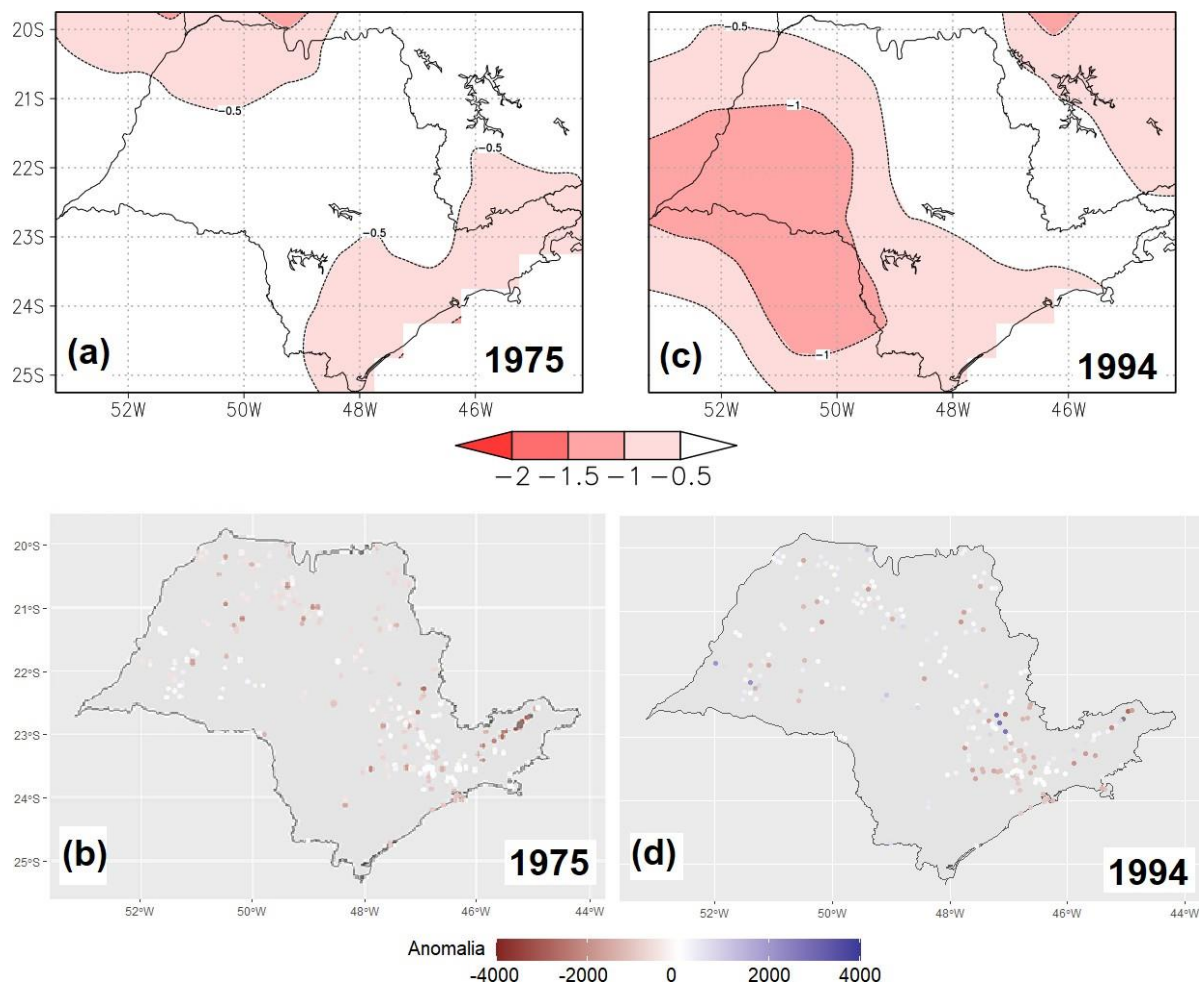


Figura 2: Distribuição espacial do índice SPEI-12 ($< -0,5$) em SP e anomalia de rendimento (tonelada/ha) para os anos de (a), (b) 1975 e (c), (d) 1994.

Em 1975, a condição seca foi registrada no leste e noroeste de SP (Fig. 2a), coincidindo com regiões produtoras relevantes do estado (Fig. 1a). Muitos municípios destas áreas registraram queda em torno de 3000-4000 toneladas/hectare no rendimento anual (Fig. 2b). Assim, tanto o SPEI como rendimento decaíram (Fig. 1c).

O mesmo comportamento não se verifica em 1994, quando houve grande diminuição no valor do balanço hídrico, porém o rendimento não acompanhou essa queda (Fig. 1c). A seca se concentrou no oeste e sul de SP (Fig. 2c). Como as principais regiões produtoras não foram afetadas por essa condição climática, muitos dos municípios com maior produção registraram rendimento próximo à média do período (Fig. 2d).

Para a confecção de um modelo de regressão linear para a previsão de futuras produções, é necessário entender como os dados de produção se comportam e se distribuem na série temporal e analisar se tais dados cumprem os requisitos necessários para a construção de tal modelo (resíduos com distribuição normal, homoscedasticidade e ausência de *outliers*).

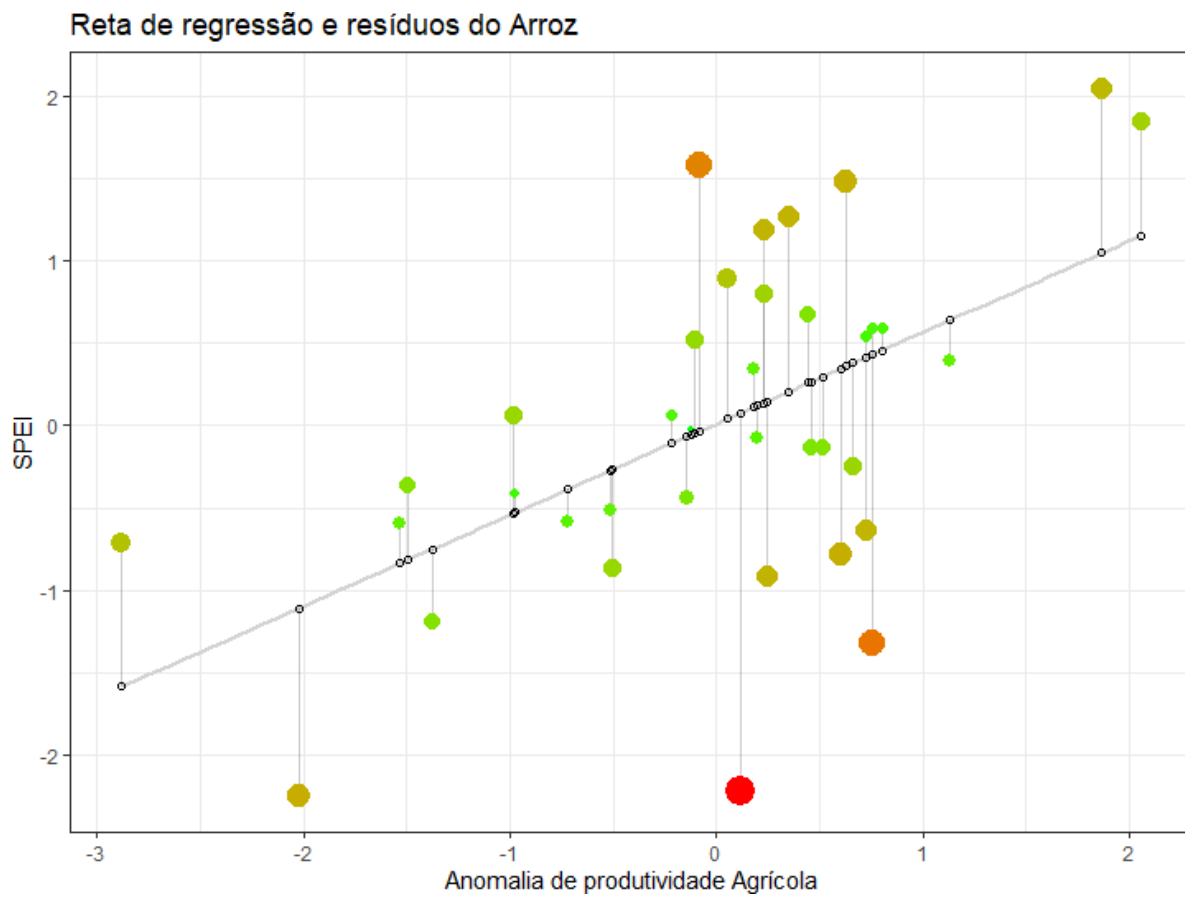


Figura 3: Reta de regressão linear e resíduos para arroz. A magnitude do resíduo é representada pelo tamanho e cor dos pontos (verde, próximo da reta, e vermelho, distante da reta)

A regressão linear do arroz nos forneceu os coeficientes linear e angular da regressão de 0.06 e 0.47 ($y = a + b.x$), respectivamente, com uma significância maior que 99% ($p\text{-val} = 0.0012$). Para uma caracterização mais detalhada do modelo, foi realizada a análise de resíduos.

