

GERAÇÃO DE CENÁRIOS E MÉTODOS DE PREDIÇÃO NA AVALIAÇÃO DA NEGOCIAÇÃO DA SOJA

Matheus Vanzela,

Instituto Federal de Mato Grosso do Sul, Câmpus Nova Andradina,
matheus.vanzela@ifms.edu.br

Hugo Monteiro Aguiar,

Instituto Federal de Mato Grosso do Sul, Câmpus Nova Andradina,
hugomonteeiro@gmail.com

RESUMO

Este artigo levanta quais as principais decisões tomadas pelo produto de soja depois que todos os obstáculos operacionais foram ultrapassados e sua produção está pronta para entrega na cooperativa, que será vendida à vista ou custeado o armazenamento para venda no mercado futuro. Realizou-se 3 estratégias de simulação de Monte Carlo e 2 de previsão de séries temporais com o algoritmo *Prophet* desenvolvido pelo *Facebook*®. Tanto com a geração de Monte Carlo quanto *Prophet* obteve-se erros relativos menores que 8% e avaliações corretas quanto aos movimentos de altas e baixas do mercado. Notou-se em *Prophet* 9 movimentos corretos e com o método de Monte Carlo, 6 movimentos corretos de 11 possíveis dentro do período de um ano. Ambas estratégias têm pontos positivos, mostrando-se assertivas para utilização como ferramentas de apoio à tomada de decisão. Todos os testes foram tratados programaticamente, com a linguagem *Python 3.7*, permitindo maior segurança no tratamento dos dados.

Palavras-chave: Séries Temporais; Monte Carlo; Prophet; Tomada de decisão.

1 INTRODUÇÃO

O objetivo principal deste trabalho foi comparar métodos de geração de cenários e predição de séries temporais para apoiar as principais decisões do período de venda da safra tomadas pelos produtores de soja. A etapa produtiva considerada foi o momento que a soja está disponível para entrega na cooperativa ou indústria. Para levantamento dos parâmetros foi feito um trabalho exploratório de custos com a produção da soja, para que se pudesse construir os patamares de custos envolvidos.

Foram avaliados três níveis de administração – Operacional, Tático e Estratégico – presentes no processo de tomada de decisão em qualquer ambiente produtivo, seja ele industrial, agrícola ou artesanal. A partir desta exploração foi possível levantar algumas decisões importantes, em cada nível e, qual delas seria a concentração desta pesquisa.

Para levantamento dos custos operacionais foram utilizadas duas pequenas propriedades da região de Nova Andradina, durante a safra 2017/2018. Também foram levantados os preços da soja no mercado futuro, análise dos estoques internacionais e algumas políticas setoriais. Tais levantamentos permitiram identificar decisões passíveis de serem aprofundadas em todos os níveis gerenciais. Este trabalho concentrar-se-á nas decisões de nível tático, ou seja, consideramos que todos os obstáculos operacionais da produção foram atendidos e o produtor precisa decidir o que fazer com a produção. Há duas principais decisões que o produtor pode tomar nesta etapa (considerando os arranjos regionais):

- a) Venda direta na cooperativa;
- b) Custear o armazenamento e realizar a venda no futuro;

É importante considerar, antes de descrever quais metodologias a se utilizar no apoio a tomada de decisão, o cenário econômico que alimenta os mercados compradores, bem como, quais são os principais elementos que compõe as altas e baixas dessa *commoditie*.

Levantando os principais países produtores de soja frente aos principais mercados compradores é possível contabilizar uma estimativa dos estoques mundiais de soja, permitindo avaliar o impacto nas decisões de compra e venda à vista ou ainda a venda no mercado de futuros. Similarmente à bolsa de valores, o produtor tem a opção de venda direta da sua produção na cooperativa ou custear o armazenamento e realizar a venda da produção em períodos de melhores preços. Essa decisão pelo mercado futuro esbarra na análise econômica

de países, acordos comerciais, avaliação de estoques e safras dos principais produtores mundiais. A Tabela 1 exibe a concentração da produção mundial de soja (P), o consumo (C) e a diferença entre produção e consumo ($P - C$) para avaliar a balança comercial dos países analisados.

Tabela 1: Produção em milhões de toneladas de soja

Países	Produção (P)	Consumo (C)	P-C
EUA	123,7	61,2	62,5
Brasil	117,0	45,9	71,11
Argentina	56,0	47,9	8,1
China	15,9	103,1	-87,2
Demais	50,3	89,6	-39,3
Mundo	362,9	347,7	15,2

Fonte: Adaptado de FIESP (2019) *apud* USDA

Considerando a soma da produção de Brasil, Argentina e EUA obtêm-se mais de 80% da produção mundial. Tanto Brasil quanto EUA são superavitários quanto à essa *commoditie*, porém, apesar de EUA ser o maior produtor, o Brasil é o maior exportador. Como nota-se, China é o maior consumidor, portanto assenta-se aí uma das principais relações comerciais internacionais do Brasil.