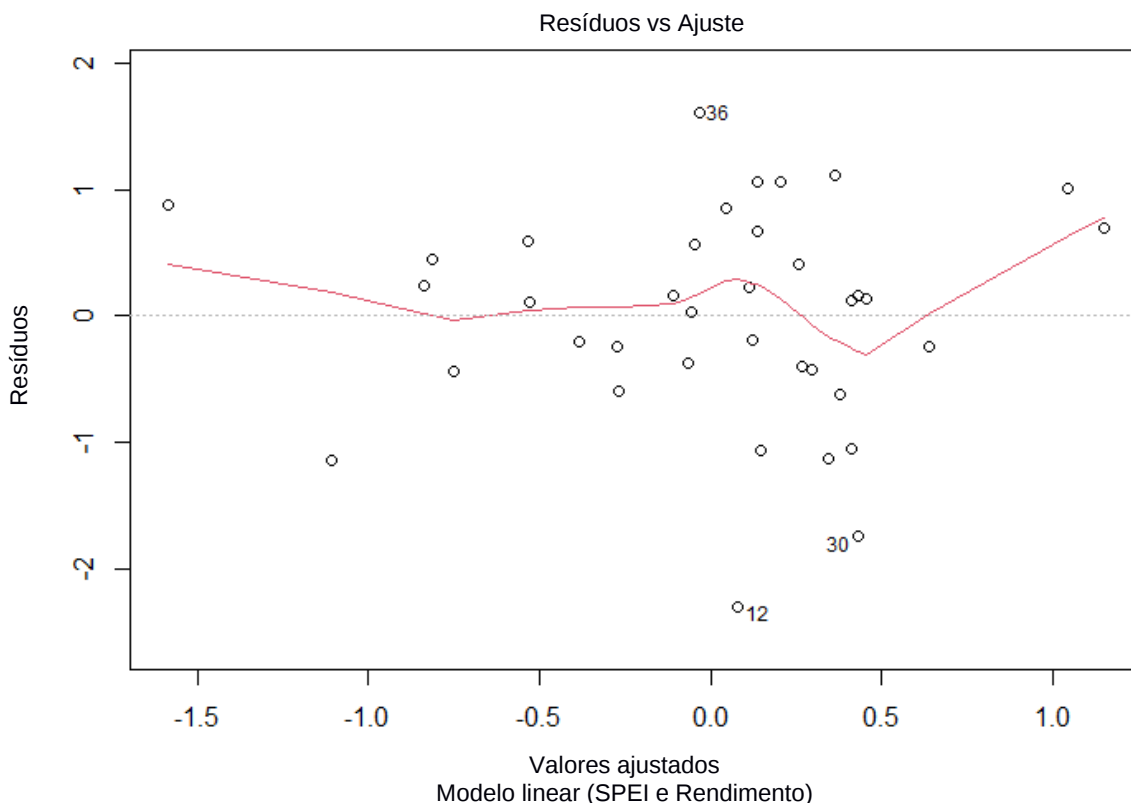


Figura 4: Gráfico de resíduos do modelo linear entre SPEI e rendimento normalizado do arroz para o estado de SP, no período de 1974 a 2000. Os resíduos estão na mesma escala que a variável normalizada.

O gráfico de resíduos possibilita verificar se a hipótese da relação linear entre os dados pode ser assumida neste modelo. Num caso ideal onde as variáveis possuíssem uma relação perfeitamente linear todos os resíduos seriam encontrados sobre a reta horizontal de valor 0 na Figura 4. Se os dois conjuntos de dados analisados possuíssem outra relação que não fosse linear (como quadrática, por exemplo), este padrão apareceria na distribuição dos pontos no gráfico de resíduos. A Fig. 4 mostra que no caso do SPEI e rendimento do arroz, eles não apresentaram nenhum padrão claramente observado, indicando certa linearidade, mas para representar um modelo linear de melhor qualidade, os pontos deveriam ser distribuídos mais homogeneamente acima e abaixo da reta de resíduo = 0, e com uma maior concentração na parte central do gráfico.

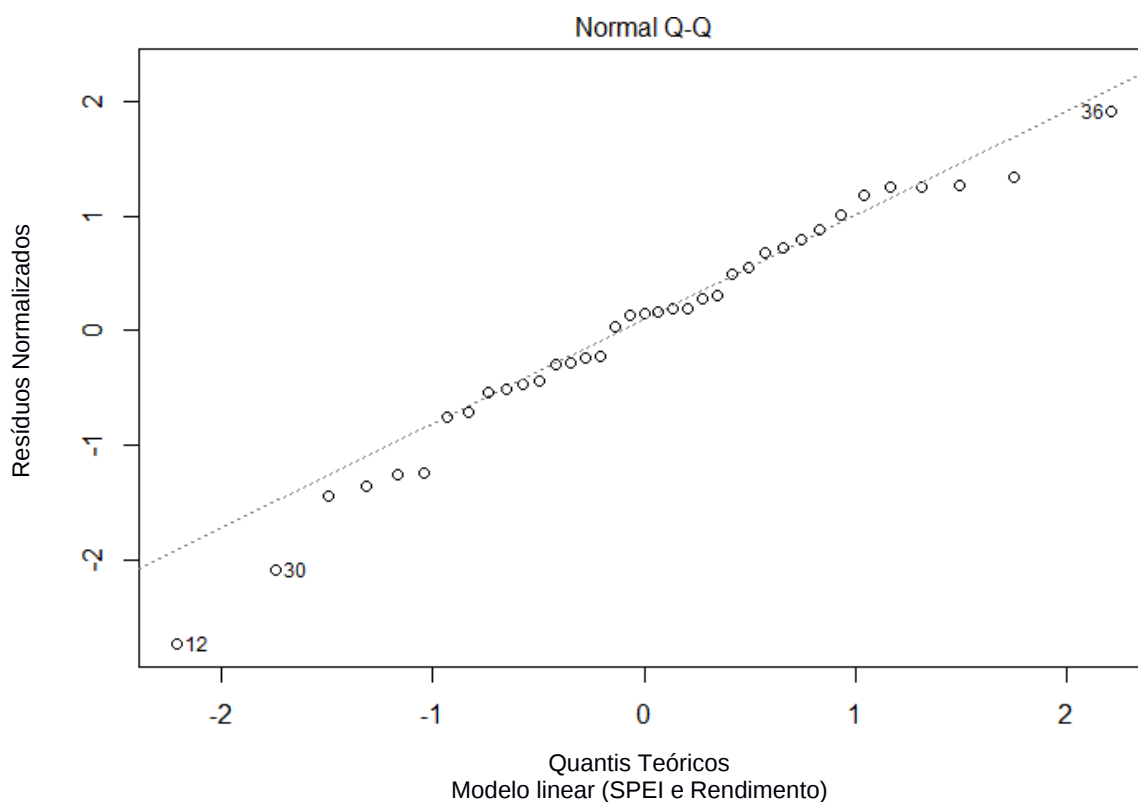


Figura 5: QQ-plot dos resíduos do modelo entre SPEI e rendimento do arroz.

Os gráficos quantil-quantil (*QQ plots*) são uma ferramenta estatística que essencialmente responde se a distribuição que estamos analisando (no caso os resíduos do modelo) é normal, com os pontos distribuindo-se sobre uma reta em caso afirmativo. No caso do modelo linear construído aqui, o *QQ plot* com os pontos desviando de uma reta nas extremidades (Fig. 5) ocorre porque os resíduos apresentam um número grande de dados nas extremidades da distribuição, que não pode ser considerada perfeitamente normal.

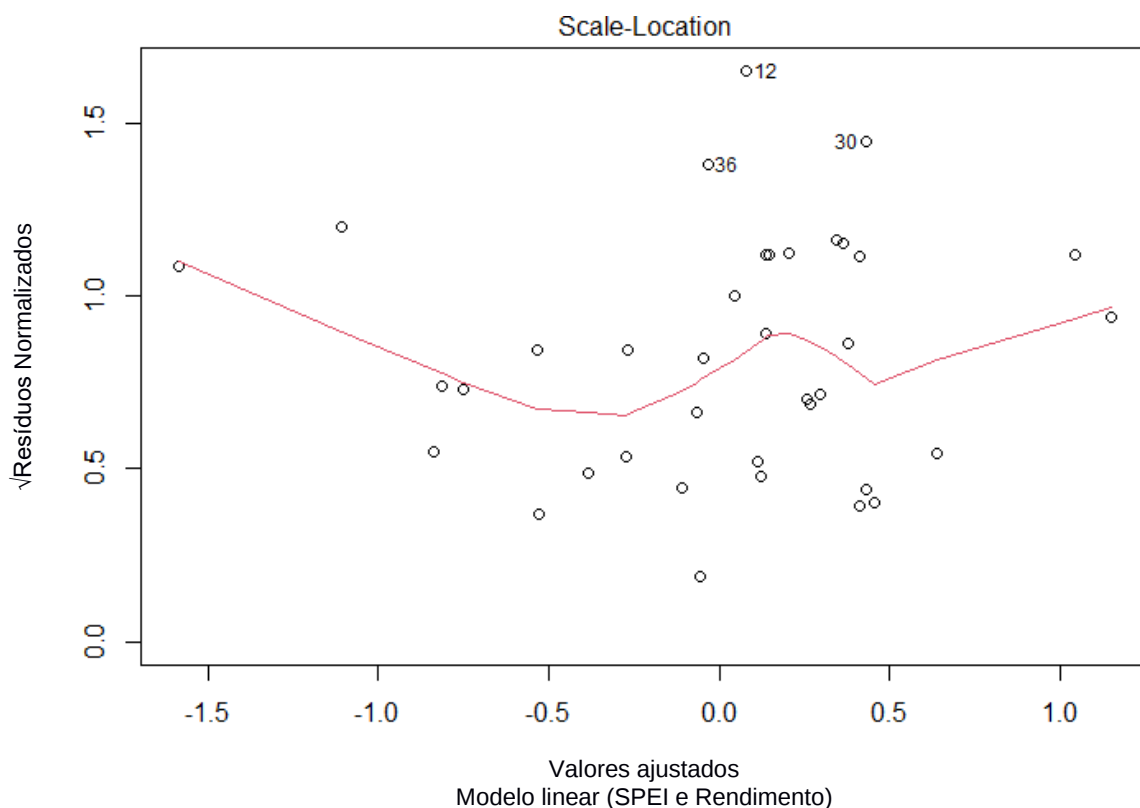


Figura 6: Variância dos resíduos do modelo linear entre SPEI e rendimento do arroz

O gráfico de variância dos resíduos deveria mostrar a variância distribuída igualmente em torno da linha vermelha horizontal ao longo de sua extensão, indicando que o modelo possui homocedasticidade. No presente estudo (Fig. 6) linha tem orientação horizontal e os pontos se distribuem de forma aproximadamente simétrica acima e abaixo dela. Porém, não se distribuem igualmente ao longo da mesma, mostrando que os resíduos não possuem variância constante, ou seja, não apresentam homocedasticidade. Essa característica já podia ser inferida do comportamento da Fig. 4, e é confirmada por esta análise adicional.

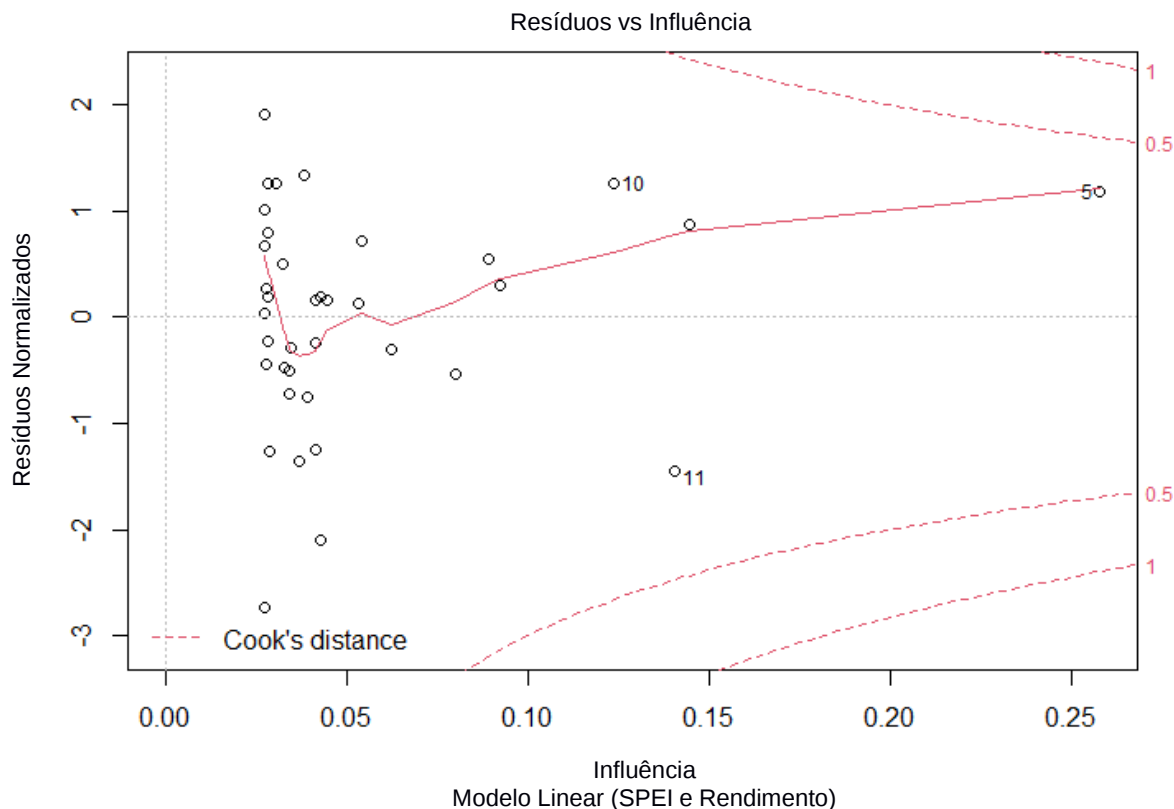


Figura 7: Resíduos padronizados e influência dos resíduos do arroz. A linha vermelha tracejada é a distância de Cook.

O gráfico da Fig. 7 foi usado para identificar dados influenciadores do conjunto, ou seja, aqueles cuja remoção poderia modificar o modelo. Quando estes dados ocorrem, sua inclusão ou exclusão deve ser considerada a fim de obter maior qualidade no modelo. Eles seriam indicados graficamente pela sua posição além da linha da distância de Cook (linha vermelha tracejada). No modelo entre SPEI e arroz construído aqui, nenhum resíduo foi encontrado além da distância de Cook, indicando que nenhum dado ao ser excluído exerceria influência significativa no modelo linear resultante.

A partir dos dados obtidos a respeito da correlação entre o rendimento dos produtos e o SPEI, e com a análise de resíduos de uma regressão linear, foi produzido um modelo linear para a previsão do rendimento do arroz a partir do SPEI. Para que este modelo, elaborado com dados de 1974-2000, fosse validado, tal foi aplicado nos anos de 2001-2011 para que seu desempenho fosse examinado, e então foi testada a correção estatística (MOS) do modelo no período de 2011-2021 buscando melhorar o desempenho do mesmo.

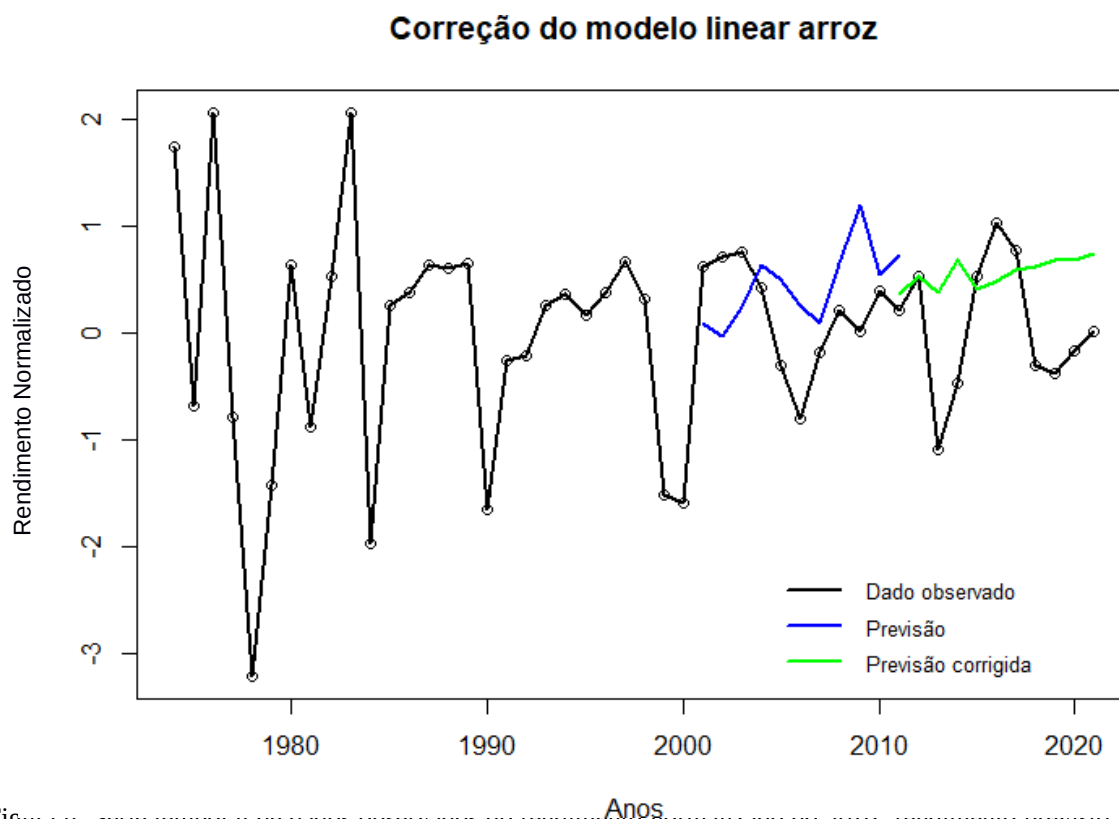


Figura 8. Série temporal de dados observados do rendimento normalizado do arroz, rendimento previsto a partir do modelo linear e rendimento previsto a partir do modelo linear corrigido pelo método MOS.