São vários aspectos que guiam o preço da soja no Brasil, desde acordos comerciais internacionais, flutuações do dólar frente ao real (grande parte dos insumos são importados), índices e ocorrências climáticas em toda a América e mesmo alguns entraves políticos entre China e EUA, como têm ocorrido nos últimos períodos. Destas observações há que se esperar alta volatilidade nos preços e grande dificuldade de administrar a negociação da produção.

Para apoiar a escolha dentre as 2 opções de entrega da produção levantadas, este estudo realiza 3 simulações, baseadas no Método de Monte Carlo (MMC), vastamente discutido na literatura de gerenciamento de projetos, como apresenta o trabalho de Fernandes (2005). Também foram realizados 2 testes de predição de séries temporais com a biblioteca de algoritmos *Prophet*, detalhado em Taylor e Letham (2017).

2 REVISÃO DA LITERATURA

Para realizar os testes utilizamos a literatura relacionada a processos estocásticos e os textos clássicos com as estratégias que discutem a geração de cenários e predição de resultados. Matemáticos como Henri Poincaré, Edward Norton Lorenz e Alan Turing debruçaram-se sobre

a previsão de incertezas em suas carreiras. Nos trabalhos com sistemas dinâmicos para previsão do clima, Lorenz observou que mesmo pequenas perturbações na atmosfera podem produzir variações climáticas, cunhando assim a expressão - efeito borboleta - : “O bater de asas de uma borboleta no Brasil pode causar um tornado no Texas”. A metodologia de geração de cenários para representar possíveis soluções desses processos estocásticos foi levantada por Von Neumann e Stanislaw Ulam, durante as simulações da difusão do coeficiente de neutrons para o Projeto Manhattan. É possível que haja divergência sobre a criação das técnicas de geração de cenários, devido ao seu caráter intuitivo. Em Hammersley (2013) encontra-se referência a outros estudiosos que apresentavam elementos de simulações probabilística em seus trabalhos.

Neste estudo foram utilizados métodos de simulação e de predição. Em situações estocásticas ou com muitas variáveis determinísticas de difícil modelagem, os métodos de simulação permitem realizar um estudo probabilístico, produzindo cenários críveis dentro do ambiente produtivo em estudo. O método de Monte Carlo (MC) é definido em Taha (2008) como um esquema de modelagem que estima parâmetros estocásticos ou determinísticos com base em amostragem aleatória. As simulações de MC dependem de bom conhecimento do problema em estudo, saber quais as variáveis envolvidas, os pontos de sazonalidade (quando há) e outros elementos que guiam as variáveis. Esta *expertise* é fundamental para planejar quais distribuições de probabilidade tem coerência com problema. Quando a situação de planejamento envolve várias variáveis aleatórias, é condição *sine qua non* a independência entre as variáveis, Fernandes (2005) menciona que, essa condição, frequentemente é esquecida nos trabalhos que realizam simulação de MC.

Uma outra abordagem de tratamento de incertezas são os métodos de predição, algumas técnicas estatísticas possibilitam realizar aproximações por meio de regressões. Em problemas com variáveis inter-dependentes as técnicas de regressão e interpolação polinomial podem aproximar os resultados com um erro calculado. Em séries temporais, frequentemente os modelos auto-regressivos integrados de médias móveis conhecidos como ARIMA são elegidos para realizar as predições. Esses modelos assumem que as séries são geradas por processos lineares. Já para situações com parametrização subjacente não linear existem os modelos baseados em redes neurais artificiais (ANN), Zhang et al. (1998) fazem uma revisão ampla sobre os dois métodos.

Neste artigo utiliza-se a biblioteca de códigos *Prophet* programada em 2017/18 dentro do núcleo de pesquisa do *Facebook*®. Esta biblioteca é de licença livre, podendo ser utilizada

com as linguagens *Python* e *R*. Os algoritmos implementados no *Prophet* são probabilísticos, consideram efeitos de sazonalidade, tendência e datas-chave (feriados). O modelador tem alta flexibilidade de ajustar os parâmetros do algoritmo para melhorar a performance das previsões. O único regressor utilizado é o tempo, a lógica de regressão é potencialmente distinta de modelos que consideram a dependência temporal na estrutura dos dados. *Prophet* transforma o problema de regressão em um exercício de ajuste de curvas. Como a estrutura para interpretar os pontos de tendência, sazonalidade e datas representativas é baseado no método aditivo de Hastie; Tibshirani (1987) o algoritmo traz grande flexibilidade, tornando mais personalizado ao analista inserir novos pontos de inflexão ou tendência incorporando irregularidades não capturadas pelos modelos auto regressivos (Taylor e Letham (2017)).