O Erro Médio (Viés) entre os dados de rendimento previstos pelo modelo linear e documentados de rendimento do arroz foi de +0.44, indicando que de forma geral o modelo baseado no SPEI tende a superestimar o rendimento, com uma magnitude (Raiz Quadrada do Erro Médio Quadrático) de +0.67. Além do erro na magnitude do dado, pode-se ver também (Fig. 8) que a série prevista não acompanha bem a série observada: embora o padrão esteja presente (aumento do rendimento, seguido de queda e novo aumento), a série prevista aparece defasada em relação aos dados, indicando que o modelo não representa satisfatoriamente a realidade. Quando aplicada a correção estatística a este modelo, no período de 2011-2021, foi obtida uma série temporal com Erro Quadrático Médio (Viés) de +0.645 e Raiz Quadrada do Erro médio de +0.803 em relação a série temporal do rendimento observados nestes anos. Fica claro que tal correção não é a ideal neste caso, pois amplifica o erro do modelo original.

5.2 MILHO

O milho é uma cultura amplamente produzida em SP, com a maior parte de sua produção concentrada no norte e centro-leste do estado (Fig. 9a). A série temporal de rendimento bruta (Fig. 9b) mostra um crescimento evidente ao longo do tempo. A tendência de incremento tecnológico foi retirada utilizando o mesmo método aplicado para retirar a tendência do arroz, e o rendimento sem tendência e normalizada está expressa na Fig. 9c, juntamente com a série temporal do SPEI.

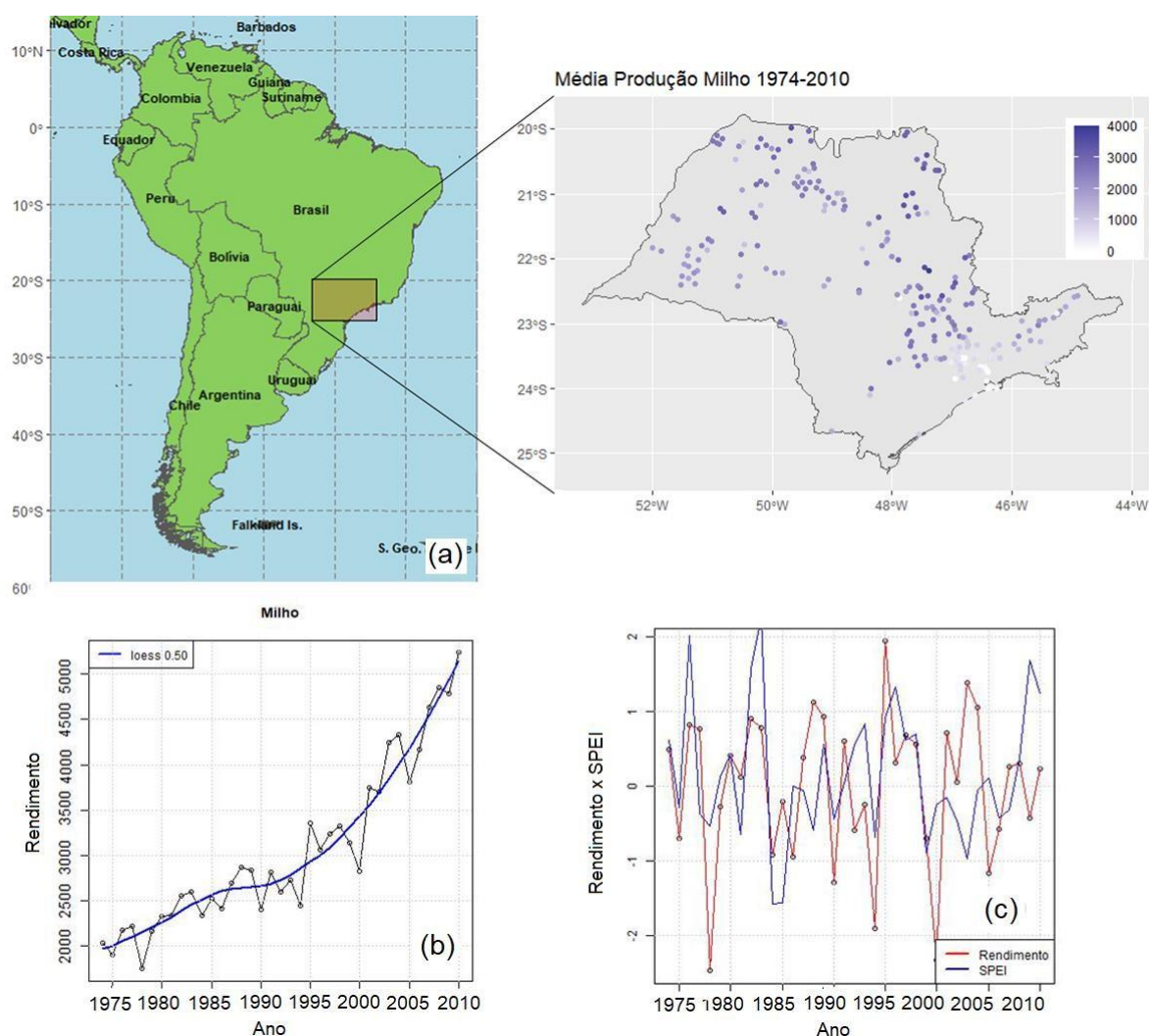


Figura 9: (a) Área de estudo e média anual do rendimento (toneladas/ha) dos municípios produtores de milho em SP no período 1975-2010; (b) rendimento anual total bruto de milho no estado e curva de regressão polinomial; (c) séries temporais de rendimento normalizado e SPEI. (Fonte: De autoria Própria)

O coeficiente de correlação linear de Pearson das séries temporais da Fig. 9c foi de +0,35 com significância estatística de 97% (p -valor = 0,03), sugerindo uma proporcionalidade moderada entre o balanço hídrico e o rendimento do milho. É possível notar na série temporal

que em vários anos com déficit (excedente) de balanço hídrico, o rendimento também é deficitário (superior à média) (por exemplo: 1975, 1976, 1978, 1980, 1989, 1994, 1999, 2009). No entanto, outros anos mostram que esse comportamento nem sempre é reproduzido (1985, 1986, 1988, 1992, 2000, 2005, 2010). e as análises da seção 3.1 indicam que a distribuição espacial da seca na área do estado de São Paulo pode ser uma causa dessa discrepância.

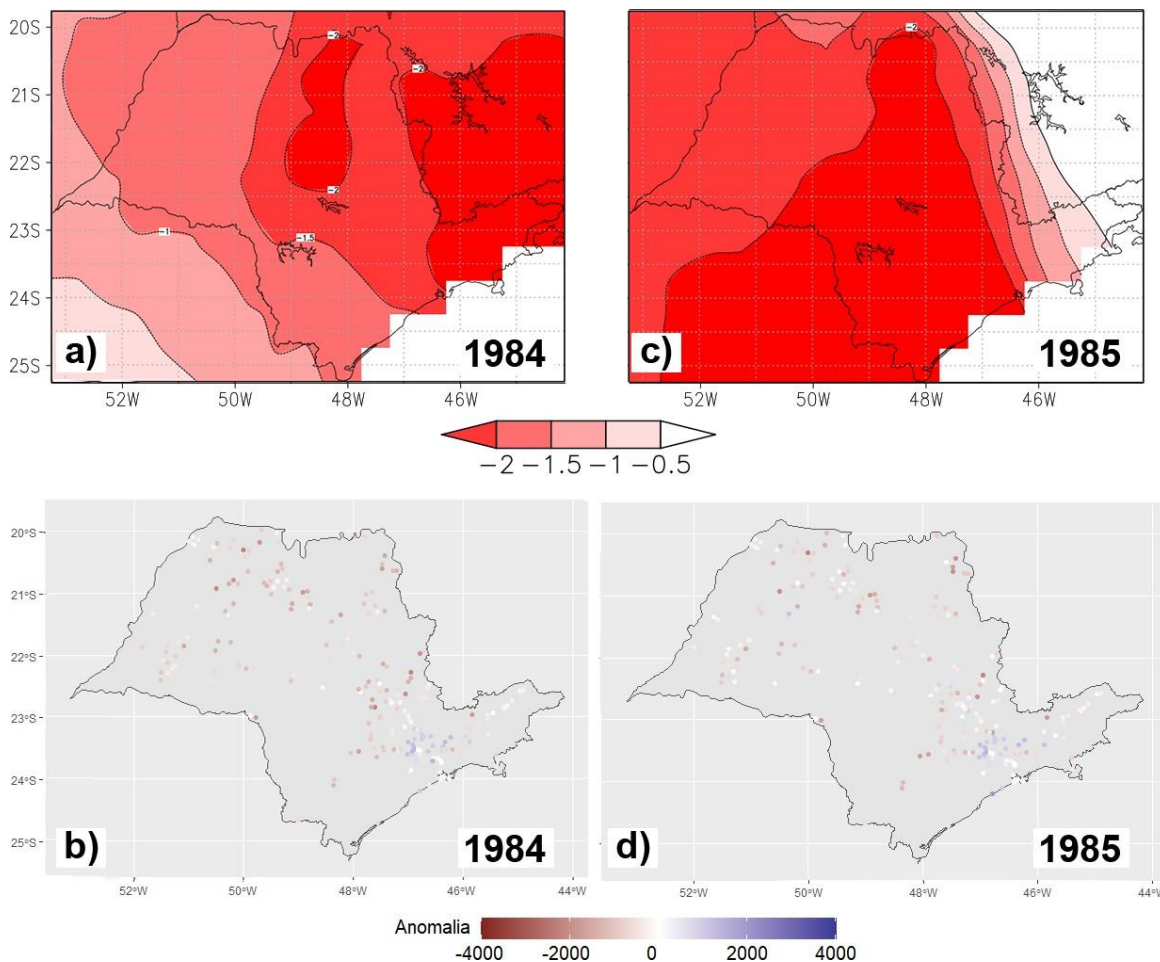


Figura 10: Distribuição espacial do índice SPEI-12 ($< -0,5$) em SP e anomalia de rendimento (tonelada/ha) para os anos de (a), (b) 1984 e (c), (d) 1985.

Tomemos o exemplo dos anos 1984/1985 (Fig. 10) Não houve alterações significativas de SPEI nestes anos, todo o estado encontrava-se em uma condição de seca, apenas alterando de um ano para outro a posição e a área abrangida pela seca mais extrema. Porém, mesmo com uma área maior de seca extrema, e englobando regiões produtoras importantes do estado (Fig. 10a), o rendimento de 1985 foi mais elevada do que a de 1984.

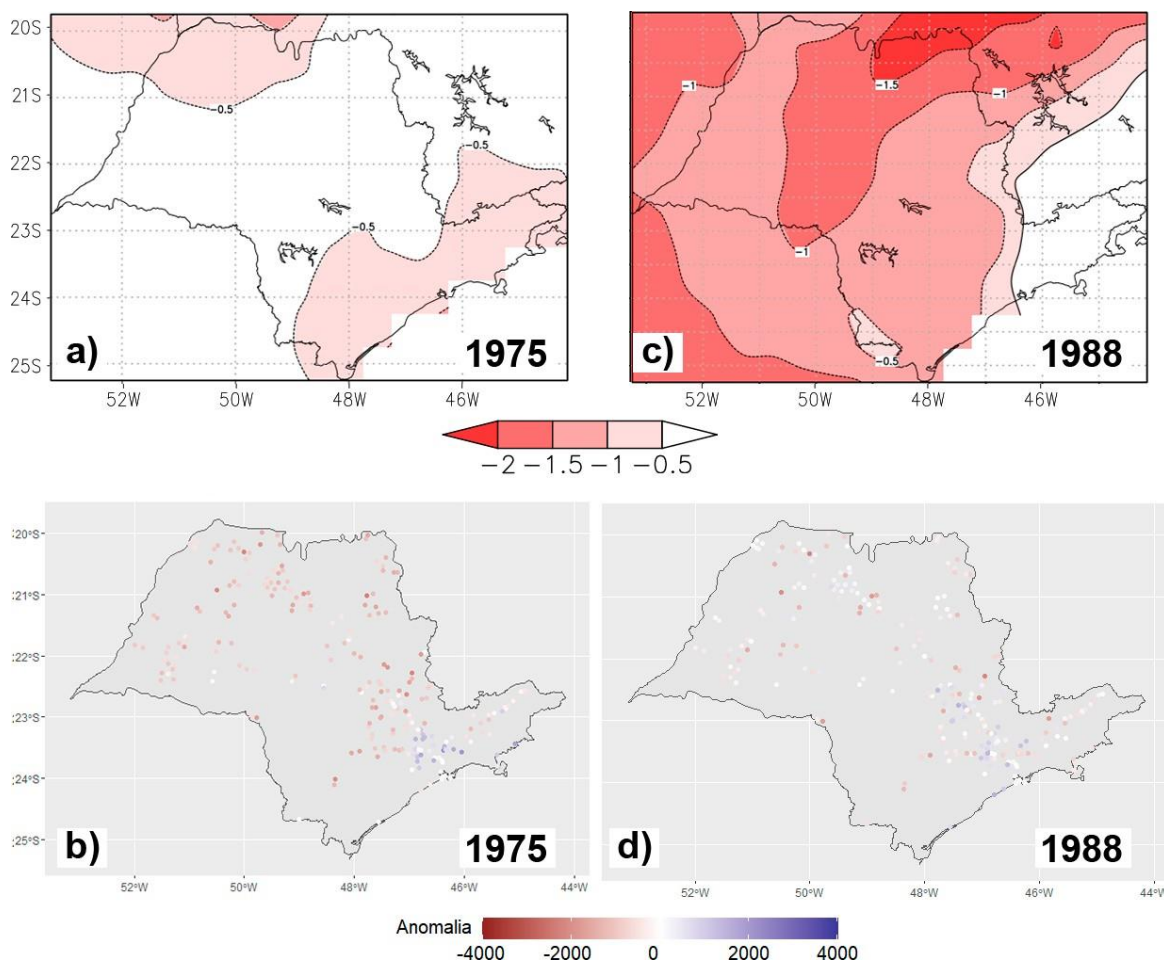


Figura 11: Distribuição espacial do índice SPEI-12 ($< -0,5$) em SP e anomalia de rendimento (tonelada/ha) para os anos de (a), (b) 1975 e (c), (d) 1988.

A Fig. 11 apresenta um cenário onde o SPEI de 1988 se agrava em relação ao de 1975 e passa a abranger grande parte do estado, até mesmo as regiões produtoras importantes, porém o rendimento resultante de 1988 foi maior do que a de 1975, mesmo que numericamente o SPEI de 1975 tenha valor mais neutro de 1988 valor mais negativo.

Comparando-se os resultados do arroz e do milho, pode-se notar que a influência da seca na produção parece ainda menos determinante na produção de milho que na de arroz. Isso se reflete matematicamente no fato do arroz apresentar um coeficiente de correlação maior e mais significativo do que o do milho.

É apresentado, a seguir, um modelo de regressão linear calculado para a cultura do milho, assim como a análise de resíduos para verificação da qualidade deste modelo, e sua validação.

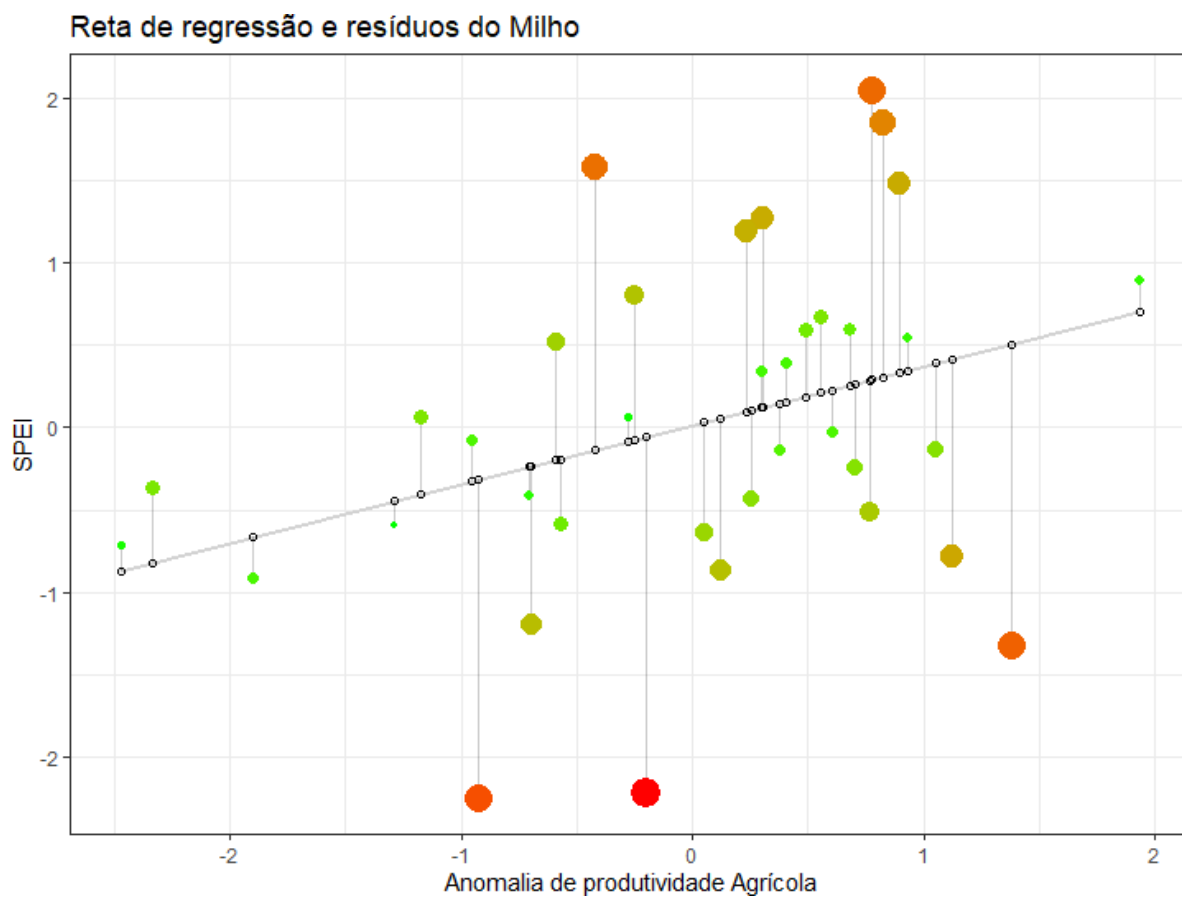


Figura 12: Reta de regressão linear e resíduos do milho, onde o tamanho do resíduo é representado por seu tamanho e cor (verde, próximo da reta, e vermelho, distante da reta)

A regressão linear do milho nos forneceu os coeficientes linear e angular da regressão de 0.06 e 0.32, respectivamente, com uma significância de aproximadamente 97% ($p\text{-val}=0.033$).