3 METODOLOGIA E DADOS DE ENTRADA

Para planejar a geração de possíveis cenários foi considerada uma série histórica de preços (em R\$) dos anos de 2011 a 2017 no porto de Paranaguá-PR, principal ponto de entrega da soja para Mato Grosso do Sul. Apesar da maioria dos produtores vender a soja no mercado doméstico, ou seja, cooperativas da região, os preços praticados nessa praça são os preços praticados no porto decrescido em média de 10%, segundo levantamento *in loco* para que se incorpore o frete até Paranaguá. Para efeito de comparação dos resultados gerados não utilizou-se a série de preços de 2018 que servirá de métrica na avaliação de performance dos modelos.

A Tabela 2 apresenta, de acordo com o Centro de Pesquisas de Engenharia Agrônômica da Esalq-SP (CEPEA) os valores médios (R\$) praticados mês a mês, durante os anos de 2011 e 2018. É conveniente destacar que a metodologia utilizada pelo CEPEA (2019) ao fornecer os resultados mensais é de exibir a média entre os valores negociados diariamente dentro dos meses base e não o valor do último dia negociado. As linhas Avg, Std e Med, representam respectivamente os valores das médias, desvios padrão e medianas mensais, entre os meses de 2011 e 2017. Estes valores serão posteriormente utilizados nas distribuições de probabilidade escolhidas para compor o processo de geração de cenários. O ano de 2018 está disposto na última linha da Tabela 2, pois como será o ano de referência, os resultados não fizeram parte dos cálculos apresentados de média, mediana e desvio padrão.

Tabela 2- Valores médios da saca de soja dentro de cada mês da série histórica

	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
2011	50,78	51,39	49,54	47,19	47,83	47,88	48,5	49,38	51,94	48,47	47,74	47,7
2012	49,55	49,32	54,93	60,35	63,79	68,05	79,36	85,58	86,82	75,73	75,73	75,73
2013	71,99	64,01	61,84	59,45	61,89	68,72	69,32	69,88	73,47	73,84	76,35	77,25
2014	72,29	69,71	72,27	71,11	70,74	70,86	67,3	67,11	63,06	61,17	61,17	61,17
2015	61,14	63,72	67,9	69,53	66,61	67,88	72,89	77,33	81,35	81,98	79,97	80,76
2016	82,75	77,83	74,53	78,04	86,43	95,19	87,46	81,69	79,5	76,7	78,27	78,43
2017	76,03	73,86	70,01	65,82	68,94	68,95	72,24	69,83	70,41	71,47	73,87	74,24
Avg.	66,36	64,26	64,43	64,50	66,60	69,65	71,01	71,54	72,36	69,91	70,44	70,75
Med.	71,91	66,86	68,96	67,68	67,78	68,84	72,57	73,61	76,49	74,79	76,04	76,49
Std.	12,79	10,76	9,36	9,97	11,53	13,75	12,03	11,90	11,90	11,40	11,75	11,98
2018	71,83	74,72	79,39	85,53	86,12	84,83	88,29	89,91	95,48	90,53	84,16	81,10

Fonte: CEPEA (2019)

Neste estudo foram estabelecidas 3 estratégias de geração de cenários, utilizando distribuições de probabilidade alimentadas com medidas estatísticas da série histórica, considerando os anos de 2011 a 2017. A Tabela 3 exibe um resumo estatístico com os números que alimentaram as variáveis de entrada na geração dos cenários de MC.

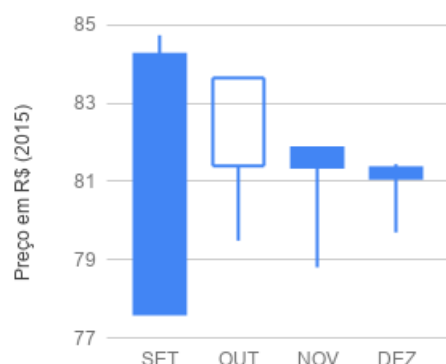
Tabela 3 - Resumo estatístico dos preços exibidos na Tabela 2

Ano	Soma	Mínimo	Máximo	Amplitude	Média	Mediana	DesvPad	Ano_{i+1}/Ano_i
2011	588,34	47,19	51,94	4,75	49,03	48,49	1,74	N.A
2012	824,94	49,32	86,82	37,50	68,75	71,89	14,16	0,40
2013	828,01	59,45	77,25	17,80	69,00	69,60	6,52	0,00
2014	807,96	61,17	72,29	11,12	67,33	68,51	4,70	-0,02
2015	871,06	61,14	81,98	20,84	72,59	71,21	7,97	0,08
2016	976,82	74,53	95,19	20,66	81,40	78,97	6,81	0,12
2017	855,67	65,82	76,03	10,21	71,31	70,94	3,32	-0,12
Média	821,83	59,80	77,36	17,55	68,49	68,51	6,46	0,08
2018	1011,89	71,83	95,48	23,65	84,32	85,18	6,73	0,18

Fonte: Autores

Destaca-se a última linha, representando o ano de 2018, com as estatísticas correspondentes e o crescimento de 18% com relação 2017. A última coluna da Tabela 3 refere-se à taxa de crescimento ou decréscimo ano a ano. Para uma perspectiva mais ampla da volatilidade desse mercado apresenta-se um gráfico de *candlestick*, Figura 2, muito utilizado nas negociações de bolsa de valores e mercadorias. Os *candles* vazados significam que o fechamento foi menor do que a abertura, *candles* preenchidos representam fechamento maior do que a abertura. Os pavios representam quais são os limites inferiores e superiores das negociações dentro do período.