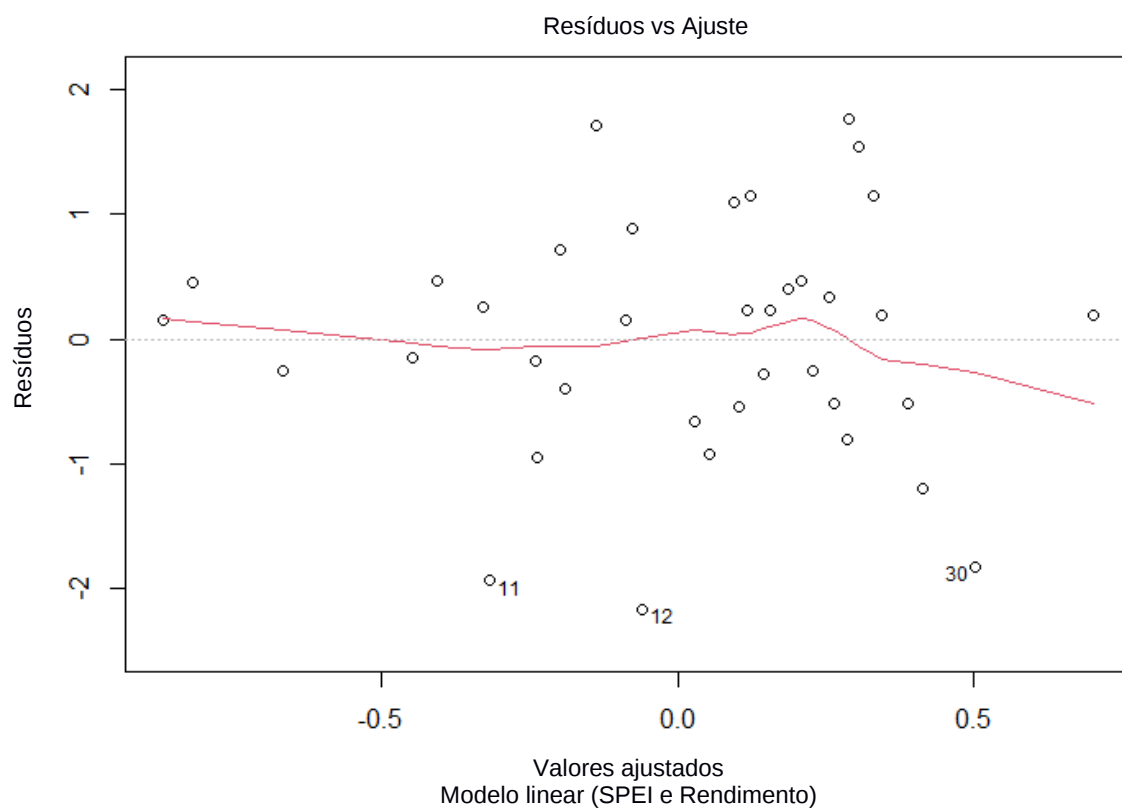


Figura 13: Gráfico de resíduos com valores ajustados e resíduos do modelo linear entre SPEI e rendimento do milho para o estado de SP, no período de 1974 a 2000. Os resíduos estão na mesma escala que a variável normalizada.

O gráfico de resíduos e reta do modelo linear ajustado (Fig. 13) também apresenta uma distribuição homogênea, assim como no caso do arroz.

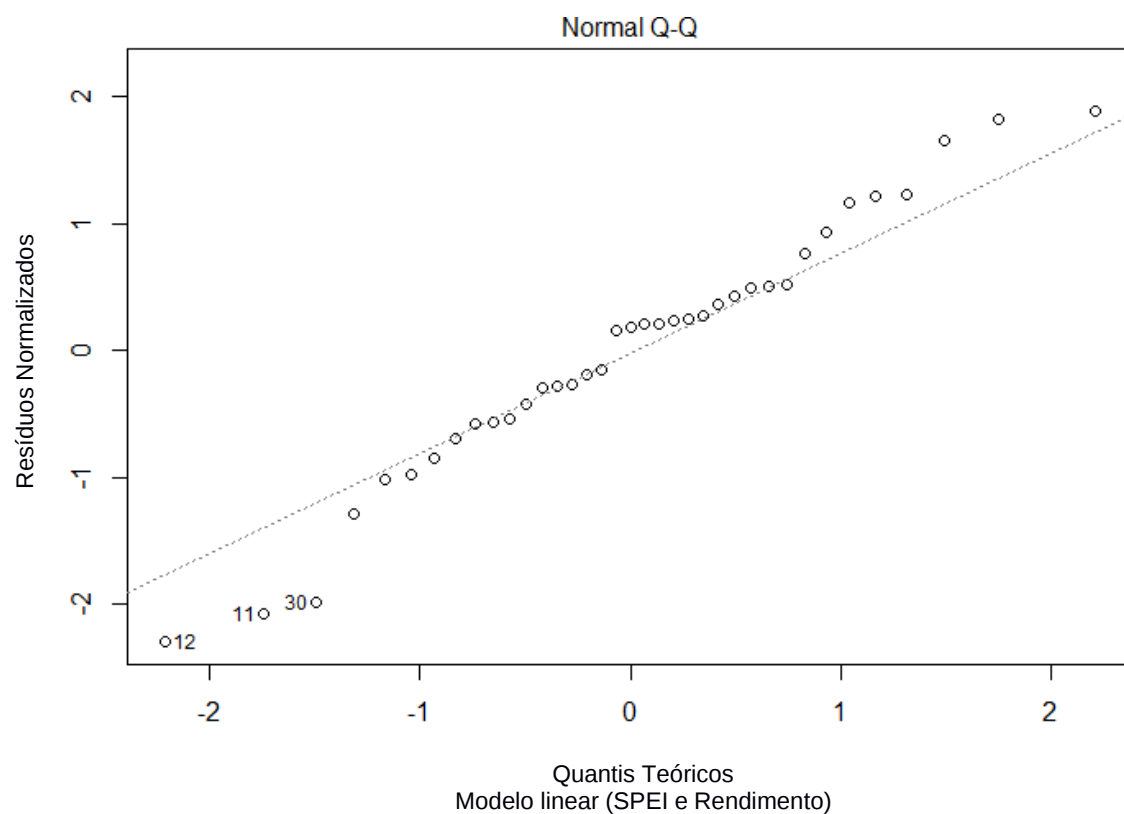


Figura 14: QQ-plot dos resíduos do modelo entre SPEI e rendimento do milho.

Assim como no caso do arroz, os resíduos do modelo linear para o milho (Fig. 14) também apresentam muitos pontos na extremidade da distribuição, indicando que também não se trata de uma distribuição perfeitamente normal.

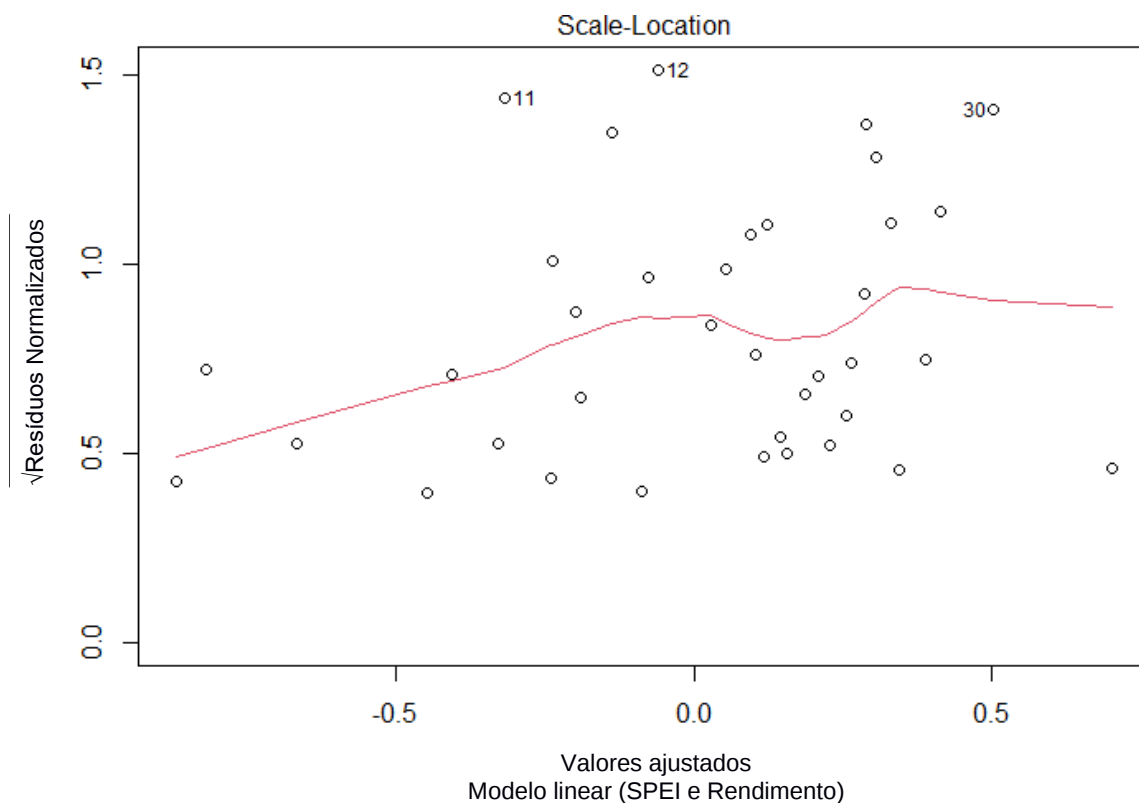


Figura 15: Variância dos resíduos do modelo linear entre SPEI e rendimento do milho

A distribuição da variância dos resíduos (Fig. 15) mostra heterocedasticidade, assim como havia ocorrido com o arroz. Uma diferença entre os gráficos dos dois modelos é que no caso do milho, a reta de distribuição das variâncias apresenta inclinação ligeiramente maior, enquanto no caso do arroz ela é um pouco mais horizontal (Fig. 6) indicando heterocedasticidade um pouco menor.