Figura 1 - *Candlestick* do 3º quadrimestre de 2015



Fonte: Autores

Para exemplificar o impacto desta volatilidade para um pequeno produtor de 50 *ha* de área que produz aproximadamente 2750 sacas em uma safra, apresenta-se as diferenças para 3 cálculos de receita bruta da produção em setembro de 2015. Um cálculo tomando o valor mínimo mensal, outro tomando a média e o último pelo máximo mensal:

- $2750 \times R\$ 77,62 = R\$ 213.455,00$ (R\$ 77,62: mínimo da saca em set/15)
- $2750 \times R\$ 81,35 = R\$ 223.712,50$ (R\$ 81,35: média da saca em set/15)
- $2750 \times R\$ 84,74 = R\$ 233.035,00$ (R\$ 84,74: máximo da saca durante set/15)

Este exemplo pedagógico mostra o impacto dessa decisão mesmo para um pequeno produtor. Nota-se uma variação de aproximadamente R\$ 10.000,00 entre mínimo e média e média e máximo, respectivamente dentro de set-15. Estas características reforçam a necessidade de se estabelecer um planejamento detalhado para apoiar a decisão da escolha do período de venda da produção. Também aponta uma especificidade não observada nos trabalhos de predição de preços mensais da soja, que é a projeção sobre os valores médios mensais, por conta de que essa é a metodologia utilizada pelo CEPEA. Todavia na prática o produtor nunca irá vender sua safra pela média dos preços mensais, mas sim pelo preço do dia.

Nas 5 estratégias foi utilizada a linguagem de programação *Python 3.7* e as bibliotecas científicas *NumPy*, *SciPy*, *Pandas* e *Prophet*. O tratamento dos dados via linguagem de programação permite maior estabilidade e menor índice de erros ao manusear resultados com muitas saídas. Foram gerados 10.000 cenários para cada uma das 3 distribuições: distribuição normal, distribuição triangular e distribuição uniforme, bem como 2 rodadas de predição variando os parâmetros de sazonalidade e tendência do *Prophet*.

Considere o conjunto $meses = \{1, \dots, 12\}$ e os anos entre 2011 e 2017. Na primeira

rodada de testes foi utilizada a distribuição uniforme de probabilidade com os parâmetros variando entre os valores mínimos e máximos para cada mês, incrementados do crescimento médio da série histórica que foi de aproximadamente 8% como observa-se na Tabela 3. Assim, os intervalos de cada um dos referidos meses, no método de MC desta distribuição foram $\{[min(i), max(i)] \text{ para } i \in meses\}$, iterados entre os anos de 2011 a 2017. Na segunda rodada de testes utilizou-se a distribuição normal com os parâmetros de média e desvio padrão dos respectivos meses obtidos da série histórica. Na terceira rodada de geração de MC foi considerada a distribuição triangular de probabilidade, que atribui como parâmetros, um valor mínimo, máximo e valor esperado. A distribuição triangular é comumente utilizada em situações que não se tem muitas informações sobre os dados, uma vez que valores mínimo, máximo e esperado podem ser inferidos, com razoável coerência, por um gestor que conheça o ambiente em análise. Nos testes com a distribuição triangular, os valores de referência foram: $\{[min(i), máx.(i), med.(i)] \text{ para } i \in meses\}$, em que $min(i)$, $máx.(i)$ e $med.(i)$ são os valores mínimo, máximo e mediana entre cada um dos meses da série histórica. Com a biblioteca *Prophet* foram realizados ajustes nos parâmetros de tendência ($Tend = [0, 1]$) quanto maior o valor mais flexível à variações de tendência ficará o modelo, outro elemento personalizável é a sazonalidade ($Saz = [0, 1]$) que representa a porcentagem da série que será pesquisada para atribuição de sazonalidade. Por padrão a biblioteca trabalha com $Tend = 0,05$ e $Saz = 0,8$. No primeiro teste utilizou-se a configuração padrão da biblioteca e no segundo teste atribuímos $Tend = 0,5$, mantendo $Saz=0,8$ no valor padrão.

4 DISCUSSÃO E ANÁLISE DOS DADOS

A discussão e análise dos resultados foi organizada iniciando por apresentar os histogramas das somas anuais dos cenários gerados pelo método de MC. Realiza-se assim a exibição das medianas, como medida para entender o ponto de equilíbrio de cada método. Na prática o fazendeiro não tem interesse nas somas anuais, porém, para as análises deste artigo, esse resultado norteia a interpretação das distribuições de probabilidade e dos parâmetros utilizados para rodar os modelos de acordo com cada distribuição. Para comparação com os resultados do ano referência de 2018 selecionou-se o 3º quartil correspondente à 75% da função de distribuição acumulada de probabilidade.

Na Figuras 2 a 4 apresentam-se os histogramas do número de cenários gerados contabilizados pelo somatório dos valores mensais.

Figura 2: Histograma distribuição normal

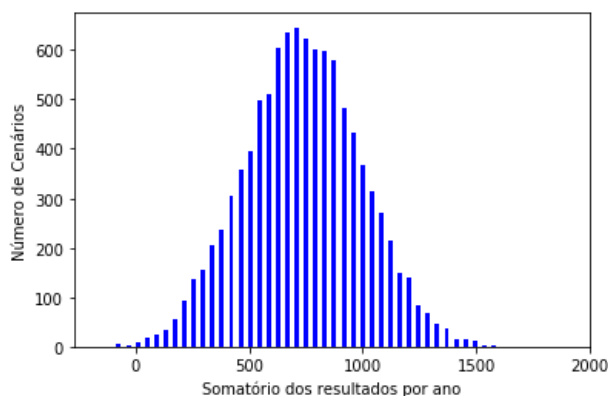


Figura 3: Histograma da distribuição uniforme

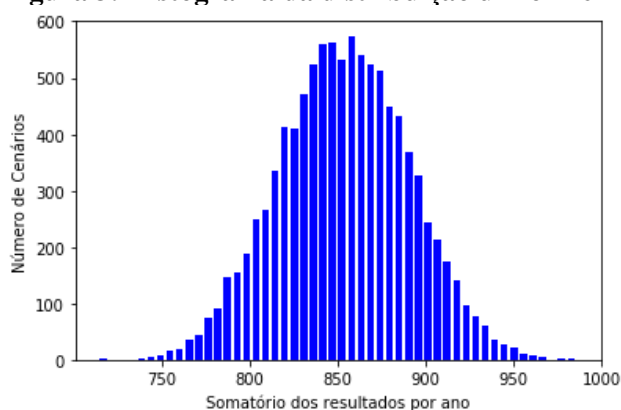
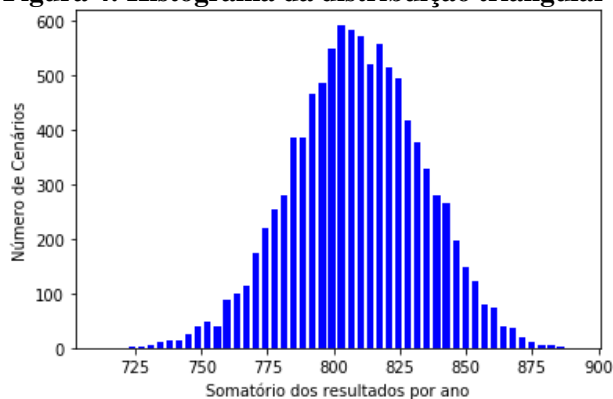


Figura 4: Histograma da distribuição triangular



Respectivamente para as distribuições normal, uniforme e triangular foram encontrados os valores 745, 854, 809 correspondente à mediana das distribuições. Notam-se valores bem inferiores as somatória referente a 2018, muito se deve ao ano de 2011 que foi um ano com recordes de produções de Argentina e Estados Unidos. Outro ponto relevante, agora para justificar os altos valores de 2018 é a escalada do dólar frente ao real nos últimos 5 anos, como as cotações do trabalho foram em reais podemos ter valores distorcidos na comparação com a heurística proposta. Em relação aos testes com o *Prophet* não foi realizada a comparação com a somatória anual e suas estatísticas, uma vez que os testes foram de predição e não de geração massiva de cenários.

Na sequência das análises é exibida a tabela referente ao 3º quartil das gerações de MC e os valores de tendência obtidos com a biblioteca *Prophet*.

Tabela 4: Valores gerados em comparação com o ano de 2018

	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
Nor	74,54	71,85	71,80	71,97	74,54	78,30	79,42	79,97	80,87	78,11	78,75	79,13
Uni	80,60	76,37	73,66	75,87	82,90	89,81	83,73	82,58	84,46	79,31	77,70	78,43
Triang.	73,27	68,06	68,11	68,23	72,71	77,18	75,34	73,68	75,96	73,76	73,60	74,07
Pro05	61,66	57,3	68,64	64,95	57,66	54,07	57,74	62,13	65,9	66,9	56,75	51,36
Pro005	75,46	76,32	81,92	79,54	75,44	74,3	78,97	83,63	87,37	87,54	78,52	74,19
2018	71,83	74,72	79,39	85,53	86,12	84,83	88,29	89,91	95,48	90,53	84,16	81,10