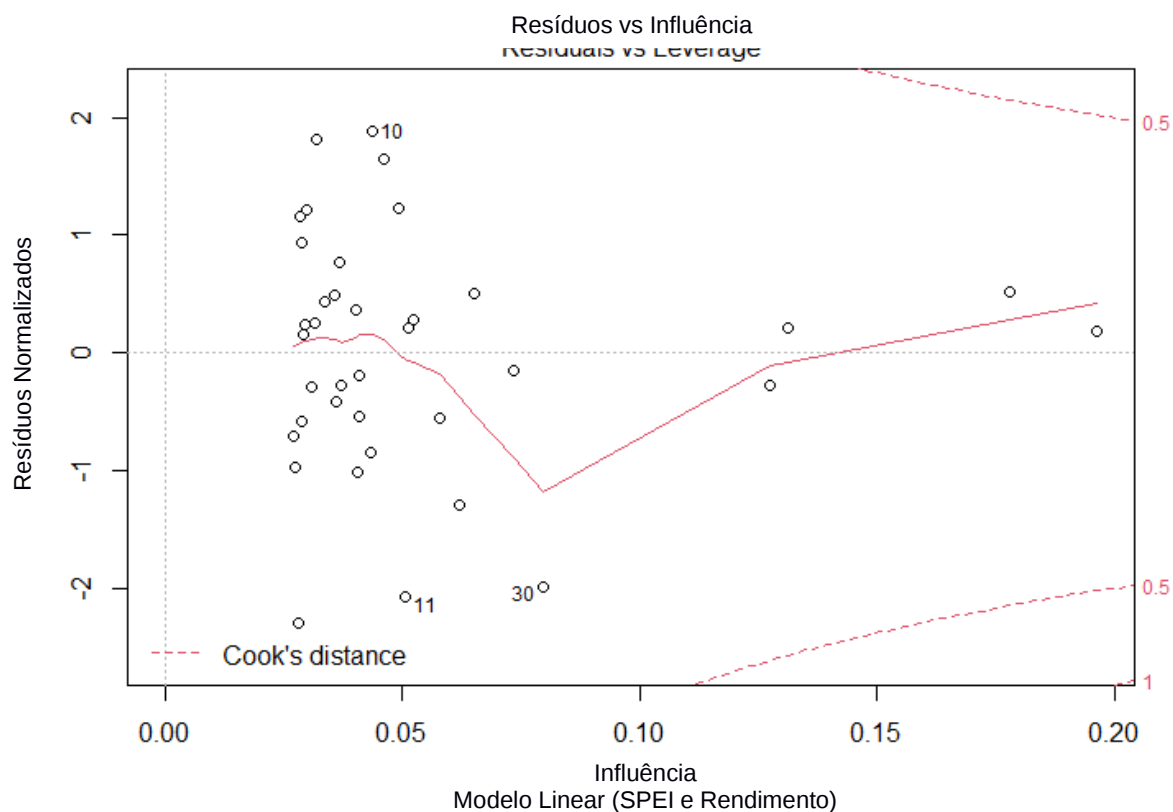


Fig. 16: Resíduos padronizados e influência dos resíduos do milho. A linha vermelha tracejada é a distância de Cook.

Assim como ocorreu com o arroz, o modelo linear entre o SPEI e o rendimento do milho também não apresentou dados influenciadores (Fig. 16).

A Figura 17 mostra a série temporal original dos dados de rendimento do milho, a previsão para o período 2000-2010 e a previsão com correção 2011-2021.

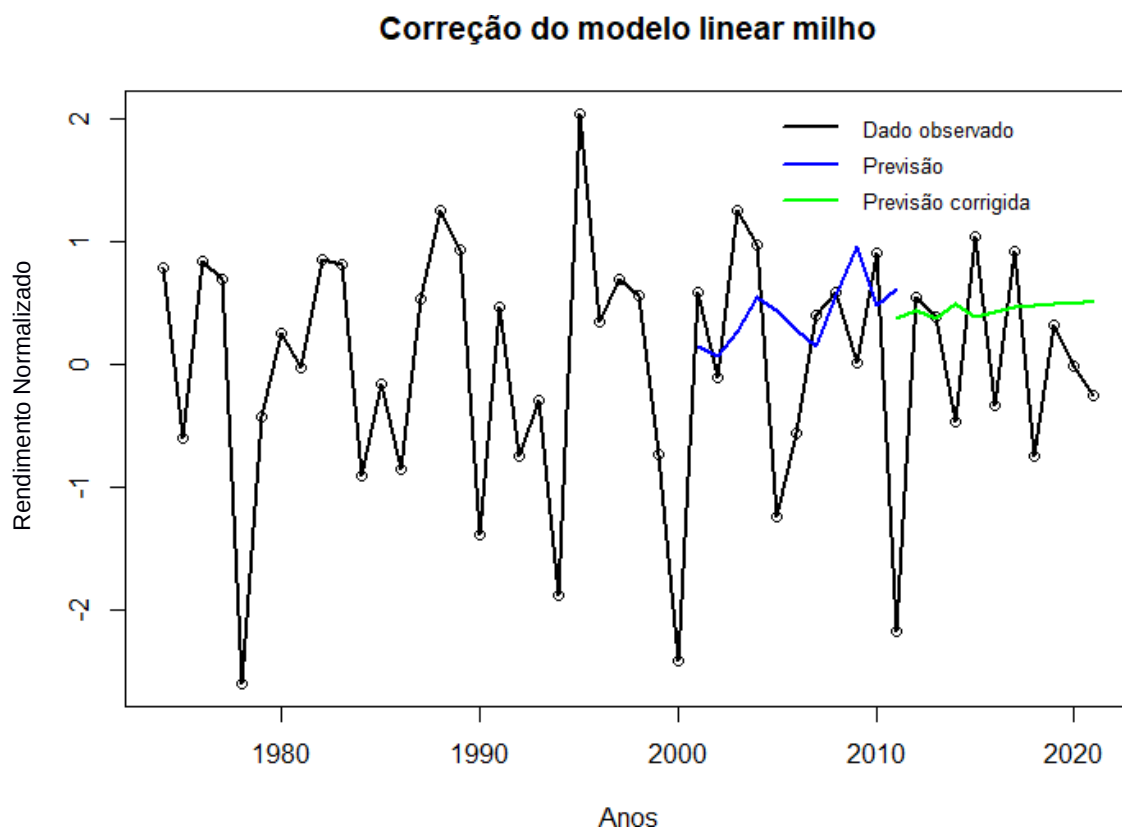


Figura 17: Série temporal de dados observados do rendimento do milho, rendimento previsto a partir do modelo linear e rendimento previsto a partir do modelo linear corrigido pelo método MOS.

A correlação entre o SPEI e o milho foi menor e menos significativa do que a do SPEI com o arroz, e a análise de resíduos do modelo gerado para o milho também apresentou maiores discrepâncias. Sendo assim, já era esperado que a previsão para esta cultura apresentasse erros maiores do que o modelo do arroz, e isso foi o que efetivamente aconteceu: o Erro Médio (Viés) entre a previsão e os dados observados neste caso foi de +1.264, e a Raiz Quadrada do Erro Médio Quadrático de +1.124, mostrando um modelo ainda menos eficiente para previsão nesta localidade. Quando aplicada a correção estatística a este modelo no período de 2011-2021, foi obtido um viés de +1.055 e uma Raiz Quadrada do Erro médio de +1.027 em relação à série temporal do rendimento observados nestes anos. Apesar da ligeira diminuição dos erros, a correção não se mostrou eficiente, e a previsão corrigida não conseguiu de maneira alguma captar as variações reais de rendimento no período (Fig. 17).

5.3 TRIGO

As séries temporais de SPEI e rendimento normalizada e sem tendência para a cultura do trigo são apresentadas na Fig. 18.