Fonte: Autores

Os valores em destaque representam meses em que os números simulados no estudo foram menores que um erro relativo de 10% em comparação ao ano de 2018. As duas estratégias com maior quantidade de meses assertivos foram MC na distribuição uniforme com 8 meses de assertividade e Pro005, com 9 meses de resultados próximos por menos de 10% dos valores de 2018. A taxa de erro aceitável (10%) foi escolhida com base na variação do preço da soja dentro de um mesmo ano. O parâmetro de comparação é, aproximadamente, menor do que a metade da média da amplitude observada na série histórica. Como medidas estatísticas de erro utilizou-se os principais indicadores:

- A média dos erros relativos absolutos totais: $MAPE = \frac{100}{12} \sum_{i=1}^{12} \frac{|Y_i - \hat{Y}_i|}{\hat{Y}_i}$;
- A média das diferenças absolutas calculado como $MAD = \frac{1}{12} \sum_{i=1}^{12} |Y_i - \hat{Y}_i|$;
- A raiz da média dos quadrados dos erros calculado como $RMSE = \sqrt{\frac{1}{12} \sum_{i=1}^{12} (Y_i - \hat{Y}_i)^2}$

Com \hat{Y}_t e Y_t respectivamente os valores gerados e do ano referência no i – ésimo mês

Tabela 5: Resumo dos erros estatísticos

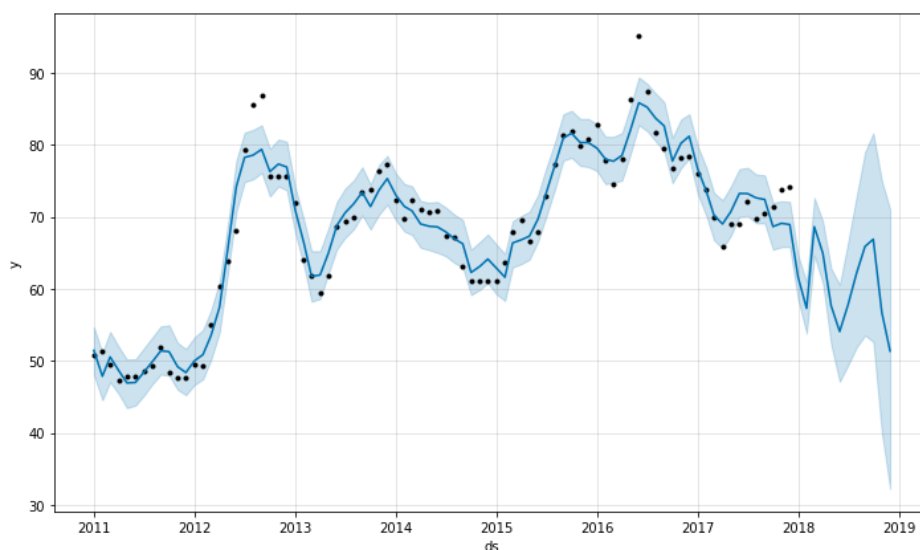
	Norm	Uni	Triang.	Pro05	Pro005
MAPE	9,4%	7,6%	13,6%	28%	7,2%
MAD	8,172	6,439	11,733	23,901	6,184
RMSE	9,185	7,140	12,798	24,959	6,853

As estratégias normal, uniforme e Pro005 apresentaram erros relativos (MAPE) menores que 10% considerando um período de 1 ano. O resultado MAD mostra em linhas gerais quanto, em reais cada modelo está diferindo dos valores reais por cada mês, tanto nesta métrica quanto em RMSE (que identifica períodos com *gap* muito alto) notam-se valores muito similares para *Prophet 5%* e a distribuição uniforme.

Notadamente utilizando a regulagem de tendência do algoritmo *Prophet* em 50% obtiveram-se os piores resultados (Figura 5), uma vez que, ao flexibilizar a tendência, o modelo tende a capturar possíveis tendências na série. Observando a Figura 5, é possível notar que da metade de 2016 em diante a soja entra em tendência de baixa, justificando a lógica dos resultados fornecidos pelo algoritmo calibrado com alta sensibilidade à tendência.

Na Figura 5 pode-se perceber ainda que o algoritmo trabalhou próximo aos valores da base de treinamento, considerando os pontos como valores reais da soja, a linha azul os resultados de treinamento e a sombra os limites máximos e mínimos aceitáveis. Note que após 2018 os pontos cessam, pois é o ano base na comparação dos resultados.

Figura 5: Gráfico dos valores previstos com Prophet em 50% de sensibilidade à tendência



Considerando o limite de 10%, foi possível perceber que apenas para as gerações de MC utilizando como heurística com a distribuição uniforme e o algoritmo *Prophet* calibrado em 5% de sensibilidade à tendência obteve-se valores competitivos. Assim é importante observar alguns pormenores para escolher qual das estratégias tem maior eficiência para apoiar a tomada de decisão sobre o momento mais oportuno para vender a produção da *commoditie*. Assim apresentam-se os gráficos comparativos, na Figura 6 correspondente à simulação com a distribuição uniforme e na Figura 7 correspondente à Pro005. Além de analisar os erros, também é importante perceber a capacidade que cada uma das estratégias é sensível à movimentos de alta ou baixa, uma vez que não estamos considerando as hipóteses em estudo como guias absolutos, mas sim como ferramentas de apoio à tomada de decisão. Nessas circunstâncias é mais pertinente avaliar o modelo que tem capacidade de apontar os movimentos de altas e baixas de mercado.

Figura 6: Distribuição Uniforme x 2018

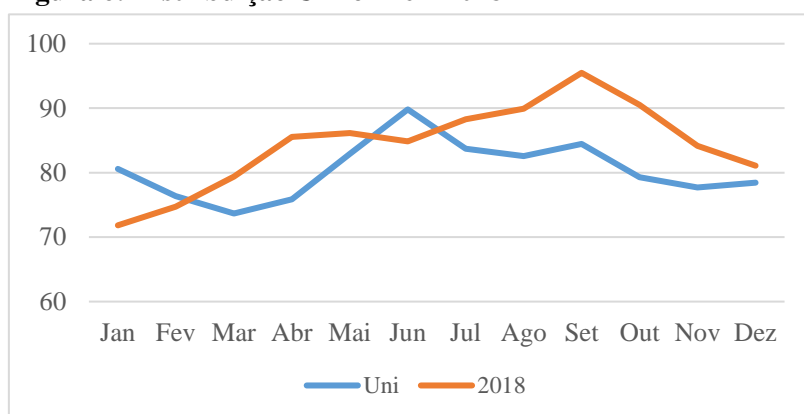
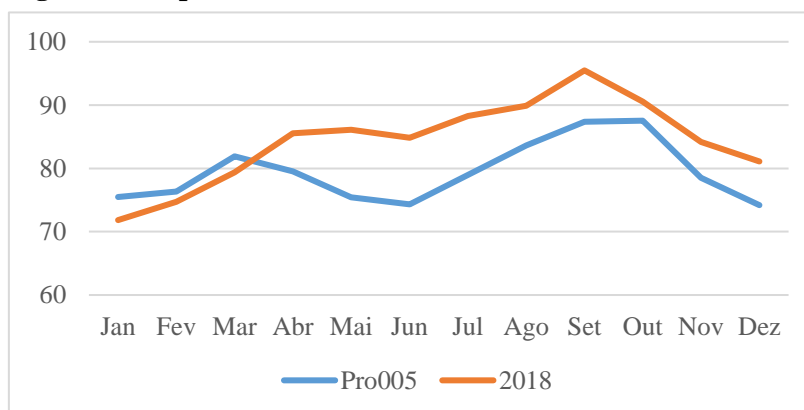
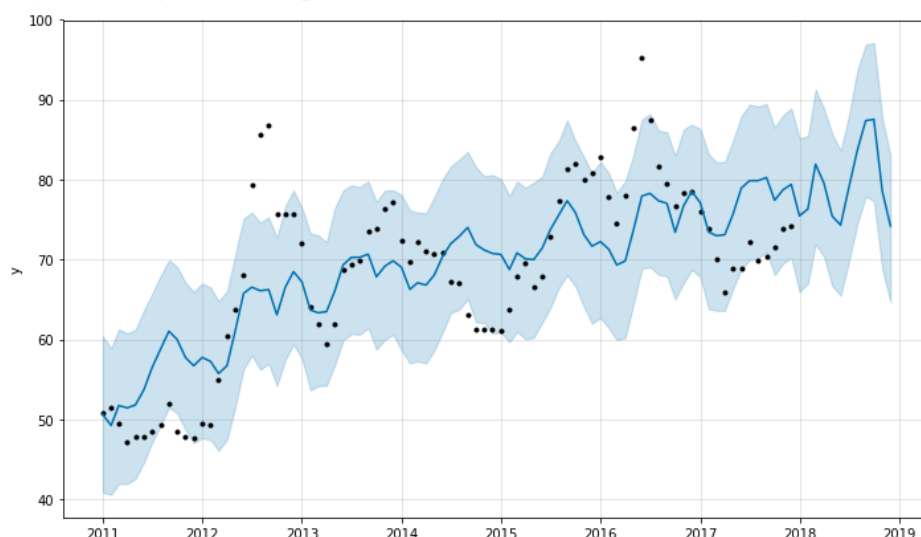


Figura 7: Prophet 5% x 2018



Considerando as oscilações dos preços no período de 1 ano, o algoritmo de *Prophet 5%* conseguiu capturar 8 vezes durante os 11 intervalos entre os meses e a distribuição uniforme acompanhou corretamente a direção dos preços em 6 meses. Essa análise com relação à sensibilidade direcional de mercado é de vigor interesse aos gestores, pois pode indicar não apenas números que por estarem distribuídos ao longo de um horizonte de planejamento extenso (1 ano), precisarão ser constantemente repensados, agora os movimentos assertivos quanto a direção dos valores é muito mais próximo da informação que o gestor deseja no apoio à tomada de decisão sobre o melhor período de negociar a produção. Na Figura apresenta-se o gráfico dos resultados da etapa de treinamento de *Pro005* disposto junto com a previsão dos resultados.