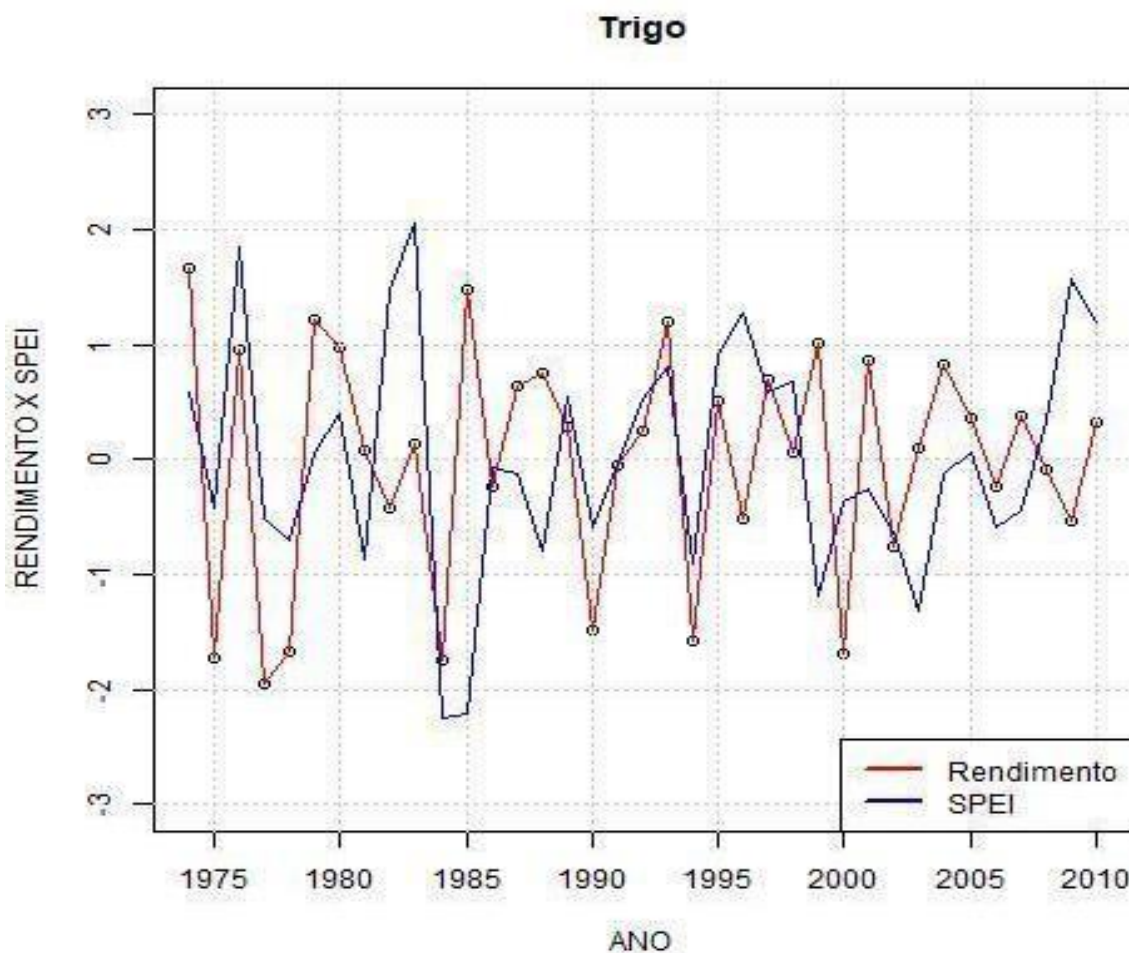


Figura 18: Séries temporais de rendimento normalizado e SPEI do trigo no período 1974-2010

O coeficiente de correlação entre o rendimento do trigo e o SPEI apresentou valor $+0.23$, com $p\text{-valor} = 0.167$. Tais resultados, de correlação muito baixa e pouco significativa, indicaram que um modelo linear de relação entre secas e rendimento de trigo não traria resultados relevantes. Por esta razão, e como os modelos para as culturas do arroz e milho já apresentaram resultados pouco consistentes, foi decidido desconsiderar a cultura do trigo para construção e validação de modelo linear.

6 CONCLUSÕES E PERSPECTIVAS

O objetivo deste trabalho foi verificar a viabilidade de um modelo linear que pudesse fornecer uma previsão das culturas de arroz, milho e trigo no estado de SP a partir de um índice de precipitação e evapotranspiração (SPEI). O modelo foi construído com dados de 1974 a 2000, e testado nos períodos de 2001-2010 e 2011-2021. Os coeficientes de correlação lineares entre o rendimento da produção anual de arroz, milho e trigo no estado e o SPEI foram, respectivamente, +0,54, +0.35 e +0.23 com significância estatística superior a 99%, 97% e 83% respectivamente. Como o trigo apresentou baixa correlação e menor significância, o modelo desta cultura não foi construído.

A área do estado onde ocorreu a seca foi identificado como fator importante para o seu impacto sobre a cultura de arroz. Quando o fenômeno é registrado em todo o estado, ou concentra-se nas regiões mais produtivas, o impacto tende a ser maior, e assim a série de rendimento acompanha o SPEI; porém, quando a seca é mais localizada, ou se restringe a áreas com menor participação no rendimento anual, o rendimento não acompanha a diminuição do índice SPEI. Já para o milho, a relação entre o SPEI e o rendimento da cultura foi baixo independente da atuação da seca nas principais regiões produtoras.

Os modelos de previsão gerados para o arroz e o milho mostraram fraca performance, com superestimativas de, respectivamente 0,67 e 1,12 ton/ha, além de não conseguir capturar a variabilidade dos dados observados de rendimento. A fraca relação encontrada pode ter a ver com o fato de não ser possível assumir uma relação linear entre o déficit hídrico representado pelo SPEI anual e o rendimento das culturas estudadas, no estado de SP. A deficiência da proposta do modelo linear fica evidente nas análises de resíduos, as quais indicaram heterocedasticidade e desvio da normalidade para os resíduos. A tentativa de correção pelo método MOS não surtiu efeito, o que possivelmente também tem a ver com a fraca linearidade das variáveis relacionadas neste estudo.

A partir dos resultados apresentados de que os efeitos da seca têm forte variabilidade espacial, sugere-se que resultados mais satisfatórios poderiam ser obtidos com uma análise mais setORIZADA do estado, em vez de abranger toda a área. Tal sugestão baseia-se em trabalhos como Mohammed et al. (2022), que obteve grande variabilidade espacial no coeficiente de correlação do SPEI com a produção de milho na Hungria. Considerando que o estado de SP é cerca de três vezes maior que a Hungria, e ao contrário daquele país, localiza-se nos subtrópicos (uma região do globo com complexas interações de forçantes atmosféricas), o efeito espacial deve ser ainda mais marcante.

Assim, sugerem-se como estudos futuros uma análise mais setorizada do estado, talvez restringindo a região de interesse às maiores zonas produtoras. Além disso, e mais importante, pode ser preciso procurar modelos mais complexos, não-lineares, para representar a relação do balanço hídrico anual com a produção de culturas na região.

AGRADECIMENTOS

Agradeço à Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) pelo financiamento deste trabalho (Processo número 2022/01326-6)

REFERÊNCIAS

- BATISTA, D. F.; ALBUQUERQUE, T. M. A. Impacto da Seca na Agricultura dos Territórios Agreste Central, Alto Sertão e Centro-Sul de Sergipe. **Revista Brasileira de Meteorologia**, v. 37, p. 81–88, 24 jun. 2022.
- GLAHN, H. R.; LOWRY, D. A. The Use of Model Output Statistics (MOS) in Objective Weather Forecasting. **Journal of Applied Meteorology**, v. 11, n. 8, p. 1203–1211, dez. 1972.
- HARRIS, I. et al. Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. **Scientific Data**, v. 7, n. 1, 3 abr. 2020.
- LU, J.; CARBONE, G. J.; GAO, P. Detrending crop yield data for spatial visualization of drought impacts in the United States, 1895–2014. **Agricultural and Forest Meteorology**, v. 237-238, p. 196–208, maio 2017.
- MEZA, I. et al. Global-scale drought risk assessment for agricultural systems. **Natural Hazards and Earth System Sciences**, v. 20, n. 2, p. 695–712, 2 mar. 2020.
- MOURA, L.; LANDAU, E. C. **Evolução da produção de arroz (*Oryza spp.*, Poaceae)**. [s.l.] In: LANDAU, E. C.; SILVA, G. A. da; MOURA, L.; HIRSCH, A.; GUIMARAES, D. P. (Ed.). Dinâmica da produção agropecuária e da paisagem natural no Brasil nas últimas décadas: produtos de origem vegetal. Brasília, DF: Embrapa, 2020. v. 2, cap. 13, p. 379-405., 2020.
- MOHAMMED, S. et al. Assessing the impacts of agricultural drought (SPI/SPEI) on maize and wheat yields across Hungary. **Scientific Reports**, v. 12, n. 1, 25 maio 2022.
- THORNTHWAITE, C. W. An Approach Toward a Rational Classification of Climate. **Soil Science**, v. 66, n. 1, p. 77, jul. 1948.
- VICENTE-SERRANO, S. M.; BEGUERÍA, S.; LÓPEZ-MORENO, J. I. A Multiscalar Drought Index Sensitive to Global Warming: The Standardized Precipitation Evapotranspiration Index. **Journal of Climate**, v. 23, n. 7, p. 1696–1718, abr. 2010.