Figura 8: Algoritmo *Prophet* calibrado em 5% na tendência



Fonte: Autores

Nesta configuração o algoritmo teve a melhor performance, além de capturar melhor os movimentos direcionais de mercado. É possível avaliar que neste teste são evitados os *outliers* durante o processo de treinamento e o modelo se mostra mais estável. Diante desse prisma analítico podemos apontar para a estratégia com o algoritmo *Prophet* configurado em 5% no parâmetro que flexibiliza a sensibilidade à movimentos de tendência como o mais próximo da realidade do ano de 2018. A geração de cenários com o método de MC teve como sua melhor estratégia a heurística proposta para a distribuição uniforme, perdendo para o algoritmo *Prophet* na percepção dos movimentos direcionais de mercado, porém ao analisar cenários de mercado futuro é mais seguro realizar testes com mais de uma estratégia. Nestes testes o pior resultado

se deu com a configuração de *Prophet* em 50%, caso o decisor decida comparar apenas a eficiência, avaliando apenas os erros clássicos dos resultados, essa estratégia seria totalmente descartada em avaliações subsequentes, porém efeitos de *pullback* (reversão) são recorrentes e imprevisíveis no mercado. Diante dessas incertezas essa configuração de *Prophet* pode ser usada na produção de um plano, caso esse efeito ocorra, como geração de cenários de pior caso.

5 CONCLUSÕES

Com este trabalho foi possível testar duas estratégias distintas de formulação de predições futuras, a simulação de Monte Carlo e o algoritmo *Prophet* desenvolvido pelo Facebook® em 2018. O trabalho com análise de cenários futuros deve tornar-se um guia nas tomadas de decisão e não resultados de “caixa-preta” que prevê números cabalísticos. Neste trabalho foi analisado os movimentos direcionais de mercado, não somente a proximidade entre os valores previstos e da série histórica os valores encontrados foram críveis com a realidade enfrentada por produtores e cooperativas, permitindo uma tomada de decisão mais robusta. As estratégias apresentadas foram testadas dentro da realidade de produção da soja, porém podem ser facilmente replicadas para outras culturas, porém faz-se necessários se debruçar sobre os detalhes, adquirir alguma *expertise* no ambiente produtivo em questão antes de valer-se apenas de resultados de uma série histórica. Os códigos implementados neste trabalho estão disponíveis no versionador *online* de códigos *GitHub* podendo ser utilizado pela comunidade para trabalhos futuros.

REFERÊNCIAS

- CEPEA, Centro de Pesquisas Econômicas da Escola Superior de Agricultura ESALQ - **Relatórios Agrícolas**. Piracicaba, 2019. Disponível em: <https://www.cepea.esalq.usp.br/br/busca.aspx?busca=RELATÓRIOS>. Acesso: Junho 2019
- FERNANDES, C. A. B. DE A. Gerenciamento de Riscos em Projetos: como usar o Microsoft Excel para realizar a simulação de Monte Carlo. **B&B Brothers**, p. 1–6, 2005. Disponível em: http://www.bbbrothers.com.br/files/pdfs/artigos/simul_monte_carlo.pdf. .
- FIESP; Federação da Indústria e Comércio do Estado de São Paulo. **Informativo Safra Mundial de Soja 2019 / 20 - 3º Levantamento do USDA**. São Paulo, 2019. Disponível em: <https://www.fiesp.com.br/indices-pesquisas-e-publicacoes/safra-mundial-de-soja/>
- HAMMERSLEY, J. **Monte Carlo Methods**. Springer Netherlands, 2013.

III Encontro Internacional de Gestão, Desenvolvimento e Inovação

10 a 13 de setembro de 2019 | Naviraí - MS



HASTIE, T. .; TIBSHIRANI, R. Generalized additive models: some applications. **Journal of the American Statistical Association**, v. 82, p. 371–386, 1987.

TAHA, H. A. **Pesquisa Operacional: Uma visão geral**. 8º ed. São Paulo, 2008.

TAYLOR, S. J.; LETHAM, B. Forcast at Scale Prophet. **PeerJ Preprints**, p. 1–25, 2017.

ZHANG, G.; PATUWO, B. E.; HU, M. Y. [10] -zhang1998 - Forecasting with artificial neural networks: The state of the art. **International Journal of Forecasting**, v. 14, p. 35–62, 1998.

Disponível em: encurtador.com.br/bhsxz